

Olasılığa Giriş — Harvard Stat 110 (Blitzstein)'dan

ML Builder için Türkçe Notlar

Phase 1

2026-05-29

İçindekiler

1	Önsöz	1
2	Bu kitap nedir?	3
3	Nasıl Okumalı	5
4	34 Ders	7
5	Notasyon	9
6	Builder Eksen — Olasılık → ML	11
7	Yazım Kuralları	13
8	Olasılık ve Sayma	15
8.1	Bu Derste Ne Var?	15
8.2	Olasılık Nereye Uygulanır? Belirsizliğin Mantığı	16
8.3	Örnek Uzay ve Olay	16
8.4	Olasılığın Naif Tanımı	18
8.5	Naif Tanımın İki Varsayımı (ve Nerede Çöker)	18
8.6	Neden Saymak? Çarpma Kuralı	19
8.7	Binom Katsayısı	21
8.8	Worked Example: Poker’de Full House	23
8.9	Örnekleme Tablosu (2×2)	24
8.10	Bu Dersin Özeti	25
8.11	Kontrol Soruları	26
8.12	Egzersizler	26
8.13	Sonraki Ders İçin Hazırlık	27
8.14	Anahtar Kavramlar (Cheat Sheet)	27
8.15	ML Bağlantıları Özeti	28
9	Story Proof’lar ve Olasılık Aksiyomları	31
9.1	Bu Derste Ne Var?	31
9.2	Etiketleme: Önce Numaralandır	31
9.3	İki Yolla Saymak: Takımlara Bölme	33
9.4	Örnekleme Tablosunun Dördüncü Hücre: Stars and Bars	34
9.5	Ayırt Edilemezlik ve Bose-Einstein	35
9.6	Story Proof Nedir?	35
9.7	Story Proof: Başkanlık Komite	36
9.8	Story Proof: Vandermonde Özdeşliği	37
9.9	Olasılığın Aksiyomatik Tanımı	38

9.10 Bu Dersin Özeti	39
9.11 Kontrol Soruları	39
9.12 Egzersizler	40
9.13 Sonraki Ders İçin Hazırlık	40
9.14 Anahtar Kavramlar (Cheat Sheet)	41
9.15 ML Bağlantıları Özeti	41
10 Doğum Günü Problemi ve Olasılığın Özellikleri	43
10.1 Bu Derste Ne Var?	43
10.2 Doğum Günü Problemi: Kurulum	43
10.3 Doğum Günü: Tümleyenle Hesap	44
10.4 Neden 23? Sezgi: $\binom{K}{2}$ Çiftler	45
10.5 Olasılığı Alan Gibi Düşün + Tümleyen Kuralı	46
10.6 Tekdüzelik: $A \subseteq B \Rightarrow P(A) \leq P(B)$	46
10.7 İçerme-Dışarma (Inclusion-Exclusion)	47
10.8 Matching Problem (de Montmort, 1713) ve $1/e$	48
10.9 Bu Dersin Özeti	49
10.10 Kontrol Soruları	49
10.11 Egzersizler	50
10.12 Sonraki Ders İçin Hazırlık	50
10.13 Anahtar Kavramlar (Cheat Sheet)	51
10.14 ML Bağlantıları Özeti	51
11 Koşullu Olasılık	53
11.1 Bu Derste Ne Var?	53
11.2 Bağımsızlık: Tanım ve Ayrıklıkla Karıştırma	53
11.3 Çoklu Bağımsızlık: İkili \neq Karşılıklı	54
11.4 Newton-Pepys Problemi (1693) + Binom Olasılığı	55
11.5 Koşullu Olasılık: İstatistiğin Ruhu	55
11.6 İki Sezgi: Çakıl Dünyası ve Frekansçı Dünya	56
11.7 Üç Teorem: Çarpım, Zincir Kuralı, Bayes	57
11.8 Bu Dersin Özeti	58
11.9 Kontrol Soruları	58
11.10 Egzersizler	59
11.11 Sonraki Ders İçin Hazırlık	59
11.12 Anahtar Kavramlar (Cheat Sheet)	59
11.13 ML Bağlantıları Özeti	60
12 Koşullamaya Devam ve Toplam Olasılık Yasası	61
12.1 Bu Derste Ne Var?	61
12.2 Toplam Olasılık Yasası (LOTP)	61
12.3 İki Kart Paradoksu: “Bir As” vs “Maça Ası”	62
12.4 Hastalık Testi: Bayes + LOTP El Ele	63
12.5 Biyohazard 1: $P(A B) \neq P(B A)$ — Savcı Yanılgısı	63
12.6 Biyohazard 2 ve 3: Prior \leftrightarrow Posterior ve Koşullu Bağımsızlık	64
12.7 Explaining Away: Bağımsız ama Koşullu Bağımlı	65
12.8 Bu Dersin Özeti	66
12.9 Kontrol Soruları	66

12.10Egzersizler	67
12.11Sonraki Ders İçin Hazırlık	67
12.12Anahtar Kavramlar (Cheat Sheet)	67
12.13ML Bağlantıları Özeti	68
13 Monty Hall ve Simpson Paradoksu	69
13.1 Bu Derste Ne Var?	69
13.2 Monty Hall: Kurulum ve Gizli Varsayımlar	69
13.3 Monty Hall: Ağaç Diyagramı	70
13.4 Monty Hall: LOTP ve Milyon Kapı	70
13.5 Simpson Paradoksu: İki Doktor	71
13.6 Simpson: Confounder ve Berkeley Davası	72
13.7 Bu Dersin Özeti	73
13.8 Kontrol Soruları	74
13.9 Egzersizler	74
13.10Sonraki Ders İçin Hazırlık	74
13.11Anahtar Kavramlar (Cheat Sheet)	75
13.12ML Bağlantıları Özeti	75
14 Kumarbazın İflası ve Rastgele Değişkenler	77
14.1 Bu Derste Ne Var?	77
14.2 Kumarbazın İflası: Random Walk Olarak	77
14.3 First-Step Analysis: Fark Denklemi	78
14.4 Fark Denklemini Çözmek	79
14.5 Moral: Minik Bias Felakettir	79
14.6 Rastgele Değişken Nedir?	80
14.7 Bernoulli ve Binom Dağılımları	80
14.8 Bu Dersin Özeti	81
14.9 Kontrol Soruları	82
14.10Egzersizler	82
14.11Sonraki Ders İçin Hazırlık	83
14.12Anahtar Kavramlar (Cheat Sheet)	83
14.13ML Bağlantıları Özeti	83
15 Rastgele Değişkenler ve Dağılımları	85
15.1 Bu Derste Ne Var?	85
15.2 Binom'a Üç Bakış + IID	85
15.3 $RV \neq$ Dağılım; $\{X = x\}$ Bir Olaydır	86
15.4 Dağılımı Tanımlamak: CDF ve PMF	86
15.5 İki Binom'un Toplamı: Konvolüsyon = Vandermonde	87
15.6 Hipergeometrik: Binom Sandığın Ama Olmayan	88
15.7 Hipergeometrik \leftrightarrow Binom: Büyük Popülasyon	88
15.8 Bu Dersin Özeti	88
15.9 Kontrol Soruları	89
15.10Egzersizler	90
15.11Sonraki Ders İçin Hazırlık	90
15.12Anahtar Kavramlar (Cheat Sheet)	90
15.13ML Bağlantıları Özeti	91

16 Beklenti, Gösterge Değişkenler ve Doğrusallık	93
16.1 Bu Derste Ne Var?	93
16.2 Ortalama → Beklenen Değer	93
16.3 Fundamental Bridge	94
16.4 Doğrusallık	94
16.5 Indicator + Linearity: Hipergeometrik Beklentisi	95
16.6 Geometrik Dağılım	96
16.7 Bu Dersin Özeti	96
16.8 Kontrol Soruları	97
16.9 Egzersizler	97
16.10 Sonraki Ders İçin Hazırlık	98
16.11 Anahtar Kavramlar (Cheat Sheet)	98
16.12 ML Bağlantıları Özeti	98
17 Beklentiye Devam	101
17.1 Bu Derste Ne Var?	101
17.2 Doğrusallığın İspatı	101
17.3 Negatif Binom ve First Success	102
17.4 Putnam Yerel Maksimumlar	103
17.5 St. Petersburg Paradoksu	103
17.6 Kritik Ders: $E[g(X)] \neq g(E[X])$	103
17.7 Bu Dersin Özeti	104
17.8 Kontrol Soruları	105
17.9 Egzersizler	105
17.10 Sonraki Ders İçin Hazırlık	106
17.11 Anahtar Kavramlar (Cheat Sheet)	106
17.12 ML Bağlantıları Özeti	106
18 Poisson Dağılımı	109
18.1 Bu Derste Ne Var?	109
18.2 Poisson Dağılımı: PMF ve Beklenti	109
18.3 Poisson Paradigması ve Binom Limiti	110
18.4 Nadir Olay Olasılığı: $P(\geq 1) \approx 1 - e^{-\lambda}$	111
18.5 Bu Dersin Özeti	111
18.6 Kontrol Soruları	111
18.7 Egzersizler	112
18.8 Sonraki Ders İçin Hazırlık	112
18.9 Anahtar Kavramlar (Cheat Sheet)	112
18.10 ML Bağlantıları Özeti	113
19 Kesikli vs Sürekli ve Uniform Dağılım	115
19.1 Bu Derste Ne Var?	115
19.2 PDF Yoğunluktur — Olasılık Değil	115
19.3 PDF \leftrightarrow CDF (FTC)	116
19.4 Varyans ve Standart Sapma	116
19.5 Uniform Dağılım	116
19.6 LOTUS: $E[g(X)]$	117
19.7 Uniform'un Evrenselliği: Inverse-Transform Sampling	117

19.8 Bu Dersin Özeti	118
19.9 Kontrol Soruları	119
19.10Egzersizler	119
19.11Sonraki Ders İçin Hazırlık	119
19.12Anahtar Kavramlar (Cheat Sheet)	120
19.13ML Bağlantıları Özeti	120
20 Normal Dağılım	121
20.1 Bu Derste Ne Var?	121
20.2 Probability Integral Transform	121
20.3 Uniform Simetrisi ve Doğrusal-Olmayan Tuzak	122
20.4 RV Bağımsızlığı: İkili \neq Tam	122
20.5 Standart Normal $N(0,1)$: PDF ve Gauss İntegrali	122
20.6 $E(Z) = 0, \text{Var}(Z) = 1$	123
20.7 Φ Notasyonu	124
20.8 Bu Dersin Özeti	124
20.9 Kontrol Soruları	124
20.10Egzersizler	125
20.11Sonraki Ders İçin Hazırlık	125
20.12Anahtar Kavramlar (Cheat Sheet)	125
20.13ML Bağlantıları Özeti	126
21 Konum, Ölçek ve LOTUS	127
21.1 Bu Derste Ne Var?	127
21.2 Konum-Ölçek: $X = \mu + \sigma Z$	127
21.3 Varyansın Özellikleri	128
21.4 Genel Normal PDF (Zincir Kuralı)	128
21.5 Bağımsız Normal Toplamı = Normal	128
21.6 68-95-99,7 Kuralı	129
21.7 LOTUS ile Varyans: Poisson ve Binom	129
21.8 LOTUS İspatı (Kesikli)	130
21.9 Bu Dersin Özeti	130
21.10Kontrol Soruları	131
21.11Egzersizler	131
21.12Sonraki Ders İçin Hazırlık	132
21.13Anahtar Kavramlar (Cheat Sheet)	132
21.14ML Bağlantıları Özeti	132
22 Ara Sınav İncelemesi	135
22.1 Bu Derste Ne Var?	135
22.2 Kupon Toplayıcı	135
22.3 Uniform'un Evrenselliği: Geometrik Sezgi	136
22.4 Lojistik Dağılım ve Logit	136
22.5 Simetri + Doğrusallık	137
22.6 LOTUS Tuzacı: İki Yol, Bir Cevap	137
22.7 Story Proof: $n - X \sim \text{Bin}(n, q)$	137
22.8 Poisson \rightarrow Üstel: İlk Bekleme Süresi	138
22.9 Üç Nesneyi Karıştırma	138

22.10	Bu Dersin Özeti	138
22.11	Kontrol Soruları	139
22.12	Egzersizler	139
22.13	Sonraki Ders İçin Hazırlık	139
22.14	Anahtar Kavramlar (Cheat Sheet)	140
22.15	ML Bağlantıları Özeti	140
23	Üstel (Exponential) Dağılım	141
23.1	Bu Derste Ne Var?	141
23.2	Üstel Dağılım	141
23.3	Belleksizlik	142
23.4	Koşullu Beklenti: $E(X X > a)$	142
23.5	Bu Dersin Özeti	143
23.6	Kontrol Soruları	143
23.7	Egzersizler	144
23.8	Sonraki Ders İçin Hazırlık	144
23.9	Anahtar Kavramlar (Cheat Sheet)	144
23.10	ML Bağlantıları Özeti	145
24	Moment Üreten Fonksiyonlar (MGF)	147
24.1	Bu Derste Ne Var?	147
24.2	Belleksizlik = Üstel (Fonksiyonel Denklem)	147
24.3	MGF Tanımı ve Üç Önem	148
24.4	Örnek MGF'ler	149
24.5	Laplace'ın Ardışıklık Kuralı	150
24.6	Bu Dersin Özeti	150
24.7	Kontrol Soruları	151
24.8	Egzersizler	151
24.9	Sonraki Ders İçin Hazırlık	151
24.10	Anahtar Kavramlar (Cheat Sheet)	152
24.11	ML Bağlantıları Özeti	152
25	MGF'lere Devam ve Joint Dağılımlar	153
25.1	Bu Derste Ne Var?	153
25.2	Üstel MGF ve Tüm Momentler	153
25.3	Normal'in Tüm Momentleri	154
25.4	Poisson MGF ve Toplamı	154
25.5	Joint Dağılımlar	154
25.6	Kare (Bağımsız) vs Disk (Bağımlı)	155
25.7	Bu Dersin Özeti	156
25.8	Kontrol Soruları	156
25.9	Egzersizler	157
25.10	Sonraki Ders İçin Hazırlık	157
25.11	Anahtar Kavramlar (Cheat Sheet)	157
25.12	ML Bağlantıları Özeti	158
26	Birleşik, Koşullu ve Marjinal Dağılımlar	159
26.1	Bu Derste Ne Var?	159

26.2	Joint \rightarrow Marjinal \rightarrow Koşullu	159
26.3	Sürekli Bağımsızlık	160
26.4	Disk Örneği: Marjinal ve Koşullu	160
26.5	2D LOTUS ve Bağımsız \Rightarrow İlişkisiz	161
26.6	İki Uniform Uzaklık: Max/Min Triki	161
26.7	Poisson Splitting (Tavuk-Yumurta)	162
26.8	Bu Dersin Özeti	162
26.9	Kontrol Soruları	163
26.10	Egzersizler	163
26.11	Sonraki Ders İçin Hazırlık	163
26.12	Anahtar Kavramlar (Cheat Sheet)	164
26.13	ML Bağlantıları Özeti	164
27	Multinomial ve Cauchy	165
27.1	Bu Derste Ne Var?	165
27.2	Bağımsız Normal Toplamı = Normal (MGF İspatı)	165
27.3	İki Normal Uzaklık (Yapıyı Kullan)	166
27.4	Multinomial Dağılım	166
27.5	Marjinal (Binom) ve Lumping	166
27.6	Multinomial Koşullu	166
27.7	Cauchy Dağılımı	167
27.8	Bu Dersin Özeti	167
27.9	Kontrol Soruları	168
27.10	Egzersizler	168
27.11	Sonraki Ders İçin Hazırlık	169
27.12	Anahtar Kavramlar (Cheat Sheet)	169
27.13	ML Bağlantıları Özeti	169
28	Kovaryans ve Korelasyon	171
28.1	Bu Derste Ne Var?	171
28.2	Kovaryans: Tanım ve Bilineerlik	171
28.3	Toplamın Varyansı	172
28.4	Bağımsız \Rightarrow İlişkisiz (Tersi YANLIŞ)	172
28.5	Korelasyon: $[-1, 1]$ ve Cauchy-Schwarz	172
28.6	Multinomial Kovaryansı	173
28.7	Binom + Hipergeometrik Varyansı (Göstergelerle)	173
28.8	Bu Dersin Özeti	174
28.9	Kontrol Soruları	174
28.10	Egzersizler	175
28.11	Sonraki Ders İçin Hazırlık	175
28.12	Anahtar Kavramlar (Cheat Sheet)	175
28.13	ML Bağlantıları Özeti	176
29	Dönüşümler ve Konvolüsyonlar	177
29.1	Bu Derste Ne Var?	177
29.2	Hipergeometrik Varyansı	177
29.3	Change of Variables (1D)	178
29.4	Log-Normal	178

29.5 Çok Boyutlu (Jacobian)	178
29.6 Konvolüsyon	178
29.7 Olasılıksal Yöntem	179
29.8 Bu Dersin Özeti	180
29.9 Kontrol Soruları	180
29.10Egzersizler	180
29.11Sonraki Ders İçin Hazırlık	181
29.12Anahtar Kavramlar (Cheat Sheet)	181
29.13ML Bağlantıları Özeti	181
30 Beta Dağılımı	183
30.1 Bu Derste Ne Var?	183
30.2 Beta Dağılımı	183
30.3 Beta-Binom Eşleniği	184
30.4 Pseudocount Yorumu	184
30.5 Bayes Bilardosu	185
30.6 Momentler ve Shrinkage	185
30.7 Finans LOTUS (Konuk)	185
30.8 Bu Dersin Özeti	186
30.9 Kontrol Soruları	186
30.10Egzersizler	186
30.11Sonraki Ders İçin Hazırlık	187
30.12Anahtar Kavramlar (Cheat Sheet)	187
30.13ML Bağlantıları Özeti	187
31 Gamma Dağılımı ve Poisson Süreci	189
31.1 Bu Derste Ne Var?	189
31.2 Gamma Fonksiyonu	189
31.3 Gamma Dağılımı	190
31.4 Poisson Süreci → Gamma	190
31.5 Momentler	191
31.6 Bu Dersin Özeti	191
31.7 Kontrol Soruları	191
31.8 Egzersizler	192
31.9 Sonraki Ders İçin Hazırlık	192
31.10Anahtar Kavramlar (Cheat Sheet)	192
31.11ML Bağlantıları Özeti	193
32 Sıra İstatistikleri ve Koşullu Beklenti	195
32.1 Bu Derste Ne Var?	195
32.2 Banka-Postane: Gamma → Beta	195
32.3 E(oran) = oran(E) Tuzağı	196
32.4 Sıra İstatistikleri	196
32.5 Uniform Sıra İstatistikleri = Beta	197
32.6 Koşullu Beklenti ve LOTE	197
32.7 Bu Dersin Özeti	198
32.8 Kontrol Soruları	198
32.9 Egzersizler	198

32.10	Sonraki Ders İçin Hazırlık	199
32.11	Anahtar Kavramlar (Cheat Sheet)	199
32.12	ML Bağlantıları Özeti	199
33	Koşullu Beklentiye Devam	201
33.1	Bu Derste Ne Var?	201
33.2	İki Zarf Paradoksu	201
33.3	Yazı-Tura Örüntüleri	202
33.4	$E(Y X)$ Bir Rastgele Değişken	202
33.5	Koşullu Beklenti Özellikleri	202
33.6	Poisson Örneği	203
33.7	Adam Yasası (Yineli Beklenti)	203
33.8	Bu Dersin Özeti	204
33.9	Kontrol Soruları	204
33.10	Egzersizler	205
33.11	Sonraki Ders İçin Hazırlık	205
33.12	Anahtar Kavramlar (Cheat Sheet)	205
33.13	ML Bağlantıları Özeti	206
34	Bir RD Verildiğinde Koşullu Beklenti	207
34.1	Bu Derste Ne Var?	207
34.2	4 Özellik	207
34.3	Projeksiyon Yorumu	208
34.4	Adam Yasası İspatı	208
34.5	Koşullu Varyans	208
34.6	Eve Yasası	208
34.7	Beta-Binom Hiyerarşik Örnek	209
34.8	Bu Dersin Özeti	210
34.9	Kontrol Soruları	210
34.10	Egzersizler	210
34.11	Sonraki Ders İçin Hazırlık	211
34.12	Anahtar Kavramlar (Cheat Sheet)	211
34.13	ML Bağlantıları Özeti	211
35	Eşitsizlikler	213
35.1	Bu Derste Ne Var?	213
35.2	Rassal Toplam: Adam + Eve	213
35.3	Cauchy-Schwarz	214
35.4	Jensen Eşitsizliği	214
35.5	Markov Eşitsizliği	214
35.6	Chebyshev Eşitsizliği	215
35.7	Bu Dersin Özeti	216
35.8	Kontrol Soruları	216
35.9	Egzersizler	216
35.10	Sonraki Ders İçin Hazırlık	217
35.11	Anahtar Kavramlar (Cheat Sheet)	217
35.12	ML Bağlantıları Özeti	218

36 Büyük Sayılar Yasası ve Merkezi Limit Teoremi	219
36.1 Bu Derste Ne Var?	219
36.2 Kurulum	219
36.3 Büyük Sayılar Yasası	220
36.4 Merkezi Limit Teoremi	220
36.5 MLT İspatı (MGF + L'Hôpital)	221
36.6 Binom Normal Yaklaşımı	221
36.7 Bu Dersin Özeti	221
36.8 Kontrol Soruları	222
36.9 Egzersizler	222
36.10 Sonraki Ders İçin Hazırlık	222
36.11 Anahtar Kavramlar (Cheat Sheet)	223
36.12 ML Bağlantıları Özeti	223
37 Ki-Kare, Student-t ve Çok Değişkenli Normal	225
37.1 Bu Derste Ne Var?	225
37.2 Ki-Kare (χ^2)	225
37.3 Student-t	226
37.4 Çok Değişkenli Normal (MVN)	227
37.5 MVN: Korelasyonsuz \implies Bağımsız	227
37.6 Bu Dersin Özeti	228
37.7 Kontrol Soruları	228
37.8 Egzersizler	228
37.9 Sonraki Ders İçin Hazırlık	229
37.10 Anahtar Kavramlar (Cheat Sheet)	229
37.11 ML Bağlantıları Özeti	229
38 Markov Zincirleri	231
38.1 Bu Derste Ne Var?	231
38.2 Markov Özelliği	231
38.3 Geçiş Matrisi Q	232
38.4 n -Adım Geçişler	232
38.5 Durağan Dağılım	232
38.6 Tarihsel İki Kullanım	233
38.7 Bu Dersin Özeti	233
38.8 Kontrol Soruları	234
38.9 Egzersizler	234
38.10 Sonraki Ders İçin Hazırlık	234
38.11 Anahtar Kavramlar (Cheat Sheet)	235
38.12 ML Bağlantıları Özeti	235
39 Markov: Sınıflandırma ve Tersinirlik	237
39.1 Bu Derste Ne Var?	237
39.2 İndirgenemezlik	237
39.3 Durum Türleri	238
39.4 Durağan Dağılım Teoremleri	238
39.5 Tersinirlik (Detailed Balance)	238
39.6 Yönsüz Ağda Rassal Yürüyüş	239

39.7 Bu Dersin Özeti	239
39.8 Kontrol Soruları	240
39.9 Egzersizler	240
39.10 Sonraki Ders İçin Hazırlık	241
39.11 Anahtar Kavramlar (Cheat Sheet)	241
39.12 ML Bağlantıları Özeti	241
40 Markov: Ağırlıklı Yürüyüş ve Google PageRank	243
40.1 Bu Derste Ne Var?	243
40.2 Ağırlıklı Rassal Yürüyüş	243
40.3 Her Tersinir Zincir = Ağırlıklı Yürüyüş	244
40.4 PageRank Markov Zinciri	244
40.5 Google Zinciri ve Işınlanma	244
40.6 Power Iteration	245
40.7 Bu Dersin Özeti	245
40.8 Kontrol Soruları	246
40.9 Egzersizler	246
40.10 Sonraki Ders İçin Hazırlık	247
40.11 Anahtar Kavramlar (Cheat Sheet)	247
40.12 ML Bağlantıları Özeti	247
41 İleriye Bakış (A Look Ahead)	249
41.1 Bu Derste Ne Var?	249
41.2 Top Ten: Stat 110'un Özü	249
41.3 Regresyon: β_1 Türetmesi	250
41.4 Horvitz-Thompson (IPW)	251
41.5 Basu'nun Fili: Yansız \neq İyi	251
41.6 Sonrası Nereye?	251
41.7 Kurs Kapanışı 🎓	252
41.8 Kontrol Soruları	252
41.9 Egzersizler	253
41.10 Anahtar Kavramlar (Cheat Sheet)	253
41.11 ML Bağlantıları Özeti	254
41.12 🎓 Kurs Sonu	254

1 Önsöz

2 Bu kitap nedir?

Bu, **Harvard Stat 110 — Introduction to Probability** dersinin (Joe Blitzstein, 34 ders) Türkçe ders notlarıdır. Hedef, izlerken paralel okunabilecek; sonradan tek başına da yeterli olabilecek bir referans seti üretmek.

Her bölüm bir “Builder Notu” katmanı taşır: kavramın **makine öğrenmesi** ile köprüsü. Koşullu olasılık → attention; Bayes teoremi → posterior + diffusion ters adımı; beklenen değer → Monte Carlo + loss; CLT → Gaussian gürültü; Markov zinciri → MCMC + RL value iteration. Olasılığı “tek başına matematik” olarak değil, ML’i türeten alet kutusu olarak okuyoruz.

Kaynak

- **Web sitesi:** stat110.net — ders sayfaları, slaytlar, strategic practice
- **Video dizisi:** [Statistics 110 — Probability \(YouTube playlist\)](#) (34 + 1 bonus video)
- **Yazar:** Joe Blitzstein — Professor of the Practice, Statistics, Harvard
- **Kitap:** *Introduction to Probability* — Blitzstein & Hwang (2014)
- **Çeviri ve genişletme:** Phase 1 (TR + ML köprüleri)

3 Nasıl Okumalı

Sıralı oku. Her ders bir öncekinin **dilini** kullanır. Atlamak istersen, en azından *Ders 1 (Olasılık ve Sayma)*, *Ders 4 (Koşullu Olasılık)* ve *Ders 9 (Beklenti, Doğrusallık)* zorunlu — bu üçü tüm seriyi taşıyor.

Pratik bir tavsiye

Her bölüm sonundaki **egzersizleri** atlama. Özellikle Python egzersizleri (simülasyon ile π yaklaşımı, koşullu olasılık histogramı, Markov sabit dağılımı) olasılık sezgisini parmaklarına yerleştirir. ML'de aynı kodu yıllarca farklı kılıklarda yazacaksın.

4 34 Ders

#	Ders	Ana Fikir
1	Olasılık ve Sayma	Örnek uzayı, çarpma kuralı, binom katsayısı
2	Story Proof'lar, Aksiyomlar	Kombinatorik özdeşliklerin hikâye ispatı
3	Doğum Günü, Özellikler	%50,7 mucize; içirme-dışarma
4	Koşullu Olasılık	Bayes teoreminin DNA'sı; bağımsızlık
5	LOTP, Bayes Detayı	Total Probability Law; explaining away
6	Monty Hall, Simpson	Sezgiyi yenen klasikler
7	Kumarbazın İflası, RD	Markov zincirinin ilk hâli
8	RD ve Dağılımlar	PMF, CDF, Bernoulli, Binom, Geometrik
9	Beklenti, Doğrusallık	Gösterge değişkenler — büyü
10	Beklentiye Devam	Hipergeometrik, Negatif Binom
11	Poisson	Nadir olayların matematiksel sınırı
12	Kesikli vs Sürekli, Uniform	PDF, inverse-transform sampling
13	Normal	CLT'nin ucu; Φ ve Gauss integrali
14	Konum-Ölçek, LOTUS	Standardizasyon; "kanunsuz istatistikçi"
15	Ara Sınav İncelemesi	Kupon toplayıcı; logit; süre
16	Exponential	Belleksizlik; sürekli geometrik
17	MGF	Moment üreten fonksiyon; tek belirleme
18	MGF Devamı, Joint	Bağımsız Poisson; karma türev
19	Marjinal, Koşullu	2D LOTUS; tavuk-yumurta
20	Multinomial, Cauchy	Lumping; oran dağılımı
21	Kovaryans, Korelasyon	İç çarpım yorumu; Cauchy-Schwarz
22	Dönüşümler, Konvolüsyon	Jacobian; bağımsız toplamlar
23	Beta	Bayes önsel-sonsalı; sıra istatistiği
24	Gamma, Poisson Süreci	Bekleme süreleri; çağrı merkezi
25	Sıra İstatistikleri	Min/max; koşullu beklenti girişi
26	Koşullu Beklenti Devamı	Adam ve Eve yasaları
27	Bir RD Verildiğinde	$E(Y X)$ — RD olarak beklenti
28	Eşitsizlikler	Cauchy-Schwarz, Jensen, Markov, Chebyshev
29	BSY ve MLT	Büyük Sayılar Yasası, Merkezi Limit
30	Ki-Kare, t, MVN	Normal'in türevleri; Mahalanobis
31	Markov Zincirleri	Geçiş matrisi; durağan dağılım
32	Markov Sınıflandırma	İndirgenemezlik; tersinirlik
33	Markov + PageRank	Ağırlıklı yürüyüş; Google'ın milyar dolarlık fikri
34	İleriye Bakış	Top Ten; regresyon; IPW

5 Notasyon

- **Olasılık:** $P(A)$ — olay A 'nın olasılığı
- **Koşullu olasılık:** $P(A | B)$ — B verildiğinde A 'nın olasılığı
- **Rastgele değişken:** X, Y, Z (büyük harf) — değerleri x, y, z (küçük)
- **Beklenti:** $E(X)$ veya $\mathbb{E}[X]$ — ortalama değer
- **Varyans:** $\text{Var}(X) = E((X - E(X))^2)$
- **Dağılım gösterimi:** $X \sim \text{Bin}(n, p)$ — “ X binom dağılımıdır”
- **PDF / PMF:** $f_X(x)$ (sürekli) ve $P(X = x)$ (kesikli)
- **CDF:** $F_X(x) = P(X \leq x)$

Tüm matematik [KaTeX](#) ile render ediliyor.

6 Builder Eksen — Olasılık → ML

💡 Her ders bu bağlantı katmanını taşıyor


Olasılık kavramı	ML karşılığı
Örnek uzay / olay	State space (RL), olay örnekleme (Monte Carlo)
Sayma / kombinatorik	Softmax paydası, partition function
Koşullu olasılık	Attention $p(\text{token} \mid \text{bağlam})$, nedensel çıkarım
Bayes teoremi	Posterior, MAP, ELBO, diffusion ters adımı
Bağımsızlık	Koşullu bağımsızlık (PGM, GNN), Naive Bayes
Beklenen değer	Loss $E[L]$, Monte Carlo integrasyon
Varyans	Belirsizlik kestirimi, MC dropout
Dağılımlar	VAE prior/posterior, GMM, mixture-of-experts
Normal & CLT	Gaussian gürültü varsayımı, diffusion ileri süreci, weight init
KL ıraksaması	Variational inference (ELBO), MLE = -KL
Markov zinciri	MCMC, diffusion geri süreci , RNN, MDP

! Bir tek şey

Olasılık iki büyük dilden ibarettir: **koşullama** ile **biriktirme**. Bayes teoremi bilgiyi yeniden tartar (koşullama); beklenti ve LOTUS ortalama davranışı biriktirir. Modern ML'in çoğunluğu — Bayes posterior, attention, ELBO, MCMC, diffusion — bu iki fikrin matematiksel sonucudur.

7 Yazım Kuralları

- **Türkçe terminoloji + parantez içinde İngilizce orijinal** ilk geçtiğinde: “üstel (exponential) dağılım”, “koşullu (conditional) bağımsızlık”.
- **Blitzstein’den alıntılar** İngilizce orijinal hâliyle, blockquote içinde verilir. Zaman damgası yanında.
- **Builder Notu** callout’ları her ana bölüm sonunda; ML köprüsünü buraya yazıyoruz.
- **Kontrol Soruları** collapse’lu — cevap kapalı başlar, okur kendi düşündükten sonra açar.
- **Egzersizler** cevapsız — Python görsel/simülasyon egzersizleri en az 1 tane.

 Bu kitap Blitzstein’in yerine geçmez

Tek başına bu set yetmez — Blitzstein’in tahta anlatımı ve sezgisi yerine geçemez. Önce videoyu izle, sonra ilgili dersi oku, son olarak kontrol sorularını **kapalıyken** cevapla. Set videoyu **destekler**, ikame etmez.

8 Olasılık ve Sayma

Belirsizlik nasıl sayılır: örnek uzay, naif tanım, çarpma kuralı, binom

i Bölüm bilgisi

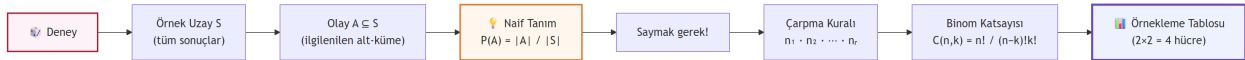
- **Blitzstein’in videosu:** [YouTube — Lecture 1: Probability and Counting](#) (≈46 dk)
- **Kaynak:** [Stat 110 ana sayfası](#) — Joe Blitzstein, Harvard
- **Okuma süresi:** ≈25 dk

8.1 Bu Derste Ne Var?

Bu, kursun ilk gerçek konu dersi. Blitzstein iki şey kuruyor: (1) olasılığın **ne** olduğu — yani belirsizliği nasıl sayıya çeviririz, ve (2) bunu hesaplayabilmek için ihtiyacımız olan ilk araç — **sayma (counting)**.

Üç temel fikir:

1. **Örnek uzay ve olay** — bir deneyin tüm olası sonuçlarının kümesi (örnek uzay) ve ilgilendiğimiz sonuç alt-kümesi (olay).
2. **Olasılığın naif tanımı** — “lehte sonuç / toplam sonuç”. Basit ama iki güçlü varsayıma dayanıyor: tüm sonuçlar eşit olası ve sonlu sayıda.
3. **Sayma** — naif tanımın payını ve paydasını bulabilmek için çarpma kuralı, binom katsayısı ve örnekleme tablosu.



Şekil 8.1: Ders 1'in kavram haritası: örnek uzay → naif tanım → sayma araçları

“*Math is the logic of certainty, statistics is the logic of uncertainty.*” — Blitzstein, 14:12

💡 Builder Notu — ML Köprüleri

- **Softmax çıktısı = olasılık dağılımı.** Örnek uzay = sınıflar, $P(\text{sınıf} \mid \text{girdi})$.
- **Cross-entropy loss** = $-\log P(\text{doğru etiket})$. Olasılığı maksimize etmek = loss'u minimize etmek (MLE).
- **Generative modeller** (LLM, diffusion, VAE) doğrudan olasılık dağılımı öğrenir ve ondan örnekler üretir.
- **Saymanın disiplini** = kombinatorik patlama. Bir LLM neden bütün cümleleri tek tek deneyemiyor:

V^L olası dizi var. Bu yüzden örnekleme, beam search, MCMC.

- Tek cümle: **olasılık, bir modelin “emin değilim ama şu kadar eminim” demesinin matematiğidir.**

8.2 Olasılık Nereye Uygulanır? Belirsizliğin Mantığı

Blitzstein dersin başında olasılığın nerelerde kullanıldığını sıralıyor — fizik (kuantum mekaniği baştan sona olasılık), genetik, ekonomi/ekonometri, oyun teorisi. Sonra daha az bariz örnekler: **tarih**. Mosteller ve Wallace, *The Federalist Papers*'ın (ABD Anayasası'nın onaylanmasıyla ilgili kritik metinler) tartışmalı yazarlığını **Bayes kuralıyla** çözmeye çalıştı — kim hangi makaleyi yazdı sorusunu olasılıkla yanıtladılar.

Olasılığın tarihsel kökü ise **kumar**. Konu buradan doğdu:

- 1650'lerin ortasında **Fermat** ve **Pascal**, kumar oyunlarını analiz eden uzun mektuplar yazıştılar. O zamana kadar kimse bu kuralları matematiksel türetmemiştir — ikisi mektuplaşarak olasılık teorisini sıfırdan kurdular.
- **Isaac Newton**'a bile kumarbazlar zar soruları soruyordu, çünkü o devirde olasılığı bilen başka kimse yoktu. İlginç olan: Newton hesabı doğru yaptı ama **sezgisi** bir zar probleminde yanıldı.

Bu son nokta dersin ruhunu veriyor: olasılık **derinden sezgi-karşıtı**. Newton'ın bile sezgisi yanlılabiliyorsa, bizimki de yanılacak — bu yüzden işi matematiksel kesinliğe bağlamak zorundayız.

“we’re going to do a lot of things that are deeply, deeply counterintuitive to almost everyone. ... to me, that makes this more fun than calculus.” — Blitzstein, 19:03

İstatistik, kalkülüsten farklı olarak insanı şaşırtır. Blitzstein'in özlü tanımı: matematik kesinliğin mantığı, istatistik belirsizliğin mantığı. Herkesin belirsizliği vardır; olasılık ve istatistik, inançlarımızı **niceliklendirme** ve **güncelleme** yöntemidir.

💡 Builder Notu — İnanç Güncelleme

“İnançları güncelleme” ifadesini aklında tut — bu, birazdan göreceğin **Bayes kuralının** tam tanımı ve modern Bayesian ML'in (posterior güncelleme, belief propagation) çekirdeği. Mosteller-Wallace'ın 1960'larda yazar tespiti için yaptığı şey, bugün bir spam filtresinin (naive Bayes) yaptığıyla matematiksel olarak aynı: metin kanıtından yazar/sınıf olasılığını güncellemek. **Diffusion** modellerin ters adımı da bir Bayes güncellemesi: gürültülü görüntü x_t verildiğinde temiz x_0 'ın posterior'ı.

8.3 Örnek Uzay ve Olay

Olasılığı tanımlamadan önce iki kavram gerekiyor.

Örnek uzay (sample space), bir deneyin tüm olası sonuçlarının kümesidir. Blitzstein “deney” kelimesini olabildiğince geniş yorumlamamızı istiyor — herhangi bir süreç, yeter ki farklı olası sonuçları olsun ve hangisinin gerçekleşeceğini önceden bilmeyelim.

“you should interpret the word *experiment* in an extremely broad manner. ... as long as there are certain possible outcomes.” — Blitzstein, 15:28

Örnek uzayı genelde büyük S harfiyle gösteririz. Bir parayı iki kez attığımız deneyde örnek uzay dört sonuçtan oluşur (H = yazı/heads, T = tura/tails — Blitzstein’in notasyonunu koruyoruz):

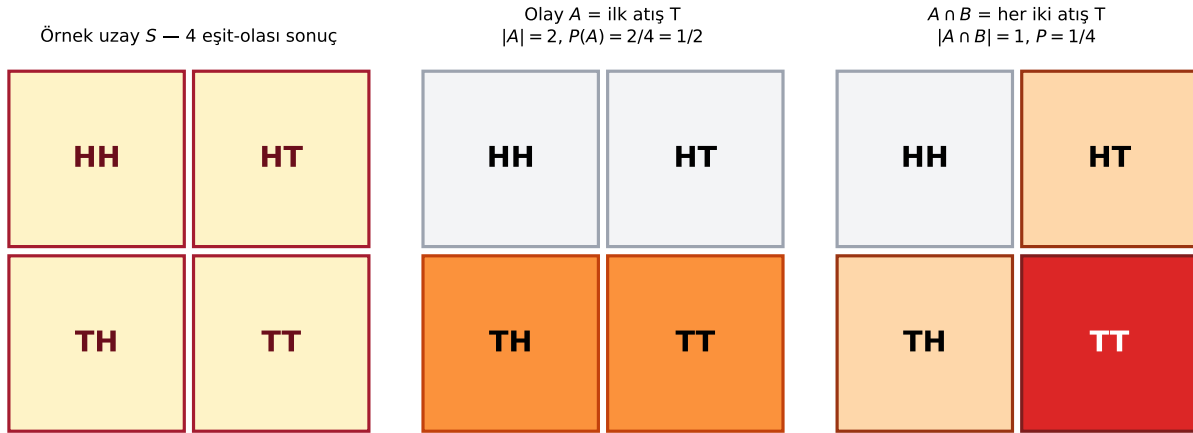
$$S = \{HH, HT, TH, TT\}$$

Olay (event), örnek uzayın bir **alt-kümesidir**. Yani ilgilendiğimiz sonuçların bir koleksiyonu. “Her iki atış da tura” olayı:

$$A = \{TT\}, \quad A \subseteq S$$

“An event is a subset of the sample space.” — Blitzstein, 16:33

Olasılık teorisinin matematiksel bir konu olabilmesini sağlayan büyük atılım tam da buydu: olayları **küme** olarak düşünmek. Bundan önce insanlar olasılık problemlerini sezgiyle çözmeye çalışıyordu ve çoğu sezgisel kural yanlış çıkıyordu. Kümeler (birleşim \cup , kesişim \cap , tümleyen) sayesinde “ A olayı **veya** B olayı”, “ A ve B ” gibi sezgisel ifadeler kesin matematiğe dönüştü. Bu kursta çok sayıda **Venn diyagramı** göreceksin.



Şekil 8.2: İki para atışının örnek uzayı $S = \{HH, HT, TH, TT\}$ ve üç ilginç olay: A = ilk atış tura, B = ikinci atış tura, $A \cap B$ = her ikisi tura.

💡 Builder Notu — Destek ve State Space

Örnek uzay, bir olasılık dağılımının **desteği (support)** ile aynı kavram. Bir dil modelinde örnek uzay = **sözlük (vocabulary)**; bir sonraki token’ın olası tüm değerleri. “Olay” ise bu sözlüğün bir alt-kümesi — örneğin “üretilen token bir noktalama işareti” bir olaydır. RL’de örnek uzayın karşılığı **state space**’tir. Olayları küme işlemleriyle birleştirme alışkanlığı, ileride birleşik / koşullu olasılıkları (joint / conditional) doğru kurmanın temeli.

8.4 Olasılığın Naif Tanımı

Artık olasılığı tanımlayabiliriz. Tüm kurs boyunca büyük P harfi “olasılık” demektir: $P(A)$, yani “ A olayının olasılığı”.

Naif tanım şöyle der: bir A olayının olasılığı, A 'ya **lehte (favorable)** sonuçların sayısının, **toplam** olası sonuç sayısına oranıdır.

$$P(A) = \frac{|A|}{|S|}$$

Burada $|A|$ = A 'ya lehte sonuç sayısı (A kümesinin eleman sayısı), $|S|$ = örnek uzaydaki toplam sonuç sayısı. Pay “kaç tanesi istediğimiz şey”, payda “toplam kaç olasılık var”.

Örnek: parayı iki kez at. Örnek uzay $S = \{HH, HT, TH, TT\}$, yani $|S| = 4$. “Her iki atış da tura” olayı $A = \{TT\}$, yani $|A| = 1$. O hâlde:

$$P(A) = \frac{1}{4}$$

Bu, lisede gördüğün olasılık tanımının ta kendisi: kaç olasılık var say, kaç istediğin durum say, böl.

“we just count how many of those happen, divide by a number of things, that’s it. That’s the naive definition.” — Blitzstein, 24:34

Builder Notu — Uniform = Maksimum Entropi

Naif tanım aslında **sonlu bir küme üzerindeki düzgün (uniform) dağılımdır** — her sonuca eşit ağırlık. Bilgi-teorik olarak bu, “elimde sadece sonlu sonuç var, başka hiçbir şey bilmiyorum” durumunun **maksimum entropili** dağılımıdır. ML’de karşılığı: bir sınıflandırıcıyı eğitmeye başlamadan önceki **uniform prior**, ya da rastgele tahmin eden **baseline** (10 sınıf varsa şans seviyesi 1/10). Payı/paydayı sayarak olasılık bulmak, dağılım gerçekten uniform olduğunda yapılan **kesin (exact) çıkarımdır** — Monte Carlo gibi yaklaşıklığa gerek kalmaz.

8.5 Naif Tanımın İki Varsayımı (ve Nerede Çöker)

Naif tanım bedava değil. İki güçlü varsayıma yaslanır:

1. **Tüm sonuçlar eşit olası.**
2. **Sonlu sayıda sonuç var** — yoksa payda sonsuz olur ve oran anlamını yitirir.

“It assumes that all outcomes are equally likely. ... it also assumes that there are finitely many outcomes.” — Blitzstein, 24:48

Birinci varsayım ne zaman makul? **Simetri** olduğunda. Altı yüzü simetrik bir zar atıyorsan, her yüzün 1/6 olması mantıklı. Ama zar **hileli (loaded)** ise — bir tarafa ağırlık verilmişse — eşit olasılık varsayımı çöker. Yani “eşit olası” bedava bir varsayım değil; onu **gerekçelendirmen** gerekir (genelde simetriyle).

Gerekçesiz uygulanırsa naif tanım saçmalığa götürür. Blitzstein’in klasik örneği: “Neptün’de yaşam olma olasılığı nedir? Ya vardır ya yoktur, iki olasılık, demek ki 1/2.”

“either there is or there isn’t, that’s two possibilities. ... so it’d be 1/2.” — Blitzstein, 26:19

Bu açıkça saçma. Daha da kötüsü: “Neptün’de **akıllı** yaşam olma olasılığı?” Yine ya vardır ya yoktur → 1/2. Oysa akıllı yaşam, herhangi bir yaşamdan **kesinlikle daha az olası** olmalı (akıllı yaşam \subseteq yaşam). Naif tanım bu kesin eşitsizliği yakalayamıyor — çünkü “iki seçenek var” demek onları eşit olası yapmaz.

Ders bu yüzden hızla naif tanımın ötesine geçecek. Ama naif tanım hâlâ önemli: hem konunun tarihsel doğuşu, hem de simetrinin geçerli olduğu (zar, kart, yazı-tura) bir sürü problemde doğru araç.

! Builder Notu — Base-Rate Neglect

“Eşit olası varsayımı gerekçe ister” dersi, ML’de **uniform prior’ın tehlikesidir**. Neptün safsatası = kendi uydurduğu bir bölümlenme üzerine uniform dağılım dayatmak. Pratik karşılığı: sınıflar dengesizken (örneğin %1 dolandırıcılık) modelin uniform sınıf prior’ı varsayması onu **kalibrasyonsuz** yapar — gerçek **base rate**’i veriden öğrenmen gerekir. “İki seçenek var, demek ki 1/2” hatası, naive sınıflandırıcıları vuran **base-rate neglect**’in ta kendisidir; Bayesian prior tam da bunu düzeltir. Kısaca: dağılımları **varsayma**, veriden **öğren**.

8.6 Neden Saymak? Çarpma Kuralı

Naif tanımda payı ve paydayı bulmak için sonuçları saymamız gerekiyor. Dört sonuçlu para örneğini elle yazdık, ama gerçek problemlerde sonuçları tek tek listelemek imkânsız hâle gelir. Bu yüzden Blitzstein kursun ilk büyük teknik konusunu sayma olarak koyuyor:

“the first major topic in this class is, how do we count?” — Blitzstein, 28:20

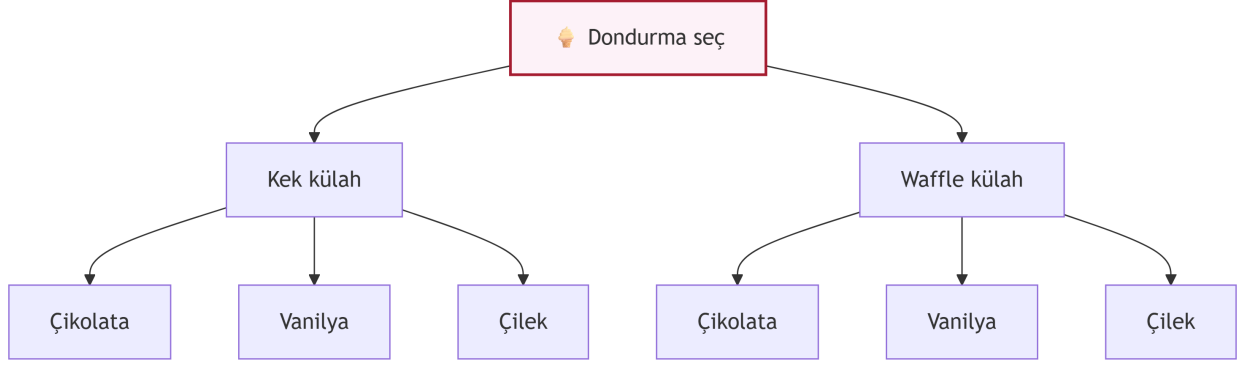
İlk ve en temel araç **çarpma kuralı (multiplication rule)**. Kural şöyle: bir deney sırayla r aşamadan oluşuyorsa ve

- Aşama 1’de n_1 olası sonuç,
- her Aşama 1 sonucu için Aşama 2’de n_2 olası sonuç,
- ve önceki aşamalardan bağımsız olarak Aşama r ’de n_r olası sonuç varsa,

birleşik deneyin toplam sonuç sayısı bu çarpımdır:

$$n_1 \times n_2 \times \cdots \times n_r$$

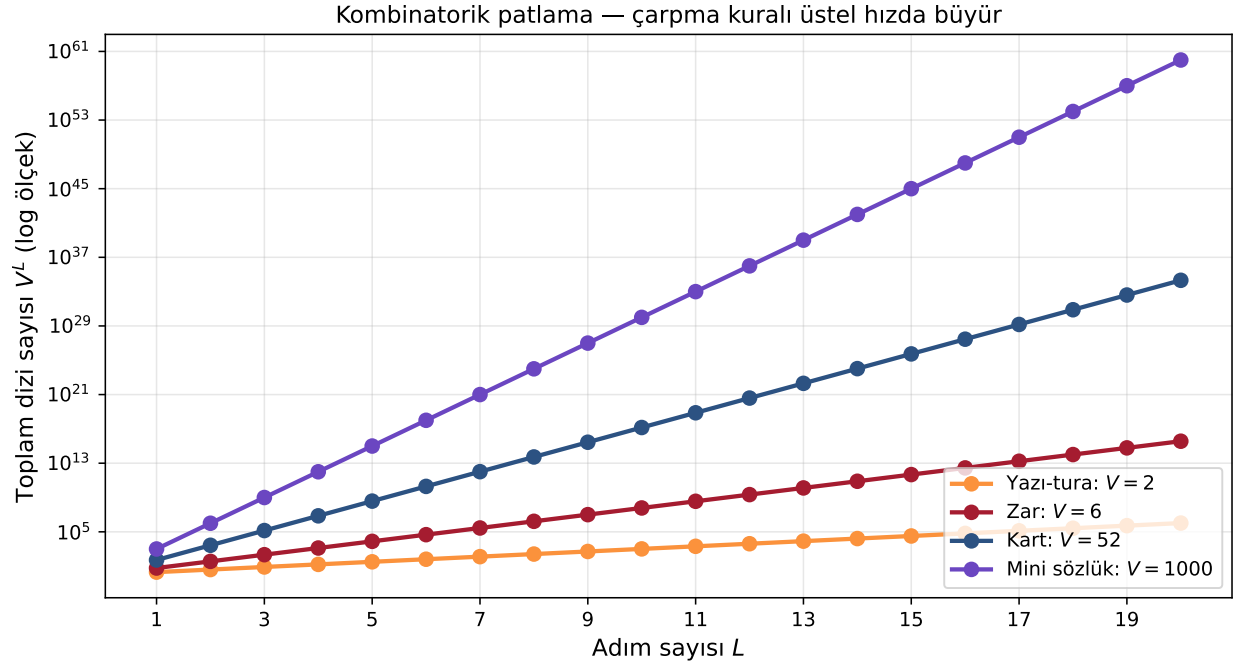
Blitzstein’in dondurma örneği bunu somutlaştırır: 3 çeşit (çikolata, vanilya, çilek) ve 2 tip külah (kek, waffle). Kaç seçenek var? Bir **ağaç diyagramı** çiz — önce külah dalı 2’ye ayrılır, sonra her dal çeşit için 3’e ayrılır: $2 \times 3 = 6$. Önce çeşidi seçsen de sonuç değişmez: $3 \times 2 = 6$.



Şekil 8.3: Dondurma ağacı: 2 külah \times 3 çeşit = 6 sonuç. Hangi sırayla seçtiğin önemsiz.

“once you understand this example completely, then all this stuff I wrote here becomes obvious.”
— Blitzstein, 32:17

Kritik sezgi: bu çarpımlar **üstel** hızla büyür. Her adımda 2 seçenek ve 10 adım olsa $2^{10} = 1024$ sonuç eder. İşte bu yüzden sonuçları elle listelemek umutsuz — saymanın kurallarına ihtiyacımız var.



Şekil 8.4: Çarpma kuralı üstel büyür. Her adımda V seçenek ve L adım olunca V^L olası dizi. Bir LLM’de $V \approx 50,000$, $L = 50$ token $\rightarrow 50000^{50} \approx 10^{234}$ dizi. Bütünü deneyemezsin.

💡 Builder Notu — Zincir Kuralı ve Beam Search

Çarpma kuralı, olasılığın **zincir kuralının (chain rule)** iskeletidir:

$$p(x_1, \dots, x_n) = p(x_1) \cdot p(x_2 | x_1) \cdots p(x_n | x_1, \dots, x_{n-1})$$

Bir **autoregressive dil modeli** tam da budur — her token bir aşama, sözlük boyutu V kadar seçenek; L token'lık bir dizi için V^L olası dizi vardır. Bu kombinatorik patlama modern üretimin neden **bütün dizileri deneyemediğini** ve neden örnekleme / beam search yaptığını açıklar. Blitzstein'in ağaç diyagramı, beam search'ün gezdiği **arama ağacının** ta kendisidir.

8.7 Binom Katsayısı

Çarpma kuralından çıkan en önemli sayma aracı **binom katsayısı**: “ n 'den k seç”. n nesneden, **sırası önemsiz** olacak şekilde k tanesini seçmenin kaç yolu olduğunu verir.

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

($k > n$ ise tanım gereği 0'dır — 10 kişiden 11 tanesini seçemezsin.)

“We'll be seeing a lot of these in this course, those are called a binomial coefficient.” — Blitzstein, 36:01

Bu formül nereden geliyor? Doğrudan çarpma kuralından. Önce n kişiden k tanesini **sıralı** seçtiğimizi düşün: ilki için n , ikincisi için $n - 1$, üçüncüsü için $n - 2$ seçenek... k . seçim için $n - k + 1$ seçenek:

$$n(n-1)(n-2)\dots(n-k+1)$$

Ama biz sırayı umursamıyoruz. Aynı k kişilik grubu $k!$ farklı sırada seçebilirdik — yani her grubu $k!$ kez fazladan saydık. Bu fazlalığı düzeltmek için $k!$ 'e böleriz:

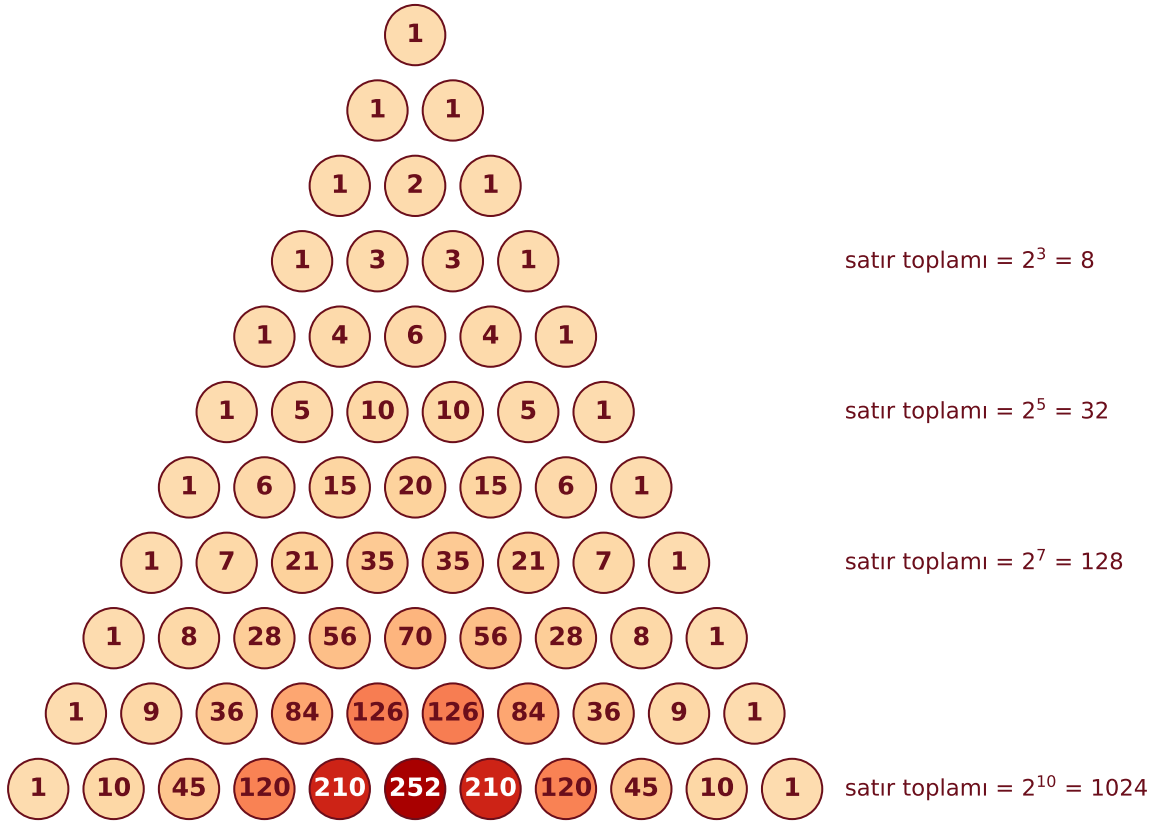
$$\binom{n}{k} = \frac{n(n-1)\dots(n-k+1)}{k!} = \frac{n!}{(n-k)!k!}$$

İki ifade aynıdır: faktöriyelleri açıp sadeleştirirsen $(n-k)!$ kısmı iptal olur ve ortadaki çarpım kalır.

💡 Builder Notu — Alt-Küme Sayma

Binom katsayısı “kaç farklı alt-küme” sorusunun cevabıdır. **Dropout**'ta her adımda hangi nöronları söndüreceğini seçmek bir alt-küme seçimidir; n nöronun k tanesini söndürmenin $\binom{n}{k}$ yolu vardır. Olası konfigürasyonları sayma, istatistiksel fizikteki **partition function** ve bilgi teorisindeki entropi hesaplarının temelidir. Ayrıca binom katsayıları, **binom dağılımının** (n bağımsız denemede k başarı olasılığı) çekirdeğidir — onu [Ders 8](#)'de göreceğiz.

Pascal üçgeni — $C(n, k)$ değerleri (renk yoğunluğu = büyüklük)



Şekil 8.5: Pascal üçgeni: her hücre $\binom{n}{k}$. Yatay simetri $\binom{n}{k} = \binom{n}{n-k}$ ve toplam $\sum_k \binom{n}{k} = 2^n$ (her satırın toplamı bir sonraki 2'nin kuvveti).

8.8 Worked Example: Poker'de Full House

Naif tanımını gerçek bir problemde çalıştıralım. 52 kartlık standart desteden 5 kart çekiyorsun ve deste tam karılmış, yani **tüm 5'li eller eşit olası** (simetri var, naif tanım meşru). Bir **full house**, üç tanesi bir değerden (rank), iki tanesi başka bir değerden olan eldir — örneğin üç tane 7 ve iki tane 10.

Naif tanım: payda = olası el sayısı, pay = full house el sayısı.

Payda: 52 karttan 5 kart seçmenin yolu, sıra önemsiz:

$$\binom{52}{5}$$

Pay (ağaç gibi düşün, çarpma kuralı):

- Üç tanesine sahip olduğumuz değeri seç: 13 olasılık (as, 2, ..., papaz).
- O değer için 4 kartından 3'ünü seç: $\binom{4}{3}$.
- İki tanesine sahip olduğumuz değeri seç: ilk değer hariç 12 olasılık.
- O değer için 4 kartından 2'sini seç: $\binom{4}{2}$.

Hepsini çarpma kuralıyla birleştir:

$$P(\text{full house}) = \frac{13 \cdot \binom{4}{3} \cdot 12 \cdot \binom{4}{2}}{\binom{52}{5}}$$

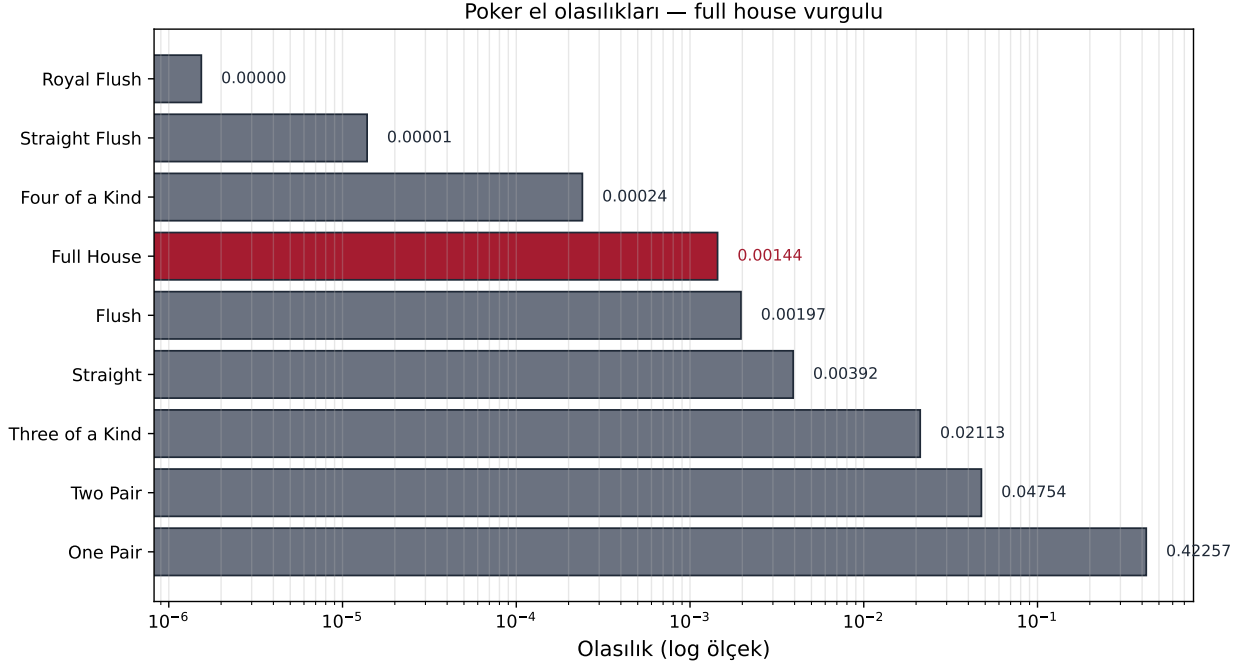
Sayıları koyalım: $\binom{4}{3} = 4$, $\binom{4}{2} = 6$, yani pay = $13 \cdot 4 \cdot 12 \cdot 6 = 3744$. Payda $\binom{52}{5} = 2,598,960$. Sonuç $\approx 0,00144$ — yani yaklaşık her 694 elde bir full house.

“it helps to think in terms of the tree — think in terms of the multiplication rule.” — Blitzstein, 40:59

Blitzstein'in vurgusu: başka doğru yollar da var, ama ağaç/çarpma kuralı üzerinden düşünmek daha **yapısal** ve hata yapma olasılığın daha düşük.

Builder Notu — Konfigürasyon Sayma ve Permütasyon Simetrisi

Bu problem aslında **konfigürasyon sayma** — belirli bir yapının kaç farklı şekilde gerçekleşebileceğini saymak. İstatistiksel fizikte aynı sayma “mikrodurum sayısı”dır ve doğrudan **entropiye** (log of count) bağlanır. Payı çarpma kuralıyla aşamalara bölmek, bir **birleşik dağılımı koşullu parçalara çarpanlamakla** ($p(a, b, c) = p(a) \cdot p(b | a) \cdot p(c | a, b)$) aynı disiplindir. “Sırayı umursama, $k!$ 'e böl” hamlesi ise **permütasyon simetrisidir**; Deep Sets gibi permütasyon-değişmez (permutation-invariant) mimariler tam olarak bu simetriyi modele gömer.



Şekil 8.6: Poker el olasılıkları log ölçekte. Full house (kırmızı) $\approx 0,144\%$, four-of-a-kind ondan altı kat daha nadir. Royal flush 0,00015%.

8.9 Örnekleme Tablosu (2×2)

Tüm sayma problemlerini tek çatıda toplayalım. **Örnekleme (sampling)**: n nesneli bir popülasyondan k nesne çekiyoruz. İki ikili seçim var:

- **Yerine koyarak mı, koymadan mı?** (with/without replacement) — çektiğimizi geri koyup tekrar seçebilir miyiz?
- **Sıra önemli mi, değil mi?** (order matters or not)

Bu iki ikili seçim 2×2 tablo verir:

Örnekleme	Sıra önemli	Sıra önemsiz
Yerine koyarak	n^k	$\binom{n+k-1}{k}$
Yerine koymadan	$n(n-1) \cdots (n-k+1)$	$\binom{n}{k}$

Üç tanesi çarpma kuralından **anında** çıkar:

- **Yerine koyarak, sıralı:** her seçimde n seçenek, k kez $\rightarrow n^k$.
- **Yerine koymadan, sıralı:** $n, n-1, \dots$ azalan $\rightarrow n(n-1) \cdots (n-k+1)$.
- **Yerine koymadan, sırasız:** tam da binom katsayısı $\rightarrow \binom{n}{k}$.

Dördüncüsü (**yerine koyarak, sırasız**) çok daha incedir — Blitzstein “bir derece daha zor” diyor ve sonucu $\binom{n+k-1}{k}$ olarak veriyor, ispatını bir sonraki derse bırakıyor.

“I don’t want you to memorize this table, I want you to understand.” — Blitzstein, 44:15

Yani amaç tabloyu ezberlemek değil; her hücrenin neden öyle olduğunu çarpma kuralından **türetebilmek**.

💡 Builder Notu — Bootstrap, SGD, Set Modelleri

Bu 2×2 tablo ML’de her gün karşına çıkar:

- **Yerine koyarak örnekleme = bootstrap**: bagging ve random forest, veriden yerine koyarak tekrar örnekleyerek çeşitlilik üretir.
- **Yerine koymadan örnekleme = SGD’de her epoch’ta** veriyi karıştırıp minibatch’leri yerine koymadan çekmek.
- **Sıra önemli/önemsiz** eksenini ise mimari seçiminin kalbidir: sıraya duyarlı modeller (RNN, positional encoding’li transformer) vs. sıra-değişmez küme modelleri (Deep Sets, pozisyonuz attention).

Aynı k nesneyi farklı sırada “aynı” saymak, modele **permütasyon simetrisi** öğretmektir.

8.10 Bu Dersin Özeti

1. **Örnek uzay**, bir deneyin tüm olası sonuçlarının kümesidir (S); **olay**, örnek uzayın bir alt-kümesidir ($A \subseteq S$).
2. **Olasılığın naif tanımı**: $P(A) = |A|/|S|$ — lehte sonuç sayısı bölü toplam sonuç sayısı.
3. Naif tanım iki varsayımı ister: **tüm sonuçlar eşit olası** ve **sonlu** sayıda. “Eşit olası” bedava değil, gerekçe (genelde simetri) ister; “Neptün’de yaşam $\rightarrow 1/2$ ” örneği gerekçesiz kullanımın saçmalığını gösterir.
4. **Sayma**, kursun ilk büyük aracıdır — çünkü payı ve paydayı elle listelemek çabuk imkânsızlaşır.
5. **Çarpma kuralı**: r aşamalı bir deneyde toplam sonuç sayısı $n_1 \cdot n_2 \cdots n_r$. Ağaç diyagramıyla görselleştir; **üstel** hızla büyür.
6. **Binom katsayısı** $\binom{n}{k} = \frac{n!}{(n-k)!k!}$: n nesneden sırasız k seçme. Sıralı seçip $k!$ ’e bölerek türetilir.
7. **Örnekleme tablosu** (2×2): yerine koyarak/koymadan \times sıra önemli/önemsiz $\rightarrow n^k, n(n-1) \cdots (n-k+1), \binom{n}{k}, \binom{n+k-1}{k}$.

! Tek bir cümle

Olasılık, belirsizliği saymaya indirger: eşit-olası ve sonlu bir dünyada bir olayın olasılığı “kaç yol lehte / kaç yol toplam”dır — ve bu yolları doğru saymak **çarpma kuralı** ile **binom katsayısının** işidir.

8.11 Kontrol Soruları

i Soru 1: Adil parayı 3 kez at. Örnek uzayda kaç sonuç var ve tam olarak iki tura (T) gelme olasılığı nedir?

Cevap: Çarpma kuralıyla $|S| = 2 \times 2 \times 2 = 2^3 = 8$ sonuç. “Tam iki T” olayının lehte sonuçları: TTH, THT, HTT — yani 3 tane (ya da $\binom{3}{2} = 3$, üç atıştan T gelecek ikisini seçmek). Hepsini eşit olası olduğundan $P = 3/8$.

i Soru 2: 52 kartlık desteden 5 kart çekersen ‘dört aynı (four of a kind)’ olma olasılığı nedir?

Cevap: Payda yine $\binom{52}{5}$. Pay için çarpma kuralı: dört kartın değerini seç (13 olasılık) \times o değer için 4 kartın hepsini al ($\binom{4}{4} = 1$) \times kalan 5. kartı diğer 48 kart arasından seç (48). Pay = $13 \cdot 1 \cdot 48 = 624$. Yani $P = 624/2,598,960 \approx 0,00024$ — full house’tan çok daha nadir, mantıklı, çünkü dört aynı daha güçlü bir el.

i Soru 3: Sıralı seçim $n(n-1) \dots (n-k+1)$, sırasız seçim $\binom{n}{k}$ yol verir. $k = n$ özel durumunda ikisi ne verir, ne doğrular?

Cevap: $k = n$ iken sıralı seçim $n(n-1) \dots 1 = n!$ verir (n nesnenin tüm permütasyonları). Sırasız seçim $\binom{n}{n} = 1$ verir (n nesnenin tümünü almanın tek bir yolu var — hepsini al). Aralarındaki oran $n!/1 = n!$, yani “sırayı umursamayınca $k! = n!$ ’e böl” kuralıyla birebir tutarlı: aynı tek grubu $n!$ farklı sırada dizebilirsin.

i Soru 4: (Builder) Naif tanımın ‘eşit olası’ varsayımı bir sınıflandırıcının hangi varsayımına denk düşer ve ne zaman tehlikelidir?

Cevap: Uniform sınıf prior’a denk düşer — her sınıfı eşit olası saymak. Sınıflar dengesizse tehlikelidir: örneğin %99 sağlam, %1 arızalı bir veride modelin (veya naif bir kuralın) sınıfları eşit olası varsayması **base rate**’i yok sayar ve kalibrasyonsuz, yanıltıcı olasılıklar üretir. Tıpkı “Neptün’de yaşam ya var ya yok $\rightarrow 1/2$ ” safsatası gibi. Doğrusu: prior’ı veriden (gerçek sınıf frekanslarından) öğrenmek, gerekirse class weighting veya olasılık kalibrasyonu uygulamak.

8.12 Egzersizler

Egzersiz 1. Bir adil zarı iki kez atıyorsun. (a) Örnek uzayda kaç sonuç var? (b) İki atışın toplamının 7 olma olasılığı nedir? (c) Toplamın 7 olmama olasılığı nedir? (İpucu: önce $|S|$ ’yi çarpma kuralıyla bul, sonra lehte sonuçları say.)

Egzersiz 2. 52 kartlık desteden 5 kartlık bir el çekiyorsun. Beş kartın da aynı türden (suit) olma olasılığını naif tanımla, binom katsayıları cinsinden yaz. (İki aşama: önce türü seç, sonra o türün 13 kartından 5’ini seç.) İfadeyi sadeleştirmeden $\binom{\cdot}{\cdot}$ cinsinden bırakabilirsin.

Egzersiz 3. Bir sayma argümanı (formülü açmadan, tek cümleyle) şunu açıkla: $\binom{n}{k} = \binom{n}{n-k}$. (İpucu: k tanesini “seçmek”, aslında dışarıda bırakılacak $n - k$ tanesini seçmekle aynı şeydir.)

Egzersiz 4. (*Python — full house Monte Carlo*) Full house olasılığını hem formülle hem de simülasyonla doğrula. İkinci kısım, bu dersin “örnekleme” fikrinin pratiğidir.

formül: 0.00144
simülasyon: 0.00136 ($N = 200,000$)

Egzersiz 5. (*Sonraki dersin habercisi*) Örnekleme tablosunun en zor hücresi “yerine koyarak, sırasız” $\rightarrow \binom{n+k-1}{k}$. $n = 2, k = 3$ için tüm olasılıkları elle listele (yani $\{1, 2\}$ kümesinden, sırası önemsiz, tekrarlı 3 seçim) ve bunların sayısının formülün verdiği $\binom{4}{3} = 4$ ile eşleştiğini göster. Ders 2’de bunun neden böyle olduğunu (“stars and bars” / Bose-Einstein) ispatlayacağız.

8.13 Sonraki Ders İçin Hazırlık

Ders 2: Story Proof’lar ve Olasılık Aksiyomları

Ders 2’de naif tanımın ötesine geçiyoruz. Örnek uzayın sonlu ve eşit-olası olma zorunluluğunu kaldıran **genel (aksiyomatik) olasılık tanımını** kuracağız. Ayrıca formül cebiriyle değil, doğrudan bir **sayma hikâyesiyle** ispat yapma tekniği olan “story proof”u öğreneceğiz — ve örnekleme tablosunun bıraktığımız zor hücrelerini $\binom{n+k-1}{k}$ bu yöntemle ispatlayacağız.

Ana konular:

- Olasılığın **aksiyomları**: negatif olmama, $P(S) = 1$, ve ayrık olaylar için toplanabilirlik.
- **Story proof**: bir kombinatorik kimliği iki farklı sayma yoluyla kanıtlanma.
- “Yerine koyarak, sırasız” örnekleme ispatı (**stars and bars / Bose-Einstein**).

⚠ Ders 2 öncesi yapılacak

- Bu dersin egzersizlerini çöz — özellikle 4 (Python) ve 5 (elle listeleme).
- Python’da `math.comb` ve `itertools` ile küçük sayma deneyleri yap.
- Ana cümleyi tekrar oku: “*Olasılık, belirsizliği saymaya indirger.*”

8.14 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Blitzstein’de
Örnek uzay (S)	Bir deneyin tüm olası sonuçlarının kümesi	15m20
Olay	Örnek uzayın bir alt-kümesi ($A \subseteq S$)	16m33
Naif tanım	$P(A) = A / S $	24m34
İki varsayım	Eşit olası + sonlu; eşit olası simetriyle gerekçelenir	24m48

Kavram	Tanım	Blitzstein'de
Neptün safsatası	Gerekçesiz “ya var ya yok $\rightarrow 1/2$ ” hatası	26m19
Çarpma kuralı	r aşamalı deneyde $n_1 \cdot n_2 \cdots n_r$	29m05
Ağaç diyagramı	Çarpma kuralının görseli (dondurma: $2 \times 3 = 6$)	32m11
Binom katsayısı	$\binom{n}{k} = n! / ((n-k)! k!)$, sırasız k seçme	36m01
$k!$ 'e bölme	Sıralı seçim $\div k! =$ sırasız seçim	38m26
Full house	$13 \cdot \binom{4}{3} \cdot 12 \cdot \binom{4}{2} / \binom{52}{5} \approx 0,00144$	39m01
Yerine koyarak, sıralı	n^k	43m47
Koymadan, sıralı	$n(n-1) \cdots (n-k+1)$	44m23
Koymadan, sırasız	$\binom{n}{k}$	44m03
Yerine koyarak, sırasız	$\binom{n+k-1}{k}$ (Ders 2'de ispat)	45m26

8.15 ML Bağlantıları Özeti

💡 7 köprü

- Örnek uzay / olay** \rightarrow bir olasılık dağılımının desteği; dil modelinde sözlük, RL'de state space, olay = ilgilenilen alt-küme.
- Naif tanım (uniform)** \rightarrow uniform prior ve rastgele baseline; sonlu destek üzerinde maksimum entropili dağılım.
- “Eşit olası gerçeğe ister”** \rightarrow dengesiz veride **base-rate neglect** tehlikesi; prior'ı veriden öğren ve modeli kalibre et.
- Çarpma kuralı** \rightarrow olasılığın **zincir kuralı** $p(x_1, \dots, x_n) = \prod_i p(x_i | x_{<i})$; autoregressive dil modelinin token-token üretimi.
- Kombinatorik patlama** (n^k, V^L) \rightarrow bütün dizileri deneyemezsin; bu yüzden sampling, beam search, MCMC gibi yaklaşık yöntemler.
- Binom katsayısı** \rightarrow alt-küme sayma; dropout maskeleri, partition function, ve binom dağılımı (Ders 8).
- Örnekleme tablosu** \rightarrow **bootstrap** (yerine koyarak) vs **SGD minibatch** karıştırma (koymadan); sıra önemi \rightarrow sequence modelleri (RNN, positional transformer) vs set modelleri (Deep Sets, pozisyonuz attention).

! Tek bir Őey alıp gideceksen

Olasılık, “**kaç yol?**” sorusudur. EŐit-olası ve sonlu bir dñnyada her olasılık bir sayma problemine indirgenir — ve saymanın iki temel aleti **çarpma kuralı** ile **binom katsayısıdır**. Ders 2’den itibaren yapacaęımız Őey, bu “eŐit-olası” varsayımını gevŐetip gerçek dñnyaya aılmak.

9 Story Proof'lar ve Olasılık Aksiyomları

Aynı şeyi iki yolla say + olasılığı iki kurala indir

i Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 2: Story Proofs, Axioms of Probability](#) (≈46 dk)
- **Okuma süresi:** ≈25 dk

9.1 Bu Derste Ne Var?

Ders 1'de naif tanımı ve saymanın temel araçlarını kurduk. Burada iki büyük adım atıyoruz:

1. **Etiketleme + story proof** — sayma problemlerini doğru kurmanın ve kombinatorik özdeşlikleri cebirsiz, “hikâyeyle” ispatlamanın yolu.
2. **Stars and bars** — “yerine koyarak, sırasız” örneklemenin neden $\binom{n+k-1}{k}$ olduğunun ispatı.
3. **Olasılık uzayı (S, P) + iki aksiyom** — tüm olasılık teorisinin dayandığı temel.

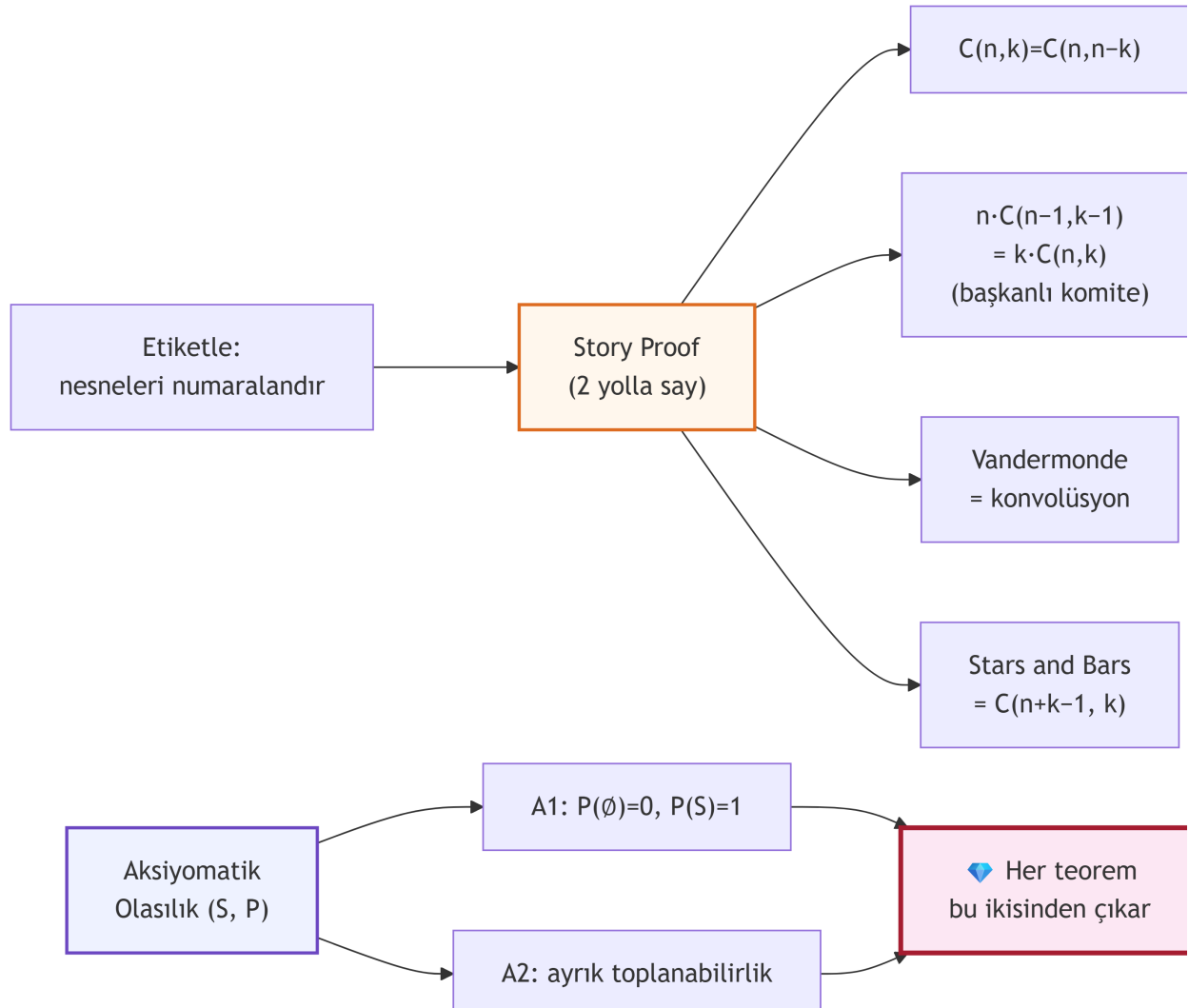
“every single theorem and result in probability eventually follows from these two rules.” — Blitzstein, 45:34

💡 Builder Notu — ML Köprüleri

- **Story proof = “aynı şeyi iki yolla say”** bijektif akıl yürütmedir; parametre sayma, kombinatorik kimlikler ve olasılık türetmelerinde sürekli görülür.
- **Vandermonde özdeşliği bir konvolüsyondur** — bağımsız dağılımların toplamının dağılımı. Ders 22'de FFT ile akraba.
- **Aksiyomlar**, bir modelin çıktısının uymak zorunda olduğu kurallardır: softmax toplamı 1 (= $P(S) = 1$), olasılıklar negatif olamaz, ayrık olaylar toplanır (= marjinalleştirme).
- **Ayırt edilemezlik (exchangeability)** — i.i.d. varsayımının, bag-of-words'ün, set modellerinin (Deep Sets) temeli.

9.2 Etiketleme: Önce Numaralandır

Blitzstein'in en pratik tavsiyesi: bir problemde “ n kişi”, “ n top” veya “ n geyik” geçiyorsa, onları zihninde 1'den n 'e numaralandır (her birine ID ver) — görünüşte özdeş olsalar bile.



Şekil 9.1: Ders 2 yapısı: sayma → story proof → stars and bars; ve eşit-olası varsayımının ötesine geçen aksiyomatik tanım

“assume that they’re labeled with the numbers 1 up to n.” — Blitzstein, 4:56

Neden? Çünkü doğa, toplar sana tıpatıp aynı görünse de onları **ayrı nesnelermiş gibi** sayar. Bir kavanozdaki 10 yeşil top gözüne özdeş gelir; ama olasılık hesabı, sanki üzerlerinde 1–10 yazılıymış gibi doğru sonuç verir. “Bunlar aynı, ayırt edilemez” deyip etiketlemezsen başın derde girer.

“as far as nature is concerned, it behaves as if they are distinguishable and labeled.” — Blitzstein, 6:40

Bu, naif tanım için kritik: eşit-olası sonuçları doğru kurmak ancak nesnelere etiketleyince mümkün olur.

💡 Builder Notu — İndeksleme vs Exchangeability

Etiketleme = veri noktalarını **indeksleme**. Bir batch’teki örnekler ayrı ayrıdır (indeksli), ama bazı yapılar **değiştirilebilir (exchangeable)** — sırası önemsiz, etiketleri atılabilir. “Ayırt edilebilir mi, edilemez mi” sorusu [Ders 1 §sec-ornekleme](#)’deki “sıra önemli mi”nin ta kendisidir ve sequence modeli (RNN, positional transformer) ile set modeli (Deep Sets, pozisyonuz attention) arasındaki seçimi belirler.

9.3 İki Yolla Saymak: Takımlara Bölme

İlk story proof örneği basit bir sayma sorusundan çıkıyor: 10 kişiyi 6 kişilik ve 4 kişilik iki takıma kaç şekilde bölersin?

Yol A: 4 kişilik takımı seç, geri kalan otomatik 6 kişilik takım — $\binom{10}{4}$. **Yol B:** Önce 6 kişilik takımı seç — $\binom{10}{6}$.

Aynı şeyi iki farklı yolla saydık, demek ki eşitler:

$$\binom{10}{4} = \binom{10}{6}$$

Genel hâli, önemli bir özdeşlik:

$$\binom{n}{k} = \binom{n}{n-k}$$

Bir **incelik:** ya iki takımın da 5’er kişi olmasını isteseydik? Cevap $\binom{10}{5}$ **değil**. Çünkü “Takım A / Takım B” ayrımı yok; {1–5} ile {6–10} bölmesi, {6–10} ile {1–5} bölmesi **aynı** bölmedir. Her bölmeyi iki kez saydık, 2’ye böleriz:

$$\frac{1}{2} \binom{10}{5}$$

“there is a clear difference between a team of 4 and a team of 6. While 2 teams of 5 ... it’s equivalent.” — Blitzstein, 10:15

Ders: “sıra önemli mi” demek fazla basit; asıl mesele **neyin gerçekten farklı sayıldığını** düşünmek.

 Builder Notu — Dropout ve Label Switching

$\binom{n}{k} = \binom{n}{n-k}$ doğrudan **dropout simetrisidir**: n nöronun k tanesini söndürmek = $n - k$ tanesini aktif bırakmak. “Etiketsiz iki gruba bölüp 2’ye bölme” hamlesi clustering’deki **label switching** problemiyle aynı: küme etiketleri keyfi olduğundan, aynı bölümlmeyi $k!$ kez sayarsın ve düzeltmen gerekir (mixture modellerin posterior’unda).

9.4 Örneklem Tablosunun Dördüncü Hücre: Stars and Bars

Ders 1’de örneklem tablosunun en zor hücreğini “yerine koyarak, sırasız” $\rightarrow \binom{n+k-1}{k}$ olarak vermiş ama ispatlamamıştık. Blitzstein’in iki tavsiyesini uyguluyoruz: önce basit/uç durumlarla kontrol et, sonra diyagram çiz.

Uç durum kontrolleri:


- $k = 0$: $\binom{n-1}{0} = 1$. Hiçbir şey seçmemenin tek bir yolu. ✓
- $k = 1$: $\binom{n}{1} = n$. Bir kez seçiyorsun; yerine koyma/sıra fark etmez. ✓
- $n = 2$: $\binom{k+1}{k} = k + 1$. İki kutu, k nokta: birinci kutuda kaç nokta var söylemek yeterli ($0, 1, \dots, k \rightarrow k + 1$ seçenek). ✓

Genel ispat (stars and bars). Asıl numara, problemin başka bir probleme **denk** olduğunu görmek: “ n ayırt edilebilir kutuya k ayırt edilemez parçacığı kaç şekilde yerleştirirsin?”

“recognizing when two problems are equivalent even if they sound different. ... that’s a thinking thing.” — Blitzstein, 19:33

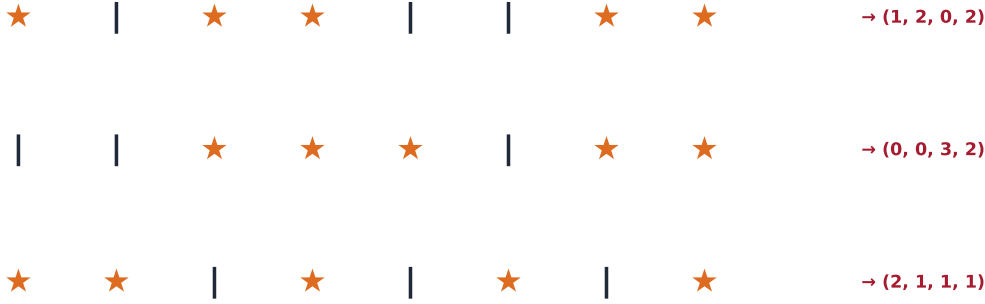
Parçacıkları **nokta** (\cdot), kutu sınırlarını **ayraç** ($|$) olarak kodla. n kutu için $n - 1$ ayraç, artı k nokta var \rightarrow toplam $n + k - 1$ konum. Bir yerleşimi belirlemek için tek yapman gereken, bu konumlardan hangilerinin **nokta** olduğunu seçmek:

$$\binom{n+k-1}{k} = \binom{n+k-1}{n-1}$$

 Builder Notu — Bag-of-Words ve Olasılık Simpleksi

“ n kutuya k ayırt edilemez parçacık” tam olarak bir **sayım vektörü / histogram** saymaktır. Bag-of-words’te V kelimelik sözlükten uzunluğu k olan bir belgenin kelime-sayım vektörü tam bu yapıdır — $\binom{V+k-1}{k}$ olası histogram. Aynı kombinatorik **Dirichlet-multinomial** modellerin temelidir. Stars and bars’ in sürekli karşılığı, softmax çıktılarının yaşadığı **olasılık simpleksidir** ($\sum p_i = 1, p_i \geq 0$).

$$n = 4 \text{ kutu, } k = 5 \text{ parçacık} \rightarrow \binom{4+5-1}{5} = \binom{8}{5} = 56 \text{ yerleşim}$$



Şekil 9.2: Stars and bars: $n = 4$ kutuya $k = 5$ ayırt edilemez parçacığı yerleştirme. Toplam $n + k - 1 = 8$ konumdan $k = 5$ tanesi nokta seçilir. Üç örnek yerleşim: $(1, 2, 0, 2)$, $(0, 0, 3, 2)$, $(2, 1, 1, 1)$.

9.5 Ayırt Edilemezlik ve Bose-Einstein

$n = 2$ durumunun fiziksel yorumu var. İki parayı at: normalde 4 eşit-olası sonuç (HH, HT, TH, TT). Peki paralar sana tıpatıp aynı görünüyorsa?

1920’lerde fizikçi Bose, parçacıklar için (paralar için değil) yalnızca **3 eşit-olası sonuç** olduğunu öne sürdü: HH, TT ve “bir yazı-bir tura” (HT ile TH ayırt edilemiyor). Önce alay edildi; Einstein fikri beğendi → Bose-Einstein istatistiği doğdu ve **Bose-Einstein yoğunlaşması** öngörüldü (70 yıl sonra deneysel gözlemlendi).

Blitzstein’in kritik uyarısı: **paralar için doğru model etiketlidir** (4 sonuç), Bose’unki (3 sonuç) değil. Stars-and-bars *sayma* için doğrudur, ama naif olasılık tanımıyla birlikte kullanma — gerçek nesnelere (zar, kart, top) etiketliymiş gibi davranır.

“you have to be careful about using this with the naive definition of probability.” — Blitzstein, 29:37

💡 Builder Notu — Exchangeability ve de Finetti

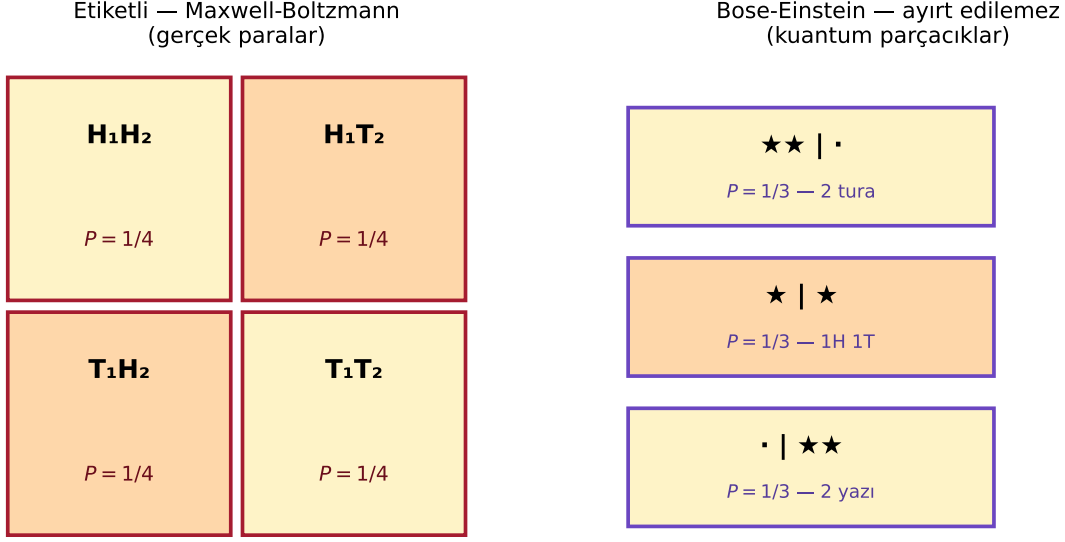
Bose-Einstein durumu tam **değiştirilebilirlik (exchangeability)** demektir — parçacıklar öyle ayırt edilemez ki “Tanrı bile” sırayı söyleyemez. ML’de bu, **de Finetti teoreminin** kalbidir: sonsuz değiştirilebilir bir dizi, gizli bir parametreye koşullandığında i.i.d. olur. Yani “veriyi i.i.d. varsay” cümlesinin teorik temeli exchangeability’dır. Pratikte set modelleri (Deep Sets, set transformer) bu ayırt-edilemezliği mimariye gömer.

9.6 Story Proof Nedir?

Bir **story proof**, bir kombinatorik özdeşliği cebirle değil, **yorumla** ispatlamaktır — “proof by interpretation”. Sembolleri faktöriyel olarak açıp sadeleştirmek yerine, her iki tarafın *neyi saydığını* anlatırsın.

“proof by interpretation ... rather than proof by algebra or calculus.” — Blitzstein, 30:20

9 Story Proof'lar ve Olasılık Aksiyomları



Şekil 9.3: İki para atışında etiketli (Maxwell-Boltzmann) ve ayırt edilemez (Bose-Einstein) modeller. Sol: 4 sonuç eşit-olası, $P(\text{karişık}) = 1/2$. Sağ: 3 sonuç eşit-olası, $P(\text{karişık}) = 1/3$. Gerçek paralar SOLA uyar.

İlk örneği gördük: $\binom{n}{k} = \binom{n}{n-k}$. Cebirle de gösterilebilir ama hikâyesi daha aydınlatıcı: n kişiden k tanesini **seçmek**, aslında dışarıda bırakılacak $n - k$ tanesini seçmekle aynı şeydir. Faktöriyel manipülasyonu sana *neden* doğru olduğunu söylemez; hikâye söyler.

9.7 Story Proof: Başkanlı Komite

Çok kullanışlı bir özdeşlik:

$$n \binom{n-1}{k-1} = k \binom{n}{k}$$

Blitzstein bunu ezberlemediğini, her seferinde hikâyeyle **türettiğini** söylüyor. Hikâye: n kişiden, biri **başkan** olarak atanmış bir k kişilik komite seçmek istiyoruz.

- **Yol A:** Önce komiteyi seç ($\binom{n}{k}$), sonra k üye arasından başkanı seç (k yol) $\rightarrow k \cdot \binom{n}{k}$.
- **Yol B:** Önce başkanı seç (n yol), sonra kalan $n - 1$ kişiden $k - 1$ üyeyi seç ($\binom{n-1}{k-1}$) $\rightarrow n \cdot \binom{n-1}{k-1}$.

İkisi de aynı şeyi (başkanlı komite) sayar, eşittir.

“count the same thing in two ways. ... if both ways are correct, they must agree.” — Blitzstein, 34:46

💡 Builder Notu — Gösterge Hilesi

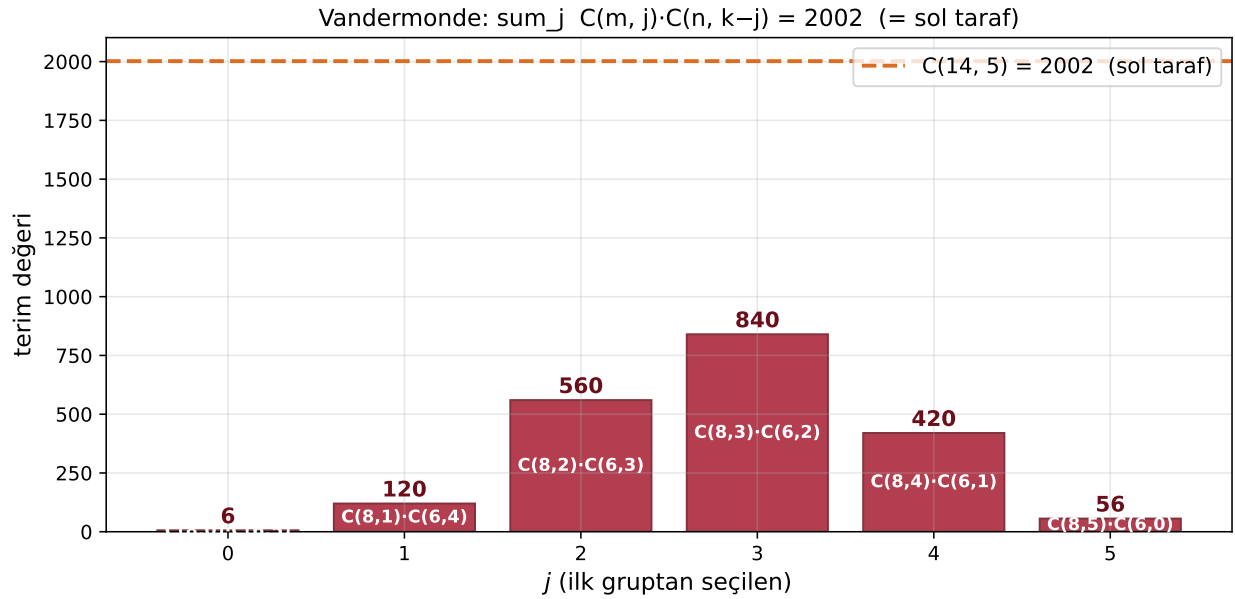
“Özel bir eleman işaretle” (başkanı seç) numarası, beklenen değer hesaplarının kalbindeki **gösterge (indicator) hilesidir** — Ders 9’da göreceğiz. Bir toplama özel bir terim ekleyip simetriden yararlanmak, hipergeometrik ve binom beklentilerini tek satırda çıkarmanı sağlar.

9.8 Story Proof: Vandermonde Özdeşliği

Cebirsel olarak korkunç, hikâyeyle çocuk oyuncağı olan ünlü bir özdeşlik:


$$\binom{m+n}{k} = \sum_{j=0}^k \binom{m}{j} \binom{n}{k-j}$$

Hikâye: $m + n$ kişiyi, biri m kişilik diğeri n kişilik iki gruba ayır. Toplam k kişi seçeceğiz. İlk gruptan j kişi seçersek, ikinci gruptan $k - j$ kişi seçmek zorundayız. j kişiyi ve $k - j$ kişiyi seçmenin yolu $\binom{m}{j} \cdot \binom{n}{k-j}$ (çarpma kuralı). j sıfırdan k 'ye kadar tüm değerleri alabilir ve bu durumlar **ayrık** (örtüşmüyor), o yüzden toplarız.



Şekil 9.4: Vandermonde özdeşliğinin sayısal doğrulaması: $m = 8, n = 6, k = 5$ için sol taraf $\binom{14}{5} = 2002$ ve sağ taraftaki j üzerinde toplam birebir eşleşir.

sol = sağ = 2002 ✓

 Builder Notu — Konvolüsyon ve Bağımsız Toplamlar

Vandermonde bir **konvolüsyondur** — sağ taraf iki dizinin konvolüsyonu. Bu, **bağımsız rastgele değişkenlerin toplamının dağılımına** karşılık gelir: $X \sim \text{Binom}(m, p)$ ve $Y \sim \text{Binom}(n, p)$ bağımsızsa, $X + Y \sim \text{Binom}(m + n, p)$ ve ispatı tam olarak Vandermonde'dur. Konvolüsyon, [Ders 22](#)'de tekrar gelecek; FFT konvolüsyonun ve diffusion'da gürültü eklemenin matematiği aynı yerden çıkar.

9.9 Olasılığın Aksiyomatik Tanımı

Şimdi naif tanımın ötesine geçiyoruz. Artık ne sonuçların eşit-olası olmasını, ne de sonlu sayıda olmasını varsaymak istiyoruz. Bunun için bir **olasılık uzayı** kuruyoruz — iki bileşenli: (S, P) .

- S örnek uzaydır (Ders 1, ama artık sonlu/eşit-olası olmak zorunda değil).
- P bir **fonksiyondur**. Ama alışık olduğun $f(x) = x^2$ gibi değil: girdisi bir **olay** (S 'nin alt-kümesi), çıktısı 0 ile 1 arasında bir sayıdır.

Bir olay A 'nın “**gerçekleştiğini**” şöyle tanımlarız: deney sonunda gözlenen sonuç s_0 , A 'nın içindeyse A gerçekleşmiştir; dışındaysa gerçekleşmemiştir.

P fonksiyonunun uyması gereken **yalnızca iki aksiyom** var:

Aksiyom 1. Boş kümenin olasılığı 0, tüm uzayın olasılığı 1:

$$P(\emptyset) = 0, \quad P(S) = 1$$

Aksiyom 2 (sayılabilir toplanabilirlik). Ayrık (örtüşmeyen) olaylar A_1, A_2, \dots için:

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

(A_n 'ler ayrık olmak şartıyla.)

“*actually we only need two axioms, two rules.*” — Blitzstein, 42:38

Şaşırtıcı: olasılığın her teoremi eninde sonunda bu iki kuraldan türetilir.

 Builder Notu — Geçerli Bir Dağılım Ne Demek?

Bu iki aksiyom, bir **softmax çıktısını geçerli dağılım** yapan tam koşullardır: değerler negatif değil ($P \geq 0$), hepsinin toplamı 1 ($= P(S) = 1$). Sayılabilir toplanabilirlik, **marjinalleştirmenin** (bir değişkeni toplayıp/integralleyip atmanın) temelidir — bir modelde $p(x) = \sum_y p(x, y)$ yazabilmenin nedeni. Ölçü teorisi bu aksiyomları sürekli uzaylara genelleştirir (Ders 12+). Bir sinir ağının olasılıksal başlığı bu aksiyomlara uymazsa, ürettiği şey olasılık dağılımı değildir.

9.10 Bu Dersin Özeti

1. **Etiketleme:** Görünüşte özdeş nesnelere 1 ... n diye numaralandır; doğa onları ayırt edilebilirmiş gibi sayar.
2. **Story proof:** Aynı şeyi iki yolla say — iki sayım doğruysa eşit olmak zorundadır.
3. $\binom{n}{k} = \binom{n}{n-k}$: k seçmek = $n - k$ 'yi dışarıda bırakmak.
4. $n\binom{n-1}{k-1} = k\binom{n}{k}$: başkanlı komite, iki yolla.
5. **Vandermonde:** $\binom{m+n}{k} = \sum_j \binom{m}{j}\binom{n}{k-j}$ — bir konvolüsyon, bağımsız toplamların habercisi.
6. **Stars and bars:** “yerine koyarak, sırasız” = n kutuya k ayırt edilemez parçacık = $\binom{n+k-1}{k}$.
7. **Ayırt edilemezlik:** Bose-Einstein (3 sonuç) sayma ve fizik için geçerli; naif olasılıkta nesnelere etiketli (4 sonuç) davranır.
8. **Olasılık uzayı (S, P) + iki aksiyom:** $P(\emptyset) = 0$, $P(S) = 1$ ve ayrık olaylar için sayılabilir toplanabilirlik.

! Tek bir cümle

“Aynı şeyi iki farklı yoldan say” hem kombinatorik özdeşliklerin en zarif ispatıdır (story proof) hem de olasılığı sayımın ötesine taşıyan **iki aksiyomun** ortak ruhudur: olasılık, tutarlı sayma kurallarından ibarettir.

9.11 Kontrol Soruları

i Soru 1: 4 çeşitli dondurmacıda 6 toplu kâse — sıra önemsiz, tekrar serbest. Kaç farklı kâse?

Cevap: “Yerine koyarak, sırasız”: $n = 4, k = 6$. Stars and bars: $\binom{n+k-1}{k} = \binom{9}{6} = \binom{9}{3} = 84$ kâse.

i Soru 2: Vandermonde özdeşliğini $m = n = 2, k = 2$ için elle doğrula.

Cevap: Sol: $\binom{4}{2} = 6$. Sağ: $\binom{2}{0}\binom{2}{2} + \binom{2}{1}\binom{2}{1} + \binom{2}{2}\binom{2}{0} = 1 + 4 + 1 = 6$. Eşit. ✓

i Soru 3: 10 kişiyi 5-5 takıma bölmek neden $\binom{10}{5}/2$, ama 6-4 bölmek $\binom{10}{4}$?

Cevap: 6’lık ve 4’lük takımlar boyutlarıyla ayırt edilir, her bölme bir kez sayılır. İki 5’erli takım ayırt edilemez ($\{1-5\} \cup \{6-10\} = \{6-10\} \cup \{1-5\}$): $\binom{10}{5}$ bunu iki kez sayar, 2’ye böleriz.

i Soru 4: (Builder) Bir model üç sınıfa [0.5, 0.3, 0.4] atıyor. Hangi aksiyomu ihlal eder ve softmax nasıl düzeltir?

Cevap: Toplam $0,5 + 0,3 + 0,4 = 1,2 \neq 1$, yani $P(S) = 1$ aksiyomu (normalizasyon) ihlal. Softmax tam bunu düzeltir: exp’ler ve toplama böler — çıktı negatif olmaz ve toplam 1.

9.12 Egzersizler

Egzersiz 1. 5 çeşit çaylı dükkânda 3 paketlik set — sıra önemsiz, tekrar serbest. Kaç set mümkün? (Stars and bars.)

Egzersiz 2. Story proof yaz: komite (k kişi) + alt-komite (j kişi) seçmeyi iki yolla sayarak şunu kanıtla:

$$\binom{n}{k} \binom{k}{j} = \binom{n}{j} \binom{n-j}{k-j}$$

Egzersiz 3. “PEPPER” kelimesinin harfleri kaç farklı şekilde dizilebilir? (6 harf, 3 P, 2 E — fazla saymayı düzelt.)

Egzersiz 4. (Python — Vandermonde doğrula)

```
m=11, n= 2, k= 0 → True
m=12, n= 5, k= 7 → True
m= 4, n= 3, k= 1 → True
m=11, n=12, k=17 → True
m= 2, n=10, k= 6 → True
m= 1, n= 1, k= 0 → True
m= 4, n= 4, k= 8 → True
m=10, n= 1, k= 8 → True
m= 4, n=12, k=13 → True
m= 4, n= 8, k= 9 → True
```

Egzersiz 5. (Sonraki ders) Yalnızca iki aksiyomu kullanarak $P(A^c) = 1 - P(A)$ olduğunu göster. (İpucu: A ile A^c ayrıktır ve birleşimleri S 'dir; Aksiyom 2'yi uygula, sonra Aksiyom 1'i.)

9.13 Sonraki Ders İçin Hazırlık

Ders 3: Doğum Günü Problemi, Olasılığın Özellikleri

Aksiyomlardan **pratik özellikler** türetiyoruz: tümleyen kuralı $P(A^c) = 1 - P(A)$, tekdüzelik, içerme-dışarma. Sonra olasılığın en ünlü sezgi-karşıtı sonucu: **doğum günü problemi** — bir odada 23 kişi varsa, ikisinin aynı gün doğmuş olma olasılığı %50'yi geçer.

 Ders 3 öncesi yapılacak

- Egzersizleri çöz — özellikle 5 (tümleyen kuralının aksiyomlardan ispatı).
- Python'da bir “doğum günü” simülasyonu kurmayı dene.
- Ana cümleyi tekrar oku: “Aynı şeyi iki farklı yoldan saymak...”

9.14 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Blitzstein'de
Etiketleme	Özdeş nesnelere 1 ... n diye numaralandır	4m56
$\binom{n}{k} = \binom{n}{n-k}$	k seçmek = $n - k$ 'yı dışarıda bırakmak	9m05
İki eşit takım	$10 \rightarrow 5-5 = \binom{10}{5}/2$ (ayrıt edilemez)	10m03
Stars and bars	n kutuya k parçacık = $\binom{n+k-1}{k}$	26m08
Bose-Einstein	Tam ayrıt edilemezlik; 2 para \rightarrow 3 sonuç (fizik)	28m13
Story proof	Aynı şeyi iki yolla say — yorumla ispat	30m20
Başkanlı komite	$n \binom{n-1}{k-1} = k \binom{n}{k}$	32m17
Vandermonde	$\binom{m+n}{k} = \sum_j \binom{m}{j} \binom{n}{k-j}$	35m06
Olasılık uzayı	(S, P) ; P : olaylar $\rightarrow [0, 1]$	40m13
Aksiyom 1	$P(\emptyset) = 0$ ve $P(S) = 1$	42m50
Aksiyom 2	Ayrık olaylar için sayılabilir toplanabilirlik	44m51

9.15 ML Bağlantıları Özeti

💡 7 köprü

- Etiketleme / ayrıt edilemezlik** \rightarrow i.i.d. ve de Finetti (exchangeability); set modelleri (Deep Sets).
- Stars and bars** \rightarrow bag-of-words sayım vektörü, Dirichlet-multinomial, olasılık simpleksi.
- Story proof** \rightarrow bijektif akıl yürütme; parametre/yol sayma, kombinatorik türetmeler.
- Vandermonde = konvolüsyon** \rightarrow bağımsız toplamlar (Binom+Binom), FFT konvolüsyon, diffusion'da gürültü ekleme.
- Başkanlı komite (işaretleme)** \rightarrow gösterge hilesi, beklenti hesapları (Ders 9).
- İki aksiyom** \rightarrow softmax'ı geçerli dağılım yapar; sayılabilir toplanabilirlik = marjinalleştirme.
- $\binom{n}{k} = \binom{n}{n-k}$ \rightarrow dropout simetrisi (k söndür = $n - k$ aktif).

! Tek bir şey alıp gideceksen

“Aynı şeyi iki farklı yoldan say” — en zarif kombinatorik ispatın (story proof) ve olasılığı sayımın ötesine taşıyan iki aksiyomun ortak ruhudur. **Olasılık, tutarlı sayma kurallarından doğar — ve bu kurallar yalnızca iki tanedir.**

10 Doğum Günü Problemi ve Olasılığın Özellikleri

23 kişide %50,7 — hash collision, içerme-dışarma, $1/e$

i Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 3: Birthday Problem, Properties of Probability](#) (≈49 dk)
- **Okuma süresi:** ≈25 dk

10.1 Bu Derste Ne Var?

Ders 2'de olasılığın iki aksiyomunu kurduk. Burada üç şey:

1. **Doğum günü problemi** — sadece 23 kişide iki kişinin aynı gün doğma olasılığı neden %50'yi geçer.
2. **Aksiyomlardan özellikler** — tümlen $P(A^c) = 1 - P(A)$ ve tekdüzelik $A \subseteq B \Rightarrow P(A) \leq P(B)$.
3. **İçerme-dışarma** — ayrık olmayan olayların birleşim olasılığı; ve ünlü $1/e$ sonucuyla biten matching problemi.

“very surprising to most people the first time that they see this result.” — Blitzstein, 0:29

💡 Builder Notu — ML Köprüleri

- **Doğum günü problemi = hash collision.** Kriptografide **birthday attack** ($2^{n/2}$ denemede bir hash'i kırmak) ve ML'de **feature hashing** aynı matematiğe dayanır.
- $\binom{K}{2}$ **çiftler = çoklu karşılaştırma.** Çok hipotez/çift test edince bazıları şans eseri anlamlı çıkar — **p-hacking** ve **Bonferroni** düzeltmesinin kökü.
- **İçerme-dışarma** → **union bound** $P(\cup A_i) \leq \sum P(A_i)$ — PAC öğrenmenin iş atı.
- $1/e$ rastgele permütasyonlarda, **secretary problem**'in optimal durmasında, **bootstrap OOB**'de karşına çıkar.

10.2 Doğum Günü Problemi: Kurulum

Problem basit: bir partide K kişi var; **en az iki kişinin aynı gün doğmuş** olma olasılığı nedir? (Belirli iki kişi değil — *herhangi* bir çift.)

Varsayımlar:

10 Doğum Günü Problemi ve Olasılığın Özellikleri

- Yılda 365 gün (29 Şubat dışlandı).
- 365 gün **eşit olası**.
- Doğum günleri **bağımsız**.

Kolay durum: $K > 365$ ise olasılık tam 1. Çünkü 365 kutu (gün) ve 365'ten fazla kişi varsa, en az bir kutuda mutlaka birden çok kişi olur — **güvercin yuvası ilkesi (pigeonhole)**.

Asıl ilginç durum $K \leq 365$. Sezgi ne diyor? Blitzstein yıllarca insanlara sormuş: “50/50 şans için kaç kişi gerekir?” Tipik tahmin 150–180. **Gerçek cevap: 23.**

💡 Builder Notu — Hash Collision

Bu problem doğrudan **hash tablosu collision**'dır. Kriptografideki **birthday attack** de buradan gelir — m bit'lik bir hash'te çakışma bulmak $\approx 2^{m/2}$ deneme sürer (kaba kuvvetin 2^m 'inden çok daha az), bu yüzden güvenli hash'ler uzun olmalı. Aynı sezgi, ML'de **feature hashing** kullanırken çakışma sıklığını tahmin eder.

10.3 Doğum Günü: Tümleyenle Hesap

Doğrudan “en az bir eşleşme” olasılığını saymak zor; **tümleyenle** (hiç eşleşme yok) çalışmak çok daha kolay.

“*it's a little bit easier to work with the complement first.*” — Blitzstein, 9:06

Naif tanım. Payda çarpma kuralından: her kişinin 365 olası günü $\rightarrow 365^K$. Pay için kişileri ID sırasıyla düşün: birincinin 365, ikincinin (birinciyle çakışmayan) 364, ..., k . kişinin $365 - k + 1$ seçeneği:

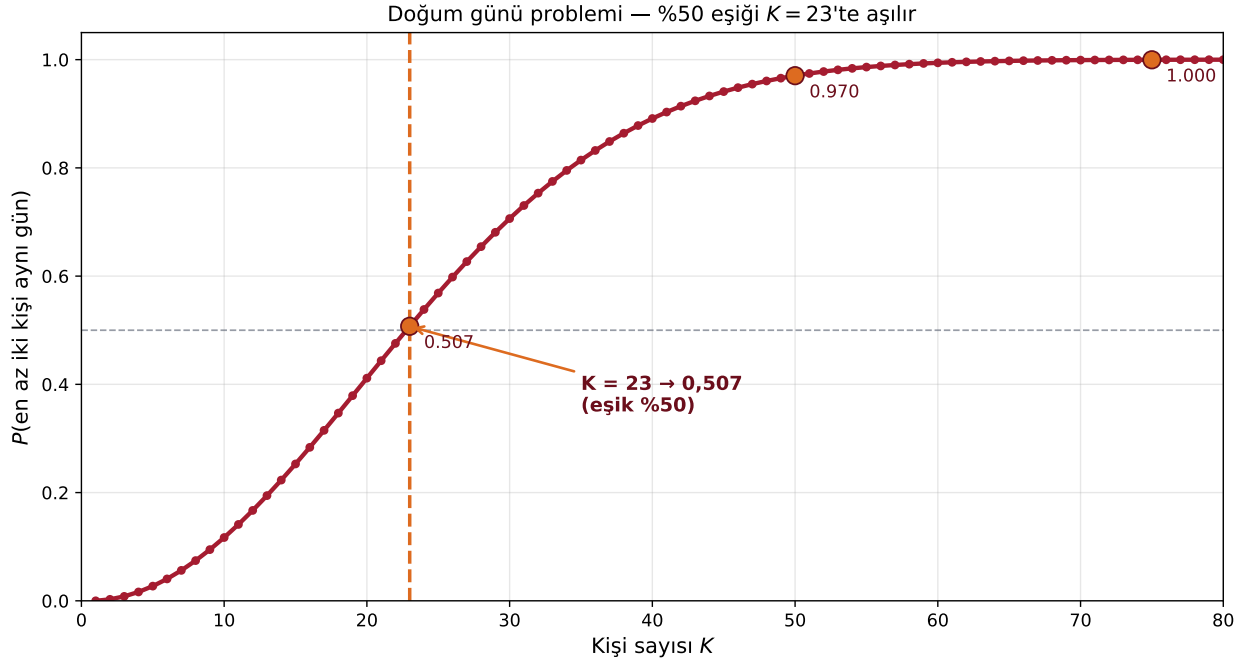
$$P(\text{hiç eşleşme yok}) = \frac{365 \cdot 364 \cdots (365 - K + 1)}{365^K}$$

(Tam K **terim** olmalı; son terimde $+1$ 'i unutmak en sık off-by-one hatasıdır.) Eşleşme olasılığı:

$$P(\text{en az bir eşleşme}) = 1 - \frac{365 \cdot 364 \cdots (365 - K + 1)}{365^K}$$

💡 Builder Notu — ‘En Az Bir’ Refleksi

“ $P(\text{en az bir}) = 1 - P(\text{hiç})$ ” kalıbı ML'de her yerde: **dropout**'ta bir nöronun en az bir adımda aktif kalma olasılığı, n denemede en az bir başarı $(1 - (1 - p)^n)$, bir veri artırma zincirinde en az bir dönüşümün uygulanması. Zor olan “en az bir”i, kolay olan “hiç”in tümleyeniyle hesaplamak temel bir reflektir.



Şekil 10.1: Doğum günü eğrisi. $K = 23 \rightarrow 0,507$ (eşik 50%); $K = 50 \rightarrow 0,97$; $K = 100 \rightarrow 0,99999+$.
Lineer büyümeyen, gözden geçirilemeyecek kadar dik bir eğri.

10.4 Neden 23? Sezgi: $\binom{K}{2}$ Çiftler

23 sayısı neden bu kadar küçük? Çünkü sezgimiz yanlış niceliğe bakıyor. Önemli olan **kaç kişi** (K) değil, **kaç çift** ($\binom{K}{2}$).

$$\binom{K}{2} = \frac{K(K-1)}{2}, \quad \binom{23}{2} = \frac{23 \cdot 22}{2} = 253$$

“The more relevant quantity is not K , but K choose 2.” — Blitzstein, 13:43

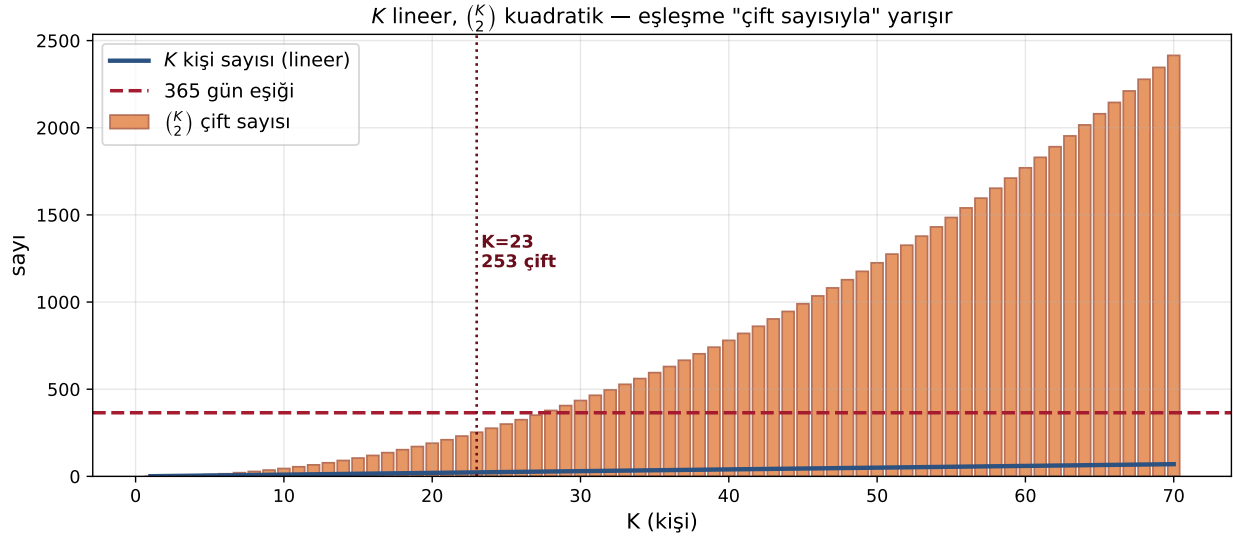
23 kişi az gibi ama aralarında **253 çift** var ve bu çiftlerin *herhangi* biri eşleşebilir. 253, 365’in mertebesine yakın — işte o zaman %50 makul görünüyor.

“the biggest coincidence of all would be if there were no coincidences.” — Blitzstein, 16:19

💡 Builder Notu — Çoklu Karşılaştırma ve P-Hacking

Bu, **çoklu karşılaştırma (multiple comparisons)** probleminin ta kendisi. Çok sayıda hipotez veya çift test edersen, bir kısmı saf şans eseri “anamlı” çıkar — çünkü önemli olan tekil testlerin değil **karşılaştırma sayısının** büyüklüğüdür. **P-hacking** tam budur. **Bonferroni düzeltmesi** anlamlılık eşiğini test sayısına bölerek bunu telafi eder — doğum günü sezgisinin istatistiksel karşılığı.

10 Doğum Günü Problemi ve Olasılığın Özellikleri



Şekil 10.2: Sezgi-karşıtı eksen: K lineer ama $\binom{K}{2}$ ikinci dereceden büyür. $K=23 = 253$ çift, $K=50 = 1225$ çift. Eşleşme olasılığı çift sayısıyla hizalanır.

10.5 Olasılığı Alan Gibi Düşün + Tümleyen Kuralı

Aksiyomları hatırla: $P(\emptyset) = 0$, $P(S) = 1$ ve ayrık olaylar için toplanabilirlik. Bu iki kuralı yazıya dökmek, büyük ölçüde **Kolmogorov**'un eseri.

Sezgi için olasılığı **alan** gibi düşün: Venn diyagramında tüm S 'nin alanı 1'dir; bir olayın olasılığı, kapladığı alandır.

Özellik 1 (tümleyen kuralı):

$$P(A^c) = 1 - P(A)$$

İspat (3 satır): Aksiyom 1'den $1 = P(S)$. $S = A \cup A^c$ ayrık. Aksiyom 2 $\rightarrow P(S) = P(A) + P(A^c)$. Yani $P(A^c) = 1 - P(A)$. ■

💡 Builder Notu — Ölçü ve Normalizasyon

“Olasılık = alan” sezgisi, olasılığın bir **ölçü (measure)** olmasının habercisi (Ders 12+). İspattaki kalıp — “ S 'yi ayrık parçalara böl, toplamları 1” — normalizasyonun kendisidir. Pratikte: $P(\text{hata}) = 1 - P(\text{doğru})$ kalıbı her yerde.

10.6 Tekdüzelik: $A \subseteq B \Rightarrow P(A) \leq P(B)$

Özellik 2 (tekdüzelik / monotonluk): A olayı B 'nin içindeyse (A gerçekleşince B de gerçekleşiyor), A 'nın olasılığı B 'ninkinden büyük olamaz.

$$A \subseteq B \Rightarrow P(A) \leq P(B)$$

İspat: $B = A \cup (B \cap A^c)$, ayrık. Aksiom 2: $P(B) = P(A) + P(B \cap A^c)$. Olasılıklar negatif olamayacağından $P(B) \geq P(A)$. ■

💡 Builder Notu — CDF Monotonluğu

Tekdüzelik bir ölçünün temel özelliğidir. Birikimli dağılım fonksiyonunun (CDF) artan olmasının nedeni budur. Sınır kurarken de kritik: bir hata olayını, onu içeren daha büyük bir olayla üstten sınırlayabilirsin — union bound'un mantığı.

10.7 İçerme-Dışarma (Inclusion-Exclusion)

Ayrık olaylarda $P(A \cup B) = P(A) + P(B)$. Peki ayrık değilse? Venn diyagramında A ile B 'nin alanlarını toplarsan, **kesişimi iki kez** sayarsın; bir kez geri çıkarırsın:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

İspat (disjointification). Aksiom 2 yalnızca ayrık olaylara uygulanır. $A \cup B = A \cup (B \cap A^c)$, ayrık. $P(A \cup B) = P(A) + P(B \cap A^c)$. Ayrıca $B = (A \cap B) \cup (A^c \cap B)$ ayrık, $P(B \cap A^c) = P(B) - P(A \cap B)$. Birleştir, sonuç çıkar. ■

“disjointification” — Blitzstein, 30:11

Üç olaya genelleme (işaretler dönüşümlü):

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

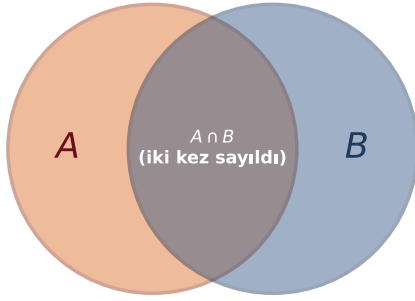
💡 Builder Notu — Union Bound ve PAC

İçerme-dışarmayı ilk terimde keserek **union bound (Boole eşitsizliği)** elde edilir:

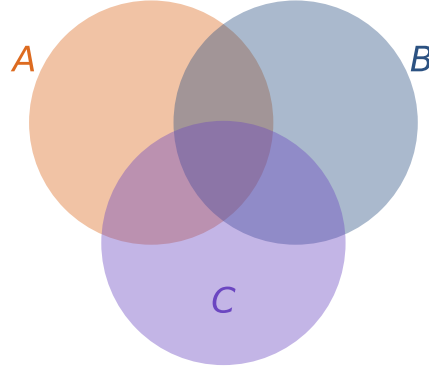
$$P\left(\bigcup_i A_i\right) \leq \sum_i P(A_i)$$

Bu eşitsizlik ML teorisinin iş atıdır: **PAC öğrenme** ve **genelleme sınırları**, çok sayıda “kötü olay”ı union bound ile üstten sınırlar; **Bonferroni** çoklu-test düzeltmesi de aynı sınırdır. İçerme-dışarma kesin değeri verir; union bound her yerde kullanılan ucuz üst sınırdır.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



3 olay: tekleri topla, çiftleri çıkar, üçlüyü ekle



Şekil 10.3: Venn ile içerme-dışarma: A ve B alanlarını toplarsan kesişimi iki kez sayarsın — bir kez çıkar. Üç olay için: tek olayları topla, çiftleri çıkar, üçlüyü ekle (dönüşümlü işaretler).

10.8 Matching Problem (de Montmort, 1713) ve $1/e$

İçerme-dışarmanın güzel uygulaması, 1713'ten kalma kumar oyunu. Deste $1 \dots n$ numaralı n kart. Karıştır, sonra teker teker aç ve $1, 2, 3, \dots$ diye say. Eğer j . açılan kartın üzerinde j yazıyorsa “eşleşme” olur. En az bir eşleşme olasılığı?

$A_j = “j$. karttaki sayı $j”$ olayı. İstedüğümüz $P(A_1 \cup A_2 \cup \dots \cup A_n)$. Simetri her şeyi kolaylaştırıyor:

- $P(A_j) = 1/n$.
- $P(A_1 \cap A_2) = (n-2)!/n! = 1/(n(n-1))$.
- Genel: $P(A_1 \cap \dots \cap A_k) = (n-k)!/n!$.

İçerme-dışarma. Her terimde n 'ler sadeleşir:

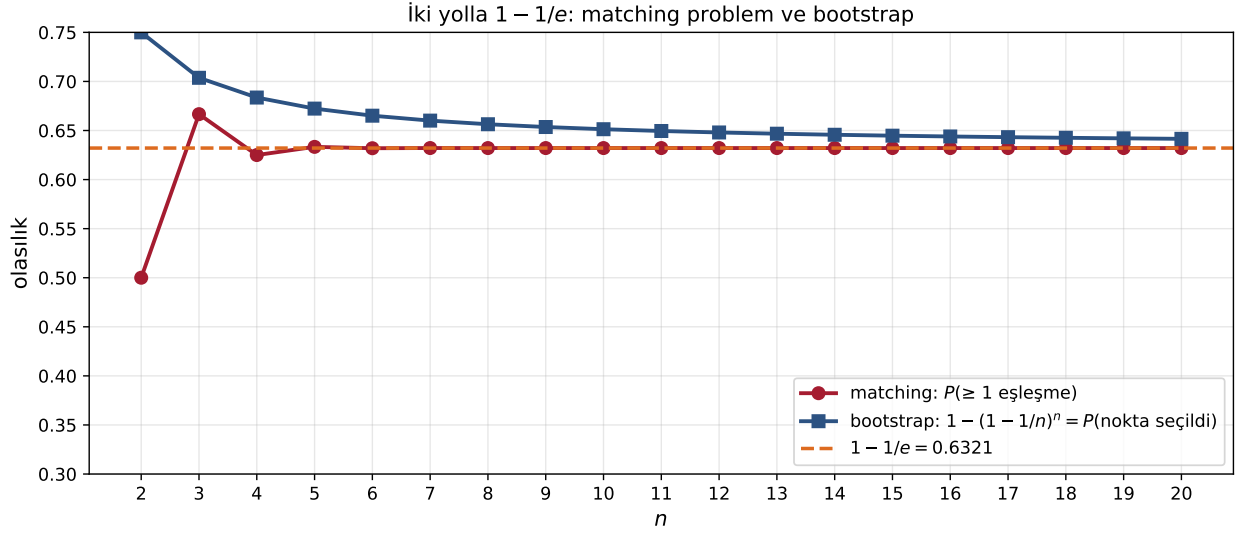
$$P\left(\bigcup_{j=1}^n A_j\right) = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \dots \approx 1 - \frac{1}{e} \approx 0,632$$

“This should remind you of the Taylor series for e to the $x \dots$ this is approximately one minus one over e .” — Blitzstein, 48:43

Şaşırtıcı: en az bir eşleşme olasılığı, n 'e neredeyse **bağımsız** olarak $\approx 0,632$ 'ye oturur.

! Builder Notu — Bootstrap OOB ve $1/e$

$1/e$ ML'de doğrudan **bootstrap**'ta karşına çıkar. n örnekten yerine koyarak n örnek çektiğinde, belirli bir noktanın hiç seçilmeme olasılığı $(1 - 1/n)^n \rightarrow 1/e \approx 0,368$. Yani her bootstrap örneği verinin $\approx \%63,2$ 'sini kullanır, kalan $\approx \%36,8$ **out-of-bag (OOB)** olur — **random forest**'ın OOB hata kestiriminin tam temeli. Aynı $1/e$ **derangement**'larda (hiç sabit noktası olmayan permütasyonlar) ve **secretary problem**'in $1/e$ durma kuralında da belirir.



Şekil 10.4: Matching problem: en az bir eşleşme olasılığı n 'den hızla bağımsızlaşır ve $1 - 1/e \approx 0,632$ 'ye oturur. Aynı $1/e$, bootstrap OOB oranında da çıkar.

10.9 Bu Dersin Özeti

- Doğum günü problemi:** $K = 23 \rightarrow \%50,7$; $K = 50 \rightarrow \%97$; $K = 100 \rightarrow \%99,999+$.
- Tümleyenle hesap:** “en az bir” yerine “hiç yok”u say; $P(\text{eşleşme}) = 1 - 365 \cdot 364 \cdots (365 - K + 1)/365^K$.
- Sezgi** $\binom{K}{2}$: Önemli olan kişi K değil, çift sayısı; 23 kişi \rightarrow 253 çift.
- Tümleyen kuralı:** $P(A^c) = 1 - P(A)$.
- Tekdüzelik:** $A \subseteq B \Rightarrow P(A) \leq P(B)$.
- İçerme-dışarma:** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$; genelde dönüşümlü işaretli toplam. İlk terimde kesince **union bound** $P(\cup A_i) \leq \sum P(A_i)$.
- Matching problem:** En az bir eşleşme $\approx 1 - 1/e \approx 0,632$ (n 'den bağımsız).

! Tek bir cümle

İki basit aksiyomdan tümleyen, tekdüzelik ve içerme-dışarma gibi tüm pratik kurallar türer; olasılığın en şartırtıcı sonuçları (doğum günü, $1/e$) aslında “**kaç çift / kaç eşleşme**” diye doğru niceliği saymaktan ibarettir.

10.10 Kontrol Soruları

i Soru 1: 30 kişilik sınıfta en az iki kişinin aynı gün doğma olasılığı $\%50$ 'den büyük mü?

Cevap: Büyük. 23 kişide 0,507 ve K arttıkça olasılık **tekdüze artar** ($A_{23} \subseteq A_{30}$). Gerçekte $\approx \%70$.

i Soru 2: Adil zar 4 kez atılıyor. En az bir 6 gelme olasılığı?

Cevap: Tümleyen (“hiç 6 yok”): $(5/6)^4 = 625/1296$. $P(\text{en az bir 6}) = 1 - 625/1296 = 671/1296 \approx 0,518$.

i Soru 3: 1-100 arası rastgele tam sayı. 2’ye veya 3’e bölünme olasılığı?

Cevap: İçerme-dışarma. $|A| = 50$ (2’nin katları), $|B| = 33$ (3’ün katları), $|A \cap B| = 16$ (6’nın katları). $P = (50 + 33 - 16)/100 = 67/100 = 0,67$.

i Soru 4: (Builder) Bootstrap’ta n örnekten n çektiğinde belirli bir noktanın hiç seçilmeme olasılığı? Hangi ders kavramıyla bağlantılı?

Cevap: Bir nokta her çekimde seçilmeme olasılığı $(1 - 1/n)$; n bağımsız çekim $\rightarrow (1 - 1/n)^n \rightarrow 1/e \approx 0,368$. Bootstrap verinin $\approx \%63,2$ ’sini kullanır, $\approx \%36,8$ **OOB**. Matching’deki $1 - 1/e$ ile **aynı** e ; random forest’in OOB kestiriminin temeli.

10.11 Egzersizler

Egzersiz 1. “Aynı ay” sürümü: 12 ay, eşit olası, bağımsız. $\%50$ ’yi geçmek için kaç kişi yeter? Tahmin et, sonra formülü 365 yerine 12 ile kur.

Egzersiz 2. 52 kartlık desteden 5 kart. En az bir as gelme olasılığını **tümleyenle** yaz. (İpucu: 48 karttan 5 seçmek = “hiç as yok”.)

Egzersiz 3. $P(A) = P(B) = P(C) = 0,5$; ikili kesişimler 0,3; üçlü 0,1. İçerme-dışarma ile $P(A \cup B \cup C)$?

Egzersiz 4. (Python — doğum günü simülasyon)


$K=23$ simülasyon: 0.5090 (formül beklenen: ~ 0.5073)

Egzersiz 5. (Sonraki ders) Matching problem sonucundan: $n \rightarrow \infty$ iken “hiç eşleşme yok” olasılığı $1/e$ ’ye yaklaşır. Hiç sabit noktası olmayan permütasyonlara **derangement** denir; $n = 4$ için tüm permütasyonları listeleyp derangement sayısını elle doğrula.

10.12 Sonraki Ders İçin Hazırlık

Ders 4: Koşullu Olasılık

Kursun belki de en önemli kavramı: **koşullu olasılık** $P(A | B)$. Blitzstein bunu “istatistiğin ruhu” sayar; öğrenmek, yeni kanıt geldikçe inançları güncellemektir. Bayes teoreminin ve tüm Bayesian çıkarımın kapısı.

 Ders 4 öncesi yapılacak

- Egzersizleri çöz — özellikle 4 (simülasyon) ve 5 (derangement / $1/e$).
- Python'da “en az bir” olasılığını her zaman tümleyenle hesaplamayı alışkanlık edin.
- Ana cümleyi tekrar oku: “İki basit aksiyomdan tüm pratik kurallar türer...”

10.13 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Blitzstein'de
Doğum günü problemi	$K = 23 \rightarrow 0,507$ eşleşme	8m16
Pigeonhole	$K > 365 \rightarrow P = 1$	6m37
Tümleyenle hesap	$P(\text{en az bir}) = 1 - P(\text{hiç})$	9m06
$\binom{K}{2}$ sezgisi	23 kişi = 253 çift	13m43
Tümleyen kuralı	$P(A^c) = 1 - P(A)$	22m50
Tekdüzelik	$A \subseteq B \Rightarrow P(A) \leq P(B)$	25m02
İçerme-dışarma	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$	28m08
Disjointification	Birleşimi ayrık yaz, Aksiyom 2 uygula	30m11
Union bound	$P(\cup A_i) \leq \sum P(A_i)$	türetme
Matching problem	$P(\text{eşleşme}) = 1 - 1/2! + 1/3! - \dots$	39m50
$1/e$ sonucu	$\approx 0,632$; bootstrap OOB, derangement	48m43

10.14 ML Bağlantıları Özeti

 7 köprü

1. **Doğum günü problemi** → hash collision, kriptografide birthday attack ($\approx 2^{m/2}$), ML'de feature hashing.
2. $\binom{K}{2}$ çiftler → çoklu karşılaştırma, p-hacking, **Bonferroni** düzeltmesi.
3. **Tümleyen kuralı** → $P(\text{en az bir}) = 1 - P(\text{hiç})$; dropout, başarı denemeleri.
4. **Tekdüzelik** → ölçü; **CDF**'nin artan olması (Ders 8).
5. **İçerme-dışarma** → **union bound** → **PAC öğrenme** / genelleme sınırları, Bonferroni.
6. $1/e$ (matching) → **bootstrap OOB** $\approx \%36,8$, derangement, secretary problem.
7. **Olasılık = alan / aksiyomlar** → ölçü teorisi, softmax geçerliliği, marjinalleştirme.

! Tek bir şey alıp gideceksen

Olasılığın sürprizleri (doğum günü, $1/e$) sezgiyi değil **doğru niceliği saymayı** test eder — kaç çift, kaç eşleşme. Ve “**en az bir**” sorusunu her zaman tümleyeniyile (“**hiç**”) yenmeye bak.

11 Koşullu Olasılık

İstatistiğin ruhu: bağımsızlık, Bayes, zincir kuralı

i Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 4: Conditional Probability](#) (≈50 dk)
- **Okuma süresi:** ≈28 dk

11.1 Bu Derste Ne Var?

Bu ders kursun kalbi. Önce **bağımsızlığı** tanımlıyoruz, sonra Blitzstein'in "istatistiğin ruhu" dediği kavrama geçiyoruz: **koşullu olasılık** — yeni kanıt geldikçe inançları nasıl güncelleriz.

1. **Bağımsızlık:** $P(A \cap B) = P(A)P(B)$. "Bağımsız = çarp" — ayrıklıkla **karıştırma** (felaket bir hata).
2. **Koşullu olasılık:** $P(A | B) = P(A \cap B) / P(B)$ — öğrenmenin matematiği.
3. **Üç teorem:** çarpım, zincir kuralı, **Bayes** — ispatları tek satır, sonuçları bir bilim dalı.

"conditioning is the soul of statistics." — Blitzstein, 33:13

💡 Builder Notu — ML Köprüleri

- **Koşullu olasılık = öğrenme.** Bir modelin yaptığı her şey koşulludur: sınıflandırıcı $p(y | x)$, dil modeli $p(\text{token} | \text{bağlam})$, Bayesian posterior $p(\theta | \text{veri})$.
- **Zincir kuralı** → **autoregressive dil modelinin** matematiği. $n!$ farklı sıralama → any-order / XLNet.
- **Bağımsız = çarp** → **naive Bayes**'in çekirdeği.
- **İkili ≠ karşılıklı bağımsız (XOR)** → tek katmanlı perceptron neden XOR öğrenemez.
- **Çakıl dünyası (yeniden normalleştirme)** → **masked / causal softmax attention**.
- **Bayes kuralı** → tüm Bayesian çıkarımın, generative ↔ discriminative ilişkisinin ve **diffusion ters sürecinin** temeli.

11.2 Bağımsızlık: Tanım ve Ayrıklıkla Karıştırma

A ve B **bağımsızdır** ancak ve ancak:

$$P(A \cap B) = P(A) P(B)$$

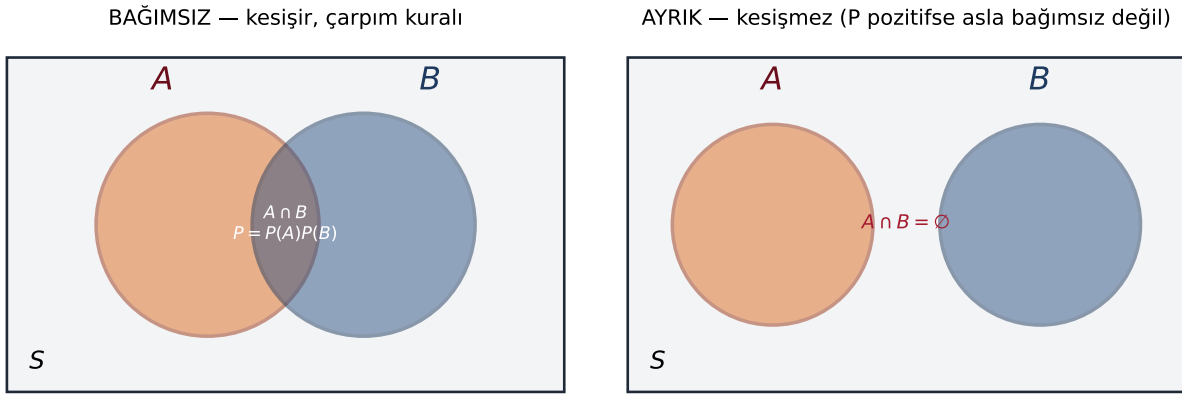
Kısa slogan: **bağımsız = çarp.**

KRİTİK UYARI: bağımsızlık ile **ayrıklık (disjoint)** bambaşka şeylerdir.

“this is completely different from disjointness ... a terrible, disastrous blunder.” — Blitzstein, 12:06

- **Ayrık:** A olursa B **olamaz** ($P(A \cap B) = 0$). Biri diğeri hakkında çok şey söyler.
- **Bağımsız:** A olması B hakkında **hiçbir şey** söylemez.

(Olasılıkları pozitif) ayrık olaylar asla bağımsız olamaz.



Şekil 11.1: Bağımsız vs ayrık olayların Venn'i. Sol: bağımsız ($P(A \cap B) = P(A)P(B) > 0$). Sağ: ayrık ($P(A \cap B) = 0$, kesişim boş). İki tamamen farklı kavram.

💡 Builder Notu — Factorization

“Bağımsız = çarp” kuralı, olasılıksal modellerde **çarpanlara ayırmanın (factorization)** temelidir: bağımsız gürültü bileşenleri, köşegen kovaryanslı VAE prior'ı, naive Bayes'teki $p(\text{özellikler} \mid \text{sınıf}) = \prod p(\text{özellik}_i \mid \text{sınıf})$. Bağımsızlık sayısız parametreyi birkaç çarpana indirir.

11.3 Çoklu Bağımsızlık: İkili \neq Karşılıklı

Üç olay A, B, C için bağımsızlık ne demek? Önce **her ikisi** bağımsız olmalı. Ama bu yetmez. Bir denklem daha gerekir:

$$P(A \cap B \cap C) = P(A) P(B) P(C)$$

Yalnızca ikili denklemler sağlanıyorsa **ikili (pairwise) bağımsızlık** denir — ve ikili bağımsızlık **karşılıklı (mutual) bağımsızlığı** gerektirmez.

! Builder Notu — XOR ve Gizli Katman

İkili-ama-karşılıklı-değil'in en güzel örneği **XOR**. A ve B bağımsız adil bit'ler, $C = A \oplus B$. Her ikili bağımsızdır (C 'yi tek başına A veya B belirlemez), ama üçü karşılıklı bağımsız değildir: A ile B 'yi **birlikte** bilince C tamamen belli. Bu, tek katmanlı perceptron'un XOR'u neden öğrenemediğinin (**Minsky-Papert**) olasılıksal gölgesidir: ikili/marjinal istatistikler bilgisizdir, etiketi birlikte belirlerler — gizli katman gerekir.

11.4 Newton-Pepys Problemi (1693) + Binom Olasılığı

1693'te Samuel Pepys, Newton'a bir zar sorusu yazdı:

- **A:** 6 zarda **en az bir** 6.
- **B:** 12 zarda **en az iki** 6.
- **C:** 18 zarda **en az üç** 6.

Çoğu kişi C der. Doğru cevap: A.

$$P(A) = 1 - \left(\frac{5}{6}\right)^6 \approx 0,665$$

“Tam k tane 6” olasılığı — **binom olasılığı**:

$$P_k = \binom{n}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{n-k}$$

Tümleyenle: $P(B) \approx 0,619$; $P(C) \approx 0,597$.

“in 1693 ... you had to go to a genius, like Isaac Newton. Whereas now, all you have to do is take STAT 110.” — Blitzstein, 28:03

💡 Builder Notu — Stigler Sanity Check

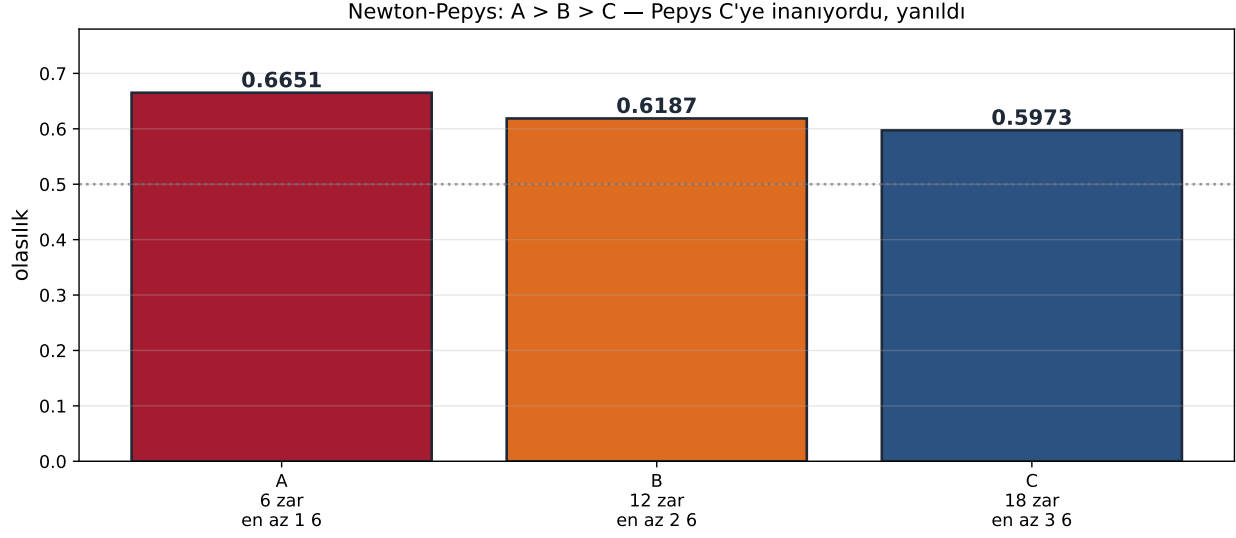
İstatistikçi Stigler'in argümanı: bir yöntem, sonucun bağlı olduğu bir şeyi (zar olasılıkları) **görmezden geliyorsa**, bozuktur. Sanity check refleksi tam budur — önemli bir girdiyi **değişmez (invariant)** kalan bir model, o girdiyi yakalamıyordu.

11.5 Koşullu Olasılık: İstatistiğin Ruhu

Asıl soru: elinde inançlar var ve yeni şeyler öğreniyorsun. **Yeni kanıt geldikçe inançlarını nasıl güncelleyelisin?**

“Everything relates to conditioning in one way or another.” — Blitzstein, 33:34

11 Koşullu Olasılık



Şekil 11.2: Newton-Pepys: $P(A) > P(B) > P(C)$. Pepys C'ye inanıyordu ama yanılıyordu — “en az bir” hep en olası, zar sayısı eşik oranını korusa da.

Tanım:

$$P(A | B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) > 0$$

Eğer A ile B bağımsızsa $P(A | B) = P(A)$ — yani B 'yi öğrenmek A hakkında hiçbir şey değiştirmez.

💡 Builder Notu — Öğrenmenin Adı

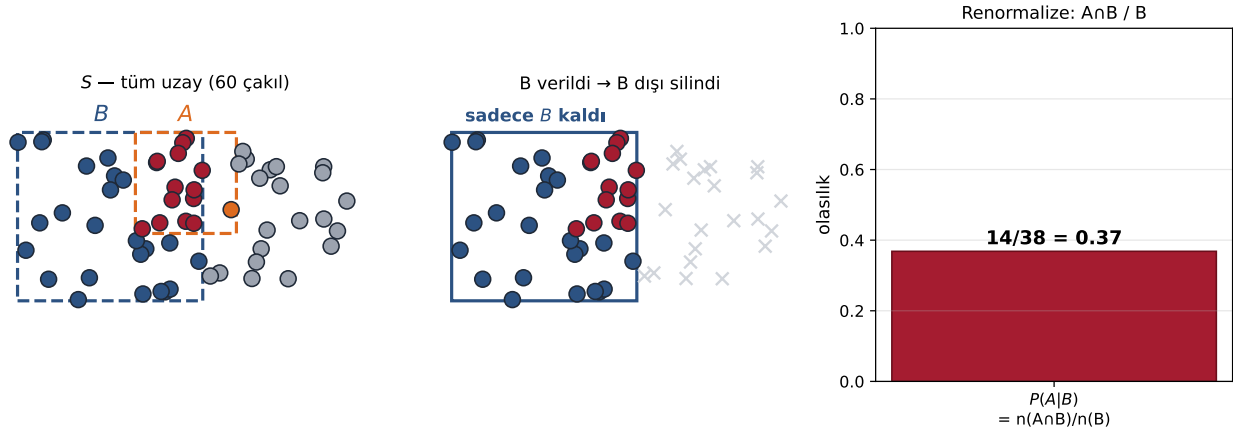
Koşullu olasılık ML'in yaptığı işin **tanımıdır**. Discriminative sınıflandırıcı doğrudan $p(y | x)$ öğrenir. Bayesian çıkarımda posterior $p(\theta | \text{veri})$, prior'ın kanıtlarla güncellenmesidir. Attention, bir sorguya koşullu ağırlık dağılımıdır.

11.6 İki Sezgi: Çakıl Dünyası ve Frekansçı Dünya

Sezgi 1 — Çakıl Dünyası. Örnek uzay S 'yi toplam kütlesi 1 olan çakıllar gibi düşün; her çakıl bir sonuç, kütlesi olasılığı. “ B verildi” demek: B 'nin dışındaki çakıllar **ilgisiz** — onları sil. Kalanların kütlesi $P(B)$, toplam 1 olsun diye **yeniden normalleştir**.

“renormalize to make total mass 1 again.” — Blitzstein, 40:33

Sezgi 2 — Frekansçı Dünya. Deneyi çok kez tekrarla. B 'nin gerçekleştiği tekrarları işaretle, sonra **yalnızca onların** içinde A 'nin oranını al.



Şekil 11.3: Çakıl dünyası: B verildiğinde dışarıları silinir, kalan çakılların kütleleri $P(B)$. $P(B)$ 'ye bölmek toplam kütleleri tekrar 1 yapar $\rightarrow P(A|B) = P(A \cap B) / P(B)$.

💡 Builder Notu — Masked Softmax

Çakıl dünyasının yeniden normalleştirilmesi doğrudan **masked softmax**'tır: izin verilmeyen seçenekleri maskele, kalanları toplamı 1 olacak şekilde normalleştir — bu birebir koşullamadır (causal attention, geçersiz token maskeleme). Frekansçı dünya ise **Monte Carlo kestirimi**: B sağlanan örnekleri süz, içlerinde A 'nın oranını hesapla.

11.7 Üç Teorem: Çarpım, Zincir Kuralı, Bayes

Koşullu olasılığın tanımını yeniden düzenleyerek üç teoremi dakikalar içinde çıkarıyoruz.

Teorem 1 (çarpım kuralı):

$$P(A \cap B) = P(B) P(A | B) = P(A) P(B | A)$$

Teorem 2 (zincir kuralı):

$$P(A_1 \cap \dots \cap A_n) = P(A_1) P(A_2 | A_1) P(A_3 | A_1 \cap A_2) \dots P(A_n | A_1 \cap \dots \cap A_{n-1})$$

Bunu $n!$ farklı sırayla yazabilirsin.

“this is not just one theorem, but n factorial theorems.” — Blitzstein, 47:47

Teorem 3 (Bayes kuralı):

$$P(A | B) = \frac{P(B | A) P(A)}{P(B)}$$

İspat tek satır (sadece böldük), ama bu eşitlik koca bir bilim dalının temeli.

“The proof is completely obvious. ... The implications are extremely deep.” — Blitzstein, 49:22

! Builder Notu — Autoregressive ve Diffusion

Zincir kuralı = autoregressive çarpanlama: dil modeli $p(w_1, \dots, w_n) = \prod p(w_i | w_{<i}). n!$ sıralama özgürlüğü → any-order üretim (**XLNet**, diffusion dil modelleri). **Bayes kuralı** = posterior \propto olabilirlik \times prior; bir generative sınıflandırıcı $p(x | y)p(y)$ modelleyip Bayes ile $p(y | x)$ 'e çevirir, ve **diffusion**'ın ters adımı bir Bayes güncellemesidir.

11.8 Bu Dersin Özeti

1. **Bağımsızlık:** $P(A \cap B) = P(A)P(B)$. “Bağımsız = çarp.”
2. **Bağımsızlık \neq ayrıklık:** Karıştırma felakettir.
3. **İkili \neq karşılıklı bağımsızlık:** Üçlü denklem ayrı gerekir (XOR).
4. **Newton-Pepys:** 6 zarda en az bir 6 ($\approx 0,665$) > 12’de en az iki ($\approx 0,619$) > 18’de en az üç ($\approx 0,597$).
5. **Koşullu olasılık:** $P(A | B) = P(A \cap B)/P(B)$.
6. **İki sezgi:** çakıl dünyası (kısıtla + normalleştir) ve frekansçı dünya.
7. **Üç teorem:** çarpım, zincir, Bayes.

! Tek bir cümle

Koşullu olasılık $P(A | B) = P(A \cap B)/P(B)$, yeni kanıtla **inançları güncellemenin matematiğidir** — “istatistiğin ruhu” — ve ondan tek satırlık cebirle çıkan Bayes kuralı, koca bir çıkarım biliminin temelini atar.

11.9 Kontrol Soruları

i Soru 1: İki adil zar. ‘İlk zar çift’ ile ‘toplam 7’ bağımsız mı?

Cevap: $P(\text{ilk çift}) = 1/2$. $P(\text{toplam 7}) = 6/36 = 1/6$. Kesişim (ilk çift, toplam 7): (2, 5), (4, 3), (6, 1) → $3/36 = 1/12$. Çarpım: $(1/2)(1/6) = 1/12$. Eşit → **bağımsız**. Sezgiye aykırı ama doğru.

i Soru 2: $P(A)=P(B)=0.5$ ve A, B ayrık. Bağımsızlar mı?

Cevap: Hayır. Ayrık → $P(A \cap B) = 0$. Bağımsız olsalardı $P(A)P(B) = 0,25$. $0 \neq 0,25$ → bağımsız değil. Pozitif olasılıklı ayrık olaylar asla bağımsız olamaz.

i Soru 3: Hastalık %1, test hassasiyeti %99, yanlış pozitif %5. Pozitif çıkan gerçekten hasta mı?

Cevap: Bayes. $P(+)=0,99 \cdot 0,01 + 0,05 \cdot 0,99 = 0,0594$. $P(\text{hasta} | +) = 0,99 \cdot 0,01 / 0,0594 \approx 0,167 = \%16,7$. Test iyi gibi ama düşük base rate yüzünden pozitiflerin çoğu **yanlış pozitif**. Ders 1’in base-rate neglect’inin Bayes hesabı.

i Soru 4: (Builder) ‘kara kedi’ dizisi nasıl çarpanlanır ve neden autoregressive üretim mümkün?

Cevap: $p(\text{kara, kedi}) = p(\text{kara}) \cdot p(\text{kedi} \mid \text{kara})$. Genelde $p(w_1, \dots, w_n) = \prod p(w_i \mid w_{<i})$. Model her adımda yalnızca bir sonraki token’ın koşullu dağılımını üretir; zincir kuralı tüm dizinin olasılığını yerel koşullulardan kurar.

11.10 Egzersizler

Egzersiz 1. 52 karttan tek kart. “Kupa” ile “Kız (Q)” bağımsız mı? Tanımla doğrula.

Egzersiz 2. Torba: 5 kırmızı, 3 mavi. Yerine koymadan 2 top. (a) İlk kırmızı verildiğinde ikincinin mavi olma olasılığı? (b) İkincinin mavi olma (koşulsuz) olasılığı?

Egzersiz 3. Üçü ikili bağımsız ama karşılıklı bağımsız **olmayan** üç olay kur. (İpucu: XOR.)

Egzersiz 4. (*Python — Newton-Pepys Monte Carlo*)

A: $P = 0.6648$ (n=6, ≥ 1)

B: $P = 0.6162$ (n=12, ≥ 2)

C: $P = 0.5978$ (n=18, ≥ 3)

Egzersiz 5. (*Sonraki ders*) Soru 3’teki hastalığı, yanlış pozitif %1 olsaydı yeniden hesapla. Paydadaki $P(+)=P(+|H)P(H)+P(+|S)P(S)$ ifadesi **toplam olasılık yasası (LOTP)** — Ders 5’in ana konusu.

11.11 Sonraki Ders İçin Hazırlık

Ders 5: Koşullamaya Devam, Toplam Olasılık Yasası

Toplam olasılık yasası (LOTP) ile bir olayın olasılığını, ayrık durumlara koşullayıp birleştirerek hesaplamayı öğreneceğiz. Bayes paydasının kendisi.

⚠ Ders 5 öncesi yapılacak

- Egzersizleri çöz — özellikle 4 (Monte Carlo) ve 5 (LOTP keşfi).
- “En az bir” refleksini pekiştir.
- Ana cümleyi tekrar oku: “*Koşullu olasılık, inançları güncellemenin matematiğidir.*”

11.12 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Blitzstein’de
Bağımsızlık	$P(A \cap B) = P(A)P(B)$	11m25

Kavram	Tanım	Blitzstein'de
Bağımsız ≠ ayrık	Zıt kavramlar	12m06
İkili ≠ karşılıklı	Üçlü denklem ayrı (XOR)	14m09
Newton-Pepys	$A \approx 0,665 > B \approx 0,619 > C \approx 0,597$	28m03
Binom olasılığı	$\binom{n}{k}(1/6)^k(5/6)^{n-k}$	26m45
Koşullu olasılık	$P(A B) = P(A \cap B)/P(B)$	34m08
Çakıl dünyası	Kısıtla + $P(B)$ 'ye böl	36m11
Frekansçı dünya	Uzun vadede oran	41m25
Çarpım kuralı	$P(A \cap B) = P(B)P(A B)$	45m00
Zincir kuralı	$\prod P(A_i \text{öncekiler})$	47m12
Bayes kuralı	$P(A B) = P(B A)P(A)/P(B)$	49m02

11.13 ML Bağlantıları Özeti

💡 7 köprü

1. **Koşullu olasılık** → tüm öğrenme: $p(y | x)$, posterior, attention.
2. **Zincir kuralı** → autoregressive dil modeli; $n!$ sıra → any-order (XLNet, diffusion LM).
3. **Bayes kuralı** → posterior \propto olabilirlik \times prior; generative \leftrightarrow discriminative.
4. **Bağımsız = çarp** → naive Bayes, faktörizasyon.
5. **İkili ≠ karşılıklı (XOR)** → gizli katmanın gerekliliği.
6. **Çakıl yeniden normalleştirme** → masked / causal softmax attention.
7. **Binom olasılığı** → Binom dağılımı (Ders 8), sayım ve sınıflandırma.

❗ Tek bir şey alıp gideceksen

Koşullu olasılık $P(A | B) = P(A \cap B)/P(B)$, öğrenmenin matematiğidir. Tek satırlık Bayes kuralı bu tanımdan doğar. “**Conditioning is the soul of statistics.**”

12 Koşullamaya Devam ve Toplam Olasılık Yasası

LOTP, hastalık testi, savcı yanlılığı, explaining away

Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 5: Conditioning Continued, Law of Total Probability](#) (≈50 dk)
- **Okuma süresi:** ≈28 dk

12.1 Bu Derste Ne Var?

Ders 4'te koşullu olasılığı ve Bayes'i kurduk. Burada onu **kullanıyoruz**:

1. **Toplam olasılık yasası (LOTP):** karmaşık $P(B)$ 'yi ayrık durumlara bölüp birleştir.
2. **Bayes + LOTP:** hastalık testi — %95 “doğru” test, ama pozitifin sadece %16'sı gerçekten hasta.
3. **Biyohazardlar:** $P(A | B)$ ile $P(B | A)$ (savcı yanlılığı), prior ile posterior, bağımsızlık ile koşullu bağımsızlık.

“you can't really think clearly except under the condition that you understand how to think conditionally.” — Blitzstein, 1:33

Builder Notu — ML Köprüleri

- **LOTP = marjinalleştirme:** $p(x) = \sum_z p(x | z)p(z)$. Latent değişkenli modellerin (mixture, VAE, EM, HMM) temeli.
- **Hastalık testi = precision vs recall.** Dengesiz veride accuracy neden yanıltır.
- **Savcı yanlılığı** = $p(\text{veri} | \text{model})$ ile $p(\text{model} | \text{veri})$ 'yi karıştırmak.
- **Koşullu bağımsızlık** = Bayesian ağların ve naive Bayes'in çekirdeği; **explaining away** = collider yapısı.

12.2 Toplam Olasılık Yasası (LOTP)

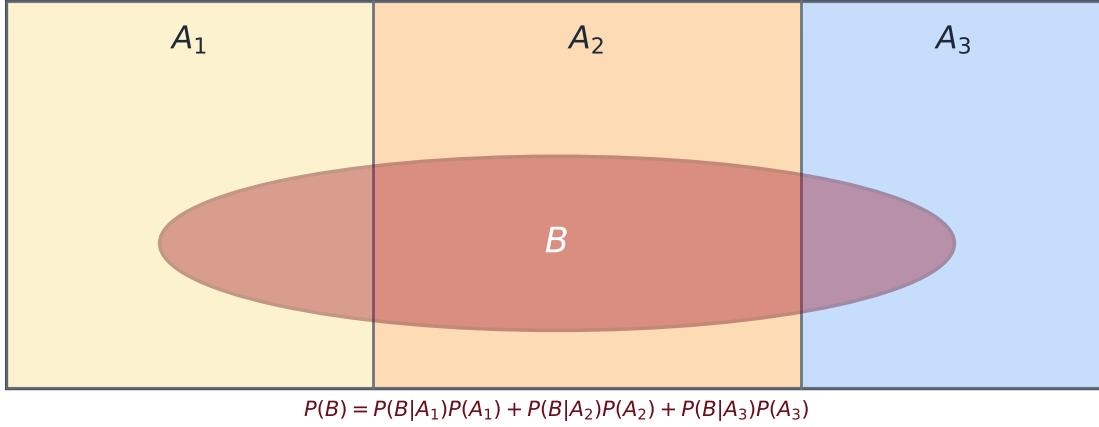
Örnek uzay S 'yi **böler (partition)** ayır: A_1, A_2, \dots, A_n hem **ayrık** hem de **birleşimleri** S . Bir B olayını her parça keser:

$$P(B) = \sum_{i=1}^n P(B \cap A_i) = \sum_{i=1}^n P(B | A_i) P(A_i)$$

“I prefer to just think of it as breaking up a problem into simpler pieces.” — Blitzstein, 7:28

Kritik nokta: yararı, **partition’ı ne kadar iyi seçtiğine** bağlı. Kötü bölme n eşit zor problem verir; iyi bölme her parçayı kolaylaştırır.

LOTP: B 'yi partition $\{A_1, A_2, A_3\}$ ile parçalara böl, koşullu olasılıkları ağırlıkla topla



Şekil 12.1: LOTP geometrisi: S 'yi A_1, A_2, A_3 'e böl (ayrık + birleşim S). B olayını her parça keser; $P(B) = \sum P(B|A_i)P(A_i)$.

💡 Builder Notu — Marjinalleştirme ve Latent Modeller

LOTP, olasılıksal modellerde **marjinalleştirmenin** ta kendisi: $p(x) = \sum_z p(x | z)p(z)$. Sınıflandırıcının marjinal tahmini, mixture'da bileşenleri toplama, HMM'de gizli durumları toplama — hepsi LOTP. “İyi partition seç” tavsiyesi de gizli değişkeni doğru tasarlamaya denk düşer.

12.3 İki Kart Paradoksu: “Bir As” vs “Maça Ası”

52 kartlık desteden 2 kart. İki koşullu olasılığı karşılaştır:

(a) **İki as, en az bir as verildiğinde:**

$$P(\text{iki as} | \text{en az bir as}) = \frac{\binom{4}{2}}{\binom{52}{2} - \binom{48}{2}} = \frac{6}{198} = \frac{1}{33}$$

(b) **İki as, maça ası verildiğinde:** Bir kart maça ası, diğer kart 51 kartın eşit olası biri. Bunların 3'ü as:

$$P(\text{iki as} \mid \text{maça ası}) = \frac{3}{51} = \frac{1}{17}$$

Şaşırtıcı: **hangi as** olduğunu söylemek olasılığı neredeyse **iki katına** çıkardı. “Bir as” = varoluşsal (belirsiz); “maça ası” = belirli bir kartı sabitlemek.

💡 Builder Notu — Kanıtın Belirliliği

Kanıtın **belirliliği** posterior’u kökten değiştirir. “En az bir” tipi (varoluşsal) ile “belirli bir öge” tipi farklı sonuç verir — **selection effect / gözlem çerçevesi sorunudur**. Bir modele kanıt verirken “belirli mi, varoluşsal mı” ayrımı çıkarılan sonucu değiştirir.

12.4 Hastalık Testi: Bayes + LOTP El Ele

Hastalık nüfusun %1’inde. Test %95 doğru:

$$P(T \mid D) = 0,95 = P(T^c \mid D^c)$$

Hastanın merak ettiği $P(D \mid T)$. Bayes + LOTP:

$$P(D \mid T) = \frac{P(T \mid D)P(D)}{P(T \mid D)P(D) + P(T \mid D^c)P(D^c)} = \frac{0,95 \cdot 0,01}{0,95 \cdot 0,01 + 0,05 \cdot 0,99} \approx 0,16$$

%95 doğru test pozitif verdiğiğinde bile, hasta olma olasılığı sadece **%16**.

“there’s only a 16% chance that the patient has the disease.” — Blitzstein, 26:15

! Builder Notu — Precision vs Recall

Bu örnek **precision vs recall**’un ta kendisi. $P(T \mid D)$ = recall/sensitivity; $P(D \mid T)$ = precision. Test “%95 accurate” görünse de, dengesiz veride (%1 pozitif) precision çöker — **accuracy’nin neden yanıltıcı bir metrik** olduğunun kanıtı. Aynı hesap, bir dolandırıcılık dedektörünün alarmlarının çoğunun neden yanlış pozitif olduğunu açıklar; çözüm base rate’i hesaba katmak ve **kalibrasyon**.

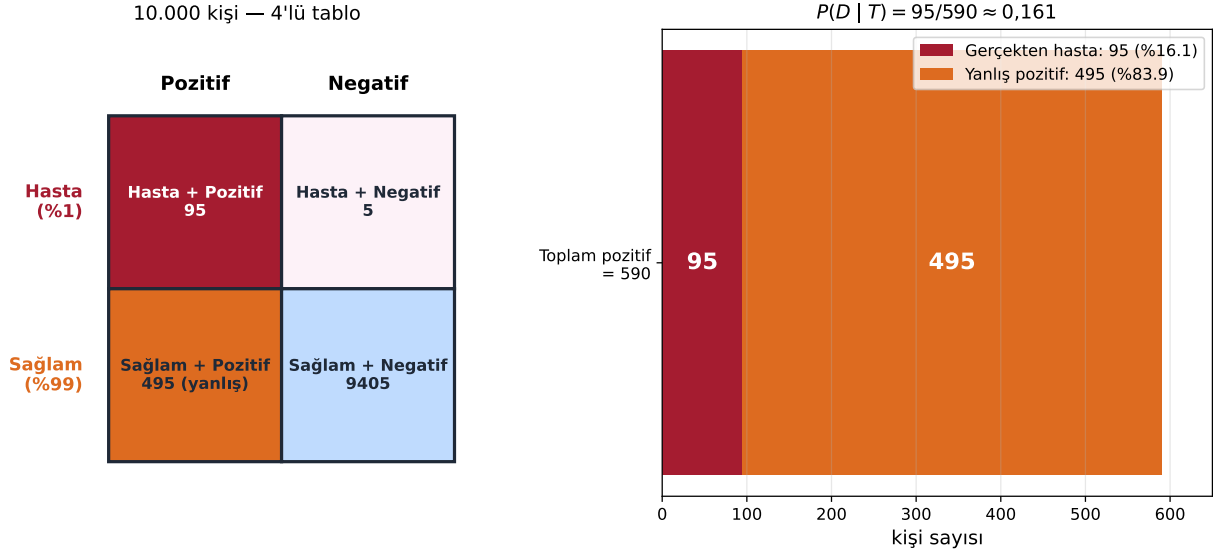
12.5 Biyohazard 1: $P(A|B) \neq P(B|A)$ — Savcı Yanılgısı

Blitzstein bu yaygın hataları “biyohazard” diye adlandırıyor. Birincisi: $P(A \mid B)$ ile $P(B \mid A)$ ’yı karıştırmak. Hukukta buna **savcı yanılgısı** denir.

Sally Clark davası. İki bebeği SIDS’le ölen kadın, “uzman” tanığa göre $(1/8500)^2 \approx 1/73$ milyon olasılıkla “masum”. İki hata:

- **Bağımsızlık varsayımı**: iki ölüm bağımsız sayıldı — ortak genetik faktör olabilir.

12 Koşullamaya Devam ve Toplam Olasılık Yasası



Şekil 12.2: Hastalık testi frekansçı sezgi: 10.000 kişide 100 hasta (95'i pozitif) ve 9900 sağlam (495'i yanlış pozitif). Toplam 590 pozitif → $95/590 \approx \%16,1$ gerçekten hasta.

- **Savcı yanılması:** $1/73$ milyon = $P(\text{kanıt} | \text{masum})$, ama önemli olan $P(\text{masum} | \text{kanıt})$. Bayes ile çevirince paya $P(\text{masum})$ girer; çoğu anne masum olduğundan bu prior ≈ 1 'e yakın, hesap kökten değişir.

Clark yıllarca hapis yattı; karar sonradan bozuldu.

! Builder Notu — Likelihood \neq Posterior

Savcı yanılması = **olabilirlik** $p(\text{kanıt} | \text{hipotez})$ ile **posterior** $p(\text{hipotez} | \text{kanıt})$ 'u karıştırmak. Bir modelde “ $p(\text{veri} | \text{null})$ çok düşük” demek “ $p(\text{null} | \text{veri})$ çok düşük” demek **değildir** — prior olmadan posterior'a geçemezsin. P-değeri yanlış-yorumlarının kökü budur.

12.6 Biyohazard 2 ve 3: Prior \leftrightarrow Posterior ve Koşullu Bağımsızlık

Biyohazard 2: prior ile posterior'u karıştırmak. $P(A)$ prior (kanıttan önce); $P(A | B)$ posterior. Sık hata: “A gerçekleşti deniyor, o hâlde $P(A) = 1$.” Yanlış. Doğrusu:

$$P(A | A) = 1$$

Verilen bilgi koşul çubuğunun **sağına** gider, $P(A)$ 'yı 1 yapmaz.

Biyohazard 3: koşullu bağımsızlığı bağımsızlıkla karıştırmak. A ve B , C verildiğinde koşullu bağımsızdır:

$$P(A \cap B | C) = P(A | C) P(B | C)$$

Bağımsızlık tanımının her yerine “| C” ekle.

💡 Builder Notu — PGM ve Naive Bayes

Koşullu bağımsızlık = Bayesian ağların (graphical models) çekirdeği — bir değişken ebeveynlerine koşullu olarak gerisinden bağımsızdır; dev birleşik dağılım küçük çarpanlara iner. **Naive Bayes:** “özellikler, sınıfa koşullu bağımsızdır” — çoğu zaman yanlış ama şaşırtıcı işe yarar.

12.7 Explaining Away: Bağımsız ama Koşullu Bağımlı

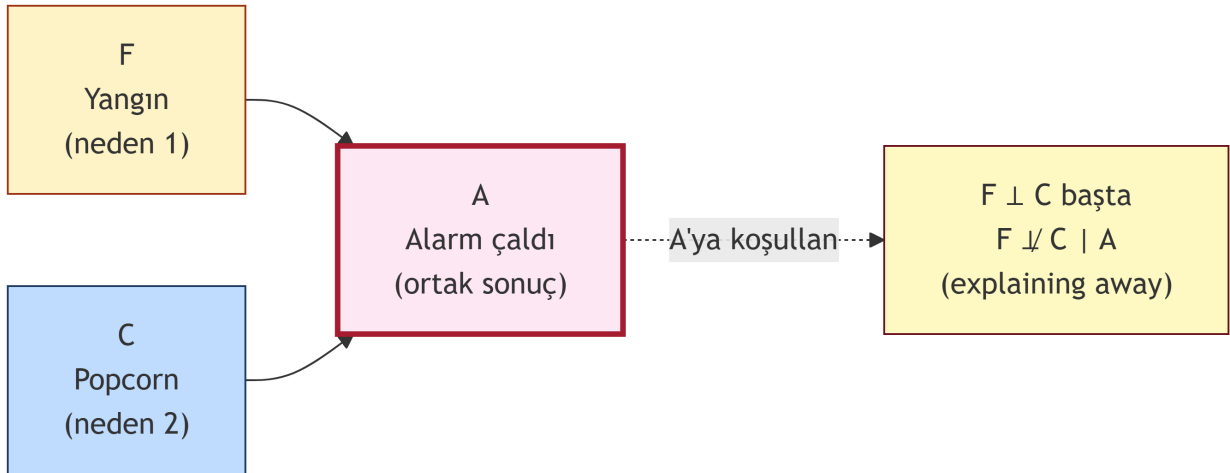
İki örnek bağımsızlık \leftrightarrow koşullu bağımsızlığın birbirini gerektirmediğini gösterir.

(a) **Koşullu bağımsız \Rightarrow bağımsız.** Bilmediğin güçlü/zayıf bir satranç rakibiyle bir dizi maç. Güce **koşullu** olarak maçlar bağımsız (güç sabit). Ama koşulsuz **bağımsız değil:** ilk beş galibiyet, sonrakileri tahmin etmeye yarar.

(b) **Bağımsız \Rightarrow koşullu bağımsız (explaining away).** Yangın alarmı: F = gerçek yangın, C = patlamış mısır. F ile C **bağımsız**. Alarmın çaldığını (A) öğrenince:

$$P(F | A \cap C^c) = 1$$

Alarm çaldı + popcorn yok \rightarrow yangın olmak **zorunda**. F ve C başta bağımsız, ama ortak sonuçlarına (A) koşullandığında **bağımlı** hale gelirler.



Şekil 12.3: Collider (v-yapısı): iki bağımsız neden (F, C) \rightarrow ortak sonuç (A). A'ya koşullanmak F ve C'yi bağımlı yapar — explaining away. Nedensel çıkarımda collider'a koşullanmak sahte korelasyon üretir (Berkson paradoksu).

! Builder Notu — Collider Bias / Berkson

Bu doğrudan **Bayesian ağlardaki collider (v-yapısı)**. Collider'a **koşullanmak** nedenleri bağımlı yapar — nedensel çıkarımda **sahte korelasyon** üretir (collider bias / **Berkson paradoksu**). Veri bir collider'a koşullanarak toplandıysa (sadece “hastaneye başvuranlar”), gerçekte bağımsız özellikler korele görünür ve model bu **sahte ilişkiyi** öğrenir. Çözüm: collider'a koşullanma/kontrol etme.

12.8 Bu Dersin Özeti

1. **Koşullu düşünmek** kursun ana teması; LOTP olasılıkta “böl ve birleştir”.
2. **LOTP**: $P(B) = \sum_i P(B | A_i)P(A_i)$, ayrık $\{A_i\}$ partition'ı. Yararı iyi bölmeye bağlı.
3. **İki kart paradoksu**: $P(\text{iki as} | \text{bir as}) = 1/33$, $P(\text{iki as} | \text{maça ası}) = 1/17$.
4. **Hastalık testi**: %95 doğru test, %1 base $\rightarrow P(D | T) \approx 0,16$.
5. **Savcı yanılması**: Sally Clark (Bayes'siz hesap masumu mahkum eder).
6. **Prior \leftrightarrow posterior**: $P(A) \neq P(A | B)$; $P(A | A) = 1$.
7. **Koşullu bağımsızlık** sıradanı gerektirmez; **explaining away** = collider'a koşullanmak nedenleri bağımlı yapar.

! Tek bir cümle

LOTP karmaşık bir olayı ayrık durumlara koşullayıp birleştirir; koşullamanın en derin dersi şudur: $P(A | B) \neq P(B | A)$, **prior \neq posterior**, ve **bağımsızlık koşula göre ortaya çıkıp kaybolabilir**.

12.9 Kontrol Soruları

i Soru 1: 2 adil para + 1 hileli (iki yüzü tura). Rastgele biri atılıyor. Tura olasılığı? (LOTP)

Cevap: $P(\text{tura}) = (2/3)(1/2) + (1/3)(1) = 1/3 + 1/3 = 2/3$.

i Soru 2: Tura geldi. Hileli olma olasılığı? (Bayes)

Cevap: $P(\text{hileli} | \text{tura}) = (1)(1/3)/(2/3) = 1/2$. Prior $1/3 \rightarrow$ posterior $1/2$.

i Soru 3: ‘NBA oyuncusu \rightarrow uzun’ doğru, ama ‘uzun \rightarrow NBA oyuncusu’ yanlış. Hangi koşulluklar?

Cevap: $P(\text{uzun} | \text{NBA}) \approx 1$; $P(\text{NBA} | \text{uzun}) \approx 0$. Savcı yanılması + base rate: NBA oyuncusu nadir, $P(\text{NBA})$ küçük, Bayes hesabını çökertir.

i Soru 4: (Builder) Collider'a koşullanmak neden sahte korelasyon ve ML için ne demek?

Cevap: Explaining away. Veri collider'a koşullanarak toplandıysa (Berkson), gerçekte bağımsız özellikler korele görünür ve model bu sahte ilişkiyi öğrenir. **Çözüm:** collider'a koşullanmadan örnekle, veya nedensel grafik bilgisiyle düzelt.

12.10 Egzersizler

Egzersiz 1. Fabrika: M1 (%50, %2 hatalı), M2 (%30, %3 hatalı), M3 (%20, %5 hatalı). Hatalı olma olasılığını LOTP ile bul.

Egzersiz 2. (Devam) Hatalı ürün bulundu. M3'ten gelme olasılığı? (Bayes.)

Egzersiz 3. Kendi "explaining away" örneğini kur — bağımsız iki neden, ortak sonuç.

Egzersiz 4. (Python — hastalık testi simülasyon)

$P(D | T) \approx 0.1615$ (Bayes beklenti: ~ 0.1610)

Egzersiz 5. (Sonraki ders) Monty Hall: 3 kapı, biri arabaya. Bir kapı seç, sunucu boş bir kapıyı açar, değiştirme şansı verir. Değiştirmeli misin? Önce sezgini yaz.

12.11 Sonraki Ders İçin Hazırlık

Ders 6: Monty Hall ve Simpson Paradoksu

İki ünlü paradoks. **Monty Hall:** değiştirmek kazanma olasılığını $1/3$ 'ten $2/3$ 'e çıkarır. **Simpson paradoksu:** gruplar tek tek bir yönde, birleştirilince ters yönde sonuç verebilir.

⚠ Ders 6 öncesi yapılacak

- Egzersizleri çöz — özellikle 4 (simülasyon) ve 5 (Monty Hall sezgisi).
- "Bayes paydası her zaman LOTP" refleksini pekiştir.
- Ana cümleyi tekrar oku: "LOTP karmaşık bir olayı parçalara böler..."

12.12 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Blitzstein'de
LOTP	$P(B) = \sum_i P(B A_i)P(A_i)$	6m00
Partition	Ayrık + birleşim S	5m15
İki kart paradoksu	$1/33$ vs $1/17$	15m43

Kavram	Tanım	Blitzstein'de
Hastalık testi	%95 test + %1 base \rightarrow \approx %16	26m15
Bayes + LOTP	Bayes paydası LOTP ile açılır	24m45
Savcı yanılgısı	$P(A B) \neq P(B A)$	32m45
Prior vs posterior	$P(A) \neq P(A B);$ $P(A A) = 1$	39m48
Koşullu bağımsızlık	$P(A \cap B C) =$ $P(A C)P(B C)$	41m14
Explaining away	Bağımsız nedenler, collider'a koşullu \rightarrow bağımlı	48m01

12.13 ML Bağlantıları Özeti

💡 7 köprü

1. **LOTP** \rightarrow marjinalleştirme; mixture, latent değişken, EM, HMM.
2. **Hastalık testi** \rightarrow **precision vs recall**; dengesiz veride accuracy yanılır.
3. **Savcı yanılgısı** \rightarrow likelihood \neq posterior; prior olmadan çıkarım yok.
4. **Prior** \leftrightarrow **posterior** \rightarrow Bayesian ML'in tanımı.
5. **Koşullu bağımsızlık** \rightarrow **Bayesian ağlar (PGM)**, naive Bayes faktörizasyonu.
6. **Explaining away (collider)** \rightarrow **selection bias**, Berkson paradoksu.
7. **Kanıt belirliliği** \rightarrow "belirli mi, varoluşsal mı"; prompt/koşul tasarımı.

❗ Tek bir şey alıp gideceksen

Zor bir olasılığı, ayrık durumlara koşullayıp birleştirerek (LOTP) parçala. Üç tuzaktan sakın: $P(A | B) \neq P(B | A)$, **prior** \neq **posterior**, **bağımsızlık koşula göre belirip kaybolur**. Koşullamanın yönü ve neye koşullandığın **her şeyi** değiştirir.

13 Monty Hall ve Simpson Paradoksu

Kanıtın tamamına koşulla; confounder'ı görmezden gelme

Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 6: Monty Hall, Simpson's Paradox](#) (≈49 dk)
- **Okuma süresi:** ≈26 dk

13.1 Bu Derste Ne Var?

Koşullamanın gücünü ve tuzaklarını iki paradoksta sınıyoruz:

1. **Monty Hall:** üç kapı, bir araba. Sunucu bir keçi kapısı açtıktan sonra **değiştirmek** kazanma olasılığını $1/3$ 'ten $2/3$ 'e çıkarır.
2. **Simpson paradoksu:** bir doktor her ameliyat türünde daha başarılı olabilir ama **toplamda** daha başarısız.

“the answer is that you should switch ... if you switch, your probability of success is two-thirds.”
— Blitzstein, 6:55

Builder Notu — ML Köprüleri

- **Monty Hall = veri-üretim sürecini modellemek.** “Kapı 2 boş” ile “Monty kapı 2'yi açtı” farklı kanıtlardır. **Missing not at random (MNAR)** çıkarımı değiştirir.
- **Simpson paradoksu = confounding** — nedensel çıkarımın merkez problemi. Korelasyon \neq nedensellik.
- **Ağırlıklar değişir = covariate shift.**
- Toplam vs alt-grup metrikleri = **fairness** analizi (Berkeley davası).

13.2 Monty Hall: Kurulum ve Gizli Varsayımlar

Üç kapı; birinin ardında **araba**, ikisinde **keçi**. Kapı 1'i seçiyorsun. Monty kalan iki kapıdan **keçili birini açıyor** ve değiştirme şansı veriyor.

Gizli varsayımlar (Blitzstein vurguluyor):

13 Monty Hall ve Simpson Paradoksu

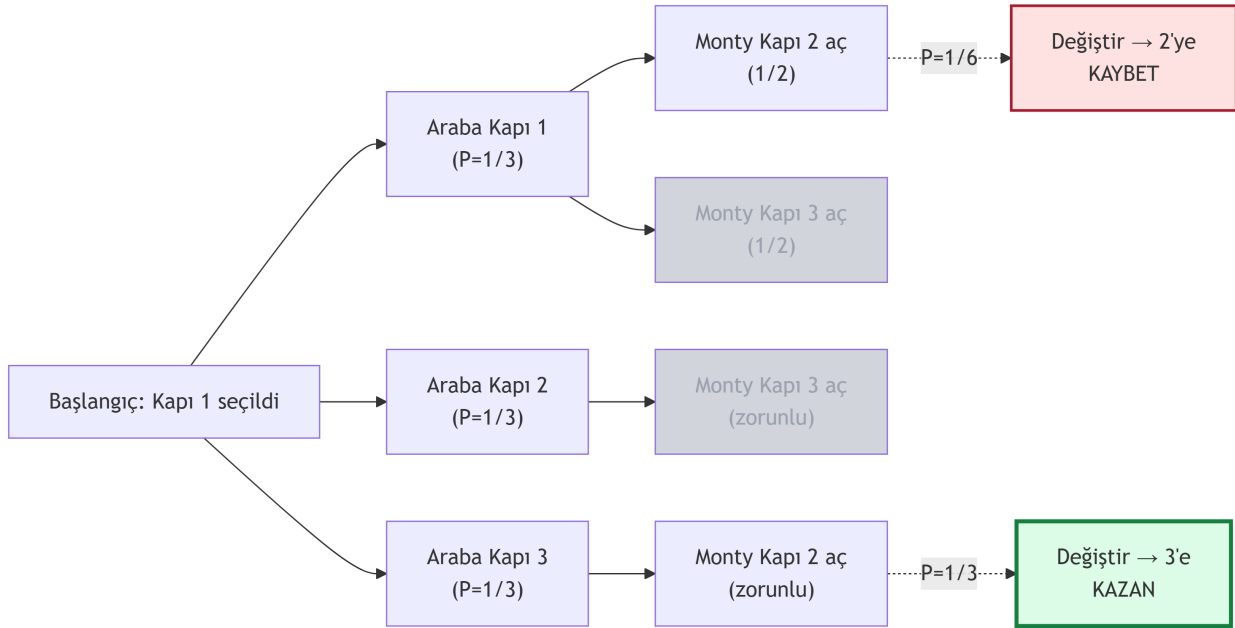
- Araba başta her kapıda **eşit olası** ($1/3$).
- Monty arabanın yerini **biliyor** ve her zaman bir keçi kapısı açıyor.
- Seçim varsa (senin seçtiğin kapıda araba), eşit olasılıkla seçiyor.

Naif sezgi: “İki kapı kaldı, 50/50.” Bu, naif tanımın kötüye kullanımı.

💡 Builder Notu — Data-Generating Process

Gizli varsayımları açık yazma disiplini, bir modelin **veri-üretim sürecini** tanımlama disiplindir. Eksik verinin neden eksik olduğu (MCAR/MAR/MNAR) çıkarımı değiştirir.

13.3 Monty Hall: Ağaç Diyagramı



Şekil 13.1: Monty Hall ağacı. Kapı 1 seçildi. Monty kapı 2’yi açtıysa kalan iki dal: (araba 1) olasılık $1/6$, (araba 3) olasılık $1/3$. Normalleştir: araba 1 $\rightarrow 1/3$, araba 3 $\rightarrow 2/3$.

Monty Kapı 2 açtıysa: (araba 1, Monty 2) = $1/3 \cdot 1/2 = 1/6$; (araba 3, Monty 2) = $1/3 \cdot 1 = 1/3$. Normalleştir: araba Kapı 1 $\rightarrow 1/3$, Kapı 3 $\rightarrow 2/3$.

“we wanna condition on all the evidence.” — Blitzstein, 8:38

Kritik: sadece “Kapı 2’de keçi var”a değil, “**Monty Kapı 2’yi açtı**” kanıtının tamamına koşulluyoruz.

13.4 Monty Hall: LOTP ve Milyon Kapı

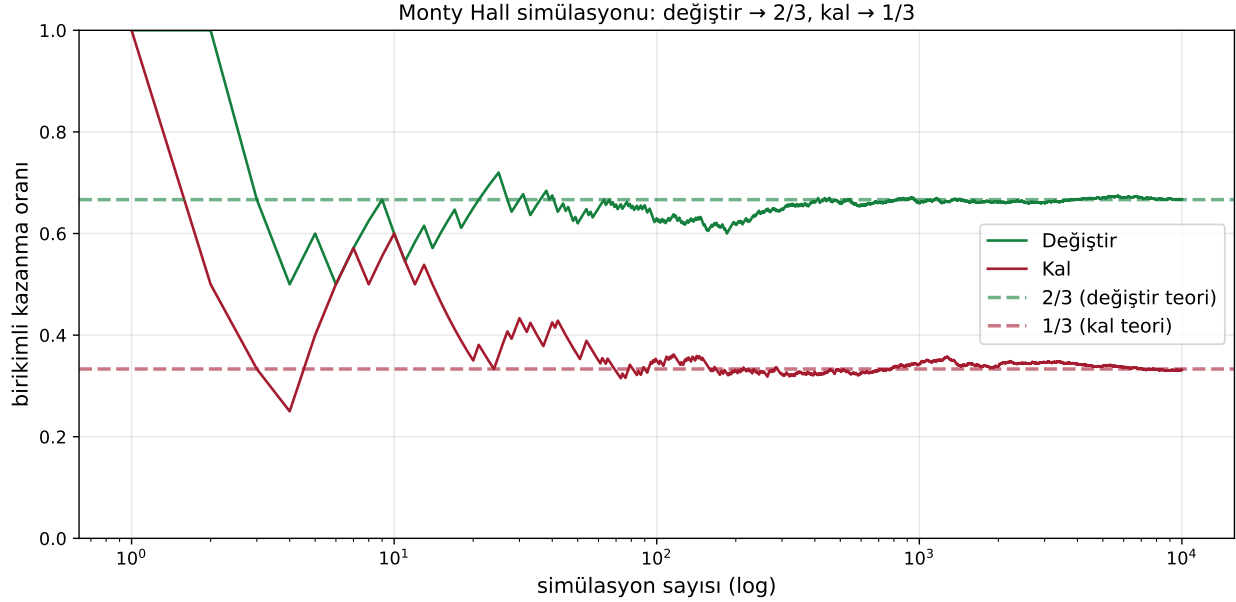
LOTP ile aynı sonuç — Blitzstein LOTP’ye “**wishful thinking**” demeyi öneriyor: neyi bilmeyi dilersin? Tabii ki **arabanın yerini**.

S = (değiştirme stratejisiyle) kazanma, D_j = araba Kapı j 'de:

$$P(S) = P(S | D_1)\frac{1}{3} + P(S | D_2)\frac{1}{3} + P(S | D_3)\frac{1}{3} = 0 + \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$$

Sezgi: 1/3 zamanda ilk tahmin doğru (değiştirme kaybettirir); 2/3 zamanda yanlış (Monty keçiyi açar, değiştirme kazandırır).

Milyon kapı: 1.000.000 kapıdan birini seç; Monty 999.998 keçi kapısı açsın. Değiştirir misin? Tabii ki — ilk tahminin neredeyse kesin yanlış.



Şekil 13.2: Monty Hall 10.000 simülasyon: ‘değiştir’ stratejisi 2/3’e, ‘kal’ 1/3’e yakınsar. Eğri Büyük Sayılar Yasası: simülasyon arttıkça frekans olasılığa yaklaşır.

Final: değiştir = 0.6672, kal = 0.3317

💡 Builder Notu — Bilgi Değeri ve Aktif Öğrenme

Monty Hall’ın asıl dersi: bir gözlemin **bilgi değeri**, onun nasıl ortaya çıktığına bağlıdır. Monty’nin bilerek keçi açması, kalan kapıya olasılık “pompalar”. ML’de bu **aktif öğrenme** sezgisidir — bilgilendirici bir gözlem posterior’u beklenenden çok kaydırabilir.

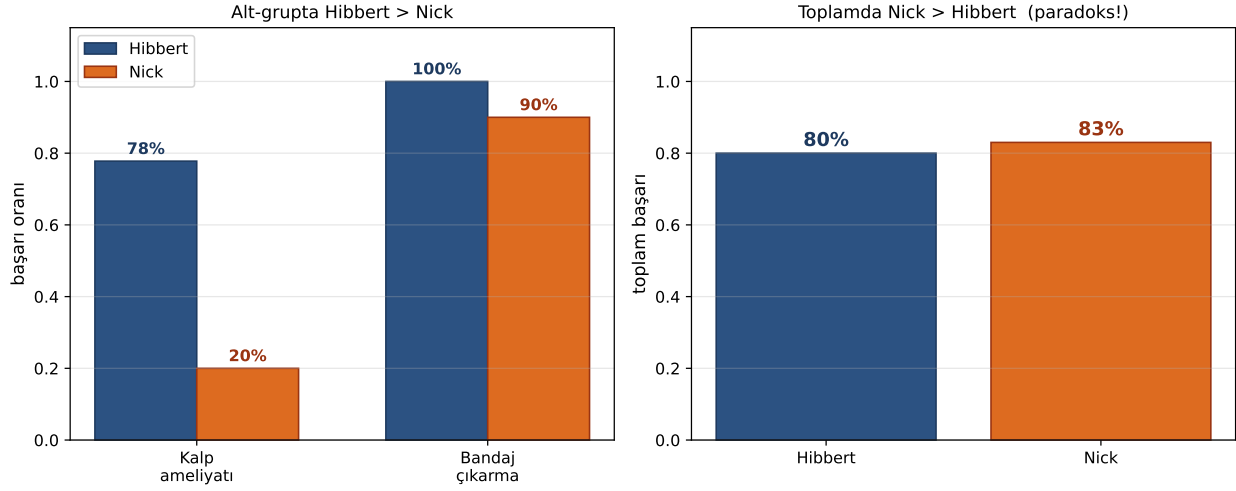
13.5 Simpson Paradoksu: İki Doktor

Bir doktor **her ameliyat türünde** diğerinden iyi olabilir mi, ama **toplamda** kötü? Evet.

“the signs of inequalities can flip when you aggregate data together.” — Blitzstein, 27:50

Doktor	Kalp ameliyatı	Bandaj çıkarma	Toplam
Hibbert	70/90 (%78)	10/10 (%100)	80/100 (%80)
Nick	2/10 (%20)	81/90 (%90)	83/100 (%83)

Hibbert her iki türde de iyi, ama **toplamda** Nick yüksek. Sebep: Nick'in ameliyatlarının %90'ı **kolay** (bandaj).



Şekil 13.3: Simpson paradoksu: Hibbert (mavi) her iki türde Nick'ten yüksek başarı, ama Nick'in case mix'i (mostly easy bandaj) toplam oranı yukarı çekiyo. Toplam metrik yanlış kararla götürür.

“there is no such thing as a true paradox ... the universe would explode.” — Blitzstein, 26:26

! Builder Notu — Sliced Metrics

Simpson paradoksu **aggregation yanlılığının** örneği: A/B testte veya metrikte, **alt-gruplara bakmadan** toplam sonuca güvenmek seni ters yöne götürür. Model değerlendirmede daima **dilimlenmiş (sliced) metrikler** ve fairness için alt-grup analizi gerekir.

13.6 Simpson: Confounder ve Berkeley Davası

Olaylarla: A = başarı, B = Nick, C = kalp ameliyatı. Tablo şunu der:

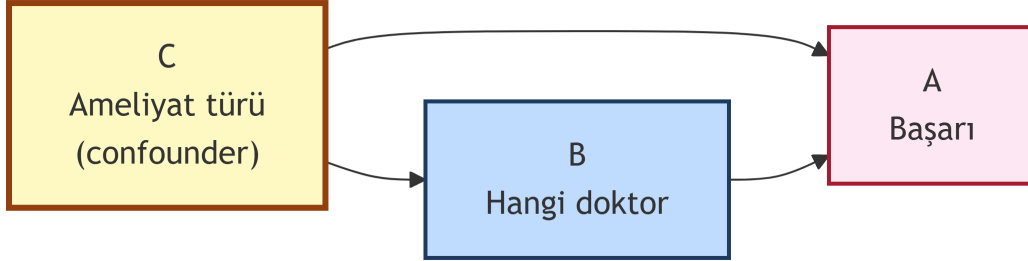
$$P(A | B, C) < P(A | B^c, C), \quad P(A | B, C^c) < P(A | B^c, C^c)$$

Ama toplamda ters döner:

$$P(A | B) > P(A | B^c)$$

Neden? Koşullu LOTP: $P(A | B) = P(A | B, C)P(C | B) + P(A | B, C^c)P(C^c | B)$. Buradaki **ağırlıklar** $P(C | B) \neq P(C | B^c)$ farklı; bu yüzden eşitsizlik aktarılmaz.

C bir **confounder** — hem B 'yi (hangi doktora gittin) hem A 'yı (başarı) etkiler. Doğru karşılaştırma C 'ye koşullamakla (tabakalama) elde edilir.



Şekil 13.4: Confounder grafiği: ameliyat türü C , hem doktor seçimini (B) hem başarıyı (A) etkiler. C 'yi kontrol etmeden $B \rightarrow A$ yorumu yanlıdır.

Berkeley davası (1973). Lisansüstü kabullerde cinsiyet ayrımcılığı iddiası: toplam oranlar erkekler lehine. Ama **bölüm bölüm** bakınca ayrımcılık kayboldu — kadınlar daha rekabetçi bölümlere başvuruyordu (**bölüm** = confounder).

! Builder Notu — Backdoor Adjustment

Berkeley **fairness ve nedensel çıkarımın** kanonik vakası. Doğru confounder'ları kontrol etmek (**backdoor adjustment**) şart — ama **Ders 5** collider uyarısını unutma: yanlış değişkene koşullamak sahte ilişki üretir. Neyi kontrol edeceğini istatistik değil **nedensel yapı** söyler.

13.7 Bu Dersin Özeti

1. **Monty Hall**: değiştir $\rightarrow 2/3$, kal $\rightarrow 1/3$. “Monty Kapı 2’yi açtı” kanıtının tamamına koşulla.
2. **Gizli varsayımlar**: Monty bilir + hep keçi açar + seçim varsa eşit. Değişirse problem değişir.
3. **Çözüm yolları**: ağaç (sil + yeniden normalleştir) ve LOTP (“wishful thinking” — arabaya koşulla).
4. **Milyon kapı**: uç durum, sezgiyi netleştirir.
5. **Simpson paradoksu**: her alt-grupta $A > B$, toplamda $B > A$ olabilir.
6. **Confounder**: ağırlıklar (case mix) gruplar arası farklı; LOTP’nin ağırlıkları eşitsizliği taşımaz.
7. **Berkeley davası**: toplam ayrımcılık, bölüm bazında kaybolur.

! Tek bir cümle

Monty Hall kanıtın **tamamına** (nasıl ortaya çıktığına) koşullamayı öğretir; Simpson paradoksu ise doğru **confounder’a** koşullamadan toplanan veriye güvenmenin tehlikesini — ikisi de “**neye koşulladığım her şeyi değiştirir**” dersinin iki yüzü.

13.8 Kontrol Soruları

i Soru 1: Monty Hall'da değiştirmek neden $1/2$ değil $2/3$?

Cevap: İlk tahminin $1/3$ doğru (değiştir kaybettirir); $2/3$ yanlış, o durumda Monty zorunlu olarak öteki keçiyi açar ve kalan kapı arabayı taşır.

i Soru 2: Monty arabamın yerini BİLMESEYDİ, şans eseri keçi açsaydı, hâlâ $2/3$ mü?

Cevap: Hayır — bu durumda $1/2$. Monty bilmeden açtığında kalan iki kapı simetrik kalır. Fark, Monty'nin **bilerek** keçi açmasının taşıdığı bilgide.

i Soru 3: İki oyuncu, sezonun her iki yarısında $A > B$ isabet, tüm sezonda $B > A$ mümkün mü?

Cevap: Evet — Simpson paradoksu. Şut hacimleri yarıdan yarıya çok farklıysa (ağırlıklar dengesiz), toplamda yön döner.

i Soru 4: (Builder) B varyantı mobilde ve masaüstünde iyi, toplamda kötü. Ne oldu?

Cevap: Simpson paradoksu. Cihaz karışımı varyantlar arasında farklı (B'ye düşen trafiğin çoğu mobil, mobil dönüşüm zaten düşük) — cihaz tipi **confounder**. Çözüm: **dilimlenmiş metrikler**, randomizasyon dengesi.

13.9 Egzersizler

Egzersiz 1. 4 kapılı Monty Hall: 1 araba, 3 keçi. Bir kapı seçtin, Monty bir keçi açtı. Kal mı, geçenlerden birine geç mi? Her seçeneğin olasılığı?

Egzersiz 2. Kendi Simpson paradoksu örneğini somut sayılarla kur.

Egzersiz 3. Milyon kapı argümanı neden 3 kapılı probleme de uygulanır? “50/50” sezgisinin çöktüğü yeri bir paragrafla açıkla.

Egzersiz 4. (Python — Monty Hall simülasyonu) Yukarıdaki kod hücrelerine bak — *değiştir* $\approx 0,667$ ve *kal* $\approx 0,333$ çıkmalı.

Egzersiz 5. (Sonraki ders) “Rastgele değişken” sezgisi: iki zar atışında “toplam”ı bir sayıya eşleyen fonksiyon. Kendi örneğini bir cümleyle yaz.

13.10 Sonraki Ders İçin Hazırlık

Ders 7: Kumarbazın İflası ve Rastgele Değişkenler

İki konu: **kumarbazın iflası (gambler's ruin)** — random walk sezgisi, fark denklemleri; ve kursun ikinci bölümüne giriş: **rastgele değişkenler** — sonuçları sayılara eşleyen fonksiyonlar.

⚠ Ders 7 öncesi yapılacak

- Egzersizleri çöz — özellikle 4 (Monty simülasyon) ve 5 (RD sezgisi).
- “Tüm kanıtla koşulla” + “confounder’a dikkat” reflekslerini pekiştir.
- Ana cümleyi tekrar oku: “*Neve koşulladığım her şeyi değiştirir.*”

13.11 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Blitzstein’de
Monty Hall	Değiştir $\rightarrow 2/3$, kal $\rightarrow 1/3$	6m55
Tüm kanıtla koşulla	“Monty açtı” \neq sadece “kapı boş”	8m38
Monty varsayımları	Bilir + hep keçi + eşit seçim	5m03
Ağaç çözümü	Tutarsız dalları sil + normalleştir	12m03
LOTP / wishful thinking	Arabanın yerine koşulla	16m26
Milyon kapı	Uç durum sezgiyi netleştirir	24m00
Simpson paradoksu	Alt-grup $A > B$, toplam $B > A$	27m50
Confounder	C hem grubu hem başarıyı etkiler	40m17
Ağırlıklar	$P(C B) \neq P(C B^c)$; LOTP aktarmaz	43m56
Berkeley davası	Toplam ayrımcılık, bölümde kaybolur	45m12

13.12 ML Bağlantıları Özeti

💡 7 köprü

1. **Monty Hall** \rightarrow veri-üretim sürecini modelle; MNAR; likelihood’un tam hâli.
2. **Tüm kanıtla koşulla** \rightarrow masked attention, posterior güncelleme.
3. **Milyon kapı** \rightarrow bilgilendirici gözlem; **aktif öğrenme** / deney tasarımı.
4. **Simpson paradoksu** \rightarrow **aggregation bias**; dilimlenmiş metrikler.
5. **Confounder** \rightarrow **nedensel çıkarım**, backdoor adjustment.
6. **Ağırlıklar / case mix** \rightarrow **covariate shift**, importance weighting.
7. **Berkeley** \rightarrow fairness alt-grup analizi; doğru değişkene koşulla (collider’a değil).

! Tek bir şey alıp gideceksen

Neye koşulladığın her şeyi değiştirir. Monty Hall, kanıtın tamamına (nasıl ortaya çıktığına) koşullamayı; Simpson paradoksu doğru confounder'a koşullamadan toplanan veriye güvenmemeyi öğretir.

14 Kumarbazın İflası ve Rastgele Değişkenler

First-step analysis, fark denklemi, Bernoulli, Binom

i Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 7: Gambler's Ruin and Random Variables](#) (≈52 dk)
- **Okuma süresi:** ≈28 dk

14.1 Bu Derste Ne Var?

Koşullamadan kursun ikinci büyük temasına geçiş. Blitzstein'e göre dönemin iki büyük fikri **koşullama** ve **rastgele değişkenler**.

1. **Kumarbazın iflası:** iki kumarbaz, \$1'lık bahisler. **First-step analysis + fark denklemi**.
2. **Rastgele değişken:** örnek uzaydaki sonuçları sayılara eşleyen bir **fonksiyon**.
3. **Bernoulli ve Binom** — ilk iki önemli dağılım.

“the two most important ideas for the entire semester ... conditioning and random variables and their distributions.” — Blitzstein, 0:23

💡 Builder Notu — ML Köprüleri

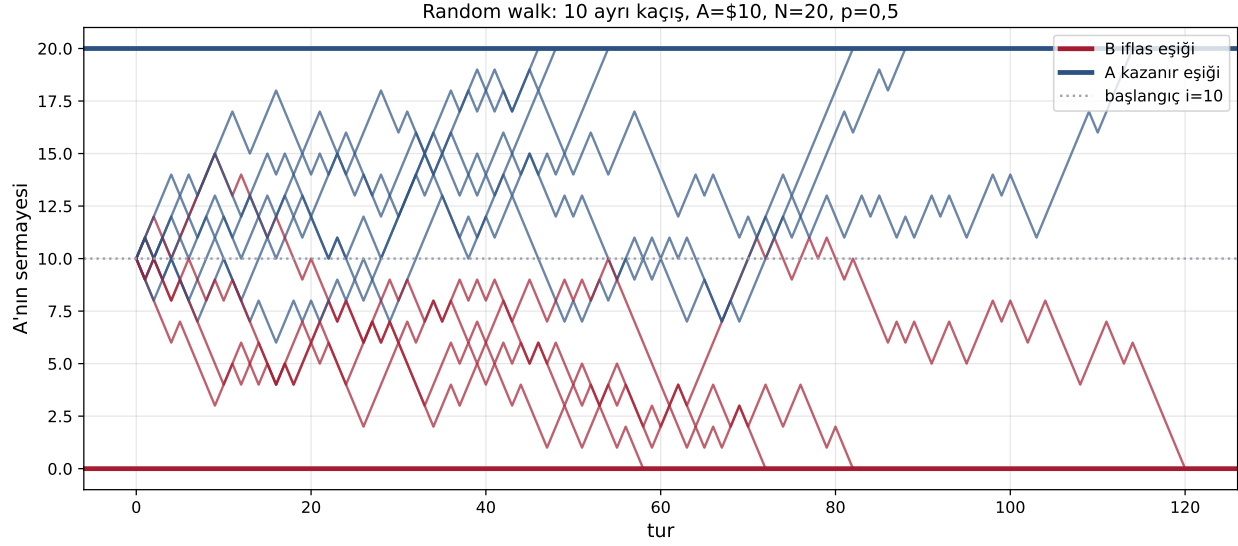
- **First-step analysis = Bellman denklemi.** $p_i = p \cdot p_{i+1} + q \cdot p_{i-1}$ özinyelemesi RL'deki **value iteration**'ın iskeleti.
- **Random walk = SGD'nin gürültülü gezintisi;** difüzyon ileri süreci; MCMC. Drift ($p \neq q$) = küçük biasın uzun ufukta nasıl biriktiği.
- **Fark denklemi = linear RNN / karakteristik kök;** **vanishing/exploding** olgusunun ayrık kardeşi.
- **Rastgele değişken = verinin fonksiyonu.** Feature, istatistik, **loss** — hepsi RV. **Bernoulli** = dropout maskesi / ikili etiket; **Binom** = başarı sayısı.

14.2 Kumarbazın İflası: Random Walk Olarak

İki kumarbaz A ve B, \$1'lık bahisler oynar. Her turda A, p olasılıkla kazanır, $q = 1 - p$ kaybeder. A elinde $\$i$, B elinde $\$(N - i)$. Oyun biri **iflas edene** kadar.

Random walk olarak: parçacık i 'de başlar, p ile sağa, q ile sola. 0 ve N **yutucu durum (absorbing state)**.

14 Kumarbazın İflası ve Rastgele Değişkenler



Şekil 14.1: Kumarbazın iflası random walk olarak: 10 ayrı kaçış, A=\$10, B=\$10 (N=20), p=0.5 adil oyun. Bazı yollar A=20’de (mavi → A kazanır), bazıları 0’da (kırmızı → B kazanır) yutulur.

💡 Builder Notu — Random Walk Her Yerde

Random walk ML’in her yerinde: **SGD** parametre uzayında gürültülü yürüyüş; **diffusion**’ın ileri süreci; **MCMC** örnekleyiciler. “Yutucu durum” kavramı erken durdurma, mutlak bariyerler ve RL’de **terminal state** ile aynı sezgi.

14.3 First-Step Analysis: Fark Denklemleri

p_i = A’nın $\$i$ ile başlayıp tüm oyunu kazanma olasılığı. **İlk adıma koşulla** (first-step analysis):

$$p_i = p p_{i+1} + q p_{i-1}, \quad 1 \leq i \leq N - 1$$

Sınır koşulları:

$$p_0 = 0, \quad p_N = 1$$

“first step analysis.” — Blitzstein, 11:48

❗ Builder Notu — Bellman Denklemi

$p_i = p \cdot p_{i+1} + q \cdot p_{i-1}$ denklemi, **RL’deki Bellman denkleminin** birebir iskeleti: bir durumun değeri = komşu durumların değerlerinin olasılıkla ağırlıklı ortalaması. First-step analysis = dinamik programlamada “bir adım aç, gerisini özyineleme bırak”. **Value iteration** ve **TD öğrenmenin** temeli.

14.4 Fark Denklemini Çözmek

Tahmin: $p_i = x^i$. Denkleme koy, x^{i-1} 'e böl:

$$px^2 - x + q = 0$$

$q = 1 - p$ olduğundan $1 - 4pq = (2p - 1)^2$. Kökler:

$$x = 1, \quad x = \frac{q}{p}$$

$p \neq q$ için sınır koşullarıyla:

$$p_i = \frac{1 - (q/p)^i}{1 - (q/p)^N}$$

Adil oyunda ($p = q = 1/2$), L'Hôpital:

$$p_i = \frac{i}{N}$$

💡 Builder Notu — Karakteristik Kök

$px^2 - x + q = 0$ bir **karakteristik denklem**. Doğrusal bir fark denklemini (ve linear RNN'i) köklerle çözenin yolu budur. Köklerin büyüklüğü kararlılığı belirler: $|\text{kök}| < 1$ söner, > 1 patlar — **kaybolan/patlayan gradyan** olgusunun ayrık kardeşi.

14.5 Moral: Minik Bias Felakettir

Adil oyun ($p = q$): $P(\text{A kazanır}) = i/N = \text{sermaye payı}$.

Adaletsiz oyun: çarpıcı. Eşit başlangıçla ($i = N/2$) ve $p = 0,49$ (turda %1 dezavantaj):

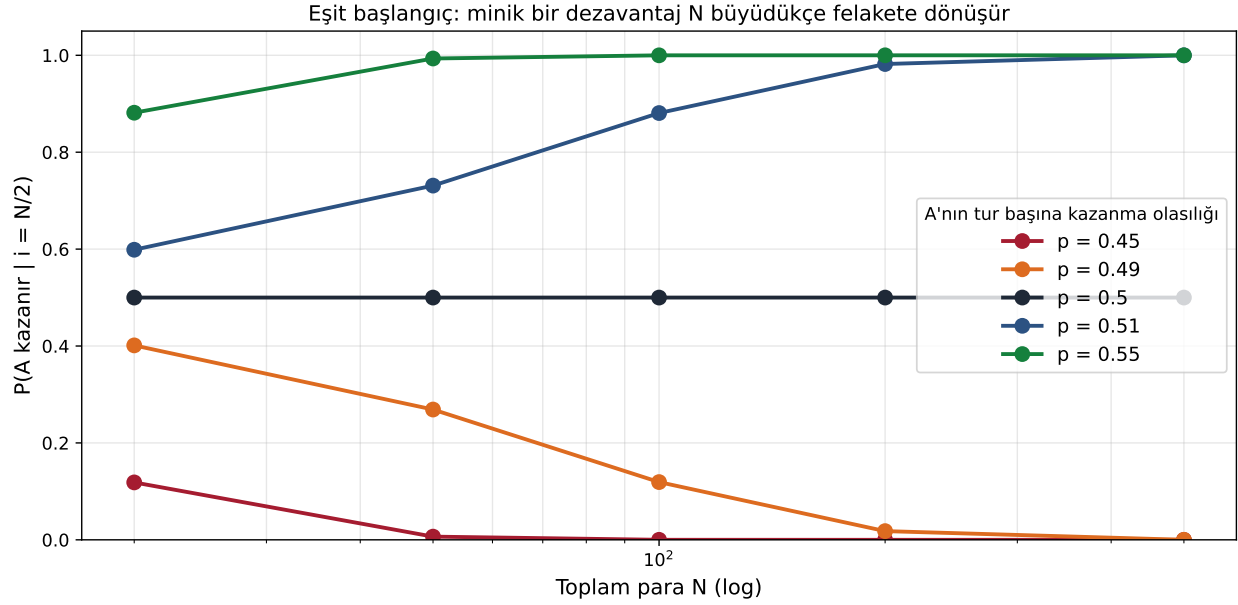
Las Vegas'ta kasa hem **daha çok paraya** sahip (büyük N) hem oyunlar gerçekten **hafifçe aleyhinedir**.

Salınım yok: $P(\text{A kazanır}) + P(\text{B kazanır}) = i/N + (N - i)/N = 1$. Oyunun sonsuza dek sürmesi olasılık 0.

💡 Builder Notu — Risk of Ruin ve Kelly

“Minik bias uzun ufukta felaket” dersi: ML'de **yansız (unbiased) kestiriciler** ve dikkatli adım boyutlarının önemi. “Risk of ruin” sezgisi → bankroll yönetimi, **Kelly kriteri**, RL'de keşif/risk dengesi. Küçük sürekli dezavantajla yeterince uzun oynayan bir ajan neredeyse kesin iflas eder.

14 Kumarbazın İflası ve Rastgele Değişkenler



Şekil 14.2: Adaletsiz oyun: $p=0.49$ (turda sadece %1 dezavantaj) bile N büyüdükçe felakete dönüşür. $N=200$ 'de yarı-yarıya başlayan A %2 kazanır. 'Risk of ruin' sezgisinin matematiksel kanıtı.

14.6 Rastgele Değişken Nedir?

Yaygın (ve işe yaramaz) tanım: "rastgele değerler alan değişken." Doğru tanım: bir **fonksiyon**

$$X : S \rightarrow \mathbb{R}$$

Deneyin her sonucu s 'yi bir sayıya eşler. Deneyin **bir yönünün sayısal özeti**.

"the idea is that we should think of a random variable just as a numerical summary ... of an aspect of the experiment." — Blitzstein, 41:08

Rastlantı deneyden gelir: deneyi yaparız, s gözleriz, X onu bir sayıya götürür.

💡 Builder Notu — Loss Bir RV'dir

"RV = (rastgele) verinin fonksiyonu" tanımı ML'in temelini netleştirir: bir **feature**, bir **istatistik**, hatta **loss** $L(\theta; \text{batch})$ bile rastgele girdinin fonksiyonudur — birer rastgele değişkendir. "Loss bir RV'dir" demek, onun beklenen değeri ve varyansı hakkında düşünebilmenin (sonraki dersler) anahtarıdır.

14.7 Bernoulli ve Binom Dağılımları

Bernoulli(p). En basit RV: yalnızca 0 ve 1. $P(X = 1) = p$, $P(X = 0) = 1 - p$.

Binom(n, p). n **bağımsız** Bernoulli(p) denemesindeki başarı sayısı. PMF:

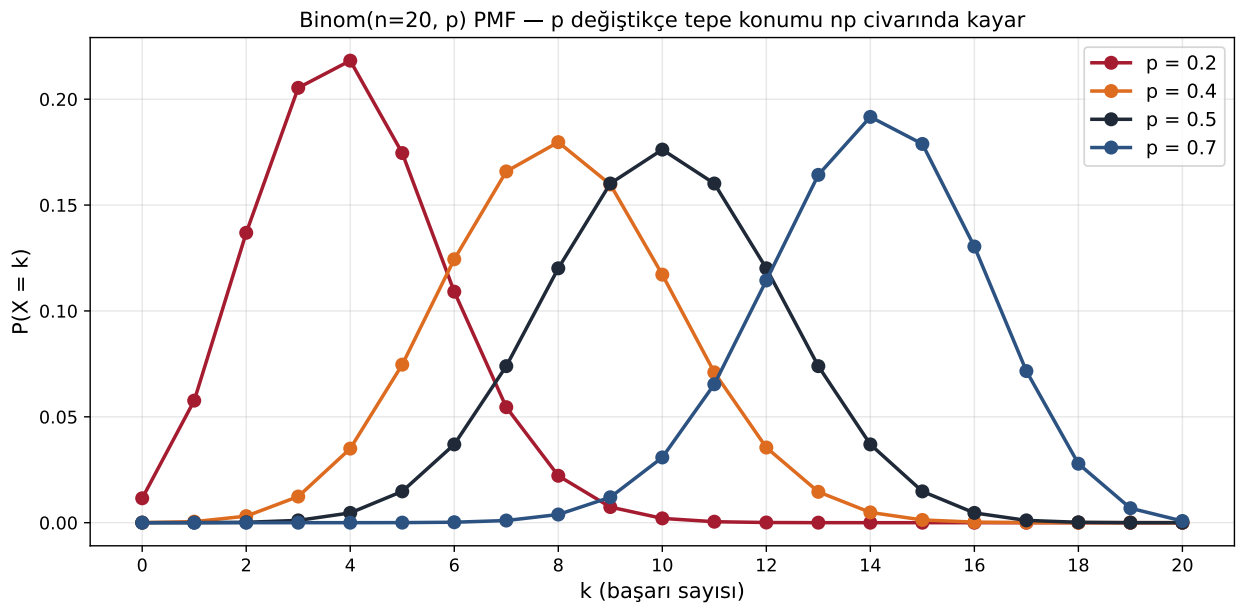
$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n$$

Nereden geliyor? k başarı + $n - k$ başarısızlık içeren **belirli** bir dizinin olasılığı $p^k(1-p)^{n-k}$; başarıların yerini seçmenin $\binom{n}{k}$ yolu var.

Kapanış özelliği (bağımsız binomlar):

$$X \sim \text{Bin}(n, p), \quad Y \sim \text{Bin}(m, p) \text{ bağımsız} \Rightarrow X + Y \sim \text{Bin}(n + m, p)$$

Story proof: $n + m$ bağımsız Bernoulli(p) denemesi \rightarrow toplam başarı $X + Y \sim \text{Bin}(n + m, p)$.



Şekil 14.3: Binom($n=20, p$) PMF'i farklı p için. $p=0.5$ simetrik; $p<0.5$ sola, $p>0.5$ sağa kayar. Tepe değeri np civarında (sonraki ders: $E[X] = np$).

! Builder Notu — Dropout, Binary Cross-Entropy

Bernoulli = dropout maskesi (her birim p olasılıkla tutulur), ikili sınıf etiketi, gösterge değışkeni. **Binom** = başarı sayısı; Binom/Bernoulli'nin **log-olabilirliđi** tam olarak **binary cross-entropy** loss'tur. Bağımsız binomların toplamının kapalıliđı, iki batch'in başarı sayılarını birleřtirmek gibi pratik durumlarda işine yarar — [Ders 2 Vandermonde](#)'un olasılıksal yüzü.

14.8 Bu Dersin Özeti

1. **Kumarbazın iflası:** A $\$i$, B $\$(N - i)$. Random walk ($0, N$ yutucu).
2. **First-step analysis:** $p_i = p \cdot p_{i+1} + q \cdot p_{i-1}$; $p_0 = 0, p_N = 1$.
3. **Çözüm:** $p_i = x^i \rightarrow$ kökler 1 ve q/p . $p \neq q$: $p_i = (1 - (q/p)^i) / (1 - (q/p)^N)$; $p = q$: $p_i = i/N$.

14 Kumarbazın İflası ve Rastgele Değişkenler

4. **Moral:** adil oyunda kazanma = sermaye payı; minik dezavantaj büyük N 'de felaket. Salınım olasılığı 0.
5. **Rastgele değişken:** fonksiyon $X : S \rightarrow \mathbb{R}$; deneyin sayısal özeti.
6. **Bernoulli(p):** $\{0, 1\}$. **Binom(n, p):** n bağımsız Bernoulli'de başarı sayısı; bağımsız toplam yine Binom.

! Tek bir cümle

Bir adıma koşullayıp özyineleme bırakmak (first-step analysis) kumarbazın iflasını bir fark denklemine indirger; rastgele değişken örnek uzayı sayılara eşleyen bir **fonksiyon** olarak, önümüzdeki tüm dağılım dünyasını açar.

14.9 Kontrol Soruları

i Soru 1: Adil oyun, $A=\$30$, $B=\$10$. A'nın tüm parayı kazanma olasılığı?

Cevap: $p_i = i/N$. $i = 30$, $N = 40 \rightarrow 30/40 = \mathbf{3/4}$. Adil oyunda kazanma = sermaye payı.

i Soru 2: $p_i = p \cdot p_{i+1} + q \cdot p_{i-1}$ 'in RL Bellman ile ilişkisi?

Cevap: Aynı yapı: durumun değeri = komşu değerlerin geçiş olasılıklarıyla ağırlıklı ortalaması. First-step analysis = value iteration + TD. Sınır koşulları terminal state değerlerine karşılık.

i Soru 3: 10 adil para atışında tam 3 tura?

Cevap: $X \sim \text{Bin}(10, 1/2)$. $P(X = 3) = \binom{10}{3}/2^{10} = 120/1024 \approx \mathbf{0,117}$.

i Soru 4: (Builder) Dropout'ta n nöron, her biri p olasılıkla tutulur. Aktif nöron sayısı?

Cevap: Her nöron Bernoulli(p), bağımsız. Aktif sayı = toplam = Binom(n, p). Beklenen aktif np (sonraki ders).

14.10 Egzersizler

Egzersiz 1. Adil oyun: $A=\$1$, $B=\$9$. A kazanma olasılığı? Sonucun küçüklüğüne dikkat.

Egzersiz 2. Adaletsiz: $p = 0,6$, $A=\$5$, $N=10$. $q/p = 2/3$, formülü kullan.

Egzersiz 3. Zar atışı + X = atılan sayının karesi. (a) S ? (b) X her sonucu neye eşler? (c) X değerleri kümesi?

Egzersiz 4. (Python — kumarbazın iflası simülasyonu)


simülasyon: 0.5017 (teori $i/N = 0.5$)

Egzersiz 5. (Sonraki ders) Binom PMF toplamının 1 olduğunu binom teoremiyle göster: $\sum_k \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1$.

14.11 Sonraki Ders İçin Hazırlık

Ders 8: Rastgele Değişkenler ve Dağılımları

PMF ve CDF, hipergeometrik dağılım, RV bağımsızlığı.


 Ders 8 öncesi yapılacak

- Egzersizleri çöz — özellikle 4 (kumar simülasyon).
- “First-step analysis = Bellman” + “RV = fonksiyon” sezgilerini pekiştir.
- Ana cümleyi tekrar oku: “Bir adıma koşullayıp özyineleme bırakmak...”

14.12 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Blitzstein'de
Kumarbazın iflası	A \$i\$, B \$(N - i)\$; kazanma olasılığı	2m42
Random walk	0 ve \$N\$ yutucu durum	7m43
First-step analysis	İlk tura koşulla → özyineleme	11m48
Fark denklemi	$p_i = p \cdot p_{i+1} + q \cdot p_{i-1}$	15m52
Çözüm (\$p \neq q\$)	$(1 - (q/p)^i)/(1 - (q/p)^N)$	24m04
Çözüm (\$p = q\$)	i/N	26m56
Rastgele değişken	$X : S \rightarrow \mathbb{R}$ fonksiyon	39m55
Bernoulli(\$p\$)	$\{0, 1\}, P(X = 1) = p$	43m06
Binom(\$n, p\$)	$\binom{n}{k} p^k (1 - p)^{n-k}$	46m01
PMF	$P(X = k)$; ayrık dağılım tarifi	49m49
Binom toplamı	$\text{Bin}(n, p) + \text{Bin}(m, p)$ $= \text{Bin}(n + m, p)$	50m47

14.13 ML Bağlantıları Özeti

 7 köprü

1. **First-step analysis** → **Bellman denklemi**, value iteration, TD.
2. **Random walk** → **SGD**, diffusion ileri süreci, MCMC.
3. **Fark denklemi** → linear RNN, karakteristik kök, **vanishing/exploding**.
4. **Minik bias birikir** → yansız kestirici, **Kelly**, risk of ruin.
5. **RV = veri fonksiyonu** → feature, istatistik, **loss** birer RV.
6. **Bernoulli** → dropout, ikili etiket, **binary cross-entropy**.

7. **Binom** → başarı sayısı, batch birleştirme, Vandermonde olasılıksal yüzü.

! Tek bir şey alıp gideceksen

Zor bir özyinelemeli problemi **ilk adıma koşullayarak** bir fark denklemine indir (Bellman'ın atası); rastgele değişkeni “değer alan değişken” değil, **örnek uzayı sayılara eşleyen bir fonksiyon** olarak gör.

15 Rastgele Değişkenler ve Dağılımları

PMF/CDF, IID, hipergeometrik vs binom

Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 8: Random Variables and Their Distributions](#) (≈50 dk)
- **Okuma süresi:** ≈28 dk

15.1 Bu Derste Ne Var?

Rastgele değişkenleri derinleştiriyoruz:

1. **Binom'a üç bakış:** story, **gösterge (indicator)** toplamı, PMF + **IID**.
2. **Dağılımı tanımlamak:** PMF (kesikli) ve CDF (her tür).
3. **Hipergeometrik:** yerine koymadan örneklemede başarı sayısı.

“there's actually three important ways to think of it ... the first one is the most important, because that's the story.” — Blitzstein, 1:28

Builder Notu — ML Köprüleri

- **Indicator toplamı** $X = X_1 + \dots + X_n$, beklenen değeri **doğrusallıkla** hesaplamamızın anahtarı (Ders 9) — karmaşık sayımları 0/1'lere indirir.
- **IID** = ML'in temel varsayımı; eğitim verisi ve minibatch örnekleri i.i.d.
- **CDF** → **inverse-transform sampling**, ROC, kalibrasyon.
- **Hipergeometrik vs Binom** = yerine koymadan vs koyarak; **minibatch** ≈ **binom** (büyük popülasyon).

15.2 Binom'a Üç Bakış + IID

(1) **Story:** $X = n$ bağımsız Bernoulli(p) denemesindeki başarı sayısı.

(2) **Gösterge toplamı:**

$$X = X_1 + X_2 + \dots + X_n, \quad X_j = \begin{cases} 1 & j. \text{ başarı} \\ 0 & \text{aksi} \end{cases}$$

Her X_j bir **gösterge (indicator)**. X_j 'ler **IID** — bağımsız ve özdeş dağılımlı.

(3) **PMF**: $P(X = k) = \binom{n}{k} p^k q^{n-k}$.

“IID means independent and identically distributed.” — Blitzstein, 5:48

 Builder Notu — Gösterge Altın Değerinde


Indicator toplamı ML'de altın değerinde: karmaşık bir sayımı (kaç doğru tahmin, kaç çakışma) 0/1'lerin toplamı olarak yazıp **beklentinin doğrusallığıyla** (Ders 9) tek satırda çözersin. IID ise tüm ML teorisinin zeminidir.

15.3 RV \neq Dağılım; $\{X = x\}$ Bir Olaydır

Rastgele değişken bir **fonksiyon** ($X : S \rightarrow \mathbb{R}$); dağılım onun olasılık profili. Çok sayıda farklı RV aynı dağılıma sahip olabilir.

“A very, very common confusion is to confuse random variables with distributions.” — Blitzstein, 6:35

“ $X = 7$ ” bir **denklem değil, olay**: $\{s : X(s) = 7\}$. Benzer şekilde “ $X \leq x$ ” da bir olaydır — CDF’i tanımlamamıza izin verir.

 Builder Notu — Model vs Dağılım

“RV bir fonksiyon, dağılım onun profili” ayrımı, ML’de **model** (somut ağ) ile o modelin temsil ettiği/öğrendiği **dağılım** ayrımına denk düşer. “ $\{X = x\}$ bir olaydır” görüşü, logit’leri olay olasılıklarına (softmax) bağlamanın temeli.

15.4 Dağılımı Tanımlamak: CDF ve PMF

CDF (birikimli dağılım fonksiyonu). Her RV için tanımlı:

$$F(x) = P(X \leq x)$$

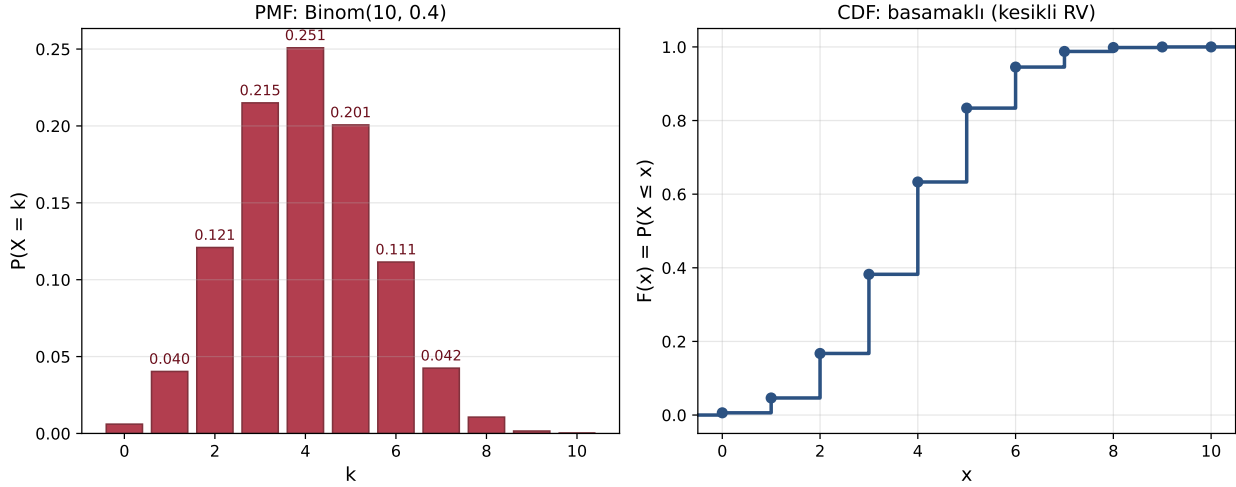
F artan; $x \rightarrow -\infty$ 'da 0'a, $+\infty$ 'da 1'e yaklaşır.

PMF (olasılık kütle fonksiyonu). Yalnızca **kesikli** için:

$$p_j = P(X = a_j) \geq 0, \quad \sum_j p_j = 1$$

Binom PMF toplamı = 1, binom teoremiyle:

$$\sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = (p + q)^n = 1$$



Şekil 15.1: Binom(10, 0.4) için PMF (sol, çubuklar) ve CDF (sağ, basamaklı). Kesikli RV'de CDF her değerde sıçrar, aralarda düz kalır. PMF'in toplam alanı = 1; CDF $-\infty$ 'dan $+\infty$ 'a $0 \rightarrow 1$.

! Builder Notu — Inverse-Transform Sampling

CDF'in evrenselliği: inverse-transform sampling — $U \sim \text{Uniform}(0, 1)$ üret, $F^{-1}(U)$ hesapla $\rightarrow X$ dağılımından örnek. Herhangi bir dağılımdan örnekleme bunun üzerine kurulu. CDF ayrıca **kuantiller**, **ROC eğrisi** ve kalibrasyon eğrilerinin temeli. PMF'in iki koşulu (negatif değil, toplam 1) bir **softmax'ı geçerli dağılım** yapan koşullar.

15.5 İki Binom'un Toplamı: Konvolüsyon = Vandermonde

$X \sim \text{Bin}(n, p)$, $Y \sim \text{Bin}(m, p)$ bağımsız $\rightarrow X + Y \sim \text{Bin}(n + m, p)$. Üç yol:

- **Story:** $n + m$ bağımsız Bernoulli(p) $\rightarrow \text{Bin}(n + m, p)$.
- **Indicator toplamı:** $X + Y = n + m$ tane IID Bernoulli(p).
- **PMF / konvolüsyon:**

$$P(X+Y = k) = \sum_j \binom{n}{j} p^j q^{n-j} \binom{m}{k-j} p^{k-j} q^{m-k+j} = p^k q^{n+m-k} \sum_j \binom{n}{j} \binom{m}{k-j} = \binom{n+m}{k} p^k q^{n+m-k}$$

Bu Vandermonde özdeşliğinin olasılıksal yüzü.

"in statistics, this is called a convolution." — Blitzstein, 25:48

💡 Builder Notu — Konvolüsyon Her Yerde

Bağımsız RV toplamı = dağılımların konvolüsyonu. ML’de her yerde: bağımsız gürültü eklemek (**diffusion** ileri süreci), bağımsız sayım birleştirme, **FFT ile hızlı konvolüsyon** ($O(n \log n)$). **Ders 2 Vandermonde**’un olasılıksal yüzü budur.

15.6 Hipergeometrik: Binom Sandığının Ama Olmayan

Binom’un kilit varsayımı: **bağımsız** denemeler + **aynı** p . İkisi de bozulursa binom değil.

Karşı örnek: 52 kartlık desteden 5 karttaki **as sayısı**. Yerine koymadan çekiyorsun → bağımlı. PMF:

$$P(X = k) = \frac{\binom{4}{k} \binom{48}{5-k}}{\binom{52}{5}}$$

Genel hâli (kavanozda w beyaz, b siyah, n çek):

$$P(X = k) = \frac{\binom{w}{k} \binom{b}{n-k}}{\binom{w+b}{n}}$$

Bu **hipergeometrik** dağılım. PMF’i ezberleme; **hikâyesini** (yerine koymadan, başarı sayısı) tanı.

“*This distribution is called the hypergeometric.*” — Blitzstein, 43:51

15.7 Hipergeometrik ↔ Binom: Büyük Popülasyon

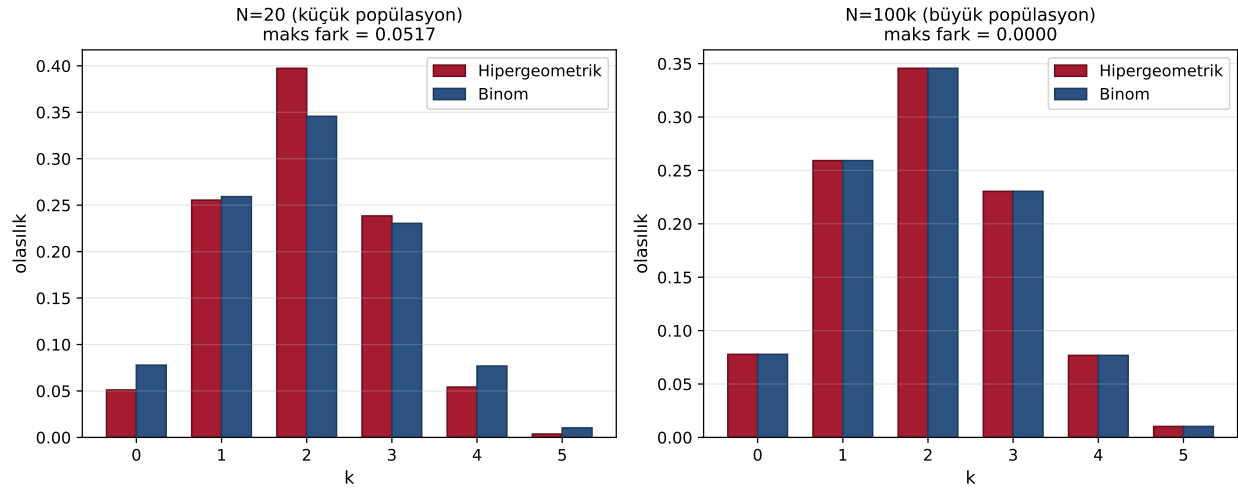
Popülasyon **çok büyük** ve örneklem **küçük** olduğunda **hipergeometrik** \approx **binom**. Bir milyar misketten 10 çekersen yerine koyup koymamak fark etmez.

! Builder Notu — Minibatch IID Varsayımı

“Büyük popülasyonda hipergeometrik \approx binom” sonucu, ML’de bir epoch’taki **minibatch’leri** (sonlu veri kümesinden **yerine koymadan**) pratikte **i.i.d.** (yerine koyarak) gibi ele alabilmemizin nedeni: veri kümesi batch’e göre çok büyükse sonlu-popülasyon düzeltmesi ihmal edilebilir.

15.8 Bu Dersin Özeti

1. **Binom’a üç bakış:** story, indicator toplamı, PMF. **IID** = bağımsız + özdeş.
2. **RV \neq dağılım:** Fonksiyon vs olasılık profili. “ $X = x$ ” bir olaydır.
3. **CDF** $F(x) = P(X \leq x)$ her RV için; **PMF** kesikli için ($p_j \geq 0, \sum p_j = 1$).
4. **Binom toplamı** = 1 (binom teoremi). $\text{Bin}(n, p) + \text{Bin}(m, p) = \text{Bin}(n + m, p) =$ konvolüsyon (Vandermonde).



Şekil 15.2: Hipergeometrik ↔ Binom yakınsaması. Aynı oran $p=0.4$, $n=5$ örneklem. Küçük popülasyonda ($N=20$) farklar belirgin; büyük popülasyonda ($N=100k$) birbirine oturur — minibatch’i ‘i.i.d. binom’ saymanın matematiksel temeli.

5. **Hipergeometrik:** yerine koymadan, $\binom{w}{k} \binom{b}{n-k} / \binom{w+b}{n}$. Büyük popülasyonda \approx binom.
6. **CDF şekilleri:** sürekli düzgün, kesikli basamaklı.

! Tek bir cümle

Rastgele değişken örnek uzayı sayılara eşleyen bir fonksiyondur; dağılımı **PMF veya CDF** ile tarif edilir. Aynı dağılımı **story, gösterge toplamı, PMF** gibi farklı temsillerle görmek ve “**yerine koyarak mı koymadan mı**” sorusu binom ile hipergeometriği ayırır.

15.9 Kontrol Soruları

i Soru 1: 10 atış, X = tura sayısı. X ’i gösterge toplamı olarak yaz, dağılımını söyle.

Cevap: $X = X_1 + \dots + X_{10}$, $X_i \sim \text{Bernoulli}(1/2)$ IID. $X \sim \text{Binom}(10, 1/2)$.

i Soru 2: 6 kırmızı 4 mavi, 3 çek. Kırmızı sayısı (a) yerine koyarak, (b) koymadan?

Cevap: (a) **Binom**(3, 0.6). (b) **Hipergeometrik:** $\binom{6}{k} \binom{4}{3-k} / \binom{10}{3}$.

i Soru 3: $P(X=0)=0.2$, $P(X=1)=0.5$, $P(X=2)=0.3$. $F(1.5)$?

Cevap: $F(1.5) = P(X = 0) + P(X = 1) = 0.7$. CDF $x = 1$ ’de 0,7’ye sıçrar, $x = 2$ ’ye kadar düz.

i Soru 4: (Builder) 1M örnekten 32 batch yerine koymadan. Sınıf sayısı tam olarak ne, neden binom gibi?

Cevap: Tam **hipergeometrik**. Popülasyon (1M) batch'e (32) göre devasa $\rightarrow \approx$ **Binom**(32, sınıf oranı). Sonlu-popülasyon düzeltmesi ihmal edilebilir.

15.10 Egzersizler

Egzersiz 1. 30 öğrenci, her biri %20 olasılıkla geç. (a) X 'i gösterge toplamı yaz. (b) Dağılımı?

Egzersiz 2. 20 kişi (8 K, 12 E), 5'li komite, tam 3 kadın olasılığı (hipergeometrik).

Egzersiz 3. Adil zar, $P(X = k) = 1/6$. (a) $F(3)$? (b) $P(2 \leq X \leq 4)$?

Egzersiz 4. (Python — hiper vs binom karşılaştırma)

N= 20: maks |hiper - binom| = 0.0517

N= 100000: maks |hiper - binom| = 0.0000

Egzersiz 5. (Sonraki ders) $\text{Bin}(n, p) = n$ Bernoulli(p) toplamı. Her göstergenin “ortalama” değeri p ise, toplamın ortalaması ne olmalı? (np tahminini gerekçelendir.)

15.11 Sonraki Ders İçin Hazırlık

Ders 9: Beklenti, Gösterge RD'ler, Doğrusallık

Beklentinin doğrusallığı: $E[X + Y] = E[X] + E[Y]$, **bağımsız olmasalar bile**. Bu, gösterge toplamı fikriyle birleşince binom, hipergeometrik, eşleşme beklentilerini tek satırda verir.

⚠ Ders 9 öncesi yapılacak

- Egzersizleri çöz — özellikle 1 (gösterge ayrışımı) ve 4 (hiper vs binom).
- “RV bir fonksiyon, dağılım onun profili” + “ $X+Y =$ konvolüsyon” sezgilerini pekiştir.
- Ana cümleyi tekrar oku: “Rastgele değişken bir fonksiyondur...”

15.12 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Blitzstein'de
Binom'a üç bakış	story / indicator / PMF	1m28
IID	Bağımsız + özdeş dağılımlı	5m48
Indicator RV	$X_j = 1$ (başarı), 0 (değil)	3m55
RV \neq dağılım	Fonksiyon vs profili	6m35
{X=x} olaydır	$\{s : X(s) = x\}$	9m59

Kavram	Tanım	Blitzstein'de
CDF	$F(x) = P(X \leq x)$	11m05
PMF	$p_j \geq 0, \sum p_j = 1$	16m36
Binom toplamı	$\text{Bin}(n, p) + \text{Bin}(m, p) = \text{Bin}(n + m, p)$	22m48
Konvolüsyon	RV toplamı; Vandermonde	25m48
Hipergeometrik	Yerine koymadan başarı sayısı	43m51
Hiper \approx Binom	Büyük popülasyon, küçük örneklem	46m13

15.13 ML Bağlantıları Özeti

💡 7 köprü

1. **Indicator toplamı** → beklentinin doğrusallığı; karmaşık sayım → 0/1 (Ders 9).
2. **IID** → ML temel varsayımı; minibatch.
3. **RV \neq dağılım** → model (fonksiyon) vs öğrenilen dağılım.
4. **CDF** → inverse-transform sampling, kuantiller, ROC.
5. **PMF koşulları** → softmax geçerliliği.
6. **Konvolüsyon** → bağımsız toplam, **diffusion gürültü**, FFT.
7. **Hiper \approx Binom** → minibatch'i i.i.d. ele almanın matematiksel temeli.

! Tek bir şey alıp gideceksen

Bir dağılımı tek bir formülle değil, **birden çok temsille** (story, gösterge toplamı, PMF/CDF) tut — ve “**yerine koyarak mı, koymadan mı**” binom ile hipergeometriği ayırır (büyük popülasyonda birleşirler).

16 Beklenti, Gösterge Değişkenler ve Doğrusallık

Fundamental bridge: $E[\mathbb{1}_A]=P(A)$; doğrusallık bağımlılıkta bile

Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 9: Expectation, Indicator Random Variables, Linearity](#) (≈50 dk)
- **Okuma süresi:** ≈28 dk

16.1 Bu Derste Ne Var?

Kursun en çok kullanacağın aracı: **beklenen değer** ve süper gücü **doğrusallık**.

1. **Beklenen değer:** $E(X) = \sum_x x \cdot P(X = x)$.
2. **Fundamental bridge:** $E(\mathbb{1}_A) = P(A)$.
3. **Doğrusallık:** $E(X + Y) = E(X) + E(Y)$ — **bağımlılıkta bile**.

“the single most important property of expectation is ... linearity.” — Blitzstein, 32:39

Builder Notu — ML Köprüleri

- **Beklenen değer = loss'un kendisi.** Minimize ettiğin $E[\text{loss}]$; **empirik risk** = örneklem ortalaması.
- **Fundamental bridge** → **Monte Carlo**; $\text{accuracy} = E[\mathbb{1}\{\text{doğru}\}]$.
- **Doğrusallık (bağımlıda bile)** → **minibatch gradyanı yansız:** $E[\sum \nabla \ell] = \sum E[\nabla \ell]$.
- **Geometrik dağılım** → RL'de **iskonto** (effective horizon $1/(1 - \gamma)$), belleksizlik.

16.2 Ortalama → Beklenen Değer

Ağırlıklı ortalamayı RV'ye taşı; ağırlıklar olasılıklar:

$$E(X) = \sum_x x \cdot P(X = x)$$

Olası her x değerini, olasılığıyla ağırlıkla. Bernoulli(p): $E(X) = 1 \cdot p + 0 \cdot q = p$.

💡 Builder Notu — Empirik Risk

Beklenen değer ML'de **amaç fonksiyonunun** kendisi: minimize ettiğin $E[\text{loss}]$ (gerçek risk). Pratikte hesaplayamazsın; **empirik risk** $= \frac{1}{n} \sum \text{loss}$, ağırlıkların hepsi $1/n$ olan ağırlıksız ortalama. Tüm denetimli öğrenme “ $E[\text{loss}]$ 'u küçült” der.

16.3 Fundamental Bridge

Bir A olayı için **gösterge** $\mathbb{1}_A = 1$ (A gerçekleşirse), 0 (aksi). Bu Bernoulli($P(A)$):

$$E(\mathbb{1}_A) = P(A)$$

Blitzstein buna **fundamental bridge** diyor — herhangi bir $P(A)$ 'yı bir göstergenin beklentisi olarak yazabilirsin.

“I call this thing the fundamental bridge ... between expected values and probabilities.” — Blitzstein, 26:34

💡 Builder Notu — Monte Carlo'nun Kalbi

Fundamental bridge, **Monte Carlo kestiriminin kalbi**: $P(A)$ 'yı göstergenin örneklem ortalamasıyla kestir — “kaç kez oldu / kaç deneme”. ML'de **accuracy** $= E[\mathbb{1}\{\text{doğru}\}] = P(\text{doğru})$; tıklama oranı, hit rate — hepsi göstergelerin ortalaması.

16.4 Doğrusallık

Beklentinin tek en önemli özelliği:

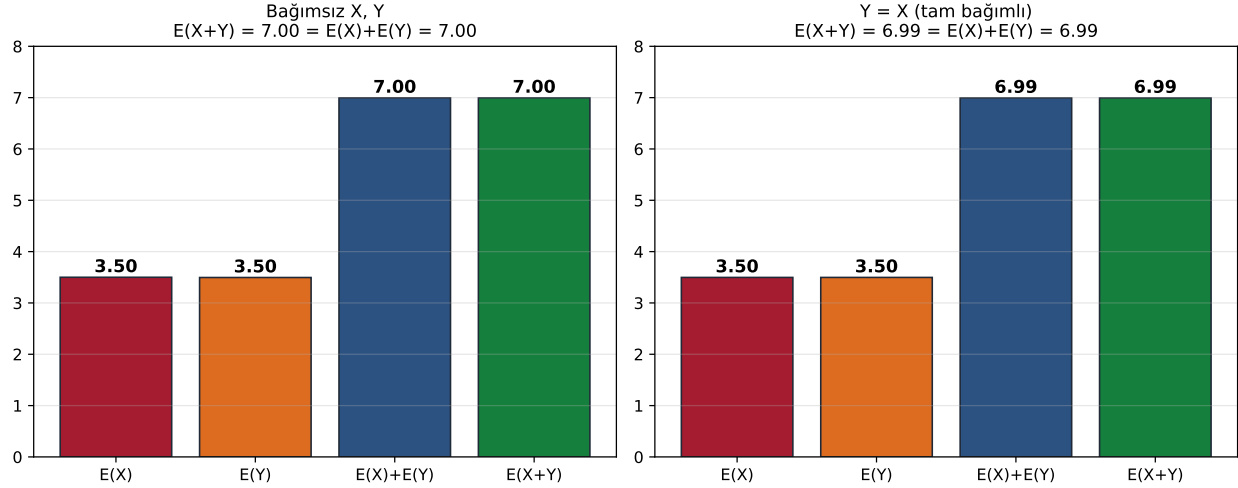
$$E(X + Y) = E(X) + E(Y), \quad E(cX) = c E(X)$$

Bağımlılık olsa bile doğrudur.

“the expected value of x plus y ... is always true even if x and y are dependent.” — Blitzstein, 33:07

Binom beklentisi (kazanç):

$$X = X_1 + \dots + X_n, \quad X_i \sim \text{Bern}(p) \Rightarrow E(X) = \underbrace{p + \dots + p}_n = np$$



Şekil 16.1: Doğrusallık bağımlılıkta da geçerli. Sol: bağımsız iki zar. Sağ: $Y=X$ (tam bağımlı). İki durumda da $E(X+Y) = E(X) + E(Y) = 7$. Bağımsızlık beklenti için gerekli değil — varyans için gerekli (Ders 10).

! Builder Notu — Yansız Minibatch Gradyanı

“Bağımlılıkta bile doğrusallık”, **minibatch gradyanının yansız** olmasının nedeni: $E[\sum \nabla \ell] = \sum E[\nabla \ell]$, örnekler korelasyonlu olsa bile. RL’de toplam ödülün beklentisi = adım ödülllerinin beklentilerinin toplamı. **“Beklentiği parçala, ayrı ayrı hesapla, topla”** refleksi sayısız türetmeyi tek satıra indirir.

16.5 Indicator + Linearity: Hipergeometrik Beklentisi

52 karttan 5 kart, X = as sayısı. PMF ile uğraşmak yerine **göstergelere ayır**:

$X_j = j$. kart as göstergesi. $X = X_1 + \dots + X_5$. Zincir:

$$E(X) = \sum_{j=1}^5 E(X_j) \stackrel{\text{simetri}}{=} 5 E(X_1) \stackrel{\text{bridge}}{=} 5 \cdot P(\text{1. kart as}) = 5 \cdot \frac{4}{52} = \frac{5}{13}$$

Üç adım: **doğrusallık + simetri + fundamental bridge**. X_j ’ler **bağımsız değil** — ama doğrusallık umursamıyor.

Genel hipergeometrik: $E(X) = n \cdot w / (w + b)$. Tıpkı binom (np) gibi görünür.

💡 Builder Notu — Beklenen Sayım Tekniği

“Karmaşık sayımı göstergelere ayır, doğrusallıkla topla” ML’de **beklenen sayıların** standart hesabı: hash collision sayısı, beklenen doğru tahmin, beklenen kapsama. Bağımsızlık gerektirmeden tek satırda. Kursun en pratik tekniklerinden biri.

16.6 Geometrik Dağılım

Geometrik(p): bağımsız Bernoulli(p) denemelerinde **ilk başarıdan önceki başarısızlık sayısı**.

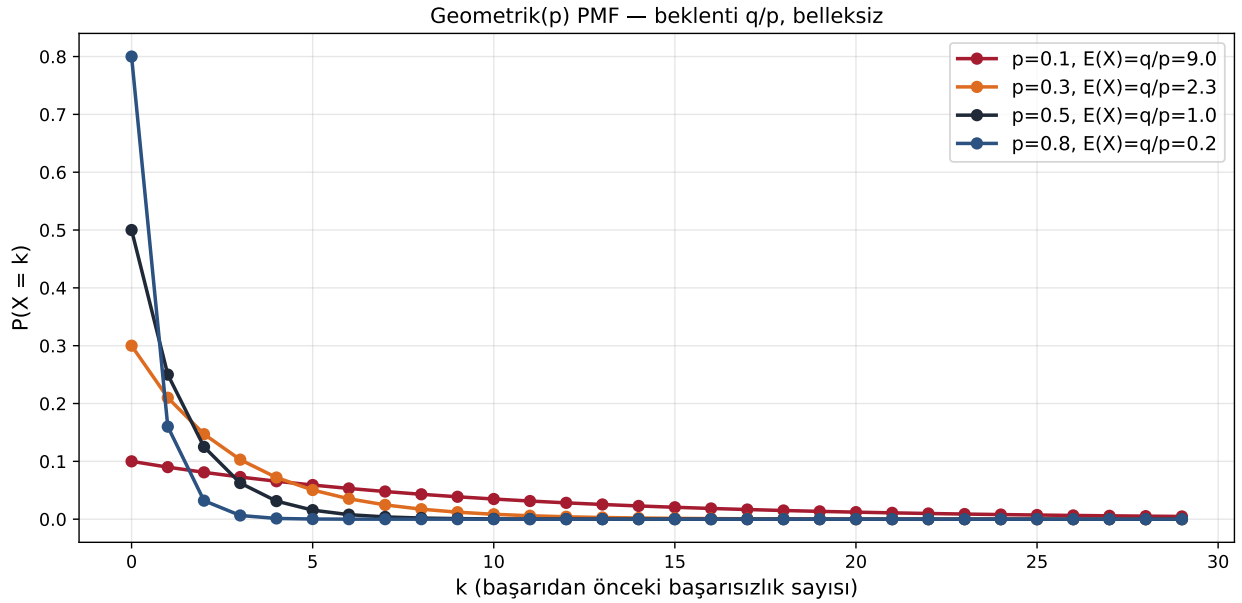
$$P(X = k) = q^k p, \quad k = 0, 1, 2, \dots$$

“the number of failures before the first success.” — Blitzstein, 41:53

Beklenti (story proof, first-step):

$C = E(X)$. İlk deneme başarı (p) $\rightarrow X = 0$; başarısızlık (q) $\rightarrow 1 + C$ (belleksiz):

$$C = p \cdot 0 + q(1 + C) \Rightarrow C = \frac{q}{p}$$



Şekil 16.2: Geometrik(p) PMF’i — p küçükse uzun kuyruk (çok deneme gerekir), p büyükse hızlı düşer. RL’de $p=1-\gamma$ olunca beklenen ufuk $q/p = \gamma/(1-\gamma) \approx 1/(1-\gamma)$.

! Builder Notu — RL İskonto ve Belleksizlik

Geometriğin **belleksizliği** sürekli karşılığı **üstel dağılımın** (Ders 16) imzası ve kuyruk teorisinin temeli. RL’de **geometrik iskonto**: her adımda γ olasılıkla devam \rightarrow bölüm uzunluğu Geometrik($1 - \gamma$); “effective horizon $\approx 1/(1 - \gamma)$ ” tam da q/p sezgisi.

16.7 Bu Dersin Özeti

1. **Beklenen değer** $E(X) = \sum x \cdot P(X = x)$: olasılıkla ağırlıklı ortalama.
2. **Fundamental bridge**: $E(\mathbb{1}_A) = P(A)$.

3. **Doğrusallık:** $E(X + Y) = E(X) + E(Y)$, **bağımlıda bile.**
4. **Indicator + doğrusallık + simetri + bridge:** binom (np), hipergeometrik ($n \cdot w / (w + b)$) tek satırda.
5. **Geometrik(p):** $P(X = k) = q^k p$, $E(X) = q/p$, belleksiz.

! Tek bir cümle

Beklentinin gerçek gücü **doğrusallıktır** — bağımlılık olsa bile $E(X + Y) = E(X) + E(Y)$. Karmaşık beklentiyi göstergelere ayır, doğrusallıkla topla, fundamental bridge ($E[\mathbb{1}_A] = P(A)$) ile olasılığa bağla.

16.8 Kontrol Soruları

i Soru 1: Adil zar, $E(X)$?

Cevap: $E(X) = (1 + 2 + \dots + 6)/6 = 21/6 = 3,5$. Beklenen değer mümkün bir değer olmak zorunda değil.

i Soru 2: 10 kişi şapkalarını karıştırır. Kendi şapkasını alan kişi sayısının beklentisi?

Cevap: $X = X_1 + \dots + X_{10}$, $X_j = j$. kişi kendi şapkasını aldı göstergesi. $E(X_j) = 1/10$. Doğrusallık: $E(X) = 10 \cdot 1/10 = 1$. X_j 'ler **bağımsız değil**, fark etmez. n 'den bağımsız hep 1 (**Ders 3** matching).

i Soru 3: (Builder) Model 100 örnekte 87 doğru. Accuracy %87. Hangi beklenti?

Cevap: Her tahmin $\mathbb{1}\{\text{doğru}\} \sim \text{Bern}(P(\text{doğru}))$. Accuracy = $(1/100) \sum \mathbb{1}\{\text{doğru}\}$ — göstergelerin ortalaması, **Monte Carlo kestirimi** $P(\text{doğru})$.

i Soru 4: (Builder) RL ajanında her adımda $\gamma=0,99$ devam. Bölüm uzunluğu dağılımı + E ?

Cevap: Geometrik($p = 1 - \gamma = 0,01$). $E = q/p = 0,99/0,01 = 99$ (effective horizon $\approx 1/(1 - \gamma) = 100$).

16.9 Egzersizler

Egzersiz 1. Üç top 1, 2, 6 (eşit olası). (a) $E(X)$? (b) $E(X^2)$? (c) $E(X^2) \neq (E(X))^2$ olduğunu göster.

Egzersiz 2. 5 adil zar. Gelen 6'ların beklenen sayısını gösterge + doğrusallıkla bul (tek satır).

Egzersiz 3. Bağımsız denemelerde $p = 0,1$. İlk kazanca kadar beklenen kayıp?

Egzersiz 4. (Python — bağımlılıkta doğrusallık)

$E(X+Y) \approx 6.9971$


$E(X) + E(Y) \approx 6.9971$ (eşit olmalı — bağımlıda bile)

Egzersiz 5. (Sonraki ders) $E(X^2) \neq (E(X))^2$ — adil zar için göster. Ders 10'da bunun **varyansa** götürdüğünü göreceğiz.

16.10 Sonraki Ders İçin Hazırlık

Ders 10: Beklentiye Devam

LOTUS (law of the unconscious statistician): $E[g(X)] = \sum g(x)P(X = x)$. Doğrusallık ispatı, daha fazla örnek, **varyansa** giden yol.


 Ders 10 öncesi yapılacak

- Egzersizleri çöz — özellikle 4 (bağımlılıkta doğrusallık).
- “Gösterge + doğrusallık” ve “fundamental bridge” reflekslerini pekiştir.
- Ana cümleyi tekrar oku: “*Beklenti olasılıkla ağırlıklı ortalamadır...*”

16.11 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Blitzstein'de
Beklenen değer	$E(X) = \sum x \cdot P(X = x)$	22m10
Fundamental bridge	$E(\mathbb{1}_A) = P(A)$	26m34
Doğrusallık	$E(X + Y) = E(X) + E(Y)$, bağımlıda bile	32m39
Sabit dışarı	$E(cX) = c \cdot E(X)$	33m42
Binom beklentisi	$E = np$	34m18
Hiper beklentisi	$E = n \cdot w / (w + b)$	35m47
Indicator tekniği	Sayım \rightarrow göstergeler toplamı	38m02
Geometrik(p)	$q^k p$; ilk başarı öncesi başarısızlık	41m17
Geometrik E	q/p	48m02

16.12 ML Bağlantıları Özeti

 7 köprü

1. **Beklenen değer** \rightarrow loss/risk; empirik risk = örneklem ortalaması.
2. **Fundamental bridge** \rightarrow **Monte Carlo**; accuracy/hit rate.
3. **Doğrusallık (bağımlıda bile)** \rightarrow **yansız minibatch gradyanı**; beklenen toplam ödül.
4. **Indicator + doğrusallık** \rightarrow beklenen sayım, bağımsızlık gerekmez.
5. **Geometrik / belleksizlik** \rightarrow üstel (Ders 16), **RL iskonto**, retry.
6. **CDF özellikleri** \rightarrow normalizing flow, ampirik CDF.
7. **RV bağımsızlığı** \rightarrow faktörize birleşik dağılım; naive Bayes, mean-field.

! Tek bir Őey alıp gideceksen

Beklentinin gerek gc **doęrusallıktır** — baęımlılık olsa bile. KarmaŐık beklentiyi **gstergelere ayır**, doęrusallıkla topla, **fundamental bridge** ile olasılıęa baęla — oęu hesap tek satıra iner.

17 Beklentiye Devam

Doğrusallık ispatı, negatif binom, FS, St. Petersburg, Jensen habercisi

Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 10: Expectation Continued](#) (≈50 dk)
- **Okuma süresi:** ≈28 dk

17.1 Bu Derste Ne Var?

Beklentiye derinleştiriyoruz:

1. **Doğrusallığın ispatı** — sonuçlar üzerinden toplam.
2. **Negatif binom** (r başarıya kadar) ve **first success** ($1/p$).
3. **Putnam yerel maksimum** — indicator + simetriyle tek satır.
4. **St. Petersburg paradoksu** — sonsuz beklenti; ve $E[g(X)] \neq g(E[X])$ uyarısı (**Jensen habercisi**).

Builder Notu — ML Köprüleri

- **Negatif binom = aşırı-dağınık** sayım modeli; varyans > ortalama olduğunda Poisson yerine NB regresyon.
- $1/p$ = **rejection sampling**'de beklenen deneme = 1/kabul oranı.
- **St. Petersburg** = beklenen değer sonsuz/yanıltıcı olabilir → sınırlı fayda, **Kelly kriteri**, risk-duyarlılık.
- $E[g(X)] \neq g(E[X])$ = **Jensen eşitsizliği** — ELBO, log-sum-exp; “E’yi içeri taşıma” hatasının kökü.

17.2 Doğrusallığın İspatı

İki yazım: **gruplu** (değerler üzerinden) ve **grupsuz** (sonuçlar üzerinden):

$$E(X) = \sum_x x P(X = x) = \sum_s X(s) P(s)$$

Grupsuzu kullanarak $T = X + Y$:

$$E(X + Y) = \sum_s (X(s) + Y(s))P(s) = \sum_s X(s)P(s) + \sum_s Y(s)P(s) = E(X) + E(Y)$$

Her şey aynı s üzerinden toplandı — bağımlılık hiç devreye girmedir. ■

💡 Builder Notu — Monte Carlo'nun Temeli

“Sonuçlar üzerinden toplam” görüşü **Monte Carlo**'nun temelidir: $E(X)$ 'i, örneklenen sonuçlardaki $X(s)$ değerlerinin ortalamasıyla kestir. Doğrusallığın bağımlılıktan bağımsız olması, derin ağlarda korelasyonlu gradyan beklentilerini bile rahatça toplamamızı sağlar.

17.3 Negatif Binom ve First Success

Negatif binom(r, p): ne negatif ne binom. Hikâye: r . başarıdan önceki başarısızlık sayısı (geometrik = $r = 1$ özel hâli).

“negative binomial is actually non-negative and it's not a binomial.” — Blitzstein, 15:00

$$P(X = n) = \binom{n+r-1}{r-1} p^r (1-p)^n$$

Beklenti: r bağımsız Geometrik(p) toplamı → doğrusallıkla

$$E(X) = r \cdot \frac{q}{p}$$

First Success FS(p): başarıyı **dahil eden** ilk başarı zamanı. $X \sim \text{FS}(p) \Rightarrow X - 1 \sim \text{Geometrik}(p)$.

$$E(X) = \frac{1}{p}$$

Çok sezgisel: $p = 1/10$ ise ortalama 10 deneme.

💡 Builder Notu — Rejection Sampling Maliyeti

$1/p$ = ilk başarıya kadar beklenen deneme = **rejection sampling**'in maliyeti: kabul oranı p ise, bir kabul edilen örnek için ortalama $1/p$ deneme. Aynı sezgi **retry mantığı** ve NLP'de **aşırı-dağınık** kelime sayımlarında (NB regresyon) karşına çıkar.

17.4 Putnam Yerel Maksimumlar

1 ... n permütasyonunda **yerel maksimum** sayısının beklentisi?

Gösterge $I_j = j$. konum yerel maksimum. Sayı = $\sum I_j$. Doğrusallık + bridge + simetri:

- **İç konumlar** ($n-2$ tane): üçlü içinde en büyük ortada $\rightarrow P = 1/3$ (1/4 değil — komşu karşılaştırmaları bağımlı).
- **Uç konumlar** (2 tane): tek komşu $\rightarrow P = 1/2$.

$$E = (n-2) \cdot \frac{1}{3} + 2 \cdot \frac{1}{2} = \frac{n+1}{3}$$

“1/3, 1/4 değil.” — Blitzstein, 36:07

💡 Builder Notu — Indicator + Simetri

“Indicator + simetri” bağımlılık olsa bile çalışır — **sıralama istatistikleri** ve **rekor sayısı** problemlerinin standart yöntemi (bir dizide beklenen rekor sayısı $\approx \ln n$).

17.5 St. Petersburg Paradoksu

Adil parayı **ilk tura** gelene kadar at; ödül 2^X dolar ($X \sim \text{FS}(1/2)$).

$$E(2^X) = \sum_{k=1}^{\infty} 2^k \cdot \frac{1}{2^k} = \sum_{k=1}^{\infty} 1 = \infty$$

Beklenen değer **sonsuz**. Ama kimse \$20-100'den fazla ödemez. Çözüm: ödülü trilyonla (2^{40}) sınırla \rightarrow toplam \approx \$40.

! Builder Notu — Reward Hacking ve Kelly

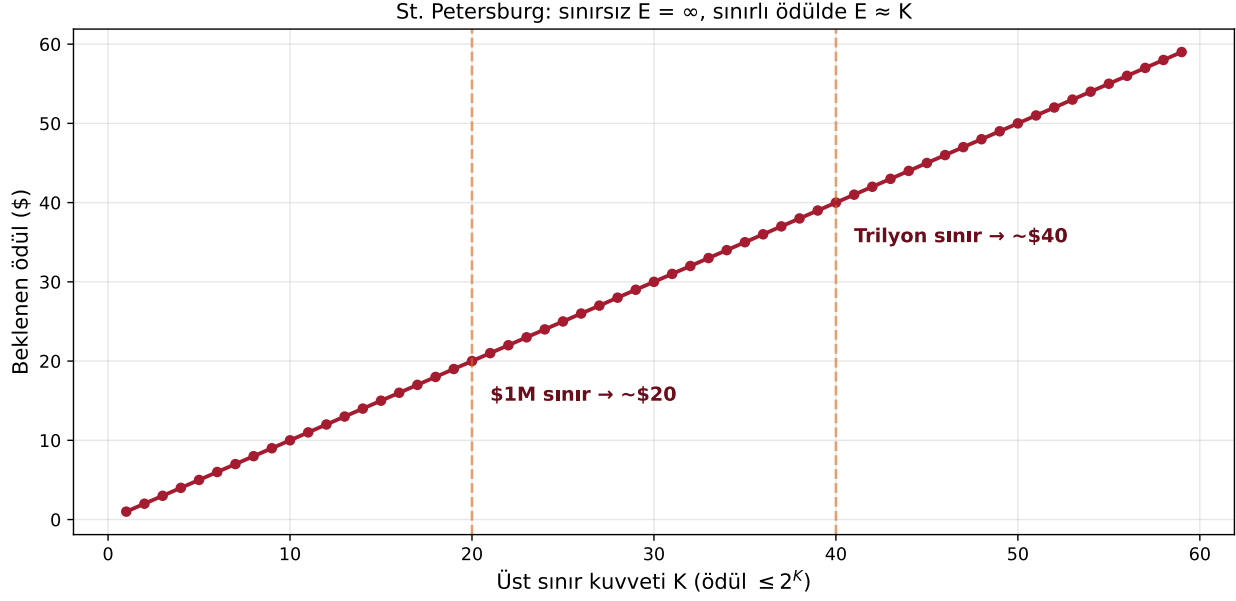
St. Petersburg, beklenen değer tek başına **yanıltıcı** olabileceğinin uyarısı: nadir ama devasa ödüller ortalamayı sonsuza taşır. ML/RL karşılığı: **sınırlı fayda**, **Kelly kriteri (log-servet)**, kuantil-temelli hedef. Ağır kuyruklu ödüllerde ortalama kötü bir özetir — sınırsız hedefler **reward hacking**'e davetiyedir.

17.6 Kritik Ders: $E[g(X)] \neq g(E[X])$

St. Petersburg gizli uyarı: E'yi **doğrusal olmayan** fonksiyonun içine taşıyamazsın. $X \sim \text{FS}(1/2)$ için $E(X) = 2$ idi. Karşılaştır:

$$E(2^X) = \infty \quad \neq \quad 2^{E(X)} = 2^2 = 4$$

Dünyalar kadar fark. **Yalnızca doğrusallık güvenli**. Genel olarak dışbükey g için:



Şekil 17.1: St. Petersburg paradoksu: beklenen ödül teorik olarak sonsuz, ama sonlu üst sınırla makulleşir. Üst sınır $2^K \rightarrow E \approx K$. Trilyon sınır (2^{40}) $\rightarrow E \approx 40$ \$. Nadir ama devasa ödüller beklenen değeri yanıltıcı yapar.

$$E[g(X)] \geq g(E[X]) \quad (\text{Jensen})$$

! Builder Notu — ELBO ve Log-Sum-Exp

$E[g(X)] \neq g(E[X])$, ML'deki en sık olasılıksal hatanın kökü ("E'yi içeri taşıma"). $E[\log p] \neq \log E[p]$ olduğu için **ELBO** vardır: $\log E[\cdot]$ hesaplanamadığından onun bir **alt sınırını** (Jensen ile) optimize ederiz. Aynı olgu **log-sum-exp**, cross-entropy ve "log-uzayda ortalama \neq ortalamanın logu" durumlarında karşına çıkar. **E'yi asla doğrusal-olmayan için sokma.**

17.7 Bu Dersin Özeti

1. **Doğrusallık ispatı:** $E(X) = \sum_s X(s)P(s)$; bağımlılık hiç gerekmez.
2. **Negatif binom**(r, p): r . başarısızlık; $E = rq/p$.
3. **FS**(p): başarı dahil ilk-başarı; $E = 1/p$.
4. **Putnam:** $E = (n + 1)/3$, indicator + simetri.
5. **St. Petersburg:** $E(2^X) = \infty$, ama sınırlı ödül makul.
6. $E[g(X)] \neq g(E[X])$: yalnız doğrusallık güvenli; **Jensen** habercisi.

! Tek bir cümle

Beklenti **sonuçlar üzerinden bir toplamdır**; doğrusallık bağımlılıktan bağımsızdır — ama beklentiyi **doğrusal olmayan** bir fonksiyonun içine taşıyamazsın ($E[g(X)] \neq g(E[X])$), ve sonsuz/ağır kuyruklu beklentiler tek başına aldatıcıdır.

17.8 Kontrol Soruları

i Soru 1: Basketbol, $p=0.8$. 3 sayıya kadar beklenen kaçırma?

Cevap: Negatif Binom($r = 3, p = 0,8$). $E = rq/p = 3 \cdot 0,25 = 0,75$.

i Soru 2: İlk 6 gelene kadar beklenen atış (dahil)?

Cevap: FS($p = 1/6$). $E = 1/p = 6$.

i Soru 3: (Builder) Neden $\log E[p]$ yerine ELBO ($E[\log p]$)?

Cevap: $E[g(X)] \neq g(E[X])$ (Jensen). \log içbükey $\rightarrow E[\log p] \leq \log E[p]$. $\log E[p]$ latent-değişkenli modellerde hesaplanamaz (integral logun içinde); $E[\log p]$ örneklenebilir. **ELBO** = $\log E[p]$ 'nin Jensen alt sınırı.

i Soru 4: (Builder) Ağır kuyruklu ödülde neden ortalama güvensiz?

Cevap: St. Petersburg gibi: nadir devasa ödüller ortalamayı şişirir, ajan onu kovalayıp **reward hacking** yapabilir. Çözüm: ödül clip, log-fayda (Kelly), kuantil/medyan, risk-duyarlı hedef.

17.9 Egzersizler

Egzersiz 1. %10 kusurlu, 2. kusura kadar beklenen kusursuz? (NB, $r = 2, p = 0,1$.)

Egzersiz 2. 1 ... 10 permütasyonunda beklenen yerel maksimum?

Egzersiz 3. Ödül 3^X olsaydı ($X \sim \text{FS}, p = 1/2$). E sonlu mu? (İpucu: $\sum (3/2)^k$.)


Egzersiz 4. (Python — NB beklenti)

simülasyon: 4.498 formül rq/p : 4.500

Egzersiz 5. (Sonraki ders) Bin(n, p)'de $n \rightarrow \infty, p \rightarrow 0, np = \lambda$ sabit. PMF neye yaklaşır? (Ders 11: Poisson = nadir olaylar limiti.)

17.10 Sonraki Ders İçin Hazırlık

Ders 11: Poisson Dağılımı — nadir olayların dağılımı; binom limiti; PMF + beklenti + toplama.


 Ders 11 öncesi yapılacak

- Egzersizleri çöz.
- “Doğrusallık + indicator” ve “ $E[g(X)] \neq g(E[X])$ ” reflekslerini pekiştir.
- Ana cümleyi tekrar oku: “*Beklenti sonuçlar üzerinden bir toplamdır...*”

17.11 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Blitzstein'de
Doğrusallık ispatı	$E(X) = \sum_s X(s)P(s)$	7m07
Negatif binom (r, p)	$\binom{n+r-1}{r-1} p^r q^n$; r . başarı öncesi	14m51
NB beklenti	rq/p	24m47
FS (p)	Başarı dahil ilk-başarı; $E = 1/p$	26m51
Putnam	$E = (n + 1)/3$	33m35
1/3 değil 1/4	Karşılaştırmalar bağımlı	36m07
St. Petersburg	$E(2^X) = \infty$; sınırlıda makul	39m15
$E[g(X)] \neq g(E[X])$	Yalnız doğrusallık güvenli; Jensen	49m55

17.12 ML Bağlantıları Özeti

 6 köprü

1. **Doğrusallık (sonuçlar)** → Monte Carlo; korelasyonlu gradyan beklentileri.
2. **Negatif binom** → aşırı-dağınık sayım; **NB regresyon**.
3. **First success** / $1/p$ → **rejection sampling** maliyeti, retry.
4. **Indicator + simetri** → beklenen sayım; rekor istatistikleri ($\approx \ln n$).
5. **St. Petersburg** → sınırlı fayda, **Kelly**, **reward hacking** riski.
6. $E[g(X)] \neq g(E[X])$ → **Jensen**, **ELBO**, log-sum-exp.

! Tek bir Őey alıp gideceksen

Beklenti sonular zerinden bir toplam → **doęrusallık baęımlılıęa aldırmaz**. Ama beklentiye doęrusal olmayan bir fonksiyonun iine taŐıyamazsın (**Jensen**), ve sonsuz/aęır kuyruklu beklentiler tek baŐına aldaticıdır.

18 Poisson Dağılımı

Nadir olaylar limiti; binomdan Poisson'a

Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 11: The Poisson distribution](#) (≈43 dk)
- **Okuma süresi:** ≈25 dk

18.1 Bu Derste Ne Var?

Son temel kesikli dağılım — olasılıkla en çok kullanılan.

1. **Poisson(λ):** PMF $e^{-\lambda} \lambda^k / k!$, $E = \lambda$. “Nadir olaylar”.
2. **Poisson paradigması:** çok sayıda, küçük olasılıklı, (zayıf) bağımsız olay \rightarrow sayıları \approx Poisson.
3. **Binom limiti:** $n \rightarrow \infty$, $p \rightarrow 0$, $np = \lambda \rightarrow$ Binom \rightarrow Poisson.

Builder Notu — ML Köprüleri

- **Poisson** = sayım modellemenin temel taşı: **bag-of-words / topic model**, olay akışları, tıklama sayıları, **Poisson regresyon**.
- **Var = ortalama** = λ (equidispersion); ihlal edilirse \rightarrow **negatif binom** (Ders 10).
- “ $P(\geq 1) \approx 1 - e^{-\lambda}$ ” kalıbı = nadir olay olasılıklarını (hash collision, near-duplicate, nadir hata) hızlı kestirmenin yolu.

18.2 Poisson Dağılımı: PMF ve Beklenti

$\lambda > 0$ tek parametrelili (rate). Üst sınırı yok:

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

Geçerli mi? Toplam = $e^{-\lambda} \cdot \sum \lambda^k / k! = e^{-\lambda} \cdot e^{\lambda} = 1$ (Taylor serisi).

Beklenti:

$$E(X) = \lambda$$

💡 Builder Notu — Poisson Regresyon

Poisson, **sayım verisinin** varsayılan ilk modeli ve **Poisson regresyonun** temeli. Kritik özellik: varyansı da λ (equidispersion). Gerçek veri çoğu zaman daha **dağınıktır** (varyans > ortalama); o zaman Poisson yetmez → **negatif binom**. “Ortalama = varyans mı?” = Poisson mu NB mi?

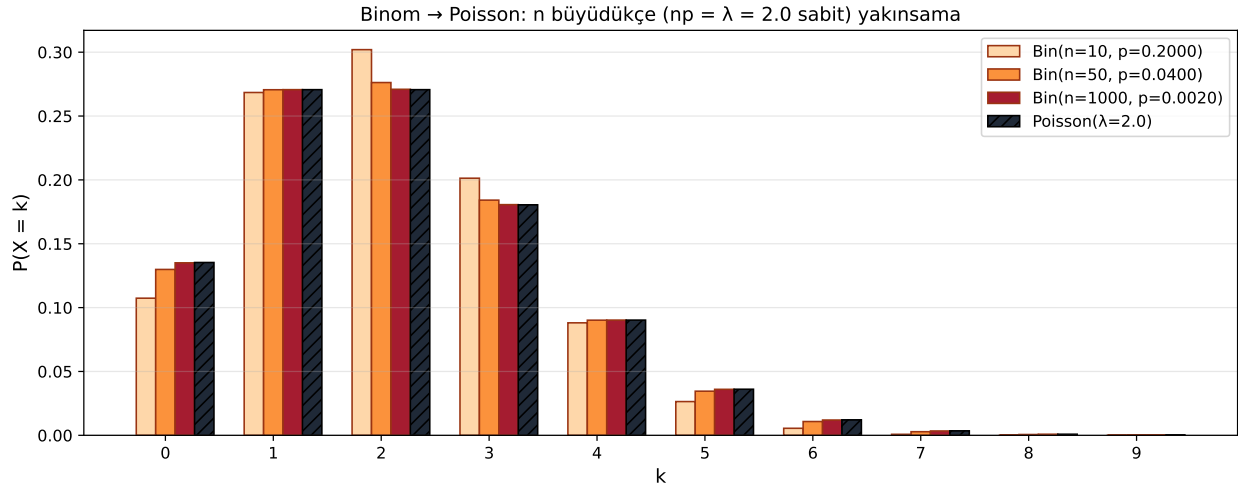
18.3 Poisson Paradigması ve Binom Limiti

Olaylar $A_1, \dots, A_n, P(A_j) = p_j$. n büyük, p_j küçük, **bağımsız ya da zayıf bağımlı** → gerçekleşen sayı \approx Poisson(λ), $\lambda = \sum_j p_j$ (doğrusallık).

Binom limiti (ispatla): $X \sim \text{Bin}(n, p)$, $np = \lambda$ sabit, $n \rightarrow \infty$:

$$\binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \rightarrow \frac{e^{-\lambda} \lambda^k}{k!}$$

Anahtar: $(1 + x/n)^n \rightarrow e^x$ (sürekli bileşik faiz).



Şekil 18.1: Binom → Poisson yakınsaması. $\lambda=2$ sabit, $n=10/50/1000$. n büyüdükçe binom PMF Poisson(2)’ye yaklaşır. Pratik kazanç: trilyon n’de bile Poisson tek parametreyle hesaplanır.


! Builder Notu — Hesaplanabilirlik Kazancı

“Küçük sayılar yasası” (Poisson limit teoremi) pratik kazanç sağlar: n trilyon, p minik olsa binom katsayıları rahat hesaplanmaz — Poisson(λ) tek parametreyle anında verir.

18.4 Nadir Olay Olasılığı: $P(\geq 1) \approx 1 - e^{-\lambda}$

Beklenen sayıyı doğrusallıkla bul $\rightarrow P(\text{en az bir}) \approx 1 - e^{-\lambda}$. Bağımlılık olsa bile $\lambda = \sum p_j$ formülü çalışır.


Örnek (üçlü doğum günü): $\lambda = \binom{n}{3}/365^2$. 100 kişide $\lambda \approx 1,214 \rightarrow P(\geq 1) \approx 1 - e^{-1,214} \approx 0,70$.

 Builder Notu — Nadir Olay Refleksi

“Beklenen sayıyı doğrusallıkla bul, $1 - e^{-\lambda}$ ” = **hash collision, near-duplicate, nadir hata** olasılıklarını saniyede kestirmenin standart yolu. Zayıf bağımlılık altında bile çalışır.

18.5 Bu Dersin Özeti

1. **Poisson(λ):** $e^{-\lambda} \lambda^k / k!$; üst sınırsız.
2. **Beklenti** = λ , varyans da λ (equidispersion).
3. **Poisson paradigması:** n büyük, p_j küçük, zayıf bağımlı \rightarrow Poisson($\sum p_j$).
4. **Binom limiti:** $n \rightarrow \infty$, $np = \lambda \rightarrow$ Poisson(λ).
5. $P(\geq 1) \approx 1 - e^{-\lambda}$.


 Tek bir cümle

Poisson(λ) “**çok sayıda nadir, (zayıf) bağımsız olayın sayısı**” dağılımıdır (binomun limit hâli); ortalama ve varyans λ , $P(\geq 1) \approx 1 - e^{-\lambda}$ ile nadir olay olasılıklarını saniyede kestirirsin.


18.6 Kontrol Soruları

 Soru 1: Saatte 3 çağrı (Poisson). Hiç çağrı yok olasılığı?

Cevap: $P(X = 0) = e^{-3} \approx 0,0498$.

 Soru 2: 100 kişide en az bir ÜÇLÜ doğum günü eşleşmesi?

Cevap: $\lambda = \binom{100}{3}/365^2 \approx 1,214$. $P(\geq 1) \approx 1 - e^{-1,214} \approx 0,70$.

 Soru 3: (Builder) Hash: 1M anahtar, 10M kova. Belirli bir kova boş?

Cevap: $\lambda = 0,1 \rightarrow P(0) = e^{-0,1} \approx 0,905$.

 Soru 4: (Builder) Her istek 1/10000 nadir hata, günde 50k istek. En az bir hata?

Cevap: $\lambda = 5 \rightarrow P(\geq 1) = 1 - e^{-5} \approx 0,993$.

18.7 Egzersizler

Egzersiz 1. Her 100 sayfada 2 hata. 50 sayfada hatasız olasılığı?

Egzersiz 2. 1000 kişi, belirli güne doğanların dağılımı + beklenti?

Egzersiz 3. 1000 oyun, $p = 1/500$. Poisson yaklaşımıyla $P(X = 0, 1, 2)$.


Egzersiz 4. (Python — Binom vs Poisson)

```
k=0: binom=0.13506  poisson=0.13534  fark=0.000271
k=1: binom=0.27067  poisson=0.27067  fark=0.000000
k=2: binom=0.27094  poisson=0.27067  fark=0.000271
k=3: binom=0.18063  poisson=0.18045  fark=0.000181
k=4: binom=0.09022  poisson=0.09022  fark=0.000000
k=5: binom=0.03602  poisson=0.03609  fark=0.000072
```

Egzersiz 5. (Sonraki ders) Sürekli RV: bir aralıktaki **her** reel sayıyı alabilen değişken. $P(X = x)$ ne olur?
Ders 12: PDF.

18.8 Sonraki Ders İçin Hazırlık

Ders 12: Kesikli vs Sürekli, Uniform — PDF, integral, ilk sürekli dağılım.

 Ders 12 öncesi yapılacak

- Egzersizleri çöz — özellikle 4.
- “ $P(\geq 1) \approx 1 - e^{-\lambda}$ ” refleksi pekiştir.
- Ana cümle: “Poisson, çok sayıda nadir olayın sayısıdır...”

18.9 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Blitzstein'de
Poisson PMF	$e^{-\lambda} \lambda^k / k!$	7m08
Poisson E	λ (varyans da)	12m04
Poisson kullanımı Paradigma	Çok fırsat, nadir olay $\approx \text{Poisson}(\sum p_j)$, zayıf bağımlıda	12m34 18m43
Binom \rightarrow Poisson	$n \rightarrow \infty, np = \lambda$	22m59
$(1 + x/n)^n \rightarrow e^x$	Limit anahtarı	29m03
$P(\geq 1) \approx 1 - e^{-\lambda}$	Nadir olay refleksi	42m01
Üçlü doğum günü	$\lambda = \binom{n}{3} / 365^2$	33m46

18.10 ML Bağlantıları Özeti

💡 6 köprü

1. **Poisson** → sayım modelleri, Poisson regresyon, topic model.
2. **Var = ortalama = λ** → overdispersion testi; ihlal → NB.
3. **Paradigma** → nadir olay sayımı, zayıf bağımlılık toleransı.
4. **Binom** → **Poisson** → küçük sayılar yasası; büyük n 'de hesaplanabilirlik.
5. $1 - e^{-\lambda}$ → hash collision, near-duplicate, nadir hata.
6. **Sympathetic magic** → model/örnek \neq dağılım.

! Tek bir şey alıp gideceksen

Poisson(λ) “çok sayıda nadir olayın sayısı”dır. Beklenen sayıyı doğrusallıkla bul, $P(\geq 1) \approx 1 - e^{-\lambda}$ ile nadir olay olasılıklarını saniyede kestir.

19 Kesikli vs Sürekli ve Uniform Dağılım

PDF, LOTUS, varyans, inverse-transform sampling

Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 12: Discrete vs. Continuous, the Uniform](#) (≈50 dk)
- **Okuma süresi:** ≈28 dk

19.1 Bu Derste Ne Var?

Büyük geçiş: **sürekli** rastgele değişkenler. Toplam yerine integral, PMF yerine PDF.

1. **PDF (olasılık yoğunluk fonksiyonu):** $P(X = x) = 0$; olasılık integralle.
2. **PDF \leftrightarrow CDF:** integral ve türev (FTC).
3. **Varyans:** $\text{Var}(X) = E(X^2) - (EX)^2$.
4. **Uniform, LOTUS, Uniform'un evrenselliği** (inverse-transform sampling).

Builder Notu — ML Köprüleri

- **PDF = yoğunluk (>1 olabilir!)** \rightarrow density estimation, **normalizing flows**, log-density.
- **Varyans \rightarrow gradyan varyansı** (eğitim kararlılığı), **bias-variance**, $\text{Var} \geq 0 \equiv E(X^2) \geq (EX)^2$ (Jensen).
- **LOTUS** $E[g(X)] = \int g(x)f(x)dx \rightarrow$ **Monte Carlo**'nun temeli; **reparameterization trick**.
- **Uniform'un evrenselliği** $X = F^{-1}(U) \rightarrow$ **inverse-transform sampling**: RNG'ler ve normalizing flows'un çekirdeği.

19.2 PDF Yoğunluktur — Olasılık Değil

Sürekli RV'de $P(X = x) = 0$. Olasılık integralle:

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

Geçerli PDF: $f(x) \geq 0$ ve $\int f = 1$. f **1'i aşabilir** (alan 1 kaldıkça sorun yok).

“the keyword here is density ... it’s not a probability, it’s a probability density.” — Blitzstein, 5:13

! Builder Notu — Yoğunluk Olasılık Değil

“PDF olasılık değil, >1 olabilir” pratikte kritik: **normalizing flow** log-density döndürür; bu pozitif/negatif olabilir, f kendisi 1’i aşar. Bir yoğunluğu olasılıkmış gibi “%120” demek tipik hatadır. **Olasılık her zaman bir aralık üzerinden integraldir.**

19.3 PDF \leftrightarrow CDF (FTC)

CDF, PDF’in integrali; PDF, CDF’in türevidir:

$$F(x) = \int_{-\infty}^x f(t) dt, \quad f(x) = F'(x)$$

Aralığın olasılığı: $P(a \leq X \leq b) = F(b) - F(a)$. Süreklide $<$ ile \leq fark etmez.

19.4 Varyans ve Standart Sapma

$$\text{Var}(X) = E[(X - EX)^2] = E(X^2) - (EX)^2$$

$$\text{SD}(X) = \sqrt{\text{Var}(X)}$$

$E(X^2) \geq (EX)^2$ her zaman ($\text{Var} \geq 0$); eşitlik yalnız X sabitse — [Ders 10 Jensen](#)’in özel hali.

💡 Builder Notu — Bias-Variance ve Gradyan Gürültüsü

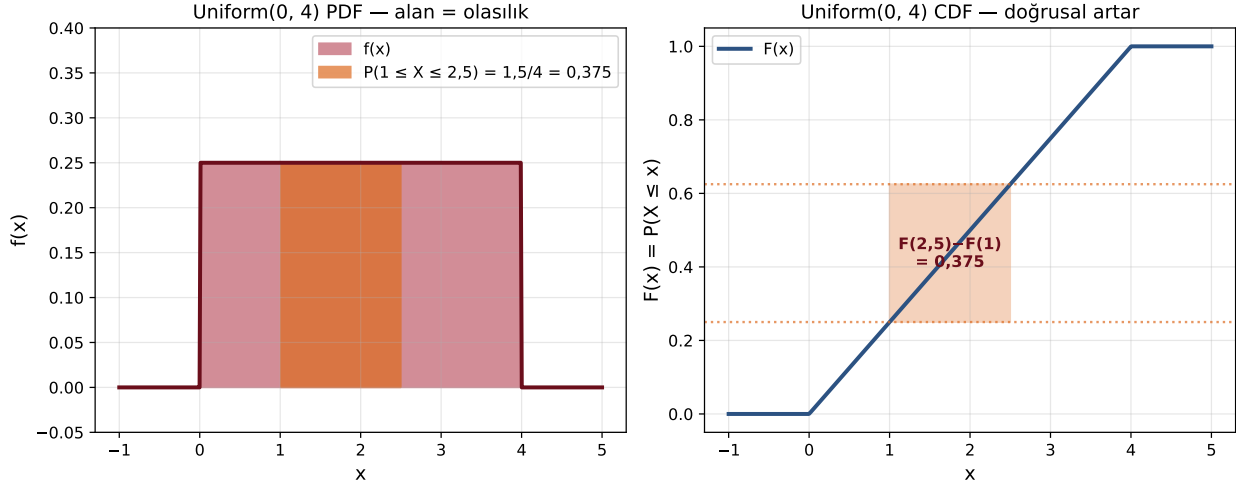
Varyans ML’de her yerde: **gradyan varyansı** (yüksekse eğitim kararsız \rightarrow variance reduction, baseline), **bias-variance ayrışımı**, minibatch gradyan gürültüsü. $\text{Var} = E(X^2) - (EX)^2$ formülü Jensen’in en sık kullanılan halidir.

19.5 Uniform Dağılım

Uniform(a, b): olasılık uzunlukla orantılı.

$$f(x) = \frac{1}{b-a}, \quad F(x) = \frac{x-a}{b-a}, \quad E(X) = \frac{a+b}{2}$$

LOTUS ile $\text{Var}(\text{Unif}(0, 1)) = E(U^2) - (EU)^2 = 1/3 - 1/4 = 1/12$.



Şekil 19.1: Uniform(0, 4) PDF (sol) ve CDF (sağ). PDF aralıkta sabit 1/4; CDF doğrusal artar 0'dan 1'e. Olasılık = PDF altındaki alan = CDF farkı.

19.6 LOTUS: $E[g(X)]$

Naif: $Y = g(X)$ 'in dağılımını bul, sonra $\int y f_Y(y) dy$. **LOTUS** der: gerek yok.

$$E[g(X)] = \int g(x) f(x) dx$$

! Builder Notu — Monte Carlo ve Reparameterization

LOTUS = Monte Carlo'nun teorik temeli: $E[g(X)]$ 'i, X 'ten örnekleyip g 'yi uygulayıp ortalayarak kestirirsin — $g(X)$ 'in dağılımını bulmana gerek yok. $E[\text{loss}]$, $E[\text{ödül}]$ böyle hesaplanır. **Reparameterization trick** de LOTUS: $X = g(\epsilon)$ yazıp $E[\text{loss}(X)] = E_\epsilon[\text{loss}(g(\epsilon))]$ ile gradyan alınır (VAE eğitiminin kalbi).

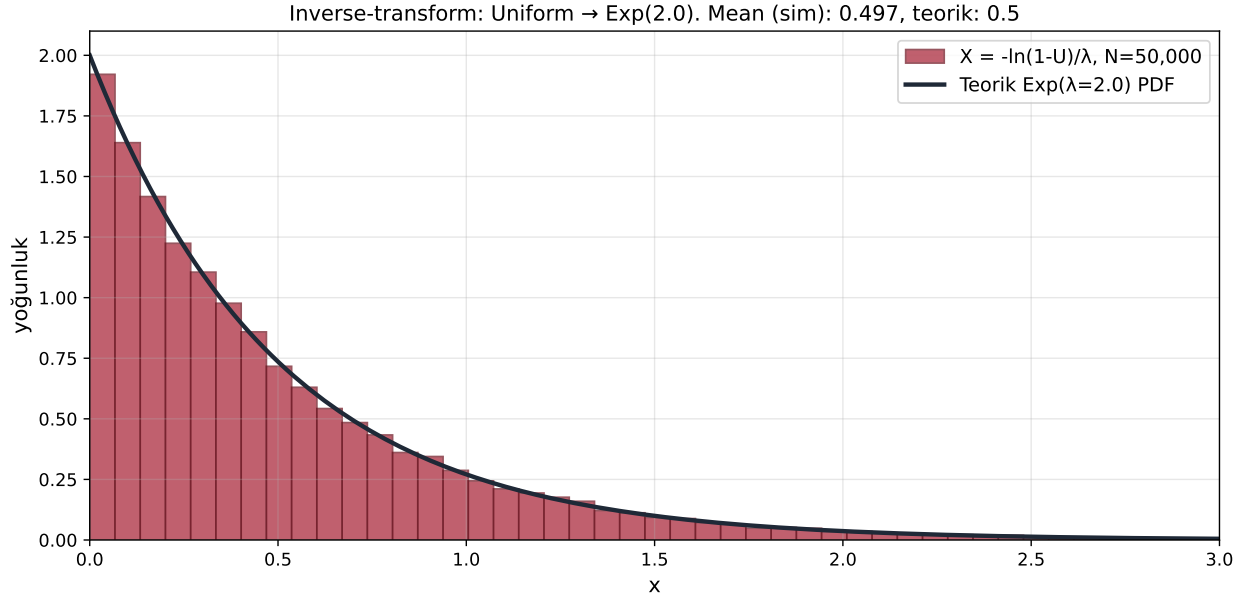
19.7 Uniform'un Evrenselliği: Inverse-Transform Sampling

Şaşırtıcı güç: tek bir Uniform(0,1)'den **istediğin dağılımı** üretebilirsin.

Teorem: F bir CDF (kesin artan, sürekli). $U \sim \text{Uniform}(0, 1)$ ile

$$X = F^{-1}(U) \Rightarrow X \sim F$$

İspat: $P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x)$.



Şekil 19.2: Inverse-transform sampling: Uniform(0,1)'den U çek, $F^{-1}(U)$ ile Exp(2) üret. Histogram teorik PDF'e oturur. Bu, herhangi bir dağılımdan örnekleme yapmanın evrensel yöntemidir.

! Builder Notu — Normalizing Flows ve Gumbel-Max

Inverse-transform sampling = tüm RNG-tabanlı örneklemin temeli ve RNG'lerin çalışma mantığı. **Normalizing flows** bunun öğrenilmiş, çok boyutlu, tersinir genellemesidir (basit tabandan örnekle, dönüşümle hedef yoğunluğa git). Kategorik için **Gumbel-max** de aynı ruhta. “Uniform'dan her şey üretilir.”

19.8 Bu Dersin Özeti

1. **Sürekli RV:** $P(X = x) = 0$; PDF $f(x)$; olasılık = integral.
2. **PDF yoğunluktur:** $f \geq 0$, $\int f = 1$, $f > 1$ olabilir.
3. **PDF \leftrightarrow CDF:** $F = \int f$, $f = F'$ (FTC).
4. **Varyans:** $E(X^2) - (EX)^2 \geq 0$; SD = $\sqrt{\text{Var}}$.
5. **Uniform(a, b):** $f = 1/(b - a)$, $E = (a + b)/2$, $\text{Var}(\text{Unif}(0, 1)) = 1/12$.
6. **LOTUS:** $E[g(X)] = \int gf dx$.
7. **Uniform'un evrenselliği:** $X = F^{-1}(U) \rightarrow$ her dağılım.

! Tek bir cümle

Süreklilikte olasılık bir noktada değil, bir **aralıkta** yaşar — PDF yoğunluktur, integral alınca olasılık olur. **LOTUS** beklentiyi dağılımı bulmadan verir (Monte Carlo'nun temeli), ve **Uniform'un evrenselliği** ($X = F^{-1}(U)$) tek bir uniform'dan her dağılımı doğurur. Bu üçü — yoğunluk, LOTUS, inverse-transform — modern üretici ML'in (VAE, normalizing flows, diffusion) omurgasıdır.

19.9 Kontrol Soruları

i Soru 1: $X \sim \text{Unif}(0, 10)$. $P(2 \leq X \leq 5)$?

Cevap: Uzunluk 3, toplam 10 $\rightarrow 0,3$.

i Soru 2: $X \sim \text{Unif}(0, 1)$. $E(X^3)$ (LOTUS)?

Cevap: $\int_0^1 x^3 \cdot 1 dx = 1/4$.

i Soru 3: $f(x) = 2$ ($0 \leq x \leq 0.5$). Geçerli PDF mi? $f(0.3) = 2$ olasılık mı?

Cevap: Geçerli ($\int = 1$). $f(0,3) = 2$ bir **yoğunluk**, olasılık değil — zaten 1'i aşıyor. $P(X = 0,3) = 0$.

i Soru 4: (Builder) $\text{Uniform}(0,1)$ 'den $\text{Exp}(\lambda)$ nasıl?

Cevap: Inverse-transform. $F(x) = 1 - e^{-\lambda x} \rightarrow F^{-1}(u) = -\ln(1 - u)/\lambda$. $X = -\ln(U)/\lambda \sim \text{Exp}(\lambda)$.

19.10 Egzersizler

Egzersiz 1. $X \sim \text{Unif}(-2, 4)$. (a) PDF, CDF. (b) $P(0 \leq X \leq 3)$? (c) $E(X)$, $\text{Var}(X)$.

Egzersiz 2. $f(x) = cx$, $0 \leq x \leq 2$. (a) c ? (b) CDF? (c) $E(X)$?

Egzersiz 3. $U \sim \text{Unif}(0, 1)$, $Y = U^2$. (a) CDF F_Y . (b) PDF f_Y . (c) $f_Y(y) \rightarrow \infty$ olabilir mi? Neden çelişki değil?

Egzersiz 4. (Python — Varyans + inverse-transform)

$\text{Var}(\text{Unif}(0,1)) \approx 0.08340$ teorik $1/12 = 0.08333$

$\text{Mean}(X) \approx 0.5006$ teorik $1/\lambda = 0.5$

KS test p-değeri: 0.314 (büyükse uyum iyi)

Egzersiz 5. (Sonraki ders) Standart Normal PDF: $\varphi(z) = e^{-z^2/2}/\sqrt{2\pi}$. (a) Kapalı-form CDF'i neden yok? Inverse-transform'u nasıl zorlaştırır? (b) **Box-Muller**: iki Uniform'dan iki Normal nasıl?

19.11 Sonraki Ders İçin Hazırlık

Ders 13: Normal Dağılım — Gauss integrali, 68-95-99,7, Φ , Box-Muller.

⚠ Ders 13 öncesi yapılacak

- Egzersizleri çöz — özellikle 4 ve 5.
- “LOTUS = Monte Carlo” ve “Uniform = her şeyin tohumu” sezgilerini pekiştir.
- Ana cümleyi tekrar oku.

19.12 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Blitzstein'de
PDF	$f \geq 0, \int f = 1, f > 1$ olabilir	5m13
CDF	$F(x) = P(X \leq x)$	5m13
FTC	$F = \int f, f = F'$	15m28
Varyans	$E(X^2) - (EX)^2$	20m24
Uniform(a, b)	$f = 1/(b - a),$ $E = (a + b)/2, \text{Var}$ $= (b - a)^2/12$	29m06
LOTUS	$E[g(X)] = \int g \cdot f$	39m38
Var(Unif(0,1))	$1/12$	39m38
Evrensellik	$X = F^{-1}(U)$	44m07

19.13 ML Bağlantıları Özeti

💡 7 köprü

1. **PDF = yoğunluk** → log-likelihood, density estimation; $f > 1$ olabilir.
2. **LOTUS** → **Monte Carlo**; $E[g(X)]$ örnekleme.
3. **Reparameterization** → LOTUS'un kendisi; VAE.
4. **Evrensellik** → **inverse-transform, normalizing flows, Gumbel-max.**
5. **Varyans** → bias-variance, gradyan gürültüsü, variance reduction.
6. **CDF türevlenebilirliği** → normalizing flows, diffusion.
7. **Uniform** → ağırlık init, keşif/baseline.

! Tek bir şey alıp gideceksen

Süreklilikte olasılık **aralıkta** yaşar. **Yoğunluk + LOTUS + inverse-transform** = modern üretici ML'in (VAE, normalizing flows, diffusion) olasılıksal omurgası.

20 Normal Dağılım

İstatistiğin yıldızı: Gauss integrali, Φ , CLT habercisi

Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 13: Normal Distribution](#) (≈51 dk)
- **Okuma süresi:** ≈32 dk

20.1 Bu Derste Ne Var?

1. **Probability integral transform:** $X \sim F \Rightarrow F(X) \sim \text{Uniform}(0, 1)$.
2. **Uniform dönüşümleri:** $1 - U, a + bU$ Uniform kalır; doğrusal-olmayan (örn. U^2) bozar.
3. **RV bağımsızlığı:** joint = marjinalerin çarpımı. İkili \neq tam.
4. **Standart Normal $N(\mathbf{0}, \mathbf{1})$:** PDF, Gauss integrali ($\sqrt{2\pi}$), $E = 0$, $\text{Var} = 1$, Φ .

“from the uniform you can get everything.” — Blitzstein, 4:38

Builder Notu — ML Köprüleri

- $F(X) \sim \text{Uniform} \rightarrow$ model **kalibrasyon kontrolü**, QQ-plot, copula.
- **Bağımsızlık = faktörizasyon** \rightarrow PGM, naive Bayes.
- **İkili \neq tam bağımsızlık** \rightarrow XOR-tipi gizli yapı; korelasyon matrisine güvenme.
- **Normal + CLT** \rightarrow Gaussian her yerde: ağırlık init, VAE prior $N(\mathbf{0}, \mathbf{I})$, **diffusion** taban, batch norm.
- $\Phi \rightarrow$ **GELU**: $\text{GELU}(x) = x \cdot \Phi(x)$, transformer aktivasyonu.
- **Gauss normalleştirme** \rightarrow **partition function**; softmax, enerji-tabanlı modeller.

20.2 Probability Integral Transform

Uniform'un evrenselliğinin **ters yönü**:

$$X \sim F \implies F(X) \sim \text{Uniform}(0, 1)$$

Notasyon tuzağı: $F(X) \neq P(X \leq X) = 1$. Önce F 'i fonksiyon olarak yaz, sonra X koy. Örn. $F(x) = 1 - e^{-x}$ ise $F(X) = 1 - e^{-X}$.

💡 Builder Notu — Kalibrasyon

Probability integral transform model testinin temeli: X 'in karmaşık dağılımı varsa, $F(X)$ 'leri hesapla, **Uniform görünüp görünmediğini** kontrol et. ML'de **kalibrasyon eğrileri**, QQ-plot, **copula**'lar bu fikre dayanır.

20.3 Uniform Simetrisi ve Doğrusal-Olmayan Tuzak

- $1 - U \sim \text{Uniform}(0, 1)$ (simetri).
- $a + bU \sim \text{Uniform}(a, a + b)$ (doğrusal).
- U^2 **Uniform DEĞİL**: $P(U^2 \leq y) = P(U \leq \sqrt{y}) = \sqrt{y} \neq y$.

“Nonlinear usually leads to nonuniform.” — Blitzstein, 14:51

💡 Builder Notu — Change of Variables

Normalizing flows: basit tabandan, doğrusal-olmayan tersinir dönüşümlerle karmaşık yoğunluklar. Jacobian ile yoğunluk yeniden şekillenir.

20.4 RV Bağımsızlığı: İkili \neq Tam

X_1, \dots, X_n bağımsız \Leftrightarrow joint = marjinalerin çarpımı:

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod_i P(X_i \leq x_i)$$

Matching pennies: $X_1, X_2 \sim \text{Bern}(1/2)$ IID, $X_3 = \mathbb{1}\{X_1 = X_2\}$. **İkili bağımsız** (X_1, X_3 kontrol et: $P = 1/4 = 1/2 \cdot 1/2$) ama **tam bağımsız değil** (X_1, X_2 bilinince X_3 belli).

❗ Builder Notu — Korelasyon Matrisine Güvenme

İkili korelasyonsuzluk tam bağımsızlığı garanti etmez — XOR-tipi yapı ikili testlerden kaçır. Sadece korelasyon matrisine bakıp “bağımsız” demek yanlış model varsayımdır.

20.5 Standart Normal $N(0,1)$: PDF ve Gauss İntegrali

$$f(z) = c e^{-z^2/2}$$

Normalleştirme sabiti c için ünlü **Gauss integrali**:

$$I = \int_{-\infty}^{\infty} e^{-z^2/2} dz$$

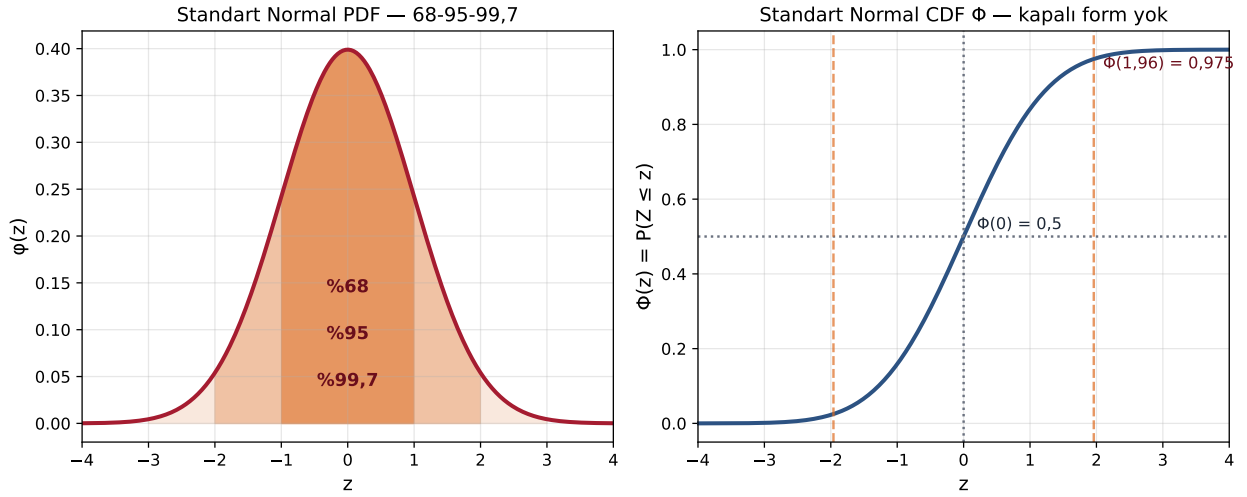
Numara: Kareyi al + kutupsal koordinat. $I^2 = \iint e^{-(x^2+y^2)/2} dx dy$. Polar: $r^2 = x^2 + y^2$, Jacobian r :

$$I^2 = \int_0^{2\pi} \int_0^{\infty} e^{-r^2/2} r dr d\theta = \int_0^{2\pi} 1 d\theta = 2\pi$$

Demek $I = \sqrt{2\pi}$, $c = 1/\sqrt{2\pi}$.

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

“where did the pi come from? Where did the circle come?” — Blitzstein, 40:21



Şekil 20.1: Standart Normal $N(0,1)$ PDF (sol) ve CDF Φ (sağ). 68-95-99,7 kuralı: %68 $[-1,1]$, %95 $[-2,2]$, %99,7 $[-3,3]$ arasında. PDF tepe = $1/\sqrt{2\pi} \approx 0,399$; σ küçülürse 1’i aşar (yoğunluk!).

20.6 $E(Z) = 0$, $\text{Var}(Z) = 1$

Ortalama: $z \cdot e^{-z^2/2}$ tek fonksiyon, simetrik aralıkta integral = 0:

$$E(Z) = 0$$

Varyans: $E(Z^2)$ için LOTUS + parçalı integrasyon ($u = z$, $dv = ze^{-z^2/2} dz$):

$$E(Z^2) = \frac{1}{\sqrt{2\pi}} \int z^2 e^{-z^2/2} dz = 1$$

$$\text{Var}(Z) = 1$$

20.7 Φ Notasyonu

Standart Normal CDF'in **kapalı formu yoktur** — adı Φ :

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt$$

Simetri: $\Phi(-z) = 1 - \Phi(z)$.

! Builder Notu — GELU ve Probit

Φ ML'de doğrudan: **GELU** aktivasyonu = $x \cdot \Phi(x)$ (transformer'larda yaygın). **Probit regresyon** bağlantı fonksiyonu olarak Φ kullanır (sigmoid'in Normal karşılığı). Φ ile **erf** birbirinin yeniden ölçeklenmiş hâli.

20.8 Bu Dersin Özeti

1. **PIT:** $F(X) \sim$ Uniform; notasyon tuzağı.
2. **Uniform:** $1 - U$, $a + bU$ Uniform kalır; U^2 değil.
3. **Bağımsızlık:** joint = \prod marjinal; ikili \neq tam.
4. **Normal + CLT:** iid toplam \rightarrow Normal.
5. **N(0,1):** $f = e^{-z^2/2}/\sqrt{2\pi}$; Gauss integrali polar + kareyi al $\rightarrow \sqrt{2\pi}$.
6. **E(Z) = 0** (simetri), **Var(Z) = 1** (LOTUS + parçalı).
7. **Φ :** kapalı formsuz; $\Phi(-z) = 1 - \Phi(z)$.

! Tek bir cümle

Uniform'un evrenselliği bir dağılımı kendi uniform'una bağlayan **iki yönlü köprüdür** (örnekleme + kalibrasyon). Standart Normal $N(0, 1)$ 'in PDF'i $(1/\sqrt{2\pi})e^{-z^2/2}$ — normalleştirme sabiti Gauss integralinden çembersel numaraya $\sqrt{2\pi}$ doğar; $E = 0$, $\text{Var} = 1$, $\text{CDF} = \Phi$. **GELU'dan VAE'ye, diffusion'a, batch norm'a kadar uzanan dağılım.**

20.9 Kontrol Soruları

i Soru 1: $F(x) = x^2$ ($0 \leq x \leq 1$). $U \rightarrow X$ nasıl?

Cevap: $F^{-1}(u) = \sqrt{u}$. $X = \sqrt{U}$.

i Soru 2: $X \sim \text{Exp}(1)$. $Y = 1 - e^{-X}$ dağılımı?

Cevap: PIT: $Y = F(X) \sim \text{Uniform}(0, 1)$.

i Soru 3: Matching pennies X_1, X_3 ikili bağımsız mı, üçü tam mı?

Cevap: $P(X_1 = 1, X_3 = 1) = P(X_1 = 1, X_2 = 1) = 1/4 = (1/2)(1/2) \checkmark$. Ama $P(X_1 = 1, X_2 = 1, X_3 = 0) = 0 \neq 1/8 \rightarrow$ tam bağımsız değil.

i Soru 4: (Builder) VAE'de $z \sim N(0,1)$. Φ^{-1} neden kullanışsız? Box-Muller?

Cevap: Φ kapalı formsuz $\rightarrow \Phi^{-1}$ analitik yazılamaz. **Box-Muller:** iki Uniform'dan iki bağımsız Normal: $Z_1 = \sqrt{-2 \ln U_1} \cos(2\pi U_2)$, $Z_2 = \sin(\dots)$. Polar yapı (Gauss integralindeki!) Normal örneklemeyi kapalı formda mümkün kılar.

20.10 Egzersizler

Egzersiz 1. Lojistik $F(x) = 1/(1 + e^{-x})$. (a) F^{-1} ? (b) U'dan örnek formülü.

Egzersiz 2. PIT iki yön: (a) $F(X) \sim \text{Uniform}$. (b) $1 - F(X) \sim \text{Uniform}$. (c) Model kontrolünde nasıl?

Egzersiz 3. $Z \sim N(0,1)$. (a) $E(Z^3)$? (b) $E(Z^4) = 3$? (c) Tek momentler neden 0?

Egzersiz 4. (Python — Gauss integrali + Box-Muller)

Gauss integrali: 2.5066 $\text{sqrt}(2\pi) = 2.5066$

Box-Muller mean: 0.0021 var: 1.0022

KS p-değeri: 0.424

Egzersiz 5. (Sonraki ders) $X = \mu + \sigma Z$. (a) $E(X)$, $\text{Var}(X)$? (b) Change of variables ile PDF. (c) CDF Φ cinsinden.

20.11 Sonraki Ders İçin Hazırlık

Ders 14: Konum, Ölçek ve LOTUS — $X = \mu + \sigma Z$, 68-95-99,7, batch norm matematiği.

⚠ Ders 14 öncesi yapılacak

- Egzersizleri çöz, özellikle 5 ($\mu + \sigma Z$).
- Python'da σ küçüldükçe $\varphi(0) > 1$ olduğunu (yoğunluk) gör.
- Ana cümleyi tekrar oku.

20.12 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Blitzstein'de
PIT iki yön	$X = F^{-1}(U); F(X) \sim \text{Unif}$	2m09
Notasyon tuzağı	$F(X) \neq P(X \leq X) = 1$	7m29
1 - U simetrisi	$-\ln(U) \sim \text{Exp}(1)$	13m17
Doğrusal-olmayan	U^2 Uniform değil	14m36
RV bağımsızlığı	joint = \prod marjinal	16m00
İkili \neq tam	Matching pennies	20m16
Normal + CLT	iid toplam \rightarrow Normal	25m42
N(0,1) PDF	$e^{-z^2/2}/\sqrt{2\pi}$	27m20
Gauss integrali	$\sqrt{2\pi}$; polar + kare	33m19
E, Var	0, 1	41m16
Φ	Kapalı formsuz; $\Phi(-z) = 1 - \Phi(z)$	49m42

20.13 ML Bağlantıları Özeti

💡 7 köprü

1. **PIT** \rightarrow örnekleme + kalibrasyon, QQ-plot, copula.
2. **Bağımsızlık = faktörizasyon** \rightarrow PGM, naive Bayes.
3. **İkili \neq tam** \rightarrow XOR-tipi gizli yapı; korelasyon \neq bağımsızlık.
4. **Normal + CLT** \rightarrow Gaussian her yerde: VAE prior, **diffusion**, ağırlık init.
5. **Gauss normalleştirme** \rightarrow **partition function**; softmax, EBM.
6. **Φ** \rightarrow GELU, probit, erf.
7. **Box-Muller** \rightarrow polar numara yeniden kullanımı.

❗ Tek bir şey alıp gideceksen

Uniform'un evrenselliği iki yönlü köprü (örnekleme + kalibrasyon). $N(0, 1)$ 'in $\sqrt{2\pi}$ 'si **çembersel numarayla** doğar; Φ kapalı formsuz ama her yerde. **GELU'dan diffusion'a, batch norm'a kadar uzanan dağılım.**

21 Konum, Ölçek ve LOTUS

$X = \mu + \sigma Z$, varyans kuralları, Poisson(λ) ve Binom(npq) varyansı

Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 14: Location, Scale, and LOTUS](#) (≈ 49 dk)
- **Okuma süresi:** ≈ 34 dk

21.1 Bu Derste Ne Var?

1. **Konum-ölçek:** $X = \mu + \sigma Z$ ile her Normal standart Normal'e indirgenir.
2. **Varyansın özellikleri:** $\text{Var}(X + c) = \text{Var}(X)$; $\text{Var}(cX) = c^2 \text{Var}(X)$; doğrusal değil.
3. **Bağımsız Normal toplamı = Normal** (kapanış).
4. **LOTUS ile varyans:** Poisson(λ) \rightarrow Var = λ ; Binom(n, p) \rightarrow Var = npq .

Builder Notu — ML Köprüleri

- **Standardizasyon \rightarrow batch / layer normalization;** feature scaling; z-skoru boyutsuz.
- **$\text{Var}(cX) = c^2 \text{Var}(X)$ \rightarrow gradyan/gürültü ölçeklemesi;** $\text{Var}(\bar{X}) = \sigma^2/n$.
- **Bağımsız Normal kapanışı \rightarrow diffusion** ileri süreci tek hamlede; Kalman, GP.
- **LOTUS ispatı = Monte Carlo garantisi.**

21.2 Konum-Ölçek: $X = \mu + \sigma Z$

$$X = \mu + \sigma Z, \quad Z \sim N(0, 1), \quad \sigma > 0 \Rightarrow X \sim N(\mu, \sigma^2)$$

- μ konum (kaydırma).
- σ ölçek (genişlik).

Standardizasyon: $Z = (X - \mu)/\sigma \sim N(0, 1)$. **Boyutsuz** (birimler sadeleşir).

Builder Notu — Batch Norm

Batch / layer normalization tam olarak standardizasyon: aktivasyonları $(x - \mu)/\sigma$ ile normalize, sonra öğrenilen γ, β ile yeniden ölçekle. **VAE reparameterization:** $x = \mu + \sigma \epsilon, \epsilon \sim N(0, 1)$.

21.3 Varyansın Özellikleri

$$\text{Var}(X + c) = \text{Var}(X), \quad \text{Var}(cX) = c^2 \text{Var}(X)$$

Kareyi unutma — c negatifse karesiz negatif varyans çıkar! Varyans ≥ 0 her zaman; $= 0 \Leftrightarrow X$ sabit.

Doğrusal değil:

$$\text{Var}(X + Y) \neq \text{Var}(X) + \text{Var}(Y) \quad (\text{genelde})$$

Eşitlik **bağımsızsa**. X kendisiyle bağımsız değil:

$$\text{Var}(X + X) = \text{Var}(2X) = 4 \text{Var}(X)$$

“*x is not IID with itself.*” — Blitzstein, 11:46

! Builder Notu — Negatif Varyans Yok

“Negatif varyans = bir yerde hata” sanity check. Kovaryans matrislerinin **pozitif yarı-tanımlı** olması gerektiğinin habercisidir. Minibatch ortalamasının $\text{Var}(\bar{X}) = \sigma^2/n$ — gradyan gürültüsünün n ile küçülmesinin temeli.

21.4 Genel Normal PDF (Zincir Kuralı)

$$F(x) = P(X \leq x) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

Türev (zincir kuralı):

$$f(x) = \frac{1}{\sigma} \varphi\left(\frac{x - \mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

Ezberleme — standardizasyon + zincir kuralıyla iki satırda türet.

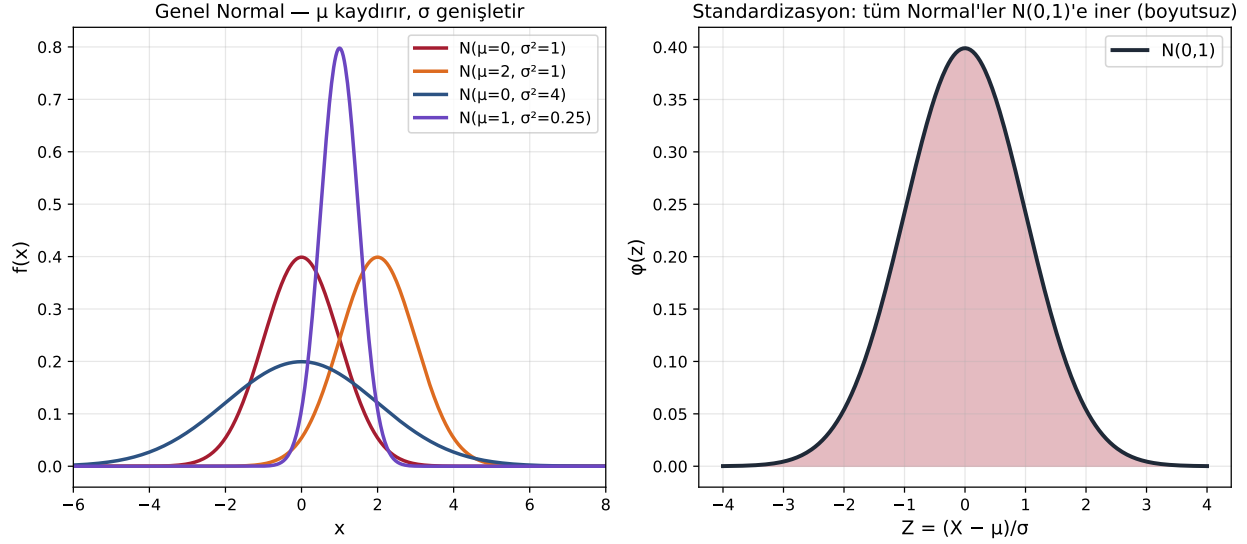
21.5 Bağımsız Normal Toplamı = Normal

$X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$ bağımsız:

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Fark için de varyanslar TOPLANIR (çünkü $-X_2$ ’nin varyansı hâlâ σ_2^2):

$$X_1 - X_2 \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$$



Şekil 21.1: Genel Normal $N(\mu, \sigma^2)$ farklı parametrelerle. σ küçüldükçe tepe yükselir (yoğunluk 1'i aşar!), σ büyüdüğüçe yayılır. Standardizasyon (sağ) hepsini $N(0,1)$ 'e indirger — batch norm'un olasılıksal temeli.

! Builder Notu — Diffusion Kapanışı

Normal'in kapanışı ML'de muazzam: Gaussian + Gaussian = Gaussian. **Diffusion ileri süreci** her adımda bağımsız Gaussian gürültü ekler; varyanslar biriktiğinden **herhangi bir gürültü seviyesine tek hamlede atlanabilir**. Kalman filtreleri, Gaussian process'ler hep bu kapanışa dayanır.

21.6 68-95-99,7 Kuralı

$$P(|X - \mu| < \sigma) \approx 68\%, \quad P(|X - \mu| < 2\sigma) \approx 95\%, \quad P(|X - \mu| < 3\sigma) \approx 99,7\%$$

💡 Builder Notu — Outlier ve Sigma Dili

Aykırı değer tespiti ve güven aralıkları: 3σ dışı şüpheli (%0,3). Fizik keşiflerinde “ 5σ ” dili, süreç kontrolü, anomali tespiti. **Yalnızca Normal için**; ağır kuyrukta yanlıtır.

21.7 LOTUS ile Varyans: Poisson ve Binom

Poisson(λ) varyansı. Faktöriyel moment $E(X(X-1))$ ile temiz yol:

$$E(X(X-1)) = \sum k(k-1) \frac{e^{-\lambda} \lambda^k}{k!} = e^{-\lambda} \lambda^2 \sum_{j \geq 0} \frac{\lambda^j}{j!} = \lambda^2$$

21 Konum, Ölçek ve LOTUS

$$E(X^2) = \lambda^2 + \lambda \rightarrow \text{Var}(X) = \lambda.$$


Poisson: ortalama = varyans = λ .

Binom(n, p) varyansı (göstergeler): $X = I_1 + \dots + I_n$.

$$E(X^2) = nE(I_1^2) + n(n-1)E(I_1 I_2) = np + n(n-1)p^2$$

$E(I_1^2) = p$ (gösterge!), $E(I_1 I_2) = p^2$ (bağımsız). Çıkarınca:

$$\text{Var}(X) = np - np^2 = np(1-p) = npq$$

 **Builder Notu — Overdispersion ve A/B Test**

Poisson: ortalama = varyans sayım modellerinin temel varsayımı. Varyans > ortalama (**overdispersion**)
→ negatif binom. **A/B testlerinde** dönüşüm oranı belirsizliği \sqrt{npq} .

21.8 LOTUS İspatı (Kesikli)

$$\sum_x g(x)P(X=x) = \sum_s g(X(s))P(s)$$

Sağdan başla, $X(s) = x$ olan tüm s 'leri grupla:

$$\sum_s g(X(s))P(s) = \sum_x \sum_{s: X(s)=x} g(x)P(s) = \sum_x g(x)P(X=x)$$

Gruplu = grupsuz.

 **Builder Notu — Monte Carlo Garantisi**

LOTUS ispatı **Monte Carlo'nun teorik garantisi**: $E[g(X)]$ 'i, X 'ten örnekleyip g 'yi ortalayarak kestirmek geçerlidir. $E[\text{loss}]$, $E[\text{ödül}]$, **policy gradient** hep bu eşitliğe dayanır — $g(X)$ 'in dağılımını bilmeden.

21.9 Bu Dersin Özeti

1. **Konum-ölçek**: $X = \mu + \sigma Z$, standardizasyon $Z = (X - \mu)/\sigma$.
2. **Varyans**: $\text{Var}(cX) = c^2 \text{Var}(X)$; ≥ 0 ; doğrusal değil; $\text{Var}(X + X) = 4 \text{Var}(X)$.
3. **Genel Normal**: $E = \mu$, $\text{Var} = \sigma^2$; PDF zincir kuralıyla.
4. **Kapanış**: bağımsız Normal toplamı Normal (varyanslar toplanır, farkta bile).
5. **68-95-99,7** kuralı.
6. **Poisson**: $\text{Var} = \lambda$. **Binom**: $\text{Var} = npq$.
7. **LOTUS ispatı**: gruplu = grupsuz toplam.

! Tek bir cümle

Her Normal tek bir standart Normal'in kaydırılıp ölçeklenmiş hâlidir ($X = \mu + \sigma Z$); varyans sabitle **kareyle** ölçeklenir, doğrusal değildir (yalnız bağımsızlıkta toplanır); ve **LOTUS** (gruplu = grupsuz toplam) hem Poisson'un (λ) hem Binom'un (npq) varyansını verir hem de **modern Monte Carlo'nun teorik temelidir**.

21.10 Kontrol Soruları

i Soru 1: $X \sim N(5, 4)$. (a) $\text{Var}(3X-1)$? (b) Standardize. (c) $P(1 < X < 9)$?

Cevap: (a) $9 \cdot 4 = 36$. (b) $Z = (X - 5)/2$. (c) $1 = \mu - 2\sigma, 9 = \mu + 2\sigma \rightarrow \approx 95\%$.

i Soru 2: X, Y bağımsız. $\text{Var}(X-Y)$? $\text{Var}(X-X)$ neden 0?

Cevap: $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$. $X - X = 0$, sabit $\rightarrow \text{Var} = 0$. X kendisiyle bağımsız değil!

i Soru 3: $X \sim \text{Pois}(\lambda)$. $E(X(X-1))$?

Cevap: Faktöriyel moment $\rightarrow \lambda^2$. $E(X^2) = \lambda^2 + \lambda$, $\text{Var}(X) = \lambda$.

i Soru 4: (Builder) Diffusion: $x \sim N(\mu, \sigma^2)$, $\varepsilon \sim N(0, \tau^2)$ bağımsız. $x + \varepsilon$?

Cevap: $N(\mu, \sigma^2 + \tau^2)$. Kapanış sayesinde t adım sonraki dağılım kapalı formda, **tek hamlede atlanabilir**.

21.11 Egzersizler

Egzersiz 1. $X \sim N(-2, 9)$. (a) PDF aç. (b) Standardize. (c) $P(-5 < X < 1)$? (d) $\text{Var}(2X - 7)$?

Egzersiz 2. $\text{Var}(X) = 4$, $\text{Var}(Y) = 9$ bağımsız. (a) $\text{Var}(2X + 3Y)$? (b) $\text{Var}(X - Y)$? (c) $\text{Var}(7)$?

Egzersiz 3. $X \sim \text{Geom}(p)$. Türev numarasıyla $E(X^2)$, sonra $\text{Var}(X) = q/p^2$.

Egzersiz 4. (Python — Var doğrulama)

`Var(Pois(4.0)) sim: 4.011 teorik: 4.0`

`Var(Bin(20,0.3)) sim: 4.179 teorik npq: 4.1999999999999999`

`P(|Z|<1) ≈ 0.6830`

`P(|Z|<2) ≈ 0.9547`


`P(|Z|<3) ≈ 0.9973`

`N(1,4)+N(-2,9) → mean=-1.00 (teori -1), var=13.01 (teori 13)`

Egzersiz 5. (Sonraki ders) Genel formül $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$. (a) Türet. (b) Bağımsızlıkta $\text{Cov} = 0 \rightarrow$ varyans toplamı. (c) $\text{Var}(X + X) = 4\text{Var}(X)$, $\text{Cov}(X, X) = \text{Var}(X)$.

21.12 Sonraki Ders İçin Hazırlık

Ders 15: Ara Sınav İncelemesi — yeni konu yok, tekrar oturumu (Ders 1-14).


 Ders 15 öncesi yapılacak

- Ders 1-14 “Tek Bir Cümle” + Cheat Sheet’leri arka arkaya oku.
- Bernoulli/Binom/Poisson/Geometrik/Uniform/Normal — ortalama + varyans ezberden.
- Egzersizleri çöz, özellikle 5 (Cov habercisi).

21.13 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Blitzstein’de
Konum-ölçek	$X = \mu + \sigma Z \Rightarrow N(\mu, \sigma^2)$	4m27
Standardizasyon	$Z = (X - \mu)/\sigma$; boyutsuz	13m15
Var($X + c$)	$= \text{Var}(X)$	7m43
Var(cX)	$= c^2 \text{Var}(X)$	8m09
Var ≥ 0	$= 0 \iff X$ sabit	8m56
Var doğrusal değil	Yalnız bağımsızsa toplanır	9m58
Genel Normal PDF	$(1/\sigma)\varphi((x - \mu)/\sigma)$	16m06
Bağımsız Normal toplam	$N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$; farkta bile	18m37
68-95-99,7	$\pm\sigma, \pm 2\sigma, \pm 3\sigma$	20m32
Var(Poisson λ)	λ	31m17
Var(Binom n, p)	npq	40m21
LOTUS ispatı	Gruplu = grupsuz	44m06

21.14 ML Bağlantıları Özeti

 7 köprü

1. **Standardizasyon** \rightarrow **batch / layer normalization**, feature scaling.
2. **Reparameterization** \rightarrow **VAE**: $x = \mu + \sigma\epsilon$.
3. **Var(cX) = c^2 Var** \rightarrow gradyan ölçekleme; $\text{Var}(\bar{X}) = \sigma^2/n$.
4. **Kapanış** \rightarrow **diffusion**, Kalman, GP.

5. **68-95-99,7** → outlier, güven aralıęı, sigma dili.
6. **Poisson Var** = λ → overdispersion → NB; A/B test \sqrt{npq} .
7. **LOTUS ispatı** → **Monte Carlo garantisi**; policy gradient.

! Tek bir şey alıp gideceksen

Tüm Normal dünyası tek bir şablonun (Z) kaydırılıp ölçeklenmesi ($X = \mu + \sigma Z$). Varyans **kareyle** ölçeklenir, doğrusal değildir, yalnız bağımsızlıkta toplanır. **LOTUS** modern Monte Carlo'nun temelidir.

22 Ara Sınav İncelemesi

Kupon toplayıcı, logit, simetri, Poisson→Üstel

Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 15: Midterm Review](#) (≈38 dk)
- **Okuma süresi:** ≈30 dk
- **Not:** Yeni teori yok — Ders 1-14'ün araçlarını 6 problemde canlandırıyor.

22.1 Bu Derste Ne Var?

Altı problem:

1. **Kupon toplayıcı** ($nH_n \approx n \ln n$).
2. **Uniform'un evrenselliği** — geometrik sezgi.
3. **Lojistik dağılım** — logit ile örnekleme.
4. **Simetri + doğrusallık** — $E(X/(X + Y + Z)) = 1/3$.
5. **LOTUS tuzağı** — örüntü, isim değil.
6. **Story proof + Poisson → Üstel** — sürekli bekleme.

Builder Notu — ML Köprüleri

- **Kupon toplayıcı** ($n \ln n$) → keşif/kapsama, örneklem karmaşıklığı.
- **Logit / sigmoid** → **logistic regression**, sınıflandırma başlığı.
- **Simetri (exchangeability)** → i.i.d.'de eşit katkı (Shapley benzeri).
- **Poisson → Üstel** → bekleme süreleri, kuyruk teorisi, **Poisson süreci**.

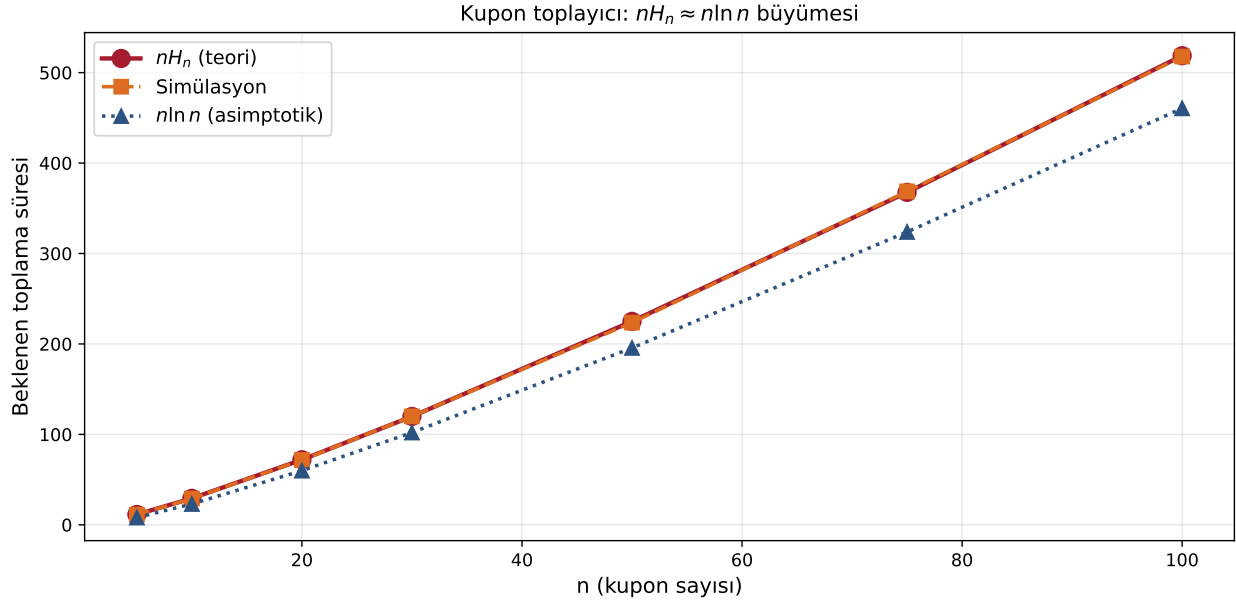
22.2 Kupon Toplayıcı

n farklı kupon. Tam seti toplama süresi T . Parçala:

$$T = T_1 + T_2 + \dots + T_n$$

$T_j = (j - 1)$. yeni'den j . yeni'ye kadar. $T_j - 1 \sim \text{Geom}((n - j + 1)/n)$, $E(T_j) = n/(n - j + 1)$:

$$E(T) = \sum_{j=1}^n \frac{n}{n-j+1} = n \sum_{k=1}^n \frac{1}{k} = nH_n \approx n \ln n$$



Şekil 22.1: Kupon toplayıcı: nH_n teori vs simülasyon. $n=10 \rightarrow \sim 29.3$, $n=50 \rightarrow \sim 225$, $n=100 \rightarrow \sim 519$. $n \ln n$ büyümesi: tüm türleri görmek herhangi birini görmekten çok daha pahalı.

💡 Builder Notu — Keşif Maliyeti

Kupon toplayıcı **ML/CS'de örneklem karmaşıklığının** klasik modeli: tüm n şeyi görmek herhangi birini görmekten **logaritmik faktör** kadar pahalı. RL keşfi, distinct-eleman tahmini, cache/streaming algoritmaları hep bu $n \ln n$ ölçeğinde.

22.3 Uniform'un Evrenselliği: Geometrik Sezgi

$P(F(X) \leq c) = P(X \leq x_c) = F(x_c) = c$. Çünkü F artan, olay $F(X) \leq c$ ile $X \leq x_c$ aynı.

Bu, **kalibrasyon eğrileri** ve PIT testlerinin temeli.

22.4 Lojistik Dağılım ve Logit

CDF: $F(x) = e^x / (1 + e^x) = \text{sigmoid}$. F^{-1} :

$$u = \frac{e^x}{1 + e^x} \Rightarrow x = \log \frac{u}{1 - u} = \text{logit}(u)$$

$U \sim \text{Uniform}(0, 1)$, $X = \log(U/(1 - U)) \sim \text{Logistic}$.

! Builder Notu — Logistic Regression

Lojistik CDF **sigmoid**'in ta kendisi; tersi **logit**. **Logistic regression** log-odds'u doğrusal modeller, sigmoid olasılığı verir. Tüm ikili sınıflandırma başlıkları bu dağılımdan gelir.

22.5 Simetri + Doğrusallık

X, Y, Z iid pozitif:

$$3E\left(\frac{X}{X+Y+Z}\right) = E\left(\frac{X+Y+Z}{X+Y+Z}\right) = 1 \Rightarrow E\left(\frac{X}{X+Y+Z}\right) = \frac{1}{3}$$

Hiç integral yok — sadece simetri + doğrusallık.

💡 Builder Notu — Exchangeability

“**Simetriyle topla, doğrusallıkla çöz**” tekniği **exchangeable** değişkenlerde her yerde: dikkat ağırlıklarında, kaynak paylaşımında, **Shapley değeri** atıflarında hesabı kısaltır.

22.6 LOTUS Tuzağı: İki Yol, Bir Cevap

$U \sim \text{Uniform}(0,1)$, $X = U^2$, $Y = e^X$.

Yol 1 (X üzerinden): $f_X(x) = \frac{1}{2}x^{-1/2}$, $E(Y) = \int_0^1 e^x f_X(x) dx$.

Yol 2 (U üzerinden, doğrudan): $Y = e^{U^2}$, $E(Y) = \int_0^1 e^{u^2} du$.

İkisi de doğru, aynı sayı. **Önemli olan örüntü, isim değil.**

💡 Builder Notu — Monte Carlo Sadeleştirme

Yol 2 = Monte Carlo'nun doğal biçimi: $E[g(U)]$ 'yu **en alttaki basit değişken** üzerinden örnekle, ara dağılımları (X) bulmaya gerek yok. Reparameterization zincirini sadeleştirir.

22.7 Story Proof: $n - X \sim \text{Bin}(n, q)$

$X \sim \text{Bin}(n, p)$. $n - X$ ne?

Story: X = başarı sayısı $\rightarrow n - X$ = başarısızlık sayısı. Başarı/başarısızlık etiketlerini değiştir, başarısızlık olasılığı q :

$$n - X \sim \text{Bin}(n, q), \quad q = 1 - p$$

Hiç hesap yok.

22.8 Poisson → Üstel: İlk Bekleme Süresi

$N_t \sim \text{Pois}(\lambda t) = [0, t]$ aralığındaki olay sayısı. T = ilk olayın zamanı.

$$P(T > t) = P(N_t = 0) = e^{-\lambda t}$$

$$F(t) = 1 - e^{-\lambda t}, \quad f(t) = \lambda e^{-\lambda t}, \quad t > 0$$

İşte **üstel (exponential) dağılım** — Poisson sayımından (kesikli) sürekli bekleme süresi doğdu.

! Builder Notu — Poisson Süreci

Poisson sürecinin kalbi: olay sayısı Poisson, aralarındaki bekleme Üstel. Olay-tabanlı simülasyon, kuyruk teorisi (istek varışları), sağkalım analizi, “memoryless” varsayımlı modeller.

22.9 Üç Nesneyi Karıştırma

- **Dağılım** (plan) — “ev planı”.
- **Rastgele değişken** — “rastgele ev”.
- **Sabit** ($E(X)$ gibi) — “belirli ev”.

Kodda: `Normal(0,1)` (dist nesnesi) \neq `dist.sample()` (tensor) \neq `dist.mean` (skaler).

22.10 Bu Dersin Özeti

1. **Kupon toplayıcı:** $E(T) = nH_n \approx n \ln n$; parçala + doğrusallık.
2. **Universality sezgi:** $P(F(X) \leq c) = c$.
3. **Lojistik:** $F^{-1}(u) = \text{logit}(u) = \log(u/(1-u))$.
4. **Simetri:** iid $\rightarrow E(X/(X+Y+Z)) = 1/3$.
5. **LOTUS:** $E(g(U))$ doğrudan U üzerinden.
6. **Story proof:** $n - X \sim \text{Bin}(n, q)$.
7. **Poisson → Üstel:** $T \sim \text{Exp}(\lambda)$.

! Tek bir cümle

Zor bir problemi araçlarının tanıdığı parçalara böl — kupon toplayıcıyı geometriklere, simetrik oranı eşit terimlere, sürekli bekleme süresini Poisson sayım olayına. **Ezber formül değil, yapıyı görmek kazandırır.**

22.11 Kontrol Soruları

i Soru 1: Zar atışı, 6 yüzün hepsini görmek?

Cevap: $E(T) = 6H_6 = 6 \cdot 49/20 = 14,7$ atış.

i Soru 2: $F(x) = 1 - e^{-(x^2)}$. $U \rightarrow X$?

Cevap: $X = \sqrt{-\ln(1-U)}$.

i Soru 3: X, Y, Z, W iid pozitif. (a) $E(X/\Sigma)$? (b) $E((X+Y)/\Sigma)$?

Cevap: (a) $1/4$. (b) $1/2$.

i Soru 4: (Builder) $\lambda=3/\text{saat}$ Poisson. (a) İlk 30 dk hiç? (b) İlk e-postaya kadar E ?

Cevap: (a) $e^{-1,5} \approx 0,223$. (b) $T \sim \text{Exp}(3)$, $E = 1/3$ saat = **20** dk.

22.12 Egzersizler

Egzersiz 1. Kupon varyantı — $n = 10$, $H_{10} \approx 2,93$.

Egzersiz 2. $F(x) = x^3$ (0..1). Inverse-transform formülü.

Egzersiz 3. X_1, \dots, X_n iid pozitif, $S = \sum X_i$. (a) $E(X_1/S)$? (b) $E((X_1 + X_2)/S)$? (c) $E(X_1^2/S^2)$ neden doğrudan $1/n^2$ değil?

Egzersiz 4. (Python — Üç problem doğrulaması)

Kupon $n=10$: sim=29.06 teori=29.29
 $P(\text{Logistic} < 0) \approx 0.5003$ teori 0.5
 $E(X/\Sigma) \approx 0.3333$ teori 0.3333

Egzersiz 5. (Sonraki ders) $T \sim \text{Exp}(\lambda)$, $f = \lambda e^{-\lambda t}$. (a) $E(T) = 1/\lambda$. (b) $\text{Var}(T) = 1/\lambda^2$. (c) **Belleksizlik:** $P(T > s + t | T > s) = P(T > t)$.

22.13 Sonraki Ders İçin Hazırlık

Ders 16: Üstel Dağılım — belleksizlik, sürekli geometrik, Poisson süreci.

⚠ Ders 16 öncesi yapılacak

- Egzersizleri çöz, özellikle 5 (Üstel özellikleri).
- Geometrik belleksizliği (Ders 9) hatırla.
- Poisson → Üstel köprüsünü tekrar oku.

22.14 Anahtar Kavramlar (Cheat Sheet)

Problem	Sonuç / Teknik	Blitzstein'de
Kupon toplayıcı	$nH_n \approx n \ln n$; geometrik + doğrusallık	0m47
Parçala	$T = \sum T_j$; küçük parçalar	8m48
Universality sezgi	$P(F(X) \leq c) = c$	11m53
Lojistik	$F = \sigma, F^{-1} = \text{logit}$	14m33
Simetri	$E(X/\Sigma) = 1/n$	18m29
LOTUS tuzak	Örüntü, isim değil	27m05
Story proof	$n - X \sim \text{Bin}(n, q)$	31m20
Poisson → Üstel	$P(T > t) = P(N_t = 0)$	37m01
Üç nesne	Plan/RD/sabit	37m44

22.15 ML Bağlantıları Özeti

💡 7 köprü

1. $n \ln n$ → keşif, örneklem karmaşıklığı.
2. **Logit / sigmoid** → **logistic regression**.
3. **Simetri** → **exchangeability**, Shapley.
4. **LOTUS doğrudan** → Monte Carlo sadeleştirme.
5. **Story proof** → bijektif/simetrik akıl.
6. **Poisson süreci** → kuyruk, sağkalım, olay-tabanlı sim.
7. **Plan/RD/sabit** → olasılıksal programlama (Pyro/NumPyro).

! Tek bir şey alıp gideceksen

Ezber formül değil, yapı görmek kazandırır — kupon toplayıcıyı geometriklere, simetrik oranı eşit terimlere, sürekli beklemeyi Poisson sayımına böl.

23 Üstel (Exponential) Dağılım

Belleksizlik: sürekli zamanın Markov çekirdeği

Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 16: Exponential Distribution](#) (≈18 dk)
- **Okuma süresi:** ≈20 dk

23.1 Bu Derste Ne Var?

Üçüncü önemli sürekli dağılım — Üstel(λ).

1. **PDF/CDF + standardizasyon:** $f = \lambda e^{-\lambda x}$, $Y = \lambda X \sim \text{Üstel}(1)$.
2. **Ortalama ve varyans:** $E = 1/\lambda$, $\text{Var} = 1/\lambda^2$.
3. **Belleksizlik:** $P(X \geq s + t \mid X \geq s) = P(X \geq t)$ — Üstel bunu sağlayan **tek** sürekli dağılım.

Builder Notu — ML Köprüleri

- λ + Poisson süreci \rightarrow istek/paket varışları, **kuyruk teorisi (M/M/1)**.
- Belleksizlik \rightarrow **Markov özelliğinin** sürekli-zaman çekirdeği; CTMC.
- Survival $S(t) = e^{-\lambda t} \rightarrow$ **sağkalım analizi**, Kaplan-Meier, **churn tahmini**, sabit hazard.
- Üstel = sürekli geometrik (kesikli \leftrightarrow sürekli ikiz).
- $E(X \mid X > a) = a + 1/\lambda \rightarrow$ **tail latency** analizi.

23.2 Üstel Dağılım

$$f(x) = \lambda e^{-\lambda x}, \quad F(x) = 1 - e^{-\lambda x}, \quad x > 0$$

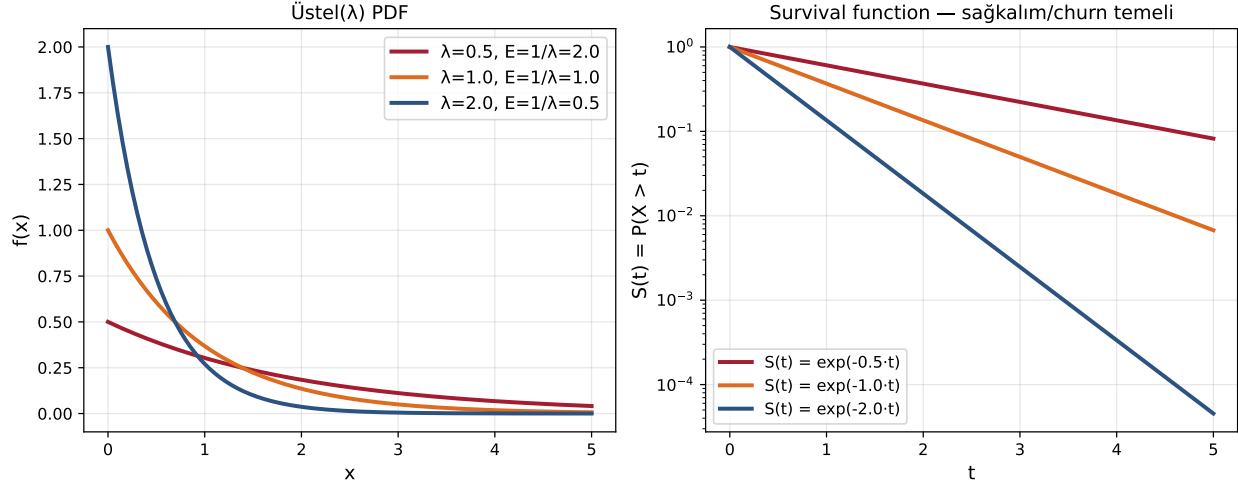
λ = oran (rate) parametresi.

Standardizasyon: $Y = \lambda X \sim \text{Üstel}(1)$.

Beklenti ve varyans (Üstel(1) için parçalı integrasyonla):

$$E(X) = \frac{1}{\lambda}, \quad \text{Var}(X) = \frac{1}{\lambda^2}$$

23 Üstel (Exponential) Dağılım



Şekil 23.1: Üstel(λ) PDF (sol) ve survival function $S(t) = e^{-\lambda t}$ (sağ). λ büyüdükçe olaylar sıklaşır, bekleme süresi kısalmır ($E = 1/\lambda$). Survival = hala beklemekte olma olasılığı; sabit hazard λ .

23.3 Belleksizlik

$$P(X \geq s + t \mid X \geq s) = P(X \geq t)$$

İspat (survival function):

$$P(X \geq s + t \mid X \geq s) = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(X \geq t)$$

“no matter how long you’ve waited then it’s like you’re starting over from fresh.” — Blitzstein, 12:03

Üstel, belleksizliği sağlayan **TEK sürekli dağılımdır** (Ders 17’de ispat).

! Builder Notu — Markov ve Sağkalım

Belleksizlik **Markov özelliğinin** sürekli-zaman çekirdeği: gelecek geçmişten bağımsız. **Sürekli-zaman Markov zincirleri (CTMC)**, kuyruk ağları, “yaşlanmayan” arızalar Üstel bekleme sürelerine dayanır. **Survival** $S(t) = e^{-\lambda t}$ sağkalım analizinin (**Kaplan-Meier, Cox**) ve **churn tahmininin** temel nesnesidir; sabit hazard oranı λ tam da belleksizliğin sonucudur.

23.4 Koşullu Beklenti: $E(X \mid X > a)$

Belleksizlik gereği $X > a$ verildiğinde **kalan süre** $X - a$ **taze bir Üstel(λ)**:

$$E(X \mid X > a) = a + E(X - a \mid X > a) = a + \frac{1}{\lambda}$$

💡 Builder Notu — Tail Latency

Kuyruk gecikmesi/tail latency: bir istek zaten uzun sürdüyse, belleksiz model kalan süreyi yine $1/\lambda$ olarak tahmin eder. Gerçek sistemlerde sapma \rightarrow belleksizlik ihlali (**ağır kuyruklu** gecikme).

23.5 Bu Dersin Özeti

1. **Üstel(λ):** $f = \lambda e^{-\lambda x}$, $F = 1 - e^{-\lambda x}$.
2. **Standardizasyon:** $Y = \lambda X \sim \text{Üstel}(1)$.
3. **Moments:** $E = 1/\lambda$, $\text{Var} = 1/\lambda^2$.
4. **Belleksizlik:** $P(X \geq s + t \mid X \geq s) = P(X \geq t)$; TEK sürekli dağılım.
5. **Koşullu E:** $E(X \mid X > a) = a + 1/\lambda$.

! Tek bir cümle

Üstel(λ), Poisson sürecinin **sürekli bekleme süresi** ve sürekli dünyanın **belleksiz** dağılımıdır — $E = 1/\lambda$, $\text{Var} = 1/\lambda^2$, $S(t) = e^{-\lambda t}$. Belleksizlik onu hem matematiksel olarak eşsiz (bu özelliği taşıyan tek sürekli) hem de modellemede güçlü kılar: **Markov, sağkalım, kuyruk, tail latency**.

23.6 Kontrol Soruları

i Soru 1: $X \sim \text{Exp}(2)$. (a) E, Var? (b) $P(X > 1)$? (c) Medyan?

Cevap: (a) $E = 1/2$, $\text{Var} = 1/4$. (b) $e^{-2} \approx 0,135$. (c) $(\ln 2)/2 \approx 0,347$ — medyan < ortalama (sağa çarpık).

i Soru 2: $P(X > E(X)) = ?$ λ 'ya bağlı mı?

Cevap: $P(X > 1/\lambda) = e^{-1} \approx 0,368$. λ 'dan **bağımsız!** Sağa çarpık göstergesi.

i Soru 3: Ampul Üstel, ortalama 1000 saat. 500 saat çalıştı, kalan beklenen?

Cevap: **1000 saat** (belleksizlik). Ampul “hiç eskimemiş” gibi. (Gerçek ampuller eskir — idealize model.)

i Soru 4: (Builder) Sunucu Exp, ort 100ms. İstek 200ms sürdü. Toplam E? Tail latency için?

Cevap: $E(X \mid X > 200) = 200 + 100 = 300$ ms. Belleksizlik \rightarrow kalan hep $1/\lambda$. Gerçek sapma \rightarrow ağır kuyruk.

23.7 Egzersizler

Egzersiz 1. $X \sim \text{Exp}(0.5)$. (a) E , Var, SD? (b) $P(X > 3)$, $P(1 < X < 2)$? (c) Medyan?

Egzersiz 2. $E(X^n)$ standardizasyonla. (Üstel(1)'de $E(Y^n) = n!$.)

Egzersiz 3. Survival ile belleksizliği göster.

Egzersiz 4. (Python — Üstel doğrulama)

$E(X) \approx 1.999$ teorik $1/\lambda = 2.0$
 $\text{Var}(X) \approx 4.012$ teorik $1/\lambda^2 = 4.0$
 $P(X > 3) \approx 0.2232$ teorik $e^{-1.5} = 0.2231$
 $P(X > 5 | X > 3) \approx 0.3681$ $P(X > 2) = 0.3679$

Egzersiz 5. (Sonraki ders) **MGF:** $M(t) = E(e^{tX})$. Üstel için $M(t) = \lambda/(\lambda - t)$ ($t < \lambda$).

23.8 Sonraki Ders İçin Hazırlık

Ders 17: Moment Üreten Fonksiyonlar (MGF) — momentleri tek fonksiyonda paketle.

⚠ Ders 17 öncesi yapılacak

- Egzersiz 5 (Üstel MGF) çöz.
- $e^{tX} = \sum (tX)^k / k!$ Taylor serisini hatırla.

23.9 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Blitzstein'de
Üstel PDF/CDF	$\lambda e^{-\lambda x}, 1 - e^{-\lambda x}$	2m00
Oran λ	Birim zamanda olay hızı	1m31
Standardizasyon	$Y = \lambda X \sim \text{Üstel}(1)$	4m54
E / Var	$1/\lambda, 1/\lambda^2$	10m49
Survival	$S(t) = e^{-\lambda t}$	14m33
Belleksizlik	$P(X \geq s + t X \geq s) = P(X \geq t)$; TEK sürekli	11m17
Koşullu E	$E(X X > a) = a + 1/\lambda$	17m43

23.10 ML Bağlantıları Özeti

💡 6 köprü

1. **Oran λ / Poisson süreci** → bekleme süreleri, **kuyruk teorisi**.
2. **Belleksizlik** → **Markov özelliği**, CTMC.
3. **Survival** → sağkalım analizi, **Kaplan-Meier**, churn.
4. **Üstel = sürekli geometrik** → kesikli ↔ sürekli.
5. **Koşullu E** → **tail latency**, SLA.
6. **Üstel + Gamma eşleniği** → Bayesçi oran tahmini (Ders 24).

❗ Tek bir şey alıp gideceksen

Üstel(λ) sürekli dünyanın **belleksiz** dağılımıdır — $S(t) = e^{-\lambda t}$, $E = 1/\lambda$. Belleksizlik onu **eşsiz** (tek sürekli) ve **modellemede güçlü** kılar: Markov, sağkalım, kuyruk, tail latency.

24 Moment Üreten Fonksiyonlar (MGF)

Belleksizlik = Üstel; $M(t) = E(e^{tX})$ — dağılımın parmak izi

Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 17: Moment Generating Functions](#) (≈51 dk)
- **Okuma süresi:** ≈36 dk

24.1 Bu Derste Ne Var?

1. **Belleksizlik = Üstel:** Tek belleksiz sürekli dağılım, fonksiyonel denklem ispatı.
2. **MGF** $M(t) = E(e^{tX})$: momentleri üretir, dağılımı belirler, bağımsız toplamı çarpıma çevirir.
3. **Örnekler:** Bern, Binom, Normal MGF.
4. **Laplace ardışıklık kuralı:** $(n + 1)/(n + 2)$.

“All the MGF is, is a fancy book keeping device for keeping track of the moments of a distribution.”
— Blitzstein, 20:16

Builder Notu — ML Köprüleri

- **Belleksizlik karakterizasyonu** → Weibull (üstelin kuvveti); sağkalım, sabit hazard testi.
- **MGF momentleri** → **kümülanlar**; CLT (Ders 29) MGF yakınsamasıyla.
- **Bağımsız toplam = MGF çarpımı** → Normal/Poisson kapanışı tek satırda.
- **Normal MGF / kareye tamamlama** → **Gaussian conjugacy** (Bayesçi regresyon, GP, VAE KL).
- **Laplace ardışıklık** $(n + 1)/(n + 2) =$ **Laplace smoothing (add-one)** — naive Bayes, n-gram.

24.2 Belleksizlik = Üstel (Fonksiyonel Denklem)

Teorem: Pozitif, sürekli, belleksiz bir RV → Üstel(λ).

Survival $G(x) = P(X > x)$ ile belleksizlik:

$$G(s + t) = G(s) G(t)$$

24 Moment Üreten Fonksiyonlar (MGF)

Değerler takarak: tamsayı k için $G(kt) = G(t)^k$; rasyonel için $G(t \cdot m/n) = G(t)^{m/n}$; süreklilikle her reel $x > 0$ için:

$$G(xt) = G(t)^x$$

$$t = 1: G(x) = G(1)^x = e^{-\lambda x}, \lambda = -\ln G(1) > 0. \blacksquare$$

“we’re solving for a function, not solving for the variable.” — Blitzstein, 9:33

24.3 MGF Tanımı ve Üç Önem

$$M(t) = E(e^{tX})$$

(0 etrafında bir aralıkta sonlu olmalı.)

Taylor: $e^{tX} = \sum (tX)^n / n! \rightarrow$

$$M(t) = \sum_{n=0}^{\infty} \frac{E(X^n)}{n!} t^n$$

Üç önem:

1. **Moment üretir:** $E(X^n) = M^{(n)}(0)$.
2. **Dağılımı belirler:** aynı MGF \Rightarrow aynı dağılım (**parmak izi**).
3. **Bağımsız toplam \rightarrow çarpım:**

$$M_{X+Y}(t) = M_X(t) M_Y(t) \quad (X \perp Y)$$

“Once you know the MGF, you know the distribution, at least in principle.” — Blitzstein, 27:24

! Builder Notu — CLT’nin Motoru

Sebepler 2 + 3 birlikte Normal/Poisson kapanışını tek satırda kanıtlar. **CLT** (Ders 29) MGF yakınsamasıyla: iid toplamın standardize MGF’i $e^{t^2/2}$ ’ye yakınsar.

24.4 Örnek MGF'ler

Bernoulli(p):

$$M(t) = pe^t + q$$

Binom(n, p) — n Bernoulli'nin toplamı, MGF kuvveti:

$$M(t) = (pe^t + q)^n$$

Standart Normal $N(0, 1)$ — kareye tamamlama:

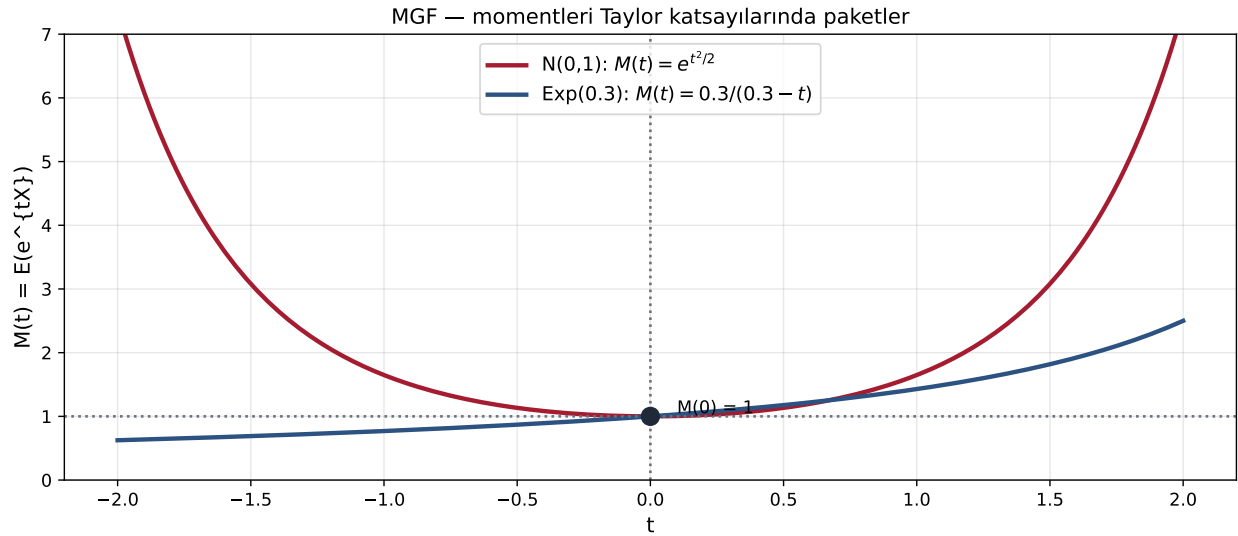
$$M(t) = \int e^{tz} \cdot \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz = e^{t^2/2} \cdot \int \frac{e^{-(z-t)^2/2}}{\sqrt{2\pi}} dz = e^{t^2/2}$$

Üstel(λ):

$$M(t) = \frac{\lambda}{\lambda - t}, \quad t < \lambda$$

Poisson(λ):

$$M(t) = e^{\lambda(e^t - 1)}$$



Şekil 24.1: MGF'ten momentler: $N(0,1)$ 'in $M(t) = \exp(t^2/2)$. Sayısal türev $M'(0) = 0 = E(Z)$, $M''(0) = 1 = \text{Var}(Z)$. MGF dağılımın parmak izi — aynı MGF aynı dağılım anlamına gelir.

! Builder Notu — Gaussian Conjugacy

Kareye tamamlama numarası Gaussian'larla yapılan hemen her hesabın altında: **Bayesçi lineer regresyon, Gaussian process, VAE'nin KL terimi**. “Küadratik üstel + kareye tamamlama → kapalı form Gaussian” deseni Gaussian conjugacy'nin motorudur.

24.5 Laplace'ın Ardışıklık Kuralı

Güneş yarın doğacak mı? n gün üst üste doğduysa?

$p \sim \text{Uniform}(0, 1)$ prior + $S_n | p \sim \text{Bin}(n, p)$:

$$f(p | S_n = n) \propto p^n \cdot 1 \Rightarrow f(p | S_n = n) = (n + 1)p^n$$

$$P(X_{n+1} = 1 | S_n = n) = \int_0^1 p(n + 1)p^n dp = \frac{n + 1}{n + 2}$$

“if the sun rose 100 days in a row, then it would probably be 101 over 102 for the next day.” — Blitzstein, 50:37

! Builder Notu — Laplace Smoothing

Laplace smoothing (add-one): k gözlem, n deneme → olasılığı $(k + 1)/(n + 2)$. **Naive Bayes / n-gram dil modellerinde** sıfır-frekans problemini çözer. Beta-Binom eşleniği (Ders 23) — pseudocount'lar prior'dan gelir.

24.6 Bu Dersin Özeti

1. **Belleksizlik = Üstel** (fonksiyonel denklem).
2. **MGF**: $M(t) = E(e^{tX})$.
3. **Üç önem**: moment üretir, dağılım belirler, bağımsız toplam çarpılır.
4. **Örnekler**: Bern, Binom, Normal ($e^{t^2/2}$), Üstel, Poisson.
5. **Laplace ardışıklık**: $(n + 1)/(n + 2)$.

! Tek bir cümle

Belleksizlik Üstel'i tam karakterize eder ($G(s+t)=G(s)G(t) \rightarrow e^{(-\lambda x)}$); ve **MGF** $M(t) = E(e^{tX})$, momentleri paketleyen, dağılımı belirleyen, bağımsız toplamı **çarpıma** çeviren bir parmak izidir — moment hesabının ve **CLT'nin motoru**.

24.7 Kontrol Soruları

i Soru 1: $N(0,1)$ MGF $e^{t^2/2}$. $M'(0)$, $M''(0)$ ile $E(Z)$, $E(Z^2)$?

Cevap: $M'(t) = te^{t^2/2}$, $M'(0) = 0 = E(Z)$. $M''(t) = (1 + t^2)e^{t^2/2}$, $M''(0) = 1 = E(Z^2)$. $\text{Var} = 1$.

i Soru 2: $X \sim \text{Pois}(\lambda_1)$, $Y \sim \text{Pois}(\lambda_2)$ bağımsız. $X+Y$?

Cevap: $M_{X+Y}(t) = e^{\lambda_1(e^t-1)} e^{\lambda_2(e^t-1)} = e^{(\lambda_1+\lambda_2)(e^t-1)} \rightarrow \text{Pois}(\lambda_1 + \lambda_2)$.

i Soru 3: $\text{Exp}(\lambda)$ MGF $\lambda/(\lambda-t)$. (a) $M'(0)$ ile $E(X)$? (b) Neden $t < \lambda$?

Cevap: (a) $M'(0) = 1/\lambda$. (b) İntegral $\int_0^\infty e^{-(\lambda-t)x} dx$ yalnız $t < \lambda$ 'da yakınsar.

i Soru 4: (Builder) 1000 dokümanda hiç geçmeyen kelime. MLE = 0. Laplace?

Cevap: $(0 + 1)/(1000 + 2) = 1/1002 \approx 0,001$. **MLE zero-frequency problem**'i çözer; naive Bayes/n-gram'da tek 0 tüm çarpımı sıfırlar. **Laplace smoothing** = bu kural.

24.8 Egzersizler

Egzersiz 1. $\text{Exp}(\lambda)$ MGF'ten $E(X^2)$, $E(X^n) = n!/\lambda^n$.

Egzersiz 2. Genel Normal MGF: $X = \mu + \sigma Z \rightarrow M_X(t) = e^{\mu t + \sigma^2 t^2/2}$.

Egzersiz 3. Bağımsız Normal toplamı MGF ile.

Egzersiz 4. (Python — MGF + Laplace)

$E(e^{0.5*Z}) \approx 1.1339$ teorik $e^{0.125} = 1.1331$

$\text{Pois}(2)+\text{Pois}(3)$: mean=4.997, var=5.006 (teori 5, 5)

$k=0$, $n=1000 \rightarrow P(\text{next}) = 0.000998$

Egzersiz 5. (Sonraki ders) $M_{S_n}(t) = M(t)^n$ — CLT habercisi.

24.9 Sonraki Ders İçin Hazırlık

Ders 18: MGF'lere Devam ve Joint Dağılımlar — Üstel momentleri, Normal momentleri, Poisson toplamı, joint kavramı.

⚠ Ders 18 öncesi yapılacak

- Egzersiz 5 ($M(t)^n$) çöz.
- Üç MGF ezberle: Bern, Normal, Üstel.

24.10 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Blitzstein'de
Belleksizlik = Üstel	Pozitif sürekli memoryless \rightarrow Exp	0m27
Fonksiyonel denklem	$G(s+t) = G(s)G(t)$ $\rightarrow e^{-\lambda x}$	9m33
MGF	$M(t) = E(e^{tX})$	18m03
Moment üretir	$E(X^n) = M^{(n)}(0)$	20m32
Dağılım belirler	Aynı MGF \Rightarrow aynı dist.	26m12
Bağımsız toplam	$M_{X+Y} = M_X \cdot M_Y$	29m52
Bern / Binom	$pe^t + q / (pe^t + q)^n$	31m07
N(0,1)	$e^{t^2/2}$ (kareye tamamla)	36m51
Exp(λ)	$\lambda/(\lambda - t), t < \lambda$	—
Pois(λ)	$e^{\lambda(e^t-1)}$	—
Laplace	$(n+1)/(n+2)$	50m37

24.11 ML Bağlantıları Özeti

💡 6 köprü

1. **Belleksizlik karakterizasyonu** \rightarrow Üstel, **Weibull**, sağkalım.
2. **MGF momentleri** \rightarrow **kümülanlar**; CLT motoru.
3. **MGF dağılımı belirler** \rightarrow parmak izi; CLT ispatı.
4. **Bağımsız toplam** \rightarrow **çarpım** \rightarrow Normal/Poisson kapanışı.
5. **Kareye tamamlama** \rightarrow **Gaussian conjugacy**, VAE KL.
6. **Laplace ardışıklık** \rightarrow **Laplace smoothing** (naive Bayes, n-gram).

❗ Tek bir şey alıp gideceksen

Belleksizlik **Üstel'i tam belirler**. MGF dağılımın parmak izidir — momentleri paketler, dağılımı belirler, bağımsız toplamı **çarpıma** çevirir; moment hesabının ve **CLT'nin motoru**.

25 MGF'lere Devam ve Joint Dağılımlar

Üstel/Normal/Poisson MGF; joint, marjinal, bağımsızlık

Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 18: MGFs Continued; Joint Distributions](#) (≈50 dk)
- **Okuma süresi:** ≈34 dk

25.1 Bu Derste Ne Var?

1. **Üstel MGF + tüm momentleri:** $1/(1-t)$ geometrik seri $\rightarrow E(X^n) = n!/\lambda^n$.
2. **Normal momentleri (Taylor):** $E(Z^{2n}) = (2n)!/(2^n n!)$.
3. **Poisson MGF:** $e^{\lambda(e^t-1)}$; bağımsız toplam Poisson.
4. **Joint dağılımlar:** PMF/PDF, marjinal, bağımsızlık = çarpım.
5. **Kare (bağımsız) vs disk (bağımlı)** — destek kısıtı.

Builder Notu — ML Köprüleri

- **Normal çift momentleri = Wick/Isserlis teoremi** — fizik, GP, derin öğrenme analizi.
- **Poisson toplam kapanışı** \rightarrow sayım birleştirme; Poisson regresyon log-link.
- **Joint + faktörizasyon** \rightarrow tüm üretici modeller; **bağımsızlık = çarpım** PGM/naive Bayes temeli.
- **Marjinalleştirme** \rightarrow **inference**, marginal likelihood, EM.
- **Destek geometrisi** \rightarrow manifold/kısıtlı dağılımlar, normalizing flows.

25.2 Üstel MGF ve Tüm Momentler

$X \sim \text{Exp}(1)$, LOTUS:

$$M(t) = \int_0^{\infty} e^{tx} e^{-x} dx = \frac{1}{1-t}, \quad t < 1$$

Geometrik seri olarak:

$$\frac{1}{1-t} = \sum_n t^n = \sum_n n! \frac{t^n}{n!} \Rightarrow E(X^n) = n!$$

25 MGF'lere Devam ve Joint Dağılımlar

Genel $\text{Exp}(\lambda)$: $E(Y^n) = n!/\lambda^n$.


25.3 Normal'in Tüm Momentleri

$M(t) = e^{t^2/2}$ Taylor:

$$e^{t^2/2} = \sum_n \frac{(t^2/2)^n}{n!} = \sum_n \frac{t^{2n}}{2^n n!}$$

$$E(Z^{2n}) = \frac{(2n)!}{2^n n!}, \quad E(Z^{2n+1}) = 0$$

Örüntü: 1, 3, 15, 105 = 1, 1 · 3, 1 · 3 · 5, 1 · 3 · 5 · 7 (çift faktöriyel).

 **Builder Notu** — Wick / Isserlis

$(2n - 1)!! = 2n$ kişiyi ikili eşlemelere ayırma sayısı! **Isserlis / Wick teoremi**: Gaussian'ın yüksek momentleri çift-eşleme + kovaryans çarpımıyla hesaplanır. Gaussian integraller, diyagramatik açılımlar.

25.4 Poisson MGF ve Toplamı

$$M(t) = e^{-\lambda} \sum_k \frac{(\lambda e^t)^k}{k!} = e^{\lambda(e^t-1)}$$

Bağımsız toplam:

$$M_{X+Y}(t) = e^{(\lambda+\mu)(e^t-1)} \Rightarrow X + Y \sim \text{Pois}(\lambda + \mu)$$

X = Y tuzağı: $X + X = 2X$ Poisson **değil** (yalnız çift değerler; $\text{Var} = 4\lambda \neq 2\lambda$).

25.5 Joint Dağılımlar

2x2 PMF tablosu: 4 sayı ≥ 0 , toplam 1.

$$F(x, y) = P(X \leq x, Y \leq y) \quad (\text{joint CDF})$$

$$P((X, Y) \in B) = \iint_B f(x, y) dx dy$$

Bağımsızlık:

$$P(X = x, Y = y) = P(X = x)P(Y = y) \quad \forall x, y$$

Marjinalleştirme: $f_Y(y) = \int f(x, y)dx$. Joint \rightarrow marjinal **olur**, marjinal \rightarrow joint **olmaz**.

! Builder Notu — Marjinalleştirme = Inference

Marjinalleştirme olasılıksal çıkarımın motorudur: latent değişkenleri \sum / \int ile atmak **marginal likelihood / evidence** verir. **EM, variational inference**, Bayeşçi model karşılaştırması hep buna dayanır. “Aynı marjinalli farklı joint” = **copula**’ların var olma nedeni.

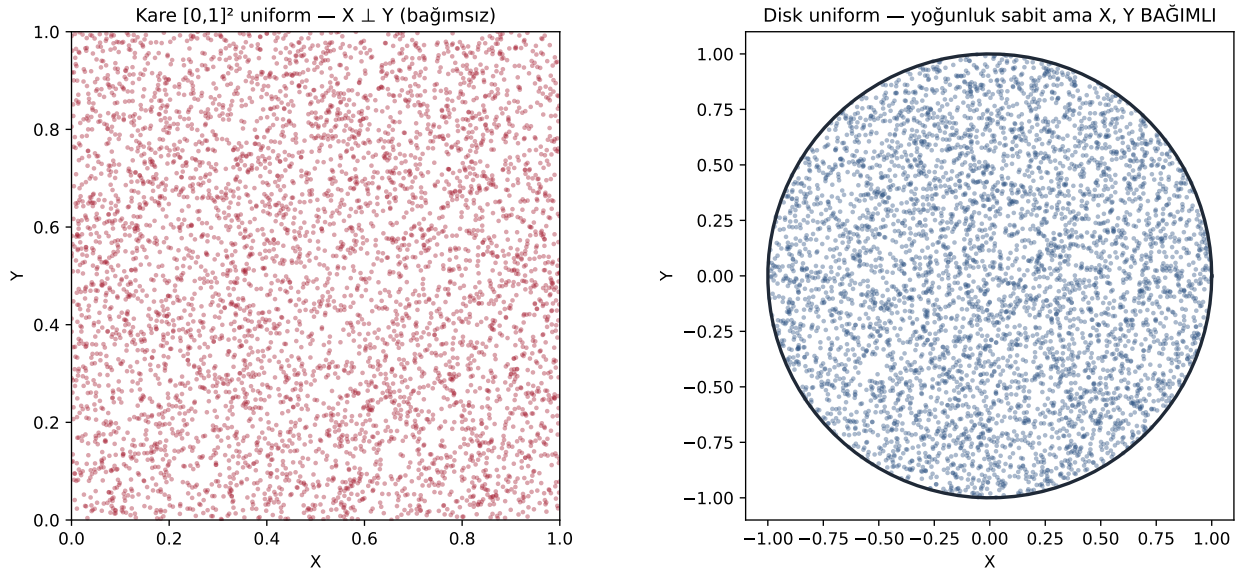
25.6 Kare (Bağımsız) vs Disk (Bağımlı)

Kare $[0, 1]^2$: $f = 1$. Marjinaler $f_X = f_Y = 1 \rightarrow$ **bağımsız** Uniform.

Disk $x^2 + y^2 \leq 1$: $f = 1/\pi$ sabit. **Ama X, Y BAĞIMLI!** Destek dikdörtgen değil:

$$x^2 + y^2 \leq 1 \Rightarrow -\sqrt{1-x^2} \leq Y \leq \sqrt{1-x^2}$$

X’i bilmek Y’nin aralığını **kısıtlıyor**.



Şekil 25.1: Kare vs disk uniform. Sol: kare — X ve Y bağımsız (her yatay/dikey kesit aynı dağılım). Sağ: disk — sabit yoğunluk $1/\pi$, ama $|X|$ büyüdükçe Y aralığı daralır = bağımlılık. Korelasyon 0 olabilir ama bağımsızlık değil.

! Builder Notu — Manifold ve Kısıt

“Sabit yoğunluk \neq bağımsız” — bağımsızlık hem yoğunluk faktörizasyonu hem **destek = product set (dikdörtgen)** ister. **Manifold üzerindeki** dağılımlar (normalizing flows, kısıtlı latent uzaylar) bileşenleri kaçınılmaz olarak bağlar.

25.7 Bu Dersin Özeti

1. **Üstel MGF** $\rightarrow E(X^n) = n!/\lambda^n$.
2. **Normal momentleri** $\rightarrow (2n)!/(2^n n!) = (2n - 1)!!$ (Wick).
3. **Poisson MGF** $\rightarrow e^{\lambda(e^t-1)}$; bağımsız toplam Poisson.
4. **Joint:** bağımsız = çarpım (tüm x, y).
5. **Marjinal:** \sum / \int ile diğerini at.
6. **Kare vs disk:** destek kritik.

! Tek bir cümle

MGF momentleri integral yerine **örüntü tanımayla** verir ve bağımsız toplamları çarpıma çevirir (Poisson toplamı yine Poisson); **joint dağılımlar** birden fazla değişkeni birlikte taşır — **bağımsızlık = faktörizasyon + dikdörtgen destek**.

25.8 Kontrol Soruları

i Soru 1: $\text{Pois}(\lambda)$ MGF $e^{\lambda(e^t-1)}$. Türevlerle E, Var?

Cevap: $M'(0) = \lambda, M''(0) = \lambda + \lambda^2 \rightarrow \text{Var} = \lambda$.

i Soru 2: $E(Z^8)$?

Cevap: $8!/(2^4 \cdot 4!) = 40320/384 = 105 = 1 \cdot 3 \cdot 5 \cdot 7$.

i Soru 3: 2×2 PMF $P = [[.1, .4], [.3, .2]]$. Marjinaller? Bağımsız mı?

Cevap: $P(X = 0) = 0,5, P(X = 1) = 0,5; P(Y = 0) = 0,4, P(Y = 1) = 0,6. P(X = 0, Y = 0) = 0,1 \neq 0,5 \cdot 0,4 = 0,2 \rightarrow$ **bağımlı**.

i Soru 4: (Builder) Disk yoğunluğu sabit. Neden bağımsız değil?

Cevap: Yoğunluk faktörize gibi (sabit), ama **destek** $x^2 + y^2 \leq 1$ dikdörtgen değil. $\mathbb{1}(x^2 + y^2 \leq 1)$ x ve y 'nin fonksiyonu olarak ayrılamaz. Bağımsızlık iki şey ister.

25.9 Egzersizler

Egzersiz 1. Bin(n, p) MGF'ten $E(X)$, $\text{Var}(X) = npq$.

Egzersiz 2. $E(Z^{10})$ + çift faktöriyel doğrula.

Egzersiz 3. 2x2 bağımlı joint kur, koşullu $P(Y | X = 0)$.

Egzersiz 4. (Python — toplam + disk)

Pois+Pois: mean=5.002, var=4.990 (Pois(5): 5, 5)
 corr(X,Y) = -0.0010 (≈ 0)
 std(Y | |X|<0.1) = 0.575
 std(Y | |X|>0.9) = 0.197 (çok daha küçük → bağımlı!)

Egzersiz 5. (Sonraki ders) Diskte $Y | X = x \sim \text{Uniform}(-\sqrt{1-x^2}, \sqrt{1-x^2})$.

25.10 Sonraki Ders İçin Hazırlık

Ders 19: Birleşik, Koşullu ve Marjinal Dağılımlar — koşullu PDF, sürekli Bayes.

⚠ Ders 19 öncesi yapılacak

- Egzersizleri çöz.
- “Bağımsız = çarpım” + “destek dikdörtgen olmalı” reflekslerini pekiştir.

25.11 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Blitzstein'de
Üstel MGF	$1/(1-t)$; $E(X^n) = n!/\lambda^n$	0m59
Normal moment	$(2n)!/(2^n n!) = (2n-1)!!$	12m06
Poisson MGF	$e^{\lambda(e^t-1)}$	17m49
Pois toplam	$\text{Pois}(\lambda + \mu)$	20m50
Joint	$f(x, y)$; $\iint = P$	30m04
Bağımsızlık	$f(x, y) = f_X(x)f_Y(y)$	30m43
Marjinal	$f_Y(y) = \int f(x, y)dx$	38m12
Kare vs disk	Destek kritik	46m21

25.12 ML Bağlantıları Özeti

💡 7 köprü

1. **MGF momentleri** → integral yerine türev/Taylor.
2. **Wick/Isserlis** → Gaussian momentleri = çift-eşleme.
3. **Poisson toplam** → sayım birleştirme.
4. **Joint** → tüm üretici modeller.
5. **Bağımsızlık = çarpım** → PGM, naive Bayes.
6. **Marjinalleştirme** → inference, EM, evidence.
7. **Destek geometrisi** → manifold, normalizing flows.

! Tek bir şey alıp gideceksen

MGF momentleri örüntü tanımayla verir, bağımsız toplamı çarpıma çevirir; **joint dağılımlar**'da bağımsızlık = **faktörizasyon + dikdörtgen destek**.

26 Birleşik, Koşullu ve Marjinal Dağılımlar

2D LOTUS, joint = koşullu × marjinal, Poisson splitting

i Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 19: Joint, Conditional, and Marginal Distributions](#) (≈50 dk)
- **Okuma süresi:** ≈35 dk

26.1 Bu Derste Ne Var?

1. **Üçlü:** joint → marjinal (integral) → koşullu (joint/marjinal). **Joint = koşullu × marjinal.**
2. **2D LOTUS:** $E(g(X, Y)) = \iint g \cdot f \, dx \, dy$.
3. **Bağımsız ⇒ ilişkisiz:** $X \perp Y \Rightarrow E(XY) = E(X)E(Y)$.
4. **İki uniform uzaklık = 1/3; Poisson splitting** (tavuk-yumurta).

💡 Builder Notu — ML Köprüleri

- **Koşullu** $f(y | x)$ → denetimli öğrenme, **koşullu üretim** (cVAE, cDiffusion).
- **Joint = \prod koşullu** → **autoregressive faktörizasyon** (GPT, PixelCNN).
- **2D LOTUS** → çok değişkenli Monte Carlo.
- **Bağımsız ⇒ ilişkisiz (tersi değil)** → **PCA vs ICA**.
- **Poisson splitting (thinning)** → yük dengeleme, A/B trafik ayırma, nokta süreçleri.

26.2 Joint → Marjinal → Koşullu

Joint PDF: $f(x, y) = \partial^2 F / \partial x \partial y$. Olasılık = çift integral.

Marjinal: $f_X(x) = \int f(x, y) dy$.

Koşullu:

$$f_{Y|X}(y | x) = \frac{f(x, y)}{f_X(x)}$$

Sürekli çarpım kuralı (Bayes):

$$f(x, y) = f_{Y|X}(y | x) f_X(x) = f_{X|Y}(x | y) f_Y(y)$$

“think of it as the PDF where we get to pretend that we know what X is.” — Blitzstein, 9:05

! Builder Notu — Autoregressive ve cGAN

“joint = \prod koşullu” ayrışımı **modern üretici modellemenin omurgası**: $p(x_1, \dots, x_d) = \prod p(x_i | x_{<i})$
— **GPT, PixelCNN**. Koşullu yoğunluk $f(y | x)$ tüm denetimli öğrenmenin ve koşullu üretimin hedefi.

26.3 Sürekli Bağımsızlık

$$X \perp Y \iff f(x, y) = f_X(x) f_Y(y) \quad \forall x, y$$

Eşdeğer: koşullu = marjinal (X 'i öğrenmek Y hakkında bilgi vermez).

26.4 Disk Örneği: Marjinal ve Koşullu

Birim disk uniform, $f = 1/\pi$.

Marjinal:

$$f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} dy = \frac{2}{\pi} \sqrt{1-x^2}$$

Uniform değil — yarım daire eğrisi.

Koşullu:

$$f_{Y|X}(y | x) = \frac{1/\pi}{(2/\pi)\sqrt{1-x^2}} = \frac{1}{2\sqrt{1-x^2}}$$

y yok \rightarrow **sabit** $\rightarrow Y | X = x \sim \text{Uniform}(-\sqrt{1-x^2}, \sqrt{1-x^2})$. Aralık x 'e bağlı \rightarrow bağımlı.

💡 Builder Notu — Heteroskedastik

“Koşullu uniform ama parametresi x 'e bağlı” = **heteroskedastik** modellerin temel sezgisi — girdiye bağlı yayılım.

26.5 2D LOTUS ve Bağımsız \Rightarrow İlişkisiz

$$E(g(X, Y)) = \iint g(x, y) f(x, y) dx dy$$

Bağımsız $\Rightarrow E(XY) = E(X)E(Y)$: $g(x, y) = xy$, $f = f_X f_Y$, integral ayrılır.

“independent implies uncorrelated” — Blitzstein, 25:07

Uyarı: Ters **DOĞRU DEĞİL**. **İlişkisiz olup bağımlı** olmak mümkün.

! Builder Notu — PCA vs ICA

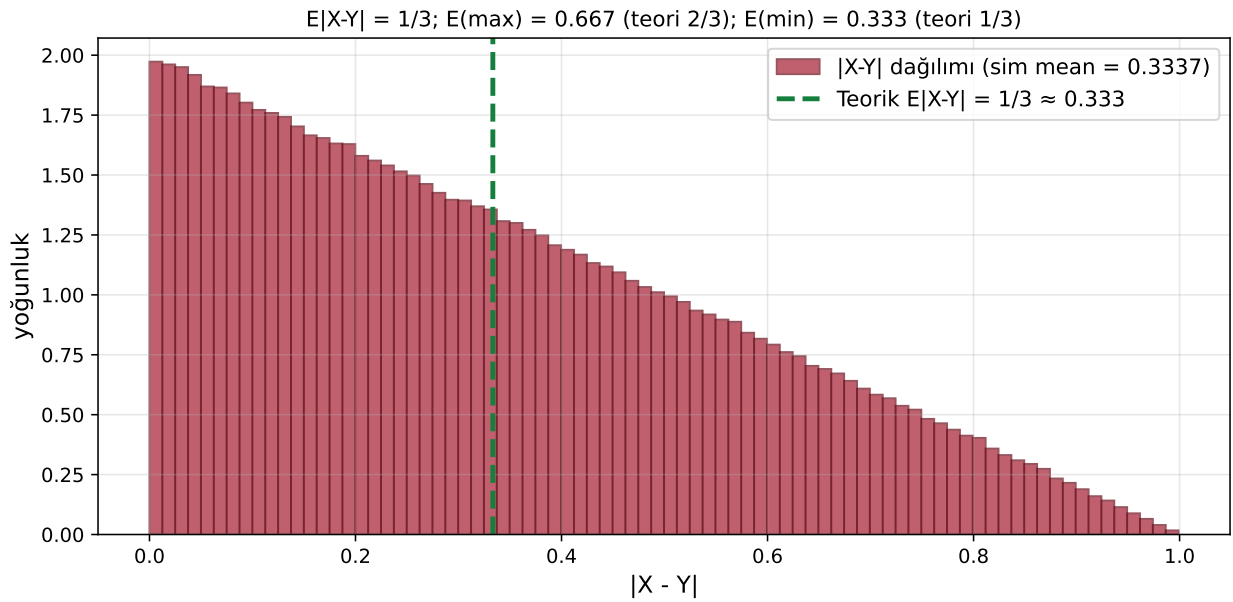
PCA / whitening değişkenleri **ilişkisiz** yapar, ama **bağımsız** yapmaz. **ICA** bağımsızlığı hedefler. Korelasyon yalnızca doğrusal ilişkiyi yakalar.

26.6 İki Uniform Uzaklık: Max/Min Triki

$X, Y \sim \text{Uniform}(0,1)$ bağımsız. $E|X - Y| = ?$

Triki: $M = \max, L = \min$. $|X - Y| = M - L$, $X + Y = M + L$:

$$E(M - L) = \frac{1}{3}, \quad E(M + L) = 1 \Rightarrow E(M) = \frac{2}{3}, \quad E(L) = \frac{1}{3}$$



Şekil 26.1: İki bağımsız $\text{Uniform}(0,1)$ 'in beklenen uzaklığı $= 1/3$. Max ortalama $2/3$, min ortalama $1/3$ — “ortalama düzende” noktalar $1/3$ ve $2/3$ 'te oturur. Sıra istatistiklerinin en basit hali.

26.7 Poisson Splitting (Tavuk-Yumurta)

$N \sim \text{Pois}(\lambda)$ yumurta, her biri bağımsız p olasılıkla çıkıyor. $X =$ çıkan, $Y =$ çıkmayan, $X + Y = N$.

$$P(X = i, Y = j) = \binom{i+j}{i} p^i q^j \cdot \frac{e^{-\lambda} \lambda^{i+j}}{(i+j)!}$$

Sadeleştir:

$$= \frac{e^{-\lambda p} (\lambda p)^i}{i!} \cdot \frac{e^{-\lambda q} (\lambda q)^j}{j!}$$

Joint çarpanlara ayrıldı!

$$X \perp Y, \quad X \sim \text{Pois}(\lambda p), \quad Y \sim \text{Pois}(\lambda q)$$

Yalnızca Poisson'a özgü.

“this is actually a very special property of the Poisson.” — Blitzstein, 49:45

! Builder Notu — Thinning ve A/B Trafik

Poisson thinning/splitting = bir Poisson akışını bağımsız alt-akışlara bölmek. Nokta süreçleri, ağ trafiği modellemesi, **yük dengeleme** (istekleri bağımsız kuyruklara), **A/B test trafik ayırma**. Bağımsızlık, alt-akışları ayrı analiz etmeyi kolaylaştırır — kovaryans düzeltilmesi yok.

26.8 Bu Dersin Özeti

1. **Üçlü:** joint \rightarrow marjinal (f) \rightarrow koşullu (joint/marjinal); $f = f_{Y|X} \cdot f_X$.
2. **2D LOTUS:** $E(g) = \iint g \cdot f$.
3. **Bağımsız \Rightarrow ilişkisiz:** $E(XY) = E(X)E(Y)$.
4. **Disk:** marjinal yarım daire, koşullu uniform \rightarrow bağımlı.
5. **İki uniform:** $E|X - Y| = 1/3$, max/min triki.
6. **Poisson splitting:** $X \perp Y$, ikisi Poisson.

! Tek bir cümle

Joint, marjinal ve koşullu dağılımlar bir **üçlüdür** — joint = koşullu \times marjinal. **2D LOTUS** beklentileri dağılım bulmadan verir; **bağımsız \Rightarrow ilişkisiz ama tersi değil**; **Poisson splitting** alt-akışları bağımsız yapar.

26.9 Kontrol Soruları

i Soru 1: $f(x,y) = x+y$ (0..1)². (a) Marjinal? (b) Koşullu? (c) Bağımsız mı?

Cevap: (a) $f_X(x) = x + 1/2$. (b) $f_{Y|X} = (x + y)/(x + 1/2)$. (c) $f_X f_Y \neq f \rightarrow$ **bağımsız değil**.

i Soru 2: $f(x,y) = 4xy$ (0..1)². Bağımsız mı?

Cevap: $f_X = 2x, f_Y = 2y$, çarpım $4xy = f \rightarrow$ **bağımsız** (Beta(2,1) marjinalleri).

i Soru 3: $X, Y \sim \text{Unif}(0,1)$ bağımsız. (a) $E(XY)$? (b) $E(\max)$?

Cevap: (a) $(1/2)(1/2) = 1/4$. (b) $2/3$.

i Soru 4: (Builder) $N \sim \text{Pois}(100)$, $p=0.02$ dönüşüm. X dağılımı? Y 'den bağımsız mı?

Cevap: **Pois(2)**, $X \perp Y \sim \text{Pois}(98)$. A/B trafik ayırmanın temeli.

26.10 Egzersizler

Egzersiz 1. $f = c(x + y^2)$ (0..1)². (a) c ? (b) Marjinaler? (c) Bağımsız mı?

Egzersiz 2. $f = 6e^{-2x-3y}$ ($x, y > 0$). Marjinaler? Bağımsız mı? Oranlar?

Egzersiz 3. $X, Y \sim \text{Unif}(0,1)$ bağımsız. $E(X + Y)$, $E(XY)$, $E((X - Y)^2)$?

Egzersiz 4. (Python — Uzaklık + Splitting)

$E|X-Y| \approx 0.3337$ teori $1/3$

$E(\max) \approx 0.6670$ $E(\min) \approx 0.3333$

$X \sim \text{Pois}$? mean=1.998, var=1.997 (Pois(2): 2, 2)

corr(X,Y) = 0.0009 (bağımsız $\rightarrow \approx 0$)

Egzersiz 5. (Sonraki ders) **Multinomial**: binomun k kategoriye genellemesi. $X_i \sim \text{Bin}(n, p_i)$ marjinal; X_i, X_j **negatif ilişkili**.

26.11 Sonraki Ders İçin Hazırlık

Ders 20: Multinomial ve Cauchy — joint sayım, kategorik genelleme; oran dağılımı.

⚠ Ders 20 öncesi yapılacak

- Egzersiz 5 (Multinomial sezgisi) çöz.
- Binom + MGF'i hatırla.
- 2D LOTUS'u tekrar oku.

26.12 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Blitzstein'de
Joint PDF	$\partial^2 F / \partial x \partial y$	2m45
Marjinal	$\int f(x, y) dy$	6m20
Koşullu	$f(x, y) / f_X(x)$	8m30
Joint = K × M	$f = f_{Y X} f_X$	10m38
Bağımsızlık	$f = f_X f_Y$	11m48
Disk	Marjinal \neq uniform; koşullu uniform	16m13
2D LOTUS	$\iint g \cdot f$	22m09
Bağımsız \Rightarrow ilişkisiz	$E(XY) = E(X)E(Y)$; tersi yanlış	24m48
İki uniform	$E\ X - Y\ = 1/3$	34m29
Poisson splitting	$X \perp Y$ Poisson	49m45

26.13 ML Bağlantıları Özeti

💡 7 köprü

1. **Koşullu** $f(y | x) \rightarrow$ denetimli, koşullu üretim.
2. **Joint = \prod koşullu** \rightarrow autoregressive (GPT).
3. **2D LOTUS** \rightarrow çok değişkenli Monte Carlo.
4. **Bağımsız vs ilişkisiz** \rightarrow PCA vs ICA.
5. **Heteroskedastik** \rightarrow girdiye bağlı yayılım.
6. **Max/min** \rightarrow sıra istatistikleri (Ders 25).
7. **Poisson splitting** \rightarrow thinning, A/B trafik.

❗ Tek bir şey alıp gideceksen

Joint/marjinal/koşullu üçlüsü ayrılmaz — $f = f_{Y|X} f_X$. **2D LOTUS** beklentileri dağılım bulmadan verir. **Bağımsız \Rightarrow ilişkisiz ama tersi değil** — korelasyon yalnız doğrusal yakalar.

27 Multinomial ve Cauchy

Softmax dağılımı + ağır kuyrukların habercisi

Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 20: Multinomial and Cauchy](#) (≈49 dk)
- **Okuma süresi:** ≈35 dk

27.1 Bu Derste Ne Var?

1. **Yapıyı kullan:** $Z_1 - Z_2 \sim N(0, 2) \rightarrow E|Z_1 - Z_2| = 2/\sqrt{\pi}$ (2D LOTUS değil).
2. **Multinomial:** binomun k kategoriye genellemesi.
3. **Lumping + koşullu:** birleştir \rightarrow yine Mult; yeniden normalleştir.
4. **Cauchy = X/Y:** ortalama/varyans **YOK**; LLN başarısız.

Builder Notu — ML Köprüleri

- **Yapıyı tamı** \rightarrow reparameterization, kapalı form KL.
- **Multinomial** \rightarrow softmax + cross-entropy, **topic modelleri (LDA)**, sayma.
- **Koşullu yeniden-normalleştirme** ($p_j/(1 - p_1)$) = **masked softmax** (attention'da maskelenen tokenlardan sonra).
- **Cauchy ağır kuyruk** \rightarrow **robust istatistik** (medyan), **Student-t**, finansal getiriler.

27.2 Bağımsız Normal Toplamı = Normal (MGF İspatı)

$X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$ bağımsız:

$$M_{X+Y}(t) = e^{\mu_1 t + \sigma_1^2 t^2 / 2} \cdot e^{\mu_2 t + \sigma_2^2 t^2 / 2} = e^{(\mu_1 + \mu_2)t + (\sigma_1^2 + \sigma_2^2)t^2 / 2}$$

$\rightarrow X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

27.3 İki Normal Uzaklık (Yapıyı Kullan)

$Z_1 - Z_2 \sim N(0, 2) = \sqrt{2} \cdot Z$. Tek boyutlu LOTUS:

$$E|Z| = \sqrt{2/\pi} \Rightarrow E|Z_1 - Z_2| = \sqrt{2} \cdot \sqrt{2/\pi} = \frac{2}{\sqrt{\pi}}$$

“it’s better to stop and think about the structure of the problem” — Blitzstein, 1:36

27.4 Multinomial Dağılım

$\vec{X} = (X_1, \dots, X_k) \sim \text{Mult}(n, \vec{p}), \sum p_j = 1$.

$$P(X_1 = n_1, \dots, X_k = n_k) = \frac{n!}{n_1! n_2! \dots n_k!} p_1^{n_1} \dots p_k^{n_k}$$

! Builder Notu — Softmax + Cross-Entropy

Multinomial **çok-sınıflı sınıflandırmanın doğal dağılımı**: softmax bir olasılık vektörü üretir, gözlemler Multinomial. **Cross-entropy kaybı** = Multinomial log-likelihood’unun negatifi. LDA topic modelleri, dil modelleri.

27.5 Marjinal (Binom) ve Lumping

Marjinal (hikâyeden):

$$X_j \sim \text{Bin}(n, p_j), \quad E(X_j) = np_j, \quad \text{Var}(X_j) = np_j(1 - p_j)$$

Lumping: Kategorileri birleştir \rightarrow yine Multinomial (olasılıklar toplanır).

27.6 Multinomial Koşullu

$(X_2, \dots, X_k) \mid X_1 = n_1 \sim \text{Mult}(n - n_1, (p'_j))$, burada:

$$p'_j = \frac{p_j}{1 - p_1}$$

! Builder Notu — Masked Softmax

$p_j/(1 - p_1)$ yeniden-normalleştirilmesi **tam olarak masked softmax**: attention’da bazı tokenları maskeleyince (0’a çekince), kalanların olasılıkları $(1 - \text{maskelenenin payı})$ ’na bölünür. Kısıtlı üretim, causal attention.

27.7 Cauchy Dağılımı

$T = X/Y$, $X, Y \sim N(0, 1)$ bağımsız.

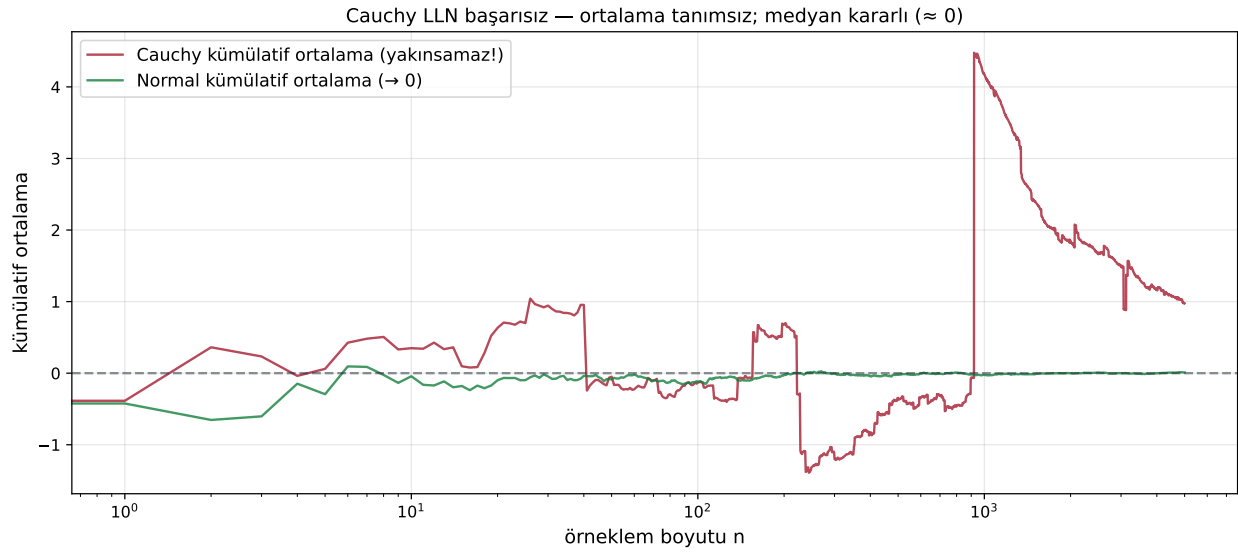
PDF türetimi (Leibniz, integral işareti altında türev):

$$f(t) = \frac{1}{\pi(1+t^2)}, \quad t \in \mathbb{R}$$

Tuhaf özellikler:

- **Ortalama YOK** ($\int t/(\pi(1+t^2))dt$ iraksar).
- **Varyans YOK.**
- **iid Cauchy ortalaması yine Cauchy** — LLN başarısız!

“You can average a million IID Cauchy it’s still gonna be Cauchy.” — Blitzstein, 32:24



Şekil 27.1: Cauchy LLN başarısız: kümülatif ortalama yakınsamaz (mavi). Normal’in kümülatif ortalaması 0’a oturur (yeşil). Cauchy’de ortalama tanımsız; çözüm: medyan (kararlı).

! Builder Notu — Ağır Kuyruk Uyarısı

Cauchy **ağır kuyruklu dağılımların afişidir**: ortalama tanımsız → örneklem ortalaması yanıltıcı → **robust istatistik (medyan)** gerekir. **Student-t** (Ders 30) Cauchy’yi içerir; ağır-kuyruklu gradyan/ödül, finansal getiriler, aykırı-değer-bol verilerde “ortalama al” refleksi tehlikelidir.

27.8 Bu Dersin Özeti

1. **Normal toplam** → MGF çarpımı.
2. **Yapıyı kullan**: $Z_1 - Z_2 = \sqrt{2}Z$.

3. **Multinomial:** binom \rightarrow k kategori.
4. **Marjinal:** $\text{Bin}(n, p_j)$. **Lumping:** birleřtir.
5. **Kořullu:** yeniden normalleřtir = **masked softmax**.
6. **Cauchy:** ortalama/varyans yok, LLN bařarisız.

! Tek bir cümle

Yapıyı tanı — $Z_1 - Z_2$ Normal'dir; **Multinomial** binomun k-kategori genellemesi (marjinal Bin, lumping, kořullu yeniden-normalleřtirme = **masked softmax**); **Cauchy** ortalaması olmayan ağır-kuyruk canavarıdır — robust ölçüler gerekir.

27.9 Kontrol Soruları

i Soru 1: $X, Y \sim N(5, 4)$ bağımsız. (a) $X+Y, X-Y$ dağılımı? (b) $E|X-Y|$?

Cevap: (a) $X + Y \sim N(10, 8), X - Y \sim N(0, 8)$. (b) $\sqrt{8} \cdot \sqrt{2/\pi} = 4/\sqrt{\pi}$.

i Soru 2: Zar 12 atıř. (a) X_6 dağılımı? (b) Çift sayı (2,4,6)?

Cevap: (a) $\text{Bin}(12, 1/6)$. (b) Lumping $\rightarrow \text{Bin}(12, 1/2)$.

i Soru 3: $X_1 = n_1$ verildiğinde X_2 ?

Cevap: $\text{Bin}(n - n_1, p_2/(1 - p_1))$.

i Soru 4: (Builder) Cauchy veride merkez tahmini? Ortalama neden?

Cevap: Ortalama yok, iid ortalaması yine Cauchy \rightarrow yakınsamaz. **Medyan** kullan (simetri sayesinde konum parametresine yakınsar). Finansal getiriler, ağır kuyruklu veride **robust**.

27.10 Egzersizler

Egzersiz 1. $X \sim N(2, 1), Y \sim N(-1, 3)$ bağımsız. $2X - 3Y$ dağılımı?

Egzersiz 2. Anket: $n = 100, p = (0,5, 0,3, 0,2)$. (a) Joint (50, 30, 20). (b) Parti 1 marjinal. (c) Parti 2+3 lumping.

Egzersiz 3. $X_1 = 45$ verildiğinde (X_2, X_3) ?

Egzersiz 4. (Python — *Multinomial + Cauchy*)

X_1 mean=50.01 (teori 50), var=24.86 (teori 25)


Cauchy 'ortalama' (kararsız): 0.185

Cauchy medyan (kararlı): -0.007

Egzersiz 5. (Sonraki ders) Multinomial’de $\text{Cov}(X_i, X_j) < 0$ (toplam n sabit). $k = 2$ (binom): $\text{Cov}(X_1, X_2)$ işareti?

27.11 Sonraki Ders İçin Hazırlık

Ders 21: Kovaryans ve Korelasyon — $\text{Cov}(X, Y)$, korelasyon, Cauchy-Schwarz.


 Ders 21 öncesi yapılacak

- Egzersizleri çöz.
- “Bağımsız $\Rightarrow E(XY) = E(X)E(Y)$ ” hatırla.

27.12 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Blitzstein’de
Normal toplam	$N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$	5m09
Z_1-Z_2	$\sqrt{2}Z, E\ \cdot\ = 2/\sqrt{\pi}$	5m22
Multinomial PMF	$\frac{n!}{\prod n_j!} \prod p_j^{n_j}$	13m08
Marjinal	$\text{Bin}(n, p_j)$	16m40
Lumping	Birleştir \rightarrow yine Mult	19m37
Koşullu	Yeniden normalleştir	23m06
Cauchy	$p_j/(1 - p_1)$ $T = X/Y$	30m24
PDF	$1/(\pi(1 + t^2))$; E/Var	45m12
LLN başarısız	YOK	
	iid Cauchy ortalaması	32m24
	yine Cauchy	

27.13 ML Bağlantıları Özeti

 7 köprü

1. **Normal kapanışı** \rightarrow diffusion, Kalman.
2. **Yapıyı tanı** \rightarrow reparameterization.
3. **Multinomial** \rightarrow softmax, cross-entropy, LDA.
4. **Marjinal Bin** \rightarrow one-vs-rest, sınıf-başına.
5. **Koşullu** \rightarrow masked softmax.
6. **Lumping** \rightarrow coarse-graining, hiyerarşik sınıf.
7. **Cauchy** \rightarrow robust istatistik (medyan), Student-t (Ders 30).

! Tek bir şey alıp gideceksen

Yapıyı tanı. **Multinomial** softmax/cross-entropy temelidir; **koşullu yeniden-normalleştirme = masked softmax**. **Cauchy** ortalaması olmayan ağır kuyruğun uyarısıdır — **medyan kullan.**

28 Kovaryans ve Korelasyon

Toplamın varyansı, $[-1,1]$ = Cauchy-Schwarz = kosinüs benzerliği

Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 21: Covariance and Correlation](#) (≈49 dk)
- **Okuma süresi:** ≈35 dk

28.1 Bu Derste Ne Var?

1. **Kovaryans:** $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$; **bilineer**.
2. **Toplamın varyansı:** $\text{Var}(\sum X_i) = \sum \text{Var}(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j)$.
3. **Bağımsız** \Rightarrow **ilişkisiz** ($\text{Cov} = 0$); **TERSİ YANLIŞ** — Z, Z^2 ilişkisiz ama tam bağımlı.
4. **Korelasyon:** $\text{Cov}/(\sigma_X \sigma_Y) \in [-1, 1]$ (Cauchy-Schwarz) = **kosinüs benzerliği**.

Builder Notu — ML Köprüleri

- **Kovaryans matrisi** $\Sigma \rightarrow$ PCA özvektörleri, **Mahalanobis**, whitening.
- **Bilineerlik** $\rightarrow \text{Cov}(AX) = A\Sigma A^T$; portföy varyansı $w^T \Sigma w$.
- **Toplamın varyansı** \rightarrow **ensemble / bagging** variance reduction; korelasyon bir **taban** koyar.
- **İlişkisiz \neq bağımsız** \rightarrow **PCA vs ICA**; korelasyon yalnız doğrusal.
- **Korelasyon** $[-1, 1]$ = **Cauchy-Schwarz** = **kosinüs benzerliği** (cosine similarity).

28.2 Kovaryans: Tanım ve Bilineerlik

$$\text{Cov}(X, Y) = E[(X - EX)(Y - EY)] = E(XY) - E(X)E(Y)$$

Özellikler:

$$\text{Cov}(X, X) = \text{Var}(X), \quad \text{Cov}(X, c) = 0, \quad \text{Cov}(cX, Y) = c \text{Cov}(X, Y)$$

Bilineerlik:

$$\text{Cov}\left(\sum_i a_i X_i, \sum_j b_j Y_j\right) = \sum_i \sum_j a_i b_j \text{Cov}(X_i, Y_j)$$

💡 Builder Notu — Σ Matrisi

Bilineerlik = **kovaryans matrisi cebirinin** temeli. Doğrusal $Y = AX$ altında $\text{Cov}(Y) = A \text{Cov}(X) A^T$.
Portföy varyansı, doğrusal model belirsizliği, whitening.

28.3 Toplamın Varyansı

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_i \text{Var}(X_i) + 2 \sum_{i<j} \text{Cov}(X_i, X_j)$$

Kovaryanslar 0 \rightarrow varyanslar toplanır. **Ensemble/bagging'in kalbi.**

28.4 Bağımsız \Rightarrow İlişkisiz (Tersi YANLIŞ)

Karşı-örnek: $Z \sim N(0, 1)$, $X = Z$, $Y = Z^2$.

$$\text{Cov}(X, Y) = E(Z^3) - 0 \cdot 1 = 0 \quad (\text{ilişkisiz})$$

Ama $Y = X^2 \rightarrow$ tam bağımlı! Korelasyon yalnızca **doğrusal** ilişkiyi ölçer.

“a common mistake is to show the covariance is 0, and then just leap to the conclusion that they're independent.” — Blitzstein, 19:23

! Builder Notu — PCA vs ICA

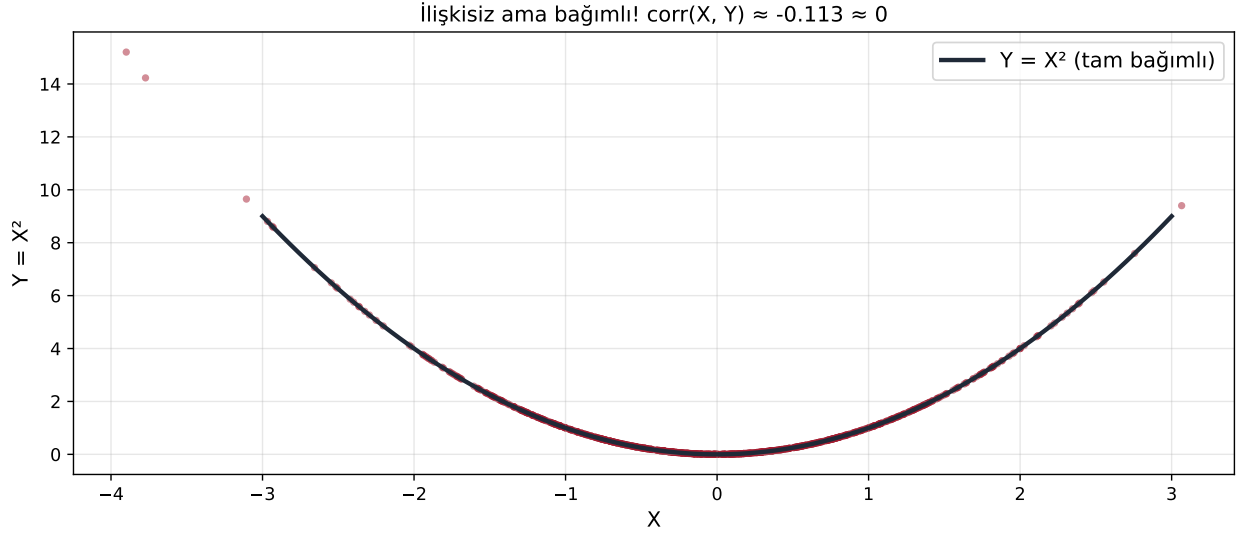
PCA / whitening bileşenleri **ilişkisiz** yapar ama bağımsız yapmaz. **ICA** gerçek bağımsızlığı hedefler. Doğrusal-olmayan bağ için **mutual information, HSIC**, distance correlation gerekir.

28.5 Korelasyon: $[-1, 1]$ ve Cauchy-Schwarz

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Birimsiz. $[-1, 1]$ sınırı = **Cauchy-Schwarz**:

$$\text{Var}(X \pm Y) = 2 \pm 2\rho \geq 0 \Rightarrow -1 \leq \rho \leq 1$$



Şekil 28.1: İlişkiz ama bağımlı: $Y = X^2$. Soldaki saçılım simetrik (negatif/pozitif X eşit, ortalama $X \cdot Y = 0 = \text{Cov}$). Ama $Y = X^2$ 'ye tam oturuyor — bağımlılığın en güçlü hali. Korelasyon doğrusal değil → görmez.

! Builder Notu — Kosinüs Benzerliği

Korelasyon birimsiz + $[-1, 1]$ = **kosinüs benzerliği**: merkezlenmiş vektörlerin normalize iç çarpımı. **Cauchy-Schwarz** ile aynı. Embedding benzerliği, öneri sistemleri, feature selection.

28.6 Multinomial Kovaryansı

$X \sim \text{Mult}(n, \vec{p})$. Toplam n **sabit** → kategoriler yarışır → negatif:

$$\text{Cov}(X_i, X_j) = -np_i p_j \quad (i \neq j)$$

İspat (lumping): $X_1 + X_2 \sim \text{Bin}(n, p_1 + p_2)$. Var'lar eşitle, çöz.

💡 Builder Notu — Softmax Yarışması

Multinomial negatif kovaryansı **softmax çıktılarının doğası**: olasılıklar 1'e toplandığından sınıflar yarışır. Çok-sınıflı modellerin gradyanlarında ve kalibrasyonunda görülür.

28.7 Binom + Hipergeometrik Varyansı (Göstergelerle)

Gösterge kimlikleri: $I_A^2 = I_A$, $I_A I_B = I_{A \cap B}$.

Binom: $X = \sum I_j$ bağımsız → $\text{Var}(X) = \sum \text{Var}(I_j) = npq$.

Hipergeometrik (yerine koymadan): Göstergeler **bağımlı** → negatif Cov + sonlu-popülasyon düzeltmesi.

$$\text{Cov}(I_1, I_2) = \frac{w(w-1)}{(w+b)(w+b-1)} - \left(\frac{w}{w+b}\right)^2 < 0$$

28.8 Bu Dersin Özeti

1. **Cov:** $E(XY) - E(X)E(Y)$; bilineer.
2. **Toplamın varyansı:** $\sum \text{Var} + 2 \sum \text{Cov}$.
3. **Bağımsız** ⇒ **ilişkisiz**, tersi yanlış.
4. **Corr** ∈ [-1, 1] (Cauchy-Schwarz).
5. **Mult Cov:** $-np_i p_j$ (negatif).
6. **Binom Var:** npq . **Hipergeometrik:** bağımlı göstergeler.

! Tek bir cümle

Kovaryans birlikte değişimi ölçer, **toplamın varyansına** izin verir; **korelasyon** birimsiz [-1, 1] = Cauchy-Schwarz = **kosinüs benzerliği**. **Korelasyon yalnız doğrusal** — ilişkisiz ≠ bağımsız; gerçek bağımsızlık (ICA, MI) daha güçlüdür.

28.9 Kontrol Soruları

i Soru 1: $\text{Var}(X)=4, \text{Var}(Y)=9, \text{Cov}=2$. $\text{Var}(2X-3Y)$?

Cevap: $4 \cdot 4 + 9 \cdot 9 + 2 \cdot 2 \cdot (-3) \cdot 2 = 16 + 81 - 24 = 73$.

i Soru 2: $X \in \{-1, 0, 1\}, Y=X^2$. İlişkili mi, bağımlı mı?

Cevap: $E(X) = 0, E(XY) = E(X^3) = 0 \rightarrow \text{Cov} = 0 \rightarrow$ **ilişkisiz**. Ama $Y = X^2 \rightarrow$ tam **bağımlı**.

i Soru 3: $\text{Corr}(X, aX+b)$? $a > 0$ vs $a < 0$?

Cevap: $a/|a| = \pm 1$. Mükemmel doğrusal.

i Soru 4: (Builder) n tahminci, σ^2, ρ . Ortalamanın varyansı? $n \rightarrow \infty$?

Cevap: $\sigma^2/n + (n-1)\rho\sigma^2/n \rightarrow \rho\sigma^2$. $\rho = 0 \rightarrow \sigma^2/n \rightarrow 0$. Ama $\rho > 0 \rightarrow$ **taban $\rho\sigma^2$! Random forest ağaçlarını “decorrelate” etmenin nedeni.**

28.10 Egzersizler

Egzersiz 1. $\text{Var}(X) = 1$, $\text{Var}(Y) = 4$, $\text{Corr} = 0,5$. (a) Cov ? (b) $\text{Var}(X \pm Y)$? (c) $\text{Cov}(X, 2X + Y)$?

Egzersiz 2. Mult: $\text{Corr}(X_i, X_j)$. $k = 2$ (binom): $\text{Corr} = -1$ neden?

Egzersiz 3. $U \sim \text{Unif}(0, 2\pi)$, $X = \cos U$, $Y = \sin U$. (a) $\text{Cov} = 0$. (b) $X^2 + Y^2 = 1 \rightarrow$ bağımlı.

Egzersiz 4. (Python — Cov + İlişkisiz/Bağımlı)

Mult $\text{Cov}(X_1, X_2) \approx -7.492$ teorik -7.5
 $\text{corr}(X, Y) \approx -0.0011$ (≈ 0 , ilişkisiz)
 $X^2 + Y^2 = 1.0000$ ($= 1$, tam bağımlı!)

Egzersiz 5. (Sonraki ders) Bağımsız $X + Y$ PDF'i = **konvolüsyon**: $f_{X+Y}(t) = \int f_X(x)f_Y(t-x)dx$.

28.11 Sonraki Ders İçin Hazırlık

Ders 22: Dönüşümler ve Konvolüsyonlar — change of variables, Jacobian.

⚠ Ders 22 öncesi yapılacak

- Egzersiz 5 (konvolüsyon sezgisi) çöz.
- “Bağımsız toplam MGF = $M_X \cdot M_Y$ ” hatırla.

28.12 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Blitzstein'de
Cov	$E(XY) - E(X)E(Y)$	0m59
Bilineerlik	$\sum \sum a_i b_j \text{Cov}$	10m20
Toplamın Var	$\sum \text{Var} + 2 \sum \text{Cov}$	16m51
Bağımsız \Rightarrow ilişkisiz	Tersi yanlış (Z, Z^2)	18m24
Corr	$\text{Cov}/(\sigma_X \sigma_Y)$, birimsiz	24m28
Sınırlar	$-1 \leq \text{Corr} \leq 1$ (Cauchy-Schwarz)	28m31
Mult Cov	$-np_i p_j$ (negatif)	39m36
Gösterge	$I_A^2 = I_A, I_A I_B = I_{A \cap B}$	42m38
HGeom Var	Sonlu-popülasyon düzeltmesi	44m36

28.13 ML Bağlantıları Özeti

💡 7 köprü

1. Σ matrisi \rightarrow PCA, Mahalanobis, whitening.
2. **Bilineerlik** $\rightarrow A\Sigma A^T$, portföy.
3. **Toplam Var** \rightarrow ensemble; korelasyon = varyans tabanı.
4. **İlişkiziz \neq bağımsız** \rightarrow PCA vs ICA.
5. **Corr = kosinüs benzerliği** \rightarrow embedding, öneri.
6. **Mult yarışması** \rightarrow softmax bağı.
7. **HGeom düzeltme** \rightarrow yerine koymadan örnekleme (anket).

❗ Tek bir şey alıp gideceksen

Cov bilineer; **toplamın varyansı** $= \sum \text{Var} + 2\sum \text{Cov}$; **Corr** = standartlaştırılmış = kosinüs benzerliği.
“**Korelasyon yok**” \neq **bağımsız** — yalnız doğrusal yakalar.

29 Dönüşümler ve Konvolüsyonlar

Change of variables, Jacobian, log-normal, olasılıksal yöntem

Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 22: Transformations and Convolutions](#) (≈48 dk)
- **Okuma süresi:** ≈35 dk

29.1 Bu Derste Ne Var?

1. **Hipergeometrik Var:** $\frac{N-n}{N-1}npq$ — sonlu-popülasyon düzeltmesi.
2. **Change of variables:** $f_Y = f_X|dx/dy|$, çok boyutta **Jacobian**.
3. **Log-normal:** $Y = e^Z$ (Z Normal).
4. **Konvolüsyon:** bağımsız $X + Y \rightarrow \int f_X(x)f_Y(t-x)dx$.
5. **Olasılıksal yöntem:** varlık ispatı.

“Convolution is just the fancy word for sums.” — Blitzstein, 27:02

Builder Notu — ML Köprüleri

- **Change of variables + log-det Jacobian** → **normalizing flows** (RealNVP, Glow) tam log-likelihood.
- **Log-normal** → çarpımsal gürültü, pozitif sağa-çarpık (fiyat, gelir, gecikme); log-dönüşüm ön işleme.
- **Konvolüsyon** → CNN, diffusion gürültü, MGF/Fourier uzayında çarpıma döner.
- **Olasılıksal yöntem** → **Shannon kodlama**, **Johnson-Lindenstrauss**, randomized algoritmalar.

29.2 Hipergeometrik Varyansı

$$\text{Var}(X) = \frac{N-n}{N-1}np(1-p)$$

$np(1-p)$ = binom; öndeki $\frac{N-n}{N-1}$ = **sonlu-popülasyon düzeltmesi**. $N \gg n \rightarrow \approx$ binom.

29.3 Change of Variables (1D)

$Y = g(X)$, g türevlenebilir + kesin monoton:

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|, \quad x = g^{-1}(y)$$


İspat (CDF + zincir kuralı): $F_Y(y) = F_X(g^{-1}(y))$, türev al.

29.4 Log-Normal

$Y = e^Z$, $Z \sim N(0, 1)$. $z = \ln y$, $|dz/dy| = 1/y$:

$$f_Y(y) = \frac{1}{y\sqrt{2\pi}} e^{-(\ln y)^2/2}, \quad y > 0$$


Çarpımsal CLT: bağımsız pozitif faktörlerin çarpımı \rightarrow log normal.

 **Builder Notu — Pozitif Sağa-Çarpık**

Log-normal çarpımsal süreçlerin dağılımı: fiyatlar, gelirler, dosya boyutları, gecikme süreleri. ML'de **log1p** ön işleme yaygın.

29.5 Çok Boyutlu (Jacobian)

$$f_Y(\mathbf{y}) = f_X(\mathbf{x}) \left| \det \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right|$$

 **Builder Notu — Normalizing Flows**

$\log f_Y(\mathbf{y}) = \log f_X(\mathbf{x}) + \log |\det J|$. **RealNVP, Glow, coupling layers** tam olarak bu log-det Jacobian'ı verimli (üçgensel matris \rightarrow köşegen) hesaplayacak şekilde tasarlanır. **Üretici modellemede tam log-likelihood**'u mümkün kılan tek araç (VAE/GAN aksine).

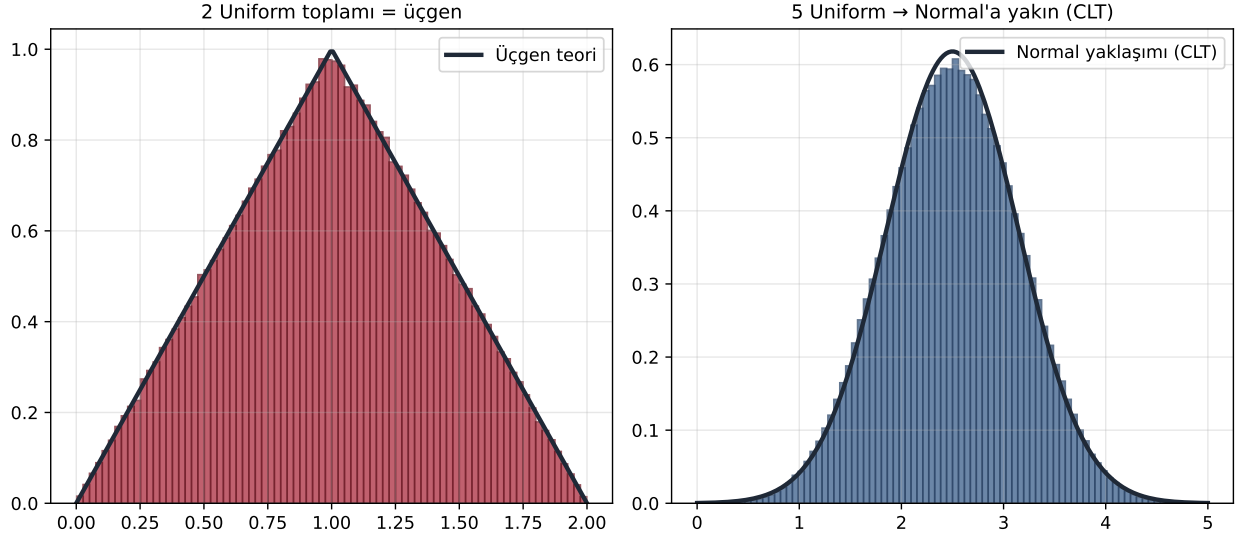
29.6 Konvolüsyon

Bağımsız $T = X + Y$:

Kesikli: $P(T = t) = \sum_x P(X = x)P(Y = t - x)$.

Sürekli:

$$f_T(t) = \int_{-\infty}^{\infty} f_X(x) f_Y(t-x) dx$$



Şekil 29.1: İki bağımsız Uniform(0,1) toplamının PDF’i: üçgen [0,2] üzerinde, tepe t=1’de. Konvolüsyon iki düz PDF’i üçgene düzleştirir. CLT’nin başlangıcı — daha çok uniform topla → Normal.

💡 Builder Notu — Convolution Theorem

MGF/Fourier uzayında konvolüsyon = çarpım (convolution theorem). [Ders 17](#) MGF çarpımı konvolüsyon integralinden kaçınmanın yolu. **Diffusion** ileri süreci konvolüsyondur; **CNN’lerin adı buradan**.

29.7 Olasılıksal Yöntem

Yöntem 1: $P(A) > 0 \Rightarrow \exists \omega : A$.

Yöntem 2 (ortalama): Skoru $\geq E(X)$ olan bir nesne **vardır** (hepsi altında olamaz).

“you can use probability to prove existence” — Blitzstein, 33:38

Örnek — Max-Cut: m kenarlı graf, köşeleri rastgele iki gruba böl. Bir kenarın kesilme olasılığı 1/2. $E(\text{kesilen}) = m/2 \rightarrow$ **kesilen $\geq m/2$ olan bir bölme var** = randomized 1/2-yaklaşım.

❗ Builder Notu — Shannon ve JL

Shannon kodlama teoremi: “iyi kod var”ı rastgele kod seçerek kanıtlar. **Johnson-Lindenstrauss** rastgele projeksiyonun boyut indirgemeyi koruduğunu gösterir. **Randomized algoritmalar, LSH, random search hiperparametre optimizasyonu** hep bu fikre dayanır.

29.8 Bu Dersin Özeti

1. **HGeom Var:** binom $\times (N - n)/(N - 1)$.
2. **CoV (1D):** $f_Y = f_X |dx/dy|$.
3. **Çok boyut:** $|\det J|$.
4. **Log-normal:** $Y = e^Z$.
5. **Konvolüsyon:** $\int f_X(x) f_Y(t - x) dx$.
6. **Olasılıksal yöntem:** $P > 0$ veya $E \geq$ eşik.

! Tek bir cümle

Change of variables ($f_Y = f_X \cdot |J|$) bir dönüşümün tüm dağılımını verir — **normalizing flows'un kalbi**. **Konvolüsyon** bağımsız toplam (MGF'in çarpıma çevirdiği). **Olasılıksal yöntem** nesneyi bulmadan varlığı kanıtlar (Shannon kodlama, JL).

29.9 Kontrol Soruları

i Soru 1: $U \sim \text{Unif}(0,1)$, $Y = -\ln U$. Dağılım?

Cevap: $u = e^{-y}$, $|du/dy| = e^{-y}$. $f_Y = e^{-y} \mathbf{Exp}(1)$ (inverse-transform).

i Soru 2: $X \sim N(0,1)$, $Y = aX + b$ ($a > 0$). PDF?

Cevap: $f_Y(y) = \frac{1}{a\sqrt{2\pi}} e^{-(y-b)^2/(2a^2)} = N(b, a^2)$.

i Soru 3: $X, Y \sim \text{Unif}(0,1)$ bağımsız. $X+Y$ PDF?

Cevap: Üçgen: t ($0 \leq t \leq 1$), $2 - t$ ($1 \leq t \leq 2$).

i Soru 4: (Builder) m kenarlı graf rastgele bölme. Kesilen sayı?

Cevap: $E = m/2$. **Max-Cut** $\geq m/2$ olan bölme var = randomized 1/2-yaklaşım.

29.10 Egzersizler

Egzersiz 1. $X \sim \text{Exp}(1)$, $Y = X^2$. PDF (Weibull).

Egzersiz 2. $X, Y \sim \text{Exp}(1)$ bağımsız. $X + Y$ PDF (Gamma(2, 1)).

Egzersiz 3. $X, Y \sim N(0, 1)$ bağımsız. Kutupsal: R^2, θ — Box-Muller.

Egzersiz 4. (Python — CoV + Convolution + Max-Cut)

Log-normal medyan ≈ 1.001 ($e^{\theta=1}$); ortalama ≈ 1.653 ($e^{(1/2)}=1.649$)
 $P(0.95 < X+Y < 1.05)$ yoğunluk ≈ 0.975 (teori 1)
 $E(\text{kesilen}) \approx 2.999$ (teori $m/2 = 3.0$)

Egzersiz 5. (Sonraki ders) **Beta**(a, b) PDF $\propto x^{a-1}(1-x)^{b-1}$, $[0, 1]$. Uniform = Beta(1,1)? Laplace ardışıklık posteriori Beta?

29.11 Sonraki Ders İçin Hazırlık

Ders 23: Beta Dağılımı — $[0, 1]$, Beta-Binom eşleniği.

⚠ Ders 23 öncesi yapılacak

- Egzersiz 5 (Beta sezgisi) çöz.
- Laplace ardışıklık (Ders 17) hatırla.

29.12 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Blitzstein'de
HGeom Var	$(N - n)/(N - 1) \cdot npq$	5m13
CoV 1D	$f_Y = f_X \ dx/dy\ $	10m12
Jacobian	$f_X \ \det J \ $	25m02
Log-normal	$Y = e^Z$	16m38
Konvolüsyon	$\int f_X(x) f_Y(t - x) dx$	27m02
Olasılıksal yöntem	$P > 0 \Rightarrow \exists$	35m05
Ortalama	Skor $\geq E$ olan var	37m51

29.13 ML Bağlantıları Özeti

💡 7 köprü

1. **CoV + Jacobian** → **normalizing flows**.
2. **Log-normal** → çarpımsal süreçler, $\log 1p$.
3. **Konvolüsyon** → CNN, diffusion.
4. **MGF** ↔ **konvolüsyon** → convolution theorem.
5. **Olasılıksal yöntem** → **Shannon, JL**, randomized algoritma.
6. **Max-Cut** → randomized approximation.
7. **HGeom düzeltmesi** → yerine koymadan örnekleme.

! Tek bir şey alıp gideceksen

CoV / Jacobian = normalizing flows'un matematiği. **Konvolüsyon** = bağımsız toplam. **Olasılıksal yöntem** = nesneyi bulmadan varlık ispatı (Shannon, JL).

30 Beta Dağılımı

Conjugate prior + Thompson sampling

i Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 23: Beta Distribution](#) (≈50 dk)
- **Okuma süresi:** ≈33 dk

30.1 Bu Derste Ne Var?

1. **Beta(a, b):** $f(x) \propto x^{a-1}(1-x)^{b-1}$, $[0, 1]$.
2. **Conjugate prior:** Beta + Bin \rightarrow Beta posterior.
3. **Pseudocount:** a = prior başarı, b = prior başarısızlık.
4. **Bayes bilardosu:** $\int_0^1 \binom{n}{k} x^k (1-x)^{n-k} dx = 1/(n+1)$.
5. **Finans LOTUS:** türev fiyatı = $E(g(S))$.

💡 Builder Notu — ML Köprüleri

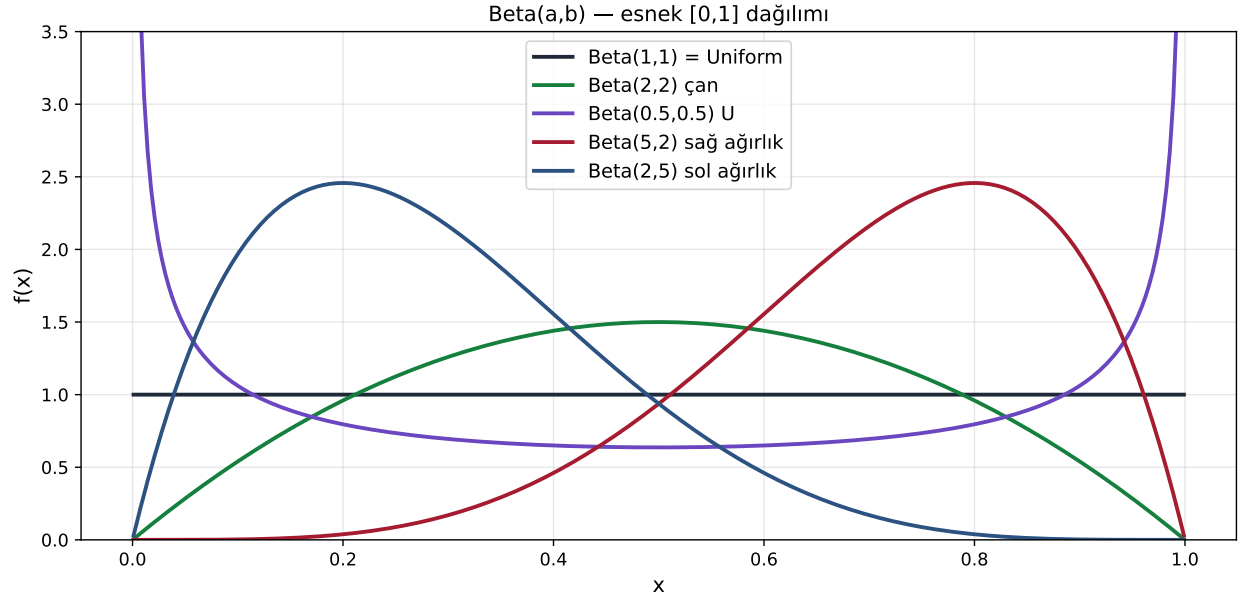
- **Beta-Binom conjugacy** \rightarrow **Thompson sampling, Beta-Bernoulli bandit, Bayesçi A/B.**
- **Pseudocount = Laplace smoothing** (add- α) Bayesçi yüzü; düzenleme.
- **Posterior ortalaması** $(a + X)/(a + b + n)$ = **shrinkage**; küçük-örneklem CTR.
- **Dirichlet** = Beta'nın çok-boyutlusu, **Multinomial**'in eşleniği \rightarrow **LDA topic model.**

30.2 Beta Dağılımı

$$f(x) = c x^{a-1} (1-x)^{b-1}, \quad 0 < x < 1, \quad a, b > 0$$

Şekiller:

- $a = b = 1 \rightarrow$ **Uniform.**
- $a = b = 1/2 \rightarrow$ **U-şekli.**
- $a = b = 2 \rightarrow$ **çan.**
- $a > b \rightarrow$ sağa kayar.



Şekil 30.1: Beta(a,b) esnek aile: a=b=1 uniform, a=b=2 çan, a=b=0.5 U, a=5/b=2 sağa, a=2/b=5 sola. Bayeşçi prior'da “ne kadar eminim” + “hangi yönde” — pseudocount yorumu.

30.3 Beta-Binom Eşleniği

$X | p \sim \text{Bin}(n, p)$, $p \sim \text{Beta}(a, b)$. Bayes (orantı):

$$f(p | X = k) \propto p^k (1-p)^{n-k} \cdot p^{a-1} (1-p)^{b-1} = p^{a+k-1} (1-p)^{b+n-k-1}$$

$$p | X \sim \text{Beta}(a + X, b + n - X)$$

! Builder Notu — Thompson Sampling

Bayeşçi online öğrenmenin motoru: her başarı $a += 1$, her başarısızlık $b += 1$ — kapalı form.
Thompson sampling: posterior'dan örnekle, **arg max** seç. **Beta-Bernoulli bandit + Bayeşçi A/B test**'in çekirdeği. Keşif-sömürü dengesi otomatik.

30.4 Pseudocount Yorumu

- a = prior başarı sayısı (hayalî).
- b = prior başarısızlık sayısı.
- Veri X başarı, $n - X$ başarısızlık ekler.

Ders 17 Laplace ardışıklık prior'ı Beta(1,1) = Uniform: $(n + 1)/(n + 2) = \text{Beta}(1,1)$ ile aynı yöntem.

💡 Builder Notu — Laplace Smoothing

Add- α smoothing = Beta(α , α) prior. $a + b$ = “etkin örnek sayısı” = ne kadar veri ön bilgiye eşdeğer.
Düzenleştirimin olasılıksal yüzü.

30.5 Bayes Bilardosu

$\int_0^1 \binom{n}{k} x^k (1-x)^{n-k} dx = ?$ — **kalkülüssüz:**

$n + 1$ top, birini pembe boya. $[0,1]$ 'e bağımsız uniform at. Pembenin solundaki beyaz sayısı X .

- **Hikâye 1:** Pembe konumuna koşulla $\rightarrow X \sim \text{Bin}(n, p) \rightarrow P(X = k) = \int \binom{n}{k} p^k (1-p)^{n-k} dp$.
- **Hikâye 2:** Topları at, sonra rastgele birini pembe boya $\rightarrow X$ üzerinde uniform $\rightarrow P(X = k) = 1/(n + 1)$.

İkisi aynı:

$$\int_0^1 \binom{n}{k} x^k (1-x)^{n-k} dx = \frac{1}{n+1}$$

30.6 Momentler ve Shrinkage

$$E(X) = \frac{a}{a+b}$$

Posterior ortalaması:

$$E(p | X) = \frac{a + X}{a + b + n}$$

Prior + veri harmanı (**shrinkage**) — az veride prior'a, çok veride veriye yakın.

30.7 Finans LOTUS (Konuk)

Türev fiyatı = $E(g(S))$ — LOTUS! **Alım opsiyonu:** $g(s) = \max(s - K, 0)$. Log-normal S altında integral = **Black-Scholes**.

“you can get a Nobel Prize in economics just for knowing LOTUS.” — Steven (konuk), 49:35

💡 Builder Notu — Monte Carlo Pricing

Türev fiyatı = $E(g(S))$ = Monte Carlo fiyatlama: S 'yi örnekle, $g(S)$ 'yi ortalama. RL'deki $E[\text{ödül}]$ ile aynı mekanizma.

30.8 Bu Dersin Özeti

1. **Beta(a, b):** esnek $[0, 1]$.
2. **Conjugate:** Beta + Bin \rightarrow Beta($a + X, b + n - X$).
3. **Pseudocount** yorumu.
4. **Bayes bilardosu:** $1/(n + 1)$.
5. **Momentler:** $E = a/(a + b)$, posterior shrinkage.
6. **Finans:** fiyat = $E(g(S))$ (LOTUS).

! Tek bir cümle

Beta(a, b) $[0, 1]$ üzerinde esnek + **binom'un eşleniği**: posterior = Beta($a + X, b + n - X$). $a, b =$ **pseudocount** — **Laplace smoothing, Thompson sampling, Bayesçi A/B**'nin temeli.

30.9 Kontrol Soruları

i Soru 1: Beta(1,1), Beta(0.5, 0.5), Beta(5,2), Beta(2,5) şekilleri?

Cevap: Uniform, U-şekli, sağa ağırlık ($E=5/7$), sola ağırlık ($E=2/7$).

i Soru 2: Beta(3,3) prior + 8/2 \rightarrow posterior? Ortalama?

Cevap: Beta(11, 5), $E = 11/16 = \mathbf{0,6875}$ (prior 0.5 + veri 0.8 harmanı).

i Soru 3: Uniform prior + $k/n \rightarrow$ posterior + sonraki başarı?

Cevap: Beta($1 + k, 1 + n - k$); $E = (k + 1)/(n + 2) =$ **Laplace ardışıklık**.

i Soru 4: (Builder) A=5/50, B=8/50 \rightarrow Thompson?

Cevap: A \sim Beta(6, 46), B \sim Beta(9, 43). Posterior'lardan örnekle \rightarrow max'ı seç. B "muhtemelen daha iyi" ama belirsizlik nedeniyle A bazen seçilir — **keşif-sömürü dengesi otomatik**.

30.10 Egzersizler

Egzersiz 1. Beta(2, 8) prior. (a) E? (b) 20/12 \rightarrow posterior + E. (c) Yorumla.

Egzersiz 2. $\int_0^1 x^3(1 - x)^2 dx$ Bayes bilardosu ile = 1/60.

Egzersiz 3. Dönüşüm prior'ı: $E = 0,3$. $a + b$ büyütme belirsizliği nasıl etkiler?

Egzersiz 4. (Python — Conjugacy + Thompson)

Posterior $E = 0.6875$ (teori 11/16)

$P(B > A) = 0.8045$ (B daha iyi görünüyor ama belirsiz)

Egzersiz 5. (Sonraki ders) **Gamma fonksiyonu:** $\Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x} dx$, $\Gamma(n) = (n-1)!$. $\text{Gamma}(n, \lambda)$ = n bağımsız $\text{Exp}(\lambda)$ toplamı.

30.11 Sonraki Ders İçin Hazırlık

Ders 24: Gamma Dağılımı + Poisson Süreci — Exp 'in genellemesi, çağrı merkezi.

⚠ Ders 24 öncesi yapılacak

- Egzersiz 5 (Γ fonksiyonu) çöz.
- Exp + konvolüsyon hatırla.

30.12 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Blitzstein'de
Beta(a,b)	$x^{a-1}(1-x)^{b-1}$	3m31
Esneklik	Uniform/U/çan/çarpık	4m47
Conjugacy	Beta + Bin \rightarrow Beta($a + X, b + n - X$)	7m30
Pseudocount	a, b = prior başarı/başarısızlık	15m48
Bayes bilardosu	$f = 1/(n+1)$	19m17
E(Beta)	$a/(a+b)$	16m50
Finans LOTUS	Türev = $E(g(S))$	34m15

30.13 ML Bağlantıları Özeti

💡 7 köprü

1. **Conjugacy** \rightarrow **Thompson sampling**, Bayesçi A/B.
2. **Pseudocount** \rightarrow **Laplace smoothing**.
3. **Posterior shrinkage** \rightarrow küçük-örneklem CTR.
4. **Esnek** $[0, 1]$ \rightarrow kalibrasyon, belirsizlik.
5. **Bayes bilardosu** \rightarrow exchangeability, de Finetti.
6. **Finans LOTUS** \rightarrow Monte Carlo pricing.
7. **Dirichlet** \rightarrow **LDA**, kategorik prior.

! Tek bir şey alıp gideceksen

Beta binomun eşleniği — posterior kapalı form. **Pseudocount** = Laplace smoothing. **Thompson sampling** keşif-sömürü dengesini otomatik sağlar.

31 Gamma Dağılımı ve Poisson Süreci

Üstel'in genellemesi, n. olaya kadar süre

Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 24: Gamma distribution and Poisson process](#) (≈49 dk)
- **Okuma süresi:** ≈34 dk

31.1 Bu Derste Ne Var?

1. **Gamma fonksiyonu:** $\Gamma(n) = (n - 1)!$, $\Gamma(a + 1) = a\Gamma(a)$, $\Gamma(1/2) = \sqrt{\pi}$.
2. **Gamma(a, λ):** Üstel'in genellemesi.
3. **Poisson süreci:** iid Üstel toplamı = Gamma.
4. **Momentler:** $E = a/\lambda$, $\text{Var} = a/\lambda^2$.

“continuous time analog of negative binomial.” — Blitzstein, 31:31

Builder Notu — ML Köprüleri

- **iid Üstel toplamı = Gamma** → **Erlang** (kuyruk teorisi M/Er/1, n-aşamalı servis).
- **Gamma = conjugate prior** Poisson λ ve Normal precision ($1/\sigma^2$) için.
- **Poisson süreci** → olay varışları, nokta süreçleri.
- $\chi^2 = \text{Gamma'nın özel hâli}$ (Ders 30).
- **Stirling + log-Gamma** → log-faktoriyel, softmax/Dirichlet sabitleri (lgamma).

31.2 Gamma Fonksiyonu

$$\Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x} dx, \quad a > 0$$

Özellikler:

$$\Gamma(n) = (n - 1)!, \quad \Gamma(a + 1) = a\Gamma(a), \quad \Gamma(1/2) = \sqrt{\pi}$$

Stirling formülü:

$$n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$$

31.3 Gamma Dağılımı

$$X \sim \text{Gamma}(a, \lambda) : f(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}, \quad x > 0$$

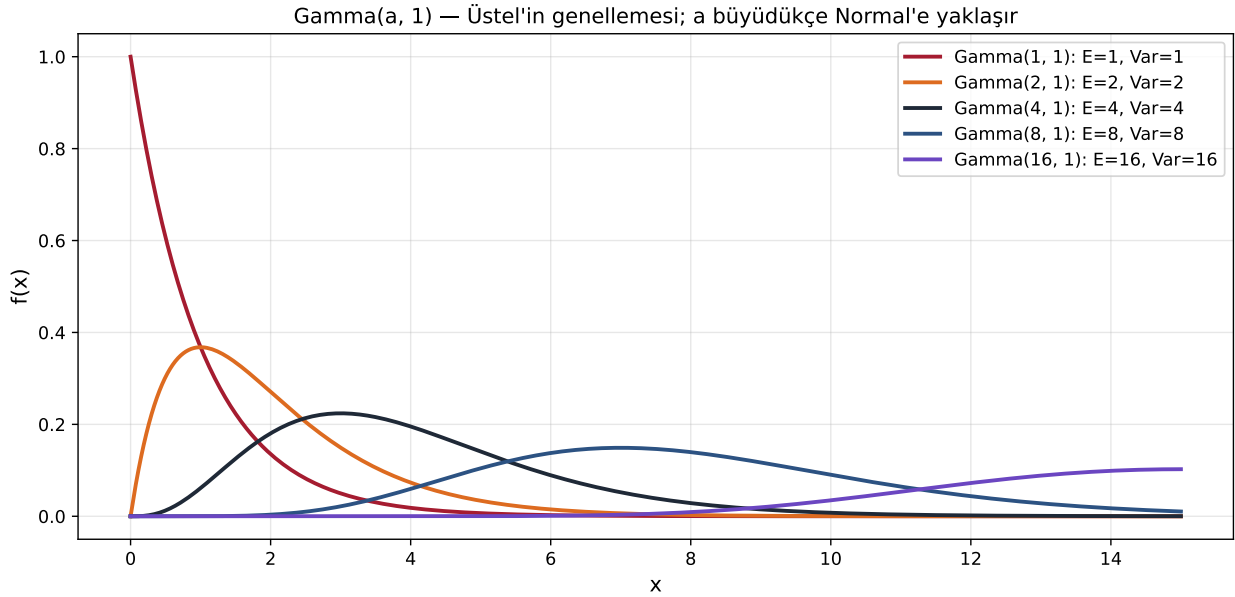
Özel: $\text{Gamma}(1, \lambda) = \text{Üstel}(\lambda)$.

31.4 Poisson Süreci → Gamma

$N_t \sim \text{Pois}(\lambda t)$. Inter-arrival iid $\text{Üstel}(\lambda)$. n . olaya kadar süre:

$$T_n = \sum_{j=1}^n X_j, \quad X_j \sim \text{Exp}(\lambda) \Rightarrow T_n \sim \text{Gamma}(n, \lambda)$$

MGF ispatı: $M_{X_j}(t) = 1/(1-t) \rightarrow M_{T_n}(t) = (1-t)^{-n} = \text{Gamma}(n, 1)$ MGF.



Şekil 31.1: $\text{Gamma}(a, \lambda=1)$ PDF. $a=1$ Üstel; a büyüdükçe tepe sağa kayar ve şekil simetrikleşir (CLT'ye doğru). n bağımsız Üstel toplamı Gamma'nın hikâyesidir.

! Builder Notu — Erlang ve Kuyruk Teorisi

Erlang dağılımı (tam sayı a için Gamma): n -aşamalı servisin toplam süresi. **M/Er/1 kuyrukları**, **telefon trafiği**, **güvenilirlik analizi**. Aşamalara bölmek varyansı $\frac{1}{n}$ kadar küçültür → daha öngörülebilir

servis.

31.5 Momentler

$$E(X^c) = \frac{\Gamma(a+c)}{\Gamma(a)}$$

$c = 1$: $E(X) = a$ (Gamma(a,1)). $c = 2$: $E(X^2) = a(a+1)$, $\text{Var} = a$.

Genel Gamma(a, λ):

$$E(X) = \frac{a}{\lambda}, \quad \text{Var}(X) = \frac{a}{\lambda^2}$$

31.6 Bu Dersin Özeti

1. $\Gamma(a)$: $(n-1)!$, $\Gamma(1/2) = \sqrt{\pi}$, Stirling.
2. **Gamma(a, λ)**: Üstel genelleme.
3. **Poisson süreci**: $T_n \sim \text{Gamma}(n, \lambda)$.
4. **MGF**: $(1 - t/\lambda)^{-a}$.
5. **Momentler**: a/λ , a/λ^2 .

! Tek bir cümle

Gamma(a, λ) Üstel'in genellemesi ve Poisson sürecinde **n. olaya kadar süre**. $\Gamma(a)$ faktöriyeli reel sayılara taşır, momentleri kapalı form $(\Gamma(a+c)/\Gamma(a))$. **Erlang, χ^2 , Bayesçi conjugate prior**'ların temeli.

31.7 Kontrol Soruları

i Soru 1: $\Gamma(5)$? $\Gamma(7/2)$?

Cevap: $4! = 24$; $\Gamma(7/2) = (5/2)(3/2)(1/2)\sqrt{\pi} = (15/8)\sqrt{\pi} \approx 3,32$.

i Soru 2: $\lambda=2/\text{saat}$. T_3 ?

Cevap: Gamma(3, 2). $E = 1,5$ saat, $\text{Var} = 0,75$.

i Soru 3: $E(1/X)$, $X \sim \text{Gamma}(a, 1)$?

Cevap: $\Gamma(a-1)/\Gamma(a) = 1/(a-1)$. $a > 1$ gerekli.

i Soru 4: (Builder) 3 aşamalı servis, Exp(1) her aşama. Toplam?

Cevap: Gamma(3, 1) = Erlang(3). $E = 3$ sn, $\text{Var} = 3$. $\text{Variance/mean}^2 = 1/3 < 1$ (Üstel). Aşamalara bölmek **daha öngörülebilir** servis.

31.8 Egzersizler

Egzersiz 1. $\Gamma(6), \Gamma(9/2)$. $\Gamma(a + 1) = a\Gamma(a)$ parçalı integrasyonla.

Egzersiz 2. Web $\lambda = 10/\text{dk}$. (a) 5. isteğe süre. (b) İlk 2 dakikadaki sayı. (c) Inter-arrival.

Egzersiz 3. $E(X^c)$ genel Gamma(a, λ).

Egzersiz 4. (Python — Gamma)

T mean ≈ 1.999 teorik $n/\lambda = 2.0$

T var ≈ 1.004 teorik $n/\lambda^2 = 1.0$

Gamma(3.5,1): mean=3.498, var=3.498 (teori 3.5)

Egzersiz 5. (Sonraki ders) $\max(U_1, \dots, U_n)$ CDF = x^n , PDF = $nx^{n-1} \rightarrow \text{Beta}(n, 1)$. $\min \sim \text{Beta}(1, n)$.

31.9 Sonraki Ders İçin Hazırlık

Ders 25: Sıra İstatistikleri ve Koşullu Beklenti — min, max, medyan; Beta bağlantısı.

⚠ Ders 25 öncesi yapılacak

- Egzersiz 5 (min/max Beta) çöz.
- Beta'yı + CDF \rightarrow PDF hatırla.

31.10 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Blitzstein'de
$\Gamma(a)$	$\int x^{a-1} e^{-x} dx;$ $\Gamma(n) = (n-1)!$	8m44
$\Gamma(1/2)$	$\sqrt{\pi}$	14m41
Stirling	$n! \approx \sqrt{2\pi n} (n/e)^n$	5m27
Gamma(a, λ)	Üstel genelleme	20m08
Poisson süreci	$T_n \sim \text{Gamma}(n, \lambda)$	30m50
Gamma MGF	$(1 - t/\lambda)^{-a}$	35m35
Momentler	$\Gamma(a+c)/\Gamma(a);$ $E = a/\lambda, \text{Var} = a/\lambda^2$	45m32

31.11 ML Bağlantıları Özeti

💡 7 köprü

1. **Üstel toplamı = Gamma** → **Erlang**, kuyruk teorisi.
2. **Gamma conjugate prior** → Poisson λ , Normal precision.
3. **Poisson süreci** → olay varışları.
4. **Stirling / lgamma** → softmax/Dirichlet, entropi.
5. **Gamma toplanabilirliği** → pseudocount toplama.
6. χ^2 = yarım-tam parametre Gamma (Ders 30).
7. **2-parametre esnekliği** → overdispersion modelleme.

! Tek bir şey alıp gideceksen

Gamma = Üstel genellemesi = Poisson sürecinde n. olaya süre. **Erlang**, χ^2 , **Bayesçi conjugate prior**'ların temeli.

32 Sıra İstatistikleri ve Koşullu Beklenti

Banka-postane, kuantiller, $E(X|A)$, LOTE

i Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 25: Order Statistics and Conditional Expectations](#) (≈48 dk)
- **Okuma süresi:** ≈35 dk

32.1 Bu Derste Ne Var?

1. **Banka-postane:** Gamma → Beta + bağımsızlık.
2. **Beta sabiti:** $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a + b)$.
3. **Sıra istatistikleri:** min, max, medyan; CDF, PDF.
4. **Uniform sıra:** $U_{(j)} \sim \text{Beta}(j, n - j + 1)$.
5. **Koşullu beklenti:** $E(X|A)$, LOTE.

💡 Builder Notu — ML Köprüleri

- **Gamma → Beta/Dirichlet (toplam \perp oran) → LDA örnekleme.**
- **E(orani) \neq oran(E) → ratio estimator bias** (self-normalized importance sampling).
- **Sıra istatistikleri** → kuantil regresyon, **robust** (medyan), top-k, k-NN.
- **Koşullu beklenti** → regresyon $E(Y|X)$, **RL value fonksiyonu** $V(s)$.
- **LOTE** → Bellman denklemi $V(s) = r + E(V(s'))$.

32.2 Banka-Postane: Gamma → Beta

$X \sim \text{Gamma}(a, \lambda)$, $Y \sim \text{Gamma}(b, \lambda)$ bağımsız. $T = X + Y$, $W = X/(X + Y)$.

Jacobian ile joint:

$$f_{T,W}(t, w) = \underbrace{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} w^{a-1} (1-w)^{b-1}}_{\text{Beta}(a,b)} \cdot \underbrace{\frac{1}{\Gamma(a+b)} t^{a+b-1} e^{-t}}_{\text{Gamma}(a+b)}$$

Çarpanlara ayrıldı $\rightarrow T \perp W$ (şaşırtıcı!).

Beta normalleştirme bedavaya:

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

32.3 E(oran) = oran(E) Tuzağı

$W \perp T$ olduğundan:

$$E(W) \cdot E(T) = E(W \cdot T) = E(X) \Rightarrow E(W) = \frac{E(X)}{E(X+Y)} = \frac{a}{a+b}$$

Genelde YANLIŞ! Burada **istisnai** olarak bağımsızlık sayesinde.

“E of numerator over E of denominator, usually that’s completely wrong.” — Blitzstein, 20:34

! Builder Notu — Ratio Estimator Bias

$E(\text{pay/payda}) \neq E(\text{pay})/E(\text{payda})$ genelde. **Self-normalized importance sampling**, normalize ödüller — pay/payda ilişkiliyse yanlılık. **Jensen / delta-method** düzeltmesi gerekir.

32.4 Sıra İstatistikleri

$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$, iid'den doğsa da **bağımlı**.

CDF (binom):

$$F_{X_{(j)}}(x) = P(\text{en az } j \text{ tane } \leq x) = \sum_{k=j}^n \binom{n}{k} F(x)^k (1-F(x))^{n-k}$$

PDF (resimle, doğrudan):

$$f_{X_{(j)}}(x) = n \binom{n-1}{j-1} F(x)^{j-1} (1-F(x))^{n-j} f(x)$$

💡 Builder Notu — Kuantil & Robust

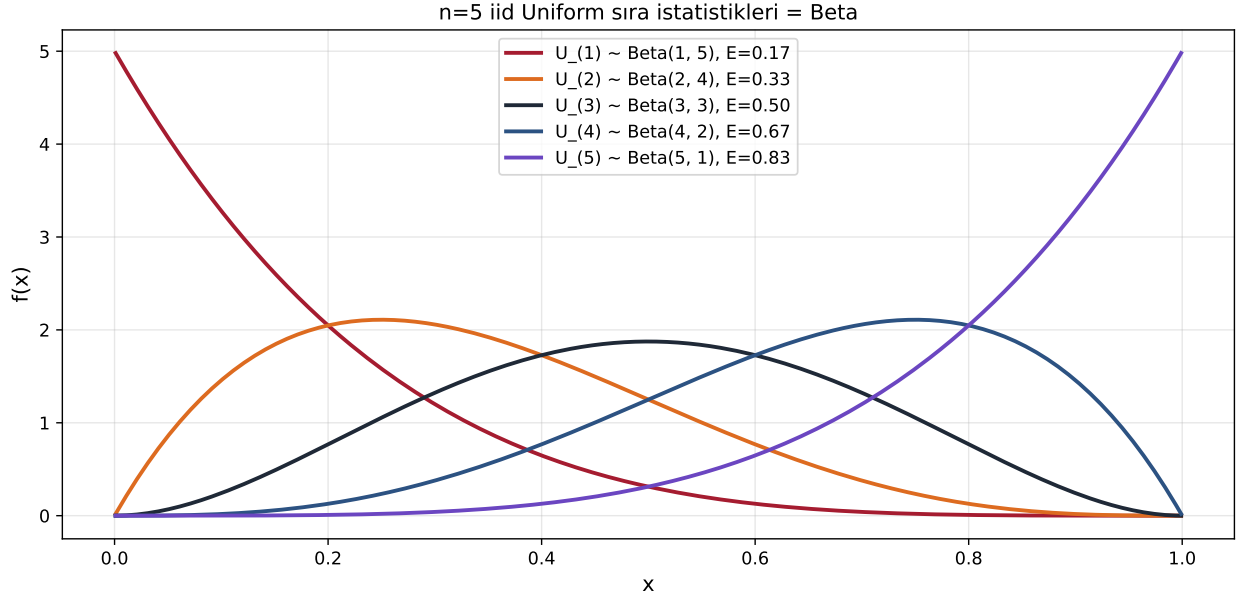
Medyan, IQR, kuantil regresyon (pinball loss), top-k, k-NN mesafeleri. Ekstrem değer teorisi (nadir olay, risk). “iid \rightarrow bağımlı” uyarısı: sıralanmış verilerde kovaryans!

32.5 Uniform Sıra İstatistikleri = Beta

$U \sim \text{Unif}(0,1)$, $F(x) = x$, $f(x) = 1$:

$$f_{U_{(j)}}(x) = n \binom{n-1}{j-1} x^{j-1} (1-x)^{n-j} \Rightarrow U_{(j)} \sim \text{Beta}(j, n-j+1)$$

Max $\sim \text{Beta}(n, 1)$, **min** $\sim \text{Beta}(1, n)$.



Şekil 32.1: n=5 iid Uniform sıra istatistikleri. $U_{(1)}$ min (Beta(1,5)) sola yakın, $U_{(3)}$ medyan (Beta(3,3)) ortada simetrik, $U_{(5)}$ max (Beta(5,1)) sağa. Sıralı veride kuantil dağılımı = Beta.

32.6 Koşullu Beklenti ve LOTE

$E(X|A) = A$ verildiğinde X 'in beklentisi. **Toplam beklenti yasası (LOTE):**

$$E(X) = E(X | A)P(A) + E(X | A^c)P(A^c)$$

Örnek (Geometrik): Zar atışı, ilk 6'ya kadar X . İlk atışa koşulla:

$$E(X) = 1 \cdot \frac{1}{6} + (1 + E(X)) \cdot \frac{5}{6} \Rightarrow E(X) = 6$$

! Builder Notu — Bellman'ın Temeli

Koşullu beklenti $E(Y|X) = \text{ML'de regresyon}$ (en iyi kare-hata tahmincisi). **RL'de değer fonksiyonu** $V(s) = E(\text{getiri}|s)$. **LOTE / ilk-adım analizi = Bellman denklemleri** $V(s) = r + E(V(s'))$ — değer

iterasyonu.

32.7 Bu Dersin Özeti

1. **Banka-postane:** $T \perp W, W \sim \text{Beta}, T \sim \text{Gamma}$.
2. **Beta sabiti:** $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a + b)$.
3. **Sıra ist.:** PDF = $n \binom{n-1}{j-1} F^{j-1} (1 - F)^{n-j} f$.
4. **Uniform:** $U_{(j)} \sim \text{Beta}(j, n - j + 1)$.
5. **LOTE:** $E(X) = \sum E(X|A_i)P(A_i)$.

! Tek bir cümle

Bağımsız Gamma'lar **toplama** \perp **oran** \rightarrow Beta'nın Γ -temelli sabiti. **Sıra istatistikleri** kuantillerin matematiği (Uniform \rightarrow Beta). **Koşullu beklenti + LOTE = Bellman'ın atası**, regresyonun tanımı.

32.8 Kontrol Soruları

i Soru 1: $X \sim \Gamma(3, \lambda), Y \sim \Gamma(2, \lambda)$. (a) T ? (b) W ? (c) $E(W)$? $T \perp W$?

Cevap: (a) $\Gamma(5, \lambda)$. (b) $\text{Beta}(3, 2)$. (c) $3/5$. $T \perp W \checkmark$.

i Soru 2: iid F . (a) max CDF? (b) min CDF?

Cevap: (a) $F(x)^n$. (b) $1 - (1 - F(x))^n$.

i Soru 3: U_1, \dots, U_5 iid Unif. $U_{(3)}$ (medyan)?

Cevap: $\text{Beta}(3, 3)$. $E = 1/2$.

i Soru 4: (Builder) Zar, ilk 6'ya kadar atış? LOTE.

Cevap: $E(X) = (1/6) \cdot 1 + (5/6)(1 + E(X)) \rightarrow E(X) = 6$. **First-step = Bellman**.

32.9 Egzersizler

Egzersiz 1. $X \sim \Gamma(2, 1), Y \sim \Gamma(3, 1)$. (a) $X/(X + Y)$ dağılımı + E . (b) $X + Y$. (c) $\text{Cov}(T, W)$?

Egzersiz 2. iid $\text{Exp}(\lambda)$. $\text{Min} \sim \text{Exp}(n\lambda)$? “En hızlı sunucu” sezgisi.

Egzersiz 3. U_1, \dots, U_{10} iid Unif. $U_{(3)}, U_{(10)}$ E ?


Egzersiz 4. (Python — Banka-postane + Uniform sıra)

$$\begin{aligned} \text{corr}(T, W) &= 0.0015 (\approx 0, \text{bağımsız}) \\ E(W) &= 0.4005 (\text{teori } 2/5 = 0.4) \\ E(U_{(3)}) &= 0.2726 (\text{teori } 3/11 = 0.2727) \end{aligned}$$

Egzersiz 5. (Sonraki ders) $E(Y|X)$ bir RV (X 'in fonksiyonu). **Tower:** $E(E(Y|X)) = E(Y)$.

32.10 Sonraki Ders İçin Hazırlık

Ders 26: Koşullu Beklentiye Devam — $E(Y|X)$ rastgele değişken; tower property.


 Ders 26 öncesi yapılacak

- Egzersiz 5 ($E(Y|X)$ RV) çöz.
- LOTE + first-step analizini hatırla.

32.11 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Blitzstein'de
Banka-postane	$T \perp W$; Gamma + Beta	0m57
B(a,b)	$\Gamma(a)\Gamma(b)/\Gamma(a+b)$	15m39
E(oran) tuzağı	İstisna: $W \perp T$	17m22
Sıra ist.	iid \rightarrow bağımlı	24m41
CDF $X_{(j)}$	$\sum \binom{n}{k} F^k (1-F)^{n-k}$	34m32
PDF $X_{(j)}$	$n \binom{n-1}{j-1} F^{j-1} (1-F)^{n-j} f$	39m05
$U_{(j)}$	Beta($j, n-j+1$)	42m24
LOTE	$E = \sum E(X A_i)P(A_i)$	45m01

32.12 ML Bağlantıları Özeti

 7 köprü

1. **Gamma \rightarrow Beta/Dirichlet \rightarrow LDA.**
2. **$E(\text{oran}) \neq \text{oran}(E) \rightarrow$ ratio estimator bias.**
3. **Sıra ist.** \rightarrow kuantil, robust, top-k.
4. **$U_{(j)}$ \rightarrow Beta \rightarrow kuantil belirsizliği.**
5. **Ekstrem değer** \rightarrow risk, nadir olay.
6. **$E(Y|X)$ \rightarrow regresyon, RL value.**
7. **LOTE = Bellman atası.**

! Tek bir şey alıp gideceksen

Banka-postane ($T \perp W$) Beta sabitini verir. **Sıra istatistikleri** kuantillerin matematiği. **LOTE** Bellman'ın atası.

33 Koşullu Beklentiye Devam

$E(Y|X)$ bir RV; Adam yasası; iki zarf paradoksu

i Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 26: Conditional Expectation Continued](#) (≈50 dk)
- **Okuma süresi:** ≈36 dk

33.1 Bu Derste Ne Var?

1. **İki zarf paradoksu** — koşulu düşürme tuzağı.
2. **Yazı-tura:** $E(W_{HT}) = 4 \neq E(W_{HH}) = 6$.
3. **$E(Y|X) = g(X)$** — bir RV!
4. **Adam yasası:** $E(E(Y|X)) = E(Y)$.

💡 Builder Notu — ML Köprüleri

- **$E(Y|X) = \text{regresyon fonksiyonu}$** = MSE'yi minimize eden tahminci.
- **$E(Y|X)$ bir RV** → **RL değer fonksiyonu** $V(s)$, conditional generation.
- **Koşulu düşürme = bağımsızlık** → selection/collider bias.
- **Yazı-tura örüntüleri** → dizi motifleri (DNA, n-gram), Markov/otomat (KMP).
- **Adam yasası** → **Bellman**, nested Monte Carlo.

33.2 İki Zarf Paradoksu

İki zarf, biri diğerinin 2 katı. Seç (içinde X), diğeri Y . “Argüman 2” (hatalı):

$$E(Y) = \frac{1}{2}E(2X) + \frac{1}{2}E(X/2) = \frac{5}{4}E(X)$$

Çelişki: simetri $E(Y) = E(X)$ doğru. **Hata:** $Y = 2X$ bilgisini koyduk, sonra koşulu unuttuk. X ile “hangi zarf büyük” göstergesi **bağımlı**.

“The only time when we can get rid of the stuff we’re conditioning on is when we know we have independence.” — Blitzstein, 12:11

! Builder Notu — Selection / Collider Bias

“Bilgiyi yerine koy ama koşulu düşür” hatası = ML’de **gizli seçim yanlılığı**, **collider’a koşullama** (Berkson), **sağkalım yanlılığı**. Koşul yalnız **bağımsızlık** varsa düşer.

33.3 Yazı-Tura Örüntüleri

$E(W_{HT})$: İlk H bekle (2) + ilk T bekle (2) = 4.

$E(W_{HH})$: HH kendiyle örtüşür — H sonrası T silinir!

$$E(W_{HH}) = 6$$

“since the HHs are more clumped, those clumps must be further apart.” — Blitzstein, 27:12

💡 Builder Notu — String Matching

Dizi motifleri (DNA, n-gram). Örüntü bekleme = **Markov zinciri/otomat** (durum = eşleşen önek). HH overlap = **KMP** önek-fonksiyonu.

33.4 $E(Y|X)$ Bir Rastgele Değişken

$E(Y|X = x) = g(x)$, x ’in fonksiyonu. **Big X** ile değiştirince:

$$E(Y | X) = g(X) \quad (\text{rastgele değişken!})$$

“ E of Y given X is a random variable. It’s a random variable that’s a function of x .” — Blitzstein, 37:15

! Builder Notu — Regresyon Fonksiyonu

$E(Y|X) =$ **kare-hata MSE’yi minimize eden tahminci**. MSE eğitilmiş modeller (linear regression, sinir ağları) örtük olarak bunu öğrenir. **RL değer fonksiyonu** $V(s) = E(\text{getiri}|s)$ da bir RV.

33.5 Koşullu Beklenti Özellikleri

Doğrusallık (her zaman):

$$E(Y_1 + Y_2|X) = E(Y_1|X) + E(Y_2|X)$$

Bilineni dışarı al:

$$E(h(X)|X) = h(X), \quad E(h(X) \cdot Y|X) = h(X) \cdot E(Y|X)$$

Bağımsızlık:

$$X \perp Y \Rightarrow E(Y|X) = E(Y) \text{ (sabit)}$$

33.6 Poisson Örneği

$X, Y \sim \text{Pois}(\lambda)$ bağımsız.

$$E(X + Y|X) = X + \lambda$$

Ters yön (simetri): $E(X|X + Y) = E(Y|X + Y)$, toplam = $X + Y$:

$$E(X|X + Y) = \frac{X + Y}{2}$$

33.7 Adam Yasası (Yineli Beklenti)

Koşullu beklentinin en önemli özelliği:

$$E(E(Y|X)) = E(Y)$$

İki adım: önce X verildiğinde Y ortalaması (X 'in fonksiyonu), sonra X üzerinden ortalama \rightarrow toplam.

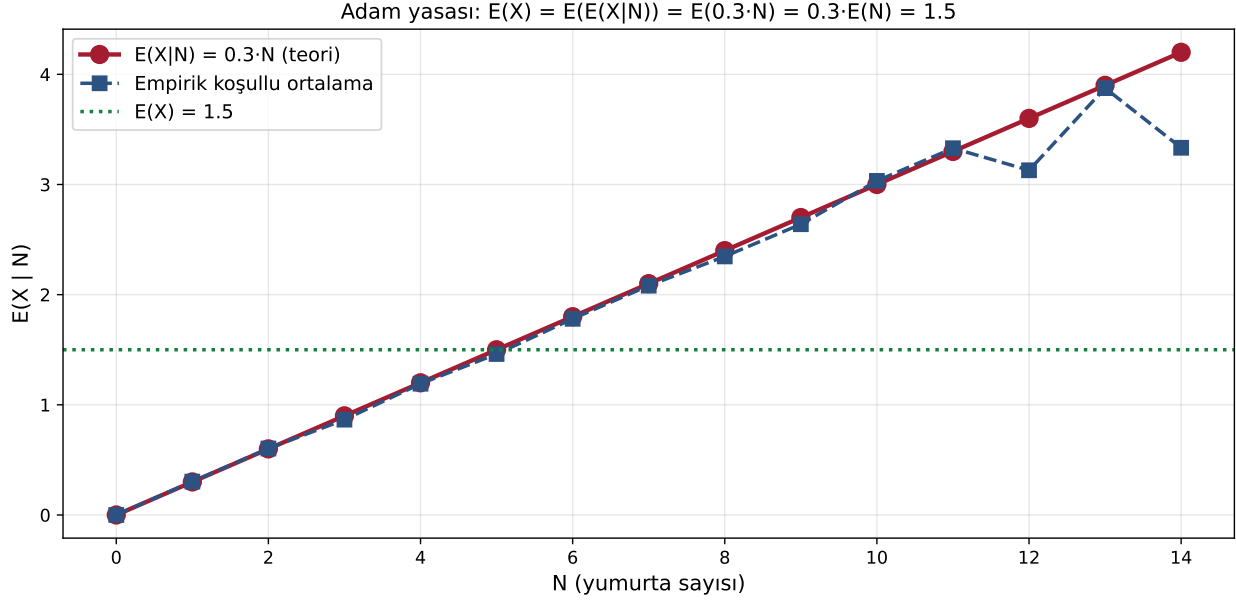
$$E(X) \text{ direkt} = 1.4782$$

$$E(E(X|N)) = E(0.3 \cdot N) = 1.5026$$

$$\text{Teori } 0.3 \cdot 5 = 1.5$$

! Builder Notu — Bellman'ın Atası

Adam yasası ML'de her yerde: değer iterasyonu / Bellman $V(s) = r + E(V(s'))$, **iterated Monte Carlo**, hiyerarşik/karışım modellerinde “önce koşulla, sonra ortalama”. Bir sonraki ders **Eve yasası** (varyans için).



Şekil 33.1: Adam yasası: $E(E(Y|X)) = E(Y)$. Önce X 'i sabitleyip Y 'nin koşullu ortalamasını al ($g(X)$), sonra X üzerinden tekrar ortalama \rightarrow tüm beklenti çıkar. Bellman denkleminin matematik atası.

33.8 Bu Dersin Özeti

1. **İki zarf:** koşulu yalnız bağımsızlıkta düşür.
2. **HT vs HH:** 4 vs 6 (overlap).
3. $E(Y|X) = g(X) = RV$.
4. **Özellikler:** doğrusallık, $h(X)$ dışarı, bağımsızlıkta sabit.
5. **Adam yasası:** $E(E(Y|X)) = E(Y)$.

! Tek bir cümle

$E(Y|X)$ bir **rastgele değişkendir** (X 'in fonksiyonu) = **regresyon fonksiyonu** (MSE minimize eder). **Adam yasası** $E(E(Y|X)) = E(Y) = \text{LOTE'nin kompakt hâli}$; **Bellman'ın atası**.

33.9 Kontrol Soruları

i Soru 1: X, Y bağımsız, $E(Y) = \mu$. (a) $E(3X|X)$? (b) $E(X+Y|X)$? (c) $E(X^2Y|X)$?

Cevap: (a) $3X$. (b) $X + \mu$. (c) $X^2\mu$.

i Soru 2: X, Y iid Pois. $E(X-Y|X+Y)$?

Cevap: Simetri: $E(X|X+Y) = E(Y|X+Y) \rightarrow \text{fark} = 0$.

i Soru 3: $N \sim \text{Pois}(\lambda)$, $X|N \sim \text{Bin}(N, p)$. $E(X)$?

Cevap: Adam: $E(X) = E(E(X|N)) = E(Np) = p\lambda$.

i Soru 4: (Builder) $E(Y|X)$ neden en iyi tahmin?

Cevap: $E[(Y - g(X))^2]$ 'yi minimize eden $g(X) = E(Y|X)$. **MSE eğitilmiş regresyon** bu fonksiyonu örtük öğrenir.

33.10 Egzersizler

Egzersiz 1. X, Y, Z iid, μ . (a) $E(2X + 3Y|X)$. (b) $E(XY|X)$. (c) $E(X + Y + Z|X + Y)$.

Egzersiz 2. $E(W_{TH})$? $E(W_{HHH})$?

Egzersiz 3. Rassal toplam: $N \sim \text{Pois}(\lambda)$, X_i iid μ . $E(S) = \lambda\mu$.

Egzersiz 4. (Python — Örüntüler + Koşullu)

$E(W_{HT}) \approx 3.99$ (teori 4)

$E(W_{HH}) \approx 6.02$ (teori 6)

$E(X | X+Y=6) \approx 3.007$ (teori 3)

Egzersiz 5. (Sonraki ders) **Eve yasası:** $\text{Var}(Y) = E(\text{Var}(Y|X)) + \text{Var}(E(Y|X))$ — açıklanamayan + açıklanan.

33.11 Sonraki Ders İçin Hazırlık

Ders 27: Bir RD Verildiğinde Koşullu Beklenti — Adam ispatı + Eve yasası.

⚠ Ders 27 öncesi yapılacak

- Egzersiz 5 (Eve) çöz.
- Adam'ı ve "E(Y|X) bir RV"yi sağlamlaştırdı.

33.12 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Blitzstein'de
İki zarf	Koşulu yalnız bağımsızlıkta düşür	5m53
HT vs HH	4 vs 6 (overlap)	18m00
$E(Y X=x)$	Koşullu dağılım beklenti = regresyon	30m05
$E(Y X) = g(X)$	RV	34m55

Kavram	Tanım	Blitzstein'de
Bilineni dışarı	$E(h(X)Y X) = h(X)E(Y X)$	41m02
Bağımsızlık	$X \perp Y \Rightarrow E(Y X) = E(Y)$	40m38
$E(X X+Y) = T/2$	Simetri	47m54
Adam yasası	$E(E(Y X)) = E(Y)$	49m21

33.13 ML Bağlantıları Özeti

💡 7 köprü

1. **$E(Y|X)$ = regresyon** → MSE minimizer.
2. **$E(Y|X)$ bir RV** → value, conditional generation.
3. **Koşulu düşürme = bağımsızlık** → collider bias.
4. **Bilineni dışarı** → durum bilgisi ayrıştırma.
5. **Örüntüler** → DNA motifleri, n-gram, KMP.
6. **$E(X|X+Y)$** → yeterli istatistik, **Rao-Blackwell**.
7. **Adam yasası** → **Bellman**, nested MC.

❗ Tek bir şey alıp gideceksen

$E(Y|X)$ bir RV = regresyon fonksiyonu = MSE minimizer. Adam yasası Bellman'ın atası. Koşulu yalnız bağımsızlıkta düşür (iki zarf!).

34 Bir RD Verildiğinde Koşullu Beklenti

Projeksiyon, Adam + Eve yasaları

Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 27](#) (≈51 dk)
- **Okuma süresi:** ≈36 dk

34.1 Bu Derste Ne Var?

1. **4 özellik:** bilenei dışarı, bağımsızlık, Adam, artık diklik.
2. **Projeksiyon yorumu:** $E(Y|X)$ = en küçük kareler.
3. **Koşullu varyans** $\text{Var}(Y|X)$ — bir RV.
4. **Eve yasası:** $\text{Var}(Y) = E(\text{Var}(Y|X)) + \text{Var}(E(Y|X))$ — **within + between**.

Builder Notu — ML Köprüleri

- $E(Y|X)$ = **izdüşüm** → **en küçük kareler**; artık \perp özellikler (normal denklemler).
- **Adam yasası** → Bellman, nested MC, **EM/VI**.
- $\text{Var}(Y|X)$ = **heteroskedastik + aleatoric** belirsizlik.
- **Eve yasası** = **law of total variance** = ANOVA; ML'de **aleatoric + epistemic** ayrışımı.
- **Beta-Binom hiyerarşik** → **random effects**, overdispersion.

34.2 4 Özellik

1. **Bilenei dışarı al:** $E(h(X)Y|X) = h(X)E(Y|X)$.
2. **Bağımsızlık:** $X \perp Y \Rightarrow E(Y|X) = E(Y)$.
3. **Adam yasası:** $E(E(Y|X)) = E(Y)$.
4. **Artık diklik:** $E((Y - E(Y|X))h(X)) = 0$ tüm h için.

34.3 Projeksiyon Yorumu

RV'leri vektör gibi düşün; iç çarpım $\langle X, Y \rangle = E(XY)$. “X’in tüm fonksiyonları” = bir düzlem.

$$E(Y|X) = Y \text{’nin bu düzleme ortogonal izdüşümü}$$

Artık $Y - E(Y|X)$ düzleme **dik** — Özellik 4.

“ $E(Y|X)$ is defined to be the point in this plane that’s closest to Y .” — Blitzstein, 18:05

! Builder Notu — En Küçük Kareler

OLS regresyonun geometrisi: $\hat{y} = y$ ’nin sütun uzayına ortogonal izdüşümü; artık \perp özellikler = **normal denklemler, Gauss-Markov**. Sınır ağları doğrusal-olmayan $E(Y|X)$ ’i yaklaşık öğrenir.

34.4 Adam Yasası İspatı

Kesikli: $g(X) = E(Y|X) \rightarrow$ LOTUS \rightarrow koşullu tanım \rightarrow **toplam sırası değiş** \rightarrow marjinal:

$$E(g(X)) = \sum_x \sum_y y \underbrace{P(Y = y|X = x)P(X = x)}_{P(X,Y)} = \sum_y yP(Y = y) = E(Y) \quad \blacksquare$$

34.5 Koşullu Varyans

$$\text{Var}(Y|X) = E(Y^2|X) - (E(Y|X))^2$$

Bir RV! (X’in fonksiyonu.) Son “ $|X$ ” ’i unutma.

💡 Builder Notu — Aleatoric Belirsizlik

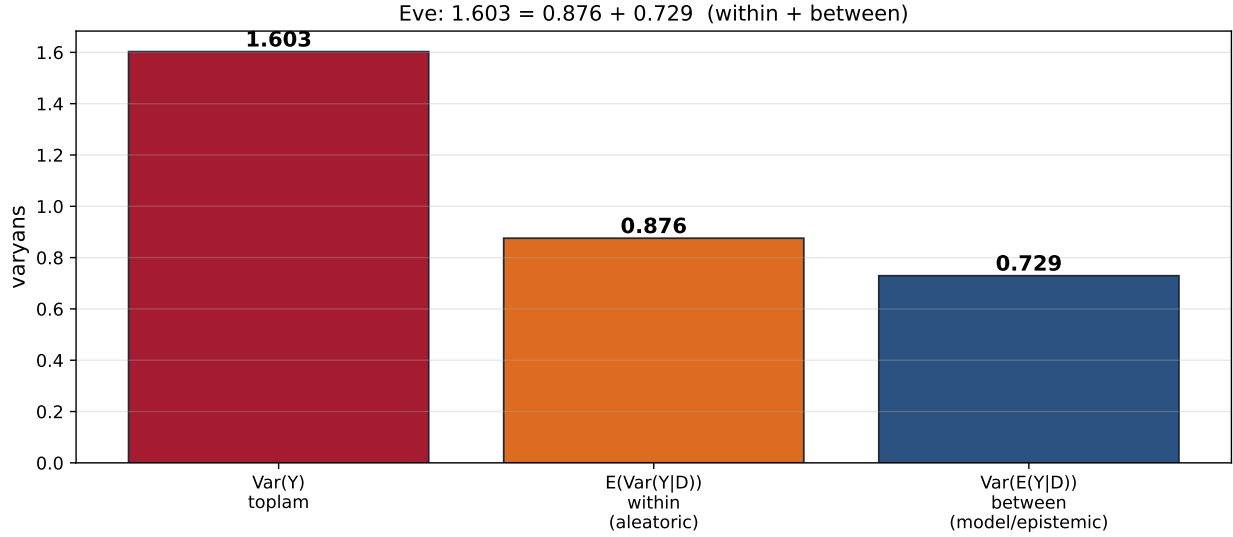
$\text{Var}(Y|X) =$ **girdiye bağlı belirsizlik** = **heteroskedastik** gürültü = **aleatoric belirsizlik** (verinin doğasından, indirilemez). Modelin $\sigma^2(x)$ çıktısı bunu öğrenir.

34.6 Eve Yasası

$$\text{Var}(Y) = E(\text{Var}(Y|X)) + \text{Var}(E(Y|X))$$

Within (grup-içi) + between (gruplar-arası) varyans.

“*Eve’s Law because it’s EVVE*” — Blitzstein, 34:25



Şekil 34.1: Eve yasası: zar (D) atışı + D para. $Y = \text{tura sayısı}$. Toplam $\text{Var}(Y) = E(\text{Var}(Y|D))$ [within = $E(D/4)$] + $\text{Var}(E(Y|D))$ [between = $\text{Var}(D/2)$]. ANOVA + aleatoric/epistemic ayrışımının olasılıksal temeli.

! Builder Notu — Aleatoric + Epistemic

Eve = law of total variance = ANOVA. ML'de **belirsizlik ayrışımı**: $E(\text{Var}(Y|X)) \approx$ **aleatoric** (indirilemez veri gürültüsü), $\text{Var}(E(Y|X)) \approx$ **between/epistemic** (model değişimi). Bayeşçi derin öğrenmede toplam tahmin belirsizliği aynı şekilde ayrışır.

34.7 Beta-Binom Hiyerarşik Örnek

$Q \sim \text{Beta}(a, b)$ (şehir oranı), $X|Q \sim \text{Bin}(n, Q)$.

Adam: $E(X) = E(nQ) = na/(a + b)$.

Eve:

$$\text{Var}(X) = E(nQ(1 - Q)) + \text{Var}(nQ) = n \cdot \frac{ab}{(a + b)(a + b + 1)} + n^2 \cdot \text{Var}(Q)$$

İki belirsizlik: şehir-içi binom + şehirler-arası Q.

! Builder Notu — Random Effects ve Overdispersion

Bayeşçi hiyerarşik / random effects modellerin ders kitabı örneği: between-city varyansı binom'un veremeyeceği **overdispersion**'ı açıklar. Epidemiyoloji, A/B testi grup heterojenliği, çok-merkezli çalışmalarda standart.

34.8 Bu Dersin Özeti

1. 4 özellik.
2. **Projeksiyon:** $E(Y|X) = \text{OLS geometrisi}$.
3. **Adam ispatı:** sırayı değiştir + marjinal.
4. **Var(Y|X)** RV.
5. **Eve:** within + between.

! Tek bir cümle

$E(Y|X) = Y$ 'nin "X'in fonksiyonları" uzayına ortogonal izdüşümü = regresyon. Adam beklentiyi, Eve varyansı parçalar (within + between) — ANOVA ve ML'de aleatoric + epistemic belirsizlik ayrışımının temeli.

34.9 Kontrol Soruları

i Soru 1: $N \sim \text{Pois}(\lambda)$, X_i iid μ . $E(S)$?

Cevap: Adam: $E(S) = E(N\mu) = \lambda\mu$.

i Soru 2: $\text{Var}(X_i) = \sigma^2$. $\text{Var}(S)$?

Cevap: Eve: $\lambda\sigma^2 + \mu^2\lambda = \lambda(\sigma^2 + \mu^2) = \lambda E(X^2)$.

i Soru 3: $X \perp Y$, $E(Y) = 5$. (a) $E(\sin X|X)$? (b) $E(XY|X)$?

Cevap: (a) $\sin X$. (b) $5X$.

i Soru 4: (Builder) Aleatoric vs epistemic Eve ayrışımı?

Cevap: $E(\text{Var}(Y|X)) = \text{aleatoric}$ (indirilemez gürültü). $\text{Var}(E(Y|X)) = \text{between/model}$ (epistemic). Bayeşçi DL'de aynı.

34.10 Egzersizler

Egzersiz 1. Zar $D + D$ para, X tura. (a) $E(X|D) = D/2$. (b) Adam $\rightarrow E(X) = 1,75$.

Egzersiz 2. (a) $\text{Var}(X|D) = D/4$. (b) Eve $\rightarrow \text{Var}(X)$?

Egzersiz 3. Artık özellikleri: $E(\text{artık}) = 0$, $\text{Cov}(\text{izdüşüm}, \text{artık}) = 0$.

Egzersiz 4. (Python — Adam + Eve)

$E(X)$ sim = 1.7519 teorik $3.5/2 = 1.75$

$\text{Var}(X)$ sim = 1.6028 = $\text{within}(0.8756) + \text{between}(0.7294) = 1.6049$

Egzersiz 5. (Sonraki ders) **Jensen:** dışbükey $g \rightarrow E(g(X)) \geq g(E(X))$. **Markov:** $X \geq 0 \rightarrow P(X \geq a) \leq E(X)/a$.

34.11 Sonraki Ders İçin Hazırlık

Ders 28: Eşitsizlikler — Jensen, Markov, Chebyshev, Cauchy-Schwarz.

⚠ Ders 28 öncesi yapılacak

- Egzersiz 5 (Jensen + Markov sezgisi) çöz.
- Jensen + Cauchy-Schwarz hatırla.

34.12 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Blitzstein'de
Bileneni dışarı	$E(h(X)Y X) = h(X)E(Y X)$	8m37
Bağımsızlık	$X \perp Y \Rightarrow E(Y X) = E(Y)$	9m55
Adam	$E(E(Y X)) = E(Y)$	10m44
Artık diklik	$E((Y - E(Y X))h(X)) = 0$	12m46
Projeksiyon	$E(Y X) = \text{OLS izdüşümü}$	17m42
Var(Y X)	$E(Y^2 X) - (E(Y X))^2$, RV	30m17
Eve	$E(\text{Var}) + \text{Var}(E)$	33m28

34.13 ML Bağlantıları Özeti

💡 7 köprü

1. **$E(Y|X) = \text{izdüşüm}$** → OLS, normal denklemler.
2. **$E(Y|X) = \text{regresyon}$** → MSE.
3. **Adam** → Bellman, EM/VI.
4. **$\text{Var}(Y|X)$** → heteroskedastik, aleatoric.
5. **Eve** → law of total variance, ANOVA.
6. **Beta-Binom hiyerarşik** → overdispersion, random effects.
7. **Compound Poisson** → toplam ödül/hasar (sigorta, RL).

! Tek bir şey alıp gideceksen

$E(Y|X)$ = **ortogonal izdüşüm** = **regresyon**. **Adam** beklentiyi, **Eve** varyansı parçalar (**within + between**) → ANOVA + aleatoric/epistemic ayrışımı.

35 Eşitsizlikler

Cauchy-Schwarz, Jensen, Markov, Chebyshev

i Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 28](#) (≈47 dk)
- **Okuma süresi:** ≈22 dk

35.1 Bu Derste Ne Var?

1. **Rassal toplam:** $X = \sum_{j=1}^N X_j$, Adam + Eve.
2. **Cauchy-Schwarz:** $|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}$.
3. **Jensen:** dışbükey $g \rightarrow E(g(X)) \geq g(E(X))$.
4. **Markov:** $P(|X| \geq a) \leq E|X|/a$.
5. **Chebyshev:** $P(|X - \mu| \geq a) \leq \sigma^2/a^2$.

💡 Builder Notu — ML Köprüleri

- **Eşitsizlik ≠ yaklaşım** → PAC, generalization bound; kanıtlanmış sınır > tahmin.
- **Markov** → **Chebyshev** → **Chernoff** → **Hoeffding** zinciri = **konsantrasyon eşitsizlikleri** ailesi.
- **Jensen** → **ELBO** = variational inference temeli (VAE, diffusion).
- **Cauchy-Schwarz** → **kosinüs benzerliği** (attention dot product).
- **Chebyshev** → **BSY** → **örneklem büyüklüğü** $\sigma^2/(n\epsilon^2)$.

35.2 Rassal Toplam: Adam + Eve

$$X = \sum_{j=1}^N X_j, N \perp X_j, E(X_j) = \mu, \text{Var}(X_j) = \sigma^2.$$

Naif hata: $E(X) = N\mu$ — kategori hatası (N RV!).

Adam: $E(X|N) = \mu N \rightarrow E(X) = \mu E(N)$.

Eve:

$$\text{Var}(X) = \sigma^2 E(N) + \mu^2 \text{Var}(N)$$

35.3 Cauchy-Schwarz

$$|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}$$

= **korelasyon** $\in [-1, 1]$ (yeniden).

💡 Builder Notu — Embedding Benzerlik

$E(XY)$ = iç çarpım, $\sqrt{E(X^2)}$ = norm \rightarrow RV'ler **Hilbert uzayı**. **Attention** nokta çarpımı, **kosinüs benzerliği** $[-1, 1]$ 'de kalmasının garantisi.

35.4 Jensen Eşitsizliği

Dışbükey g ($g'' \geq 0$):

$$E(g(X)) \geq g(E(X))$$

İçbükeyse ($\ln, \sqrt{\cdot}$) yön döner.

İspat (teğet doğrusu): $g(x) \geq a + bx$ tüm x (dışbükeylik). Beklenti al.

Örnekler:

- $g(x) = x^2$: $E(X^2) \geq (E(X))^2$ (varyans ≥ 0).
- $g(x) = 1/x$ ($x > 0$): $E(1/X) \geq 1/E(X)$ (oran tahmincisi yanlılığı).
- \ln : $E(\ln X) \leq \ln E(X)$.

❗ Builder Notu — ELBO ve Cross-Entropy

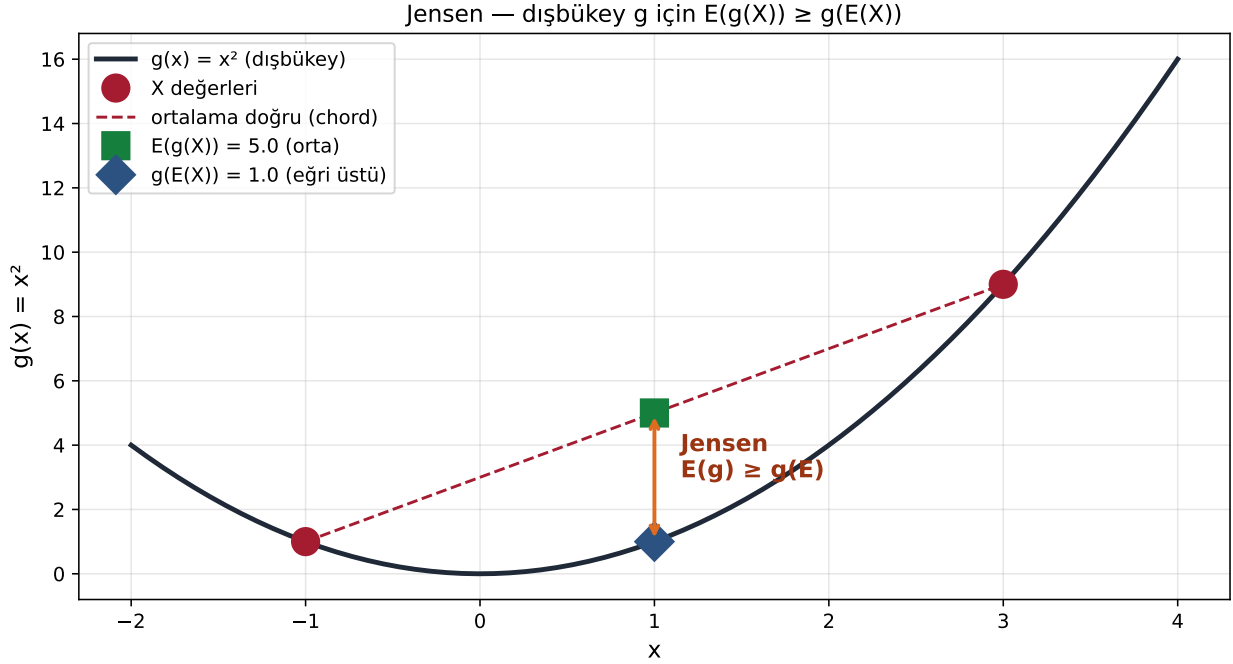
Jensen ML'de her yerde: (1) **ELBO:** $\log E[\cdot] \geq E[\log \cdot] \rightarrow$ variational inference (**VAE, diffusion**); (2) AM-GM eşitsizliği; (3) oran tahmincisi yanlılığı; (4) karar ağacı bölme kriterleri (Gini, entropi). **ELBO = Jensen.**

35.5 Markov Eşitsizliği

$$P(|X| \geq a) \leq \frac{E|X|}{a}, \quad a > 0$$

Hiçbir varsayım yok ($X \geq 0$ dışında).

İspat (gösterge): $a \cdot \mathbb{1}_{|X| \geq a} \leq |X|$ her durumda. Beklenti al.



Şekil 35.1: Jensen: dışbükey $g(x) = x^2$ için $E(g(X)) \geq g(E(X))$. İki nokta + olasılıkla seç \rightarrow ortalama nokta eğrinin üstünde. ELBO/cross-entropy/aritmetik-geometrik eşitsizlik hep Jensen.

💡 Builder Notu — Konsantrasyonun Atası

Markov = tüm konsantrasyon eşitsizliklerinin atası. Chebyshev (kareyle), **Chernoff** (MGF'yle), **Hoeffding**, Bernstein — hepsi “negatif olmayan dönüşüme Markov uygula” tarifi. ML’de **SGD yakınsama garantisi**, **diferansiyel mahremiyet** gürültü kuyruğu hep Markov ile başlar.

35.6 Chebyshev Eşitsizliği

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$$

Veya $a = c\sigma$:

$$P(|X - \mu| \geq c\sigma) \leq \frac{1}{c^2}$$

İspat: Markov’u $(X - \mu)^2 \geq a^2$ olayına uygula.

! Builder Notu — BSY Motoru

Chebyshev = BSY’nin motoru. \bar{X}_n ’in varyansı $\sigma^2/n \rightarrow P(|\bar{X}_n - \mu| \geq \epsilon) \leq \sigma^2/(n\epsilon^2) \rightarrow 0$. **Monte Carlo yakınsama hızı, A/B test örneklem büyüklüğü, minibatch gradyan kalitesi** hep bu

ölçeklemeden.

35.7 Bu Dersin Özeti

1. **Rassal toplam:** $E = \mu E(N)$, $\text{Var} = \sigma^2 E(N) + \mu^2 \text{Var}(N)$.
2. **Cauchy-Schwarz:** korelasyon $[-1, 1]$.
3. **Jensen:** dışbükey \rightarrow yönü hatırla.
4. **Markov:** $P \leq E/a$.
5. **Chebyshev:** $P \leq \sigma^2/a^2$.

! Tek bir cümle

Dağılımı bilmesen bile dört eşitsizlik kanıtlanmış sınır verir. **Markov \rightarrow Chebyshev = BSY motoru** = ML'de **konsantrasyon ailesinin atası**; **Jensen \rightarrow ELBO** = variational inference temeli; **Cauchy-Schwarz** = embedding benzerlik garantisi.

35.8 Kontrol Soruları

i Soru 1: $E(X) = N\mu$ neden yanlış?

Cevap: Kategori hatası — $E(X)$ sayı, $N\mu$ RV. Doğru: $\mu E(N)$.

i Soru 2: Rassal toplam Var iki terimi neyi temsil eder?

Cevap: Aleatorik (her müşterinin gürültüsü) + yapısal (N belirsizliği).

i Soru 3: $X > 0$. $E(1/X)$ vs $1/E(X)$?

Cevap: $E(1/X) \geq 1/E(X)$ (Jensen, $1/x$ dışbükey).

i Soru 4: Chebyshev $c=3$ vs Normal?


Cevap: Chebyshev $\leq 1/9 \approx 0,111$ her dağılım için. Normal $\approx 0,003$. Evrenselliğin bedeli kabalık.

35.9 Egzersizler

Egzersiz 1. Bileşik Poisson: $N \sim \text{Pois}(\lambda)$, $E(X)$, $\text{Var}(X)$.

Egzersiz 2. $X \geq 0$, $E(X) = 2$. Markov $P(X \geq 8) \leq ?$ vs $\text{Exp}(1/2)$ gerçek.

Egzersiz 3. Adil para n atış. Chebyshev ile $P(|\bar{X}_n - 0,5| \geq 0,1) \leq 0,05$ için n ?

Egzersiz 4. (Python — Markov + Chebyshev + Jensen)Markov: $E|X| = 1.999$ a=4: $P(X \geq a)$ gerçek=0.1353 ≤ 0.4999 a=6: $P(X \geq a)$ gerçek=0.0500 ≤ 0.3332 a=8: $P(X \geq a)$ gerçek=0.0184 ≤ 0.2499 Chebyshev: $\mu=2.00$, $\sigma=2.00$ c=1: $P(|X-\mu| \geq c\sigma)$ gerçek=0.1351 ≤ 1.0000 c=2: $P(|X-\mu| \geq c\sigma)$ gerçek=0.0499 ≤ 0.2500 c=3: $P(|X-\mu| \geq c\sigma)$ gerçek=0.0183 ≤ 0.1111 Jensen: $E(1/X) = 7.6485 \geq 1/E(X) = 0.5001$ **Egzersiz 5.** (Sonraki ders) **BSY:** Chebyshev ile $\bar{X}_n \rightarrow \mu$. **MLT:** $\bar{X}_n - \mu$ normal şekilde dalgalanır.**35.10 Sonraki Ders İçin Hazırlık****Ders 29: BSY ve MLT** — Chebyshev kullanılacak. Ders 29 öncesi yapılacak

- Egzersiz 5 (BSY/MLT sezgi) çöz.
- Chebyshev tekrar oku.

35.11 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Not
Rassal toplam E	$\mu E(N)$	Adam
Rassal toplam Var	$\sigma^2 E(N) + \mu^2 \text{Var}(N)$	Eve
Cauchy-Schwarz	$\ E(XY)\ \leq \sqrt{E(X^2)E(Y^2)}$	Corr $\in [-1, 1]$
Jensen	g dışbükey $\rightarrow E(g) \geq g(E)$	ELBO
Markov	$P(\ X\ \geq a) \leq E\ X\ /a$	Atası
Chebyshev	$P(\ X - \mu\ \geq a) \leq \sigma^2/a^2$	BSY motoru

35.12 ML Bağlantıları Özeti

💡 6 köprü

1. **Eşitsizlik \neq yaklaşım** \rightarrow PAC, garanti.
2. **Markov \rightarrow Chernoff \rightarrow Hoeffding** \rightarrow konsantrasyon ailesi.
3. **Jensen \rightarrow ELBO** \rightarrow VAE, diffusion.
4. **Cauchy-Schwarz** \rightarrow embedding, attention.
5. **Eve** \rightarrow aleatoric + epistemic.
6. **Chebyshev \rightarrow BSY \rightarrow örneklem** $\sigma^2 / (n\epsilon^2)$.

! Tek bir şey alıp gideceksen

Dağılımı bilmesen bile dört eşitsizlik **kanıtlanmış sınır** verir. Markov-Chebyshev = BSY + konsantrasyon ailesi. Jensen = ELBO. Cauchy-Schwarz = embedding garantisi.

36 Büyük Sayılar Yasası ve Merkezi Limit Teoremi

$1/\sqrt{n}$ hata ölçeği; istatistiğin iki büyük teoremi

Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 29](#) (≈ 51 dk)
- **Okuma süresi:** ≈ 24 dk

36.1 Bu Derste Ne Var?

1. **BSY:** $\bar{X}_n \rightarrow \mu$ — *nereye* (sabit).
2. **MLT:** $\sqrt{n}(\bar{X}_n - \mu)/\sigma \rightarrow N(0, 1)$ — *nasıl* (şekil + \sqrt{n} hız).
3. **Var**(\bar{X}_n) = σ^2/n — $1/\sqrt{n}$ hata ölçeği.
4. **MLT ispatı:** MGF + L'Hôpital.
5. **Binom normal yaklaşım** + süreklilik düzeltmesi.

Builder Notu — ML Köprüleri

- $1/\sqrt{n}$ ölçeği → **Monte Carlo, minibatch gradyan, A/B test örneklem.**
- σ/\sqrt{n} → **ensemble/bagging** (n bağımsız model varyansı $1/n$).
- $\sqrt{n_{in}}$ **init** → **Xavier/He** ağırlık ilklendirme.
- \sqrt{d} **attention** → QK^T/\sqrt{d} skorları, d bağımsız terim SD'si.
- **MLT** → **Gauss varsayımı her yerde:** diffusion, VAE prior, gürültü modelleri.
- **iid kırılması** → distribution shift, RL'de policy değişikliği.

36.2 Kurulum

X_1, X_2, \dots iid, $E(X_j) = \mu$, $\text{Var}(X_j) = \sigma^2$ (sonlu).

$$\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$$

Anahtar:

$$\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}, \quad \text{SD}(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}$$

36.3 Büyük Sayılar Yasası

Güçlü BSY: $\bar{X}_n \rightarrow \mu$ olasılık 1 ile (noktasal).

Zayıf BSY: Her $c > 0$ için $P(|\bar{X}_n - \mu| > c) \rightarrow 0$ (olasılıkta).

İspat (Chebyshev, tek satır):

$$P(|\bar{X}_n - \mu| > c) \leq \frac{\text{Var}(\bar{X}_n)}{c^2} = \frac{\sigma^2}{nc^2} \rightarrow 0$$

“This theorem is crucial for science to be possible.” — Blitzstein, 9:59

💡 Builder Notu — Tutarlılık (Consistency)

“Olasılıkta yakınsama” = tahmincinin tutarlılığı. MLE, SGD yakınsaması hep bu dilde. **Kumarbazın yanlışlığı** (“telafi olur”) BSY’yi yanlış yorumlar — BSY **swamping** ile çalışır, telafi ile değil.

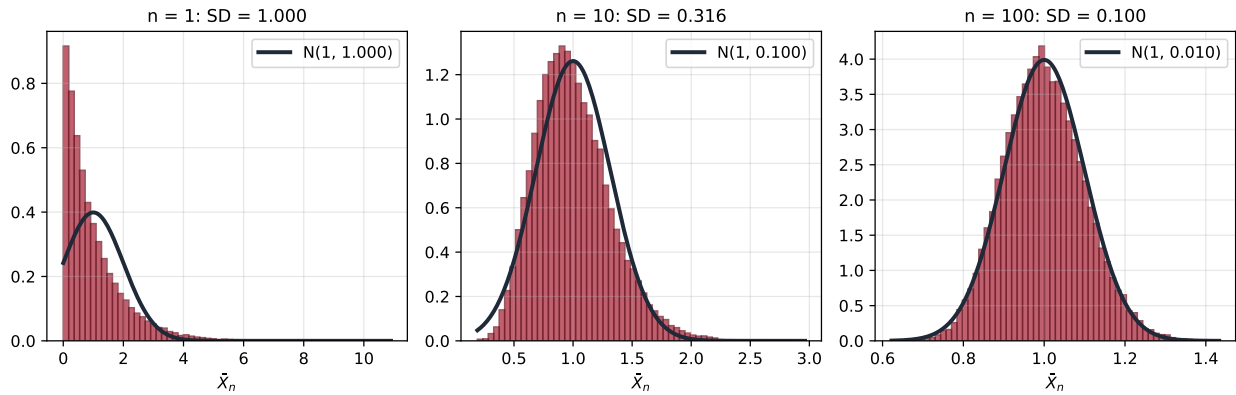
36.4 Merkezi Limit Teoremi

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0, 1)$$

Şaşırtıcı: sadece sonlu varyans yeter. X_j ’ler ne kadar çirkin olsa da ortalaması normale gider.

Neden \sqrt{n} ? $\text{SD}(\bar{X}_n) = \sigma/\sqrt{n} \rightarrow \sqrt{n}$ ile çarp \rightarrow SD sabitlenir.

MLT: Üstel(1) çarpık dağılım \rightarrow n büyüdükçe ortalama Normal’e gider



Şekil 36.1: MLT canlı: Üstel(1) çarpık dağılımdan örneklem ortalamasının dağılımı. $n=1$ çarpık; $n=10, 100$ hızla Normal’e oturuyor. $\text{SD} = \sigma/\sqrt{n}$ ölçeği. “Yeterince çok şey topla, çana dönüş”

! Builder Notu — \sqrt{d} Her Yerde

\sqrt{n} ölçeği DL'de her yerde: (1) Xavier/He init $1/\sqrt{n_{in}}$ → katman varyansı sabit; (2) Attention QK^T/\sqrt{d} → d bağımsız terim SD'si \sqrt{d} ; (3) Layer norm sezgisi. MLT bilmeden bu $\sqrt{\cdot}$ bölenlerin anlaşılması zor.

36.5 MLT İspatı (MGF + L'Hôpital)

Standartlaştır: $\mu = 0, \sigma = 1$. $S_n = \sum X_j$, $E(e^{tS_n/\sqrt{n}}) = [M(t/\sqrt{n})]^n$.

Log + $y = 1/\sqrt{n}$:

$$\lim_{y \rightarrow 0} \frac{\ln M(yt)}{y^2}, \quad 0/0$$

L'Hôpital $\times 2 + M(0) = 1$, $M'(0) = 0$, $M''(0) = 1$:

$$\frac{t}{2} \cdot t \cdot M''(0) = \frac{t^2}{2}$$

Üs al → $e^{t^2/2} = N(0, 1)$ MGF'si. ■

36.6 Binom Normal Yaklaşımı

$X \sim \text{Bin}(n, p)$: $E = np$, $SD = \sqrt{npq}$.

$$P(a \leq X \leq b) \approx \Phi\left(\frac{b - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np}{\sqrt{npq}}\right)$$

$p \approx 1/2$ iyi. Süreklilik düzeltmesi:

$$P(X = a) \approx P(a - 0,5 \leq X \leq a + 0,5)$$

Poisson vs Normal: $p \rightarrow 0$ (nadir) → Poisson; $p \approx 1/2$ → Normal.

36.7 Bu Dersin Özeti

1. **BSY**: $\bar{X}_n \rightarrow \mu$.
2. **Var**(\bar{X}_n) = σ^2/n .
3. **Zayıf BSY**: Chebyshev tek satır.
4. **MLT**: dağılımda \sqrt{n} ölçeği.
5. **Binom normal + süreklilik düzeltmesi**.

! Tek bir cümle

Bağımsız özdeş şeyleri ortalarsan, **gerçeğe gider** (BSY) ve etrafındaki dalgalanma $1/\sqrt{n}$ ölçüğünde çana dönüşür (MLT) — dağılım ne olursa olsun. **Bilim ve istatistik bu iki garanti üzerine kuruludur.**

36.8 Kontrol Soruları

i Soru 1: BSY vs MLT?

Cevap: BSY nereye (sabit μ). MLT nasıl (\sqrt{n} ölçüğü + Normal). Farklı yakınsama türleri.

i Soru 2: $\text{Var}(\bar{X}_n)$? Zayıf BSY?

Cevap: σ^2/n . Chebyshev $\rightarrow \sigma^2/(nc^2) \rightarrow 0$.

i Soru 3: Neden \sqrt{n} ?

Cevap: $\text{SD}(\bar{X}_n) = \sigma/\sqrt{n}$. \sqrt{n} ile çarpılırsa SD sabitlenir \rightarrow dejenere olmayan limit.

i Soru 4: $\text{Bin}(100, 0.5)$, $P(X=50)$ normal yaklaşım?

Cevap: $\Phi(0.1) - \Phi(-0.1) \approx 0,0796$ (gerçek değer 0.0796, mükemmel).

36.9 Egzersizler

Egzersiz 1. $\sigma^2 = 9$, $\text{SD} \leq 0,1$ için n ? Hatayı yarıya n kaç kat?

Egzersiz 2. $\text{Bern}(0.5)$, $P(|\bar{X}_n - 0,5| > 0,05) \leq 0,01$ için n (Chebyshev).

Egzersiz 3. $\text{Exp}(1)$ iid, $n = 100$. \bar{X}_{100} yaklaşık dağılımı?

Egzersiz 4. $\text{Bin}(400, 0.5)$, $P(X \geq 220)$ normal yaklaşım + süreklilik.

Egzersiz 5. (Python — MLT gözlemele)

$n= 1$: emp SD=1.0051 teorik $\sigma/\sqrt{n}=1.0000$ çarpıklık=+2.018 ($\rightarrow 0$)

$n= 10$: emp SD=0.3177 teorik $\sigma/\sqrt{n}=0.3162$ çarpıklık=+0.659 ($\rightarrow 0$)

$n=100$: emp SD=0.1001 teorik $\sigma/\sqrt{n}=0.1000$ çarpıklık=+0.211 ($\rightarrow 0$)

36.10 Sonraki Ders İçin Hazırlık

Ders 30: Ki-Kare, Student-t, MVN — Normal'in türevleri.

⚠ Ders 30 öncesi yapılacak

- Egzersizleri çöz.
- \sqrt{n} ölçeği + binom normal yaklaşımını içselleştir.

36.11 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Not
\bar{X}_n	$\frac{1}{n} \sum X_j$	RV
$\text{Var}(\bar{X}_n)$	σ^2/n	SD = σ/\sqrt{n}
Güçlü BSY	$\bar{X}_n \rightarrow \mu$ (ols 1)	Noktasal
Zayıf BSY	$P \rightarrow 0$	Olasılıkta; Chebyshev
MLT	$\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{d} N(0, 1)$	Dağılımda
Binom normal	Φ farkı	$p \approx 1/2$
Süreklilik düzeltmesi	$P(X = a) \approx P(a \pm 0.5)$	Kesikli \rightarrow sürekli

36.12 ML Bağlantıları Özeti

💡 7 köprü

1. $1/\sqrt{n}$ ölçeği \rightarrow Monte Carlo, minibatch.
2. **Xavier/He init** $1/\sqrt{n_{in}}$.
3. **Attention** $1/\sqrt{d}$.
4. **Tutarlılık** (consistency).
5. **Ensemble/bagging** σ^2/n .
6. **Gauss her yerde** \rightarrow diffusion, VAE.
7. **iid kırılması** \rightarrow distribution shift, RL.

❗ Tek bir şey alıp gideceksen

Yeterince çok şey topla, gerçeğe git (μ , **BSY**) + **çana dön** ($N(0, 1)$, **MLT**), dağılım ne olursa olsun. $1/\sqrt{n}$ ölçeği ML'de her yerde.

37 Ki-Kare, Student-t ve Çok Değişkenli Normal

Normalin türevleri: χ^2 , t, MVN

Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 30](#) (≈47 dk)
- **Okuma süresi:** ≈23 dk

37.1 Bu Derste Ne Var?

1. **Ki-Kare** (χ^2): $\sum Z_j^2 = \text{Gamma}(n/2, 1/2)$.
2. **Student-t:** $Z/\sqrt{V/n}$; $t_1 = \text{Cauchy}$.
3. **MVN:** her doğrusal kombinasyon normal.
4. **MVN'de korelasyonsuz** \implies **bağımsız** (yalnız orada!).

Builder Notu — ML Köprüleri

- $\chi^2 \rightarrow$ **Mahalanobis uzaklığı**, **L2 düzenleme**, aykırı değer tespiti.
- **t** \rightarrow **t-SNE** (ağır kuyruk crowding'i çözer); **robust regresyon**.
- **MVN** \rightarrow **Gaussian Process (GP)**, **VAE latent**, **Kalman filter**, **GMM**.
- **Korelasyonsuz** \implies **bağımsız (MVN)** \rightarrow **PCA = ICA** Gaussian'da.
- **Marjinal vs ortak normallik** \rightarrow **copula**.

37.2 Ki-Kare (χ^2)

Z_1, \dots, Z_n iid $N(0, 1)$:

$$V = \sum_{j=1}^n Z_j^2 \sim \chi_n^2$$

Gamma bağlantısı: $\chi_1^2 = Z^2 \sim \text{Gamma}(1/2, 1/2)$. Toplam:

$$\chi_n^2 = \text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$$

$$E(\chi_n^2) = n, \text{Var} = 2n.$$

! Builder Notu — Mahalanobis

Karelerin toplamı = χ^2 ML'de her yerde: **MSE, L2 düzenleme**, Gaussian'ın **negatif log-likelihood**'u. **Mahalanobis uzaklığı** $\|x - \mu\|_{\Sigma}^2 \sim \chi^2 \rightarrow$ aykırı değer skoru.

37.3 Student-t

$$Z \sim N(0,1), V \sim \chi_n^2, Z \perp V:$$

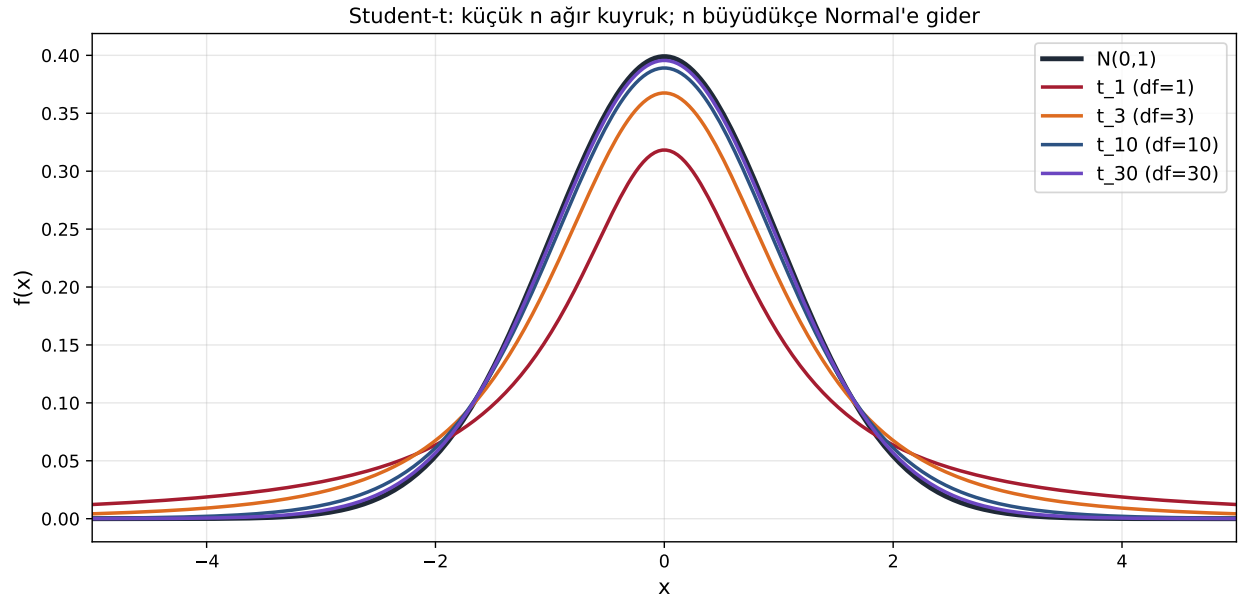
$$T = \frac{Z}{\sqrt{V/n}} \sim t_n$$

Tarih: William Gosset (Guinness) takma ad "Student".

Özellikler:

- Simetri: $-T \sim t_n$.
- $t_1 = \text{Cauchy}$ (ortalama yok).
- t_n ortalaması 0 ($n \geq 2$); t_n 'in n . ve üstü momenti yok.
- **Ağır kuyruk:** Cauchy $\propto 1/x^2$, normal $e^{-x^2/2}$.

$n \rightarrow \infty \rightarrow \text{Normal: BSY} \rightarrow V_n/n \rightarrow E(Z^2) = 1$, payda 1'e $\rightarrow T \rightarrow Z$.



Şekil 37.1: Student-t kuyruğu: t_1 (Cauchy) en kalın, t_{30} neredeyse Normal. Küçük n 'de ağır kuyruk \rightarrow outlier'lara dayanıklılık (t-SNE crowding, robust regresyon). BSY ile $t \rightarrow$ Normal.

! Builder Notu — t-SNE ve Robust

t-SNE: düşük-boyutlu gömmede **Student-t çekirdeği** kullanır; ağır kuyruk **crowding problem**'i çözer, kümeler ayrışır. **Robust regresyon:** Gaussian gürültü yerine t-gürültü → aykırı değerler “şok” yerine “olası” görülür.

37.4 Çok Değişkenli Normal (MVN)

Tanım: $\mathbf{X} = (X_1, \dots, X_k)$ MVN \iff her $\sum t_j X_j$ normaldir.

Karşı-örnek: $Z \sim N(0, 1)$, $S \in \{\pm 1\}$ (Z 'den bağımsız). (Z, SZ) :

- Marjinaller her ikisi $N(0,1)$.
- Ama $Z + SZ$: yarı zaman 0 (kesikli) + yarı sürekli → **MVN değil!**

MGF:

$$M(\mathbf{t}) = \exp\left(\sum t_j \mu_j + \frac{1}{2} \text{Var}\left(\sum t_j X_j\right)\right)$$

MVN tamamen **ortalama vektör + kovaryans matrisi** ile belirlenir.

37.5 MVN: Korelasyonsuz \implies Bağımsız

Genelde: bağımsız \implies korelasyonsuz, **tersi yanlış.**

MVN içinde: tersi de doğru!

Örnek: $X, Y \sim N(0, 1)$ iid. $(X + Y, X - Y)$ MVN.

$$\text{Cov}(X + Y, X - Y) = \text{Var}(X) - \text{Var}(Y) = 0$$

→ $X + Y \perp X - Y$. (Bu sadece Gauss'a özgü.)

! Builder Notu — PCA = ICA Gaussian'da

Korelasyonsuz \implies bağımsız (MVN) Gauss varsayımlı modellerin işlenebilirliğinin **özüdür**. **PCA:** kovaryansı köşegenleştir → Gaussian'da **bağımsız** bileşenler. **Gaussian Process** koşullama, **Kalman** güncelleme, **VAE** köşegen latent — kapalı form bu sayede.

37.6 Bu Dersin Özeti

1. $\chi_n^2 = \sum Z_j^2 = \mathbf{Gamma}(n/2, 1/2)$.
2. $t_n = Z/\sqrt{V/n}$; $t_1 = \text{Cauchy}$.
3. $t \rightarrow \mathbf{Normal}$ (BSY).
4. **MVN**: her kombinasyon normal.
5. **Korelasyonsuz** \implies **bağımsız** (MVN'de).

! Tek bir cümle

Ki-kare karelerle, **t** oranlarla, **MVN** vektörlerle normali genişletir. Yalnız MVN'de **“korelasyonsuz = bağımsız”** — Gauss'un derin armağanı; **GP, VAE, Kalman, PCA**'nın işlenebilirliğinin kaynağı.

37.7 Kontrol Soruları

i Soru 1: χ^2_n neden yeni değil?

Cevap: $\mathbf{Gamma}(n/2, 1/2)$. Tüm Gamma araçları geçerli.

i Soru 2: t_1 neden Cauchy?

Cevap: $T = Z/|Z_1| = \text{iki Normal oranı} = \text{Cauchy}$. Ortalama yok.

i Soru 3: $t_n \rightarrow \text{Normal}$ nasıl?

Cevap: BSY: $V_n/n \rightarrow 1$, payda sabitleşir $\rightarrow T \rightarrow Z$.

i Soru 4: (Z, SZ) MVN mi? Korelasyon?

Cevap: MVN **değil** ($Z + SZ$ kesikli-sürekli karışım). **Korelasyonsuz** ($\text{Cov} = 0$) ama **bağımlı** (S, Z 'nin işaretini belirler).

37.8 Egzersizler

Egzersiz 1. $E(\chi_n^2)$, $\text{Var}(\chi_n^2)$. Hem Gamma hem $\sum Z_j^2$ ile.

Egzersiz 2. t_2 momentleri var mı? Hangi mertebeye kadar?

Egzersiz 3. MVN testi: (a) $(Z + W, Z - W)$, (b) (Z, Z^2) , (c) $(2Z + 3W, Z - W, 5W)$.

Egzersiz 4. X, Y iid $N(0, \sigma^2)$. $X + Y \perp X - Y$? $\sigma_X \neq \sigma_Y$ olsa?


Egzersiz 5. (Python — *t* kuyruğu + MVN)

$P(T > 3)$:
 t_{1} : 0.1024
 t_{2} : 0.0477
 t_{5} : 0.0150
 t_{10} : 0.0067
 t_{30} : 0.0027
 t_{100} : 0.0017
 $N(0,1)$: 0.0013

$\text{corr}(X+Y, X-Y) = 0.0017$ ($\approx 0 \rightarrow$ MVN'de bağımsız)

37.9 Sonraki Ders İçin Hazırlık

Ders 31: Markov Zincirleri — zaman içinde bağımlılığın temel modeli.

 Ders 31 öncesi yapılacak

- Egzersizleri çöz.
- **MVN korelasyonsuz \implies bağımsız**'ı içselleştir.

37.10 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Not
χ_n^2	$\sum Z_j^2$	Gamma($n/2, 1/2$)
t_n	$Z/\sqrt{V/n}$	$t_1 =$ Cauchy
t moment	$n.$ ve üstü yok	Simetri 0
t kuyruk	Cauchy $\propto 1/x^2$	Normal'den kalın
$t \rightarrow N$	BSY ile	n büyük
MVN	Her $\sum t_j X_j$ normal	Bileşen-normal yetmez
MVN MGF	$\mu + \Sigma$	Belirler
Korelasyonsuz \implies bağımsız	Yalnız MVN	Gauss armağanı

37.11 ML Bağlantıları Özeti

 6 köprü

1. $\chi^2 \rightarrow$ Mahalanobis, aykırı değer, L2.
2. **t-SNE** \rightarrow ağır kuyruk crowding.
3. **Robust regresyon** \rightarrow t-gürültü.
4. **MVN** \rightarrow GP, VAE, Kalman, GMM.
5. **Korelasyonsuz \implies bağımsız** \rightarrow PCA = ICA Gaussian'da.

6. **Marjinal \neq ortak \rightarrow copula.**

! Tek bir şey alıp gideceksen

Ki-kare karelerle, **t** oranlarla, **MVN** vektörlerle normal genişler. Yalnız MVN'de **korelasyonsuz = bağımsız** — Gauss'un işlenebilirliğinin sırrı.

38 Markov Zincirleri

Markov özelliği, geçiş matrisi, durağan dağılım

Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 31](#) (≈47 dk)
- **Okuma süresi:** ≈22 dk

38.1 Bu Derste Ne Var?

1. **Markov özelliği:** geçmiş \perp gelecek | şimdi.
2. **Geçiş matrisi Q :** $q_{ij} = P(i \rightarrow j)$, satır toplamı 1.
3. **n -adım:** $X_n \sim s \rightarrow X_{n+m} \sim sQ^m$.
4. **Durağan dağılım:** $sQ = s$.

Builder Notu — ML Köprüleri

- **MCMC** (Markov Chain Monte Carlo) — Bayesçi çıkarımın motoru, **Metropolis-Hastings, Gibbs, HMC**.
- **Diffusion modelleri** — ileri/geri Markov zinciri.
- **RL = MDP** (Markov Decision Process) — Bellman'ın temeli.
- **PageRank** — web grafının durağan dağılımı.
- **n-gram dil modelleri** = $(n - 1)$. dereceden Markov; **Transformer** Markov varsayımını kırar.
- **HMM forward** = sQ belief güncellemesi.

38.2 Markov Özelliği

$$P(X_{n+1} = j \mid X_n = i, X_{n-1}, \dots, X_0) = P(X_{n+1} = j \mid X_n = i)$$

Yalnız son durum önemli. Geçmiş ve gelecek **şimdi verildiğinde koşullu bağımsız**.

“Past and future are conditionally independent given the present.” — Blitzstein, 11:43

Homojen: $q_{ij} = P(X_{n+1} = j \mid X_n = i)$, n 'den bağımsız.

38.3 Geçiş Matrisi Q

Örnek (4 durum):

$$Q = \begin{pmatrix} 1/3 & 2/3 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \\ 1/2 & 0 & 1/4 & 1/4 \end{pmatrix}$$

Her satır 1'e toplanır (stokastik matris).

38.4 n -Adım Geçişler

X_n dağılımı satır vektörü s . LOTP:

$$P(X_{n+1} = j) = \sum_i s_i q_{ij} = (sQ)_j$$

$$X_n \sim s \Rightarrow X_{n+m} \sim sQ^m$$

$$P(X_{n+m} = j \mid X_n = i) = (Q^m)_{ij}$$

38.5 Durağan Dağılım

$$sQ = s$$

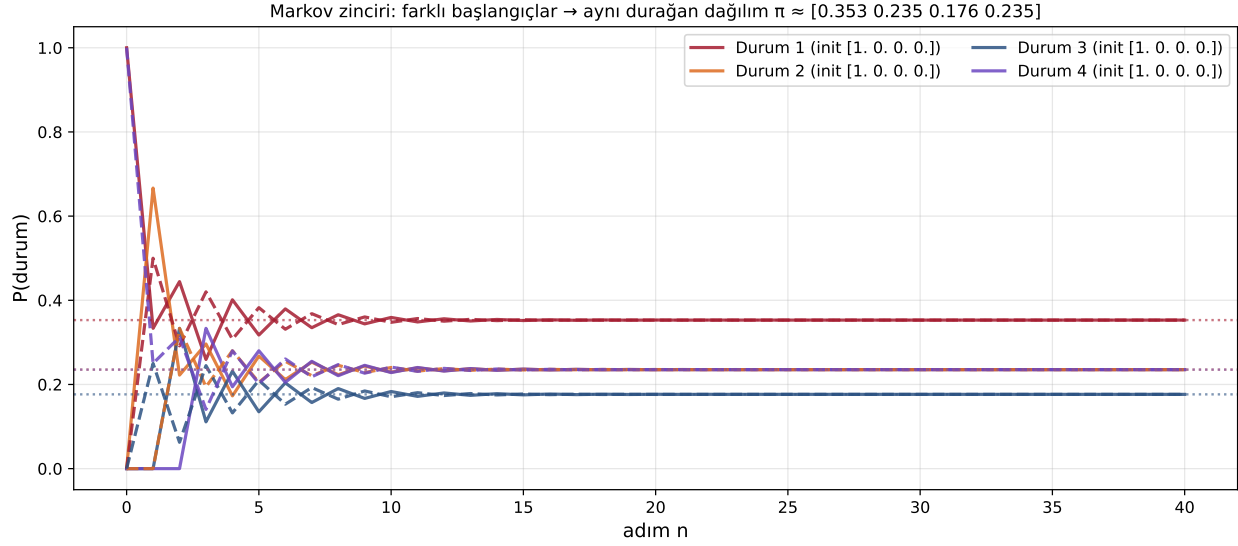
Bir kez girilince **sonsuz dek** s . Q 'nun **özdeğer-1 sol özvektörü**.

Dört soru: Var mı? Tek mi? Yakınsar mı? Nasıl hesaplanır?

Hafif koşullarda → evet, evet, evet.

! Builder Notu — MCMC ve Diffusion

MCMC: durağan dağılımı hedef posterior olan zincir kur (**Metropolis-Hastings, Gibbs, HMC**), çalıştır, örnek. Bayeşçi ML'in motoru. **Diffusion modelleri:** ileri süreç gürültü ekleyen zincir, geri süreç onu çözen — üretim = zinciri tersine çalıştırma.



Şekil 38.1: 4-durumlu Markov zinciri farklı başlangıçlardan aynı durağan dağılıma yakınsar. PageRank, MCMC, diffusion'ın temel mekanizması: “yeterince çalıştır, durağana git”.

38.6 Tarihsel İki Kullanım

1. **Modelleme:** gerçek sistem (Pushkin'in sesli/sessiz harfleri, hava durumu).
2. **MCMC:** kendi zincirini kur, durağan dağılımı hedeflediğin dağılım.

“You can construct your own Markov chain that will converge to a distribution you're interested in.” — Blitzstein, 23:19

38.7 Bu Dersin Özeti

1. **Markov özelliği:** koşullu bağımsızlık.
2. Q : stokastik, satır toplamı 1.
3. sQ^m : m adım sonra.
4. $sQ = s$: durağan.

! Tek bir cümle

Markov zinciri: gelecek geçmişe **yalnız şimdi üzerinden** bağlı. Tüm davranış Q 'da; kuvvetleri zamanı sarar; durağan $sQ = s$ uzun-vade dengesi. **MCMC, diffusion, RL (MDP), PageRank, n-gram, HMM** — hepsi.

38.8 Kontrol Soruları

i Soru 1: Markov özelliği = bağımsızlık mı?

Cevap: Koşullu bağımsızlık (şimdi verildiğinde). X_n bilinmiyorsa X_{n-1} hala bilgi taşır.

i Soru 2: Q'nun yapısal özelliği?

Cevap: Her satır 1'e toplanır (negatif olmayan). Sütun toplamak zorunda değil.

i Soru 3: $X_n \sim s$ 3 adım sonra?

Cevap: sQ^3 dağılım; $(Q^3)_{ij}$ nokta olasılığı.

i Soru 4: $sQ = s$ neden durağan?

Cevap: Bir adım sonra dağılım değişmez → sonsuza dek s . Özdeğer-1 sol özvektör.

38.9 Egzersizler

Egzersiz 1. Geçerli Q ? (a) (.3,.7;.6,.4), (b) (.5,.5;.8,.3), (c) (1,0;.2,.8).

Egzersiz 2. $Q = (.9,.1;.5,.5)$. $(Q^2)_{11}$?

Egzersiz 3. $s = (1, 0)$. sQ , sQ^2 hangi yöne?

Egzersiz 4. Yukarıdaki Q için durağan dağılım çöz.

Egzersiz 5. (Python — Markov + durağan)

Satır toplamı: [1. 1. 1. 1.]

[1. 0. 0. 0.] → [0.3529 0.2353 0.1765 0.2353]

[0. 0. 0. 1.] → [0.3529 0.2353 0.1765 0.2353]

Durağan (özvektör): [0.3529 0.2353 0.1765 0.2353]

38.10 Sonraki Ders İçin Hazırlık

Ders 32: Markov Sınıflandırma + Tersinirlik — irreducibility, periyot, **detailed balance** (MCMC'nin temeli).

⚠ Ders 32 öncesi yapılacak

- Egzersizleri çöz.
- Durağan denklemi içselleştir.

38.11 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Not
Markov özelliği	$P(X_{n+1} \ X_n, \dots) = P(X_{n+1} \ X_n)$	Geçmiş \perp gelecek şimdi
q_{ij}	$P(X_{n+1} = j X_n = i)$	Homojen
Q	(q_{ij})	Satır toplamı 1
Evrim	sQ	Satır vektörü
m-adım	$sQ^m, (Q^m)_{ij}$	Matris kuvveti
Durağan	$sQ = s$	Özdeğer-1 sol özvektör

38.12 ML Bağlantıları Özeti

💡 6 köprü

1. **MCMC** → Metropolis-Hastings, Gibbs, HMC.
2. **Diffusion** → ileri/geri zincir.
3. **RL (MDP)** → Bellman, value iteration.
4. **PageRank** → durağan dağılım.
5. **n-gram** = $(n - 1)$. Markov; Transformer kırar.
6. **HMM forward** = sQ .

! Tek bir şey alıp gideceksen

Markov: **gelecek yalnız şimdi üzerinden** bağlı. Q davranışı, Q^n zaman, $sQ = s$ uzun-vade. **MCMC + diffusion + RL + PageRank** hepsi.

39 Markov: Sınıflandırma ve Tersinirlik

Detailed balance = MCMC'nin temeli

Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 32](#) (≈48 dk)
- **Okuma süresi:** ≈23 dk

39.1 Bu Derste Ne Var?

1. **İndirgenemezlik (irreducibility):** her durumdan her duruma.
2. **Geri dönüşlü / geçici / yutucu / periyodik** durumlar.
3. **Durağan teoremleri:** var + tek + $s_i = 1/r_i$ + yakınsama (aperiyodik).
4. **Tersinirlik (detailed balance):** $s_i q_{ij} = s_j q_{ji}$.
5. **Yönsüz ağ rassal yürüyüşü:** $s_i = d_i / \sum d_j$ (dereceyle orantılı).

Builder Notu — ML Köprüleri

- **Detailed balance = Metropolis-Hastings** tasarım ilkesi.
- **İndirgenemez + aperiodyk = ergodiklik** (MCMC neden işler).
- **Rassal yürüyüş \propto derece \rightarrow PageRank, degree centrality.**
- **Graf gömme (DeepWalk, node2vec)** = ağda rassal yürüyüş.
- **GNN difüzyonu** = komşu ortalaması rassal yürüyüş.

39.2 İndirgenemezlik

Her durumdan her duruma **ulaşılabilir** (pozitif olasılıkla, sonlu adımda).

“Irreducible means you can get from anywhere to anywhere.” — Blitzstein, 5:32

Builder Notu — MCMC İndirgenemezlik

MCMC zincirinin **örnek uzayını tamamen dolaşabilmesi** koşulu. **PageRank**'te **ışınlanma (teleportation)** indirgenemezliği garanti eder (yoksa çıkışsız sayfalar tuzak olur).

39.3 Durum Türleri

- **Geri dönüşlü:** olasılık 1 ile döner (sonsuz kez). İndirgenemez sonlu zincirde **tüm durumlar** geri dönüşlü.
- **Geçici:** eninde sonunda dönmez.
- **Yutucu:** girince hep orada (kumarbazın iflası).
- **Periyodik:** belirli periyod katlarında dönüş.

39.4 Durağan Dağılım Teoremleri

İndirgenemez sonlu zincir:

1. **Var.**
2. **Tek.**
3. $s_i = 1/r_i$ (r_i = ortalama geri dönüş süresi).
4. **Aperiyodikse** $tQ^n \rightarrow s$ (başlangıçtan bağımsız).

! Builder Notu — Ergodiklik

İndirgenemez + aperiyodik = ergodiklik. Zaman-ortalaması (tek uzun yürüyüş) = uzay-ortalaması (durağan beklenti). **Tek MCMC zincirinden örnek toplayabilmenin gerekçesi.**

39.5 Tersinirlik (Detailed Balance)

$$s_i q_{ij} = s_j q_{ji}$$

Teorem: s detailed balance sağlıyorsa s **durağandır**.

İspat: \sum_i uygula:

$$\sum_i s_i q_{ij} = \sum_i s_j q_{ji} = s_j \sum_i q_{ji} = s_j \cdot 1 = s_j$$

Sol = $(sQ)_j \rightarrow sQ = s$. ■

! Builder Notu — Metropolis-Hastings'in Sırrı

Detailed balance = MH'in tasarım ilkesi. Kabul olasılığı ($\min(1, \dots)$) tam $s_i q_{ij} = s_j q_{ji}$ sağlanacak şekilde — **normalizasyon sabiti gerekmez**, sadece **oran** s_j/s_i ! Bu, Bayesçi posterior'un (sabit hesaplanamayan) örneklenmesini mümkün kılan **can alıcı numara**.

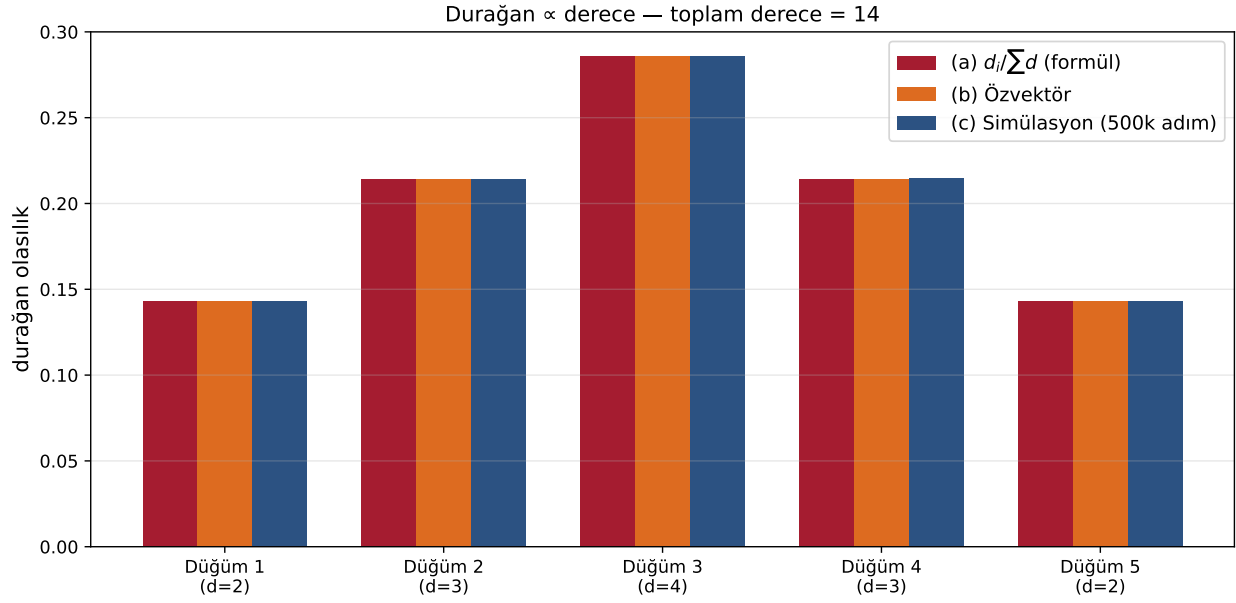
39.6 Yönsüz Ağda Rassal Yürüyüş

d_i = derece, $q_{ij} = 1/d_i$ (komşuysa).

Detailed balance kontrolü: $d_i \cdot 1/d_i = 1 = d_j \cdot 1/d_j \checkmark$

$$s_i = \frac{d_i}{\sum_j d_j}$$

Durağan dağılım dereceyle orantılı — matris çözmeden.



Şekil 39.1: Yönsüz ağda rassal yürüyüş: durağan dağılım dereceyle orantılı. Üç farklı hesaplama yolu (formül, özvektör, simülasyon) aynı sonuca varır. PageRank, GNN, DeepWalk’un temel sezgisi.

! Builder Notu — PageRank, GNN, DeepWalk

“Durağan \propto derece” = **PageRank’in çekirdeği**. Web grafında sayfa önemi = rassal gezginin durağan dağılımı. **Degree centrality, DeepWalk / node2vec** (rassal yürüyüş örnekleme), **GNN’in komşu ortalaması adımları** — hepsi bu zincirden.

39.7 Bu Dersin Özeti

1. **İndirgenemez** + sonlu \rightarrow tüm durumlar geri dönüşlü.
2. **Durağan teoremi:** var + tek + $s_i = 1/r_i$ + aperiodykse yakınsar.
3. **Detailed balance:** $s_i q_{ij} = s_j q_{ji} \implies$ durağan.
4. **Rassal yürüyüş:** $s_i \propto d_i$.

! Tek bir cümle

İndirgenemez + aperiodyik → durağan var/tek + yakınsama (**ergodiklik = MCMC işler**). **Detailed balance** ($s_i q_{ij} = s_j q_{ji}$) durağanı matris çözmeden verir — **Metropolis-Hastings'in tasarım ilkesi**. **Yönsüz ağ rassal yürüyüş** = derece orantılı = **PageRank**.

39.8 Kontrol Soruları

i Soru 1: İndirgenemez sonluda geçici olabilir mi?

Cevap: Hayır. Hepsi geri dönüşlü.

i Soru 2: $s_i = 0.2 \rightarrow r_i$?

Cevap: $r_i = 5$. Ters orantı.

i Soru 3: Detailed balance → durağan ispatı?

Cevap: i üzerinden topla: $\sum s_i q_{ij} = s_j \cdot 1 = s_j \rightarrow sQ = s$.

i Soru 4: $d = (2, 2, 3, 1)$ durağan?

Cevap: $s = (2/8, 2/8, 3/8, 1/8)$.

39.9 Egzersizler

Egzersiz 1. Zinciri sınıflandır: $1 \rightarrow 2, 2 \rightarrow 1/2 \rightarrow 3, 3 \rightarrow 4, 4 \rightarrow 3$.

Egzersiz 2. İki durumlu Q, detailed balance kontrolü.

Egzersiz 3. C_4 döngüsü. Periyodik mi?

Egzersiz 4. Yıldız grafi: merkez vs yaprak.

Egzersiz 5. (Python — Üç yolla durağan)

Formül: $[0.25 \ 0.25 \ 0.375 \ 0.125]$

Özvektör: $[0.25 \ 0.25 \ 0.375 \ 0.125]$

Detailed balance: True

39.10 Sonraki Ders İçin Hazırlık

Ders 33: Markov + PageRank — ağırlıklı yürüyüş, Google'ın milyar dolarlık fikri.


 Ders 33 öncesi yapılacak

- Egzersizleri çöz.
- Detailed balance + derece formülünü içselleştir.


39.11 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Not
İndirgenemez	Her \rightarrow her	“Connected”
Geri dönüşlü	$P(\text{dönüş}) = 1$	Sonlu indirgenemez = hepsi
Yutucu	Hep orada	Iflas
Periyodik	Periyot katları	Yakınsamayı bozar
Durağan var+tek	$sQ = s$	İndirgenemez sonlu
$s_i = 1/r_i$	Ort. geri dönüş	Ters orantı
Yakınsama	$tQ^n \rightarrow s$	Aperiyodik
Detailed balance	$s_i q_{ij} = s_j q_{ji}$	\Rightarrow durağan
Rassal yürüyüş	$s_i = d_i / \sum d_j$	Dereceyle orantılı

39.12 ML Bağlantıları Özeti

 6 köprü

1. **Metropolis-Hastings** \rightarrow detailed balance.
2. **Ergodiklik** \rightarrow tek zincirden örnek.
3. **PageRank** \rightarrow durağan dağılım.
4. **DeepWalk / node2vec** \rightarrow rassal yürüyüş örnekleme.
5. **GNN** \rightarrow rassal yürüyüş difüzyonu.
6. **RL terminal** \rightarrow yutucu durum.

 Tek bir şey alıp gideceksen

İndirgenemez + aperiyodik = MCMC'nin işlemesinin sebebi. Detailed balance = Metropolis-Hastings'in sırrı. Rassal yürüyüş \propto derece = PageRank.

40 Markov: Ağırlıklı Yürüyüş ve Google PageRank

Power iteration + ışınlanma; olasılığın milyar dolarlık fikri

i Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 33](#) (≈48 dk)
- **Okuma süresi:** ≈23 dk

40.1 Bu Derste Ne Var?

1. **Ağırlıklı rassal yürüyüş:** $q_{ij} = w_{ij} / \sum_k w_{ik}$, $w_{ij} = w_{ji}$.
2. **Her tersinir zincir** ağırlıklı ağ yürüyüşüdür.
3. **PageRank:** $s = sQ$, web önem skoru.
4. **Google zinciri:** $G = \alpha Q + (1 - \alpha)J/M$.
5. **Power iteration:** $tG^n \rightarrow s$.

💡 Builder Notu — ML Köprüleri

- **PageRank** → eigenfactor, **eigenvector centrality**, öneri sistemleri.
- **Power iteration** → büyük-ölçek özvektör, **spektral kümeleme**.
- **Işınlanma** → **düzenleştirme** (label smoothing, ϵ -greedy).
- **Ağırlıklı graf** → spektral, **GNN**, attention-weighted komşu.
- **DeepWalk / node2vec** → rassal yürüyüş gömme.

40.2 Ağırlıklı Rassal Yürüyüş

Yönsüz, $w_{ij} = w_{ji}$:

$$q_{ij} = \frac{w_{ij}}{\sum_k w_{ik}}$$

Detailed balance kontrolü → durağan ağırlık-toplamıyla orantılı:

$$s_i \propto \sum_k w_{ik}$$

40.3 Her Tersinir Zincir = Ağırlıklı Yürüyüş

Teorem: Tersinir zincir $\rightarrow w_{ij} = s_i q_{ij}$ koy. Tersinirlik ($s_i q_{ij} = s_j q_{ji}$) $\rightarrow w_{ij} = w_{ji}$ (simetrik). Geçiş hesapla:

$$\frac{w_{ij}}{\sum_k w_{ik}} = \frac{s_i q_{ij}}{s_i \sum_k q_{ik}} = q_{ij}$$

40.4 PageRank Markov Zinciri

İçgörü: Sayfa önemi = bağlantı verenlerin öneminden doğar (sabit nokta).

$$s_j = \sum_i s_i q_{ij} \Rightarrow s = sQ$$

Seyreltme: $q_{ij} = 1/(\text{çıkış sayısı})$. **Sarkan sayfa:** $1/M$ tüm sayfalara.

“The importance of a page is the long-run fraction of time that you spend at that page, randomly surfing the web.” — Blitzstein, 31:34

40.5 Google Zinciri ve Işınlanma

$$G = \alpha Q + (1 - \alpha) \frac{J}{M}, \quad \alpha = 0,85$$

%85 bağlantı takip + %15 rastgele sayfaya ışınlan.

Neden ışınlanma?

- **İndirgenemezlik garantisi.**
- **Aperiyodiklik / sıfırsızlık** \rightarrow durağan teoremleri uygulanır (tersinir olmasa bile).

! Builder Notu — Düzenleştirme

Işınlanma = düzenleştirme: ana modele küçük düzgün bileşen karıştırmak. Aynı kalıp: **label smoothing**, ϵ -greedy (RL), **Laplace düzeltmesi**, **dropout**. Patolojileri eler.

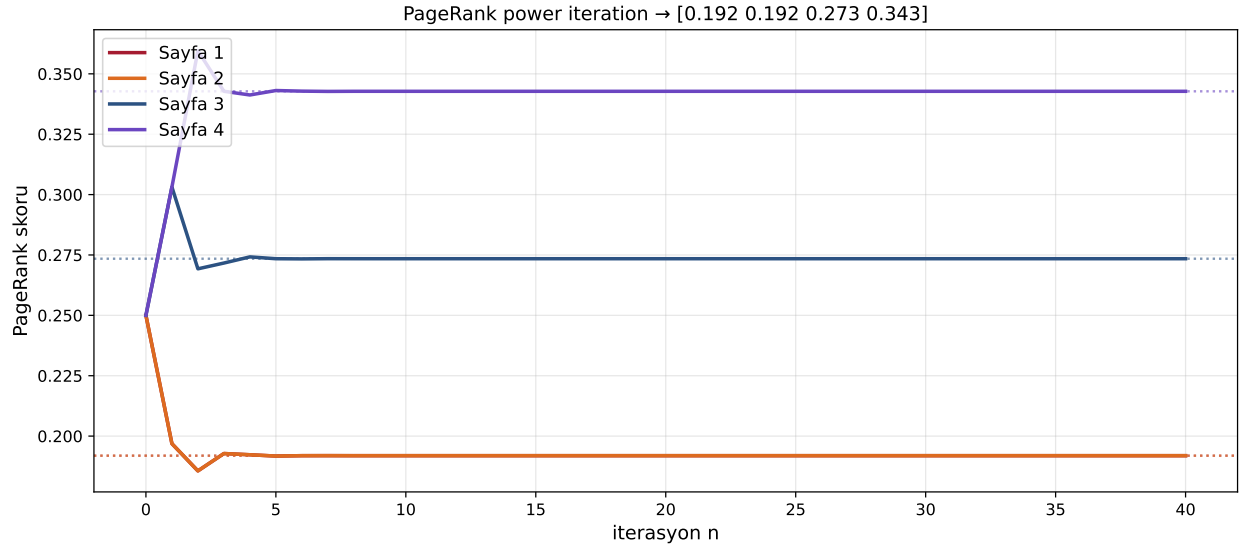
40.6 Power Iteration

$$t, tG, tG^2, \dots, tG^n \rightarrow s$$

Neden ucuz?

- Q çok seyrek (tipik sayfa az bağlantı).
- tJ = hepsi-1 vektörü (toplamı 1).

Her adım $O(\text{kenar sayısı})$ — milyarlarca sayfada uygulanabilir.



Şekil 40.1: PageRank power iteration: 4 sayfalı web (sayfa 4 sarkan). tG^n hızla durağan dağılıma yakınsar. $\alpha = 0.85$, ışınlanma minik ama indirgenemezlik + aperiodyklik garantisi. Google’ın milyar dolarlık fikri tek bir Markov zincirinin durağan dağılımı.

Final sıralama (yüksek→düşük): [4 3 1 2]

! Builder Notu — Spektral Yöntemler

Power iteration = en büyük özdeğerin özvektörünü bulmanın standart yöntemi. **Spektral kümeleme**, **latent faktör çıkarımı**, hatta bazı verimli **attention** yaklaşımları (linear attention) aynı desen. “Devasa ama seyrek operatörün baskın özvektörü, operatörü tekrar tekrar uygulayarak” — büyük-ölçek ML’in her yerinde.

40.7 Bu Dersin Özeti

1. **Ağırlıklı yürüyüş:** $s_i \propto \sum w_{ik}$.
2. **Her tersinir = ağırlıklı ağ yürüyüşü** ($w_{ij} = s_i q_{ij}$).

3. **PageRank:** $s = sQ$.
4. **Google:** $G = \alpha Q + (1 - \alpha)J/M$.
5. **Power iteration** + seyreklik.

! Tek bir cümle

Her tersinir Markov zinciri ağırlıklı ağ yürüyüşüdür (durağan \propto ağırlık-derece); **PageRank** ise tersinir-olmayan dev zincir — web sayfalarını **ışınlanmayla sağlamlaştırılmış** Google zincirinin durağan dağılımını **power iteration** ile sıralar. **Önem = uzun vadede o sayfada geçirilen zaman.**

40.8 Kontrol Soruları

i Soru 1: Her tersinir \rightarrow ağ yürüyüşü ispatı?

Cevap: $w_{ij} = s_i q_{ij}$. Tersinirlik $\rightarrow w_{ij} = w_{ji}$. Geçiş: $w_{ij} / \sum w_{ik} = q_{ij}$ (s_i sadeleşir).

i Soru 2: Seyreltme + sarkan sayfa?

Cevap: $q_{ij} = 1/\text{çıkış sayısı}$. Sarkan $\rightarrow 1/M$ tüm sayfalara \rightarrow satır toplamı 1.

i Soru 3: Işınlanma niye?

Cevap: İndirgenemezlik + sıfırsız \rightarrow durağan teoremleri uygulanır.

i Soru 4: Neden power iteration?

Cevap: $O(m^3)$ imkânsız. $tG^n \rightarrow s$ ve seyreklik + $tJ = \text{hepsi-1} \rightarrow$ her adım ucuz.

40.9 Egzersizler

Egzersiz 1. Üçgen, $w_{12} = 1, w_{13} = 2, w_{23} = 3$. Durağan?

Egzersiz 2. Detailed balance kontrolü.

Egzersiz 3. Küçük web: $1 \rightarrow 2, 3; 2 \rightarrow 3; 3 \rightarrow 1$. Q matrisi.

Egzersiz 4. Egzersiz 3'te $\alpha = 0,85$ ile G_{11} ?


Egzersiz 5. (Python — PageRank)

PageRank: [0.1919 0.1919 0.2734 0.3428]

Sıralama: [4 3 2 1]

40.10 Sonraki Ders İçin Hazırlık

Ders 34: İleriye Bakış — kursun son dersi! Top Ten + ötesi.


 Ders 34 öncesi yapılacak

- Tüm kursu gözden geçir.
- Markov + PageRank'i içselleştir.


40.11 Anahtar Kavramlar (Cheat Sheet)

Kavram	Tanım	Not
Ağırlıklı yürüyüş	$q = w / \sum w$	Simetrik
Durağan	$s_i \propto \sum w$	Ağırlık-derece
Tersinir = ağ	$w_{ij} = s_i q_{ij}$	Hepsi
PageRank	$s = sQ$	Önem
Seyreltme	$1/\text{çıkış}$	Sarkan: $1/M$
Google	$G = \alpha Q + (1 - \alpha)J/M$	$\alpha = 0.85$
İşinlanma	Düzenleştirme	Patoloji eler
Power iteration	tG^n	Seyrek + tJ ucuz

40.12 ML Bağlantıları Özeti

 6 köprü

1. **PageRank** → eigenfactor, eigenvector centrality.
2. **Power iteration** → spektral kümeleme.
3. **İşinlanma** → label smoothing, ϵ -greedy.
4. **Ağırlıklı graf** → GNN, attention.
5. **Tersinirlik** → MCMC spektral analiz.
6. **Random walk gömme** → DeepWalk, node2vec.

 Tek bir şey alıp gideceksen

Her tersinir zincir = ağırlıklı ağ yürüyüşü. PageRank = devasa Markov zincirinin durağan dağılımı + power iteration. Olasılığın milyar dolarlık fikri.

41 İleriye Bakış (A Look Ahead)

Top Ten, regresyon, Horvitz-Thompson, Basu'nun Fili

i Bölüm bilgisi

- **Blitzstein'in videosu:** [YouTube — Lecture 34](#) (≈37 dk)
- **Okuma süresi:** ≈18 dk
- **Not:** Stat 110'un **son dersi**. Toparlama + ileriye bakış. Ders 35 (kapsam dışı bonus konuşma) çevrilmedi.

41.1 Bu Derste Ne Var?

1. **Top Ten** — 3 tema: rassallık / beklenti / uzun-vade.
2. **Sonrası:** çıkarım, regresyon, stokastik süreçler.
3. **Regresyon türetmesi:** $\beta_1 = \text{Cov}(X, Y) / \text{Var}(X)$.
4. **Horvitz-Thompson (IPW):** $\hat{T} = \sum Y_j I_j / p_j$.
5. **Basu'nun Fili:** yansız \neq iyi.

💡 Builder Notu — ML Köprüleri

- **Regresyon = en küçük kareler** → tüm denetimli öğrenme.
- **IPW** → **nedensel çıkarım** (propensity score), **off-policy RL** (importance sampling).
- **Basu'nun Fili** → **bias-variance tradeoff**, ridge/lasso düzenleme.
- **Olasılık vs çıkarım** → **generative vs discriminative**.
- **Koşullama** → istatistiğin ruhu, modern ML'in temeli.

41.2 Top Ten: Stat 110'un Özü

3 tema:

(1-4) Rassallık nedir?

1. **Koşullama** — istatistiğin ruhu.
2. **Simetri** — güçlü ama tehlikeli.
3. **Rastgele değişkenler ve dağılımları**.

4. **Hikâyeler (stories)** — story proofs.

(5-7) Beklenti hesaplama.

5. **Doğrusallık** — bağımlılıkta bile.
6. **Gösterge** — favori numara.
7. **LOTUS** — vazgeçilmez.

(8-10) Uzun-vade.

8. **BSY.**
9. **MLT.**
10. **Markov zincirleri** — iid'nin bir adım ötesi.

💡 Builder Notu — ML Temeli

10 fikir = ML'in olasılıksal temeli. Koşullama → Bayesian/attention; hikâyeler → üretici tasarım; doğrusallık+LOTUS → $E[L]$ + Monte Carlo; BSY/MLT → SGD + belirsizlik; Markov → MCMC/diffusion/RL.

41.3 Regresyon: β_1 Türetmesi

$$Y = \beta_0 + \beta_1 X + \varepsilon, E(\varepsilon|X) = 0.$$

İki tarafın X ile Cov'unu al:

$$\text{Cov}(Y, X) = \beta_1 \text{Var}(X) + \text{Cov}(\varepsilon, X)$$

$\text{Cov}(\varepsilon, X) = 0$ ispatı (Adam + bilineni dışarı):

$$\text{Cov}(\varepsilon, X) = E(\varepsilon X) = E(X \cdot E(\varepsilon|X)) = E(X \cdot 0) = 0$$

$$\beta_1 = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}$$

Koşullu beklenti = izdüşüm → en küçük karelerin geometrik temeli.

❗ Builder Notu — Sınır Ağlarının Son Katmanı

$\beta_1 = \text{Cov}/\text{Var} = \text{en küçük kareler çözümü}$. “Koşullu beklenti = izdüşüm” görüşü doğrusal regresyondan **kernel regresyona, sınır ağlarının son katmanına** (öğrenilen özelliklerin hedefe izdüşümü) uzanır.

41.4 Horvitz-Thompson (IPW)

Sonlu nüfus Y_1, \dots, Y_N sabit. p_j = kişi j 'nin dahil olma olasılığı.

Kestirici:

$$\hat{T} = \sum_{j=1}^N \frac{Y_j}{p_j} I_j$$

Yansız ($E(I_j) = p_j$ temel köprü):

$$E(\hat{T}) = \sum \frac{Y_j}{p_j} p_j = \sum Y_j$$

! Builder Notu — IPW Her Yerde

IPW modern ML'in devasa aracı: Nedensel çıkarım (propensity score weighting), **off-policy RL** (importance sampling), **örneklem yanlılığı** düzeltme. “Gözlemi olasılığına böl” = Ders 9 gösterge + Adam yasasının doğrudan uygulaması.

41.5 Basu'nun Fili: Yansız ≠ İyi

50 fil, sahip “ortalama görünen” Stampy'yi tartmak istiyor. $p_{\text{Stampy}} = 0,99$, kalan 49 file 0,01 paylaşılır.

Sonuç: %99 olasılıkla sadece Stampy çıkar $\rightarrow \hat{T} = Y_{\text{Stampy}} \cdot 100/99 \approx$ bir filin ağırlığı! 50 fil için **berbat**. Yansız ama **varyans felaket**.

! Builder Notu — Bias-Variance Tradeoff

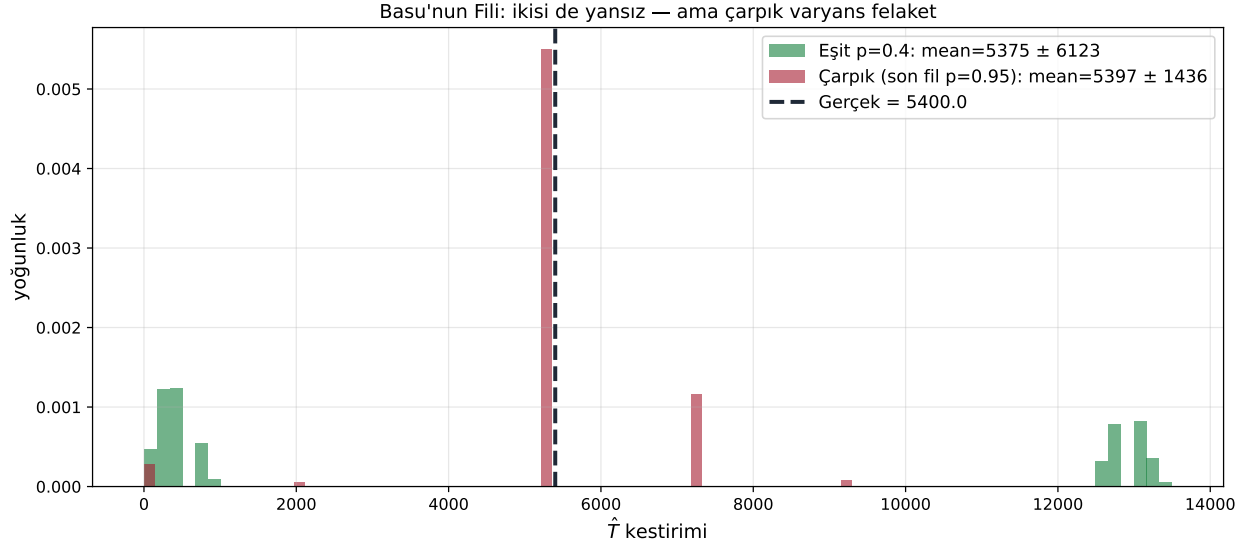
Basu = ML'in en sinsi dersi: bir metriği (yansızlık) optimize etmek hedefe ulaştırmayabilir. **Bias-variance tradeoff** — ridge/lasso bilerek yanlılık ekler ama varyansı düşürür. **Off-policy RL'de IPW patlar** \rightarrow “clipped/weighted importance sampling” ile evcilleştirilir.

41.6 Sonrası Nereye?

- **Çıkarım** (model \rightarrow veri ile veri \rightarrow parametre, ML'in iki yarısı).
- **Regresyon / lineer modeller.**
- **Stokastik süreçler** (Markov'un devamı).
- **R / Python** öğren.

“110 is actually a recurrent state in this chain... revisiting the material over and over again is a good thing.” — Blitzstein, 36:18

41 İleriye Bakış (A Look Ahead)



Şekil 41.1: Basu'nun Fili: yansız ama yüksek varyans. Çarpık olasılık ile (son file $p=0.95$) ortalama doğru ama büyük dalgalanma. Eşit olasılıkta varyans düşük. ML dersi: bias-variance tradeoff.

41.7 Kurs Kapanışı 🎓

Stat 110 tamamlandı! 34 derste olasılığı sıfırdan — sayma + koşullamadan dağılımlar, beklenti, eşitsizlikler, BSY/MLT, Markov zincirleri ve PageRank'a — kurduk.

Karpathy: “Olasılık, makine öğrenmesinin dilidir.” Bu setteki tüm Builder Notları o köprüyü kurdu.

Buradan ileri: çıkarım, regresyon, stokastik süreçler, modern ML — hepsi bu temelin üzerine.

! Tek bir cümle

Stat 110'un özü **koşullamadır**; 10 temel fikir “rassallık nedir / beklenti nasıl / uzun-vade ne” diye üçe ayrılır; regresyondan IPW'ye her uygulamada Adam yasası + gösterge + izdüşüm; ama **Basu'nun Fili** hatırlatır: **yansız \neq iyi, düşünmek gerekir.**

41.8 Kontrol Soruları

i Soru 1: Koşullama neden istatistiğin ruhu?

Cevap: Belirsizlik altında bilgi güncellenmenin temel mekanizması — koş. olasılık/beklenti, Bayes, Markov (geçmiş \perp gelecek | şimdi). Tüm kursun temasıdır.

i Soru 2: β_1 türetmesinde $E(\epsilon|X)=0$ nasıl?

Cevap: İki yer: (1) Adam: $E(\epsilon) = E(E(\epsilon|X)) = 0$. (2) $\text{Cov}(\epsilon, X) = E(X \cdot E(\epsilon|X)) = 0$. Hata açıklayıcıya **ortogonal** = en küçük kareler.

i Soru 3: HT yansızlığı?

Cevap: Gösterge $E(I_j) = p_j$. $E(\hat{T}) = \sum(Y_j/p_j)p_j = \sum Y_j$. Gösterge numarası rassal paydayı sabite çevirir.

i Soru 4: Basu'nun Fili neyi gösterir?

Cevap: Yansızlık \neq iyilik. Varyans kritik. ML karşılığı **bias-variance tradeoff** — düzenleme, ridge/lasso bilerek yanlılık ekler.

41.9 Egzersizler

Egzersiz 1. β_0 'ı $E(Y)$, β_1 , $E(X)$ ile bul.

Egzersiz 2. $E(\varepsilon|X) = 0 \Rightarrow \text{Cov}(\varepsilon, g(X)) = 0$ tüm g için.

Egzersiz 3. $Y = (10, 20, 30)$, $p = (0,5, 0,5, 1)$. Tüm örneklemeler üzerinden $E(\hat{T})$.

Egzersiz 4. Basu, iki olası değer arası fark? Varyans hakkında ne?

Egzersiz 5. (Python — IPW + Basu)

Eşit $p=0.4$ | gerçek=5400 ort=5349 (yansız) std=6118
Çarpık (son $p=0.95$) | gerçek=5400 ort=5390 (yansız) std=1441

41.10 Anahtar Kavramlar (Cheat Sheet)

Kavram	Formül	Not
İstatistiğin ruhu	Koşullama	Her yer
3 tema	Rassallık / beklenti / uzun-vade	(1-4)/(5-7)/(8-10)
β_1	$\text{Cov}(X, Y)/\text{Var}(X)$	İzdüşüm
Ortogonalite	$\text{Cov}(\varepsilon, X) = 0$	Adam yasası
Horvitz-Thompson	$\sum Y_j I_j / p_j$	IPW; yansız
Basu	Yansız \neq iyi	Bias-variance
Sonrası	Çıkarım, regresyon, süreçler	R/Python

41.11 ML Bağlantıları Özeti

💡 6 köprü

1. **Regresyon** → tüm denetimli öğrenme; sinir ağı son katmanı.
2. **IPW** → **nedensel çıkarım**, off-policy RL, dengesiz veri.
3. **Basu** → **bias-variance tradeoff**, düzenleme.
4. **Koşullama** → Bayesian ML, **attention**.
5. **Olasılık vs çıkarım** → generative vs discriminative.
6. **Stokastik süreçler** → RL, **diffusion**.

! Tek bir şey alıp gideceksen

Stat 110'un özü koşullamadır. Adam yasası + gösterge + izdüşüm = ML'in olasılıksal motoru. Yansızlık ≠ iyilik (Basu) → bias-variance her yerde. Olasılık makine öğrenmesinin dilidir (Karpathy).

41.12 🎓 Kurs Sonu

34/34 ✅. Stat 110 Türkçe ders notları tamamlandı.

Bu set Blitzstein'in Harvard Stat 110 dersinin Türkçe paralel okumasıdır. Her dersin **Builder Notları ML/AI** için yapı taşlarını gösterir.

Bundan sonra:

- stat110.net — strategic practice, kitap (Blitzstein & Hwang 2014).
- Çıkarım: Stat 111, regresyon: Stat 139, stokastik süreçler: Stat 171.
- R + Python.

“Tuning is physics, mathematics, and logic — and so is probability.”

Karpathy: “Olasılık, makine öğrenmesinin dilidir.”

Bu setin tüm Builder Notları o dili öğretmek içindi.