Olasılık ve Dağılımlar

Sinan Yıldırım

MDBF, Sabancı Üniversitesi

3 Şubat 2021

Olasılık uzayı $(\Omega, \mathcal{A}, P)$

- **Örneklem uzayı** Ω

- **Olay uzayı** $\mathcal{A}$

  Olay $A \in \mathcal{A}$

- **Olasılık** $P : \mathcal{A} \mapsto [0, 1]$.

Olasılık uzayı $(\Omega, \mathcal{A}, P)$

Rassal değişken

- **Hedef uzay** $\mathcal{T}$ (çoğu zaman gerçel sayılardan oluşur.)
- **Rassal değişken**

$$X : \Omega \mapsto \mathcal{T}$$

- Bir $S \subseteq \mathcal{T}$ için

$$P_X(X \in S) = P(X^{-1}(S)) = P \circ X^{-1}(S)$$

X'in **olasılık dağılımı (kanunu)**: $P \circ X^{-1}$.

- Ayrık rassal değişken: $\mathcal{T}$ sayılabilir.
- Sürekli rassal değişken: Örnek $\mathcal{T} = \mathbb{R}$.

Kümülatif dağılım fonsksiyonu

- Tek boyutlu değişken:

$$F(x) := P(X \leq x), \quad x \in \mathbb{R}.$$

- Çok boyutlu değişken: $X = (X_1, \ldots, X_D)$.
  $\boldsymbol{x} = (x_1, \ldots, x_D)$ için,

$$F(\boldsymbol{x}) := P(X_1 \leq x_1, \ldots, X_D \leq x_D), \quad \boldsymbol{x} \in \mathbb{R}^D.$$

Ayrık dağılımlar

$\mathcal{T}$ sayılabilir.

- Tek boyutlu değişkenler:
  Olasılık kütle fonksiyonu:

$$p(x) = P(X = x), \quad x \in \mathcal{T}$$

- Çok boyutlu değişkenler:
  $\mathcal{T}$ kartezyen çarpımı
  Bileşik olasılık kütle fonksiyonu (örn. iki değişken için)

$$p(x, y) = P(X = x, Y = y), \quad (x, y) \in \mathcal{T}$$

Sürekli dağılımlar

Olasılık yoğunluk fonksiyonu: $f : \mathbb{R}^D \mapsto [0, \infty)$

X'in olasılık yoğunluk fonksiyonu f ise,

Bir aralığın olasılığı:

$P(a \leq X \leq b)$

Bir değerin olasılığı:

$P(X = a)$

Toplam kuralı

Bileşik dağılım: $p(\boldsymbol{x}, \boldsymbol{y})$

Marjinal dağılım:

- $p(\boldsymbol{x})$
- $p(\boldsymbol{y})$

Koşullu olasılık ve çarpım kuralı

Olasılık uzayı: $(\Omega, \mathcal{A}, P)$

$A, B \in \mathcal{A}$ kümeleri için,

$P(A|B)$: A olayının B olayına **koşullu olasılığı**:

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

Çarpım kuralı:

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

Koşullu olasılık ve çarpım kuralı

Dağılımlar için:

$p(\boldsymbol{y}|\boldsymbol{x})$: Y'nin $X = \boldsymbol{x}$'e koşullu dağılımı

$$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}, \boldsymbol{y})}{p(\boldsymbol{y})}$$

Çarpım kuralı:

$$p(\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x})$$

Örnek

$p(x,y)$	$y=1$	$y=2$	$y=3$	$p(x)$
$x=1$	0.0	0.1	0.2	
$x=2$	0.4	0.2	0.1	
$p(y)$				

$p(x\|y)$	$y=1$	$y=2$	$y=3$
$x=1$			
$x=2$			

Bayes teoremi

Çarpım kuralının basit bir uygulaması:

Bayesci istatistik

Beklenti (Ortalama)

Beklenti doğrusal bir operatördür:

Ortalama, Medyan, Doruk

Varyans (tek boyutta)

Kovaryans

- Tek boyutta
- Çok boyutta

Kovaryans matrisi

Korelasyon

Ampirik ortalama

- Tek boyutta

$$\bar{x} = \frac{1}{N} \sum_{n=1}^{N} x_n$$

- Çok boyutta

Ampirik varyans/kovaryans matrisi

- Tek boyutta

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{n=1}^{N}(x_i - \bar{x})^2$$

- Çok boyutta

Toplamlar ve dönüşümler

- $\mathbb{E}(\boldsymbol{x}+\boldsymbol{y})$
- $\mathbb{E}(\boldsymbol{x}-\boldsymbol{y})$
- $\mathbb{V}(\boldsymbol{x}+\boldsymbol{y})$
- $\mathbb{V}(\boldsymbol{x}-\boldsymbol{y})$

Afin dönüşümler

$\boldsymbol{x}$: ortalaması $\boldsymbol{\mu}$ ve kovaryans matrisi $\boldsymbol{\Sigma}$

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$$

- $\mathbb{E}(\boldsymbol{y})$
- $\mathbb{V}(\boldsymbol{y})$
- $\mathrm{Cov}(\boldsymbol{x}, \boldsymbol{y})$

Bağımsızlık

X ve Y'nin bağımsızlığı için gerekli ve yeterli koşul

$$p(\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{x})p(\boldsymbol{y}), \quad \forall \boldsymbol{x}, \boldsymbol{y}$$

Gösterim: $X \perp\!\!\!\perp Y$

X ve Y bağımsız ise,

- $p(\boldsymbol{y}|\boldsymbol{x})$
- $p(\boldsymbol{x}|\boldsymbol{y})$
- $\mathbb{V}_{X,Y}[\boldsymbol{x} + \boldsymbol{y}]$
- $\mathrm{Cov}_{X,Y}[\boldsymbol{x}, \boldsymbol{y}]$

Koşullu bağımsızlık

X ve Y'nin Z'ye **koşullu bağımsızlığı** için gerekli ve yeterli koşul

$$p(\boldsymbol{x}, \boldsymbol{y} | \boldsymbol{z}) = p(\boldsymbol{x} | \boldsymbol{z}) p(\boldsymbol{y} | \boldsymbol{z}), \quad \forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$$

Gösterim: $X \perp\!\!\!\perp Y \mid Z$

Gauss dağılımı (Normal dağılım)

Tek boyutlu değişken: $X \sim \mathcal{N}(\mu, \sigma^2)$.

$$p(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad x \in \mathbb{R}$$

Yoğunluk fonksiyonu $\mathcal{N}(x|\mu, \sigma^2)$ olarak da gösterilir.

Gauss dağılımı (Normal dağılım)

Çok boyutlu değişken: $X \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

$$p(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{|2\pi\boldsymbol{\Sigma}|^{-1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right\}, \quad \boldsymbol{x} \in \mathbb{R}^D$$

Yoğunluk fonksiyonu $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ olarak da gösterilir.

Gauss dağılımı: marjinal ve koşullu dağılımlar

$$p(\boldsymbol{x}, \boldsymbol{y}) = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} \middle| \begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{bmatrix} \right)$$

- $p(\boldsymbol{x})$
- $p(\boldsymbol{y})$
- $p(\boldsymbol{x}|\boldsymbol{y})$

Gauss dağılımı: toplam

$X \sim \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$, $Y \sim \mathcal{N}(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y)$; X, Y bağımsız ve aynı boyda.

$Z = aX + bY$

- $\mathbb{E}[\boldsymbol{z}]$
- $\mathbb{V}[\boldsymbol{z}]$
- $p(\boldsymbol{z})$

Gauss dağılımı: doğrusal dönüşüm

$X \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

$\boldsymbol{A} \in \mathbb{R}^{M \times N}$, $Y = \boldsymbol{A}X$.

Y'nin dağılımı?

Gauss dağılımı: doğrusal dönüşüm

$Y \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{A} \in \mathbb{R}^{M \times N}$

$Y = \boldsymbol{A}X.$

X'in dağılımı?

- $M = N$, $\boldsymbol{A}$ terslenebilir:
- $M > N$, $\boldsymbol{A}$'nın kertesi N:

Gauss dağılımı: karışım

$p_1(x) = \mathcal{N}(x, \mu_1, \sigma_1^2)$, $p_2(x) = \mathcal{N}(x, \mu_2, \sigma_2^2)$ ve $0 < \alpha < 1$ olsun.

Bir karışım dağılımı:

$$p(x) = \alpha p_1(x) + (1 - \alpha) p_2(x)$$

- $\mathbb{E}(x)$

- $\mathbb{V}(x)$

Bu karışım dağılımı bir Gauss dağılımı mıdır?

Gauss dağılımı: örnekleme

Bernoulli dağılımı

Başarı olasılığı $\mu \in [0, 1]$

$$p(x|\mu) = \mu^x (1-\mu)^{1-x}, \quad x = 0, 1.$$

- $\mathbb{E}[x]$
- $\mathbb{V}[x]$

Binom dağılımı

N bağımsız deneme, her bir denemede başarı olasılığı $\mu \in [0, 1]$

$$p(x|N, \mu) = \binom{N}{x} \mu^x (1 - \mu)^{N-x}, \quad x = 0, \ldots, N.$$

- $\mathbb{E}[x]$
- $\mathbb{V}[x]$

Beta dağılımı

Rassal değişken: $\mu \in [0, 1]$,
Parametreler: $\alpha > 0$, $\beta > 0$

$$p(\mu|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \mu^{\alpha-1}(1-\mu)^{\beta-1}, \quad \mu \in [0, 1].$$

- $\mathbb{E}[\mu]$
- $\mathbb{V}[\mu]$

Eşleniklik

Bayes teoremi:

$$\text{sonsal dağılım} = \frac{\text{öncül dağılım} \times \text{olabilirlik}}{\text{kanıt}}$$
$$\propto \text{öncül dağılım} \times \text{olabilirlik}$$

Eşlenik dağılım: Öncül dağılım ile sonsal dağılım aynı forma sahipse, o öncül dağılım olabilirlik fonksiyonu için eşlenik bir dağılımdır.

Beta-Binom eşlenikliği

$$p(x|N,\mu) = \binom{N}{x} \mu^x (1-\mu)^{N-x}, \quad x = 0, \ldots, N.$$

$$p(\mu|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \mu^{\alpha-1} (1-\mu)^{\beta-1}, \quad \mu \in [0,1].$$

Yeterli istatistik

Bir veya birden fazla rassal değişkenin herhangi bir belirlenimci fonksiyonuna istatistik denir.

Yeterli istatistik: X'in dağılımı $p(x|\theta)$ olsun. Bir $\phi(x)$ istatistiğinin θ için yeterli istatistik olması için yeterli ve gerekli koşul:

$$p(x|\theta) = h(x)g_\theta(\phi(x))$$

$h(x) \geq 0$: θ'dan bağımsız bir fonksiyon.

Hangi dağılımlarda (x'in kendisinden başka) yeterli istatistiğe rastlanabilir?

Üstel dağılım aileleri

Bir üstel dağılım ailesinin üyeleri $\boldsymbol{\theta} \in \mathbb{R}^D$ ile şu şekilde belirlenir

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = h(\boldsymbol{x}) \exp(\boldsymbol{\theta}^T \phi(\boldsymbol{x}) - A(\boldsymbol{\theta}))$$

- $\phi(\boldsymbol{x}) \in \mathbb{R}^D$: yeterli istatistikler
- $\boldsymbol{\theta}$: doğal parametre
- $A(\boldsymbol{\theta})$: düzgeleştirici (normalize edici) katsayı (log-bölüntüleme fonksiyonu)
- $h(\boldsymbol{x}) \geq 0$: $\boldsymbol{\theta}$'dan bağımsız bir fonksiyon.

Gauss dağılımı

Bernoulli dağılımı

Üstel dağılımlar ve eşleniklik

Olabilirlik fonksiyonu üstel dağılım ise, mutlaka eşlenik dağılımı vardır.

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = h(\boldsymbol{x}) \exp\left(\boldsymbol{\theta}^T \phi(\boldsymbol{x}) - A(\boldsymbol{\theta})\right)$$

olabilirlik fonksiyonu için,

$$p(\boldsymbol{\theta}|\gamma) = h_c(\boldsymbol{\theta}) \exp\left(\left[\boldsymbol{\gamma}_1^T \boldsymbol{\theta} - \gamma_2 A(\boldsymbol{\theta})\right] - A_c(\gamma)\boldsymbol{\theta}\right)$$

öncül dağılımı eşleniktir.

Öncül dağılım için:

- Doğal parametre $\gamma = \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \gamma_2 \end{bmatrix}$
- Yeterli istatistik: $\begin{bmatrix} \boldsymbol{\theta} \\ -A(\boldsymbol{\theta}) \end{bmatrix}$.

Bernoulli dağılımı için eşlenik öncül

Değişken dönüşümü

Diyelim ki X'in dağılımını biliyoruz.
$Y = U(X)$.
Y'nin dağılımı nedir?

Değişken dönüşümü - ayrık dağılımlar

$$Y = U(X)$$

U tersi alınabilir bir fonksiyon ise

$$\begin{aligned} P(Y = y) &= P(U(X) = y) \\ &= P(X = U^{-1}(y)). \end{aligned}$$

Değişken dönüşümü - sürekli dağılımlar

$$Y = U(X)$$

Kümülatif dağılım fonksiyonu tekniği:

$$F_Y(y) = P(Y \leq y)$$

Örnek

X'in dağılımı

$$f_X(x) = 3x^2, \quad x \in [0, 1].$$

$Y = X^2$'nin dağılımı nedir?

Değişken dönüşümü - genel

U tersi alınabilir ve artan ise

$$\begin{aligned} F_Y(y) &= P(U(X) \leq y) \\ &\leq P(X \leq U^{-1}(y)) \\ &= \int_{-\infty}^{U^{-1}(y)} f(x)dx \end{aligned}$$

İfadenin y'ye göre türevi Y'nin olasılık dağılım fonksiyonunu verir.

U tersi alınabilir ve azalan olsaydı da aynı ifadeyi elde edecektik.

Değişken dönüşümü - çok boyutlu

$X \in \mathbb{R}^D$

$Y = U(X).$

$\boldsymbol{y} = U(\boldsymbol{x})$ tersi alınabilir ve türevlenebilir ise

$$f_Y(\boldsymbol{y}) = f_X(U^{-1}(\boldsymbol{y})) \left| \det \left(\frac{\partial}{\partial \boldsymbol{y}} U^{-1}(\boldsymbol{y}) \right) \right| .$$