

Beta Distribution

Beta Function:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

$$= \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx$$

$$= \int_0^1 \int_0^\infty u^{\alpha-1} v^{\beta-1} e^{-(u+v)} du dv$$

$$\text{Let } u = zx, v = z(1-x)$$

$$\Rightarrow \frac{\alpha\beta}{\alpha+\beta} \binom{\alpha+\beta}{\alpha} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

(exercise) $\checkmark$ For $\alpha, \beta > 0$, the function $f(x)$ is a pdf (exercise).

binomial pmf: $\binom{n}{x} p^x (1-p)^{n-x}$ $\leftrightarrow$ cf. $f(x)$

binomial coefficient: $\frac{1}{\binom{n}{x}} = \frac{x!(n-x)!}{n!}$

if α, β : integers $\rightarrow \frac{(\alpha-1)!(\beta-1)!}{(\alpha+\beta-1)!}$

check LNp.6-22~23

possible values of X

The distribution of a random variable X with this pdf is called the beta distribution with parameters α and β .

The cdf of beta distribution can be expressed in terms of the incomplete beta function, i.e., $F(x)=0$ for $x<0$, $F(x)=1$ for $x>1$, and for $0 \leq x \leq 1$,

$$F(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^x y^{\alpha-1} (1-y)^{\beta-1} dy \equiv \frac{B(x; \alpha, \beta)}{B(\alpha, \beta)}$$

incomplete beta function

$$= \sum_{i=\alpha}^{\alpha+\beta-1} \frac{(\alpha+\beta-1)!}{i!(\alpha+\beta-1-i)!} x^i (1-x)^{\alpha+\beta-1-i}$$

(exercise) Integration by parts for integer values of α and β

$$= \sum_{i=\alpha}^{\alpha+\beta-1} \binom{\alpha+\beta-1}{i} x^i (1-x)^{(\alpha+\beta-1)-i}$$

pmf of binomial $(\alpha+\beta-1, x)$

binomial & negative binomial in LNp.6-24

α, β : integers $\rightarrow$ Beta (α, β) $\rightarrow P(X \leq x) = P(Y \geq \alpha)$ $\rightarrow$ binomial $(\alpha+\beta-1, x)$

Theorem. The mean and variance of a beta distribution with parameters α and β are

$$\mu = \frac{\alpha}{\alpha+\beta} \quad \text{and} \quad \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Interpretation

$$E(X^2) - [E(X)]^2 = \frac{\alpha}{\alpha+\beta} \times \frac{\beta}{\alpha+\beta} \times \frac{1}{\alpha+\beta+1}$$

cf.

Proof.

$$E(X) = \int_0^1 x \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} dx$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+1)\Gamma(\beta)} \int_0^1 \frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+1)\Gamma(\beta)} x^{(\alpha+1)-1} (1-x)^{\beta-1} dx$$

$$= \frac{\alpha}{\alpha+\beta}$$

pdf of Beta $(\alpha+1, \beta)$

$$\text{Var}\left(\frac{X}{n}\right) = \frac{p(1-p)}{n}$$

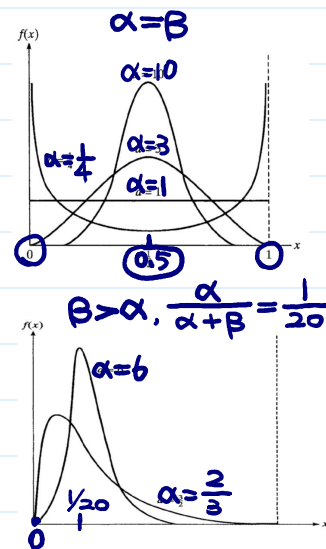
$$X \sim \text{binomial}(n, p)$$

$$\begin{aligned}
 E(X^2) &= \int_0^1 x^2 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} dx \\
 &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+2)\Gamma(\beta)}{\Gamma(\alpha+2)\Gamma(\beta+2)} \int_0^1 \frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+2)\Gamma(\beta)} x^{(\alpha+2)-1} (1-x)^{\beta-1} dx \\
 &= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} \cdot \text{pdf of Beta}(\alpha+2, \beta)
 \end{aligned}$$

Some properties

- When $\alpha=\beta=1$, the beta distribution is the same as the uniform(0, 1).

- Whenever $\alpha=\beta$, the beta distribution is symmetric about $x=0.5$, i.e.,



When $\alpha=\beta$

$$E(X) = \frac{\alpha}{\alpha+\beta} = \frac{1}{2}$$

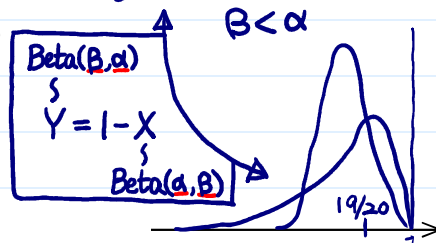
$$Var(X) = \frac{1}{4(2\alpha+1)}$$

$Var(X) \downarrow \alpha \uparrow$

$$f(0.5-\Delta) = f(0.5+\Delta).$$

$$x^{\alpha-1}(1-x)^{\alpha-1} \Rightarrow \begin{matrix} (0.5-\Delta)^{\alpha-1}(0.5+\Delta)^{\alpha-1} \\ (0.5+\Delta)^{\alpha-1}(0.5-\Delta)^{\alpha-1} \end{matrix}$$

- As the common value of α and β increases, the distribution becomes more peaked at $x=0.5$ and there is less probability outside of the central portion.



- When $\beta > \alpha$, values close to 0 become more likely than those close to 1; when $\beta < \alpha$, values close to 1 are more likely than those close to 0.

skewed

(Q: How to connect it with $E(X)$?) $\frac{\alpha}{\alpha+\beta}$

$\therefore$ pdf larger

Summary for $X \sim \text{Beta}(\alpha, \beta)$

- Pdf: $f(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, & \text{if } 0 \leq x \leq 1, \\ 0, & \text{otherwise,} \end{cases}$

- Cdf: $F(x) = B(x; \alpha, \beta) / B(\alpha, \beta)$.

- Parameters: $\alpha, \beta > 0$.

- Mean: $E(X) = \alpha / (\alpha + \beta)$.

- Variance: $Var(X) = (\alpha\beta) / [(\alpha + \beta)^2(\alpha + \beta + 1)]$.

$$E\left(\frac{X-\mu}{\sigma}\right) = 0, \quad Var\left(\frac{X-\mu}{\sigma}\right) = 1$$

standardization (標準化)

高斯

Normal (Gaussian) Distribution

常態

- For $\mu \in \mathbb{R}$ and $\sigma > 0$, the function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty,$$

x deviates from μ by $y \cdot \sigma$

$$y = \frac{x-\mu}{\sigma} \Rightarrow x = \sigma y + \mu$$

$$\frac{dx}{dy} = \sigma \Rightarrow dx = \sigma dy$$

functions as a unit how many σ (*)

possible values of X

is a pdf since (1) $f(x) \geq 0$ for all $x \in \mathbb{R}$, and (2)

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \equiv \frac{I}{\sqrt{2\pi}},$$

and

$$I^2 = \left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \right) \left(\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \right)$$

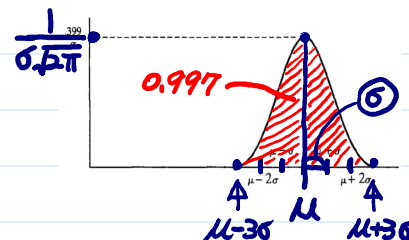
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dx dy = \int_0^{\infty} \int_0^{2\pi} e^{-\frac{r^2}{2}} r d\theta dr$$

$$= 2\pi \int_0^{\infty} r e^{-\frac{r^2}{2}} dr = -2\pi e^{-\frac{r^2}{2}} \Big|_0^{\infty} = 2\pi.$$

$$\begin{aligned} x &= r \cdot \cos\theta \\ y &= r \cdot \sin\theta \end{aligned} \Rightarrow dx dy = r dr d\theta$$

$$\begin{vmatrix} dx/dr & dx/d\theta \\ dy/dr & dy/d\theta \end{vmatrix} = r$$

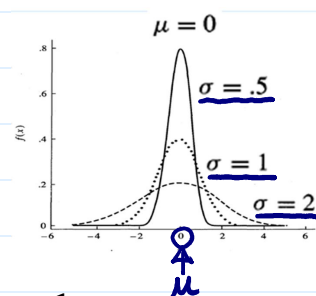
- The distribution of a random variable X with this pdf is called the normal (Gaussian) distribution with parameters μ and σ , denoted by $N(\mu, \sigma^2)$.



- The normal pdf is a bell-shaped curve.

- It is symmetric about the point μ , i.e.,

$$f(\mu+\Delta) = f(\mu-\Delta)$$

and falls off in the rate determined by σ .

$$\begin{aligned} \text{solve } \frac{df}{dx} &= 0 \Rightarrow x = \mu \text{ or} \\ \frac{d}{dx} \ln(f) &= 0 \Rightarrow x = \mu \end{aligned}$$

□ The pdf has a maximum at μ (can be shown by differentiation) and the maximum height is $1/(\sigma\sqrt{2\pi})$.