2022 SUMMER

업데이터 통계학 스터디

Chapter 5 – Continuous Random Variables 3

Beta Distribution

Beta function

If $\alpha > 0$ and $\beta > 0$,

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

Which is called the beta function.

*There's a high probability that it won't reach you right now.

If you study prior/posterior r.v. in Bayes Statistics in the future, the meaning of beta distribution may will be realized.

Beta Distribution

Definition

The continuous random variable X follows a Beta distribution if its PDF is:

$$f_X(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}\, I_{(0,1)}(x)$$

We denote $X \sim Beta(\alpha, \beta)$

Example 5.1

Suppose that continuous random variable $X \sim Beta(\alpha, \beta)$. Find

1. $E(X^k) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \underbrace{x^{k+\alpha-1}(1-x)^{\beta-1}}_{Beta(\alpha+k,\beta)} dx = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cancel{\Gamma(\beta)}} \cdot \frac{\Gamma(\alpha+k)\cancel{\Gamma(\beta)}}{\Gamma(\alpha+\beta+k)} = \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+k)}{\Gamma(\alpha)\Gamma(\alpha+\beta+k)}$

2. $E(X) = \frac{\Gamma(\alpha+1)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha+\beta+1)} = \frac{\alpha!\,(\alpha+\beta-1)!}{(\alpha-1)!\,(\alpha+\beta)!} = \frac{\alpha}{\alpha+\beta}$

3. $Var(X) = \underbrace{E(X^2)}_{\frac{\Gamma(\alpha+2)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha+\beta+2)} = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}} - \left(\frac{\alpha}{\alpha+\beta}\right)^2 = \frac{\alpha(\alpha+1)(\alpha+\beta) - \alpha^2(\alpha+\beta+1)}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Normal Distribution

Gaussian Integral

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}, \quad \int_{-\infty}^{\infty} e^{-ax^2} dx = \sqrt{\frac{\pi}{a}}$$

Proof)

$$\int_{\mathbb{R}^2} e^{-(x^2+y^2)}\,dx\,dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)}\,dx\,dy$$

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

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

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)}\,dx\,dy$$

$$= \int_{0}^{2\pi}\int_{0}^{\infty} re^{-r^2}\,dr\,d\theta$$

$$= \int_{0}^{2\pi} \frac{1}{2}\,d\theta$$

$$= \pi$$

$$\therefore \int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$$

Normal Distribution

Definition

The continuous random variable X follows a Normal distribution if its PDF is:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad x \in \mathbb{R}$$

We denote $X \sim N(\mu, \sigma^2)$

NOTE: $E(X) = \mu, Var(X) = \sigma^2$

Normal Distribution

1. All normal curves are bell-shaped with points of inflection at $\mu \pm \sigma$
2. All normal curves are symmetric about the mean μ
3. The area under an entire normal curve is 1
4. All normal curves are positive for all x. That is, $f_X(x) > 0$ for all x.
5. Convergence; $\lim_{x \to \infty} f_X(x) = 0$, and $\lim\limits_{x \to -\infty} f_X(x) = 0$
6. The height of any normal curve is maximized at $x = \mu$

Standard Normal Distribution

Definition

If $X \sim N(\mu, \sigma^2)$, then

$$Z = \frac{X - \mu}{\sigma} \sim N(0,1)$$

Which is called the Standard Normal Distribution.

Standardization makes it easy to obtain probabilities using a normal distribution table.

Example 5.2

Suppose that continuous random variable $X \sim N(\mu, \sigma^2)$. Find

1. $M_X(t)$ $= E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} \cdot \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx$

$$\left(\begin{array}{l} z = \frac{x-\mu}{\sigma} \\ \sigma dz = dx \end{array} \right.$$

2. $E(X)$ $= \mu$

$$\rightsquigarrow \int_{-\infty}^{\infty} e^{t(\mu+\sigma z)} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\,dz$$

3. $Var(X)$ $= \sigma^2$

$$= e^{\mu t} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z^2 - 2\sigma t z + \sigma^2 t^2)} \cdot e^{\frac{1}{2}(\sigma t)^2}\,dz$$

$$= e^{\mu t + \frac{1}{2}\sigma^2 t^2} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-\sigma t)^2}{2}}\,dz}_{\text{pdf of } N(\sigma t, 1)}$$

$$\therefore M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

CDF Method

CDF Method

Let X be a c.r.v.. If we consider $Y = u(X)$ then Y must also be a r.v. that has its own distribution.

We can find the PDF of the random function $Y = u(X)$ by:

1. Find the CDF of Y

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

2. Differentiate the CDF of Y

$$f_Y(y) = F'_Y(y)$$

Example 5.3

Let $X \sim N(0,1)$ What is the distribution of $Y = X^2$?

Sol.

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

$$= P(X^2 \le y)$$

$$= \begin{cases} P(-\sqrt{y} \le X \le \sqrt{y}) & (y \ge 0) \\ 0 & (y < 0) \end{cases}$$

for $y \ge 0$,

$$F_Y(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

$$f_Y(y) = f_X(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} + f_X(-\sqrt{y}) \frac{1}{2\sqrt{y}}$$

$$= \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{2\sqrt{y}} \cdot 2 \cdot e^{-\frac{y}{2}}$$

$$= \underbrace{\frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} \cdot e^{-\frac{y}{2}}}_{Gamma(\frac{1}{2},2)}$$

$\therefore Y \sim Gamma(\frac{1}{2}, 2)$

$\sim \chi^2(1)$

Example 5.4

$X_i \sim Unif(0, \theta)$ independently. Let $Y = max(X_1, \cdots, X_n)$.

What is the distribution of $\frac{Y}{\theta}$?

수정 : $Y|\theta$

Sol.

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

$$= P(\max(x_i) \le y)$$

$$= P(\text{all } x_i\text{'s} \le y)$$

$$= \{P(X \le y)\}^n = \{F_X(y)\}^n$$

$$\therefore F_Y(y) = \begin{cases} 0 & (y<0) \\ (y/\theta)^n & (0 \le y \le \theta) \\ 1 & (y > \theta) \end{cases}$$

$$f_Y(y) = \frac{ny^{n-1}}{\theta^n} I_{(0,\theta)}(y)$$

if $\theta = 1$: $f_Y(y) = \underbrace{ny^{n-1}}_{\text{Beta}(n,1)}$ $(0<y<1)$

Change of Variable

Change of Variable

[Continuous, Uni-variate, 1:1 function]

$F_U(u) = F_Y(h^{-1}(u))$

$f_U(u) = f_Y(h^{-1}(u)) \frac{d}{du} h^{-1}(u)$ 양수

Suppose $Y \sim f_Y(y)$, $U = h(Y)$ where $h(y)$ increasing in y. Then,

$$P(U \leq u) = P(h(Y) \leq u) = P(Y \leq h^{-1}(u))$$

Suppose $Y \sim f_Y(y)$, $U = h(Y)$ where $h(y)$ decreasing in y. Then,

수정

$$P(U \leq u) = P(h(Y) \leq u) = P(Y \geq h^{-1}(u))$$

부호 상관 X

$$\therefore f_U(u) = f_Y\left(h^{-1}(u)\right) \left| \frac{dh^{-1}(u)}{du} \right|$$

$F_U(u) = 1 - F_Y(h^{-1}(u))$

$f_U(u) = - f_Y(h^{-1}(u)) \frac{d}{du} h^{-1}(u)$

음수

Example 5.5

Let Y be a c.r.v. with the following PDF:

$$f_Y(y) = 2y\, I_{(0,1)}(y)$$

What is the PDF of $U = 8y^3$?

$u = 8y^3.$

$y = \frac{1}{2}u^{\frac{1}{3}} = h^{-1}(u)$

$f_U(u) = f_Y(\frac{1}{2}u^{\frac{1}{3}}) \cdot |\frac{1}{6}u^{-\frac{2}{3}}|$

$= \begin{cases} \frac{1}{6}u^{-\frac{1}{3}} & (0<u<8) \\ 0 & (\text{o.w}) \end{cases}$

$0<y<1$

$0<\frac{1}{2}u^{\frac{1}{3}}<1$

$\vdots$

$0<u<8$

Example 5.6

Let Y be a c.r.v. with the following PDF:

$$f_Y(y) = 4y^3\, I_{(0,1)}(y)$$

What is the PDF of $U = e^Y$?

$u = e^y$

$y = \ln u = h^{-1}(u)$

$f_U(u) = f_Y(\ln u) \left| \frac{1}{u} \right|$

$= 4(\ln u)^3 \cdot \frac{1}{u}\, I_{(1,e)}(u)$

$0 < y < 1$

$0 < \ln u < 1$

$\underline{1 < u < e}$

MGF Method

MGF Method

As we saw earlier, MGF uniquely determines the distribution.

Let X be a c.r.v.. If we consider $Y = u(X)$ then Y must also be a r.v. that has its own distribution.

We can find the distribution of the random function $Y = u(X)$ by proving that the MGF of Y matches the MGF of a particular distribution.

Example 5.7

Find a distribution;

X ≔ Sum of i.i.d. $Exponential(\theta)$

$$X := \sum_{i=1}^{n} Y_i \,, \qquad Y_i \overset{iid}{\sim} Exp(\theta)$$

$$\begin{aligned} M_X(t) &= E(e^{tX}) \\ &= E(e^{t\Sigma Y_i}) \\ &= E(e^{tY_1} \cdot e^{tY_2} \cdots e^{tY_n}) \\ &= \{E(e^{tY})\}^n \quad (\because iid) \end{aligned}$$

$$\begin{aligned} &= \{M_Y(t)\}^n \\ &= (1-\theta t)^{-n} \; ; \text{ MGF of Gamma}(n, \theta) \end{aligned}$$

$\therefore$ By MGF method, $X \sim Gamma(n, \theta)$

Example 5.8

Find a distribution;

Y ≔ Sum of independent $X_i \sim Bin(n_i, p)$

$$Y = \sum_{i=1}^{n} X_i$$

$$M_Y(t) = E(e^{tY})$$

$$= E(e^{tX_1} \cdot e^{tX_2} \cdots e^{tX_n})$$

$$= E(e^{tX_1}) E(e^{tX_2}) \cdots E(e^{tX_n}) \quad (\because \text{independent})$$

$$= M_{X_1}(t) M_{X_2}(t) \cdots M_{X_n}(t)$$

$$= (pe^t + 1 - p)^{n_1} (pe^t + 1 - p)^{n_2} \cdots$$

$$= (pe^t + 1 - p)^{\Sigma n_i}$$

$$\therefore Y \sim Bin(\Sigma n_i, p)$$

Chi-Square Distribution (Recall)

Chi-Square Distribution

$P(U \le u) = P(Z^2 \le u)$

$= P(-\sqrt{u} \le Z \le \sqrt{u})$

$= F_Z(\sqrt{u}) - F_Z(-\sqrt{u})$

$f_U(u) = \frac{1}{2\sqrt{u}}\left(f_Z(\sqrt{u}) + f_Z(-\sqrt{u})\right)$

$= \frac{1}{2\sqrt{u}} \cdot \frac{2}{\sqrt{2\pi}} e^{-\frac{u}{2}} \sim \text{Gamma}(\frac{1}{2}, 2)$

$\sim \chi^2(1)$

<u>Definition</u>

If $X \sim N(\mu, \sigma^2)$, then:

$$U = \left(\frac{X-\mu}{\sigma}\right)^2 = Z^2$$

Is distributed as a Chi-Square r.v. with 1 $d.o.f.$ $\Leftrightarrow U \sim \chi^2(1)$

Generalization; $U = {Z_1}^2 + \cdots + {Z_n}^2 \sim \chi^2(n)$ (where Z_i are indep)

Furthermore; $U_1, \cdots, U_n$: independent $\chi^2(r_i) \Rightarrow \sum U_i \sim \chi^2(\sum r_i)$

Chi-Square Distribution

Proof) $U = Z_1^2 + \cdots + Z_n^2$

$$M_U(t) = E(e^{tZ_1^2}) \cdots E(e^{tZ_n^2})$$

$$= M_{Z_1^2}(t) \cdots M_{Z_n^2}(t)$$

$$= (1-2t)^{-\frac{n}{2}} \sim \text{Gamma}\left(\frac{n}{2}, 2\right)$$

$$\sim \chi^2(n)$$

$X := \Sigma U_i$

$$M_X(t) = E(e^{tX})$$

$$= E(e^{tU_1})E(e^{tU_2}) \cdots E(e^{tU_n})$$

$$= (1-2t)^{\frac{\Sigma r_i}{2}}$$

$$\sim \chi^2(\Sigma r_i)$$

t Distribution

t Distribution

Definition

$Z \sim N(0,1)$ and $W \sim \chi^2(v)$, independently. Then

$$T := \frac{Z}{\sqrt{W/v}} \sim t(v)$$

$$f_T(t) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{\pi v}(\frac{v}{2})}(1 + \frac{t^2}{v})^{-\frac{v+1}{2}} I_{\mathbb{R}}(t)$$

t Distribution

$E(T) = 0, Var(T) = \frac{v}{v-2} \; for \; v > 2$

If v$(d.o.f.) \; increase < \infty \;\; \Rightarrow \;\; T \longrightarrow N(0,1)$

F Distribution

F Distribution

Definition

$W_1 \sim \chi^2(v_1)$ and $W_2 \sim \chi^2(v_2)$, independently. Then

$$\mathrm{F} \coloneqq \frac{W_1/V_1}{W_2/V_2} \sim F(v_1, v_2)$$

$$PDF?\ You\ don't\ have\ to\ know.$$

$$f(y)=\frac{\Gamma((\nu_1+\nu_2)/2)(\nu_1/\nu_2)^{\nu_1/2}}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)}y^{(\nu_1/2)-1}\left(1+\frac{\nu_1 y}{\nu_2}\right)^{-(\nu_1+\nu_2)/2}, \quad 0<y<\infty$$

F Distribution

$$E(F) = \frac{v_2}{v_2 - 2} \text{ for } v_2 > 2$$

$$Var(F) = \frac{2{v_2}^2(v_1 + v_2 - 2)}{v_1(v_2 - 2)^2(v_2 - 4)} \; for \;\; v_2 > 4$$

- $F \sim F(v_1, v_2) \Longrightarrow \frac{1}{F} \sim F(v_2, v_1)$

$F := \frac{W_1/v_1}{W_2/v_2} \rightarrow \frac{1}{F} := \frac{W_2/v_2}{W_1/v_1}$ 이죠!