2

Random Variable

$\Rightarrow$ A random variable is a mapping

$$X : \Omega \to \mathbb{R}$$

that assigns a real number $X(\omega)$ to each outcome ω.

$\Rightarrow$ Most of the probability course, the sample space is rarely mentioned, we work directly with random variables.

$\Rightarrow$ Given a random variable X and a subset A of a real line, define $X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\}$

$$\mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\})$$

$$\mathbb{P}(X = x) = \mathbb{P}(X^{-1}(x)) = \mathbb{P}(\{\omega \in \Omega ; X(\omega) = x\})$$

$X \Rightarrow$ denotes random variable

$x \Rightarrow$ denotes a particular value of X.

★ Distribution Function and Probability Functions

⇒ The cumulative distribution function or CDF is the function $F_X : \mathbb{R} \to [0,1]$ defined by

$$F_X(x) = \mathbb{P}(X \leq x)$$

⇒ <u>Theorem:</u> Let X have CDF F and let Y have CDF G. If $F(x) = G(x)$ $\forall x$, then $\mathbb{P}(X \in A) = \mathbb{P}(Y \in A)$ $\forall A$.

⇒ X is discrete if it takes countably many values $\{x_1, x_2 \dots\}$. We define the probability function or probability mass function for X by $f_X(x) = \mathbb{P}(X = x)$.

⇒ A random variable X is continuous if there exist a function f_X such that $f_X(x) \geq 0$ $\forall x$, $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$ and for every $a \leq b$

$$\mathbb{P}(a < X < b) = \int_a^b f_X(x)\,dx$$

The function f_X is called Probability density function (PDF).

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$$

and $f_X(x) = \frac{d}{dx}(F_X(x))$ at all point at which F_X is differentiable.

⇒ Let X be a random variable with CDF F. The inverse CDF or quantile function is defined by

$$F^{-1}(q) = \inf\{x : F(x) > q\}$$

Infimum $\quad \forall\, q \in [0,1]$

$$X \overset{d}{=} Y \text{ — if } F_X(x) = F_Y(x) \quad \forall x$$

This does not mean that X and Y are equal, Rather, it means that all probability statement about X and Y will be the same.

★ Some Important discrete random Variables

$X \sim F \Rightarrow X$ has distribution F

Point Mass Distribution

$$f(x) = 1 \quad \forall\, x = a$$
$$= 0 \quad \forall\, x \neq a$$

Discrete Uniform Distribution

$$f(x) = \begin{cases} 1/k & \forall\, x = 1, \dots k \\ 0 & \text{Otherwise} \end{cases}$$

Bernoulli distribution

- Let X be binary
- If $P(X=1) = p$ & $P(X=0) = 1-p$ $\forall$ some $p \in [0,1]$. We say X has a bernoulli distribution.
- $X \sim \text{Bernoulli}(p)$
- $f(x) = p^x (1-p)^{1-x} \quad \forall x \in \{0,1\}$

Binomial distribution

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

- $$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \forall x = 0, \dots n \\ 0 & \text{otherwise} \end{cases}$$

Geometric Distribution

$X \sim \text{Geom}(p)$

- $f(x) = p(1-p)^{x-1} \quad \forall K \geq 1$

Poisson distribution

$X \sim \text{Poisson}(\lambda)$

- $f(x) = e^{-\lambda} \dfrac{\lambda^x}{x!} \quad x \geq 0$

★ Some important Continuous Random Variables

Uniform distribution $X \sim \text{Uniform}(a, b)$

$$f(x) = \begin{cases} \frac{1}{b-a} & \forall\, x \in [a,b] \\ 0 & \text{otherwise.} \end{cases}$$

Normal (Gaussian) $X \sim N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

$x \in \mathbb{R}$

$\mu \in \mathbb{R}$ & $\sigma > 0$

μ → mean

σ → standard deviation

⇒ We say that X has a standard Normal distribution if $\mu = 0$ & $\sigma = 1$.

- Standard normal random variable is denoted by Z
- PDF and CDF of a standard Normal are denoted by $\varphi(z)$ and $\Phi(z)$

⇒ Important facts about normal distribution:

→ If $X \sim N(\mu, \sigma^2)$ then,
$$Z = (X-\mu)/\sigma \sim N(0,1)$$

→ If $Z \sim N(0,1)$, then
$$X = \mu + \sigma Z \sim N(\mu, \sigma^2)$$

→ If $X_i \sim N(\mu_i, \sigma_i^2)$, $i=1, \dots n$ are independent then
$$\sum_{i=1}^{n} X_i \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

Exponential distribution $X \sim Exp(\beta)$

→ $f(x) = \frac{1}{\beta} e^{-x/\beta}$ $x > 0$
(where $\beta > 0$)

→ used to model the lifetime of electronic components and the waiting times between rare events.

Gamma distribution $X \sim Gamma(\alpha, \beta)$

~~$X \sim Gamma(\alpha)$~~ ⇒ $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$.
→ Gamma function

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \quad x > 0$$

where $\alpha, \beta > 0$

$\Rightarrow$ The exponential distribution is just a Gamma$(1, \beta)$ distribution.

$\Rightarrow$ If $X_i \sim$ Gamma(α_i, β) are independent then $\sum_{i=1}^{n} X_i \sim$ Gamma$\left(\sum_{i=1}^{n} \alpha_i, \beta\right)$

Beta distribution $\quad X \sim$ Beta(α, β)

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

$0 < x < 1$

t and Cauchy distribution $\quad X \sim t_v$

$\rightarrow$ $$f(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma(v/2)} \frac{1}{\left(1+\frac{x^2}{v}\right)^{\left(\frac{v+1}{2}\right)}}$$

$\rightarrow$ The t distribution is similar to a Normal but it has thicker tails.

$\rightarrow$ Normal corresponds to t with $V = \infty$.

$\rightarrow$ Cauchy distribution is a special case of t corresponding to $V = 1$.

$$f(x) = \frac{1}{\pi(1+x^2)}$$

χ^2 Distribution $\quad X \sim \chi^2_p$

→ X has a χ^2 distribution with p degree of freedom if

$$f(x) = \frac{1}{\Gamma(p/2)\,2^{p/2}} x^{(p/2)-1} e^{-x/2}, \quad x>0$$

★ Bivariate Distributions

⇒ Given a pair of discrete random variable X and Y, define the joint mass function by $f(x,y) = \mathbb{P}(X=x \ \& \ Y=y)$.

$$\Updownarrow$$

$$\mathbb{P}(X=x, Y=y)$$

→ We write f as f_{XY} when we want to be more explicit.

⇒ In the continuous case, we call a function $f(x,y)$ a PDF for the random variable (X,Y) if

(i) $f(x,y) \geq 0 \quad \forall (x,y)$

(ii) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$

(iii) $\forall$ set $A \subset \mathbb{R}\times\mathbb{R}$, $\mathbb{P}((X,Y)\in A)$
$$= \iint_A f(x,y)\,dx\,dy$$

⇒ In the discrete or continuous case we define the joint CDF as $F_{XY}(x,y)$
$= \mathbb{P}(X \le x, Y \le y)$

★ Marginal Distribution

⇒ If (X,Y) have joint distribution with mass function f_{XY}, then the marginal mass function for X is defined by

$$\boxed{f_X(x) = \mathbb{P}(X=x) = \sum_y \mathbb{P}(X=x, Y=y) = \sum_y f(x,y)}$$

⇒ For continuous random variables, the marginal densities are

$$\boxed{f_X(x) = \int f(x,y)\,dy}$$

★ Independent Random Variable

⇒ Two random variables X and Y are independent if for every A and B,

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\,\mathbb{P}(Y \in B)$$

and we write $X \amalg Y$.

↳ Otherwise we say that X and Y are dependent and we write $X \not\amalg Y$.

Theorem

Theorem

★ C

⇒

⇒

Theorem: Let X and Y have joint PDF f_{XY}. Then $X \perp\!\!\!\perp Y$ if and only if $f_{XY}(x,y) = f_X(x) f_Y(y)$ $\forall\, x \,\&\, y$.

Theorem: Suppose that the range of X and Y is a (Possibly infinite) rectangle. If $f(x,y) = g(x)\,h(y)$ for some functions g and h (not necessarily probability density function) then X and Y are independent.

★ **Conditional Distribution**

⇒ The conditional probability mass function is

$$f_{X|Y}(x|y) = \mathbb{P}(X=x \mid Y=y) = \frac{\mathbb{P}(X=x, Y=y)}{\mathbb{P}(Y=y)}$$

$$= \frac{f_{XY}(x,y)}{f_Y(y)} \quad \text{if } f_Y(y) > 0$$

⇒ For continuous random variables, the conditional probability density function is

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

assuming that $f_Y(y) > 0$. Then,

$$\mathbb{P}(X \in A \mid Y=y) = \int_A f_{X|Y}(x|y)\,dx$$

★ Multivariate Distributions and IID Samples

⇒ Let $X = (X_1, \dots X_n)$ where $X_1 \dots X_n$ are random variables.

- Let $f(x_1, x_2, \dots x_n)$ denote the PDF.
- We call X a random vector.
- We say that $X_1, \dots X_n$ are independent if, for every $A_1, \dots A_n$,

$$\mathbb{P}(X_1 \in A_1, \dots X_n \in A_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i)$$

- It suffices to check that

$$f(x_1, x_2, \dots x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

⇒ If $X_1, \dots X_n$ are independent and each has the same marginal distribution with CDF F, we say that $X_1, \dots X_n$ are IID (independent and identically distributed) and we write

$$X_1, \dots X_n \sim F$$

- If F has density f we also write $X_1, \dots X_n \sim f$. We also call $X_1, \dots X_n$ a random sample of size n from F.

* Two important Multivariate Distribution

Multinomial

→ The multivariate version of a Binomial is called a Multinomial.

→ $$\boxed{f(x) = \binom{n}{x_1 \cdots x_k} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}}$$

⇒ Suppose that $X \sim \text{Multinomial}(n, P)$ where $X = (X_1 \cdots X_k)$ and $P = (P_1 \cdots P_k)$ The marginal distribution of X_j is Binomial (n, P_j)

Multivariate Normal

→ Here μ is a vector and σ is replaced by a matrix Σ.

→ Let $Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_k \end{pmatrix}$

→ $Z_1 \cdots Z_k \sim N(0,1)$ are independent

→ The density of Z is

$$f(z) = \prod_{i=1}^{k} f(z_i) = \frac{1}{(2\pi)^{k/2}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{k} z_j^2\right\}$$

$$= \frac{1}{(2\pi)^{k/2}} \exp\left\{-\frac{1}{2} z^T z\right\}$$

⇒ We say that Z has a standard multivariate Normal distribution

↳ $Z \sim N(\underbrace{0}_{\text{Vector of zeros}}, \underbrace{I}_{K \times K \text{ Identity matrix}})$

⇒ More generally, a vector X has a multivariate Normal distribution, denoted by $X \sim N(\mu, \Sigma)$, if it has density

$$f(x; \mu, \Sigma) = \frac{1}{(2\pi)^{K/2} |(\Sigma)|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right\}$$

where, $|\Sigma|$ denotes the determinant of Σ

→ μ is a vector of length K

→ Σ is $K \times K$ symmetric, positive definite matrix.

⇒ It can be shown that there exists a matrix $\Sigma^{1/2}$ (called the square root of Σ) with the following properties:

- ↳ $\Sigma^{1/2}$ is symmetric
- ↳ $\Sigma = \Sigma^{1/2} \Sigma^{1/2}$
- ↳ $\Sigma^{1/2} \Sigma^{-1/2} = I$ when $\Sigma^{-1/2} = (\Sigma^{1/2})^{-1}$

Theorem: If $Z \sim N(0, I)$ and $X = \mu = \Sigma^{1/2} Z$

then $X \sim N(\mu, \Sigma)$.

↳ Conversely, if $X \sim N(\mu, \Sigma)$ then

$\Sigma^{-1/2}(X - \mu) \sim N(0, I)$.

⇒ Suppose we partition a random Normal vector X as $X = (X_a, X_b)$

↳ we can similarly partition

$$\mu = (\mu_a, \mu_b)$$

$$\Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$$

Theorem: Let $X \sim N(\mu, \Sigma)$. Then:

(1) The marginal distribution of X_a is $X_a \sim N(\mu_a, \Sigma_{aa})$

(2) The conditional distribution of X_b given $X_a = x_a$ is

$$X_b | X_a = x_a \sim N\left(\mu_b + \Sigma_{ba}\Sigma_{aa}^{-1}(x_a - \mu_a), \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab}\right)$$

(3) If a is a vector then $a^T X \sim N(a^T\mu, a^T\Sigma a)$

(4) $V = (X-\mu)^T \Sigma^{-1} (X - \mu) \sim \chi^2_K$

★ Transformations of Random Variable

⇒ Suppose that X is a random variable with PDF f_X and CDF F_X.

↳ Let $Y = g(X)$ be a function of X

↳ We call $Y = g(X)$ a transformation of X.

⇒ The mass function of Y is given by

$$f_Y(y) = \mathbb{P}(Y=y) = P(g(X)=y)$$
$$= P(\{x; g(x)=y\}) = \mathbb{P}(X \in g^{-1}(y))$$

For Continuous Case

1) For each y, find set $A_y = \{x : g(x) \le y\}$

2) Find the CDF

$$F_Y(y) = \mathbb{P}(Y \le y) = P(g(X) \le y)$$
$$= \mathbb{P}(\{x; g(x) \le y\})$$
$$= \int_{A_y} f_X(x)\, dx$$

3) The PDF is $f_Y(y) = F_Y'(y)$

→ When g is strictly monotonic increasing or strictly monotonic decreasing then g has an inverse $s = g^{-1}$.

↳ In this case one can show that
$$f_Y(y) = f_X(s(y)) \left| \frac{ds(y)}{dy} \right|$$

★ Transformations of several Random Variable

1) For each z, find the set $A_z = \{(x,y) : g(x,y) \leq z\}$

2) Find the CDF $\mathbb{P}(g(X,Y) \leq z)$
$$F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}(g(X,Y) \leq z)$$
$$= \mathbb{P}(\{(x,y) ; g(x,y) \leq z\})$$
$$= \iint_{A_z} f_{XY}(x,y)\, dx\, dy.$$

3) Then $f_Z(z) = F_Z'(z)$