15: Distributions and parameters

1 A continuous random variable

(1) Probability density function (pdf)
X is a random variable if it is a variable and its values are associated with probabilities.
The relation of its values and probabilities is called the distribution of X.
The distribution can be depicted by its pdf $f(x)$, a function with properties
(i) $f(x) \geq 0$ (ii) $P(X \in A) = \int_A f(x)dx$.
The domain of $f(x)$ is where X can assume its values and is called the support of X.
One can always extend the support to R with $f(x) = 0$ outside its original support.
$g(x)$ can be used as a pdf to define a distribution if (i) $g(x) \geq 0$ (ii) $\int_R g(x)dx = 1$.

(2) Parameters of X
$E[h(X)] = \int_R h(x)f(x)\,dx$, the expectation of $g(X)$, is the "average" value of $h(X)$.
$\mu = E(X) = \int_R xf(x)\,dx$ is the mean of X.
$E(X^2) = \int_R x^2 f(x)\,dx$ is the second moment of X.
$\sigma^2 = \text{var}(X) = E(X-\mu)^2 = E(X^2) - [E(X)]^2$ is the variance of X that gives the magnitude of the fluctuation of the value of X.
μ and σ^2 are two important parameters for X. $X \sim (\mu, \sigma^2)$ is often written to indicate the two parameters.

(3) pdf of $h(X)$

(i) $y = h(x)$ is a 1-1 function with inverse $x = h^{-1}(y)$. For $Y = h(X)$
$P(Y \in A) = P(X \in h^{-1}(A)) = \int_{h^{-1}(A)} f(x)\,dx \overset{y=h(x)}{=\!=\!=} \int_A \left.\frac{f(x)}{h'(x)}\right|_{x=h^{-1}(y)} dy$.
So $f_Y(y) = \left.\frac{f(x)}{h'(x)}\right|_{x=h^{-1}(y)}$ is pdf for Y.

(ii) $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx$ is the cumulative distribution function (cdf) for X. Clearly $f(x) = F'(x)$. So $f_Y(y) = F_Y'(y) = [P(Y \leq y)]_y'$.

Ex1: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] \geq 0$ and $\int_R f(x)dx = 1$. Let $f(x)$ be the pdf for X.
Then $E(X) = \int_R xf(x)dx = \mu$, $E(X^2) = \int_R x^2 f(x)dx = \sigma^2 + \mu^2$. So $E(X-\mu)^2) = \sigma^2$.
This distribution is called a normal distribution denoted by $X \sim N(\mu, \sigma^2)$.

Ex2: For $X \sim N(\mu, \sigma^2)$ let $Y = aX + b$ where $a \neq 0$. Then the pdf of Y is
$f_Y(y) = \left.\frac{f(x)}{a}\right|_{x=\frac{y-b}{a}} = \frac{1}{\sqrt{2\pi a^2\sigma^2}} \exp\left[-\frac{1}{2a^2\sigma^2}(y - a\mu - b)^2\right]$. Thus $Y \sim N(a\mu + b, a^2\sigma^2)$.

Ex3 For $Z \sim N(0, 1^2)$ let $Y = Z^2$. Then
$f_Y(y) = [P(Y \leq y)]_y' = [P(-\sqrt{y} \leq Z \leq \sqrt{y})]_y' = \left[2\int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\,dz\right]_y' = \frac{1}{\sqrt{2\pi y}}\, e^{-\frac{y}{2}}$.
So $Y \sim \text{Gamma}\left(\frac{1}{2}, 2\right) = \chi^2(1)$.

2. A continuous random vector

(1) Joint pdf and marginal pdf
Random vector $X = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$ has joint pdf $f(x)$ if $f(x) = f(x_1, .., x_p) \geq 0$ and

$P(X \in A) = \iint_A f(x)dx_1,..,dx_p$ for all $A \subset R^p$. Let X_I contains some of the components of X, and X_{II} contains the other components. For example $X_I = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$. Then $f_{X_I}(x_1,\, x_2) = \iint_{R^{p-2}} f(x)dx_3,..,dx_p \geq 0$ is the pdf for X_I, one of marginal pdfs of X.

$$\begin{aligned} P(X_I \in A) &= P(X_I \in A, X_{II} \in R^{p-2}) = \iint_A [\iint_{R^{p-2}} \; f(x)dx_3,...,dx_p]\, dx_1dx_2 \\ &= \iint_A f_{X_I}(x_1,x_2)\, dx_1dx_2. \end{aligned}$$

(2) Parameters

X has joint pdf $f(x)$. $E[h(X)] = \iint_{R^p} hx)f(x)\, dx_1,..,dx_p$, is the expectation of $h(X)$.

So $\quad E(X_i) = \iint_{R^p} x_i f(x)\, dx_1,..dx_p = \int_R x_i f_{X_i}(x_i)\, dx_i = \mu_i,$

$\quad E(X_i^2) = \iint_{R^p} x_i^2 f(x)\, dx_1,...,dx_p = \int_R x_i^2 f_{X_i}(x_i)\, dx_i$ and

$\quad \mathrm{var}(X_i) = E(X_i - \mu_i)^2 = E(X_i^2) - [E(X_i)]^2 = \sigma_i^2 = \sigma_{ii}$

can be calculated with either joint pdf of X or marginal pdf of X_i, $f_{X_i}(x_i)$.

$\quad \mathrm{cov}(X_i,\, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)] = E(X_iX_j) - E(X_i)\, E(X_j) = \sigma_{ij}$ and

$\quad \rho(X_i,\, X_j) = \frac{\mathrm{cov}(X_i,\, X_j)}{\sqrt{\mathrm{var}(X_i)\ \mathrm{var}(X_j)}} = \rho_{ij}$

can be calculated with either joint pdf of X or marginal pdf of $(X_i,\, X_j)$.

It is well known that $-1 \leq \rho_{ij} \leq 1$ and $\rho_{ii} = 1$.

(3) Parameter vectors and matrices

For random $X \in R^p$, $\begin{pmatrix} E(X_1) \\ \vdots \\ E(X_p) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix} \overset{den}{=\!=} \mu$ is the mean vector.

$(\mathrm{cov}(X_i,\, X_j))_{p\times p} = (\sigma_{ij})_{p\times p} \overset{den}{=\!=} \Sigma$ is the variance-covariance matrix.

$V = \mathrm{diag}(\sigma_1^2, ..., \sigma_p^2)$ is the variance matrix.

$\rho = (\rho_{ij})_{p\times p}$ is the correlation matrix.

From Σ one can have V and $\rho = V^{-1/2}\Sigma V^{-1/2}$. $X \sim (\mu,\, \Sigma)$ is often written to indicate $E(X) = \mu$ and $\mathrm{Cov}(X,\, X) = \Sigma$.

Ex4: $f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right] \geq 0$ where $x \in R^p$, $\mu \in R^p$, $\Sigma \in R^{p\times p}$ and $\Sigma > 0$. Then $f(x)$ is a pdf for a random vector $X \in R^p$. This distribution is denoted as $X \sim N(\mu,\, \Sigma)$.

$f(x) > 0$ and by substitution $z = \Sigma^{-1/2}(x-\mu)$

$$\iint_{R^p} f(x)dx_1,..dx_p = \prod_{i=1}^p \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z_i^2}{2}\right)\, dz_i = 1.$$

Ex5: For X in Ex4, $f_{X_i}(x_i) = \iint_{R^{p-1}} f(x)dx_1,...,dx_{i-1},dx_{i+1},..,dx_p = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}\right]$.

Thus $X_i \sim N(\mu_i,\, \sigma_i^2)$. So $E(X_i) = \mu_i$ and $\mathrm{var}(X_i) = \sigma_i^2$.

With $x_I = (x_i,\, x_j)'$, $f_{X_i,X_j}(x_I) = \frac{1}{2\pi|\Sigma_I|^{1/2}} \exp\left[-\frac{1}{2}(x_I - \mu_I)'\Sigma_I^{-1}(x_I - \mu_I)\right]$ where $\mu_I = (\mu_i,\, \mu_j)'$ and $\Sigma_I = \begin{pmatrix} \sigma_i^2 & \sigma_{ij} \\ \sigma_{ji} & \sigma_j^2 \end{pmatrix}$. Consequently $\mathrm{cov}(X_i,\, X_j) = \sigma_{ij}$.

Thus in $X \sim N(\mu,\, \Sigma)$, $\mu = E(X)$, $\Sigma = \mathrm{Cov}(X,\, X)$, and $N(\mu,\, \Sigma)$ is called a normal distribution.