MA204: Mathematical Statistics

Tutorial 3

T3.1 Mixture Technique (The Fifth Method)

- Statistically, we can represent $f_X(x)$ as the marginal pdf of the joint pdf $f_{(X,Y)}(x,y)$ in the form
$$f_X(x) = \int_{\mathcal{S}_Y} f_{(X,Y)}(x,y)\,\mathrm{d}y, \tag{T3.1}$$
where $\mathcal{S}_Y$ denotes the support of the r.v. Y, and the joint density $f_{(X,Y)}(x,y)$ exists.

- Alternatively, we can rewrite (T3.1) in the mixture form
$$f_X(x) = \int_{\mathcal{S}_Y} f_Y(y) f_{(X|Y)}(x|y)\,\mathrm{d}y \quad \text{or} \quad f_X(x) = \sum_{k\in\mathcal{S}_Y} p_k f_k(x), \tag{T3.2}$$
depending on if Y is continuous or discrete.

Example T3.1 (Normal and Cauchy distributions). Let $Y, Z \overset{\text{iid}}{\sim} N(0,1)$.

(a) Find the distribution of $X = Y + Z$.

(b) Find the distribution of $W = Y/Z$.

(c) Find the distribution of $V = Y/|Z|$, can we claim that $W \overset{\text{d}}{=} V$?

Solution: (a) From (T3.2), we first find the conditional distribution of $X|(Y = y)$. Note that
$$X|(Y=y) = y + Z \sim N(y,1), \quad \text{i.e.,} \quad f_{(X|Y)}(x|y) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x-y)^2}{2}\right].$$

Hence, from (T3.2), we have

$$
\begin{aligned}
f_X(x) &= \int_{\mathcal{S}_Y} f_Y(y) f_{(X|Y)}(x|y)\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) \cdot \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x-y)^2}{2}\right] \mathrm{d}y \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{2y^2 - 2xy + x^2}{2}\right) \mathrm{d}y \\
&= \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{4}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{2(y-0.5x)^2}{2}\right] \mathrm{d}y \\
& \quad [\text{Let } \sigma^2 = 0.5 \quad \text{and} \quad \sigma_*^2 = 2 \Rightarrow \sigma\sigma_* = 1] \\
&= \frac{1}{\sqrt{2\pi}\sigma_*} \mathrm{e}^{-\frac{x^2}{2\sigma_*^2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y-0.5x)^2}{2\sigma^2}\right] \mathrm{d}y \\
&= \frac{1}{\sqrt{2\pi}\sigma_*} \mathrm{e}^{-\frac{x^2}{2\sigma_*^2}}, \quad -\infty < x < \infty.
\end{aligned}
$$

indicating that $X \sim N(0, \sigma_*^2) = N(0, 2)$.

(b) From (T3.2), we first find the conditional distribution of $W|(Z = z)$. Note that

$$W|(Z = z) = z^{-1} \cdot Y \sim N(0, z^{-2}),$$

i.e.,

$$f_{(W|Z)}(w|z) = \frac{1}{\sqrt{2\pi}|z|^{-1}} \exp\left(-\frac{w^2}{2z^{-2}}\right) = \frac{|z|}{\sqrt{2\pi}} \exp\left(-\frac{z^2 w^2}{2}\right).$$

Hence, from (T3.2), we have

$$
\begin{aligned}
f_W(w) &= \int_{\mathcal{S}_Z} f_Z(z) f_{(W|Z)}(w|z)\,\mathrm{d}z \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \cdot \frac{|z|}{\sqrt{2\pi}} \exp\left(-\frac{z^2 w^2}{2}\right) \mathrm{d}z \\
&= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} \frac{z}{\sqrt{2\pi}} \exp\left[-\frac{z^2(1+w^2)}{2}\right] \mathrm{d}z
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{2\pi}\int_0^\infty \frac{1}{1+w^2}\cdot\exp\left[-\frac{z^2(1+w^2)}{2}\right]\,\mathrm{d}[z^2(1+w^2)] \\
&\quad [\text{Let } u = z^2(1+w^2)] \\
&= \frac{1}{2\pi}\cdot\frac{1}{1+w^2}\int_0^\infty \exp(-0.5u)\,\mathrm{d}u \\
&= \frac{1}{\pi(1+w^2)}, \quad -\infty < w < \infty.
\end{aligned}
$$

indicating that W follows the standard Cauchy distribution, i.e., $W \sim \text{Cauchy}(0,1)$. As a by-product, we have

$$
\frac{2}{\sqrt{2\pi}}\int_0^\infty \frac{z}{\sqrt{2\pi}}\exp\left[-\frac{z^2(1+w^2)}{2}\right]\mathrm{d}z = \frac{1}{\pi(1+w^2)}, \quad -\infty < w < \infty. \tag{T3.3}
$$

(c) From (T3.2), we first find the conditional distribution of $V|(|Z|=z)$, where $z \geqslant 0$. Note that

$$
V|(|Z|=z) = z^{-1}\cdot Y \sim N(0, z^{-2}),
$$

i.e.,

$$
f_{(V||Z|)}(v|z) = \frac{1}{\sqrt{2\pi}z^{-1}}\exp\left(-\frac{v^2}{2z^{-2}}\right) = \frac{z}{\sqrt{2\pi}}\exp\left(-\frac{z^2v^2}{2}\right)\cdot I(z \geqslant 0), \quad v \in \mathbb{R}.
$$

Since $Z \sim N(0,1)$, from Example 2.9, we know that

$$
f_{|Z|}(z) = [f_Z(-z) + f_Z(z)]\cdot I(z \geqslant 0) = \frac{2}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)\cdot I(z \geqslant 0).
$$

Hence, from (T3.2), we have

$$
\begin{aligned}
f_V(v) &= \int_{\mathcal{S}_{|Z|}} f_{|Z|}(z) f_{(V||Z|)}(v|z)\,\mathrm{d}z \\
&= \int_0^\infty \frac{2}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)\cdot\frac{z}{\sqrt{2\pi}}\exp\left(-\frac{z^2v^2}{2}\right)\mathrm{d}z \\
&= \frac{2}{\sqrt{2\pi}}\int_0^\infty \frac{z}{\sqrt{2\pi}}\exp\left[-\frac{z^2(1+v^2)}{2}\right]\mathrm{d}z \overset{\text{(T3.3)}}{=} \frac{1}{\pi(1+v^2)}\cdot I(v \in \mathbb{R}),
\end{aligned}
$$

i.e., $V \sim \text{Cauchy}(0,1)$. Therefore, we can claim that $W \overset{\mathrm{d}}{=} V$. $\|$

Example T3.2 (t distribution). Let $Y \sim \chi^2(\nu)$ with $\nu > 0$, $Z \sim N(0,1)$ and $Y \perp\!\!\!\perp Z$. Find the distribution of $T = Z/\sqrt{Y/\nu}$.

Solution: From (T3.2), we first find the conditional distribution of $T|(Y=y)$. Note that

$$T|(Y=y) = \sqrt{\frac{\nu}{y}} \cdot Z \sim N\left(0, \frac{\nu}{y}\right), \quad \text{i.e.,} \quad f_{(T|Y)}(t|y) = \frac{1}{\sqrt{2\pi\nu/y}} \exp\left(-\frac{t^2}{2\nu/y}\right).$$

Hence, from (T3.2), we have

$$\begin{aligned}
f_T(t) &= \int_{\mathcal{S}_Y} f_Y(y) f_{(T|Y)}(t|y)\,\mathrm{d}y \\
&= \int_0^\infty \frac{(1/2)^{\nu/2}}{\Gamma(\nu/2)} y^{\nu/2-1} \mathrm{e}^{-y/2} \cdot \frac{1}{\sqrt{2\pi\nu/y}} \exp\left(-\frac{yt^2}{2\nu}\right) \mathrm{d}y \\
&= \frac{(1/2)^{\nu/2}}{\Gamma(\nu/2)} \cdot \frac{1}{\sqrt{2\pi\nu}} \cdot \underbrace{\int_0^\infty y^{\frac{\nu+1}{2}-1} \exp\left[-y\left(\frac{1}{2}+\frac{t^2}{2\nu}\right)\right] \mathrm{d}y}_{\text{Using (1.41) in the textbook}} \\
&= \frac{(1/2)^{\nu/2}}{\Gamma(\nu/2)} \cdot \frac{1}{\sqrt{2\pi\nu}} \cdot \frac{\Gamma((\nu+1)/2)}{\left(\frac{1}{2}+\frac{t^2}{2\nu}\right)^{(\nu+1)/2}} \\
&= \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}} \left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad -\infty < t < \infty.
\end{aligned} \tag{T3.4}$$

indicating that T follows the t distribution with ν degrees of freedom, i.e., $T \sim t(\nu)$. ‖

Remark T3.1 (Standard Cauchy distribution). Show that $t(1) = \text{Cauchy}(0,1)$.

Proof: In (T3.4), let $\nu = 1$, we obtain

$$\text{Cauchy}(t|0,1) = \frac{\Gamma(1)}{\Gamma(\frac{1}{2})\sqrt{\pi}}(1+t^2)^{-1} = \frac{1}{\pi(1+t^2)}, \quad -\infty < t < \infty, \tag{T3.5}$$

which is the density of the standard Cauchy distribution. ‖

Remark T3.2 (The general t distribution). Let $T \sim t(\nu)$ with $\nu > 0$. Define $X = \mu + \sigma T$, where $\mu \in \mathbb{R}$ and $\sigma > 0$. From (T3.4), we know that the pdf of X is given by

$$t(x|\mu, \sigma^2, \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}\sigma}\left[1 + \frac{(x-\mu)^2}{\nu\sigma^2}\right]^{-\frac{\nu+1}{2}}, \quad x \in \mathbb{R}. \tag{T3.6}$$

We say that X follows the general t distribution with location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma > 0$, and ν degrees of freedom, denoted by $X \sim t(\mu, \sigma^2, \nu)$.

Especially, when $\nu = 1$, $t(\mu, \sigma^2, \nu)$ is called the general Cauchy distribution, denoted by $\text{Cauchy}(\mu, \sigma^2)$. Its pdf is

$$\text{Cauchy}(x|\mu, \sigma^2) = \frac{1}{\pi} \cdot \frac{\sigma}{\sigma^2 + (x-\mu)^2}, \quad x \in \mathbb{R}. \tag{T3.7}$$

‖

Example T3.3 (Example T2.11 revisited). Let $X_1, X_2 \overset{\text{iid}}{\sim} \text{Cauchy}(0,1)$, find the distribution of $Y = X_1 + X_2$.

Solution: From (T3.2), we first find the conditional distribution of $Y|(X_1 = x_1)$ with $x_1 \in \mathbb{R}$. Note that

$$Y|(X_1 = x_1) = \underbrace{x_1}_{\mu} + \underbrace{1}_{\sigma} X_2 \sim \text{Cauchy}(x_1, 1),$$

i.e,

$$f_{(Y|X_1)}(y|x_1) \overset{\text{(T3.7)}}{=} \frac{1}{\pi[1 + (y-x_1)^2]} \cdot I(x_1 \in \mathbb{R}), \quad y \in \mathbb{R}.$$

Hence, from (T3.2), we have

$$\begin{aligned} f_Y(y) &= \int_{\mathcal{S}_{X_1}} f_{X_1}(x_1) f_{(Y|X_1)}(y|x_1)\,\mathrm{d}x_1 \\ &= \int_{-\infty}^{\infty} \frac{1}{\pi(1+x_1^2)} \cdot \frac{1}{\pi[1+(y-x_1)^2]}\,\mathrm{d}x_1 \\ &= \frac{1}{\pi} \cdot \underbrace{\int_{-\infty}^{\infty} \frac{1}{1+x_1^2} \cdot \frac{1}{\pi[1+(x_1-y)^2]}\,\mathrm{d}x_1 = \frac{1}{\pi} \cdot \frac{2}{4+y^2}}_{\text{Using (T2.6) in Tutorial 2}}, \end{aligned}$$

indicating that $Y \sim \text{Cauchy}(0, 2)$ from (T3.7). ‖

Example T3.4 (F distribution). Let $U \sim \chi^2(\nu_1)$ with $\nu_1 > 0$, $V \sim \chi^2(\nu_2)$ with $\nu_2 > 0$, and $U \perp\!\!\!\perp V$. Find the distribution of $W = (U/\nu_1)/(V/\nu_2)$.

Solution: Note that the following two facts:

$$\chi^2(\nu) = \text{Gamma}\left(\frac{\nu}{2}, \frac{1}{2}\right) \quad \text{and} \quad c \cdot \text{Gamma}(\alpha, \beta) = \text{Gamma}\left(\alpha, \frac{\beta}{c}\right) \text{ for any } c > 0.$$

From (T3.2), we first find the conditional distribution of $W|(V = v)$. Note that

$$W|(V = v) = \frac{\nu_2}{\nu_1 v} \cdot U = \frac{\nu_2}{\nu_1 v} \cdot \text{Gamma}\left(\frac{\nu_1}{2}, \frac{1}{2}\right) \sim \text{Gamma}\left(\frac{\nu_1}{2}, \frac{\nu_1 v}{2\nu_2}\right),$$

i.e.,

$$f_{(W|V)}(w|v) = \frac{\left(\frac{\nu_1 v}{2\nu_2}\right)^{\nu_1/2}}{\Gamma(\nu_1/2)} \cdot w^{\nu_1/2-1} \exp\left(-w\frac{\nu_1 v}{2\nu_2}\right) \cdot I(v \geqslant 0), \quad w \geqslant 0.$$

Hence, from (T3.2), we have

$$\begin{aligned}
f_W(w) &= \int_{\mathcal{S}_V} f_V(v) f_{(W|V)}(w|v)\, \mathrm{d}v \\
&= \int_0^\infty \frac{(1/2)^{\nu_2/2}}{\Gamma(\nu_2/2)} v^{\nu_2/2-1} \mathrm{e}^{-v/2} \cdot \frac{\left(\frac{\nu_1 v}{2\nu_2}\right)^{\nu_1/2}}{\Gamma(\nu_1/2)} \cdot w^{\nu_1/2-1} \exp\left(-w\frac{\nu_1 v}{2\nu_2}\right) \mathrm{d}v \\
&= \frac{(1/2)^{\nu_2/2}}{\Gamma(\nu_2/2)} \cdot \frac{\left(\frac{\nu_1}{2\nu_2}\right)^{\nu_1/2}}{\Gamma(\nu_1/2)} \cdot w^{\nu_1/2-1} \underbrace{\int_0^\infty v^{\frac{\nu_1+\nu_2}{2}-1} \exp\left[-v\left(\frac{1}{2} + \frac{w\nu_1}{2\nu_2}\right)\right] \mathrm{d}v}_{\text{Using (1.41) in the textbook}} \\
&= \frac{(1/2)^{\nu_2/2}}{\Gamma(\nu_2/2)} \cdot \frac{\left(\frac{\nu_1}{2\nu_2}\right)^{\nu_1/2}}{\Gamma(\nu_1/2)} \cdot w^{\nu_1/2-1} \cdot \frac{\Gamma((\nu_1+\nu_2)/2)}{\left(\frac{1}{2} + \frac{w\nu_1}{2\nu_2}\right)^{(\nu_1+\nu_2)/2}} \\
&= \frac{\Gamma(\frac{\nu_1+\nu_2}{2})}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})} \left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} w^{\nu_1/2-1} \left(1 + \frac{\nu_1}{\nu_2} w\right)^{-\frac{\nu_1+\nu_2}{2}}, \quad w > 0,
\end{aligned}$$

indicating that W follows the F distribution with ν_1 and ν_2 degrees of freedom, i.e., $W \sim F(\nu_1, \nu_2)$. $\|$

T3.2 Stochastic Representation (SR) Technique (The Sixth Method)

Stochastic representation technique is a probability method which directly deals with random variables/vectors instead of treating their cdfs, pdfs (or pmfs), mgfs.

Example T3.5 (Example T3.1 revisited). Let $Y, Z \overset{\text{iid}}{\sim} N(0,1)$. Find the distribution of $V = Y/|Z|$ via the SR technique.

Solution: Since $Z^2 \sim \chi^2(1)$, by, we have

$$V = \frac{Y}{|Z|} = \frac{Y}{\sqrt{Z^2}} = \frac{N(0,1)}{\sqrt{\chi^2(1)/1}} \sim \underbrace{t(1) = \text{Cauchy}(0,1)}_{\text{Remark T3.1}}. \qquad \|$$

Example T3.6 (Relationship between t and F distributions). Let $T \sim t(\nu)$. Find the distribution of T^2 via the SR technique.

Solution: Let $Y \sim N(0,1)$, then

$$T^2 = \left[\frac{Y}{\sqrt{\chi^2(\nu)/\nu}}\right]^2 = \frac{\chi^2(1)/1}{\chi^2(\nu)/\nu} \sim F(1,\nu). \qquad \|$$

Example T3.7 (The gamma–integral distribution). Let X have the gamma–integral distribution with parameter $a > 0$. Its density is given by

$$f_X(x) = \int_x^\infty \frac{y^{a-2}\mathrm{e}^{-y}}{\Gamma(a)}\,\mathrm{d}y, \quad x > 0.$$

Find $E(X)$ and $\text{Var}(X)$.

Solution: The analytic method is rather tedious because

$$E(X) = \int_0^\infty x\left[\int_x^\infty \frac{y^{a-2}\mathrm{e}^{-y}}{\Gamma(a)}\,\mathrm{d}y\right]\mathrm{d}x.$$

Note that

$$\begin{aligned} f_X(x) &= \int_x^\infty \frac{y^{a-2}\mathrm{e}^{-y}}{\Gamma(a)}\,\mathrm{d}y = \int_0^\infty \frac{y^{a-2}\mathrm{e}^{-y}}{\Gamma(a)} \cdot I(y > x)\,\mathrm{d}y \\ &= \int_0^\infty \frac{y^{a-1}\mathrm{e}^{-y}}{\Gamma(a)} \cdot \frac{I(0 < x < y)}{y}\,\mathrm{d}y \\ &\overset{(\mathrm{T3.2})}{=} \int_0^\infty f_Y(y) \cdot f_{(X|Y)}(x|y)\,\mathrm{d}y. \end{aligned}$$

Therefore, $Y \sim \mathrm{Gamma}(a, 1)$ and $X|(Y = y) \sim U(0, y)$ or

$$\left.\frac{X}{y}\right|(Y = y) \sim U(0, 1),$$

which is free from y. Let $U \sim U(0, 1)$, we have

$$\left.\frac{X}{y}\right|(Y = y) \overset{\mathrm{d}}{=} U$$

where U is independent of Y. Thus the unconditional distribution

$$\frac{X}{Y} \overset{\mathrm{d}}{=} U \quad \Rightarrow \quad X \overset{\mathrm{d}}{=} UY.$$

Hence,

$$\begin{aligned} E(X) &= E(U) \cdot E(Y) \quad = a/2, \\ E(X^2) &= E(U^2) \cdot E(Y^2) = a(a+1)/3, \end{aligned}$$

and $\mathrm{Var}(X) = a(a+4)/12$. ‖

Example T3.8 (Dirichlet, its marginal and conditional distributions). The traditional way of defining a Dirichlet distribution is through density function. An n-dimensional random vector $\mathbf{x} = (X_1, \dots, X_n)^\top$ with realization values $\boldsymbol{x} = (x_1, \dots, x_n)^\top$ is said to have a *Dirichlet distribution* with parameters $a_1, \dots, a_{n+1}$ if its joint pdf is

$$\frac{\Gamma(\sum_{i=1}^{n+1} a_i)}{\prod_{i=1}^{n+1} \Gamma(a_i)} \left(\prod_{i=1}^n x_i^{a_i - 1}\right) \left(1 - \sum_{i=1}^n x_i\right)^{a_{n+1}-1} \cdot I_{\mathbb{V}_n}(\boldsymbol{x}), \tag{T3.8}$$

where

$$\mathbb{V}_n \hat{=} \left\{ \boldsymbol{x} = (x_1, \dots, x_n)^\top \colon \boldsymbol{x} \in \mathbb{R}_+^n, \ \sum_{i=1}^n x_i \leqslant 1 \right\}.$$

We will write $\mathbf{x} \sim \text{Dirichlet}(a_1, \dots, a_n; a_{n+1})$.

The density-based definition requires to verify the following integral identity:

$$\int_{\mathbb{V}_n} \frac{\Gamma(\sum_{i=1}^{n+1} a_i)}{\prod_{i=1}^{n+1} \Gamma(a_i)} \left(1 - \sum_{i=1}^n x_i\right)^{a_{n+1}-1} \prod_{i=1}^n x_i^{a_i - 1} \mathrm{d}x_i = 1.$$

Moreover, to derive a marginal distribution, we need to perform an integration with respect to the variables that are not of interest.

SR-based definition: Alternatively, a SR-based definition can be described as follows. An n-dimensional r.v. $(X_1, \dots, X_n)^\top$ is said to follow a Dirichlet distribution if

$$X_i \stackrel{\mathrm{d}}{=} \frac{Y_i}{Y_1 + \cdots + Y_n + Y_{n+1}}, \quad i = 1, \dots, n, \tag{T3.9}$$

where $Y_i \sim \text{Gamma}(a_i, 1)$ for $i = 1, \dots, n+1$, and $\{Y_i\}_{i=1}^{n+1}$ be mutually independent.

Marginal distributions: Now, marginal distributions are easy to obtain without performing any integrations. For any $s < n$, we have

$$X_i \stackrel{\mathrm{d}}{=} \frac{Y_i}{Y_1 + \cdots + Y_s + \underbrace{Y_{s+1} + \cdots + Y_{n+1}}_{Y'_{s+1}}}, \quad i = 1, \dots, s,$$

where $Y_i \sim \text{Gamma}(a_i, 1)$, $i = 1, \dots, s$, $Y'_{s+1} \hat{=} \sum_{j=s+1}^{n+1} Y_j \sim \text{Gamma}(\sum_{j=s+1}^{n+1} a_j, 1)$, and $Y_1, \dots, Y_s, Y'_{s+1}$ are mutually independent. It follows immediately that

$$(X_1, \dots, X_s)^\top \sim \text{Dirichlet}\left(a_1, \dots, a_s; \sum_{j=s+1}^{n+1} a_j\right).$$

Conditional distribution: The SR (T3.9) can also be employed to derive conditional distribution of $(X_{s+1}, \dots, X_n)^\top$ given $X_1 = x_1, \dots, X_s = x_s$. Set

$$X'_i = \frac{X_i}{1 - \sum_{j=1}^s X_j}, \qquad i = s+1, \dots, n.$$

By (T3.9), we have

$$X_i' = \frac{Y_i}{Y_{s+1} + \cdots + Y_n + Y_{n+1}}, \quad i = s+1, \ldots, n,$$

so that

$$(X_{s+1}', \ldots, X_n')^\top \Big|_{X_1=x_1,\ldots,X_s=x_s} \sim \text{Dirichlet}(a_{s+1}, \ldots, a_n; a_{n+1}),$$

which is independent of $X_1, \ldots, X_s$, i.e.,

$$(X_{s+1}', \ldots, X_n')^\top \sim \text{Dirichlet}(a_{s+1}, \ldots, a_n; a_{n+1}).$$

Therefore, given $X_1 = x_1, \ldots, X_s = x_s$, we have

$$X_i \stackrel{\text{d}}{=} (1 - \textstyle\sum_{j=1}^s x_j) X_i', \qquad i = s+1, \ldots, n.$$

In other words, we have

$$(X_{s+1}, \ldots, X_n)^\top \Big|_{X_1=x_1,\ldots,X_s=x_s} \stackrel{\text{d}}{=} \underbrace{(1 - \textstyle\sum_{j=1}^s x_j)}_{\text{a positive number}} \cdot (X_{s+1}', \ldots, X_n')^\top. \qquad \|$$

T3.3 Order Statistics

3.3.1 Definition

Let $X_1, \ldots, X_n$ be a random sample from a population with cdf $F(\cdot)$ and pdf $f(\cdot)$. Then, $X_{(1)} \leqslant \cdots \leqslant X_{(n)}$ are called the *order statistics*.

How to understand $X_{(1)} = \min(X_1, \ldots, X_n)$?

```
> x <- rnorm(4, mean=2, sd=0.2)                            min(x)
Sample 1: 1.792966 1.935051 2.205505 1.717259           1.717259
Sample 2: 1.736903 2.092398 1.955097 2.292306           1.736903
Sample 3: 3.100611 2.144726 1.855688 1.862706           1.855688
Sample 4: 2.087154 1.733278 2.185363 1.823968           1.733278
Sample 5: 1.821756 1.861677 1.785434 1.997736           1.785434
Sample 6: 2.209436 1.836078 2.413644 1.998032           1.836078
Sample 7: 1.862467 1.938282 2.011536 2.223131           1.862467
Sample 8: 1.927599 2.012464 2.003386 1.899905           1.899905
```

3.3.2 Single order statistic

Let $G_r(x)$ and $g_r(x)$ be the cdf and pdf of the r-th order statistic $X_{(r)}$, respectively. Then

$$
\begin{aligned}
G_r(x) &= \sum_{i=r}^{n} \binom{n}{i} F^i(x)[1-F(x)]^{n-i} = \frac{1}{B(r, n-r+1)} \int_0^{F(x)} t^{r-1}(1-t)^{n-r}\, \mathrm{d}t, \\
g_r(x) &= \frac{n!}{(r-1)!(n-r)!} f(x) F^{r-1}(x)[1-F(x)]^{n-r}.
\end{aligned}
$$

3.3.3 Multiple order statistics

Let $g_{r_1,\ldots,r_k}(x_1,\ldots,x_k)$ be the joint pdf of $X_{(r_1)},\ldots,X_{(r_k)}$ $(1 \leqslant r_1 < \cdots < r_k \leqslant n;\ 1 \leqslant k \leqslant n)$,

$$
\begin{aligned}
g_{r_1,\ldots,r_k}(x_1,\ldots,x_k) &= n! \left[\prod_{i=1}^{k} f(x_i)\right] \cdot \prod_{i=0}^{k} \left\{ \frac{[F(x_{i+1}) - F(x_i)]^{r_{i+1}-r_i-1}}{(r_{i+1}-r_i-1)!} \right\}, \\
g_{1,\ldots,r}(x_1,\ldots,x_r) &= \frac{n!}{(n-r)!} f(x_1)\cdots f(x_r)[1-F(x_r)]^{n-r}, \\
g_{r,s}(x,y) &= \frac{n!}{(r-1)!(s-r-1)!(n-s)!} f(x) f(y) \\
&\quad \times F^{r-1}(x)[F(y)-F(x)]^{s-r-1}[1-F(y)]^{n-s}.
\end{aligned}
$$

Example T3.9 (An exponential distribution). Let $\{X_i\}_{i=1}^n \stackrel{\text{ind}}{\sim} \text{Exponential}(\lambda_i)$. Show that

(a) $Y = \min(X_1,\ldots,X_n) \sim \text{Exponential}(\lambda_+)$, where $\lambda_+ = \lambda_1 + \cdots + \lambda_n$.

(b) For a fixed i, $\Pr(Y = X_i) = \Pr(X_i \leqslant X_j,\ \forall j = 1,\ldots,n; j \neq i) = \lambda_i/\lambda_+$.

Proof: (a) Although $X_1,\ldots,X_n$ are independent and have the same family of distribution, but they have different parameters. Thus, the formula (2.18) of $G_1(\cdot)$ on page 82 of the textbook cannot be applied. Since $F_{X_i}(x) = (1 - \mathrm{e}^{-\lambda_i x}) \cdot I_{[0,\infty)}(x)$, we have

$$\begin{aligned}
\Pr(Y \leqslant y) &= 1 - \Pr(Y > y) = 1 - \Pr(\text{all } X_i > y) \\
&= 1 - \prod_{i=1}^{n} \Pr(X_i > y) \qquad [\text{since all } X_i \text{ are independent}] \\
&= 1 - \prod_{i=1}^{n} [1 - \Pr(X_i \leqslant y)] = 1 - \prod_{i=1}^{n} [1 - F_{X_i}(y)] \\
&= 1 - \mathrm{e}^{-(\sum_{i=1}^{n} \lambda_i) y} = 1 - \mathrm{e}^{-\lambda_+},
\end{aligned}$$

indicating that $Y \sim \text{Exponential}(\lambda_+)$.

(b) Let $Y_i = \min(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ and $\lambda_{-i} = \lambda_+ - \lambda_i$, then we have $Y_i \sim \text{Exponential}(\lambda_{-i})$. Furthermore, let $Z_i = X_i - Y_i$, then

$$\Pr(Y = X_i) = \Pr(X_i \leqslant X_j, \forall j \neq i) = \Pr(X_i \leqslant Y_i) = \Pr(Z_i \leqslant 0),$$

and

$$\begin{aligned}
F_{Z_i}(z) &= \Pr(X_i - Y_i \leqslant z) \\
&= \int_0^\infty \int_{x-z}^\infty f_{(X_i,Y_i)}(x,y)\mathrm{d}y\mathrm{d}x \qquad [x - y \leqslant z \Rightarrow x - z \leqslant y] \\
&= \int_0^\infty \int_{x-z}^\infty f_{X_i}(x) f_{Y_i}(y)\mathrm{d}y\mathrm{d}x \qquad [\because X_i \perp\!\!\!\perp Y_i\,] \\
&= \int_0^\infty \lambda_i \mathrm{e}^{-\lambda_i x} \left(\int_{x-z}^\infty \lambda_{-i} \mathrm{e}^{-\lambda_{-i} y} \mathrm{d}y \right) \mathrm{d}x \\
&= \int_0^\infty \lambda_i \mathrm{e}^{-\lambda_i x} \left(-\mathrm{e}^{-\lambda_{-i} y} \Big|_{x-z}^{\infty} \right) \mathrm{d}x \\
&= \int_0^\infty \lambda_i \mathrm{e}^{-\lambda_i x} \cdot \mathrm{e}^{-\lambda_{-i}(x-z)} \mathrm{d}x \\
&= \lambda_i \mathrm{e}^{\lambda_{-i} z} \int_0^\infty \mathrm{e}^{-\lambda_+ x} \mathrm{d}x = \frac{\lambda_i \mathrm{e}^{\lambda_{-i} z}}{\lambda_+}.
\end{aligned}$$

Thus, $\Pr(Y = X_i) = \Pr(Z_i \leqslant 0) = F_{Z_i}(0) = \lambda_i/\lambda_+$.

Comments on the key point: When we meet **several** inequalities such as

$$X_i \leqslant X_j,\ \forall j \neq i$$

we should seek for **one** inequality like $X_i \leqslant Y_i$ with

$$Y_i = \min(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n).$$

If these inequalities are given by $X_i \geqslant X_j$, $\forall j \neq i$, then we have $X_i \geqslant Y_i^*$ with

$$Y_i^* = \max(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n). \qquad \|$$

T3.4 Convergence in Distribution

3.4.1 Definition

Let the cdf $F_n(x)$ of the r.v. X_n depend upon n ($n = 1, 2, \ldots$). If $F(x)$ is a distribution function of an r.v. X and

$$\lim_{n\to\infty} F_n(x) = F(x), \quad \forall x \text{ s.t. } F(x) \text{ is continuous},$$

then $\{X_n\}_{n=1}^{\infty}$ *converge in distribution* to X and denote it by $X_n \xrightarrow{\text{L}} X$.

3.4.2 Theorem

Let $\{X_n\}_{n=1}^{\infty}$ have mgfs $M(t;n)$, $t \in (-h, h)$. We say $X_n \xrightarrow{\text{L}} X$, if there exists an mgf $M(t)$ for the r.v. X with cdf $F(x)$ such that

$$M(t) = \lim_{n\to\infty} M(t;n), \quad \forall t \in (-h, h).$$

Example T3.10 (Normal distribution). Let the sequence of r.v.s $\{X_n\}_{n=1}^{\infty} \overset{\text{ind}}{\sim} N(0, 1/n)$. Show that $X_n \xrightarrow{\text{L}} X$, where $X \sim$ Degenerate(0), i.e., $\Pr(X = 0) = 1$, and the cdf of X is $F(x) = 0 \cdot I(x < 0) + 1 \cdot I(x \geqslant 0)$.

Solution: Let $\phi(t) = \exp(-0.5t^2)/\sqrt{2\pi}$ and $\Phi(t)$ denote the pdf and cdf of $N(0,1)$, respectively, then the cdf of X_n is

$$F_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi/n}} \mathrm{e}^{-\frac{t^2}{2/n}} \mathrm{d}t = \int_{-\infty}^{\sqrt{n}x} \phi(s)\mathrm{d}s = \Phi(\sqrt{n}x) \qquad (\text{let } s = \sqrt{n}t),$$

so that

$$\begin{aligned}
\lim_{n\to\infty} F_n(x) &= \lim_{n\to\infty} \Phi(\sqrt{n}x) \\
&= \begin{cases} \Phi(-\infty) = 0, & \text{if } x < 0, \\ \Phi(+\infty) = 1, & \text{if } x > 0, \\ \Phi(0) = 0.5, & \text{if } x = 0 \end{cases} \\
&\to F(x) \\
&= \begin{cases} 0, & \text{if } x < 0, \\ 1, & \text{if } x > 0, \\ 1, & \text{if } x = 0 \end{cases}
\end{aligned}$$

for any $x \neq 0$, where $x = 0$ is the discontinuous point of $F(x)$. Therefore, $X_n \xrightarrow{\mathrm{L}} X$. ‖

T3.5 Central Limit Theorem (CLT)

3.5.1 Definition and understanding

If $\{X_n\}_{n=1}^{\infty}$ is a sequence of i.i.d. r.v.'s with the mean μ and the variance $\sigma^2 < +\infty$, then

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{\mathrm{L}} N(0,1), \quad \text{where} \quad \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

How to understand the CLT? First, there are two conditions in the CLT: (i) $\{X_n\}_{n=1}^{\infty}$ are i.i.d.; (ii) the variance σ^2 exists. Next, the following statement

$$\bar{X}_n \xrightarrow{\mathrm{L}} N(\mu, \sigma^2/n)$$

is not good because n is in the $N(\mu, \sigma^2/n)$ distribution. When $n \to \infty$, we have $N(\mu, \sigma^2/n) \to \text{Degenerate}(\mu)$. $\|$

3.5.2 Negative binomial distribution

Let $Y \sim \text{NBinomial}\ (r, p)$ with pmf

$$\Pr(Y = y) = \text{NBinomial}(y|r, p) \mathrel{\hat{=}} \binom{y+r-1}{y} p^r (1-p)^y, \qquad y = 0, 1, \ldots, \infty.$$

When r is a positive integer, the negative binomial distribution is also called the *Pascal distribution* (after Blaise Pascal).

— The negative binomial r.v. Y can be viewed as the number of failures in a sequence of i.i.d. Bernoulli trials before r successes occur, where p is the probability of success.

— Define $\pi = 1 - p$, we have

$$p^{-r} = (1-\pi)^{-r} = \sum_{y=0}^{\infty} \binom{-r}{y} (-\pi)^y = \sum_{y=0}^{\infty} \binom{y+r-1}{y} (1-p)^y.$$

Thus $\sum_{y=0}^{\infty} \text{NBinomial}(y|r, p) = 1$.

— The *geometric distribution* is the special case of the negative-binomial with $r = 1$.

— Let $X_1, \ldots, X_r \overset{\text{iid}}{\sim} \text{NBinomial}(1, p)$, then $Y = \sum_{i=1}^{r} X_i \sim \text{NBinomial}(r, p)$.

Example T3.11 (Negative binomial distribution). Let $Y \sim \text{NBinomial}\ (r, p)$ with $r = 20$ and $p = 0.7$.

(a) Exactly calculate the value of $\Pr(Y = 12)$.

(b) Approximately calculate $\Pr(Y = 12)$ by the Central Limit Theorem.

Solution: (a) In the definition (see, the last row of Table 1.2 in Textbook Chapter 1) of the pmf for a negative binomial distribution, let $X - r = Y$, then, the pmf of $Y \sim \text{NBinomial}(r, p)$ is

$$\Pr(Y = y) = \binom{y + r - 1}{y} p^r (1-p)^y, \qquad y = 0, 1, 2, \ldots, \infty,$$

where $E(Y) = E(X) - r = r(1-p)/p$ and $\text{Var}(Y) = \text{Var}(X) = r(1-p)/p^2$. We directly compute

$$\Pr(Y = 12) = \binom{12 + 20 - 1}{12} \times 0.7^{20} \times 0.3^{12} = 0.0598.$$

(b) Now,

$$E(Y) = \frac{20 \times (1 - 0.7)}{0.7} = 8.5714 \quad \text{and} \quad \text{Var}(Y) = \frac{20 \times (1 - 0.7)}{0.7^2} = 12.2449.$$

Let $X_1, \ldots, X_r \overset{\text{iid}}{\sim} \text{NBinomial}(1, p)$, then $Y = \sum_{i=1}^r X_i \sim \text{NBinomial}(r, p)$. By CLT,

$$Z = \sqrt{r} \cdot \frac{\bar{X} - E(X_i)}{\sqrt{\text{Var}(X_i)}} = \sqrt{r} \cdot \frac{Y/r - E(Y)/r}{\sqrt{\text{Var}(Y)/r}} = \frac{Y - E(Y)}{\sqrt{\text{Var}(Y)}} \xrightarrow{\text{L}} N(0, 1).$$

We obtain

$$\begin{aligned}
\Pr(Y = 12) &= \Pr(12 - 0.5 < Y < 12 + 0.5) \\
&= \Pr\left(\frac{11.5 - 8.5714}{\sqrt{12.2449}} < \frac{Y - 8.5714}{\sqrt{12.2449}} < \frac{12.5 - 8.5714}{\sqrt{12.2449}}\right) \\
&= \Pr(0.8369 < Z < 1.1227) \\
&= \Phi(1.1227) - \Phi(0.8369) = 0.8692 - 0.7987 = 0.0705.
\end{aligned}$$

The error is $0.0705 - 0.0598 = 0.0107$, and the percentage error is $\frac{0.0107}{0.0598} = 17.87\%$. $\|$

Example T3.12 (Poisson distribution). Let $\{X_i\}_{i=1}^{50} \overset{\text{iid}}{\sim} \text{Poisson}(0.03)$ and $Y = \sum_{i=1}^{50} X_i$. Calculate $\Pr(Y \geqslant 3)$.

Solution: $E(X_i) = 0.03$ and $\mathrm{Var}(X_i) = 0.03$. By CLT,

$$
\begin{aligned}
Z &= \sqrt{n} \cdot \frac{\bar{X} - E(X_i)}{\sqrt{\mathrm{Var}(X_i)}} = \sqrt{50} \cdot \frac{Y/50 - 0.03}{\sqrt{0.03}} \\
&= \frac{Y - 50 \times 0.03}{\sqrt{50 \times 0.03}} = \frac{Y - 1.5}{\sqrt{1.5}} \xrightarrow{\mathrm{L}} N(0,1).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\Pr(Y \geqslant 3) &= \Pr\left(Z = \frac{Y - 1.5}{\sqrt{1.5}} \geqslant \frac{3 - 1.5}{\sqrt{1.5}}\right) = 1 - \Pr(Z < 1.224745) \\
&= 1 - \Phi(1.224745) = 1 - 0.8896643 = 0.1103357,
\end{aligned}
$$

which is desired. $\|$