Chapter 2

Sampling Distributions

2.1 Distribution of the Function of Random Variables

1• Aim of this section

- Given a set of r.v.'s $X_1, \ldots, X_n$ with the cdf $F(x_1, \ldots, x_n)$ or the pdf $f(x_1, \ldots, x_n)$, we seek the distribution of $Y = h(X_1, \ldots, X_n)$ for some function $h(\cdot)$.
- In this section, we will introduce three commonly used methods.

1.1• Three techniques

— Cumulative distribution function technique.

— Transformation technique.

— Moment generating function technique.

2.1.1 Cumulative distribution function technique

2• The continuous case

- A set of r.v.'s $X_1, \ldots, X_n$ can define a new r.v. $Y = h(X_1, \ldots, X_n)$ via the function $h(\cdot)$.
- The distribution of Y can be determined by the transformation $h(\cdot)$ together with the joint distribution of $X_1, \ldots, X_n$.

2.1• The procedure of cdf

— If $X_1, \ldots, X_n$ are continuous r.v.'s, then the cdf of Y can be determined by integrating $f(x_1, \ldots, x_n)$ over the domain

$$\mathbb{D} = \{(x_1, \ldots, x_n)\colon\ h(x_1, \ldots, x_n) \leqslant y\};$$

that is

$$\begin{aligned} G(y) &= \Pr(Y \leqslant y) \\ &= \Pr\{h(X_1, \ldots, X_n) \leqslant y\} \\ &= \int_{\mathbb{D}} f(x_1, \ldots, x_n)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n. \end{aligned}$$

— Then by differentiating it with respect to y, we obtain the density of Y as $g(y) = G'(y)$.

Example 2.1 (Beta distribution). Suppose that $X \sim \text{Beta}(2, 2)$, then its pdf is $f(x) = 6x(1-x)$, $0 \leqslant x \leqslant 1$. Find the pdf of $Y = X^3$.

Solution. The distribution function of Y for $0 \leqslant y \leqslant 1$ is

$$\begin{aligned} G(y) &= \Pr(X^3 \leqslant y) \\ &= \Pr(X \leqslant y^{1/3}) \\ &= \int_0^{y^{1/3}} 6x(1-x)\,\mathrm{d}x \\ &= 3y^{2/3} - 2y. \end{aligned}$$

Then, the pdf of Y is $g(y) = 2y^{-1/3} - 2$, $0 \leqslant y \leqslant 1$.

The corresponding densities and distribution functions of $X \sim \text{Beta}(2, 2)$ and $Y = X^3$ are shown in Figure 2.1. ‖

Example 2.2 (Bivariate exponential distribution). Let

$$(X_1, X_2) \sim f(x_1, x_2) = 6\exp(-3x_1 - 2x_2), \quad x_1 \geqslant 0, \quad x_2 \geqslant 0.$$

Find the pdf of $Y = X_1 + X_2$.

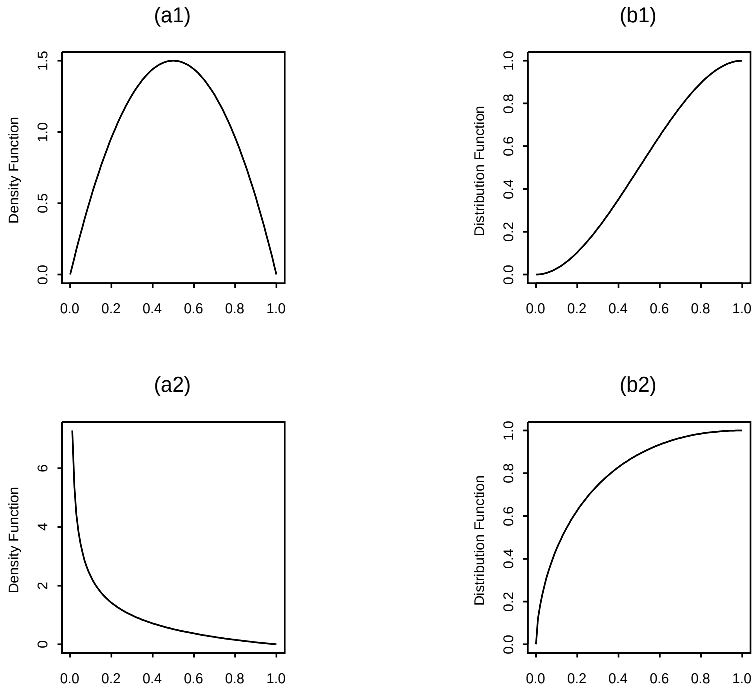

Figure 2.1 The densities and distribution functions of $X \sim \text{Beta}(2,2)$ and $Y = X^3$. (a1) The density $f(x)$ of X; (b1) The cdf $F(x)$ of X; (a2) The density $g(y)$ of Y; (b2) The cdf $G(y)$ of Y.

Solution. The cdf of Y is

$$\begin{aligned}
G(y) &= \int\int_{\mathbb{D}} 6\exp(-3x_1 - 2x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 \\
&= \int_0^y \left\{ \int_0^{y-x_2} 6\exp(-3x_1 - 2x_2)\,\mathrm{d}x_1 \right\} \mathrm{d}x_2 \\
&= \int_0^y 2\,\mathrm{e}^{-2x_2}\{1 - \mathrm{e}^{-3(y-x_2)}\}\,\mathrm{d}x_2 \\
&= 1 + 2\,\mathrm{e}^{-3y} - 3\,\mathrm{e}^{-2y}, \quad y \geqslant 0,
\end{aligned}$$

where $\mathbb{D} = \{(x_1, x_2)\colon x_1 \geqslant 0, x_2 \geqslant 0, x_1 + x_2 \leqslant y\}$ with $y \geqslant 0$ denotes the integration region. Figure 2.2 gives an illustration for the $\mathbb{D}$.

Therefore, the density of Y is

$$g(y) = 6(\mathrm{e}^{-2y} - \mathrm{e}^{-3y}), \quad y \geqslant 0.$$

Figure 2.3 shows the pdf $g(y)$ and the cdf $G(y)$ of $Y = X_1 + X_2$.

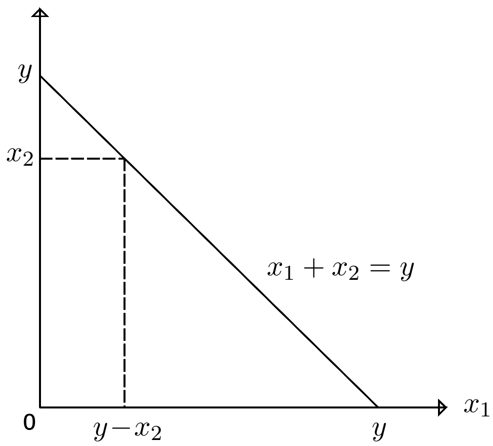

Figure 2.2 The integration region $\mathbb{D} = \{(x_1, x_2)\colon x_1 \geqslant 0, x_2 \geqslant 0, x_1 + x_2 \leqslant y\}$.

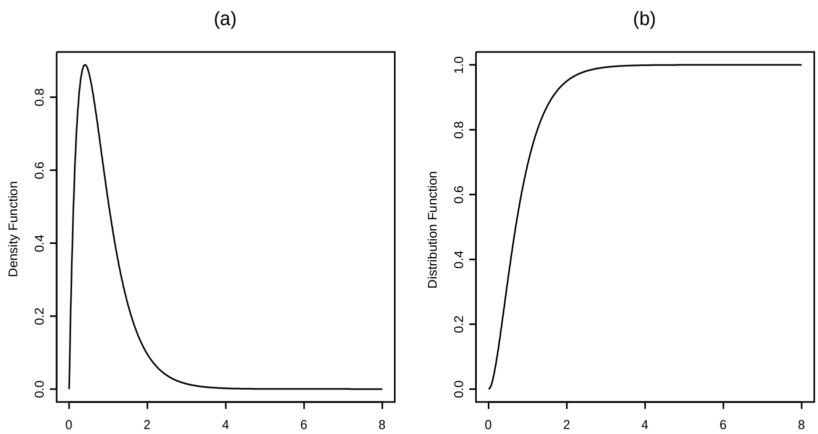

Figure 2.3 The density function and distribution function of $Y = X_1 + X_2$. (a) The pdf $g(y)$ of Y; (b) The cdf $G(y)$ of Y. ‖

3• The discrete case

- For the purpose of illustration, first we let $n = 1$.
- If X is a discrete r.v. taking values $\{x_i\}$ with probabilities $\{p_i\}$, then the distribution of $Y = h(X)$ is determined directly by the laws of probability.

- It may be that several values of X give rise to the same value of Y.
- The probability that Y takes on a given value, say y_j, is

$$\Pr(Y = y_j) = \sum_{\{i:\, h(x_i)=y_j\}} p_i.$$

Example 2.3 (Finite discrete distribution). Suppose that X takes the values of 0, 1, 2, 3, 4, 5 with the corresponding probabilities p_0, p_1, p_2, p_3, p_4 and p_5. Find the pmf of $Y = h(X) = (X-2)^2$.

Solution. From the following table

X	0	1	2	3	4	5
$p_i = \Pr(X = x_i)$	p_0	p_1	p_2	p_3	p_4	p_5
$Y = (X-2)^2$	4	1	0	1	4	9

we note that Y can take on values 0, 1, 4 and 9; then

$$\begin{aligned} \Pr(Y=0) &= p_2, & \Pr(Y=1) = p_1 + p_3, \\ \Pr(Y=4) &= p_0 + p_4, & \Pr(Y=9) = p_5. \end{aligned}$$

‖

Example 2.4 (Joint discrete distribution). Let (X_1, X_2, X_3) have a joint discrete distribution given by

(X_1, X_2, X_3)	$(0,0,0)$	$(0,0,1)$	$(0,1,1)$	$(1,0,1)$	$(1,1,0)$	$(1,1,1)$
Probability	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

Find the pmf of $Y = h(X_1, X_2, X_3) = X_1 + X_2 + X_3$.

Solution. We note that Y can take on values 0, 1, 2 and 3; then

$$\begin{aligned} \Pr(Y=0) &= \frac{1}{8}, \\ \Pr(Y=1) &= \frac{3}{8}, \\ \Pr(Y=2) &= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}, \\ \Pr(Y=3) &= \frac{1}{8}. \end{aligned}$$

‖

Example 2.5 (Poisson distribution). Let $X_i \sim \text{Poisson}(\lambda_i)$, $i = 1, 2$, and $X_1 \perp\!\!\!\perp X_2$, find the pmf of $Y = X_1 + X_2$.

Solution. The pmf of $Y = X_1 + X_2$ is

$$\begin{aligned}
\Pr(Y = y) &= \Pr(X_1 + X_2 = y) \\
&= \sum_{x=0}^{y} \Pr(X_1 = x, X_2 = y - x) \\
&= \sum_{x=0}^{y} \Pr(X_1 = x) \cdot \Pr(X_2 = y - x) \\
&= \sum_{x=0}^{y} \frac{\lambda_1^x}{x!} \mathrm{e}^{-\lambda_1} \cdot \frac{\lambda_2^{y-x}}{(y-x)!} \mathrm{e}^{-\lambda_2} \\
&= \frac{1}{y!} \mathrm{e}^{-(\lambda_1+\lambda_2)} \sum_{x=0}^{y} \binom{y}{x} \lambda_1^x \lambda_2^{y-x} \\
&= \frac{(\lambda_1 + \lambda_2)^y}{y!} \mathrm{e}^{-(\lambda_1+\lambda_2)}, \quad y = 0, 1, 2, \ldots.
\end{aligned}$$

Therefore, $X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$. ‖

2.1.2 Transformation technique

4• Monotone transformation

- Let $f(x)$ and $F(x)$ denote the corresponding pdf and cdf of an r.v. X.
- If $y = h(x)$ is a differentiable and monotone function and the inverse function is $x = h^{-1}(y)$, then the pdf of $Y = h(X)$ is given by

$$g(y) = f(x) \times \left| \frac{\mathrm{d}x}{\mathrm{d}y} \right| = f(h^{-1}(y)) \times \left| \frac{\mathrm{d}h^{-1}(y)}{\mathrm{d}y} \right|. \tag{2.1}$$

Proof. We first assume that $y = h(x)$ is increasing. Thus, $\mathrm{d}h(x)/\mathrm{d}x \geqslant 0$ and $\mathrm{d}h^{-1}(y)/\mathrm{d}y \geqslant 0$. Since

$$\begin{aligned}
G(y) &= \Pr(Y \leqslant y) = \Pr\{h^{-1}(Y) \leqslant h^{-1}(y)\} \\
&= \Pr\{X \leqslant h^{-1}(y)\} = F(h^{-1}(y)),
\end{aligned}$$

by differentiating, we have

$$\begin{aligned} g(y) &= \frac{\mathrm{d}G(y)}{\mathrm{d}y} \\ &= \frac{\mathrm{d}F(h^{-1}(y))}{\mathrm{d}y} \qquad \text{let } x = h^{-1}(y) \\ &= \left.\frac{\mathrm{d}F(x)}{\mathrm{d}x}\right|_{x=h^{-1}(y)} \times \frac{\mathrm{d}x}{\mathrm{d}y} \\ &= f(h^{-1}(y)) \times \frac{\mathrm{d}h^{-1}(y)}{\mathrm{d}y}. \end{aligned}$$

When $y = h(x)$ is decreasing, the proof is similar. □

Example 2.6 (Pareto distribution). Suppose that X has the Pareto density $f(x) = \theta x^{-\theta-1}$, $x \geqslant 1$, $\theta > 0$, find the pdf of $Y = \log(X)$.

Solution. Because $y = \log(x)$ is increasing with inverse $x = \mathrm{e}^y$, we have

$$\begin{aligned} g(y) &= f(x) \times \left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| \\ &= \theta x^{-\theta-1} \cdot \mathrm{e}^y \;=\; \theta\,\mathrm{e}^{-\theta y}, \quad y \geqslant 0. \end{aligned}$$

Thus, Y follows an exponential distribution with mean $1/\theta$. Figure 2.4 shows the density functions of X and Y.

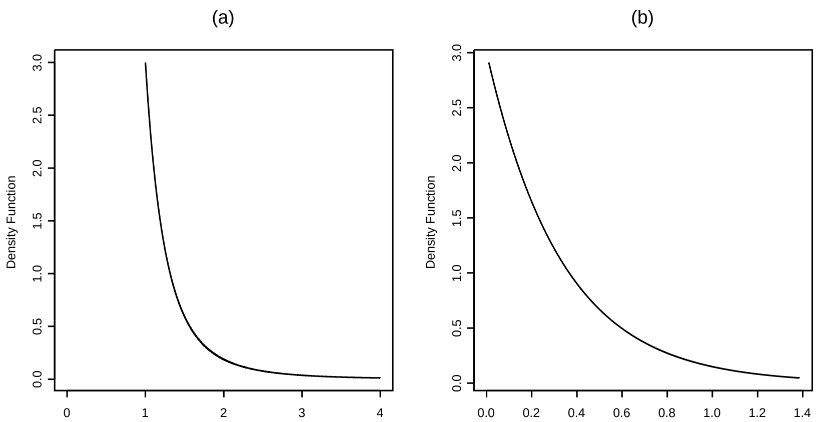

Figure 2.4 (a) The Pareto density $f(x) = \theta x^{-\theta-1} I_{[1,\infty)}(x)$; (b) The density of $Y = \log(X) \sim \text{Exponential}(\theta)$. ‖

5• Piecewise monotone transformation

- Let $\mathbb{A}_1, \ldots, \mathbb{A}_n$ be a partition of the real line $\mathbb{R} = (-\infty, \infty)$, i.e., they are mutually exclusive and $\cup_{i=1}^n \mathbb{A}_i = \mathbb{R}$.
- If $y = h(x)$ is monotone on each $\mathbb{A}_i$, then $h_i(x) \mathrel{\hat{=}} h(x) I_{\mathbb{A}_i}(x)$ has a unique inverse h_i^{-1} on $\mathbb{A}_i$, and the pdf of Y is given by

$$g(y) = \sum_{i=1}^{n} f(h_i^{-1}(y)) \times \left| \frac{\mathrm{d}h_i^{-1}(y)}{\mathrm{d}y} \right|. \tag{2.2}$$

Example 2.7 (Standard normal distribution). Let $X \sim N(0,1)$, find the pdf of $Y = X^2$.

Solution. The function $y = x^2$ is decreasing on $\mathbb{A}_1 = (-\infty, 0]$ and increasing on $\mathbb{A}_2 = (0, \infty)$. For $y \geqslant 0$, the inverse in $\mathbb{A}_1$ is $x = -\sqrt{y}$ and the inverse in $\mathbb{A}_2$ is $x = \sqrt{y}$. We apply (2.2) to get

$$\begin{aligned} g(y) &= \sum_{i=1}^{2} f(h_i^{-1}(y)) \times \left| \frac{\mathrm{d}h_i^{-1}(y)}{\mathrm{d}y} \right| \\ &= \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-y/2} \cdot \frac{y^{-1/2}}{2} + \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-y/2} \cdot \frac{y^{-1/2}}{2} \\ &= \frac{1}{\sqrt{2\pi}} y^{-1/2} \mathrm{e}^{-y/2} \\ &= \frac{(1/2)^{1/2}}{\Gamma(1/2)} y^{-1/2} \mathrm{e}^{-y/2}. \end{aligned}$$

Then, $Y = X^2 \sim \mathrm{Gamma}(1/2, 1/2) = \chi^2(1)$.

Figure 2.5 shows the density functions of the standard normal distribution and the chi-squared distribution with 1 degree of freedom. ‖

6• Bivariate transformation

- Let $(X_1, X_2) \sim f(x_1, x_2)$.
- Let the functions $y_i = h_i(x_1, x_2)$ for $i = 1, 2$ are differentiable and their inverse functions

$$x_i = h_i^{-1}(y_1, y_2) \quad \text{for} \quad i = 1, 2$$

  exist.

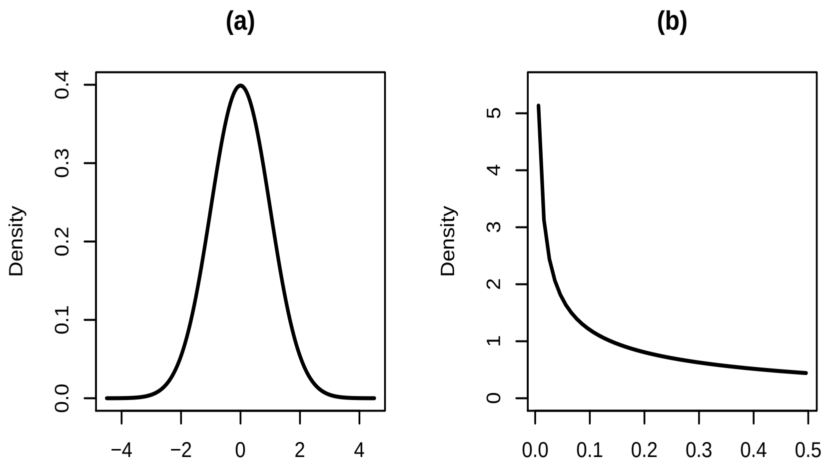

Figure 2.5 (a) The pdf of $X \sim N(0,1)$; (b) The pdf of $Y = X^2 \sim \chi^2(1)$.

- Then, the joint pdf of $Y_1 = h_1(X_1, X_2)$ and $Y_2 = h_2(X_1, X_2)$ is

$$\begin{aligned} g(y_1, y_2) &= f(x_1, x_2) \times |J(x_1, x_2 \to y_1, y_2)| \\ &= f(h_1^{-1}(y_1, y_2), h_2^{-1}(y_1, y_2)) \\ &\quad \times |J(x_1, x_2 \to y_1, y_2)|, \end{aligned} \tag{2.3}$$

  where

$$J(x_1, x_2 \to y_1, y_2) = \left| \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} \right| = \det \begin{pmatrix} \dfrac{\partial x_1}{\partial y_1} & \dfrac{\partial x_1}{\partial y_2} \\ \dfrac{\partial x_2}{\partial y_1} & \dfrac{\partial x_2}{\partial y_2} \end{pmatrix}$$

  denotes the Jacobian determinant of the transformation from (x_1, x_2) to (y_1, y_2).

Example 2.8 (Quotient of two independent normal variables). Let X_1 and X_2 be two independent standard normal random variables. Define

$$Y_1 = X_1 + X_2 \quad \text{and} \quad Y_2 = \frac{X_1}{X_2}.$$

1) Find the joint density of Y_1 and Y_2.

2) Find the marginal density of Y_2.

Solution. 1) From $y_1 = x_1 + x_2$ and $y_2 = x_1/x_2$, we have

$$x_1 = \frac{y_1 y_2}{1+y_2} \quad \text{and} \quad x_2 = \frac{y_1}{1+y_2}.$$

The Jacobian determinant is

$$\begin{aligned}
J(x_1, x_2 \to y_1, y_2) &= \left| \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} \right| \\
&= \det \begin{pmatrix} \dfrac{y_2}{1+y_2} & \dfrac{y_1}{(1+y_2)^2} \\ \dfrac{1}{1+y_2} & -\dfrac{y_1}{(1+y_2)^2} \end{pmatrix} = -\frac{y_1}{(1+y_2)^2}
\end{aligned}$$

so that

$$\begin{aligned}
g(y_1, y_2) &= f(x_1, x_2) \times |J(x_1, x_2 \to y_1, y_2)| \\
&= \frac{1}{2\pi} \exp\left[-\frac{1}{2}\left\{\frac{(y_1 y_2)^2}{(1+y_2)^2} + \frac{y_1^2}{(1+y_2)^2}\right\}\right] \times \frac{|y_1|}{(1+y_2)^2} \\
&= \frac{1}{2\pi} \frac{|y_1|}{(1+y_2)^2} \exp\left[-\frac{1}{2}\left\{\frac{(1+y_2^2) y_1^2}{(1+y_2)^2}\right\}\right].
\end{aligned}$$

2) The marginal density of Y_2 is given by

$$\begin{aligned}
h(y_2) &= \int_{-\infty}^{\infty} g(y_1, y_2)\, \mathrm{d}y_1 \\
&= \frac{1}{2\pi} \frac{1}{(1+y_2)^2} \int_{-\infty}^{\infty} |y_1| \exp\left[-\frac{1}{2}\left\{\frac{(1+y_2^2) y_1^2}{(1+y_2)^2}\right\}\right] \mathrm{d}y_1
\end{aligned}$$

Let

$$u = \frac{1}{2} \frac{(1+y_2^2) y_1^2}{(1+y_2)^2},$$

then $u \geqslant 0$ and

$$\mathrm{d}u = \frac{(1+y_2^2) y_1}{(1+y_2)^2}\, \mathrm{d}y_1,$$

so

$$h(y_2) = \frac{1}{2\pi(1+y_2)^2} \cdot 2 \int_0^{\infty} \mathrm{e}^{-u} \frac{(1+y_2)^2}{(1+y_2^2)}\, \mathrm{d}u = \frac{1}{\pi(1+y_2^2)},$$

which is a Cauchy density. ‖

Example 2.9 (Uniform distribution on the unit square). Let

$$(X_1, X_2)^\top \sim f(x_1, x_2) = 1, \quad 0 \leqslant x_1 \leqslant 1, \quad 0 \leqslant x_2 \leqslant 1,$$

1) Find the joint pdf of $Y = X_1 + X_2$ and $Z = X_2$.

2) Find the marginal density of Y.

Solution. 1) Make the transformation $y = x_1 + x_2$ and $z = x_2$, where

$$(x_1, x_2) \in \mathcal{S}_{(X_1,X_2)} = \{(x_1, x_2)\colon\ 0 \leqslant x_i \leqslant 1,\ i = 1, 2\},$$

then the corresponding inverse transformation is given by $x_1 = y - z$ and $x_2 = z$, where

$$(y, z) \in \mathcal{S}_{(Y,Z)} = \{(y, z)\colon\ z \leqslant y \leqslant z + 1,\ 0 \leqslant z \leqslant 1\}.$$

Figure 2.6 shows the two regions.

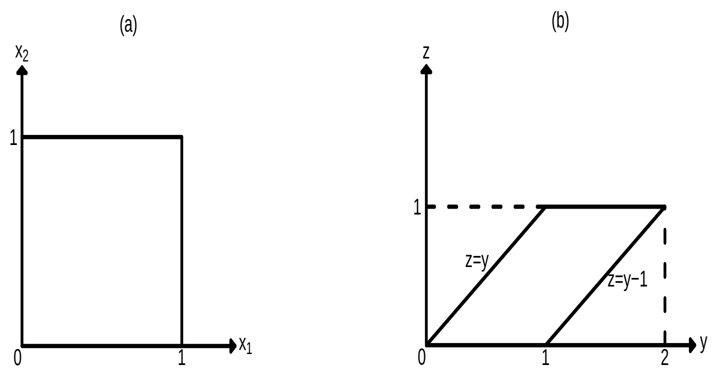

Figure 2.6 (a) $\mathcal{S}_{(X_1,X_2)} = \{(x_1, x_2)\colon\ 0 \leqslant x_i \leqslant 1,\ i = 1, 2\}$; (b) $\mathcal{S}_{(Y,Z)} = \{(y, z)\colon\ z \leqslant y \leqslant z + 1,\ 0 \leqslant z \leqslant 1\}$.

Hence, we have

$$J(x_1, x_2 \to y, z) = \det \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = 1.$$

By using (2.3), we obtain the joint pdf of (Y, Z) as

$$g(y,z) = f(x_1, x_2) \times |J(x_1, x_2 \to y, z)| = 1 \cdot I_{\mathcal{S}_{(Y,Z)}}(y,z);$$

that is, $(Y, Z)^\top \sim U(\mathcal{S}_{(Y,Z)})$.

2) The marginal density of Y is given by

$$\begin{aligned} g(y) &= \int_{-\infty}^{\infty} g(y,z)\,\mathrm{d}z \\ &= \begin{cases} \int_0^y \mathrm{d}z, & \text{if } 0 \leqslant y \leqslant 1 \\ \int_{y-1}^1 \mathrm{d}z, & \text{if } 1 < y \leqslant 2 \\ 0, & \text{otherwise} \end{cases} \\ &= \begin{cases} y, & \text{if } 0 \leqslant y \leqslant 1 \\ 2-y, & \text{if } 1 < y \leqslant 2 \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

Figure 2.7 shows this density function. The key point for the transformation technique is to determine the image domain $\mathcal{S}_{(Y,Z)}$.

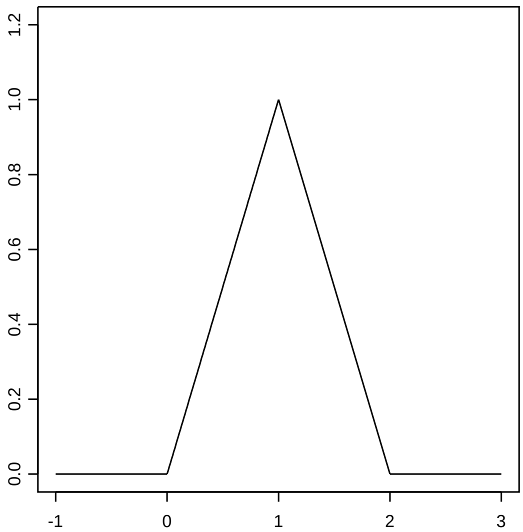

Figure 2.7 The density function of $Y = X_1 + X_2$, where $X_1, X_2 \overset{\text{iid}}{\sim} U[0,1]$. $\|$

7• Multivariate transformation

- Let $(X_1, \ldots, X_n)^\top \sim f(x_1, \ldots, x_n)$.
- If the functions $y_i = h_i(x_1, \ldots, x_n)$ for $i = 1, \ldots, n$ are differentiable, then the joint pdf of $Y_i = h_i(X_1, \ldots, X_n)$ for $i = 1, \ldots, n$ is given by

$$g(y_1, \ldots, y_n) = f(x_1, \ldots, x_n) \times |J(x_1, \ldots, x_n \to y_1, \ldots, y_n)|. \tag{2.4}$$

Example 2.10 (Multivariate t-distribution). Let $Z \sim \chi^2(\nu)$, $Z \perp\!\!\!\perp \mathbf{y}$, and $\mathbf{y} = (Y_1, \ldots, Y_d)^\top \sim N_d(\mathbf{0}, \mathbf{\Sigma})$. Define

$$X_i = \mu_i + \frac{Y_i}{\sqrt{Z/\nu}}, \quad i = 1, \ldots, d, \tag{2.5}$$

then $\mathbf{x} = (X_1, \ldots, X_d)^\top$ is said to follow a d-dimensional t-distribution with location parameter vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)^\top \in \mathbb{R}^d$, dispersion matrix $\mathbf{\Sigma} > 0$ and degree of freedom $\nu > 0$, denoted by $\mathbf{x} \sim t_d(\boldsymbol{\mu}, \mathbf{\Sigma}, \nu)$.

1) Find the joint density of $\mathbf{x}$ and Z.

2) Find the joint density of $\mathbf{x}$.

3) Find the marginal density of X_i for $i = 1, \ldots, d$.

4) When $\mathbf{\Sigma} = \mathbf{I}_d$, are X_i and X_j $(i \neq j)$ independent?

Solution. 1) Making the following transformation

$$\begin{cases} x_i &= \mu_i + \dfrac{y_i}{\sqrt{z/\nu}}, \quad i = 1, \ldots, d, \\ z &= z, \end{cases}$$

we have

$$\begin{cases} y_i &= \sqrt{z/\nu}\,(x_i - \mu_i), \quad i = 1, \ldots, d, \\ z &= z, \end{cases}$$

or

$$\begin{cases} \boldsymbol{y} &= (y_1, \ldots, y_d)^\top = \sqrt{z/\nu}\,(\boldsymbol{x} - \boldsymbol{\mu}), \\ z &= z, \end{cases}$$

where $\boldsymbol{x} = (x_1, \ldots, x_d)^\top \in \mathbb{R}^d$ and $z > 0$. The Jacobian determinant is

$$\begin{aligned}
& J(y_1, \ldots, y_d, z \to x_1, \ldots, x_d, z) \\
= \; & \det \begin{pmatrix} \dfrac{\partial y_1}{\partial x_1} & \cdots & \dfrac{\partial y_1}{\partial x_d} & \dfrac{\partial y_1}{\partial z} \\ \vdots & \ddots & \vdots & \vdots \\ \dfrac{\partial y_d}{\partial x_1} & \cdots & \dfrac{\partial y_d}{\partial x_d} & \dfrac{\partial y_d}{\partial z} \\ \dfrac{\partial z}{\partial x_1} & \cdots & \dfrac{\partial z}{\partial x_d} & \dfrac{\partial z}{\partial z} \end{pmatrix} \\
= \; & \det \begin{pmatrix} \sqrt{z/\nu} & 0 & \cdots & 0 & 0.5(x_1 - \mu_1)/\sqrt{z/\nu} \\ 0 & \sqrt{z/\nu} & \cdots & 0 & 0.5(x_2 - \mu_2)/\sqrt{z/\nu} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \sqrt{z/\nu} & 0.5(x_d - \mu_d)/\sqrt{z/\nu} \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix} \\
= \; & (z/\nu)^{d/2}.
\end{aligned}$$

Thus, the joint pdf of $\mathbf{x}$ and Z is

$$\begin{aligned}
& f(x_1, \ldots, x_d, z) \\
= \; & f(y_1, \ldots, y_d, z) \times |J(y_1, \ldots, y_d, z \to x_1, \ldots, x_d, z)| \\
= \; & f(y_1, \ldots, y_d) \times f(z) \times (z/\nu)^{d/2} \\
= \; & N_d(\boldsymbol{y}|\mathbf{0}, \boldsymbol{\Sigma}) \times \chi^2(z|\nu) \times (z/\nu)^{d/2} \\
= \; & \frac{1}{(\sqrt{2\pi}\,)^d |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \boldsymbol{y}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{y} \right) \times \frac{2^{-\nu/2}}{\Gamma(\nu/2)} z^{\frac{\nu}{2}-1}\, \mathrm{e}^{-z/2} \times (z/\nu)^{\frac{d}{2}} \\
= \; & c \cdot \exp\left\{ -\frac{z}{2\nu} (\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right\} \times z^{\frac{\nu+d}{2}-1}\, \mathrm{e}^{-z/2} \\
= \; & c \cdot z^{\frac{\nu+d}{2}-1} \exp\left[-z \left\{ \frac{1}{2} + \frac{(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})}{2\nu} \right\} \right],
\end{aligned}$$

where $\boldsymbol{x} \in \mathbb{R}^d$, $z > 0$ and

$$c = \frac{2^{-\frac{\nu}{2}}}{(2\pi\nu)^{\frac{d}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}} \Gamma(\frac{\nu}{2})} = \frac{1}{2^{\frac{\nu+d}{2}} \Gamma(\frac{\nu}{2}) (\sqrt{\pi\nu}\,)^d |\boldsymbol{\Sigma}|^{\frac{1}{2}}}.$$

2) By using (1.41) in Chapter 1, we obtain the joint pdf of $\mathbf{x}$ given by

$$\begin{aligned}
& f(x_1,\ldots,x_d) \\
= & \int_0^\infty f(x_1,\ldots,x_d,z)\,\mathrm{d}z \\
= & c\cdot\int_0^\infty z^{\frac{\nu+d}{2}-1}\exp\left[-z\left\{\frac{1}{2}+\frac{(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{2\nu}\right\}\right]\mathrm{d}z \\
\overset{(1.41)}{=} & c\cdot\frac{\Gamma(\frac{\nu+d}{2})}{\left\{\frac{1}{2}+\frac{(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{2\nu}\right\}^{\frac{\nu+d}{2}}} \\
= & \frac{\Gamma(\frac{\nu+d}{2})}{\Gamma(\frac{\nu}{2})(\sqrt{\pi\nu}\,)^d|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\left\{1+\frac{(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{\nu}\right\}^{-\frac{\nu+d}{2}}, \quad \boldsymbol{x}\in\mathbb{R}^d,
\end{aligned}$$

which is the density of d-dimensional t-distribution.

3) In particular, let $d=1$ and denote $\boldsymbol{\Sigma}$ by σ^2. The density of X_1 is

$$\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}\,\sigma}\left\{1+\frac{(x_1-\mu)^2}{\nu\sigma^2}\right\}^{-\frac{\nu+1}{2}}, \quad x_1\in\mathbb{R},$$

which is the density of the univariate t-distribution with location parameter $\mu\in\mathbb{R}$, dispersion parameter $\sigma^2>0$ and degree of freedom $\nu>0$. We denote it by $X_1\sim t(\mu,\sigma^2,\nu)$.

4) When $d=2$ and $\boldsymbol{\Sigma}=\mathbf{I}_2$, it is easy to show that

$$f_{(X_1,X_2)}(x_1,x_2)\neq f_{X_1}(x_1)\times f_{X_2}(x_2),$$

So X_1 and X_2 are not independent. From (2.5), it is clear that X_i and X_j $(i\neq j)$ share a common r.v. Z, so they are not independent. ‖

2.1.3 Moment generating function technique

8• The procedure of mgf

- Let $Y=\sum_{i=1}^n X_i$.
- If $\{X_i\}_{i=1}^n$ are independent r.v.'s, then

$$M_Y(t)=\prod_{i=1}^n M_{X_i}(t). \tag{2.6}$$

Example 2.11 (Sum of independent binomial r.v.'s with a common p). Let $\{X_i\}_{i=1}^n$ be independent r.v.'s and $X_i \sim \text{Binomial}(m_i, p)$ for $i = 1, \ldots, n$, find the distribution of $Y = \sum_{i=1}^n X_i$.

Solution. From (2.6) and Table 1.2, we have

$$M_Y(t) = \prod_{i=1}^n M_{X_i}(t) = \prod_{i=1}^n (p\,\mathrm{e}^t + 1 - p)^{m_i} = (p\,\mathrm{e}^t + 1 - p)^{\sum_{i=1}^n m_i},$$

indicating that $\sum_{i=1}^n X_i \sim \text{Binomial}(\sum_{i=1}^n m_i, p)$. This result means that binomial distribution is additive. ‖

Example 2.12 (Sum of independent Poisson r.v.'s). Let $\{X_i\}_{i=1}^n$ be independent r.v.'s and $X_i \sim \text{Poisson}(\lambda_i)$ for $i = 1, \ldots, n$, find the distribution of $Y = \sum_{i=1}^n X_i$.

Solution. From (2.6) and Table 1.2, we have

$$M_Y(t) = \prod_{i=1}^n M_{X_i}(t) = \prod_{i=1}^n \exp\{\lambda_i(\mathrm{e}^t - 1)\} = \exp\left\{\sum_{i=1}^n \lambda_i(\mathrm{e}^t - 1)\right\},$$

which means $\sum_{i=1}^n X_i \sim \text{Poisson}(\sum_{i=1}^n \lambda_i)$; i.e., Poisson distribution is also additive. This result is a generalization of the result in Example 2.5. ‖

Example 2.13 (Sum of independent chi-squared r.v.'s). Let $\{X_i\}_{i=1}^n$ be independent r.v.'s and $X_i \sim \chi^2(m_i)$ for $i = 1, \ldots, n$, find the distribution of $Y = \sum_{i=1}^n X_i$.

Solution. Note that $\chi^2(m) = \text{Gamma}(\frac{m}{2}, \frac{1}{2})$. From (2.6) and Table 1.3, we have

$$\begin{aligned} M_Y(t) &= \prod_{i=1}^n M_{X_i}(t) \\ &= \prod_{i=1}^n \left(\frac{0.5}{0.5 - t}\right)^{m_i/2} \\ &= \left(\frac{0.5}{0.5 - t}\right)^{\sum_{i=1}^n m_i/2}, \end{aligned}$$

which means $\sum_{i=1}^n X_i \sim \chi^2(\sum_{i=1}^n m_i)$. ‖

2.2 Statistics, Sample Mean and Sample Variance

9• What is a random sample?

- Let $F(x)$ be the cdf of an r.v. X.
- If $\{X_i\}_{i=1}^n \stackrel{\text{iid}}{\sim} F(x)$, then $\{X_i\}_{i=1}^n$ is said to be a *random sample* of X, or $\{X_i\}_{i=1}^n$ is a random sample from $F(x)$.

10• What is a statistic?

Definition 2.1 (Function of random variables). Let $\{X_i\}_{i=1}^n \stackrel{\text{iid}}{\sim} F(x)$. An arbitrary function $T(X_1, \ldots, X_n)$ of $\{X_i\}_{i=1}^n$ is called a *statistic*. ‖

10.1• The sample mean and sample variance

— For example,

$$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i \quad \text{and} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2 \tag{2.7}$$

are two statistics.

— They are called the sample mean and sample variance, respectively.

2.2.1 Distribution of the sample mean

11• Basic properties of the sample mean

- Let $\{X_i\}_{i=1}^n \stackrel{\text{iid}}{\sim} F(x)$ with $E(X_1) = \mu$ and $\text{Var}(X_1) = \sigma^2$.
- For any $F(x)$, we have $E(\bar{X}) = \mu$ and $\text{Var}(\bar{X}) = \sigma^2/n$.
- If $F(x)$ is the cdf of the normal distribution $N(\mu, \sigma^2)$, then

$$\bar{X} \sim N(\mu, \sigma^2/n). \tag{2.8}$$

Proof. In fact, by the mgf technique, we have

$$\begin{aligned} M_{\bar{X}}(t) &= M_{\sum_{i=1}^n X_i/n}(t) = \prod_{i=1}^n M_{X_i/n}(t) = \prod_{i=1}^n M_{X_i}\left(\frac{t}{n}\right) \\ &= \left\{M_{X_1}\left(\frac{t}{n}\right)\right\}^n = \left\{\exp\left(\mu\frac{t}{n} + 0.5\sigma^2\frac{t^2}{n^2}\right)\right\}^n \end{aligned}$$

$$= \exp\left\{\mu t + 0.5\left(\frac{\sigma^2}{n}\right)t^2\right\},$$

indicating that $\bar{X} \sim N(\mu, \sigma^2/n)$. □

2.2.2 Distribution of the sample variance

To prove (2.10) below, we need the following theorem with proof given in Section 2.6.

Theorem 2.1 (Linear combination of normal components). Let $\mathbf{A}_{m\times n}$ and $\mathbf{B}_{r\times n}$ be two scalar matrices and $\mathbf{x} = (X_1, \dots, X_n)^\top \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then

1) $\mathbf{Ax} \sim N_m(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$.

2) $\mathbf{Bx} \sim N_r(\mathbf{B}\boldsymbol{\mu}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^\top)$.

3) $\mathbf{Ax} \perp\!\!\!\perp \mathbf{Bx}$ iff $\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}^\top = \mathbf{O}_{m\times r}$. ‖

12• Basic properties of the sample variance

- Let $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} F(x)$ with $E(X_1) = \mu$ and $\mathrm{Var}(X_1) = \sigma^2$.
- For any $F(x)$, the sample variance is an unbiased estimator of the variance, i.e.,
$$E(S^2) = \sigma^2. \tag{2.9}$$

  Proof. Since

$$(n-1)S^2 = \sum_{i=1}^n [X_i - \mu - (\bar{X} - \mu)]^2 = \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2,$$

  taking expectation on both sides, we have

$$(n-1)E(S^2) = n\sigma^2 - n \cdot \frac{\sigma^2}{n},$$

  which means (2.9). □

- If $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$, then
$$S^2 \perp\!\!\!\perp \bar{X} \quad \text{and} \quad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1). \tag{2.10}$$

Proof. Define $\mathbf{Q}_n = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top$, it is easy to show that

$$\mathbf{Q}_n = \mathbf{Q}_n^\top = \mathbf{Q}_n^2 \quad \text{and} \quad \mathbf{Q}_n\mathbf{1}_n = \mathbf{0}_n. \tag{2.11}$$

Let $\mathbf{x} = (X_1, \ldots, X_n)^\top$, then $\mathbf{x} \sim N_n(\mu\mathbf{1}_n, \sigma^2\mathbf{I}_n)$. From the result 1) of Theorem 2.1 and (2.11), we have

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\mathbf{1}_n^\top\mathbf{x} \sim N(\mu, \sigma^2/n)$$

and

$$\begin{pmatrix} X_1 - \bar{X} \\ \vdots \\ X_n - \bar{X} \end{pmatrix} = \mathbf{x} - \bar{X}\mathbf{1}_n = \mathbf{x} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\mathbf{x} = \mathbf{Q}_n\mathbf{x} \sim N_n(\mathbf{0}, \sigma^2\mathbf{Q}_n).$$

Note that $\mathbf{Q}_n \cdot \sigma^2\mathbf{I}_n \cdot \mathbf{1}_n = \mathbf{0}$, by the result 3) of Theorem 2.1, we can conclude that $\mathbf{Q}_n\mathbf{x} \perp\!\!\!\perp \mathbf{1}_n^\top\mathbf{x}$. Since

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}(\mathbf{Q}_n\mathbf{x})^\top\mathbf{Q}_n\mathbf{x}$$

is a function of $\mathbf{Q}_n\mathbf{x}$ and $\bar{X}$ is a function of $\mathbf{1}_n^\top\mathbf{x}$, we have $S^2 \perp\!\!\!\perp \bar{X}$.

Since

$$\begin{aligned} \sum_{i=1}^{n}(X_i - \mu)^2 &= \sum_{i=1}^{n}(X_i - \bar{X} + \bar{X} - \mu)^2 \\ &= \sum_{i=1}^{n}(X_i - \bar{X})^2 + n(\bar{X} - \mu)^2, \end{aligned}$$

we have

$$W \mathrel{\hat{=}} \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 = \frac{(n-1)S^2}{\sigma^2} + \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)^2 \mathrel{\hat{=}} U + V,$$

where $W \sim \chi^2(n)$, $V \sim \chi^2(1)$, and $U \perp\!\!\!\perp V$. Then

$$M_W(t) = M_U(t) \cdot M_V(t),$$

or

$$(1 - 2t)^{-n/2} = M_U(t) \cdot (1 - 2t)^{-1/2}.$$

Hence

$$M_U(t) = (1 - 2t)^{-(n-1)/2}.$$

This implies that $U = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$. □

2.3 The t and F Distributions

2.3.1 The t distribution

13• DEFINITION

- Let $Y \sim \chi^2(n)$, $Z \sim N(0,1)$ and $Y \perp\!\!\!\perp Z$.
- The distribution of
$$T = \frac{Z}{\sqrt{Y/n}} \tag{2.12}$$
is called the t distribution with n degrees of freedom and is written as $T \sim t(n)$.

13.1• The name of the t distribution

— The t distribution was introduced originally by W. S. Gosset, who published his scientific writings under the pen name "Student" since the company for which he worked, a brewery, did not permit publication by employees.

— Thus, the t distribution is also known as the *Student t distribution*, or *Student's t distribution*.

Theorem 2.2 (Density of the t distribution). The density of $T \sim t(n)$ is given by

$$f(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\,\Gamma(\frac{n}{2})}\left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < t < \infty. \qquad \|$$

Proof. Let $\phi(z) = \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-z^2/2}$ denote the pdf of $Z \sim N(0,1)$ and $g(y)$ denote the pdf of $Y \sim \chi^2(n)$. The cdf of T is

$$\begin{aligned}
F(x) &= \Pr(T \leqslant x) = \Pr\left(\frac{Z}{\sqrt{Y/n}} \leqslant x\right) \\
&\overset{(1.33)}{=} \int \Pr\left(\frac{Z}{\sqrt{Y/n}} \leqslant x \,\middle|\, Y = y\right) \cdot g(y)\,\mathrm{d}y \\
&= \int_0^\infty \Pr\left(Z \leqslant x\sqrt{y/n}\right) \cdot g(y)\,\mathrm{d}y \\
&= \int_0^\infty \left\{\int_{-\infty}^{x\sqrt{y/n}} \phi(z)\,\mathrm{d}z\right\} \cdot g(y)\,\mathrm{d}y.
\end{aligned}$$

Let $t = \frac{z}{\sqrt{y/n}}$, then $-\infty < t \leqslant x$, $\mathrm{d}z = \sqrt{y/n}\,\mathrm{d}t$, and $F(x)$ becomes

$$\begin{aligned} F(x) &= \int_0^\infty \left\{ \int_{-\infty}^x \phi\Big(t\sqrt{y/n}\Big) \cdot \sqrt{y/n}\,\mathrm{d}t \right\} \cdot g(y)\,\mathrm{d}y \\ &= \int_{-\infty}^x \left\{ \int_0^\infty \phi\Big(t\sqrt{y/n}\Big) \cdot \sqrt{y/n} \cdot g(y)\,\mathrm{d}y \right\} \mathrm{d}t \\ &= \int_{-\infty}^x f(t)\,\mathrm{d}t. \end{aligned}$$

Hence, the density of T is given by

$$\begin{aligned} f(t) &= \int_0^\infty \phi\Big(t\sqrt{y/n}\Big) \cdot \sqrt{y/n} \cdot g(y)\,\mathrm{d}y \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-t^2 y/(2n)} \cdot \sqrt{y/n} \cdot \frac{(1/2)^{n/2}}{\Gamma(n/2)} y^{\frac{n}{2}-1}\,\mathrm{e}^{-y/2}\,\mathrm{d}y \\ &= \frac{1}{\sqrt{2\pi n}} \cdot \frac{(1/2)^{n/2}}{\Gamma(n/2)} \cdot \int_0^\infty y^{\frac{n+1}{2}-1}\,\mathrm{e}^{-y(\frac{1}{2}+\frac{t^2}{2n})}\,\mathrm{d}y \\ &\overset{(1.39)}{=} \frac{(1/2)^{(n+1)/2}}{\sqrt{\pi n}\,\Gamma(n/2)} \cdot \frac{\Gamma(\frac{n+1}{2})}{(\frac{1}{2}+\frac{t^2}{2n})^{\frac{n+1}{2}}} \\ &= \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\,\Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}. \end{aligned}$$

This completes the proof of Theorem 2.2. □

13.2• The usefulness of the t distribution

— The t distribution is an important distribution in statistical inferences on the mean of the normal population.

— Figure 2.8 compares the $t(4)$ density with the standard normal density.

— Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$. From (2.8), we obtain

$$\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim N(0,1). \tag{2.13}$$

— By using (2.10), we have

$$T = \frac{\sqrt{n}(\bar{X}-\mu)/\sigma}{\sqrt{\frac{(n-1)S^2}{\sigma^2}\Big/(n-1)}} = \frac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t(n-1). \tag{2.14}$$

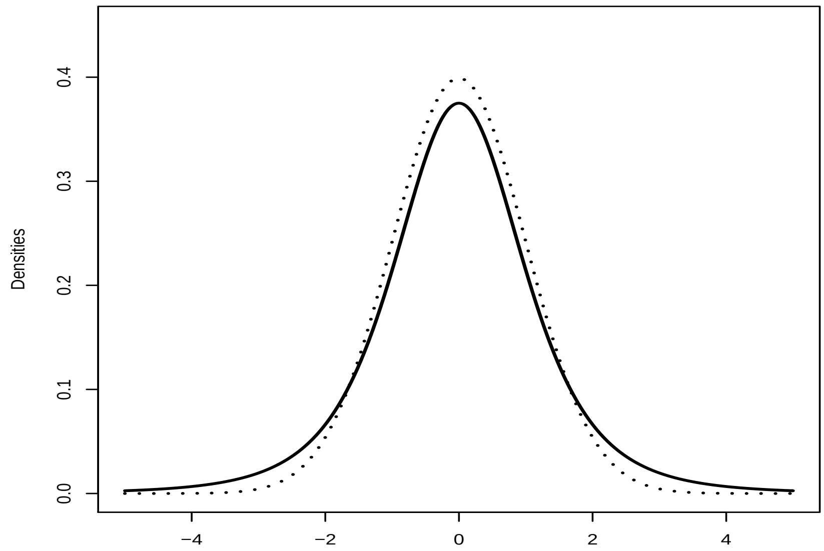

Figure 2.8 The comparison between the $t(4)$ density (solid curve) and the standard normal density (dotted curve).

2.3.2 The *F* distribution

14• Definition

- Let $U \sim \chi^2(m)$, $V \sim \chi^2(n)$ and $U \perp\!\!\!\perp V$.
- The distribution of the r.v.

$$W = \frac{U/m}{V/n} \tag{2.15}$$

  is said to have an F distribution with m and n degrees of freedom. We write $W \sim F(m, n)$.

14.1• The name of the F distribution

— Besides the t distribution, another distribution that plays an important role in connection with sampling from normal populations is the F distribution, named after Sir Ronald A. Fisher, one of the most prominent statisticians of the last century.

— The F distribution is also known as Snedecor's F distribution (after George W. Snedecor) or the Fisher–Snedecor distribution.

Theorem 2.3 (Density of the F distribution). The density of $W \sim F(m,n)$ is given by

$$f(w) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})}\left(\frac{m}{n}\right)^{m/2} w^{m/2-1}\left(1+\frac{m}{n}w\right)^{-\frac{m+n}{2}}, \quad w > 0. \qquad \|$$

Proof. Let $h(u)$ and $g(v)$ denote the densities of $U \sim \chi^2(m)$ and $V \sim \chi^2(n)$, respectively. Since $U \perp\!\!\!\perp V$, the cdf of W is

$$\begin{aligned} F(x) &= \Pr(W \leqslant x) = \Pr\left(\frac{U/m}{V/n} \leqslant x\right) \\ &= \int \Pr\left(\frac{U/m}{V/n} \leqslant x \middle| V = v\right) \cdot g(v)\,\mathrm{d}v \\ &= \int_0^\infty \Pr\Big(U \leqslant xvm/n\Big) \cdot g(v)\,\mathrm{d}v \\ &= \int_0^\infty \left\{\int_0^{xvm/n} h(u)\,\mathrm{d}u\right\} \cdot g(v)\,\mathrm{d}v. \end{aligned}$$

Let $w = \frac{u/m}{v/n}$, then $0 \leqslant w \leqslant x$, $\mathrm{d}u = \frac{mv}{n}\,\mathrm{d}w$, and $F(x)$ becomes

$$\begin{aligned} F(x) &= \int_0^\infty \left\{\int_0^x h\Big(\frac{mv}{n}w\Big) \cdot \frac{mv}{n}\,\mathrm{d}w\right\} \cdot g(v)\,\mathrm{d}v \\ &= \int_0^x \left\{\int_0^\infty h\Big(\frac{mv}{n}w\Big) \cdot \frac{mv}{n} \cdot g(v)\,\mathrm{d}v\right\}\mathrm{d}w = \int_0^x f(w)\,\mathrm{d}w. \end{aligned}$$

Hence, the density of W is given by

$$\begin{aligned} f(w) &= \int_0^\infty h\Big(\frac{mv}{n}w\Big) \cdot \frac{mv}{n} \cdot g(v)\,\mathrm{d}v \\ &= \int_0^\infty \frac{(\frac{1}{2})^{m/2}}{\Gamma(\frac{m}{2})}\Big(\frac{mv}{n}w\Big)^{\frac{m}{2}-1} \mathrm{e}^{-\frac{mvw}{2n}} \cdot \frac{mv}{n} \cdot \frac{(\frac{1}{2})^{n/2}}{\Gamma(\frac{n}{2})} v^{\frac{n}{2}-1}\,\mathrm{e}^{-v/2}\,\mathrm{d}v \\ &= \frac{(\frac{1}{2})^{(m+n)/2}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})}\left(\frac{m}{n}\right)^{m/2} w^{m/2-1} \cdot \int_0^\infty v^{\frac{m+n}{2}-1}\,\mathrm{e}^{-v(\frac{1}{2}+\frac{mw}{2n})}\,\mathrm{d}v \\ &\overset{(1.39)}{=} \frac{(\frac{1}{2})^{(m+n)/2}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})}\left(\frac{m}{n}\right)^{m/2} w^{m/2-1} \cdot \frac{\Gamma(\frac{m+n}{2})}{(\frac{1}{2}+\frac{mw}{2n})^{\frac{m+n}{2}}} \\ &= \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})}\left(\frac{m}{n}\right)^{m/2} w^{m/2-1}\left(1+\frac{m}{n}w\right)^{-\frac{m+n}{2}}. \end{aligned}$$

This completes the proof of Theorem 2.3. $\square$

Theorem 2.4 (Ratio of two normal sample variances). If S_1^2 and S_2^2 are the sample variances of independent random samples of size n_1 and n_2 from normal populations $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively, then

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} \sim F(n_1 - 1, n_2 - 1). \qquad \|$$

Proof. Note that

$$\frac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2(n_1 - 1) \quad \text{and} \quad \frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2(n_2 - 1)$$

are independent, then

$$F = \frac{\frac{(n_1-1)S_1^2}{\sigma_1^2}/(n_1-1)}{\frac{(n_2-1)S_2^2}{\sigma_2^2}/(n_2-1)} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1). \qquad \square$$

14.2• The usefulness of the F distribution

— If $X \sim F(m, n)$, then $Y = 1/X \sim F(n, m)$.

— The densities of $F(m, n)$ with various degrees of freedom are shown in Figure 2.9.

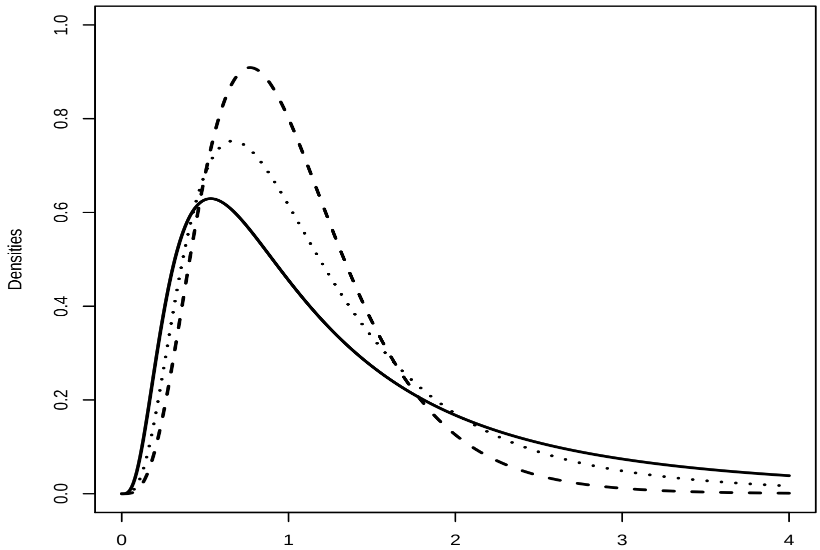

Figure 2.9 Plots of the densities of $W \sim F(m, n)$ with $m = 10$ and $n = 4$ (solid curve), $n = 10$ (dotted curve), $n = 50$ (broken curve).

2.4 Order Statistics

15• Definition

- Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} F(\cdot)$, and $f(\cdot)$ is the pdf.
- Let

  — $X_{(1)} = \min(X_1, \ldots, X_n)$ be the smallest of $X_1, \ldots, X_n$;

  — $X_{(2)}$ be the second smallest of $X_1, \ldots, X_n$;

  $\vdots$

  — $X_{(n)} = \max(X_1, \ldots, X_n)$ be the largest of $X_1, \ldots, X_n$.
- Then $X_{(1)}, \ldots, X_{(n)}$ are called the *order statistics* and $X_{(r)}$ is called the r-th *order statistic* for $r = 1, \ldots, n$.
- We use $x_{(1)}, \ldots, x_{(n)}$ to denote the realizations of $X_{(1)}, \ldots, X_{(n)}$.

15.1• An example

— Let $\{x_1, \ldots, x_5\} = \{2, 5, -1, 0, 6\}$, then we have $x_{(1)} = -1$, $x_{(2)} = 0$, $x_{(3)} = 2$, $x_{(4)} = 5$, and $x_{(5)} = 6$.

15.2• Remarks

— The $X_{(r)}$'s are statistics since they are functions of the random sample $X_1, \ldots, X_n$ and are in order.

— Unlike the random sample themselves, the order statistics are clearly *not independent*, because if $X_{(r)} \geqslant x$, then $X_{(r+1)} \geqslant x$.

2.4.1 Distribution of a single order statistic

16• The distribution of the largest order statistic

- Let $G_r(x)$ denote the cdf of the r-th order statistic $X_{(r)}$.
- Then the cdf of the largest order statistic $X_{(n)}$ is

$$\begin{aligned} G_n(x) &= \Pr\{\max(X_1, \ldots, X_n) \leqslant x\} \\ &= \Pr(X_1 \leqslant x, \ldots, X_n \leqslant x) = F^n(x). \end{aligned} \tag{2.16}$$

- The pdf of $X_{(n)}$ is

$$g_n(x) = \frac{\mathrm{d}G_n(x)}{\mathrm{d}x} = nf(x)F^{n-1}(x). \tag{2.17}$$

17• The distribution of the smallest order statistic

- Similarly, we have

$$\begin{aligned} G_1(x) &= \Pr(X_{(1)} \leqslant x) \\ &= 1 - \Pr\{\min(X_1, \ldots, X_n) > x\} \\ &= 1 - \Pr(X_1 > x, \ldots, X_n > x) \\ &= 1 - \{1 - F(x)\}^n. \end{aligned} \tag{2.18}$$

- The pdf of $X_{(1)}$ is

$$g_1(x) = \frac{\mathrm{d}G_1(x)}{\mathrm{d}x} = nf(x)\{1 - F(x)\}^{n-1}. \tag{2.19}$$

18• The distribution of the r-th order statistic

18.1• The cdf of $X_{(r)}$

— Let $G_r(x)$ denote the cdf of $X_{(r)}$, then

$$G_r(x) = \frac{1}{B(r, n-r+1)} \int_0^{F(x)} t^{r-1}(1-t)^{n-r}\,\mathrm{d}t. \tag{2.20}$$

Proof. The formulae (2.16) and (2.18) are important special cases of the general result:

$$\begin{aligned} G_r(x) &= \Pr(X_{(r)} \leqslant x) \\ &= \Pr(\text{at least } r \text{ of } X_1, \ldots, X_n \leqslant x) \\ &= \sum_{i=r}^{n} \Pr(\text{exact } i \text{ of } X_1, \ldots, X_n \leqslant x) \\ &= \sum_{i=r}^{n} \binom{n}{i} \Pr(X_1, \ldots, X_i \leqslant x) \cdot \Pr(X_{i+1}, \ldots, X_n > x) \\ &= \sum_{i=r}^{n} \binom{n}{i} F^i(x)\{1 - F(x)\}^{n-i}. \end{aligned} \tag{2.21}$$

By using the identity

$$\sum_{i=r}^{n} \binom{n}{i} p^i (1-p)^{n-i} = \frac{1}{B(r, n-r+1)} \int_0^p t^{r-1}(1-t)^{n-r}\,\mathrm{d}t \qquad (2.22)$$

for any $p \in [0,1]$, we can rewrite (2.21) into (2.20) and hence complete the proof. $\square$

18.2• Proof of (2.22)

— Let $f(p)$ denote the left-hand side of (2.22), we have

$$\begin{aligned} f'(p) &= \sum_{i=r}^{n} \binom{n}{i} \left\{ i p^{i-1}(1-p)^{n-i} - (n-i)p^i(1-p)^{n-i-1} \right\} \\ &= \sum_{i=r}^{n} \frac{n!}{i!(n-i)!} \left\{ i p^{i-1}(1-p)^{n-i} - (n-i)p^i(1-p)^{n-i-1} \right\} \\ &= \sum_{i=r}^{n} \frac{n! p^{i-1}(1-p)^{n-i}}{(i-1)!(n-i)!} - \sum_{i=r}^{n} \frac{n! p^i (1-p)^{n-i-1}}{i!(n-i-1)!} \\ &= \frac{n!}{(n-r)!(r-1)!} p^{r-1}(1-p)^{n-r} \end{aligned}$$

— Let $g(p)$ denote the right-hand side of (2.22), we obtain

$$\begin{aligned} g'(p) &= \frac{1}{B(r, n-r+1)} p^{r-1}(1-p)^{n-r} \\ &= \frac{n!}{(r-1)!(n-r)!} p^{r-1}(1-p)^{n-r}, \end{aligned}$$

so that $f'(p) = g'(p)$.

— This implies $f(p) = g(p) + c$ for any $p \in [0,1]$, where c is a constant.

— In particular, let $p = 0$, we have

$$c = f(0) - g(0) = 0.$$

Thus $f(p) = g(p)$. $\square$

18.3• The pdf of $X_{(r)}$

— Let $g_r(x)$ denote the pdf of $X_{(r)}$, from (2.20), we obtain

$$\begin{aligned} g_r(x) &= \frac{\mathrm{d}}{\mathrm{d}x} G_r(x) \\ &= \frac{1}{B(r, n-r+1)} \cdot \frac{\mathrm{d}}{\mathrm{d}x} \int_0^{F(x)} t^{r-1}(1-t)^{n-r}\,\mathrm{d}t \\ &= \frac{n!}{(r-1)!(n-r)!} f(x) F^{r-1}(x)\{1-F(x)\}^{n-r}. \end{aligned} \tag{2.23}$$

— In (2.23), we utilized the following formula:

$$\frac{\mathrm{d}}{\mathrm{d}x} \int_0^{A(x)} g(t)\,\mathrm{d}t = \frac{\mathrm{d}}{\mathrm{d}x}\{G(A(x)) - G(0)\} = A'(x) \cdot g(A(x)),$$

where $G'(t) = g(t)$.

Example 2.14 (Distribution of sample median). In a random sample of size $n = 2m+1$, the *sample median* is $X_{(m+1)}$, whose sampling distribution is

$$\frac{(2m+1)!}{m!m!} f(x) F^m(x)\{1-F(x)\}^m, \quad -\infty < x < \infty.$$

For a random sample of size $n = 2m$, the median is defined as $\frac{1}{2}(X_{(m)} + X_{(m+1)})$. ‖

2.4.2 Joint distribution of more order statistics

19• The general case

- The joint density of $X_{(r_1)}, \ldots, X_{(r_k)}$ $(1 \leqslant r_1 \leqslant \cdots \leqslant r_k \leqslant n;\ 1 \leqslant k \leqslant n)$ is, for $x_1 \leqslant \cdots \leqslant x_k$ (or $x_{(r_1)} \leqslant \cdots \leqslant x_{(r_k)}$),

$$\begin{aligned} & g_{r_1 \cdots r_k}(x_1, \ldots, x_k) \\ = {} & n! \left\{ \prod_{i=1}^k f(x_i) \right\} \cdot \prod_{i=0}^k \frac{\{F(x_{i+1}) - F(x_i)\}^{r_{i+1}-r_i-1}}{(r_{i+1}-r_i-1)!}, \end{aligned} \tag{2.24}$$

where $x_0 = -\infty$, $x_{k+1} = +\infty$, $r_0 = 0$ and $r_{k+1} = n+1$.

19.1• Three special cases

— The joint pdf of $X_{(r)}$ and $X_{(s)}$ $(1 \leqslant r < s \leqslant n)$ is, for $x \leqslant y$,

$$\begin{aligned} g_{rs}(x,y) &= \frac{n!}{(r-1)!(s-r-1)!(n-s)!} f(x)f(y) \\ &\quad \times F^{r-1}(x)\{F(y)-F(x)\}^{s-r-1}\{1-F(y)\}^{n-s}. \end{aligned} \tag{2.25}$$

— The joint pdf of $X_{(1)}, \ldots, X_{(r)}$ $(1 \leqslant r \leqslant n)$ is, for $x_1 \leqslant \cdots \leqslant x_r$,

$$g_{1\cdots r}(x_1,\ldots,x_r) = \frac{n!}{(n-r)!} f(x_1)\cdots f(x_r)\{1-F(x_r)\}^{n-r}. \tag{2.26}$$

— The joint pdf of $X_{(1)}, \ldots, X_{(n)}$ is, for $x_1 \leqslant \cdots \leqslant x_n$,

$$g_{1\cdots n}(x_1,\ldots,x_n) = n! f(x_1)\cdots f(x_n). \tag{2.27}$$

Example 2.15 (Distribution of $X_{(s)} - X_{(r)}$ for uniform population). Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} U[0,1]$.

1) Find the distribution of $X_{(r)}$.

2) Find the distribution of $X_{(s)} - X_{(r)}$, where $1 \leqslant r < s \leqslant n$.

Solution. 1) Obviously, the corresponding cdf is

$$F(x) = \begin{cases} 0, & \text{if } x < 0, \\ x, & \text{if } 0 \leqslant x \leqslant 1, \\ 1, & \text{if } x > 1. \end{cases}$$

From (2.23), we have at once

$$g_r(x) = \frac{1}{B(r, n-r+1)} x^{r-1}(1-x)^{n-r}, \quad 0 \leqslant x \leqslant 1.$$

Thus $X_{(r)} \sim \text{Beta}(r, n-r+1)$.

2) From (2.25), the joint density of $X_{(r)}$ and $X_{(s)}$ is

$$g_{rs}(x_{(r)}, x_{(s)}) = c \cdot x_{(r)}^{r-1}\{x_{(s)} - x_{(r)}\}^{s-r-1}\{1-x_{(s)}\}^{n-s},$$

where $0 \leqslant x_{(r)} \leqslant x_{(s)} \leqslant 1$ and

$$c \;\hat{=}\; \frac{n!}{(r-1)!(s-r-1)!(n-s)!}$$

Making the transformation $z = x_{(s)} - x_{(r)}$ and $x = x_{(r)}$, we have

$$\begin{aligned} J(z, x \to x_{(r)}, x_{(s)}) &= \left| \frac{\partial(z, x)}{\partial(x_{(r)}, x_{(s)})} \right| \\ &= \det \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix} = -1. \end{aligned}$$

Hence, the joint density of $Z = X_{(s)} - X_{(r)}$ and $X = X_{(r)}$ is

$$\begin{aligned} h(z, x) &= g_{rs}(x_{(r)}, x_{(s)}) / |J(z, x \to x_{(r)}, x_{(s)})| \\ &= c \cdot x^{r-1} z^{s-r-1} (1 - x - z)^{n-s}, \end{aligned}$$

where $0 \leqslant x \leqslant 1$, $0 \leqslant z \leqslant 1$, and $0 \leqslant x + z \leqslant 1$. The marginal density of $Z = X_{(s)} - X_{(r)}$ is given by

$$\begin{aligned} h(z) &= \int_0^{1-z} h(z, x)\, \mathrm{d}x \\ &= c \cdot z^{s-r-1} \int_0^{1-z} x^{r-1} (1 - z - x)^{n-s}\, \mathrm{d}x \\ &= c \cdot z^{s-r-1} (1-z)^{n-s} \int_0^{1-z} x^{r-1} \left(1 - \frac{x}{1-z}\right)^{n-s} \mathrm{d}x. \end{aligned}$$

Let $w = x/(1-z)$, note that

$$\begin{aligned} \int_0^{1-z} x^{r-1} \left(1 - \frac{x}{1-z}\right)^{n-s} \mathrm{d}x &= \int_0^1 (1-z)^r w^{r-1} (1-w)^{n-s}\, \mathrm{d}x \\ &= (1-z)^r \cdot B(r, n-s+1), \end{aligned}$$

we obtain $h(z) \propto z^{s-r-1}(1-z)^{n-s+r}$, i.e.,

$$X_{(s)} - X_{(r)} \sim \text{Beta}(s - r, n - s + r + 1). \qquad \|$$

2.5 Limit Theorems

2.5.1 Convergency of a sequence of distribution functions

20• A MOTIVATION EXAMPLE

- Consider a sequence of i.i.d. r.v.'s $\{Y_i\}_{i=1}^{\infty}$ each having a uniform distribution on the unit interval $(0, 1)$.

- The mgf of $Y_1 \sim U(0,1)$ is

$$M_{Y_1}(t) = \begin{cases} 1, & \text{if } t = 0, \\ (\mathrm{e}^t - 1)/t, & \text{if } t \neq 0. \end{cases} \tag{2.28}$$

- Let $X_n \hat{=} \bar{Y} = \sum_{i=1}^n Y_i/n$. Since $X_1 = Y_1$ and $X_2 = (Y_1 + Y_2)/2 = (X_1 + Y_2)/2$, $\{X_n\}_{n=1}^\infty$ are dependent. The mgf of X_n is

$$M_{X_n}(t) = \begin{cases} 1, & \text{if } t = 0, \\ \{n(\mathrm{e}^{t/n} - 1)/t\}^n \to \mathrm{e}^{t/2} \text{ as } n \to \infty, & \text{if } t \neq 0. \end{cases} \tag{2.29}$$

- Since $\mathrm{e}^{t/2}$ is the mgf of the degenerate r.v. Z with all mass at 0.5; i.e., $\Pr(Z = 0.5) = 1$, we may expect the cdf F_n of X_n has the following limitation distribution

$$F_n(x) \to F_Z(x) = \begin{cases} 0, & x < 0.5, \\ 1, & x \geqslant 0.5. \end{cases}$$

20.1• Proof of (2.28)

— The pdf of $Y_1 \sim U(0,1)$ is $f(y_1) = 1 \cdot I_{(0,1)}(y_1)$.

— The mgf of Y_1 is defined by $M_{Y_1}(t) = E(\mathrm{e}^{tY_1})$.

— If $t = 0$, we have $M_{Y_1}(t) = M_{Y_1}(0) = E(\mathrm{e}^0) = 1$.

— If $t \neq 0$, we obtain

$$M_{Y_1}(t) = \int_0^1 \mathrm{e}^{ty_1}\, \mathrm{d}y_1 = \frac{1}{t}\mathrm{e}^{ty_1}\Big|_0^1 = \frac{1}{t}(\mathrm{e}^t - 1),$$

which completes the proof of (2.28). □

20.2• Proof of (2.29)

— We have

$$M_{X_n}(t) = M_{\bar{Y}}(t) = E\left\{\exp\left(\sum_{i=1}^n tY_i/n\right)\right\} = \left\{M_{Y_1}\left(\frac{t}{n}\right)\right\}^n.$$

— If $t = 0$, from the first one of (2.28), we have $M_{X_n}(t) = \{M_{Y_1}(0)\}^n = 1$.

— If $t \neq 0$, from the second formula of (2.28), we have

$$M_{X_n}(t) = \left(\frac{\mathrm{e}^{\frac{t}{n}} - 1}{\frac{t}{n}}\right)^{\frac{n}{t}\cdot t} \to \mathrm{e}^{t/2}, \quad \text{as } n \to \infty, \tag{2.30}$$

which completes the proof of (2.29). □

20.3• Proof of (2.30)

— To prove (2.30), we need to prove that

$$\lim_{x \to 0} \left(\frac{\mathrm{e}^x - 1}{x}\right)^{\frac{1}{x}} = \mathrm{e}^{1/2}. \tag{2.31}$$

Proof. Note that $\mathrm{e}^x = 1 + x + x^2/2! + x^3/3! + \cdots$, we have

$$\frac{\mathrm{e}^x - 1}{x} = 1 + \frac{x}{2} + \frac{x^2}{6} + \cdots. \tag{2.32}$$

Define

$$y = \left(\frac{\mathrm{e}^x - 1}{x}\right)^{\frac{1}{x}},$$

we obtain

$$\log(y) = \frac{1}{x} \log\left(\frac{\mathrm{e}^x - 1}{x}\right) \stackrel{(2.32)}{=} \frac{\log(1 + x/2 + x^2/6 + \cdots)}{x},$$

so that

$$\lim_{x \to 0} \log(y) = \lim_{x \to 0} \frac{\frac{1/2 + x/3 + \cdots}{1 + x/2 + x^2/6 + \cdots}}{1} = \frac{1}{2}.$$

Hence,

$$\lim_{x \to 0} y = \lim_{x \to 0} \mathrm{e}^{\log(y)} = \mathrm{e}^{1/2},$$

which completes the proof of (2.31). □

21• Convergence in distribution via cdf

Definition 2.2 (Convergence in distribution). Given a sequence of r.v.'s $\{X_n\}_{n=1}^{\infty}$. Let $F_n(x)$ be the cdf of X_n, if there exists an r.v. X with cdf $F(x)$ such that

$$\lim_{n \to \infty} F_n(x) = F(x)$$

for all points x at which $F(x)$ is continuous, then we say that $\{X_n\}_{n=1}^{\infty}$ converges *in distribution* or *in law* to X and write $X_n \xrightarrow{\mathrm{D}} X$ or $X_n \xrightarrow{\mathrm{L}} X$. ‖

21.1• Remarks on Definition 2.2

— It is possible that $\lim_{n\to\infty} F_n(x_0) \neq F(x_0)$ for such points x_0 at which $F(x)$ is discontinuous.

— $X_n \xrightarrow{\mathrm{L}} X \iff$ as $n \to \infty$, $X_n \overset{\mathrm{d}}{=} X$.

— The procedure for proving $X_n \xrightarrow{\mathrm{L}} X$ is as follows:

Step 1: Find $F_n(x)$.

Step 2: Find $F(x)$.

Step 3: Prove $F_n(x) \to F(x)$ as $n \to \infty$.

Example 2.16 (Uniform distribution). Let $\{Y_i\}_{i=1}^n \overset{\text{iid}}{\sim} U(0,\theta)$ and $X_n = Y_{(n)}$ be the n-th order statistic of $Y_1, \dots, Y_n$. Show that $X_n \xrightarrow{\mathrm{L}} X$, where X is an r.v. with $\Pr(X = \theta) = 1$.

Solution. The pdf and cdf of $Y \sim U(0,\theta)$ are $g(y) = 1/\theta$, $0 < y < \theta$, and

$$G(y) = \begin{cases} 0, & y < 0, \\ y/\theta, & 0 \leqslant y < \theta, \\ 1, & y \geqslant \theta, \end{cases}$$

respectively. From (2.17), we know that the pdf of X_n is

$$f_n(x) = ng(x)G^{n-1}(x) = nx^{n-1}/\theta^n, \quad 0 < x < \theta.$$

Thus, the cdf of X_n is

$$F_n(x) = \begin{cases} 0, & x < 0, \\ x^n/\theta^n, & 0 \leqslant x < \theta, \\ 1, & x \geqslant \theta, \end{cases} \quad \to \quad F(x) = \begin{cases} 0, & x < \theta, \\ 1, & x \geqslant \theta. \end{cases}$$

Therefore, $X_n \xrightarrow{\mathrm{L}} X$. ‖

Example 2.17 (Degenerate distribution). Let $\{X_n\}_{n=1}^\infty$ be a sequence of r.v.'s with $\Pr(X_n = 2 + \frac{1}{n}) = 1$. Show that $X_n \xrightarrow{\mathrm{L}} X$, where X is an r.v. with $\Pr(X = 2) = 1$.

Solution. The cdf of X_n is

$$F_n(x) = \begin{cases} 0, & x < 2 + 1/n, \\ 1, & x \geqslant 2 + 1/n, \end{cases}$$

$$\to F(x) = \begin{cases} 0, & x < 2, \\ 1, & x \geqslant 2, \end{cases} \qquad \text{as } n \to \infty.$$

Thus, $\lim_{n\to\infty} F_n(x) = F(x)$ for $x \neq 2$; i.e., all points where $F(x)$ is continuous. Thus $X_n \xrightarrow{\mathrm{L}} X$. ‖

22• Convergence in distribution via mgf

Theorem 2.5 (Equivalent result). Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of r.v's. Assume that the mgf $M_{X_n}(t) = M(t;n)$ of X_n exists for $|t| < h$ for all n, and there exists an r.v. X with mgf $M(t)$ that exists for $|t| < h_1 < h$. If

$$\lim_{n\to\infty} M(t;n) = M(t),$$

then $X_n \xrightarrow{\mathrm{L}} X$. ‖

Example 2.18 (Binomial distribution). Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of r.v.'s and $X_n \sim \text{Binomial}(n,p)$ with $np = \mu$, then $X_n \xrightarrow{\mathrm{L}} X$, where $X \sim \text{Poisson}(\mu)$.

Solution. The mgf of $X_n \sim \text{Binomial}(n,p)$ is

$$\begin{aligned} M(t;n) = (p\,\mathrm{e}^t + q)^n &= \left\{1 + \frac{\mu(\mathrm{e}^t - 1)}{n}\right\}^n \\ &\to \exp\{\mu(\mathrm{e}^t - 1)\} \quad \text{as } n \to \infty. \end{aligned} \tag{2.33}$$

for all real t. Since $\exp\{\mu(\mathrm{e}^t - 1)\}$ is the mgf of Poisson r.v. X, we have $X_n \xrightarrow{\mathrm{L}} X$. ‖

22.1• Proof of (2.33). To prove (2.33), we need to prove that

$$\lim_{n\to\infty}\left(1 + \frac{a}{n}\right)^n = \mathrm{e}^a \quad \text{or} \quad \lim_{x\to 0}(1 + ax)^{\frac{1}{x}} = \mathrm{e}^a. \tag{2.34}$$

— Proof. Define $y = (1+ax)^{\frac{1}{x}}$, we have $\log(y) = (1/x)\log(1+ax)$ so that

$$\lim_{x\to 0}\log(y) = \lim_{x\to 0}\frac{\log(1+ax)}{x} = \lim_{x\to 0}\frac{\frac{a}{1+ax}}{1} = a.$$

Therefore, $\lim_{x\to 0} y = \mathrm{e}^a$, which completes the proof of (2.34). □

2.5.2 Convergence in probability

Definition 2.3 (Weak convergence). A sequence of r.v.'s $\{X_n\}_{n=1}^{\infty}$ is said to *weakly* converge *in probability* to an r.v. X, denoted by $X_n \xrightarrow{\mathrm{P}} X$, if for any $\varepsilon > 0$,

$$\lim_{n\to\infty}\Pr(|X_n - X| \geqslant \varepsilon) = 0. \qquad \|$$

Theorem 2.6 (Markov inequality). Let $E|X|^r < \infty$, $r > 0$, $\varepsilon > 0$. Then

$$\Pr(|X| \geqslant \varepsilon) \leqslant \frac{E|X|^r}{\varepsilon^r}. \tag{2.35}$$

In particular, let $r = 2$, then $\mathrm{Var}(X) < \infty$ and

$$\begin{aligned}\Pr(|X-\mu| \geqslant \varepsilon) &\leqslant \frac{\mathrm{Var}(X)}{\varepsilon^2} \quad \text{or} \\ \Pr(|X-\mu| < \varepsilon) &\geqslant 1 - \frac{\mathrm{Var}(X)}{\varepsilon^2},\end{aligned} \tag{2.36}$$

where $\mu = E(X)$. ‖

Proof. If $|x| \geqslant \varepsilon$, then $|x|^r \geqslant \varepsilon^r$; i.e.,

$$1 \leqslant \frac{|x|^r}{\varepsilon^r}.$$

Let $X \sim F(x)$, we have

$$\begin{aligned}\Pr(|X| \geqslant \varepsilon) &= \int_{|x|\geqslant\varepsilon} \mathrm{d}F(x) \\ &\leqslant \int_{|x|\geqslant\varepsilon} \frac{|x|^r}{\varepsilon^r}\,\mathrm{d}F(x) \\ &\leqslant \int_{-\infty}^{\infty} \frac{|x|^r}{\varepsilon^r}\,\mathrm{d}F(x) \\ &= \frac{E|X|^r}{\varepsilon^r},\end{aligned}$$

which implies (2.35). □

2.5.3 Relationship of four classes of convergency

Definition 2.4 (Strong convergence). A sequence of r.v.'s $\{X_n\}_{n=1}^{\infty}$ is said to *strongly* converge *almost surely* to an r.v. X, denoted by $X_n \stackrel{\text{a.s.}}{\to} X$, if

$$\Pr\left(\lim_{n\to\infty} X_n = X\right) = 1. \qquad \|$$

Definition 2.5 (Convergence in mean square). A sequence of r.v.'s $\{X_n\}_{n=1}^{\infty}$ is said to converge *in mean square* to an r.v. X, denoted by $X_n \stackrel{\text{m.s.}}{\to} X$, if

$$\lim_{n\to\infty} E(X_n - X)^2 = 0. \qquad \|$$

The relationship of the four classes of convergency can be summarized by

$$\begin{array}{l} X_n \stackrel{\text{a.s.}}{\to} X \\ X_n \stackrel{\text{m.s.}}{\to} X \end{array} \Longrightarrow X_n \stackrel{\text{P}}{\to} X \Longrightarrow X_n \stackrel{\text{L}}{\to} X.$$

Property 2.1 $X_n \stackrel{\text{P}}{\to} X \Longrightarrow X_n \stackrel{\text{L}}{\to} X.$ $\|$

Proof. We first prove the following facts: (i) $\forall\, x' < x$, if $X_n \stackrel{\text{P}}{\to} X$, then

$$\Pr(X_n \geqslant x, X < x') \to 0. \tag{2.37}$$

(ii) $\forall\, x < x''$, if $X_n \stackrel{\text{P}}{\to} X$, then

$$\Pr(X_n < x, X \geqslant x'') \to 0. \tag{2.38}$$

In fact, $\{X_n \geqslant x, X < x'\} \Longrightarrow X_n - X \geqslant x - x' > 0$, then

$$|X_n - X| = X_n - X \geqslant x - x' > 0.$$

Thus,

$$0 \leqslant \Pr\{X_n \geqslant x, X < x'\} \leqslant \Pr\{|X_n - X| \geqslant x - x'\} \to 0,$$

which implies (2.37). Similarly, we can prove (2.38).

On the one hand, for $x' < x$, since

$$\begin{aligned} \{X < x'\} &= \{X_n < x, X < x'\} + \{X_n \geqslant x, X < x'\} \\ &\subset \{X_n < x\} + \{X_n \geqslant x, X < x'\}, \end{aligned}$$

we have

$$F(x') \leqslant F_n(x) + \Pr\{X_n \geqslant x, X < x'\} \leqslant \underline{\lim}_{n\to\infty} F_n(x).$$

On the other hand, for $x < x''$, since

$$\begin{aligned}\{X \geqslant x''\} &= \{X_n \geqslant x, X \geqslant x''\} + \{X_n < x, X \geqslant x''\} \\ &\subset \{X_n \geqslant x\} + \{X_n < x, X \geqslant x''\},\end{aligned}$$

we have

$$1 - F(x'') \leqslant \underline{\lim}_{n\to\infty} \Pr\{X_n \geqslant x\} = 1 - \overline{\lim}_{n\to\infty} F_n(x),$$

i.e., $F(x'') \geqslant \overline{\lim}_{n\to\infty} F_n(x)$.

Therefore, for $x' < x < x''$, we have

$$F(x') \leqslant \underline{\lim}_{n\to\infty} F_n(x) \leqslant \overline{\lim}_{n\to\infty} F_n(x) \leqslant F(x'').$$

Let x be a point at which $F(x)$ is continuous. Let $x' \to x$ and $x'' \to x$, then $F(x) = \lim_{n\to\infty} F_n(x)$. □

Property 2.2 $X_n \xrightarrow{\text{L}} c \Longleftrightarrow X_n \xrightarrow{\text{P}} c$, where c is a constant. ‖

Proof. Property 2.1 indicates that we only need to prove "$\Longrightarrow$". Note that the cdf of $X = c$ is

$$F_X(x) = \begin{cases} 0, & \text{if } x \leqslant c, \\ 1, & \text{if } x > c, \end{cases}$$

hence, as $n \to \infty$,

$$\begin{aligned}\Pr(|X_n - c| \geqslant \varepsilon) &= \Pr(X_n \geqslant c + \varepsilon) + \Pr(X_n \leqslant c - \varepsilon) \\ &= 1 - F_n(c + \varepsilon) + F_n(c - \varepsilon) \\ &\to 1 - F_X(c + \varepsilon) + F_X(c - \varepsilon) \\ &\to 1 - 1 + 0 = 0,\end{aligned}$$

which completes the proof. □

Property 2.3 $X_n \xrightarrow{\text{m.s.}} X \Longrightarrow X_n \xrightarrow{\text{P}} X$. ‖

Proof. If $X_n \xrightarrow{\text{m.s.}} X$, by using (2.35), then

$$\Pr(|X_n - X| \geqslant \varepsilon) \leqslant \frac{E(X_n - X)^2}{\varepsilon^2} \to 0, \quad \text{as } n \to \infty.$$

This means that $X_n \xrightarrow{\text{P}} X$. □

2.5.4 Law of large number

Theorem 2.7 (Weak law of large number). Assume that $\{X_n\}_{n=1}^{\infty}$ is a sequence of i.i.d. random variables with $E(X_n) = \mu < \infty$. Let $\bar{X}_n = \sum_{i=1}^{n} X_i/n$, then $\bar{X}_n \xrightarrow{\text{P}} \mu$. ‖

Proof. We prove it under an additional assumption $\text{Var}(X_n) = \sigma^2 < \infty$. By using (2.35), we have

$$\Pr(|\bar{X}_n - \mu| \geqslant \varepsilon) \leqslant \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0, \quad \text{as } n \to \infty.$$

This means that $\bar{X}_n \xrightarrow{\text{P}} \mu$. □

Theorem 2.8 (Strong law of large number). Assume that $\{X_n\}_{n=1}^{\infty}$ is a sequence of i.i.d. random variables with $E(X_n) = \mu < \infty$. Let $\bar{X}_n = \sum_{i=1}^{n} X_i/n$, then $\bar{X}_n \xrightarrow{\text{a.s}} \mu$. ‖

2.5.5 Central limit theorem

23• Proof of the central limit theorem via mgf

Theorem 2.9 (Central limit theorem). Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of i.i.d. random variables with common mean μ and common variance $\sigma^2 > 0$. Let $\bar{X}_n = \sum_{i=1}^{n} X_i/n$ and $Y_n = \sqrt{n}(\bar{X}_n - \mu)/\sigma$, then $Y_n \xrightarrow{\text{L}} Z$ as $n \to \infty$, where $Z \sim N(0, 1)$. ‖

Proof. Assume that the mgf of X exists for $|t| < h$. Let

$$m(t) = E\{\,\text{e}^{t(X-\mu)}\}.$$

Then $m(0) = 1$, $m'(0) = E(X - \mu) = 0$, $m''(0) = E(X - \mu)^2 = \sigma^2$. By Maclaurin's expansion,

$$m(t) = m(0) + m'(0)t + \frac{1}{2}m''(\xi)t^2 = 1 + \frac{m''(\xi)}{2}t^2, \quad 0 < \xi < t,$$

where $m''(\xi) \to m''(0) = \sigma^2$ as $t \to 0$. Now

$$\begin{aligned}
M(t; n) &= E(\,\text{e}^{tY_n}) \\
&= E[\exp\{t\sqrt{n}(\bar{X}_n - \mu)/\sigma\}] \\
&= E[\exp\{t\textstyle\sum_{i=1}^{n}(X_i - \mu)/(\sqrt{n}\sigma)\}]
\end{aligned}$$

$$
\begin{aligned}
&= \prod_{i=1}^{n} E[\exp\{t(X_i-\mu)/(\sqrt{n}\sigma)\}] \\
&= \{m(t/(\sqrt{n}\sigma))\}^n \\
&= \left\{1+\frac{m''(\xi(n))}{2}(t/(\sqrt{n}\sigma))^2\right\}^n \\
&= \left\{1+\frac{m''(\xi(n))}{2n\sigma^2}t^2\right\}^n, \quad 0<\xi(n)<t/(\sqrt{n}\sigma) \\
&\to \mathrm{e}^{t^2/2} \quad \text{as } n\to\infty,
\end{aligned}
$$

since $\xi(n) \to 0$ and $m''(\xi(n)) \to m''(0) = \sigma^2$. Because $\mathrm{e}^{t^2/2}$ is the mgf of $Z \sim N(0,1)$, this means that $Y_n \xrightarrow{\text{L}} Z$. □

Example 2.19 (Bernoulli distribution). Let $X_1, \ldots, X_n$ be a random sample from Bernoulli(θ). Let $Z_n = \sum_{i=1}^n X_i$, then

$$\frac{Z_n - n\theta}{\sqrt{n\theta(1-\theta)}} \xrightarrow{\text{L}} N(0,1) \quad \text{as } n\to\infty. \tag{2.39}$$

Solution. Because $\mu = \theta$ and $\sigma^2 = \theta(1-\theta)$, by the central limit theorem, we have

$$\frac{Z_n - n\theta}{\sqrt{n\theta(1-\theta)}} = \frac{n\bar{X}_n - n\theta}{\sqrt{n\theta(1-\theta)}} = \frac{\sqrt{n}(\bar{X}_n-\mu)}{\sigma} \xrightarrow{\text{L}} Z \quad \text{as } n\to\infty,$$

where $Z \sim N(0,1)$. ‖

23.1• Remarks on normal approximation

— Since $Z_n \sim \text{Binomial}(n,\theta)$, we have $E(Z_n) = n\theta$ and $\text{Var}(Z_n) = n\theta(1-\theta)$. Then (2.39) means

$$\frac{Z_n - E(Z_n)}{\sqrt{\text{Var}(Z_n)}} \xrightarrow{\text{L}} Z \sim N(0,1) \quad \text{as } n\to\infty.$$

— If n is large, approximately we have

$$Z_n \mathrel{\dot\sim} N(n\theta, n\theta(1-\theta)).$$

That is, Binomial(n,θ) can be approximated by $N(n\theta, n\theta(1-\theta))$.

— If Z_n is a discrete r.v., in using normal approximation, we should use

$$\Pr(Z_n = k) = \Pr(k - 0.5 < Z_n < k + 0.5),$$

and number 0.5 here is called the continuity correction.

Example 2.20 (Binomial distribution). Let $X \sim \text{Binomial}(10, 0.5)$, directly calculate $\Pr(X = 4)$ and compute $\Pr(X = 4)$ by normal approximation.

Solution. First, we directly compute

$$\Pr(X = 4) = \binom{10}{4} 0.5^4 0.5^6 = 0.2051.$$

Second, we use normal approximation $X \sim N(5, 2.5)$ and obtain

$$\begin{aligned}
\Pr(X = 4) &= \Pr(4 - 0.5 < X < 4 + 0.5) \\
&= \Pr(3.5 < X < 4.5) \\
&= \Pr\left(\frac{3.5 - 5}{\sqrt{2.5}} < \frac{X - 5}{\sqrt{2.5}} < \frac{4.5 - 5}{\sqrt{2.5}}\right) \\
&\doteq \Pr(-0.9487 < Z < -0.3162) \\
&= \Phi(-0.3162) - \Phi(-0.9487) \\
&= \Phi(0.9487) - \Phi(0.3162) \\
&= 0.8286 - 0.6241 = 0.2045.
\end{aligned}$$

The error is $0.2051 - 0.2045 = 0.0006$ and the percentage error is

$$\frac{|0.2051 - 0.2045|}{0.2051} = 0.29\%. \qquad \|$$

2.6 Some Challenging Questions

24• Dependency and correlation

- Let r.v. $X \sim N(0, 1)$ and we define a new random variable $Y = X^2$.
- In Example 2.7, we know that $Y \sim \chi^2(1)$.

24.1• Dependency and correlation between X and Y

— It is clear that X and Y are *dependent* because $Y = X^2$ is uniquely determined when X is given.

— Let $\phi(x)$ be the pdf of $N(0,1)$. Since $x^3\phi(x)$ is an odd function, we have

$$E(XY) = E(X^3) = \int_{-\infty}^{\infty} x^3\phi(x)\,\mathrm{d}x = 0.$$

— Note that $E(X) = 0$, we obtain

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{E(XY) - E(X)E(Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = 0.$$

— In other words, X and Y are uncorrelated but surely dependent.

24.2• Conditional distributions of $Y|(X = x)$ and $X|(Y = y)$

— The conditional distribution of $Y|(X = x)$ is

$$\Pr(Y = x^2|X = x) = 1;$$

i.e., $Y|(X = x) \sim \text{Degenerate}(x^2)$.

— The conditional distribution of $X|(Y = y > 0)$ is given by

$$\Pr(X = -\sqrt{y}|Y = y) = \Pr(X = \sqrt{y}|Y = y) = 0.5;$$

that is, $X|(Y = y > 0)$ follows a uniform two-point distribution.

— The conditional distribution of $X|(Y = y = 0)$ is

$$\Pr(X = 0|Y = 0) = 1;$$

that is, $X|(Y = y = 0) \sim \text{Degenerate}(0)$.

24.3• The joint cdf of X and Y

— Let $F(x,y)$ denote the cdf of (X,Y), we have

$$\begin{aligned} F(x,y) &= \Pr(X \leqslant x, X^2 \leqslant y) = \Pr(X \leqslant x, -\sqrt{y} \leqslant X \leqslant \sqrt{y}) \\ &= \Pr\{-\sqrt{y} \leqslant X \leqslant \min(x, \sqrt{y})\} \\ &= \Phi(\min\{x, \sqrt{y}\}) - \Phi(-\sqrt{y}), \quad -\infty < x < \infty,\ y > 0, \end{aligned}$$

where $\Phi(\cdot)$ is the cdf of the standard normal distribution.

24.4• Can the identities

$$f_{(X,Y)}(x,y) = f_X(x) f_{(Y|X)}(y|x) = f_Y(y) f_{(X|Y)}(x|y) \tag{2.40}$$

be used to derive the joint density function of X and Y?

— No.

24.5• Comment on the existence of $f_{(X,Y)}(x,y)$ in the xy-plane

— The joint pdf of (X,Y) does *not exist* in the xy-plane because the support of (X,Y) is

$$\mathbb{S}_{(X,Y)} = \{(x,y)\colon\ -\infty < x < \infty,\ y = x^2\},$$

which is a curve and the *measure/area of* $\mathbb{S}_{(X,Y)}$ *is zero.*

25• Proof of Theorem 2.1

- In **41.2•** of Chapter 1, it was shown that the mgf of $\mathbf{x} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is

$$M_{\mathbf{x}}(\boldsymbol{t}) = \exp(\boldsymbol{t}^\top \boldsymbol{\mu} + 0.5\boldsymbol{t}^\top \boldsymbol{\Sigma} \boldsymbol{t}). \tag{2.41}$$

25.1• $\mathbf{Ax} \sim N_m(\mathbf{A}\boldsymbol{\mu}, \mathbf{A\Sigma A}^\top)$ and $\mathbf{Bx} \sim N_r(\mathbf{B}\boldsymbol{\mu}, \mathbf{B\Sigma B}^\top)$

— Let $\boldsymbol{s} = (s_1, \ldots, s_m)^\top$ and define

$$\underset{m\times 1}{\mathbf{y}} = \underset{m\times n}{\mathbf{A}}\,\underset{n\times 1}{\mathbf{x}},$$

then the mgf of $\mathbf{y}$ is

$$\begin{aligned}
M_{\mathbf{y}}(\boldsymbol{s}) &= E\{\exp(\boldsymbol{s}^\top \mathbf{y})\} = E\{\exp(\boldsymbol{s}^\top \mathbf{Ax})\} \\
&= E[\exp\{(\mathbf{A}^\top \boldsymbol{s})^\top \mathbf{x}\}] \\
&= M_{\mathbf{x}}(\mathbf{A}^\top \boldsymbol{s}) \qquad [\text{Let } \boldsymbol{t} = \mathbf{A}^\top \boldsymbol{s}] \\
&= M_{\mathbf{x}}(\boldsymbol{t}) \\
&\overset{(2.41)}{=} \exp(\boldsymbol{t}^\top \boldsymbol{\mu} + 0.5\boldsymbol{t}^\top \boldsymbol{\Sigma} \boldsymbol{t}) \\
&= \exp(\boldsymbol{s}^\top \mathbf{A}\boldsymbol{\mu} + 0.5\boldsymbol{s}^\top \mathbf{A\Sigma A}^\top \boldsymbol{s}) \\
&= \exp\{\boldsymbol{s}^\top (\mathbf{A}\boldsymbol{\mu}) + 0.5\boldsymbol{s}^\top (\mathbf{A\Sigma A}^\top) \boldsymbol{s}\},
\end{aligned}$$

implying $\mathbf{y} \sim N_m(\mathbf{A}\boldsymbol{\mu},\ \mathbf{A\Sigma A}^\top)$.

— Similarly, we can prove $\mathbf{Bx} \sim N_r(\mathbf{B}\boldsymbol{\mu}, \mathbf{B\Sigma B}^\top)$.

25.2• $\mathbf{Ax} \perp\!\!\!\perp \mathbf{Bx}$ iff $\mathbf{A\Sigma B}^\top = \mathbf{O}_{m\times r}$

— Define

$$\underset{(m+r)\times 1}{\mathbf{z}} = \begin{pmatrix} \mathbf{Ax} \\ \mathbf{Bx} \end{pmatrix} = \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} \mathbf{x} \mathrel{\hat{=}} \underset{(m+r)\times n}{\mathbf{C}} \underset{n\times 1}{\mathbf{x}},$$

then, we have $\mathbf{z} \sim N_{m+r}(\mathbf{C}\boldsymbol{\mu},\ \mathbf{C\Sigma C}^\top)$.

— Note that

$$\mathbf{C\Sigma C}^\top = \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} \boldsymbol{\Sigma}(\mathbf{A}^\top\, \mathbf{B}^\top) = \begin{pmatrix} \mathbf{A\Sigma A}^\top & \mathbf{A\Sigma B}^\top \\ \mathbf{B\Sigma A}^\top & \mathbf{B\Sigma B}^\top \end{pmatrix},$$

we can see that $\mathbf{Ax} \perp\!\!\!\perp \mathbf{Bx}$ iff $\mathbf{A\Sigma B}^\top = \mathbf{O}_{m\times r}$. □

Exercise 2

2.1 Calculate the expectation and variance of the $T \sim t(n)$ via the stochastic representation (SR):

$$T \stackrel{\mathrm{d}}{=} \frac{Z}{\sqrt{Y/n}},$$

where $Z \sim N(0,1)$, $Y \sim \chi^2(n)$ and $Z \perp\!\!\!\perp Y$.

2.2 Let $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{Beta}(3,2)$. Find the sampling distributions of $X_{(1)} = \min\{X_1, \ldots, X_n\}$ and $X_{(n)} = \max\{X_1, \ldots, X_n\}$.

2.3 Let $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ be the order statistics of a random sample of size n from the exponential distribution with pdf $f(x) = \mathrm{e}^{-x}$, $x > 0$, zero elsewhere.

(a) Show that $Z_1 = nX_{(1)}$, $Z_2 = (n-1)[X_{(2)} - X_{(1)}]$, $Z_3 = (n-2)[X_{(3)} - X_{(2)}], \ldots, Z_n = X_{(n)} - X_{(n-1)}$ are independent and that each Z_i has the exponential distrbution.

(b) Demonstrate that all linear functions of $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$, such as $\sum_{i=1}^n a_i X_{(i)}$, can be expressed as linear functions of independent random variables.

2.4 Let $X_i \sim \text{Gamma}(a_i, 1)$, $i = 1, \ldots, n$, and $X_1, \ldots, X_n$ are mutually independent. Define

$$Y_i = \frac{X_i}{X_1 + \cdots + X_n}, \quad i = 1, \ldots, n-1.$$

(a) Find the joint density of $(Y_1, \ldots, Y_{n-1})$.

(b) Find the density of $X_1 + \cdots + X_n$.

2.5 Let $X \sim \text{Gamma}(p, 1)$, $Y \sim \text{Beta}(q, p-q)$, and $X \perp\!\!\!\perp Y$, where $0 < q < p$. Find the distribution of XY.

2.6 Let $Z \sim \text{Bernoulli}\,(1-\phi)$, $\mathbf{x} = (X_1, \ldots, X_m)^\top$, $X_i \sim \text{Poisson}\,(\lambda_i)$ for $i = 1, \ldots, m$, and $(Z, X_1, \ldots, X_m)$ be mutually independent. Define $\mathbf{y} = (Y_1, \ldots, Y_m)^\top = Z\mathbf{x}$. Find the joint pmf of $\mathbf{y}$.