Appendix B:

A Unified Expectation Technique

1• The issue

- Let $\mathbf{x} = (X_1, \ldots, X_n)^\top$ be an $n \times 1$ random vector with the joint pdf $f_{\mathbf{x}}(\boldsymbol{x})$ or joint pmf $p_{\mathbf{x}}(\boldsymbol{x})$, and we want to find the distribution of the continuous or discrete r.v. $Y = h(\mathbf{x})$ for some function $h(\cdot)$.

- Especially, when $n = 1$, the problem can be formulated as follows. Let $X \sim f_X(x)$ or $X \sim p_X(x)$, it is desired to find the distribution of $Y = h(X)$ for some function $h(\cdot)$.

- For continuous random variables, most of the statistics textbooks generally introduced the following three commonly used methods (see Section 2.1):

  — The *cumulative distribution function* (cdf) technique.

  — The transformation technique.

  — The *moment generating function* (mgf) technique.

- For discrete random variables, we have introduced the following three methods:

  — The *probability generating function* (pgf) technique (see **32.2•** of Chapter 1).

  — The *law-of-probability* technique (see Examples 2.3–2.5).

  — The mgf technique (see Examples 2.11–2.12).

2• The aim

- This appendix will introduce the fourth method, called as *expectation technique*.

- For continuous cases, the expectation technique was originally proposed by Professor Yaoting ZHANG at Wuhan University (Wuhan City, Hubei Province, P. R. China). As one of his postgraduates, the first author of this textbook learned this useful method around 1986 from a statistics seminar organized by Professor Yaoting ZHANG.
- For discrete cases, the expectation technique is totally new and we did not publish it anywhere.
- The expectation technique is a unified method with some advantages over the existing three methods.
- To our best knowledge, the expectation technique did not appear in general textbooks of statistics.

B.1 Continuous Random Variables

B.1.1 The one-dimensional case

3• Formulation of the expectation technique

3.1• The motivation

— We know that if $X \sim f_X(x)$, then we have

$$E\{g(X)\} = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) \, \mathrm{d}x \tag{B.1}$$

for an arbitrary function $g(\cdot)$ provided that $E\{|g(X)|\} < +\infty$.

— Inversely, if (B.1) is true for any function $g(\cdot)$, we wonder whether $f_X(x)$ is the pdf of X?

3.2• The expectation technique

— Let $X \sim f_X(x)$ and $Y = h(X)$. For any nonnegative and measurable $g(\cdot)$, of course, we have

$$E\{g(Y)\} = E\{g(h(X))\} \overset{(B.1)}{=} \int_{-\infty}^{\infty} g(h(x)) \cdot f_X(x) \, \mathrm{d}x. \tag{B.2}$$

— Make the transformation of $y = h(x)$, if we can express the last integral in (B.2) as $\int g(y) f_Y(y)\,\mathrm{d}y$; i.e.,

$$E\{g(Y)\} = \int_{-\infty}^{\infty} g(h(x)) \cdot f_X(x)\,\mathrm{d}x = \int g(y) \cdot f_Y(y)\,\mathrm{d}y, \tag{B.3}$$

then we can claim that $f_Y(y)$ is the pdf of Y. ‖

3.3• The first proof of the expectation technique

— Let $g(Y) = I(Y \leqslant y)$, we have

$g(Y)$	0	1
Probability	$1 - \Pr(Y \leqslant y)$	$\Pr(Y \leqslant y)$

i.e., $g(Y) \sim \text{Bernoulli}(\Pr(Y \leqslant y))$. Thus

$$E\{g(Y)\} = \Pr(Y \leqslant y), \tag{B.4}$$

which is the cdf of Y.

— On the other hand, (B.3) can be rewritten as

$$\begin{aligned} E\{g(Y)\} &= \int g(t) \cdot f_Y(t)\,\mathrm{d}t \qquad [\text{since } g(t) = I(t \leqslant y)] \\ &= \int I(t \leqslant y) \cdot f_Y(t)\,\mathrm{d}t = \int_{-\infty}^{y} f_Y(t)\,\mathrm{d}t. \end{aligned} \tag{B.5}$$

— By combining (B.4) with (B.5), we obtain

$$\Pr(Y \leqslant y) = \int_{-\infty}^{y} f_Y(t)\,\mathrm{d}t,$$

indicating that $f_Y(\cdot)$ is the pdf of Y. □

3.4• The second proof of the expectation technique

— Let $g(Y) = \mathrm{e}^{tY}$. On the one hand, we have

$$E\{g(Y)\} = E(\mathrm{e}^{tY}) = M_Y(t), \tag{B.6}$$

which is the mgf of Y.

— On the other hand, (B.3) can be rewritten as

$$\begin{aligned} E\{g(Y)\} &= \int g(y)\cdot f_Y(y)\,\mathrm{d}y \qquad [\text{since } g(y)=\mathrm{e}^{ty}] \\ &= \int \mathrm{e}^{ty}\cdot f_Y(y)\,\mathrm{d}y. \end{aligned} \tag{B.7}$$

— By combining (B.6) with (B.7), we obtain

$$M_Y(t) = \int \mathrm{e}^{ty}\cdot f_Y(y)\,\mathrm{d}y,$$

indicating that $f_Y(y)$ is the pdf of Y. □

3.5• Comments

— From the process of the first proof, we know that the cdf technique is a special case of the expectation technique with $g(Y) = I(Y \leqslant y)$.

— From the process of the second proof, we know that the mgf technique is also a special case of the expectation technique with $g(Y) = \mathrm{e}^{tY}$.

4• Two examples

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

Solution. For any nonnegative and measurable $g(\cdot)$, from (B.2), we have

$$\begin{aligned} E\{g(Y)\} &= \int_{-\infty}^{\infty} g(h(x))\cdot f_X(x)\,\mathrm{d}x \\ &= \int_0^1 g(x^3)\cdot 6x(1-x)\,\mathrm{d}x \qquad [\text{let } y = x^3 \Rightarrow x = y^{1/3}] \\ &= \int_0^1 g(y)\cdot 6y^{1/3}(1-y^{1/3})\frac{1}{3}y^{-2/3}\,\mathrm{d}y \\ &= \int_0^1 g(y)\cdot 2(y^{-1/3}-1)\,\mathrm{d}y, \end{aligned}$$

implying that $Y \sim f_Y(y) = 2(y^{-1/3}-1)$ for $0 < y < 1$.

Remark. The advantage of using the expectation technique in this example is not so obvious when comparing with the cdf technique. ‖

Example B.2 (Example 2.7 revisited). Let $X \sim N(0,1)$, find the pdf of $Y = h(X) = X^2$.

Solution. For any nonnegative and measurable $g(\cdot)$, from (B.2), we have

$$
\begin{aligned}
E\{g(Y)\} &= \int_{-\infty}^{\infty} g(h(x)) \cdot f_X(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} g(x^2) \cdot \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-x^2/2}\,\mathrm{d}x \\
&= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} g(x^2) \cdot \mathrm{e}^{-x^2/2}\,\mathrm{d}x \qquad [\text{let } y = x^2 \Rightarrow x = y^{1/2}] \\
&= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} g(y) \cdot \mathrm{e}^{-y/2} \frac{1}{2} y^{-1/2}\,\mathrm{d}y \\
&= \int_0^{1} g(y) \cdot \frac{1}{\sqrt{2\pi}} y^{-1/2}\,\mathrm{e}^{-y/2}\,\mathrm{d}y,
\end{aligned}
$$

implying that $Y \sim \chi^2(1) = \text{Gamma}(1/2, 1/2)$.

Remark. The merit of using the expectation technique is apparent when comparing with the transformation technique, where two piecewise monotone functions must be considered. ‖

B.1.2 The two-dimensional case

5• The expectation technique

- Let $f_{(X_1,X_2)}(x_1, x_2)$ be the joint density of (X_1, X_2), and $Y = h(X_1, X_2)$ for some function $h(\cdot,\cdot)$.
- For any nonnegative and measurable $g(\cdot)$, if

$$
\begin{aligned}
& E\{g(Y)\} = E\{g(h(X_1, X_2))\} \\
=\ & \int\!\!\int g(h(x_1, x_2)) \cdot f_{(X_1,X_2)}(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 \qquad \text{(B.8)} \\
=\ & \int g(y) \cdot f_Y(y)\,\mathrm{d}y,
\end{aligned}
$$

then we can claim that $f_Y(y)$ is the pdf of Y. ‖

6• Several examples

Example B.3 (Example 2.2 revisited). Let

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

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

Solution. For any nonnegative and measurable $g(\cdot)$, from (B.8), we have

$$\begin{aligned}
E\{g(Y)\} &= \int_0^\infty \int_0^\infty g(h(x_1, x_2)) \cdot f_{(X_1,X_2)}(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 \\
&= \int_0^\infty \int_0^\infty g(x_1 + x_2) \cdot 6\exp(-3x_1 - 2x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 \\
&= \int_0^\infty 6\,\mathrm{e}^{-2x_2} \left\{ \int_0^\infty g(x_1 + x_2)\,\mathrm{e}^{-3x_1}\,\mathrm{d}x_1 \right\} \mathrm{d}x_2 \qquad [\text{let } y = x_1 + x_2] \\
&= \int_0^\infty 6\,\mathrm{e}^{-2x_2} \left\{ \int_{x_2}^\infty g(y)\,\mathrm{e}^{-3(y-x_2)}\,\mathrm{d}y \right\} \mathrm{d}x_2 \qquad [\text{exchange } y \text{ and } x_2] \\
&= \int_0^\infty g(y) \cdot \mathrm{e}^{-3y} \left(\int_0^y 6\,\mathrm{e}^{x_2}\,\mathrm{d}x_2 \right) \mathrm{d}y \\
&= \int_0^\infty g(y) \cdot \mathrm{e}^{-3y} \cdot 6(\mathrm{e}^y - 1)\,\mathrm{d}y \\
&= \int_0^\infty g(y) \cdot 6(\mathrm{e}^{-2y} - \mathrm{e}^{-3y})\,\mathrm{d}y
\end{aligned}$$

implying that $Y \sim f_Y(y) = 6(\mathrm{e}^{-2y} - \mathrm{e}^{-3y})$ for $y \geqslant 0$.

Remark. In Example 2.2, we use the cdf technique and need to change the integration region $\mathbb{D} = \{(x_1, x_2)\colon x_1 \geqslant 0, x_2 \geqslant 0, x_1 + x_2 \leqslant y\}$.

— When the transformation technique is applied to this example, we first need to find the joint density of (Y, Z) by making the transformations $Y = X_1 + X_2$ and $Z = X_2$, where the Jacobian determinant $J(x_1, x_2 \to y, z)$ must be calculated. Second, we need to integrate $f_{(Y,Z)}(y, z)$ about z to obtain the pdf of Y.

— The merit of using the expectation technique can avoid the calculation of the Jacobian determinant of a 2×2 matrix. ‖

Example B.4 (Example 2.8 revisited). Let $X_1, X_2 \stackrel{\text{iid}}{\sim} N(0,1)$. Find the distribution of $Y = h(X_1, X_2) = X_1/X_2$.

Solution. Let $\phi(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}$ denote the density of $N(0,1)$. For any nonnegative and measurable $g(\cdot)$, from (B.8), we have

$$
\begin{aligned}
E\{g(Y)\} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(h(x_1,x_2)) \cdot f_{(X_1,X_2)}(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1/x_2) \cdot \phi(x_1)\phi(x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-0.5x_2^2}\left\{\int_{-\infty}^{\infty} g(x_1/x_2)\,\mathrm{e}^{-0.5x_1^2}\,\mathrm{d}x_1\right\}\mathrm{d}x_2 \\
&\qquad [\text{let } y = x_1/x_2 \;\Rightarrow\; x_1 = x_2 y \;\Rightarrow\; \mathrm{d}x_1 = x_2\,\mathrm{d}y \;\Rightarrow\; \mathrm{d}x_1 = |x_2|\,\mathrm{d}y] \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-0.5x_2^2}\left\{\int_{-\infty}^{\infty} g(y)\,\mathrm{e}^{-0.5x_2^2y^2}|x_2|\,\mathrm{d}y\right\}\mathrm{d}x_2 \quad [\text{exchange } y \,\&\, x_2] \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} g(y)\left\{\int_{-\infty}^{\infty} \underbrace{|x_2|\,\mathrm{e}^{-0.5x_2^2(1+y^2)}}_{\text{even function on } x_2}\,\mathrm{d}x_2\right\}\mathrm{d}y \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} g(y)\left\{2\int_{0}^{\infty} x_2\,\mathrm{e}^{-0.5x_2^2(1+y^2)}\,\mathrm{d}x_2\right\}\mathrm{d}y \\
&\qquad [\text{let } z = 0.5x_2^2(1+y^2) \;\Rightarrow\; \mathrm{d}z = x_2(1+y^2)\,\mathrm{d}x_2] \\
&= \frac{1}{\pi}\int_{-\infty}^{\infty} g(y)\left\{\int_{0}^{\infty} \frac{1}{1+y^2}\,\mathrm{e}^{-z}\,\mathrm{d}z\right\}\mathrm{d}y \\
&= \int_{-\infty}^{\infty} g(y) \cdot \frac{1}{\pi(1+y^2)}\,\mathrm{d}y,
\end{aligned}
$$

implying that Y follows the standard Cauchy distribution. ‖

Example B.5 (Linear combination of two i.i.d. standard normal r.v.s). Let $X_1, X_2 \stackrel{\text{iid}}{\sim} N(0,1)$. Show that

$$Y \hat{=} \mu + a_1X_1 + a_2X_2 \sim N(\mu, a_1^2 + a_2^2). \tag{B.9}$$

Proof. Let $\sigma^2 \hat{=} a_1^2 + a_2^2$, it is easy to verify that

$$\begin{aligned} x_2^2 + \frac{(y-\mu-a_2x_2)^2}{a_1^2} &= \frac{[x_2 - a_2(y-\mu)/\sigma^2]^2}{a_1^2/\sigma^2} + \frac{(y-\mu)^2}{\sigma^2} \\ &\hat{=} A + B. \end{aligned} \tag{B.10}$$

For any nonnegative and measurable $g(\cdot)$, from (B.8), we have

$$\begin{aligned} E\{g(Y)\} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(\mu + a_1x_1 + a_2x_2)\cdot\phi(x_1)\phi(x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-0.5x_2^2}\left\{\int_{-\infty}^{\infty} g(\mu + a_1x_1 + a_2x_2)\,\mathrm{e}^{-0.5x_1^2}\,\mathrm{d}x_1\right\}\mathrm{d}x_2 \\ &\quad \Big[\text{ let } y = \mu + a_1x_1 + a_2x_2 \;\Rightarrow\; x_1 = (y-\mu-a_2x_2)/a_1 \\ &\quad\quad \Rightarrow\; \mathrm{d}x_1 = \frac{1}{|a_1|}\,\mathrm{d}y \quad \Big] \\ &= \frac{1}{2\pi|a_1|}\int_{-\infty}^{\infty} \mathrm{e}^{-0.5x_2^2}\left\{\int_{-\infty}^{\infty} g(y)\,\mathrm{e}^{-0.5(y-\mu-a_2x_2)^2/a_1^2}\,\mathrm{d}y\right\}\mathrm{d}x_2 \\ &= \frac{1}{2\pi|a_1|}\int_{-\infty}^{\infty} g(y)\left\{\int_{-\infty}^{\infty} \mathrm{e}^{-0.5[x_2^2 + (y-\mu-a_2x_2)^2/a_1^2]}\,\mathrm{d}x_2\right\}\mathrm{d}y \\ &\overset{(B.10)}{=} \frac{1}{2\pi|a_1|}\int_{-\infty}^{\infty} g(y)\left\{\int_{-\infty}^{\infty} \exp[-0.5(A+B)]\,\mathrm{d}x_2\right\}\mathrm{d}y \\ &= \frac{1}{2\pi|a_1|}\int_{-\infty}^{\infty} g(y)\cdot\exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}\cdot\frac{\sqrt{2\pi}|a_1|}{\sigma}\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty} g(y)\cdot\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}\mathrm{d}y, \end{aligned}$$

indicating (B.9). □

Corollary B.1 (Linear combination of two independent normal r.v.s). If $X_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1, 2$ and $X_1 \perp\!\!\!\perp X_2$, then

$$Y \hat{=} a_0 + a_1X_1 + a_2X_2 \sim N(a_0 + a_1\mu_1 + a_2\mu_2,\; a_1^2\sigma_1^2 + a_2^2\sigma_2^2). \tag{B.11}$$

Proof. In fact, set $Z_i = (X_i - \mu_i)/\sigma_i$ for $i = 1, 2$, we have $X_i = \mu_i + \sigma_i Z_i$

and $Z_1, Z_2 \overset{\text{iid}}{\sim} N(0,1)$. Now, we can rewrite

$$\begin{aligned} Y &= a_0 + a_1(\mu_1 + \sigma_1 Z_1) + a_2(\mu_2 + \sigma_2 Z_2) \\ &= a_0 + a_1\mu_1 + a_2\mu_2 + a_1\sigma_1 Z_1 + a_2\sigma_2 Z_2 \\ &\overset{(B.9)}{\sim} N(a_0 + a_1\mu_1 + a_2\mu_2,\ a_1^2\sigma_1^2 + a_2^2\sigma_2^2), \end{aligned}$$

which completes the proof of (B.11) □

Example B.6 (Example 2.9 revisited). Let $X_1, X_2 \overset{\text{iid}}{\sim} U[0,1]$. Find the distribution of $Y = h(X_1, X_2) = X_1 + X_2$.

Solution. For any nonnegative and measurable $g(\cdot)$, from (B.8), we have

$$\begin{aligned} E\{g(Y)\} &= \int_0^1 \int_0^1 g(h(x_1, x_2)) \cdot f_{(X_1,X_2)}(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 \\ &= \int_0^1 \int_0^1 g(x_1 + x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int_0^1 \left\{ \int_0^1 g(x_1 + x_2)\,\mathrm{d}x_1 \right\} \mathrm{d}x_2 \\ &\qquad [\text{let } y = x_1 + x_2 \ \Rightarrow\ \mathrm{d}x_1 = \mathrm{d}y \quad \text{and} \quad x_2 \leqslant y \leqslant x_2 + 1] \\ &= \int_0^1 \left\{ \int_{x_2}^{x_2+1} g(y)\,\mathrm{d}y \right\} \mathrm{d}x_2 \qquad [\text{exchange } y\ \&\ x_2] \\ &= \int g(y) \left[\int \left\{ I(0 \leqslant y \leqslant 1) + I(1 < y \leqslant 2) \right\} \mathrm{d}x_2 \right] \mathrm{d}y \\ &= \int g(y) \left\{ \int_0^y \mathrm{d}x_2 \cdot I(0 \leqslant y \leqslant 1) + \int_{y-1}^1 \mathrm{d}x_2 \cdot I(1 < y \leqslant 2) \right\} \mathrm{d}y \\ &= \int g(y) \cdot \left\{ y \cdot I(0 \leqslant y \leqslant 1) + (2 - y) \cdot I(1 < y \leqslant 2) \right\} \mathrm{d}y, \end{aligned}$$

implying that Y follows the triangle distribution. ‖

Example B.7 (Theorem 2.2 revisited). Let $Z \sim N(0,1)$, $Y \sim \chi^2(n)$ and $Z \perp\!\!\!\perp Y$. Find the distribution of $T = h(Z, Y) = Z/\sqrt{Y/n}$.

Solution. Let $\phi(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-z^2/2}$ and $f_Y(y)$ denote the densities of Z and Y, respectively. For any nonnegative and measurable $g(\cdot)$, from (B.8), we have

$$E\{g(T)\} = \int_0^\infty \int_{-\infty}^\infty g(h(z, y)) \cdot f_{(Z,Y)}(z, y)\,\mathrm{d}z\,\mathrm{d}y$$

$$= \int_0^\infty f_Y(y) \left\{ \int_{-\infty}^\infty g(z/\sqrt{y/n}) \cdot \phi(z)\, \mathrm{d}z \right\} \mathrm{d}y \qquad [\text{let } t = z/\sqrt{y/n}]$$

$$= \int_0^\infty f_Y(y) \left\{ \int_{-\infty}^\infty g(t) \cdot \phi\Big(t\sqrt{y/n}\Big) \sqrt{y/n}\, \mathrm{d}t \right\} \mathrm{d}y \quad [\text{exchange } t \text{ and } y]$$

$$= \int_{-\infty}^\infty g(t) \cdot \left\{ \int_0^\infty \phi\Big(t\sqrt{y/n}\Big) \cdot \sqrt{y/n} \cdot f_Y(y)\, \mathrm{d}y \right\} \mathrm{d}t.$$

Hence, the density of T is given by

$$\begin{aligned}
f_T(t) &= \int_0^\infty \phi\Big(t\sqrt{y/n}\Big) \cdot \sqrt{y/n} \cdot f_Y(y)\, \mathrm{d}y \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-t^2 y/(2n)} \cdot \sqrt{y/n} \cdot \frac{(1/2)^{n/2}}{\Gamma(n/2)} y^{\frac{n}{2}-1} \mathrm{e}^{-y/2}\, \mathrm{d}y \\
&= \frac{1}{\sqrt{2\pi n}} \cdot \frac{(1/2)^{n/2}}{\Gamma(n/2)} \cdot \int_0^\infty y^{\frac{n+1}{2}-1} \mathrm{e}^{-y(\frac{1}{2}+\frac{t^2}{2n})}\, \mathrm{d}y \\
&= \frac{(1/2)^{(n+1)/2}}{\sqrt{\pi n}\,\Gamma(n/2)} \cdot \frac{\Gamma(\frac{n+1}{2})}{(\frac{1}{2}+\frac{t^2}{2n})^{\frac{n+1}{2}}} \qquad [\text{use } \Gamma(\cdot) \text{ function}] \\
&= \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\,\Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}},
\end{aligned}$$

which completes the derivation and we denote it by $T \sim t(n)$. ‖

Example B.8 (Theorem 2.3 revisited). Let $U \sim \chi^2(m)$, $V \sim \chi^2(n)$ and $U \perp\!\!\!\perp V$. Find the distribution of $W = h(U, V) = (U/m)/(V/n)$.

Solution. Let $f_U(u)$ and $f_V(v)$ denote the densities of U and V, respectively. For any nonnegative and measurable $g(\cdot)$, from (B.8), we have

$$E\{g(W)\} = \int_0^\infty \int_0^\infty g(h(u,v)) \cdot f_{(U,V)}(u,v)\, \mathrm{d}u\, \mathrm{d}v$$

$$= \int_0^\infty f_V(v) \left\{ \int_0^\infty g(nu/(mv)) \cdot f_U(u)\, \mathrm{d}u \right\} \mathrm{d}v \qquad [\text{let } w = nu/(mv)]$$

$$= \int_0^\infty f_V(v) \left\{ \int_0^\infty g(w) \cdot f_U\Big(\frac{mv}{n} w\Big) \frac{mv}{n}\, \mathrm{d}w \right\} \mathrm{d}v \quad [\text{exchange } w \text{ and } v]$$

$$= \int_0^\infty g(w) \left\{ \int_0^\infty f_U\Big(\frac{mv}{n} w\Big) \cdot \frac{mv}{n} \cdot f_V(v)\, \mathrm{d}v \right\} \mathrm{d}w.$$

Hence, the density of W is given by

$$
\begin{aligned}
f_W(w) &= \int_0^\infty f_U\Big(\frac{mv}{n}w\Big)\cdot\frac{mv}{n}\cdot f_V(v)\,\mathrm{d}v \\
&= \int_0^\infty \frac{(\frac{1}{2})^{m/2}}{\Gamma(\frac{m}{2})}\Big(\frac{mv}{n}w\Big)^{\frac{m}{2}-1}\mathrm{e}^{-\frac{mvw}{2n}}\cdot\frac{mv}{n}\cdot\frac{(\frac{1}{2})^{n/2}}{\Gamma(\frac{n}{2})}v^{\frac{n}{2}-1}\,\mathrm{e}^{-v/2}\,\mathrm{d}v \\
&= \frac{(\frac{1}{2})^{(m+n)/2}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})}\Big(\frac{m}{n}\Big)^{m/2}w^{m/2-1}\cdot\int_0^\infty v^{\frac{m+n}{2}-1}\,\mathrm{e}^{-v(\frac{1}{2}+\frac{mw}{2n})}\,\mathrm{d}v \\
&= \frac{(\frac{1}{2})^{(m+n)/2}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})}\Big(\frac{m}{n}\Big)^{m/2}w^{m/2-1}\cdot\frac{\Gamma(\frac{m+n}{2})}{(\frac{1}{2}+\frac{mw}{2n})^{\frac{m+n}{2}}} \\
&= \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})}\Big(\frac{m}{n}\Big)^{m/2}w^{m/2-1}\Big(1+\frac{m}{n}w\Big)^{-\frac{m+n}{2}},
\end{aligned}
$$

which completes the derivation and we denote it by $F \sim F(m,n)$. ‖

B.1.3 The multiple-dimensional case

7• The expectation technique

- Let $f_{\mathbf{x}}(\boldsymbol{x})$ be the joint density of $\mathbf{x} = (X_1,\dots,X_n)^\top$, where $\boldsymbol{x} = (x_1,\dots,x_n)^\top$ are realizations of $\mathbf{x}$. Define $Y = h(X_1,\dots,X_n) = h(\mathbf{x})$ for some function $h(\cdot)$.
- For any nonnegative and measurable $g(\cdot)$, if

$$
\begin{aligned}
E\{g(Y)\} &= E\{g(h(X_1,\dots,X_n))\} = E\{g(h(\mathbf{x}))\} \\
&= \int g(h(\boldsymbol{x}))\cdot f_{\mathbf{x}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \qquad \text{(B.12)} \\
&= \int g(y)\cdot f_Y(y)\,\mathrm{d}y,
\end{aligned}
$$

then we can claim that $f_Y(y)$ is the pdf of Y. ‖

8• The first useful integral identity with $\boldsymbol{x} \in \mathbb{R}_+^n$

- Let $\boldsymbol{x} = (x_1,\dots,x_n)^\top \in \mathbb{R}_+^n \,\hat{=}\, \{(x_1,\dots,x_n)^\top\colon x_i > 0,\ i = 1,\dots,n\}$. Let $\boldsymbol{a} = (a_1,\dots,a_n)^\top$, $a_i > 0$ for $i = 1,\dots,n$ and $a = \sum_{i=1}^n a_i$. For any nonnegative function $f(\cdot)$, we have

$$\begin{aligned} I_n(\boldsymbol{a}; f) &\hat{=} \int_0^\infty \cdots \int_0^\infty f\left(\sum_{i=1}^n x_i\right)\left(\prod_{i=1}^n x_i^{a_i-1}\right) \mathrm{d}x_1 \cdots \mathrm{d}x_n \\ &= \int_{\mathbb{R}_+^n} f(\mathbf{1}_n^\top \boldsymbol{x})\left(\prod_{i=1}^n x_i^{a_i-1}\right) \mathrm{d}\boldsymbol{x} \\ &= \frac{\prod_{i=1}^n \Gamma(a_i)}{\Gamma(a)} \int_0^\infty f(y) y^{a-1}\,\mathrm{d}y \qquad \text{(B.13)} \\ &= B(a_1, \ldots, a_n) \cdot I_1(a; f), \end{aligned}$$

where $B(a_1, \ldots, a_n) = \prod_{i=1}^n \Gamma(a_i)/\Gamma(a)$ denotes the multivariate beta function, see **8.2**$^\bullet$ below.

8.1$^\bullet$ Proof of (B.13)

— Let $y_i = x_i$ for $i = 1, \ldots, n-1$ and $y_n = \sum_{i=1}^n x_i$, then we have

$$x_i = y_i, \quad i = 1, \ldots, n-1 \quad \text{and} \quad x_n = y_n - \sum_{i=1}^{n-1} y_i.$$

— Since the Jacobian determinant

$$J(x_1, \ldots, x_n \to y_1, \ldots, y_n) = \det\left(\frac{\partial x_i}{\partial y_j}\right) = 1,$$

we obtain

$$\begin{aligned} &I_n(\boldsymbol{a}; f) \\ &= \int_0^\infty f(y_n)\left\{\int_{\mathbb{V}_{n-1}(y_n)}\left(\prod_{i=1}^{n-1} y_i^{a_i-1}\right)\left(y_n - \sum_{i=1}^{n-1} y_i\right)^{a_n-1} \mathrm{d}\boldsymbol{y}_{-n}\right\} \mathrm{d}y_n, \end{aligned}$$

where $\boldsymbol{y}_{-n} = (y_1, \ldots, y_{n-1})^\top$ and for any $c > 0$,

$$\mathbb{V}_{n-1}(c) \;\hat{=}\; \{(y_1, \ldots, y_{n-1})^\top\!: \ y_i > 0,\ \textstyle\sum_{i=1}^{n-1} y_i < c\}.$$

— Furthermore, let $u_i = y_i y_n^{-1}$ for $i = 1, \ldots, n-1$, then

$$J(y_1, \ldots, y_{n-1} \to u_1, \ldots, u_{n-1}) = J(\boldsymbol{y}_{-n} \to \boldsymbol{u}_{-n}) = y_n^{n-1},$$

so that

$$
\begin{aligned}
& I_n(\boldsymbol{a}; f) \\
= \; & \int_0^\infty f(y_n) y_n^{a-1} \left\{ \int_{\mathbb{V}_{n-1}} \left(\prod_{i=1}^{n-1} u_i^{a_i - 1} \right) \left(1 - \sum_{i=1}^{n-1} u_i \right)^{a_n - 1} \mathrm{d}\boldsymbol{u}_{-n} \right\} \mathrm{d}y_n \\
\stackrel{(B.14)}{=} \; & B(a_1, \ldots, a_n) \int_0^\infty f(y) y^{a-1} \, \mathrm{d}y,
\end{aligned}
$$

indicating (B.13), where $\mathbb{V}_{n-1} \hat{=} \mathbb{V}_{n-1}(1)$. □

8.2• The multivariate beta function and Dirichlet distribution

— The multivariate beta function is defined as

$$
\begin{aligned}
B(a_1, \ldots, a_n) \; & \hat{=} \; \int_{\mathbb{V}_{n-1}} \left(\prod_{i=1}^{n-1} u_i^{a_i - 1} \right) \left(1 - \sum_{i=1}^{n-1} u_i \right)^{a_n - 1} \mathrm{d}u_1 \cdots \mathrm{d}u_{n-1} \\
& = \; \frac{\Gamma(a_1) \cdots \Gamma(a_n)}{\Gamma(a_1 + \cdots + a_n)}, \qquad a_i > 0. \qquad \text{(B.14)}
\end{aligned}
$$

This integral identity can be used to define the Dirichlet distribution.

— A random vector $\mathbf{u}_{-n} = (U_1, \ldots, U_{n-1})^\top \sim \text{Dirichlet}(a_1, \ldots, a_{n-1}; a_n)$ or $\mathbf{u}_{-n} \sim \text{Dirichlet}_n(\boldsymbol{a})$ if its joint density is given by

$$
\frac{1}{B(a_1, \ldots, a_n)} \left(\prod_{i=1}^{n-1} u_i^{a_i - 1} \right) \left(1 - \sum_{i=1}^{n-1} u_i \right)^{a_n - 1}, \qquad \text{(B.15)}
$$

for $\boldsymbol{u}_{-n} = (u_1, \ldots, u_{n-1})^\top \in \mathbb{V}_{n-1}$. In particular, when $n = 2$, we have $\text{Dirichlet}(a_1; a_2) = \text{Beta}(a_1, a_2)$.

8.3• A special case of (B.13)

— In (B.13), let all $a_i = 1$, we have

$$
\int_0^\infty \cdots \int_0^\infty f\left(\sum_{i=1}^n x_i \right) \mathrm{d}x_1 \cdots \mathrm{d}x_n = \frac{1}{\Gamma(n)} \int_0^\infty f(y) y^{n-1} \, \mathrm{d}y. \qquad \text{(B.16)}
$$

8.4• Constructing multivariate distributions via (B.13)

— We can rewrote the integral identity (B.13) as

$$\int_{\mathbb{R}_+^n} \frac{f(\mathbf{1}_n^\top \boldsymbol{x}) \prod_{i=1}^n x_i^{a_i-1}}{B(a_1,\ldots,a_n)\int_0^\infty f(y)y^{a-1}\,\mathrm{d}y}\,\mathrm{d}\boldsymbol{x} = 1,$$

indicating that

$$\frac{f(\mathbf{1}_n^\top \boldsymbol{x}) \prod_{i=1}^n x_i^{a_i-1}}{B(a_1,\ldots,a_n)\int_0^\infty f(y)y^{a-1}\,\mathrm{d}y}, \qquad \boldsymbol{x} \in \mathbb{R}_+^n, \tag{B.17}$$

defines a density family indexed by a constant vector $\boldsymbol{a} \in \mathbb{R}_+^n$ and a nonnegative function $f(\cdot)$, provided that $\int_0^\infty f(y)y^{a-1}\,\mathrm{d}y < +\infty$.

— Especially, let $f(y) = \mathrm{e}^{-y}$, $y > 0$, then (B.17) becomes $\prod_{i=1}^n \frac{x_i^{a_i-1}}{\Gamma(a_i)}\,\mathrm{e}^{-x_i}$; i.e., the product of n independent gamma r.v.s with $X_i \sim \mathrm{Gamma}(a_i, 1)$.

— In addition, let $f(y) = (1-y)^{a_{n+1}-1}$, $0 < y < 1$, then (B.17) reduces to the density of $\mathrm{Dirichlet}(a_1,\ldots,a_n; a_{n+1})$ for $\boldsymbol{x} \in \mathbb{V}_n$.

Example B.9 (Sum of independent gamma r.v.s with a common β). Let $\{X_i\}_{i=1}^n \overset{\text{ind}}{\sim} \mathrm{Gamma}(a_i, \beta)$ with $a_i > 0$ and $\beta > 0$. Find the distribution of $Y = h(X_1,\ldots,X_n) = \sum_{i=1}^n X_i$.

Solution. For any nonnegative and measurable $g(\cdot)$, we have

$$\begin{aligned}
E\{g(Y)\} &\overset{\text{(B.12)}}{=} \int_{\mathbb{R}_+^n} g\left(\sum_{i=1}^n x_i\right)\left\{\prod_{i=1}^n \frac{\beta^{a_i}}{\Gamma(a_i)} x_i^{a_i-1}\,\mathrm{e}^{-\beta x_i}\right\}\mathrm{d}\boldsymbol{x} \\
&= \frac{\beta^a}{\prod_{i=1}^n \Gamma(a_i)} \int_{\mathbb{R}_+^n} \underbrace{g(\mathbf{1}_n^\top \boldsymbol{x})\exp(-\beta \mathbf{1}_n^\top \boldsymbol{x})}_{f(y)\,\hat{=}\,g(y)\exp(-\beta y)} \left(\prod_{i=1}^n x_i^{a_i-1}\right)\mathrm{d}\boldsymbol{x} \\
&\overset{\text{(B.13)}}{=} \frac{\beta^a}{\prod_{i=1}^n \Gamma(a_i)} \cdot \frac{\prod_{i=1}^n \Gamma(a_i)}{\Gamma(a)} \int_0^\infty g(y)\,\mathrm{e}^{-\beta y} y^{a-1}\,\mathrm{d}y \\
&= \int_0^\infty g(y) \cdot \frac{\beta^a}{\Gamma(a)} y^{a-1}\,\mathrm{e}^{-\beta y}\,\mathrm{d}y,
\end{aligned}$$

implying that $Y \sim \mathrm{Gamma}(a, \beta)$ with $a = \sum_{i=1}^n a_i$. ‖

9• The second useful integral identity with $\boldsymbol{x} \in \mathbb{R}^n$

- Let $\boldsymbol{x} = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n \,\hat{=}\, \{(x_1, \ldots, x_n)^\top \colon x_i \in \mathbb{R},\ i = 1, \ldots, n\}$. Let $\boldsymbol{a} = (a_1, \ldots, a_n)^\top$, $a_i > 0$ for $i = 1, \ldots, n$ and $a = \sum_{i=1}^n a_i$. For any nonnegative function $f(\cdot)$, we have

$$\begin{aligned} J_n(\boldsymbol{a}; f) &\,\hat{=}\, \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f\left(\sum_{i=1}^n x_i^2\right) \left(\prod_{i=1}^n |x_i|^{2a_i - 1}\right) \mathrm{d}x_1 \cdots \mathrm{d}x_n \\ &= \int_{\mathbb{R}^n} f(\boldsymbol{x}^\top \boldsymbol{x}) \left(\prod_{i=1}^n |x_i|^{2a_i - 1}\right) \mathrm{d}\boldsymbol{x} \\ &= B(a_1, \ldots, a_n) \int_0^\infty f(y) y^{a-1} \,\mathrm{d}y = I_n(\boldsymbol{a}; f). \end{aligned} \tag{B.18}$$

9.1• Proof of (B.18)

— Let $y_i = x_i^2$ for $i = 1, \ldots, n$. When $x_i > 0$, we have $\mathrm{d}x_i = \frac{1}{2\sqrt{y_i}} \,\mathrm{d}y_i$. Thus

$$\begin{aligned} J_n(\boldsymbol{a}; f) &= 2^n \int_{\mathbb{R}_+^n} f\left(\sum_{i=1}^n x_i^2\right) \left(\prod_{i=1}^n |x_i|^{2a_i - 1}\right) \mathrm{d}x_1 \cdots \mathrm{d}x_n \\ &= 2^n \int_{\mathbb{R}_+^n} f\left(\sum_{i=1}^n y_i\right) \left(\prod_{i=1}^n y_i^{a_i - 1/2}\right) \cdot \prod_{i=1}^n \left(\frac{1}{2\sqrt{y_i}} \,\mathrm{d}y_i\right) \\ &= \int_{\mathbb{R}_+^n} f(\mathbf{1}_n^\top \boldsymbol{y}) \left(\prod_{i=1}^n y_i^{a_i - 1}\right) \mathrm{d}\boldsymbol{y} \overset{\text{(B.13)}}{=} I_n(\boldsymbol{a}; f), \end{aligned}$$

which completes the proof of (B.18). □

9.2• A special case of (B.18)

— In (B.18), let all $a_i = 1/2$, we have

$$\int_{\mathbb{R}^n} f(\boldsymbol{x}^\top \boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} = \frac{\pi^{n/2}}{\Gamma(n/2)} \int_0^\infty f(y) y^{n/2 - 1} \,\mathrm{d}y. \tag{B.19}$$

9.3• Constructing multivariate distributions via (B.18)

— We can rewrote the integral identity (B.18) as

$$\int_{\mathbb{R}^n} \frac{f(\boldsymbol{x}^\top \boldsymbol{x}) \prod_{i=1}^n |x_i|^{2a_i-1}}{B(a_1,\ldots,a_n)\int_0^\infty f(y) y^{a-1}\,\mathrm{d}y}\,\mathrm{d}\boldsymbol{x} = 1,$$

indicating that

$$\frac{f(\boldsymbol{x}^\top \boldsymbol{x}) \prod_{i=1}^n |x_i|^{2a_i-1}}{B(a_1,\ldots,a_n)\int_0^\infty f(y) y^{a-1}\,\mathrm{d}y}, \qquad \boldsymbol{x} \in \mathbb{R}^n, \tag{B.20}$$

defines a density family indexed by a constant vector $\boldsymbol{a} \in \mathbb{R}_+^n$ and a nonnegative function $f(\cdot)$, provided that $\int_0^\infty f(y)y^{a-1}\,\mathrm{d}y < +\infty$.

— Especially, let all $a_i = 1/2$, then (B.20) has the form of $f(\boldsymbol{x}^\top \boldsymbol{x})$. Note that the density of $N_n(\mathbf{0}, \boldsymbol{I}_n)$ is of the form of $\exp(-\boldsymbol{x}^\top \boldsymbol{x}/2)$.

— An $n \times 1$ random vector $\mathbf{x} = (X_1,\ldots,X_n)^\top$ is said to have a spherically symmetric distribution (or simply *spherical distribution*) iff $\boldsymbol{\Gamma}\mathbf{x} \overset{\mathrm{d}}{=} \mathbf{x}$, for every $\boldsymbol{\Gamma} \in \mathcal{O}(n)$, where $\mathcal{O}(n) = \{\boldsymbol{\Gamma}_{n\times n}\colon \boldsymbol{\Gamma}^\top\boldsymbol{\Gamma} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top = \boldsymbol{I}_n\}$ denotes the set of $n \times n$ orthogonal matrices.

— An n-vector $\mathbf{x}$ has a spherical distribution, denoted by $\mathbf{x} \sim S_n(\varphi)$ iff its characteristic function $\psi(\boldsymbol{t}) = E\{\exp(\mathrm{i}\boldsymbol{t}^\top\mathbf{x})\}$ satisfies that there exists a scalar function $\varphi(\cdot)$ such that $\psi(\boldsymbol{t}) = \varphi(\boldsymbol{t}^\top\boldsymbol{t})$.

— In general, $\mathbf{x} \sim S_n(\varphi)$ does not necessarily possess a density. If the density of $\mathbf{x}$ exists, then it must be of the form $f(\boldsymbol{x}^\top\boldsymbol{x})$ for some nonnegative function $f(\cdot)$ of scalar variable.

Example B.10 (The volume of an n-dimensional sphere). Let

$$\mathbb{B}_n(r) = \{\boldsymbol{x} = (x_1,\ldots,x_n)^\top\colon\ \boldsymbol{x} \in \mathbb{R}^n,\ \boldsymbol{x}^\top\boldsymbol{x} \leqslant r^2\}$$

denote the n-dimensional sphere with radius r. Find the volume of $\mathbb{B}_n(r)$.

Solution. Define $f(y) = I(y \leqslant r^2)$, then the volume of $\mathbb{B}_n(r)$ is given by

$$\begin{aligned}\int_{\boldsymbol{x}^\top\boldsymbol{x}\leqslant r^2} \mathrm{d}\boldsymbol{x} &= \int_{\mathbb{R}^n} I(\boldsymbol{x}^\top\boldsymbol{x} \leqslant r^2)\,\mathrm{d}\boldsymbol{x} \\ &= \int_{\mathbb{R}^n} f(\boldsymbol{x}^\top\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\end{aligned} \tag{B.21}$$

$$\overset{(B.19)}{=} \frac{\pi^{n/2}}{\Gamma(n/2)} \int_0^{\infty} f(y) y^{n/2-1}\, \mathrm{d}y \;=\; \frac{\pi^{n/2}}{\Gamma(n/2)} \int_0^{r^2} y^{n/2-1}\, \mathrm{d}y$$

$$= \frac{\pi^{n/2}}{\Gamma(n/2)} \cdot \frac{2r^n}{n} \;=\; \frac{(\pi r^2)^{n/2}}{\Gamma(n/2+1)}.$$

In particular, when $n = 2$, the area of $\mathbb{B}_2(r)$ is πr^2. ‖

Example B.11 (The density of $\chi^2(n)$). Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(0,1)$, find the distribution of $Y = h(\mathbf{x}) = \sum_{i=1}^n X_i^2$.

Solution. For any nonnegative and measurable $g(\cdot)$, we have

$$E\{g(Y)\} \overset{(B.12)}{=} \int_{\mathbb{R}^n} g(\boldsymbol{x}^\top \boldsymbol{x}) \cdot \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp(-\boldsymbol{x}^\top \boldsymbol{x}/2)\, \mathrm{d}\boldsymbol{x}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \int_{\mathbb{R}^n} \underbrace{g(\boldsymbol{x}^\top \boldsymbol{x}) \exp(-\boldsymbol{x}^\top \boldsymbol{x}/2)}_{f(y) \,\hat{=}\, g(y)\mathrm{e}^{-y/2}}\, \mathrm{d}\boldsymbol{x}$$

$$\overset{(B.19)}{=} \frac{\pi^{n/2}}{\Gamma(n/2)} \int_0^{\infty} g(y)\, \mathrm{e}^{-y/2} y^{n/2-1}\, \mathrm{d}y$$

$$= \int_0^{\infty} g(y) \cdot \frac{\pi^{n/2}}{\Gamma(n/2)} y^{n/2-1}\, \mathrm{e}^{-y/2}\, \mathrm{d}y,$$

implying that $Y \sim \text{Gamma}(n/2, 1/2) = \chi^2(n)$. ‖

10• The third useful integral identity with $\boldsymbol{x} \in \mathbb{R}^n$

- Let $\boldsymbol{x} = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n$. Let $\boldsymbol{a} = (a_1, \ldots, a_n)^\top \neq \mathbf{0}_n$ and $\|\boldsymbol{a}\|_2 = \sqrt{\boldsymbol{a}^\top \boldsymbol{a}} = (\sum_{i=1}^n a_i^2)^{1/2}$. For any nonnegative function $f(\cdot)$, we have

$$\int_{\mathbb{R}^n} f(\boldsymbol{a}^\top \boldsymbol{x}, \boldsymbol{x}^\top \boldsymbol{x})\, \mathrm{d}\boldsymbol{x} = \int_{\mathbb{R}^n} f(\|\boldsymbol{a}\|_2 x_1, \boldsymbol{x}^\top \boldsymbol{x})\, \mathrm{d}\boldsymbol{x}. \tag{B.22}$$

10.1• Proof of (B.22)

— Given $\boldsymbol{a} \neq \mathbf{0}_n$, we define $\boldsymbol{\gamma}_1 = \boldsymbol{a}/\|\boldsymbol{a}\|_2$ so that $\boldsymbol{\gamma}_1^\top \boldsymbol{\gamma}_1 = 1$; i.e., $\boldsymbol{\gamma}_1$ is a unit vector.

— Thus, we can find standard orthogonal bases $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_n$ such that $\boldsymbol{\gamma}_i^\top \boldsymbol{\gamma}_i = 1$ $(i = 1, \ldots, n)$ and $\boldsymbol{\gamma}_i^\top \boldsymbol{\gamma}_j = 0$ $(i \neq j)$.

— Let $\boldsymbol{\Gamma}^\top = (\gamma_1, \ldots, \gamma_n)$ or

$$\boldsymbol{\Gamma} = \begin{pmatrix} \gamma_1^\top \\ \vdots \\ \gamma_n^\top \end{pmatrix},$$

then $\boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top = \boldsymbol{\Gamma}^\top\boldsymbol{\Gamma} = \boldsymbol{I}_n$, i.e., $\boldsymbol{\Gamma} \in \mathcal{O}(n)$.

— Make the transformation $\boldsymbol{y} = \boldsymbol{\Gamma}\boldsymbol{x}$, we have $\boldsymbol{y}^\top\boldsymbol{y} = \boldsymbol{x}^\top\boldsymbol{x}$ and $J(\boldsymbol{x} \to \boldsymbol{y}) = 1$. Next, we have $y_1 = \boldsymbol{\gamma}_1^\top\boldsymbol{x} = \boldsymbol{a}^\top\boldsymbol{x}/\|\boldsymbol{a}\|_2$ so that $\boldsymbol{a}^\top\boldsymbol{x} = \|\boldsymbol{a}\|_2 y_1$. Thus,

$$\int_{\mathbb{R}^n} f(\boldsymbol{a}^\top\boldsymbol{x}, \boldsymbol{x}^\top\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \int_{\mathbb{R}^n} f(\|\boldsymbol{a}\|_2 y_1, \boldsymbol{y}^\top\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}. \qquad \square$$

Example B.12 (Linear combination of n i.i.d. standard normal r.v.'s). Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(0,1)$. Show that

$$Y \hat{=} \mu + \sum_{i=1}^n a_i X_i \sim N\left(\mu, \sum_{i=1}^n a_i^2\right). \tag{B.23}$$

Proof. Let $\boldsymbol{a} = (a_1, \ldots, a_n)^\top \neq \mathbf{0}_n$, $\boldsymbol{x} = (x_1, \ldots, x_n)^\top$ and $\sigma \hat{=} (\sum_{i=1}^n a_i^2)^{1/2} = \|\boldsymbol{a}\|_2$. For any nonnegative and measurable $g(\cdot)$, we have

$$\begin{aligned}
E\{g(Y)\} &\overset{\text{(B.12)}}{=} \left(\frac{1}{\sqrt{2\pi}}\right)^n \int_{\mathbb{R}^n} \underbrace{g(\mu + \boldsymbol{a}^\top\boldsymbol{x}) \cdot \exp(-\boldsymbol{x}^\top\boldsymbol{x})}_{f(\boldsymbol{a}^\top\boldsymbol{x},\, \boldsymbol{x}^\top\boldsymbol{x})}\,\mathrm{d}\boldsymbol{x} \\
&\overset{\text{(B.22)}}{=} \left(\frac{1}{\sqrt{2\pi}}\right)^n \int_{\mathbb{R}^n} g(\mu + \|\boldsymbol{a}\|_2 x_1) \cdot \exp(-\boldsymbol{x}^\top\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\
&= \int_{-\infty}^{\infty} g(\mu + \sigma x_1) \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-x_1^2/2} \left\{\int_{\mathbb{R}^{n-1}} \prod_{i=2}^n \phi(x_i)\,\mathrm{d}x_2 \cdots \mathrm{d}x_n\right\} \mathrm{d}x_1 \\
&\qquad \left[\text{ let } y = \mu + \sigma x_1 \;\Rightarrow\; x_1 = (y-\mu)/\sigma \;\Rightarrow\; \mathrm{d}x_1 = \frac{1}{\sigma}\,\mathrm{d}y \right] \\
&= \int_{-\infty}^{\infty} g(y) \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\} \mathrm{d}y,
\end{aligned}$$

indicating (B.23). $\square$

B.1.4 Joint distributions of order statistics

11• The fourth useful integral identity with $\boldsymbol{x} \in \mathbb{A}_n(a,b)$

- Let $\boldsymbol{x} = (x_1, \ldots, x_n)^\top \in \mathbb{A}_n(a,b) \,\hat{=}\, \{\boldsymbol{x}\colon\ a < x_1 < \cdots < x_n < b\}$, we have

$$I_1 \,\hat{=} \int_{\mathbb{A}_n(a,b)} \mathrm{d}x_1 \cdots \mathrm{d}x_n = \int_{\mathbb{A}_n(a,b)} \mathrm{d}\boldsymbol{x} = \frac{(b-a)^n}{n!}. \tag{B.24}$$

- Let $F(\cdot)$ be a cdf, we have

$$I_2 \,\hat{=} \int_{\mathbb{A}_n(a,b)} \mathrm{d}F(x_1) \cdots \mathrm{d}F(x_n) = \frac{\{F(b) - F(a)\}^n}{n!}. \tag{B.25}$$

11.1• Proof of (B.24)

— Method I: With direct integration one by one, we obtain

$$\begin{aligned}
I_1 &= \int_{\mathbb{A}_{n-1}(a,b)} \left(\int_a^{x_2} \mathrm{d}x_1 \right) \mathrm{d}x_2 \cdots \mathrm{d}x_n \\
&= \int_{\mathbb{A}_{n-1}(a,b)} (x_2 - a)\, \mathrm{d}x_2 \cdots \mathrm{d}x_n \\
&= \int_{\mathbb{A}_{n-2}(a,b)} \left\{ \int_a^{x_3} (x_2 - a)\, \mathrm{d}x_2 \right\} \mathrm{d}x_3 \cdots \mathrm{d}x_n \\
&= \int_{\mathbb{A}_{n-2}(a,b)} \left\{ \frac{(x_2 - a)^2}{2} \bigg|_a^{x_3} \right\} \mathrm{d}x_3 \cdots \mathrm{d}x_n \\
&= \frac{1}{2!} \int_{\mathbb{A}_{n-2}(a,b)} (x_3 - a)^2\, \mathrm{d}x_3 \cdots \mathrm{d}x_n \\
&= \frac{1}{3!} \int_{\mathbb{A}_{n-3}(a,b)} (x_4 - a)^3\, \mathrm{d}x_4 \cdots \mathrm{d}x_n \\
&\ \vdots \\
&= \frac{1}{(n-1)!} \int_a^b (x_n - a)^{n-1}\, \mathrm{d}x_n = \frac{(b-a)^n}{n!},
\end{aligned}$$

indicating (B.24). □

— Method II: Define $(a,b)^n = \{(x_1,\dots,x_n)^\top\colon a < x_i < b,\ i = 1,\dots,n\}$. By symmetry, we have

$$I_1 = \frac{1}{n!}\int_{(a,b)^n} \mathrm{d}x_1\cdots\mathrm{d}x_n = \frac{(b-a)^n}{n!}. \qquad \square$$

11.2• Proof of (B.25)

— Make the transformations $u_i = F(x_i)$ for $i = 1,\dots,n$.

— From $a < x_1 < \cdots < x_n < b$, we have $F(a) < F(x_1) < \cdots < F(x_n) < F(b)$, i.e., $F(a) < u_1 < \cdots < u_n < F(b)$. Thus

$$I_2 = \int_{\mathbb{A}_n(F(a),F(b))} \mathrm{d}u_1\cdots\mathrm{d}u_n \stackrel{\text{(B.24)}}{=} \frac{\{F(b)-F(a)\}^n}{n!}. \qquad \square$$

12• Two examples

Example B.13 (Distributions of order statistics of n i.i.d. uniform r.v.s). Let $U_1,\dots,U_n \stackrel{\text{iid}}{\sim} U(0,1)$ and $U_{(1)},\dots,U_{(n)}$ be their order statistics.

1) Find the density of $U_{(r)}$ for $r = 1,\dots,n$.

2) Find the joint density of $U_{(r_1)},\dots,U_{(r_k)}$ for $1 \leqslant r_1 < \cdots < r_k \leqslant n$.

Solution. 1) Let $\mathbf{u} = (U_1,\dots,U_n)^\top$ and $\boldsymbol{u} = (u_1,\dots,u_n)^\top$. The joint density of $\mathbf{u}$ is $f_{\mathbf{u}}(\boldsymbol{u}) = 1$ for $\boldsymbol{u} \in (0,1)^n$. Note that $U_{(r)}$ is a function of $\mathbf{u}$. For any nonnegative and measurable $g(\cdot)$, we have

$$\begin{aligned}
& E\{g(U_{(r)})\} \stackrel{\text{(B.12)}}{=} \int_{(0,1)^n} g(u_{(r)})\cdot f_{\mathbf{u}}(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} \\
= {} & \int_{(0,1)^n} g(u_{(r)})\,\mathrm{d}u_1\cdots\mathrm{d}u_n \qquad \text{[by symmetry]} \\
= {} & n!\int_{\mathbb{A}_n(0,1)} g(u_{(r)})\,\mathrm{d}u_{(1)}\cdots\mathrm{d}u_{(n)} = n!\int_{\mathbb{A}_n(0,1)} g(u_r)\,\mathrm{d}u_1\cdots\mathrm{d}u_n \\
= {} & n!\int_0^1 g(u_r)\left\{\int_{\mathbb{A}_{r-1}(0,u_r)} \mathrm{d}u_1\cdots\mathrm{d}u_{r-1}\right\}\left\{\int_{\mathbb{A}_{n-r}(u_r,1)} \mathrm{d}u_{r+1}\cdots\mathrm{d}u_n\right\}\mathrm{d}u_r \\
\stackrel{\text{(B.24)}}{=} {} & n!\int_0^1 g(u_r)\cdot\frac{u_r^{r-1}}{(r-1)!}\cdot\frac{(1-u_r)^{n-r}}{(n-r)!}\,\mathrm{d}u_r \\
= {} & \int_0^1 g(u)\cdot\left\{\frac{n!}{(r-1)!(n-r)!}u^{r-1}(1-u)^{n-r}\right\}\mathrm{d}u_r,
\end{aligned}$$

indicating that $U_{(r)} \sim \text{Beta}(r, n-r+1)$.

2) For any nonnegative and measurable $g(\cdot)$, we have

$$
\begin{aligned}
E\{g(U_{(r_1)}, \ldots, U_{(r_k)})\} &= \int_{(0,1)^n} g(u_{(r_1)}, \ldots, u_{(r_k)}) f_{\mathbf{u}}(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u} \\
&\overset{(B.12)}{=} \int_{(0,1)^n} g(u_{(r_1)}, \ldots, u_{(r_k)})\, \mathrm{d}u_1 \cdots \mathrm{d}u_n \qquad \text{[by symmetry]} \\
&= n! \int_{\mathbb{A}_n(0,1)} g(u_{(r_1)}, \ldots, u_{(r_k)})\, \mathrm{d}u_{(1)} \cdots \mathrm{d}u_{(n)} \\
&= n! \int_{\mathbb{A}_n(0,1)} g(u_{r_1}, \ldots, u_{r_k})\, \mathrm{d}u_1 \cdots \mathrm{d}u_n \\
&= n! \int_{\mathbb{A}_k(0,1)} g(u_{r_1}, \ldots, u_{r_k}) \left\{ \int_{\mathbb{A}_{r_1-1}(0,u_{r_1})} \mathrm{d}u_1 \cdots \mathrm{d}u_{r_1-1} \right\} \\
&\quad \times \left\{ \int_{\mathbb{A}_{r_2-r_1-1}(u_{r_1},u_{r_2})} \mathrm{d}u_{r_1+1} \cdots \mathrm{d}u_{r_2-1} \right\} \times \cdots \\
&\quad \times \left\{ \int_{\mathbb{A}_{r_k-r_{k-1}-1}(u_{r_{k-1}},u_{r_k})} \mathrm{d}u_{r_{k-1}+1} \cdots \mathrm{d}u_{r_k-1} \right\} \\
&\quad \times \left\{ \int_{\mathbb{A}_{n-r_k}(u_{r_k},1)} \mathrm{d}u_{r_k+1} \cdots \mathrm{d}u_n \right\} \mathrm{d}u_{r_1} \cdots \mathrm{d}u_{r_k} \\
&\overset{(B.24)}{=} n! \int_{\mathbb{A}_k(0,1)} g(u_{r_1}, \ldots, u_{r_k}) \cdot \frac{u_{r_1}^{r_1-1}}{(r_1-1)!} \cdot \frac{(u_{r_2}-u_{r_1})^{r_2-r_1-1}}{(r_2-r_1-1)!} \\
&\quad \times \frac{(u_{r_k}-u_{r_{k-1}})^{r_k-r_{k-1}-1}}{(r_k-r_{k-1}-1)!} \cdot \frac{(1-u_{r_k})^{n-r_k}}{(n-r_k)!}\, \mathrm{d}u_{r_1} \cdots \mathrm{d}u_{r_k} \\
&= \int_{\mathbb{A}_k(0,1)} g(u_1, \ldots, u_k) \left\{ n! \frac{u_1^{r_1-1}}{(r_1-1)!} \cdot \frac{(u_2-u_1)^{r_2-r_1-1}}{(r_2-r_1-1)!} \right. \\
&\quad \left. \times \frac{(u_k-u_{k-1})^{r_k-r_{k-1}-1}}{(r_k-r_{k-1}-1)!} \cdot \frac{(1-u_k)^{n-r_k}}{(n-r_k)!} \right\} \mathrm{d}u_1 \cdots \mathrm{d}u_k,
\end{aligned}
$$

indicating that the joint density of $U_{(r_1)}, \ldots, U_{(r_k)}$ is

$$
n! \prod_{i=1}^{k+1} \frac{(u_i - u_{i-1})^{r_i - r_{i-1} - 1}}{(r_i - r_{i-1} - 1)!}, \quad 0 < u_1 < \cdots < u_k < 1, \tag{B.26}
$$

where $u_0 = 0$, $u_{k+1} = 1$, $r_0 = 0$ and $r_{k+1} = n + 1$. ‖

Example B.14 (Distribution of order statistics of n i.i.d. random variables). Let $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} F(x)$ and $X_{(1)}, \ldots, X_{(n)}$ be their order statistics. Find the density of $X_{(r)}$ for $r = 1, \ldots, n$.

Solution. From Q1.12, we know that $X \stackrel{\text{d}}{=} F^{-1}(U)$, where $X \sim F(\cdot)$ and $U \sim U(0, 1)$. Thus, the cdf of $X_{(r)}$ is

$$\begin{aligned}
G_r(x) &= \Pr(X_{(r)} \leqslant x) \qquad [\because X_{(r)} \stackrel{\text{d}}{=} F^{-1}(U_{(r)})\,] \\
&= \Pr\{F^{-1}(U_{(r)}) \leqslant x\} \\
&= \Pr\{U_{(r)} \leqslant F(x)\} \qquad [\because U_{(r)} \sim \text{Beta}(r, n-r+1)\,] \\
&= \int_0^{F(x)} \frac{1}{B(r, n-r+1)} u^{r-1}(1-u)^{n-r}\, \text{d}u,
\end{aligned}$$

indicating (2.20). The pdf of $X_{(r)}$ is

$$g_r(x) = G_r'(x) = \frac{n!}{(r-1)!(n-r)!} f(x) F^{r-1}(x)\{1 - F(x)\}^{n-r}, \tag{B.27}$$

which is (2.23). ‖

B.2 Discrete Random Variables

B.2.1 The one-dimensional case

13• Formulation of the expectation technique

13.1• The motivation

— If a discrete r.v. $X \sim p_X(x)$, then we have

$$E\{g(X)\} = \sum_x g(x) \cdot p_X(x) \tag{B.28}$$

for an arbitrary function $g(\cdot)$ provided that $E\{|g(X)|\} < +\infty$.

— Inversely, if (B.28) is true for any function $g(\cdot)$, we wonder whether $p_X(x)$ is the pmf of X?

13.2• The expectation technique

— Let $X \sim p_X(x)$ and $Y = h(X)$. For any nonnegative and measurable $g(\cdot)$, of course, we have

$$E\{g(Y)\} = E\{g(h(X))\} \overset{\text{(B.28)}}{=} \sum_x g(h(x)) \cdot p_X(x). \tag{B.29}$$

— Make the transformation of $y = h(x)$, if we can express the last summation in (B.29) as $\sum_y g(y)p_Y(y)$; i.e.,

$$E\{g(Y)\} = \sum_x g(h(x)) \cdot p_X(x) = \sum_y g(y) \cdot p_Y(y), \tag{B.30}$$

then we can claim that $p_Y(y)$ is the pmf of Y. ‖

13.3• The first proof of the expectation technique

— Let $g(Y) = z^Y$. On the one hand, we have

$$E\{g(Y)\} = E(z^Y) = G_Y(z), \tag{B.31}$$

which is the pgf of Y.

— On the other hand, (B.30) can be rewritten as

$$\begin{aligned} E\{g(Y)\} &= \sum_y g(y) \cdot p_Y(y) \qquad [\text{since } g(y) = z^y] \\ &= \sum_y z^y \cdot p_Y(y). \end{aligned} \tag{B.32}$$

— By combining (B.31) with (B.32), we obtain

$$G_Y(z) = \sum_y z^y \cdot p_Y(y),$$

indicating that $p_Y(y)$ is the pmf of Y. □

13.4• The second proof of the expectation technique

— Let $g(Y) = \mathrm{e}^{tY}$. On the one hand, we have

$$E\{g(Y)\} = E(\mathrm{e}^{tY}) = M_Y(t), \tag{B.33}$$

which is the mgf of Y.

— On the other hand, (B.30) can be rewritten as

$$\begin{aligned} E\{g(Y)\} &= \sum_y g(y) \cdot p_Y(y) \qquad [\text{since } g(y) = \mathrm{e}^{ty}] \\ &= \sum_y \mathrm{e}^{ty} \cdot p_Y(y). \end{aligned} \tag{B.34}$$

— By combining (B.33) with (B.34), we obtain

$$M_Y(t) = \sum_y \mathrm{e}^{ty} \cdot p_Y(y),$$

indicating that $p_Y(y)$ is the pmf of Y. □

13.5• Comments

— From the process of the first proof, we know that the pgf technique is a special case of the expectation technique with $g(Y) = z^Y$.

— From the process of the second proof, we know that the mgf technique is also a special case of the expectation technique with $g(Y) = \mathrm{e}^{tY}$.

14• Two examples

Example B.15 (Example 2.3 revisited). 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. For any nonnegative and measurable $g(\cdot)$, from (B.29), we have

$$\begin{aligned} E\{g(Y)\} &= \sum_{x=0}^{5} g(h(x)) \cdot p_X(x) = \sum_{x=0}^{5} g((x-2)^2) \cdot p_X(x) \\ &= g(4) \cdot p_X(0) + g(1) \cdot p_X(1) + g(0) \cdot p_X(2) \\ &\quad + g(1) \cdot p_X(3) + g(4) \cdot p_X(4) + g(9) \cdot p_X(5) \end{aligned}$$

$$
\begin{aligned}
&= g(0)\cdot p_2 + g(1)\cdot(p_1+p_3) + g(4)\cdot(p_0+p_4) + g(9)\cdot p_5 \\
&= \sum_{y\in\mathcal{S}_Y} g(y)\cdot p_{_Y}(y),
\end{aligned}
$$

where $\mathcal{S}_Y = \{0,1,4,9\}$ is the support of Y. Then, the pmf of Y, $p_{_Y}(y) = \Pr(Y=y)$, is given by p_2, p_1+p_3, p_0+p_4 and p_5, respectively. ‖

Example B.16 (Poisson distribution). Let $X \sim \text{Poisson}(\lambda)$. Find the pmf of $Y = h(X) = 2X$.

Solution. For any nonnegative and measurable $g(\cdot)$, from (B.29), we have

$$
\begin{aligned}
E\{g(Y)\} &= \sum_{x=0}^{\infty} g(h(x))\cdot p_{_X}(x) = \sum_{x=0}^{\infty} g(2x)\cdot \frac{\lambda^x\,\mathrm{e}^{-\lambda}}{x!} \\
&= \sum_{y\in\mathcal{S}_Y} g(y)\cdot \frac{\lambda^{y/2}\,\mathrm{e}^{-\lambda}}{(y/2)!},
\end{aligned}
$$

where $\mathcal{S}_Y = \{0,2,4,6,\ldots,\infty\}$ is the support of Y. Then, the pmf of Y is

$$
p_{_Y}(y) = \Pr(Y=y) = \frac{\lambda^{y/2}\,\mathrm{e}^{-\lambda}}{(y/2)!}, \quad y\in\mathcal{S}_Y. \qquad \|
$$

B.2.2 The two-dimensional case

15• The expectation technique

- Let $p_{(X_1,X_2)}(x_1,x_2)$ be the joint pmf of the random vector (X_1,X_2) and $Y = h(X_1,X_2)$ for some function $h(\cdot,\cdot)$.
- For any nonnegative and measurable $g(\cdot)$, if

$$
\begin{aligned}
E\{g(Y)\} &= E\{g(h(X_1,X_2))\} \\
&= \sum_{x_1}\sum_{x_2} g(h(x_1,x_2))\cdot p_{(X_1,X_2)}(x_1,x_2) \\
&= \sum_{y} g(y)\cdot p_{_Y}(y), \qquad \text{(B.35)}
\end{aligned}
$$

  then we can claim that $p_{_Y}(y)$ is the pmf of Y. ‖

16• Several examples

Example B.17 (Example 2.5 revisited). 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. For any nonnegative and measurable $g(\cdot)$, from (B.35), we have

$$\begin{aligned}
& E\{g(Y)\} \\
= & \sum_{x_1}\sum_{x_2} g(h(x_1, x_2)) \cdot p_{(X_1,X_2)}(x_1, x_2) \\
= & \sum_{x_1=0}^{\infty}\sum_{x_2=0}^{\infty} g(x_1 + x_2) \cdot \frac{\lambda_1^{x_1}\,\mathrm{e}^{-\lambda_1}}{x_1!} \cdot \frac{\lambda_2^{x_2}\,\mathrm{e}^{-\lambda_2}}{x_2!} \qquad [\text{let } y = x_1 + x_2] \\
= & \sum_{y=0}^{\infty} g(y) \left\{ \sum_{x_1=0}^{y} \frac{\lambda_1^{x_1}\,\mathrm{e}^{-\lambda_1}}{x_1!} \cdot \frac{\lambda_2^{y-x_1}\,\mathrm{e}^{-\lambda_2}}{(y-x_1)!} \right\} \\
= & \sum_{y=0}^{\infty} g(y) \cdot \frac{\mathrm{e}^{-\lambda_1+\lambda_2}}{y!} \left\{ \sum_{x_1=0}^{y} \frac{y!}{x_1!(y-x_1)!} \lambda_1^{x_1}\lambda_2^{y-x_1} \right\} \\
= & \sum_{y=0}^{\infty} g(y) \cdot \frac{(\lambda_1+\lambda_2)^y\,\mathrm{e}^{-\lambda_1+\lambda_2}}{y!},
\end{aligned}$$

implying that $Y \sim \text{Poisson}(\lambda_1 + \lambda_2)$. ‖

Example B.18 (Binomial distribution). Let $X_i \sim \text{Binomial}(m_i, p)$, $i = 1, 2$, and $X_1 \perp\!\!\!\perp X_2$, find the pmf of $Y = X_1 + X_2$.

Solution. For any nonnegative and measurable $g(\cdot)$, from (B.35), we have

$$\begin{aligned}
E\{g(Y)\} &= \sum_{x_1}\sum_{x_2} g(h(x_1, x_2)) \cdot p_{(X_1,X_2)}(x_1, x_2) \\
&= \sum_{x_1=0}^{m_1}\sum_{x_2=0}^{m_2} g(x_1 + x_2) \cdot \prod_{i=1}^{2} \binom{m_i}{x_i} p^{x_i}(1-p)^{m_i - x_i} \qquad [\text{let } y = x_1 + x_2] \\
&= \sum_{y=0}^{m_1+m_2} g(y) \left\{ \sum_{x_1} \binom{m_1}{x_1}\binom{m_2}{y-x_1} p^y (1-p)^{m_1+m_2-y} \right\}
\end{aligned}$$

$$= \sum_{y=0}^{m_1+m_2} g(y) \cdot \binom{m_1+m_2}{y} p^y (1-p)^{m_1+m_2-y} \left\{ \sum_{x_1} \frac{\binom{m_1}{x_1}\binom{m_2}{y-x_1}}{\binom{m_1+m_2}{y}} \right\}$$

$$= \sum_{y=0}^{m_1+m_2} g(y) \cdot \binom{m_1+m_2}{y} p^y (1-p)^{m_1+m_2-y},$$

implying that $Y \sim \text{Binomial}(m_1+m_2, p)$. ‖

Example B.19 (Charlier series distribution). Let $X_1 \sim \text{Binomial}(m,p)$, $X_2 \sim \text{Poisson}(\lambda)$, and $X_1 \perp\!\!\!\perp X_2$. Define $Y = X_1 + X_2$, then Y is said to follow the *Charlier series* (CS) distribution (Ong, 1988), denoted by $Y \sim \text{CS}(m,p,\lambda)$. Find the pmf of Y.

Solution. For any nonnegative and measurable $g(\cdot)$, from (B.35), we have

$$E\{g(Y)\} = \sum_{x_1}\sum_{x_2} g(h(x_1,x_2)) \cdot p_{(X_1,X_2)}(x_1,x_2)$$

$$= \sum_{x_1=0}^{m}\sum_{x_2=0}^{\infty} g(x_1+x_2) \cdot \binom{m}{x_1} p^{x_1}(1-p)^{m-x_1} \cdot \frac{\lambda^{x_2}\,\mathrm{e}^{-\lambda}}{x_2!}$$

[let $y = x_1 + x_2$]

$$= \sum_{y=0}^{\infty} g(y) \left\{ \sum_{x_1} \binom{m}{x_1} p^{x_1}(1-p)^{m-x_1} \cdot \frac{\lambda^{y-x_1}\,\mathrm{e}^{-\lambda}}{(y-x_1)!} \right\}$$

$$= \sum_{y=0}^{\infty} g(y) \cdot \left\{ \sum_{k=0}^{\min(m,y)} \binom{m}{k} p^{k}(1-p)^{m-k} \cdot \frac{\lambda^{y-k}\,\mathrm{e}^{-\lambda}}{(y-k)!} \right\}.$$

Hence, the pmf of Y is

$$p_Y(y) = \Pr(Y=y) = \sum_{k=0}^{\min(m,y)} \binom{m}{k} p^{k}(1-p)^{m-k} \cdot \frac{\lambda^{y-k}\,\mathrm{e}^{-\lambda}}{(y-k)!}$$

for $y = 0, 1, \ldots, \infty$. ‖

Example B.20 (Zero-inflated Poisson distribution). Let $Z \sim \text{Bernoulli}(1-\phi)$, $X \sim \text{Poisson}(\lambda)$ and $Z \perp\!\!\!\perp X$, find the pmf of $Y = ZX$.

Solution. For any nonnegative and measurable $g(\cdot)$, from (B.35), we have

$$\begin{aligned}
& E\{g(Y)\} \\
= & \sum_{z=0}^{1}\sum_{x=0}^{\infty} g(h(z,x)) \cdot p_{(Z,X)}(z,x) \\
= & \sum_{z=0}^{1}\sum_{x=0}^{\infty} g(zx) \cdot (1-\phi)^z \phi^{1-z} \cdot \frac{\lambda^x \,\mathrm{e}^{-\lambda}}{x!} \\
= & \sum_{x=0}^{\infty} g(0) \cdot \phi \cdot \frac{\lambda^x \,\mathrm{e}^{-\lambda}}{x!} + \sum_{x=0}^{\infty} g(x) \cdot (1-\phi) \cdot \frac{\lambda^x \,\mathrm{e}^{-\lambda}}{x!} \\
= & \; g(0) \cdot \phi + g(0) \cdot (1-\phi)\,\mathrm{e}^{-\lambda} + \sum_{x=1}^{\infty} g(x) \cdot (1-\phi) \cdot \frac{\lambda^x \,\mathrm{e}^{-\lambda}}{x!} \\
= & \; g(0)\{\phi + (1-\phi)\,\mathrm{e}^{-\lambda}\} + \sum_{y=1}^{\infty} g(y) \cdot (1-\phi)\frac{\lambda^y \,\mathrm{e}^{-\lambda}}{y!} \\
= & \sum_{y=0}^{\infty} g(y)\left[\{\phi + (1-\phi)\,\mathrm{e}^{-\lambda}\} I(y=0) + \left\{(1-\phi)\frac{\lambda^y \,\mathrm{e}^{-\lambda}}{y!}\right\} I(y>0)\right],
\end{aligned}$$

implying that $Y \sim \mathrm{ZIP}(\phi, \lambda)$. ‖

17• A class of zero-truncated distributions

- Let $X \sim p_X(x)$, $x \in \mathbb{N}_0 \,\hat{=}\, \{0\} \cup \mathbb{N}$, where $\mathbb{N} = \{1, 2, \ldots, m\}$ and m is a positive integer or $m = +\infty$.
- Let $Y \sim p_Y(y)$, $y \in \mathbb{N}$.
- The Y is said to be zero-truncated from the parent r.v. X, if $X = ZY$, where $Z \sim \mathrm{Bernoulli}(1 - p_X(0))$ and $Z \perp\!\!\!\perp Y$. Then, we have

$$p_Y(y) = \frac{p_X(y)}{1 - p_X(0)}, \quad y \in \mathbb{N}. \tag{B.36}$$

17.1• Proof of (B.36)

— For any nonnegative and measurable $g(\cdot)$, from (B.35), we have

$$\begin{aligned} E\{g(X)\} &= E\{g(ZY)\} \\ &= \sum_{z=0}^{1}\sum_{y\in\mathbb{N}} g(zy)\cdot\{1-p_X(0)\}^z p_X(0)^{1-z}\cdot p_Y(y) \\ &= \sum_{y\in\mathbb{N}} g(0)p_X(0)\cdot p_Y(y) + \sum_{y\in\mathbb{N}} g(y)\cdot\{1-p_X(0)\}\cdot p_Y(y) \\ &= g(0)p_X(0) + \sum_{y\in\mathbb{N}} g(y)\cdot\{1-p_X(0)\}p_Y(y). \qquad \text{(B.37)} \end{aligned}$$

— On the other hand, we have

$$\begin{aligned} E\{g(X)\} &= \sum_{x\in\mathbb{N}_0} g(x)\cdot p_X(x) = g(0)p_X(0) + \sum_{x\in\mathbb{N}} g(x)\cdot p_X(x) \\ &= g(0)p_X(0) + \sum_{y\in\mathbb{N}} g(y)\cdot p_X(y). \qquad \text{(B.38)} \end{aligned}$$

— By combining (B.37) with (B.38), we obtain the identity:

$$\sum_{y\in\mathbb{N}} g(y)\cdot\{1-p_X(0)\}p_Y(y) = \sum_{y\in\mathbb{N}} g(y)\cdot p_X(y),$$

for any $g(\cdot)$, indicating that (B.36) is true. □

17.2• Comments

— In fact, $Y \stackrel{\mathrm{d}}{=} X|(X>0)$.

B.2.3 The multiple-dimensional case

18• Expectation technique for a discrete random variable Y

- Let the discrete random vector $\mathbf{x} = (X_1,\dots,X_n)^\top \sim p_{\mathbf{x}}(\boldsymbol{x})$, where $\boldsymbol{x} = (x_1,\dots,x_n)^\top$ are realizations of $\mathbf{x}$.

- Define a discrete random variable $Y = h(X_1, \ldots, X_n) = h(\mathbf{x})$ for some function $h(\cdot)$. For any nonnegative and measurable $g(\cdot)$, if

$$\begin{aligned} E\{g(Y)\} &= E\{g(h(\mathbf{x}))\} \\ &= \sum_{\boldsymbol{x}} g(h(\boldsymbol{x})) \cdot p_{\mathbf{x}}(\boldsymbol{x}) = \sum_{y} g(y) \cdot p_{_Y}(y), \end{aligned} \tag{B.39}$$

then we can claim that $p_{_Y}(y)$ is the pmf of Y. ‖

19• Two examples

Example B.21 (Example 2.4 revisited). 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. For any nonnegative and measurable $g(\cdot)$, from (B.39), we have

$$\begin{aligned} E\{g(Y)\} &= \sum_{(x_1,x_2,x_3)} g(x_1 + x_2 + x_3) \cdot p_{\mathbf{x}}(x_1, x_2, x_3) \\ &= g(0) \cdot p_{\mathbf{x}}(0,0,0) + g(1) \cdot p_{\mathbf{x}}(0,0,1) + g(2) \cdot p_{\mathbf{x}}(0,1,1) \\ &\quad + g(2) \cdot p_{\mathbf{x}}(1,0,1) + g(2) \cdot p_{\mathbf{x}}(1,1,0) + g(3) \cdot p_{\mathbf{x}}(1,1,1) \end{aligned}$$

$$\begin{aligned} &= g(0) \cdot \frac{1}{8} + g(1) \cdot \frac{3}{8} + g(2) \cdot \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) + g(3) \cdot \frac{1}{8} \\ &= \sum_{y \in \mathcal{S}_Y} g(y) \cdot p_{_Y}(y), \end{aligned}$$

where $\mathcal{S}_Y = \{0, 1, 2, 3\}$ is the support of Y. Then, the pmf of Y, $p_{_Y}(y) = \Pr(Y = y)$, is given by $1/8$, $3/8$, $3/8$ and $1/8$, respectively. ‖

Example B.22 (Example 2.12 revisited). 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. For any nonnegative and measurable $g(\cdot)$, from (B.39), we have

$$
\begin{aligned}
E\{g(Y)\} &= \sum_{\boldsymbol{x}} g(h(\boldsymbol{x})) \cdot p_{\mathbf{x}}(\boldsymbol{x}) \\
&= \sum_{x_1=0}^{\infty} \cdots \sum_{x_n=0}^{\infty} g(\textstyle\sum_{i=1}^n x_i) \cdot \displaystyle\prod_{i=1}^{n} \frac{\lambda_i^{x_i}\, \mathrm{e}^{-\lambda_i}}{x_i!} \qquad [\text{let } y = \textstyle\sum_{i=1}^n x_i] \\
&= \sum_{y=0}^{\infty} g(y) \left\{ \sum_{(x_1,\ldots,x_n)} \prod_{i=1}^{n} \frac{\lambda_i^{x_i}\, \mathrm{e}^{-\lambda_i}}{x_i!} \right\} \\
&= \sum_{y=0}^{\infty} g(y) \cdot \frac{\mathrm{e}^{-\lambda}}{y!} \left\{ \sum_{(x_1,\ldots,x_n)} \frac{y!}{x_1! \cdots x_n!} \lambda_1^{x_1} \cdots \lambda_n^{x_n} \right\} \\
&= \sum_{y=0}^{\infty} g(y) \cdot \frac{\lambda^y\, \mathrm{e}^{-\lambda}}{y!},
\end{aligned}
$$

implying that $Y \sim \text{Poisson}(\lambda)$, where $\lambda = \sum_{i=1}^n \lambda_i$. ‖

20• Expectation technique for a discrete random vector $\mathbf{y}$

- Let the discrete random vector $\mathbf{x} = (X_1, \ldots, X_n)^\top \sim p_{\mathbf{x}}(\boldsymbol{x})$, where $\boldsymbol{x} = (x_1, \ldots, x_n)^\top$ are realizations of $\mathbf{x}$.

- Define a discrete random vector

$$
\mathbf{y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_m \end{pmatrix} = \begin{pmatrix} h_1(X_1, \ldots, X_n) \\ \vdots \\ h_m(X_1, \ldots, X_n) \end{pmatrix} = \begin{pmatrix} h_1(\mathbf{x}) \\ \vdots \\ h_m(\mathbf{x}) \end{pmatrix}
$$

for some functions $h_j(\cdot)$, $j = 1, \ldots, m$. For any nonnegative and measurable m-dimensional function $g(\boldsymbol{y}) = g(y_1, \ldots, y_m)$, if

$$
\begin{aligned}
E\{g(\mathbf{y})\} &= E\{g(h_1(\mathbf{x}), \ldots, h_m(\mathbf{x}))\} \\
&= \sum_{\boldsymbol{x}} g(h_1(\boldsymbol{x}), \ldots, h_m(\boldsymbol{x})) \cdot p_{\mathbf{x}}(\boldsymbol{x}) \\
&= \sum_{\boldsymbol{y}} g(y_1, \ldots, y_m) \cdot p_{\mathbf{y}}(\boldsymbol{y}) = \sum_{\boldsymbol{y}} g(\boldsymbol{y}) \cdot p_{\mathbf{y}}(\boldsymbol{y}), \qquad \text{(B.40)}
\end{aligned}
$$

then we can claim that $p_{\mathbf{y}}(\boldsymbol{y})$ is the joint pmf of $\mathbf{y}$. ‖

21• Two examples

Example B.23 (Bivariate Poisson distribution). Let $Y_1 = X_0 + X_1$ and $Y_2 = X_0 + X_2$, where $X_i \overset{\text{ind}}{\sim} \text{Poisson}(\lambda_i)$ for $i = 0, 1, 2$. Then the discrete random vector $\mathbf{y} = (Y_1, Y_2)^\top$ is said to follow a bivariate Poisson distribution with parameters $\lambda_0 \geqslant 0$, $\lambda_1 > 0$ and $\lambda_2 > 0$, denoted by $\mathbf{y} \sim \text{MP}_2(\lambda_0, \lambda_1, \lambda_2)$. Find the joint pmf of $\mathbf{y}$.

Solution. For any nonnegative and measurable 2-variate function $g(y_1, y_2)$, from (B.40), we have

$$\begin{aligned}
E\{g(\mathbf{y})\} &= E\{g(Y_1, Y_2)\} = E\{g(X_0 + X_1, X_0 + X_2)\} \\
&= \sum_{(x_0, x_1, x_2)} g(x_0 + x_1, x_0 + x_2) \cdot \prod_{i=0}^{2} \frac{\lambda_i^{x_i}\, \mathrm{e}^{-\lambda_i}}{x_i!} \\
&\qquad [\text{ let } y_1 = x_0 + x_1,\ y_2 = x_0 + x_2] \\
&= \sum_{(y_1, y_2)} g(y_1, y_2) \cdot \sum_{x_0} \frac{\lambda_0^{x_0}\, \mathrm{e}^{-\lambda_0}}{x_0!} \prod_{i=1}^{2} \frac{\lambda_i^{x_i}\, \mathrm{e}^{-\lambda_i}}{x_i!} \\
&= \sum_{(y_1, y_2)} g(y_1, y_2) \left\{ \sum_{x_0=0}^{\min(y_1, y_2)} \frac{\lambda_0^{x_0}\, \mathrm{e}^{-\lambda_0}}{x_0!} \prod_{i=1}^{2} \frac{\lambda_i^{y_i - x_0}\, \mathrm{e}^{-\lambda_i}}{(y_i - x_0)!} \right\}.
\end{aligned}$$

Hence, the joint pmf of $\mathbf{y}$ is

$$p_{\mathbf{y}}(\boldsymbol{y}) = \Pr(\mathbf{y} = \boldsymbol{y}) = \sum_{x_0=0}^{\min(y_1, y_2)} \frac{\lambda_0^{x_0}\, \mathrm{e}^{-\lambda_0}}{x_0!} \prod_{i=1}^{2} \frac{\lambda_i^{y_i - x_0}\, \mathrm{e}^{-\lambda_i}}{(y_i - x_0)!},$$

where $\boldsymbol{y} = (y_1, y_2)^\top$ are realizations of $\mathbf{y} = (Y_1, Y_2)^\top$. ‖

Example B.24 (Exercise 2.6 revisited). Let $Z \sim \text{Bernoulli}\,(1 - \phi)$, $\mathbf{x} = (X_1, \ldots, X_m)^\top$, $X_j \overset{\text{ind}}{\sim} \text{Poisson}(\lambda_j)$ for $j = 1, \ldots, m$, and $Z \perp\!\!\!\perp \mathbf{x}$. Define $\mathbf{y} = (Y_1, \ldots, Y_m)^\top = Z\mathbf{x}$. Find the joint pmf of $\mathbf{y}$.

Solution. For any nonnegative and measurable m-dimensional function

$g(y_1, \ldots, y_m)$, from (B.40), we have

$$\begin{aligned}
& E\{g(\mathbf{y})\} = E\{g(ZX_1, \ldots, ZX_m)\} \\
= & \sum_{z=0}^{1} \sum_{\boldsymbol{x}} g(zx_1, \ldots, zx_m) \cdot p_{(Z,\mathbf{x})}(z, \boldsymbol{x}) \\
= & \sum_{z=0}^{1} \sum_{\boldsymbol{x}} g(z\boldsymbol{x}) \cdot (1-\phi)^z \phi^{1-z} \cdot \prod_{j=1}^{m} \frac{\lambda_j^{x_j} \, \mathrm{e}^{-\lambda_j}}{x_j!} \\
= & \sum_{\boldsymbol{x}} g(\mathbf{0}_m) \cdot \phi \prod_{j=1}^{m} \frac{\lambda_j^{x_j} \, \mathrm{e}^{-\lambda_j}}{x_j!} + \sum_{\boldsymbol{x}} g(\boldsymbol{x}) \cdot (1-\phi) \, \mathrm{e}^{-\lambda} \prod_{j=1}^{m} \frac{\lambda_j^{x_j}}{x_j!} \\
= & g(\mathbf{0})\phi + g(\mathbf{0})(1-\phi) \, \mathrm{e}^{-\lambda} + \sum_{\boldsymbol{x},\, \boldsymbol{x} \neq \mathbf{0}} g(\boldsymbol{x}) \cdot (1-\phi) \, \mathrm{e}^{-\lambda} \prod_{j=1}^{m} \frac{\lambda_j^{x_j}}{x_j!} \\
= & g(\mathbf{0})\{\phi + (1-\phi) \, \mathrm{e}^{-\lambda}\} + \sum_{\boldsymbol{y},\, \boldsymbol{y} \neq \mathbf{0}} g(\boldsymbol{y}) \cdot (1-\phi) \, \mathrm{e}^{-\lambda} \prod_{j=1}^{m} \frac{\lambda_j^{y_j}}{y_j!} \\
= & \sum_{\boldsymbol{y}} g(\boldsymbol{y}) \Big[\{\phi + (1-\phi) \, \mathrm{e}^{-\lambda}\} I(\boldsymbol{y} = \mathbf{0}) \\
& \qquad + \left\{ (1-\phi) \, \mathrm{e}^{-\lambda} \textstyle\prod_{j=1}^{m} \frac{\lambda_j^{y_j}}{y_j!} \right\} I(\boldsymbol{y} \neq \mathbf{0}) \Big],
\end{aligned}$$

where $\boldsymbol{x} = (x_1, \ldots, x_m)^\top$ are realizations of $\mathbf{x}$, $\lambda = \sum_{j=1}^{m} \lambda_j$ and $\boldsymbol{y} = (y_1, \ldots, y_m)^\top$ are realizations of $\mathbf{y}$. Hence, the joint pmf of $\mathbf{y}$ is

$$\begin{aligned}
& p_{\mathbf{y}}(\boldsymbol{y}) = \Pr(\mathbf{y} = \boldsymbol{y}) \\
= & \{\phi + (1-\phi) \, \mathrm{e}^{-\lambda}\} I(\boldsymbol{y} = \mathbf{0}) + \left\{ (1-\phi) \, \mathrm{e}^{-\lambda} \prod_{j=1}^{m} \frac{\lambda_j^{y_j}}{y_j!} \right\} I(\boldsymbol{y} \neq \mathbf{0}). \qquad \|
\end{aligned}$$