CHAPTER

EXAMPLES OF CONTINUOUS PROBABILITY DISTRIBUTIONS

4

CHAPTER CONTENTS

In this chapter, popular continuous probability distributions and their properties such as the expectation, the variance, and the moment-generating functions are illustrated.

4.1 CONTINUOUS UNIFORM DISTRIBUTION

The *continuous uniform distribution* has a constant probability density over a finite interval $[a,b]$:

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

The continuous uniform distribution is denoted by $U(a,b)$, and its expectation and variance are given by

$$E[x] = \frac{a+b}{2} \quad \text{and} \quad V[x] = \frac{(b-a)^2}{12}.$$

4.2 NORMAL DISTRIBUTION

The *normal distribution*, also known as the *Gaussian distribution*, would be the most important continuous distribution. For $-\infty < \mu < \infty$ and $\sigma > 0$, the normal

An Introduction to Statistical Machine Learning. DOI: 10.1016/B978-0-12-802121-7.00015-7

distribution is denoted by $N(\mu,\sigma^2)$, and its probability density is given by

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

The fact that the above $f(x)$ is integrated to one can be proven by the *Gaussian integral* shown in Fig. 4.1 as

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)\mathrm{d}x &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\mathrm{d}x \\ &= \frac{\sigma\sqrt{2}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-r^2\right)\mathrm{d}r = 1.\end{aligned}$$

Here, variables of integration are changed from x to $r = \frac{x-\mu}{\sigma\sqrt{2}}$ (i.e., $\frac{\mathrm{d}x}{\mathrm{d}r} = \sigma\sqrt{2}$), as explained in Fig. 2.9.

The constants μ and σ included in normal distribution $N(\mu,\sigma^2)$ correspond to the expectation and standard deviation:

$$E[x] = \mu \quad \text{and} \quad V[x] = \sigma^2.$$

As explained in Section 2.4.3, the expectation and variance can be obtained through the moment-generating function: Indeed, the moment-generating function of $N(\mu,\sigma^2)$ is given by

$$\begin{aligned}M_x(t) &= \int_{-\infty}^{\infty} e^{tx} f(x)\mathrm{d}x \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} + tx\right)\mathrm{d}x \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2 - 2(\mu+\sigma^2 t)x + \mu^2}{2\sigma^2}\right)\mathrm{d}x \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{\left(x-(\mu+\sigma^2 t)\right)^2}{2\sigma^2} + \mu t + \frac{\sigma^2 t^2}{2}\right)\mathrm{d}x \\ &= \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right) \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\left(x-(\mu+\sigma^2 t)\right)^2}{2\sigma^2}\right)\mathrm{d}x \\ &= \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right). \end{aligned} \tag{4.1}$$

Note that, in the above derivation, *completing the square*,

$$x^2 + 2ax + b = 0 \iff (x+a)^2 - a^2 + b = 0, \tag{4.2}$$

is used to have the probability density function of $N(\mu+\sigma^2 t,\sigma^2)$, which is integrated to 1.

Probability density functions of $N(\mu,\sigma^2)$ for $\mu = 0$ are illustrated in Fig. 4.3, showing that the normal density is symmetric and bell-shaped.

The Gaussian integral bridges the two well-known *irrational numbers*, $e = 2.71828\cdots$ and $\pi = 3.14159\cdots$, as

$$\int_{-\infty}^{\infty} e^{-x^2}\mathrm{d}x = \sqrt{\pi}.$$

To prove this, let us consider change of variables in integration (see Fig. 4.2) for $f(x,y) = e^{-(x^2+y^2)}$ and $\mathcal{X} = \mathcal{Y} = [-\infty,\infty]$. Let $g(r,\theta) = r\cos\theta$ and $h(r,\theta) = r\sin\theta$. Then

$$\mathcal{R} = [0,\infty],\quad \Theta = [0,2\pi],\quad \boldsymbol{J} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix},\quad \text{and}\quad \det(\boldsymbol{J}) = r,$$

which yields

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)}\mathrm{d}x\mathrm{d}y &= \int_{0}^{2\pi}\int_{0}^{\infty} re^{-r^2}\mathrm{d}r\mathrm{d}\theta = \int_{0}^{2\pi}\mathrm{d}\theta\int_{0}^{\infty} re^{-r^2}\mathrm{d}r \\ &= \int_{0}^{2\pi}\mathrm{d}\theta\int_{0}^{\infty} re^{-r^2}\mathrm{d}r = 2\pi\left[-\frac{1}{2}e^{-r^2}\right]_0^{\infty} = \pi.\end{aligned}$$

Consequently,

$$\begin{aligned}\int_{-\infty}^{\infty} e^{-x^2}\mathrm{d}x &= \sqrt{\left(\int_{-\infty}^{\infty} e^{-x^2}\mathrm{d}x\right)^2} = \sqrt{\left(\int_{-\infty}^{\infty} e^{-x^2}\mathrm{d}x\right)\left(\int_{-\infty}^{\infty} e^{-y^2}\mathrm{d}y\right)} \\ &= \sqrt{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)}\mathrm{d}x\mathrm{d}y} = \sqrt{\pi}.\end{aligned}$$

FIGURE 4.1

Gaussian integral.

If random variable x follows $N(\mu,\sigma^2)$, its affine transformation

$$r = ax + b$$

follows $N(a\mu + b, a^2\sigma^2)$. This can be proved by the fact that the probability density function of r is given as follows (see Section 2.5):

$$g(r) = \frac{1}{|a|}f\left(\frac{r-b}{a}\right).$$

Let us extend change of variables in integration from one dimension (see Fig. 2.9) to two dimensions. Integration of function $f(x, y)$ over $\mathcal{X} \times \mathcal{Y}$ can be computed with $x = g(r, \theta)$, $y = h(r, \theta)$, $\mathcal{X} = g(\mathcal{R}, \Theta)$, and $\mathcal{Y} = h(\mathcal{R}, \Theta)$ as

$$\int_{\mathcal{X}} \int_{\mathcal{Y}} f(x, y) \mathrm{d}y \mathrm{d}x = \int_{\mathcal{R}} \int_{\Theta} f\big(g(r, \theta), h(r, \theta)\big) |\det(\boldsymbol{J})| \mathrm{d}\theta \mathrm{d}r,$$

where $\boldsymbol{J}$ is called the *Jacobian matrix* and $\det(\boldsymbol{J})$ denotes the *determinant* of $\boldsymbol{J}$:

$$\boldsymbol{J} = \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix} \quad \text{and} \quad \det(\boldsymbol{J}) = \frac{\partial x}{\partial r} \frac{\partial y}{\partial \theta} - \frac{\partial x}{\partial \theta} \frac{\partial y}{\partial r}.$$

The determinant is the product of all eigenvalues and it corresponds to the ratio of volumes when (x, y) is changed to (r, θ). $\det(\boldsymbol{J})$ is often called the *Jacobian*. The above formula can be extended to more than two dimensions.

FIGURE 4.2

Two-dimensional change of variables in integration.

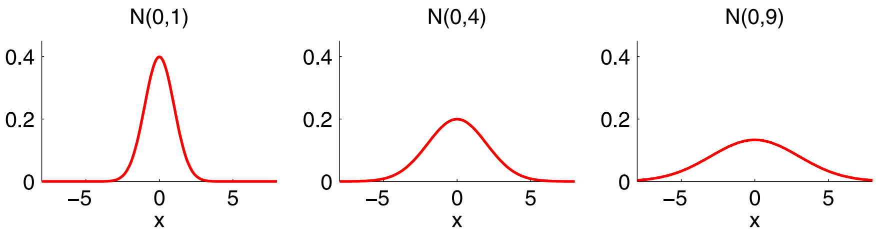

FIGURE 4.3

Probability density functions of normal density $N(\mu, \sigma^2)$.

Setting $a = \frac{1}{D[x]}$ and $b = -\frac{E[x]}{D[x]}$ yields

$$z = \frac{x}{D[x]} - \frac{E[x]}{D[x]} = \frac{x - E[x]}{D[x]},$$

which follows *standard normal distribution* $N(0, 1)$ (Fig. 4.4).

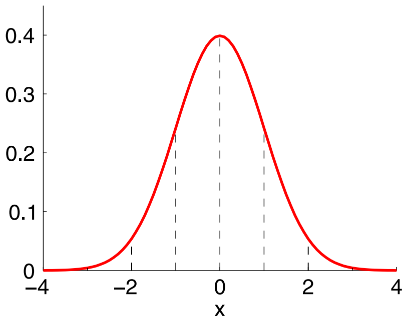

FIGURE 4.4

Standard normal distribution $N(0,1)$. A random variable following $N(0,1)$ is included in $[-1,1]$ with probability 68.27%, in $[-2,2]$ with probability 95.45%, and in $[-3,3]$ with probability 99.73%.

4.3 GAMMA DISTRIBUTION, EXPONENTIAL DISTRIBUTION, AND CHI-SQUARED DISTRIBUTION

As explained in Section 3.4, the Poisson distribution represents the probability that the event occurring λ times on average in unit time occurs x times in unit time. On the other hand, the elapsed time x that the event occurring λ times on average in unit time occurs α times follows the *gamma distribution*. The gamma distribution for positive real constants α and λ is denoted by $\mathrm{Ga}(\alpha,\lambda)$.

The probability density function of $\mathrm{Ga}(\alpha,\lambda)$ is given by

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \quad \text{for } x \geq 0,$$

where

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \mathrm{d}x \tag{4.3}$$

is called the *gamma function*. $\int_0^\infty f(x)\mathrm{d}x = 1$ can be proved by changing variables of integration as $y = \lambda x$ (see Fig. 2.9):

$$\begin{aligned}\int_{-\infty}^\infty f(x)\mathrm{d}x &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-\lambda x} \mathrm{d}x = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty \left(\frac{y}{\lambda}\right)^{\alpha-1} e^{-y} \frac{1}{\lambda} \mathrm{d}y \\ &= \frac{1}{\Gamma(\alpha)} \int_0^\infty y^{\alpha-1} e^{-y} \mathrm{d}y = 1.\end{aligned}$$

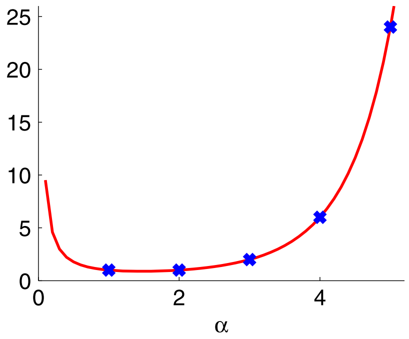

FIGURE 4.5

Gamma function. $\Gamma(\alpha + 1) = \alpha!$ holds for non-negative integer α, and the gamma function smoothly interpolates the factorials.

The gamma function fulfills

$$\Gamma(1) = \int_0^\infty e^{-x}\mathrm{d}x = \left[-e^{-x}\right]_0^\infty = 1.$$

Furthermore, *integration by parts* for functions $u(x)$ and $v(x)$ and their derivatives $u'(x)$ and $v'(x)$ given by

$$\int_a^b u(x)v'(x)\mathrm{d}x = \left[u(x)v(x)\right]_a^b - \int_a^b u'(x)v(x)\mathrm{d}x \tag{4.4}$$

yields

$$\begin{aligned}\Gamma(\alpha) &= \int_0^\infty e^{-x}x^{\alpha-1}\mathrm{d}x = \left[e^{-x}\frac{x^\alpha}{\alpha}\right]_0^\infty - \int_0^\infty (-e^{-x})\frac{x^\alpha}{\alpha}\mathrm{d}x \\ &= \frac{1}{\alpha}\int_0^\infty e^{-x}x^{(\alpha+1)-1}\mathrm{d}x = \frac{\Gamma(\alpha+1)}{\alpha}.\end{aligned}$$

Putting them together, the gamma function for non-negative integer α is shown to fulfill

$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha) = \alpha(\alpha-1)\Gamma(\alpha-1) = \cdots = \alpha!\Gamma(1) = \alpha!.$$

Thus, the gamma function can be regarded as generalization of the factorial to real numbers (see Fig. 4.5). Furthermore, change of variables $x = y^2$ in integration yields

$$\Gamma(\alpha) = \int_0^\infty y^{2(\alpha-1)}e^{-y^2}\frac{\mathrm{d}x}{\mathrm{d}y}\mathrm{d}y = 2\int_0^\infty y^{2\alpha-1}e^{-y^2}\mathrm{d}y, \tag{4.5}$$

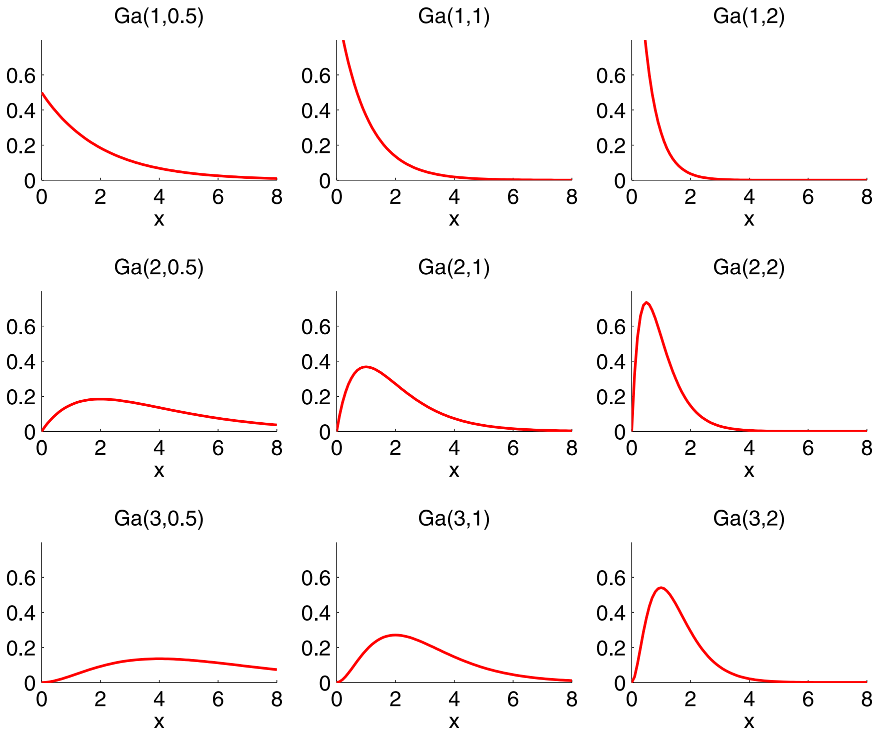

FIGURE 4.6

Probability density functions of gamma distribution $\mathrm{Ga}(\alpha, \lambda)$.

and applying the *Gaussian integral* (Fig. 4.1) to this results in

$$\Gamma\left(\frac{1}{2}\right) = 2\int_0^\infty e^{-y^2}\mathrm{d}y = \int_{-\infty}^\infty e^{-y^2}\mathrm{d}y = \sqrt{\pi}.$$

Probability density functions of $\mathrm{Ga}(\alpha, \lambda)$ are illustrated in Fig. 4.6. The probability density is monotone decreasing as x increases when $\alpha \leq 1$, while the probability density increases and then decreases as x increases when $\alpha > 1$.

By changing variables of integration as $y = (\lambda - t)x$, the moment-generating function of $\mathrm{Ga}(\alpha, \lambda)$ can be expressed as

$$\begin{aligned} M_x(t) = E[e^{tx}] &= \frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^\infty x^{\alpha-1}e^{-(\lambda-t)x}\mathrm{d}x \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^\infty \left(\frac{y}{\lambda - t}\right)^{\alpha-1} e^{-y}\frac{1}{\lambda - t}\mathrm{d}y = \frac{\lambda^\alpha}{\Gamma(\alpha)}\frac{\Gamma(\alpha)}{(\lambda-t)^\alpha} = \left(\frac{\lambda}{\lambda - t}\right)^\alpha. \end{aligned}$$

From this, the expectation and variance of $\mathrm{Ga}(\alpha, \lambda)$ are given by

$$E[x] = \frac{\alpha}{\lambda} \quad \text{and} \quad V[x] = \frac{\alpha}{\lambda^2}.$$

Gamma distribution $\mathrm{Ga}(\alpha, \lambda)$ for $\alpha = 1$ is called the *exponential distribution* and is denoted by $\mathrm{Exp}(\lambda)$. The probability density function of $\mathrm{Exp}(\lambda)$ is given by

$$f(x) = \lambda e^{-\lambda x}.$$

The elapsed time x that the event occurring λ times on average in unit time occurs for the first time follows the exponential distribution.

Gamma distribution $\mathrm{Ga}(\alpha, \lambda)$ for $\alpha = n/2$ and $\lambda = 1/2$ where n is an integer is called the *chi-squared distribution* with n degrees of freedom. This is denoted by $\chi^2(n)$, and its probability density function is given by

$$f(x) = \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})}.$$

Let $z_1, \ldots, z_n$ independently follow the standard normal distribution $N(0, 1)$. Then their squared sum,

$$x = \sum_{i=1}^{n} z_i^2,$$

follows $\chi^2(n)$. The chi-squared distribution plays an important role in hypothesis testing explained in Chapter 10.

4.4 BETA DISTRIBUTION

For positive real scalars α and β, the *beta distribution*, denoted by $\mathrm{Be}(\alpha, \beta)$, is the probability distribution whose probability density is given by

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \quad \text{for } 0 \le x \le 1,$$

where

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} \mathrm{d}x$$

is the *beta function*. When α and β are positive integers, αth smallest value (or equivalently βth largest value) among the $\alpha + \beta - 1$ random variables independently following the continuous uniform distribution $U(0, 1)$ follows the beta distribution.

Change of variables in integration as $x = (\sin\theta)^2$ allows us to express the beta function using sinusoidal functions as

$$B(\alpha,\beta) = \int_0^{\frac{\pi}{2}} (\sin\theta)^{2(\alpha-1)} \left(1-(\sin\theta)^2\right)^{\beta-1} \frac{\mathrm{d}x}{\mathrm{d}\theta}\mathrm{d}\theta$$
$$= \int_0^{\frac{\pi}{2}} (\sin\theta)^{2(\alpha-1)}(\cos\theta)^{2(\beta-1)} \cdot 2\sin\theta\cos\theta\mathrm{d}\theta$$
$$= 2\int_0^{\frac{\pi}{2}} (\sin\theta)^{2\alpha-1}(\cos\theta)^{2\beta-1}\mathrm{d}\theta. \tag{4.6}$$

Furthermore, from Eq. (4.5), the gamma product $\Gamma(\alpha)\Gamma(\beta)$ can be expressed as

$$\Gamma(\alpha)\Gamma(\beta) = \left(2\int_0^{\infty} u^{2\alpha-1}e^{-u^2}\mathrm{d}u\right)\left(2\int_0^{\infty} v^{2\beta-1}e^{-v^2}\mathrm{d}v\right)$$
$$= 4\int_0^{\infty}\int_0^{\infty} u^{2\alpha-1}v^{2\beta-1}e^{-(u^2+v^2)}\mathrm{d}u\mathrm{d}v.$$

If variables of integrations are changed as $u = r\sin\theta$ and $v = r\cos\theta$ (see Fig. 4.2) in the above equation, then

$$\Gamma(\alpha)\Gamma(\beta) = 4\int_0^{\frac{\pi}{2}}\int_0^{\infty} r^{2(\alpha+\beta)-2}e^{-r^2}(\sin\theta)^{2\alpha-1}(\cos\theta)^{2\beta-1} r\mathrm{d}r\mathrm{d}\theta$$
$$= \left(2\int_0^{\infty} r^{2(\alpha+\beta)-1}e^{-r^2}\mathrm{d}r\right)\left(2\int_0^{\frac{\pi}{2}} (\sin\theta)^{2\alpha-1}(\cos\theta)^{2\beta-1}\mathrm{d}\theta\right)$$
$$= \Gamma(\alpha+\beta)B(\alpha,\beta),$$

where Eq. (4.6) was used. Thus, the beta function can be expressed by using the gamma function as

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}. \tag{4.7}$$

This allows us to compute, e.g., the following integrals:

$$\int_0^{\frac{\pi}{2}} (\sin\theta)^{2n}\mathrm{d}\theta = \int_0^{\frac{\pi}{2}} (\sin\theta)^{2(n+\frac{1}{2})-1}(\cos\theta)^{2\frac{1}{2}-1}\mathrm{d}\theta = B\left(n+\frac{1}{2},\frac{1}{2}\right),$$
$$\int_0^{\frac{\pi}{2}} (\sin\theta)^{2n+1}\mathrm{d}\theta = \int_0^{\frac{\pi}{2}} (\sin\theta)^{2(n+1)-1}(\cos\theta)^{2\frac{1}{2}-1}\mathrm{d}\theta = B\left(n+1,\frac{1}{2}\right).$$

Probability density functions of $\mathrm{Be}(\alpha,\beta)$ are illustrated in Fig. 4.7. This shows that the profile of the beta density drastically changes depending on the values of α and β, and the beta distribution is reduced to the continuous uniform distribution when $\alpha = \beta = 1$.

The expectation and variance of $\mathrm{Be}(\alpha,\beta)$ are given by

$$E[x] = \frac{\alpha}{\alpha+\beta} \quad \text{and} \quad V[x] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

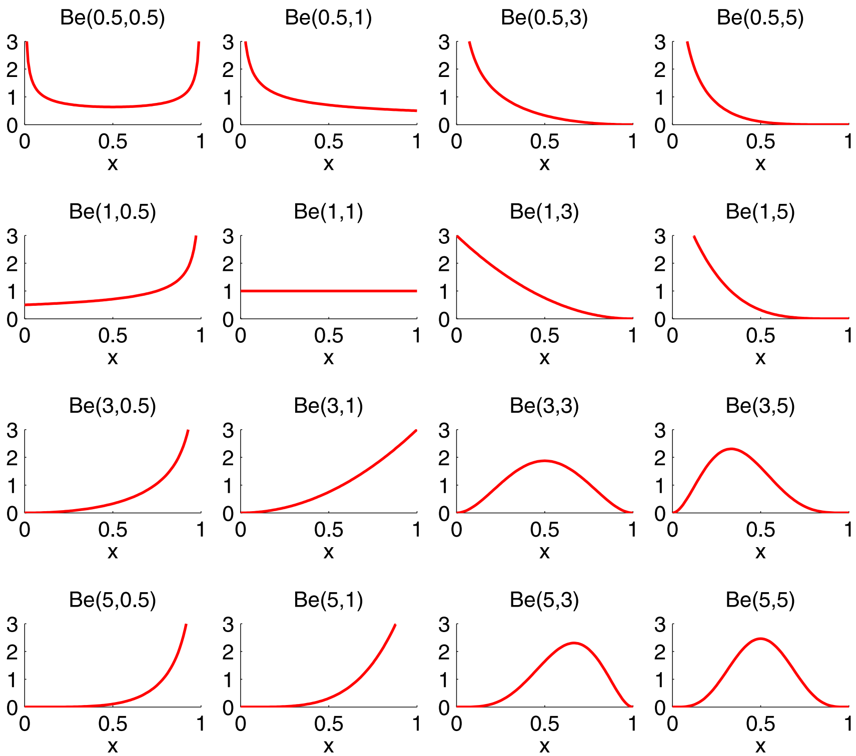

FIGURE 4.7

Probability density functions of beta distribution $\mathrm{Be}(\alpha, \beta)$.

The expectation can be proved by applying integration by parts as

$$
\begin{aligned}
E[x] &= \frac{1}{B(\alpha,\beta)} \int_0^1 x x^{\alpha-1}(1-x)^{\beta-1} \mathrm{d}x \\
&= \frac{1}{B(\alpha,\beta)} \int_0^1 x^{\alpha}(1-x)^{\beta-1} \mathrm{d}x \\
&= \frac{1}{B(\alpha,\beta)} \left\{ \left[x^{\alpha} \left(-\frac{(1-x)^{\beta}}{\beta} \right) \right]_0^1 - \int_0^1 \alpha x^{\alpha-1} \left(-\frac{(1-x)^{\beta}}{\beta} \right) \mathrm{d}x \right\} \\
&= \frac{\alpha}{\beta} \frac{1}{B(\alpha,\beta)} \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}(1-x) \mathrm{d}x \\
&= \frac{\alpha}{\beta} \frac{1}{B(\alpha,\beta)} \left\{ \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} \mathrm{d}x - \int_0^1 x x^{\alpha-1}(1-x)^{\beta-1} \mathrm{d}x \right\} \\
&= \frac{\alpha}{\beta}(1 - E[x]).
\end{aligned}
$$

Similar computation applied to $E[x^2]$ gives

$$E[x^2] = \frac{\alpha+1}{\beta}\Big(E[x] - E[x^2]\Big),$$

which yields

$$E[x^2] = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}.$$

Plugging this into

$$V[x] = E[x^2] - \big(E[x]\big)^2$$

gives $V[x] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

In Section 6.3, the beta distribution will be extended to multiple dimensions.

4.5 CAUCHY DISTRIBUTION AND LAPLACE DISTRIBUTION

Let z and z' be the random variables independently following the standard normal distribution $N(0,1)$. Then their ratio,

$$x = \frac{z}{z'},$$

follows the standard Cauchy distribution, whose probability density function is given by $f(x) = \frac{1}{\pi(x^2+1)}$. Its generalization using a real scalar a and a positive real b is given by

$$f(x) = \frac{b}{\pi((x-a)^2+b^2)}.$$

This is called the *Cauchy distribution* and is denoted by $\mathrm{Ca}(a,b)$.

Computing the expectation of the standard Cauchy distribution yields

$$\begin{aligned} E[x] &= \int_{-\infty}^{+\infty} x f(x)\mathrm{d}x = \int_{-\infty}^{+\infty} \frac{x}{\pi(x^2+1)}\mathrm{d}x \\ &= \frac{1}{2\pi}\Big[\log(1+x^2)\Big]_{-\infty}^{+\infty} = \frac{1}{2\pi}\lim_{\alpha\to+\infty,\beta\to-\infty}\log\frac{1+\alpha^2}{1+\beta^2}, \end{aligned}$$

which means that the value depends on how fast α approaches $+\infty$ and β approaches $-\infty$. For this reason, $\mathrm{Ca}(a,b)$ does not have the expectation. The limiting value when $\alpha = -\beta$ is called the *principal value*, which is given by a and represents the "location" of the Cauchy distribution. Since the expectation does not exist, the Cauchy distribution does not have the variance and all higher moments, either. The positive scalar b represents the "scale" of the Cauchy distribution.

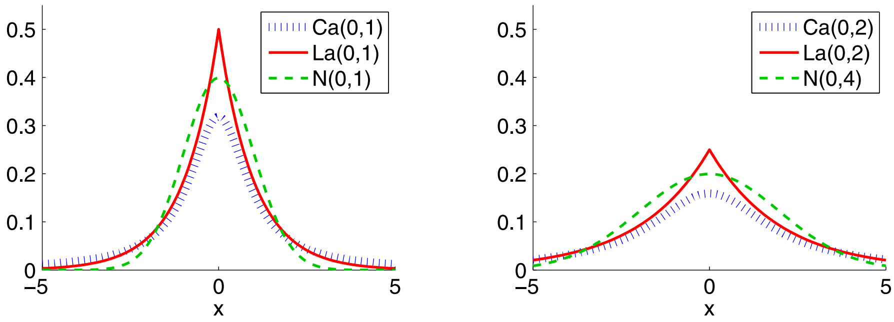

FIGURE 4.8

Probability density functions of Cauchy distribution $\mathrm{Ca}(a,b)$, Laplace distribution $\mathrm{La}(a,b)$, and normal distribution $N(a,b^2)$.

Let y and y' be random variables independently following the exponential distribution $\mathrm{Exp}(1)$. Then their difference,

$$x = y - y',$$

follows the standard Laplace distribution, whose probability density function is given by $f(x) = \frac{1}{2}\exp(-|x|)$. Its generalization using a real scalar a and a positive real b is given by

$$f(x) = \frac{1}{2b}\exp\left(-\frac{|x-a|}{b}\right).$$

This is called the *Laplace distribution* and is denoted by $\mathrm{La}(a,b)$. Since the Laplace distribution can be regarded as extending the exponential distribution to the negative domain, it is also referred to as the *double exponential distribution*.

When $|t| < 1/b$, the moment-generating function of $\mathrm{La}(a,b)$ is given by

$$\begin{aligned} M_x(t) &= \frac{1}{2b}\int_{-\infty}^{a}\exp\left(xt + \frac{x}{b} - \frac{a}{b}\right)\mathrm{d}x + \frac{1}{2b}\int_{a}^{+\infty}\exp\left(xt - \frac{x}{b} + \frac{a}{b}\right)\mathrm{d}x \\ &= \frac{1}{2}\left[\frac{1}{1+bt}\exp\left(xt + \frac{x}{b} - \frac{a}{b}\right)\right]_{-\infty}^{a} - \frac{1}{2}\left[\frac{1}{1-bt}\exp\left(xt - \frac{x}{b} + \frac{a}{b}\right)\right]_{-\infty}^{a} \\ &= \frac{\exp(at)}{1-b^2t^2}. \end{aligned}$$

From this, the expectation and variance of $\mathrm{La}(a,b)$ are given by

$$E[x] = a \quad \text{and} \quad V[x] = 2b^2.$$

Probability density functions of Cauchy distribution $\mathrm{Ca}(a,b)$, Laplace distribution $\mathrm{La}(a,b)$, and normal distribution $N(a,b^2)$ are illustrated in Fig. 4.8. Since the Cauchy

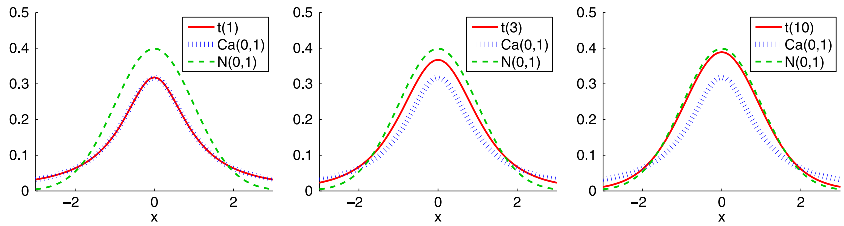

FIGURE 4.9

Probability density functions of t-distribution $t(d)$, Cauchy distribution $\mathrm{Ca}(0,1)$, and normal distribution $N(0,1)$.

and Laplace distributions have heavier tails than the normal distribution, realized values can be quite far from the origin. For this reason, the Cauchy and Laplace distributions are often used for modeling data with *outliers*. Note that the Laplace density is not differentiable at the origin.

4.6 t-DISTRIBUTION AND F-DISTRIBUTION

Let z be a random variable following the standard normal distribution $N(0,1)$ and y be a random variable following the chi-squared distribution with d degrees of freedom $\chi^2(d)$. Then their ratio,

$$x = \frac{z}{\sqrt{y/d}},$$

follows the *t-distribution* denoted by $t(d)$. Following its inventor's pseudonym, the t-distribution is also referred to as *Student's t-distribution*.

The probability density function of t-distribution $t(d)$ is given as follows (Fig. 4.9):

$$f(x) = \frac{1}{B(\frac{d}{2},\frac{1}{2})\sqrt{d}}\left(1+\frac{x^2}{d}\right)^{-\frac{d+1}{2}},$$

where B is the beta function explained in Section 4.4. The t-distribution agrees with the Cauchy distribution when the degree of freedom is $d = 1$, and it is reduced to the normal distribution when the degree of freedom tends to ∞. The expectation exists when $d \geq 2$, and the variance exists when $d \geq 3$, which are given by

$$E[x] = 0 \quad \text{and} \quad V[x] = \frac{d}{d-2}.$$

Let y and y' be random variables following the chi-squared distributions with d and d' degrees of freedom, respectively. Then their ratio,

$$x = \frac{y/d}{y'/d'},$$

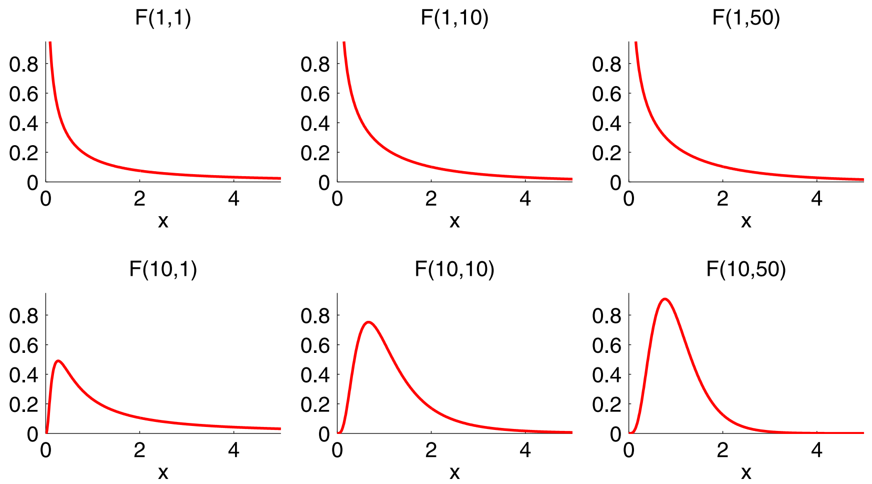

FIGURE 4.10

Probability density functions of F-distribution $F(d, d')$.

follows the *F-distribution* denoted by $F(d, d')$. Following its inventor's name, the F-distribution is also referred to as *Snedecor's F-distribution*.

The probability density function of F-distribution $F(d, d')$ is given as follows (Fig. 4.10):

$$f(x) = \frac{1}{B(d/2, d'/2)} \left(\frac{d}{d'}\right)^{\frac{d}{2}} x^{\frac{d}{2}-1} \left(1 + \frac{d}{d'}x\right)^{-\frac{d+d'}{2}} \quad \text{for } x \geq 0.$$

The expectation exists when $d' \geq 3$, and the variance exists when $d' \geq 5$, which are given by

$$E[x] = \frac{d'}{d'-2} \quad \text{and} \quad V[x] = \frac{2d'^2(d+d'-2)}{d(d'-2)^2(d'-4)}.$$

If y follows the t-distribution $t(d)$, then y^2 follows the F-distribution $F(1, d)$.

The t-distribution and the F-distribution play important roles in hypothesis testing explained in Chapter 10. The t-distribution is also utilized for deriving the confidence interval in Section 9.3.1.