Chapter 4 Important Probability Distributions

Abstract: In this chapter, we introduce a variety of discrete probability distributions and continuous probability distributions that are commonly used in economics and finance. Examples of discrete probability distributions include Bernoulli, Binomial, Negative Binomial, Geometric and Poisson distributions. Examples of continuous probability distributions include Beta, Cauchy, Chi-squares, Exponential, Gamma, generalized Gamma distribution, normal, lognormal, Weibull, and uniform distributions. The properties of these distributions as well as their applications in economics and finance are discussed. We also show some important techniques of obtaining moments and MGF's for various probability distributions.

Key words: Beta distribution, Bernoulli distribution, Binomial distribution, Double exponential distribution, Exponential distribution, Gamma distribution, Generalized Gamma distribution, Geometric distribution, Negative Binomial distribution, Normal distribution, Lognormal distribution, Poisson distribution, Weibull Distribution

4.1 Introduction

Recall that the probability space $(S, \mathbb{B}, P)$ completely characterizes a random experiment. In practice, the true probability distribution for the random experiment is usually unknown. One usually considers a class of probability measures, say PMF/PDF $f(x, \theta)$, indexed by parameter θ, where the functional form $f(\cdot, \cdot)$ is known. Each parameter value of θ yields a probability distribution, and different values of θ result in different distributions. The collection of these distributions constitute a family of probability distributions. This family of probability distributions is called a class of parametric probability distribution models. One main objective of statistics and econometrics is to use the observed economic data to estimate the true parameter value, say θ_0, under the assumption that there exists some parameter value θ_0 such that $f(x, \theta_0)$ coincides with the true probability distribution $f_X(x)$ of the random experiment. Before we come to the stage of estimating the true model parameter value θ_0 (we will do so in Chapters 6 and 8), we first introduce a number of important parametric distribution models and discuss their properties and applications in economics and finance. We emphasize that it is important to understand the meanings and roles that parameters play in each parametric distribution.

4.2 Discrete Probability Distributions

We start with discrete probability distributions.

4.2.1 Bernoulli Distribution

A DRV X follows a Bernoulli(p) distribution if its PMF

$$\begin{aligned} f_X(x) &= \begin{cases} p & \text{if } x = 1, \\ 1-p & \text{if } x = 0, \end{cases} \\ &= p^x (1-p)^{1-x} \text{ for } x = 0, 1, \end{aligned}$$

where $0 < p < 1$. This is a binary random variable, taking value 1 with probability p, and taking value 0 with probability $1 - p$.

For a Bernoulli(p) random variable, we have

$$\begin{aligned} E(X) &= p, \\ \text{var}(X) &= p(1-p). \end{aligned}$$

All moments can be derived from the MGF of the Bernoulli(p) distribution

$$M_X(t) = pe^t + 1 - p \text{ for } -\infty < t < \infty.$$

We note that all moments of X are a function of p, the only parameter for a Bernoulli distribution. Because $E(X) = P(X = 1)$ and X only takes two possible values, the mean parameter $E(X) = p$ can fully characterize the probability distribution of a Bernoulli random variable.

This binary distribution can arise when one tosses a coin whose head shows up with probability p. It has wide applications in economics and finance. For example, one can define a random variable X to take value 1 if the IBM stock price goes up, and to take value 0 if the IBM stock price goes down. Then X is the directional indicator of the IBM stock price and follows a Bernoulli distribution. Das(2002, "The Surprise Element: Jumps in Interest Rate", *Journal of Econometrics*, 106, 27-65), in a study of jumps in the Fed Funds rates to capture surprise effects, approximates the jumps by a Bernoulli distribution.

In many economic applications, the probability p will vary across individuals as a function of economic variables, say Z. For example, one may assume

$$P(X = 1|Z) = \frac{1}{1 + \exp(-\beta' Z)}.$$

This is called a logit model, which has been a popular econometric model for binary choice problems where the outcome only has two possibilities, and the interest is to explain the choice of economic agents using some economic characteristics.

4.2.2 Binomial Distribution

A DRV X follows a Binomial distribution, denoted as $B(n, p)$, if its PMF

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \cdots, n.$$

where $n \geq 1$ and $0 < p < 1$.

A binomial random variable can take $n + 1$ possible integer values $\{0, 1, \cdots, n\}$. Thus, it is a distribution with finite support. Figure 4.1 shows the probability histogram of the Binomial(n, p) distribution with different choices of n.

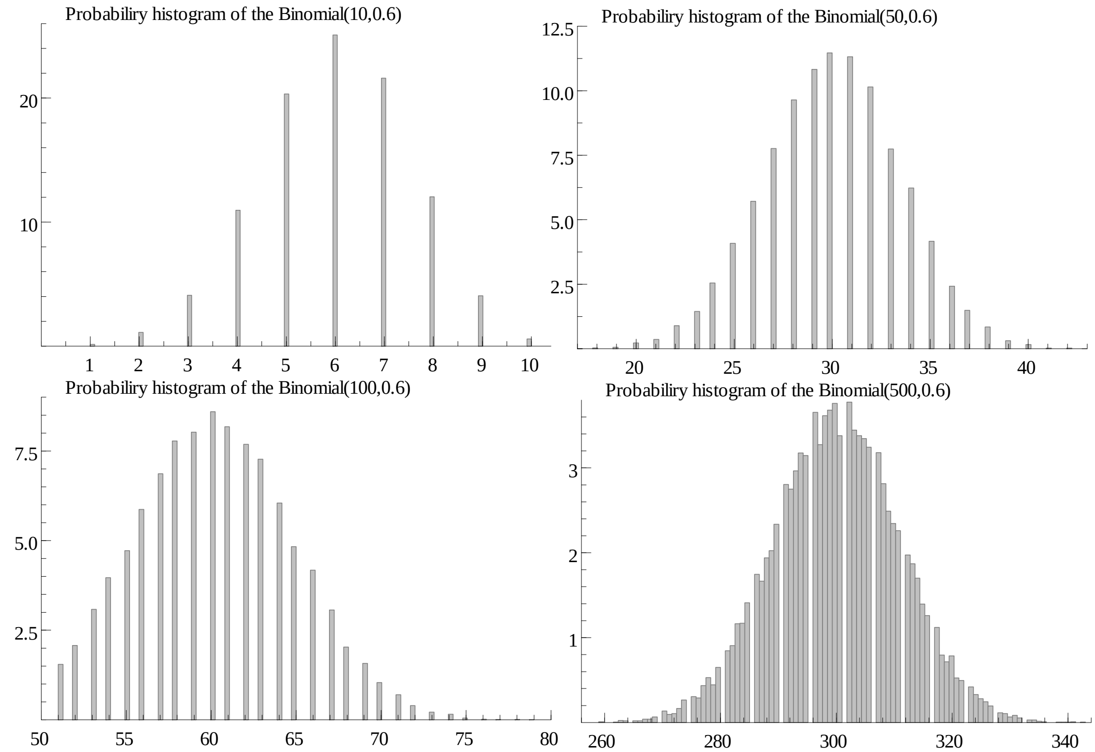

Figure 4.1: Probability Histograms of Binomial(n, p) with Different Choices of n

When can this distribution arise? Suppose one throws a coin n times independently. Each time the head has probability p to occur. How many heads can one get from these n trials? Let X_i denotes the number of heads for the i-th trial, and X denotes the total number of heads in the n trials. Then

$$X = \sum_{i=1}^{n} X_i,$$

where X_i is a sequence of independently identically distribution (IID) Bernoulli(p) random variables. It can be shown that $X = \Sigma_{i=1}^{n} X_i$ follows a $B(n, p)$ distribution (how?).

Question: How to verify

$$\sum_{x=0}^{n} f_X(x) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = 1 \text{ for all } n \geq 1 \text{ and all } p \in (0, 1)?$$

By the binomial theorem that for any real numbers x and y and integer $n \geq 0$,

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i},$$

we can obtain the identity immediately by setting $x = p$ and $y = 1 - p$. This binomial expansion is the reason why the distribution is called the binomial distribution.

We now calculate some moments for the Binomial(n,p) distribution. For the mean of $B(n,p)$, we have

$$\begin{aligned}
E(X) &= \sum_{x=0}^{n} x f_x(x) \\
&= \sum_{x=0}^{n} x\tbinom{n}{x}p^x(1-p)^{n-x} \\
&= \sum_{x=1}^{n} x\tbinom{n}{x}p^x(1-p)^{n-x} + 0\cdot\tbinom{n}{0}p^0(1-p)^{n-0} \\
&= \sum_{x=1}^{n} n\tbinom{n-1}{x-1}p^x(1-p)^{n-x} \quad \text{(by setting } y = x-1\text{)} \\
&= np\sum_{y=0}^{n-1}\tbinom{n-1}{y}p^y(1-p)^{(n-1)-y} \\
&= np\sum_{y=0}^{n-1} f_Y(y) = np,
\end{aligned}$$

where the last sum can be viewed as the sum of the PMF $f_Y(\cdot)$ values of a Binomial $(n-1,p)$ random variable Y so that $\sum_{y=0}^{n-1} f_Y(y) = 1$.

Next, we compute the variance of $B(n,p)$. The second moment

$$\begin{aligned}
E(X^2) &= \sum_{x=0}^{n} x^2 f_X(x) \\
&= \sum_{x=0}^{n} x^2\tbinom{n}{x}p^x(1-p)^{n-x} \\
&= \sum_{x=1}^{n} x\cdot x\tbinom{n}{x}p^x(1-p)^{n-x} \\
&= \sum_{x=1}^{n} xn\tbinom{n-1}{x-1}p^x(1-p)^{n-x} \text{ (setting } y = x-1\text{)} \\
&= n\sum_{y=0}^{n-1}(y+1)\tbinom{n-1}{y}p^{y+1}(1-p)^{(n-1)-y} \\
&= np\sum_{y=0}^{n-1} y\tbinom{n-1}{y}p^y(1-p)^{(n-1)-y} + np\sum_{y=0}^{n-1}\tbinom{n-1}{y}p^y(1-p)^{(n-1)-y} \\
&= npE(Y) + np\sum_{y=0}^{n-1} f_Y(y) \\
&= np(n-1)p + np \\
&= np[(n-1)p+1]
\end{aligned}$$

where the first sum can be viewed as the mean of a Binomial$(n-1,p)$ random variable Y, and the second sum is the sum of the PMF $f_Y(\cdot)$ of Y over its support. Therefore, we have the variance

$$\begin{aligned}\sigma_X^2 &= E(X^2)-\mu_X^2\\ &= \{np\cdot[(n-1)p]+np\}-(np)^2\\ &= np(1-p).\end{aligned}$$

Finally, we compute the MGF

$$\begin{aligned}M_X(t) &= E(e^{tX})\\ &= \sum_{x=0}^{n} e^{tx}\tbinom{n}{x}p^x(1-p)^{n-x}\\ &= \sum_{x=0}^{n}\tbinom{n}{x}(pe^t)^x(1-p)^{n-x}\\ &= (pe^t+1-p)^n, \qquad -\infty<t<\infty,\end{aligned}$$

where the last equality follows from the the binomial formula

$$\sum_{x=0}^{n}\tbinom{n}{x}u^x v^{n-x}=(u+v)^n.$$

The binomial distribution has been one of the oldest to have been the subject of study in statistics. It arises whenever underlying events have two possible outcomes, the chances of which remain constant. The importance of the binomial distribution has been extended from the original application of games to many other areas. The binomial distribution has wide applications in economics. Many experiments can be modeled as a sequence of Bernoulli trials, and the sum of Bernoulli trials follows the binomial distribution. For example, it can be used to approximate the distribution of the numbers of defective products in a total of n products each of which is probability p to be defective. It can also be used to model the cumulative number of jumps occurred in financial price movements during a given period of time (e.g., Das 2002, Journal of Econometrics).

4.2.3 Negative Binomial Distribution

A $B(n,p)$ random variable describes the probability distribution for the number of successes in a fixed number of n trials. Now we are interested in the probability distribution of the number of trials required to obtain a given number of successes. For this reason, we call this distribution as the negative binomial distribution, denoted as $NB(n,p)$.

Specifically, in a sequence of independent Bernoulli(p) trials, let the random variable X denote the number of trials such that at the X-th trial the r-th success occurs, where r is a fixed integer. In other words, $X-1$ is the number of trials right before the r-th success is obtained. Because there are $r-1$ successes in the first $X-1$ trials and the X-th trial is a success, the PMF of X

$$\begin{aligned}f_X(x) &= \left[\tbinom{x-1}{r-1}p^{r-1}(1-p)^{(x-1)-(r-1)}\right]\cdot p\\ &= \tbinom{x-1}{r-1}p^r(1-p)^{x-r}, \qquad x=r,r+1,\cdots,\end{aligned}$$

where $\binom{x-1}{r-1} p^{r-1}(1-p)^{(x-1)-(r-1)}$ is the probability of getting $r-1$ successes with $x-1$ trials, and p is the probability of success in the x-th trial.

The negative binomial distribution is sometimes defined in terms of the number of failures when obtaining the r-th success, $Y = X - r$. As an example, this distribution can be used to model family size when a family prefers to have a given number of boys or a given number of girls (Rao *et al.* 1973).

The support of Y is the set of all nonnegative integers $\{0, 1, \cdots\}$. The PMF of Y

$$\begin{aligned} f_Y(y) &= P(Y = y) \\ &= P(X = y + r) \\ &= \tbinom{y+r-1}{r-1} p^r (1-p)^y, \text{ for } y = 0, 1, \cdots. \end{aligned}$$

4.2.4 Geometric Distribution

The geometric distribution is the probability distribution of the number of Bernoulli trials required to obtain the first success. This is a special case of the negative binomial distribution with $r = 1$. When $r = 1$, the negative binomial distribution becomes

$$f_X(x) = p(1-p)^{x-1}, \qquad x = 1, 2, \cdots.$$

In China, many rural families stop to give birth of a baby until they have the first boy. Thus, the geometric distribution is applicable to model the Chinese rural population.

The geometric distribution is the simplest of the waiting time distributions. The random variable X can be interpreted as the number of trials required to obtain the first success, so we are "waiting for a success" .

The geometric distribution has the so-called "memoryless" or "nonaging" property in the sense that for integers $s > t$, we have

$$P(X > s | X > t) = P(X > s - t).$$

That is, the probability of getting additional $s - t$ failures, having already observed t failures, is the same as the probability of observing $s - t$ failures at the start of the sequence. In other words, the probability of getting a run of failures depends only on the length of the run, not on its position.

Question: How to show the "memoryless" property?

Using the conditional probability formula $P(A|B) = P(A \cap B)/P(B)$ and the fact that when

$s > t$, the event $\{X > t\}$ contains the event $\{X > s\}$ as a subset, we have

$$\begin{aligned} P(X > s|X > t) &= \frac{P(X > s, X > t)}{P(X > t)} \\ &= \frac{P(X > s)}{P(X > t)} \\ &= \frac{1 - P(X \leq s)}{1 - P(X \leq t)} \\ &= \frac{1 - \sum_{x=1}^{s} p(1-p)^{x-1}}{1 - \sum_{x=1}^{t} p(1-p)^{x-1}} \\ &= \frac{(1-p)^s}{(1-p)^t} \\ &= (1-p)^{s-t} \\ &= P(X > s - t). \end{aligned}$$

The geometric distribution is usually viewed as a discrete analog of the exponential distribution to be introduced below. It can be used to model births and populations. For example, many Chinese families in the rural area will stop to give births to their children until they have a boy.

4.2.5 Poisson Distribution

A DRV X follows a Poisson(λ) distribution if its PMF

$$f_X(x|\lambda) = e^{-\lambda}\frac{\lambda^x}{x!}, \qquad x = 0, 1, \cdots,$$

where $\lambda > 0$. The parameter λ is called an intensity parameter.

The support of a Poisson(λ) random variable is the set of all nonnegative integers, and thus is infinite countable. Its PMF is depicted in Figure 4.2 below, for different values of parameter λ.

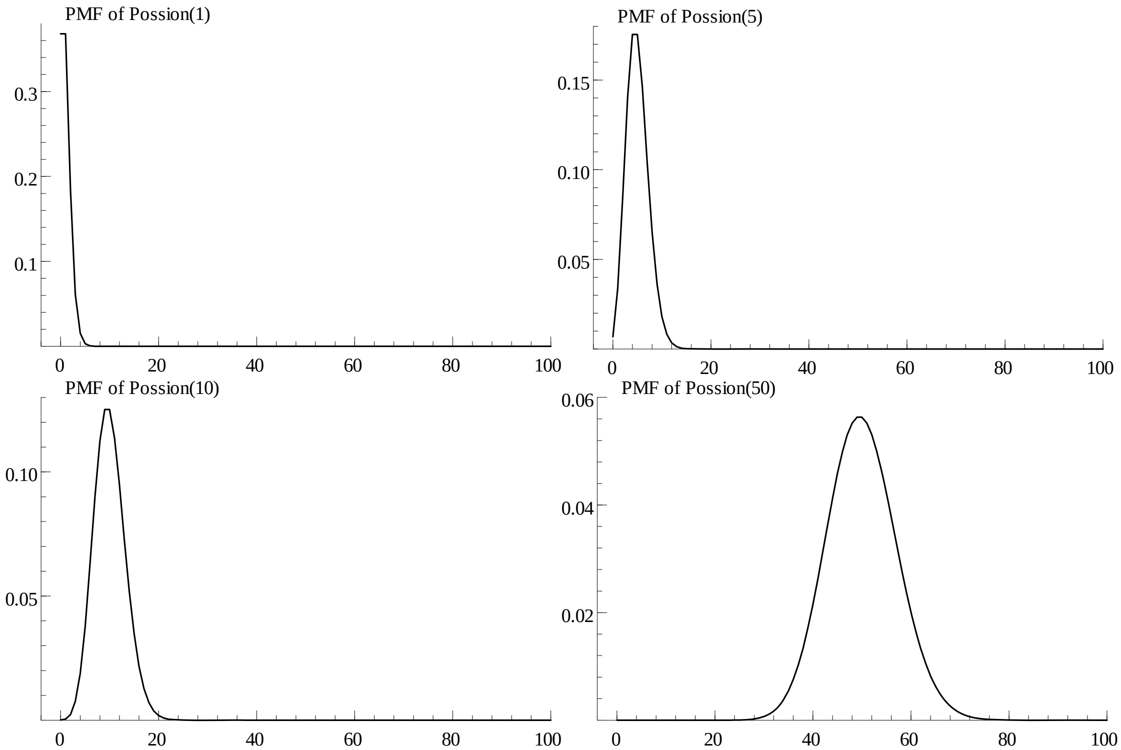

Figure 4.2: PMF of Poisson Distribution for Different Parameter Values λ

We first verify $\sum_{x=0}^{\infty} f_X(x|\lambda) = 1$ for any given $\lambda > 0$. Using MacLaurin's series expansion

$$e^{\lambda} = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!},$$

we have

$$\begin{aligned}
\sum_{x=0}^{\infty} f_X(x|\lambda) &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \\
&= e^{-\lambda} e^{\lambda} \\
&= 1 \ .
\end{aligned}$$

Next, we compute the mean of $X \sim \text{Poisson}(\lambda)$:

$$\begin{aligned}
E(X) &= \sum_{x=0}^{\infty} x e^{-\lambda} \frac{\lambda^x}{x!} \\
&= \sum_{x=1}^{\infty} x e^{-\lambda} \frac{\lambda^x}{x!} \ [\text{using } x! = x \cdot (x-1)!] \\
&= \lambda \sum_{x=1}^{\infty} e^{-\lambda} \frac{\lambda^{x-1}}{(x-1)!} \\
&= \lambda \sum_{y=0}^{\infty} e^{-\lambda} \frac{\lambda^y}{y!} \\
&= \lambda,
\end{aligned}$$

where the last sum is obtained by change of variable $y = x - 1$ and it can be viewed as the sum of the PMF $f_Y(\cdot)$ values of a Poisson(λ) random variable Y. The fact that $E(X) = \lambda$ provides an interpretation for parameter λ : it is the average number for the occurring events following the Poisson(λ) distribution. For example, if X is the number of jumps in an asset price that follow a Poisson(λ) distribution, then λ is the average number of jumps in asset prices.

To obtain the variance of Poisson(λ), we now compute the second moment:

$$\begin{aligned}
E(X^2) &= \sum_{x=0}^{\infty} x^2 e^{-\lambda} \frac{\lambda^x}{x!} \\
&= \sum_{x=1}^{\infty} x e^{-\lambda} \frac{\lambda^x}{(x-1)!} \\
&= \lambda \sum_{x=1}^{\infty} x e^{-\lambda} \frac{\lambda^{x-1}}{(x-1)!} \quad \text{(setting } y = x-1\text{)} \\
&= \lambda \sum_{y=0}^{\infty} (y+1) e^{-\lambda} \frac{\lambda^y}{y!} \\
&= \lambda \sum_{y=0}^{\infty} y e^{-\lambda} \frac{\lambda^y}{y!} + \lambda \sum_{y=0}^{\infty} e^{-\lambda} \frac{\lambda^y}{y!} \\
&= \lambda E(Y) + \lambda \sum_{y=0}^{\infty} f_Y(y) \\
&= \lambda^2 + \lambda,
\end{aligned}$$

where the first sum can be viewed as the mean of a Poisson(λ) random variable Y, and the second sum can be viewed as the sum of the PMF $f_Y(\cdot)$ values of Y over its support.

Therefore, the variance

$$\begin{aligned}
\sigma_X^2 &= E(X^2) - \mu_X^2 \\
&= (\lambda^2 + \lambda) - \lambda^2 \\
&= \lambda.
\end{aligned}$$

It is interesting to note that both mean and variance are equal to λ for a Poisson(λ) distribution.

Finally, the MGF of the Poisson(λ) distribution

$$\begin{aligned} M_X(t) &= E(e^{tX}) \\ &= \sum_{x=0}^{\infty} e^{tx} f_X(x) \\ &= \sum_{x=0}^{\infty} e^{tx} e^{-\lambda} \frac{\lambda^x}{x!} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} \qquad \left(\text{using } e^a = \sum_{x=0}^{\infty} \frac{a^x}{x!}\right) \\ &= e^{-\lambda} e^{(\lambda e^t)} \\ &= e^{\lambda(e^t-1)}, -\infty < t < \infty. \end{aligned}$$

The support of a Poisson(λ) random variable is the set of all nonnegative integers while the support of a Binomial $B(n, p)$ random variable is the set of nonnegative integers $\{0, 1, \cdots, n\}$. When $n \to \infty$, however, there exists a closed link between these two distributions.

Recall the MGF of the binomial distribution $B(n, p)$ is

$$M_B(t) = \left(pe^t + 1 - p\right)^n, \qquad -\infty < t < \infty.$$

When $n \to \infty$ but $np \to \lambda$, i.e., when the occurrence of an event is small or rare and there are many trials, we can use a Poisson(λ) distribution to approximate a binomial distribution $B(n, p)$ because

$$\begin{aligned} M_B(t) &= \left(pe^t + 1 - p\right)^n \\ &= \left[1 + \frac{np(e^t - 1)}{n}\right]^n \\ &\to e^{\lambda(e^t-1)} = M_P(t), \end{aligned}$$

where we have made use of the formula that $(1 + \frac{a}{n})^n \to e^a$ as $n \to \infty$. Therefore, the Binomial $B(n, p)$ distribution is approximately equivalent to a Poisson(λ) distribution when $n \to \infty, np \to \lambda$, i.e., when n is large and p is small. Among other things, this avoids the tedious calculation of the binomial probability formula. Indeed, Poisson (1837) derived the Poisson distribution by considering the limit of a sequence of binomial distributions. This is called the law of small numbers in Bortkiewicz (1898).

Figure 4.3 below plots the Poisson(λ) approximation to the binomial distribution $B(n, p)$, for different values of n, where $np = \lambda = 5$. One can see that even if n is not very large, this approximation works well.

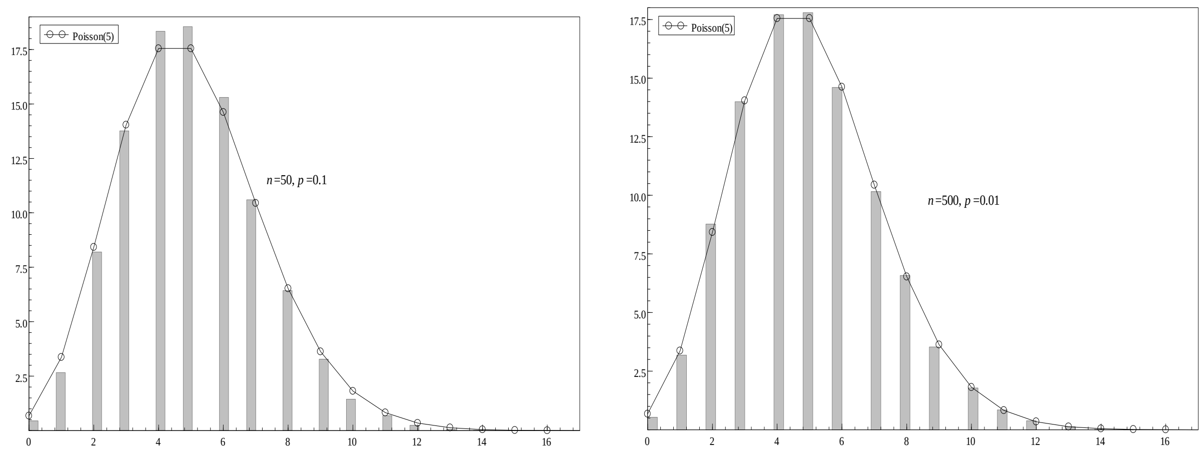

Figure 4.3: the Poisson(λ) approximation to the Binomial distribution $B(n, p)$

In fact, for a negative binomial distribution $NB(r, p)$, when $r \to \infty$ but $r(1-p) \to \lambda \in (0, \infty)$, $NB(r, p)$ will also become a Poisson(λ) distribution, because

$$\begin{aligned} f_Y(y) &= \binom{y+r-1}{r-1} p^r (1-p)^y \\ &\to \frac{e^{-\lambda}\lambda^y}{y!}, \qquad y = 0, 1, \cdots. \end{aligned}$$

We can show this by showing the convergence of the MGF of the negative binomial distribution $NB(r, p)$ to that of the Poisson(λ) distribution as $r \to \infty, r(1-p) \to \lambda$.

The Poisson approximation for the binomial distribution makes the Poisson distribution rather useful in quality control for the number of defective items. More generally, the Poisson distribution has been used to model the distribution of the number of events in a specific period of time or a specific unit of space (e.g., the number of customers passing through a cashier counter, the number of telephone calls, the number of accidents, the number of jumps in an asset price). One of the basic assumptions on which the Poisson distribution is built is that, for small time intervals, the probability of an arrival is proportional to the length of waiting time. This makes it a reasonable model for situations like those just mentioned above. Bortkiewicz (1898) considered circumstances in which Poisson's distribution might arise. In particular, he used the Poisson distribution to characterize the number of deaths from kicks by horses per annum in the Prussian Army Corps where the probability of death from this cause is small while the number of soldiers exposed to the risk was large. In finance the Poisson distribution has been widely used to model the probability distribution of the number of jumps of an asset price over a given period of time. Merton (1976), for example, proposes a model where in addition to a Brownian motion component, the price process of the underlying asset is assumed to have a Poisson jump component. The Poisson regression, namely the analysis of the relationship between an observed count and a set of explanatory variables, is also popularly used in econometrics (e.g., Hausman et al. (1984)). It has been claimed in Douglas (1980) that the Poisson distribution plays a similar role with respect to discrete distributions to that of the normal distribution to be introduced below for absolutely continuous distributions.

4.3 Continuous Probability Distributions

We now introduce a variety of popular continuous probability distributions.

4.3.1 Uniform Distribution

A CRV X follows a uniform probability distribution on the interval $[a, b]$, denoted as $X \sim U[a, b]$, if its PDF

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

Figure 4.4 shows the PDF of the $U[0, 1]$ distribution. Due to the shape of its PDF, the uniform distribution is also called a rectangular distribution.

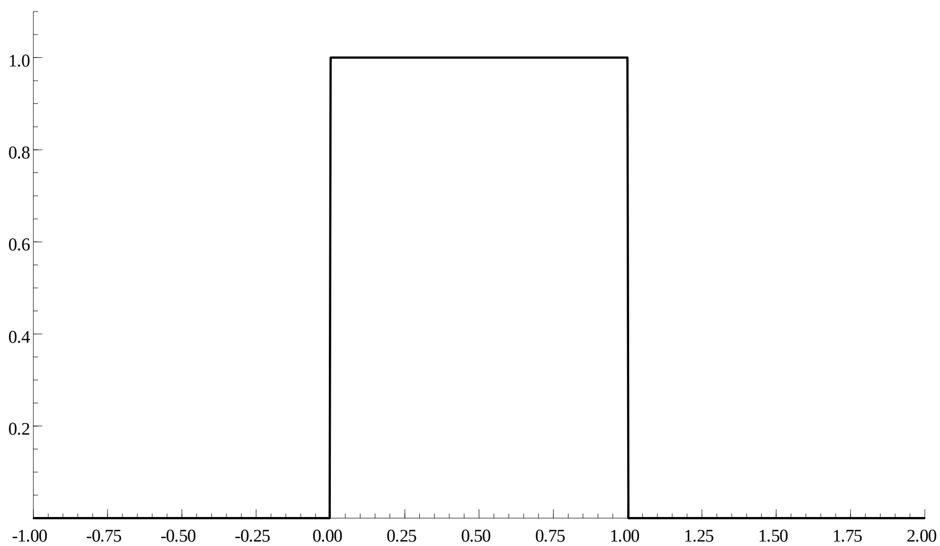

Figure 4.4: PDF of the $U[0, 1]$ Distribution

Because X is a bounded random variable, all its moments exist. The k-th moment

$$\begin{aligned} E(X^k) &= \int_{-\infty}^{\infty} x^k f_X(x) dx \\ &= \frac{1}{b-a} \int_a^b x^k dx \\ &= \frac{1}{b-a} \frac{x^{k+1}}{k+1}\Big|_a^b \\ &= \frac{1}{b-a} \frac{b^{k+1} - a^{k+1}}{k+1}. \end{aligned}$$

When $k = 1$, we obtain the mean of X,

$$\begin{aligned}
\mu_X &= \frac{1}{b-a}\frac{b^2-a^2}{2} \\
&= \frac{1}{2}\frac{b^2-a^2}{b-a} \\
&= \frac{1}{2}(a+b).
\end{aligned}$$

When $k = 2$, we obtain the second moment

$$\begin{aligned}
E(X^2) &= \frac{1}{b-a}\frac{b^3-a^3}{3} \\
&= \frac{1}{3}\frac{b^3-a^3}{b-a} \\
&= \frac{1}{3}(b^2+a^2+ab),
\end{aligned}$$

where we have made use of the formula

$$b^3 - a^3 = (b-a)(b^2+a^2+ab).$$

It follows that the variance of X,

$$\begin{aligned}
\sigma_X^2 &= E(X^2) - \mu_X^2 \\
&= \frac{1}{3}(b^2+a^2+ab) - \frac{1}{4}(a+b)^2 \\
&= \frac{1}{12}(b^2+a^2-2ab) \\
&= \frac{1}{12}(b-a)^2.
\end{aligned}$$

The MGF

$$\begin{aligned}
M_X(t) &= \int_{-\infty}^{+\infty} e^{tx} f_X(x)dx \\
&= \int_a^b e^{tx}\frac{1}{b-a}dx \\
&= \frac{1}{t(b-a)}e^{tx}\Big|_a^b \\
&= \frac{1}{t(b-a)}(e^{tb}-e^{ta}), -\infty < t < \infty.
\end{aligned}$$

When $a = 0, b = 1$, the distribution is called the standard uniform U[0,1] distribution, which has the mean $\frac{1}{2}$ and variance $\frac{1}{12}$. In Chapter 3, we have shown that the probability integral transform $Y = F_X(X)$ follows a U[0,1] distribution. The uniform distribution plays a very important role in statistics and econometrics.

4.3.2 Beta Distribution

A CRV X follows a Beta(α, β) distribution if it has a PDF

$$f_X(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}, \qquad 0 \le x \le 1,$$

where $\alpha > 0, \beta > 0$, and $B(\alpha, \beta)$ is called the Beta function defined as

$$\begin{aligned} B(\alpha, \beta) &= \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx \\ &= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}. \end{aligned}$$

The function $\Gamma(\alpha)$ is called the Gamma function, defined as

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt.$$

The following lemma states some useful properties of the Gamma function $\Gamma(\alpha)$.

Lemma 1 (4.1)**. [Properties of** $\Gamma(\alpha)$**]** :

(1) $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$;
(2) $\Gamma(k) = (k-1)!$ if k is a positive integer;
(3) $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Proof: The proof is left as an exercise.

Question: How to show the identity that

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}?$$

For a Beta(α, β) distribution, the mean

$$\begin{aligned} E(X) &= \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{(\alpha+1)-1}(1-x)^{\beta-1} dx \\ &= \frac{\alpha}{\alpha+\beta} \int_0^1 \frac{(\alpha+\beta)\Gamma(\alpha+\beta)}{\alpha\Gamma(\alpha)\Gamma(\beta)} x^{(\alpha+1)-1}(1-x)^{\beta-1} dx \\ &= \frac{\alpha}{\alpha+\beta} \int_0^1 \frac{\Gamma(\alpha+1+\beta)}{\Gamma(\alpha+1)\Gamma(\beta)} x^{(\alpha+1)-1}(1-x)^{\beta-1} dx \\ &= \frac{\alpha}{\alpha+\beta}, \end{aligned}$$

where the last integral can be viewed as the integral of the PDF of a Beta$(\alpha+1, \beta)$ random variable.

To compute the variance, we first compute the second moment:

$$
\begin{aligned}
E(X^2) &= \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{(\alpha+2)-1}(1-x)^{\beta-1}dx \\
&= \frac{\alpha}{\alpha+\beta}\int_0^1 \frac{\Gamma(\alpha+1+\beta)}{\Gamma(\alpha+1)\Gamma(\beta)} x^{(\alpha+2)-1}(1-x)^{\beta-1}dx \\
&= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}\int_0^1 \frac{\Gamma(\alpha+2+\beta)}{\Gamma(\alpha+2)\Gamma(\beta)} x^{(\alpha+2)-1}(1-x)^{\beta-1}dx \\
&= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)},
\end{aligned}
$$

where the last integral can be viewed as the integral of the PDF of a Beta$(\alpha+2,\beta)$ random variable.

Therefore, we have the variance

$$
\begin{aligned}
\text{var}(X) &= E(X^2) - E^2(X) \\
&= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - (\frac{\alpha}{\alpha+\beta})^2 \\
&= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.
\end{aligned}
$$

The MGF

$$
\begin{aligned}
M_X(t) &= E(e^{tX}) \\
&= \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} e^{tx} x^{\alpha-1}(1-x)^{\beta-1}dx \\
&= 1 + \sum_{j=1}^{\infty}\left(\prod_{i=0}^{j-1}\frac{\alpha+i}{\alpha+\beta+i}\right)\frac{t^j}{j!},
\end{aligned}
$$

where we have made use of MacLaurin's series expansion

$$
e^{tx} = \sum_{j=0}^{\infty}\frac{t^j x^j}{j!}.
$$

Like the uniform distribution, $Beta(\alpha,\beta)$ is one of the few well-known distributions that have bounded support. In fact, the standard uniform distribution on the unit interval $[0,1]$, $U[0,1]$, is a special case of Beta(α,β) with $\alpha=\beta=1$. The shape of the Beta(α,β) distribution depends on the values of parameters (α,β), as can be seen in Figure 4.5. For this reason, α and β are called shape parameters.

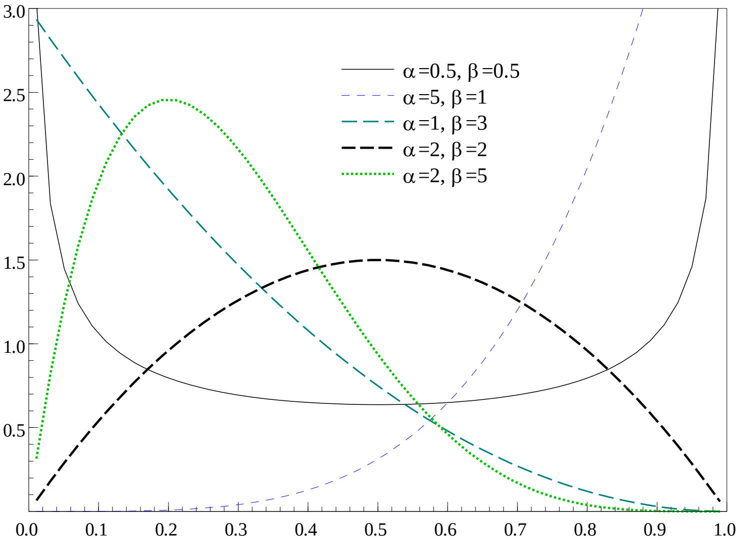

Figure 4.5: PDF of the Beta(α, β) Distribution

Because the support of a Beta(α, β) is [0,1], the Beta distribution can be used to model the probability distribution of proportions or quantities whose values fall into the interval $[0, 1]$. For example, Granger (1980) uses the Beta distribution for the marginal propensities to consume for individual consumers and show that the sum of "short memory" time series processes displays a "long memory" property the remote past consumption is still persistently correlated with the current consumption.

4.3.3 Normal Distribution

A CRV X is called to follow a normal distribution, denoted as $X \sim N(\mu, \sigma^2)$, if it has the PDF

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty < x < \infty,$$

where $-\infty < \mu < \infty$ and $\sigma^2 > 0$.

The parameters μ and σ^2 are location and scale parameters respectively. When $\mu = 0, \sigma^2 = 1$, $X \sim N(0, 1)$ is called a standard normal or unit normal distribution. Tables relating the the standard normal distribution are a necessary ingredient of any standard statistical textbook in statistical theory or its applications. Figure 4.6 shows the PDF of the standard normal and other normal distributions.

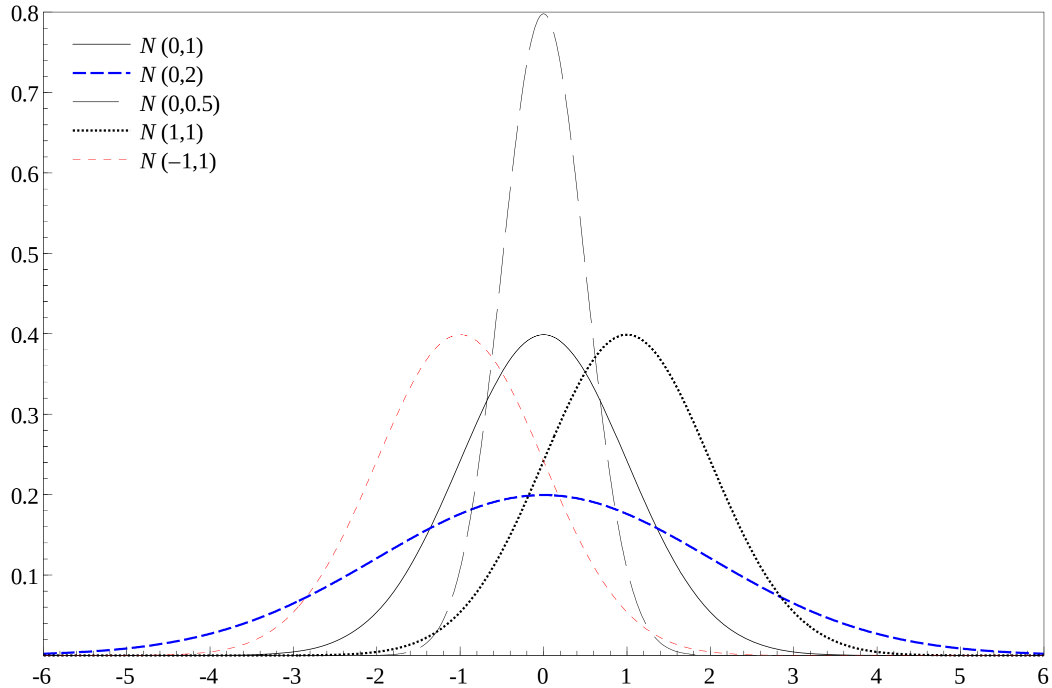

Figure 4.6: PDF's of Normal Distributions

The normal distribution was discovered in 1733 by Abraham De Moivre (1667-1754) in his investigation of approximating coin tossing probabilities. He named the PDF of his discovery the exponentially bell-shaped curve. In 1809, Carl Friedrich Gauss (1777-1855) firmly established the importance of the normal distribution by using it to predict the location of astronomical bodies. As a result, the normal distribution then became commonly known as the Gaussian distribution.

The normal distribution is the most important distribution in probability theory. For many decades the normal distribution have been holding a central position in statistics. Most theoretical arguments for the use of the normal distribution are based on the forms of the Central Limit Theorem (CLT), which says that under suitable conditions, a sample average of n independent and identically distributed (IID) random variables $\{X_1, \cdots, X_n\}$, with suitable centering and standardization, will converge in distribution to a standard normal distribution as the sample size n increases, that is,

$$\sqrt{n}\frac{\bar{X}_n - \mu_X}{\sigma_X} \to^d N(0,1) \text{ as } n \to \infty,$$

where $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$, and $\to^d$denotes convergence in distribution, that is, the distribution of $\sqrt{n}(\bar{X}_n - \mu_X)/\sigma_X$ converges to the $N(0,1)$ distribution as $n \to \infty$. This follows no matter whether X_i a discrete or continuous, is compact supported or uncompact supported. See more discussion on CLT's in Chapter 7.

Question: How to verify

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1 \text{ for all } (\mu, \sigma^2)?$$

Put $y = (x-\mu)/\sigma$. Then it becomes to check whether

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = 1.$$

Because

$$\begin{aligned}
\left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx\right)^2 &= \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dxdy \text{ (set } x = r\cos(\theta), y = r\sin(\theta)) \\
&= \int_{0}^{\infty}\int_{0}^{2\pi} e^{-\frac{r^2}{2}} rdrd\theta \\
&= 2\pi \int_{0}^{\infty} e^{-\frac{r^2}{2}} rdr \\
&= 2\pi,
\end{aligned}$$

it follows that

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = 1.$$

We now compute the moments of the normal distribution. In particular, we show that the mean and variance of X are equal to μ and σ^2 respectively.

First, the mean

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f_X(x) dx \qquad \text{(setting } x = (x-\mu)+\mu) \\
&= \int_{-\infty}^{\infty} (x-\mu) f_X(x) dx + \mu \int_{-\infty}^{\infty} f_X(x) dx \\
&= \int_{-\infty}^{\infty} (x-\mu) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \mu \qquad \text{(setting } y = x-\mu) \\
&= \int_{-\infty}^{\infty} y \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} dy + \mu \\
&= \mu,
\end{aligned}$$

where the integral in the last second equality is identically zero because the integrand $g(y) = y\frac{1}{\sqrt{2\pi}\sigma} e^{-y^2/2\sigma^2}$ is an odd function (i.e. $g(-y) = -g(y)$ for all y).

Next, using integration by part, the variance

$$
\begin{aligned}
\sigma_X^2 &= \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x) dx \\
&= \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \qquad \text{(setting } y = x-\mu\text{)} \\
&= \int_{-\infty}^{\infty} y^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}} dy \\
&= -\sigma^2 \int_{-\infty}^{\infty} y d\left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}}\right) \qquad \left(\text{setting } u = y, v = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-y^2/2\sigma^2}\right) \\
&= -\sigma^2 \left(y \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}} \Big|_{-\infty}^{+\infty} - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}} dy \right) \\
&= \sigma^2,
\end{aligned}
$$

where the last integral can be viewed as the integral of the PDF of a $N(0, \sigma^2)$ random variable Y.

Finally, we derive the MGF of $X \sim N(0, \sigma^2)$.

Theorem 1 (4.2)**.** Suppose $X \sim N(\mu, \sigma^2)$. Then

$$
M_X(t) = e^{\mu t + \frac{\sigma^2}{2} t^2}, \qquad -\infty < t < \infty.
$$

Proof: There are at least two methods to calculate the MGF of X.

(1) Method 1: For the first method, we have

$$
\begin{aligned}
M_X(t) &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{1}{2\sigma^2}[x^2 - 2\mu x + \mu^2]} dx \\
&= e^{-\frac{\mu^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}[x^2 - 2(\mu+\sigma^2 t)x + (\mu+\sigma^2 t)^2 - (\mu+\sigma^2 t)^2]} dx \\
&= e^{-\frac{\mu^2}{2\sigma^2}} e^{\frac{(\mu+\sigma^2 t)^2}{2\sigma^2}} \left\{ \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{[x-(\mu+\sigma^2 t)]^2}{2\sigma^2}} dx \right\} \\
&= e^{\frac{2\mu\sigma^2 t + \sigma^4 t^2}{2\sigma^2}} \times 1 \\
&= e^{\mu t + \frac{1}{2}\sigma^2 t^2}, \quad \text{for } t \in (-\infty, \infty),
\end{aligned}
$$

where the second to last equality follows from the fact that

$$
\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{[x-(\mu+\sigma^2 t)]^2}{2\sigma^2}} dx = 1
$$

for all μ, σ^2 and t.

(2) Method 2: For the second method, we note that $X = \mu + \sigma Y$, where $Y \sim N(0,1)$. By Theorem 3.20, we have

$$\begin{aligned} M_X(t) &= E\left(e^{tX}\right) \\ &= E\left[e^{t(\mu+\sigma Y)}\right] \\ &= e^{\mu t} E(e^{\sigma t Y}) \\ &= e^{\mu t} M_Y(\sigma t). \end{aligned}$$

It suffices to find $M_Y(t)$:

$$\begin{aligned} M_Y(t) &= E\left(e^{tY}\right) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ty} e^{-\frac{1}{2}y^2} dy \\ &= e^{\frac{1}{2}t^2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(y-t)^2} dy \\ &= e^{\frac{1}{2}t^2}. \end{aligned}$$

It follows that

$$M_X(t) = e^{\mu t} M_Y(\sigma t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}.$$

Therefore, we have

$$\begin{aligned} M_X'(t)|_{t=0} &= \left. e^{\mu t + \frac{1}{2}\sigma^2 t^2} (\mu + \sigma^2 t) \right|_{t=0} \\ &= \mu, \\ M_X''(t)|_{t=0} &= \left. \left[e^{\mu t + \frac{1}{2}\sigma^2 t^2} \sigma^2 + e^{\mu t + \frac{1}{2}\sigma^2 t^2} (\mu + \sigma^2 t)^2 \right] \right|_{t=0} \\ &= \sigma^2 + \mu^2. \end{aligned}$$

All centered odd moments $E(X-\mu)^{2k+1} = 0$ for all integers $k \geq 0$ because the normal distribution is symmetric about μ. Suppose we are interested in computing the moments $E(X-\mu)^{2k}$ for an arbitrary positive integer $k > 0$. One can of course differentiate $M_X(t)$ up to $2k$ times, but this is rather tedious when k is large. We now consider some techniques to calculate the moments $E(X-\mu)^{2k}$ by exploiting the duality between integration and differentiation to obtain the higher order moments of a normal random variable X.

For the first method, put $\beta = \frac{1}{2\sigma^2}$ or equivalently $\sigma = \frac{1}{\sqrt{2\beta}}$, we obtain

$$\begin{aligned}
E(X-\mu)^{2k} &= \int_{-\infty}^{\infty} (x-\mu)^{2k} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \\
&= \int_{-\infty}^{\infty} y^{2k} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} dy \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} y^{2k} e^{-\beta y^2} dy \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (-1)^k \frac{d^k}{d\beta^k} e^{-\beta y^2} dy \\
&= \frac{1}{\sqrt{2\pi}\sigma} (-1)^k \frac{d^k}{d\beta^k} \int_{-\infty}^{\infty} \sqrt{2\pi\sigma^2} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}y^2} dy \\
&= \frac{1}{\sqrt{2\pi}\sigma} (-1)^k \frac{d^k}{d\beta^k} \left(\sqrt{2\pi}\sigma\right) \\
&= \frac{1}{\sqrt{2\pi}\sigma} (-1)^k \sqrt{2\pi} \frac{d^k}{d\beta^k} \left(\frac{1}{\sqrt{2\beta}}\right) \quad \text{(noting } \sigma = \frac{1}{\sqrt{2\beta}}\text{)} \\
&= \frac{1}{\sqrt{2}\sigma} (-1)^k \frac{d^k}{d\beta^k} \left(\beta^{-\frac{1}{2}}\right) \\
&= \frac{1}{\sqrt{2}\sigma} (-1)^k \left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right) \cdots \left(\frac{1}{2}-k\right) \beta^{-\frac{1}{2}-k} \\
&= \frac{1}{\sqrt{\pi}} \Gamma\left(k+\frac{1}{2}\right) 2^k \sigma^{2k},
\end{aligned}$$

where the last equality is obtained by using $\beta = \frac{1}{2\sigma^2}$ and $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ from Lemma 4.1(3).

For the special case of $k = 2$, we have

$$E(X-\mu)^4 = (-1)^2 \left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right) 4\sigma^4 = 3\sigma^4.$$

It follows that the kurtosis of $N(\mu, \sigma^2)$ is

$$K = \frac{E(X-\mu)^4}{\sigma^4} = 3.$$

This method of exploiting the duality between integration and differentiation is applicable to many probability distributions, including discrete ones.

In fact, for the normal distribution, there is an alternative simpler method to compute various moments.

Lemma 2 (4.3)**. [Stein's Lemma]:** Suppose $X \sim N(\mu, \sigma^2)$, and $g(\cdot)$ is a differentiable function satisfying $E|g'(X)| < \infty$. Then

$$E\left[g(X)(X-\mu)\right] = \sigma^2 E\left[g'(X)\right].$$

Proof: Using integration by part, we have

$$\begin{aligned}
E[g(X)(X-\mu)] &= \int_{-\infty}^{\infty} g(x)(x-\mu)\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx \\
&= -\int_{-\infty}^{\infty} g(x)d\left[\frac{\sigma^2}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right] \\
&= -g(x)\frac{\sigma^2}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{\sigma^2}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dg(x) \\
&= \sigma^2\int_{-\infty}^{\infty} g'(x)\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx \\
&= \sigma^2 E[g'(X)].
\end{aligned}$$

This completes the proof.

As an example, let us apply this lemma to calculate $E(X-\mu)^4$. We write

$$\begin{aligned}
E(X-\mu)^4 &= E\left[(X-\mu)^3(X-\mu)\right] \\
&= E\left[g(X)(X-\mu)\right],
\end{aligned}$$

where $g(X)=(X-\mu)^3$. By Stein's lemma, we have

$$\begin{aligned}
E(X-\mu)^4 &= \sigma^2 E\left[3(X-\mu)^2\right] \\
&= 3\sigma^4.
\end{aligned}$$

The differentiation of $g(X)$ reduces the order of moments.

4.3.4 Cauchy and Stable Distribution

A CRV X follows a Cauchy(μ,σ) distribution if its PDF

$$f_X(x) = \frac{1}{\pi\sigma}\frac{1}{1+\left(\frac{x-\mu}{\sigma}\right)^2} \text{ for } -\infty < x < \infty,$$

where $\sigma > 0$.

The parameters μ and σ are location and scale parameters respectively. This distribution is symmetric about μ, with unbounded support. When $\mu = 0$ and $\sigma = 1$, the distribution is called a standard Cauchy distribution, denoted as Cauchy$(0,1)$.

There has been little use of the Cauchy distribution in practice. However, it is of special theoretical importance. In particular, the Cauchy distribution has some peculiar properties and could provide counter examples to some generally accepted results and concepts in statistics.

Compared to the normal distribution, the Cauchy distribution has very long and heavy tails. The most notable difference between the normal and Cauchy distributions is in the longer and flatter tails of the latter. For a Cauchy(μ,σ) distribution, the tail of the PDF decays to zero at a very slow hyperbolic rate: $f_X(x) \sim x^{-2}$ as $|x| \to \infty$. As a consequence, all finite moments of order greater than or equal to 1 do not exist, and so its MGF does not exist either. For example, for the Cauchy$(0,1)$ distribution, we have

$$
\begin{aligned}
E|X| &= \int_{-\infty}^{\infty} |x| f_X(x) dx \\
&= \int_{-\infty}^{\infty} |x| \frac{1}{\pi} \frac{1}{1+x^2} dx \\
&= \frac{2}{\pi} \int_0^{\infty} \frac{x}{1+x^2} dx \\
&= \frac{1}{\pi} \ln(1+x^2)|_0^{\infty} \\
&= \infty.
\end{aligned}
$$

This implies the mean and all higher order moments do not exist. Thus, the location parameter μ cannot be interpreted as the mean, and the scale parameter σ cannot be interpreted as the standard deviation.

The characteristic function of the Cauchy(μ,σ) distribution is

$$
\begin{aligned}
\varphi_X(t) &= E\left(e^{itX}\right) \\
&= e^{i\mu t - \sigma|t|}.
\end{aligned}
$$

This characteristic function is not differentiable with respect to t at the origin, which is consistent with the fact that all finite moments of order greater than or equal to 1 do not exist.

Question: When can a Cauchy distribution arise?

A Cauchy distribution can arise when a ratio of two independent normal random variables is considered.

In fact, both the Cauchy distribution and the normal distribution belong to the so-called stable distribution, which has very considerable importance in probability theory.

Question: What is a stable distribution?

For a stable distribution, its PDF is usually unknown in closed form. However, its characteristic function has a closed form

$$\varphi_X(t) = e^{\mathbf{i}\mu t - \sigma|t|^c[1+\mathbf{i}\lambda \operatorname{sgn}(t)\omega(|t|,c)]}$$

where $\mathbf{i} = \sqrt{-1}$, $0 < c \leq 2, -1 \leq \lambda \leq 1, \sigma > 0$, and

$$\omega(|t|,c) = \begin{cases} \tan\left(\frac{1}{2}\pi c\right), & c \neq 1, \\ -2/\pi \ln(|t|), & c = 1, \end{cases}$$

and

$$\operatorname{sgn}(t) = \begin{cases} 1, & \text{if } t > 0, \\ 0, & \text{if } t = 0, \\ -1, & \text{if } t < 0. \end{cases}$$

Intuitively, μ is a location parameter, σ is a scale parameter, c is a tail parameter, and λ is a skew parameter. The shape of the PDF is determined by both c and λ. If $\lambda = 0$, the distribution is symmetric. When $\lambda = 0$, $c = 2$ gives a normal distribution, and $c = 1$ gives a Cauchy distribution. Figure 4.7 plots the PDF's for the stable distribution with various choices of parameters $(\mu, \sigma, c, \lambda)$.

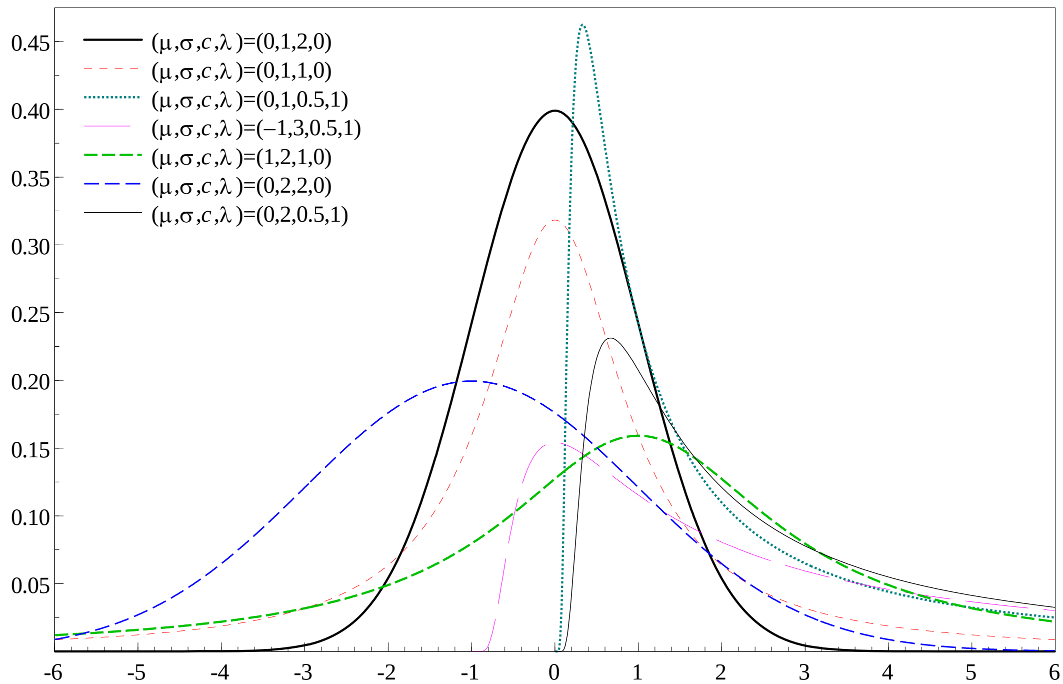

Figure 4.7: The PDF's for the Stable Distribution with Various Choices of Parameters $(\mu, \sigma, c, \lambda)$.

The moments of a stable distribution exist only when $c > 1$. When the sum of independent identically distributed stable random variables has a limiting distribution, it must be a stable distribution rather than the normal distribution. Thus nonnormal stable distributions generalize the Central Limit Theorem (CLT) to the cases where the second moment of the summed variables are infinite

The stable distributions are closely related to the Levy process which has recently become a hot topic in financial econometrics. They are more appropriate to model heavy tails which are often observed in financial data. Mandelbrot (1963) and Fama (1965) have applied the stable distributions to model stock returns.

4.3.5 Lognormal Distribution

A CRV X follows a Lognormal(μ, σ^2) distribution if its PDF

$$f_X(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma}\frac{1}{x}e^{\frac{1}{2\sigma^2}(\ln x - \mu)^2}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

Figure 4.8 shows the shape of the lognormal distribution with different values of parameters (μ, σ^2).

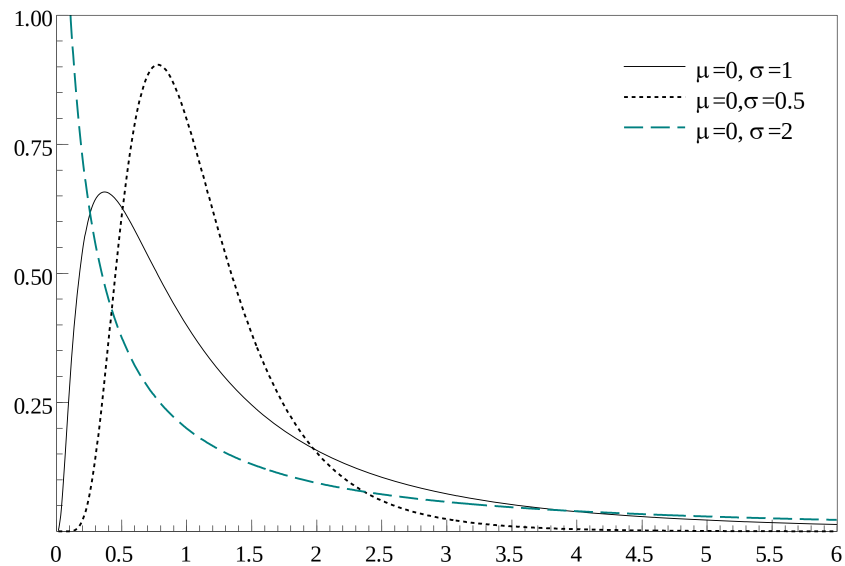

Figure 4.8: PDF of the Lognormal Distribution

Using the transformation method in Chapter 3, we can show that $Y = \ln(X) \sim N(\mu, \sigma^2)$. Indeed, X is called a lognormal random variable because its logarithm follows a normal distribution. The lognormal distribution is sometimes called the antilognormal distribution. This name has some logical basis in that it is not the distribution of the logarithm of a normal variable but of the exponential—i.e., antilogarithmic function of such a variable.

In fact, there is a more general definition of the lognormal distribution. Suppose $Y = \ln(X - \alpha) \sim N(\mu, \sigma^2)$. Then the random variable X is said to follow a Lognormal(α, μ, σ^2) distribution. Since parameter α only affects the mean of X, we consider the two-parameter lognormal(μ, σ^2) distribution here.

Recall from Theorem 4.2 that the MGF of a normal random variable $Y \sim N(\mu, \sigma^2)$ is

$$M_Y(t) = E\left(e^{tY}\right) = e^{\mu t + \frac{\sigma^2}{2} t^2}.$$

It follows that all moments of the lognormal(μ, σ^2) random variable X exist and are given by

$$\begin{aligned} E(X^k) &= E\left(e^{kY}\right) \\ &= M_Y(k) \\ &= e^{k\mu + \frac{\sigma^2}{2} k^2}, \qquad k = 1, 2, \cdots. \end{aligned}$$

In particular, we have the mean

$$\mu_X = e^{k\mu + \frac{\sigma^2}{2}},$$

and the variance

$$\begin{aligned} \sigma_X^2 &= E(X^2) - \mu_X^2 \\ &= e^{2\mu + \sigma^2}\left(e^{\sigma^2} - 1\right). \end{aligned}$$

It is important to note that parameters μ and σ^2 are not the mean and variance of the Lognormal(μ, σ^2) distribution.

Although all moments exist, the MGF does not exist for a lognormal distribution. To see this, we consider

$$
\begin{aligned}
M_X(t) &= E(e^{tX}) \\
&= \int_0^\infty e^{tx} \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{x} e^{-\frac{1}{2\sigma^2}(\ln x - \mu)^2} dx \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma} e^{te^{\ln x} - \frac{1}{2\sigma^2}(\ln x - \mu)^2} d\ln x \\
&= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\sigma} e^{te^y - \frac{1}{2\sigma^2}(y-\mu)^2} dy \text{ (by setting } y = \ln x) \\
&\geq \int_c^{c+1} \frac{1}{\sqrt{2\pi}\sigma} e^{te^y - \frac{1}{2\sigma^2}(y-\mu)^2} dy \text{ for any } c > 0 \\
&\geq \frac{1}{\sqrt{2\pi}\sigma} e^{te^c - \frac{1}{2\sigma^2}(c-\mu)^2} (c + 1 - c) \text{ for } c \text{ sufficiently large} \\
&= \frac{1}{\sqrt{2\pi}\sigma} e^{te^c - \frac{1}{2\sigma^2}(c-\mu)^2} \\
&\to \infty \text{ as } c \to \infty
\end{aligned}
$$

because $te^c - (c - \mu)^2/2\sigma^2 \to \infty$ as $c \to \infty$ and $t > 0$.

Intuitively, the lognormal distribution has a long right-tail. When $t > 0$, the exponential function e^{tX} takes rather large values with a relatively large probability, rendering the non-existence of the expectation of e^{tX}.

Question: What is the characteristic function of the Lognormal(μ, σ^2) distribution?

The lognormal distribution is very popular in modeling applications when the variable of interest is nonnegative and skewed to the right. In particular, it has been widely used to model the distribution of asset prices, commodity prices, incomes and populations. To appreciate this, we consider following nonnegative economic variable

$$X_t = X_{t-1}(1 + Y_t),$$

where $\{Y_t\}$ are a sequence of IID random variables such that Y_t is independent of X_{t-1}.

The recursive relationship leads to

$$X_n = X_0 \prod_{t=1}^n (1 + Y_t) .$$

and so

$$\ln X_n = \ln X_0 + \sum_{t=1}^n \ln(1 + Y_t).$$

By the CLT, for large n, $\Sigma_{t=1}^{n} \ln(1+Y_t)$ will be approximately normally distributed, after suitable standardization. Thus, its exponential, $\Pi_{t=1}^{n}(1+Y_t)$ is approximately lognormally distributed.

Among other things, the lognormality assumption offers a great deal of convenience in analysis. To see this, suppose a stock price $P_t \sim \text{lognormal}(\mu t, \sigma^2 t)$, where the time t changes continuously. Then $\ln P_t \sim N(\mu t, \sigma^2 t)$, and the log-return

$$\begin{aligned} R_t &= \ln(P_t/P_{t-1}) \\ &= \ln(P_t) - \ln(P_{t-1}) \end{aligned}$$

which is approximately equal to the relative price change from time $t-1$ to time t, is also normally distributed. Furthermore, the cumulative return $\sum_{t=1}^{m} R_t$ over m time periods from $t = 1$ to $t = m$ is also normally distributed. Black and Scholes (1973) use the lognormal distribution for the underlying stock price in deriving the European options prices.

Lognormal distributions have been also found useful in representing the distribution of size for varied kinds of natural economic units (e.g., Gibrat (1930, 1931)). The relationship between leaving a company and employees tenure has been described by lognormal distributions with great success (Young 1971, McClean 1976). O'Neill and Wells (1972) point out that the lognormal distribution can be effectively shed to fit the distribution for individual insurance claim payments. The lognormal distribution is also a serious competitor to the Weibull distribution to be introduced below in modeling lifetime distributions.

4.3.6 Gamma and Generalized Gamma Distributions

A nonnegative CRV X follows a Gamma distribution $G(\alpha, \beta)$ if its PDF

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, & x > 0, \\ 0, & x \le 0, \end{cases}$$

where $\alpha, \beta > 0$, and $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$ is the Gamma function.

The Gamma(α, β) distribution is a flexible family of distributions for a nonnegative random variable on $[0, \infty)$. Here, α is a shape parameter, and β is a scale parameter controlling the spread of the distribution. When $\beta = 1$, the Gamma$(\alpha, 1)$ distribution is called a standard Gamma distribution. Figure 4.9 shows the shape of the Gamma distribution with different values of parameters (α, β).

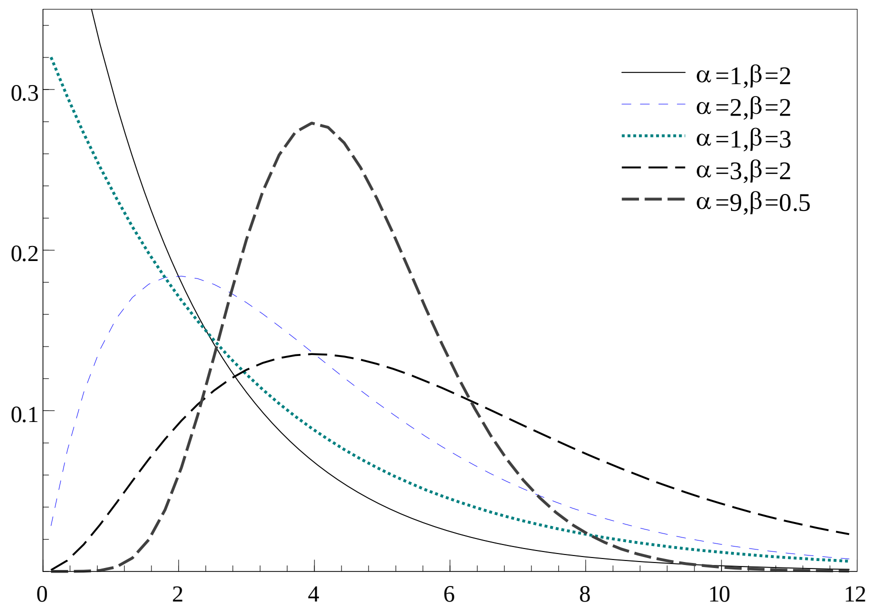

Figure 4.9: PDF of the Gamma Distribution

Question: How to verify that for any given parameters $\alpha, \beta > 0$, the integral

$$\int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx = 1?$$

By change of variable and the definition of the Gamma function, we have

$$\begin{aligned}
\int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx &= \frac{1}{\Gamma(\alpha)} \int_0^\infty (x/\beta)^{\alpha-1} e^{-(x/\beta)} d(x/\beta) \\
&= \frac{1}{\Gamma(\alpha)} \int_0^\infty y^{\alpha-1} e^{-y} dy \text{ (setting } y = x/\beta) \\
&= \frac{1}{\Gamma(\alpha)} \Gamma(\alpha) \\
&= 1.
\end{aligned}$$

We now derive the mean, variance and MGF of X. The mean

$$\begin{aligned}
\mu_X &= \int_{-\infty}^\infty x f_X(x) dx \\
&= \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^\alpha e^{-x/\beta} dx \\
&= \int_0^\infty \frac{\alpha\beta}{\alpha\Gamma(\alpha)\beta^{(a+1)}} x^{(\alpha+1)-1} e^{-x/\beta} dx \quad \text{(setting } \alpha^* = \alpha + 1 \text{ \& using } \alpha\Gamma(\alpha) = \Gamma(\alpha+1)) \\
&= \alpha\beta \int_0^\infty \frac{1}{\Gamma(\alpha^*)\beta^{\alpha^*}} x^{\alpha^*-1} e^{-x/\beta} dx \\
&= \alpha\beta,
\end{aligned}$$

where the last integral can be viewed as the integral of the PDF of a Gamma(α^*, β) random variable.

The second moment

$$\begin{aligned} E(X^2) &= \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha+2-1} e^{-x/\beta} dx \\ &= \int_0^\infty \frac{\alpha(\alpha+1)\beta^2}{\Gamma(\alpha+2)\beta^{\alpha+2}} x^{\alpha+2-1} e^{-x/\beta} dx \\ &= \alpha(\alpha+1)\beta^2 \int_0^\infty \frac{1}{\Gamma(\alpha+2)\beta^{\alpha+2}} x^{\alpha+2-1} e^{-x/\beta} dx \\ &= \alpha(\alpha+1)\beta^2, \end{aligned}$$

where the last integral can be viewed as the integral of the PDF of a Gamma($\alpha + 2, \beta$) random variable. It follows that the variance

$$\sigma_X^2 = E(X^2) - \mu_X^2 = \alpha\beta^2$$

The MGF

$$\begin{aligned} M_X(t) &= \int_0^\infty e^{tx} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx \\ &= \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x(1/\beta - t)} dx \quad \left(\text{setting } \beta^* = \frac{1}{1/\beta - t}\right) \\ &= \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta^*} dx \\ &= \frac{(\beta^*)^\alpha}{\beta^\alpha} \int_0^\infty \frac{1}{\Gamma(\alpha)(\beta^*)^\alpha} x^{\alpha-1} e^{-x/\beta^*} dx \\ &= \frac{(\beta^*)^\alpha}{\beta^\alpha} \\ &= (1 - \beta t)^{-\alpha}, \quad t < 1/\beta, \end{aligned}$$

where the last integral can be viewed as the integral of the PDF of a Gamma(α, β^*) random variable.

The Gamma distribution has a similar shape to the lognormal distribution. It has been used to model the distribution of the continuous waiting time of economic events (e.g., unemployment duration, price duration, poverty duration, etc.). It can also be used to model distribution of nonnegative random variables, such as income, population, and range. Cox, Ingersoll and Ross (1985) propose an equilibrium model for the term structure of the spot interest rate. They assume that the spot interest rate process follows a square root process, that is,

$$dr_t = k(\theta - r_t)dt + \sigma\sqrt{r_t} dB_t$$

where B_t is a Brownian motion process. For $k, \theta > 0$, the interest rate is elastically pulled toward a central location or long-term value θ, the parameter k determines the speed of adjustment.

This specification precludes negative interest rates as may occur in Vasicek's (1977) model which assumes

$$dr_t = k(\theta - r_t)dt + \sigma dB_t.$$

It can be shown that the Cox, Ingersoll and Ross (1985) model admits a steady state Gamma distribution (stationary distribution) with PDF:

$$f(r) = \frac{1}{\Gamma(\alpha)\beta^\alpha} r^{\alpha-1} e^{-r/\beta}$$

where $\alpha = \frac{2k\theta}{\sigma^2}$ and $\beta = \frac{\sigma^2}{2k}$. As a result, the steady state mean and variance of the short-term interest rate r_t are θ and $\sigma^2\theta/2k$ respectively.

There is a closely related distribution called Generalized Gamma distribution. Suppose the random variable

$$Y = \left(\frac{X-\gamma}{\beta}\right)^c$$

follows a standard Gamma distribution, i.e., $Y \sim \text{Gamma}(\alpha, 1)$ or equivalently it has the PDF

$$f_Y(y) = \frac{y^{\alpha-1}e^{-y}}{\Gamma(\alpha)}, \qquad y \geq 0.$$

Then the random variable X is said to follow a Generalized Gamma distribution with parameters where α and c are shape parameters, β is a scale parameter, and γ is a location parameter. With the univariate transformation method (Theorem 3.13), it can be shown that the PDF of X

$$f_X(x) = \frac{c}{\Gamma(\alpha)\beta^{c\alpha}}(x-\gamma)^{c\alpha-1}e^{-\left(\frac{x-\gamma}{\beta}\right)^c}, \qquad x \geq \gamma.$$

Often one can set $\gamma = 0$ in practice. Figure 4.10 plots the shape of the PDF of the Generalized Gamma distribution with various choices of parameters, particularly the choice of parameter c.

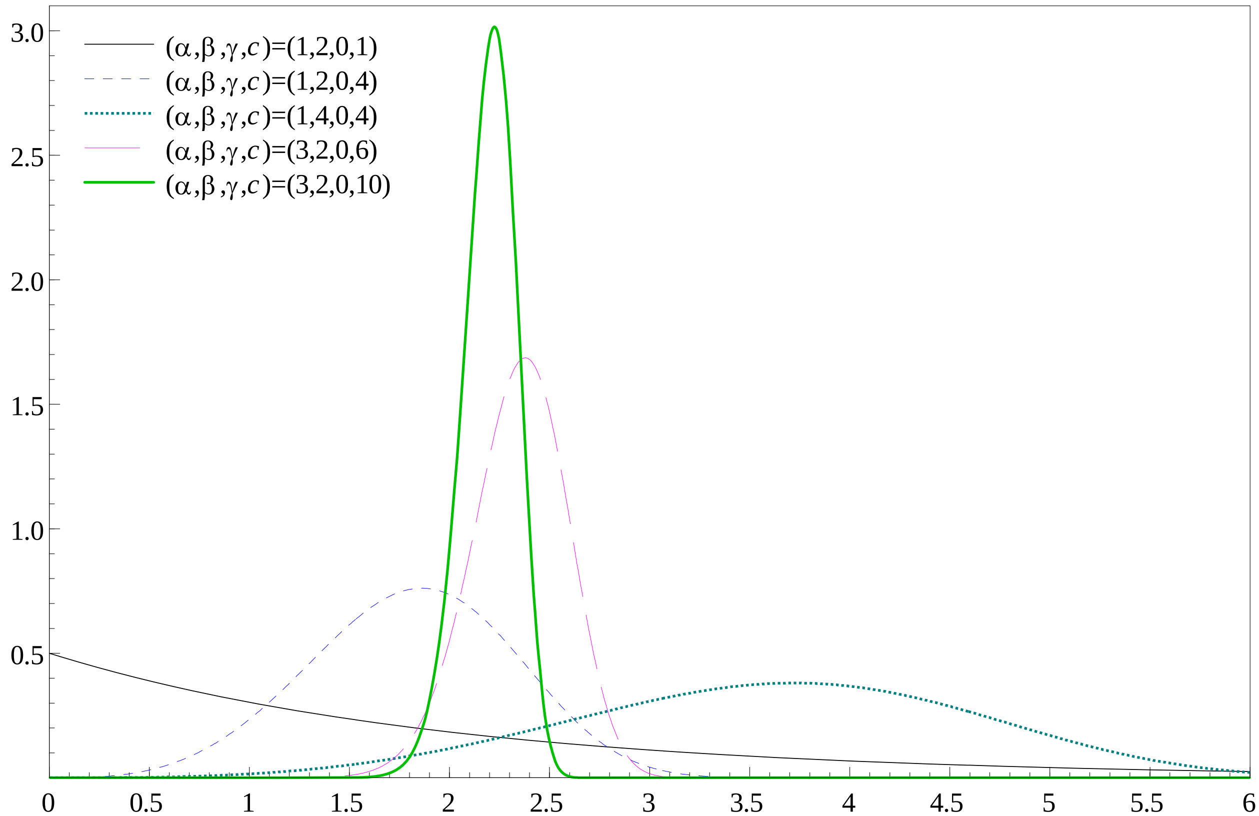

Figure 4.10: PDF of the Generalized Gamma Distribution with Various Choices of Parameter Values

We note that the moments of the Generalized Gamma distribution can be obtained from the moments of the standard Gamma$(\alpha, 1)$ distribution by observing the following relationship:

$$\begin{aligned} E\left[\left(\frac{X-\gamma}{\beta}\right)^k\right] &= E\left[\left(\frac{X-\gamma}{\beta}\right)^{c(k/c)}\right] \\ &= E\left(Y^{(k/c)}\right) \\ &= \frac{\Gamma(\alpha+k/c)}{\Gamma(\alpha)}. \end{aligned}$$

4.3.7 Chi-Square Distribution

We now consider a special case of the Gamma distribution.

A nonnegative CRV X follows a Chi-squared distribution with ν degrees of freedom, denoted as χ^2_ν, if its PDF

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(\frac{\nu}{2})\sqrt{2^\nu}} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}}, & x > 0, \\ 0, & x \le 0. \end{cases}$$

The χ^2_ν distribution is a special case of the Gamma(α, β) distribution with $\alpha = \frac{\nu}{2}$, and $\beta = 2$. Its k-th moment is

$$E(X^k) = \frac{2^k \Gamma(\frac{\nu}{2} + k)}{\Gamma(\frac{\nu}{2})}.$$

In particular, its mean is

$$E(X) = \nu;$$

its variance is

$$\text{var}(X) = 2\nu;$$

The MGF

$$M_X(t) = (1 - 2t)^{-\frac{\nu}{2}} \text{ for } t < \frac{1}{2}.$$

The χ^2_ν distribution defined this way allow ν, the degree of freedom parameter, to be a non-integer value. Later, in Chapter 5, we will see that when ν is an integer, a χ^2_ν distribution is equivalent to that of the sum of ν squared independent $N(0, 1)$ random variables.

The χ^2_ν distribution is a right-skewed distribution. When the degree of freedom $\nu \to \infty$, χ^2_ν becomes an approximately normal distribution with with mean ν and variance 2ν.

Like the normal distribution, the χ^2_ν distribution is one of the most important distributions in statistics and has occupied a central position in econometrics. Many popular test statistics in econometrics are in a quadratic form and have an asymptotic χ^2_ν distribution.

4.3.8 Exponential and Weibull Distributions

Another special case of the Gamma distribution is called an Exponential distribution.

A nonnegative CRV X follows an Exponential(β) distribution, if its PDF

$$f_X(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

where $\beta > 0$. The parameter β is a scale parameter. When $\beta = 1$, X is called to follow the standard exponential distribution, denoted as EXP(1).

The Exponential(β) distribution is a special case of the Gamma distribution $G(1, \beta)$. The MGF

$$M_X(t) = E(e^{tX}) = \frac{1}{1 - \beta t}, \qquad t < \frac{1}{\beta}.$$

The mean

$$E(X) = \beta$$

and the variance

$$\text{var}(X) = \beta^2.$$

Like the geometric distribution, the exponential distribution also has the so-called "memory-less" property in the sense that for any positive numbers x and y, $x > y$,

$$P(X > x|X > y) = P(X > x - y).$$

To see this, note that for an Exponential(β) distribution, the CDF

$$F_X(x) = 1 - e^{-x/\beta} \text{ for } x \geq 0.$$

It follows that when $x > y$,

$$\begin{aligned} P(X &> x|X > y) = \frac{P(X > x, X > y)}{P(X > y)} \\ &= \frac{P(X > x)}{P(X > y)} \\ &= \frac{1 - F_X(x)}{1 - F_Y(y)} \\ &= \frac{e^{-\beta x}}{e^{-\beta y}} \\ &= e^{-\beta(x-y)} \\ &= P(X > x - y). \end{aligned}$$

The exponential distribution can be viewed as a continuous analog of the discrete geometric distribution introduced in Section 4.2.4.

Why is the exponential distribution useful in economics and finance?

The exponential distribution is of considerable importance and widely used in statistics and econometrics. Like the Gamma distribution, the exponential distribution has been used to model time durations of economic events, such as the unemployment spell of a worker, time before a credit default, the time between two trades or price changes, and etc. There are many situations where one would expect an exponential distribution to give a useful description of observed phenomena. Below is an example in labor economics: Let X be the unemployment duration of a worker which has a PDF $f_X(x)$. Then the so-called hazard rate or hazard function is defined as

$$\begin{aligned} \lambda(x) &= \lim_{\Delta x \to 0^+} \frac{P(X \leq x + \Delta x | X \geq x)}{\Delta x} \\ &= \lim_{\Delta x \to 0^+} \frac{P(x \leq X \leq x + \Delta x)}{P(X \geq x)\Delta x} \\ &= \left[\lim_{\Delta x \to 0^+} \frac{\int_x^{x+\Delta x} f_X(u)du}{\Delta x} \right] \cdot \frac{1}{P(X \geq x)} \\ &= \frac{f_X(x)}{P(X \geq x)} \\ &= \frac{f_X(x)}{1 - F_X(x)} \\ &= -\frac{d}{dx} \ln[1 - F_X(x)]. \end{aligned}$$

Intuitively, the hazard function $\lambda(x)$ is the instantaneous probability that the unemployed worker will find a job after an unemployment duration of x. Duration analysis is to model $\lambda(x)$ using economic explanatory variables. A simplest example is to assume that the hazard rate is a constant of the unemployment duration x, that is,

$$\lambda(x) = \lambda_0 \text{ for all } x.$$

Then the corresponding distribution of the unemployment duration X follows an Exponential$(1/\lambda_0)$ distribution:

$$f_X(x) = \lambda_0 e^{-\lambda_0 x} \text{ for } x > 0.$$

In financial econometrics, an empirical stylized fact of high-frequency stock return $\{X_t\}$ is that the absolute value of the stock return $|X_t|$ approximately follows a standard exponential distribution (Ding, Granger and Engle 1993, Journal of Empirical Finance). Here, X_t is the standardized financial return in time period t.

There are many important distributions closely related to the exponential distribution. For example, suppose $Y = (X - \alpha)^c$ follows an exponential distribution with parameter β. Then X is said to have a Weibull distribution. Its PDF is

$$f_X(x) = \frac{c}{\beta}(x-\alpha)^{c-1} e^{-\frac{(x-\alpha)^c}{\beta}}, \qquad x > \alpha,$$

where α is a location parameter, β is a scale parameter, and c is a shape parameter. It is necessary that c be greater than 1 (why?). Often, one can set $\alpha = 0$ in many applications.

The Weibull distribution is more flexible than the exponential distribution. For example, the associated hazard function is no longer a constant function. Thus, it is very useful to model hazard functions. Figure 4.11 plots the shapes of the Weibull distribution with various choices of parameters (α, β, c) :

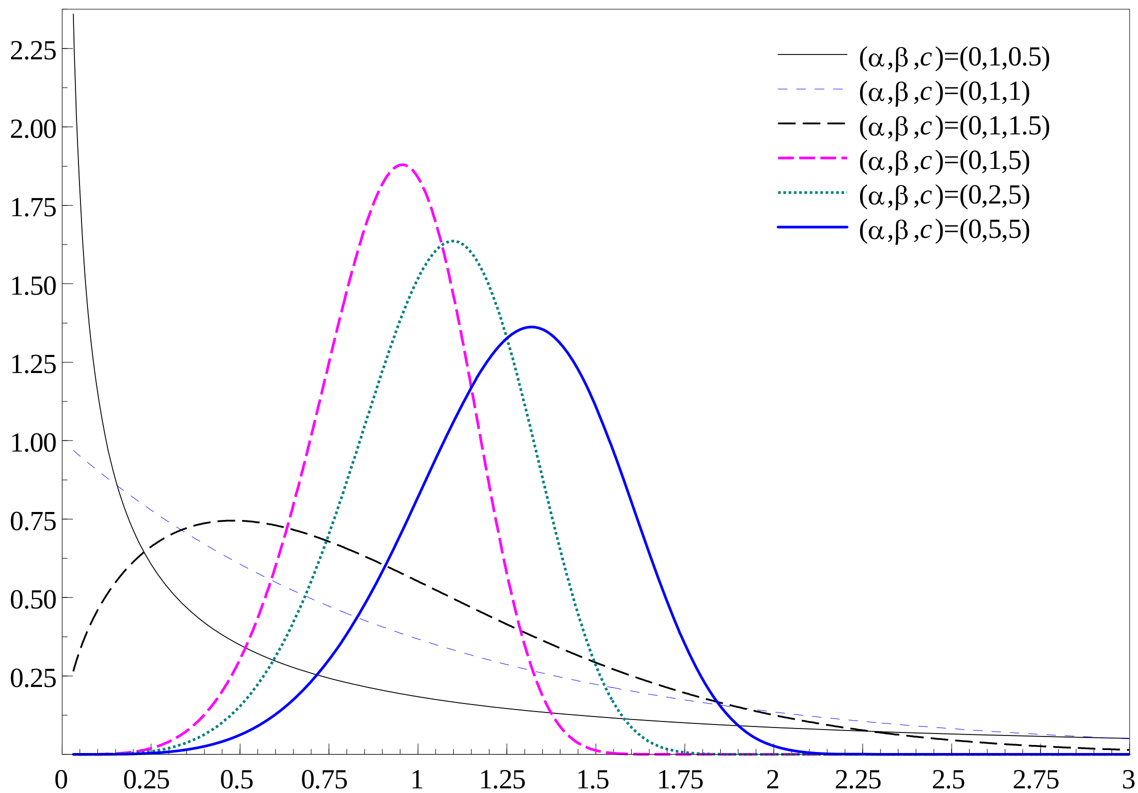

Figure 4.11: PDF of the Weibull Distribution with Carious Choices of Parameter Values

The Weibull distribution is used in Engle and Russell (1998) to model the time duration between trades or price changes in finance.

On the other hand, if $Y = e^{-X}$ follows an Exponential(β) distribution, then X is said to follow an extreme value distribution. Such a distribution is used in Nelson (1991) to model the asymmetric volatility dynamics of stock returns in a time series context.

4.3.9 Double Exponential Distribution

A continuous random variable X follows a Double Exponential(α, β) distribution if its PDF

$$f_X(x) = \frac{1}{2\beta} e^{-\frac{|x-\alpha|}{\beta}}, \qquad -\infty < x < \infty,$$

where $\beta > 0$.

The double exponential(α, β) distribution is a symmetric distribution about α, but has a fatter tail than a normal distribution. It has a peak at $x = \alpha$ where the derivative does not exist.

The mean of X,

$$E(X) = \alpha,$$

the variance

$$\text{var}(X) = 2\beta^2,$$

and the MGF

$$M_X(t) = \frac{e^{\alpha t}}{1 - \beta^2 t^2} \text{ for } |t| < \frac{1}{\beta}.$$

The double exponential distribution is also called the Laplace distribution. When $\alpha = 0$, the absolute value of X, $Y = |X|$, follows an Exponential(β) distribution (see Example 3.24 in Chapter 3).

Although the double exponential(α, β) distribution has heavier tails than the normal distribution, all its moments exist, and has lighter tails than the Cauchy distribution. Figure 4.12 shows a Cauchy$(0, 1)$ distribution, a standard normal distribution $N(0, 1)$, together with the double exponential$(0, 1)$ distribution.

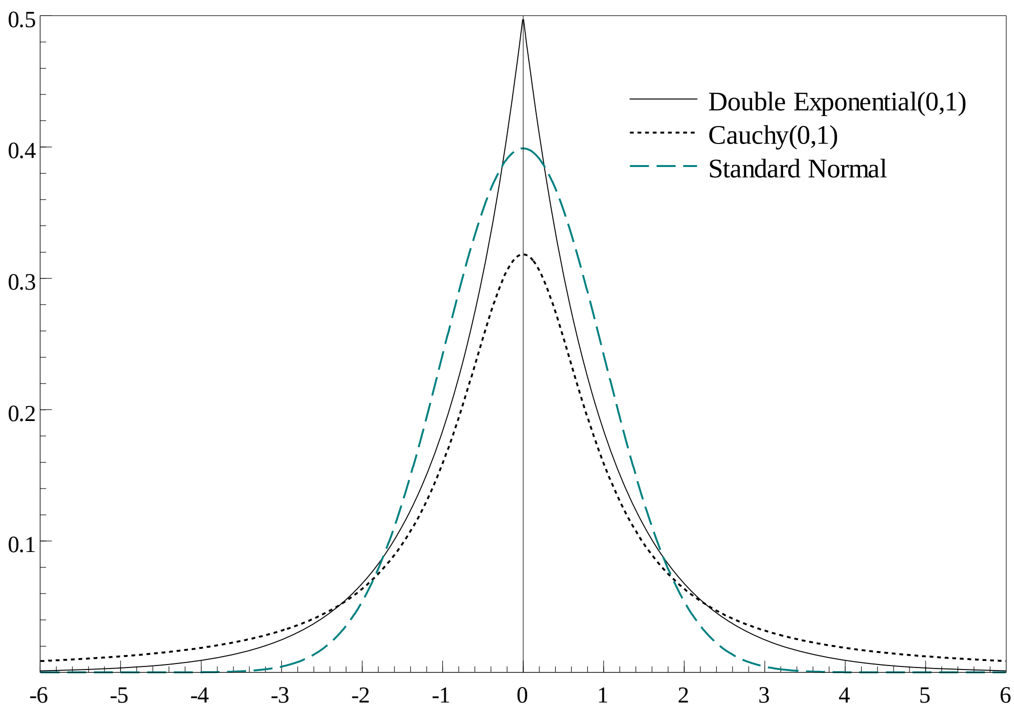

Figure 4.12: PDF's of Double Exponential$(0, 1)$, Cauchy$(0, 1)$ and $N(0, 1)$ Distributions

4.4 Conclusion

In this chapter we have introduced a variety of important parametric discrete and continuous probability distributions, and investigated their properties and their relationships. The flexibility of these parametric distributions depends on the functional form of the PDF, and the number of parameters. We have paid a closed attention to the interpretation of the parameters and their relationships. These distributions are widely used in modeling economic and financial data.

EXERCISE 4

4.1. Suppose $X_i \sim$ Bernoulli(p), and $X_1, \cdots, X_n$ are jointly independent. Define $X = \Sigma_{i=1}^n X_i$. Show $X \sim$ Binomial(n, p).

4.2. [C&B, # 2.11, P.77] Let X have the standard normal pdf $f_X(x) = \left(1/\sqrt{2\pi}\right) e^{-x^2/2}$ for $-\infty < x < \infty$.

(1) Find $E(X^2)$ directly, and then by using the pdf of $Y = X^2$ and calculating $E(Y)$.

(2) Find the PDF of $Y = |X|$, and find its mean and variance. (This distribution is sometimes called a *folded normal*.)

4.3. [C&B, # 3.16, p.130] The Gamma function is defined as

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt.$$

Verify these two identities regarding the Gamma function:

(1) $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$;

(2) $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

4.4. [C&B, # 3.20 (b), p.131] Let the random variable X have the pdf $f(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2}$, $0 < x < \infty$. Find the transformation $g(X) = Y$ and values of α and β so that $Y \sim$Gamma(α, β). The PDF of a Gamma(α, β) is given by $f_X(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$, for $x > 0$.

4.5. [C&B, # 3.23, p.131] The *Parieto distribution*, with parameters α and β, has PDF $f(x) = \frac{\beta\alpha^\beta}{x^{\beta+1}}, \alpha < x < \infty, \beta > 0, \alpha > 0$.

(1) Verify that $f(x)$ is a PDF.

(2) Derive the mean and variance of this distribution.

(3) Prove that the variance does not exist if $\beta \leq 2$.

4.6. Suppose X follows a binomial distribution, $B(n, p)$, where $p \in (0, 1)$, with PMF

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \text{ for } x = 0, 1, \cdots, n.$$

Using the uniqueness theorem of the moment generating function, show that when $n \to \infty$ but $np \to \lambda \in (0, \infty)$, the binomial distribution can be approximated by a Poisson(λ) distribution. [Hint: Show the MGF of the binomial distribution converges to the MGF of the Poisson distribution.]

4.7. Find the mean and variance of a geometric distribution with parameter $p \in (0, 1)$.

4.8. Show that the geometric distribution is characterized by the Markovian property

$$p(X = x + y | X \geq y) = P(X = x) \text{ for all positive integers } x, y.$$

4.9. (1) A electronic product can withstand a number of external shocks. However, when the K_{th} shock arrives, the product fails. That is, the lifetime of the product is the arrival time of the K_{th} shock. Suppose during the time $(0, t)$, the number of shocks follow $Possion(\lambda t)$ distribution:

$$P(X = x) = \frac{(\lambda t)^x}{x!} e^{-x}, \quad x = 0, 1, \cdots.$$

Prove that the lifetime of the product follows a $Gamma(K, \lambda)$ distribution.
(2) Generally, if $X \sim Gamma(\alpha, \beta)$ and α is an integer, then

$$P(X \leq x) = P(Y \geq \alpha), \text{ for any } x$$

where $Y \sim Poisson(x/\beta)$.

4.10. Find the mean and variance of the geometric distribution with parameter p.

4.11. The series system is built in such a way that it operates only when all its components operate(so it fails when at least one component fails). Assuming that the lifetime of each components has Exp(1) distribution and that the components operate independently, find the distribution and survival function of the system's life time T.

4.12. Suppose X is exponentially distributed. Show

$$P(X > x + y | X > x) = P(X > y) \text{ for all } 0 < x, y < \infty.$$

4.13. [C&B, # 3.24, p.131] Many "named" distributions are special cases of the more common distributions already discussed. For each of the following named distributions derive the form of the PDF and verify that it is a PDF.
(1) If $X \sim$ exponential(β),then $Y = X^{1/\gamma}$ has the $Weibull(\gamma, \beta)$ distribution, where $\gamma > 0$ is a constant.
(2) If $X \sim$ exponential(β),then $Y = (2X/\beta)^{1/2}$ has the Rayleigh distribution.
(3) If $X \sim$ Gamma(a, b), then $Y = 1/X$ has the invert gamma $IG(a, b)$ distribution.
(4) If $X \sim$ Gamma($\frac{3}{2}, \beta$), then $Y = (X/\beta)^{1/2}$ has the Maxwell distribution.
(5) If $X \sim$ exponential(1), then $Y = \alpha - \gamma \log X$ has the $Gumbel(\alpha, \gamma)$ distribution, where $-\infty < \alpha < \infty$ and $\gamma > 0$.

4.14. [C&B, # 3.49, p.135] Prove the following analogs to Stein's Lemma, assuming appropriate conditions on the function $g(\cdot)$.
(1) If $X \sim Gamma(\alpha, \beta)$, then

$$E\left[g(X)(X - \alpha\beta)\right] = \beta E\left[X g\prime(X)\right].$$

(2) If $X \sim Beta(\alpha, \beta)$, then

$$E\left\{g(X)\left[\beta - (\alpha - 1)\frac{(1 - X)}{X}\right]\right\} = E\left[(1 - X)g\prime(X)\right].$$

4.15. Suppose X follows a double exponential distribution with PDF $f_X(x) = \frac{1}{2\sigma}e^{-|x-\mu|/\sigma}$. Show that when $\mu = 0$, the absolute value of X, $Y = |X|$, follows an exponential (σ) distribution.

4.16. Suppose $Y = e^{-X}$ follows an Exponential(β) distribution. Find the PDF of X. The distribution of X is called an extreme value distribution.

4.17. A random variable is said to have a twice-piece normal distribution with parameters μ, σ_1, σ_2 if it has its PDF

$$f_X(x) = \begin{cases} Ae^{-(x-\mu)^2/(2\sigma_1^2)}, & x \le \mu, \\ Ae^{-(x-\mu)/(2\sigma_2^2)}, & x > \mu. \end{cases}$$

(1) Find the value of constant A. (2) Find the mean of X. (3) Find the variance of X.

4.18. Suppose $X \sim N(\mu, \sigma^2)$. Find the PDF of $Y = 1/X$.

4.19. Suppose $X \sim \chi^2_\nu$. Then $Y = \sqrt{X}$ is called a Chi-distribution with n degrees of freedom and is denoted as χ_ν. Show (1) the PDF of Y is $f_Y(y) = \frac{1}{2^{(\nu/2)-1}\Gamma(\nu/2)} e^{-y^2/2} y^{\nu-1}$, $y > 0$; and (2) $E(Y^k) = \frac{2^{k/2}\Gamma[(\nu+k)/2]}{\Gamma(\nu/2)}$.