Core course for *App.AI*, *Bioinfo.*, *Data Sci.&Eng.*, *Dec.Analytics*, *Q.Fin.*, *Risk Mgmt.* and *Stat.* Majors:

STAT2601A Probability and Statistics I (2023-2024 First Semester)

Chapter 6: Other Properties of Distributions

6.1 Survival Function and Hazard Rate Function

Let X be a positive continuous random variable that we interpret as being the lifetime of some item, having distribution function F and probability density function f.

Definition 6.1.
The *survival function* of a random variable X is defined by

$$S(t) = \Pr(X > t) = 1 - F(t).$$

It represents the probability that the item can survive at least for a time t.
The *hazard rate function* (*mortality function*, *failure rate function*) is defined by

$$\lambda(t) = \frac{f(t)}{S(t)}.$$

◀

Remarks

1. The hazard rate function $\lambda(t)$ represents the conditional probability intensity that a t-unit-old item will fail/die instantly at time t.

$$\Pr(t < X < t + \mathrm{d}t | X > t) = \frac{F(t + \mathrm{d}t) - F(t)}{S(t)} \approx \frac{f(t)\mathrm{d}t}{S(t)} = \lambda(t)\mathrm{d}t.$$

2. The hazard rate function $\lambda(t)$ uniquely determines the distribution F.

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = -\frac{\mathrm{d}}{\mathrm{d}t}[\ln S(t)]$$

$$\Longrightarrow \quad \int_0^x \lambda(t)\mathrm{d}t = [-\ln S(t)]_0^x = -\ln S(x) = -\ln[1 - F(x)]$$

$$\Longrightarrow \quad F(x) = 1 - \exp\left(-\int_0^x \lambda(t)\mathrm{d}t\right) = 1 - e^{-\int_0^x \lambda(t)\mathrm{d}t}.$$

3. The mathematical relationships among four functions (distribution function, survival function, density function, hazard rate function) are summarized as follows:

Functions in terms of	$F(t)$	$S(t)$	$f(t)$	$\lambda(t)$
$F(t) =$	$F(t)$	$1 - S(t)$	$\int_0^t f(s)\mathrm{d}s$	$1 - e^{-\int_0^t \lambda(s)\mathrm{d}s}$
$S(t) =$	$1 - F(t)$	$S(t)$	$\int_t^\infty f(s)\mathrm{d}s$	$e^{-\int_0^t \lambda(s)\mathrm{d}s}$
$f(t) =$	$F'(t)$	$-S'(t)$	$f(t)$	$\lambda(t)e^{-\int_0^t \lambda(s)\mathrm{d}s}$
$\lambda(t) =$	$\frac{F'(t)}{1 - F(t)}$	$-\frac{\mathrm{d}}{\mathrm{d}t}\ln S(t)$	$\frac{f(t)}{\int_t^\infty f(s)\mathrm{d}s}$	$\lambda(t)$

Example 6.1.

If $X \sim \text{Exp}(\lambda)$, then

$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{\lambda e^{-\lambda t}}{1 - (1 - e^{-\lambda t})} = \lambda.$$

Hence the exponential random variable has a constant hazard rate, i.e., old subject will be as likely to "die" as young subject, without regarding to their ages. Due to such memoryless property, the exponential distribution is generally not a reasonable model for the survival time of subjects with natural aging property.

★

Example 6.2.

Usually it would be more reasonable to model the lifetime of an item by an increasing hazard rate function rather than a constant hazard rate. ("Older" item will have higher chance to fail/die.) For example, we can use a linear hazard rate function $\lambda(t) = a + bt$. Then the distribution function is given by

$$\begin{aligned}
F(x) &= 1 - \exp\left(-\int_0^x (a + bt)\mathrm{d}t\right) = 1 - \exp\left(-\left[at + \frac{bt^2}{2}\right]_0^x\right) \\
&= 1 - \exp\left(-ax - \frac{bx^2}{2}\right), \qquad \text{for } x > 0. \\
f(x) &= F'(x) = (a + bx)\exp\left(-ax - \frac{bx^2}{2}\right), \qquad \text{for } x > 0.
\end{aligned}$$

In particular, if $a = 0$, the corresponding random variable is said to have the *Rayleigh distribution*.

★

6.2 Indicator Function

Definition 6.2.
An *indicator function* of a subset A of a set Ω is a function $\mathbf{1}_A : \Omega \to \{0,1\}$ defined as

$$\mathbf{1}_A(x) = \begin{cases} 1, & \text{if } x \in A; \\ 0, & \text{if } x \notin A. \end{cases}$$

◀

Properties of an indicator function

1. $\mathbf{1}_{A\cap B} = \min\{\mathbf{1}_A, \mathbf{1}_B\} = \mathbf{1}_A \times \mathbf{1}_B$.
2. $\mathbf{1}_{A\cup B} = \max\{\mathbf{1}_A, \mathbf{1}_B\} = \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_A \times \mathbf{1}_B$.
3. $\mathbf{1}_{A^c} = 1 - \mathbf{1}_A$.
4. $\mathrm{E}(\mathbf{1}_A) = \Pr(A)$, i.e., $\int_{-\infty}^{\infty} \mathbf{1}_A f(x)\mathrm{d}x = \int_A f(x)\mathrm{d}x$.

Remarks

1. $\mathbf{1}_A$ essentially follows a Bernoulli distribution with success probability $p = \Pr(A)$, that is, $\mathbf{1}_A \sim \mathrm{Ber}(p)$.
2. With the indicator function, some density functions can be written in a more compact way. For example,
$$f(x) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, & \text{for } 0 < x < 1; \\ 0, & \text{otherwise,} \end{cases}$$
can simply be written as
$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}\mathbf{1}_{\{x\in(0,1)\}}.$$
For another example,
$$f(x) = \begin{cases} 3x^{-4}, & \text{for } x > 1; \\ 0, & \text{otherwise,} \end{cases}$$
can simply be written as
$$f(x) = 3x^{-4}\mathbf{1}_{\{x>1\}}.$$
3. With the indicator function, a random variable can also be written in a compact way. For example,
$$X(\omega) = \begin{cases} 2016, & \text{if } \omega \in A; \\ 2601, & \text{if } \omega \in B; \\ 6210, & \text{if } \omega \in C, \end{cases}$$
can be rewritten as
$$X = 2016 \times \mathbf{1}_A + 2601 \times \mathbf{1}_B + 6210 \times \mathbf{1}_C, \text{ or } 2016^{\mathbf{1}_A} \times 2601^{\mathbf{1}_B} \times 6210^{\mathbf{1}_C}.$$

Indicator functions are also useful in solving some probability problems.

Example 6.3.

A monkey types at a 26-letter keyboard with one key corresponding to each of the upper-case English letters. Each keystroke is chosen independently and uniformly at random from the 26 possibilities. If the monkey types 1 million letters, what is the expected number of times the sequence "BOOK" appears?

Solution:

There are $1,000,000 - 4 + 1 = 999,997$ places where the string "BOOK" can appear, each with a (non-independent) probability of $\frac{1}{26^4}$ of happening. If X is the random variable that counts the number of times the string "BOOK" appears, and X_i is the indicator variable that is 1 if the string "BOOK" appears starting at the i-th letter, then

$$\begin{aligned} X &= X_1 + X_2 + \cdots + X_{999,997}, \\ \mathrm{E}(X) &= \mathrm{E}\left(X_1 + X_2 + \cdots + X_{999,997}\right) \\ &= \mathrm{E}\left(X_1\right) + \mathrm{E}\left(X_2\right) + \cdots + \mathrm{E}\left(X_{999,997}\right) \\ &= \frac{1}{26^4} + \frac{1}{26^4} + \cdots + \frac{1}{26^4} \\ &= \frac{999,997}{26^4} \\ &\approx 2.19. \end{aligned}$$

★

Example 6.4.

A building has n floors numbered $1, 2, \ldots, n$, plus a ground floor. At the ground floor, m people get on the elevator together, and each gets off at a uniformly random one of the n floors (independently of everybody else). What is the expected number of floors the elevator stops at (not counting the ground floor)?

Solution:

Let X_i be the indicator that the elevator stopped at floor i.

$$\Pr(X_i = 1) = 1 - \Pr(\text{no one gets off at floor } i) = 1 - \left(\frac{n-1}{n}\right)^m.$$

Let X be the number of floors the elevator stops at, then the required expected number is

$$\begin{aligned} \mathrm{E}(X) &= \mathrm{E}(X_1 + X_2 + \cdots + X_n) \\ &= \mathrm{E}(X_1) + \mathrm{E}(X_2) + \cdots + \mathrm{E}(X_n) \\ &= n\left[1 - \left(\frac{n-1}{n}\right)^m\right]. \end{aligned}$$

★

Example 6.5.

A coin with a probability p to get a head is flipped n times. A "run" is a maximal sequence of consecutive flips that are all the same. (Thus, for example, the sequence HTHHHTTH with $n = 8$ has five runs.) Show that the expected number of runs is $1 + 2(n-1)p(1-p)$.

Solution:

Let X_i be the indicator for the event that a run starts at the i-th toss. Let $X = X_1 + X_2 + \cdots + X_n$ be the random variable for the number of runs total. Obviously, $\mathrm{E}(X_1) = 1$. For $i > 1$,

$$\begin{aligned}
\mathrm{E}(X_i) &= \Pr(X_i = 1) \\
&= \Pr\{i\text{-th toss is H}|(i-1)\text{-th toss is T}\} \times \Pr\{(i-1)\text{-th toss is T}\} \\
&\quad\quad + \Pr\{i\text{-th toss is T}|(i-1)\text{-th toss is H}\} \times \Pr\{(i-1)\text{-th toss is H}\} \\
&= p(1-p) + (1-p)p \\
&= 2p(1-p).
\end{aligned}$$

This gives

$$\begin{aligned}
\mathrm{E}(X) &= \mathrm{E}(X_1 + X_2 + \cdots + X_n) \\
&= \mathrm{E}(X_1) + \mathrm{E}(X_2) + \cdots + \mathrm{E}(X_n) \\
&= \mathrm{E}(X_1) + [\mathrm{E}(X_2) + \cdots + \mathrm{E}(X_n)] \\
&= 1 + (n-1) \times 2p(1-p) \\
&= 1 + 2(n-1)p(1-p).
\end{aligned}$$

★

6.3 Some Inequalities

Theorem 6.1. (Markov's Inequality)
If X is a non-negative random variable ($X(\omega) \geq 0$ for all $\omega \in \Omega$) with finite mean, then for any constant $c > 0$,

$$\Pr(X \geq c) \leq \frac{\mathrm{E}(X)}{c}.$$

▲

Proof. Consider a discrete random variable X (if X is continuous, change the following summations to integrations and replace the pmf $p(x)$ by the pdf $f(x)$).

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{x \in X(\Omega)} x p(x) \\
&= \sum_{x \geq c} x p(x) + \sum_{x < c} x p(x) \\
&\geq \sum_{x \geq c} x p(x) \geq c \sum_{x \geq c} p(x) = c \Pr(X \geq c).
\end{aligned}$$

□

Another proof making use of the *indicator function* is as follows.

Proof. Define the indicator function for the event $\{X \geq c\}$ as

$$\mathbf{1}_{\{X\geq c\}}(c) = \begin{cases} 1, & \text{if } X \geq c; \\ 0, & \text{otherwise.} \end{cases}$$

Since $X \geq 0$, $\mathbf{1}_{\{X\geq c\}}(c) \leq \dfrac{X}{c}$. Therefore,

$$\Pr(X \geq c) = \mathrm{E}\left[\mathbf{1}_{\{X\geq c\}}(c)\right] \leq \mathrm{E}\left(\frac{X}{c}\right) = \frac{\mathrm{E}(X)}{c}.$$

□

Theorem 6.2. (Chebyshev's Inequality)
If a random variable X has a finite mean μ and a finite variance σ^2, then for any constant $k > 0$,

$$\Pr(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

For example, the probability that X deviates from the mean for more than 2 standard deviations is at most 0.25.

▲

Proof. By Markov's inequality,

$$\Pr(|X-\mu| \geq k\sigma) = \Pr\left(\left|\frac{X-\mu}{\sigma}\right|^2 \geq k^2\right) \leq \frac{\mathrm{E}\left[\left(\frac{X-\mu}{\sigma}\right)^2\right]}{k^2} = \frac{1}{k^2}.$$

□

Example 6.6.

Suppose that in a large class the mean and standard deviation of the midterm test scores are $\mu = 65$ and $\sigma = 6.8$ respectively. How many students (in proportion) are there having a score between 50 and 80?

Solution:

The exact proportion cannot be determined because the distribution of the scores is unknown. However, using the Chebyshev's inequality, an lower bound can be given based on the information on the mean and standard deviation:

$$\begin{aligned} \Pr(50 < X < 80) &= \Pr(-15 < X - 65 < 15) \\ &= \Pr(|X-\mu| < 15) \\ &= \Pr\left(|X-\mu| < \frac{15}{6.8}\sigma\right) \\ &\geq 1 - \frac{1}{(15/6.8)^2} = 0.7945. \end{aligned}$$

Hence, there must be at least 79.45% students having a midterm score between 50 and 80.

★

6.4 Distribution of a Mixed Type

There exist distributions which are neither discrete nor (*absolute*) continuous. We call these distributions the *mixed distributions*.

Example 6.7.

Suppose T is the lifetime of a device (in 1000 hour units) distributed as exponential with rate $\lambda = 1$. In a test of the device, we cannot wait forever, so we might terminate the test after 2000 hours and the *truncated* lifetime X is recorded, i.e.,

$$X = \begin{cases} T, & \text{if } T < 2; \\ 2, & \text{if } T \geq 2. \end{cases}$$

Therefore, we have

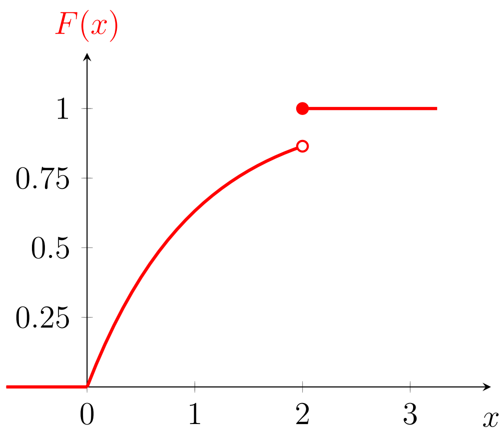
$$\Pr(X = 2) = \Pr(T \geq 2) = 1 - (1 - e^{-2}) = e^{-2},$$

and for $0 < x < 2$,

$$\Pr(X \leq x) = \Pr(T \leq x) = 1 - e^{-x}.$$

The distribution function of X is therefore given by

$$F(x) = \begin{cases} 0, & \text{for } x \leq 0; \\ 1 - e^{-x}, & \text{for } 0 < x < 2; \\ 1, & \text{for } x \geq 2. \end{cases}$$



It is discrete at $x = 2$ and continuous elsewhere. The expected lifetime recorded can be computed as

$$\mathrm{E}(X) = \int_0^2 x e^{-x} \mathrm{d}x + 2 \times e^{-2} = 1 - e^{-2} = 0.8647.$$

6.5 Quantiles

In describing the distribution of a group of data, it would be more comprehensive if some cut-off points that partition the dataset into consecutive pieces are reported. For example, the *lower quartile* is the data value such that 1/4 of the data points are lower than or equal to it; the *median* is the data value such that 1/2 of the data points are lower than or equal to it; etc. In general, these cut-off values are called *quantiles*.

6.5.1 Quantiles of Continuous Distributions

Definition 6.3.
When the distribution of a random variable X is continuous and one-to-one over the whole set of possible values of X, the *inverse distribution function* F^{-1} exists and the value $F^{-1}(p)$ is called the *p-quantile* of X for $0 < p < 1$. It is also named as the $100p$ *percentile* of X. ◀

If x_p is the $100p$-th percentile of X, then $100p$ percent of the distribution of X is at or below the value of x_p.

Example 6.8.

Suppose that the final examination scores in a large class of STAT2601 approximately follows the normal distribution with mean 69 and standard deviation 12, i.e., $X \sim \mathrm{N}(69, 144)$. Then the 85th percentile of the scores is $F^{-1}(0.85)$ and can be evaluated by solving

$$\begin{aligned} F(x_{0.85}) = 0.85 \qquad &\Longrightarrow \Pr(X \le x_{0.85}) = 0.85 \\ &\Longrightarrow \Phi\left(\frac{x_{0.85} - 69}{12}\right) = 0.85 \\ &\Longrightarrow \frac{x_{0.85} - 69}{12} = 1.04 \\ &\Longrightarrow x_{0.85} = 81.48. \end{aligned}$$

The 85th percentile of the scores is 81.48. Therefore a student with score higher than 81.5 is in the top 15% of the class.

Note that the calculation relies on the fact that the normal distribution function is a one-to-one function.

★

Remark

Certain quantiles have special names. The 1/2-quantile or the 50th percentile is a special case of what we shall call the *median*. The 1/4-quantile or the 25th percentile is the *lower quartile*. The 3/4-quantile or the 75th percentile is the *upper quartile*. These three values partition the distribution into four equal pieces.

Example 6.9.

If $X \sim \text{Exp}(\lambda)$, then

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{for } x > 0; \\ 0, & \text{for } x \leq 0. \end{cases}$$

For any $0 < p < 1$,

$$F(x_p) = p \qquad \Longrightarrow 1 - e^{-\lambda x_p} = p \qquad \Longrightarrow x_p = -\frac{1}{\lambda}\ln(1-p).$$

Since $F(x_p) = p$ has a unique solution, F^{-1} exists and is given by

$$F^{-1}(p) = -\frac{1}{\lambda}\ln(1-p).$$

In particular, if the lifetime (in 1000 hours unit) of a device is distributed as exponential with rate $\lambda = 1$, then

$$\begin{aligned} x_{0.25} &= F^{-1}(0.25) = -\frac{1}{1}\ln(1-0.25) = \ln\frac{4}{3} = 0.288, \\ x_{0.5} &= F^{-1}(0.5) = -\frac{1}{1}\ln(1-0.5) = \ln 2 = 0.693, \\ x_{0.75} &= F^{-1}(0.75) = -\frac{1}{1}\ln(1-0.75) = \ln 4 = 1.386. \end{aligned}$$

That is, the lower quartile, median, and upper quartile of the lifetime of the devices are 288, 693, and 1386 hours respectively.

★

Example 6.10.

Let X be a binomial random variable with parameters $n = 5$ and $p = 0.3$. The following table shows the pmf and cdf of X.

x	0	1	2	3	4	5
$p(x)$	0.1681	0.3602	0.3087	0.1323	0.0284	0.0024
$F(x)$	0.1681	0.5282	0.8369	0.9692	0.9976	1.0000

The p-quantile cannot be evaluated by $F^{-1}(p)$ because F^{-1} does not exist, e.g., there is no x such that $F(x) = 0.5$.

To define quantile for distribution that is not continuous, we may make use of a generalization of the inverse for non-decreasing functions.

★

6.5.2 Quantiles of Discrete Distributions

Definition 6.4.
Let X be a random variable with distribution function F. For any $0 < p < 1$, the p-quantile ($100p$ *percentile*) of X is defined as the smallest x such than $F(x) \geq p$.

Example 6.11.

From the previous example, the values of $F(x)$ for $x = 1, 2, 3, 4, 5$ are all greater than 0.5. Therefore the median of a $\mathrm{B}(5, 0.3)$ random variable is $x_{0.5} = 1$.

Similarly, the 75th percentile is $x_{0.75} = \min\{x : \; F(x) \geq 0.75\} = 2$.

Remark

This definition of the quantile applies to any distribution, no matter whether it is discrete, continuous, or of the mixed type. In fact, for continuous distribution, we have

$$\min\{x : \; F(x) \geq p\} = \min\{x : \; x \geq F^{-1}(p)\} = F^{-1}(p).$$

6.6 Transformation of Random Variables

In general, functions of random variables are also random variables.

Example 6.12.

1. If $Z \sim \mathrm{N}(0, 1)$, then $Y = Z^2 \sim \chi^2(1)$ and $X = \sigma Z + \mu \sim \mathrm{N}(\mu, \sigma^2)$.
2. If $X \sim \Gamma(\alpha, \lambda)$, then $Y = kX \sim \Gamma\left(\alpha, \dfrac{\lambda}{k}\right)$ for $k > 0$.
3. If $X \sim \mathrm{U}(0, 1)$, then $Y = -\dfrac{1}{\lambda} \ln X \sim \mathrm{Exp}(\lambda)$.

Theorem 6.3.
Let X be a continuous random variable which is distributed on a space S with pdf $f_X(x)$. Let $Y = g(X)$ where g is a function such that g^{-1} exists. Then the pdf of Y can be obtained by

$$f_Y(y) = f_X\left(g^{-1}(y)\right) \left| \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y) \right|, \qquad \text{for } y \in g(S).$$

Proof. Let $F_X(x)$ be the distribution function of X. The distribution function of Y is

$$F_Y(y) = \Pr(Y \le y) = \Pr\left(g(X) \le y\right).$$

If g is a strictly increasing function, then

$$F_Y(y) = \Pr\left(X \le g^{-1}(y)\right) = F_X\left(g^{-1}(y)\right)$$

$$\begin{aligned} \Longrightarrow \qquad f_Y(y) &= \frac{\mathrm{d}}{\mathrm{d}y} F_X\left(g^{-1}(y)\right) \\ &= f_X\left(g^{-1}(y)\right) \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y). \end{aligned}$$

On the other hand, if g is a strictly decreasing function, then

$$F_Y(y) = \Pr\left(X \ge g^{-1}(y)\right) = 1 - F_X\left(g^{-1}(y)\right)$$

$$\begin{aligned} \Longrightarrow \qquad f_Y(y) &= -\frac{\mathrm{d}}{\mathrm{d}y} F_X\left(g^{-1}(y)\right) \\ &= -f_X\left(g^{-1}(y)\right) \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y). \end{aligned}$$

Since $\frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y)$ is positive when g is increasing, and is negative when g is decreasing, the pdf of Y can be expressed as

$$f_Y(y) = f_X\left(g^{-1}(y)\right) \left| \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y) \right|, \qquad \text{for } y \in g(S).$$

□

Example 6.13.

Let $X \sim \mathrm{U}(0,1)$ and $Y = g(X) = -\frac{1}{\lambda} \ln X$ for some $\lambda > 0$.

$$\begin{aligned} f_X(x) &= 1, \qquad \text{for } 0 < x < 1. \\ S &= (0,1), \qquad g(S) = (0, \infty). \end{aligned}$$

Since $g(x) = y$, i.e., $-\frac{1}{\lambda} \ln x = y \implies x = e^{-\lambda y}$, the inverse function $g^{-1}(y) = e^{-\lambda y}$ exists.

The pdf of Y is

$$\begin{aligned} f_Y(y) &= f_X(e^{-\lambda y}) \left| \frac{\mathrm{d}}{\mathrm{d}y} e^{-\lambda y} \right| \\ &= 1 \times \lambda e^{-\lambda y} \\ &= \lambda e^{-\lambda y}, \qquad \text{for } y > 0. \end{aligned}$$

Therefore,

$$Y = -\frac{1}{\lambda} \ln X \sim \mathrm{Exp}(\lambda).$$

★

Example 6.14.

Let $X \sim \mathrm{N}(0,1)$ and $Y = \Phi(X)$.

$$S = (-\infty, \infty), \qquad \Phi(S) = (0,1).$$

Since Φ is a one-to-one function, Φ^{-1} exists. The pdf of Y is given by

$$\begin{aligned} f_Y(y) &= f_X\left(\Phi^{-1}(y)\right)\left|\frac{\mathrm{d}}{\mathrm{d}y}\Phi^{-1}(y)\right| \\ &= \phi\left(\Phi^{-1}(y)\right)\frac{1}{\phi\left(\Phi^{-1}(y)\right)} \qquad \text{(Note that } \Phi'(x) = \phi(x).\text{)} \\ &= 1, \qquad \text{for } y \in (0,1). \end{aligned}$$

Hence,

$$Y \sim \mathrm{U}(0,1).$$

Example 6.15. (Log-normal Distribution)

Let $X \sim \mathrm{N}(\mu, \sigma^2)$ and $Y = g(X) = e^X$.

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

$$S = (-\infty, \infty), \qquad g(S) = (0, \infty).$$

Since $g(x) = y$, i.e., $e^x = y \implies x = \ln y$, the inverse function $g^{-1}(y) = \ln y$ exists.

The pdf of Y is given by

$$\begin{aligned} f_Y(y) &= f_X(\ln y)\left|\frac{\mathrm{d}}{\mathrm{d}y}\ln y\right| \\ &= \frac{1}{y\sqrt{2\pi\sigma^2}}e^{-\frac{(\ln y-\mu)^2}{2\sigma^2}}, \qquad \text{for } y > 0. \end{aligned}$$

It is called the *log-normal distribution* which is commonly used to model multiplicative product of random quantities such as the growth factor (i.e., $1 +$ return rate) on stock investment.

Example 6.16.

Let $X \sim \mathrm{N}(0,1)$ and $Y = g(X) = X^2$.

$$f_X(x) = \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \qquad \text{for } -\infty < x < \infty.$$

$$S = (-\infty, \infty), \qquad g(S) = (0, \infty).$$

Since $g(x) = y$, i.e., $x^2 = y \implies x = \pm\sqrt{y}$ does not have a unique solution, $g^{-1}(y)$ does not exist. We cannot use the formula.

We may start from the distribution function. Consider

$$\begin{aligned} F_Y(y) &= \Pr(Y \le y) \\ &= \Pr(X^2 \le y) \\ &= \Pr(-\sqrt{y} \le X \le \sqrt{y}) \\ &= \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) \\ &= 2\Phi(\sqrt{y}) - 1. \end{aligned}$$

Therefore, the pdf of Y is

$$\begin{aligned} f_Y(y) &= F'_Y(y) \\ &= 2\phi(\sqrt{y}) \frac{\mathrm{d}}{\mathrm{d}y}\sqrt{y} \\ &= 2\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\sqrt{y})^2} \frac{1}{2\sqrt{y}} \\ &= \frac{1}{2^{\frac{1}{2}}\sqrt{\pi}} y^{-\frac{1}{2}} e^{-\frac{1}{2}y} \\ &= \frac{1}{2^{\frac{1}{2}}\Gamma(\frac{1}{2})} y^{\frac{1}{2}-1} e^{-\frac{1}{2}y}, \qquad \text{for } y > 0. \end{aligned}$$

Hence,

$$Y = X^2 \sim \chi^2(1).$$

★

Example 6.17. (Weibull Distribution)

Let $X \sim \text{Exp}(\lambda)$ and $Y = g(X) = X^{1/\beta}$, $\beta > 0$.

$$f_X(x) = \lambda e^{-\lambda x}, \qquad \text{for } x > 0.$$
$$S = (0, \infty), \qquad g(S) = (0, \infty).$$

Obviously, $g^{-1}(y) = y^{\beta}$ exists. The pdf of Y is given by

$$\begin{aligned} f_Y(y) &= f_X(y^{\beta}) \left| \frac{\mathrm{d}}{\mathrm{d}y} y^{\beta} \right| \\ &= \lambda e^{-\lambda y^{\beta}} \beta y^{\beta-1} \\ &= \lambda \beta y^{\beta-1} \exp(-\lambda y^{\beta}), \qquad \text{for } y > 0. \end{aligned}$$

It is called the *Weibull distribution* which is often used in the field of life data analysis. The corresponding distribution function and hazard rate function are respectively

$$F(y) = 1 - \exp(-\lambda y^{\beta}), \qquad \text{for } y > 0,$$

and

$$\lambda(y) = \lambda \beta y^{\beta-1}, \qquad \text{for } y > 0.$$

★

Example 6.18. (Maxwell-Boltzmann Distribution)

Let $X \sim \Gamma\left(\frac{3}{2}, \lambda\right)$ and $Y = g(X) = \sqrt{X}$.

$$f_X(x) = \frac{\lambda^{\frac{3}{2}}}{\Gamma(\frac{3}{2})} x^{\frac{3}{2}-1} e^{-\lambda x} = \frac{2\lambda^{\frac{3}{2}}}{\sqrt{\pi}} x^{\frac{1}{2}} e^{-\lambda x}, \qquad \text{for } x > 0.$$
$$S = (0, \infty), \qquad g(S) = (0, \infty).$$

Obviously, $g^{-1}(y) = y^2$ exists. The pdf of Y is given by

$$\begin{aligned} f_Y(y) &= f_X(y^2) \left| \frac{\mathrm{d}}{\mathrm{d}y} y^2 \right| \\ &= \frac{2\lambda^{\frac{3}{2}}}{\sqrt{\pi}} (y^2)^{\frac{1}{2}} e^{-\lambda y^2} \times 2y \\ &= \frac{4\lambda^{\frac{3}{2}}}{\sqrt{\pi}} y^2 e^{-\lambda y^2}, \qquad \text{for } y > 0. \end{aligned}$$

It is called the *Maxwell-Boltzmann distribution* which is widely used in statistical physics to model the speed of molecules in a uniform gas at equilibrium.

★

∼ End of Chapter 6 ∼