Contents

Chapter 3

SINGLE RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

3.1 What is a Random Variable?

Why random variable?:

It is easier to describe and manipulate *outcomes* and *events* of the chance experiments using *numerical values* rather than in *words*...

$\Longrightarrow$ **purpose of a random variable**

$\Longrightarrow$ maps each point in S into a point on R^1. [1]

For example,

$$x = X(\zeta)$$

where X is the random variable, x is its specific value, and ζ denotes the outcome of a chance experiment.

[1] This type of transform or mapping is called a *function*.

Figure 3.1: A r.v. is a mapping of the sample space into a real line.

NOTE: Events are now described by the random variable, rather than in words, and corresponding probability of the event can be calculated via r.v..

Example: For a chance experiment of tossing a fair coin, define a r.v. [2] $X(\omega) \ni$:

$$X(\text{head}) = 1 \quad \& \quad X(\text{tail}) = 0$$

Then, using the equally likely assignment of probability, we have:

$$P(X = 1) = \frac{1}{2}, \quad P(X = 0) = \frac{1}{2}$$

Example 3.1

Consider an experiment of rolling a pair of dice, and find the probability of all possible values of the sum.

Solution:

Define a r.v. X as the sum of the two dice, then referring the Table3.1. and applying the equally likely assignment of probability, we have:

$P(X = 2) = \frac{1}{36}$, $P(X = 3) = \frac{2}{36} = \frac{1}{18}$, $P(X = 4) = \frac{3}{36} = \frac{1}{12}$,

$P(X = 5) = \frac{4}{36} = \frac{1}{9}$, $P(X = 6) = \frac{5}{36}$, $P(X = 7) = \frac{6}{36} = \frac{1}{6}$,

$P(X = 8) = \frac{5}{36}$, $P(X = 9) = \frac{4}{36} = \frac{1}{9}$, $P(X = 10) = \frac{3}{36} = \frac{1}{12}$,

$P(X = 11) = \frac{2}{36} = \frac{1}{18}$, $P(X = 12) = \frac{1}{36}$,

[2] Value assignment of a r.v. entirely depends on the convenience, i.e. values 0 and 1 would be more convenient to handle than values π and e.

FACT: Functions of r.v. are ALSO r.v.'s, for instance:

$$Y = e^X, \;\; W = \ln X, \;\; U = \cos X, \;\; V = X^2$$

Categorization of random variables:

1. *Discrete random variable* assumes only a countable number of values.

   (e.g.) rolling dice: example3.1 (only 11 values)

2. *Continuous random variable* can assume a continuum of values.

   (e.g.) weather vane: angle of the indicator (any value from 0 to 2π radian)

3. *Mixed random variable* is a combination of above two types.

Example 3.2 *Self study*

3.2 Probability Distribution Functions

Description of r.v. using probability:

In case of discrete r.v.:

$\Longrightarrow$ *probability mass distribution*

$\Longrightarrow$ tabulate (or plot) probability of its values

Example: example3.1(equation (3-2)), table3.1

Two types of general methods of description: [3]

(1) Cumulative (probability) distribution function: cdf

(2) Probability density function: pdf

3.2.1 Cumulative Distribution Function

Definition 3.1 cumulative distribution function*:*
The cdf of a random variable X is defined as follows:

$$F_X(x) = P(X \leq x)$$

[3]These apply to and work for all three types of r.v., i.e., comtinuous, discrete, and mixed random variables.

Example 3.3

Consider a chance experiment of rolling a pair of fair dice, and define a r.v. X as the sum of the numbers showing up. Find the cdf of X. (refer to example 3.1)

Solution:

Since the cdf is defined as $F_X(x) = P(X \leq x)$, for instance if $x = 3$, we have: [4]

$$\begin{aligned}
F_X(3) &= P[(X=2) \cup (X=3)] = P[(X=2)] + P[(X=3)] \\
&= P[\{1,1\}] + P[\{1,2\} \cup \{2,1\}] = P[\{1,1\}] + P[\{1,2\}] + P[\{2,1\}] \\
&= \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{1}{12}
\end{aligned}$$

$$\vdots$$

Continuing until we cover all possible values of X, we get:

$$F_X(x) = \begin{cases}
0, & x < 2 \\
\frac{1}{36}, & 2 \leq x < 3 \\
\frac{3}{36}, & 3 \leq x < 4 \\
\frac{6}{36}, & 4 \leq x < 5 \\
\frac{10}{36}, & 5 \leq x < 6 \\
\frac{15}{36}, & 6 \leq x < 7 \\
\frac{21}{36}, & 7 \leq x < 8 \\
\frac{26}{36}, & 8 \leq x < 9 \\
\frac{30}{36}, & 9 \leq x < 10 \\
\frac{33}{36}, & 10 \leq x < 11 \\
\frac{35}{36}, & 11 \leq x < 12 \\
\frac{36}{36}, & x \geq 12
\end{cases}$$

Figure 3.2: The cdf of X, which is the sum of two dice.

[4]Notice that: $F_X(-\infty) = P(\phi) = 0$, and $F_X(\infty) = P(S) = 1$.

General properties of cdf:

1. Limiting values:

$$\begin{aligned}\lim_{x\to-\infty} F_X(x) &= 0\\ \lim_{x\to\infty} F_X(x) &= 1\end{aligned}$$

2. The cdf is *right hand continuous*, i.e.:

$$F_X(x_0) = \lim_{x\to x_0^+} F_X(x)$$

3. The cdf $F_X(x)$ is monotine non-decreasing function of x.

4. The probability of X having values b/w x_1 and x_2 is given by:

$$P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)$$

Brief verification:

1. Notice that the inverse images of X at $x = -\infty$ and $x = \infty$ are respectively:

$$X^{-1}(-\infty) = \phi$$

$$X^{-1}(\infty) = S$$

2. This is due to the fact that the cdf is defined as $F_X(x) \triangleq P(X \leq x)$ rather than $F_X(x) \triangleq P(X < x)$: *detailed proof is omitted!* [5]

3. Follows from property # 4.

4. Let $x_1 \leq x_2$, then we have:

$$P(X \leq x_2) = P[(X \leq x_1)\cup(x_1 < X \leq x_2)] \overset{\text{why?}}{=} P(X \leq x_1)+P(x_1 < X \leq x_2)$$

which is equivalent to:

$$F_X(x_1) + P(x_1 < X \leq x_2) = F_X(x_2)$$

Rearranging the above provides:

$$P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1) \geq 0$$

[5] To prove this, we need the so called **continuity axiom**, which is beyond the scope of this class: *will be discussed at the graduate level course!*.

Example 3.4

Find the probability that the sum of two dice is between 3 and 7 inclusive.

Solution:

From the definition and properties of cdf, and using figure 3.1, we have: [6]

$$P(3 \leq X \leq 7) = F_X(7) - F_X(3^-) = F_X(7) - F_X(2) = \frac{21}{36} - \frac{1}{36} = \frac{20}{36} = \frac{5}{9}$$

Example 3.5

Is $F_X(x)$ below is a valid cdf?

$$F_X(x) = \frac{1}{2}(1 + \frac{2}{\pi}\tan^{-1} x)$$

Solution:

Check if all the properties of cdf are satisfied: *self study*

Figure 3.3: Suitable cdf.

[6] Note that there $\exists$ 20 outcomes favorable to the event, thus applying the equally likely assignment probability, we get $\frac{20}{36}$.

3.2.2 Probability Density Function

Definition 3.2 probability density function*:*
The pdf of a (continuous)random variable X is defined as follows:

$$f_X(x) = \frac{dF_X(x)}{dx}$$

NOTE: interpretation of pdf!!!

Using the definition of derivative, we can re-write pdf as:

$$f_X(x) = \lim_{\Delta x \to 0} \frac{F_X(x + \Delta x) - F_X(x)}{\Delta x}$$

For Δx suffuciently small, we can remove the limit, and thus:

$$F_X(x + \Delta x) - F_X(x) = P(x < X \leq x + \Delta x) \simeq f_X(x)\Delta x$$

General properties of pdf:

Since the pdf of a r.v. X is the derivative of the cdf, cdf $F_X(x)$ can be expressed as the integration of the pdf $f_X(x)$, i.e.:

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du$$

1. The pdf is a non-negative function, i.e.:

$$f_X(x) \geq 0, \quad \forall x$$

2. The area under pdf is unity, i.e.:

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

3. The probability of X having values b/w x_1 and x_2 is given by:

$$P(x_1 < X \leq x_2) = \int_{x_1}^{x_2} f_X(x)dx$$

Brief verification:

1. Since the cdf is non-decreasing, and pdf is the derivetive (or slope) of it, it must be non-negative.
2. Using the relationship b/w the cdf and pdf, we have:

$$\int_{-\infty}^{\infty} f_X(x)dx \stackrel{\text{def}}{=} F_X(\infty) = 1$$

3. From the property of the cdf, and using the relationship b/w the cdf and pdf, we have:

$$\begin{aligned} P(x_1 < X \leq x_2) &= F_X(x_2) - F_X(x_1) \\ &= \int_{-\infty}^{x_2} f_X(u)du - \int_{-\infty}^{x_1} f_X(u)du \\ &= \int_{x_1}^{x_2} f_X(u)du \end{aligned}$$

Example 3.6

Obtain the pdf of a r.v. X whose cdf is given as: (example 3.5)

$$F_X(x) = \frac{1}{2}(1 + \frac{2}{\pi}\tan^{-1} x)$$

and find the probability of an event $\ni: \;\; 2 < X \leq 5$.

Solution:

Since $\frac{d}{dx}\tan^{-1} x = \frac{1}{1+x^2}$, we easily get:

$$f_X(x) = \frac{1/\pi}{1+x^2}$$

and using the property of the cdf, we obtain:

$$\begin{aligned} P(2 < X \leq 5) &= \frac{1}{\pi}(\tan^{-1} 5 - \tan^{-1} 2) \\ &= 0.085 \end{aligned}$$

Figure 3.4: The pdf of X in example 3.6.

Remark:
The definition of the pdf can *generally* be applied to both continuous and discrete random variables!!!

motive:

Since the cdf of a discrete r.v. can generally be expressed as the sum of the weighted & shifted unit step function $u(x)$ [7], i.e.:

$$F_X(x) = \sum_{i=1}^{N} p_i u(x - x_i)$$

and the derivative of $u(x)$ is the unit impulse function $\delta(x)$ as:

$$\frac{du(x)}{dx} = \delta(x)$$

the pdf $f_X(x)$ of a discrete r.v. can generally be described as the sum of the weighted & shifted unit impulse function $\delta(x)$, i.e.:

$$f_X(x) = \sum_{i=1}^{N} p_i \delta(x - x_i)$$

[7] Recall that the unit step function is defined as $u(x) \triangleq 1$ for $x \geq 0$ and 0 elsewhere.

Example 3.7

Express the pdf of the r.v. X discussed in example 3.3.

Solution:

The cdf of X in terms of $u(x)$ can be expressed as:

$$\begin{aligned} F_X(x) = & \tfrac{1}{36} \ [u(x-2) + 2u(x-3) + 3u(x-4) + 4u(x-5) + 5u(x-6) \\ + & \ 6u(x-7) + 5u(x-8) + 4u(x-9) + 3u(x-10) \\ + & \ 2u(x-11) + u(x-12)] \end{aligned}$$

Taking the derivative, we find the pdf to be:

$$\begin{aligned} f_X(x) = & \tfrac{1}{36} \ [\delta(x-2) + 2\delta(x-3) + 3\delta(x-4) + 4\delta(x-5) + 5\delta(x-6) \\ + & \ 6\delta(x-7) + 5\delta(x-8) + 4\delta(x-9) + 3\delta(x-10) \\ + & \ 2\delta(x-11) + \delta(x-12)] \end{aligned}$$

Figure 3.5: The pdf of X in example 3.7

3.3 Common Random Variables and their Distribution Functions

Commonly occurring and widely used r.v.'s are discussed

$\Longrightarrow$ In terms of their cdf & pdf

Uniform Random Variable:

The uniform random variable X is defined in terms of its probability density function as follows:

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

By integration, we can derive its cumulative distribution function(cdf) to be:

$$f_X(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a < x \leq b \\ 1, & x > b \end{cases}$$

Figure 3.6: Uniform r.v.: (a) pdf (b) cdf in case of $a = 0$ & $b = 5$

Example 3.8

Resistors are known to be uniformly distributed within $\pm 10\%$ tolerance range. Find the probability that a nominal 1000 Ω resistor has a value b/w 990 and 1010 Ω.

Solution:

Since the tolerance region is bounded in the interval [900, 1100] centered at the nominal value of 1000Ω, the pdf of the resistance R is given by:

$$f_R(r) = \begin{cases} \frac{1}{200}, & 900 \leq x \leq 1100 \\ 0, & \text{otherwise} \end{cases}$$

Therefore, the probability that the resistor has a resistance value in the range of [900, 1100] is then:

$$P(990\Omega < R \leq 1010\Omega) = \int_{990}^{1010} \frac{dr}{200} = 0.1$$

Gaussian Random Variable:

The Gaussian random variable X is defined in terms of its probability density function as follows:

$$f_X(x) = \frac{e^{-(x-m)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}$$

where m and σ^2 are parameters called the *mean* and *variance* respectively.

The cdf can be derived by direct integration of the pdf, which cannot be expressed in a closed form, however...

Instead, we use either of the following functions defined as:

(1) Q function:

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-u^2/2} du$$

whose numerical values are tabulated in Appendix C.

(2) Error function:

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2} du$$

In terms of the Q function, the cdf of the Gaussian r.v. is given by:

$$F_X(x) = 1 - Q(\frac{x-m}{\sigma})$$

derivation: assignment

Question: Express Q and error functions in terms of each other: *assignment*

Figure: Illustration of integration for Q and error functions.

Figure 3.7: Gaussian distribution for $m = 5$ and $\sigma = 2$: (a) pdf (b) cdf.

Example 3.9

A mechanical compopnent, whose 5mm thickness(T)is known to have a Gaussian distribution w/ $m = 5$ and $\sigma = 0.05$. Find the probability that T is less than 4.9mm OR greater than 5.1mm.

Solution: [8]

$$
\begin{aligned}
& P(T < 4.9mm \text{ OR } T > 5.1mm) \\
= \; & 1 - P(4.9mm \leq T \leq 5.1mm) \\
= \; & 1 - [F_T(5.1) - F_T(4.9)] \\
= \; & 1 - [1 - Q(\frac{5.1 - 5}{0.05}) - 1 + Q(\frac{4.9 - 5}{0.05})] \\
= \; & 1 - [Q(-2) - Q(2)] = 1 - [1 - Q(|-2|) - Q(2)] \\
= \; & 2Q(2) = 0.02275
\end{aligned}
$$

[8]We use here the relationship $Q(x) = 1 - Q(|x|)$ for $x < 0$.

Exponential Random Variable:

The exponential random variable X is defined in terms of its probability density function with parameter α as follows:

$$f_X(x) = \alpha e^{-\alpha x} u(x), \quad \alpha > 0$$

The cdf, by integration, can then be obtained as:

$$F_X(x) = (1 - e^{-\alpha x}) u(x)$$

where $u(x)$ is the unit step function.

Assignment: Plot $f_X(x)$ and $F_X(x)$ for a exponential r.v.,and verify its properties.

Example 3.10 *Self Study*

Gamma Random Variable:

The Gamma random variable X has its probability density function as follows:

$$f_X(x) = \frac{c^b}{\Gamma(b)} x^{b-1} e^{-cx} u(x), \quad b, c > 0$$

where $\Gamma(b)$ is the gamma function given by the integral [9] :

$$\Gamma(b) = \int_0^\infty y^{b-1} e^{-y} dy$$

[9] It can be shown by evaluation that $\Gamma(1) = 1$ and $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, and by replacing b by $b+1$, we can also show that $\Gamma(b+1) = b\Gamma(b)$, which in turn provides that $\Gamma(n+1) = n!$ in case of b an integer n.

Special cases:

1. Chi-square pdf: *(statistics)*

$$f_X(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{(n/2)-1} e^{-x/2} u(x)$$

   where b and c are replaced by $b = n/2$ and $c = 2$

2. Erlang pdf: *(queing theory)*

$$f_X(x) = \frac{c^n}{(n-1)!} x^{n-1} e^{-cx} u(x)$$

   where $b = n$ is an integer.

Figure 3.8: Plots of (a) chi-square pdf and (b) Erlang pdf.

Cauchy Random Variable:

The Cauchy random variable X has its probability density function as follows:

$$f_X(x) = \frac{\alpha/\pi}{x^2 + \alpha^2}$$

Corresponding cdf is given by:

$$F_X(x) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1} \frac{x}{\alpha}$$

Note: Example 3.5 with fig.3.3 is the case of Cauchy r.v. for $\alpha = 1$!

Binomial Random Variable:

The r.v. X is binomially distributed if it takes on the non-negative integer values $0, 1, 2, \ldots, n$ with probabilities:

$$P(X = k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, 2, \cdots, n$$

where $p + q = 1$.

Corresponding pdf and the cdf, using the unit impulse function, are given by:

$$f_X(x) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \delta(x - k)$$

$$F_X(x) = \sum_{k \le x} \binom{n}{k} p^k q^{n-k}$$

Remark: This is the distribution describing the number(k) of heads occurring in n tosses of a fair coin, in which case $p = q = 0.5$.

Figure 3.9: The pdf of binomial r.v. for $n = 10$: (a) $p = q = \frac{1}{2}$; (b) $p = \frac{1}{5}$, $q = \frac{4}{5}$.

Geometric Random Variable:

Suppose we flip a biased coin w/ probability of a head is p whereas the probability of tail is $1 - p$. Then, the probability of getting the head(*success*) for the first time at the $k - th$ toss is given by:

$$P(\text{first success at trial } k) = P(X = k) = (1 - p)^{k-1} p, \quad k = 1, 2, \cdots$$

where p is the probability of *success*.

$\Longrightarrow$ called the geometric distribution

$\Longrightarrow$ corresponds to the geometric mean of $k - 1$ and $k + 1$ values.

Check: assignment

Poisson Random Variable:

The r.v. X is Poisson with parameter a if it takes on the non-negative integer values $0, 1, 2, \ldots$ with probabilities:

$$P(X = k) = \frac{a^k}{k!}e^{-a}, \quad k = 0, 1, 2, \cdots$$

Corresponding cdf of a Poisson r.v. is given by:

$$F_X(x) = \sum_{k \leq x} \frac{a^k}{k!}e^{-a}$$

Figure 3.10: Poisson pdf (a) $a = 0.9$; (b) $a = 0.2$.

Limiting Forms of Bionomial and Poisson Distributions:

Theorem 3.1 *De Moivre-Laplace theorem:*

For n sufficiently large, the binomial distribution can be approximated by the samples of a Gaussian curve properly scaled and shifted as:

$$p^k q^{n-k} \simeq \frac{e^{-(k-m)^2}}{\sqrt{2\pi\sigma^2}}, \quad n \gg 1$$

where $m = np$ and $\sigma^2 = npq$ respectively.

Proof: *omit*

Corresponding cdf of a binomial r.v., based on the *De Moivre-Laplace theroem*, can be approximated as:

$$F_X(x) \simeq 1 - Q(\frac{x - np}{\sqrt{npq}})$$

where $Q(\cdot)$ is the Q-function.

Figure 3.11: Binomial cdf and *De Moivre-Laplace* approximation for $n = 10$: (a) $p = 0.2$; (b) $a = 0.5$.

Theorem 3.2 *:*

The Poisson distribution approaches to a Gaussian distribution for $a \gg 1$ as:

$$\frac{a^k}{k!}e^{-a} \simeq \frac{e^{-(k-a)^2/2a}}{\sqrt{2\pi a}}, \quad n \gg 1,\ p \ll 1,\ np = a$$

Proof:
This is due to the fact that a binomial distribution approaches the Poisson distribution with $a = np$ if $n \gg 1$ and $p \ll 1$, and applying the *De Moivre-Laplace theorem* proves it!

:READ

Example 3.11

In a digital communication system, the probability of an error is 10^{-3}. What is the probability of 3 errors in transmission of 5000 bits?

Solution:

This has a binomial distribution, and the Poisson approximation with $n = 5000$, $p = 10^{-3}$, and $k = 3$ shows:

$$P(3\text{ errors}) \simeq \frac{5^3}{3!}e^{-5} = 0.14037$$

whereas the exact value of the probability using the binomial distribution is:

$$\binom{5000}{3}(10^{-3})^3(1-10^{-3})^{4997} = 0.14036$$

Poisson Points and Exponential Probability Density Function:

Consider a finite interval T (fig 3.12), with the probability of k occurrence of an event (ni : arrival of electrons at certain point) in this interval obeys a Poisson distribution, i.e.:

$$P(X = k) = \frac{(\lambda T)^k}{k!}e^{-\lambda T}, \quad k = 0, 1, 2, \cdots$$

where λ is the numberof events per unit time.

Then, the interval from an arbitrarily selected point and the next event is a r.v. W following an exponential distribution with its pdf as:

$$f_W(w) = \lambda e^{-\lambda w}u(w)$$

Figure 3.12: Poisson points on a finite interval T.

proof:

First, find the cdf of W as [10] :

$$F_W(w) = P(W \leq w) = 1 - P(X = 0) = 1 - e^{-\lambda w}, \ \ w \geq 0$$

Differentiating it, we obtain:

$$f_W(w) = \frac{d}{dw} F_W(w) = \lambda e^{-\lambda w} u(w)$$

Example 3.12

Raindrops impinge on a tin roof at a rate of 100/sec. What is the probability that the interval between adjacent raindrops is greater than 1(msec), and 10(msec)?

Solution:

This can be approximated by a Poisson poiny process with $\lambda = 100$, and therefore:

$$\begin{aligned} P(W \geq 10^{-3}) &= 1 - P(W \leq 10^{-3}) \\ &= 1 - (1 - e^{-100 \times 0.001}) = e^{-0.1} = 0.905 \end{aligned}$$

and

$$\begin{aligned} P(W \geq 10^{-2}) &= 1 - P(W \leq 10^{-2}) \\ &= 1 - (1 - e^{-100 \times 0.01}) = e^{-1} = 0.368 \end{aligned}$$

Pascal Random Variable:

The r.v. X has a Pascal distribution if it takes on the positive integer values $1, 2, 3, \ldots$ with probabilities:

$$P(X = k) = \begin{pmatrix} n-1 \\ k-1 \end{pmatrix} p^k q^{n-k}, \quad k = 1, 2, \cdots, \quad q = 1 - p$$

[10] $\{W \leq w\}$ corresponds to the event that there is *at least* one Poisson event in the interval $(0, w)$.

Note:

1. For example, in a sequence of coin tossing, the probability of getting the $k-th$ head on the $n-th$ toss obeys a Pascal distribution, where $p =$ prob. of a head.
2. **Reasoning:** We get the first $k-1$ heads in any order in the first $n-1$ tosses, which is binomial, and then must get a head on the $n-th$ toss.
3. Geometric distribution is a special case of the Pascal distribution.

Example 3.13 *Self study*

Hypergeometric Random Variable:

Consider a box containg N items, K of which are defective. Then, the probability of obtaining k defective items in a selection of n items *without replacement* follows the hypergeometric distribution:

$$P(X=k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, 2, \cdots, n$$

Example 3.14 *Self study*

3.4 Transformation of a Single Random Variable

Consider functions of random variables, i.e.:

$$Y = g(X)$$

where $g(\cdot)$ is a (*single valued*) function.

(e.g.)

$$Y = e^X, \;\; W = \ln X, \;\; U = \cos X, \;\; V = X^2$$

Recall: If X is a random variable, Y is also a random variable!!!

Question:
Given the probability distributions ($F_X(x)$ or $f_X(x)$) of X, find corresponding probability distributions ($F_Y(y)$ or $f_Y(y)$) of the newly defined r.v. Y...

Figure 3.13: Examples of monotonic and nonmonotonic transformations of a r.v..

1. Case #1: $g(\cdot)$ is monotone increasing

For an arbitrary value y_0 of Y, there $\exists$ a unique corresponding value x_0 of $X \ni$:

$$y_0 = g(x_0)$$

Then, we have:

$$P(Y \le y_0) \triangleq F_Y(y_0) = P[g(X) \le g(x_0)] = P(X \le x_0) = F_X(x_0)$$

Changing x_0 and y_0 to arbitrary values x and y, we get

$$F_Y(y) = F_X(x) \qquad \text{where} \quad x = g^{-1}(y)$$

Differentiating the above $F_Y(y)$ w.r.t. y, we get the pdf $f_Y(y)$ as:

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{dF_X(x)}{dx} \left.\frac{dx}{dy}\right|_{x=g^{-1}(y)} = f_X(x) \left.\frac{dx}{dy}\right|_{x=g^{-1}(y)}$$

2. Case #2: $g(\cdot)$ is monotone decreasing:

For a specific value y_0 of Y, there $\exists$ also a unique corresponding value x_0 of $X \ni$:

$$y_0 = g(x_0)$$

But, in this case we have:

$$F_Y(y_0) = P[g(X) \le g(x_0)] = P(X \ge x_0) = 1 - P(X \le x_0) = 1 - F_X(x_0)$$

In general, this can be expressed as:

$$F_Y(y) = 1 - F_X(x) \qquad \text{where} \quad x = g^{-1}(y)$$

Differentiating the above $F_Y(y)$ w.r.t. y, we get the pdf $f_Y(y)$ as:

$$f_Y(y) = \frac{dF_Y(y)}{dy} = -\frac{dF_X(x)}{dx} \left.\frac{dx}{dy}\right|_{x=g^{-1}(y)} = -f_X(x) \left.\frac{dx}{dy}\right|_{x=g^{-1}(y)}$$

Remark:

Notice that the slope(or derivative) $\frac{dx}{dy} > 0$ for case #1, whereas $\frac{dx}{dy} < 0$ for case #2. Therefore, using the absolute value of the derivative, we can combine the above two cases, and express the probability density function $f_Y(y)$ as follows:

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{dF_X(x)}{dx}\left|\frac{dx}{dy}\right|_{x=g^{-1}(y)} = f_X(x)\left|\frac{dx}{dy}\right|_{x=g^{-1}(y)}$$

Example 3.15

Suppose X is an exponential r.v. w/ parameter α, i.e.:

$$f_X(x) = \alpha e^{-\alpha x}u(x), \quad \alpha > 0$$

Then find the pdf of a newly defined r.v. Y via the following transformation.

$$Y = aX + b$$

Solution:

The transformation is monotone, and solving the transformation w.r.t. x, we get:

$$x = \frac{1}{a}(y - b)$$

and the derivative has a constant value as:

$$\frac{dx}{dy} = \frac{1}{a}$$

Therefore, the pdf of Y can be derived as:

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|_{x=g^{-1}(y)} = f_Y(y) = \frac{\alpha}{|a|}e^{(\alpha/|a|)(y-b)}u(y-b)$$

Figure 3.14: The pdf for example 3.15: (a) $a > 0$; (b) $a < 0$.

3. Case #3: $g(\cdot)$ is non-monotonic:

In this case, there will $\exists$ more than one solution of $X = x$ for a given value of $Y = y$, i.e.:

$$x_i = g_i^{-1}(y), \quad i = 1, 2, \cdots, m$$

Therefore, we can generalize the formula of the newly derined r.v.'s pdf as:

$$f_Y(y) = \sum_{i=1}^{m} f_X(x) \left| \frac{dx_i}{dy} \right|_{x_i = g_i^{-1}(y)}$$

Example 3.16

Let X be a Gaussian r.v. w/ mean $m = 0$. Find the pdf of Y defined a follows:

$$Y = X^2$$

Solution:

Note that $Y > 0$, and therefore $f_Y(y) = 0$ for $y < 0$.
For the case when $y \geq 0$, solving the transformation w.r.t. x, we obtain:

$$x_1 = \sqrt{y} \quad \text{and} \quad x_2 = -\sqrt{y}$$

Thus

$$\left|\frac{dx_i}{dy}\right| = \frac{1}{2\sqrt{y}}, \quad i = 1, 2$$

Therefore, the pdf of Y becomes:

$$\begin{aligned} f_Y(y) &= \frac{1}{2\sqrt{y}}\left[\frac{e^{-x^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}\right]_{x_1=\sqrt{y}} + \frac{1}{2\sqrt{y}}\left[\frac{e^{-x^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}\right]_{x_1=-\sqrt{y}} \\ &= \frac{e^{-y/2\sigma^2}}{\sqrt{2\pi\sigma^2 y}}, \quad y \geq 0 \end{aligned}$$

Figure 3.15: The pdf of Y in example 3.16 for the case of $m = 0$ and $\sigma^2 = 1$.

Example 3.17 *Self study*

3.5 Averages of Random Variables

Expressing r.v.'s using its *representative values*!!!

: For the cases when complete description of the r.v. $\ni$: the pdf and/or cdf might not be necessary.....

Definition 3.3 Expectation of a r.v.:
The mathematical expectation of a random variable X, using the r.v.'s pdf, is defined according to the following equation:

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

where $E(\cdot)$ stands for expectation.

Note:

The above definition applies to both *continuous* and *discrete* random variables, i.e., if X is discrete, we have [11]:

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x \sum_{i=1}^{n} p_i \delta(x - x_i) dx \\ &= \sum_{i=1}^{n} p_i \int_{-\infty}^{\infty} x \delta(x - x_i) dx \\ &= \sum_{i=1}^{n} x_i p_i \end{aligned}$$

where $p_i \triangleq P(X = x_i)$, and this might be the more familiar form of the mathematical expectation for discrete r.v.'s for you.

[11] This is due to the *sifting property* of the unit impulse function, which is, $\int_{-\infty}^{\infty} g(x)\delta(x - x_i)dx = g(x_i)$.

Example 3.18

The test scores of 100 students are summarized in Table3.2. Find the average score using the mathematical expectation.

Score	# of students	Relative frequency
100	2	0.02
95	5	0.05
90	10	0.10
85	20	0.20
80	33	0.33
75	15	0.15
70	7	0.07
65	4	0.04
60	3	0.03
55	1	0.01

Table3.2 Test scores for 100 students.

Solution:

Using the relative frequency approach for probability, and by the definition of mathematical expectation, we have:

$$\begin{aligned} E(X) &= 100 \times 0.02 + 95 \times 0.05 + 90 \times 0.1 + 85 \times 0.2 + 80 \times 0.33 \\ &+ 75 \times 0.15 + 70 \times 0.07 + 65 \times 0.04 + 60 \times 0.03 + 55 \times 0.01 = 80 \end{aligned}$$

(cf) Compare the result with ordinary way of calculating averages, which you may be more accustomed to from elementary school days, below:

$$\frac{100 \times 2 + 95 \times 5 + 90 \times 10 + 85 \times 20 + \cdots + 65 \times 4 + 60 \times 3 + 55 \times 1}{100}$$

Definition 3.4 Expectation of functions of r.v.'s*:*
In general, for any function $g(X)$ of a r.v. X, we defined the expectation of this function to be [12]*:*

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

1. *m-th moment:*
If $g(X) = X^m$ where m is an integer, we call it the *m-th moment* of r.v. X, i.e.:

$$\text{m-th moment} \triangleq E\left[X^m\right] = \int_{-\infty}^{\infty} x^m f_X(x) dx$$

(cf)
The first moment($m = 1$) is called the *mean* and denoted as μ_X, whereas the second moment($m = 2$) is called its *mean squared value.*

Example 3.19

Find the mean and the eman squared value of a uniform r.v. $X \sim U[a, b]$.

Solution:
The mean is given by:

$$E(X) = \int_a^b \frac{xdx}{b-a} = \left.\frac{x^2}{2(b-a)}\right|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}$$

whereas the eman squared value is as follows:

$$E(X^2) = \int_a^b \frac{x^2dx}{b-a} = \left.\frac{x^3}{3(b-a)}\right|_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{a^2+ab+b^2}{3}$$

[12] We take this as a definition here, but it can actually be proved: *more advanced course on probability...*

2. <u>*central moment:*</u>

   If $g(X) = (X - \mu_X)^n$ where n is an integer, we call it the *n-th central moment* of r.v. X, i.e.:

$$m_n \triangleq E\left[(X - \mu_X)^n\right] = \int_{-\infty}^{\infty} (x - \mu_X)^n f_X(x)dx$$

(cf)

The second central moment is especially called the *variance*, and denoted by the symbol σ_X^2:

$$\sigma_X^2 \triangleq m_n = E\left[(X - \mu_X)^2\right] = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)dx$$

Assignment:

Show that the variance of a uniform r.v. $X \sim U[a, b]$ is $\sigma_X^2 = (b - a)^2/12$.

Note:

The square root of the variance is called the *standard deviation*, and it represents the average amount of spread around the mean [13] :

$$\sigma = \sqrt{E\{[X - E(X)]^2\}}$$

[13] Note that $E[X - E(X)]$ is NOT adequate for representing the spread about mean since positive and negative values of the difference $X - E(X)$ will cancel out, thus smaller measure of deviation may result. On the other hand, $E[|X - E(X)|]$ would cure this problem, but hard to handle the absolute value analytically...

Example 3.20

Consider a Gaussian r.v. X, whose pdf is given by:

$$f_X(x) = \frac{e^{-(x-m)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}$$

(a) Show that the mean is $\mu_X = m$.

(b) Find the central moments of X.

Solution:

(a) The mean is given by:

$$\begin{aligned}
\mu_X &= \int_{-\infty}^{\infty} x \frac{e^{-(x-m)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} dx \quad (\text{let } u = x - m) \\
&= \int_{-\infty}^{\infty} (u+m) \frac{e^{-u^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} du \\
&= \int_{-\infty}^{\infty} u \frac{e^{-u^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} du + m \int_{-\infty}^{\infty} \frac{e^{-u^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} du \\
&= 0 + m = m
\end{aligned}$$

(b) By the definition of the central moments, we have:

$$\begin{aligned}
m_n = E[(X-\mu_X)^n] &= \int_{-\infty}^{\infty} (x-\mu_X)^n \frac{e^{-(x-\mu_X)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} dx \\
&= \int_{-\infty}^{\infty} u^n \frac{e^{-u^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} du, \quad n = 1, 2, \cdots
\end{aligned}$$

which is zero when n is odd. (why?)
For the case when n being even integers,the integrand is symmetric about $u = 0$, and employing the table of integral, we obtain:

$$m_{2k} = 2\int_{-\infty}^{\infty} u^{2k} \frac{e^{-u^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} du = 1 \cdot 3 \cdot \cdots \cdot (2k-1)\sigma^{2k}, \quad k = 1, 2, \cdots$$

Note that the special case of $n = 2k = 2$ provides the variance σ^2!!!

Properties of expectation:

1. The expectation of a constant is the constant itself:

$$E[a] = a, \quad a = \text{constant}$$

2. The expectation of a constant times a function of r.v. is the constant times the expectation of the function of r.v.:

$$E[ag(X)] = aE[g(X)], \quad a = \text{constant}$$

3. The expectation of the sum of two functions of r.v. is the sum of each expectation:

$$E[g_1(X) + g_2(X)] = E[g_1(X)] + E[g_2(x)]$$

proof: assignment

Note: Combination of the properties 2 & 3 is called the *linearity* property of the expectation $\ni$: $E[ag_1(X) + bg_2(X)] = aE[g_1(X)] + bE[g_2(x)]$.

Example 3.21

Show that the variance of a r.v. X can be computed according to:

$$\sigma_X^2 = E[(x - \mu_X)^2] = E(X^2) - [E(X)]^2$$

Solution:

Using the foregoing properties,

$$\begin{aligned} \sigma_X^2 &= E[(x - \mu_X)^2] \\ &= E(X^2 - 2\mu_X X + \mu_X^2) \\ &= E(X^2) - 2\mu_X E(X) + E(\mu_X^2) \\ &= E(X^2) - 2\mu_X^2 + \mu_X^2 \\ &= E(X^2) - \mu_X^2 \end{aligned}$$

Example 3.22

Let two r.v.'s X and Y are linearly related as:

$$Y = aX + b$$

Find the mean and variance of Y in terms of those of X.

Solution:

Using the properties of expectation, we have the mean as:

$$\mu_Y = E[aX + b] = aE(X) + E(b) = a\mu_X + b$$

whereas the variance is given by:

$$\sigma_Y^2 = E[(Y - \mu_Y)^2] = E\{[(aX + b) - (a\mu_X + b)]^2\} = E[a^2(X - \mu_X)^2] = a^2\sigma_X^2$$

Example 3.23

The mean and variance of a binomial random variable.

Solution: *Self study*

3.6 Characteristic Function

Definition 3.5 The characteristic function:
The characteristic function *of a r.v. is a special case of the mathematical expectation defined as follows:*

$$M_X(j\nu) = E(e^{j\nu X}) \triangleq \int_{-\infty}^{\infty} f_X(x) e^{j\nu x} dx$$

Usefulness of characteristic function:

1. **The *m-th* moment of a r.v. can be obtained by differentiating the characteristic function w.r.t. its argument.**
2. Sometimes the characteristic function of a r.v. is easier to obtain than the pdf.
3. The characteristic function and the pdf are Fourier transform pairs.

To show the first statement, we differentiate the characteristic function w.r.t. ν to obtain:

$$\frac{dM_X(j\nu)}{d\nu} = \int_{-\infty}^{\infty} f_X(x) \frac{d}{d\nu} e^{j\nu x} dx = j \int_{-\infty}^{\infty} x f_X(x) e^{j\nu x} dx$$

Now set $\nu = 0$, and divide by j to get:

$$-j \left. \frac{dM_X(j\nu)}{d\nu} \right|_{\nu=0} = \int_{-\infty}^{\infty} x f_X(x) dx = E(X)$$

Repeating the same procedure n times, the $n-th$ moment of the r.v. X can generally be expressed as:

$$E(X^m) = (-j)^m \left. \frac{d^m M_X(j\nu)}{d\nu^m} \right|_{\nu=0}$$

Example 3.24

Find the characteristic function of a Cauchy r.v. w/ its pdf given as:

$$f_X(x) = \frac{\alpha/\pi}{x^2+\alpha^2}$$

Solution:

Applying the definition of the characteristic function, we obtain:

$$\begin{aligned} M_X(j\nu) &= \int_{-\infty}^{\infty} \frac{\alpha/\pi}{x^2+\alpha^2} e^{j\nu x} dx \\ &= \int_{-\infty}^{\infty} \frac{\alpha/\pi}{x^2+\alpha^2} [\cos(\nu x) + j\sin(\nu x)] dx \\ &= \frac{\alpha}{\pi} \int_{-\infty}^{\infty} \frac{\cos(\nu x)}{x^2+\alpha^2} dx \end{aligned}$$

which, by use of a table of indefinite integral, can be expressed as [14] :

$$M_X(j\nu) = e^{\alpha|\nu|}$$

Example 3.25

Find the characteristic function of the double sided exponential r.v.(called the Laplacian r.v.), whose pdf is given by:

$$f_X(x) = \frac{\alpha}{2} e^{-\alpha|x|}, \quad \alpha > 0$$

[14] Note that $M_X(j\nu)$ is not differentiable at $\nu = 0$, and thus we cannot use it to evaluate the moments. In fact, its moments do not exists in this case.

Solution:

By the definition of the characteristic function, we get:

$$\begin{aligned} M_X(j\nu) &= \int_{-\infty}^{\infty} \frac{\alpha}{2} e^{-\alpha|x|} e^{j\nu x} dx \\ &= \int_{-\infty}^{\infty} \frac{\alpha}{2} e^{-\alpha|x|} [\cos(\nu x) + j\sin(\nu x)] dx \\ &= \frac{\alpha}{2} \int_{-\infty}^{\infty} \cos(\nu x) e^{-\alpha|x|} dx \end{aligned}$$

which, by use of a table of indefinite integral, can be expressed as [15] :

$$M_X(j\nu) = \alpha \int_0^{\infty} \cos(\nu x) e^{-\alpha x} dx = \frac{\alpha^2}{\alpha^2 + \nu^2}$$

Assignment: Show that the first and the second moments are 0 and $2/\alpha^2$ respectively by differentiation.

[15]Note that the integrand is symmetric about $x = 0$.

3.7 Chebyshev's Inequality

Recall:
The standard deviation of a r.v. gives a measure of spread about its mean

$\Longrightarrow$ The Chebyshev's inequality provides a *bound* on the probability that a r.v. deviated more than k standard deviations from its mean [16] !!!

Chebyshev's Inequality:

For any random variable X, the probability of X being deviated from its mean more than k standard deviation must satisfy the following inequality:

$$P(|X - \mu_X| \geq k\sigma_X) \leq \frac{1}{k^2}$$

or [17]

$$P(|X - \mu_X| < k\sigma_X) > 1 - \frac{1}{k^2}$$

proof:

Let $Y = X - \mu_X$[18] and $a = k\sigma_X$. Then, the LHS of the first inequality becomes:

$$P(|Y| \geq a) = P(Y \leq -a) + P(Y \geq a)$$

which follows from the fact $|Y| \geq a$ is the union of two mutually exclusive events $Y \geq a$ and $Y \leq -a$.

[16] It is a very *loose* bound, but its merit is the fact that very little need to be known about the r.v. to obtain the bound...

[17] Note that two events $|X - \mu_X| \geq k\sigma_X$ and $|X - \mu_X| < k\sigma_X$ are mutually exclusive to each other!

[18] Note then: $E[Y^2] = \sigma_X^2$.

Now, consider the second moment of Y, which is:

$$\begin{aligned} E(Y^2) = \int_{-\infty}^{\infty} y^2 f_Y(y)dy &\geq \int_{-\infty}^{-a} y^2 f_Y(y)dy + \int_{a}^{\infty} y^2 f_Y(y)dy \\ &\geq a^2 \left[\int_{-\infty}^{-a} f_Y(y)dy + \int_{a}^{\infty} f_Y(y)dy\right] \\ &= a^2[P(Y \leq -a) + P(Y \geq a)], \quad a > 0 \end{aligned}$$

Solving, we obtain:

$$P(Y \leq -a) + P(Y \geq a) = P(|Y| \geq a) \leq \frac{E(Y^2)}{a^2}$$

Replacing $Y = X - \mu_X$ with $E[Y^2] = \sigma_X^2$, and $a = k\sigma_X$, we have the Chebyshev's inequality as:

$$P(|X - \mu_X| \geq k\sigma_X) \leq \frac{1}{k^2}$$

Q.E.D.

Example 3.26

(a) Find a bound on the probability that a r.v. is within three standard deviations of its mean.

(b) Find the exact probability of this event, if the r.v. is a Gaussian, and compare with the bound.

Solution:

(a) From the Chebyshev's inequality, we have:

$$P(|X - \mu_X| < 3\sigma_X) > 1 - \frac{1}{3^2} = 0.889$$

(b) The probability of the given event for a Gaussian r.v. is:

$$\begin{aligned} P(|X - \mu_X| < 3\sigma_X) &= \int_{\mu_X - 3\sigma_X}^{\mu_X + 3\sigma_X} \frac{e^{-(x-\mu_X)^2/2\sigma_X^2}}{\sqrt{2\pi\sigma_X^2}} dx \\ &= \int_{-3}^{3} \frac{e^{-u^2/2}}{\sqrt{2\pi}} du = 2\int_{0}^{3} \frac{e^{-u^2/2}}{\sqrt{2\pi}} du \\ &= 1 - 2Q(3) = 1 - 2 \times 0.00135 \\ &= 0.9973 \end{aligned}$$

(cf) Note that the Chebyshev's inequality does NOT provide a *tight* bound in this case!!!

3.8 Computer Generation of Random Variables

Recall:

1. Generation of uniform pseudorandom numbers $X \sim U[0,1]$:

$$X = rand(1, 1000);$$

2. Generation of Gaussian pseudorandom numbers $Y \sim N(0,1)$ [19] :

$$Y = randn(p, q)$$

3. Generation of Gaussian pseudorandom numbers $Z \sim N(m, \sigma^2)$ [20] :

$$Z = \sigma Y + m$$

Generation of random numbers with an arbitrary distribution:

Let U be a r.v. uniformly distributed in $[0,1]$, and define a new r.v. V as:

$$V = g(U)$$

where $g(\cdot)$ is assumed to br monotonic.

Then, the pdf of the newly defined r.v. V is given by:

$$\begin{aligned} f_V(v) &= f_U(u)\left|\frac{du}{dv}\right|_{u=g^{-1}(v)} \\ &= \begin{cases} \left|\frac{du}{dv}\right| = \left|\frac{dg^{-1}(v)}{dv}\right|, & 0 \le u \le 1 \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

where the last equation follows because $f_U(u)$ is unity in $[0,1]$ and zero elsewhere.

[19] This generates an array of Gaussian pseudorandom numbers with p rows and q columns.
[20] By way of transformation.

Re-writing the above result for the case of $0 \le u \le 1$:

$$f_V(v) = \begin{cases} \frac{dg^{-1}(v)}{dv}, & \frac{dg^{-1}(v)}{dv} \ge 0 \\ \\ -\frac{dg^{-1}(v)}{dv}, & \frac{dg^{-1}(v)}{dv} < 0 \end{cases}$$

Integrating and solving for $g^{-1}(v)$, we obtain:

$$g^{-1}(v) = \begin{cases} \int_{-\infty}^{v} f_V(\lambda)d\lambda = F_V(v), & \frac{dg^{-1}(v)}{dv} \ge 0 \\ \\ -\int_{-\infty}^{v} f_V(\lambda)d\lambda = -F_V(v), & \frac{dg^{-1}(v)}{dv} < 0 \end{cases}$$

where $F_V(v)$ represents the *desired* cdf of r.v. V.

Example 3.27

Using a uniform r.v. U uniformly distributed in $[0, 1]$, find the required transformation $V = g(U)$ so that it will generate an exponential pdf given by:

$$f_V(v) = 2e^{-2v}u(v)$$

Solution:

The cdf of the desired exponential r.v. is:

$$F_V(v) = \int_{-\infty}^{v} f_V(\lambda)d\lambda = \begin{cases} 0, & v < 0 \\ \\ 1 - e^{-2v}, & v \ge 0 \end{cases}$$

From which, we obtain [21]:

$$u = g^{-1}(v) = 1 - e^{-2v}, \quad v \ge 0$$

Solving for v, expressing it into the relationship between two r.v.'s U and V [22]:

$$\begin{aligned} V &= -0.5\ln(1 - U) \\ &= -0.5\ln(U) \end{aligned}$$

which means that the required transformation is $V = g(U) = -0.5\ln(U)$.

[21] Note that this inverse transformation always has positive slope.

[22] Here we use that fact: if U is uniform on $[0, 1]$, so is $1 - U$.