MATH4427 Loss Models and its application

Lecture Note 1

A quick review on basic probability theory

Random variables

There are two uncertainties in a typical loss model: (1) Frequency and timing of payments and (2) Amount of each payment.

Mathematically, these two uncertainties can be modelled by random variables with suitable probability distributions.

Definition (Random variable)

We let Ω be a sample space (collection of all possible outcomes). A random variable X is a real-valued function that maps an outcome ω in the sample space to a real number. That is,

$$X = X(\omega) \in \mathbb{R}.$$

- We say X is a *discrete* random variable if the set of possible values is finite or countably infinite.
- We say X is a *continuous* random variable if the set of possible values is uncountable and $P(X = x) = 0$.

Some examples of discrete and continuous random variables in loss models

In studying insurance problem, discrete random variable is often used to model the number of claims/loss events involved in the policies.

- We consider a motor vehicle insurance (full insurance) that covers the loss caused by an insured vehicle. We let N_1 be the number of claim requests made by the policyholder. Then N_1 is discrete random variable which takes the values $0,1,2,\ldots$.
- Suppose that an insurance company sold n life insurances recently, then the total number of claims made N_2 is also discrete random variable since it takes the value over $\{0,1,2,\ldots,n\}$.

On the other hand, continuous random variable is used in the following contexts.

- We let X_i be the loss amount (or claim amount) in a claim event, then X_i is a continuous random variable which takes value over $[0,\infty)$.
- We let T be the life expectancy of a person who has pursued a life insurance, then T is a continuous random variable which takes value over $[0,M]$, where M is a large number.

Probability mass function and Probability density function of random variable

When X is a discrete random variable, the distribution of X can be described by *probability mass function* $f_X(x)$. For any value of x, $f_X(x)$ simply indicates the probability that the random variable X equals x. That is,

$$f_X(x) \overset{def}{\triangleq} P(X = x).$$

When X is a continuous random variable, the distribution of X is described by a *probability density function* $f(x)$ (where $f(x) \geq 0$). For any real number a, b such that $a \leq b$, the probability that X lies between a and b is given by

$$P(a \leq X \leq b) \overset{def}{\triangleq} \int_a^b f_X(x)dx.$$

*Note: When $a = b = x$, we have $P(X = x) = \int_x^x f_X(y)dy = 0$.

Cumulative distribution function

Given a probability mass function/probability density function of a random variable, the cumulative distribution function (c.d.f in short) is defined as the probability that the value of random variable is less than x. That is,

$$F_X(x) = P(X \leq x) = \begin{cases} \sum_{z \leq x} f_X(z) & if\ \ X\ \ is\ discrete \\ \int_{-\infty}^{x} f_X(z)dz & if\ \ X\ is\ continuous \end{cases}.$$

Depending on the nature of the random variable, the cumulative distribution function can be either continuous or piecewise continuous.

Example 1 (c.d.f. of Discrete random variable)

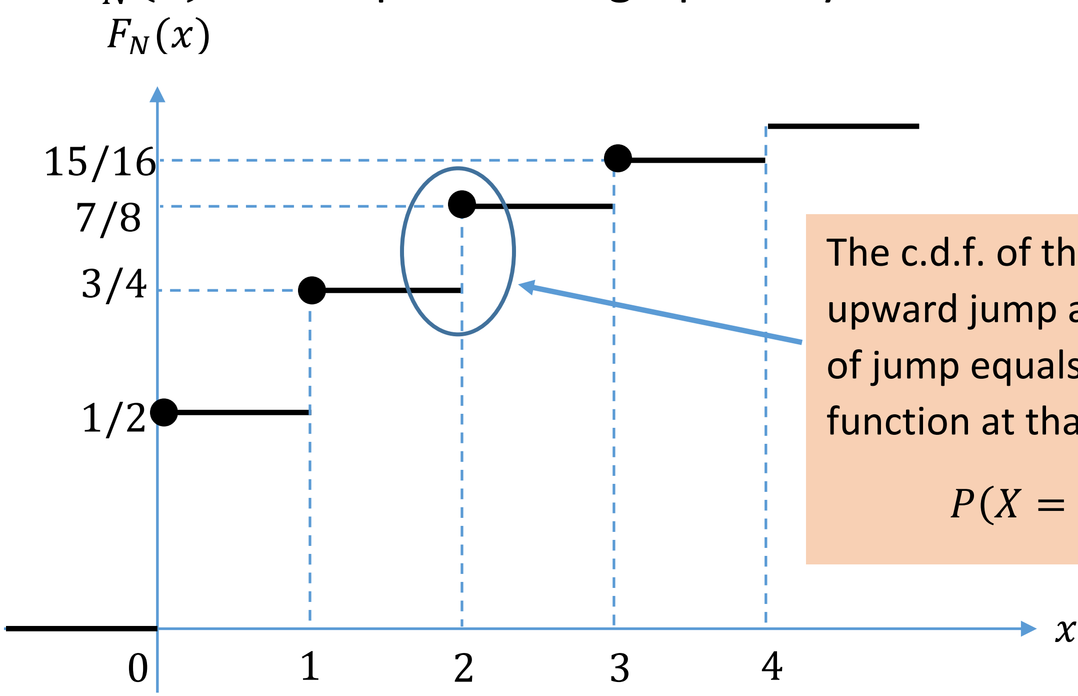
We let N be a discrete random variable with the probability mass function:

$$f_N(n) = P(N = n) = \frac{1}{2^{n+1}}, \quad n = 0, 1, 2, \cdots.$$

The cumulative distribution function of the random variable N is seen to be

$$F_N(x) = P(N \le x) = P(N \le \lfloor x \rfloor) = \sum_{n=0}^{\lfloor x \rfloor} f_N(n) = \sum_{n=0}^{\lfloor x \rfloor} \frac{1}{2^{n+1}} = 1 - \frac{1}{2^{\lfloor x \rfloor + 1}}.$$

The c.d.f $F_N(x)$ can be presented graphically as follows:



The c.d.f. of the random variable has upward jump at some point. The size of jump equals the probability mass function at that time. That is,

$$P(X = 2) = \frac{7}{8} - \frac{3}{4} = \frac{1}{8}$$

We observe the c.d.f. is *piecewise continuous*.

Example 2 (c.d.f. of continuous random variable)

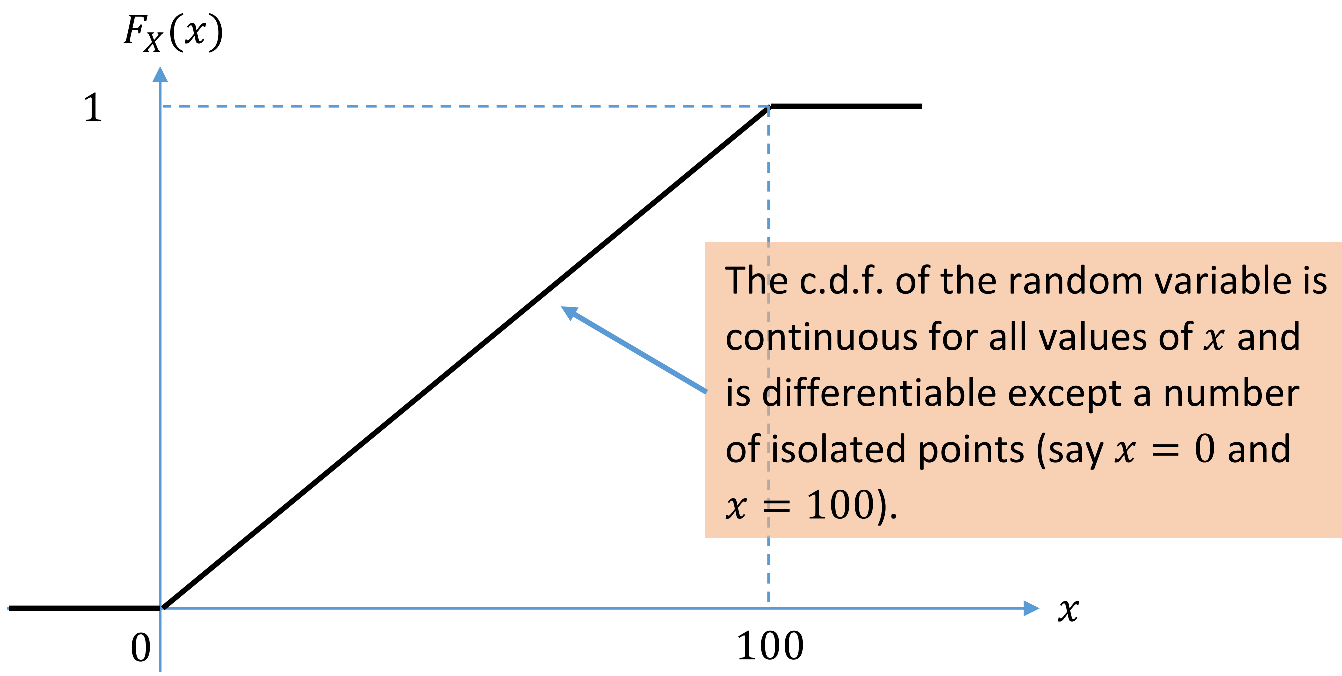
We let Y be a continuous random variable with the following probability density function (i.e. Y has uniform distribution

$$f_Y(y) = P(Y = y) = \begin{cases} \dfrac{1}{100} & if \;\; 0 \leq y \leq 100 \\ 0 & if \; otherwise \end{cases}.$$

The cumulative distribution of Y can be computed as

$$F_Y(y) = P(Y \leq y) = \begin{cases} 0 & if \;\; y < 0 \\ \displaystyle\int_0^y \frac{1}{100} dz = \frac{y}{100} & if \;\; 0 \leq y \leq 100 \\ \displaystyle\int_0^{100} \frac{1}{100} dz = 1 & if \;\; y > 100 \end{cases}.$$

The function can be presented graphically as follows:



Remark: Relation between $f_X(x)$ and $F_X(x)$ for continuous random variable

Provided that $F_X(x)$ is differentiable, then we have

$$f_X(x) = F_X'(x).$$

This equation is useful in studying the distribution of new random variables.

Example 3

We let X, Y be two *independent* random variables such that each of the random variables has exponential distribution with mean θ. We define $Z = X + Y$.
Identify the distribution of Z by find the probability density function of Z.

☺Solution

Using law of total probability, we deduce that for any $z > 0$

$$F_Z(z) = P(X + Y \le z) = \int_0^\infty P(X + Y \le z | Y = y) \underbrace{\left(\frac{1}{\theta} e^{-\frac{y}{\theta}}\right)}_{f_Y(y)} dy$$

$$= \int_0^\infty P(X + y \le z | Y = y) \left(\frac{1}{\theta} e^{-\frac{y}{\theta}}\right) dy = \int_0^\infty P(X \le z - y) \left(\frac{1}{\theta} e^{-\frac{y}{\theta}}\right) dy$$

$$= \int_0^{z} P(X \le z - y) \left(\frac{1}{\theta} e^{-\frac{y}{\theta}}\right) dy = \int_0^z \left(\int_0^{z-y} \frac{1}{\theta} e^{-\frac{x}{\theta}} dx\right) \left(\frac{1}{\theta} e^{-\frac{y}{\theta}}\right) dy$$

$$= \int_0^z \left(1 - e^{-\frac{z-y}{\theta}}\right) \left(\frac{1}{\theta} e^{-\frac{y}{\theta}}\right) dy = \int_0^z \frac{1}{\theta} e^{-\frac{y}{\theta}} dy - \int_0^z \frac{1}{\theta} e^{-\frac{z}{\theta}} dy = \left(1 - e^{-\frac{z}{\theta}}\right) - \frac{1}{\theta} z e^{-\frac{z}{\theta}}.$$

So the p.d.f. is given by

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{1}{\theta} e^{-\frac{z}{\theta}} - \frac{1}{\theta} e^{-\frac{z}{\theta}} + \frac{1}{\theta^2} z e^{-\frac{z}{\theta}} = \frac{1}{\theta^2} z e^{-\frac{z}{\theta}}.$$

Example 4

We let Z be a standard normal variable. Find the distribution of $Y = Z^2$.

☺Solution

We first compute the c.d.f. for Y. For any $y \geq 0$, we consider

$$F_Y(y) = P(Y \leq y) = P(Z^2 \leq y) = P\left(-\sqrt{y} \leq Z \leq \sqrt{y}\right) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz.$$

Hence, the p.d.f. of Y is found to be

$$f_Y(y) = F_Y'(y) = \frac{d}{dy}\left(\int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz\right)$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{(\sqrt{y})^2}{2}} \left(\frac{d}{dy}\sqrt{y}\right) - \frac{1}{\sqrt{2\pi}} e^{-\frac{(-\sqrt{y})^2}{2}} \left(\frac{d}{dy}\left(-\sqrt{y}\right)\right) = \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}}.$$

(*Note: In fact, Y has chi-squared distribution with degree of freedom 1.)

Mixed random variable

To motivate this concept, we consider the following scenario:

We let X be a continuous random variable (loss) which is uniformly distributed on the interval $[0,500]$ (That is, the probability density function of X is $f(x) = \begin{cases} \frac{1}{500} & if\ \ 0 \leq x \leq 500 \\ 0 & if\ otherwise \end{cases}$).

We consider an insurance that pays insured person $X - 100$ if the loss X is greater than 100 and nothing of otherwise.

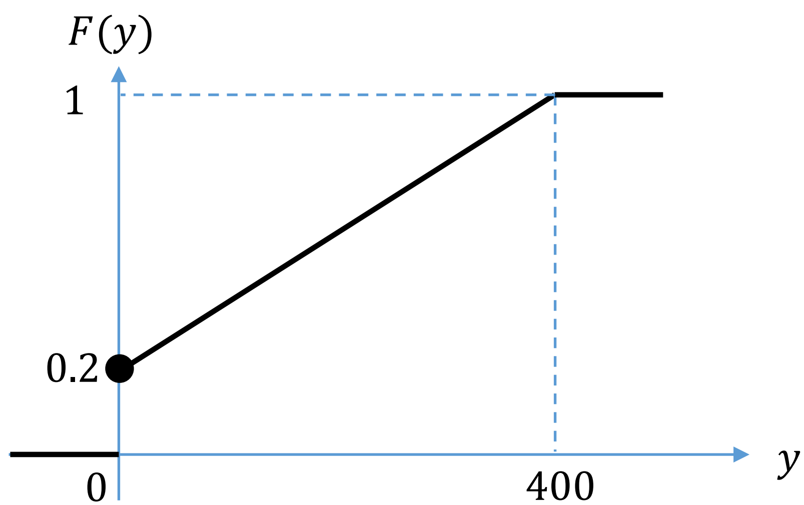
We let Y be the amount paid by the insurance company in a single claim. Using the above information, Y can be expressed as

$$Y = \begin{cases} X - 100 & if\ X > 100 \\ 0 & if\ X \leq 100 \end{cases}.$$

Although Y appears to be a continuous random variable, it is not continuous random variable since $P(Y = 0) = P(0 \leq X \leq 100) = \int_0^{100} \frac{1}{500} dx = \frac{1}{5} \neq 0.$

The cumulative distribution function of Y is found to be

$$F(y) = P(Y \leq y) = \begin{cases} 0 & if \;\; y < 0 \\ \underbrace{\int_0^{100} \frac{1}{500} dx}_{=P(X \leq 100+y)} & if \; y = 0 \\ \underbrace{\int_0^{100+y} \frac{1}{500} dx}_{=P(X \leq 100+y)} & if \; 0 < y < 400 \\ 1 & if \;\; y \geq 400 \end{cases} = \begin{cases} 0 & if \;\; y < 0 \\ \frac{1}{5} & if \; y = 0 \\ \frac{100+y}{500} & if \; 0 < y < 400 \\ 1 & if \;\; y \geq 400 \end{cases}.$$



We observe from the above graph that

- the left portion of the function behaves like the cumulative distribution function of discrete random variable (i.e. a upward jump at $y = 0$) and
- the right portion of the function behaves like the cumulative distribution function of continuous random variable (i.e. continuous everywhere and differentiable almost everywhere).

Hence, one observes that this random variable is a "mixture" of discrete and continuous random variables. It is called *mixed random variable.*

Definition (Mixed random variable)

A random variable X is said to be *mixed* if it satisfies the following conditions:

- It is not discrete random variable;
- The corresponding cumulative distribution function $F(x)$ is continuous everywhere with the exception of at least one value and at most a countable number of values.
  (*In other words, $F(x)$ has upward jumps at some points)

Example 5 (Obtaining probability density function for a mixed random variable)

We let X be the amount of loss (in thousands) to the car owner in a traffic accident. X is usually modelled as a continuous, non-negative random variable.

Assume that the probability density function of the random variable X is given by

$$f_X(x) = \begin{cases} 0 & if \;\; x < 1 \\ x^{-2} & if \;\; x \geq 1 \end{cases}.$$

The car owner has bought a car insurance from an insurance company. The terms of the contract states that the car owner needs to pay for the loss up to \$5 and the company will pay the remaining.

We let Y be the amount paid by the car owner. Determine the probability density function of Y.

☺Solution

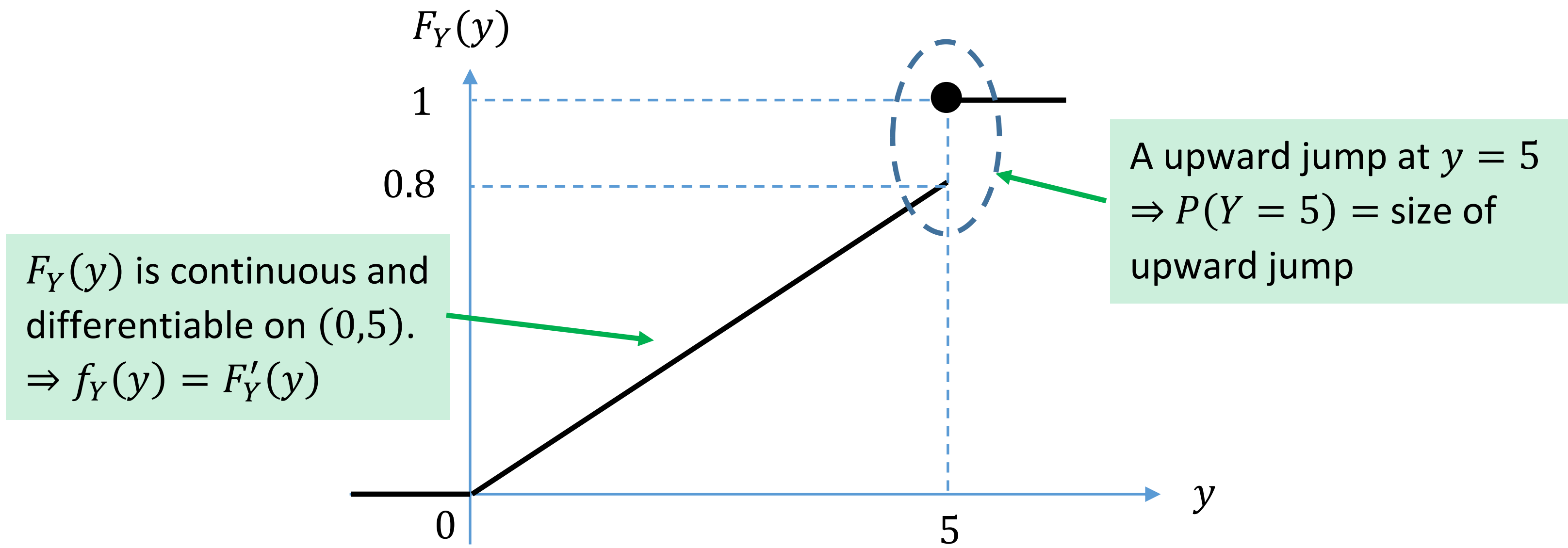
Note that the random variable Y can be expressed as

$$Y = \min(X, 5) = \begin{cases} X & if\ X \le 5 \\ 5 & if\ \ X > 5 \end{cases}.$$

The cumulative density function of Y can be computed as

$$F_Y(y) = P(Y \le y) = \begin{cases} 0 & if\ \ y < 1 \\ P(X \le y) & if\ 1 \le y < 5 \\ 1 & if\ y \ge 5 \end{cases} = \begin{cases} 0 & if\ \ y < 1 \\ \displaystyle\int_1^y x^{-2} dx & if\ 1 \le y < 5 \\ 1 & if\ y \ge 5 \end{cases}$$

$$= \begin{cases} 0 & if\ \ y < 1 \\ -x^{-1}|_1^y & if\ 1 \le y < 5 \\ 1 & if\ y \ge 5 \end{cases} = \begin{cases} 0 & if\ \ y < 1 \\ 1 - \dfrac{1}{y} & if\ 1 \le y < 5. \\ 1 & if\ y \ge 5 \end{cases}$$



Hence, the probability density function of Y is given by

$$f_Y(y) = \begin{cases} \frac{d}{dy} F_Y(y) & if \ y < 5 \\ 1 - 0.8 & if \ \ y = 5 \end{cases} = \begin{cases} \frac{1}{y^2} & if \ y < 5 \\ 0.2 & if \ \ y = 5 \end{cases}.$$

Survival function and Hazard rate function

Survival function

Given a random variable X, we define the *survival function* (denoted by $S_X(x)$ or $S(x)$) to be the probability that the random variable X is *strictly* greater than x. That is,

$$S_X(x) = P(X > x) = 1 - P(X \le x) = 1 - F_X(x),$$

where $F_X(x)$ is the cumulative distribution function of X.

This function is quite useful in studying various phenomena in loss model:

- Study the *tail distribution* of loss in severity model. That is, the probability that the insurance company faces a big claim from the client.
- Study the life expectancy of an insured person (mortality study). By taking X be life span (age at death) of the person, then $S_X(x)$ represents the probability that the person lives for at least x years (or is still alive after x years). This is important in life insurance or medical insurance.

Hazard rate function

We consider the following examples.

We let T (measured in years) be the life-time of a randomly selected person. Here, T is assumed to be a continuous random variable with probability density function $f(t)$).

Suppose that the person has lived for t years, one would like to examine the probability that this person will die in the next time period $[t, t + \Delta t]$, where Δt is a small positive constant.

We let $F(t)$ be the cumulative distribution function of T. Using the definition of conditional probability, the required probability can be computed as

$$P(t < T \leq t + \Delta t | T \geq t) = \frac{P(t < T \leq t + \Delta t \;\; and \;\; T \geq t)}{P(T \geq t)}$$

$$= \frac{P(t < T \leq t + \Delta t)}{P(T \geq t)} = \frac{\int_t^{t+\Delta t} f(u)du}{\int_t^{\infty} f(u)du} = \frac{\int_t^{t+\Delta t} f(u)du}{1 - F(t)} \ldots \ldots (*)$$

Since Δt is very small, one would expect that $f(u)$ is *almost constant* over the interval $[t, t + \Delta t]$. So we have $f(u) \approx f(t)$ for all $u \in [t, t + \Delta t]$.

From eq. $(*)$, we can deduce that

$$P(t < T \leq t + \Delta t | T \geq t) \approx \frac{f(t) \int_t^{t+\Delta t} du}{1 - F(t)} = \frac{f(t)}{1 - F(t)} \Delta t.$$

We observe that the probability is proportional to the length of time period Δt with the proportionality constant $\frac{f(t)}{1-F(t)}$. The function $\frac{f(t)}{1-F(t)}$ is known as the hazard rate function (or force of mortality or failure rate function) at time t and is denoted by $h(t)$. That is,

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)}.$$

Recovering the probability distribution from hazard rate function

Given the hazard rate function, one can deduce the probability distribution of the random variable as follows:

Using the fact that $S'(t) = -F'(t) = -f(t)$, we get

$$h(t) = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = -\frac{d}{dt}[\ln S(t)].$$

Integrate the above equation with respect to t and use the fact that $S(0) = P(T > 0) = 1$, we deduce that

$$\int_0^t h(u)du = -\int_0^t \frac{d}{du}[\ln S(u)]du \Rightarrow \ln S(t) - \underbrace{\ln S(0)}_{=\ln 1=0} = -\int_0^t h(u)du$$

$$\Rightarrow S(t) = e^{-\int_0^t h(u)du}.$$

So we get

$$F(t) = 1 - e^{-\int_0^t h(u)du} \; and \; f(t) = F'(t) = h(t)e^{-\int_0^t h(u)du}, \qquad t \geq 0.$$

Example 6 (Hazard rate function of some special random variables)

(a) We let X be an exponentially distributed random variable with parameter λ. That is, the probability density function of X is given by

$$f_X(x) = \lambda e^{-\lambda x}, \qquad x \geq 0.$$

The hazard rate function for X can be computed as

$$h_X(x) = \frac{f_X(x)}{S_X(x)} \overset{S_X(x)=P(X>x)}{\triangleq} \frac{\lambda e^{-\lambda x}}{\int_x^\infty \lambda e^{-\lambda z} dz} \overset{\int \lambda e^{-\lambda z} dz = -e^{-\lambda z}}{\triangleq} \frac{\lambda e^{-\lambda x}}{e^{-\lambda x}} = \lambda.$$

We observe that the hazard rate function of X is constant.

(b) We let Y be a Pareto random variable with parameters α and θ (where $\alpha > 0$ and $\theta > 0$). The probability density function of Y is given by

$$f_Y(y) = \frac{\alpha\theta^\alpha}{(y+\theta)^{\alpha+1}}, \qquad y \geq 0.$$

The hazard rate function for Y can be computed as

$$h_Y(y) = \frac{f_Y(y)}{S_Y(y)} = \frac{\frac{\alpha\theta^\alpha}{(y+\theta)^{\alpha+1}}}{\int_y^\infty \frac{\alpha\theta^\alpha}{(z+\theta)^{\alpha+1}}\,dz} = \frac{\frac{\alpha\theta^\alpha}{(y+\theta)^{\alpha+1}}}{\left[\frac{-\theta^\alpha}{(z+\theta)^\alpha}\right]_y^\infty} = \frac{\frac{\alpha\theta^\alpha}{(y+\theta)^{\alpha+1}}}{\frac{\theta^\alpha}{(y+\theta)^\alpha}} = \frac{\alpha}{y+\theta}.$$

The hazard rate function is decreasing with respect to y.

(c) We let Z be a random variable which has uniform distribution over $[0, c]$. The probability density function of Y is given by

$$f_Z(z) = \frac{1}{c}, \qquad z \in [0, c].$$

The hazard rate function of Z is found to be

$$h_Z(z) = \frac{f_Z(z)}{S_Z(z)} = \frac{\frac{1}{c}}{\int_z^c \frac{1}{c}\,dx} = \frac{\frac{1}{c}}{\frac{c-z}{c}} = \frac{1}{c-z}, \quad z \in [0, c).$$

The hazard rate function is increasing with respect to z. It is also worthwhile to point out that $h_Z(z)$ is defined over the interval $[0, c)$ so that $S_Z(z) = \frac{c-z}{c} > 0$.

Example 7

It is given that the hazard rate function of a non-negative random variable T is $h(x) = \frac{ax}{1+x^2}$. Suppose that $S(1) = P(T > 1) = 0.25$, find the value of $S(0.5)$.

☺Solution

We first recover the survival function/ cumulative distribution function of X. Note that

$$S(x) = e^{-\int_0^x h(y)dy} = e^{-\int_0^x \frac{ay}{1+y^2}dy} = e^{-\frac{a}{2}\ln(1+y^2)|_0^x} = e^{-\frac{a}{2}\ln(1+x^2)}$$
$$= (1+x^2)^{-\frac{a}{2}}.$$

Since $F(1) = \frac{3}{4}$, we have $S(1) = 1 - F(1) = \frac{1}{4}$ and

$$\underbrace{(1+1^2)^{-\frac{a}{2}}}_{S(1)} = \frac{1}{4} \Rightarrow a = 4.$$

Hence, we conclude that $S_{0.5} = (1+(0.5)^2)^{-\frac{4}{2}} \approx 0.64.$

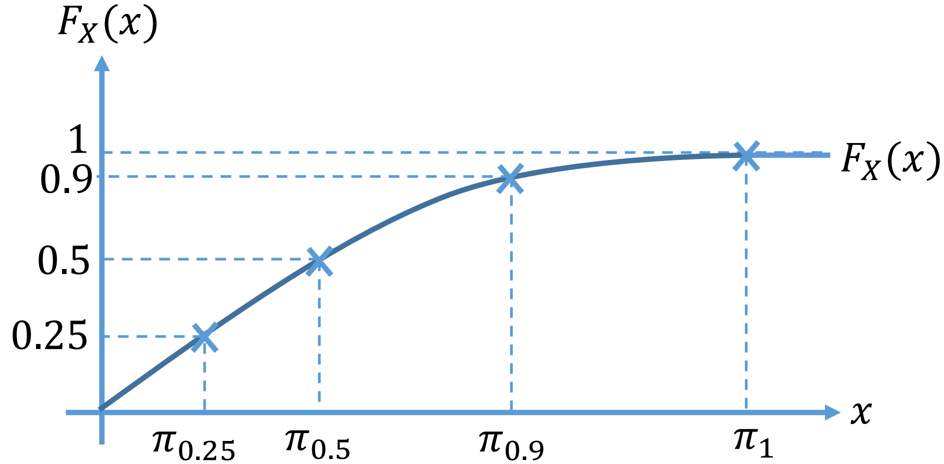
Percentiles of a distribution

Definition (Percentiles for continuous random variable)

We let X be a continuous random variable. For any $p \in [0,1]$, we define $(100p)^{th}$ *percentile* of X, denoted by π_p, be a number satisfying

$$F_X(\pi_p) = P(X \le \pi_P) = p.$$

Roughly speaking, $(100p)^{th}$ percentile describes the value of X at a particular position of the probability distribution. A set of percentiles can roughly describe the shape of the probability distribution (i.e. c.d.f).



Example 8

A random variable X (loss variable) has Pareto-distribution with parameter α and θ (both α, θ are positive). The probability density function of X is given by

$$f_X(x) = \frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}}, \qquad x \geq 0.$$

It is given that 10th percentile of X of $\theta - k$ and 90th percentile of X is $5\theta - 3k$, where k is some constant. Find the value of α.

😊Solution

The cumulative distribution function of X is given by

$$F_X(x) = P(X \leq x) = \int_0^x \frac{\alpha\theta^\alpha}{(y+\theta)^{\alpha+1}} dy = \alpha\theta^\alpha \frac{(y+\theta)^{-\alpha}}{-\alpha}\Big|_0^x = 1 - \frac{\theta^\alpha}{(x+\theta)^\alpha}.$$

Using the given information, we have

$$\begin{cases} F_X(\theta - k) = 0.1 \\ F_X(5\theta - 3k) = 0.9 \end{cases} \Rightarrow \begin{cases} 1 - \dfrac{\theta^\alpha}{(2\theta - k)^\alpha} = 0.1 \\ 1 - \dfrac{\theta^\alpha}{(6\theta - 3k)^\alpha} = 0.9 \end{cases} \Rightarrow \begin{cases} \dfrac{\theta^\alpha}{(2\theta - k)^\alpha} = 0.9 \\ \dfrac{\theta^\alpha}{(6\theta - 3k)^\alpha} = 0.1 \end{cases} \overset{(1)\div(2)}{\Rightarrow} \frac{(6\theta - 3k)^\alpha}{(2\theta - k)^\alpha}$$

$$= 9 \Rightarrow 3^\alpha = 9 \Rightarrow \alpha = 2.$$

Some remarks about percentiles

- When $p = 0.5$, the corresponding percentile $\pi_{0.5}$ is called *median* which represents the value of X in the *middle* of the population.
- If X is continuous random variable, π_p exists for all $p \in [0,1]$ due to the continuity of the c.d.f..
- It may happen that π_p is not unique given a value of p. For example, we consider c.d.f given by $F_X(x) = \begin{cases} 0.1x & if\ \ 0 \leq x < 6 \\ 0.6 & if\ \ 6 \leq x < 9 \\ 0.6 + 0.4(x-9) & if\ \ 9 \leq x < 10 \\ 1 & if\ x \geq 10 \end{cases}$.
  If we take $p = 0.6$, we see that $F_X(x) = 0.6$ for all $x \in [6,9]$. Thus, π_p can be any real number between 6 and 9.
- One cannot apply the above definition to *find the percentiles for discrete random variable* (or mixed random variable) since the c.d.f. of the discrete random variable is *not* continuous and $F_X(\pi_p) = p$ may not have solution for some value of p.

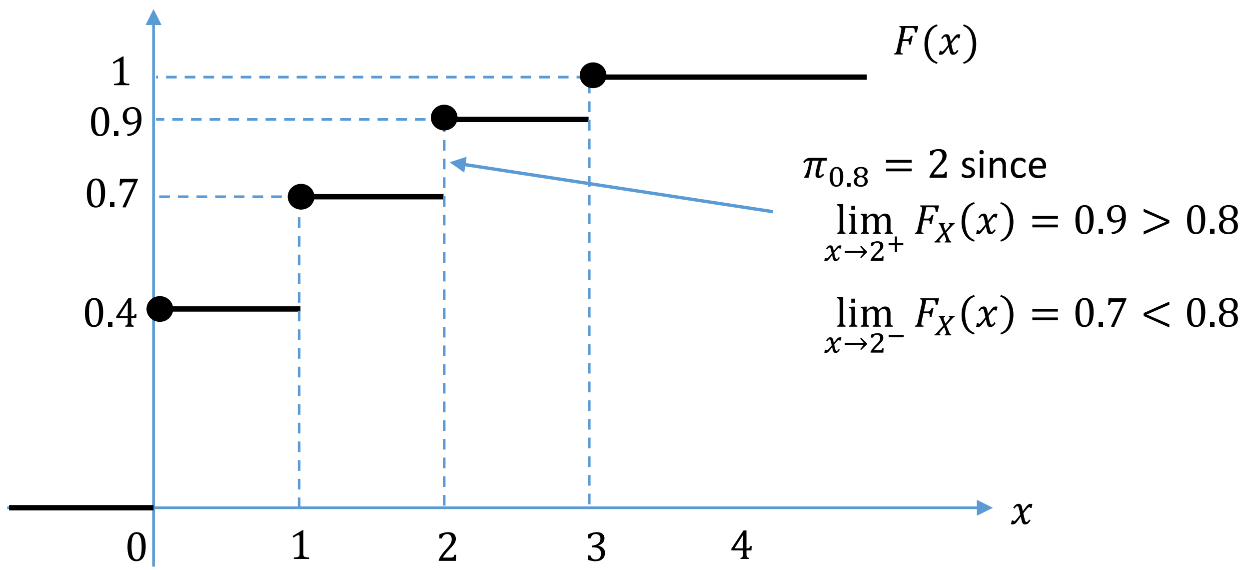
Definition (Percentiles for general random variable)

We let X be a random variable. For any $p \in [0,1]$, we define $(100p)^{th}$ *percentile* of X, denoted by π_p, be a number satisfying

$$\lim_{x \to \pi_p^-} F_X(x) \leq p \leq \lim_{x \to \pi_p^+} F_X(x).$$

(*Note: Since the c.d.f. is right-continuous, so we have $\lim_{x \to \pi_p^+} F_X(x) = F_X(\pi_p)$.

As an example, the c.d.f. of a random variable is given as follows and we would like to find $\pi_{0.8}$



Expected value of random variable

Given a random variable X, we define the expected value of X as the average value of X (weighted by its probability distribution). That is,

$$\mathbb{E}[X] = \begin{cases} \sum_{x} xP(X = x) & if\ X\ is\ discrete \\ \int_{-\infty}^{\infty} xf_X(x)dx & if\ X\ is\ continuous \end{cases}.$$

In some case, we may wish to evaluate the expected value of function of X (i.e. $g(X)$, where g is some function), then the expected value of $g(X)$ can be expressed as

$$\mathbb{E}[g(X)] = \begin{cases} \sum_{k=1}^{\infty} g(a_k)P(X = a_k) & if\ \ X\ is\ discrete \\ \int_{-\infty}^{\infty} g(x)f(x)dx & if\ X\ is\ continuous \end{cases}$$

Moments of a random variable

Definition (k^{th} moment of random variable)

We let X be a random variable. The k^{th} raw moment of random variable is defined as the *expected value* of k^{th} power of X. That is,

$$k^{th}\ moment = \mathbb{E}[X^k] = \begin{cases} \displaystyle\sum_x x^k P(X = x) & if\ X\ is\ discrete \\ \displaystyle\int_{-\infty}^{\infty} x^k f(x)dx & if\ X\ is\ continuous \end{cases}.$$

Remark

- When $k = 1$, the 1st moment $\mathbb{E}[X]$ is simply the expected (average) value of X which is very useful in estimating the value of X.
- When $k = 2$, the 2nd moment $\mathbb{E}[X^2]$ is useful in calculating the *variance* of X since $Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

Example 9

We let N be a Poisson random variable with parameter λ which the probability mass function is

$$f_N(k) = P(N = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \qquad k = 0,1,2,\cdots.$$

Find the first and second moments of N.

☺Solution

The first moment of X can be expressed as

$$\mathbb{E}[N] = \sum_{k=0}^{\infty} k f_N(k) = \sum_{k=0}^{\infty} k\left(\frac{e^{-\lambda}\lambda^k}{k!}\right) = \sum_{k=1}^{\infty} k\left(\frac{e^{-\lambda}\lambda^k}{k!}\right) = \sum_{k=1}^{\infty} \frac{e^{-\lambda}\lambda^k}{(k-1)!}$$

$$= \lambda \sum_{k=1}^{\infty} \frac{e^{-\lambda}\lambda^{k-1}}{(k-1)!} \overset{m=k-1}{\widehat{=}} \lambda \sum_{m=0}^{\infty} \frac{e^{-\lambda}\lambda^m}{m!} = \lambda \underbrace{\sum_{m=0}^{\infty} f_N(m)}_{=1} = \lambda.$$

Using similar trick, the second moment of X is found to be

$$\mathbb{E}[N^2] = \sum_{k=0}^{\infty} k^2 f_N(k) = \sum_{k=1}^{\infty} k^2 \left(\frac{e^{-\lambda}\lambda^k}{k!}\right)$$

$$\overset{k^2=k(k-1)+k}{\triangleq} \sum_{k=1}^{\infty} k(k-1)\left(\frac{e^{-\lambda}\lambda^k}{k!}\right) + \sum_{k=1}^{\infty} k\left(\frac{e^{-\lambda}\lambda^k}{k!}\right)$$

$$= \sum_{k=2}^{\infty} \frac{e^{-\lambda}\lambda^k}{(k-2)!} + \sum_{k=1}^{\infty} \frac{e^{-\lambda}\lambda^k}{(k-1)!} = \lambda^2 \sum_{k=2}^{\infty} \frac{e^{-\lambda}\lambda^{k-2}}{(k-2)!} + \lambda \sum_{k=1}^{\infty} \frac{e^{-\lambda}\lambda^{k-1}}{(k-1)!}$$

$$\overset{\substack{m=k-2 \\ p=k-1}}{\triangleq} \lambda^2 \underbrace{\sum_{m=0}^{\infty} \frac{e^{-\lambda}\lambda^m}{m!}}_{=1} + \lambda \underbrace{\sum_{p=0}^{\infty} \frac{e^{-\lambda}\lambda^p}{p!}}_{=1}$$

$$= \lambda^2 + \lambda.$$

Example 10 (Existence of moments)

The probability density functions of two random variables X and Y are

$$f_X(x) = \frac{1}{x^2}, \quad x \geq 1 \quad and \quad f_Y(y) = \frac{4}{x^5}, \quad y \geq 1$$

respectively.

(a) Show that $\mathbb{E}[X^k]$ does not exist for any positive integer k.

(b) Show that $\mathbb{E}[Y^k]$ exists for all positive integer $k \leq 3$.

☺Solution

(a) Note that

$$\mathbb{E}[X^k] = \int_1^\infty x^k \underbrace{\left(\frac{1}{x^2}\right)}_{=f_X(x)} dx = \int_1^\infty x^{k-2} dx$$

$$= \begin{cases} \frac{x^{k-1}}{k-1}\Big|_1^\infty & if \;\; k \geq 2 \\ \ln x \,\Big|_1^\infty & if \;\; k = 1 \end{cases}$$

$$= \begin{cases} 0 & if \;\; k \geq 2 \\ \infty & if \;\; k = 1 \end{cases}.$$

Hence, we conclude that $\mathbb{E}[X^k]$ does not exist for all positive integer k.

(b) Using similar method, one can deduce that

$$\mathbb{E}[Y^k] = \int_1^\infty y^k \underbrace{\left(\frac{4}{y^5}\right)}_{=f_Y(y)} dy = \int_1^\infty 4y^{k-5} dy = \begin{cases} \dfrac{4y^{k-4}}{k-4}\Big|_1^\infty & if \;\; k > 4 \\ 4\ln y\,\Big|_1^\infty & if \;\; k = 4 \\ \dfrac{4y^{k-4}}{k-4}\Big|_1^\infty & if \;\; k < 4 \end{cases}$$

$$= \begin{cases} \infty & if \;\; k > 4 \\ \infty & if \;\; k = 4 \\ \dfrac{4}{4-k} & if \;\; k < 4 \end{cases}$$

Hence, we conclude that $\mathbb{E}[Y^k]$ exists when $k \leq 3$.

Existence of moments and its application

- Since the formula of the k^{th} moment may involve infinite sum (discrete case) or improper integral (continuous case), it is possible that the integral or sum may not *converge* to a finite number so that the k^{th} moment $\mathbb{E}[X^k]$ may not exist. If X is a *non-negative* random variable, $\mathbb{E}[X^k]$ either converges to a finite number (i.e. $\mathbb{E}[X^k] < \infty$) or tends to infinity (i.e. $\mathbb{E}[X^k] = \infty$).
- In the previous example, the non-existence of k^{th} moment is caused by the "heavy tail" of the density function. That is, the distribution assigns a significant probability to larger value of X so that $x^k f(x)$ is large for large value of X.
- As observed from the previous example, $\mathbb{E}[X^k]$ exists for larger range of k if the distribution assigns smaller probability to large value of x (i.e. light tail). This observation is useful in comparing the *tail distributions* between different distribution functions.

Measuring the deviation of a random variable: k^{th} central moment

Although the expected value $\mu = \mathbb{E}[X]$ can indicate the "average value" of X, the actual value of X can be very different from μ. Thus, one would like to measure the "distance" between the actual value of X and expected value of X.

Definition (k^{th} central moment of random variable)

We let X be a random variable. The k^{th} central moment of a random variable, denoted by $\mu_d^{(k)}$, to be the expected value of the k^{th} power of the deviation of the variable from its expected value. That is,

$$\mu_d^{(k)} = \mathbb{E}[(X-\mu)^k] = \begin{cases} \sum_x (x-\mu)^k P(X=x) & if\ X\ is\ discrete \\ \int_{-\infty}^{\infty} (x-\mu)^k f(x)dx & if\ X\ is\ continuous \end{cases}.$$

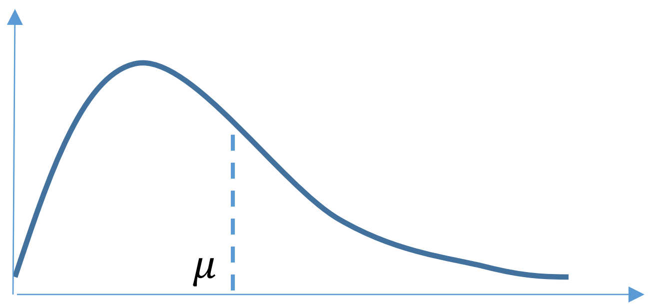
- When $k = 2$, the above quantity reduces to $\mu_d^{(2)} = \mathbb{E}[(X-\mu)^2] = Var(X)$ which is the usual variance of X.

- When $k = 3$, $\mu_d^{(3)}$ is used to measure the *symmetry* of the probability density function $f(x)$ about the mean μ. More formally, we define the *skewness*, denoted by γ_1 to be

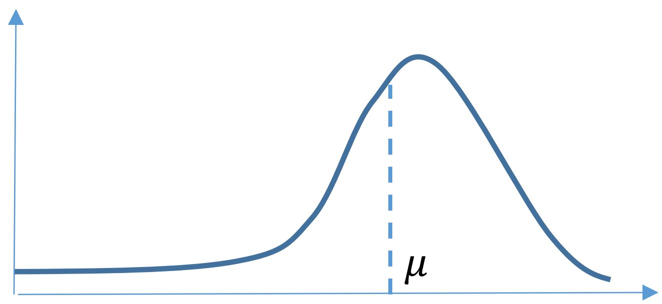
$$\gamma_1 = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mathbb{E}[(X-\mu)^3]}{\sigma^3} = \frac{\mu_d^{(3)}}{\sigma^3}.$$

  ✓ If the probability density function $f(x)$ is *symmetric* about the mean, one can show that $\gamma_1 = 0$.

  ✓ If $\gamma_1 > 0$ (resp. $\gamma_1 < 0$), we say the density function has positive (resp. negative) skew in the sense that the tail of the distribution lies on the right (resp. left) of the distribution (*Note: This is *not* a general rule)



(Positive skew)



(Negative skew)

- When $k = 4$, $\mu_d^{(4)}$ is used to measure *heaviness of tail* of a distribution relative to the normal distribution. More formally, we define the *kurtosis*, denoted by γ_2 to be

$$\gamma_2 = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{\mathbb{E}[(x-\mu)^4]}{\sigma^4} = \frac{\mu_d^{(4)}}{\sigma^4}.$$

  Note: The kurtosis of normal distribution is known to be 3 (see Example 10).

  ✓ If $\gamma_2 < 3$, this implies that the distribution has thinner tails relative to the normal distribution (assign less probability to the values that far away from the mean) since $\left(\frac{X-\mu}{\sigma}\right)^4$ is small.

  ✓ If $\gamma_2 > 3$, this happens when $\frac{X-\mu}{\sigma}$ is large so that X is likely to be far away from the mean. This reveal that the distribution function has heavier tail compared with normal distribution.

Example 11 (Kurtosis of normal random variable)

A random variable X is normally distributed with mean μ and standard deviation σ. Calculate the kurtosis of X.

😊Solution

Using the definition of kurtosis and integration by parts twice, we get

$$\gamma_2 = \frac{\mathbb{E}[(X-\mu)^4]}{\sigma^4} = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \int_{-\infty}^{\infty} \left(\frac{x-\mu}{\sigma}\right)^4 \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) dx$$

$$\stackrel{y=\frac{x-\mu}{\sigma}}{=} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^4 e^{-\frac{y^2}{2}} dy = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^3 d\left(e^{-\frac{y^2}{2}}\right)$$

$$= -\frac{1}{\sqrt{2\pi}} y^3 e^{-\frac{y^2}{2}} \Big|_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} 3y^2 e^{-\frac{y^2}{2}} dy = -\frac{3}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y d\left(e^{-\frac{y^2}{2}}\right)$$

$$= -\frac{3}{\sqrt{2\pi}} y e^{-\frac{y^2}{2}} \Big|_{-\infty}^{\infty} + \frac{3}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \approx 3 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = 3.$$

Example 12

It is given that the probability density function of loss variable X is

$$f_X(x) = \frac{9}{2}x^{-\frac{11}{2}}, \quad x \geq 1.$$

Determine the skewness and kurtosis of X.

😊Solution

For any positive integer k, k^{th} moment of X can be computed as

$$\mathbb{E}[X^k] = \int_1^\infty x^k \left(\frac{9}{2}x^{-\frac{11}{2}}\right) dx = \frac{9}{2}\int_1^\infty x^{k-\frac{11}{2}} dx = \frac{9}{2\left(k-\frac{9}{2}\right)} x^{k-\frac{9}{2}}|_1^\infty$$

$$= \begin{cases} \dfrac{9}{9-2k} & if \;\; k < \dfrac{9}{2} \\ \infty & if \; k > \dfrac{9}{2} \end{cases}.$$

(*Note: k^{th} moment exists for $k \leq 4$)

Then the first four moments are computed as

$$\mu = \mathbb{E}[X] = \frac{9}{7}, \qquad \mathbb{E}[X^2] = \frac{9}{5}, \qquad \mathbb{E}[X^3] = \frac{9}{3} = 3, \qquad \mathbb{E}[X^4] = \frac{9}{1} = 9$$

The standard deviation of X is $\sigma = \sqrt{Var(X)} = \sqrt{\mathbb{E}[X^2] - (\mathbb{E}[X])^2} = 0.383326$. Hence, the skewness and kurtosis can be computed as

$$\gamma_1 = \frac{\mathbb{E}[(X-\mu)^3]}{\sigma^3} = \frac{\mathbb{E}[X^3] - 3\mu\mathbb{E}[X^2] + 3\mu^2\mathbb{E}[X] - \mu^3}{\sigma^3} \approx 5.465944.$$

$$\gamma_2 = \frac{\mathbb{E}[(X-\mu)^4]}{\sigma^4} = \frac{\mathbb{E}[X^4] - 4\mu\mathbb{E}[X^3] + 6\mu^2\mathbb{E}[X^2] - 4\mu^3\mathbb{E}[X] + \mu^4}{\sigma^4} \approx 149.4444.$$

Remark of Example 12

Since $\gamma_1 > 0$, we observe that the right portion of distribution (i.e. $x > \mu = 9/7$) assigns value further from the mean $\mu = 9/7$.

As $\gamma_2 > 3$, so the probability distribution of X has heavier tail relative to normal distribution.

Conditional expectation and its properties

We let X be a random variable which the distribution depends on the occurrence of an event A. Given that the event A has happened, one can define the *conditional expectation* of X as

.

$$\mathbb{E}[X|A] = \begin{cases} \sum_j x_j P(X = x_j|A) & if \;\; X \; is \; discrete \\ \int_{-\infty}^{\infty} x f_{X|A}(x) dx & if \;\; X \; is \; continuous \end{cases}$$

(*Note: Here, $f_{X|A}(x)$ denotes the *conditional* probability density function of X which can be computed using $f_{X|A}(x) = \frac{d}{dx} F_{X|A}(x) = \frac{d}{dx} P(X \le x|A)$.

Suppose that the value of X depends on the value of another random variable Y, then we can define the conditional expected value of X, conditional on $Y = y$ as

$$\mathbb{E}[X|Y] = \mathbb{E}[X|Y = y].$$

Here, we take $A = \{Y = y\}$ as an event

Remark of conditional expectation

- Note that the exact value of the conditional expectation $\mathbb{E}[X|Y]$ depends on the value of random variable Y. Thus, the $\mathbb{E}[X|Y]$ is a function Y and is a random variable in general.
- As an example, we let N be a random variable which has *Poisson distribution* with mean λ. Suppose that λ is also a random variable, then the conditional expected value of N (given λ) is a random variable since
$$\mathbb{E}[N|\lambda] = \lambda.$$
- Another example is as follows: We consider the sum $S = X_1 + X_2 + \cdots + X_N$, where X_i are independent and identically distributed random variables and N, then we have
$$\mathbb{E}[S|N] = N\mathbb{E}[X_1].$$
So that $\mathbb{E}[S|N]$ is also a random variable.

An useful properties of conditional expectation

It is clear that the conditional expectation satisfies all properties that the standard expected value has. In addition, one can establish the following important properties:

We let X and Y be two random variables defined in the common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then we have

$$\mathbb{E}[X] = \mathbb{E}\big[\mathbb{E}[X|Y]\big]$$

$$Var(X) = \mathbb{E}[Var(X|Y)] + Var\big(\mathbb{E}[X|Y]\big).$$

(*Note: Since $\mathbb{E}[X|Y] = \mathbb{E}[X|Y=y]$ depends on the value of the random variable Y, so $\mathbb{E}[X|Y]$ is a *random variable* in this case).

(*Note 2: The first property is also known as *double expectation formula* or *tower property*)

Proof of the properties

To illustrate the idea of the derivation, we focus on the case when both X and Y are discrete. The derivation for the case when X, Y are continuous is similar (just more tedious). To prove the first property, we apply *law of total probability* and obtain

$$\mathbb{E}[X] = \sum_{all\ x_i} x_i P(X = x_i) = \sum_{all\ x_i} x_i \left(\sum_{all\ y_j} P(X = x_i | Y = y_j) P(Y = y_j) \right)$$
$$= \sum_{all\ x_i} \sum_{all\ y_j} x_i P(X = x_i | Y = y_j) P(Y = y_j)$$
$$= \sum_{all\ y_j} \sum_{all\ x_i} x_i P(X = x_i | Y = y_j) P(Y = y_j)$$
$$= \sum_{all\ y_j} P(Y = y_j) \sum_{all\ x_i} x_i P(X = x_i | Y = y_j) = \sum_{j} P(Y = y_j) \underbrace{\mathbb{E}[X | Y = y_j]}_{function\ of\ Y}$$
$$= \mathbb{E}[\mathbb{E}[X|Y]]$$

To prove the second property, we apply the same trick (left as exercise) and obtain

$$\mathbb{E}[X^2] = \sum_i x_i^2 P(X = x_i) = \cdots = \mathbb{E}\big[\mathbb{E}[X^2|Y]\big].$$

Then the variance of X can be computed as

$$Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

$$= \mathbb{E}\big[\mathbb{E}[X^2|Y]\big] - \big(\mathbb{E}\big[\mathbb{E}[X|Y]\big]\big)^2$$

$$=^{(*)} \mathbb{E}\big[\mathbb{E}[X^2|Y]\big] - \big(\mathbb{E}[(\mathbb{E}[X|Y])^2] - Var(\mathbb{E}[X|Y])\big)$$

$$= \mathbb{E}\left[\underbrace{\mathbb{E}[X^2|Y] - (\mathbb{E}[X|Y])^2}_{=Var(X|Y)}\right] + Var(\mathbb{E}[X|Y])$$

$$= \mathbb{E}[Var(X|Y)] + Var(\mathbb{E}[X|Y]).$$

Example 13

A random variable X has exponential distribution with p.d.f. $f_X(x) = \lambda e^{-\lambda x}$ for $x > 0$. Here, λ is also a random variable is uniformly distributed over $[a, b]$, where $0 < a < b$. Compute $\mathbb{E}[X]$.

☺Solution

Given the value of the parameter λ, we have

$$\mathbb{E}[X|\lambda] = \int_0^\infty x\left(\lambda e^{-\lambda x}\right)dx = \frac{1}{\lambda}.$$

Using double expectation formula, we deduce that

$$\mathbb{E}[X] = \mathbb{E}\big[\mathbb{E}[X|\lambda]\big] = \mathbb{E}\left[\frac{1}{\lambda}\right] \overset{\substack{\lambda\ is\ uniformly \\ distributed}}{\widehat{=}} \int_a^b \frac{1}{\lambda}\left(\frac{1}{b-a}\right) d\lambda = \frac{\ln b - \ln a}{b-a}.$$

Example 14 (Expected value and variance of random sum)

We let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with $\mathbb{E}[X_n] = 3$ and $Var(X_n) = 9$ and let N be a random variable which has binomial distribution with parameters $n = 100$ and $p = 0.2$.

We define a random variable S to be

$$S = X_1 + X_2 + \cdots + X_N,$$

(*Note: We take $S = 0$ if $N = 0$.) . We also assume that <u>X_i and N are independent.</u>

Find the values of $\mathbb{E}[S]$ and $Var(S)$.

☺Solution

Since N has binomial distribution, we recall that

$$\mathbb{E}[N] = np = 20, \qquad Var(N) = np(1-p) = 16.$$

We first compute $\mathbb{E}[S]$. By applying double expectation formula, we obtain

$$\mathbb{E}[S] = \mathbb{E}\big[\mathbb{E}[S|N]\big] \ldots (*)$$

Given the value of N, the number of terms of S is fixed so that

$$\mathbb{E}[S|N] = \mathbb{E}[X_1 + X_2 + \cdots + X_N|N] = \sum_{i=1}^{N} \mathbb{E}[X_i|N] = \sum_{i=1}^{N} \underbrace{\mathbb{E}[X_i]}_{=3} = 3N.$$

Then we can deduce from equation $(*)$ that

$$\mathbb{E}[S] = \mathbb{E}\big[\mathbb{E}[S|N]\big] = \mathbb{E}[3N] = 3(20) = 60.$$

Next, we compute $Var(S)$. Using similar method, we get

$$Var(S) = \mathbb{E}[Var(S|N)] + Var(\mathbb{E}[S|N]).$$

By assumption, $X_i s$ are independent, so we have

$$Var(S|N) = \sum_{i=1}^{N} Var(X_i|N) = NVar(X_i|N) = NVar(X_i) = 9N.$$

Combining the result, we deduce that

$$Var(S) = \mathbb{E}\left[\underbrace{9N}_{Var(S|N)}\right] + Var\left(\underbrace{3N}_{\mathbb{E}[S|N]}\right) = 9(20) + 9(16) = 324.$$

Moment generating function (MGF)

In probability theory, the moment generating function is very useful in studying the moments and probability distribution of some complex random variables.

For any random variable X, the MGF of X, denoted by $M_X(t)$, is defined as

$$M_X(t) = \mathbb{E}[e^{tX}].$$

where $t \in (-\infty, \infty)$ is any real number.

Example 15 (Some examples of MGFs)

- If X has exponential distribution with mean θ, then

$$M_X(t) = \mathbb{E}[e^{tX}] = \int_0^\infty e^{tx}\left(\frac{1}{\theta}e^{-\frac{x}{\theta}}\right)dx$$

$$= \frac{1}{\theta}\int_0^\infty e^{-\left(\frac{1}{\theta}-t\right)x}dx \overset{t\neq 1/\theta}{\triangleq} -\frac{1}{\theta\left(\frac{1}{\theta}-t\right)}e^{-\left(\frac{1}{\theta}-t\right)x}\Big|_0^\infty = \begin{cases} \infty & if\ \frac{1}{\theta}-t \le 0 \\ \frac{1}{1-\theta t} & if\ \frac{1}{\theta}-t > 0 \end{cases}.$$

(*Note: When $t = \frac{1}{\theta}$, $\mathbb{E}[e^{tX}] = \frac{1}{\theta}\int_0^\infty 1dx = \infty$.)

- If X has Poisson distribution with mean λ, we have

$$M_X(t) = \mathbb{E}[e^{tX}] = \sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=0}^{\infty} \frac{(e^t\lambda)^k e^{-\lambda}}{k!} \overset{y=e^t\lambda}{\triangleq} e^{-\lambda} \sum_{k=0}^{\infty} \frac{y^k}{k!}$$

$$= e^{-\lambda} e^y = e^{e^t\lambda - \lambda} = e^{\lambda(e^t - 1)}.$$

- If Z has standard normal distribution (with $\mu = 0$ and $\sigma = 1$), we have

$$M_Z(t) = \mathbb{E}[e^{tZ}] = \int_{-\infty}^{\infty} e^{tz} \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \right) dz = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2} + tz} dz$$

$$= e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-t)^2}{2}} dx \overset{y=z-t}{\triangleq} e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = e^{\frac{t^2}{2}}.$$

(General case) If X has normal distribution with mean μ and standard deviation σ ($X \sim N(\mu, \sigma^2)$), we can write X as $X = \mu + \sigma Z$, where Z has standard normal distribution. Then we have

$$M_X(t) = \mathbb{E}[e^{tX}] = \mathbb{E}\left[e^{t(\mu + \sigma Z)}\right] = e^{\mu t} \mathbb{E}\left[e^{(\sigma t)Z}\right] = e^{\mu t} M_Z(\sigma t) = e^{\mu t} e^{\frac{(\sigma t)^2}{2}}$$

$$= e^{\mu t + \frac{\sigma^2 t^2}{2}}.$$

Why do we need MGF?

There are two important applications of MGF.

1. *Finding moment of a random variable*

   We let X be a random variable and $M_X(t)$ be the corresponding MGF. By differentiating the MGF with respect to t, we have

$$M_X'(t) = \frac{d}{dt}\mathbb{E}[e^{tX}] = \frac{d}{dt}\int_{-\infty}^{\infty} e^{tx} f(x)dx = \int_{-\infty}^{\infty}\left(\frac{d}{dt}e^{tx}f(x)\right)dx$$

$$= \int_{-\infty}^{\infty} xe^{tx} f(x)dx = \mathbb{E}[Xe^{tX}].$$

   By repeating the differentiation, we get

$$M_X^{(k)}(t) = \frac{d^k}{dt^k}\mathbb{E}[e^{tX}] = \mathbb{E}[X^k e^{tX}], \quad k = 1,2,\ldots..$$

   By taking $t = 0$ (assuming $M_X^{(k)}$ exists near $t = 0$), we get

$$M_X^{(k)}(0) = \mathbb{E}[X^k].$$

   So k^{th} moment of X can be found by differentiating MGF for k times.

Example 16 (Example 11 revisited)

Suppose that X has normal distribution with mean μ and variance σ^2. Find the kurtosis of X, i.e. $\mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]$.

☺Solution

We take $Y=\frac{X-\mu}{\sigma}$. Since Y has standard normal distribution, the MGF of Y is $M_Y(t)=e^{\frac{t^2}{2}}$. By differentiating MGF for 4 times, we get

$$M_Y'(t)=te^{\frac{t^2}{2}},\qquad M_Y^{(2)}(t)=e^{\frac{t^2}{2}}+t^2e^{\frac{t^2}{2}},\qquad M_Y^{(3)}(t)=3te^{\frac{t^2}{2}}+t^3e^{\frac{t^2}{2}},$$

$$M_Y^{(4)}(t)=3e^{\frac{t^2}{2}}+6t^2e^{\frac{t^2}{2}}+t^4e^{\frac{t^2}{2}}.$$

Then the kurtosis of X can now be computed as

$$\gamma_2=\mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]=\mathbb{E}[Y^4]=M_Y^{(4)}(0)=3.$$

So we can find the kurtosis without doing integration by parts.

Example 17

The random variables $X_1, X_2, \dots, X_n$ are independent and identically distributed with probability density function

$$f(x) = \frac{e^{-\frac{x}{\theta}}}{\theta}, \quad x \geq 0.$$

Determine $\mathbb{E}[\bar{X}^4]$. Here, $\bar{X}$ denotes the mean of $X_1, X_2, \dots., X_n$.

☺Solution

The MGF of X_i is $M_{X_i}(t) = \mathbb{E}[e^{tX}] = \frac{1}{1-\theta t}$ since X is exponentially distributed.

So the MGF of $\bar{X}$ is given by

$$M_{\bar{X}}(t) = \mathbb{E}\left[e^{\frac{t}{n}(X_1+X_2+\cdots+X_n)}\right] = \prod_{i=1}^{n} \mathbb{E}\left[e^{\left(\frac{t}{n}\right)X_i}\right] = \prod_{i=1}^{n} M_{X_i}\left(\frac{t}{n}\right) = \prod_{i=1}^{n}\left(\frac{1}{1-\theta t/n}\right)$$
$$= \left(\frac{n}{n-\theta t}\right)^n.$$

Since $M_{\bar{X}}^{(4)}(t) = \theta^4 n(n+1)(n+2)(n+3)\frac{n^n}{(n-\theta t)^{n+4}}$, so we have $\mathbb{E}[\bar{X}^4] = M_{\bar{X}}^{(4)}(0) = \frac{\theta^4 n(n+1)(n+2)(n+3)n^n}{n^{n+4}} = \frac{\theta^4(n+1)(n+2)(n+3)}{n^3}$.

2. *Determine the probability distribution of X*

Another important property of MGF is that the MGF can uniquely determine the probability distribution of a random variable.

Property of MGF (One-to-one correspondence between MGF and distribution)

Let X, Y be two random variables. If $M_X(t) = M_Y(t)$ for all t in the open interval containing 0, then X and Y have *same* probability distribution.

As an example, suppose that the MGF of a random variable X be $M_X(t) = e^{3t+2t^2}$. One can observe that the MGF of X is simply the MGF of normal random variable by taking $\mu = 3$ and $\sigma = 2$. That is,

$$M_X(t) = e^{3t+2t^2} = \underbrace{e^{3t+\frac{2^2t^2}{2}} = e^{\mu t+\frac{\sigma^2 t^2}{2}}}_{MGF\ of\ Y \sim N(3,2^2)}.$$

Since X and $N(3, 2^2)$ has same MGF, we can conclude from the above property that $X \sim N(3, 2^2)$.

Example 18

We let $X_1, X_2, \dots, X_n$ be n independent random variables with $X_k \sim N\left(\mu_k, \sigma_k^2\right)$ for $k = 1,2, \dots., n$. What is the probability distribution of $X_1 + X_2 + \cdots + X_n$?

😊Solution

The MGF of random variable X_k ($X_k \sim N\left(\mu_k, \sigma_k^2\right)$) is given by

$$M_{X_k}(t) = e^{\mu_k t + \frac{\sigma_k^2 t^2}{2}}.$$

Then the MGF of $S = X_1 + X_2 + \cdots + X_n$ can be computed as

$$M_S(t) = \mathbb{E}\left[e^{t(X_1+X_2+\cdots+X_n)}\right] \overset{X_k s \text{ are independent}}{\widehat{=}} \mathbb{E}[e^{tX_1}]\mathbb{E}[e^{tX_2}] \dots \mathbb{E}[e^{tX_n}]$$
$$= \left(e^{\mu_1 t + \frac{\sigma_1^2 t^2}{2}}\right)\left(e^{\mu_2 t + \frac{\sigma_2^2 t^2}{2}}\right) \dots \left(e^{\mu_n t + \frac{\sigma_n^2 t^2}{2}}\right) = e^{\left(\sum_{k=1}^n \mu_k\right)t + \frac{\left(\sum_{k=1}^n \sigma_k^2\right)t^2}{2}}.$$

The function of R.H.S is the MGF of normal random variable Y with $\mu = \sum_{k=1}^n \mu_k$ and $\sigma^2 = \sum_{k=1}^n \sigma_k^2$. Thus, we conclude that $X \sim N(\mu, \sigma^2)$.

Example 19 (Harder)

We let $X_1, X_2, X_3, \dots, X_{100}$ be independent and identically distributed random variables such that the sum $S = X_1 + X_2 + \cdots + X_{100}$ is Poisson distributed with mean 350.

Assuming that the MGF of X_i exists, find the probability that X_1 is less than 3.5.

☺Solution

Since $X_1, X_2, \dots, X_{100}$ are identically distributed, we have $M_{X_1}(t) = M_{X_2}(t) = \cdots = M_{X_{100}}(t)$. Together with independence assumption, we get

$$e^{350(e^t-1)} = M_S(t) = \mathbb{E}\left[e^{t(X_1+\cdots+X_{100})}\right] = \prod_{i=1}^{100} \underbrace{\mathbb{E}[e^{tX_i}]}_{M_{X_i}(t)} = \left(M_{X_1}(t)\right)^{100}$$

This implies $M_{X_1}(t) = e^{3.5(e^t-1)}$. It follows that X_1 has Poisson distribution with mean $\lambda = 3.5$. Hence, we deduce that

$$P(X_1 < 3.5) = \sum_{k=0}^{3} \underbrace{\frac{e^{-3.5}(3.5)^k}{k!}}_{P(X_1=k)} \approx 0.536633.$$

The following property reveals that MGF is also useful in predicting the *limiting distribution* of some random variables:

Property of MGF 2

We let $Y_1, Y_2, Y_3, \ldots$ be a sequence of random variables such that $\lim_{n\to\infty} M_{Y_n}(t) = M_Y(t)$ for some random variable Y. Then

$$\lim_{n\to\infty} F_{Y_n}(y) = F_Y(y),$$

for all y where $F_Y(y)$ is continuous.

Remark

- In practice, the statement "$\lim_{n\to\infty} F_{Y_n}(y) = F_Y(y)$" means that of X and we can approximate the distribution of X_n by the distribution of X.
- Two well-known examples about this approximation are
  1. Normal approximation to binomial distribution
  2. Central Limit Theorem (i.e. $\bar{X} = \frac{X_1+X_2+\cdots+X_n}{n}$ is approximately normal when $n \to \infty$))

Example 20 (Normal approximation to binomial distribution)

A random variable X_n has binomial distribution with parameter n and p. In this example, we would like to demonstrate that the distribution of X_n converges to normal distribution when $n \to \infty$.

Using moment generating function, show that

$$\frac{X_n - \mathbb{E}[X_n]}{\sqrt{Var(X_n)}} = \frac{X_n - np}{\sqrt{np(1-p)}} \to^d Z \sim N(0,1), \qquad when\, n \to \infty.$$

☺Solution

We let $Y_n = \frac{X_n - np}{\sqrt{np(1-p)}}$, the MGF of Y_n can be expressed as

$$M_{Y_n}(t) = \mathbb{E}[e^{tY_n}] = \mathbb{E}\left[e^{t\left(\frac{X_n - np}{\sqrt{np(1-p)}}\right)}\right] = e^{\frac{-npt}{\sqrt{np(1-p)}}}\mathbb{E}\left[e^{\frac{t}{\sqrt{np(1-p)}}X_n}\right]$$

$$= e^{\frac{-npt}{\sqrt{np(1-p)}}} M_{X_n}\left(\frac{t}{\sqrt{np(1-p)}}\right) = e^{\frac{-npt}{\sqrt{np(1-p)}}}\left(pe^{\frac{t}{\sqrt{np(1-p)}}} + 1 - p\right)^n$$

$$x^n = e^{\ln x^n} = e^{n \ln x} \quad \stackrel{\triangle}{=} \quad e^{\frac{-npt}{\sqrt{np(1-p)}} + n\ln\left(pe^{\frac{t}{\sqrt{np(1-p)}}} + 1 - p\right)} \quad \ldots\ldots (*)$$

To proceed, we consider the Taylor series of e^x and $\ln(1+x)$, i.e.

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \quad and \quad \ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots.$$

Using these formula, we deduce that

$$M_{Y_n}(t) = e^{\frac{-npt}{\sqrt{np(1-p)}} + n\ln\left(p\left[1 + \frac{t}{\sqrt{np(1-p)}} + \frac{1}{2!}\left(\frac{t}{\sqrt{np(1-p)}}\right)^2 + O\left(\frac{1}{n^{3/2}}\right)\right] + 1 - p\right)}$$

$$= e^{\frac{-npt}{\sqrt{np(1-p)}} + n\ln\left(1 + \frac{pt}{\sqrt{np(1-p)}} + \frac{pt^2}{2(np(1-p))} + O\left(\frac{1}{n^{3/2}}\right)\right)}$$

$$= e^{\frac{-npt}{\sqrt{np(1-p)}} + n\left[\frac{pt}{\sqrt{np(1-p)}} + \frac{pt^2}{2(np(1-p))} + O\left(\frac{1}{n^{3/2}}\right) - \frac{1}{2}\left(\frac{pt}{\sqrt{np(1-p)}} + \frac{pt^2}{2(np(1-p))} + O\left(\frac{1}{n^{3/2}}\right)\right)^2 + O\left(\frac{1}{n^{3/2}}\right)\right]}$$

$$= e^{\frac{npt^2}{2np(1-p)} - \frac{np^2t^2}{np(1-p)} + nO\left(\frac{1}{n^{3/2}}\right)} = e^{\frac{t^2}{2} + nO\left(\frac{1}{n^{3/2}}\right)}.$$

(*Here, $O\left(\frac{1}{n^{3/2}}\right)$ denotes the collection of all terms involving $\frac{1}{n^{\frac{3}{2}}}, \frac{1}{n^2}, \frac{1}{n^{\frac{5}{2}}}, \ldots$)

By taking $n \to \infty$ and noting that $nO\left(\frac{1}{n^{3/2}}\right) \to 0$, we deduce that

$$\lim_{n\to\infty} M_{Y_n}(t) = e^{\frac{t^2}{2}}.$$

Note that $e^{\frac{t^2}{2}}$ is simply the MGF of standard normal distribution, therefore we conclude that $Y_n = \frac{X_n - np}{\sqrt{np(1-p)}} \to^d Z \sim N(0,1)$.

Remark of Example 20

Therefore, it follows that X_n is approximately normal with mean np and variance $np(1-p)$ if n is sufficiently large.

To see the efficiency of the approximation, we consider the following scenario:

Suppose that a random variable N has binomial distribution with parameters $n = 100$ and $p = 0.6$ and we would like to compute the probability $P(45 \leq N \leq 65)$

- If we use ad-hoc approach, the required probability can be computed as

$$P(45 \leq N \leq 65) = \sum_{k=45}^{65} C_k^{100}(0.6)^k(1-0.6)^{100-k} = 0.868782.$$

- If we adopt normal approximation, the required probability is found to be

$$P(45 \leq N \leq 65) \overset{(*)}{\cong} P(44.5 \leq N \leq 65.5)$$

$$\overset{\substack{np=60 \\ np(1-p)=24}}{\cong} P\left(\frac{44.5-60}{\sqrt{24}} \leq Z \leq \frac{65.5-60}{\sqrt{24}}\right)$$

$$= P(-3.16392 \leq Z \leq 1.122683) \approx 0.868436.$$

((*) This technique is known as *end correction*.)

Example 21 (A special case of central limit theorem)

We let $X_1, X_2, \ldots$ be a sequence of independent and identically distribution (i.i.d.) random variables which each X_i has exponential distribution with mean θ.
Show that when $n \to \infty$, the sum $X_1 + X_2 + \cdots + X_n$ is approximately normal with mean $n\theta$ and variance $n\theta^2$.
☺Solution

We let $Y_n = \frac{X_1+X_2+\cdots+X_n-n\theta}{\sqrt{n\theta^2}}$ (standard score). Our goal is to show $\lim_{n\to\infty} M_{Y_n}(t) = M_Z(t)$, where Z is standard normal random variable (i.e. $Z \sim N(0,1)$).

Note that the MGF of X_i is $M_{X_i}(t) = \frac{1}{1-\theta t}$. So we have

$$M_{Y_n}(t) = \mathbb{E}[e^{tY_n}] = \mathbb{E}\left[e^{t\left(\frac{X_1+X_2+\cdots+X_n-n\theta}{\sqrt{n\theta^2}}\right)}\right] = e^{-\sqrt{n}t}\prod_{i=1}^{n}\underbrace{\mathbb{E}\left[e^{\frac{t}{\theta\sqrt{n}}X_i}\right]}_{M_{X_i}\left(\frac{t}{\theta\sqrt{n}}\right)}$$

$$= e^{-t\sqrt{n}}\prod_{i=1}^{n}\frac{1}{1-\theta\left(\frac{t}{\theta\sqrt{n}}\right)} = e^{-t\sqrt{n}}\left(1-\frac{t}{\sqrt{n}}\right)^{-n} = e^{-t\sqrt{n}-n\ln\left(1-\frac{t}{\sqrt{n}}\right)}.$$

To compute the limit $\lim_{n\to\infty} M_{Y_n}(t)$, we recall the Taylor series of $\ln(1+x)$:

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{x^n}{n}.$$

Using this fact, the limit can be computed as

$$\lim_{n\to\infty} M_{Y_n}(t) = \lim_{n\to\infty} e^{-t\sqrt{n} - n\left(-\frac{t}{\sqrt{n}} - \frac{t^2}{2n} - \frac{t^3}{3n^{\frac{3}{2}}} - \frac{t^4}{4n^2} - \cdots\right)} = e^{\lim_{n\to\infty}\left(\frac{t^2}{2} + \frac{t^3}{3n^{\frac{1}{2}}} + \frac{t^4}{4n} + \cdots\right)}$$

$$= e^{\frac{t^2}{2}}.$$

(*Note: The last step can be justified by *Abel's limit theorem*, see MATH3033)

Note that $e^{\frac{t^2}{2}}$ is the MGF of standard normal distribution.

Hence we conclude that $Y_n = \frac{X_1 + X_2 + \cdots + X_n - n\theta}{\sqrt{n\theta^2}}$ is approximately standard normal when $n \to \infty$ and the sum $X_1 + X_2 + \cdots + X_n$ is approximately normal (with mean $n\theta$ and variance $n\theta^2$) when $n \to \infty$.

Some comment about moment generating function

Although the moment generating function is very efficient in studying the distribution of a random variable, it has several limitations.

- Firstly, it is possible that the moment generating function may not exist for some random variable. For example, we consider a random variable with following probability density function:
$$f_X(x) = \frac{1}{x^2}, \qquad x > 1.$$
One can show that for $t > 0$, we have
$$M_X(t) = \mathbb{E}[e^{tX}] = \int_1^\infty \frac{e^{tx}}{x^2} dx \overset{\lim_{n\to\infty} \frac{e^{tx}}{x^2} = \infty > 0}{\cong} +\infty.$$
So that the moment generating function does not exist.
- Secondly, there is no explicit formula which allows us to get the probability distribution (p.d.f. or c.d.f) of a random variable directly using the moment generating function $M_X(t)$.

Probability generating function (for discrete random variable only)

We let X be a discrete random variable which takes values $0,1,2,3,\ldots$ (non-negative integer value). The probability generating function of X, denoted by $P_X(t)$, is defined as

$$P_X(t) = \mathbb{E}[t^X] = \sum_{k=0}^{\infty} \underbrace{P(X = k)}_{p_k} t^k.$$

Example 22

- If X is a binomial random variable with parameter n and p, then the probability generating function of X is

$$P_X(t) = \sum_{k=0}^{n} t^k \underbrace{[C_k^n p^k (1-p)^{n-k}]}_{p_k = P(X=k)} = \sum_{k=0}^{n} C_k^n (tp)^k (1-p)^{n-k}$$
$$= [tp + (1-p)]^n = [1 + (t-1)p]^n.$$

- If X is a Poisson random variable with parameter λ, then the probability generating function of X is

$$P_X(t) = \sum_{k=0}^{n} t^k \underbrace{\left(\frac{\lambda^k e^{-\lambda}}{k!}\right)}_{P(X=k)} = e^{-\lambda} \sum_{k=0}^{n} \frac{(\lambda t)^k}{k!} = e^{-\lambda} e^{\lambda t} = e^{\lambda(t-1)}.$$

Importance of PGF in analyzing discrete random variable

1. *PGF can generate the probability distribution for random variable*

   We let $P_X(t) = \sum_{n=0}^{\infty} p_n t^n$ be the PGF of random variable X.

   ✓ By taking $t = 0$, we get $P_X(0) = p_0$.

   ✓ By differentiating $P_X(t)$ with respect to t for k times, we get

   $$P_X^{(k)}(t) = k!\, p_k + (??)t + (??)t^2 + \cdots.$$

   By taking $t = 0$, we let $P_X^{(k)}(0) = k!\, p_k$

   $$\Rightarrow P(X = k) = p_k = \frac{P_X^{(k)}(0)}{k!}.$$

2. *PGF can generate the moments of X*

By differentiating $P_X(t)$ with respect to t, we get

$$P_X'(t) = \sum_{k=1}^{\infty} k p_k t^{k-1} = \sum_{k=0}^{\infty} k p_k t^{k-1} = \mathbb{E}[X t^{X-1}]$$

Put $t = 1$, we get $P_X'(1) = \mathbb{E}[X]$.

By differentiating $P_X'(t)$ one more time, we get

$$P_X''(t) = \sum_{k=2}^{\infty} k(k-1) p_k t^{k-2} = \sum_{k=0}^{\infty} k(k-1) p_k t^{k-2} = \mathbb{E}[X(X-1)t^{X-2}].$$

Put $t = 1$, we get $P_X''(1) = \mathbb{E}[X(X-1)]$.

By repeating the process, we get

$$P_X^{(k)}(1) = \mathbb{E}[X(X-1)\dots(X-k+1)].$$

3. *PGF can determine the probability distribution*

Similar to PGF, PGF can also uniquely determine the probability distribution of a random variable.

Example 23

We let $N_1, N_2, \dots, N_n$ be n independent random variable such that each N_i has Poisson distribution with mean λ_i.

Find the distribution of $N = N_1 + N_2 + \cdots + N_n$.

☺Solution

Using the result in Example, the probability generating function of S is

$$P_N(t) = \mathbb{E}[t^N] = \mathbb{E}[t^{N_1+N_2+\cdots+N_n}] = \prod_{k=1}^{n} \mathbb{E}[t^{N_k}] = \prod_{k=1}^{n} P_{N_k}(t) = \prod_{k=1}^{n} e^{\lambda_k(t-1)}$$

$$= e^{\left(\sum_{k=1}^{n}\lambda_k\right)(t-1)} \overset{\lambda=\sum_{k=1}^{n}\lambda_k}{\triangleq} e^{\lambda(t-1)},$$

By direct differentiation, one can deduce that

$$P(N=k) = \frac{P_N^{(k)}(0)}{k!} = \frac{1}{k!}\left(\lambda^k e^{\lambda(t-1)}\right)\Big|_{t=0} = \frac{\lambda^k e^{-\lambda}}{k!}.$$

So we conclude that N has Poisson distribution with mean $\lambda = \sum_{k=1}^{n}\lambda_k$.

Example 24

We let N be a Poisson random variable with mean λ. We assume that λ is a random variable which has exponential distribution with mean $\theta = 1$.

Compute the probability $P(N > 1)$.

☺Solution

We first obtain the PGF of N. Using double expectation formula, we get

$$P_N(t) = \mathbb{E}[t^N] = \mathbb{E}\big[\mathbb{E}[t^N|\lambda]\big] = \mathbb{E}\big[P_{N|\lambda}(t)\big] = \mathbb{E}\big[e^{\lambda(t-1)}\big]$$
$$= \int_0^\infty e^{x(t-1)}\left(\frac{1}{1}e^{-x(1)}\right)dx = \int_0^\infty e^{x(t-2)}dx = \left[\frac{1}{t-2}e^{x(t-2)}\right]_0^\infty$$
$$= \frac{1}{2-t}, \qquad for\ t < 2.$$

So the required probability can be computed as

$$P(N > 1) = 1 - \underbrace{P(N = 0)}_{P_N(0)} - \underbrace{P(N = 1)}_{\frac{P_N'(1)}{1!}} = \frac{1}{4}.$$

Characteristic function: An introduction (Optional)

One limitation of moment generating function is that it may not exist for some random variable X (such as lognormal distribution or Pareto distribution). Hence, one needs to develop a better function which has **(i)** similar functions as MGF and **(ii)** is guaranteed to exist for all random variables. This function is known as *characteristic function*.

We let X be a random variable. The characteristic function of X, denoted by $\varphi_X(t)$, is defined as

$$\varphi_X(t) = \mathbb{E}\left[e^{itX}\right] = \underbrace{\mathbb{E}[\cos tX]}_{\in[-1,1]} + i\,\underbrace{\mathbb{E}[\sin tX]}_{\in[-1,1]},$$

where $i = \sqrt{-1}$. In other words, $\varphi(t)$ is a *complex valued function*.

Since $\left|e^{itX}\right| = \sqrt{(\cos tX)^2 + (\sin tX)^2} = 1$, so the expectation does not tend to ∞ so that the characteristic function always exist. So one may use it to study some distributions which the moment generating function does not exist.

Properties of characteristic function: A summary

- Computing moments of X

  If k^{th} moment $\mathbb{E}[X^k]$ exists, then $\mathbb{E}[X^k]$ can be expressed as
  $$\mathbb{E}[X^k] = (-i)^k \frac{d^k}{dt^k} \varphi_X(t)|_{t=0}.$$

- Finding probability distribution of X

  In fact, the characteristic function allows us to obtain the probability of X directly using characteristic function. Such formula is known as *Levy's inversion formula*.

  ✓ For any random variable X, we have
  $$\lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi_X(t)dt = P(a < X < b) + \frac{P(X = a) + P(X = b)}{2}$$

  ✓ If X is continuous random variable and $\int_{-\infty}^{\infty} |\varphi_X(t)|dt < \infty$, then we have
  $$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t)dt,$$
  where $f_X(x)$ is the probability density function of X.