7. Properties of expectation

7.1. **Expectation of sums of random variables.** Recall that if X is discrete with p.m.f. p, then

$$E(X) = \sum_{x:p(x)>0} xp(x).$$

If X is continuous with p.d.f. f, then

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

Moreover, if $X_1, ..., X_n$ are discrete with joint p.m.f. p and g is a real-valued (Borel) function on $\mathbb{R}^n$, then

$$E(g(X_1, ..., X_n)) = \sum_{x_1,...,x_n:p(x_1,...,x_n)>0} g(x_1, ..., x_n)p(x_1, ..., x_n).$$

If $X_1, ..., X_n$ are jointly continuous with p.d.f. f, then

$$E(g(X_1, ..., X_n)) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, ..., x_n)f(x_1, ..., x_n)dx_1 \cdots dx_n.$$

Moreover, if g is nonnegative, then

$$E(g(X_1, ..., X_n)) = \int_0^{\infty} P(g(X_1, ..., X_n) > t)dt.$$

If g takes values on nonnegative integers, then

$$E(g(X_1, ..., X_n)) = \sum_{k=0}^{\infty} P(g(X_1, ..., X_n) > k).$$

Proposition 7.1. *Let $X_1, ..., X_n$ be random variables. Then,*

$$E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n).$$

Proof. It suffices to consider the case $n = 2$. We prove this proposition by assuming that X_1, X_2 are jointly continuous here, while the proof of the general case requires a generalized definition of expectations. Let f be the joint p.d.f. of X_1, X_2. Then,

$$\begin{aligned}
E(X_1 + X_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 + x_2)f(x_1, x_2)dx_1dx_2 \\
&= \int_{-\infty}^{\infty} x_1f(x_1, x_2)dx_1dx_2 + \int_{-\infty}^{\infty} x_2f(x_1, x_2)dx_1dx_2 \\
&= \int_{-\infty}^{\infty} x_1f_{X_1}(x_1)dx_1 + \int_{-\infty}^{\infty} x_2f_{X_2}(x_2)dx_2 \\
&= E(X_1) + E(X_2).
\end{aligned}$$

□

Example 7.1. Let X be a hypergeometric random with parameters n, N, m. Recall that if n balls are randomly selected from a box of N balls with m white and $N - m$ black, then X denotes the number of white balls selected. Let $w_1, ..., w_m$ be the indices of all white balls and let A_i be the event that w_i is selected for $1 \le i \le n$. Then, we may write

$$X = X_1 + \cdots + X_m,$$

where

$$X_i(s) = \begin{cases} 1 & \text{if } s \in A_i \\ 0 & \text{if } s \in A_i^c \end{cases}.$$

It is easy to see that

$$E(X_i) = P(A_i) = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}.$$

Using the above proposition, we obtain

$$E(X) = \sum_{i=1}^{m} E(X_i) = \sum_{i=1}^{m} \frac{n}{N} = \frac{nm}{N}.$$

Example 7.2 (Coupon collecting problem). Recall the problem of collecting coupons. Suppose that there are N different coupons distributed uniformly in a box. Each time, a coupon is selected at random with replacement. Let T be the random variable denoting the first time that a complete set of N different coupons is finished. After the completion of i distinct types, let T_i be the additional time that is required for the appearance of a new type of coupon other than those collected. Clearly, $T_0 = 1$ and T_i is geometric with parameter $1 - i/N$ for $1 \leq i < N$. If the collection starts with no coupon, then

$$T = T_0 + T_1 + \cdots + T_{N-1}.$$

This implies

$$E(T) = \sum_{i=0}^{N-1} E(T_i) = 1 + \sum_{i=1}^{N-1} \frac{N}{i} = N \sum_{i=1}^{N} \frac{1}{i} \sim N \log N \quad \text{as } N \to \infty.$$

Example 7.3. Consider the quick-sort algorithm. Suppose that we have a set of n distinct numbers, say $x_1, ..., x_n$, with $n > 2$. We plan to arrange them in increasing order. The quick-sort algorithm works in the following way. First, randomly select a number, say x_i, and compare x_i and x_j for all $j \neq i$. If $x_j < x_i$, put x_j in the bracket that left to x_i. If $x_j > x_i$, put x_j in the bracket that right to x_i. Next, randomly select a number other than x_i, say x_j and do the same thing as before for the numbers that are in the same bracket as x_j. Next, randomly select a number other than x_i, x_j and follow the same process as before. Iterate such a sorting until no bracketed set contains more than one number. Denote by X the total number of direct comparisons made before the algorithm stops. To study X, let $A_{i,j}$ be the event that a direct comparison of $x_{(i)}$ and $x_{(j)}$ is made, where $x_{(1)} < \cdots < x_{(n)}$ is the increasing sequence of $x_1, ..., x_n$. Define

$$I_{i,j}(s) = \begin{cases} 1 & \text{if } s \in A_{i,j} \\ 0 & \text{if } s \notin A_{i,j} \end{cases}.$$

Then, $X = \sum_{j=2}^{n} \sum_{i=1}^{j-1} I_{i,j}$ and

$$E(X) = \sum_{j=2}^{n} \sum_{i=1}^{j-1} E(I_{i,j}) = \sum_{j=2}^{n} \sum_{i=1}^{j-1} P(A_{i,j}).$$

Observe that, for $i < j$, if $x_{(k)}$ is selected with $k < i$ or $k > j$, then the numbers $x_{(i)}, ..., x_{(j)}$ are addressed in the same bracket. Since $i < j$, such a bracketed set contains more than 1 number and the sorting keeps going on. If $x_{(k)}$ is selected with $i < k < j$, then $x_{(i)}$ and $x_{(j)}$ will be separated into different bracketed sets and, thus, $A_{i,j}$ can not happen. If $x_{(k)}$ is selected with

$k \in \{i, j\}$, then $A_{i,j}$ happens. Since each number is selected at random, one may prove by induction that

$$P(A_{i,j}) = \frac{2}{j-i+1}.$$

This implies

$$E(X) = \sum_{j=2}^{n} \sum_{i=1}^{j-1} \frac{2}{j-i+1}$$

Note that

$$\ln j - 1 < \ln j - \ln 2 = \int_2^j \frac{dt}{t} \leq \sum_{i=1}^{j-1} \frac{1}{j-i+1} \leq \int_1^j \frac{dt}{t} = \ln j$$

and

$$n \ln n - n + 1 = \int_1^n \ln t dt \leq \sum_{j=2}^{n} \ln j \leq \ln n + \int_1^n \ln t dt = n \ln n - n + \ln n + 1 \leq n \ln n.$$

Consequently, putting all estimations together yields

$$|E(X) - 2n \log n| \leq 3n, \quad \forall n > 2.$$

7.2. **Covariances and correlations.**

Definition 7.1. The **covariance** of two random variables X, Y is defined by

$$\mathrm{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))].$$

Remark 7.1. Note that the variance must be nonnegative but the covariance can be negative.

Proposition 7.2. *Let X, Y, Z be random variables.*

1. $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X) = E(XY) - E(X)E(Y)$.
2. $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$.
3. $\mathrm{Cov}(aX, Y) = a\mathrm{Cov}(X, Y)$.
4. $\mathrm{Cov}(X + Y, Z) = \mathrm{Cov}(X, Z) + \mathrm{Cov}(Y, Z)$.

Corollary 7.3. *Let $X_1, ..., X_n$ and $Y_1, ..., Y_m$ be random variables. Then,*

$$\mathrm{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i,j} \mathrm{Cov}(X_i, Y_j)$$

and

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j).$$

Proposition 7.4. *Let X, Y be independent random variables and g, h be real-valued functions. Then,*

$$E(g(X)h(Y)) = E(g(X))E(h(Y)).$$

In particular, $\mathrm{Cov}(g(X), h(Y)) = 0$.

Proof. We consider the case that X, Y are jointly continuous in this proof. Let f be the joint density function of X, Y. Then, $f(x,y) = f_X(x)f_Y(y)$, which implies

$$\begin{aligned}
E(g(X)h(Y)) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f(x,y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)f_X(x)h(y)f_Y(y)dxdx \\
&= \left(\int_{-\infty}^{\infty} g(x)f_X(x)dx\right)\left(\int_{-\infty}^{\infty} h(y)f_Y(y)dy\right) \\
&= E(g(X))E(h(Y)).
\end{aligned}$$

□

Corollary 7.5. *Let $X_1, ..., X_n$ be independent random variables. Then,*

$$\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n).$$

Example 7.4. Let $X_1, ..., X_n$ be i.i.d. random variables. The **sample mean** is defined to be $\overline{X} = (X_1 + \cdots + X_n)/n$ and the **sample variance** is defined to be

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Suppose that $E(X_1) = \mu$ and $\mathrm{Var}(X_1) = \sigma^2$. Then,

$$E(\overline{X}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu, \quad \mathrm{Var}(\overline{X}) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{\sigma^2}{n}.$$

For the expectation of the sample variance, observe that

$$\begin{aligned}
(n-1)S^2 &= \sum_{i=1}^{n}(X_i - \mu + \mu - \overline{X})^2 \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 + 2(\mu - \overline{X})\sum_{i=1}^{n}(X_i - \mu) + n(\overline{X} - \mu)^2 \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 - n(\overline{X} - E(\overline{X}))^2.
\end{aligned}$$

This implies

$$E(S^2) = \frac{1}{n-1}\left(\sum_{i=1}^{n}\mathrm{Var}(X_i) - n\mathrm{Var}(\overline{X})\right) = \sigma^2.$$

Definition 7.2. For any two random variables X, Y, the **correlation** of X, Y is defined to be

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{E(XY) - E(X)E(Y)}{\sigma_X\sigma_Y}.$$

Remark 7.2. $\rho(X,Y) \in [-1, 1]$.

7.3. **Conditional expectation.** Recall that if X, Y are jointly continuous with joint p.d.f. f and $f_Y(y) > 0$, then the conditional density of X given $Y = y$ is

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}.$$

If X, Y are discrete with joint p.m.f. p and $p_Y(y) > 0$, then the conditional mass function of X given $Y = y$

$$p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)}.$$

It is worthwhile to note that given $Y = y$, $f_{X|Y}(x|y)$ is a probability density function and $p_{X|Y}(x|y)$ is a probability mass function. That is,

$$\int_{-\infty}^{\infty} f_{X|Y}(x|y)dx = 1, \quad \sum_y p_{X|Y}(x|y) = 1 \quad \forall y.$$

Thus, it makes sense to define the expectation of $f_{X|Y}(x|y)$ and $p_{X|Y}(x|y)$ with respect to x.

Definition 7.3. Let X, Y be random variables.

(1) If X, Y are jointly continuous with joint p.d.f. f and $f_Y(y) > 0$, then the **conditional expectation of X given $Y = y$** is defined to be

$$E(X|Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx.$$

(2) If X, Y are discrete with joint p.m.f. p and $p_Y(y) > 0$, then the **conditional expectation of X given $Y = y$** is defined to be

$$E(X|Y = y) = \sum_x x p_{X|Y}(x|y).$$

Remark 7.3. For a more general setting, note that if X has distribution function F, then $E(X) = \int_{-\infty}^{\infty} x dF(x)$ whatever X is continuous or discrete. Similarly, if the conditional distribution of X given $Y = y$ is $F_{X|Y}(x|y)$, then

$$E(X|Y = y) = \int_{-\infty}^{\infty} x dF_{X|Y}(x|y).$$

Example 7.5. Let X, Y be jointly continuous with joint p.d.f.

$$f(x,y) = \begin{cases} \frac{e^{-x/y}e^{-y}}{y} & \text{if } x > 0, y > 0 \\ 0 & \text{otherwise} \end{cases}.$$

It has been shown before that

$$f_{X|Y}(x|y) = \begin{cases} e^{-x/y}/y & \text{if } x > 0, y > 0 \\ 0 & \text{otherwise} \end{cases}.$$

Following the definition, one has

$$E(X|Y = y) = \int_0^{\infty} x f_{X|Y}(x|y)dx = \int_0^{\infty} \frac{x}{y} e^{-x/y} dx = y \quad \forall y > 0.$$

Proposition 7.6. *If X, Y are independent. Then, $E(X|Y = y) = E(X)$.*

Proof. We deal with the continuous case. Assume that X, Y are jointly continuous with joint p.d.f. f. The independence of X, Y implies $f(x,y) = f_X(x)f_Y(y)$ and $f_{X|Y}(x|y) = f_X(x)$. Thus,

$$E(X|Y=y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx = \int_{-\infty}^{\infty} x f_X(x)dx = E(X).$$

□

Remark 7.4. Let X, Y be random variables and g be a real valued function. Then,

$$E(g(X)|Y=y) = \int_{-\infty}^{\infty} g(x)dF_{X|Y}(x|y).$$

In particular, if X, Y are jointly continuous with conditional p.d.f. $f_{X|Y}(x|y)$, then

$$E(g(X)|Y=y) = \int_{-\infty}^{\infty} g(x)f_{X|Y}(x|y)dx.$$

If X, Y are discrete with conditional p.m.f. $p_{X|Y}(x|y)$, then

$$E(g(X)|Y=y) = \sum_x g(x)p_{X|Y}(x|y).$$

Remark 7.5. Regarding $E(X|Y)$ as a function of Y whose value at $Y = y$ is $E(X|Y=y)$, one may treat $E(X|Y)$ as a random variable satisfying $E(X|Y(s)) = E(X|Y=y)$ if $Y(s) = y$.

Proposition 7.7. *For any random variables X, Y, $E(E(X|Y)) = E(X)$. In particular, if X, Y are jointly continuous with joint p.d.f. f, then*

$$E(X) = \int_{-\infty}^{\infty} E(X|Y=y)f_Y(y)dy.$$

If X, Y are discrete with joint p.m.f. p, then

$$E(X) = \sum_y E(X|Y=y)p_Y(y).$$

Proof. We deal with the continuous case. Following the definition, we have

$$E(X|Y=y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx = \int_{-\infty}^{\infty} \frac{xf(x,y)}{f_Y(y)}dx.$$

This implies

$$E(E(X|Y)) = \int_{-\infty}^{\infty} E(X|Y=y)f_Y(y)dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)dxdy = \int_{-\infty}^{\infty} xf_X(x)dx = E(X).$$

□

Example 7.6. Let N be the number of costumers entering a store in one day and let X_n denote the total expense of the nth customer in this store. Then, the total money received in a day is $G = X_1 + X_2 + \cdots + X_N$. Suppose $P(N < \infty) = 1$, $X_1, X_2, ...$ are i.i.d., and $X_1, X_2, ...$ and N are independent. Then,

$$\begin{aligned} E(G) = E(E(G|N)) &= \sum_{n=1}^{\infty} E(G|N=n)P(N=n) \\ &= \sum_{n=1}^{\infty} E(X_1 + \cdots + X_n)P(N=n) = E(X_1)E(N). \end{aligned}$$

Example 7.7. Let $X_1, X_2, \ldots$ be i.i.d. random variables uniformly distributed over $(0,1)$. Let N be integer-valued random variable defined by

$$N = \min\left\{ n \ge 1 \,\middle|\, \sum_{i=1}^{n} X_i > 1 \right\}.$$

To find the expectation of N, we consider a more general setting. For $x \ge 0$, define

$$N(x) = \min\left\{ n \ge 1 \,\middle|\, \sum_{i=1}^{n} X_i > x \right\}.$$

Then, $N = N(1)$. By setting $m(x) = E(N(x))$, one may compute

$$E(N(x)|X_1 = y) = \begin{cases} 1 & \text{if } x < y \\ 1 + m(x-y) & \text{if } x \ge y \end{cases}.$$

This implies

$$m(x) = \int_0^1 E(N(x)|X_1 = y)dy = 1 + \int_0^x m(x-y)dy = 1 + \int_0^x m(y)dy.$$

Clearly, m is differentiable and $m'(x) = m(x)$. Along with the boundary condition of $m(0) = 1$, we obtain $m(x) = e^x$ and, in particular, $E(N) = e$.

Example 7.8. Let X, Y be jointly continuous with joint p.d.f.

$$\begin{aligned} f(x,y) =& \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \\ &\times \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} \right] \right\}, \end{aligned}$$

for $x, y \in \mathbb{R}$. f is also called a bivariate normal joint density function since

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \frac{1}{\sqrt{2\pi}\sigma_X} e^{-(x-\mu_X)^2/(2\sigma_X^2)}$$

and

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \frac{1}{\sqrt{2\pi}\sigma_Y} e^{-(y-\mu_Y)^2/(2\sigma_Y^2)}.$$

Using the above identity, it is obvious that, given $Y = y$, X is normal with mean $\mu_X + \rho(y - \mu_Y)\sigma_X/\sigma_Y$ and variance $(1-\rho^2)\sigma_X^2$. This implies

$$E(XY|Y=y) = yE(X|Y=y) = \left(y\mu_X + \frac{\rho(y^2 - y\mu_Y)\sigma_X}{\sigma_Y} \right)$$

and

$$\begin{aligned} E(XY) = E(E(XY|Y)) &= E\left(Y\mu_X + \frac{\rho(Y^2 - Y\mu_Y)\sigma_X}{\sigma_Y} \right) \\ &= \mu_X E(Y) + \rho\frac{\sigma_X}{\sigma_Y}\left[E(Y^2) - \mu_Y E(Y)\right] \\ &= \mu_X\mu_Y + \rho\frac{\sigma_X}{\sigma_Y}(\sigma_Y^2 + \mu_Y^2 - \mu_Y^2) = \mu_X\mu_Y + \rho\sigma_X\sigma_Y. \end{aligned}$$

Hence,

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)\mathrm{Var}(Y)} = \frac{E(XY) - E(X)E(Y)}{\sigma_X\sigma_Y} = \rho.$$

Definition 7.4. The conditional variance of X given $Y = y$ is defined to be

$$\mathrm{Var}(X|Y = y) = E[(X - E(X|Y = y))^2|Y = y] = E(X^2|Y = y) - [E(X|Y = y)]^2.$$

Remark 7.6. Note that

$$E(\mathrm{Var}(X|Y)) = E(X^2) - E[E(X|Y)]^2.$$

In general, the latter term is not equal to $[E(X)]^2$.

Proposition 7.8. *For any random variables X, Y,*

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

Example 7.9. Recall the example that $X_1, X_2, ...$ are i.i.d. and are independent of an nonnegative integer-valued random variable N and $G = X_1 + \cdots + X_N$. Using a similar argument as before, one has

$$E(G|N = n) = E(X_1)n, \quad E(G^2|N = n) = nE(X_1^2) + n(n-1)[E(X_1)]^2.$$

This implies $\mathrm{Var}(G|N = n) = n\mathrm{Var}(X_1)$ and, hence,

$$\mathrm{Var}(G) = E[\mathrm{Var}(G|N)] + \mathrm{Var}(E(G|N)) = \mathrm{Var}(X_1)E(N) + [E(X_1)]^2\mathrm{Var}(N).$$

7.4. **Moment generating functions.**

Definition 7.5. Let X be a random variable and n be a positive integer.

(1) The nth moment of X is defined to be $E(X^n)$.
(2) The moment generating function of X is the function $M(t) = E(e^{tX})$.

Remark 7.7. Formally, one can expect the following computation

$$M'(t) = \frac{d}{dt}E(e^{tX}) = E\left(\frac{d}{dt}e^{tX}\right) = E(Xe^{tX}).$$

This yields $E(X) = M'(0)$. Inductively, we have $M^{(n)}(0) = E(X^n)$ for $n \geq 1$. In fact, from the Taylor expansion, one has

$$M(t) = E\left(\sum_{n=0}^{\infty}\frac{t^n}{n!}X^n\right) = \sum_{n=0}^{\infty}\frac{t^n}{n!}E(X^n).$$

Remark 7.8. The moment generating function is exactly the Laplace transform.

Proposition 7.9. *Let X, Y be random variables with moment generating functions M_X, M_Y. If X, Y are independent, then $M_{X+Y}(t) = M_X(t)M_Y(t)$.*

Example 7.10. Let X be a binomial random variable with parameters (n, p) and $M(t)$ be its moment generating function. For $n = 1$, one has

$$M(t) = e^t p + (1 - p) \quad \forall t \in \mathbb{R}.$$

For $n > 1$, let $X_1, ..., X_n$ be i.i.d. Bernoulli random variables with parameters p. Then, the above result implies

$$M(t) = E(e^{tX}) = E(e^{t(X_1+\cdots+X_n)}) = \prod_{i=1}^{n} E(e^{tX_i}) = (e^t p + 1 - p)^n, \quad \forall t \in \mathbb{R}.$$

Example 7.11. If X is Poisson(λ), then

$$M(t) = \sum_{n=0}^{\infty} e^{tn}e^{-\lambda}\frac{\lambda^n}{n!} = e^{-\lambda}\sum_{n=0}^{\infty}\frac{(e^t\lambda)^n}{n!} = \exp\{-\lambda + e^t\lambda\} = \exp\{\lambda(e^t - 1)\}.$$

Example 7.12. If X is an exponential random variable with parameter λ, then

$$M(t) = \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx = \lambda \int_0^\infty e^{(t-\lambda)x} dx = \frac{\lambda}{\lambda - t}, \quad \forall t < \lambda.$$

Example 7.13. Let X be a normal random variable with mean 0 and variance 1. Then,

$$M(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{tx} e^{-x^2/2} dx = \frac{e^{t^2/2}}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-(x-t)^2/2} dx = e^{t^2/2}, \quad \forall t \in \mathbb{R}.$$

For the general case, let Y be normal with mean μ and variance σ^2. By writing $Y = \mu + \sigma X$, we have

$$M_Y(t) = E(e^{tY}) = E(e^{t(\mu+\sigma X)}) = e^{\mu t} M_X(t\sigma) = \exp\{\mu t + \sigma^2 t^2/2\}, \quad \forall t \in \mathbb{R}.$$

Theorem 7.10. *Let X, Y be nonnegative random variables with distribution functions F_X, F_Y and moment generating functions M_X, M_Y. Then, $F_X(a) = F_Y(a)$ for all $a \in \mathbb{R}$ if and only if $M_X(t) = M_Y(t)$ for all $t \leq 0$.*

Example 7.14. Suppose that X is a nonnegative random variable with moment generating function $M(t) = e^{3(e^t-1)}$. As a Poisson random variable with parameter 3 has its moment generating function $M(t)$, X must be a Poisson random variable with parameter 3.