

Probability Theory

Chapter 7—Properties of Expectation

Nan Liu¹

¹These slides are based on Prof. Mui Li's. Any mistakes are my own. If you find an error, please send an email at nanliu@xmu.edu.cn

7.1 Introduction

- Discrete random variable with probability mass function $p(x)$,

$$E[X] = \sum_x xp(x)$$

- Continuous random variable with probability density function $f(x)$

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

- $E[X]$ is a weighted average of the possible values of X ,
- If

$$P\{a \leq X \leq b\} = 1$$

then

$$a \leq E[X] \leq b$$

7.2 Expectation of Function of Random Variables

- If X and Y have a joint probability mass function $p(x, y)$, then

$$E[g(X, Y)] = \sum_y \sum_x g(x, y)p(x, y)$$

- If X and Y have a joint probability density function $f(x, y)$, then

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f(x, y)dx dy$$

- Let $g(X, Y) = X$, then for D.R.V.

$$E[X] = \sum_y \sum_x xp(x, y).$$

For C.R.V,

$$E[X] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xf(x, y)dx dy.$$

7.2 Example

Let X and Y have the PDF

$$f(x, y) = \begin{cases} 8xy & 0 < x < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

- Find $E[X]$, $E[Y]$.
- Let $Z = \frac{X}{Y}$. Find $E[Z]$.

7.2 Expectation of Sums of Random Variables

- Suppose $E[X]$ and $E[Y]$ are both finite and let $g(X, Y) = X + Y$. Then, (in the continuous case)

$$\begin{aligned} E[X + Y] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y) f(x, y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y) dy dx + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x, y) dx dy \\ &= \int_{-\infty}^{\infty} x f_X(x) dx + \int_{-\infty}^{\infty} y f_Y(y) dy \\ &= E[X] + E[Y] \end{aligned}$$

- If $E[X_i]$ is finite for all $i = 1, \dots, n$, then

$$E[X_1 + \dots + X_n] = E[X_1] + \dots + E[X_n]$$

7.2 Example: The sample mean

- Let $X_1, \dots, X_n$ be independent and identically distributed random variables having distribution function F and expected value μ . Such a sequence of random variables is said to constitute a sample from the distribution F . The quantity

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

is called the sample mean.

- $$\begin{aligned} E[\bar{X}] &= E \left[\frac{1}{n} \sum_{i=1}^n X_i \right] = \frac{1}{n} E \left[\sum_{i=1}^n X_i \right] \\ &= \frac{1}{n} \sum_{i=1}^n E[X_i] = \mu \end{aligned}$$

- When the distribution mean μ is unknown, the sample mean is often used in statistics to estimate it.

7.2 Example: Expected Number of Matches

- Suppose that N people throw their hats into the center of a room. The hats are mixed up, and each person randomly selects one. Find the expected number of people who select their own hat.

•

$$X = X_1 + X_2 + \cdots + X_N$$

where

$$X_i = \begin{cases} 1 & \text{if the } i\text{th person selects his own hat} \\ 0 & \text{otherwise} \end{cases}$$

- Since, for each i , the i th person is equally likely to select any of the N hats,

$$E[X_i] = P(X_i = 1) = \frac{1}{N}$$

$$E[X] = E[X_1] + E[X_2] + \cdots + E[X_n] = \frac{1}{N} \cdot N = 1$$

- On the average, exactly one person selects his own hat.

7.3 Moments of the Number of Events that Occur

- For given events $A_1, \dots, A_n$, find $E[X]$, where X is the number of these events that occur. The solution then involved defining an indicator variable I_i for event A_i such that

$$I_i = \begin{cases} 1, & \text{if } A_i \text{ occurs} \\ 0, & \text{otherwise} \end{cases}$$

Because

$$X = \sum_{i=1}^n I_i$$

we obtained the result

$$E[X] = E \left[\sum_{i=1}^n I_i \right] = \sum_{i=1}^n E[I_i] = \sum_{i=1}^n P(A_i)$$

7.3 Moments of the Number of Events that Occur

- The number of pairs of events that occur is $\binom{X}{2}$. Consequently,

$$\binom{X}{2} = \sum_{i < j} I_i I_j$$

Taking expectations yields

$$E \left[\binom{X}{2} \right] = \sum_{i < j} E [I_i I_j] = \sum_{i < j} P(A_i A_j)$$

or

$$E \left[\frac{X(X-1)}{2} \right] = \sum_{i < j} P(A_i A_j)$$

giving that

$$E[X^2] - E[X] = 2 \sum_{i < j} P(A_i A_j)$$

which yields $E[X^2]$, and thus $\text{Var}(X) = E[X^2] - (E[X])^2$.

7.3 Moments of the Number of Events that Occur

- Moreover, by considering the number of distinct subsets of k events that all occur, we see that

$$\binom{X}{k} = \sum_{i_1 < i_2 < \dots < i_k} I_{i_1} I_{i_2} \cdots I_{i_k}$$

Taking expectations gives the identity

$$E \left[\binom{X}{k} \right] = \sum_{i_1 < i_2 < \dots < i_k} E [I_{i_1} I_{i_2} \cdots I_{i_k}] = \sum_{i_1 < i_2 < \dots < i_k} P(A_{i_1} A_{i_2} \cdots A_{i_k})$$

7.3 Example: Moments of binomial random variables

- Consider n independent trials, with each trial being a success with probability p . Let A_i be the event that trial i is a success. When $i \neq j$, $P(A_i A_j) = p^2$. Consequently, Equation (3.2) yields

$$E \left[\binom{X}{2} \right] = \sum_{i < j} p^2 = \binom{n}{2} p^2$$

or

$$E[X(X-1)] = n(n-1)p^2$$

or

$$E[X^2] - E[X] = n(n-1)p^2$$

Now, $E[X] = \sum_{i=1}^n P(A_i) = np$, so, from the preceding equation

$$\text{Var}(X) = E[X^2] - (E[X])^2 = n(n-1)p^2 + np - (np)^2 = np(1-p)$$

7.3 Moments of the Number of Events that Occur

- Moments in the match problem For $i = 1, \dots, N$, let A_i be the event that person i selects his or her own hat in the match problem. Then

$$P(A_i A_j) = P(A_i) P(A_j | A_i) = \frac{1}{N} \frac{1}{N-1}$$

which follows because, conditional on person i selecting her own hat, the hat selected by person j is equally likely to be any of the other $N-1$ hats, of which one is his own. Consequently, with X equal to the number of people who select their own hat, it follows from Equation (3.2) that

$$E \left[\binom{X}{2} \right] = \sum_{i < j} \frac{1}{N(N-1)} = \binom{N}{2} \frac{1}{N(N-1)}$$

thus showing that

$$E[X(X-1)] = 1$$

Therefore, $E[X^2] = 1 + E[X]$. Because $E[X] = \sum_{i=1}^N P(A_i) = 1$, we obtain that $\text{Var}(X) = E[X^2] - (E[X])^2 = 1$.

7.4 Covariance, Variance of Sums, and Correlations

- Consider two r.v. X and Y (either discrete or continuous), then the **covariance** of (X, Y) is defined as

$$\begin{aligned}\text{Cov}(X, Y) &= E(X - E(X))(Y - E(Y)) \\ &= E(XY) - E(X)E(Y)\end{aligned}$$

- The covariance measures the degree to which X and Y vary together. If the covariance is positive, X tends to be larger than its mean when Y is larger than its mean. The covariance of a variable with itself is the variance of that variable.

7.4 Example: Two Stocks

- Let X and Y denote the values of two stocks at the end of a five-year period. X is uniformly distributed on $(0, 12)$. Given $X = x$, Y is uniformly distributed on the interval $(0, x)$. Determine $\text{Cov}(X, Y)$.
- We have for $0 < x < 12$ and $0 < y < x$

$$f(x, y) = f_X(x) f_{Y|X}(y|x) = \frac{1}{12} \frac{1}{x}$$

so

$$E(X) = \int_0^{12} x \cdot \frac{1}{12} dx = 6,$$

$$E(Y) = \int_0^{12} \int_0^x y \cdot \frac{1}{12} \frac{1}{x} dy dx = 3,$$

$$E(XY) = \int_0^{12} \int_0^x xy \cdot \frac{1}{12} \frac{1}{x} dy dx = 24.$$

Hence we have

$$\text{Cov}(X, Y) = 24 - 3 \times 6 = 6.$$

7.4 Proposition of Covariance

- (i) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;
- (ii) $\text{Cov}(X, X) = \text{Var}(X)$;
- (iii) $\text{Cov}(aX, Y) = a\text{Cov}(X, Y)$;
- (iv) $\text{Cov}(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j) = \sum_{i=1}^n \sum_{j=1}^m \text{Cov}(X_i, Y_j)$.

7.4 Variance of Sums of R.V.s



$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var} (X_i) + 2 \sum_{i < j} \sum_i \text{Cov} (X_i, X_j)$$



$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)$$

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2 \text{Cov}(X, Y)$$

- If $X_1, \dots, X_n$ are pairwise independent, then

$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var} (X_i)$$

7.4 Example: Spreading your risk optimally

- You have 2 financial products whose returns can be modelled by the r.v. X and Y such that $E(X) = E(Y) = \mu$, $\text{Var}(X) = \sigma_x^2$, $\text{Var}(Y) = \sigma_y^2$ and $\text{Cov}(X, Y) = \sigma_{xy}$. (These two products are equally priced). You want to buy a proportion λ of product 1 and $(1 - \lambda)$ of product 2 where $\lambda \in [0, 1]$ to spread the risk.
 - (a) What is the expectation of the total return $Z = \lambda X + (1 - \lambda) Y$?
 - (b) What is the variance of the total return?
 - (c) How should you select λ to minimize this variance?
 - (d) What is the minimum variance of the return if X and Y are independent?

7.4 Example: Spreading your risk optimally

- (a) The total return is given by $Z = \lambda X + (1 - \lambda) Y$ so

$$E(Z) = \lambda E(X) + (1 - \lambda) E(Y) = \mu.$$

- (b) We have

$$\begin{aligned} \text{Var}(Z) &= \lambda^2 \sigma_x^2 + (1 - \lambda)^2 \sigma_y^2 + 2\lambda(1 - \lambda) \sigma_{xy} \\ &= \lambda^2 (\sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}) + 2\lambda (\sigma_{xy} - \sigma_y^2) + \sigma_y^2 \end{aligned}$$

- (c) We just differentiate $\text{Var}(Z)$ w.r.t. λ and obtain

$$\lambda_{\text{opt}} = \frac{\sigma_y^2 - \sigma_{xy}}{\sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}}$$

- (d) For X, Y independent, we have $\sigma_{xy} = 0$ so $\lambda_{\text{opt}} = \sigma_y^2 / (\sigma_x^2 + \sigma_y^2)$

$$\begin{aligned} \text{Var}(Z) &= \frac{\sigma_y^4}{(\sigma_x^2 + \sigma_y^2)^2} \sigma_x^2 + \frac{\sigma_x^4}{(\sigma_x^2 + \sigma_y^2)^2} \sigma_y^2 \\ &= \frac{\sigma_x^2 \sigma_y^2 (\sigma_x^2 + \sigma_y^2)}{(\sigma_x^2 + \sigma_y^2)^2} = \frac{\sigma_x^2 \sigma_y^2}{\sigma_x^2 + \sigma_y^2} \end{aligned}$$

7.4 Example

- Let $X_1, \dots, X_n$ be independent and identically distributed random variables having variance σ^2 .

$$\begin{aligned}\text{Var}(\bar{X}) &= \left(\frac{1}{n}\right)^2 \text{Var}\left(\sum_{i=1}^n X_i\right) \\ &= \left(\frac{1}{n}\right)^2 \sum_{i=1}^n \text{Var}(X_i) \quad \text{by independence} \\ &= \frac{\sigma^2}{n}\end{aligned}$$

$$\begin{aligned}\text{Cov}(X_i - \bar{X}, \bar{X}) &= \text{Cov}(X_i, \bar{X}) - \text{Cov}(\bar{X}, \bar{X}) \\ &= \text{Cov}\left(X_i, \frac{1}{n} \sum_{j=1}^n X_j\right) - \text{Var}(\bar{X}) \\ &= \frac{1}{n} \sum_{j=1}^n \text{Cov}(X_i, X_j) - \frac{\sigma^2}{n} = \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0\end{aligned}$$

7.4 Correlation

- The correlation of (X, Y) is defined as

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$$

- The correlation is a measure of “linear dependence” between X and Y . It is a uniteless measure which takes values in $[-1, 1]$. Proof can be established using Cauchy-Schwartz inequality:

$$\left(\int \alpha(u) \beta(u) du \right)^2 \leq \left(\int \alpha^2(u) du \right) \left(\int \beta^2(u) du \right).$$

- If X and Y are two independent r.v. then $\rho(X, Y) = 0$ as $\text{Cov}(X, Y) = 0$.

7.4 Example: Two Stocks

- We want to compute the correlation in the previous example:

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$$

- We have computed $\text{Cov}(X, Y) = 6$ so we need to compute $\text{Var}(X)$ and $\text{Var}(Y)$. We have

$$E(X^2) = \int_0^{12} x^2 \cdot \frac{1}{12} dx = \frac{1}{12} \left[\frac{x^3}{3} \right]_0^{12} = \frac{144}{3} = 48,$$

$$E(Y^2) = \int_0^{12} \frac{1}{12} \frac{1}{x} \left(\int_0^x y^2 \cdot dy \right) dx = \frac{1}{36} \int_0^{12} x^2 dx = \frac{144}{9} = 16.$$

Hence

$$\rho(X, Y) = \frac{6}{\sqrt{48 - 6^2} \sqrt{16 - 3^2}} = 0.3631$$

7.4 Independence and Uncorrelated

- If X and Y are two independent r.v. then

$$\text{Cov}(X, Y) = 0, \text{ and } \rho(X, Y) = 0.$$

- **Proof.** We are going to show that $E(XY) = E(X)E(Y)$ if X and Y are independent

$$\begin{aligned} E(XY) &= \int \int xy \cdot f(x, y) \, dx dy \\ &= \int \int xy \cdot f_X(x) f_Y(y) \, dx dy \quad (\text{independence}) \\ &= \left[\int x \cdot f_X(x) \, dx \right] \left[\int y \cdot f_Y(y) \, dy \right] \\ &= E(X) E(Y) \end{aligned}$$

7.4 Independence and Uncorrelated

- Independence does imply uncorrelation but the reverse is NOT true.
- Counter example:
 - (a) for discrete r.v.: let X be such that

$$P(X = -1) = P(X = 0) = P(X = 1) = \frac{1}{3}$$

and $Y = X^2$ then X and Y are dependent but X and Y are uncorrelated as

$$\begin{aligned}\text{Cov}(X, Y) &= E(XY) - E(X)E(Y) \\ &= \underbrace{E(X^3)}_{=0} - \underbrace{E(X)}_{=0}E(X^2) = 0\end{aligned}$$

- (b) for continuous r.v.: let X be a standard normal and $Y = X^2$ then clearly X and Y are dependent but X and Y are uncorrelated as

$$\begin{aligned}\text{Cov}(X, Y) &= E(XY) - E(X)E(Y) \\ &= \underbrace{E(X^3)}_{=0} - \underbrace{E(X)}_{=0}E(X^2) = 0\end{aligned}$$

7.5 Properties of Conditional Expectation

- We have

$$E_Y [E_X (X|Y)] = E [X].$$

- **Proof.** We have

$$\begin{aligned} & E_Y [E_X (X|Y)] \\ &= \int \left[\int x \cdot f_{X|Y} (x|y) dx \right] f_Y (y) dy \\ &= \int \int \underbrace{x \cdot f_{X|Y} (x|y) f_Y (y)}_{=f(x,y)} dx dy \\ &= \int \int x \cdot f_X (x) f_{Y|X} (y|x) dx dy = \int \left[\int x \cdot f_X (x) dx \right] f_{Y|X} (y|x) dy \\ &= \int x \cdot f_X (x) dx = E [X] \end{aligned}$$

- This is valid for any function $E [E (g(X)|Y)] = E [g(X)]$.

7.5 Example: Workplace Injuries

- The number of workplace injuries N occurring in a factory on any given day is Poisson distributed of parameter λ . The parameter λ is a random variable determined by the level of activity in the factory and is uniformly distributed on the interval $[0, 3]$. What is the expectation of N ?

- We have

$$E[N] = E[E[N|\lambda]]$$

where

$$E[N|\lambda] = \lambda.$$

- Hence we have

$$E[N] = E[\lambda] = \frac{3}{2}.$$

7.5 Properties of Conditional Variance

- **Variance decomposition**

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

hence as $E [\text{Var} (X| Y)] \geq 0$, we have

$$\text{Var} [E (X| Y)] \leq \text{Var} [X]$$

- **Proof.** We have

$$\begin{aligned}\text{Var} [E (X| Y)] &= E \left(\{E (X| Y)\}^2 \right) - E (E (X| Y))^2 \\ &= E \left(\{E (X| Y)\}^2 \right) - E (X)^2\end{aligned}$$

and

$$\text{Var} (X| Y) = E (X^2| Y) - \{E (X| Y)\}^2$$

so

$$\begin{aligned}E [\text{Var} (X| Y)] &= E (E (X^2| Y)) - E \left(\{E (X| Y)\}^2 \right) \\ &= E (X^2) - E \left(\{E (X| Y)\}^2 \right)\end{aligned}$$

7.5 Example: Workplace Injuries

- The number of workplace injuries N occurring in a factory on any given day is Poisson distributed of parameter λ . The parameter λ is a random variable determined by the level of activity in the factory and is uniformly distributed on the interval $[0, 3]$. What is the variance of N ?

- We have

$$\text{Var} [N] = \text{Var} [E (N | \lambda)] + E [\text{Var} (N | \lambda)]$$

where

$$E (N | \lambda) = \text{Var} (N | \lambda) = \lambda$$

as, given λ , N is a Poisson random variable of param λ .

- Hence

$$\begin{aligned} \text{Var} [N] &= \text{Var} [\lambda] + E [\lambda] \\ &= \frac{3^2}{12} + \frac{3}{2} = \frac{9}{4} \end{aligned}$$

as λ follows an uniform distribution on $[0, 3]$.

7.6 Conditional Expectation and Prediction

- Let us consider two r.v. X and Y . Assume we observe X and want to find a way to estimate Y based on X . Then in some sense, $E(Y|X)$ is the best possible estimate of Y .
- **Proposition.** Consider an arbitrary function $g(X)$ then, we have

$$E[(Y - E(Y|X))^2] \leq E[(Y - g(X))^2],$$

that is the expected square “distance” between $g(X)$ and Y is minimized for $g(X) = E(Y|X)$.

- The proof is valid for both discrete and continuous r.v.

7.6 Conditional Expectation and Prediction

- We have

$$\begin{aligned} & E \left[(Y - g(X))^2 \middle| X = x \right] \\ &= E \left[(Y - E(Y|X) + E(Y|X) - g(X))^2 \middle| X = x \right] \\ &= E \left[(Y - E(Y|X))^2 \middle| X = x \right] + E \left[(E(Y|X) - g(X))^2 \middle| X = x \right] \\ &\quad + 2E \left[(Y - E(Y|X))(E(Y|X) - g(X)) \middle| X = x \right] \end{aligned}$$

- We have

$$\begin{aligned} & E \left[(Y - E(Y|X))(E(Y|X) - g(X)) \middle| X = x \right] \\ &= \underbrace{E \left[(Y - E(Y|X)) \middle| X = x \right]}_{=0} (E(Y|X = x) - g(x)) \end{aligned}$$

7.6 Conditional Expectation and Prediction

- Hence

$$\begin{aligned} E \left[(Y - g(X))^2 \middle| X = x \right] &= E \left[(Y - E(Y|X))^2 \middle| X = x \right] \\ &\quad + \underbrace{E \left[(E(Y|X) - g(X))^2 \middle| X = x \right]}_{\geq 0} \end{aligned}$$

- We can conclude that

$$E \left[(Y - g(X))^2 \middle| X = x \right] \geq E \left[(Y - E(Y|X))^2 \middle| X = x \right].$$

- Now by taking the expectation on both sides with respect to X , we obtain

$$E \left[(Y - g(X))^2 \right] \geq E \left[(Y - E(Y|X))^2 \right].$$

7.6 Conditional Expectation and Prediction

- **Best Linear Predictor.** Suppose X and Y are two random variables with finite second moments. Consider a linear regression function

$$Y = \alpha + \beta X + u,$$

where u is called an *error term*. Then the optimal coefficients (α^*, β^*) that minimizes the mean squared error

$$\text{MSE}(\alpha, \beta) = E[Y - \alpha - \beta X]^2$$

is

$$\begin{aligned}\beta^* &= \frac{\text{Cov}(X, Y)}{\text{Var}(X)}, \\ \alpha^* &= \mu_Y - \beta^* \mu_X.\end{aligned}$$

7.6 Conditional Expectation and Prediction

- **Investment Beta.** Let R_p be the return on a portfolio during a certain horizon, r_f is the risk-free interest rate, R_m is the return on the market portfolio (i.e., return on S&P500 index). The *investment beta* is

$$\beta = \frac{\text{Cov}(R_p - r_f, R_m - r_f)}{\text{Var}(R_m - r_f)}.$$

- $R_p - r_f$ is called the excess return on portfolio, and $R_m - r_f$ is the excess return on the market portfolio.
- If $\beta = 1$, the portfolio is equally risky to the market portfolio.
- If $\beta > 1$, the portfolio is more risky than the market portfolio.
- If $\beta < 1$, the portfolio is less risky than the market portfolio.

7.7 Moment Generating Function

- **Definition: Moment Generating Function (mgf).** The *moment generating function* of a r.v. X is defined as

$$\begin{aligned}M_X(t) &= E[\exp(tX)] \\&= \begin{cases} \sum_x \exp(tx) f_X(x) & \text{d.r.v.,} \\ \int_{-\infty}^{\infty} \exp(tx) f_X(x) dx & \text{c.r.v..} \end{cases}\end{aligned}$$

- **Remarks:**

- $M_X(t)$ may not exist for some $t \in \mathbb{R}$.
- The existence of the MGF $M_X(t)$ imply the existence of an infinite set of moments.

7.7 Moment Generating Function

- **Theorem:**

- If $M_X(t)$ exists for t in some neighborhood of 0, then for $k = 1, 2, \dots$,

$$M_X^{(k)}(0) = E(X^k).$$

- If a and b are two constants, then

$$M_{bX+a} = e^{at} M_X(bt).$$

- **Example:** Let $Y = \frac{X-\mu}{\sigma}$, then $E(Y) = 0$, $\text{Var}(Y) = 1$, and

$$M_Y(t) = \exp\left\{-\frac{\mu}{\sigma}t\right\} M_X\left(\frac{t}{\sigma}\right).$$

7.7 Moment Generating Function : D.R.V.

- **Example: Binomial Distribution.** For the Binomial distribution $B(n, p)$,

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

Find

- (1) μ_X by direct formula.
- (2) σ_X^2 by direct formula.
- (3) $M_X(t)$.
- (4) Use $M_X(t)$ to find μ_X and σ_X^2 .

7.7 Moment Generating Function : D.R.V.

- **Solution:**

(1) $E(X) = np.$

(2) $E(X^2) = n(n-1)p^2 + np.$

(3) $M_X(t) = (pe^t + 1 - p)^n.$

7.7 Moment Generating Function : D.R.V.

- **Example: Poisson Distribution** The pmf of Poisson distribution is as follows

$$f_X(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \dots$$

Then

- (1) $E(X) = \lambda.$
- (2) $E(X^2) = \lambda^2 + \lambda.$
- (3) $M_X(t) = \exp\{e^t \lambda - \lambda\}.$

7.7 Moment Generating Function : C.R.V.

- **Example: Uniform Distribution.** A c.r.v. X follows a uniform distribution on $[a, b]$ if its pdf

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

- (1) $E(X^k) = \frac{b^{k+1} - a^{k+1}}{(k+1)(b-a)}.$
- (2) $E(X) = \frac{a+b}{2}.$
- (3) $\sigma_X^2 = \frac{1}{12}(b-a)^2.$
- (4) $M_X(t) = \frac{1}{(b-a)t} [e^{tb} - e^{ta}].$

7.7 Moment Generating Function : C.R.V.

- **Example: Normal Distribution.** A normally distributed random variable, $X \sim N(\mu, \sigma^2)$, has the pdf

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\},$$

where $-\infty < \mu < \infty$, $\sigma > 0$.

- (1) $E(X) = \mu$.
- (2) $\text{Var}(X) = \sigma^2$.
- (3) $M_X(t) = \exp\left\{\frac{\sigma^2 t^2}{2} + \mu t\right\}$.

7.7 Moment Generating Function : C.R.V.

- **Example: Gamma Distribution** A c.r.v. X follows a $\text{Gamma}(\alpha, \beta)$ ($\alpha, \beta > 0$) distribution if

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(\alpha)} (\lambda x)^{\alpha-1} \lambda e^{-\lambda x} & 0 < x < \infty, \\ 0 & \text{otherwise,} \end{cases}$$

where $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$. The gamma function Γ satisfies $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$.

- (1) $E(X) = \alpha / \lambda$.
- (2) $E(X^2) = \alpha(\alpha + 1) / \lambda^2$.
- (3) $\sigma_X^2 = E(X^2) - E^2(X) = \alpha / \lambda^2$.
- (4) $M_X(t) = (1 - t/\lambda)^{-\alpha}$ for $t < \lambda$.

7.7 Moment Generating Function

- **Definition: Identical Distribution.** Let X and Y be two r.v. with cdf's $F_X(x)$ and $F_Y(y)$, respectively. If two cdf's are the same, i.e. $F_X(x) = F_Y(x)$ for all $x \in \mathbb{R}$, then we say that X and Y are *identically distributed*.

- **Remarks:**

- If X and Y are identically distributed, then for any function $g(\cdot)$,

$$E[g(X)] = E[g(Y)].$$

- Identity of the distributions of X and Y does not imply $X = Y$.
 - **Example:** Suppose a fair coin is tossed n times, and let X be the number of heads obtained, Y be the number of tails obtained. Then $F_X = F_Y$, but $X + Y = n$ and $X \neq Y$.

7.7 Moment Generating Function

- **Theorem: Uniqueness of MGF.** Suppose two r.v. X and Y with $M_X(t)$ and $M_Y(t)$ existing in a neighborhood of 0. Then X and Y have the same $M_X(t)$ and $M_Y(t)$ for all t in $N_\varepsilon(0) = (-\varepsilon, \varepsilon)$, if and only if $F_X(u) = F_Y(u)$ for all $u \in \mathbb{R}$.
- Given some $M_X(t)$ in neighborhood of 0, suppose we can find some distribution $F_X(x)$, then $F_X(x)$ must be the only distribution that generates $M_X(t)$.

7.7 Moment Generating Function

- **Example:** A d.r.v. X has $M_X(t) = 1/2 + 1/4e^{-t} + 1/4e^t$. Then its pmf is

$$f_X(x) = \begin{cases} 1/4 & x = 1, \\ 1/2 & x = 0, \\ 1/4 & x = -1. \end{cases}$$

- **Example:** If a D.R.V. X has $M_X(t) = \frac{1-r}{1-re^t}$ for $|e^t| < 1/r$, where $r > 0$. Then what is the probability distribution of X ?

Ans:

$$\begin{aligned} M_X(t) &= (1-r) \sum_{x=0}^{\infty} (re^t)^x \\ &= \sum_{x=0}^{\infty} (1-r)r^x e^{tx} \\ &= \sum_{x=0}^{\infty} f_X(x) e^{tx}, \end{aligned}$$

7.7 Moment Generating Function

- **Example:** Suppose $M_X(t) = e^{2t^2}$. Find the distribution of X .
- **ANS:** Recall the normal distribution $N(\mu, \sigma^2)$ has mgf

$$M_X(t) = \exp\left\{\mu t + \frac{1}{2}\sigma^2 t^2\right\}.$$

Let $\mu = 0$ and $\sigma^2 = 4$, then $M_X(t) = e^{2t^2}$. So $X \sim N(0, 4)$.

7.7 Moment Generating Function

- **Definition: Converge in Distribution (Weak Convergence).** Let $\{X_n\}$ be a sequence of r.v.s with cdf's $\{F_n(x)\}$. Let X be a r.v. with cdf F_X . If $F_n(x) \rightarrow F_X(x)$ as $n \rightarrow \infty$ for all x 's where $F_X(x)$ is **continuous**, we say X_n *converges in distribution* to X , denoted by

$$X_n \xrightarrow{d} X.$$

- **Remark:** The convergence is required only for the continuous points of $F(x)$.
- **Example:** Let $X_n \sim N(0, \frac{1}{n})$ and $X = 0$. Then

$$F_X(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$$

Because for any $x < 0$, $\lim_n F_n(x) \rightarrow 0$ and for any $x > 0$, $\lim_n F_n(x) \rightarrow 1$, so $X_n \xrightarrow{d} X$. However, $F_n(0) = 0.5 \neq F_X(0)$.

7.7 Moment Generating Function

- **Theorem : Convergence of MGF.** Suppose $\{X_n, n = 1, 2, \dots\}$ is a sequence of random variables, each with mgf $M_n(t)$ and cdf $F_n(x)$. Furthermore, suppose that

$$\lim_{n \rightarrow \infty} M_n(t) = M_X(t)$$

for all t in a neighborhood of 0, and $M_X(t)$ is a mgf of some random variable X with cdf F_X . Then

$$\lim_{n \rightarrow \infty} F_n(x) = F_X(x)$$

or, denoted by

$$X_n \xrightarrow{d} X.$$

7.7 Moment Generating Function

- **Example: Poisson Approximation.** The mgf of the binomial distribution $B(n, p)$ is

$$M_B(t) = (pe^t + 1 - p)^n,$$

and the mgf of the Poisson distribution $P(\lambda)$ is

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

- Because

$$M_B(t) = \left(1 + \frac{np(e^t - 1)}{n}\right)^n.$$

When $n \rightarrow \infty$ and $np \rightarrow \lambda$, we can use the Poisson(λ) distribution to approximate the binomial distribution.

7.7 Joint Moment Generating Function

- **Definition: Joint Moment Generating Function.** The joint mgf of (X, Y) is defined as

$$M_{XY}(t_1, t_2) = E[\exp(t_1 X + t_2 Y)].$$

- **Theorem:** Suppose $M_{XY}(t_1, t_2)$ exists in a neighborhood of $(0, 0)$. Then

$$E(X^k Y^l) = \frac{\partial^{k+l}}{\partial t_1^k \partial t_2^l} M_{XY}(0, 0) = M_{XY}^{(k,l)}(0, 0), \quad k, l \geq 0,$$

and

$$\text{Cov}(X^k, Y^l) = M_{XY}^{(k,l)}(0, 0) - M_X^{(k)}(0)M_Y^{(l)}(0).$$

7.7 Joint Moment Generating Function

- **Remarks:**

- The joint mgf may not exist for some joint distributions.
- When $M_{XY}(t_1, t_2)$ exists in a neighborhood of $(0, 0)$, it can be used to uniquely characterize the joint distribution of (X, Y) .
- More generally, for n -dimensional random vector $(X_1, X_2, \dots, X_n)$, we can define the joint mgf

$$M(t_1, \dots, t_n) = E \left[\exp \left(\sum_{i=1}^n t_i X_i \right) \right].$$

7.7* Implications of Independence

- **Corollary:** If X and Y are independent, and their mgfs exist in a neighborhood of 0. Let $Z = X + Y$, then

$$M_Z(t) = M_X(t)M_Y(t)$$

exists in a neighborhood of 0.

- **Remark:** This property of mgf is useful in characterizing the distribution for the sum of independent random variables.

7.7* Implications of Independence

- **Example:** Suppose $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$ and X, Y are independent. Then

$$X \pm Y \sim N(\mu_1 \pm \mu_2, \sigma_1^2 + \sigma_2^2).$$

- **Proof.**

- The mgfs of X and Y are

$$M_X(t) = e^{\mu_1 t + \frac{\sigma_1^2}{2} t^2}$$

$$M_Y(t) = e^{\mu_2 t + \frac{\sigma_2^2}{2} t^2}$$

- $M_{X+Y}(t) = M_X(t)M_Y(t) = e^{(\mu_1 + \mu_2)t + \frac{\sigma_1^2 + \sigma_2^2}{2} t^2}.$
- $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$

7.7* Implications of Independence

- **Example:** Suppose $X_1, \dots, X_n$ are independent following Poisson distributions with parameters $\lambda_1, \dots, \lambda_n$, respectively. Find the distribution of $Y = \sum_{i=1}^n X_i$.
- **ANS:** Because $M_Y(t) = \prod_{i=1}^n M_{X_i}(t) = \prod_{i=1}^n e^{\lambda_i(e^t-1)}$, Y follows Poisson distributions with parameter $\lambda = \sum_{i=1}^n \lambda_i$.
- **Example:** Suppose $X_1, \dots, X_n$ are independent random variables having exponential distributions with the same parameter θ . Find the distribution of $Y = \sum_{i=1}^n X_i$.
- **ANS:** $M_Y(t) = \prod_{i=1}^n M_{X_i}(t) = (1 - t/\theta)^{-n}$. This implies that $Y \sim \text{Gamma}(n, \theta)$

7.7* Implications of Independence

- **Theorem:** Suppose $M_{XY}(t_1, t_2)$ exists in a neighborhood of $(0, 0)$. Then (X, Y) are independent if and only if

$$M_{XY}(t_1, t_2) = M_X(t_1)M_Y(t_2)$$

for all (t_1, t_2) in the neighborhood of $(0, 0)$.

7.7* Implications of Independence

- Let X and Y be independent normal random variables, each with mean μ and variance σ^2 . Show that $X + Y$ and $X - Y$ are independent. Compute their joint MGF:

$$\begin{aligned} E \left[e^{t(X+Y)+s(X-Y)} \right] &= E \left[e^{(t+s)X+(t-s)Y} \right] \\ &= E \left[e^{(t+s)X} \right] E \left[e^{(t-s)Y} \right] \\ &= e^{\mu(t+s)+\sigma^2(t+s)^2/2} e^{\mu(t-s)+\sigma^2(t-s)^2/2} \\ &= e^{2\mu t+\sigma^2 t^2} e^{\sigma^2 s^2} \end{aligned}$$

- This is the joint MGF of the sum of a normal random variable with mean 2μ and variance $2\sigma^2$ and an independent normal random variable with mean 0 and variance $2\sigma^2$.

7.7* Implications of Independence

- **Theorem:** Suppose X and Y are independent, then for any measurable function $h(\cdot)$ and $g(\cdot)$,

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

- **Corollary:** If X and Y are independent, then $\text{Cov}(X, Y) = 0$.
- **Remarks:**
 - Uncorrelated (orthogonal) random variables may not be independent.
 - Suppose $X \sim N(0, \sigma^2)$, $Y = X^2$. Then $\text{Cov}(X, Y) = 0$, but X and Y are not independent.

7.7* Implications of Independence

- **Problem:** Suppose for any measurable functions $h(\cdot)$ and $q(\cdot)$, we have

$$E[h(X)q(Y)] = E[h(X)]E[q(Y)].$$

Are X and Y independent?

- **ANS:** For any $a, b \in \mathbb{R}$, let $h(x) = I_{(-\infty, a]}(x)$, $g(y) = I_{(-\infty, b]}(y)$, where $I(\cdot)$ is the indicator function. Then

$$F_{XY}(a, b) = E[h(X)g(Y)] = E[h(X)]E[g(Y)] = F_X(a)F_Y(b).$$

X and Y are independent.

7.7* Implications of Independence

- **Question:** Suppose for some $h(X)$ and $g(Y)$ are independent, X and Y are independent? If not, can you give a counter example?
- **Solution:** Suppose $h(X) = X^2$, $g(Y) = Y^2$, and the joint pdf of (X, Y) are as follows:

$$f_{XY}(x, y) = \begin{cases} \frac{1+xy}{4}, & |x| < 1, |y| < 1 \\ 0, & \text{otherwise} \end{cases}$$

You can prove that X^2 and Y^2 are independent, but X and Y are not independent.

7.8 Additional Properties of Normal Random Variables

- Bivariate Normal Distribution - We say that the random variables X and Y have a bivariate normal distribution if, for constants $\mu_x, \mu_y, \sigma_x > 0, \sigma_y > 0, -1 < \rho < 1$, their joint density function is given, for all $-\infty < x, y < \infty$, by

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_x}{\sigma} \right)^2 + \left(\frac{y-\mu_y}{\sigma} \right)^2 - 2\rho \frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} \right] \right\}.$$

- The MGF of bivariate normal distribution $N(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$ is

$$\begin{aligned} M_{XY}(t_1, t_2) &= E[\exp(t_1X + t_2Y)] \\ &= \exp \left\{ t_1\mu_x + t_2\mu_y + \frac{1}{2}t_1^2\sigma_x^2 + \frac{1}{2}t_2^2\sigma_y^2 + t_1t_2 \text{Cov}(X, Y) \right\} \end{aligned}$$

- The marginal distributions of (X, Y) are normal.
- The distribution of any linear combination of X and Y is normal.

7.8 Additional Properties of Normal Random Variables

- **Multivariate Normal Distribution.** The random vector $\mathbf{X} = (X_1, X_2, \dots, X_n)$ follow a joint normal distribution if their joint pdf

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{\det[2\pi\Sigma]}} \exp \left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right\},$$

where $\mathbf{x} = (x_1, \dots, x_n)'$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$, and Σ is a $n \times n$ symmetric and positive definite matrix.

- **Theorem:** The mgf of multivariate normal distribution $N(\boldsymbol{\mu}, \Sigma)$ is

$$M_{\mathbf{X}}(t_1, \dots, t_n) = \exp\{\boldsymbol{\mu}'\mathbf{t} + \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}\},$$

where $\mathbf{t} = (t_1, \dots, t_n)'$.

7.8 Additional Properties of Normal Random Variables

- **Proof.**

$$\begin{aligned}M_X(\mathbf{t}) &= \int e^{\mathbf{t}'\mathbf{x}} f_X(\mathbf{x}) d\mathbf{x} \\&= \int \frac{1}{\sqrt{\det[2\pi\Sigma]}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) + \mathbf{t}'\mathbf{x}\right\} d\mathbf{x} \\&= \int \frac{1}{\sqrt{\det[2\pi\Sigma]}} \\&\quad \times \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu} - \Sigma\mathbf{t})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu} - \Sigma\mathbf{t}) + \boldsymbol{\mu}'\mathbf{t} + \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}\right\} d\mathbf{x} \\&= \exp\left\{\boldsymbol{\mu}'\mathbf{t} + \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}\right\}.\end{aligned}$$

7.8 Additional Properties of Normal Random Variables

- **Remarks: Properties of multivariate normal distribution.**

Suppose random vector $(X_1, \dots, X_n)$ follow multivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then

- Marginal distributions of $(X_1, \dots, X_n)$ are normal.
- $E(X_i) = \mu_i$.
- $\text{Var}(X_i) = \Sigma_{i,i}$.
- $\text{Cov}(X_i, X_j) = \Sigma_{i,j}$.
- Pairwise independence is equivalent to joint independence in multivariate normal distribution.
- If $(X, Y) \sim$ bivariate normal distribution, then the best linear predictor (based on the linear function of X (Y) for Y (X)) is also the best predictor (based on all types functions of X (Y) for Y (X)).

7.8 Additional Properties of Normal Random Variables

- **Remark:** The joint distribution of two normal random variables may not be normal.
- **Example:** Let X and Y are random variables with joint pdf

$$f_{XY}(x, y) = \begin{cases} 2f_X(x)f_Y(y) & \text{if } xy > 0, \\ 0 & \text{otherwise,} \end{cases}$$

where

$$\begin{aligned} f_X(x) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \\ f_Y(y) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}. \end{aligned}$$

Then X and Y are normal random variables, but they are not jointly normally distributed.