Chapter 5 Expected Values

Jingrui Sun

Email: sunjr@sustech.edu.cn

Southern University of Science and Technology

September 8, 2024

Outline I

1. Part 5.1: The Expected Value of a Random Variable
   - Part 5.1.1: Idea
   - Part 5.1.2: Definition of Expected Values: Discrete Case
   - Part 5.1.3: Basic Examples (Discrete Case)
   - Part 5.1.4: Further Examples (Discrete Case)
   - Part 5.1.5: Expected Value of Continuous R.V.s
   - Part 5.1.6: Basic Examples (Continuous Case)
   - Part 5.1.7: Further Examples (Continuous Case)

2. Part 5.2: Expectation of Functions of Random Variables
   - Part 5.2.1: Problem
   - Part 5.2.2: Conclusion
   - Part 5.2.3: Examples
   - Part 5.2.4: The Second Moment
   - Part 5.2.5: Expectations of Function of Several r.v.s

Outline II

3. Part 5.3: Probability Theory Properties of Expectation, Variance and Standard Deviation
   - Part 5.3.1: Properties of Expectation
   - Part 5.3.2: Variance and Standard Deviation
   - Part 5.3.3: Examples
4. Part 5.4: Covariance and Correlation
   - Part 5.4.1: Covariance
   - Part 5.4.2: Variance Formula and Application in Statistics
   - Part 5.4.3: Correlation Coefficients
   - Part 5.4.4: Examples
   - Part 5.4.5: Covariance Matrix and Correlation Matrix
5. Part 5.5: Moment Generating Functions
   - Part 5.5.1: Moment Generating Functions
   - Part 5.5.2: Examples

Outline III

6 Part 5.6: Applications of Moment Generating Functions
- Part 5.6.1: Find Moments
- Part 5.6.2: Determine the Distribution
- Part 5.6.3: Statistic Application – Weak Law of Large Number
- Part 5.6.4: Summary of Application of m.g.f.s.

7 Part 5.7: Three Special Distributions
- Part 5.7.1: Chi-Square (χ^2)-Distribution
- Part 5.7.2: Student's (t)-Distribution
- Part 5.7.3: F-Distribution

8 Part 5.8: Summary of Mathematical Expectation
- Part 5.8.1: Mixing Case
- Part 5.8.2: Summary of Chapter 5: Mathematical Expectation

Outline

1. Part 5.1: The Expected Value of a Random Variable
2. Part 5.2: Expectation of Functions of Random Variables
3. Part 5.3: Probability Theory Properties of Expectation, Variance and Standard Deviation
4. Part 5.4: Covariance and Correlation
5. Part 5.5: Moment Generating Functions
6. Part 5.6: Applications of Moment Generating Functions
7. Part 5.7: Three Special Distributions

Part 5.1.1: Idea

1. Example: Play a game with a "computer", say.

If win, get \$10; if lose, get $\$-2$; if tie, get \$1.

Probabilities: win: $1/4$, loss: $1/2$, tie: $1/4$.

Let X denote the amount of cash you get after the play. Then X is a (discrete) random variable.

Question: How much one could get, averagely, after a play?
(This might decide whether one should try)

Not suitable just simply $[10 + (-2) + 1]/3$.
Why? (The probability of winning is not considered)

Part 5.1.1: Idea

In order to get the correct "average", consider as follows:

Suppose one play n times (n is very large, and for convenience, assume that $n = 4m$ where m is a positive integer.), then one can expect to win $n/4$ times and thus get $n/4 \times 10$.

Similarly one can expect to lose $n/2$ times and get $n/2 \times (-2)$, and expect to get tie $n/4$ times and thus get $n/4 \times 1$.

So, after play n times, one reasonably expects to get $n/4 \times 10 + n/2 \times (-2) + n/4 \times 1$.

Hence, "on the average", for each play one would get $[n/4 \times 10 + n/2 \times (-2) + n/4 \times 1]/n = 1/4 \times 10 + 1/2 \times (-2) + 1/4 \times 1 = 1.75$.

Part 5.1.1: Idea

The value 1.75 is the value one could expect to get – this value is $10 \times P(X = 10) + (-2) \times P(X = -2) + 1 \times P(X = 1)$,
i.e. summing over all the possible values $\times$ the corresponding probability,
i.e. "weighted average" with weighted function of probability. We shall call this value "1.75" the "expected value" of the r.v. X.

2. Notes:

(1) The "expected value" is a real number, NOT a r.v.
(2) This value may not be a possible value of X (In the above example, one gets either $\$10, \$1$ or $\$(-2)$ after a play. One never gets "1.75" after a play).
(3) However, this value is important in "understanding" the r.v. X.

Part 5.1.2: Definition of Expected Values: Discrete Case

1. Definition: Suppose X is a discrete r.v. with all possible values $x_1, x_2, \cdots, x_n, \cdots$ (finite or a sequence) together with the p.m.f.

$$p(x_i) = P(X = x_i).$$

Then the expected value of X, denoted by $E(X)$, is defined as

$$\begin{aligned} E(X) &= \sum_i x_i \cdot p(x_i) \\ &\equiv \sum_i x_i \cdot P(X = x_i). \end{aligned} \tag{5.1.1}$$

The expected value is also called "the mean value", "the mean", "the expectation", "the mathematical expectation", etc.

2. Note:

If all the possible value of X is a sequence, then there is a "convergence" problem in Definition (5.1.1). Hence we need the condition that

$$\sum_i |x_i| \cdot p(x_i) < +\infty. \tag{5.1.2}$$

Otherwise, "expected value" is undefined.

Part 5.1.3: Basic Examples (Discrete Case)

1. Let X be the outcome when rolling a fair dice. $E(X) = ?$
2. Binomial distribution: $X \sim B(n; p)$; $E(X) = ?$
3. Poisson Distribution: $X \sim \text{Poisson } (\lambda)$; $E(X) = ?$

Part 5.1.4: Further Examples (Discrete Case)

1. The Geometric Random Variable X:

X : Independent Bernoulli trials; the No. required until the 1st success!

2. The Negative Binomial Distribution

Recall $X \sim NB(p, r)$ (Negative binomial distribution with parameters p and r). The p.m.f. is

$$p_n \triangleq P\{X = n\} = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \quad n = r, r+1, \cdots$$

Part 5.1.4: Further Examples (Discrete Case)

Recall: X denotes the No. required to get rth successes in the independent Bernoulli trials, and thus

$$P\{X = n\} = p \cdot \binom{n-1}{r-1} p^{r-1}(1-p)^{(n-1)-(r-1)}$$
$$\equiv \binom{n-1}{r-1} p^r (1-p)^{n-r}$$

Therefore,

$$E(X) = \sum_{n=r}^{\infty} n \cdot P\{X = n\} = \sum_{n=r}^{\infty} n \binom{n-1}{r-1} p^r (1-p)^{n-r}. \quad (5.1.3)$$

but

$$n \binom{n-1}{r-1} = r \binom{n}{r}. \quad (5.1.4)$$

Substituting (5.1.4) into (5.1.3) yields

$$E(X) = \sum_{n=r}^{\infty} r \binom{n}{r} p^r (1-p)^{n-r}$$
$$= \frac{r}{p} \sum_{n=r}^{\infty} \binom{n}{r} p^{r+1} (1-p)^{n-r}.$$

[letting $m = n + 1$, then $n : r \to \infty \Leftrightarrow m : r + 1 \to \infty$ and thus]

$$E(X) = \frac{r}{p} \sum_{m=r+1}^{\infty} \binom{m-1}{r} p^{r+1} (1-p)^{m-1-r} = \frac{r}{p}.$$

Part 5.1.4: Further Examples (Discrete Case)

Why $\sum_{m=r+1}^{\infty} \begin{pmatrix} m-1 \\ r \end{pmatrix} p^{r+1}(1-p)^{m-1-r} = 1$?

Recall if

$$Y \sim NB(p, r+1)$$

then the p.m.f. is: for $m = r+1, r+2, \cdots$

$$\begin{aligned} P(Y = m) &= \begin{pmatrix} m-1 \\ r+1-1 \end{pmatrix} p^{r+1}(1-p)^{m-(r+1)} \\ &= \begin{pmatrix} m-1 \\ r \end{pmatrix} p^{r+1}(1-p)^{m-1-r}. \end{aligned}$$

and thus

$$\sum_{m=r+1}^{\infty} \binom{m-1}{r} p^{r+1}(1-p)^{m-1-r} = 1.$$

Therefore, if $X \sim NB(p, r)$, then

$$E(X) = \frac{r}{p}.$$

In particular, if $r = 1$ (geometric distribution), then

$$E(X) = \frac{1}{p}.$$

Part 5.1.5: Expected Value of Continuous R.V.s

1. Definition:

(Compare with the discrete case: summation should be replaced by integration and the p.m.f. should be replaced by p.d.f.)

Suppose X is a continuous r.v. with p.d.f. $f(x)$, then the expected value of X, denoted by $E(X)$, is defined by

$$E(X) = \int_{-\infty}^{+\infty} x \cdot f(x) dx. \tag{5.1.5}$$

2. Notes:

(1) Again, we need the condition that

$$\int_{-\infty}^{+\infty} |x| \cdot f(x)dx < +\infty. \tag{5.1.6}$$

(2) If X is a non-negative r.v. then

$$E(X) = \int_{0}^{+\infty} x \cdot f(x)dx.$$

Indeed, for this case, $f(x) \equiv 0$ if $x < 0$ and thus

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} xf(x)dx = \int_{-\infty}^{0} xf(x)dx + \int_{0}^{+\infty} xf(x)dx \\ &= \int_{-\infty}^{+\infty} 0 \cdot xdx + \int_{0}^{+\infty} xf(x)dx \\ &= \int_{0}^{+\infty} xf(x)dx. \end{aligned}$$

Part 5.1.6: Basic Examples (Continuous Case)

1. Uniform distribution:

Easy, also see below.

2. Exponential distribution:

Recall: X is exponentially distributed with parameter λ if the p.d.f. is

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0. \end{cases}$$

Then

Part 5.1.6: Basic Examples (Continuous Case)

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{+\infty} xf(x)dx = \int_{0}^{+\infty} x \cdot f(x)dx + 0 = \int_{0}^{+\infty} x \cdot \lambda e^{-\lambda x} dx \\
&= \int_{0}^{+\infty} xd\left(-e^{-\lambda x}\right) = x\left(-e^{-\lambda x}\right)\Big|_{0}^{+\infty} - \int_{0}^{+\infty} \left(-e^{-\lambda x}\right) dx \\
&= 0 - 0 + \int_{0}^{+\infty} e^{-\lambda x} dx = \left[-\frac{1}{\lambda} e^{-\lambda x}\right]\Big|_{0}^{+\infty} = 0 - \left(-\frac{1}{\lambda} e^{0}\right) \\
&= \frac{1}{\lambda}.
\end{aligned}
$$

Conclusion: $E(X) = 1/\lambda$.

3. Normal Distribution:

(1) Standard Normal Distribution:

Recall $X \sim N(0,1)$, then

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Thus

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x \cdot e^{-\frac{x^2}{2}} dx.$$

It is easy to see then $E(X) = 0$.

Part 5.1.6: Basic Examples (Continuous Case)

Indeed,

$$
\begin{aligned}
E(X) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{x^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} x \cdot e^{-\frac{x^2}{2}} dx + \frac{1}{\sqrt{2\pi}} \int_{0}^{+\infty} x \cdot e^{-\frac{x^2}{2}} dx.
\end{aligned}
$$

For the former, let $y = -x$, then

$$
\begin{aligned}
x: &\quad -\infty \to 0 \\
y: &\quad +\infty \to 0
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow \quad \int_{-\infty}^{0} x \cdot e^{-\frac{x^2}{2}} dx &= \int_{+\infty}^{0} (-y) \cdot e^{-\frac{(-y)^2}{2}} d(-y) \\
&= \int_{+\infty}^{0} y e^{-\frac{y^2}{2}} dy \\
&= -\int_{0}^{+\infty} y e^{-\frac{y^2}{2}} dy \\
&= -\int_{0}^{+\infty} x e^{-\frac{x^2}{2}} dx.
\end{aligned}
$$

Part 5.1.6: Basic Examples (Continuous Case)

Thus

$$E(X) = \frac{1}{\sqrt{2\pi}}(-1) \cdot \int_0^{+\infty} xe^{-\frac{x^2}{2}} dx + \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} xe^{-\frac{x^2}{2}} dx = 0.$$

Conclusion: If $X \sim N(0,1)$, then $E(X) = 0$.

Intuitively, clear!! Just recall the graph of $f(x)$: is symmetric with $x = 0$.

Easy to see $\int_{-\infty}^{+\infty} xf(x)dx = 0$.

However, one needs to check $\int_{-\infty}^{+\infty} |x| \cdot f(x)dx < \infty$. Since otherwise, the above arguments may not be valid. But

Part 5.1.6: Basic Examples (Continuous Case)

$$
\begin{aligned}
\int_{-\infty}^{+\infty} |x| f(x) dx &= \int_{-\infty}^{0} |x| f(x) dx + \int_{0}^{+\infty} |x| f(x) dx \\
&= \int_{-\infty}^{0} (-x) f(x) dx + \int_{0}^{+\infty} x f(x) dx \\
&= \int_{+\infty}^{0} y f(-y)(-1) dy + \int_{0}^{+\infty} x f(x) dx \\
&= \int_{0}^{+\infty} y f(y) dy + \int_{0}^{+\infty} x f(x) dx \\
&= 2 \int_{0}^{+\infty} x f(x) dx.
\end{aligned}
$$

Hence only need to check that

$$\int_0^{+\infty} xf(x)dx = \int_0^{+\infty} x \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx < +\infty,$$

or, only need to check $\int_0^{+\infty} xe^{-\frac{x^2}{2}} dx < +\infty$?

But for $0 < y < +\infty$,

$$\int_0^y xe^{-\frac{x^2}{2}} dx = \int_0^y (-1)de^{-\frac{x^2}{2}} = (-1)\left[e^{-\frac{x^2}{2}}\right]\Big|_0^y = -e^{-\frac{y^2}{2}} + 1.$$

Letting $y \to +\infty$ we get $\int_0^{+\infty} xe^{-\frac{x^2}{2}} dx = 1 < +\infty$.

Part 5.1.6: Basic Examples (Continuous Case)

(2) General Normal Distribution:

Suppose $X \sim N\left(\mu, \sigma^2\right), E(X) = ?$

Recall: $X \sim N\left(\mu, \sigma^2\right)$, then

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x \cdot f(x) dx = \int_{-\infty}^{+\infty} x \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} x \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx. \end{aligned}$$

Part 5.1.6: Basic Examples (Continuous Case)

Let $y = \frac{x-\mu}{\sigma}$, then $x = \mu + \sigma y$, $dx = \sigma dy$, and

$$x : -\infty \to +\infty \quad \Leftrightarrow \quad y : -\infty \to +\infty \quad (\because \sigma > 0).$$

$$\begin{aligned}
E(X) &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}(\mu+\sigma y)e^{-\frac{y^2}{2}}\sigma dy = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}(\mu+\sigma y)e^{-\frac{y^2}{2}}dy \\
&= \frac{\mu}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-\frac{y^2}{2}}dy + \frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}ye^{-\frac{y^2}{2}}dy \\
&= \frac{\mu}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-\frac{y^2}{2}}dy + 0 \\
&= \mu.
\end{aligned}$$

Conclusion: If $X \sim N(\mu, \sigma^2)$, then $E(X) = \mu$.

Part 5.1.7: Further Examples (Continuous Case)

1. Gamma Distribution: $E(X) = ?$

Recall $X \sim \Gamma(\alpha, \lambda)$ (where $\alpha > 0, \lambda > 0$) if p.d.f. is given by

$$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, & \text{if } x \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

where the gamma function $\Gamma(\alpha)$ is defined as

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

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_{0}^{+\infty} xf(x)dx = \int_{0}^{+\infty} x \cdot \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} dx$$
$$= \frac{1}{\Gamma(\alpha)} \int_{0}^{+\infty} e^{-\lambda x} \cdot (\lambda x)^{\alpha} dx = \frac{1}{\Gamma(\alpha) \cdot \lambda} \int_{0}^{+\infty} \lambda e^{-\lambda x} (\lambda x)^{\alpha} dx$$
$$= \frac{1}{\lambda \cdot \Gamma(\alpha)} \int_{0}^{+\infty} \lambda e^{-\lambda x} \cdot (\lambda x)^{\alpha+1-1} dx$$

let $\lambda x = y$ then $dy = \lambda dx$ or $dx = \frac{1}{\lambda} dy$ and

$$x : 0 \to +\infty \quad \Leftrightarrow \quad y : 0 \to +\infty \quad (\because \lambda > 0)$$

Hence

$$\begin{aligned}
E(X) &= \frac{1}{\lambda \cdot \Gamma(\alpha)} \int_0^{+\infty} \lambda e^{-y} \cdot y^{\alpha+1-1} \cdot \frac{1}{\lambda} dy \\
&= \frac{1}{\lambda \Gamma(\alpha)} \int_0^{+\infty} e^{-y} \cdot y^{\alpha+1-1} dy \\
&= \frac{1}{\lambda \Gamma(\alpha)} \cdot \Gamma(\alpha+1) = \frac{1}{\lambda \Gamma(\alpha)} \cdot \alpha \Gamma(\alpha) \\
&= \frac{\alpha}{\lambda}.
\end{aligned}$$

2. The standard Cauchy distribution: $E(X)$ does not exist.

Recall X obey the Cauchy distribution with parameters μ if

$$\text{p.d.f.}: \quad f(x) = \frac{1}{\pi} \cdot \frac{1}{1 + (x-\mu)^2} \quad (-\infty < x < +\infty).$$

If $\mu = 0$, then the p.d.f. of the standard Cauchy distribution is

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1 + x^2} \quad (-\infty < x < +\infty)$$

Note that $f(-x) = f(x)$, but we can not say

$$E(Z) = \int_{-\infty}^{+\infty} x f(x) dx = 0,$$

since we haven't checked the condition $\int_{-\infty}^{+\infty} |x| f(x) dx < +\infty$ yet.

Now,

$$\begin{aligned}\int_{-\infty}^{+\infty}|x|f(x)dx &= \int_{-\infty}^{0}(-x)f(x)dx+\int_{0}^{+\infty}xf(x)dx\\ &= 2\int_{0}^{+\infty}xf(x)dx = 2\int_{0}^{+\infty}x\cdot\frac{1}{\pi}\cdot\frac{1}{1+x^2}dx\\ &= \frac{1}{\pi}\int_{0}^{+\infty}\frac{d(1+x^2)}{1+x^2} = \left[\ln\left(1+x^2\right)\right]\Big|_{0}^{+\infty}\\ &= +\infty.\end{aligned}$$

Hence $E(X)$ does not exist!!! Similarly for other Cauchy r.v.s.

3. A Remark:

Note that, for both discrete and continuous r.v., we have

$$X(\omega) \geq 0 \ (\forall \omega) \quad \Rightarrow \quad E(X) \geq 0.$$

This is clear either from the definitions or from the meaning of Expectation.

Indeed, $E(X) = \sum_x xp(x) \geq 0$ (for non-negative discrete r.v.) is clear since all $x \geq 0$.

For continuous r.v. (non-negative)

$$\begin{aligned} \int_{-\infty}^{+\infty} xf(x)dx &= \int_{-\infty}^{0} xf(x)dx + \int_{0}^{+\infty} xf(x)dx \\ &= \int_{0}^{+\infty} xf(x)dx \geq 0. \end{aligned}$$

Hence, in the future, if we could prove that

$$E(X - Y) = E(X) - E(Y),$$

then we would be able to get the conclusion that:

$$X(\omega) \geq Y(\omega)\ (\forall \omega) \quad \Rightarrow \quad E(X) \geq E(Y).$$

Indeed, if $X(\omega) \geq Y(\omega)$ then

$$X(\omega) - Y(\omega) \geq 0 \quad \Rightarrow \quad E(X - Y) \geq 0.$$

Now if $E(X - Y) = E(X) - E(Y)$, then we obtain $E(X) \geq E(Y)$.

Outline

1. Part 5.1: The Expected Value of a Random Variable
2. Part 5.2: Expectation of Functions of Random Variables
3. Part 5.3: Probability Theory Properties of Expectation, Variance and Standard Deviation
4. Part 5.4: Covariance and Correlation
5. Part 5.5: Moment Generating Functions
6. Part 5.6: Applications of Moment Generating Functions
7. Part 5.7: Three Special Distributions

Part 5.2.1: Problem

1. Function: Suppose X is a r.v., then easy to see X^2, for example, is also a random variable.

Let $Y = X^2$, then Y is a r.v.. This random variable Y is called a function of this r.v. X. [The function form is $y = x^2$.]

Similarly, $Y = X^3$, $Y = e^X$, $Y = c_1X + c_2X^2 + c_3 \cdot X^5 \cdot e^X$, (where c_1, c_2, c_3 are constants) are all functions of the r.v. X.

Corresponding function forms are $y = x^3$, $y = e^x$, $y = c_1x + c_2x^2 + c_3x^5 \cdot e^x$, etc.

Part 5.2.1: Problem

In general, if $y = g(x)$ is an ordinary function, X is a r.v., then $g(X)$ is also a r.v., denoted by, say,

$$Y = g(X).$$

Question: $E(Y) = E(g(X)) = ?$

The key thing here is that we <u>only know</u> the p.m.f. (or p.d.f.) of X (but not Y!!)

Part 5.2.1: Problem

2. Expectation of the Function of the Random Variable:

To calculate $E(Y) = E(g(X))$, one method is:

First try to find the p.m.f. (or p.d.f.) of Y and then use the definition.

However, this method is usually quite difficult or even impossible.

Fortunately, we have the following alternative method.

Part 5.2.2: Conclusion

1. **Theorem:** Suppose that X is a r.v. and that $Y = g(X)$ where $y = g(x)$ is a function.

   (i) If X is discrete with p.m.f. $p(x)$, then

   $$E(Y) \equiv E(g(X)) = \sum_{x} g(x)p(x). \tag{5.2.1}$$

   (ii) If X is continuous with p.d.f. $f(x)$, then

   $$E(Y) \equiv E(g(X)) = \int_{-\infty}^{+\infty} g(x)f(x)dx. \tag{5.2.2}$$

2. Meaning: (5.2.1) means

$$E(Y) = \sum_i g(x_i) \cdot p(x_i),$$

where the summation is taken for all possible values of X.

[$g(x_i)$ then: all the possible values of the r.v. $Y = g(X)$.]

(5.2.2): Integrand: $f(x) \cdot g(x)$.

Part 5.2.3: Examples

Example 1. Find $E(X^2)$ with X having p.m.f.

$$P(X=-1)=0.2, \quad P(X=0)=0.5, \quad P(X=1)=0.3.$$

Solution.

$$\begin{aligned} E\left(X^2\right) &= (-1)^2 \cdot P(X=-1) + 0^2 \cdot P(X=0) + 1^2 \cdot P(X=1) \\ &= P(X=-1) + P(X=1) \\ &= 0.2 + 0.3 = 0.5. \end{aligned}$$

Note that

$$\begin{aligned} E(X) &= (-1) \cdot P(X=-1) + 0 \cdot P(X=0) + 1 \cdot P(X=1) \\ &= -0.2 + 0.3 = 0.1 \quad \Rightarrow \quad (E(X))^2 = 0.01 \end{aligned}$$

So, usually, $E\left(X^2\right) \neq (E(X))^2$.

Part 5.2.3: Examples

Example 2. $X \sim B(n, p)$. $E(X^2) =?$

Solution.

$$E(X^2) = \sum_{k=0}^{n} k^2 \cdot P(X = k) = \sum_{k=1}^{n} k^2 \cdot \begin{pmatrix} n \\ k \end{pmatrix} p^k q^{n-k}.$$

Similarly, $E(X^m) = \sum_{k=1}^{n} k^m \begin{pmatrix} n \\ k \end{pmatrix} p^k q^{n-k}$.

Part 5.2.3: Examples

Example 3. $X \sim \text{Poisson}(\lambda)$. $E(X^2) = ?$

Solution.

$$E\left(X^2\right) = \sum_{k=0}^{n} k^2 \cdot P(X = k)$$

$$= \sum_{k=1}^{n} k^2 \cdot \binom{n}{k} p^k q^{n-k}.$$

Similarly, $E(X^m) = \sum_{k=1}^{n} k^m \cdot e^{-\lambda} \cdot \frac{\lambda^k}{k!}$.

Note: For any r.v. X, the quantity $E(X^m)$ $(m \geq 1)$ is called the m'th moment of X.

Example 4. Normal Distribution.

Suppose $X \sim N(\mu, \sigma^2)$, i.e. p.d.f.: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.

We know that $E(X) = \mu$.

Now, let $Y = a \cdot X + b$, where $a \neq 0$ and b are two constants.

$$E(Y) = ?$$

Part 5.2.3: Examples

$$
\begin{aligned}
E(Y) &= \int_{-\infty}^{+\infty} (ax+b)\cdot f(x)dx = \int_{-\infty}^{+\infty} [axf(x) + b\cdot f(x)]dx \\
&= \int_{-\infty}^{+\infty} ax\cdot f(x)dx + \int_{-\infty}^{+\infty} b\cdot f(x)dx \\
&= a\int_{-\infty}^{+\infty} x\cdot f(x)dx + b\cdot \int_{-\infty}^{+\infty} f(x)dx \\
&= a\cdot E(X) + b.
\end{aligned}
$$

Part 5.2.4: The Second Moment

1. Binomial Random Variables

Recall $X \sim B(n, p)$, then

$$P\{X = k\} = \begin{pmatrix} n \\ k \end{pmatrix} p^k (1-p)^{n-k}, \quad k = 0, 1, \cdots, n,$$

$$E(X) = np.$$

What is $E(X^2)$?

Part 5.2.4: The Second Moment

Solution: By the law of calculating the mean value of a function of random variables,

$$
\begin{aligned}
E(X^2) &= \sum_{k=0}^{n} k^2 P\{X = k\} \\
&= \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k^2 \binom{n}{k} p^k q^{n-k},
\end{aligned}
$$

where $q = 1 - p$. Write $k^2 = k(k-1+1) = k(k-1) + k$ and thus

Part 5.2.4: The Second Moment

$$E(X^2) = \sum_{k=1}^{n} [k(k-1)+k] \binom{n}{k} p^k q^{n-k}$$
$$= \sum_{k=1}^{n} k(k-1) \binom{n}{k} p^k q^{n-k} + \sum_{k=1}^{n} k \binom{n}{k} p^k q^{n-k}. \tag{5.2.3}$$

But $\sum_{k=1}^{n} k \binom{n}{k} p^k q^{n-k} = E(X) = np.$

Hence only need to consider

$$\sum_{k=1}^{n} k(k-1) \binom{n}{k} p^k q^{n-k}. \tag{5.2.4}$$

Part 5.2.4: The Second Moment

Consider $(px+q)^n = \sum_{k=0}^{n} \binom{n}{k} (px)^k q^{n-k}$. $\boxed{p+q=1}$

$$\Rightarrow \; n(px+q)^{n-1} \cdot p = \frac{d}{dx}(px+q)^n$$
$$= \sum_{k=1}^{n} \binom{n}{k} k(px)^{k-1} \cdot pq^{n-k}.$$

$$\Rightarrow \; n(px+q)^{n-1} = \sum_{k=1}^{n} k \binom{n}{k} (px)^{k-1} \cdot q^{n-k}.$$

$$\Rightarrow \; \frac{d}{dx}\left(n(px+q)^{n-1}\right) = \sum_{k=2}^{n} k \binom{n}{k} (k-1)(px)^{k-2} \cdot pq^{n-k},$$

or equivalently,

$$n(n-1)(px+q)^{n-2} \cdot p = \sum_{k=2}^{n} k(k-1) \binom{n}{k} (px)^{k-2} \cdot pq^{n-k}$$

Letting $x = 1$ and noticing $(px+q)^{n-2} = (p+q)^{n-2} = 1$ then yields

$$n(n-1)p = \sum_{k=2}^{n} k(k-1) \binom{n}{k} p^{k-1} \cdot q^{n-k}$$

Timing p on both side yields

$$n(n-1)p^2 = \sum_{k=2}^{n} k(k-1) \binom{n}{k} p^k q^{n-k}, \tag{5.2.5}$$

which yields the result (5.2.4).

Part 5.2.4: The Second Moment

Now substituting (5.2.4) into (5.2.3) yields

$$E(X^2) = n(n-1)p^2 + np = [(n-1)p + 1] \cdot np.$$

Conclusion: If $X \sim B(n, p)$, then

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

2. Poisson Random Variables

Recall $X \sim$ Poisson (λ), if

$$p_k = P\{X = k\} = e^{-\lambda} \cdot \frac{\lambda^k}{k!}, \quad k = 0, 1, \cdots$$

We have already known $E(X) = \sum_{k=0}^{\infty} kp_k = \sum_{k=1}^{\infty} ke^{-\lambda} \cdot \frac{\lambda^k}{k!} = \lambda$.

$$\begin{aligned} E(X^2) &= \sum_{k=0}^{\infty} k^2 e^{-\lambda} \cdot \frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} k^2 e^{-\lambda} \cdot \frac{\lambda^k}{k!} \\ &= \sum_{k=1}^{\infty} ke^{-\lambda} \cdot \frac{\lambda^k}{(k-1)!} = \lambda \sum_{k=1}^{\infty} ke^{-\lambda} \cdot \frac{\lambda^{k-1}}{(k-1)!}. \end{aligned}$$

Part 5.2.4: The Second Moment

Let $j = k - 1$. Then

$$
\begin{aligned}
E(X^2) &= \lambda \sum_{j=0}^{\infty} (j+1) e^{-\lambda} \cdot \frac{\lambda^j}{j!} \\
&= \lambda \left[\sum_{j=0}^{\infty} j e^{-\lambda} \cdot \frac{\lambda^j}{j!} + \sum_{j=0}^{\infty} e^{-\lambda} \cdot \frac{\lambda^j}{j!} \right] \\
&= \lambda [E(X) + 1] \\
&= \lambda(\lambda + 1) \qquad (\because E(X) = \lambda)
\end{aligned}
$$

In short: If $X \sim$ Poisson (λ), then

$$
\boxed{E(X) = \lambda, \quad E(X^2) = \lambda^2 + \lambda}
$$

Part 5.2.4: The Second Moment

3. Geometric Random Variables

Recall p.m.f.:

$$P\{X = n\} = p(1-p)^{n-1} = pq^{n-1}, \quad n = 1, 2, \cdots ,$$

and

$$E(X) = \frac{1}{p} \equiv \sum_{n=1}^{\infty} npq^{n-1}.$$

$E(X^2) =?$

Part 5.2.4: The Second Moment

$$
\begin{aligned}
E(X^2) &= \sum_{n=1}^{\infty} n^2 p q^{n-1} = p \sum_{n=1}^{\infty} n^2 q^{n-1} = p \sum_{n=1}^{\infty} \frac{d}{dq}(nq^n) \\
&= p \frac{d}{dq} \left(\sum_{n=1}^{\infty} n q^n \right) = p \frac{d}{dq} \left(\frac{q}{p} \sum_{n=1}^{\infty} n p q^{n-1} \right) \\
&= p \frac{d}{dq} \left(\frac{q}{1-q} E(X) \right) \quad \left(\because E(X) = \sum_{n=1}^{\infty} n p q^{n-1} \right) \\
&= p \frac{d}{dq} \left(\frac{q}{1-q} \cdot \frac{1}{p} \right) \quad \left(\because E(X) = \frac{1}{p} \right) \\
&= p \frac{d}{dq} \left(\frac{q}{(1-q)^2} \right) \quad (\because p = 1-q)
\end{aligned}
$$

Part 5.2.4: The Second Moment

Therefore (noting $q = 1 - p$),

$$
\begin{aligned}
E(X^2) &= p\frac{d}{dq}\left[q(1-q)^{-2}\right] \\
&= p\left[(1-q)^{-2} + q \cdot (-2)(1-q)^{-3} \cdot (-1)\right] \\
&= \frac{2}{p^2} - \frac{1}{p}.
\end{aligned}
$$

Part 5.2.4: The Second Moment

In short, if $X \sim Geo(p)$, then

$$\boxed{E(X) = \frac{1}{p}, \quad E(X^2) = \frac{2}{p^2} - \frac{1}{p}} \tag{5.2.6}$$

Note: By (5.2.6) we get

$$E(X^2) - [E(X)]^2 = \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} = \frac{1}{p^2} - \frac{1}{p} = \frac{1-p}{p^2}.$$

In the future we shall show that this is the variance.

Part 5.2.4: The Second Moment

4. The Uniform Random Variable (over $[a, b]$)

Recall p.d.f.:

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

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} xf(x)dx = \int_a^b xf(x)dx \\ &= \int_a^b x \cdot \frac{1}{b-a} dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]\Bigg|_a^b \\ &= \frac{a+b}{2}. \end{aligned}$$

Part 5.2.4: The Second Moment

$$
\begin{aligned}
E(X^2) &= \int_{-\infty}^{+\infty} x^2 f(x)dx = \int_a^b x^2 f(x)dx \\
&= \int_a^b \frac{x^2}{b-a}dx = \frac{1}{b-a}\left[\frac{x^3}{3}\right]\Bigg|_a^b \\
&= \frac{a^2+ab+b^2}{3}.
\end{aligned}
$$

Part 5.2.4: The Second Moment

In short

$$\boxed{E(X) = \frac{a+b}{2}, \quad E(X^2) = \frac{a^2+ab+b^2}{3}}$$

$$\Rightarrow: \quad E(X^2) - [E(X)]^2 = \frac{(b-a)^2}{12}.$$

In particular, if $X \sim U(0,1)$, then

$$E(X) = \frac{1}{2}, \quad E(X^2) = \frac{1}{3}, \quad E(X^2) - [E(X)]^2 = \frac{1}{12}.$$

Part 5.2.4: The Second Moment

5. Exponential Random Variables

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0. \end{cases} \qquad (\lambda > 0)$$

$$E(X) = \frac{1}{\lambda}, \quad E(X^2) = ?$$

Part 5.2.4: The Second Moment

Solutions:

$$
\begin{aligned}
E(X^2) &= \int_{-\infty}^{+\infty} x^2 f(x)dx = \int_0^{+\infty} x^2 \lambda e^{-\lambda x} dx \\
&= (-1)\int_0^{+\infty} x^2 de^{-\lambda x} \\
&= (-1)\left[x^2 e^{-\lambda x}\Big|_{x=0}^{x=+\infty} - \int_0^{+\infty} e^{-\lambda x} d(x^2) \right]
\end{aligned}
$$

[Integration by Parts!!]

Part 5.2.4: The Second Moment

$$E(X^2) = 0 + \int_0^{+\infty} e^{-\lambda x} \cdot 2x dx = \frac{2}{\lambda} \cdot \int_0^{+\infty} \lambda x e^{-\lambda x} dx = \frac{2}{\lambda} E(X)$$
$$= \frac{2}{\lambda} \cdot \frac{1}{\lambda} = \frac{2}{\lambda^2}.$$

In short,

$$E(X) = \frac{1}{\lambda}, \quad E(X^2) = \frac{2}{\lambda^2}, \quad E(X^2) - [E(X)]^2 = \frac{1}{\lambda^2}.$$

Part 5.2.4: The Second Moment

6. Normal: $X \sim N(\mu, \sigma^2)$. $E(X) = \mu$, $E(X^2) = ?$

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx = \int_{-\infty}^{+\infty} x^2 \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Let $\frac{x-\mu}{\sigma} = y$, then $x = \sigma y + \mu, dx = \sigma dy$

Part 5.2.4: The Second Moment

$$E(X^2) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} (\sigma y + \mu)^2 e^{-\frac{y^2}{2}} \sigma dy$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \left(\sigma^2 y^2 + 2\mu\sigma y + \mu^2\right) e^{-\frac{y^2}{2}} dy$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y^2 e^{-\frac{y^2}{2}} dy + \frac{2\mu\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y e^{-\frac{y^2}{2}} dy + \frac{\mu^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} dy$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y^2 e^{-\frac{y^2}{2}} dy + 0 + \mu^2.$$

Part 5.2.4: The Second Moment

If we could show that

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} y^2 e^{-\frac{y^2}{2}}\,dy = 1 \tag{5.2.7}$$

then

$$E(X^2) = \sigma^2 + \mu^2. \tag{5.2.8}$$

But (5.2.7) is easy. Indeed, by Integration by parts

Part 5.2.4: The Second Moment

$$
\begin{aligned}
\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} y^2 e^{-\frac{y^2}{2}}\,dy &= \frac{1}{\sqrt{2\pi}}(-1)\int_{-\infty}^{+\infty} y d\left(e^{-\frac{y^2}{2}}\right)\\
&= \frac{-1}{\sqrt{2\pi}}\left[ye^{-\frac{y^2}{2}}\Big|_{y=-\infty}^{y=+\infty} - \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}}\,dy\right]\\
&= \frac{-1}{\sqrt{2\pi}}\left[0 - 0 - \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}}\,dy\right]\\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}}\,dy = 1.
\end{aligned}
$$

Part 5.2.4: The Second Moment

Hence we have shown that if $X \sim N(\mu, \sigma^2)$, then

$$E(X) = \mu, \quad E(X^2) = \sigma^2 + \mu^2, \quad E(X^2) - [E(X)]^2 = \sigma^2.$$

In the future, we shall reobtain these conclusions more easily!

In particular, if $X \sim N(0, 1)$, then

$$E(X^2) = 1, \quad E(X^2) - (E(X))^2 = 1.$$

Part 5.2.5: Expectations of Function of Several r.v.s

Question: Suppose that $X_1, X_2, \cdots, X_n$ are r.v.s and

$$Y = g(X_1, X_2, \cdots, X_n),$$

where g is a function of n variables. How to find $E(Y)$?

Theorem. Suppose that $X_1, X_2, \cdots, X_n$ are r.v.s and

$$Y = g(X_1, X_2, \cdots, X_n),$$

where g is a function of n variables. Then

(i) If $X_1, X_2, \cdots, X_n$ are all discrete with joint p.m.f. p, then

$$E(Y) = \sum_{x_1} \cdots \sum_{x_n} g(x_1, \cdots, x_n) p(x_1, \cdots, x_n).$$

(ii) If $X_1, X_2, \cdots, X_n$ are all continuous with joint p.d.f. f, then

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

Example 1.

(1) The joint p.m.f. of (X, Y) is given by

x	0	1	2
0	$1/6$	$1/3$	$1/12$
1	$2/9$	$1/6$	0
2	$1/36$	0	0

Find $E(X)$, $E(Y)$, and $E(X+Y)$.

(2) The joint p.d.f. $f(x, y)$ of X and Y is given by

$$f(x,y) = \begin{cases} \frac{2}{7}(x+2y), & \text{if } 0 < x < 1,\ 1 < y < 2, \\ 0, & \text{otherwise.} \end{cases}$$

Find $E(X/Y^3)$.

Outline

1. Part 5.1: The Expected Value of a Random Variable
2. Part 5.2: Expectation of Functions of Random Variables
3. Part 5.3: Probability Theory Properties of Expectation, Variance and Standard Deviation
4. Part 5.4: Covariance and Correlation
5. Part 5.5: Moment Generating Functions
6. Part 5.6: Applications of Moment Generating Functions
7. Part 5.7: Three Special Distributions

Proposition 1.

- If X is a constant C, then $E(C) = C$.
- (Linearity) If X, Y are r.v.s and a, b are real numbers, then $E(a \cdot X + b \cdot Y) = a \cdot E(X) + b \cdot E(Y)$.

Remark 2. Generally, for r.v.s X and Y,

$$\begin{aligned} E(X \cdot Y) &\neq Y \cdot E(X), & E(X^2) &\neq X \cdot E(X), \\ E(X \cdot Y) &\neq E(X) \cdot E(Y), & E(X \cdot X) &\neq E(X) \cdot E(X). \end{aligned}$$

Proposition 3. If X and Y are independent r.v.s, then

$$E(X \cdot Y) = E(X) \cdot E(Y).$$

Corollary 4. If X and Y are independent r.v.s and g and h are functions, then

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

Meaning: if X and Y are independent, then for example,

$$E\left[X^3 \cdot e^Y\right] = E\left[X^3\right] \cdot E\left[e^Y\right].$$

We summarize the above results as follows.

Theorem 5. The expectation satisfies the following properties:

(i) $E(C) = C$ for all constants C.

(ii) $E(aX + bY) = aE(X) + bE(Y)$ for any r.v.s X, Y and any constants a, b.

(iii) If X and Y are independent, then $E(XY) = E(X)E(Y)$.

Part 5.3.2: Variance and Standard Deviation

A motivating example. Consider two r.v.s X and Y:

$$X = \begin{cases} -1, & \text{with probability } \frac{1}{2}, \\ +1, & \text{with probability } \frac{1}{2}. \end{cases} \quad Y = \begin{cases} -100, & \text{with probability } \frac{1}{2}, \\ +100, & \text{with probability } \frac{1}{2}. \end{cases}$$

Easy to see $E(X) = E(Y) = 0$.

However, the "spread" of values for X and Y is different. There is <u>much greater spread</u> in the possible values of Y.

How to measure the "spread": $E\left[(X - E(X))^2\right]$. Why so?

Definition 1. The variance of the r.v. X, denoted by $\text{Var}(X)$, is defined by

$$\text{Var}(X) = E\left[(X - E(X))^2\right].$$

$\sqrt{\text{Var}(X)}$ is called the standard deviation of X.

Remark 2.

(1) Note: $X - E(X)$ is a function of X. So $X - E(X)$, and hence $(X - E(X))^2$, is a r.v.. The expected value of $(X - E(X))^2$ is called the variance of the original r.v. X.

(2) Note that $Y \triangleq (X - E(X))^2$ is a non-negative r.v. So $E[(X - E(X))^2] \geq 0$ and thus $\sqrt{\text{Var}(X)}$ is meaningful.

Proposition 3. For discrete or continuous r.v. X,

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

Remark 4. If $\text{Var}\,(X) = 0$, i.e., $E\left[(X - E(X))^2\right] = 0$, then $[X - E(X)]^2 \geq 0$ would imply that $X(\omega) = E(X)$ for almost all ω.

Proposition 5. Let X, Y be a r.v.s and a, b constants. Then

(i) $\operatorname{Var}(X) \geq 0$;

(ii) $\operatorname{Var}(X) = 0$ iff X is a constant r.v. and in this case, $X = E(X)$;

(iii) $\operatorname{Var}(a \cdot X + b) = a^2 \cdot \operatorname{Var}(X)$;

(iv) If X and Y are independent, then

$$\operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y).$$

Note that $\operatorname{Var}(a \cdot X) \neq a \cdot \operatorname{Var}(X)$ in general.

1. Bernoulli r.v.: X takes on values 0 and 1, with probability $1-p$ and p, respectively.

$$E(X) = \sum_i x_i \cdot p(x_i) = 0 \times P(X=0) + 1 \times P(X=1) = p,$$

$$E(X^2) = \sum_i x_i^2 \cdot p(x_i) = 0^2 \times P(X=0) + 1^2 \times P(X=1) = p,$$

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

2. Binomial r.v. $X \sim B(n,p)$: Recall that

$$E(X) = np, \quad E(X^2) = np[n-1)p+1].$$

Thus, $\mathrm{Var}\,(X) = E(X^2) - [E(X)]^2 = np(1-p)$.

3. Poisson r.v. $X \sim$ Poisson (λ): Recall

$$E(X) = \lambda, \quad E(X^2) = \lambda^2 + \lambda.$$

Hence,

$$\text{Var}\,(X) = E(X^2) - [E(X)]^2 = \lambda.$$

4. Geometric r.v. $X \sim$ Geo (p): Recall

$$E(X^2) - [E(X)]^2 = \frac{1-p}{p^2}.$$

So $\text{Var}\,(X) = \frac{1-p}{p^2}$.

5. Uniform r.v. $X \sim U[a, b]$: Recall

$$\operatorname{Var}(X) = E(X^2) - [E(X)]^2 = \frac{(b-a)^2}{12}.$$

6. Exponential r.v. $X \sim \operatorname{Exp}(\lambda)$: Recall

$$E(X) = \frac{1}{\lambda}, \quad E(X^2) = \frac{2}{\lambda^2}.$$

So $\operatorname{Var}(X) = E(X^2) - [E(X)]^2 = \frac{1}{\lambda^2}$.

7. Normal r.v. $X \sim N(\mu, \sigma^2)$:

$$E(X) = \mu, \quad \operatorname{Var}(X) = \sigma^2.$$

8. The p.d.f. of X is $f(x) = \frac{1}{\sqrt{2\pi}\cdot\sqrt{5}} e^{-\frac{(x+3)^2}{10}}$. Find $E(X)$ and $\mathrm{Var}(X)$.

9. T p.d.f. of X is given by

$$f(x) = \begin{cases} \frac{1}{x(\ln 3)}, & 1 < x < 3, \\ 0, & \text{otherwise.} \end{cases}$$

   Find $E(X)$, $E(X^2)$, $\mathrm{Var}(X)$, and $E(2X^2 - 3X + 1)$.

10. X and Y have joint p.d.f.

$$f(x, y) = \begin{cases} x + y, & 0, \le x \le 1,\ 0 \le y \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

   Find $E(X + Y)^2$.

Outline

1. Part 5.1: The Expected Value of a Random Variable
2. Part 5.2: Expectation of Functions of Random Variables
3. Part 5.3: Probability Theory Properties of Expectation, Variance and Standard Deviation
4. Part 5.4: Covariance and Correlation
5. Part 5.5: Moment Generating Functions
6. Part 5.6: Applications of Moment Generating Functions
7. Part 5.7: Three Special Distributions

Single r.v.: expectation and variance.

Several r.v.s: need a quantity to "measure" the relationship among them.

Definition 1. The covariance of two r.v.s X and Y, denoted by $\mathrm{Cov}(X, Y)$, is defined by

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

Proposition 2. We have

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

Proposition 3. Let X, Y, Z, W be r.v.s and a, b, c, d constants.

(i) $\operatorname{Cov}(X, Y) = \operatorname{Cov}(Y, X)$ and $\operatorname{Cov}(X, X) = \operatorname{Var}(X)$.

(ii) If X and Y are independent, then $\operatorname{Cov}(X, Y) = 0$. In particular, $\operatorname{Cov}(X, a) = \operatorname{Cov}(a, X) = 0$.

(iii) (Bilinearity) $\operatorname{Cov}(aX + bY, Z) = a\operatorname{Cov}(X, Z) + b\operatorname{Cov}(Y, Z)$. Consequently,

$$\begin{aligned}
&\operatorname{Cov}(aX + bY, cZ + dW) \\
&= ac\operatorname{Cov}(X, Z) + bc\operatorname{Cov}(Y, Z) \\
&\quad + ad\operatorname{Cov}(X, W) + bd\operatorname{Cov}(Y, W).
\end{aligned}$$

Corollary 4. Let $X_1, X_2, \cdots, X_n; Y_1, Y_2, \cdots, Y_m$ be r.v.s and $a_0, a_1, \cdots, a_n; b_0, b_1, \cdots, b_m$ constants. Then

$$\mathrm{Cov}\left(a_0 + \sum_{i=1}^{n} a_i X_i, b_0 + \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \mathrm{Cov}\,(X_i, Y_j).$$

Remark 5. $\mathrm{Cov}\,(X, Y)$ can be any real number.

if $\mathrm{Cov}\,(X, Y) = 0$, we say X and Y are uncorrelated;

if $\mathrm{Cov}\,(X, Y) > 0$, we say X and Y are positively correlated;

if $\mathrm{Cov}\,(X, Y) < 0$, we say X and Y are negatively correlated.

Proposition 1. Let $X_1, X_2, \ldots, X_n$ be r.v.s. Then

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

Note that if the r.v.s $X_1, X_2, \ldots, X_n$ are (mutual) independent, then the above formula yields

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

Application in Statistics.

Suppose $X_1, X_2, \cdots, X_n$ are i.i.d. r.v.s having the finite common expected value μ and variance σ^2. Let

$$\bar{X} \triangleq \frac{1}{n} \sum_{i=1}^{n} X_i \quad \text{(sample mean)}$$

$$S^2 \triangleq \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 \quad \text{(sample variance)}$$

Find $E(\bar{X})$, $\mathrm{Var}\left(\bar{X}\right)$, and $E\left[\frac{S^2}{n-1}\right]$.

Proposition 2. Suppose $X_1, X_2, \cdots, X_n$ are i.i.d. r.v.s with common mean μ and common variance σ^2 which are assumed to be finite. Let

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S^2 = \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2.$$

Then

$$E(\bar{X}) = \mu, \quad \mathrm{Var}\left(\bar{X}\right) = \frac{1}{n}\sigma^2, \quad E\left(\frac{S^2}{n-1}\right) = \sigma^2.$$

In statistics, $\frac{S^2}{n-1}$ is called the unbiased estimator of σ^2.

Part 5.4.3: Correlation Coefficients

Definition 1. The correlation coefficient between two r.v.s X and Y is defined as

$$\rho = \rho(X, Y) \triangleq \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X) \cdot \operatorname{Var}(Y)}}.$$

Note: We assume that neither X nor Y is a constant so that $\operatorname{Var}(X) > 0, \operatorname{Var}(Y) > 0$ and thus the above definition is well-defined.

Proposition 2. For any r.v.s X and Y,

$$-1 \leq \rho(X, Y) \leq 1.$$

Remark 3. By Proposition 2, we have

$$[\mathrm{Cov}\,(X, Y)]^2 \leq \mathrm{Var}\,(X) \cdot \mathrm{Var}\,(Y).$$

$$\begin{array}{lcll}
\rho = 0 & \Leftrightarrow & \mathrm{Cov}\,(X, Y) = 0: & \text{uncorrelated;} \\
\rho > 0 & \Leftrightarrow & \mathrm{Cov}\,(X, Y) > 0: & \text{positively correlated;} \\
\rho < 0 & \Leftrightarrow & \mathrm{Cov}\,(X, Y) < 0: & \text{negatively correlated.}
\end{array}$$

Remark 4. With $\mathrm{Var}(X) = \sigma_X^2$ and $\mathrm{Var}(Y) = \sigma_Y^2$, that

$$0 \leq \mathrm{Var}\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) = 2\Big[1 + \rho(X, Y)\Big],$$
$$0 \leq \mathrm{Var}\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) = 2\Big[1 - \rho(X, Y)\Big].$$

Thus,

$$\rho(X, Y) = -1 \quad \Leftrightarrow \quad Y = a + bX \quad \text{with } b = -\frac{\sigma_Y}{\sigma_X} < 0,$$

$$\rho(X, Y) = 1 \quad \Leftrightarrow \quad Y = a + bX \quad \text{with } b = \frac{\sigma_Y}{\sigma_X} > 0.$$

If X and Y are independent, then $\rho = 0$ and thus X and Y are uncorrelated, but the <u>converse is usually not true</u>.

X and Y obeys a bivariate normal distribution if the joint p.d.f. is

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\right.$$
$$\left.\cdot\left[\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right]\right\}.$$

We can prove that $\operatorname{Cov}(X, Y) = \rho \cdot \sigma_1 \cdot \sigma_2$. Recall that $\operatorname{Var}(X) = \sigma_1^2$, $\operatorname{Var}(Y) = \sigma_2^2$, then $\rho(X, Y) = \frac{\rho \cdot \sigma_1 \cdot \sigma_2}{\sigma_1 \cdot \sigma_2} = \rho$.

We know that, for bivariate normal distribution, X and Y are independent iff $\rho = 0$ (it is not true for general case).

Theorem. Suppose (X, Y) is bivariate normal distributed. Then X and Y are independent iff X and Y are uncorrelated, i.e., $\rho = 0$.

For a n-dimiensional random (column) vector $X = (X_1, X_2, \cdots, X_n)^\top$, we define

$$E(X) \triangleq (E(X_1), E(X_2), \cdots, E(X_n))^\top.$$

Also we define the covariance and correlation matrices, resp., as

$$C = \begin{bmatrix} \operatorname{Cov}(X_1, X_1) & \operatorname{Cov}(X_1, X_2) & \cdots & \operatorname{Cov}(X_1, X_n) \\ \operatorname{Cov}(X_2, X_1) & \operatorname{Cov}(X_2, X_2) & \cdots & \operatorname{Cov}(X_2, X_n) \\ \cdots & \cdots & & \cdots \\ \operatorname{Cov}(X_n, X_1) & \operatorname{Cov}(X_n, X_2) & \cdots & \operatorname{Cov}(X_n, X_n) \end{bmatrix},$$

$$R = \begin{bmatrix} \rho(X_1, X_1) & \rho(X_1, X_2) & \cdots & \rho(X_1, X_n) \\ \rho(X_2, X_1) & \rho(X_2, X_2) & \cdots & \rho(X_2, X_n) \\ \cdots & \cdots & & \cdots \\ \rho(X_n, X_1) & \rho(X_n, X_2) & \cdots & \rho(X_n, X_n) \end{bmatrix}.$$

Note that C and R are symmetric and all the diagonal elements are $\mathrm{Cov}\,(X_i, X_i) = \mathrm{Var}\,(X_i)$ and $\rho(X_i, X_i) = 1$, resp., $i = 1, 2, \cdots, n$.

Proposition. C is a nonnegative definite matrix.

Key point: covariance matrix has the representation

$$C = E\left[(\bar{X}_1, \bar{X}_2, \cdots, \bar{X}_n)^\top (\bar{X}_1, \bar{X}_2, \cdots, \bar{X}_n)\right],$$

where $\bar{X}_i = X_i - E\,(X_i)\,, i = 1, \cdots, n$.

Moreover, if $X_1, X_2, \ldots, X_n$ are (mutual) independent and $E(X_i^2) > 0$ (i.e., $X_i \not\equiv 0$) for all $i = 1, 2, \cdots, n$, then C is positive definite.

Outline

1. Part 5.1: The Expected Value of a Random Variable
2. Part 5.2: Expectation of Functions of Random Variables
3. Part 5.3: Probability Theory Properties of Expectation, Variance and Standard Deviation
4. Part 5.4: Covariance and Correlation
5. Part 5.5: Moment Generating Functions
6. Part 5.6: Applications of Moment Generating Functions
7. Part 5.7: Three Special Distributions

Part 5.5.1: Moment Generating Functions

Definition 1. The moment generating function (m.g.f.) of a r.v. X, denoted by $M_X(t)$, is defined (for those $t \in \mathbb{R}$ s.t. $E\left[e^{tX}\right] < \infty$) as

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

Remark 2. If X is a continuous r.v. with p.d.f. $f(x)$, then

$$M_X(t) = \int_{-\infty}^{+\infty} e^{tx} f(x) dx.$$

If X is a discrete r.v. with p.m.f. $p(x_i)$, then

$$M_X(t) = \sum_i e^{tx_i} p(x_i).$$

Note: $M_X(t)$ may not exist for all real values t.

Proposition 3. The m.g.f. has the following properties:

(i) $M_X(0) = 1$.

(ii) If $Y = aX + b$ $(a, b \in \mathbb{R})$, then $M_Y(t) = e^{bt} M_X(at)$.

(iii) If X and Y are independent, then $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$.

(iv) If $X_1, X_2, \cdots, X_n$ are independent, then

$$M_{X_1+X_2+\cdots+X_n}(t) = M_{X_1}(t) M_{X_2}(t) \cdots M_{X_n}(t) = \prod_{i=1}^{n} M_{X_i}(t).$$

(v) If $Y = b_0 + \sum_{i=1}^{n} b_i X_i$, where $X_1, X_2, \cdots, X_n$ are independent r.v.s and b_i are constants, then $M_Y(t) = e^{b_0 t} \cdot \prod_{i=1}^{n} M_{X_i}(b_i t)$.

(vi) Two r.v.s X and Y have the same m.g.f. iff they have the same p.d.f. (the proof of this part is vary complicated!).

1. Binomial Distribution: $X \sim B(n, p)$.

p.m.f.: $P\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n.$

$$
\begin{aligned}
M_X(t) = E\left[e^{tX}\right] &= \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} \binom{n}{k} \left(pe^t\right)^k (1-p)^{n-k} \\
&= \left(pe^t + 1 - p\right)^n .
\end{aligned}
$$

If $M_X(t) = \left(pe^t + 1 - p\right)^n$, then $X \sim B(n, p)$.

2. Poisson Distribution

p.m.f.: $p_k = e^{-\lambda} \cdot \frac{\lambda^k}{k!}, \quad k = 0, 1, \cdots$

$$M_X(t) = E\left[e^{tX}\right] = \sum_{k=0}^{\infty} e^{tk} e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$
$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\left(\lambda e^t\right)^k}{k!} = e^{\lambda(e^t - 1)}.$$

If $M_X(t) = e^{\lambda(e^t-1)}$, then $X \sim$ Poisson (λ).

3. Exponential Distribution

p.d.f.: $f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases}$

$$M_X(t) = E\left[e^{tx}\right] = \int_0^\infty e^{tx} f(x)dx = \int_0^\infty e^{tx}\lambda e^{-\lambda x}dx$$
$$= \lambda \int_0^\infty e^{-(\lambda - t)x}dx = \lambda \cdot \left[-\frac{e^{-(\lambda-t)x}}{\lambda - t}\right]\Bigg|_0^\infty = \frac{\lambda}{\lambda - t}, \quad \text{for } t < \lambda.$$

Note that $M_X(t)$ does not exist for $t \geq \lambda$. Hence the m.g.f. of the exponential distribution is only defined for $t < \lambda$.

If $M_X(t) = \frac{\lambda}{\lambda - t}$ for $t < \lambda$, then $X \sim$ Poisson (λ).

4. Standard Normal Distribution

p.d.f.: $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

$$
\begin{aligned}
M_X(t) = E\left[e^{tx}\right] &= \int_{-\infty}^{+\infty} e^{tx} f(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{tx} e^{-\frac{x^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2-2tx}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2-2tx+t^2}{2}} \cdot e^{\frac{t^2}{2}} dx \\
&= \frac{e^{\frac{t^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-t)^2}{2}} dx = e^{\frac{t^2}{2}}.
\end{aligned}
$$

5. General Normal Distribution

Recall: if $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$, $M_Z(t) = e^{\frac{t^2}{2}}$. So,

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

Important formula:

$$\boxed{X \sim N(\mu, \sigma^2) \quad \Leftrightarrow \quad M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}}$$

If you find the m.g.f. of some r.v. X is in the form of e^{at+bt^2}, where $a > 0$, then you may immediately claim that $X \sim N(a, 2b)$.

6. Other Examples

(1) If $X \sim \Gamma(\lambda, \alpha)$ (where $\alpha > 0$), then

$$M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha}, \quad \text{for } t < \lambda.$$

(2) If $X \sim \text{Geo}(p)$, then for $t < \ln[1/(1-p)]$,

$$M_X(t) = \frac{pe^t}{1-(1-p)e^t}.$$

(3) If $X \sim NB(p, r)$ (where r is a positive integer), then

$$M_X(t) = \left[\frac{pe^t}{1-(1-p)e^t}\right]^r.$$

(4) If $X \sim U[a, b]$, then $M_X(t) = \frac{e^{bt}-e^{at}}{(b-a)t}$.

Outline

1. Part 5.1: The Expected Value of a Random Variable
2. Part 5.2: Expectation of Functions of Random Variables
3. Part 5.3: Probability Theory Properties of Expectation, Variance and Standard Deviation
4. Part 5.4: Covariance and Correlation
5. Part 5.5: Moment Generating Functions
6. Part 5.6: Applications of Moment Generating Functions
7. Part 5.7: Three Special Distributions

Part 5.6.1: Find Moments

If we know the m.g.f. $M_X(t) = E\left[e^{tX}\right]$ of X, then all m-th moments $E\left(X^m\right)$ (m is a positive integer) can be obtained, by successively differentiating the m.g.f. with t:

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

Letting $t = 0$ yields $E(X) = M'(0)$. Similarly,

$$M_X''(t) = \frac{d}{dt}M_X'(t) = \frac{d}{dt}E\left[Xe^{tX}\right] = E\left[\frac{d}{dt}\left(Xe^{tX}\right)\right] = E\left[X^2e^{tX}\right].$$

Letting $t = 0$ yields $E(X^2) = M_X''(0)$.

Proposition 1. If the m.g.f. $M_X(t)$ is known for some r.v. X, then for all $n \in \mathbb{N}_+$

$$E(X^n) = M_X^{(n)}(0),$$

$$E(X - EX)^n = \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} (EX)^{n-k} EX^k$$

$$= \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} [M_X'(0)]^{n-k} M_X^{(k)}(0),$$

$$E(X^n) = \sum_{k=0}^{n} \binom{n}{k} E(X - EX)^k (EX)^{n-k}.$$

Example 1: Binomial Distribution

Recall $X \sim B(n,p)$ if p.m.f.

$$p(k) = P\{X = k\} = \left(\begin{array}{c} n \\ k \end{array} \right) p^k (1-p)^{n-k}, \quad k = 1, 2, \ldots n.$$

$$M_X(t) = \left(pe^t + q\right)^n,$$

$$M_X'(t) = n\left(pe^t + q\right)^{n-1} pe^t, \quad E(X) = M_X'(0) = np(p+q)^{n-1} = np,$$

$$M_X''(t) = np\left[e^t\left(pe^t + q\right)^{n-1} + e^t(n-1)\left(pe^t + q\right)^{n-2} \cdot pe^t\right],$$

$$E(X^2) = M_X''(0) = np + n(n-1)p^2, \quad \mathrm{Var}\left(X\right) = np(1-p).$$

Example 2: Poisson Distribution

Recall $X \sim \text{Poisson}(\lambda)$ if p.m.f.

$$p_k = e^{-\lambda} \cdot \frac{\lambda^k}{k!}, \quad k = 0, 1, \cdots$$

$$M_X(t) = e^{\lambda(e^t - 1)},$$

$$M_X'(t) = e^{\lambda(e^t-1)} \cdot \lambda e^t, \quad E(X) = M_X'(0) = \lambda,$$

$$M_X''(t) = \left(\lambda e^t\right)^2 e^{\lambda(e^t-1)} + \lambda e^t e^{\lambda(e^t-1)},$$

$$E(X^2) = M_X''(0) = \lambda^2 + \lambda, \quad \text{Var}\,(X) = \lambda.$$

Example 3: Exponential Distribution

Recall $X \sim \text{Exp}(\lambda)$ if p.d.f.

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

$$M_X(t) = E\left[e^{tX}\right] = \frac{\lambda}{\lambda - t}, \quad t < \lambda,$$

$$M_X'(t) = \frac{\lambda}{(\lambda - t)^2}, \quad E(X) = M_X'(0) = \frac{1}{\lambda},$$

$$M_X''(t) = \frac{2\lambda}{(\lambda - t)^3}, \quad E(X^2) = M_X''(0) = \frac{2}{\lambda^2}, \quad \text{Var}(X) = \frac{1}{\lambda^2}.$$

Example 4: General Normal Distribution

Recall: if $X \sim N(\mu, \sigma^2)$, then

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

Hence

$$M_X'(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}} \cdot \left(\mu + \sigma^2 t\right), \quad E(X) = M_X'(0) = \mu,$$

$$M_X''(t) = \left[\left(\mu + \sigma^2 t\right)^2 + \sigma^2\right] e^{\mu t + \frac{\sigma^2 t^2}{2}},$$

$$E(X^2) = \mu^2 + \sigma^2, \quad \text{Var}(X) = \sigma^2.$$

Example 5: Negative Binomial Distribution

Recall: $X \sim NB(p, r)$, then $M_X(t) = \left(\frac{pe^t}{1-qe^t}\right)^r$.

$$M_X(t) = \left(\frac{p}{e^{-t} - q}\right)^r = p^r \left(e^{-t} - q\right)^{-r},$$

$$M_X'(t) = \frac{rp^r}{e^t \left(e^{-t} - q\right)^{r+1}}, \quad E(X) = M_X'(0) = \frac{r}{p},$$

$$M_X''(t) = \frac{rp^r}{e^t \left(e^{-t} - q\right)^{r+1}} \left[\frac{r+1}{1 - qe^t} - 1\right],$$

$$E(X^2) = M_X''(0) = \frac{r(r+1-p)}{p^2}, \quad \mathrm{Var}\,(X) = \frac{r(1-p)}{p^2}.$$

In particular, if $r = 1$, then we obtain that for Geometric r.v.:

$$E(X) = \frac{1}{p}, \quad \mathrm{Var}\,(X) = \frac{1-p}{p^2}.$$

Use Maclaurin Series to Find Moments:

Recall if a function $f(x)$ can be expanded into a Maclaurin power series, then

$$f(x) = \sum_{n=0}^{\infty} c_n x^n, \quad c_n = \frac{f^{(n)}(0)}{n!}, \quad n \geq 0.$$

Therefore, if we can expand the m.g.f. of X into power series

$$M_X(t) = C_0 + C_1 t + C_2 t^2 + \cdots + C_n t^n + \cdots$$

then $C_n = \frac{M_X^{(n)}(0)}{n!}$, and thus $E(X^n) = M_X^{(n)}(0) = C_n \cdot n!$.

For example, if $X \sim N(0,1)$, then

$$M_X(t) = e^{\frac{t^2}{2}} = 1 + \frac{t^2}{2} + \frac{\left(\frac{t^2}{2}\right)^2}{2!} + \cdots + \frac{\left(\frac{t^2}{2}\right)^n}{n!} + \cdots = \sum_{n=0}^{\infty} \frac{1}{2^n \cdot n!} t^{2n}.$$

Therefore,

$$E(X^{2m+1}) = 0, \quad E(X^{2m}) = \frac{(2m)!}{2^m m!} = (2m-1)!!.$$

In particular,

$$E(X^2) = 1, \quad E(X^4) = 3, \quad E(X^6) = 15, \quad E(X^8) = 105.$$

If $X \sim N(0,t)$ with $t > 0$, then for any $p > 0$,

$$E|x|^p = C_p t^{\frac{p}{2}} < \infty, \text{ with } C_p = \frac{1}{\sqrt{2\pi}} \int_0^\infty x^p e^{-\frac{x^2}{2}} dx < \infty.$$

Higher Moments of the Exponential Distribution:

Recall if $X \sim \text{Exp}(\lambda)$, then for $\lambda > t$,

$$\begin{aligned} M_X(t) &= \frac{\lambda}{\lambda - t} = \frac{1}{1 - \frac{t}{\lambda}} = \sum_{n=0}^{\infty} \left(\frac{t}{\lambda}\right)^n \quad \left(\frac{t}{\lambda} < 1\right) \\ &= 1 + \frac{t}{\lambda} + \left(\frac{t}{\lambda}\right)^2 + \cdots + \left(\frac{t}{\lambda}\right)^n + \cdots \\ &= 1 + \frac{1}{\lambda} t + \frac{1}{\lambda^2} t^2 + \cdots + \frac{1}{\lambda^n} t^n + \cdots \end{aligned}$$

Hence

$$E(X^n) = n! \cdot \frac{1}{\lambda^n} = \frac{n!}{\lambda^n}, \quad n = 0, 1, 2, \cdots$$

Part 5.6.2: Determine the Distribution

General principle: "m.g.f." and "distribution" is "one-to-one".

For example, if you find the m.g.f. of some r.v. takes the form of

$$M_X(t) = e^{at+bt^2},$$

then you may immediately obtain that $X \sim N(a, 2b)$.

Example 1: Sum of Independent Binomial r.v.s

If $X \sim B(n, p)$, $Y \sim B(m, p)$, and X and Y are independent, then what is the p.m.f. of $Z \triangleq X + Y$?

Solution: (1) "Old method": The possible values of Z: $0, 1, 2, \cdots, n + m$. For $r = 0, 1, 2, \cdots, n + m$,

$$\begin{aligned} p_r &= P\{Z = r\} = P\{X + Y = r\} \\ &= P\left\{\bigcup_{k=0}^{r}[(X = k) \cap (Y = r - k)]\right\} \\ &= \sum_{k=0}^{r} \binom{n}{k} \binom{m}{r-k} p^r (1-p)^{m+n-r} = \cdots \end{aligned}$$

Workable, but quite complicated!

(2) New method: with the same p and thus same $q = 1 - p$,

$$X \sim B(n, p) \quad \Rightarrow \quad M_X(t) = \left(pe^t + q\right)^n,$$
$$Y \sim B(m, p) \quad \Rightarrow \quad M_Y(t) = \left(pe^t + q\right)^m.$$

Now $Z = X + Y$ and X and Y are independent,

$$M_Z(t) = M_X(t) \cdot M_Y(t) = (pe^t + q)^{n+m}.$$

Since we recognize that $(pe^t + q)^{n+m}$ is the m.g.f. of Binomial distribution with parameters $n + m$ and p, we conclude that

$$Z \sim B(n + m, p).$$

This particularly shows that:

if $Y_i : 1 \leq i \leq r$ are i.i.d. Geo(p) r.v.s, then $X \sim NB(p, r)$.

Example 2: Sum of Independent Normal r.v.s

If $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, and X and Y are independent, what is the distribution of $Z \triangleq X + Y$?

Solution: Since X and Y are independent and

$$X \sim N(\mu_1, \sigma_1^2) \quad \Rightarrow \quad M_X(t) = \exp\left[\mu_1 t + \frac{\sigma_1^2 t^2}{2}\right],$$

$$Y \sim N(\mu_2, \sigma_2^2) \quad \Rightarrow \quad M_Y(t) = \exp\left[\mu_2 t + \frac{\sigma_2^2 t^2}{2}\right].$$

So the m.g.f. of $Z = X + Y$ is

$$M_Z(t) = M_X(t) \cdot M_Y(t) = e^{(\mu_1+\mu_2)t + \frac{\sigma_1^2+\sigma_2^2}{2} t^2}.$$

Hence by checking the m.g.f. table, we immediately obtain

$$Z \sim N\left(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2\right).$$

Example 3: Sum of Independent exponential r.v.s

If $X_1, X_2, \cdots, X_n$ are i.i.d. with a common exponential distribution $\text{Exp}(\lambda)$, what is the distribution of $Z \triangleq X_1 + X_2 + \cdots + X_n$?

Solution: Each X_i obeys $\text{Exp}(\lambda)$ and they are independent, so

$$M_{X_i}(t) = \frac{\lambda}{\lambda - t} \quad (\text{for } t < \lambda) \quad \Rightarrow \quad M_Y(t) = \left(\frac{\lambda}{\lambda - t}\right)^n.$$

Checking the m.g.f. Table, we see that

$$Y \sim \Gamma(\lambda; n).$$

Let $X_1, X_2, \cdots$ be i.i.d. (independent observations of some r.v. X and thus are called a "Sample"). We are interested in the following two r.v.s:

$$S_n = X_1 + X_2 + \cdots + X_n,$$
$$\bar{X}_n = \frac{1}{n} S_n = \frac{X_1 + X_2 \cdots + X_n}{n}.$$

Usually, it is very difficult to find the c.d.f.s or p.d.f.s for them, but it is easy to find their m.g.f.s if we know the m.g.f. of X. Indeed, suppose the m.g.f. of X is $M_X(t)$, then all $X_1, \cdots, X_n$ have the same $M_X(t)$. Then by the independent property,

$$M_{S_n}(t) = [M_X(t)]^n, \quad M_{\bar{X}_n}(t) = \left[M\left(\frac{t}{n}\right)\right]^n.$$

Let's consider the case that the r.v. $X \sim \mathrm{Exp}(\lambda)$, then

$$M_{S_n}(t) = \left(\frac{\lambda}{\lambda - t}\right)^n,$$
$$M_{\bar{X}_n}(t) = \left(\frac{\lambda}{\lambda - \frac{t}{n}}\right)^n = \left(\frac{1}{1 - \frac{t}{n\lambda}}\right)^n \to e^{\frac{t}{\lambda}} \text{ as } n \to \infty.$$

Note $e^{\frac{t}{\lambda}}$ is the m.g.f. of a **constant** r.v. $\frac{1}{\lambda} = E(X)$. This property also holds for general cases.

Theorem (Weak Law of Large Number). If $X_1, X_2, \cdots$ are i.i.d with common mean $E(X) < \infty$, then

$$\lim_{n\to\infty} M_{\bar{X}_n}(t) = e^{tE(X)}.$$

Part 5.6.4: Summary of Application of m.g.f.s.

(1) For any r.v. X, the m.g.f. $M_X(t)$ is defined as $M_X(t) = E\left[e^{tX}\right]$ for those t s.t. the expectation exists

(2) M.g.f.s and distributions are uniquely determined with each other - if we find the m.g.f. of X and recognize it as the m.g.f. of a known distribution, then X must have that distribution

(3) If X and Y are independent r.v.s then the m.g.f. of $Z = X + Y$ is the product of their m.g.f.: $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$. In particular, $M_{aX+b}(t) = e^{bt} M_X(at)$ for $a, b \in \mathbb{R}$

(4) If $M_X(t)$ exists in a neighbourhood of 0 for a r.v. X, then the moments of X exist and can be found from the power series expansion of $M_X(t)$ with $E\left(X^k\right) = M^{(k)}(0)$:

$$M_X(t) = 1 + t \cdot E(X) + \frac{t^2}{2} \cdot E\left(X^2\right) + \cdots + \frac{t^n}{n!} E\left(X^n\right) + \cdots$$

Outline

1. Part 5.1: The Expected Value of a Random Variable
2. Part 5.2: Expectation of Functions of Random Variables
3. Part 5.3: Probability Theory Properties of Expectation, Variance and Standard Deviation
4. Part 5.4: Covariance and Correlation
5. Part 5.5: Moment Generating Functions
6. Part 5.6: Applications of Moment Generating Functions
7. Part 5.7: Three Special Distributions

Recall that if $X \sim \Gamma(\lambda, \alpha)$, then

$$M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha} = \left(1 - \frac{1}{\lambda}t\right)^{-\alpha}, \quad t < \lambda.$$

Let $X \sim N(0, 1)$. Then

$$M_{X^2}(t) = Ee^{tX^2} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx^2} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(1-2t)x^2}{2}} dx$$
$$= \frac{1}{\sqrt{1-2t}} = (1-2t)^{-\frac{1}{2}}, \quad t < \frac{1}{2}.$$

Thus, $X^2 \sim \Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$.

Proposition 1. Let $X_1, \cdots, X_n$ be i.i.d. with common $N(0,1)$-distribution. Then

$$Y := X_1^2 + X_2^2 + \cdots + X_n^2 \sim \Gamma\left(\frac{1}{2}, \frac{n}{2}\right).$$

The distribution of the above Y is called $\chi^2(n)$-distribution, i.e., $\chi^2(n) = \Gamma\left(\frac{1}{2}, \frac{n}{2}\right)$, then the p.d.f. $f(x)$ of the χ^2-Distribution is

$$f(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Proposition 2. Let $X \sim \chi^2(n)$. Then

$$EX^k = n(n+2)(n+4)\cdots(n+2k-2)$$

for any positive integer k. In particular,

$$E(Y) = n, \quad E(Y^2) = n(n+2), \quad \mathrm{Var}\,(Y) = 2n.$$

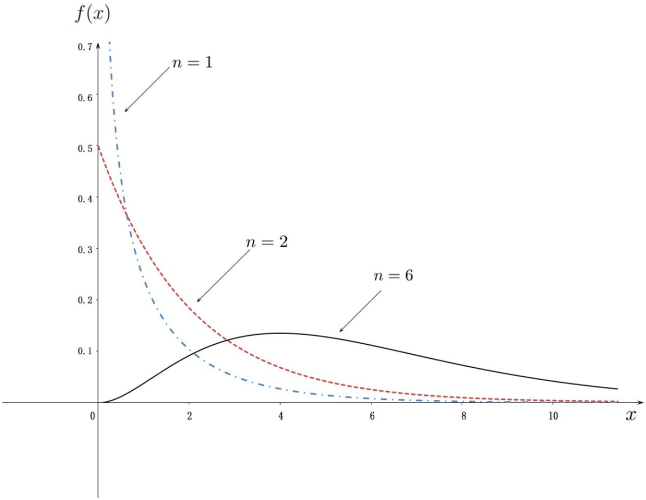
Proof. Recall the general formula (with $m \in \mathbb{R}$)

$$(1+x)^m = \sum_{k=0}^{m} \binom{k}{m} x^k = 1+mx+\cdots+\frac{m(m-1)(m-2)\cdots(m-k+1)}{k!}x^k+\cdots$$

Taking $x = -2t$ and $m = -\frac{n}{2}$, we obtain

$$M_X(t) = \sum_{k=0}^{\infty} \frac{n(n+2)(n+4)\cdots(n+2k-2)}{k!} t^k.$$

The p.d.f. of $\chi^2(n)$–Distribution



Part 5.7.2: t–Distribution

Definition 1. If $X \sim N(0,1)$, $Y \sim \chi^2(n)$, X and Y are independent, then the distribution of $T = \frac{X}{\sqrt{Y/n}}$ is called t-distribution with n degree of freedom, denoted by $T \sim t(n)$.

Proposition 2. Let $T \sim t(n)$. Then T has p.d.f. f given by

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}}\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x \in \mathbb{R}.$$

Recall the Γ-function is given by

$$\Gamma(z) \triangleq \int_0^{\infty} x^{z-1} e^{-x} dx, \quad z > 0.$$

It has the following properties:

$$\Gamma(z+1) = z\Gamma(z).$$

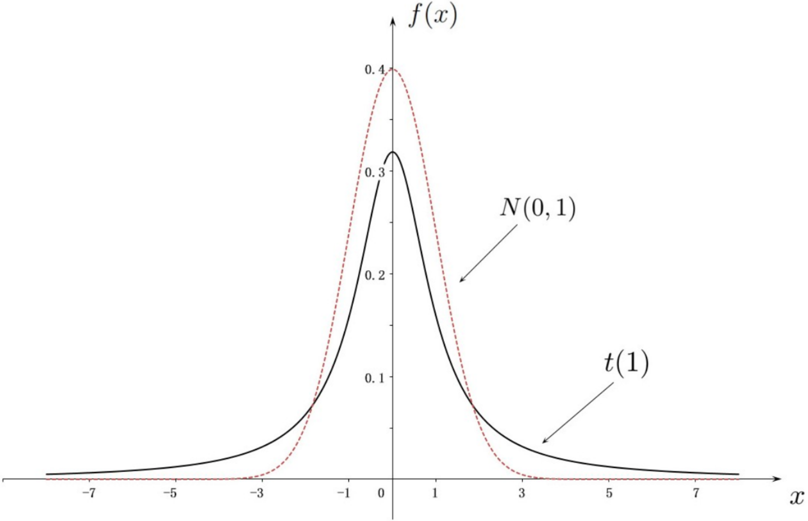
$$\Gamma\left(\frac{1}{2}+n\right) = \frac{(2n)!}{4^n n!}\sqrt{\pi} = \frac{(2n-1)!!}{2^n}\sqrt{\pi} = \begin{pmatrix} n-\frac{1}{2} \\ n \end{pmatrix} n!\sqrt{\pi}.$$

Try to prove

$$\lim_{n\to\infty} t(n) \to N(0,1)$$

using the above properties.

The p.d.f. of t–Distribution



Definition 1. If $X \sim \chi^2(m)$, $Y \sim \chi^2(n)$, X and Y are independent, then the distribution of $F = \frac{X/m}{Y/n}$ is called F-distribution with (m, n) degrees of freedom, denoted by $F \sim F(m, n)$.

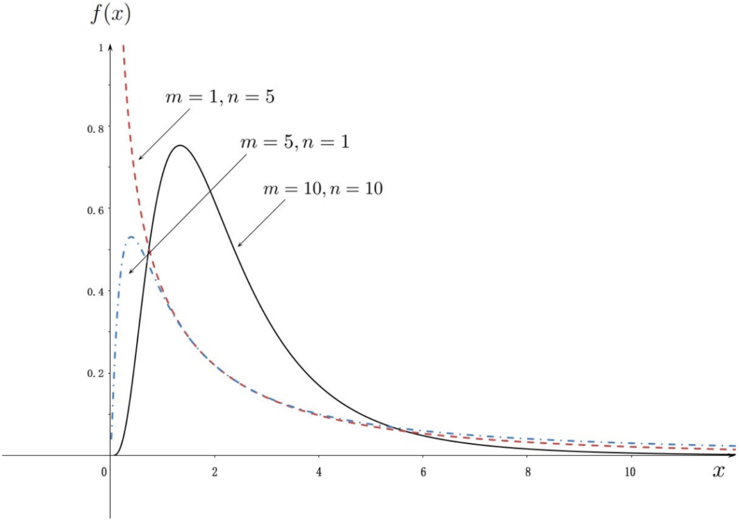
Proposition 2.

(i) Let $T \sim F(m, n)$. Then T has p.d.f. f given by

$$f(x) = \begin{cases} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \cdot \Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right) \left(\frac{m}{n}x\right)^{\frac{m}{2}-1} \left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}}, & \text{if } x > 0, \\ 0, & \text{if } x \leq 0. \end{cases}$$

(ii) If $X \sim F(m, n)$, then $X^{-1} \sim F(n, m)$; if $T \sim t(n)$, then $T^2 \sim F(1, n)$.

The p.d.f. of F-Distribution



Outline

1. Part 5.1: The Expected Value of a Random Variable
2. Part 5.2: Expectation of Functions of Random Variables
3. Part 5.3: Probability Theory Properties of Expectation, Variance and Standard Deviation
4. Part 5.4: Covariance and Correlation
5. Part 5.5: Moment Generating Functions
6. Part 5.6: Applications of Moment Generating Functions
7. Part 5.7: Three Special Distributions

Part 5.8.1: Mixing Case

Suppose X is a discrete r.v. but Y is a continuous r.v., then the joint c.d.f. and margin c.d.f.s of X and Y are defined as

$$F_{(X,Y)}(x,y) = P\{X \leq x, Y \leq y\},$$
$$F_X(x) = P\{X \leq x\},$$
$$F_Y(y) = P\{Y \leq y\}.$$

It is clear that

$$F_X(x) = \lim_{y \to +\infty} F_{(X,Y)}(x,y),$$
$$F_Y(y) = \lim_{x \to +\infty} F_{(X,Y)}(x,y).$$

We say that X and Y are independent if the joint c.d.f. is the product of two marginal c.d.f.s:

$$F_{(X,Y)}(x,y) = F_X(x) \cdot F_Y(y).$$

We can also define the so-called "joint p.d.f.-p.m.f." function

$$f_{(X,Y)}(x,y) \quad \left[\begin{array}{l} x : \{x_1, x_2, \cdots\} \\ y : (-\infty, +\infty) \end{array} \right]$$

such that for any real numbers a, b, we have

$$F_{(X,Y)}(a,b) = \sum_{x \le a} \int_{-\infty}^{b} f(x,y)dy$$

It can be easily seen that

- $f(x,y) \ge 0 \quad (\forall x = x_i; \forall y \in \mathbb{R});$
- $\sum_x \int_{-\infty}^{+\infty} f(x,y)dy = \int_{-\infty}^{+\infty} \left(\sum_x f(x,y)\right) dy = 1;$
- $p_X(x) = \int_{-\infty}^{+\infty} f(x,y)dy, \quad f_Y(y) = \sum_x f(x,y);$
- X and Y are independent iff
  joint p.d.f.-p.m.f. =(marginal p.d.f.) $\times$ (marginal p.m.f.).

Summary of Chapter 5: Basic Concepts

(1) Expected value: $E(X)$ (Weighted average).

(2) Variance: $\mathrm{Var}\,(X) = E\left[(X - E(X))^2\right]$.

(3) Standard Deviation: $\sqrt{\mathrm{Var}\,(X)}$.

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

(5) Correlation: $\rho(X, Y) = \frac{\mathrm{Cov}\,(X,Y)}{\sqrt{\mathrm{Var}\,(X) \cdot \mathrm{Var}\,(Y)}}$.

(6) Function of r.v.s: $g(X), g(X, Y)$ etc.

Summary of Chapter 5: Basic properties

1. Expectation:
   (1) $E(C) = C \quad$ for constant C.
   (2) $E(a \cdot X + b \cdot Y) = a \cdot E(X) + b \cdot E(Y)$ (linear property).
   (3) If X and Y are independent, then $E(X \cdot Y) = E(X) \cdot E(Y)$.
2. Variance:
   (1) $\text{Var}\,(C) = 0 \quad$ for constant C.
   (2) $\text{Var}\,(a \cdot X + b) = a^2 \cdot \text{Var}\,(X)$.
   (3) If X and Y are independent, then
   $\text{Var}\,(X + Y) = \text{Var}\,(X) + \text{Var}\,(Y)$.

3. Covariance:
   (1) $\mathrm{Cov}\,(a, b) = 0$ for constants a and b.
   (2) $\mathrm{Cov}\,(a_1 X + b_1 Y + c_1, a_2 U + b_2 V + c_2) = a_1 \cdot a_2 \mathrm{Cov}\,(X, U) + a_1 b_2 \mathrm{Cov}\,(X, V) + a_2 \cdot b_1 \mathrm{Cov}\,(Y, U) + b_1 b_2 \mathrm{Cov}\,(Y, V)$.
   (3) If X and Y are independent, then $\mathrm{Cov}\,(X, Y) = 0$ (the converse is not true; true for bivariate normal r.v.)
   (4) Correlation: for any two r.v.s, $-1 \leq \rho(X, Y) \leq 1$.

Summary of Chapter 5: Calculation

1. Function of r.v.s

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x) \cdot f(x) dx.$$

$$E[g(X)] = \sum_{i} g\left(x_i\right) \cdot p\left(x_i\right).$$

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

$$E[g(X, Y)] = \sum_{i} \sum_{i} g\left(x_i, y_j\right) \cdot f\left(x_i, y_j\right).$$

2. Variance: $\operatorname{Var}(X) = E\left(X^2\right) - (EX)^2$.
3. Covariance: $\operatorname{Cov}(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$.

Summary of Chapter 5: Important Facts

1. Normal distribution:
   (1) If $X \sim N\left(\mu, \sigma^2\right)$, then $E(X)=\mu, \operatorname{Var}(X)=\sigma^2$.
   (2) If $X \sim N\left(\mu, \sigma^2\right)$, a, b are constants, $a \neq 0$, then $a \cdot X+b \sim N\left(a\mu+b, a^2\sigma^2\right)$.
   (3) If $X \sim N\left(\mu_1, \sigma_1^2\right)$, $Y \sim N\left(\mu_2, \sigma_2^2\right)$, and X and Y are independent, then $X+Y \sim N\left(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2\right)$.
   (4) If $X \sim N\left(\mu, \sigma^2\right)$, let $Y=\frac{X-\mu}{\sigma}$, then $Y \sim N(0,1)$.
2. Poisson distribution:
   If $X \sim \operatorname{Poisson}(\lambda)$, then $E(X)=\lambda, \operatorname{Var}(X)=\lambda$.
3. Binomial distribution:
   If $X \sim B(n, p)$, then $E(X)=np, \operatorname{Var}(X)=npq=np(1-p)$.
4. Exponential distribution:
   If $X \sim \operatorname{Exp}(\lambda)$, then $E(X)=\frac{1}{\lambda}, \operatorname{Var}(X)=\frac{1}{\lambda^2}$.

Summary of Chapter 5: Examples

Example 1. Suppose the m.g.f of X is $M_X(t) = \frac{2}{4-t}, t < 4$.

(1) Find $E(X), E\left(X^2\right)$ and $\operatorname{Var}(X)$.

(2) If X and Y are independent and both with this m.g.f. Then find the m.g.f and also identify the distribution of $X + Y$.

Example 2. Let $X_1, \cdots, X_n$ be i.i.d., each having the $N\left(\mu, \sigma^2\right)$ distribution.

(1) Find the m.g.f. of the sample mean $\bar{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i$

(2) What is the distribution of the above sample mean?