Chapter 4*

September 11, 2017

1 Expected value

The distribution of a random variable X contains all of the probabilistic information about X. The entire distribution of X, however, is usually too cumbersome for presenting this information.

Thus we need some *summary* measures, like the center of the distribution, spread of the distribution.

Intuition: The *expected value* of a random variable indicates its (weighted) average.

Example: How many heads would you expect if you flipped a coin twice?

Let X = number of heads. Then $X = \{0, 1, 2\}$. Here $p(0) = \frac{1}{4}, p(1) = \frac{1}{2}, p(2) = \frac{1}{4}$. [Draw p.m.f!]

Weighted average $= 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1$.

Definition: Let X be a *bounded* random variable assuming the values $x_1, x_2, x_3, \ldots$ with corresponding probabilities $p(x_1), p(x_2), p(x_3), \ldots$. The *mean* or *expected value* of X is defined by

$$\mathbb{E}(X) = \sum_k x_k \cdot p(x_k).$$

Interpretations:

(i) The expected value measures the center of the probability distribution – center of mass.

*Notes for Chapter 4 of DeGroot and Schervish adapted from Giovanni Motta's, Bodhisattva Sen's and Martin Lindquists notes for STAT W4109/W4105.

Table 1: Distribution of r.v X

X=1	$\mathbb{P}(X=1)=\mathbb{P}(\{1\})=1/6$
X=3	$\mathbb{P}(X=3)=\mathbb{P}(\{3\})=1/6$
X=5	$\mathbb{P}(X=5)=\mathbb{P}(\{5\})=1/6$
X=-4	$\mathbb{P}(X=-4)=\mathbb{P}(\{2,4,6\})=3/6$

(ii) Long term frequency (law of large numbers we'll get to this soon).

Expectations can be used to describe the potential gains and losses from games.

Example: Roll a die. If the side that comes up is odd, you win the USD equivalent of that side. If it is even, you lose USD 4.

Solution: Let X = your earnings. Thus,

$$\mathbb{E}(X) = 1\cdot\frac{1}{6} + 3\cdot\frac{1}{6} + 5\cdot 16 + (-4)\cdot\frac{1}{2} = \frac{1}{6}+\frac{3}{6}+\frac{5}{6}-2 = -\frac{1}{2}.$$

Example: (Lottery) You pick 3 different numbers between 1 and 12. If you pick all the numbers correctly you win USD 100. What are your expected earnings if it costs USD 1 to play?

Solution: Let X = your earnings. Then X takes values in $\{99,-1\}$. Here

$$\mathbb{P}(X=99) = \frac{1}{\binom{12}{3}} = \frac{1}{220}, \qquad \& \qquad \mathbb{P}(X=-1) = 1-\frac{1}{220} = \frac{219}{220}.$$

Thus,

$$\mathbb{E}(X) = 99\cdot\frac{1}{220} + (-1)\cdot\frac{219}{220} = -\frac{120}{220} = -0.55.$$

Definition: Let X be a discrete random variable whose p.m.f is p. Suppose that at least one of the following sums is finite:

$$\sum_{x:x\geq 0} x\cdot p(x), \qquad \sum_{x:x<0} x\cdot p(x). \tag{1}$$

Then the *mean, expectation, expected value* of X is said to exist and is defined to be

$$\mathbb{E}(X) = \sum_x x\cdot p(x). \tag{2}$$

If both the sums in (1) are infinite, then $\mathbb{E}(X)$ does not exist. The sum in (2) is not well-defined then.

Example: Let X be a random variable whose p.m.f is

$$p(x) = \begin{cases} \frac{1}{2|x|(|x|+1)}, & \text{if } x = \pm 1, \pm 2, \dots, \\ 0 & \text{otherwise.} \end{cases}$$

Then p is indeed a p.m.f (verify at home). The two sums in (1) are

$$\sum_{x=-1}^{-\infty} x \cdot p(x) = -\infty, \qquad \sum_{x=1}^{\infty} x \cdot p(x) = \infty;$$

and hence $\mathbb{E}(X)$ does not exist.

Note that the expectation of a random variable X depends **only** on the distribution of X.

1.1 Expectation of a continuous distribution

Idea: Use integrals instead of sums!

Definition: Let X be a bounded random variable whose p.d.f is f. The *expectation* of X, denoted by $\mathbb{E}(X)$, is defined as follows:

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

Example: An appliance has a maximum lifetime of one year. The time X until it fails is a random variable with a continuous distribution having p.d.f

$$f(x) = \begin{cases} 2x & \text{for } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then,

$$\mathbb{E}(X) = \int_0^1 x(2x) dx = \int_0^1 2x^2 dx = \frac{2}{3}.$$

Definition: Let X be a continuous random variable whose p.d.f is f. Suppose that at least one of the following integrals is finite:

$$\int_0^{\infty} x \cdot f(x) dx, \qquad \int_{-\infty}^{0} x \cdot f(x) dx. \tag{3}$$

Then the *mean, expectation, expected value* of X is said to exist and is defined to be

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

If both the sums in (3) are infinite, then $\mathbb{E}(X)$ does not exist.

1.2 Expectation of a function of a random variable

Let X be a r.v assuming values $x_1, x_2, \ldots$ with corresponding probabilities $p(x_1), p(x_2), \ldots$.

For any function g, what is $\mathbb{E}[g(X)]$?

The mean or expected value of $g(X)$ can be found by applying the definition of expectation to the distribution of $g(X)$, i.e., let $Y = g(X)$, determine the probability distribution of Y, and then determine $\mathbb{E}(Y)$ by applying the definition (e.g., (2)).

Example: Roll a fair die. Let $X = \#$ of dots on the side that comes up. Let $Y = I_{\{X \text{ is odd}\}}$. Calculate $\mathbb{E}(Y)$.

Solution: Y takes two values 0,1 with probability 1/2 each. Thus, $\mathbb{E}(Y) = 1/2$.

Theorem 1.1. *We have*

$$\mathbb{E}[g(X)] = \sum_{k=1}^{\infty} g(x_k) \cdot p(x_k),$$

if the mean exists.

Proof. Let $Y = g(X)$ and let p_Y be the p.m.f of Y. Then,

$$\begin{aligned}
\sum_y y\, p_Y(y) &= \sum_y y\, \mathbb{P}(g(X) = y) = \sum_y y \sum_{x:g(x)=y} \mathbb{P}(X = x) \\
&= \sum_y \sum_{x:g(x)=y} g(x)p(x) = \sum_x g(x)p(x).
\end{aligned}$$

Thus, $\mathbb{E}(Y) = \sum_y y\, p_Y(y) = \sum_x g(x)p(x)$. $\square$

Similarly, if X has a continuous distribution with p.d.f f, then

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

if the mean exists.

Example: Roll a fair die. Let $X = \#$ of dots on the side that comes up. Calculate $\mathbb{E}(X^2)$.

Solution: $\mathbb{E}(X^2) = \sum_{i=1}^{6} i^2 p(i) = 1^2 p(1) + 2^2 p(2) + 3^2 p(3) + 4^2 p(4) + 5^2 p(5) + 6^2 p(6) = \frac{1}{6} \cdot (1 + 4 + 9 + 16 + 25 + 36) = \frac{91}{6}$.

Calculate $\mathbb{E}(\sqrt{X}) = \sum_{i=1}^{6} \sqrt{i} p(i)$. Calculate $\mathbb{E}(e^X) = \sum_{i=1}^{6} e^i p(i)$. (Do at home)

Note that in general, $\mathbb{E}[g(X)] \neq g(\mathbb{E}(X))$.

$\mathbb{E}(X)$ is the expected value or 1-st *moment* of X. Then, $\mathbb{E}(X^n)$ is called the *n-th moment* of X.

If $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ is a random vector such that $\mathbb{E}(X_i)$ exists for all $i = 1, 2, \ldots, d$, then $\mathbb{E}(\mathbf{X}) = (\mathbb{E}(X_1), \mathbb{E}(X_2), \ldots, \mathbb{E}(X_d))$.

Example: An indicator variable for the event A is defined as the random variable that takes on the value 1 when event A happens and 0 otherwise, i.e.,

$$I_A = \begin{cases} 1, & \text{if } A \text{ occurs} \\ 0, & \text{if } A^c \text{ occurs.} \end{cases}$$

Thus, $\mathbb{P}(I_A = 1) = \mathbb{P}(A)$ and $\mathbb{P}(I_A = 0) = \mathbb{P}(A^c)$.

The expectation of this indicator (noted I_A) is

$$\mathbb{E}(I_A) = 1 \cdot \mathbb{P}(A) + 0 \cdot \mathbb{P}(A^c) = \mathbb{P}(A).$$

One-to-one correspondence between expectations and probabilities.

1.3 Properties of expectation

Proposition 1.2. *If a and b are constants, then $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$.*

Proof. $\mathbb{E}(aX + b) = \sum_k [(ax_k + b) \cdot p(x_k)] = a \sum_k x_k p(x_k) + b \sum_k p(x_k) = a\mathbb{E}(X) + b$. □

Proposition 1.3. *If there exists a constant such that $\mathbb{P}(X \geq a) = 1$ then $\mathbb{E}(X) \geq a$.*

Theorem 1.4. *Let X be a random variable that can take only the values $0, 1, 2, \ldots$. Then $\mathbb{E}(X) = \sum_{i=1}^{\infty} \mathbb{P}(X \geq i)$.*

Proof. Note that

$$\begin{aligned} \mathbb{E}(X) &= \sum_{n=0}^{\infty} n\mathbb{P}(X = n) = \sum_{n=1}^{\infty} n\mathbb{P}(X = n) \\ &= \sum_{n=1}^{\infty} \sum_{i=1}^{n} \mathbb{P}(X = n) = \sum_{i=1}^{\infty} \sum_{n=i}^{\infty} \mathbb{P}(X = n) \\ &= \sum_{i=1}^{\infty} \mathbb{P}(X \geq i). \end{aligned}$$

$\square$

Theorem 1.5. *Let X be a nonnegative random variable with c.d.f F. Then*

$$\mathbb{E}(X) = \int_0^\infty [1 - F(x)] \; dx.$$

Proof. If X is continuous, the proof follows as

$$\mathbb{E}(X) = \int_0^\infty x f(x) dx = \int_0^\infty \left(\int_0^x 1 \; dy \right) f(x) dx = \int_0^\infty \left(\int_y^\infty f(x) dx \right) dy = \int_0^\infty [1 - F(y)] \; dy.$$

$\square$

Theorem 1.6. *If X and Y have a joint p.m.f $p_{XY}(\cdot, \cdot)$, then*

$$\mathbb{E}[g(X, Y)] = \sum_x \sum_y g(x, y) p_{XY}(x, y).$$

If X and Y have a joint p.d.f $f_{XY}(\cdot, \cdot)$, then

$$\mathbb{E}[g(X, Y)] = \int_{-\infty}^\infty \int_{-\infty}^\infty g(x, y) f_{XY}(x, y) \; dx dy.$$

It is important to note that if the function $g(x, y)$ is only dependent on either x or y the formula above reverts to the 1-dimensional case.

Example: Suppose that X and Y have a joint p.d.f $f_{XY}(\cdot, \cdot)$. Calculate $\mathbb{E}(X)$.

Solution:

$$\mathbb{E}(X) = \int_{-\infty}^\infty \int_{-\infty}^\infty x f_{XY}(x, y) \; dy dx = \int_{-\infty}^\infty x \left(\int_{-\infty}^\infty f_{XY}(x, y) \; dy \right) dx = \int_{-\infty}^\infty x f_X(x) dx.$$

Example: An accident occurs at a point X that is uniformly distributed on a road of length L. At the time of the accident an ambulance is at location Y that is also uniformly distributed on the road. Assuming that X and Y are independent, find the expected distance between the ambulance and the point of the accident.

Solution: We have to compute $\mathbb{E}(|X - Y|)$ when both X and Y are uniform on the interval $(0, L)$. The joint p.d.f is $f_{XY}(x, y) = \frac{1}{L^2} I\{0 < x < L, 0 < y < L\}$. Then,

$$\mathbb{E}(|X - Y|) = \int_0^L \int_0^L |x - y| \frac{1}{L^2} dy dx = \frac{1}{L^2} \int_0^L \int_0^L |x - y| dy dx.$$

Note that

$$\int_0^L |x - y| dy = \int_0^x (x - y) dy + \int_x^L (y - x) dy = \frac{L^2}{2} + x^2 - xL.$$

Thus,

$$\mathbb{E}(|X-Y|) = \frac{1}{L^2}\int_0^L \left(\frac{L^2}{2} + x^2 - xL\right) dx = \frac{L}{3}.$$

Theorem 1.7. *If $X_1, \ldots, X_n$ are n r.v's such that each $\mathbb{E}(X_i)$ is finite, then*

$$\mathbb{E}(X_1 + X_2 + \ldots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \ldots + \mathbb{E}(X_n).$$

Proof. We shall first assume that $n = 2$ and also, for convenience, that X_1 and X_2 have a continuous joint distribution for which the joint p.d.f is f. Then,

$$\begin{aligned}
\mathbb{E}(X_1 + X_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x_1 + x_2) f(x_1, x_2) dx_1 dx_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 f(x_1, x_2) dx_1 dx_2 + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_2 f(x_1, x_2) dx_1 dx_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 f(x_1, x_2) dx_2 dx_1 + \int_{-\infty}^{\infty} x_2 \int_{-\infty}^{\infty} f(x_1, x_2) dx_1 dx_2 \\
&= \int_{-\infty}^{\infty} x_1 f_1(x_1) dx_1 + \int_{-\infty}^{\infty} x_2 f_2(x_2) dx_2 \\
&= \mathbb{E}(X_1) + \mathbb{E}(X_2),
\end{aligned}$$

where f_1 and f_2 are the marginal p.d.f's of X_1 and X_2 respectively. The proof of the theorem can be extended for each n by an induction argument. □

Example: Let X be a *Binomial* random variable with parameters n and p. X represents the number of successes in n trials. We can write X as

$$X = X_1 + X_2 + \ldots + X_n,$$

where $X_i = \begin{cases} 1 & \text{if trial } i \text{ is a success} \\ 0 & \text{if trial } i \text{ is a failure.} \end{cases}$

The X_i's are *Bernoulli* random variables with parameter p. Thus,

$$\begin{aligned}
\mathbb{E}(X_i) &= 1 \cdot p + 0 \cdot (1-p) = p \\
\mathbb{E}(X) &= \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n) = np.
\end{aligned}$$

Example: A group of N people throw their hats into the center of a room. The hats are mixed, and each person randomly selects one. Find the expected number of people that select their own hat.

Solution: Let $X =$ the number of people who select their own hat.

Number the people from 1 to N. Let

$$X_i = \begin{cases} 1 & \text{if person } i \text{ chooses his own hat,} \\ 0 & \text{otherwise.} \end{cases}$$

Then $X = X_1 + \ldots + X_N$.

Each person is equally likely to select any of the N hats, so $\mathbb{P}(X_i = 1) = \frac{1}{N}$.

Thus, $\mathbb{E}(X_i) = 1 \cdot \frac{1}{N} + 0 \cdot \left(1 - \frac{1}{N}\right) = \frac{1}{N}$.

Hence, $\mathbb{E}(X) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_N) = N\frac{1}{N} = 1$.

Example: Twenty people, consisting of 10 married couples, are to be seated at five different tables, with four people at each table. If the seating is done at random, what is the expected number of married couples that are seated at the same table?

Solution: Let $X =$ the number of married couples at the same table.

Number the couples from 1 to 10 and let,

$$X_i = \begin{cases} 1 & \text{if couple } i \text{ is seated at the same table,} \\ 0 & \text{otherwise.} \end{cases}$$

Then $X = X_1 + \ldots + X_{10}$.

To calculate $\mathbb{E}(X)$ we need to know $\mathbb{E}(X_i)$.

Consider the table where husband i is sitting. There is room for three other people at his table. There are a total of 19 possible people which could be seated at his table. Thus,

$$\mathbb{P}(X_i = 1) = \frac{\binom{18}{2}}{\binom{19}{3}} = \frac{3}{19},$$

and $\mathbb{E}(X_i) = 1 \cdot \frac{3}{19} + 0 \cdot \frac{16}{19} = \frac{3}{19}$.

Hence, $\mathbb{E}(X) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_{10}) = \frac{30}{19}$.

Proposition 1.8. *If X and Y are independent, then for any functions h and g,*

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

Proof. Note that

$$\begin{aligned} \mathbb{E}[g(X)h(Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_{XY}(x,y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_{XY}(x,y)dxdy \\ &= \int_{-\infty}^{\infty} g(x)f_X(x)dx \int_{-\infty}^{\infty} h(y)f_Y(y)dy = \mathbb{E}[g(X)] \cdot \mathbb{E}[h(Y)]. \end{aligned}$$

$\square$

In fact, this is an equivalent way to characterize *independence*: if for any functions g and h, if $\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)]$, then X and Y are independent. To see this, just use indicator functions.

1.4 Variance

We often seek to summarize the essential properties of a random variable in as simple terms as possible.

The mean is one such property. It gives a measure of the centre of the distribution.

Let $X = 0$ with probability 1.

Let $Y = \begin{cases} -2, & \text{with prob. } \frac{1}{3} \\ -1, & \text{with prob. } \frac{1}{6} \\ 1, & \text{with prob. } \frac{1}{6} \\ 2, & \text{with prob. } \frac{1}{3}. \end{cases}$

Both X and Y have the same expected value, but are quite different in other respects. One such respect is in their spread. We would like a measure of *spread*.

Definition: If X is a random variable with mean $\mathbb{E}(X)$, then the *variance* of X, denoted by $\mathrm{Var}(X)$, is defined by

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

A small variance indicates a small spread.

Example: $\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$. This follows as

$$\begin{aligned} \mathrm{Var}(X) &= \mathbb{E}[(X - \mathbb{E}(X))^2] = \sum (x - \mathbb{E}(X))^2 p(x) \\ &= \sum [x^2 - 2x\mathbb{E}(X) + \mathbb{E}(X)^2] p(x) \\ &= \sum x^2 \cdot p(x) - 2\mathbb{E}(X) \sum x \cdot p(x) + \mathbb{E}(X)^2 \sum p(x) \\ &= \mathbb{E}(X^2) - 2[\mathbb{E}(X)]^2 + [\mathbb{E}(X)]^2 = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2. \end{aligned}$$

Example: Roll a fair die. Let $X = \#$ that comes up. What is $\mathrm{Var}(X)$?

Solution: Recall that $\mathbb{E}(X^2) = 91/6$, $\mathbb{E}(X) = (1+2+3+4+5+6)/6 = 21/6 = 7/2$. Thus,

$$\mathrm{Var}(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{91}{6} - \frac{7}{2} = \frac{182 - 147}{12} = \frac{35}{12}.$$

Proposition 1.9. *If a and b are constants then* $\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$.

Proof. Note that, $\mathbb{E}(aX+b) = a\mathbb{E}(X)+b$. Then,

$$\begin{aligned}\mathrm{Var}(aX+b) &= \mathbb{E}[(aX+b-(a\mathbb{E}(X)+b))^2] \\ &= \mathbb{E}[a^2(X-\mathbb{E}(X))^2] = a^2\mathbb{E}[(X-\mathbb{E}(X))^2] = a^2\mathrm{Var}(X).\end{aligned}$$

□

The square root of $\mathrm{Var}(X)$ is called the *standard deviation* of X, i.e.,

$$SD(X) = \sqrt{\mathrm{Var}(X)},$$

and it measures the scale of X.

Proposition 1.10. $\mathrm{Var}(X) = 0$ *if and only if there exists a constant c such that* $\mathbb{P}(X = c) = 1$.

Proof. Suppose first that there exists a constant c such that $\mathbb{P}(X = c) = 1$. Then $\mathbb{E}(X) = c$, and

$$\mathrm{Var}(X) = \mathbb{E}[(X-c)^2] = 0.$$

Conversely, suppose that $\mathrm{Var}(X) = 0$. Let $\mu = \mathbb{E}(X)$. Then, for a discrete r.v,

$$\mathbb{E}[(X-\mu)^2] = \sum_x p(x)(x-\mu)^2 = 0,$$

which implies that $(x-\mu)^2 = 0$ for all x such that $p(x) > 0$. Thus, $\mathbb{P}(X = \mu) = 1$. □

1.5 Mean and median

Mean is the best estimate under squared loss, i.e., the number θ that minimizes

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

is $\mathbb{E}(X)$. [Proof: expand and differentiate with respect to θ]

Median is the best estimate under absolute loss, i.e., median minimizes

$$\mathbb{E}[|X-\theta|].$$

Proof is given in the book. Note that median is non-unique in general.

1.6 Moment generating functions

The *moment generating function* of the random variable X, denoted $M_X(t)$, is defined for all real values of t by,

$$M_X(t) = \mathbb{E}(e^{tX}) = \begin{cases} \sum_x e^{tx} \cdot p(x) & \text{if } X \text{ is discrete with p.m.f } p(\cdot) \\ \int_{-\infty}^{\infty} e^{tx} \cdot f(x) & \text{if } X \text{ is continuous with p.d.f } f(\cdot) \end{cases}$$

The reason $M_X(\cdot)$ is called a moment generating function is because all the moments of X can be obtained by successively differentiating $M_X(t)$ and evaluating the result at $t = 0$.

First moment:

$$\begin{aligned} M_X'(t) &= \frac{d}{dt} M_X(t) = \frac{d}{dt}\mathbb{E}(e^{tX}) = \mathbb{E}\left(\frac{d}{dt}e^{tX}\right) = \mathbb{E}(Xe^{tX}) \\ M_X'(0) &= \mathbb{E}(X). \end{aligned}$$

(We will assume that we can move the derivative inside the expectation).

Second moment:

$$\begin{aligned} M_X''(t) &= \frac{d}{dt} M_X'(t) = \frac{d}{dt}\mathbb{E}(Xe^{tX}) = \mathbb{E}(\frac{d}{dt}Xe^{tX}) = \mathbb{E}(X^2e^{tX}) \\ M_X''(0) &= \mathbb{E}(X^2). \end{aligned}$$

k-th moment:

$$\begin{aligned} M_X^{(k)}(t) &= \mathbb{E}(X^k e^{tX}) \\ M_X^{(k)}(0) &= \mathbb{E}(X^k). \end{aligned}$$

Example: (Binomial random variable with parameters n and p) Suppose that X is a discrete random variable with p.m.f

$$p(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{for } k = 0, 1, 2, \ldots, n \\ 0 & \text{otherwise.} \end{cases}$$

Calculate $M_X(t)$.

Solution: Note that

$$M_X(t) = \mathbb{E}(e^{tX}) = \sum_{k=0}^{n} e^{kt} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{n} \binom{n}{k} (e^t p)^k (1-p)^{n-k} = [pe^t + (1-p)]^n.$$

Therefore,

$$\begin{aligned} M_X'(t) &= n[pe^t + (1-p)]^{n-1} pe^t \\ M_X''(t) &= n(n-1)[pe^t + (1-p)]^{n-2} (pe^t)^2 + M_X'(t). \end{aligned}$$

Thus,

$$\begin{aligned}
\mathbb{E}(X) = M'_X(0) &= n[pe^0 + (1-p)]^{n-1} pe^0 = np \\
\mathbb{E}(X^2) = M''_X(0) &= n(n-1)[pe^0 + (1-p)]^{n-2}(pe^0)^2 + M'_X(0) = n(n-1)p^2 + np \\
\mathrm{Var}(X) &= \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = n(n-1)p^2 + np - n^2p^2 = np(1-p).
\end{aligned}$$

Theorem 1.11. *Suppose that for two random variables X and Y, their moment generating functions are given by $M_X(\cdot)$ and $M_Y(\cdot)$, respectively, and are finite in an open interval around 0. If $M_X(t) = M_Y(t)$ for all values of t in the open interval (around 0), then X and Y have the same probability distribution.*

Proposition 1.12. *The moment generating function of the sum of independent random variables equals the product of the individual moment generating functions.*

Proof.

$$M_{X+Y}(t) = \mathbb{E}(e^{t(X+Y)}) = \mathbb{E}(e^{tX}e^{tY}) = \mathbb{E}(e^{tX})\mathbb{E}(e^{tY}) = M_X(t)M_Y(t).$$

□

1.7 Covariance and correlation

Previously, we have discussed the absence or presence of a relationship between two random variables, i.e., independence or dependence. But if there is in fact a relationship, the relationship may be either weak or strong.

Example: (a) Let X = weight of a sample of water, and Y = volume of the same sample of water.

There is an extremely strong relationship between X and Y.

(b) Let X = a person's weight, and Y = same person's height.

There is a relationship between X and Y, but not as strong as in (a).

We would like a measure that can quantify this difference in the strength of a relationship between two random variables.

Definition: The covariance between X and Y, denoted by $\mathrm{Cov}(X, Y)$, is defined by

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

As with the variance, we can rewrite this equation as

$$\begin{aligned}
\mathrm{Cov}(X, Y) &= \mathbb{E}\left[XY - \mathbb{E}(X)Y - X\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y)\right] \\
&= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y) \\
&= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).
\end{aligned}$$

Note that if X and Y are independent,

$$\operatorname{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) = 0.$$

The converse is however **not** true.

Counter-example: Define X and Y so that,

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

and

$$Y = \begin{cases} 0 & \text{if } X \neq 0 \\ 1 & \text{if } X = 0. \end{cases}$$

Then, X and Y are clearly dependent.

Note that $XY = 0$ so we have that $\mathbb{E}(XY) = \mathbb{E}(X) = 0$, so

$$\operatorname{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\cdot\mathbb{E}(Y) = 0.$$

Proposition 1.13. *The following hold:*

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

(ii) $\operatorname{Cov}(X,X) = \operatorname{Var}(X)$

(iii) $\operatorname{Cov}(aX,Y) = a\cdot\operatorname{Cov}(X,Y)$

(iv) $\operatorname{Cov}(\sum_{i=1}^m X_i, \sum_{j=1}^n Y_j) = \sum_{i=1}^m \sum_{j=1}^n \operatorname{Cov}(X_i, Y_j)$.

Proof. (i)–(iii) Verify yourselves.

(iv) Let $\mu_i = \mathbb{E}(X_i)$, for $i = 1,\dots,m$, and $\nu_i = \mathbb{E}(Y_j)$, for $j = 1,\dots,n$. Then

$$\mathbb{E}\left(\sum_{i=1}^m X_i\right) = \sum_{i=1}^m \mu_i \qquad \text{and} \qquad \mathbb{E}\left(\sum_{j=1}^n Y_j\right) = \sum_{j=1}^n \nu_j.$$

Now,

$$\begin{aligned} \operatorname{Cov}\left(\sum_{i=1}^m X_i, \sum_{j=1}^n Y_j\right) &= \mathbb{E}\left[\left(\sum_{i=1}^m X_i - \sum_{i=1}^m \mu_i\right)\left(\sum_{j=1}^n Y_j - \sum_{j=1}^n \nu_j\right)\right] \\ &= \mathbb{E}\left[\sum_{i=1}^m \sum_{j=1}^n (X_i - \mu_i)(Y_j - \nu_j)\right] \\ &= \sum_{i=1}^m \sum_{j=1}^n \mathbb{E}\left[(X_i-\mu_i)(Y_j-\nu_j)\right] = \sum_{i=1}^m \sum_{j=1}^n \operatorname{Cov}(X_i, Y_j). \end{aligned}$$

$\square$

Proposition 1.14. $\text{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \text{Var}(X_i) + 2\sum_{i<j}\sum \text{Cov}(X_i, X_j)$. *In particular,*

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

Proof. Note that $\text{Var}(\sum_{i=1}^{n} X_i) = \text{Cov}(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i)$. ☐

If $X_1, \ldots, X_n$ are pairwise independent or uncorrelated for $i \neq j$, then

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

Example: Let $X_1, \ldots, X_n$ be i.i.d r.v's having expected value μ and variance σ^2. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the **sample mean**. The random variable

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

is called the **sample variance**. Calculate (a) $\text{Var}(\bar{X})$ and (b) $\mathbb{E}(S^2)$.

Solution: (a) We know that $\text{Var}(X_i) = \sigma^2$. Then,

$$\text{Var}(\bar{X}) = \frac{1}{n^2}\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}(X_i) = \frac{\sigma^2}{n}.$$

(b) Hint: Use

$$\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n}\left[(X_i - \mu)^2 + (\bar{X} - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu)\right].$$

Example: A group of N people throw their hats into the center of a room. The hats are mixed, and each person randomly selects one.

Let $X = \#$ of people who select their own hat.

Number the people from 1 to N. Let

$$X_i = \begin{cases} 1 & \text{if person } i \text{ chooses his own hat,} \\ 0 & \text{otherwise.} \end{cases}$$

Then $X = X_1 + \ldots X_N$. We showed last time that $\mathbb{E}(X) = 1$. Calculate $\text{Var}(X)$.

Solution: Recall that since each person is equally likely to select any of the N hats, i.e, $\mathbb{P}(X_i = 1) = \frac{1}{N}$. Then, $\mathbb{E}(X_i) = \mathbb{E}(X_i^2) = \frac{1}{N}$ and

$$\text{Var}(X_i) = \mathbb{E}(X_i^2) - [\mathbb{E}(X_i)]^2 = \frac{1}{N} - \frac{1}{N^2} = \frac{N-1}{N^2}.$$

Now let us compute $\mathrm{Cov}(X_i, X_j)$. Observe that

$$X_i X_j = \begin{cases} 1 & \text{if persons } i \text{ and } j \text{ choose their own hats,} \\ 0 & \text{otherwise.} \end{cases}$$

So, $\mathbb{P}(X_i = 1, X_j = 1) = \mathbb{P}(X_j = 1 | X_i = 1) \cdot \mathbb{P}(X_i = 1) = \frac{1}{N(N-1)}$. Therefore, $\mathbb{E}(X_i X_j) = \frac{1}{N(N-1)}$, and thus,

$$\mathrm{Cov}(X_i, X_j) = \frac{1}{N(N-1)} - \frac{1}{N^2} = \frac{1}{N^2(N-1)}.$$

Hence,

$$\mathrm{Var}(X) = N \cdot \frac{N-1}{N^2} + 2\binom{N}{2} \frac{1}{N^2(N-1)} = \frac{N-1}{N} + \frac{1}{N} = 1.$$

1.8 Correlation coefficient

Definition: The *correlation* between X and Y, denoted by $\rho(X, Y)$, is defined (as long as $\mathrm{Var}(X)$ and $\mathrm{Var}(Y)$ are positive) by

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

It can be shown that

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

with equality if and only if $Y = aX + b$ (assuming $\mathbb{E}(X^2)$ and $\mathbb{E}(Y^2)$ are both finite). This is called the "Cauchy-Schwarz" inequality.

Proof. It suffices to prove that

$$[\mathbb{E}(XY)]^2 \leq \mathbb{E}(X^2) \cdot \mathbb{E}(Y^2).$$

The basic idea is to look at the expectations $\mathbb{E}[(aX + bY)^2]$ and $\mathbb{E}[(aX - bY)^2]$. We use the usual rules for adding and subtracting variance:

$$\begin{aligned} 0 &\leq \mathbb{E}[(aX + bY)^2] = a^2\mathbb{E}(X^2) + b^2\mathbb{E}(Y^2) + 2ab \cdot \mathbb{E}(XY) \\ 0 &\leq \mathbb{E}[(aX - bY)^2] = a^2\mathbb{E}(X^2) + b^2\mathbb{E}(Y^2) - 2ab \cdot \mathbb{E}(XY). \end{aligned}$$

Now let $a^2 = \mathbb{E}(Y^2)$ and $b^2 = \mathbb{E}(X^2)$. Then the above two inequalities read

$$\begin{aligned} 0 &\leq 2a^2b^2 + 2ab\mathbb{E}(XY) \\ 0 &\leq 2a^2b^2 - 2ab\mathbb{E}(XY); \end{aligned}$$

dividing by $2ab$ gives

$$E(XY) \geq -\sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}$$
$$E(XY) \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)},$$

and this is equivalent to the inequality $-1 \leq \rho(X,Y) \leq 1$. For equality to hold, either $\mathbb{E}[(aX+bY)^2] = 0$ or $\mathbb{E}[(aX-bY)^2] = 0$, i.e., X and Y are linearly related with a negative or positive slope, respectively. $\square$

The correlation coefficient is therefore a measure of the degree of *linear association* between X and Y. If $\rho(X,Y) = 0$ then this indicates no linearity, and X and Y are said to be *uncorrelated.*

1.9 Conditional expectation

Recall that if X and Y are discrete random variables, the *conditional mass function* of X, given $Y = y$, is defined (for all y such that $\mathbb{P}(Y = y) > 0$) by

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

Definition: If X and Y are discrete random variables, the *conditional expectation* of X, given $Y = y$, is defined (for all y such that $\mathbb{P}(Y = y) > 0$) by

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

The conditional expectation of X given $Y = y$, is just the expected value on a reduced sample space consisting only of outcomes where $Y = y$.

Similarly, if X and Y are continuous random variables, the *conditional p.d.f* of X given $Y = y$, is defined (for all y such that $f_Y(y) > 0$) by

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

Definition: If X and Y are continuous random variables, the conditional expectation of X, given $Y = y$, is defined (for all y such that $f_Y(y) > 0$) by

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

Conditional means: $\mathbb{E}(X|Y = y)$ is a function of y. Let $h(y) = \mathbb{E}(X|Y = y)$. Define the symbol $\mathbb{E}(Y|X)$ to mean $h(X)$ and call it the *conditional mean* of Y given X.

Conditional expectations are themselves random variables.

It is important to note that conditional expectations satisfy all the properties of regular expectations:

1. $\mathbb{E}[g(X)|Y=y] = \begin{cases} \sum_x g(x)\cdot p_{X|Y}(x|y) & \text{if } X \text{ and } Y \text{ are discrete} \\ \int_{-\infty}^{\infty} g(x)\cdot f_{X|Y}(x|y)dx & \text{if } X \text{ and } Y \text{ are continuous.} \end{cases}$
2. $\mathbb{E}(\sum_{i=1}^n X_i|Y=y) = \sum_{i=1}^n \mathbb{E}(X_i|Y=y)$.

Proposition 1.15. $\mathbb{E}(X) = \mathbb{E}[\mathbb{E}(X|Y)]$.

Proof. If Y is discrete then,

$$\begin{aligned}
\mathbb{E}[\mathbb{E}(X|Y)] &= \sum_y \mathbb{E}(X|Y=y)p_Y(y) = \sum_y \left(\sum_x x\cdot p_{X|Y}(x|y)\right) p_Y(y) \\
&= \sum_y \left(\sum_x x\cdot \frac{p_{XY}(x,y)}{p_Y(y)}\right) p_Y(y) \\
&= \sum_x x \sum_y p_{XY}(x,y) = \sum_x x p_X(x) = \mathbb{E}(X).
\end{aligned}$$

□

When manipulating the conditional distribution given $X=x$, it is safe to assume at X is the constant x.

Proposition 1.16. *Let X and Y be random variables, and let $Z = r(X,Y)$ for some function r. The conditional distribution of Z given $X=x$ is the same as the conditional distribution of $r(x,Y)$ given $X=x$.*

Theorem 1.17. *(Law of total variance) If X and Y are arbitrary random variables for which the necessary expectations and variances exist, then*

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

Proof. Note that

$$\begin{aligned}
\mathrm{Var}(Y) &= \mathbb{E}(Y^2) - [\mathbb{E}(Y)]^2 \\
&= \mathbb{E}(\mathbb{E}(Y^2|X)) - [\mathbb{E}(\mathbb{E}(Y|X))]^2 \\
&= \mathbb{E}[\mathrm{Var}(Y|X) + (\mathbb{E}(Y|X))^2] - [\mathbb{E}(\mathbb{E}(Y|X))]^2 \\
&= \mathbb{E}[\mathrm{Var}(Y|X)] + \mathbb{E}[\mathbb{E}(Y|X))^2] - [\mathbb{E}(\mathbb{E}(Y|X))]^2 \\
&= \mathbb{E}[\mathrm{Var}(Y|X)] + \mathrm{Var}[\mathbb{E}(Y|X)].
\end{aligned}$$

$\square$

Let N be a r.v with values in the natural numbers and finite mean $\mathbb{E}(N) < \infty$. Let $X_1, X_2, \cdots$ be a sequence of i.i.d random variables that are independent of N and have a finite mean $\mathbb{E}(X)$. Let $Y := X_1 + X_2 + \ldots + X_N$. What is the expected value of Y? What is the variance of Y if we assume that $\mathrm{Var}(N) < \infty$ and $\mathrm{Var}(X_1) < \infty$?

For example, think of N as the number of stores visited, and X_i is the money spent in store i.

We will use $\mathbb{E}\left[\sum_{i=1}^N X_i\right] = \mathbb{E}\left[\mathbb{E}\left(\sum_{i=1}^N X_i | N\right)\right]$. Notice that

$$\begin{aligned}
\mathbb{E}\left(\sum_{i=1}^N X_i | N = n\right) &= \mathbb{E}\left(\sum_{i=1}^n X_i | N = n\right) \\
&= \mathbb{E}\left(\sum_{i=1}^n X_i\right) \qquad \text{since } X_i \text{ and } N \text{ are independent} \\
&= n\mathbb{E}(X_1),
\end{aligned}$$

which implies that $\mathbb{E}\left[\mathbb{E}\left(\sum_{i=1}^N X_i | N\right)\right] = N\mathbb{E}(X_1)$. Consequently,

$$\mathbb{E}\left[\sum_{i=1}^N X_i\right] = \mathbb{E}[N\mathbb{E}(X_1)] = \mathbb{E}(N)\mathbb{E}(X_1).$$

To find $\mathrm{Var}(Y)$ we will use the law of total variance. As $\mathbb{E}(Y|N) = N\mathbb{E}(X_1)$, we have

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

We can show that $\mathrm{Var}(Y|N = n) = n\mathrm{Var}(X_1)$, and thus,

$$\mathbb{E}[\mathrm{Var}(Y|N)] = \mathbb{E}[N\mathrm{Var}(X_1)] = \mathrm{Var}(X_1)\mathbb{E}(N).$$

Therefore,

$$\begin{aligned}
\mathrm{Var}(Y) &= \mathbb{E}[\mathrm{Var}(Y|N)] + \mathrm{Var}[\mathbb{E}(Y|N)] \\
&= \mathbb{E}(N)\mathrm{Var}(X) + [\mathbb{E}(X_1)]^2 \mathrm{Var}(N).
\end{aligned}$$

---

1. Consider a random point (X, Y) in the triangle $\{(x, y) : x, y \geq 0, x + y \leq 1\}$. Find the marginal density of Y and use this to find $\mathbb{E}(Y)$. Use the *towering property* of conditional expectation to find $\mathbb{E}(Y)$.

Solution: Show that

$$f_Y(y) = 2(1-y)I(0 < y < 1),$$

and thus,

$$\mathbb{E}(Y) = \int_0^1 y f_Y(y) dy = \frac{1}{3}.$$

Given that $X = x$, Y is distributed uniformly on $[0, 1-x]$ and so

$$\mathbb{E}(Y|X = x) = \frac{1}{2}(1-x).$$

By definition, $\mathbb{E}(Y|X) = \frac{1}{2}(1-X)$, and the expectation of $\frac{1}{2}(1-X)$ must, therefore, equal the expectation of Y, and, indeed

$$\mathbb{E}(X) = \mathbb{E}(Y) = \frac{1}{3}.$$

2. The joint density of (X, Y) is given by

$$f(x,y) = \begin{cases} 3x & \text{if } 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Compute $\mathrm{Cov}(X, Y)$.

Solution: $\mathrm{Cov}(X,Y) = \frac{3}{10} - \frac{3}{4} \cdot \frac{3}{8} = \frac{3}{160}$.

3. Recall that a full deck of cards contains 52 cards, 13 cards of each of the four suits. Distribute the cards at random to 13 players, so that each gets 4 cards. Let N be the number of players whose four cards are of the same suit. Using the indicator trick, compute $E(N)$.

Solution: Let

$$I_i = I_{\{\text{player } i \text{ has four cards of the same suit}\}},$$

so that

$$N = I_1 + \ldots + I_{13}.$$

Observe that:

- the number of ways to select 4 cards from a 52 card deck is $\binom{52}{4}$;
- the number of choices of a suit is 4; and
- after choosing a suit, the number of ways to select 4 cards of that suit is $\binom{13}{4}$.

Therefore, for all i,

$$\mathbb{E}(I_i) = \frac{4 \cdot \binom{13}{4}}{\binom{52}{4}},$$

and thus,

$$\mathbb{E}(N) = 13 \cdot \frac{4 \cdot \binom{13}{4}}{\binom{52}{4}}.$$