

6 Topics: Expectation, independence of random variables, probability generating functions

6.1 Prerequisites: Lecture 14

Exercise 6- 1: (Suggested for personal/peer tutorial) Indicator variables and their expectation: Recall that we defined the indicator of an event $A \in \mathcal{F}$ in Definition 7.4.2 as follows: For an event $A \in \mathcal{F}$, we denote by

$$\mathbb{I}_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \notin A, \end{cases}$$

the indicator variable of the event A .

- (a) For events $A, B \in \mathcal{F}$, show that
- $(\mathbb{I}_A)^k = \mathbb{I}_A$ for any $k \in \mathbb{N}$,
 - $\mathbb{I}_{A^c} = 1 - \mathbb{I}_A$,
 - $\mathbb{I}_{A \cap B} = \mathbb{I}_A \mathbb{I}_B$,
 - $\mathbb{I}_{A \cup B} = \mathbb{I}_A + \mathbb{I}_B - \mathbb{I}_A \mathbb{I}_B$.
- (b) Prove the fundamental bridge between probability and expectation, i.e. show that there is a one-to-one correspondence between events and indicator random variables and for any $A \in \mathcal{F}$ we have

$$P(A) = E(\mathbb{I}_A).$$

Solution:

- (a) i. This follows from the fact that $0^k = 0$ and $1^k = 1$ for all $k \in \mathbb{N}$.
- ii. We note that

$$\mathbb{I}_{A^c}(\omega) = \begin{cases} 1, & \text{if } \omega \in A^c, \\ 0, & \text{if } \omega \notin A^c, \end{cases}$$

and

$$1 - \mathbb{I}_A(\omega) = 1 - \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \notin A, \end{cases} = \begin{cases} 0, & \text{if } \omega \in A, \\ 1, & \text{if } \omega \notin A, \end{cases} = \begin{cases} 1, & \text{if } \omega \notin A^c, \\ 0, & \text{if } \omega \in A^c, \end{cases},$$

which implies (ii).

- iii. The identity holds since

$$\mathbb{I}_{A \cap B} = \begin{cases} 1, & \text{if } \omega \in A \cap B, \\ 0, & \text{if } \omega \notin A \cap B, \end{cases} = \begin{cases} 1, & \text{if } \omega \in A \text{ and } \omega \in B, \\ 0, & \text{if } \omega \in A^c \text{ or } \omega \in B^c, \end{cases},$$

and

$$\mathbb{I}_A \mathbb{I}_B = \begin{cases} 1, & \text{if } \omega \in A \text{ and } \omega \in B, \\ 0, & \text{if } \omega \in A^c \text{ or } \omega \in B^c. \end{cases}$$

- iv.

$$\begin{aligned} \mathbb{I}_{A \cup B} &= 1 - \mathbb{I}_{(A \cup B)^c} = 1 - \mathbb{I}_{A^c \cap B^c} = 1 - \mathbb{I}_{A^c} \mathbb{I}_{B^c} \\ &= 1 - (1 - \mathbb{I}_A)(1 - \mathbb{I}_B) = \mathbb{I}_A + \mathbb{I}_B - \mathbb{I}_A \mathbb{I}_B. \end{aligned}$$

- (b) For any event $A \in \mathcal{F}$ we can define the indicator random variable \mathbb{I}_A uniquely, and vice versa given an indicator random variable \mathbb{I}_A we can get the event A back by noting that $A = \{\omega \in \Omega : \mathbb{I}_A(\omega) = 1\}$.

Also, we have that $\mathbb{I}_A \sim \text{Bern}(p)$ with $p = P(A)$, since

$$P(\mathbb{I}_A = 1) = P(A), \quad P(\mathbb{I}_A = 0) = P(A^c) = 1 - P(A), \quad P(\mathbb{I}_A = x) = 0 \text{ for } x \notin \{0, 1\}.$$

Hence

$$E(\mathbb{I}_A) = 0 \cdot P(\mathbb{I}_A = 0) + 1 \cdot P(\mathbb{I}_A = 1) = P(\mathbb{I}_A = 1) = P(A).$$

Exercise 6- 2: Prove Theorem 10.2.6: Consider a discrete/continuous random variable X with finite expectation.

- (a) If X is non-negative, then $E(X) \geq 0$.
 (b) If $a, b \in \mathbb{R}$, then $E(aX + b) = aE(X) + b$.

Solution: First, consider the case when X is discrete:

- (a) Since X is assumed to be non-negative, we have that $\text{Im}X \subseteq [0, \infty)$. Also recall that probabilities are by definition non-negative. Hence the expectation

$$E(X) = \sum_{x \in \text{Im}X} xP(X = x)$$

is given by a sum of non-negative terms. Hence it must be non-negative itself.

- (b) We apply Theorem 10.2.1:

$$E[aX + b] = \sum_x (ax + b)p_X(x) = a \sum_x xp_X(x) + b \sum_x p_X(x) = aE(X) + b.$$

Next, consider the case when X is continuous:

- (a) Since X is assumed to be non-negative, we have that $\text{Im}X \subseteq [0, \infty)$. Also recall that the density is by definition non-negative. Hence the expectation

$$E(X) = \int_0^\infty xf_X(x)dx$$

is given as an integral of non-negative terms. Hence it must be non-negative itself.

- (b) We apply Theorem 10.2.4:

$$E[aX + b] = \int_{-\infty}^\infty (ax + b)f_X(x)dx = a \int_{-\infty}^\infty xf_X(x)dx + b \int_{-\infty}^\infty f_X(x)dx = aE(X) + b.$$

Exercise 6- 3: Prove Theorem 10.3.3: Let X be a discrete random variable with finite variance and consider deterministic constants $a, b \in \mathbb{R}$. Then

$$\text{Var}(aX + b) = a^2\text{Var}(X).$$

Solution: From Theorem 10.2.6 we note that

$$E(aX + b) = aE(X) + b.$$

We apply Theorem 10.2.6 once more in the following derivation:

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

6.2 Prerequisites: Lecture 15

Exercise 6- 4: Consider a sequence of *Bernoulli* random variables X_1, \dots, X_n each with parameter θ resulting from independent binary trials, so that

$$P(X = 0) = 1 - \theta, \quad P(X = 1) = \theta.$$

Find the probability mass functions of the random variables

- (a) $Y = \text{Min} \{X_1, \dots, X_n\}$
- (b) $Z = \text{Max} \{X_1, \dots, X_n\}$

[Hint: find the ranges of Y and Z , and consider $P(Y = 1)$, $P(Z = 0)$.]

Solution:

- (a) $Y = \text{Min} \{X_1, \dots, X_n\}$, so $\text{Im}Y = \{0, 1\}$.

$$\begin{aligned} P(Y = 1) &= P(\text{Min} \{X_1, \dots, X_n\} = 1) = P(X_1 = 1, X_2 = 1, \dots, X_n = 1) = \theta^n, \\ P(Y = 0) &= 1 - \theta^n. \end{aligned}$$

Hence

$$p_Y(y) = \begin{cases} 1 - \theta^n, & y = 0 \\ \theta^n, & y = 1 \end{cases}$$

- (b) $Z = \text{Max} \{X_1, \dots, X_n\}$, so $\text{Im}Z = \{0, 1\}$.

$$\begin{aligned} P(Z = 0) &= P(\text{Max} \{X_1, \dots, X_n\} = 0) = P(X_1 = 0, X_2 = 0, \dots, X_n = 0) = (1 - \theta)^n, \\ P(Z = 1) &= 1 - (1 - \theta)^n. \end{aligned}$$

$$p_Z(z) = \begin{cases} (1 - \theta)^n, & z = 0 \\ 1 - (1 - \theta)^n, & z = 1 \end{cases}$$

6.3 Prerequisites: Lecture 16

Exercise 6- 5: Suppose that $F_{X,Y}(x, y)$ is the joint distribution function of (X, Y) . Find an expression for $P(X \leq x, Y > y)$ in terms of $F_{X,Y}$.

Solution: Since $P(Y > y) + P(Y \leq y) = 1$, we can apply the law of total probability and find that

$$P(X \leq x, Y > y) + P(X \leq x, Y \leq y) = P(X \leq x) = F_{X,Y}(x, \infty).$$

Hence $P(X \leq x, Y > y) = F_{X,Y}(x, \infty) - F_{X,Y}(x, y)$.

Exercise 6- 6: Consider a probability space given by $\Omega = \{-1, 0, 1\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ and P is defined by

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

We define two discrete random variables by $X(\omega) = \omega$ and $Y(\omega) = |\omega|$.

- (a) Compute $P(X = 0, Y = 1)$, $P(X = 0)$ and $P(Y = 1)$. What can you conclude?
 (b) Compute $E(XY)$ and $E(X)E(Y)$. What can you conclude?

Solution:

- (a) We have

$$P(X = 0, Y = 1) = P(\{\omega : X(\omega) = \omega = 0, Y(\omega) = |\omega| = 1\}) = P(\emptyset) = 0,$$

and

$$P(X = 0) = P(\{\omega : X(\omega) = \omega = 0\}) = P(\{0\}) = 1/3,$$

$$P(Y = 1) = P(\{\omega : Y(\omega) = |\omega| = 1\}) = P(\{-1, 1\}) = P(\{-1\}) + P(\{1\}) = \frac{2}{3}.$$

So, we have that

$$P(X = 0, Y = 1) = 0 \neq \frac{2}{3} = P(X = 0)P(Y = 1),$$

hence X and Y are dependent.

- (b) For the joint probability mass function, we have

$$P(X = 0, Y = 1) = P(\{\omega : X(\omega) = \omega = 0, Y(\omega) = |\omega| = 1\}) = P(\emptyset) = 0,$$

$$P(X = 1, Y = 1) = P(\{\omega : X(\omega) = \omega = 1, Y(\omega) = |\omega| = 1\}) = P(\{1\}) = \frac{1}{3},$$

$$P(X = -1, Y = 1) = P(\{\omega : X(\omega) = \omega = -1, Y(\omega) = |\omega| = 1\}) = P(\{-1\}) = \frac{1}{3},$$

$$P(X = 0, Y = 0) = P(\{\omega : X(\omega) = \omega = 0, Y(\omega) = |\omega| = 0\}) = P(\{0\}) = \frac{1}{3},$$

$$P(X = 1, Y = 0) = P(\{\omega : X(\omega) = \omega = 1, Y(\omega) = |\omega| = 0\}) = P(\emptyset) = 0,$$

$$P(X = -1, Y = 0) = P(\{\omega : X(\omega) = \omega = -1, Y(\omega) = |\omega| = 0\}) = P(\emptyset) = 0.$$

Hence

$$E(XY) = \sum_x \sum_y xyP(X = x, Y = y) = 1 \cdot \frac{1}{3} + (-1) \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} = 0.$$

Also,

$$E(X) = (-1 + 0 + 1) \frac{1}{3} = 0, \quad E(Y) = 1 \cdot \frac{2}{3} + 0 \cdot \frac{1}{3} = \frac{2}{3}.$$

Altogether, we have

$$E(XY) = 0 = E(X)E(Y),$$

although X and Y are dependent! So we conclude that from Theorem 12.7.5 we know that if X and Y are independent, then the product formula (12.7.1) for the expectations holds. If, however, the product formula holds, that does not in general imply that the random variables are independent.

Exercise 6- 7: Show that discrete random variables X and Y on (Ω, \mathcal{F}, P) are independent if and only if

$$E(g(X)h(Y)) = E(g(X))E(h(Y)), \quad (6.1)$$

for all functions $g, h : \mathbb{R} \rightarrow \mathbb{R}$ for which the expectations on the right hand side exist.

Hint: For the part, where you assume (6.1) for all g, h and want to show that X and Y are independent, proceed as follows: Write down the definition of independence of discrete random variables using the factorisation of the corresponding joint pmf. Now, choose specific functions g and h (as suitably defined indicator functions), which allow you to derive the factorisation formula for the pmf from (6.1).

Solution: Let us assume that X and Y are independent. We use Theorem 12.6.1 again. Then

$$\begin{aligned} E(g(X)h(Y)) &= \sum_x \sum_y g(x)h(y)P(X=x, Y=y) \\ &= \sum_x \sum_y g(x)h(y)P(X=x)P(Y=y) \quad (\text{by independence}) \\ &= \sum_x g(x)P(X=x) \sum_y h(y)P(Y=y) \quad (\text{using the existence of } E(g(X)), E(h(Y))) \\ &= E(g(X))E(h(Y)). \end{aligned}$$

Next we assume that (6.1) holds for all functions $g, h : \mathbb{R} \rightarrow \mathbb{R}$ for which the corresponding expectations on the right hand side exist. We want to show that for any $x, y \in \mathbb{R}$, we have

$$P(X=x, Y=y) = P(X=x)P(Y=y).$$

Consider any $x, y \in \mathbb{R}$. Let us now specify the functions g and h in a suitable way. We define

$$g(z) = \mathbb{I}_{\{x\}}(z) = \begin{cases} 1, & \text{if } z = x, \\ 0, & \text{if } z \neq x, \end{cases}, \quad h(z) = \mathbb{I}_{\{y\}}(z) = \begin{cases} 1, & \text{if } z = y, \\ 0, & \text{if } z \neq y, \end{cases}$$

Then

$$\begin{aligned} E(g(X)h(Y)) &= 0 \cdot P(g(X)h(Y)=0) + 1 \cdot P(g(X)h(Y)=1) \\ &= P(g(X)=1, h(Y)=1) = P(X=x, Y=y). \end{aligned}$$

Also,

$$E(g(X)) = P(g(X)=1) = P(X=x), \quad E(h(Y)) = P(h(Y)=1) = P(Y=y).$$

By our assumption, we have

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

hence

$$P(X = x, Y = y) = P(X = x)P(Y = y). \quad (6.2)$$

Since the identity 6.2 holds for any $x, y \in \mathbb{R}$ we have shown the independence of X and Y .

Exercise 6- 8: Convolution theorem: Consider jointly discrete/continuous random variables X and Y and define their sum as $Z = X + Y$. Find the probability mass function/density function of Z (leaving your expression as a sum/integral). If you assume that X and Y are independent, can you simplify the p.m.f./p.d.f. of Z ?

Hint: Recall that $P((X, Y) \in A) = \sum \sum_{(x,y) \in A} P(X = x, Y = y)$ in the discrete case, and $P((X, Y) \in A) = \int \int_A f_{X,Y}(x, y) dx dy$ in the jointly continuous case for "nice" sets $A \subseteq \mathbb{R}^2$.

Solution:

- (a) Discrete case: We use the law of total probability: If $Z = z$ and we know that $X = x$, then $Y = z - x$. Similarly, if $X = x$ and $Y = z - x$, then this implies that $Z = z$. So we can write for any $z \in \mathbb{R}$

$$\begin{aligned} P(Z = z) &\stackrel{\text{law of total prob.}}{=} \sum_x P(Z = z, X = x) = \sum_x P(Z = z, X = x, Y = z - x) \\ &= \sum_x P(X = x, Y = z - x). \end{aligned}$$

If we assume independence of X and Y , then

$$P(Z = z) = \sum_x P(X = x, Y = z - x) = \sum_x P(X = x)P(Y = z - x).$$

You can imagine that the above sum can be rather tedious to compute in practice!

- (b) Continuous case: We first derive the c.d.f. of Z : For any $z \in \mathbb{R}$, we have

$$\begin{aligned} F_Z(z) &= P(Z \leq z) = \int \int_{\{(x,y) \in \mathbb{R}^2: x+y \leq z\}} f_{X,Y}(x, y) dx dy \\ &= \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{z-y} f_{X,Y}(x, y) dx dy \\ &= \int_{v=-\infty}^{\infty} \int_{u=-\infty}^z f_{X,Y}(u - v, v) du dv \\ &= \int_{u=-\infty}^z \int_{v=-\infty}^{\infty} f_{X,Y}(u - v, v) dv du. \end{aligned}$$

Differentiating with respect to z gives us the density function:

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(z - v, v) dv \left(= \int_{-\infty}^{\infty} f_{X,Y}(u, z - u) du \right).$$

If X and Y are independent, then their joint density factorises and we have

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z - v) f_Y(v) dv \left(= \int_{-\infty}^{\infty} f_X(u) f_Y(z - u) du \right).$$

Exercise 6- 9: Prove Theorem 12.7.2: For jointly discrete/continuous random variables X, Y with finite expectations, we have

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

Solution: We use the linearity of the expectation (Theorem 12.6.3):

$$\begin{aligned}\text{Cov}(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] = E[XY - X\mu_Y - \mu_X Y + \mu_X \mu_Y] \\ &= E(XY) - E(X)\mu_Y - \mu_X E(Y) + \mu_X \mu_Y \\ &= E(XY) - \mu_X \mu_Y - \mu_X \mu_Y + \mu_X \mu_Y = E(XY) - \mu_X \mu_Y.\end{aligned}$$

Exercise 6- 10: Prove Theorem 12.7.7: Let X, Y denote two jointly discrete/continuous random variables with finite variances. Then

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

Solution: Using the definition of the variance and the linearity of the expectation, we have

$$\begin{aligned}\text{Var}(X + Y) &= E[(X + Y - E(X + Y))^2] \\ &= E[(X + Y)^2 - 2(X + Y)E(X + Y) + (E(X) + E(Y))^2] \\ &= E(X^2) + 2E(XY) + E(Y^2) - 2\{[E(X)]^2 + 2E(X)E(Y) + [E(Y)]^2\} \\ &\quad + [E(X)]^2 + 2E(X)E(Y) + [E(Y)]^2 \\ &= E(X^2) + 2E(XY) + E(Y^2) - [E(X)]^2 - 2E(X)E(Y) - [E(Y)]^2 \\ &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y).\end{aligned}$$

6.4 Prerequisites: Lecture 17

Exercise 6- 11: Suppose that X_1, \dots, X_n are independent $\text{Ber}(p)$ random variables. Define $S_n = \sum_{i=1}^n X_i$. Use probability generating functions to show that $S_n \sim \text{Bin}(n, p)$.

Solution: We have

$$G_{S_n}(s) = E(s^{S_n}) = E(s^{X_1 + \dots + X_n}) = E(s^{X_1} \dots s^{X_n}) = E(s^{X_1}) \dots E(s^{X_n}) = (sp + 1 - p)^n,$$

which is the pgf of a $\text{Bin}(n, p)$ random variable. Theorem 13.1.2 allows us to conclude.

Exercise 6- 12: Suppose that X_1, \dots, X_n are independent Poisson random variables – not necessarily with the same parameter, i.e. $X_i \sim \text{Poi}(\lambda_i)$. Define $S_n = \sum_{i=1}^n X_i$. Use probability generating functions to show that $S_n \sim \text{Poi}(\sum_{i=1}^n \lambda_i)$.

Solution: As above, we have

$$\begin{aligned} G_{S_n}(s) &= E(s^{S_n}) = E(s^{X_1 + \dots + X_n}) = E(s^{X_1} \dots s^{X_n}) = E(s^{X_1}) \dots E(s^{X_n}) \\ &= \prod_{i=1}^n \exp(\lambda_i(s-1)) = \exp\left(\sum_{i=1}^n \lambda_i(s-1)\right), \end{aligned}$$

which is the p.g.f. of a $\text{Poi}(\sum_{i=1}^n \lambda_i)$ random variable. Again, Theorem 13.1.2 allows us to conclude.