

7 Topics: Moment generating functions, conditional distribution and conditional expectation

7.1 Prerequisites: Lecture 18

Exercise 7- 1: Use moment generating functions to find the mean and variance of

- (a) $X \sim \text{Poi}(\lambda)$,
- (b) $X \sim \text{Bin}(n, p)$,
- (c) $X \sim \text{Exp}(\lambda)$,
- (d) $X \sim N(\mu, \sigma^2)$.

Solution:

- (a) $X \sim \text{Poi}(\lambda)$: For $t \in \mathbb{R}$, we have

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

Then

$$M'_X(t) = \exp(\lambda(e^t - 1)) \lambda e^t, \quad M''_X(t) = \exp(\lambda(e^t - 1)) \lambda e^t + \exp(\lambda(e^t - 1)) (\lambda e^t)^2,$$

and hence $E(X) = M'_X(0) = \lambda$, $E(X^2) = M''_X(0) = \lambda + \lambda^2$ and $\text{Var}(X) = \lambda + \lambda^2 - \lambda^2 = \lambda$.

- (b) $X \sim \text{Bin}(n, p)$: For $t \in \mathbb{R}$, we have

$$M_X(t) = \sum_{j=0}^n e^{jt} \binom{n}{j} p^j (1-p)^{n-j} = \sum_{j=0}^n \binom{n}{j} (e^t p)^j (1-p)^{n-j} = (e^t p + 1 - p)^n.$$

Then

$$M'_X(t) = n(e^t p + 1 - p)^{n-1} p e^t, \\ M''_X(t) = n(n-1)(e^t p + 1 - p)^{n-2} (p e^t)^2 + n(e^t p + 1 - p)^{n-1} p e^t,$$

hence $E(X) = M'_X(0) = np$ and $E(X^2) = M''_X(0) = np + n(n-1)p^2$, hence $\text{Var}(X) = np(1-p)$.

- (c) $X \sim \text{Exp}(\lambda)$: For $t < \lambda$, we have

$$M_X(t) = \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - t}.$$

Then

$$M'_X(t) = \frac{\lambda}{(\lambda - t)^2}, \quad M''_X(t) = \frac{2\lambda}{(\lambda - t)^3},$$

hence $E(X) = M'_X(0) = 1/\lambda$ and $E(X^2) = M''_X(0) = 2/\lambda^2$, hence $\text{Var}(X) = 1/\lambda^2$.

- (d) $X \sim N(\mu, \sigma^2)$. From the lectures we know that, for $t \in \mathbb{R}$, we have

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = e^{\mu t + \sigma^2 t^2 / 2}.$$

Then

$$M'_X(t) = e^{\mu t + \sigma^2 t^2 / 2} (\mu + \sigma^2 t), \quad M''_X(t) = e^{\mu t + \sigma^2 t^2 / 2} (\mu + \sigma^2 t)^2 + e^{\mu t + \sigma^2 t^2 / 2} \sigma^2,$$

hence $E(X) = M'_X(0) = \mu$ and $E(X^2) = M''_X(0) = \mu^2 + \sigma^2$, hence $\text{Var}(X) = \sigma^2$.

Exercise 7- 2: (Suggested for personal/peer tutorial) Use moment generating functions to prove that for independent random variables $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$, we have that $X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

Solution: From Theorem 13.2.9 we deduce that, for $t \in \mathbb{R}$,

$$M_{X+Y}(t) = M_X(t)M_Y(t) = e^{\mu_X t + \sigma_X^2 t^2/2} e^{\mu_Y t + \sigma_Y^2 t^2/2} = e^{(\mu_X + \mu_Y)t + \frac{1}{2}(\sigma_X^2 + \sigma_Y^2)t^2}.$$

We observe that the right hand side is the m.g.f. of a random variable with $N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$ distribution. Then Theorem 13.2.10 allows us to conclude.

Exercise 7- 3: Suppose that X_1 and X_2 are independent and identically distributed random variables, each having a standard normal distribution. Let random variable V be defined by

$$V = X_1^2 + X_2^2.$$

Find the pdf of V .

Solution: There are many possible routes for solving this problem.

- (a) The possibly simplest approach is to use m.g.f.s. We have from Exercise 5- 8 that if $X \sim N(0, 1)$ then

$$Y = X^2 \sim \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right)$$

and hence

$$\begin{aligned} M_Y(t) &= \int_0^\infty e^{ty} \left(\frac{1}{2\pi}\right)^{1/2} y^{-1/2} e^{-y/2} dy = \left(\frac{1}{2\pi}\right)^{1/2} \int_0^\infty y^{-1/2} e^{-y(1-2t)/2} dy \\ &= \left(\frac{1}{2\pi}\right)^{1/2} \frac{(2\pi)^{1/2}}{(1-2t)^{1/2}} = \frac{1}{(1-2t)^{1/2}}, \end{aligned}$$

for $\frac{1}{2} > t$, as the integrand is proportional to a $\text{Gamma}(1/2, (1-2t)/2)$ pdf.

More precisely, we note that, using a change of variable with $z = y(1-2t)/2 \Leftrightarrow 2z/(1-2t) = y$, $dy = 2/(1-2t)dz$,

$$\begin{aligned} \int_0^\infty y^{-1/2} e^{-y(1-2t)/2} dy &= \int_0^\infty \left(\frac{2z}{1-2t}\right)^{-1/2} e^{-z} dz \frac{2}{1-2t} = \left(\frac{2}{1-2t}\right)^{-1/2} \int_0^\infty z^{-1/2} e^{-z} dz \frac{2}{1-2t} \\ &= \left(\frac{2}{1-2t}\right)^{-1/2} \Gamma(1/2) \frac{2}{1-2t} = \frac{\sqrt{2\pi}}{\sqrt{1-2t}}, \end{aligned}$$

where we used that $\Gamma(1/2) = \sqrt{\pi}$. Hence

$$M_Y(t) = \left(\frac{1}{1-2t}\right)^{1/2}.$$

Now, let $Y_1 = X_1^2$ and $Y_2 = X_2^2$, so that Y_1 and Y_2 are independent $\text{Gamma}(1/2, 1/2)$ variables. Now we have from a key m.g.f. result, for $\frac{1}{2} > t$,

$$V = Y_1 + Y_2 \implies M_V(t) = M_{Y_1}(t)M_{Y_2}(t) = \left(\frac{1}{1-2t}\right)^{1/2} \left(\frac{1}{1-2t}\right)^{1/2} = \left(\frac{1}{1-2t}\right),$$

and hence, noting that this is the mgf of a Gamma random variable with parameters 1 and $\frac{1}{2}$, we conclude that

$$V \sim \text{Gamma}\left(1, \frac{1}{2}\right) = \text{Gamma}\left(\frac{2}{2}, \frac{1}{2}\right) \equiv \chi_2^2.$$

(b) By using a joint p.d.f. approach, we could also write

$$F_V(v) = P(V \leq v) = P(X_1^2 + X_2^2 \leq v) = \int_{A_v} \int f_{X_1}(x_1) f_{X_2}(x_2) dx_1 dx_2,$$

where

$$A_v = \{(x_1, x_2) : x_1^2 + x_2^2 \leq v\},$$

that is, the integral over the region A_v of the joint density function of X_1 and X_2 . Now, should reparameterize into polar coordinates in the double integral; let $x_1 = r \cos \theta$ and $x_2 = r \sin \theta$. Then

$$\begin{aligned} F_V(v) &= \int_{A_v} \int f_{X_1}(x_1) f_{X_2}(x_2) dx_1 dx_2 = \int_{A_v} \int \frac{1}{2\pi} e^{-(x_1^2 + x_2^2)/2} dx_1 dx_2 \\ &= \int_{r=0}^{\sqrt{v}} \int_0^{2\pi} \frac{1}{2\pi} e^{-r^2/2} r d\theta dr \end{aligned}$$

$$= \int_{r=0}^{\sqrt{v}} r e^{-r^2/2} dr = 1 - e^{-v/2}, \quad v > 0,$$

so $V \sim \text{Exp}(1/2) \equiv \chi_2^2$.

(c) We could also use the convolution theorem: Let $V = Y_1 + Y_2$ where $Y_1 = X_1^2$ and $Y_2 = X_2^2$. We have already shown that $Y_1 \sim \text{Gamma}(1/2, 1/2)$ so $f_{Y_1}(y) = (2\pi)^{-1/2} y^{-1/2} e^{-y/2}$, $y \geq 0$. The range of V is $(0, \infty)$. The convolution theorem is:

$$\begin{aligned} f_V(v) &= \int_{-\infty}^{\infty} f_{Y_1, Y_2}(y, v-y) dy \\ &= \int_0^v f_{Y_1}(y) f_{Y_2}(v-y) dy \quad \text{from independence} \\ &= \int_0^v \frac{1}{\sqrt{2\pi}} y^{-1/2} e^{-y/2} \frac{1}{\sqrt{2\pi}} (v-y)^{-1/2} e^{-(v-y)/2} dy \\ &= \frac{1}{2\pi} e^{-v/2} \int_0^v y^{-1/2} (v-y)^{-1/2} dy, \quad v > 0. \end{aligned}$$

Note that the kernel of this integral looks like a scaled Beta distribution. Substitute $z = y/v$ (could also use a trig substitution) then

$$\begin{aligned} f_V(v) &= \frac{1}{2\pi} e^{-v/2} \int_0^1 (vz)^{-1/2} (v-vz)^{-1/2} v dz \\ &= \frac{1}{2\pi} e^{-v/2} \int_0^1 (z)^{-1/2} (1-z)^{-1/2} dz \\ &= \frac{1}{2\pi} e^{-v/2} \frac{\Gamma(1/2)\Gamma(1/2)}{\Gamma(1)} = \frac{1}{2\pi} e^{-v/2} \sqrt{\pi}\sqrt{\pi} \\ &= \frac{1}{2} e^{-v/2}, \quad v > 0, \end{aligned}$$

which we recognise as the p.d.f. of an $\text{Exp}(1/2)$ random variable.

7.2 Prerequisites: Lecture 19

Exercise 7- 4: Consider tossing a coin repeatedly, where the probability of heads appearing in one toss is given by $p \in (0, 1)$. Let X denote the length of the initial run (i.e. if you toss heads first, how many heads do you toss before tossing tail and vice versa if you toss tail first). By conditioning on the outcome of the first coin toss and by using the law of total expectation, find $E(X)$.

Solution: Let H denote the event that the first coin toss gives heads. Then H^c is the event that the first coin toss gives tail. Then, for $x \in \mathbb{N}$, we have

$$P(X = x|H) = p^{x-1}(1 - p),$$

since after the initial toss of heads we need to toss heads $x - 1$ times and then toss tail ones to obtain a run of heads of length x .

Similarly, we get for $x \in \mathbb{N}$,

$$P(X = x|H^c) = (1 - p)^{x-1}p.$$

By the law of total probability (noting that H and H^c form a partition of the sample space), we get

$$E(X) = pE(X|H) + (1 - p)E(X|H^c).$$

Since

$$E(X|H) = \sum_{x=1}^{\infty} xP(X = x|H) = \sum_{x=1}^{\infty} xp^{x-1}(1 - p) = (1 - p) \sum_{x=1}^{\infty} xp^{x-1}.$$

Now we can use a little trick: Note that since $p \in (0, 1)$, we know that the geometric series is given by

$$\sum_{x=0}^{\infty} p^x = \frac{1}{1 - p} = f(p).$$

So, if we view this as a function in p and differentiate both the left and the right hand side (assuming we can interchange the infinite sum and the differentiation), then

$$f'(p) = \sum_{x=1}^{\infty} xp^{x-1} = \frac{1}{(1 - p)^2}.$$

Hence

$$E(X|H) = \sum_{x=1}^{\infty} xP(X = x|H) = \sum_{x=1}^{\infty} xp^{x-1}(1 - p) = \frac{(1 - p)}{(1 - p)^2} = \frac{1}{1 - p}.$$

Using the same arguments, we get

$$E(X|H^c) = \sum_{x=1}^{\infty} xP(X = x|H^c) = \sum_{x=1}^{\infty} x(1 - p)^{x-1}p = p \frac{1}{(1 - (1 - p))^2} = \frac{1}{p}.$$

So, altogether we have

$$\begin{aligned} E(X) &= pE(X|H) + (1 - p)E(X|H^c) = p \frac{1}{1 - p} + (1 - p) \frac{1}{p} = \frac{p^2 + (1 - p)^2}{(1 - p)p} \\ &= \frac{p^2 + 1 - 2p + p^2}{(1 - p)p} = \frac{2p^2 + 1 - 2p}{(1 - p)p} = -2 + \frac{1}{p(1 - p)}. \end{aligned}$$

Exercise 7- 5: Consider two discrete random variables X and Y with joint probability mass function given by

$$P(X = x, Y = y) = \begin{cases} c(2x + y), & \text{if } x \in \{0, 1, 2\} \text{ and } y \in \{0, 1, 2, 3\}, \\ 0, & \text{otherwise,} \end{cases}$$

where c is an appropriately-chosen constant.

- Find the value of c .
- Find $P(X = 2, Y = 1)$
- Find $P(X \geq 1, Y = 1)$
- Find $P(X \geq 1, Y \leq 1)$
- Find the marginal probability mass function of X .
- Find the marginal probability mass function of Y .
- Are X and Y independent random variables?
- Find the probability mass function of Y given $X = 2$.
- Compute $P(Y = 1|X = 2)$.
- Compute $E(Y|X = 2)$.

Solution:

- (a) Since the sum of the probabilities must equal 1, one can directly compute:

$$\begin{aligned} c \sum_{x=0}^2 \sum_{y=0}^3 (2x + y) &= 1 \\ \Leftrightarrow \sum_{x=0}^2 \sum_{y=0}^3 (2x + y) &= \frac{1}{c} \\ \Leftrightarrow \sum_{x=0}^2 \sum_{y=0}^3 2x + \sum_{x=0}^2 \sum_{y=0}^3 y &= \frac{1}{c} \\ \Leftrightarrow \sum_{x=0}^2 4 \cdot 2x + \sum_{x=0}^2 6 &= \frac{1}{c} \\ \Leftrightarrow 8(3) + 6(3) &= \frac{1}{c} \\ \Leftrightarrow 42 &= \frac{1}{c} \\ \Leftrightarrow c &= \frac{1}{42} \end{aligned}$$

Alternatively, one can construct the following table for the joint probability mass function:

	$Y = 0$	$Y = 1$	$Y = 2$	$Y = 3$	Totals ↓
$X = 0$	0	c	$2c$	$3c$	$6c$
$X = 1$	$2c$	$3c$	$4c$	$5c$	$14c$
$X = 2$	$4c$	$5c$	$6c$	$7c$	$22c$
Totals →	$6c$	$9c$	$12c$	$15c$	$42c$

Looking at the bottom right corner, we see that the sum of the probabilities add up to $42c$. But this implies

$$42c = 1 \Leftrightarrow c = \frac{1}{42}.$$

The Table above is very useful and will help us to quickly answer the next few questions.

(b) Reading off the table,

$$P(X = 2, Y = 1) = c(2 \cdot 2 + 1) = 5c = \frac{5}{42}.$$

(c) Again, one can use the table,

$$P(X \geq 1, Y = 1) = P(X = 1, Y = 1) + P(X = 2, Y = 1) = 3c + 5c = 8c = \frac{8}{42} = \frac{4}{21}.$$

(d) One first computes (using the table):

$$P(X \geq 1, Y = 0) = P(X = 1, Y = 0) + P(X = 2, Y = 0) = 2c + 4c = 6c = \frac{6}{42} = \frac{1}{7}.$$

Then, one can compute:

$$P(X \geq 1, Y \leq 1) = P(X \geq 1, Y = 0) + P(X \geq 1, Y = 1) = \frac{6}{42} + \frac{8}{42} = \frac{14}{42} = \frac{1}{3}.$$

(e) We can simply read the marginal totals in the right-hand "margin" of the table:

$$\begin{aligned} P(X = 0) &= 6c = \frac{6}{42} = \frac{1}{7}, \\ P(X = 1) &= 14c = \frac{14}{42} = \frac{1}{3}, \\ P(X = 2) &= 22c = \frac{22}{42} = \frac{11}{21}, \\ P(X = x) &= 0, \quad \forall x \notin \{0, 1, 2\}. \end{aligned}$$

(f) We can simply read the marginal totals in the last row of the table:

$$\begin{aligned} P(Y = 0) &= 6c = \frac{6}{42} = \frac{1}{7}, \\ P(Y = 1) &= 9c = \frac{9}{42} = \frac{3}{14}, \\ P(Y = 2) &= 12c = \frac{12}{42} = \frac{2}{7}, \\ P(Y = 3) &= 15c = \frac{15}{42} = \frac{5}{14}, \\ P(Y = y) &= 0, \quad \forall y \notin \{0, 1, 2, 3\}. \end{aligned}$$

(g) If the random variables X and Y are independent, then for all values of x and y :

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

Now, recall:

- In Part (b) we computed $P(X = 2, Y = 1) = 5c = \frac{5}{42}$.
- From the marginal distribution of X in Part (e), one can compute $P(X = 2) = \frac{11}{21}$.
- From the marginal distribution of Y in Part (f), $P(Y = 1) = \frac{3}{14}$.

Therefore,

$$P(X = 2)P(Y = 1) = \frac{11}{21} \cdot \frac{3}{14} = \frac{11}{7} \cdot \frac{1}{14} = \frac{11}{98} \neq \frac{5}{42} = P(X = 2, Y = 1),$$

and so X and Y are not independent.

(h)

$$P(Y = y|X = 2) = \frac{P(Y = y, X = 2)}{P(X = 2)} = \frac{c(2 \cdot 2 + y)}{11/21} = \frac{4 + y}{11} \cdot \frac{21}{11} = \frac{4 + y}{22},$$

for $y \in \{0, 1, 2, 3\}$ and 0 otherwise.

(i) Using Part (h), we have

$$P(Y = 1|X = 2) = \frac{4 + 1}{22} = \frac{5}{22}.$$

(j) Using Part (h), we have

$$\begin{aligned} E(Y|X = 2) &= \sum_{y=0}^3 yP(y|X = 2) = \sum_{y=0}^3 y \frac{(4 + y)}{22} \\ &= 0 + \frac{5}{22} + 2 \cdot \frac{6}{22} + 3 \cdot \frac{7}{22} = \frac{19}{11}. \end{aligned}$$

Exercise 7- 6: Consider two jointly continuous random variables (X, Y) with joint density function given by

$$f_{X,Y}(x, y) = cxy, \quad \text{for } 0 \leq x \leq 1, 0 \leq y \leq 1,$$

and $f_{X,Y}(x, y) = 0$ otherwise.

- Determine the constant c such that $f_{X,Y}$ is a valid density function.
- Find the marginal density of X .
- Find the conditional density of $Y|X = x$.
- Find the conditional distribution of $Y|X = x$.
- Are X and Y independent?

Solution:

(a) Note that

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy = \int_0^1 \int_0^1 cxy dx dy = c \int_0^1 \left(\frac{1}{2} x^2 y \Big|_0^1 \right) dy \\ &= c \int_0^1 \frac{1}{2} y dy = c \frac{1}{4} y^2 \Big|_0^1 = c \frac{1}{4} \quad \Leftrightarrow c = 4. \end{aligned}$$

(b)

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = \int_0^1 4xy dy = \frac{4}{2} xy^2 \Big|_0^1 = 2x,$$

for all $x \in [0, 1]$ and 0 otherwise.

(c) Let $x \in [0, 1]$. Then

$$f_{Y|X=x}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{4xy}{2x} = 2y,$$

for all $y \in [0, 1]$ and 0 otherwise.

(d) Let $x \in [0, 1]$. Then

$$F_{Y|X}(y|x) = \int_0^y 2v dv = y^2,$$

for $y \in [0, 1]$, $F_{Y|X}(y|x) = 0$, for $y < 0$ and $F_{Y|X}(y|x) = 1$ for $y > 1$.

(e) We observe that X and Y are independent, since their joint density factorises into the product of the marginal densities: $f_{X,Y}(x, y) = 4xy = 2x \times 2y = f_X(x)f_Y(y)$ for $x, y \in [0, 1]$ and $0 = 0$ otherwise.

Hence, the conditional distribution of $Y|X = x$ does not depend on x .

Exercise 7- 7: Consider two jointly continuous random variables (X, Y) with joint density function given by

$$f_{X,Y}(x, y) = c(x + y), \quad \text{for } 0 \leq y \leq x \leq 1,$$

and $f_{X,Y}(x, y) = 0$ otherwise.

- Determine the constant c such that $f_{X,Y}$ is a valid density function.
- Find the marginal density of X .
- Find the marginal density of Y .
- Are X and Y independent?
- Find the conditional density of $Y|X = x$.
- Find the conditional distribution of $Y|X = x$.

Solution:

(a) Note that

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy = \int_0^1 \int_y^1 c(x + y) dx dy = c \int_0^1 \left(\frac{1}{2}x^2 + xy \right) \Big|_y^1 dy \\ &= c \int_0^1 \left(\frac{1}{2} + y - \frac{3}{2}y^2 \right) dy = c \left(\frac{1}{2}y + \frac{1}{2}y^2 - \frac{3}{2} \frac{1}{3}y^3 \right) \Big|_0^1 = c \frac{1}{2} \quad \Leftrightarrow c = 2. \end{aligned}$$

(b)

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = \int_0^x 2(x + y) dy = 2xy + y^2 \Big|_0^x = 3x^2,$$

for all $x \in [0, 1]$ and 0 otherwise.

(c)

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx = \int_y^1 2(x + y) dx = x^2 + 2xy \Big|_y^1 = 1 + 2y - y^2 - 2y^2 = 1 + 2y - 3y^2,$$

for all $y \in [0, 1]$ and 0 otherwise.

(d) X and Y are not independent, since we observe from the above computations that there exists x, y (e.g. $x = y = 1$) such that $f_{X,Y}(x, y) \neq f_X(x)f_Y(y)$.

(e) Let $x \in [0, 1]$. Then

$$f_{Y|X=x}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{2(x+y)}{3x^2},$$

for all $0 \leq y \leq x$ and 0 otherwise.

(f) Let $x \in [0, 1]$. Then

$$F_{Y|X}(y|x) = \int_0^y \frac{2(x+v)}{3x^2} dv = \frac{1}{3x^2} (2xv + v^2|_0^y) = \frac{2xy + y^2}{3x^2},$$

for $0 \leq y \leq x$, $F_{Y|X}(y|x) = 0$, for $y < 0$ and $F_{Y|X}(y|x) = 1$ for $y > x$.