Sample answers

Session 7

Proposition 4.4 *Let X and Y be discrete random variables. Then X and Y are independent if and only if, for any values a_i and b_j of X and Y respectively, we have*

$$P(X = a_i \mid Y = b_j) = P(X = a_i).$$

Note that Proposition 4.4 holds only if for any b_j the probability $P(Y = b_j) > 0$

Q1. How to proof?

Q1.

"$\Rightarrow$" giving X and Y are independent, we would prove $\forall a_i, b_j$ ~~we~~ having $P(X=a_i \mid Y=b_j) = P(X=a_i)$

X and Y independent

$\Rightarrow \forall a_i, b_j$, we have

$$P(X=a_i, Y=b_j) = P(X=a_i) \cdot P(X=b_j)$$

$$\Rightarrow P(X=a_i \mid Y=b_j) P(Y=b_j) = P(X=a_i) \cdot P(Y=b_j) \quad \text{(Bayesian)}$$

$$\Rightarrow P(X=a_i \mid Y=b_j) = P(X=a_i) \quad (\because P(Y=b_j) > 0)$$

" $\Leftarrow$ " giving $\forall a_i, b_j \quad P(X=a_i \mid Y=b_j) = P(X=a_i)$, we would prove X and Y are independent

$$\forall a_i, b_j \quad P(X=a_i \mid Y=b_j) = P(X=a_i)$$

$$\Rightarrow P(X=a_i \mid Y=b_j) \cdot P(Y=b_j) = P(X=a_i) \cdot P(Y=b_j)$$

$$\Rightarrow P(X=a_i, Y=b_j) = P(X=a_i) \cdot P(Y=b_j) \quad \text{(Bayesian)}$$

$\Rightarrow$ X and Y are independent (by definition)

Q2 An urn has 2 red, 5 white, and 3 green balls. Select 3 balls at random without replacement and let X be the number of red balls and Y the number of white balls. Determine
(a) joint p. m. f. of (X, Y),
(b) marginal p. m. f.'s, (c) $P(X \geq Y)$, and (d) $P(X = 2 | X \geq Y)$.

The joint p. m. f. is given by $P(X = x, Y = y)$ for all possible x and y. In our case, x can be $0, 1$, or 2 and y can be $0, 1, 2$, or 3. The values can be given by the formula

$$P(X = x, Y = y) = \frac{\binom{2}{x}\binom{5}{y}\binom{3}{3-x-y}}{\binom{10}{3}},$$

where we use the convention that $\binom{a}{b} = 0$ if $b > a$, or in the table:

$y \backslash x$	0	1	2	$P(Y = y)$
0	$1/120$	$2 \cdot 3/120$	$3/120$	$10/120$
1	$5 \cdot 3/120$	$2 \cdot 5 \cdot 3/120$	$5/120$	$50/120$
2	$10 \cdot 3/120$	$10 \cdot 2/120$	0	$50/120$
3	$10/120$	0	0	$10/120$
$P(X = x)$	$56/120$	$56/120$	$8/120$	1

The last row and column entries are the respective column and row sums and, therefore, determine the marginal p. m. f.'s. To answer (c) we merely add the relevant probabilities,

$$P(X \geq Y) = \frac{1+6+3+30+5}{120} = \frac{3}{8},$$

and, to answer (d), we compute

$$\frac{P(X=2, X \geq Y)}{P(X \geq Y)} = \frac{\frac{8}{120}}{\frac{3}{8}} = \frac{8}{45}.$$

EX1 Two numbers X and Y are chosen independently from the uniform distribution on the unit interval [0,1]. Let Z be the maximum of the two numbers. Find the p.d.f. of Z, and hence find its expected value, variance and median.

Solution The c.d.f.s of X and Y are identical, that is,

$$F_X(x) = F_Y(x) = \begin{cases} 0 & \text{if } x < 0, \\ x & \text{if } 0 < x < 1, \\ 1 & \text{if } x > 1. \end{cases}$$

(The variable can be called x in both cases; its name doesn't matter.)

The key to the argument is to notice that

$$Z = \max(X,Y) \le x \quad \text{if and only if} \quad X \le x \text{ and } Y \le x.$$

(For, if both X and Y are smaller than a given value x, then so is their maximum; but if at least one of them is greater than x, then again so is their maximum.) For $0 \le x \le 1$, we have $P(X \le x) = P(Y \le x) = x$; by independence,

$$P(X \le x \text{ and } Y \le x) = x \cdot x = x^2.$$

Thus $P(Z \leq x) = x^2$. Of course this probability is 0 if $x < 0$ and is 1 if $x > 1$. So the c.d.f. of Z is

$$F_Z(x) = \begin{cases} 0 & \text{if } x < 0, \\ x^2 & \text{if } 0 < x < 1, \\ 1 & \text{if } x > 1. \end{cases}$$

The median of Z is the value of m such that $F_Z(m) = 1/2$, that is $m^2 = 1/2$, or $m = 1/\sqrt{2}$.

We obtain the p.d.f. of Z by differentiating:

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

Then we can find $E(Z)$ and $\mathrm{Var}(Z)$ in the usual way:

$$E(Z) = \int_0^1 2x^2 \mathrm{d}x = \frac{2}{3}, \qquad \mathrm{Var}(Z) = \int_0^1 2x^3 \mathrm{d}x - \left(\frac{2}{3}\right)^2 = \frac{1}{18}.$$

EX2 I roll a fair die bearing the numbers 1 to 6. If N is the number showing on the die, I then toss a fair coin N times. Let X be the number of heads I obtain.
(a) Write down the p.m.f. for X.
(b) Calculate $E(X)$ without using this information.

Solution (a) If we were given that $N = n$, say, then X would be a binomial $\mathrm{Bin}(n, 1/2)$ random variable. So $P(X = k \mid N = n) = {}^{n}C_k(1/2)^n$.

By the ToTP,

$$P(X = k) = \sum_{n=1}^{6} P(X = k \mid N = n)P(N = n).$$

Clearly $P(N = n) = 1/6$ for $n = 1, \ldots, 6$. So to find $P(X = k)$, we add up the probability that $X = k$ for a $\mathrm{Bin}(n, 1/2)$ r.v. for $n = k, \ldots, 6$ and divide by 6. (We start at k because you can't get k heads with fewer than k coin tosses!) The answer comes to

k	0	1	2	3	4	5	6
$P(X = k)$	$\frac{63}{384}$	$\frac{120}{384}$	$\frac{99}{384}$	$\frac{64}{384}$	$\frac{29}{384}$	$\frac{8}{384}$	$\frac{1}{384}$

For example,

$$P(X = 4) = \frac{{}^{4}C_4(1/2)^4 + {}^{5}C_4(1/2)^5 + {}^{6}C_4(1/2)^6}{6} = \frac{4 + 10 + 15}{384}.$$

(b) By Proposition 4.3,

$$E(X) = \sum_{n=1}^{6} E(X \mid (N = n))P(N = n).$$

Now if we are given that $N = n$ then, as we remarked, X has a binomial $\mathrm{Bin}(n, 1/2)$ distribution, with expected value $n/2$. So

$$E(X) = \sum_{n=1}^{6} (n/2) \cdot (1/6) = \frac{1+2+3+4+5+6}{2 \cdot 6} = \frac{7}{4}.$$

Try working it out from the p.m.f. to check that the answer is the same!

EX3 Let

$$f(x,y) = \begin{cases} c\,x^2 y & \text{if } x^2 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Determine (a) the constant c, (b) $P(X \ge Y)$, (c) $P(X = Y)$, and (d) $P(X = 2Y)$, (e) Compute marginal densities and determine whether X and Y are independent.

For (a),

$$\begin{aligned} \int_{-1}^{1} dx \int_{x^2}^{1} c\,x^2 y\,dy &= 1 \\ c \cdot \frac{4}{21} &= 1 \end{aligned}$$

and so

$$c = \frac{21}{4}.$$

For (b), let S be the region between the graphs $y = x^2$ and $y = x$, for $x \in (0,1)$. Then,

$$\begin{aligned} P(X \ge Y) &= P((X,Y) \in S) \\ &= \int_0^1 dx \int_{x^2}^{x} \frac{21}{4} \cdot x^2 y\,dy \\ &= \frac{3}{20} \end{aligned}$$

Both probabilities in (c) and (d) are 0 because a two-dimensional integral over a line is 0.

(e)

We have

$$f_X(x) = \int_{x^2}^{1} \frac{21}{4} \cdot x^2 y \, dy = \frac{21}{8} x^2 (1 - x^4),$$

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

$$f_Y(y) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{21}{4} \cdot x^2 y \, dx = \frac{7}{2} y^{\frac{5}{2}},$$

where $y \in [0, 1]$, and 0 otherwise. The two random variables X and Y are clearly not independent, as $f(x, y) \neq f_X(x) f_Y(y)$.

EX4 Assume that you are waiting for two phone calls, from Alice and from Bob. The waiting time T_1 for Alice's call has expectation 10 minutes and the waiting time T_2 for Bob's call has expectation 40 minutes. Assume T_1 and T_2 are independent exponential random variables. What is the probability that Alice's call will come first?

We need to compute $P(T_1 < T_2)$. Assuming our unit is 10 minutes, we have, for $t_1, t_2 > 0$,

$$f_{T_1}(t_1) = e^{-t_1}$$

and

$$f_{T_2}(t_2) = \frac{1}{4} e^{-t_2/4},$$

so that the joint density is

$$f(t_1, t_2) = \frac{1}{4} e^{-t_1 - t_2/4},$$

for $t_1, t_2 \geq 0$. Therefore,

$$\begin{aligned} P(T_1 < T_2) &= \int_0^\infty dt_1 \int_{t_1}^\infty \frac{1}{4} e^{-t_1 - t_2/4}\, dt_2 \\ &= \int_0^\infty e^{-t_1}\, dt_1\, e^{-t_1/4} \\ &= \int_0^\infty e^{\frac{-5t_1}{4}}\, dt_1 \\ &= \frac{4}{5}. \end{aligned}$$

EX5 The joint density of (X, Y) is given by

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

(a) Compute the conditional density of Y given $X = x$.
(b) Are X and Y independent?

(a) Assume that $x \in [0,1]$. As

$$f_X(x) = \int_0^x 3x\,dy = 3x^2,$$

we have

$$f_Y(y|X=x) = \frac{f(x,y)}{f_X(x)} = \frac{3x}{3x^2} = \frac{1}{x},$$

for $0 \le y \le x$. In other words, Y is uniform on $[0, x]$.

(b) As the answer in (a) depends on x, the two random variables are not independent.