Parametric Statistics-Recitation 3 (Solutions)

Exercise 1.

Suppose that X and Y have a continuous joint distribution for which the joint p.d.f. is defined as follows:

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

Determine (a) the value of the constant c; (b) $\Pr(X + Y > 2)$; (c) $\Pr(Y < 1/2)$; (d) $\Pr(X \le 1)$; (e) $\Pr(X = 3Y)$.

Solution.

(a) Since f is the joint p.d.f of X,Y it must be $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = 1$.
But $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = \int_0^1\int_0^2 cy^2dxdy = \frac{2c}{3}$. So $c = \frac{3}{2}$.

(b) We want to integrate the function $f(x,y)$ in the region:

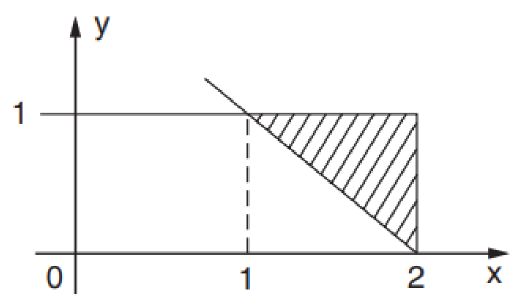


Figure 1: Region of integration

So $Pr(X+Y>2) = \int_1^2\int_{2-x}^1 \frac{3y^2}{2}dydx = \frac{3}{8}$.

(c) For $Pr(Y<\frac{1}{2})$ the region to integrate is

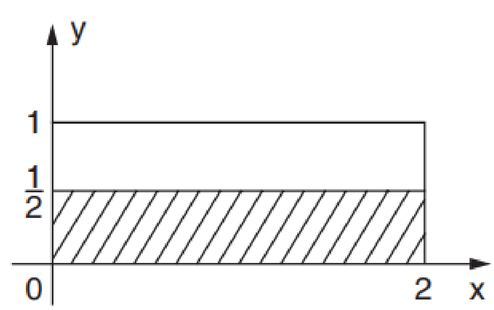


Figure 2: Region of integration

So $Pr(Y<\frac{1}{2}) = \int_0^2\int_0^{\frac{1}{2}} \frac{3y^2}{2}dydx = \frac{1}{8}$.

(d) For $Pr(X \le 1)$ the region to integrate is

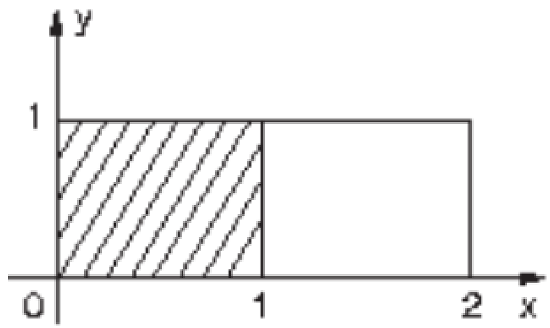


Figure 3: Region of integration

So $Pr(X \leq 1) = \int_0^1 \int_0^1 \frac{3y^2}{2} dydx = \frac{1}{2}$.

(e) The probability that (X, Y) will lie on the line $x = 3y$ is 0 for every continuous joint distribution.

Exercise 2.

Suppose that the joint p.d.f. of X and Y is as follows:

$$f(x,y) = \begin{cases} 2xe^{-y} & \text{for } 0 \leq x \leq 1 \text{ and } 0 < y < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

Are X and Y independent?

Solution.

Since $f(x, y)$ is 0 outside a rectangle and $f(x, y)$ can factorized $f(x, y) = g(x)h(y)$ inside the rectangle for some non-negative functions g, h of x, y respectively ($g(x) = 2x$ and $h(y) = e^{-y}$), it follows that X and Y are independent.

Exercise 3.

Suppose that the joint p.d.f. of X and Y is as follows:

$$f(x,y) = \begin{cases} 24xy & \text{for } x \geq 0, y \geq 0, \text{ and } x + y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Are X and Y independent?

Solution.

Although $f(x, y)$ can be factored as $f(x, y) = g(x)h(y)$ inside the triangle where $f(x, y) > 0$, the fact that $f(x, y) > 0$ inside a triangle, rather than a rectangle, implies that X and Y cannot be independent.
For example suppose $f(x, y) = g(x)h(y)$. Since $f(\frac{1}{3}, \frac{1}{4}) = 2$ and $f(\frac{1}{6}, \frac{3}{4}) = 3$ it must be $g(\frac{1}{3}) > 0$ and $h(\frac{3}{4}) > 0$. However, since $f(\frac{1}{3}, \frac{3}{4}) = 0$, it must be $g(\frac{1}{3}) = 0$ or $h(\frac{3}{4}) = 0$. But this is a contradiction.

Exercise 4.

Each student in a certain high school was classified according to her year in school (freshman, sophomore, junior, or senior) and according to the number of times that she had visited a certain museum (never, once, or more than once). The proportions of students in the various classifications are given in the following table:

	Never	Once	More than once
Freshmen	0.08	0.10	0.04
Sophomores	0.04	0.10	0.04
Juniors	0.04	0.20	0.09
Seniors	0.02	0.15	0.10

(a) If a student selected at random from the high school is a junior, what is the probability that she has never visited the museum?

(b) If a student selected at random from the high school has visited the museum three times, what is the probability that she is a senior?

Solution.

(a) We have that $Pr(Junior) = 0.04 + 0.2 + 0.09 = 0.33$. Therefore,

$$Pr(Never|Junior) = \frac{Pr(Junior, Never)}{Pr(junior)} = \frac{0.04}{0.33} = \frac{4}{33}$$

(b) We classify the student as having visited the museum more than once.
So Pr(More than once)$= 0.04 + 0.04 + 0.09 + 0.1 = 0.27$. Therefore,

$$Pr(\text{Senior}|\text{More than once}) = \frac{Pr(\text{Senior,More than once})}{Pr(\text{More than once})} = \frac{0.1}{0.27} = \frac{10}{27}$$

Exercise 5.

Let Y be the rate (calls per hour) at which calls arrive at a switchboard. Let X be the number of calls during a two-hour period. Suppose that the marginal p.d.f. of Y is

$$f_2(y) = \begin{cases} e^{-y} & \text{if } y > 0, \\ 0 & \text{otherwise,} \end{cases}$$

and that the conditional p.f. of X given $Y = y$ is

$$g_1(x \mid y) = \begin{cases} \frac{(2y)^x}{x!} e^{-2y} & \text{if } x = 0, 1, \dots, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the marginal p.f. of X. (You may use the formula $\int_0^\infty y^k e^{-y} dy = k!$!)

(b) Find the conditional p.d.f. $g_2(y \mid 0)$ of Y given $X = 0$.

(c) Find the conditional p.d.f. $g_2(y \mid 1)$ of Y given $X = 1$.

(d) For what values of y is $g_2(y \mid 1) > g_2(y \mid 0)$? Does this agree with the intuition that the more calls you see, the higher you should think the rate is?

Solution.

(a) The joint p.f/p.d.f of X and Y is the product $f_2(y)g_1(x|y)$.

$$f(x,y) = \begin{cases} \frac{(2y)^x}{x!}e^{-3y} & \text{if } y > 0, x = 0, 1, \ldots, \\ 0 & \text{otherwise} \end{cases}$$

The marginal p.f of X is obtained by integrating over y.

$$f_1(x) = \int_0^\infty \frac{(2y)^x}{x!}e^{-3y}dy = \frac{2^x}{x!3^x}\int_0^\infty (3y)^x e^{-3y}dy = \frac{2^x}{x!3^x}\frac{x!}{3} = \frac{1}{3}\left(\frac{2}{3}\right)^x$$

for $x = 0, 1, ...$

(b) The conditional p.d.f of Y given $X = 0$ is the ratio of the joint p.f/p.d.f to $f_1(0)$.

$$g_2(y|0) = \frac{(2y)^0 \frac{e^{-3y}}{0!}}{\frac{1}{3}\left(\frac{2}{3}\right)^0} = 3e^{-3y},$$

for $y > 0$.

(c) The conditional p.d.f of Y given $X = 1$ is the ratio of the joint p.f/p.d.f to $f_1(1)$.

$$g_2(y|1) = \frac{(2y)^1 \frac{e^{-3y}}{1!}}{\frac{1}{3}\left(\frac{2}{3}\right)^1} = 9ye^{-3y},$$

for $y > 0$.

(d) The ratio of the two conditional p.d.f's is

$$\frac{g_2(y|1)}{g_2(y|0)} = \frac{9ye^{-3y}}{3e^{-3y}} = 3y.$$

The ratio is greater than 1 if $y > \frac{1}{3}$. This corresponds to the intuition that if we observe more calls, then we should think the rate is higher.

Exercise 6.

Suppose that a random variable X can have each of the seven values $-3, -2, -1, 0, 1, 2, 3$ with equal probability. Determine the p.f. of $Y = X^2 - X$.

Solution.

For each possible value of x, we have the following value of $y = x^2 - x$:

x	y
-3	12
-2	6
-1	2
0	0
1	0
2	2
3	6

Since the probability of each of X is $\frac{1}{7}$, it follows that the p.f $g(y)$ is as follows:

y	$g(y)$
0	$\frac{2}{7}$
2	$\frac{2}{7}$
6	$\frac{2}{7}$
12	$\frac{1}{7}$

Exercise 7.

Suppose that the p.d.f. of X is as follows:

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

Determine the p.d.f. of $Y = X^{1/2}$.

Solution.

First we will find the c.d.f of Y, $G(y)$.

$$G(y) = P(Y \leq y) = P(X^{1/2} < y) = P(X \leq y^2) = \int_0^{y^2} e^{-x} dx = 1 - e^{-y^2}$$

So the p.d.f of Y is $g(y) = \frac{dG(y)}{dy} = 2ye^{-y^2}$.
Or we can use the theorem:
As x varies over all positive values, y also varies over all positive values and $x = y^2$. So

$$g(y) = f(y^2)\frac{dx}{dy} = 2ye^{-y^2}$$

Exercise 8.

Let X and Y be random variables for which the joint p.d.f. is as follows:

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

Find the p.d.f. of $Z = X + Y$.

Solution.

By theorem the p.d.f of $Z = X + Y$ is given by $g(y) = \int_{-\infty}^{\infty} f(z - t, t)dt$ for $-\infty < z < \infty$. The integrated is positive only for $0 \leq z - t \leq t \leq 1$ and $0 \leq z \leq 2$.

If $0 \leq z \leq 1$ then $\frac{z}{2} \leq t \leq z$ and we have

$$g(z) = \int_{\frac{z}{2}}^{z} 2zdt = z^2$$

and if $1 < z \leq 2$ then $\frac{z}{2} \leq t \leq 1$ and we have

$$g(z) = \int_{\frac{z}{2}}^{1} 2z dt = z(2-z)$$

Exercise 9.

Suppose that X_1 and X_2 are i.i.d. random variables and that the p.d.f. of each of them is as follows:

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

Find the p.d.f. of $Y = X_1 - X_2$.

Solution.

Let $Z = -X_2$. Then the p.d.f of Z is

$$f_2(z) = \begin{cases} e^{z} & \text{for } z < 0 \\ 0 & \text{for } z \geq 0 \end{cases}$$

Since X_1 and Z are independent, the joint p.d.f of X_1 and Z is

$$f(x_1, z) = \begin{cases} e^{-(x-z)} & \text{for } x > 0, z < 0 \\ 0 & \text{otherwise.} \end{cases}$$

It now follows from the previous exercise that the p.d.f of $Y = X_1 - X_2 = X_1 + Z$ is

$$g(y) = \int_{-\infty}^{\infty} f(y-z, z) dz$$

The integrand is positive only for $y - z > 0$ and $z < 0$. If $y \leq 0$,

$$g(y) = \int_{-\infty}^{y} e^{-(y-2z)} dz = \frac{1}{2} e^{y}$$

If $y > 0$,

$$g(y) = \int_{-\infty}^{0} e^{-(y-2z)} dz = \frac{1}{2} e^{-y}$$

Exercise 10.

Suppose that a point is chosen at random on a stick of unit length and that the stick is broken into two pieces at that point. Find the expected value of the length of the longer piece.

Solution.

If X denotes the point at which the stick is broken, then $X \sim U[0,1]$. If Y denotes the lenght of the longer piece, then $Y = max\{X, 1-X\}$. Therefore,

$$E(Y) = \int_{0}^{1} max\{X, 1-X\} dx = \int_{0}^{1/2} (1-x) dx + \int_{1/2}^{1} x dx = \frac{3}{4}$$

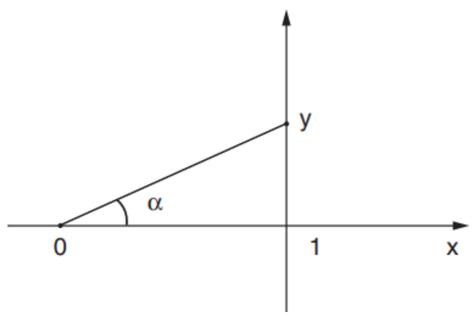
Exercise 11.

Suppose that a particle is released at the origin of the xy-plane and travels into the half-plane where $x > 0$. Suppose that the particle travels in a straight line and that the angle between the positive half of the x-axis and this line is α, which can be either positive or negative. Suppose, finally, that the angle α has the uniform distribution on the interval $[-\pi/2, \pi/2]$. Let Y be the ordinate of the point at which the particle hits the vertical line $x = 1$. Show that the distribution of Y is a Cauchy distribution.

Solution.

Since $a \sim U[-\frac{\pi}{2}, \frac{\pi}{2}]$, the p.d.f of a is

$$f(a) = \begin{cases} \frac{1}{\pi} & \text{for } -\frac{\pi}{2} < a < -\frac{\pi}{2} \\ 0 & \text{otherwise.} \end{cases}$$



Also, $Y = tan(a)$. Therefore, the inverse transformation is $a = tan^{-1}Y$ and $\frac{da}{dy} = \frac{1}{1+y^2}$. As a varies over the interval $(-\frac{\pi}{2}, \frac{\pi}{2})$, Y varies over the entire real line. Therefore, for $-\infty < y < \infty$, the p.d.f of Y is

$$g(y) = f(tan^{-1}y)\frac{1}{1+y^2} = \frac{1}{\pi(1+y^2)}$$

So Y is a Cauchy distribution.

Exercise 12.

Suppose that a class contains 10 boys and 15 girls, and suppose that eight students are to be selected at random from the class without replacement. Let X denote the number of boys that are selected, and let Y denote the number of girls that are selected. Find $E(X - Y)$.

Solution.

We observe that $Y = 8 - X$. So $E(X - Y) = E(X - (8 - X)) = E(2X - 8)$.
Also $X \sim$Hypergeometric$(10, 10 + 5, 8)$ and $E(X) = 8 \cdot \frac{10}{25} = \frac{16}{5}$. So $E(X - Y) = -\frac{8}{5}$.

Exercise 13.

Suppose that one word is selected at random from the sentence THE GIRL. PUT ON HER BEAUTIFUL RED HAT. If X denotes the number of letters in the word that is selected, what is the value of Var(X)?

Solution.

X takes the values $\{2, 3, 4, 9\}$ with probabilities $\{\frac{1}{8}, \frac{5}{8}, \frac{1}{8}, \frac{1}{8}\}$ respectively. So

$$E(X) = 2 \cdot \frac{1}{8} + 3 \cdot \frac{5}{8} + 4 \cdot \frac{1}{8} + 9 \cdot \frac{1}{8} = \frac{15}{4}$$
$$E(X^2) = 2^2 \cdot \frac{1}{8} + 3^2 \cdot \frac{5}{8} + 4^2 \cdot \frac{1}{8} + 9^2 \cdot \frac{1}{8} = \frac{73}{4}$$

Therefore

$$Var(X) = E(X^2) - [E(X)]^2 = \frac{67}{16}$$

Exercise 14.

Suppose that X is a random variable for which $E(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$. Show that $E[X(X-1)] = \mu(\mu-1) + \sigma^2$.

Solution.

$E(X(X-1)) = E(X^2 - X) = E(X^2) - E(X) = Var(X) + [E(X)]^2 - E(X) = \sigma^2 + \mu^2 - \mu = \sigma^2 + \mu(\mu-1)$

Exercise 15.

Suppose that X and Y are independent random variables whose variances exist and such that $E(X) = E(Y)$. Show that

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

Solution.

Since $E(X) = E(Y)$ we have that $E(X - Y) = 0$. Therefore

$$E\left[(X-Y)^2\right] = Var(X-Y) = Var(X + (-Y)) = Var(X) + Var(Y)$$

because X and Y are independent.

Exercise 16.

For all random variables X and Y and all constants a, b, c, and d, show that

$$\mathrm{Cov}(aX + b, cY + d) = ac\,\mathrm{Cov}(X, Y).$$

Solution.

We have that $E(aX + b) = a\mu_X + b$ and $E(cY + d) = c\mu_Y + d$. Therefore,

$\mathrm{Cov}(aX+b, cY+d) = E[(aX+b-a\mu_X-b)(cY+d-c\mu_Y-d)] = E[ac(X-\mu_X)(Y-\mu_Y)] = ac\,\mathrm{Cov}(X,Y)$

Exercise 17.

Let X, Y, and Z be three random variables such that $\mathrm{Cov}(X, Z)$ and $\mathrm{Cov}(Y, Z)$ exist, and let a, b, and c be arbitrary given constants. Show that

$$\mathrm{Cov}(aX + bY + c, Z) = a\,\mathrm{Cov}(X, Z) + b\,\mathrm{Cov}(Y, Z)$$

Solution.

We have that $E(aX + bY + c) = a\mu_X + b\mu_Y + c$. Therefore $\mathrm{Cov}(aX + bY + c, Z) =$
$E[(aX + bY + c - a\mu_X - b\mu_Y - c)(Z - \mu_Z)] = ... = a\,\mathrm{Cov}(X, Z) + b\,\mathrm{Cov}(Y, Z)$