407 Final Solutions[1]

1. Question 1

Let X and Y be independent random variables.
Suppose X is uniformly distributed in the integers $\{1,2,3,\ldots,100\}$.
Suppose Y is uniformly distributed in the integers $\{-100,-99,-98,\ldots,-1\}$.
What is the PMF of $X+Y$? Simplify your answer to the best of your ability.
Solution. We condition on the value of X. We have

$$\mathbf{P}(X+Y=k)=\sum_{j=1}^{100}\mathbf{P}(X+Y=k\,|X=j)\mathbf{P}(X=j)=\frac{1}{100}\sum_{j=1}^{100}\mathbf{P}(Y=k-j\,|X=j)$$
$$=\frac{1}{100}\sum_{j=1}^{100}\mathbf{P}(Y=k-j)$$

The last equality used the independence of X and Y. So,

$$\mathbf{P}(X+Y=k)=\frac{1}{100}\sum_{j=1}^{k}\mathbf{P}(Y=k-j)=\frac{1}{100}\sum_{j=1}^{k}\frac{1}{100}1_{-100\leq k-j\leq -1}=\frac{1}{10000}\sum_{j=1}^{k}1_{1+k\leq j\leq 100+k}. \qquad (*)$$

When k is an integer with $-99\leq k\leq -1$, we have

$$\mathbf{P}(X+Y=k)=\frac{1}{10000}\sum_{j=1}^{100+k}=\frac{100+k}{10000}.$$

When k is an integer with $0\leq k\leq 99$, we have by $(*)$

$$\mathbf{P}(X+Y=k)=\frac{1}{10000}\sum_{j=k+1}^{100}=\frac{100-k}{10000}.$$

In summary, for any $-99\leq k\leq 99$, we have

$$\mathbf{P}(X+Y=k)=\frac{100-|k|}{10000},$$

and $\mathbf{P}(X+Y=k)=0$ for any other integer k.

2. Question 2

Give an example of the joint density of two continuous random variables X and Y such that: X and Y are **NOT** independent.

Prove that the X and Y you find are not independent.

Solution. Many examples work here, e.g. Example 5.49 in the notes: Suppose X and Y have a joint PDF given by $f_{X,Y}(x,y)=\frac{1}{\pi}$ if $x^2+y^2\leq 1$, and $f_{X,Y}(x,y)=0$ otherwise. Note:

$$\iint_{x^2+y^2\leq 1} f_{X,Y}(x,y)dxdy=\frac{1}{\pi}\pi=1,$$

[1]November 22, 2020,

so $f_{X,Y}$ is a joint PDF. Let $x, y \in \mathbf{R}$ with $x^2 + y^2 \leq 1$. Using the definition of marginal,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \int_{y=-\sqrt{1-x^2}}^{y=\sqrt{1-x^2}} \frac{1}{\pi}dy = \frac{2\sqrt{1-x^2}}{\pi}.$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \int_{x=-\sqrt{1-y^2}}^{x=\sqrt{1-y^2}} \frac{1}{\pi}dx = \frac{2\sqrt{1-y^2}}{\pi}.$$

So, if $x^2 + y^2 \leq 1$, then

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{1/\pi}{2\sqrt{1-y^2}/\pi} = \frac{1}{2\sqrt{1-y^2}}.$$

This function is not equal to $f_X(x)$ on the set $x^2 + y^2 \leq 1$. Since the function $f_{X|Y}(x|y)$ is not equal to the function $f_X(x)$ on the set $x^2 + y^2 \leq 1$, we conclude that X and Y are not independent.

3. Question 3

Let X and Y be independent random variables. Suppose X is uniformly distributed in $[0,1]$. Suppose Y is an exponential random variable with parameter 1. That is, Y has density

$$f_Y(y) = \begin{cases} 0 & \text{, if } y < 0 \\ e^{-y} & \text{, if } y \geq 0. \end{cases}$$

Let $Z = \max\big(X(1-X)\,,\, Y\big)$ be the maximum of $X(1-X)$ and Y.
Find f_Z, the density function of Z.
Simplify your answer to the best of your ability.

Solution. Using the quadratic formula, the function $f(t) = t(1-t)$ takes the value $c \in [0, 1/4]$ when $x = (1/2) \pm (1/2)\sqrt{1-4c}$. So, if $x \in [0,1]$, we have

$$\begin{aligned}\mathbf{P}(X(1-X) \leq x) &= \mathbf{P}(X \in [0, 1/2 - (1/2)\sqrt{1-4x}] \text{ or } X \in [1/2 + (1/2)\sqrt{1-4x}, 1]) \\ &= (1/2) - (1/2)\sqrt{1-4x} + 1 - (1/2 + (1/2)\sqrt{1-4x}) = 1 - \sqrt{1-4x}.\end{aligned}$$

Now, using the definition of the maximum, and then using independence of X and Y,

$$\mathbf{P}(Z \leq t) = \mathbf{P}(\max(X(1-X), Y) \leq t) = \mathbf{P}(X(1-X) \leq t,\, Y \leq t) = \mathbf{P}(X(1-X) \leq t)\mathbf{P}(Y \leq t),$$

for all $t \in \mathbf{R}$. Since $X \in [0,1]$, $X(1-X) \in [0,1]$. Also $Y \geq 0$. So $\mathbf{P}(Z \leq t) = 0$ for all $t < 0$. When $0 \leq t \leq 1/4$, we have

$$\mathbf{P}(Z \leq t) = (1 - \sqrt{1-4t})(1 - e^{-t}), \qquad \forall\, 0 \leq t \leq 1/4$$

(Since $\max_{x \in [0,1]} x(1-x) = 1/4$ (with the maximum occurring at $x = 1/2$), we have $X(1-X) \leq 1/4$.) When $t > 1/4$, $\mathbf{P}(X(1-X) \leq t) = 1$, so

$$\mathbf{P}(Z \leq t) = (1 - e^{-t}), \qquad \forall\, t > 1/4.$$

So, differentiating in each case to get the density,

$$f_Z(t) = \begin{cases} \frac{d}{dt}(0) & \text{, if } t < 0, \\ \frac{d}{dt}[(1-\sqrt{1-4t})(1-e^{-t})] & \text{, if } 0 \le t \le 1/4, \\ \frac{d}{dt}(1-e^{-t}) & \text{, if } t > 1/4. \end{cases}$$

$$= \begin{cases} 0 & \text{, if } t < 0, \\ \frac{1}{2}(1-4t)^{-1/2}(1-e^{-t}) + (1-\sqrt{1-4t})e^{-t} & \text{, if } 0 \le t \le 1/4, \\ e^{-t} & \text{, if } t > 1/4. \end{cases}$$

4. Question 4

- Find a random variable X such that
$$\mathbf{P}(|X| \ge 3) = \frac{\mathbf{E}\,|X|}{3}.$$
Prove that X satisfies this property. (Hint: can X take only one value?)
- Find a random variable Y such that
$$\mathbf{P}(|Y - \mathbf{E}Y| \ge 2) = \frac{\mathrm{var}(Y)}{4}.$$
Prove that Y satisfies this property.

Solution. Suppose $X = 3$ with probability one, i.e. $\mathbf{P}(X = 3) = 1$. Then $\mathbf{E}X = \mathbf{E}(3) = 3$ and $\mathbf{P}(|X| \ge 3) = \mathbf{P}(X = 3) = 1$. So, both sides of the desired equality are one.

The same example works for the second part. Let $Y = X = 3$. Then $\mathbf{E}Y = 3$, $\mathbf{E}Y^2 = \mathbf{E}3^2 = 9$, so $\mathrm{var}(Y) = \mathbf{E}Y^2 - (\mathbf{E}Y)^2 = 9 - 9 = 0$. Meanwhile, since $\mathbf{P}(Y = 3) = 1$, $\mathbf{P}(|Y - \mathbf{E}Y| \ge 2) = \mathbf{P}(|Y - 3| \ge 2) = \mathbf{P}(Y \ge 5 \text{ or } Y \le 1) = 0$. So, both sides of the desired equality are zero.

For another example for the second part, suppose Y is uniformly distributed in $\{-2, 2\}$. Then $\mathbf{E}Y = 0$ and $\mathrm{var}(Y) = 4(1/2) + 4(1/2) = 4$. Also, $\mathbf{P}(|Y - \mathbf{E}Y| \ge 2) = \mathbf{P}(|Y| = 2) = 1$, so both sides of the desired equality are one.

5. Question 5

Let X and Y be random variables. Let t be a constant. Suppose these random variables have joint density function

$$f_{X,Y}(x,y) = \begin{cases} tx^2y^2 & \text{, if } -1 \le x \le 1 \text{ and } 0 \le y \le 1, \\ 0 & \text{, otherwise.} \end{cases}$$

- Find the constant t such that $f_{X,Y}$ is a joint probability density function.
- Let f_Y be the marginal density of Y. Show that its third derivative satisfies
$$\frac{d^3}{dy^3} f_Y(y) = 0, \qquad \forall\, y \in (0,1).$$
- Write a formula that computes $\mathbf{P}(X > Y)$ using integrals. You do **NOT** have to simplify this formula. Your final answer **must** be an integral of the following form:
$$\int_{x=(\cdots)}^{x=(\cdots)} \int_{y=(\cdots)}^{y=(\cdots)} (\text{some function})\,dy\,dx.$$

Solution. We must have $\iint_{\mathbf{R}^2} f_{X,Y}(x,y)dxdy = 1$. We have

$$1 = \iint_{\mathbf{R}^2} f_{X,Y}(x,y)dxdy = t\int_{x=-1}^{x=1}\int_{y=0}^{y=1} x^2y^2dydx = t\int_{x=-1}^{x=1}[x^2y^3/3]_{y=0}^{y=1}dx$$
$$= t\int_{x=-1}^{x=1}[x^2/3]dx = t[x^3/9]_{x=-1}^{x=1} = t(2/9).$$

Solving for t, we get $t = 9/2$.

By definition of marginal, we have, for any $y \in (0,1)$,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \int_{x=-1}^{x=1}\frac{9}{2}x^2y^2dx = y^2\int_{x=-1}^{x=1}\frac{9}{2}x^2dx$$

This function is quadratic in y, so its third derivative is zero.

By definition of joint density, we have

$$\mathbf{P}(X > Y) = \iint_{\{(x,y)\in\mathbf{R}^2\colon x>y\}} f_{X,Y}(x,y)dxdy = \int_{x=0}^{x=1}\int_{y=0}^{y=x}\frac{9}{2}x^2y^2dydx.$$

6. Question 6

Let X and Y be independent random variables. Suppose X has characteristic function (Fourier Transform)

$$\phi_X(t) = e^{-t^2}, \qquad \forall\, t \in \mathbf{R}.$$

(Recall that $\phi_X(t) = \mathbf{E}e^{itX}$ where $i = \sqrt{-1}$, for any $t \in \mathbf{R}$.) Suppose Y has moment generating function

$$M_Y(t) = 1 + t^4, \qquad \forall\, t \in \mathbf{R}.$$

(Recall that $M_Y(t) = \mathbf{E}e^{tY}$ for any $t \in \mathbf{R}$.)

Compute $\mathbf{E}\Big[(X+Y)^2\Big]$.

Solution 1. Note that $M_X(t) = \phi_X(-it) = e^{t^2}$. Since X, Y are independent, we have $M_{X+Y}(t) = M_X(t)M_Y(t) = e^{t^2}(1+t^4)$ for all $t \in \mathbf{R}$. Also, recall from the notes that

$$\frac{d^2}{dt^2}|_{t=0}M_{X+Y}(t) = \mathbf{E}\frac{d^2}{dt^2}|_{t=0}e^{t(X+Y)} = \mathbf{E}(X+Y)^2.$$

So,

$$\mathbf{E}(X+Y)^2 = \frac{d^2}{dt^2}|_{t=0}M_{X+Y}(t) = \frac{d}{dt}|_{t=0}((4t^3+2t)e^{t^2})$$
$$= [(4t^3+2t)(2t) + (12t^2+2)]|_{t=0} = 2.$$

Solution 2. As mentioned above, we can differentiate each MGF separately:

$$\frac{d}{dt}|_{t=0}M_X(t) = \mathbf{E}\frac{d}{dt}|_{t=0}e^{tX} = \mathbf{E}X.$$
$$\mathbf{E}X = \frac{d}{dt}|_{t=0}M_X(t) = \frac{d}{dt}|_{t=0}e^{t^2} = [2te^{t^2}]_{t=0} = 0.$$
$$\mathbf{E}X^2 = \frac{d^2}{dt^2}|_{t=0}M_X(t) = \frac{d^2}{dt^2}|_{t=0}e^{t^2} = \frac{d}{dt}|_{t=0}2te^{t^2} = 2.$$

$$\mathbf{E}Y = \frac{d}{dt}|_{t=0} M_Y(t) = \frac{d}{dt}|_{t=0}(1+t^4) = [4t^3]_{t=0} = 0.$$

$$\mathbf{E}Y^2 = \frac{d^2}{dt^2}|_{t=0} M_Y(t) = \frac{d^2}{dt^2}|_{t=0}(1+t^4) = 0.$$

Therefore, using also that X, Y are independent,

$$\mathbf{E}(X+Y)^2 = \mathbf{E}X^2 + \mathbf{E}Y^2 + 2\mathbf{E}(XY) = 2 + 0 + (\mathbf{E}X)(\mathbf{E}Y) = 2 + 0 \cdot 0 = 2.$$

7. Question 7

Consider a population of 30,000 people, where half of them are given a vaccine for a disease. Suppose all 30,000 people are exposed to a virus causing the disease. We observe that 90 of the unvaccinated people catch the disease, while 5 of the vaccinated people catch the disease.

Consider the following statement:

"The number of infections of vaccinated people, divided by the number of infections of unvaccinated people, is less than 15/100."

Is the statement true with greater than 90% certainty? Justify your answer.

(Assume that each person's ability to catch the disease is independent of each other person's ability to catch the disease.)

(Hint: the estimated probability of a vaccinated person getting the disease is $5/15,000$, and the estimated probability of an unvaccinated person getting the disease is $90/15,000$.)

(Hint: use the Central Limit Theorem. If Z is a standard Gaussian, then $\mathbf{P}(|Z| \leq 2) \approx .9545$. Also, $\sqrt{5} \approx 2.23$, $\sqrt{90} \approx 9.5$.)

Solution. Let X_i be the indicator random variable which is 1 if the i^{th} vaccinated person catches the disease and 0 if not, for all $i \in \{1, 2, \ldots, 15,000\}$. Let Y_i be the indicator random variable which is 1 if the i^{th} unvaccinated person catches the disease and 0 if not, for all $i \in \{1, 2, \ldots, 15,000\}$. Then we are assuming the $X_1, X_2, \ldots, Y_1, Y_2, \ldots$ are i.i.d. with $\mathbf{P}(X_1 = 1) = p = 5/15,000$ and thus $\mathbf{E}[X_1] = p$ $\text{var}(X_1) = p(1-p)$. Also, $\mathbf{P}(Y_1 = 1) = q = 90/15,000$ and thus $\mathbf{E}[Y_1] = q$ $\text{var}(Y_1) = q(1-q)$. The statement can be written as

$$\frac{X_1 + \cdots + X_{15000}}{Y_1 + \cdots + Y_{15000}} < .15. \qquad (*)$$

Then by the central limit theorem, we have

$$\mathbf{P}\left(-2 \leq \frac{X_1 + \cdots + X_{15,000} - 15,000p}{\sqrt{15,000p(1-p)}} \leq 2\right) \approx .9545$$

Since $15000p = 5$ and $150000p(1-p) = 5(1-p) \approx 5$, we have

$$\mathbf{P}\left(5 - 2\sqrt{5} \leq X_1 + \cdots + X_{15,000} \leq 5 + 2\sqrt{5}\right) \approx .9545$$

That is,

$$\mathbf{P}\left(0 \leq X_1 + \cdots + X_{15,000} \leq 10\right) \approx .9545$$

Meanwhile,

$$\mathbf{P}\left(-2 \leq \frac{Y_1 + \cdots + Y_{15,000} - 15,000q}{\sqrt{15,000q(1-q)}} \leq 2\right) \approx .9545$$

Since $15000q = 90$ and $150000q(1-q) = 90(1-q) \approx 90$, we have

$$\mathbf{P}\left(90 - 2\sqrt{90} \leq Y_1 + \cdots + Y_{15,000} \leq 90 + 2\sqrt{90}\right) \approx .9545$$

That is,

$$\mathbf{P}\left(71 \leq Y_1 + \cdots + Y_{15,000} \leq 109\right) \approx .9545$$

So, with at least 95.45% certainty, the number of vaccinated people with the disease is at most 10. Also, with at least 95.45% certainty, the number of unvaccinated people with the disease is at least 71. So, with at least 90% certainty, we have

$$\frac{X_1 + \cdots + X_{15000}}{Y_1 + \cdots + Y_{15000}} \leq \frac{10}{71} < .15.$$

So, the statement is true with at least 90% certainty.

8. Question 8

Suppose you are flipping a fair coin, so that each flip of the coin has probability 1/2 of landing heads, and probability 1/2 of landing tails. What is the expected number of coin flips that you have to make until you see two consecutive heads appear? (That is, you keep flipping the coin until you see two heads in a row, at which point you stop flipping the coin any more, and you count the total number of coin flips you have made.) (Hint: condition on the first two coin flips.)

(Simplify your final answer to the best of your ability.)

Solution. Let T be the number of coin flips that occur until two successive heads occur. Let $X_1 = 1$ if the first flip is heads and $X_1 = 0$ otherwise. Let $X_2 = 1$ if the second flip is heads and $X_2 = 0$ otherwise. From the Total Expectation Theorem,

$$\begin{aligned}\mathbf{E}T &= \mathbf{E}(T|X_1 = 0)\mathbf{P}(X_1 = 0) + \mathbf{E}(T|X_1 = 1, X_2 = 0)\mathbf{P}(X_1 = 1, X_2 = 0)\\ &\quad + \mathbf{E}(T|X_1 = 1, X_2 = 1)\mathbf{P}(X_1 = 1, X_2 = 1)\\ &= \frac{1}{2}\mathbf{E}(T|X_1 = 0) + \frac{1}{4}\mathbf{E}(T|X_1 = 1, X_2 = 0) + \frac{1}{4}\mathbf{E}(T|X_1 = 1, X_2 = 1).\end{aligned}$$

(Note that $\mathbf{P}(X_1 = 0) = 1/2$ since we are flipping a fair coin. Similarly, $\mathbf{P}(X_1 = 1, X_2 = 0) = (1/2)^2$ since this corresponds to flipping a fair coin twice.) If we condition on $X_1 = 0$, then $\mathbf{E}(T|X_1 = 0) = 1 + \mathbf{E}T$, since flipping one tail at the start results in "resetting" the number of flips it takes to observe two heads. That is, one tail at the start is like starting over again from the beginning, with one additional flip already made. By similar reasoning, $\mathbf{E}(T|X_1 = 1, X_2 = 0) = 2 + \mathbf{E}T$. Also, $\mathbf{E}(T|X_1 = 1, X_2 = 1) = 2$, since both heads occurred during the first two coin flips in this case. In summary,

$$\mathbf{E}T = \frac{1}{2}(1 + \mathbf{E}T) + \frac{1}{4}(2 + \mathbf{E}T) + \frac{1}{4}(2).$$

Rearranging, we get

$$\frac{1}{4}\mathbf{E}T = \frac{3}{2}.$$

That is, $\mathbf{E}T = 6$.