Theory of Probability HW #7

Ozaner Hansha

November 11, 2019

Problem 1

Problem: Is the following function a joint pdf?

$$f(x,y) = \begin{cases} \frac{1}{2}, & 0 < y < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Solution: If $f(x,y)$ is a joint pdf, then the following definite integral must evaluate to 1:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\, dy\, dx = \int_0^1 \int_0^x \frac{1}{2}\, dy\, dx$$
$$= \int_0^1 \left[\frac{y}{2}\right]_0^x dx$$
$$= \int_0^1 \frac{x}{2}\, dx$$
$$= \left[\frac{x^2}{4}\right]_0^1 = \frac{1}{4} \neq 1$$

And so $f(x,y)$ is *not* a joint pdf.

Problem 2

Problem: Sonia and Natasha are supposed to meet at a certain location around 5:30 pm. Sonia arrives at some time uniformly distributed between 5:00 pm and 6:00 pm, while Natasha arrives at some time uniformly distributed between 5:15 pm and 6:00 pm. Given that Natasha arrives first, what is the probability that she will not have to wait for more than 10 minutes for Sonia?

Solution: Let S be the time in minutes after 5:00 pm that Sonia arrives, and N be the time that Natasha arrives. We then have the following *independent*, as neither one's arrival time affects the other, random variables:

$$S \sim \mathcal{U}(0,60) \qquad N \sim \mathcal{U}(15,60)$$

Our desired probability is thus given by:

$$P(N - S \le 10 \,|\, N \le S) = \frac{P(N - S \le 10, N \le S)}{P(N \le S)} \qquad \text{(def. of conditional prob.)}$$

$$= \frac{\iint\limits_{\substack{s-n\le 10 \\ n\le s}} f(n,s)\,ds\,dn}{\iint\limits_{n\le s} f(n,s)\,dn\,ds} \qquad \text{(joint pdf)}$$

$$= \frac{\iint\limits_{\substack{s-n\le 10 \\ n\le s}} f_N(n)f_S(s)\,ds\,dn}{\iint\limits_{n\le s} f_N(n)f_S(s)\,dn\,ds} \qquad \text{(independence)}$$

$$= \frac{\frac{1}{45}\cdot\frac{1}{60}\iint\limits_{\substack{s-n\le 10 \\ n\le s}} ds\,dn}{\frac{1}{45}\cdot\frac{1}{60}\iint\limits_{n\le s} dn\,ds} \qquad \text{(uniform pdf)}$$

$$= \frac{\iint\limits_{\substack{s-n\le 10 \\ n\le s}} ds\,dn}{\iint\limits_{n\le s} dn\,ds}$$

Now all that remains is to find the areas of integration of both the numerator and denominator. To illustrate them, we present a graph of the relevant conditions in the sn plane:
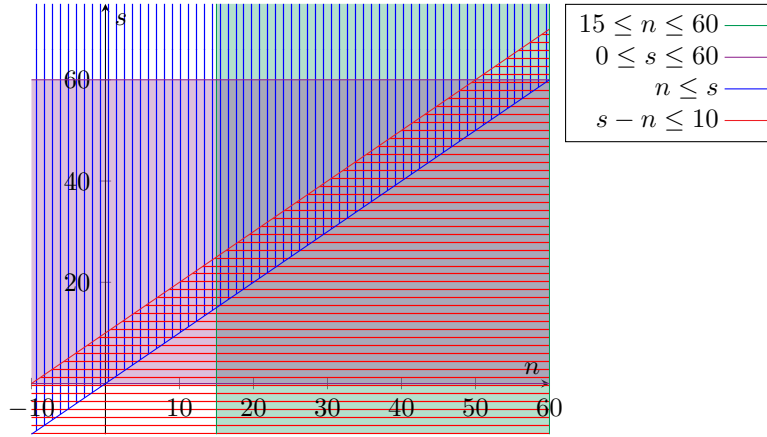
The area of the denominator is given by the blue right triangle that satisfies the first three listed conditions:

$$\iint\limits_{n\le s} dn\,ds = \frac{(60-15)^2}{2} = \frac{2025}{2}$$

The area of the numerator is the trapezoidal shape contained in aforementioned right triangle that is further bounded by the last condition. The legs of this isosceles trapezoid, i.e. the green and purple sides, are simply $c = 10$. The top base, i.e. the red side, is given by:

$$a = \|(50,60) - (15,25)\| = \|(35,35)\| = 35\sqrt{2}$$

Similarly the bottom base, i.e. the blue side, is given by:

$$b = \|(60,60) - (15,15)\| = \|(45,45)\| = 45\sqrt{2}$$

The semiperimeter s of the trapezoid is given by:

$$s = \frac{a + b + 2c}{2} = \frac{35\sqrt{2} + 45\sqrt{2} + 20}{2}$$

Plugging these into Brahmagupta's formula for an isosceles trapezoid, we find the area to be:

$$\iint\limits_{\substack{s-n \leq 10 \\ n \leq s}} ds\,dn = \sqrt{(s-a)(s-b)(s-c)^2} = 10\sqrt{30\sqrt{7} + 79}$$

And so our final answer is given by:

$$P(N - S \leq 10 \,|\, N \leq S) = \frac{\iint\limits_{\substack{s-n \leq 10 \\ n \leq s}} ds\,dn}{\iint\limits_{n \leq s} dn\,ds}$$

$$= \frac{2}{2025}\left(10\sqrt{30\sqrt{7} + 79}\right)$$

$$= \boxed{\frac{4}{405}\sqrt{30\sqrt{7} + 79} \approx 0.124292}$$

Problem 3

Problem: At a fundraiser, the individual dontations of 200 people are independent random variables, each of which is uniformly distributed over \$0 to \$200. Give an expression for the probability that exactly 5 people donate at most \$10 each.

Solution: Let X be the amount any particular individual donates. It is given by:

$$X \sim \mathcal{U}(0, 200)$$

The probability that any particular individual donates at most \$10 is given by:

$$P(X \leq 10) = F_X(10) = \frac{10}{200} = \frac{1}{20}$$

Given 200 such individuals, each independently donating money with the above distribution, the number of donors Y who donate at most \$10 is given by the following binomial distribution:

$$Y \sim B\left(200, \frac{1}{20}\right)$$

And so our desired probability is given by:

$$P(Y = 5) = p_Y(5) \hspace{3cm} \text{(def. of pmf)}$$

$$= \boxed{\binom{200}{5}\left(\frac{1}{20}\right)^5\left(\frac{19}{20}\right)^{195} \approx 0.035896} \hspace{1cm} \text{(binomial distribution)}$$

If desired, a more numerically stable approximation of the above probability could be found via the normal approximation.

Problem 4

Problem: It is observed that the scores of a typical student in Math 477 on Exam I are normally distributed with mean 50 and standard deviation 5, while their scores on Exam II are independently and normally distributed with mean 55 and standard deviation 12. Suppose you are a typical student. What is the probability that the total of your scores on Exam I and II exceeds 150?

Solution: Let the random variables E_1 and E_2 be the score of a typical student on exam 1 and 2 respectively. They are given by:

$$E_1 \sim \mathcal{N}(50, \underbrace{25}_{5^2}) \qquad E_2 \sim \mathcal{N}(55, \underbrace{144}_{12^2})$$

Let the total score $T = E_1 + E_2$. Recall that the sum of two normally distributed random variables is itself, normally distributed with $\mu = \mu_1 + \mu_2$ and $\sigma^2 = \sigma_1^2 + \sigma_2^2$. We thus have:

$$T \sim \mathcal{N}(\underbrace{105}_{50+55}, \underbrace{169}_{5^2+12^2})$$

And so, our desired probability is given by:

$$
\begin{aligned}
P(T > 150) &= 1 - P(T \le 150) &&\text{(complement)} \\
&= 1 - F_T(150) &&\text{(def. of cdf)} \\
&= 1 - \Phi\left(\frac{150 - 105}{\sqrt{169}}\right) &&\text{(unit normal distribution)} \\
&= \boxed{1 - \Phi\left(\frac{45}{13}\right) \approx 0.000269}
\end{aligned}
$$

Problem 5

Problem: A point (X, Y) in the Cartesian plane is uniformly distributed within the unit circle when X and Y have the following joint density:

$$f(x, y) = \begin{cases} \frac{1}{\pi}, & x^2 + y^2 \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Find the marginal densities f_X and f_Y and state whether X and Y are independent or not.

Solution: The marginal pdf of X, when $x^2 \le 1$, is given by:

$$
\begin{aligned}
f_X(x) &= \int_{x^2+y^2\le 1} f(x, y)\, dy &&\text{(def. of marginal pdf)} \\
&= \frac{1}{\pi} \int_{x^2+y^2\le 1} dy &&\text{(uniform distribution)} \\
&= \frac{1}{\pi} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dy \\
&= \frac{2\sqrt{1-x^2}}{\pi}
\end{aligned}
$$

By symmetry, we know that the marginal distribution of Y is the same. And so, the complete marginal pdfs of both X and Y are given by:

$$f_X(a) = f_Y(a) = \begin{cases} \frac{2\sqrt{1-a^2}}{\pi}, & a^2 \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Now we show that the two random variables are *not* independent. Recall that two random variables X and Y are only independent if the product of their pdfs is equal to their joint pdf for all inputs:

$$f(x, y) = f_X(x)f_Y(y) \qquad (x, y) \in \mathbb{R}^2$$

However, we can prove that this is not the case for our unit circle distribution via a simple counterexample. Consider the point $(0.5, 0.5)$:

$$f_X(0.5)f_Y(0.5) = \left(\frac{2\sqrt{1-(0.5)^2}}{\pi} \right)$$

$$= \left(\frac{2\sqrt{0.75}}{\pi} \right) \neq \frac{1}{\pi} = f(0.5, 0.5)$$

And so, X and Y are *not* independent.

Problem 6

Problem: There were 8 questions on Exam I. Out of these, each student was allowed to choose 2 questions for resubmission. Suppose students made their choices independently and randomly (i.e., each student chose 2 out of the 8 questions completely at random). What is the expected number of problems not chosen by any of the 70 students in the classroom?

Solution: Let the random variable X_{ij}, with $i < j$, denote the number of times both questions i and j were chosen by a student. Since each student chooses independently we have:

$$X_{ij} \sim B\left(70, \frac{1}{\binom{8}{2}} \right)$$

Now let the random variable X_i denote the number of times the question i was chosen by a student. It is defined as:

$$X_i = \sum_{\substack{j \in [1..8] \\ i<j}} X_{ij}$$

Now note that $X_{ij} > 0$ since a pair of questions can't be chosen less than 0 times. As a result, the only way for $X_i = 0$ is if each $X_{ij} = 0$ for all $i < j$. And since each X_{ij} is i.i.d., we have the following probability:

$$P(X_i = 0) = \prod_{\substack{j \in [1..8] \\ i < j}} P(X_{ij} = 0) \qquad \text{(independent)}$$

$$= P(X_{ij} = 0)^{8-i} \qquad \text{(identical)}$$

$$= \left(\binom{70}{0} \left(\frac{1}{\binom{8}{2}} \right)^0 \left(1 - \frac{1}{\binom{8}{2}} \right)^{70} \right)^{8-i} \qquad \text{(binomial distribution)}$$

$$= \left(1 - \frac{1}{\binom{8}{2}} \right)^{560 - 70i}$$

Now, letting Y be the number of questions that haven't been chosen by anyone, we have the following expectation:

$$Y = \sum_{i=1}^{8} \mathbf{1}_{X_i = 0}$$

$$E[Y] = \sum_{i=1}^{8} E[\mathbf{1}_{X_i = 0}] \qquad \text{(linearity of expectation)}$$

$$= \sum_{i=1}^{8} P(X_i = 0) \qquad \text{(expectation of indicator)}$$

$$= \boxed{\sum_{i=1}^{8} \left(1 - \frac{1}{\binom{8}{2}} \right)^{560 - 70i} \approx 1.08509}$$