Problem Set 4

Warmups

Problem 1

You roll 6 dice. How much more likely is a roll with:
$[one\ 1, one\ 2, one\ 3, one\ 4, one\ 5, one\ 6]$ than a roll with six 6s?

Solution

~0.015. We can use the multinomial random variable to calculate it:

$$P(X_1 = 1,\ X_2 = 1,\ X_3 = 1,\ X_4 = 1,\ X_5 = 1,\ X_6 = 1) \\ \implies \binom{6}{1,1,1,1,1,1}(1/6)^6 = 0.015432098765432098$$

Problem 2

The joint probability density function of continuous random variables X and Y is given by: $f(X = x, Y = y) = \frac{4y}{x}$ where 0 < y < x < 1
a. What is the marginal density function of X?
b. What is the marginal density function of Y?
c. What is $E[X]$?

Solution

(a) $f_X(x) = 2x$. Can be found by integrating over every value of y:

$$\int_0^x \frac{4y}{x} dy = \frac{2y^2}{x} \ |_0^x = 2x$$

(b) $f_Y(y) = -4y \ln y$. Similarly to part a we integrate over every value of x:

$$\int_y^1 \frac{4y}{x} dx = 4y \ln |x| \ |_y^1 = -4y \ln y$$

(c) $E[X] = \frac{2}{3}$. We need to multiply the marginal pdf of X with x and then integrate over every value of x:

$$\int_0^1 2x^2 dx = \frac{2x^3}{3} \ |_0^1 = \frac{2}{3}$$

Problem 3

Let X_i be the number of weekly visitors to a web site in week i, where $X_i \sim N$(2200, 52900) for all i. Assume that all X_i are independent of each other. What is the probability that the weekly number of visitors exceeds 2000 in at least 2 of the next 3 weeks?

Solution

0.904. Since we know the mean and the variance of the weekly visitor amount we can find the probability of getting more than 2000 visitors using the cdf of the standard normal:

$$P(X > 2000) = 1 - P(X \leq 2000) = 1 - \Phi(\frac{2000 - 2200}{230}) = 0.8077309735016883$$

And with the probability of exceeding 2000 visitors we can find the probability of exceeding 2000 visitors in at least 2 of the next 3 weeks using the binomial random variable:

$$P(Y \geq 2) = 1 - P(Y \leq 2) = \sum_{i=0}^{1} \binom{3}{i} (0.808)^i (0.192)^{3-i} = 0.903563776$$

Problem 4

You think your baby might be tired, and you estimate this prior belief to be $P(Tired) = \frac{3}{4}$. If a baby is tired, the time in minutes until they rub their eyes is distributed as $Exp(\lambda = 3)$. If a baby is not tired, the time in minutes until they rub their eyes is distributed as $Exp(\lambda = 1)$. A baby rubs their eyes after 2 mins. What is your updated belief that they are tired?

Solution

~0.135. We are trying to find out the probability of the baby being tired given the baby rubbed its eyes after 2 minutes. We can find this using the Bayes theorem:

$$P(X = 1 | Y_{\text{rubbed eye}} = 2) = \frac{P(Y = 2 | X = 1) P(X = 1)}{P(Y = 2)}$$

For continuous random variables instead of the probability we can use the pdf of the random variable:

$$\frac{f(Y=2|X=1)P(X=1)}{f(Y=2)} = \frac{f(Y=2|X=1)P(X=1)}{f(Y=2|X=1)P(X=1)+f(Y=2|X=0)P(X=0)}$$

Now we can substitute the probabilities:

$$\implies \frac{(0.007)(0.75)}{(0.007)(0.75)+(0.135)(0.25)} = 0.13461538461538464$$

Problem 5

You are developing medicine that sometimes has a desired effect, and sometimes does not. With FDA approval, you are allowed to test your medicine on 9 patients. You observe that 7 have the desired outcome. Your belief as to the probability of the medicine having an effect before running any experiments was Beta(2, 2).
a. What is the distribution for your belief of the probability of the medicine being effective after the trial?
b. Use your distribution from (a) to calculate your confidence that the probability of the drug having effect is greater than 0.5. You may use scipy.stats or an online calculator.

Solution

Music Tastes

Problem 6

Write a program that reads the data file music.csv and estimates the answers to the following questions. Each row in the csv represents one person and their corresponding ratings of different music types, on a scale of 1-5. For each question write the mathematical formula you used to compute the answer, and include the numeric estimate (see Name2Age for an example). Let R_i be a random variable for the rating a user gives to genre i. You may either use a Frequentest or a Bayesian approach to estimating probabilities from data:
a. What is $P(R_{Folk}=5)$
b. What is $P(R_{Folk}=x)$
c. What is $E[R_{Musical}]$
d. What is $P(R_{Folk}=5|R_{Musical}=5)$
e. What is $P(R_{Folk}=x|R_{Musical}=5)$
f. What is the covariance of R_{Opera} and R_{Punk}?

Formulas

(a)

$$P(R_{Folk} = 5) = \frac{count(R_{Folk} = 5)}{n}$$

(b)

$$P(R_{Folk} = x) = \begin{cases} \frac{count(R_{Folk}=x)}{n} & \text{for } 1 \leq x \leq 5 \\ 0 & \text{otherwise} \end{cases}$$

(c)

$$E[R_{Musical}] = \sum_{i=1}^{5} x f(x)$$

(d)

$$P(R_{Folk} = 5 | R_{Musical} = 5) = \frac{count(R_{Folk} = 5, R_{Musical} = 5)}{count(R_{Musical} = 5)}$$

(e)

$$P(R_{Folk} = x) = \begin{cases} \frac{count(R_{Folk}=x, R_{Musical}=5)}{count(R_{Musical}=5)} & \text{for } 1 \leq x \leq 5 \\ 0 & \text{otherwise} \end{cases}$$