Stat330: Homework #5

Due on February 19, 2014 at 3:10pm

Mr. Lanker Section A

Josh Davis

Problem 1

Problem 3.15 from Baron.

Let X and Y be the number of hardware failures in two computer labs in a given month. The joint distribution of X and Y is givin in the table below:

$P(x,y)$		$x=0$	$x=1$	$x=2$	$P_Y(y)$
y	0	0.52	0.20	0.04	0.76
	1	0.14	0.02	0.01	0.17
	2	0.06	0.01	0.00	0.07
$P_X(x)$		0.72	0.23	0.05	1.00

Part A

Compute the probability of at least one hardware failure.

Solution

This means that either one component in either lab fails. Thus it would be equal to the complement of the probability that **no** components fail.

This gives us:

$$P_X(X \geq 1) + P_Y(Y \geq 1) = 1 - P(X = 0, Y = 0) = 1 - 0.52 = 0.48$$

Part B

From the given distribution, are X and Y independent? Why or why not?

Solution

To determine if joint probabilities are independent, we just need to see if the probability of $P_{XY}(x, y)$ is equal to the separate probabilties for both $P_X(x)$ and $P_Y(y)$.

Thus using our table, we just have to check if the following equation holds $P(x, y) = P_X(x)P_Y(y)$ for each part in the table. Just by looking at the first cell, $P(0, 0) = 0.52 \neq 0.547 = P_X(0)P_Y(0)$.

Thus our variables, X and Y, aren't independent, they are dependent.

Problem 2

Two random variables X and Y have the joint distribution defined by the table:

$P(x,y)$		$x=0$	$x=1$	$x=2$	$P_Y(y)$
y	0	0.4	0.2	0.1	0.7
	1	0	0.2	0	0.2
	2	0	0	0.1	0.1
$P_X(x)$		0.4	0.4	0.2	1.00

Part A

Find the probability mass function of a new random variable $U = 2X - Y$ by looking at the possible values for u based on the defined joint distribution. Hint: the value of $u = 0$ has the probability 0.4.

Solution

Using the equation and the above table, this gives us the following:

U		$x=0$	$x=1$	$x=2$
y	0	0	2	4
	1	-1	1	3
	2	-2	0	2

u	-2	-1	0	1	2	3	4
Values	$P(0,2)$	$P(0,1)$	$P(0,0)+P(1,2)$	$P(1,1)$	$P(1,0)+P(2,2)$	$P(2,1)$	$P(2,0)$
Sum	0	0	0.4 + 0	0.2	0.2 + 0.1	0	0.1
$P(u)$	0	0	0.4	0.2	0.3	0	0.1

Part B

Find E$[U]$ using $P(U)$ probability mass function.

Solution

Since it is a discrete random variable, our expected value is:

$$\begin{aligned}
\mathrm{E}[U] &= \sum_{u=-2}^{4} uP(u) \\
&= -2(0) + -1(0) + 0(0.4) + 1(0.2) + 2(0.3) + 3(0) + 4(0.1) \\
&= 1.2
\end{aligned}$$

Problem 2 continued on next page...

Part C

The expected values for X and Y are $\mathrm{E}[X] = 0.8$ and $\mathrm{E}[Y] = 0.4$. This time calculate $\mathrm{E}[U]$ using properties of expected values (see pg. 49).

Solution

Using the properties on page 49, we get:

$$\begin{aligned}
\mathrm{E}[U] &= \mathrm{E}[2X - Y] \\
&= \mathrm{E}[2X] - \mathrm{E}[Y] \\
&= 2\mathrm{E}[X] - \mathrm{E}[Y] \\
&= 2(.8) - (.4) \\
&= 1.2
\end{aligned}$$

which agrees with part (b).

Part D

Calculate $P(X = 2 \mid U \geq 2)$.

Solution

We want to find the probability that given $U \geq 2$, that $X = 2$. Using our table, this gives us:

$$\begin{aligned}
P(X = 2 \mid U \geq 2) &= \frac{P(X = 2 \cap U \geq 2)}{P(U \geq 2)} \\
&= \frac{P(2,2) + P(2,1) + P(2,0)}{P(U = 2) + P(U = 3) + P(U = 4)} \\
&= \frac{0.1 + 0.1 + 0.1}{0.3 + 0 + 0.1} \\
&= 0.75
\end{aligned}$$

Problem 3

The time in minutes for a certain system to reboot can be modeled with a continuous random variable T that has the following pdf:

$$f(t) = \begin{cases} C(2-t)^2 & 0 \leq t \leq 2 \\ 0 & \text{any other } t \end{cases}$$

Part A

Compute C.

Solution

Since this is a probability density, we know that the total area of any probability density, is equal to one. Thus we get:

$$\begin{aligned}
F(t) &= \int_0^2 C \cdot (2-t)^2 \mathrm{d}t \\
&= C\left[\int_0^2 t^2 - 4t + 4 \ \mathrm{d}t\right] \\
&= C\left[\frac{1}{3}t^3 - 2t^2 + 4t + c\right|_0^2 \\
&= C\left[\frac{1}{3}(2)^3 - 2(2)^2 + 4(2) + c\right] - \left[\frac{1}{3}(0)^3 - 2(0)^2 + 4(0) + c\right] \\
&= C\left[(\frac{8}{3} - 8 + 8 + c) - (c)\right] \\
&= C\frac{8}{3}
\end{aligned}$$

Since this is equal to 1, we can see that $C = 3/8$.

Part B

Compute $\mathrm{E}[T]$.

Solution

Using $C = 3/8$ from the previous problem, we get:

$$\begin{aligned}
\mathrm{E}[T] &= \frac{3}{8}\int_0^2 t \cdot (2-t)^2 \mathrm{d}t \\
&= \frac{3}{8}\left[\int_0^2 t^3 - 4t^2 + 4t \ \mathrm{d}t\right] \\
&= \frac{3}{8}\left[\frac{1}{4}t^4 - \frac{4}{3}t^3 + 2t^2 \ \right|_0^2 \\
&= \frac{3}{8}\left[\frac{1}{4}t^4 - \frac{4}{3}t^3 + 2t^2 \ \right|_0^2 \\
&= \frac{3}{8}\left[\frac{1}{4}(2)^4 - \frac{4}{3}(2)^3 + 2(2)^2 \ \right|_0^2 \\
&= \frac{3}{8}\left[4 - \frac{32}{3} + 8 \ \right] = \frac{1}{2}
\end{aligned}$$

Problem 3 continued on next page. . .

Part C

Compute Var$[T]$.

Solution

We know that $\text{Var}[T] = \text{E}[(T - \text{E}[T])^2]$ or $\text{Var}[T] = \text{E}[T^2] - (\text{E}[T])^2$ which gives us:

$$\begin{aligned}
\text{Var}[T] &= \frac{3}{8}\int_0^2 t^2 \cdot (2-t)^2 \text{d}t - \left(\frac{1}{2}\right)^2 \\
&= \frac{3}{8}\left[\int_0^2 t^4 - 4t^3 + 4t^2 \ \text{d}t\right] - \frac{1}{4} \\
&= \frac{3}{8}\left[\frac{1}{5}t^5 - t^4 + \frac{4}{3}t^3 \ \Big|_0^2\right. - \frac{1}{4} \\
&= \frac{3}{8}\left[\frac{1}{5}(2)^5 - (2)^4 + \frac{4}{3}(2)^3 \ \Big|\right. - \frac{1}{4} \\
&= \frac{3}{8}\left[\frac{32}{5} - 16 + \frac{32}{3}\right] - \frac{1}{4} \\
&= \frac{3}{8}\left[\frac{16}{15}\right] - \frac{1}{4} \\
&= \frac{3}{20} \\
&= .15
\end{aligned}$$

Part D

Compute the probability that it takes between 1 and 2 minutes to reboot.

Solution

Since probability is area of our pdf, we just need to calculate the integral from 1 to 2. Thus:

$$\begin{aligned}
F(t) &= \frac{3}{8}\int_1^2 (2-t)^2 \text{d}t \\
&= \frac{3}{8}\left[\int_1^2 t^2 - 4t + 4 \ \text{d}t\right] \\
&= \frac{3}{8}\left[\frac{1}{3}t^3 - 2t^2 + 4t + c\Big|_1^2\right. \\
&= \frac{3}{8}\left[\left(\frac{1}{3}(2)^3 - 2(2)^2 + 4(2) + c\right) - \left(\frac{1}{3}(1)^3 - 2(1)^2 + 4(1) + c\right)\right] \\
&= \frac{3}{8}\left[\left(\frac{8}{3} - 8 + 8 + c\right) - \left(\frac{1}{3} - 2 + 4 + c\right)\right] \\
&= \frac{3}{8}\left[\left(\frac{8}{3}\right) - \left(\frac{7}{3}\right)\right] \\
&= \frac{3}{8}\left[\frac{1}{3}\right] \\
&= \frac{1}{8} \\
&= 0.125
\end{aligned}$$

Problem 4

Say that jobs arrive at a print queue in a busy computer lab at a rate of 1 print jobs ever two minutes. The number of jobs, X, arriving in an time interval of size t is distributed $X \sim \text{Poisson}(\lambda t)$.

Part A

Calculate the probability that less than 2 jobs arrive at the print queue in 4 minutes.

Solution

The problem description gives us $\lambda = .5$ and t is the interval size. We want the probability that less than 2 jobs arrive, or $P(X < 2)$.

Our interval time is then $t = 4$, so $X \sim \text{Poisson}(\frac{1}{2} \cdot 4) = X \sim \text{Poisson}(2)$. Our distribution function is then: $P(x) = e^{-\lambda} \frac{\lambda^x}{x!}$.

This is:

$$\begin{aligned} P[X < 2] &= P(X = 0) + P(X = 1) \\ &\approx .1353 + .2707 \\ &\approx .4060 \end{aligned}$$

Part B

If a print job just arrived at the print queue, what is the probability that the queue will wait less than 1.5 minutes for the next print job to arrive?

Solution

Since we want the time between events, if we let T be the random variable for time between jobs, then we know that $T \sim \text{Exponential}(\lambda)$. Since we know $\lambda = 0.5$, we get: $T \sim \text{Exponential}(.5)$. The cdf is: $F(x) = 1 - e^{-\lambda x}$.

We want to find the probability that T is less than 1.5 minutes, thus $P(I < 1.5)$ which is:

$$\begin{aligned} P(I < 1.5) &= 1 - e^{-.5 \cdot 1.5} \\ &= 1 - e^{-.75} \\ &\approx 0.5276 \end{aligned}$$

Problem 5

The memoryless property of the exponential.

Part A

Using your answer to (4b), if a print job as just arrived, calculate the probability that the next print job arrives after 1.5 minutes.

Solution

Now we want the probability of $P(I > 1.5)$, which is equal to:

$$\begin{aligned} P(I > 1.5) &= 1 - P(I < 1.5) \\ &= 1 - 0.5276 \\ &\approx 0.4724 \end{aligned}$$

Part B

Similarly if a print job has just arrived, calculate the probability that the next print job arrives after 3 minutes.

Solution

Now we want the probability of $P(I > 3)$, which is equal to:

$$\begin{aligned} P(I > 3) &= 1 - P(I < 3) \\ &= 1 - (1 - e^{-.5 \cdot 3}) \\ &= e^{-.5 \cdot 3} \\ &\approx 0.2231 \end{aligned}$$

Part C

Calculate the conditional probability that the next print job arrives after 3 minutes given that the job does not arrive in the first 1.5 minutes. To start you on your answer, let I represent the time until the next arrival:

$$\begin{aligned} P(I > 3.0 \mid I > 1.5) &= \frac{P(I > 3.0 \text{ and } I > 1.5)}{P(I > 1.5)} \\ &= \frac{P(I > 3.0)}{P(I > 1.5)} \end{aligned}$$

with the simplification in the numerator due to the fact if $I > 3.0$ then it is automatically greater than 1.5 minutes, so the last part can be ignored.

Solution

Using the derivation and the probability from our previous parts, we get:

$$\frac{P(I > 3.0)}{P(I > 1.5)} = \frac{0.2231}{0.4724} = 0.4724$$

THAT'S AWESOME!

Part D

Why is the exponential probability distribution called memoryless? Compare your answers for (5a) and (5c).

Problem 5 continued on next page...

Solution

The exponential probability distribution is called memoryless because given the conditional that an event happened in the past, the probability of that event doesn't depend on the conditional. Instead it just depends on the time between the two events.

Problem 6

Problem 4.16 from Baron.

For the Standard Normal random variable Z, compute the values of the following.

Solution

(a) $P(Z < 1.25) = 0.8944$

(b) $P(Z \leq 1.25) = 0.8944$

(c) $P(Z > 1.25) = 1 - P(Z \leq 1.25) = 1 - 0.8944 = 0.1056$

(d) $P(|Z| \leq 1.25) = P(Z \leq 1.25) - P(Z \leq -1.25) = 0.8944 - 0.1056 = 0.7888$

(e) $P(Z < 6.0) = 1.0$

(f) $P(Z > 6.0) = 1 - P(Z < 6.0) = 0$

(g) With probability 0.8, variable Z does not exceed what value? $P(Z < c) = 0.8$ where $c = 0.85$.

Problem 7

Problem 4.17 from Baron.

For the Standard Normal random variable Z, compute the values of the following.

Solution

(a) $P(Z \geq 0.99) = 1 - P(Z < 0.99) = 0.1611$

(b) $P(Z \geq -0.99) = 1 - P(Z < -0.99) = 0.8389$

(c) $P(Z < 0.99) = 0.8389$

(d) $P(|Z| > 0.99) = 1 - [P(Z > -0.99) - P(Z > 0.99)] = 1 - 0.6778 = 0.3222$

(e) $P(Z < 10.0) = 1$

(f) $P(Z > 10.0) = 1 - P(Z < 10.0) = 0$

(g) With probability 0.9, variable Z is less than what? $P(Z < c) = 0.9$ where $c = 1.29$.

Problem 8

If X is a continuous random variable that follows a normal probability distribution with $\mu = 100$ points and $\sigma = 15$, then find the probability that X is:

Part A

Less than 106.

Solution

$$\begin{aligned}
P(X < 106) &= P\left(\frac{X-\mu}{\sigma} < \frac{106-\mu}{\sigma}\right) \\
&= P\left(Z < \frac{106-100}{15}\right) \\
&= P(Z < 0.4) \\
&= 0.6554
\end{aligned}$$

Part B

Greater than 88.

Solution

$$\begin{aligned}
P(X > 88) &= P\left(\frac{X-\mu}{\sigma} > \frac{88-\mu}{\sigma}\right) \\
&= P\left(Z > \frac{88-100}{15}\right) \\
&= P(Z > -0.8) \\
&= 1 - P(Z \leq -0.8) \\
&= 0.7881
\end{aligned}$$

Part C

Between 85 and 115.

Solution

$$\begin{aligned}
P(85 < X < 115) &= P\left(\frac{85-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{115-\mu}{\sigma}\right) \\
&= P\left(\frac{85-100}{15} < Z < \frac{115-100}{15}\right) \\
&= P\left(-1 < Z < 1\right) \\
&= P(Z < 1) - P(Z < -1) \\
&= 0.8413 - 0.1587 \\
&= 0.6826
\end{aligned}$$

Problem 8 continued on next page...

Part D

Find the point value that separates the lower half and upper half of values of X, or the median value c where $P(X < c) = 0.50$.

Solution

Using the Z table, $c = 0.0$, un-normalizing it gives: $0 \cdot 15 + 100 = 100$

Part E

Find the point value that separates the lower quarter from the upper three quarters of values of X, or where $P(X < c) = 0.25$.

Solution

Using the Z table, $c = -0.67$, un-normalizing it gives: $-0.67 \cdot 15 + 100 = 89.95$

Part F

Find the point value that separates the lower three quarters from the upper quarter of values of X, or where $P(X < c) = 0.75$.

Solution

Using the Z table, $c = 0.68$, un-normalizing it gives: $0.68 \cdot 15 + 100 = 110.2$

Problem 9

Problem 4.22 from Baron.

Refer to the country in example 4.11 on pg. 91, where household incomes follow Normal distribution with $\mu = 900$ coins and $\sigma = 200$ coins.

Let X be the random variable for household income.

Part A

A recent economic reform made households with the income below 640 coins qualify for a free bottle of milk at every breakfast. What portion of the population qualifies for a free bottle of milk?

Solution

We are looking for $P(X < 640)$. Standardizing gives us:

$$\begin{aligned} P(X < 640) &= P\left(\frac{X-\mu}{\sigma} < \frac{640-\mu}{\sigma}\right) \\ &= P\left(Z < \frac{-260}{200}\right) \\ &= P\left(Z < -1.3\right) \\ &= 0.0968 \end{aligned}$$

Part B

Moreover, households with an income within the lowest 5% of the polulation are entitled to a free sandwich. What income qualifies a household to receive free sandwiches?

Solution

We are looking for $P(Z < c) = .05$. Using the table, we know that $c = -1.6$. Standardizing gives us $-1.6 \cdot 200 + 900 = 580$ as their income.

Problem 10

The average height of professional basketball players is around 6 feet and 7 inches (79 inches), and the standard deviation is 3.89 inches. Assuming Normal distribution of heights within this group:

Let X be the random variable for the height of the basketball players in inches.

Part A

What percent of professional backetball players are taller than 7 feet?

Solution

Solution is

$$\begin{aligned}
P(X > 84) &= 1 - P(X \leq 84) \\
&= 1 - P\left(\frac{X-\mu}{\sigma} < \frac{84-\mu}{\sigma}\right) \\
&= 1 - P\left(Z < \frac{5}{3.89}\right) \\
&= 1 - P\left(Z < 1.29\right) \\
&= 1 - 0.9015 \\
&= 0.0985
\end{aligned}$$

Part B

If your favorite player is within the tallest 20% of all players, what can his height be?

Solution

We are looking for when $P(Z > c) = 0.20$ on the standard normal table. This is $1 - P(Z < c) = 0.20$ or $P(Z < c) = 0.80$. Using the table, $c = 0.85$. Standardizing this gives us $0.85 \cdot 3.89 + 79 = 82.3$.