Homework 4

Combining Random Variables

Jaiden Atterbury

Instructions

Please answer the following questions in the order in which they are posed. Add a few empty lines below each and write your answers there. **Focus on answering in complete sentences and show work whether we ask for it or not**. You will also need scratch paper/pen to work out the answers before typing it.

For help with formatting documents in RMarkdown, please consult R Markdown: The Definitive Guide. Another option is to search using Google.

Exercises

1. (Floods) Suppose that many years of observation have confirmed that the annual maximum flood tide X (in feet) for a certain river can be modeled by the PDF

$$f(x) = \frac{1}{20} \qquad 20 < x < 40$$

The Army Corps of Engineers is planning to build a levee along a certain portion of the river and they want to make it high enough so there is only a 20% chance that the worst flood in the next thirty years will overflow the embankment. How high should the levee be?

Be sure to define random variables, show your derivations and clearly highlight important intermediate steps so we don't miss them.

Setup:
Let X denote the annual maximum flood tide of a certain river (in feet). Since X can be modeled by the PDF $f(x) = \frac{1}{20}$, $20 < x < 40$, we can clearly see that $X \sim Unif(20, 40)$. Also, let h be the height of the embankment/levee the Army Corps of Engineers is planning to build. Then to find the height h so there is only a 20% chance that the worst flood in the next thirty years will overflow the embankment, we must solve the probability problem that states $P(X_{max} > h) = 0.2$.

Since we want $P(X_{max} > h) = 0.2$ and we are dealing with X_{max} it is much easier to deal with the CDF of X_{max} than it is with the PDF of X_{max}. Thus we can rewrite our probability statement as $1 - P(X_{max} \leq h) = 0.2$. Using the fact that the CDF of X is $F(x) = \frac{x-20}{40-20} = \frac{x-20}{20}$when $x \in (a, b)$ and that the CDF of $X_{max} = F(x)^n$ we can now solve for the value of h that will make the levee high enough so there is only a 20% chance that the worst flood in the next thirty years will overflow the embankment.

Solving for h:

$$\begin{aligned}
0.2 &= 1 - P(X_{max} \leq h) \\
0.2 &= 1 - F(h)^{30} \\
0.2 &= 1 - (\frac{h-20}{20})^{30} \\
0.2 &= 1 - (\frac{1}{20})^{30}(h-20)^{30} \\
0.8 &= (\frac{1}{20})^{30}(h-20)^{30} \\
\frac{0.8}{(\frac{1}{20})^{30}} &= (h-20)^{30} \\
\sqrt[30]{\frac{0.8}{(\frac{1}{20})^{30}}} &= h-20 \\
h &= \sqrt[30]{\frac{0.8}{(\frac{1}{20})^{30}}} + 20 \\
h &\approx 39.8518 \text{ feet}
\end{aligned}$$

Thus, the height h so there is only a 20% chance that the worst flood in the next thirty years will overflow the embankment is $\boxed{h \approx 39.8518 \text{ feet}}$.

2. (Two Poissons) Suppose $X_1, X_2, \dots, X_n \overset{i.i.d.}{\sim} Pois(\lambda)$. That is, the PMF of each X is

$$f(x) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \dots$$

Show, using mathematical induction that

$$S = X_1 + X_2 + \dots + X_n \sim Pois(n\lambda).$$

Hint: Look at Example 16.1 and Theorem 16.2 for the proof for the Binomial distribution.

Proof: By induction:

1.) Base case: $n = 2$

When $n = 2$, $S_2 = X_1 + X_2$ where $f_1(x_1) = \frac{e^{-\lambda}\lambda^{x_1}}{x_1!},\ x_1 = 1, 2, \dots$ and $f_2(x_2) = \frac{e^{-\lambda}\lambda^{x_2}}{x_2!},\ x_2 = 1, 2, \dots$ are the PMFs of X_1 and X_2. In order to see what the distribution of S_2 is we must first find the PMF of S_2. To do this, we will use the convolution summation from theorem 16.1 which gives a formula for finding the PMF of the sum of two independent random variables.

$$
\begin{aligned}
P(S_2 = s_2) &= f(s_2) \\
&= \sum_{x_1} f_1(x_1) f_2(s_2 - x_1),\ s_2 = 0, 1, 2, \ldots \\
&= \sum_{x_1} \frac{e^{-\lambda}\lambda^{x_1}}{x_1!} \cdot \frac{e^{-\lambda}\lambda^{s_2 - x_1}}{(s_2 - x_1)!} \\
&= e^{-2\lambda} \sum_{x_1} \frac{\lambda^{x_1}\lambda^{s_2 - x_1}}{x_1!(s_2 - x_1)!} \\
&= \frac{e^{-2\lambda}}{s_2!} \sum_{x_1} \frac{s_2!}{x_1!(s_2 - x_1)!} \lambda^{x_1}\lambda^{s_2 - x_1} \\
&= \frac{e^{-2\lambda}}{s_2!} \sum_{x_1=0}^{s_2} \frac{s_2!}{x_1!(s_2 - x_1)!} \lambda^{x_1}\lambda^{s_2 - x_1} \quad \text{Since } f(s_2) \text{ is defined as long as } s_2 - x_1 \geq 0 \\
&= \frac{e^{-2\lambda}}{s_2!} (\lambda + \lambda)^{s_2} \quad \text{By the Binomial Theorem} \\
&= \frac{e^{-2\lambda}(2\lambda)^{s_2}}{s_2!}
\end{aligned}
$$

Thus $S_2 \sim Pois(2\lambda)$ and therefore the base case holds.

2.) Inductive step:
Assume that the result holds for some integer k. That is, $S_k = X_1 + X_2 + \cdots + X_k \sim Pois(k\lambda)$. Then to see that the result holds for n = k + 1 notice $S_{k+1} = X_1 + X_2 + \cdots + X_k + X_{k+1} = S_k + X_{k+1}$. Since S_k can be written as the sum of k X_i terms and the X_i's are independent and identically distributed, it follows that S_k and X_{k+1} are independent. Thus $S_k + X_{k+1}$ is the sum of two independent poisson random variables one with rate $k\lambda$ and one with rate λ thus we must show that this sum has a rate parameter of $(k+1)\lambda$. To do this we will follow a similar process to what we did in the base case, we will take $f_1(s_k) = \frac{e^{-k\lambda}(k\lambda)^{s_k}}{s_k!},\ s_k = 0, 1, 2, \ldots$ and $f_2(x_{k+1}) = \frac{e^{-\lambda}\lambda^{x_{k+1}}}{x_{k+1}!},\ x_{k+1} = 0, 1, 2, \ldots$ as the PMFs of S_{k+1} and X_{k+1} then use theorem 16.1 to find the distribution of S_{k+1}.

$$
\begin{aligned}
P(S_{k+1} = s_{k+1}) &= f(s_{k+1}) \\
&= \sum_{s_k} f_1(s_k) f_2(s_{k+1} - s_k),\ s_{k+1} = 0, 1, 2, \ldots \\
&= \sum_{s_k} \frac{e^{-k\lambda}(k\lambda)^{s_k}}{s_k!} \cdot \frac{e^{-\lambda}\lambda^{s_{k+1} - s_k}}{(s_{k+1} - s_k)!} \\
&= e^{-(k+1)\lambda} \sum_{s_k} \frac{(k\lambda)^{s_k}\lambda^{s_{k+1} - s_k}}{s_k!(s_{k+1} - s_k)!} \\
&= \frac{e^{-(k+1)\lambda}}{s_{k+1}!} \sum_{s_k} \frac{s_{k+1}!}{s_k!(s_{k+1} - s_k)!} (k\lambda)^{s_k}\lambda^{s_{k+1} - s_k} \\
&= \frac{e^{-(k+1)\lambda}}{s_{k+1}!} \sum_{s_k=0}^{s_{k+1}} \frac{s_{k+1}!}{s_k!(s_{k+1} - s_k)!} (k\lambda)^{s_k}\lambda^{s_{k+1} - s_k} \quad \text{Since } f(s_{k+1}) \text{ is defined as long as } s_{k+1} - s_k \geq 0 \\
&= \frac{e^{-(k+1)\lambda}}{s_{k+1}!} (k\lambda + \lambda)^{s_{k+1}} \quad \text{By the Binomial Theorem} \\
&= \frac{e^{-(k+1)\lambda}((k+1)\lambda)^{s_{k+1}}}{s_{k+1}!}
\end{aligned}
$$

Thus $S_{k+1} \sim Pois((k+1)\lambda)$. Therefore if $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} Pois(\lambda)$ then $S = X_1 + X_2 + \cdots + X_n \sim Pois(n\lambda)$. $\square$

3. (Three normals) The random variables X_1, X_2 and X_3 are independent $Norm(\mu, 1)$. Let $\bar{X}$ be the (arithmetic) average of the three random variables. That is:

$$\bar{X} = \frac{1}{3}\left(X_1 + X_2 + X_3\right).$$

What is the probability that $Y = (X_1 - \bar{X}) > 1.6$?

Be sure to write the distribution of Y clearly, cite any results you use, then show the calculation including any code you write to do your calculations.

Hint Write Y as a linear combination of the X's. No further hints on this. You are expected to figure it out.

Setup:
Since $\bar{X} = \frac{1}{3}(X_1 + X_2 + X_3)$ with $X_i \sim Norm(\mu, 1)$, it follows that we can rewrite $Y = X_1 - \bar{X}$ as $Y = X_1 - \frac{1}{3}X_1 - \frac{1}{3}X_2 - \frac{1}{3}X_3$ which equals $Y = \frac{2}{3}X_1 + \frac{-1}{3}X_2 + \frac{-1}{3}X_3$. Since Y is a linear combination of three independent and identically distributed normal distributions, we can use theorem 16.5 to see that $Y \sim Norm(\mu_y = \frac{2}{3}\mu + \frac{-1}{3}\mu + \frac{-1}{3}\mu,\ \sigma_y = \sqrt{(\frac{2}{3})^2(1)^2 + (\frac{-1}{3})^2(1)^2 + (\frac{-1}{3})^2(1)^2}$), which simplifies to $Y \sim Norm(\mu_y = 0,\ \sigma_y = \frac{\sqrt{6}}{3})$.

Solving $\mathbf{P(Y > 1.6)}$:
To solve $P(Y > 1.6)$ we can simply use R as a calculator:

```
prob <- pnorm(q = 1.6, mean = 0, sd = sqrt(6) / 3,  lower.tail = FALSE)
```

Thus, the probability that $Y = (X_1 - \bar{X}) > 1.6$ is 0.0250218.

4. (Call center) Let the Poisson random variable X be the number of calls for technical assistance received by a computer company during the firm's nine normal workday hours. Suppose the average number of calls per hour is 7.0 and that each call costs the company \$50. Let Y be a Poisson random variable representing the number of calls for technical assistance received during a day's remaining fifteen hours. Suppose the average number of calls per hour is 4.0 for that time period and that each such call costs the company \$60.

   a. Let the random variable C denote the cost associated with the calls received during a 24 hour day. Write C as a function of X and Y.

No hints on how to do this. You are expected to figure it out.

Let X denote the number of calls for technical assistance received by a computer company during the firm's nine normal workday hours. Since the average number of calls per hour is 7.0, and the firm has 9 normal workday hours, it follows from Lemma 8.1 that $E[X] = Var[X] = \lambda_x = 9 * 7 = 63$. Thus $X \sim Pois(\lambda_x = 63)$. Furthermore, let Y denote the number of calls for technical assistance received during a day's remaining fifteen hours. Since the average number of calls per hour is 4.0 during these 15 non-working hours, using Lemma 8.1 again we can see that $E[Y] = Var[Y] = \lambda_y = 15 * 4 = 60$. Thus $Y \sim Pois(\lambda_y = 60)$. Using our two random variables X and Y and the fact that calls during working hours cost the company \$50 and calls during non-working hours cost the company \$60, we can see that $\boxed{C = 50X + 60Y}$, where C denotes the cost associated with the calls received during a 24 hour day.

   b. Find the expected cost and the variance of the random variable C. Be sure to state any assumptions you make in order to do your calculation. And show your work.

Since the two periods of our random variables, which are the 9 working hours and the 15 non-working hours, are different with no other information on how calls in one period affect the likliehood of calls in the other, it is safe to assume that X and Y are independent, and we will use that fact in calculating the expected value and variance of C.

Finding E[C]:

$$\begin{aligned}
E[C] &= E[50X + 60Y] \\
&= 50E[X] + 60E[Y] \quad \text{Linearity of expectations} \\
&= 50(63) + 60(60) \quad \text{Since } E[X] = 63 \text{ and } E[Y] = 60 \\
&= 3150 + 3600 \\
&= \boxed{\$6750}
\end{aligned}$$

Finding Var[C]:

$$\begin{aligned}
Var[C] &= Var[50X + 60Y] \\
&= 50^2 Var[X] + 60^2 Var[Y] \quad \text{Since } X \text{ and } Y \text{ are assumed to be independent} \\
&= 50^2(63) + 60^2(60) \quad \text{Since } Var[X] = 63 \text{ and } Var[Y] = 60 \\
&= 157500 + 216000 \\
&= \boxed{\$373500}
\end{aligned}$$