APMA 1650

Homework 5 - Solutions

Problem 1. (18 pts) You own a moderately successful small-town grocery store. You have collected enough data to determine that the weekly demand Y for potatoes (measured in hundreds of pounds of potatoes purchased) has a PDF

$$f_Y(y) = \begin{cases} 3y^2 & 0 \leq y \leq 1, \\ 0 & \text{otherwise} \end{cases}$$

Here we are assuming that you can't stock more than 100 pounds of potatoes at once. At the beginning of each week, you purchase up to 100 pounds of potatoes at 75 cents per pound and proceed to sell them at 100 cents per pound. At the end of the week you donate any remaining potatoes to a local food bank.

a. (4 pts) Suppose that you buy a fixed number $a \in [0, 1]$ of hundreds of pounds of potatoes a week. What is your weekly profit in dollars as a function of a and the demand Y? (Keep in mind, if the demand exceeds a, you run out of stock)

b. (8 pts) What are your expected profits in a given week? This will be a function of a. Simplify your answer. (Hint: divide your integral into $0 \leq y \leq a$ and $a \leq y \leq 1$)

c. (6 pts) How many hundreds of pounds of potatoes should you buy every week to maximize your expected profits? Give an exact answer. What are the expected profits in this case?

Solution:

a. If you buy a hundreds of pounds of potatoes that will cost you $75a$ dollars. If the demand in a given week is Y, then you will sell $\min\{a, Y\}$ hundreds of pounds of potatoes that week and make $100 \min\{a, Y\}$ dollars. The profit $g(Y)$ that you make each week is how much you sell - how much you bought

$$g(Y) = 100 \min\{a, Y\} - 75a.$$

b. The expected profits are then

$$\begin{aligned} Eg(Y) &= E[100 \min\{a, Y\} - 75a] \\ &= 100E[\min\{a, Y\}] - 75a \\ &= 300 \int_0^1 \min\{a, y\} y^2 \mathrm{d}y - 75a \end{aligned}$$

Following the hint, we can solve this integral by splitting it up into $\{0 \le y \le a\}$ and $\{a \le y \le 1\}$, where $\min\{a, y\} = y$ and $\min\{a, y\} = a$ respectively. This gives

$$\begin{aligned}
\int_0^1 \min\{a, y\} y^2 \mathrm{d}y &= \int_0^a y^3 \mathrm{d}y + \int_a^1 a y^2 \mathrm{d}y \\
&= \frac{1}{4} y^4 \Big|_0^a + a \frac{1}{3} y^3 \Big|_a^1 \\
&= \frac{1}{4} a^4 + \frac{1}{3} a(1 - a^3) \\
&= -\frac{1}{12} a^4 + \frac{1}{3} a
\end{aligned}$$

Therefore the expected profits are

$$\begin{aligned}
Eg(X) &= 300 \left(-\frac{1}{12} a^4 + \frac{1}{3} a \right) - 75a \\
&= 25a(1 - a^3)
\end{aligned}$$

c. To maximize the profits we need to maximize the function $a \mapsto a(1-a^3)$ over $a \in [0, 1]$. To do this, we take the derivative to identify the critical points by solving

$$\frac{\mathrm{d}}{\mathrm{d}a} a(1 - a^3) = 1 - 4a^3 = 0, \quad \Rightarrow \quad a = \frac{1}{4^{1/3}}.$$

Since there is only one critical point and $a(1-a^3) = 0$ at the end points $a = 0$ or 1 but is positive for $a \in (0, 1)$, we deduce that $a = 1/4^{1/3}$ is indeed the value that maximizes $a(1-a^3)$. Therefore you should buy $1/4^{1/3} = 2^{1/3}/2 \approx .63$ hundred pounds of potatoes each week to maximize your expected profits at

$$25a(1 - a^3) = \frac{75(2^{1/3})}{8} \approx 11.81 \text{ dollars per week.}$$

Hopefully you are selling more than just potatoes...

Problem 2. (26 pts) Let X be a continuous random variable with range $[a, \infty)$, for some $a > 0$ and PDF given by

$$f_X(x) = \begin{cases} cx^{-p-1} & x \ge a \\ 0 & \text{otherwise} \end{cases}$$

for some $p > 0$.

a. (4 pts) Find the value of c that makes this a valid PDF.

b. (4 pts) What is the CDF of X?

c. (5 pts) Find the expectation of X.

d. (5 pts) Find $P(X > 4a|X > 3a)$.

e. (8 pts) What is distribution of the random variable $Y = \ln(X/a)$ (i.e. range and PDF)? Do you recognize this distribution? Interpret you answer to part d in terms of this.

Solution: This distribution is just the Pareto distribution. Commonly used to model income inequality in society. `https://en.wikipedia.org/wiki/Pareto_distribution`

a. We need the PDF to be normalized, that is, we need it to satisfy

$$1 = \int_{-\infty}^{\infty} f_X(x)\mathrm{d}x = c\int_a^{\infty} x^{-p-1}\mathrm{d}x = -\frac{c}{p}x^{-p}\Big|_a^{\infty} = \frac{c}{pa^p}$$

Therefore

$$c = pa^p.$$

b. The CDF for $x \geq a$ is given by

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(y)\mathrm{d}y = pa^p\int_a^x y^{-p-1}\mathrm{d}y \\ &= -a^p(x^{-p} - a^{-p}) \\ &= 1 - \left(\frac{a}{x}\right)^p, \end{aligned}$$

with the obvious extension that $F_X(x) = 0$ for $x \leq a$.

c. The expectation is given by

$$EX = \int_{-\infty}^{\infty} xf_X(x)\mathrm{d}x = pa^p\int_a^{\infty} x^{-p}\mathrm{d}x = \begin{cases} a\ln(x)\Big|_a^{\infty} & p = 1 \\ -\frac{pa^p}{p-1}x^{1-p}\Big|_a^{\infty} & p \neq 1 \end{cases}$$

The above limits are only finite then if $p > 1$, otherwise they are infinity. This gives the final answer

$$EX = \begin{cases} \infty & 0 < p \leq 1 \\ \frac{ap}{p-1} & p \geq 1 \end{cases}.$$

Yes you can have an infinite expectation!

d. To find $P(X > 4a|X > 3a)$, we compute

$$\begin{aligned} P(X > 4a|X > 3a) &= \frac{P(X > 4a, X > 3a)}{P(X > 3a)} \\ &= \frac{P(X > 4a)}{P(X > 3a)} \\ &= \frac{\left(\frac{a}{4a}\right)^p}{\left(\frac{a}{3a}\right)^p} = \left(\frac{3}{4}\right)^p \end{aligned}$$

Note that this is just

$$\left(\frac{3}{4}\right)^p = P\left(X > \frac{4}{3}a\right)$$

e. Let $Y = \ln(X/a)$, then we can see that since $X \geq a$, that the range of Y is just $[0, \infty)$. To calculate the distribution, we note that for $y \geq 0$

$$F_Y(y) = P(Y \leq y) = P(\ln(X/a) \leq y) = P(X \leq ae^y) = 1 - \left(\frac{a}{ae^y}\right)^p = 1 - e^{-py}$$

Taking the derivative gives

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) = \begin{cases} pe^{-py} & y \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

We should recognize this as the PDF of an Exponential(p) random variable. With this in mind, we can see that the result from part d

$$P(X > 4a | X > 3a) = P\left(X > \frac{4}{3}a\right)$$

can be written with respect to Y as

$$P(Y > \ln 4 | Y > \ln 3) = P(Y > \ln 4 - \ln 3),$$

which is just an example of the memory less property of the exponential RV Y.

Problem 3. (15 pts) Let $U \sim \text{Uniform}([1, 2])$ and let X be the largest root of the quadratic equation

$$x^2 - 2Ux + 1 = 0.$$

a. (5 pts) Give a formula for X in terms of U. What is the range of X?

b. (10 pts) What are the CDF and PDF of X? (Hint: use the fact that for $x \geq 1$, the event $\{X \leq x\}$ is the same as the event $\{x^2 - 2Ux + 1 \geq 0\}$)

Solution:

a. By the quadratic formula the roots of the polynomial are given by

$$U \pm \sqrt{U^2 - 1}$$

The largest one X is therefore

$$X = U + \sqrt{U^2 - 1}.$$

Note that $u \mapsto u + \sqrt{u^2 - 1}$ is increasing whenever $u \geq 1$ and therefore since U varies between 1 and 2 we can see that the range of X is just

$$[1, 2 + \sqrt{3}].$$

b. To calculate the CDF and PDF of X, following the hint, we note that if $x \geq 1$ then $X \leq x$ is equivalent to

$$x^2 - 2Ux + 1 \geq 0 \quad \Rightarrow \quad U \leq \frac{x^2+1}{2x}$$

and therefore using the CDF for Uniform$(1,2)$, when $x \in [1, 2+\sqrt{3}]$

$$F_X(x) = P(X \leq x) = P\left(U \leq \frac{x^2+1}{2x}\right) = \frac{\frac{x^2+1}{2x} - 1}{2-1} = \frac{x}{2} + \frac{1}{2x} - 1.$$

As a sanity check we can see that

$$F(1) = \frac{1}{2} + \frac{1}{2} - 1 = 0$$

and

$$\begin{aligned} F(2+\sqrt{3}) &= \frac{2+\sqrt{3}}{2} + \frac{1}{2(2+\sqrt{3})} - 1 \\ &= \frac{2+\sqrt{3}}{2} + \frac{2-\sqrt{3}}{2} - 1 \\ &= 1 \end{aligned}$$

This gives a PDF of

$$f_X(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_X(x) = \begin{cases} \frac{1}{2}\left(1 - \frac{1}{x^2}\right) & x \in [1, 2+\sqrt{3}] \\ 0 & \text{otherwise} \end{cases}.$$

Problem 4. (15 pts) The Normal distribution is commonly used in practice to approximate the Binomial distribution for large n. In fact, it also works pretty well for fairly "small" values of n as long as p isn't too close to 0 or 1. Lets see how well it does for $n = 25$ and $p = .6$. (You may use a calculator and CDF calculator in the book)

a. (2 pts) Suppose that $X \sim$ Binomial(n,p), n and p as above. Compute the probability $P_X(7)$.

b. (5 pts) Suppose that $Y \sim N(\mu, \sigma^2)$, with the same mean and variance as X, $\mu = np$ and $\sigma^2 = npq$. Compute $P(6.5 \leq Y \leq 7.5)$. Compare your answer with part a.

c. (4 pts) The general principle is that the normal approximation is good as long as the values $p \pm 3\sqrt{\frac{pq}{n}}$ are between 0 and 1. Show that this is the case if and only if

$$n > \frac{9p}{q}, \quad \text{and} \quad n > \frac{9q}{p}$$

or in other words if

$$n > 9\left(\frac{\max\{p,q\}}{\min\{p,q\}}\right).$$

d. (4 pts) How large should n be taken to approximate the Binomial distribution by a Normal for $p = .5, .8, .99, .999$?

Solution:

a. For $n = 25$ and $p = .6$, we find using the Binomial PMF (or CDF) calculator

$$P_X(7) = P(X \leq 7) - P(X \leq 6) = 0.00092$$

b. The mean $\mu = np = 25(.6) = 15$ and the variance is $\sigma^2 = npq = 25(.6)(.4) = 6$. To compute this probability we find that

$$\begin{aligned} P(6.5 \leq Y \leq 7.5) &= \Phi\left(\frac{7.5 - 15}{\sqrt{6}}\right) - \Phi\left(\frac{6.5 - 15}{\sqrt{6}}\right) \\ &\approx \Phi(-3.06186) - \Phi(-3.47011) \\ &= 0.0011 - 0.00026 = 0.00084 \end{aligned}$$

While these answers differ, they are actually very close with $0.00092 - 0.00084 = 0.00008$ as the error between them.

c. Note first that $p + 3\sqrt{\frac{pq}{n}}$ being between 0 and 1 is equivalent to

$$p + 3\sqrt{\frac{pq}{n}} < 1$$

since the lower bound by 0 is automatic. After rearranging we obtain the equivalent inequalities

$$3\sqrt{\frac{pq}{n}} < q \quad \Leftrightarrow \quad n > \frac{9p}{q}.$$

Similarly, $p - 3\sqrt{\frac{pq}{n}}$ being between 0 and 1 is equivalent to

$$0 < p - 3\sqrt{\frac{pq}{n}}$$

since the upper bound by 1 is automatic (p being a probability). Again rearranging give the equivalent inequalities

$$\sqrt{\frac{pq}{n}} < q \quad \Leftrightarrow \quad n > \frac{9q}{p}.$$

The fact that all these inequalities are equivalent implies "if and only if". Since n needs to be bigger than both $\frac{9p}{q}$ and $\frac{9q}{p}$, we see this condition is equivalent to

$$n > \max\left\{\frac{9p}{q}, \frac{9q}{p}\right\} = 9\left(\frac{\max\{p, q\}}{\min\{p, q\}}\right)$$

d. Using the formula we just showed above we see that for each value of p we need n to be

$$
\begin{aligned}
p = .5 \quad &\Rightarrow \quad n > 9\frac{.5}{.5} = 9 \\
p = .8 \quad &\Rightarrow \quad n > 9\frac{.8}{.2} = 36 \\
p = .99 \quad &\Rightarrow \quad n > 9\frac{.99}{.01} = 891 \\
p = .999 \quad &\Rightarrow \quad n > 9\frac{.999}{.001} = 8,991.
\end{aligned}
$$

Problem 5 (26 pts) Our book doesn't have a CDF calculator for the Gamma distribution (bummer!). However, we have seen that the Gamma distribution is closely related to the Poisson distribution. In fact, it is possible to calculate all the probabilities for the Gamma distribution using only the Poisson distribution. Let $X \sim \text{Gamma}(n, \lambda)$ for $\lambda > 0$ and $n \in \mathbb{N}$ and let $Y \sim \text{Poisson}(\lambda)$.

a. (8 pts) Use the following formula

$$\frac{1}{\Gamma(n)} \int_{\lambda}^{\infty} x^{n-1} e^{-x} \, \mathrm{d}x = \sum_{k=0}^{n-1} \frac{\lambda^k e^{-\lambda}}{k!}$$

to show that

$$P(X > 1) = P(Y \leq n - 1).$$

b. (5 pts) What does the result of part a give you for $P(X > 1)$ when $\lambda = 1$ and $n = 5$? (Use the Poisson CDF calculator)

c. (5 pts) Recall that the Poisson distribution can be interpreted as the number of occurrences of a certain event that occurs in the time window $[0, 1]$, where the times between each event are independent and Exponential(λ) distributed. We also know that the gamma distribution is a sum of n independent Exponential(λ) RVs. Using words, explain why $P(X > 1) = P(Y \leq n - 1)$ makes sense.

d. (8 pts) Now for each $t > 0$, let $Y^{(t)} \sim \text{Poisson}(\lambda t)$ (the number of occurrences of events in the time window $[0, t]$). Show that

$$P(X > t) = P(Y^{(t)} \leq n - 1).$$

Solution:

a. To show this, let's calculate $P(X > 1)$. Using change of variables $u = \lambda x$, we see

$$\begin{aligned} P(X > 1) &= \frac{1}{\Gamma(n)} \int_1^\infty \lambda^n x^{n-1} e^{-\lambda x} \mathrm{d}x \\ &= \frac{1}{\Gamma(n)} \int_\lambda^\infty u^{n-1} e^{-u} \mathrm{d}u \end{aligned}$$

and using the formula we have

$$P(X > 1) = \sum_{k=0}^{n-1} \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=0}^{n-1} P(Y = k) = P(Y \leq n-1)$$

b. For $\lambda = 1$ and $n = 5$ we get using the CDF calculator

$$P(X > 1) = P(Y \leq 4) \approx 0.99941$$

c. Since X is just the sum of n independent exponentials, it is the amount of "time" you have to wait until you experience n "rare" events, each separated by exponentially distributed times. Therefore $X > 1$ means that by time 1, you experienced no more than $n - 1$ of these events. Since the number of such events that occur before time 1 is distributed according to the Poisson distribution, this is the same as $Y \leq n - 1$.

d. Following a similar calculation to part a and using the same change of variables $u = \lambda x$

$$\begin{aligned} P(X > t) &= \frac{1}{\Gamma(n)} \int_t^\infty \lambda^n x^{n-1} e^{-\lambda x} \mathrm{d}x \\ &= \frac{1}{\Gamma(n)} \int_{\lambda t}^\infty u^{n-1} e^{-u} \mathrm{d}u \end{aligned}$$

Using the formula again, we find

$$P(X > t) = \sum_{k=0}^{n-1} \frac{(\lambda t)^k e^{-\lambda t}}{k!} = P(Y^{(t)} \leq n-1)$$