Tutorial 5

February 27, 2020

Question 1

The weekly demand for propane gas (in 1000s of litres) from a particular facility is a random variable, X, with pdf:

$$f(x) = \begin{cases} 2\left(1 - \frac{1}{x^2}\right) & 1 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) Compute the cdf of X.

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(t)\,dt \\ &= \int_{1}^{x} 2\left(1 - \frac{1}{t^2}\right) dt \\ &= 2\int_{1}^{x} 1 - t^{-2}\,dt \\ &= 2\left(t + t^{-1}\right)\Big|_{t=1}^{t=x} \\ &= 2\left(\left(x + \frac{1}{x}\right) - (1+1)\right) \\ &= 2\left(x + \frac{1}{x} - 2\right) \end{aligned}$$

$$F(x) = \mathbf{P}\left(X \leq x\right) = \begin{cases} 0 & x < 1 \\ 2\left(x + \frac{1}{x} - 2\right) & 1 \leq x \leq 2 \\ 1 & x > 2 \end{cases}$$

(b) Obtain an expression of the $100p^{\text{th}}$ percentile.

Let $\eta(p)$ represent the $100p^{\text{th}}$ percentile. We can think of $\eta(p)$ as some x-value on the x-axis of the graph of the cdf. From the cdf of the previous part, we have the following relationship:

$$F(\eta(p)) = \mathbf{P}\left(X \leq \eta(p)\right) = 2\left(\eta(p) + \frac{1}{\eta(p)} - 2\right) = p$$

To obtain an expression for the $100p^{\text{th}}$ percentile, we need to solve for $\eta(p)$. For clarity, we will write $\eta(p)$ as simply η.

$$2\left(\eta + \frac{1}{\eta} - 2\right) = p$$

$$2\eta + \frac{2}{\eta} - 4 - p = 0$$

$$2\eta^2 + 2 - 4\eta - p\eta = 0$$

$$\underbrace{2}_{a}\eta^2 \underbrace{-(4+p)}_{b}\eta + \underbrace{2}_{c} = 0$$

Using the quadratic formula:

$$\eta = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{4 + p \pm \sqrt{(4+p)^2 - 16}}{4}$$

$$\text{Let } \eta_1 = \frac{4 + p - \sqrt{(4+p)^2 - 16}}{4} \quad \text{and} \quad \eta_2 = \frac{4 + p + \sqrt{(4+p)^2 - 16}}{4}$$

For $0 \leq p \leq 1$, $\frac{1}{2} \leq \eta_1 \leq 1$ but $F(\eta) = p > 0$ only for $1 < \eta \leq 2$. Since this does not agree with the definition of our cdf above, η_1 is not a valid solution.

For $0 \leq p \leq 1$, $1 \leq \eta_2 \leq 2$. This agrees with the definition of our cdf above. Therefore an expression for the $100p^{\text{th}}$ percentile is:

$$\eta(p) = \frac{4 + p + \sqrt{(4+p)^2 - 16}}{4}$$

(i) What is the value of the median?

The median is the 50^{th} percentile $\implies p = 0.5$.

$$\eta(0.5) = \frac{4 + 0.5 + \sqrt{(4+0.5)^2 - 16}}{4} = 1.64$$

(ii) 75% of demand is above what value?

75% of demand above some value means that we are looking for the 25^{th} percentile $\implies p = 0.25$.

$$\eta(0.25) = \frac{4 + 0.25 + \sqrt{(4+0.25)^2 - 16}}{4} = 1.42$$

(iii) 75% of demand is below what value?

75% of demand below some value means that we are looking for the 75^{th} percentile $\implies p = 0.75$.

$$\eta(0.75) = \frac{4 + 0.75 + \sqrt{(4+0.75)^2 - 16}}{4} = 1.83$$

(c) Compute $\mathbf{E}(X)$ and $\mathbf{Var}(X)$.

$$\begin{aligned}
\mathbf{E}(X) &= \int_{-\infty}^{\infty} x \cdot f(x)\,dx \\
&= \int_{1}^{2} x \cdot 2\left(1 - \frac{1}{x^2}\right)\,dx \\
&= 2\int_{1}^{2} x - \frac{1}{x}\,dx \\
&= 2\left(\frac{1}{2}x^2 - \ln|x|\right)\Bigg|_{x=1}^{x=2} \\
&= 2\left(\left(\frac{1}{2}(2^2) - \ln(2)\right) - \left(\frac{1}{2}(1^2) - \underbrace{\ln(1)}_{=0}\right)\right) \\
&= 2\left(\frac{3}{2} - \ln(2)\right) \\
&= 3 - \ln(4) \qquad\qquad [\text{Recall: } 2\ln(2) = \ln(2^2) = \ln(4)]
\end{aligned}$$

$\mathbf{Var}(X) = \mathbf{E}\left((X-\mu)^2\right)$ but it can get quite messy to compute:

$$\int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)\,dx$$

We can however use the fact that $\mathbf{Var}(X) = \mathbf{E}\left(X^2\right) - \left(\mathbf{E}(X)\right)^2$.

We start by computing $\mathbf{E}\left(X^2\right)$:

$$\begin{aligned}
\mathbf{E}\left(X^2\right) &= \int_{-\infty}^{\infty} x^2 \cdot f(x)\,dx \\
&= \int_{1}^{2} x^2 \cdot 2\left(1 - \frac{1}{x^2}\right)\,dx
\end{aligned}$$

$$
\begin{aligned}
&= 2\int_{1}^{2} x^2 - 1\,dx \\
&= 2\left(\frac{1}{3}x^3 - x\right)\Bigg|_{x=1}^{x=2} \\
&= 2\left(\left(\frac{1}{3}(2^3) - 2\right) - \left(\frac{1}{3}(1^3) - 1\right)\right) \\
&= 2\left(\frac{8}{3} - 2 - \frac{1}{3} + 1\right) \\
&= \frac{8}{3}
\end{aligned}
$$

$$\mathbf{Var}\left(X\right) = \mathbf{E}\left(X^2\right) - \left(\mathbf{E}\left(X\right)\right)^2 = \frac{8}{3} - \left(3 - \ln\left(4\right)\right)^2 = 0.06262$$

(d) If 1.8 thousands of litres are in stock at the beginning of the week and no new supply is due in during the week, how much of the 1.8 thousand litres is expected left at the end of the week? [Hint: express the quantity left as a linear function of X.]

Let $h(X)$ be the amount of gas left at the end of the week. Then $h(X) = 1.8 - X$. The expected amount left at the end of the week is:

$$\mathbf{E}\left(h(X)\right) = \mathbf{E}\left(1.8 - X\right) = 1.8 - \mathbf{E}\left(X\right) = 1.8 - \left(3 - \ln\left(4\right)\right) = 0.18629$$

Question 2

Let $Z \sim N(0,1)$. Compute the following probabilities:

(a) $\mathbf{P}\left(Z \leq 2.15\right)$

Table A.3 Standard Normal Curve Areas *(cont.)*

z	.00	.01	.02	.03	.04	.05
0.0	.5000	.5040	.5080	.5120	.5160	.5199
0.1	.5398	.5438	.5478	.5517	.5557	.5596
0.2	.5793	.5832	.5871	.5910	.5948	.5987
0.3	.6179	.6217	.6255	.6293	.6331	.6368
0.4	.6554	.6591	.6628	.6664	.6700	.6736
0.5	.6915	.6950	.6985	.7019	.7054	.7088
0.6	.7257	.7291	.7324	.7357	.7389	.7422
0.7	.7580	.7611	.7642	.7673	.7704	.7734
0.8	.7881	.7910	.7939	.7967	.7995	.8023
0.9	.8159	.8186	.8212	.8238	.8264	.8289
1.0	.8413	.8438	.8461	.8485	.8508	.8531
1.1	.8643	.8665	.8686	.8708	.8729	.8749
1.2	.8849	.8869	.8888	.8907	.8925	.8944
1.3	.9032	.9049	.9066	.9082	.9099	.9115
1.4	.9192	.9207	.9222	.9236	.9251	.9265
1.5	.9332	.9345	.9357	.9370	.9382	.9394
1.6	.9452	.9463	.9474	.9484	.9495	.9505
1.7	.9554	.9564	.9573	.9582	.9591	.9599
1.8	.9641	.9649	.9656	.9664	.9671	.9678
1.9	.9713	.9719	.9726	.9732	.9738	.9744
2.0	.9772	.9778	.9783	.9788	.9793	.9798
2.1	.9821	.9826	.9830	.9834	.9838	.9842

We look in row **2.1** and column **0.05**.

This gives us a probability of 0.9842.

Therefore $\mathbf{P}\left(Z \leq 2.15\right) = 0.9842$.

(b) $\mathbf{P}(0 \leq Z \leq 2.15)$

The standard normal distribution is centred at 0 and is symmetric about its centre. We know that the total area under a pdf must be 1. Then the area to the left of the centre of the standard normal distribution is 0.5.

$$\begin{aligned}
\mathbf{P}(0 \leq Z \leq 2.15) &= \mathbf{P}(Z \leq 2.15) - \mathbf{P}(Z \leq 0) \\
&= 0.9842 - 0.5000 \\
&= 0.4842
\end{aligned}$$

($*$) Note that for a continuous distribution, $\mathbf{P}(X \leq x) = \mathbf{P}(X < x)$. This is not true for discrete distributions.

Question 3

Suppose the force acting on a column that helps to support a building is a normally distributed random variable, X, with mean value 15 kips and standard deviation 1.25 kips. Compute the following probabilities.

From the question, we have that $X \sim N(\mu = 15, \sigma = 1.25)$.

(a) $\mathbf{P}(X \leq 15)$

$$\begin{aligned}
\mathbf{P}(X \leq 15) &= \mathbf{P}\left(\frac{X - 15}{1.25} \leq \frac{15 - 15}{1.25}\right) \\
&= \mathbf{P}(Z \leq 0) \\
&= 0.5
\end{aligned}$$

(b) $\mathbf{P}(|X - 15| \leq 3)$

$$\begin{aligned}
\mathbf{P}(|X - 15| \leq 3) &= \mathbf{P}\left(\frac{|X - 15|}{1.25} \leq \frac{3}{1.25}\right) \\
&= \mathbf{P}\left(\left|\frac{X - 15}{1.25}\right| \leq \frac{3}{1.25}\right) \\
&= \mathbf{P}(|Z| \leq 2.40) \\
&= \mathbf{P}(-2.40 \leq Z \leq 2.40) \\
&= \mathbf{P}(Z \leq 2.40) - \mathbf{P}(Z \leq -2.40) \\
&= \mathbf{P}(Z \leq 2.40) - (1 - \mathbf{P}(Z \leq 2.40)) \\
&= 2\mathbf{P}(Z \leq 2.40) - 1 \\
&= 0.9836
\end{aligned}$$

Question 4

Suppose that only 75% of all drivers in a certain province regularly wear a seat belt. A random sample of 500 drivers is selected.

Let X be the number of people in the sample of 500 who regularly wear a seat belt. Then:

$$X \sim \text{Binom}(n = 500,\, p = 0.75)$$

The probabilities that we wish to calculate are binomial probabilities. Since n is very large, we do not want to use the binomial pmf as it would be messy and time consuming. Since p is reasonably large, we can use a normal approximation of the binomial distribution, but we must first check that $np \geq 10$ and $nq \geq 10$.

$$np = (500)(0.75) = 375 > 10 \qquad\qquad nq = (500)(0.25) = 125 > 10$$

The conditions are satisfied. Then X will have an approximate normal distribution with parameters $\mu = np = 375$ and $\sigma = \sqrt{npq} = \sqrt{93.75}$ (≈ 9.68).

$$X \,\dot{\sim}\, N(\mu = 375,\, \sigma = \sqrt{93.75})$$

When going from the binomial distribution (discrete) to the normal distribution (continuous), we must make a continuity correction. This involves "widening" our desired interval by 0.5. For example, if we are interested in $\mathbf{P}(X \leq 200)$, the interval is:

$$X \leq 200 \quad \text{or} \quad X \in (-\infty,\, 200]$$

We cannot "widen" this interval on the left side so we widen it on the right side by 0.5. Therefore:

$$\underbrace{\mathbf{P}(X \leq 200)}_{X \text{ discrete}} \quad \approx \quad \underbrace{\mathbf{P}(X \leq 200.5)}_{X \text{ continuous}}$$

(a) What is the probability that between 360 and 400 drivers (inclusive) in the sample regularly wear a seat belt?

$$\begin{aligned}
\mathbf{P}(360 \leq X \leq 400) &= \mathbf{P}(X \leq 400) - \mathbf{P}(X < 360) \\
&= \mathbf{P}(X \leq 400) - \mathbf{P}(X \leq 359) \\
&\approx \mathbf{P}(X \leq 400.5) - \mathbf{P}(X \leq 359.5) \\
&= \mathbf{P}\left(\frac{X - 375}{\sqrt{93.75}} \leq \frac{400.5 - 375}{\sqrt{93.75}}\right) - \mathbf{P}\left(\frac{X - 375}{\sqrt{93.75}} \leq \frac{359.5 - 375}{\sqrt{93.75}}\right) \\
&= \mathbf{P}(Z \leq 2.63) - \mathbf{P}(Z \leq -1.60) \\
&= \mathbf{P}(Z \leq 2.63) - (1 - \mathbf{P}(Z \leq 1.60)) \\
&= 0.9958 - (1 - 0.9453) \\
&= 0.9411
\end{aligned}$$

(b) What is the probability that fewer than 400 of those sampled wear seat belt?

$$\begin{aligned}
\mathbf{P}(X < 400) &= \mathbf{P}(X \leq 399) \\
&\approx \mathbf{P}(X \leq 399.5) \\
&= \mathbf{P}\left(\frac{X - 375}{\sqrt{93.75}} \leq \frac{399.5 - 375}{\sqrt{93.75}}\right) \\
&= \mathbf{P}(Z \leq 2.53) \\
&= 0.9943
\end{aligned}$$

Question 5

The number of arrivals at a drive-thru window of a local bank follows a Poisson distribution with rate of 0.3 per hour.

Let X represent the number of arrivals in t hours. Then $\lambda = \alpha \cdot t = 0.3t$ such that

$$X \sim \text{Poisson}(\lambda = 0.3t)$$

(a) What is the probability that at least 2 hours will elapse between two successive arrivals?

Let T be the time (in hours) between two successive occurrences. Then

$$T \sim \text{Exp}(\lambda = \alpha = 0.3)$$

The exponential cdf is:

$$F(t) = \mathbf{P}(T \leq t) = \begin{cases} 0 & t < 0 \\ 1 - e^{-\lambda t} & t \geq 0 \end{cases}$$

The probability that at least $t = 2$ hours will elapse between these two successive occurrences is:

$$\begin{aligned}
\mathbf{P}(T \geq 2) &= 1 - \mathbf{P}(T \leq 2) \\
&= 1 - (1 - e^{-0.3 \cdot 2}) \\
&= e^{-0.3 \cdot 2} \\
&= 0.5488
\end{aligned}$$

(b) What is the average time elapsed between two successive arrivals?

If T has an exponential distribution with **rate** parameter λ then:

$$\mu = \mathbf{E}(T) = \frac{1}{\lambda}$$

Therefore $\mu = \mathbf{E}(T) = \frac{1}{\lambda} = \frac{1}{0.3} = \frac{10}{3}$.