Economics 30330: Statistics for Economics
Problem Set 5 - Suggested Solutions
University of Notre Dame
Instructor: Julio Garín
Spring 2012

Continuous Probability Distributions (120 Points)

1. A management consulting firm conducted a study of service time at the drive-up window at McDonalds. They found that the average time between placing an order and receiving the order was 2.78 minutes.
   Note that $X \sim exp(\mu, \mu^2)$ *so that* $f(x) = \frac{1}{\mu}e^{-\frac{x}{\mu}}$ *with* $\mu = 2.78$.

   (a) What is the variance of the waiting time?
   $Var(x) = \sigma^2 = \mu^2 = 2.78^2 = 7.73$

   (b) What is the probability that a customer's waiting time is more than 5 minutes?

$$\begin{aligned} P(x \geq 5) &= 1 - P(x \leq 5) \\ &= 1 - F(5) \\ &= 1 - \int_0^5 \frac{1}{2.78}e^{-\frac{x}{2.78}}\,dx \\ &= e^{-\frac{5}{2.78}} \\ &= 0.1655 \end{aligned}$$

   (c) What is the probability that a customer's waiting time is less than 2.78 minutes?
   Another way is to use the definition of CDF for an exponential distribution that is given by $F(x) = 1 - e^{-\frac{x}{\mu}}$*:*

$$\begin{aligned} P(x \leq 2.78) &= F(2.78) \\ &= 1 - e^{\frac{2.78}{2.78}} \\ &= 1 - e^{-1} \\ &= 0.632 \end{aligned}$$

2. Eric's weekly expenditures at North Dining Hall are normally distributed with a mean of \$80 and a standard deviation of \$10. Eric wants to budget an amount each week he can spend at the cafeteria.

   (a) How much should Eric budget if he wants to spend less than the budgeted amount 90 percent of the time?
   Let X *be the amount spent by Eric,* $X \sim \mathcal{N}(80, 100)$.

$$\begin{aligned} P(x \leq \text{budget}) &= 0.90 \\ &= P\left(z \leq \frac{\text{budget} - \mu}{\sigma}\right) \end{aligned}$$

   From the table, $\Phi(1.285) = 0.90 \implies \frac{budget-\mu}{\sigma} = 1.285$*, which solving for budget gives us 92.85. The latter is the value such that 90% of the time Eric will spend less than that.*

3. Let $z \sim \mathcal{N}(0,1)$. Calculate,

   (a) $P(-1 \le z \le 0)$.

$$\begin{aligned} P(-1 \le z \le 0) &= P(x \le 0) - P(x \le -1) \\ &= \Phi(0) - \Phi(-1) \\ &= 0.5 - 0.1587 \\ &= 0.3413 \end{aligned}$$

   (b) $P(-2.51 \le z \le 0)$.

$$\begin{aligned} P(-2.51 \le z \le 0) &= P(x \le 0) - P(x \le -2.51) \\ &= \Phi(0) - \Phi(-2.51) \\ &= 0.5 - 0.062 \\ &= 0.494 \end{aligned}$$

   (c) $P(-1.75 \le z \le -1.04)$.

$$\begin{aligned} P(-1.75 \le z \le -1.04) &= P(x \le -1.04) - P(x \le -1.75) \\ &= \Phi(-1.04) - \Phi(-1.75) \\ &= 0.1492 - 0.0401 \\ &= 0.1091 \end{aligned}$$

   (d) Find z^* such that $P(z \ge z^*) = 0.6915$

$$\begin{aligned} P(z \ge z^*) &= 1 - P(z \le z^*) \\ &= 1 - \Phi(z^*) \\ &= 0.6915 \\ &\implies \Phi(z^*) = 0.3085 \end{aligned}$$

   From the table: $\Phi(-0.5) = 0.3085 \implies z^* = -0.5$.

4. Your classmates Luke and Shawn are decent golfers but they are always bragging about their ability to hit the monster drive. Just last week, Shawn claimed that his drives routinely go 295 yards. The average drive for a professional player on the PGA tour travels 272 yards with a standard deviation of 8 yards.

   We have that $X \sim \mathcal{N}(272, 64)$

   (a) If the distance of a professional drive is normally distributed, what fraction of drives exceed 295 yards? What can you said about Luke and Shawn's statements?

$$\begin{aligned} P(x \ge 295) &= 1 - P(x \le 295) \\ &= 1 - P\left(z \le \frac{295 - 272}{8}\right) \\ &= 1 - \Phi(2.88) = 0.002 \end{aligned}$$

   If Shawn's statement is true, he would be in the top .2% of PGA golfers in terms of average drive. Shawn and Luke are probably exaggerating about their skills.

(b) Again, assuming normality, how long would a drive have to be in the top 5 percent of drives hit on the professional tour?

We want to find the value α such that $P(x \geq \alpha) = 0.05$.

$$\begin{aligned} P(x \geq \alpha) &= 1 - P(x \leq \alpha) \\ &= 0.05 \\ &= 1 - P\left(z \leq \frac{\alpha - 272}{8}\right) \end{aligned}$$

Therefore,

$$\begin{aligned} P\left(z \leq \frac{\alpha - 272}{8}\right) &= 0.95 \\ &= \Phi\left(\frac{\alpha - 272}{8}\right) \end{aligned}$$

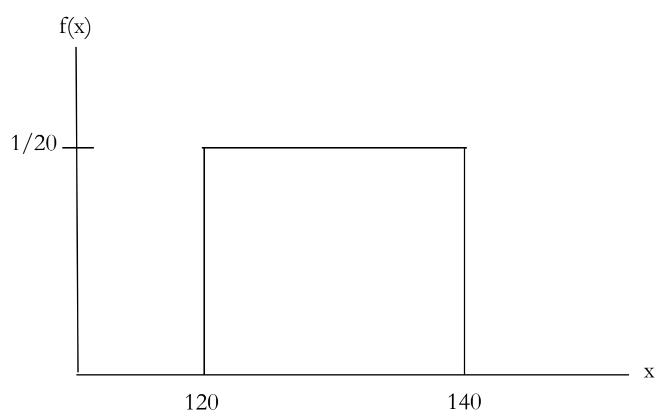
From the table, $\Phi(1.64) = 0.95 \Longrightarrow \dfrac{\alpha - 272}{8} = 1.64 \Longrightarrow \alpha = 285.12.$

5. Delta airlines quotes a flight time of 2 hours, 5 minutes for its flights from Cincinnati to Tampa. Suppose we believe that actual flight times are uniformly distributed between 2 and 2.333 hours (2 hours 20 minutes). Let X be the time flight for this route.

(a) Graph the PDF of X.

If X is uniformly distributed, its PDF is given by $f(x) = \dfrac{1}{b-a}$. *Hence, in this particular case,*

$$f(x) = \begin{cases} \frac{1}{20} & 120 \leq x \leq 140 \\ 0 & \text{otherwise} \end{cases}$$



(b) What is the probability that the flight will be no more than 5 minutes late?

$$\begin{aligned} P(x \leq 130) &= \int_{120}^{130} \frac{1}{20}\, dx \\ &= \frac{1}{20}(130 - 120) \\ &= 0.50 \end{aligned}$$

or, using the definition of the Uniform CDF:

$$
\begin{aligned}
F(x) &= \frac{x-a}{b-a} \\
&= \frac{130-120}{140-120} \\
&= 0.50
\end{aligned}
$$

(c) What is the expected flight time and what is the standard deviation of flight times?
Recall that the expected value is defined as:

$$E(x) = \mu = \int_x x f(x) dx$$

So we have,

$$
\begin{aligned}
E(x) &= \int_{120}^{140} x \frac{1}{20} d\,x \\
&= \frac{1}{40}(140^2 - 120^2) \\
&= 130
\end{aligned}
$$

Another way will be to use the definition we saw in class,

$$
\begin{aligned}
E(x) &= \frac{a+b}{2} \\
&= \frac{120+140}{2} \\
&= 130
\end{aligned}
$$

Similarly, for the variance

$$
\begin{aligned}
Var(x) &= \frac{(b-a)^2}{12} \\
&= \frac{(140-120)^2}{12} \\
&= 33.33 \\
\sigma &= \sqrt{33.33} = 5.77
\end{aligned}
$$

6. Most publications that describe colleges and universities report the 25th and 75th percentile of SAT scores for the entering students. Suppose that the SAT scores for entering students at ND are normally distributed with a mean of 1440 and a standard deviation of 120.

(a) What are the 25th and 75th percentile SAT scores?
Let X be the SAT score, then $X \sim \mathcal{N}(1440, 120^2)$

$$
\begin{aligned}
P(x \leq \Pi_{75}) &= 0.75 \\
&= P\left(z \leq \frac{\Pi_{75} - 1440}{120}\right)
\end{aligned}
$$

From the table, $\Phi(0.67) = 0.75$. Hence, $\frac{\Pi_{75}-1440}{120} = 0.67$ and solving for the percentile we have $\Pi_{75} = 1520.4 \approx 1520$. Since the normal distribution is symmetric, the 25th and 75th percentiles have to be the same distance from the mean. Therefore, $\Pi_{25} = 1440 - (1520 - 1440) = 1360$.

(b) The 25th and 75th percentile of SAT scores for incoming freshman at UVA are 1220 and 1440, respectively. If SAT scores for these students are normally distributed, what is the mean and standard deviation of SAT scores at UVA?

We know that the 25th and 75th percentiles are the same distance from the mean, therefore their average should be equal to the expected value. That is,

$$\begin{aligned}\mu_{UVA} &= \frac{1220+1440}{2} \\ &= 1330\end{aligned}$$

Now we are ready to calculate the standard deviation.

$$\begin{aligned}P(x \leq 1440) &= 0.75 \\ &= P\left(z \leq \frac{1440-1330}{\sigma_{UVA}}\right) \\ &= 0.75\end{aligned}$$

From the table, $\Phi(0.67) = 0.75$, $\implies \frac{1440-1330}{\sigma_{UVA}} = 0.67$. *Solving the latter for the standard deviation, we have that* $\sigma_{UVA} = 164.2$. *Therefore, if* X *is the r.v. that represents the SAT scores for incoming freshman students at UVA we can describe it as* $X \sim \mathcal{N}(1330, 164.2^2)$.

7. The proportion of time Y that an industrial robot is in operation during a 40-hour week is a random variable with probability density function

$$f(y) = \begin{cases} 2y & \text{for } 0 \leq y \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

(a) Find $\mathbb{E}(Y)$ and $\text{Var}(Y)$.

$$\begin{aligned}E(Y) &= \int_1^0 yf(y)\,dy \\ &= \int_1^0 y2y\,dy \\ &= \frac{2}{3}(1-0) \\ &= \frac{2}{3}\end{aligned}$$

Similarly,

$$\begin{aligned}Var(Y) &= E(Y^2) - [E(X)]^2 \\ &= 2\int_1^0 y^2y\,dy - \left(\frac{2}{3}\right)^2 \\ &= \frac{2}{4} - \left(\frac{2}{3}\right)^2 \\ &= \frac{2}{4} - \frac{4}{9} \\ &= \frac{1}{18}\end{aligned}$$

(b) For the robot under study, the profit X for a week is given by $X = 200Y - 60$. Find $\mathbb{E}(X)$ and $\text{Var}(X)$.

$$\begin{aligned} E(X) &= E(200Y - 60) \\ &= E(200Y) - E(60) \\ &= 200E(Y) - E(60) \\ &= 200\left(\frac{2}{3}\right) - 60 \\ &= \frac{220}{3} \end{aligned}$$

$$\begin{aligned} Var(x) &= Var(200Y - 60) \\ &= Var(200Y) - Var(60) \\ &= 200^2 Var(Y) - 0 \\ &= 40000\left(\frac{1}{18}\right) \\ &= \frac{20000}{9} \end{aligned}$$

(c) Find an interval in which the profit should lie for at least 75% of the weeks that the robot is in use.
Here we can use Chebyshev's Theorem:

$$\begin{aligned} P(|x - \mu| \leq k\sigma) &= P(\mu - k\sigma \leq x \leq \mu + k\sigma) \\ &\geq 1 - \frac{1}{k^2} \end{aligned}$$

And, here we would like an interval with 75% of probability of occurring, so that

$$1 - \frac{1}{k^2} = 0.75$$

which implies that $k = 2$. Now that we know the number of standard deviations, we can calculate the numeric bounds for the interval.

$$\begin{aligned} P(\mu - 2\sigma \leq x \leq \mu + 2\sigma) &= P(-20.948 \leq x \leq 167.614) \\ &= 0.75 \end{aligned}$$

Therefore, $X \in (-20.948, 167.614)$ at least 75% of the weeks in which the robot is in use.

8. The SAT and ACT college entrance exams are taken by thousands of students each year. The mathematics portions of each of these exams produce scores that are approximately normally distributed. In recent years SAT mathematics exam scores have averaged 480 with standard deviation 100. The average and standard deviation for ACT mathematics scores are 18 and 6, respectively.

   (a) An engineering school sets 550 as the minimum SAT math score for new students. What percentage of students will score below 550 in a typical year?

Let X be the SAT score. Hence, $X \sim \mathcal{N}(480, 1000)$

$$\begin{aligned} P(x \leq 550) &= P\left(z \leq \frac{550-480}{100}\right) \\ &= P(Z \leq 0.70) \\ &= \Phi(0.70) \\ &= 0.758 \end{aligned}$$

Therefore, 75.8% of students will score below 550 in a typical year.

(b) What score should the engineering school set as a comparable standard on the ACT math test?

Let X be the ACT score so that $X \sim \mathcal{N}(18, 36)$ *We want to find* α *such that* $P(x \leq \alpha) = 0.758$ *and We know that*

$$\begin{aligned} 0.758 &= \left(z \leq \frac{\alpha - 18}{6}\right) \\ &= \Phi(0.7) \end{aligned}$$

This implies that $\frac{\alpha-18}{6} = 0.7$ *which, after solving for* α*, gives us* $\alpha = 22.2$.

9. As a measure of intelligence, mice are timed when going through a maze to reach a reward of food. The time (in seconds) required for any mouse is a random variable Y with a density function given by

$$f(y) = \begin{cases} \frac{b}{y^2} & y \geq b \\ 0 & \text{elsewhere} \end{cases}$$

where b is the minimum possible time needed to traverse the maze

(a) Show that $f(y)$ has the properties of a denstiy function.

i. $f(y) \geq 0 \; \forall \; y$. *Since b is time, it is always greater than 0, so* $b/y^2 > 0$.

ii. $\int_{-\infty}^{\infty} f(y)\, dy = 1$ *So,*

$$\begin{aligned} \int_b^{\infty} \frac{b}{y^2}\, dy &= -b\left(\left.\frac{1}{y}\right|_b^{\infty}\right) \\ &= -(0-1) \\ &= 1 \end{aligned}$$

(b) Find $F(y)$.

$$\begin{aligned} F(y) &= \int_b^y f(t)\, dt \\ &= \int_b^y \frac{b}{t^2}\, dt \\ &= -\left.\frac{b}{t}\right|_b^y \\ &= -\frac{b}{y} + \frac{b}{b} \\ &= 1 - \frac{b}{y} \end{aligned}$$

(c) Find $P(Y > b + c)$ for a positive constant c.

$$\begin{aligned} P(Y > b + c) &= 1 - P(Y \leq b + c) \\ &= 1 - F(b + c) \end{aligned}$$

We have then that $F(y) = 1 - b/y$, *so* $F(b + c) = 1 - b/(b + c)$. Hence,

$$\begin{aligned} P(Y > b + c) &= 1 - F(b + c) \\ &= 1 - \left(1 - \frac{b}{b + c}\right) \\ &= \frac{b}{b + c} \end{aligned}$$

(d) If c and d are both positive constants such that $d > c$, find $P(Y > b + d|Y > b + c)$.
Here we should use the definition of conditional probability, i.e.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

.

$$\begin{aligned} P(Y > b + d|Y > b + c) &= \frac{P\left[(Y > b + d) \cap (Y > b + c)\right]}{P(Y > b + c)} \\ &= \frac{P(Y > b + d)}{P(Y > b + c)} \end{aligned}$$

Where the last step follows from the fact that $d > c$ *so that the only way of obtaining, simultaneously,* $Y > b + d$ *and* $Y > b + c$ *is if* $Y > b + d$.

$$\begin{aligned} P(Y > b + d|Y > b + c) &= \frac{1 - P(Y < b + d)}{P(Y > b + c)} \\ &= \frac{1 - F(Y < b + d)}{1 - F(Y > b + c)} \\ &= \frac{1 - \left(1 - \frac{b}{b+d}\right)}{1 - \left(1 - \frac{b}{b+c}\right)} \\ &= \frac{b + c}{b + d} \end{aligned}$$

10. Suppose that X has a density function

$$f(y) = \begin{cases} kx(1 - x) & 0 \leq x \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

(a) Find the value of k that makes $f(x)$ a probability density function.

$$\begin{aligned} \int_0^1 kx(1 - x)\, dx &= k\left(\int_0^1 x\, dx - \int_0^1 x^2\, dx\right) \\ &= k\left[\frac{1}{2}(1 - 0) - \frac{1}{3}(1 - 0)\right] \\ &= k\left(\frac{1}{2} - \frac{1}{3}\right) \\ &= \frac{1}{6}k \end{aligned}$$

For f(x) to be a proper PDF, it must be that $\int_{-\infty}^{\infty} f(x)\, dx = 1$. *That is,* $(1/6)k = 1$, *so* $k = 6$.

(b) Find $P(0.4 \leq X \leq 1)$.

$$\begin{aligned}
P(0.4 \leq x \leq 1) &= \int_{0.4}^{1} 6(x - x^2)\, dx \\
&= 6\left[\frac{1}{2}(1 - 0.4^2) - \frac{1}{3}(1 - 0.4^3)\right] \\
&= 6(0.108) \\
&= 0.648
\end{aligned}$$

(c) Find $P(X < 0.4|X \leq 0.8)$.

$$\begin{aligned}
P(X < 0.4|X \leq 0.8) = \frac{P(X \leq 0.4)}{P(X \leq 0.8)} \\
&= \frac{1 - P(0.4 \leq X \leq 1)}{P(X \leq 0.8)} \\
&= \frac{1 - 0.648}{\int_0^{0.8} 6(x - x^2)\, dx} \\
&= \frac{1 - 0.648}{0.896} \\
&= 0.393
\end{aligned}$$

(d) Find the 95th percentile of X.

$$\begin{aligned}
P(x \leq \Pi_{95}) = 0.95 &= \int_0^{\Pi_{95}} 6(x - x^2)\, dx \\
&= 6\left(\frac{1}{2}(\Pi_{95})^2 - \frac{1}{3}(\Pi_{95})^3\right) \\
&= 3(\Pi_{95})^2 - 2(\Pi_{95})^3 \\
0.95
\end{aligned}$$

Therefore, we have to solve the equation of third degree $3(\Pi_{95})^2 - 2(\Pi_{95})^3 - 0.95 = 0$ *whose only value between* $[0, 1]$ *is equal to 0.86465. Therefore, the 95th percentile is given by* $\Pi_{95} = 0.86465$

(e) Find a value of x_0 so that $P(X < x_0) = 0.95$.

Since Y *is continuous,* $P(X < x_0) = P(X \leq x_0)$*, so* $x_0 = 0.86465$ *(from part (d)).*

(f) Compare the values obtained in part $d) - e)$. Explain the relationship between these two values.

The values are exactly the same because in a continuous distribution, the probability of obtaining a particular number is equal to zero (the mass of the distribution at any given point is zero).

11. Show that the maximum value of the normal density with parameters μ and σ is $1/(\sigma\sqrt{2\pi})$ and occurs when $x = \mu$. *Hint:* You are just maximizing a function.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{\sigma^2}}$$

To make the problem a little easier, first take the natural log of the function (sometimes is easier to work with a monotonically increasing transformation of the original function):[1]

$$\begin{aligned} \ln\left[(f(x)\right] &= ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{(x-\mu)^2}{\sigma^2} \\ &= ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{(x^2 - 2x\mu + \mu^2)}{\sigma^2} \end{aligned}$$

Now maximize the function by taking the first derivative w.r.t. x and setting this equal to zero:

$$\begin{aligned} \frac{\partial \ln\left[f(x)\right]}{\partial x} &= 0 \\ &= -\frac{1}{\sigma^2}(2x - 2\mu) \end{aligned}$$

We know this will gives as a maximum since the second order condition is negative, i.e. $\frac{\partial^2 \ln[f(x)]}{\partial^2 x^2} = -2/\sigma^2 < 0$*; therefore, solving the first order condition w.r.t. x give us $x = \mu$ at the maximum, which completes the proof.*

12. Explain intuitively:

    (a) What happens to the first and second moments of a variable when every observation its doubled?

    By doubling each observation we are also increasing the dispersion by the same amount, therefore the standard deviation doubles or, in other words, the variance increase by four. Similarly, the mean is multiplied by two. The new variable is $Y = 2X$.

    (b) What happens to the first and second moments of a variable when we add a 3 to each observation?

    Adding 3 to each observation shift the distribution, without affecting its dispersion. Therefore, the mean willincrease by 3 but the variance remains the same. In this case the new variable $Y = 3 + X$.

    (c) What happens to the first and second moments of a variable when, after multiplying each observation by two, we add a 3 to each value?

    By multiplying by two we had that the mean is doubled while the variance is multiplied by four and, by adding 3, only the mean increases by the same amount. Now the new variable can be written as $Y = 3 + 2X$.

[1] Recall that $\ln(a \times b) = \ln(a) + \ln(b)$ and $\ln(e^x) = x$.