Stat S-110 Homework 4 Solutions, Summer 2015

The following problems are from Chapters 5-6 of the book.

1. (BH 5.1) The Rayleigh distribution from Example 5.1.7 has PDF

$$f(x) = xe^{-x^2/2}, \quad x > 0.$$

Let X have the Rayleigh distribution.

(a) Find $P(1 < X < 3)$.

(b) Find the first quartile, median, and third quartile of X; these are defined to be the values q_1, q_2, q_3 (respectively) such that $P(X \leq q_j) = j/4$ for $j = 1, 2, 3$.

Solution:

(a) We have

$$P(1 < X < 3) = \int_1^3 xe^{-x^2/2}dx.$$

To compute this, we can make the substitution $u = -x^2/2$, or we can use the fact from Example 5.1.7 that the CDF of X is $F(x) = 1 - e^{-x^2/2}$ for $x > 0$. Then

$$P(1 < X < 3) = F(3) - F(1) = e^{-1/2} - e^{-9/2} \approx 0.595.$$

(b) Using the CDF F from above, we can find q_j by setting $F(q_j) = j/4$ and solving for q_j. This gives

$$1 - e^{-q_j^2/2} = j/4,$$

which becomes

$$q_j = \sqrt{-2\log(1 - j/4)}.$$

Numerically,

$$q_1 \approx 0.759, q_2 \approx 1.177, q_3 \approx 1.665.$$

2. (BH 5.4) Let X be a continuous r.v. with CDF F and PDF f.

(a) Find the conditional CDF of X given $X > a$ (where a is a constant with $P(X > a) \neq 0$). That is, find $P(X \leq x | X > a)$ for all a, in terms of F.

(b) Find the conditional PDF of X given $X > a$ (this is the derivative of the conditional CDF).

(c) Check that the conditional PDF from (b) is a valid PDF, by showing directly that it is nonnegative and integrates to 1.

Solution:

(a) We have $P(X \le x | X > a) = 0$ for $x \le a$. For $x > a$,

$$P(X \le x | X > a) = \frac{P(a < X \le x)}{P(X > a)} = \frac{F(x) - F(a)}{1 - F(a)}.$$

(b) The derivative of the conditional CDF is $f(x)/(1 - F(a))$ for $x > a$, and 0 otherwise.

(c) We have $f(x)/(1 - F(a)) \ge 0$ since $f(x) \ge 0$. And

$$\int_a^\infty \frac{f(x)}{1 - F(a)} dx = \frac{1}{1 - F(a)} \int_a^\infty f(x) dx = \frac{1 - F(a)}{1 - F(a)} = 1.$$

3. (BH 5.5) A circle with a random radius $R \sim \text{Unif}(0, 1)$ is generated. Let A be its area.

(a) Find the mean and variance of A, without first finding the CDF or PDF of A.

(b) Find the CDF and PDF of A.

Solution:

(a) We have $A = \pi R^2$. By LOTUS,

$$\begin{aligned} E(A) &= \pi \int_0^1 r^2 dr = \frac{\pi}{3} \\ E(A^2) &= \pi^2 \int_0^1 r^4 dr = \frac{\pi^2}{5} \\ \text{Var}(A) &= \frac{\pi^2}{5} - \frac{\pi^2}{9} = \frac{4\pi^2}{45}. \end{aligned}$$

(b) The CDF of A is

$$P(A \le a) = P(\pi R^2 \le a) = P(R \le \sqrt{a/\pi}) = \sqrt{a/\pi},$$

for $0 < a < \pi$ (and the CDF is 0 for $a \le 0$ and 1 for $a \ge \pi$). So the PDF of A is

$$f(a) = \frac{1}{2\sqrt{\pi a}},$$

for $0 < a < \pi$ (and 0 otherwise).

4. (BH 5.6) The 68-95-99.7% rule gives approximate probabilities of a Normal r.v. being within $1, 2,$ and 3 standard deviations of its mean. Derive analogous rules for the following distributions.

(a) Unif$(0, 1)$.

(b) Expo(1).

(c) Expo$(1/2)$. Discuss whether there is one such rule that applies to all Exponential distributions, just as the 68-95-99.7% rule applies to all Normal distributions, not just to the standard Normal.

Solution:

(a) Let $U \sim$ Unif$(0, 1)$. The mean is $\mu = 1/2$ and the standard deviation is $\sigma = \frac{1}{\sqrt{12}}$. So

$$P(|U - \mu| \leq \sigma) = P(\mu - \sigma \leq U \leq \mu + \sigma) = 2\sigma \approx 0.5774.$$

But

$$P(|U - \mu| \leq 2\sigma) = 1,$$

since $\mu + 2\sigma > 1$. Therefore, the analogous rule for the Unif$(0, 1)$ is a 58-100-100% rule.

(b) Let $X \sim$ Expo(1). The mean is $\mu = 1$ and the standard deviation is 1. So

$$\begin{aligned} P(|X - \mu| \leq \sigma) &= P(X \leq 2) = 1 - e^{-2} \approx 0.8647, \\ P(|X - \mu| \leq 2\sigma) &= P(X \leq 3) = 1 - e^{-3} \approx 0.9502, \\ P(|X - \mu| \leq 3\sigma) &= P(X \leq 4) = 1 - e^{-4} \approx 0.9817. \end{aligned}$$

So the analogous rule for the Expo(1) is an 86-95-98% rule.

(c) Let $Y \sim$ Expo(λ) and $X = \lambda Y \sim$ Expo(1). Then Y has mean $\mu = 1/\lambda$ and standard deviation $\sigma = 1/\lambda$. For any real number $c \geq 1$, the probability of Y being within c standard deviations of its mean is

$$P(|Y - \mu| \leq c\sigma) = P(|\lambda Y - \lambda\mu| \leq c\lambda\sigma) = P(|X - 1| \leq c) = P(X \leq c + 1) = 1 - e^{-c-1}.$$

This probability does not depend on λ, so for any λ the analogous rule for the Expo(λ) distribution is an 86-95-98% rule.

5. (BH 5.14) Let $U_1, \ldots, U_n$ be i.i.d. Unif$(0, 1)$, and $X = \max(U_1, \ldots, U_n)$. What is the PDF of X? What is EX?

Hint: Find the CDF of X first, by translating the event $X \le x$ into an event involving $U_1, \dots, U_n$.

Solution: Note that $X \le x$ holds if and only if all of the U_j's are at most x. So the CDF of X is

$$P(X \le x) = P(U_1 \le x, U_2 \le x, \dots, U_n \le x) = (P(U_1 \le x))^n = x^n,$$

for $0 < x < 1$ (and the CDF is 0 for $x \le 0$ and 1 for $x \ge 1$). So the PDF of X is

$$f(x) = nx^{n-1},$$

for $0 < x < 1$ (and 0 otherwise). Then

$$EX = \int_0^1 x(nx^{n-1})dx = n\int_0^1 x^n dx = \frac{n}{n+1}.$$

(For generalizations of these results, see the material on *order statistics* in Chapter 8.)

6. (BH 5.18) The *Pareto distribution* with parameter $a > 0$ has PDF $f(x) = a/x^{a+1}$ for $x \ge 1$ (and 0 otherwise). This distribution is often used in statistical modeling.

(a) Find the CDF of a Pareto r.v. with parameter a; check that it is a valid CDF.

(b) Suppose that for a simulation you want to run, you need to generate i.i.d. Pareto(a) r.v.s. You have a computer that knows how to generate i.i.d. Unif$(0, 1)$ r.v.s but does not know how to generate Pareto r.v.s. Show how to do this.

Solution:

(a) The CDF F is given by

$$F(y) = \int_1^y \frac{a}{t^{a+1}}dt = (-t^{-a})\Big|_1^y = 1 - \frac{1}{y^a}$$

for $y > 1$, and $F(y) = 0$ for $y \le 1$. This is a valid CDF since it is increasing in y (this can be seen directly or from the fact that $F' = f$ is nonnegative), right continuous (in fact it is continuous), $F(y) \to 0$ as $y \to -\infty$, and $F(y) \to 1$ as $y \to \infty$.

(b) Let $U \sim$ Unif$(0, 1)$. By universality of the Uniform, $F^{-1}(U) \sim$ Pareto(a). The inverse of the CDF is

$$F^{-1}(u) = \frac{1}{(1-u)^{1/a}}.$$

So

$$Y = \frac{1}{(1-U)^{1/a}} \sim \text{Pareto}(a).$$

To check directly that Y defined in this way is Pareto(a), we can find its CDF:

$$P(Y \leq y) = P\left(\frac{1}{y} \leq (1-U)^{1/a}\right) = P\left(U \leq 1 - \frac{1}{y^a}\right) = 1 - \frac{1}{y^a},$$

for any $y \geq 1$, and $P(Y \leq y) = 0$ for $y < 1$. Then if we have n i.i.d. Unif$(0,1)$ r.v.s, we can apply this transformation to each of them to obtain n i.i.d. Pareto(a) r.v.s.

7. (BH 5.24) The distance between two points needs to be measured, in meters. The true distance between the points is 10 meters, but due to measurement error we can't measure the distance exactly. Instead, we will observe a value of $10 + \epsilon$, where the error ϵ is distributed $\mathcal{N}(0, 0.04)$. Find the probability that the observed distance is within 0.4 meters of the true distance (10 meters). Give both an exact answer in terms of Φ and an approximate numerical answer.

Solution: Standardizing ϵ (which has mean 0 and standard deviation 0.2), the desired probability is

$$\begin{aligned}
P(|\epsilon| \leq 0.4) &= P(-0.4 \leq \epsilon \leq 0.4) \\
&= P(-\frac{0.4}{0.2} \leq \frac{\epsilon}{0.2} \leq \frac{0.4}{0.2}) \\
&= P(-2 \leq \frac{\epsilon}{0.2} \leq 2) \\
&= \Phi(2) - \Phi(-2) \\
&= 2\Phi(2) - 1.
\end{aligned}$$

By the 68-95-99.7% rule, this is approximately 0.95.

8. (BH 5.28) Walter and Carl both often need to travel from Location A to Location B. Walter walks, and his travel time is Normal with mean w minutes and standard deviation σ minutes (travel time can't be negative without using a tachyon beam, but assume that w is so much larger than σ that the chance of a negative travel time is negligible).

Carl drives his car, and his travel time is Normal with mean c minutes and standard deviation 2σ minutes (the standard deviation is larger for Carl due to variability in traffic conditions). Walter's travel time is independent of Carl's. On a certain day, Walter and Carl leave from Location A to Location B at the same time.

(a) Find the probability that Carl arrives first (in terms of Φ and the parameters). For this you can use the important fact, proven in the next chapter, that if X_1 and X_2 are independent with $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, then $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

(b) Give a fully simplified criterion (*not* in terms of Φ), such that Carl has more than a 50% chance of arriving first if and only if the criterion is satisfied.

(c) Walter and Carl want to make it to a meeting at Location B that is scheduled to begin $w + 10$ minutes after they depart from Location A. Give a fully simplified criterion (*not* in terms of Φ) such that Carl is more likely than Walter to make it on time for the meeting if and only if the criterion is satisfied.

Solution:

(a) Let W be Walter's travel time and C be Carl's. Using standardization,

$$P(C < W) = P(C - W < 0) = P\left(\frac{C - W - (c - w)}{\sigma\sqrt{5}} < \frac{-(c - w)}{\sigma\sqrt{5}}\right) = \Phi\left(\frac{w - c}{\sigma\sqrt{5}}\right).$$

(b) Since Φ is a strictly increasing function with $\Phi(0) = 1/2$ (by symmetry of the Normal), $P(C < W) > 1/2$ if and only if $w > c$.

(c) For Walter, the probability of making it on time is

$$P(W \le w + 10) = P\left(\frac{W - w}{\sigma} \le \frac{10}{\sigma}\right) = \Phi\left(\frac{10}{\sigma}\right).$$

For Carl, the probability is

$$P(C \le w + 10) = P\left(\frac{C - c}{2\sigma} \le \frac{w - c + 10}{2\sigma}\right) = \Phi\left(\frac{w - c + 10}{2\sigma}\right).$$

The latter is bigger than the former if and only if $(w - c + 10)/(2\sigma) > 10/\sigma$, which simplifies to $w > c + 10$.

9. (BH 5.39) Three students are working independently on their probability homework. All 3 start at 1 pm on a certain day, and each takes an Exponential time with mean 6 hours to complete the homework. What is the earliest time when all 3 students will have completed the homework, on average? (That is, at this time all 3 students need to be done with the homework.)

Solution: Label the students as $1, 2, 3$, and let X_j be how long it takes student j to finish the homework. Let T be the time it takes for all 3 students to complete the homework, so $T = T_1 + T_2 + T_3$ where $T_1 = \min(X_1, X_2, X_3)$ is how long it takes for one student to complete the homework, T_2 is the additional time it takes for a second student to complete the homework, and T_3 is the additional time until all 3 have completed the homework. Then $T_1 \sim \text{Expo}(\frac{3}{6})$ since, as shown in Example 5.6.3, the minimum of independent Exponentials is Exponential with rate the sum of the rates. By the memoryless property, at the first time when a student completes the homework the other two students are starting from fresh, so $T_2 \sim \text{Expo}(\frac{2}{6})$. Again by the memoryless property, $T_3 \sim \text{Expo}(\frac{1}{6})$. Thus,

$$E(T) = 2 + 3 + 6 = 11,$$

which shows that on average, the 3 students will have all completed the homework at midnight, 11 hours after they started.

10. (BH 6.9) Let Y be Log-Normal with parameters μ and σ^2. So $Y = e^X$ with $X \sim \mathcal{N}(\mu, \sigma^2)$. Evaluate and explain whether or not each of the following arguments is correct.

(a) Student A: "The median of Y is e^μ because the median of X is μ and the exponential function is continuous and strictly increasing, so the event $Y \le e^\mu$ is the same as the event $X \le \mu$."

(b) Student B: "The mode of Y is e^μ because the mode of X is μ, which corresponds to e^μ for Y since $Y = e^X$."

(c) Student C: "The mode of Y is μ because the mode of X is μ and the exponential function is continuous and strictly increasing, so maximizing the PDF of X is equivalent to maximizing the PDF of $Y = e^X$."

Solution:

(a) Student A is right: e^μ is the median of Y, since

$$P(Y \le e^\mu) = P(X \le \mu) = 1/2.$$

(b) Student B is wrong. Figure 6.2 and the discussion of it give an example of a Log-Normal where the mode is clearly less than the median. It turns out that the mode of Y is $e^{\mu-\sigma^2}$, which is less than e^μ for any $\sigma > 0$.

If Z is a *discrete* r.v. and $W = e^Z$, then $P(W = w) = P(Z = z)$, where $z = \log w$, so if z_0 maximizes $P(Z = z)$ then $w_0 = e^{z_0}$ maximizes $P(W = w)$. But X is a *continuous* r.v., and it's *not* true that $f_Y(y) = f_X(x)$, where $x = \log y$; see Chapter 8 for a detailed discussion of how to handle transformations correctly.

(c) Student C is also wrong. In fact, saying that the mode of Y is μ is a *category error*, since μ could be negative, whereas Y is always positive. It's true (and useful) that if a function $f(x)$ is maximized at $x = x_0$, then $g(x) = e^{f(x)}$ is also maximized at $x = x_0$. But the PDF of Y is *not* the exponential of the PDF of X; to think so would be *sympathetic magic*, confusing a random variable with its PDF.

11. (BH 6.15) Let $W = X^2 + Y^2$, with X, Y i.i.d. $\mathcal{N}(0, 1)$. The MGF of X^2 turns out to be $(1 - 2t)^{-1/2}$ for $t < 1/2$ (you can assume this).

(a) Find the MGF of W.

(b) What famous distribution that we have studied so far does W follow (be sure to state the parameters in addition to the name)? In fact, the distribution of W is also a special case of two more famous distributions that we will study in later chapters!

Solution:

(a) The MGF of W is $(1 - 2t)^{-1}$ for $t < 1/2$, since X^2 and Y^2 are i.i.d.

(b) The MGF of W is the Expo(1/2) MGF, so $W \sim \text{Expo}(1/2)$. (It is also $\text{Gamma}(1, 1/2)$, and Chi-Square with 2 degrees of freedom; the Gamma and Chi-Square distributions are introduced in later chapters.)

12. (BH 6.25) Let $Y = X^\beta$, with $X \sim \text{Expo}(1)$ and $\beta > 0$. The distribution of Y is called the *Weibull* distribution with parameter β. This generalizes the Exponential, allowing for non-constant hazard functions. Weibull distributions are widely used in statistics, engineering, and survival analysis; there is even an 800-page book devoted to this distribution: *The Weibull Distribution: A Handbook* by Horst Rinne.

For this problem, let $\beta = 3$.

(a) Find $P(Y > s + t | Y > s)$ for $s, t > 0$. Does Y have the memoryless property?

(b) Find the mean and variance of Y, and the nth moment $E(Y^n)$ for $n = 1, 2, \ldots$.

(c) Determine whether or not the MGF of Y exists.

Solution:

(a) The CDF of Y is

$$P(Y \leq y) = P(X^3 \leq y) = P(X \leq y^{1/3}) = 1 - e^{-y^{1/3}},$$

for $y > 0$. So

$$P(Y > s + t | Y > s) = \frac{P(Y > s + t)}{P(Y > s)} = \frac{e^{-(s+t)^{1/3}}}{e^{-s^{1/3}}},$$

which is not the same as $P(Y > t) = e^{-t^{1/3}}$. Thus, Y does *not* have the memoryless property (nor could it, since it is not Exponential).

(b) Example 6.5.1 shows that the moments of X are given by $E(X^n) = n!$. This allows us to find the moments of Y without doing any additional work! Specifically, we have

$$E(Y^n) = E(X^{3n}) = (3n)!.$$

The mean and variance of Y are

$$E(Y) = 3! = 6, \mathrm{Var}(Y) = 6! - 6^2 = 684.$$

(c) By LOTUS,

$$E(e^{tY}) = E(e^{tX^3}) = \int_0^\infty e^{tx^3 - x} dx.$$

This integral diverges for $t > 0$ since the tx^3 term dominates over the x; more precisely, we have $tx^3 - x > x$ for all x sufficiently large (specifically, for $x > \sqrt{2/t}$), so this integral diverges by comparison with the divergent integral $\int_0^\infty e^x dx$. Therefore, the MGF of Y does not exist, even though all the moments of Y do exist.