Statistics 110: Introduction to Probability — Section 5 Solutions, Fall 2018

Continuous Random Variables

TFs: Justin Zhu, Michael Zhang, Credits to Timothy Kang

Practice Problems

Example 1. Birthdays.
Use Poisson approximations to investigate the following types of coincidences. The usual assumptions of the birthday problem apply, such as that there are 365 days in a year, with all days equally likely.

(a) How many people are needed to have a 50% chance that at least one of them has the same birthday as *you*?

(b) How many people are needed to have a 50% chance that there are two people who not only were born on the same day, but also were born at the same hour (e.g., two people born between 2 pm and 3 pm are considered to have been born at the same hour).

Solution

(a) Let k be the number of people there are other than you. Create an indicator variable I_i for each of the k people as to whether they have the same birthday as you. Then, $P(I_i = 1) = \frac{1}{365}$ and thus $E\left[\sum_{i=1}^{k} I_i\right] = \frac{k}{365}$. Therefore, we can model this as a Pois $\left(\frac{k}{365}\right)$ and so we just need to calculate

$$1 - e^{-k/365} = 0.5$$

It turns out that $k \approx 253$.

(b) This is the birthday problem but with $365 \cdot 24$ types instead of just 365. Creating an indicator r.v. for whether each pair of k people have the same birthday, we get that the number of pairs of people with the same birthday is distributed approximately Pois $\left(\frac{\binom{k}{2}}{365 \cdot 24}\right)$ and thus the probability of at least people having the same birthday is approximately:

$$1 - e^{-\frac{\binom{k}{2}}{365 \cdot 24}}$$

Setting it equal to $\frac{1}{2}$ gives us $k = 111$.

Example 2. Uniform Power.
Let $U \sim \text{Unif}(-1, 1)$

(a) Compute $E(U), \text{Var}(U), E(U^4)$.

(b) Find the PDF and CDF of U^2. Is it also a uniform distribution?

Solution
When trying to compute the PDF of a transformation, we usually work from the CDF first. Note that the PDF of U is $f(x) = \frac{1}{2}$.

(a) $E(U) = 0$ because the distribution is symmetric about 0. We need to calculate $E(U^2)$ for the variance, so we have:

$$E(U^2) = \int_{-1}^{1} u^2 \cdot \frac{1}{2}\,\mathrm{d}u = \left[\frac{1}{6}u^3\right]_{-1}^{1} = \frac{1}{3}$$

Therefore, $\mathrm{Var}(U) = E(U^2) - E(U)^2 = \boxed{\frac{1}{3}}$

Next, we do the same for $E(U^4)$:

$$E(U^4) = \int_{-1}^{1} u^4 \cdot \frac{1}{2}\,\mathrm{d}u = \left[\frac{1}{10}u^5\right]_{-1}^{1} = \boxed{\frac{1}{5}}$$

(b) We first find the CDF.

$$P(U^2 < k) = P(-\sqrt{k} < U < \sqrt{k}) = \frac{2\sqrt{k}}{2} = \sqrt{k}$$

, which we can easily calculate graphically. If this wasn't possible to do graphically, then we would integrate the PDF of U between $-k$ and k.

Therefore, the PDF is:

$$\frac{d}{dk}P(U^2 < k) = \boxed{\frac{1}{2\sqrt{k}}}$$

This is definitely **not** a uniform distribution. This also shows that U^k is not uniform anymore for any $k > 1$.

Example 3. Normal Squared.
Let $Z \sim N(0,1)$ with CDF Φ. The PDF of Z^2 is the function given by:

$$g(w) = \frac{1}{\sqrt{2\pi w}}e^{-w/2}$$

with a support of $w \geq 0$.

(a) Find expressions for $E(Z^4)$ as integrals in two different ways, one based on the PDF of Z and the other based on the PDF of Z^2.

(b) Find $E(Z^2 + Z + \Phi(Z))$.

(c) Find the CDF of Z^2 in terms of Φ; do not find the PDF of g.

Solution

(a) Let $W = Z^2$, so $W^2 = Z^4$. By LOTUS,

$$E(Z^4) = \int_{-\infty}^{\infty} z^4 \varphi(z) dz = \int_0^{\infty} w^2 g(w) dw,$$

where $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ is the PDF of Z, and g is as above. (Using techniques from Chapter 6, it turns out that this reduces to a very simple answer: $E(Z^4) = 3$.)

(b) By linearity, this is $E(Z^2) + E(Z) + E(\Phi(Z))$. The second term is 0 and the first term is 1 since $E(Z) = 0, \mathrm{Var}(Z) = 1$. The third term is $1/2$ since by universality of the Uniform, $\Phi(Z) \sim \mathrm{Unif}(0, 1)$. Thus, the value is $3/2$.

(c) For $w \leq 0$, the CDF of Z^2 is 0. For $w > 0$, the CDF of Z^2 is

$$P(Z^2 \leq w) = P(-\sqrt{w} \leq Z \leq \sqrt{w}) = \Phi(\sqrt{w}) - \Phi(-\sqrt{w}) = 2\Phi(\sqrt{w}) - 1.$$

Example 4. Breaking Sticks.
A stick of length 1 is broken at a uniformly random point yielding two pieces. Let X and Y be the lengths of the shorter and longer pieces, respectively, and let $R = \frac{X}{Y}$ be the ratio of the lengths X and Y.

(a) Find the CDF and PDF of R

(b) Find the expected value of R (if it exists).

Solution

(a) Let $U \sim \mathrm{Unif}(0, 1)$ represent the location of the break, and let $X = \min(U, 1-U)$. As a side note, recall that U and $1-U$ have the same distribution of $\mathrm{Unif}(0, 1)$. For some value $r \in (0, 1)$, we have that

$$P(R \leq r) = P(\frac{X}{Y} \leq r) = P(X \leq r(1-X)) = P(X \leq \frac{r}{1+r})$$

However, to fully evaluate this, we must find the CDF of X. To do so, we first recognize that, because X is the minimum of two random variables, it is easier to consider the event $\{X > x\}$, as this event is equivalent to saying $\{U > x, 1-U > x\}$ (if the minimum of two values is greater than x, then the other must also be greater than x). We thus have that

$$P(X \leq x) = 1 - P(X > x) = 1 - P(U > x, 1-U > x) = 1 - P(x < U < 1-x) = 2x$$

Therefore, we have that the CDF of R is simply

$$P(R \leq r) = \frac{2r}{1+r}$$

To find the PDF we differentiate with respect to r and find that

$$f_R(r) = \frac{2}{(1+r)^2}$$

(b) Letting $t = 1 + r$, we have

$$E(R) = 2\int_0^1 \frac{r}{(1+r)^2} dr = 2\int_1^2 \frac{t-1}{t^2} dt = 2\int_1^2 \frac{1}{t} dt - 2\int_1^2 \frac{1}{t^2} dt = 2\ln 2 - 1$$

Example 5. Universality of the Uniform.
Let $U \sim \text{Unif}(0,1)$, and let $X = -(\log(1-U))^{1/3}$. Find the CDF and PDF of X.

Solution
We calculate $P(x \leq X)$, but use the substitution given to try and solve a Uniform distribution's CDF instead:

$$\begin{aligned}
P(X \leq x) &= P(-(\log(1-U))^{1/3} \leq x) \\
&= P(\log(1-U)^{1/3}) \geq -x) \\
&= P(\log(1-U) \geq -x^3) \\
&= P(1 - e^{-x^3} \geq U) \\
&= P(U \leq 1 - e^{-x^3}) \\
&= \boxed{(1 - e^{-x^3})}
\end{aligned}$$

which is the CDF of X.

The PDF of x is the derivative, so:

$$f(x) = \frac{\partial}{\partial x}(1 - e^{-x^3}) = 3x^2 e^{-x^3}$$

Example 6. Finishing Homework.
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

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.

Example 7. Continuous RV manipulation.
Let X be a continuous r.v. with CDF F and PDF f.

(a) Find the conditional CDF X given that $X > a$ (where a is a constant with $P(X > a) \neq 0$).

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

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

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

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

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

Example 8. Pareto Distribution.
The Pareto distribution with parameter $a > 0$ has PDF $f(x) = \frac{a}{x^{a+1}}$ for $x \geq 1$ (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. $\text{Unif}(0, 1)$ 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 $y \leq 1$. 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 $U \sim \text{Unif}(0, 1)$. By universality of the Uniform, $F^{-1}(U) \sim \text{Pareto}(a)$. The inverse of the CDF is

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

So

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