

Due: Sunday, August 2, 2020 at 10:00 PM  
Grace period until Sunday, August 2, 2020 at 11:59 PM

## 1 Unreliable Servers

In a single cluster of a Google competitor, there are a huge number of servers  $n$ , each with a uniform and independent probability of going down in a given day. On average, 4 servers go down in the cluster per day. As each cluster is responsible for a huge amount of internet traffic, it is fair to assume that  $n$  is a very large number. Recall that as  $n \rightarrow \infty$ , a  $\text{Binom}(n, \lambda/n)$  distribution will tend towards a  $\text{Poisson}(\lambda)$  distribution.

- What is an appropriate distribution to model the number of servers that crash on any given day for a certain cluster?
- Compute the expected value and variance of the number of crashed servers on a given day for a certain cluster.
- Compute the probability that fewer than 3 servers crashed on a given day for a certain cluster.
- Compute the probability at least 3 servers crashed on a given day for a certain cluster.

## 2 Class Enrollment

Lydia has just started her CalCentral enrollment appointment. She needs to register for a marine science class and CS 70. There are no waitlists, and she can attempt to enroll once per day in either class or both. The CalCentral enrollment system is strange and picky, so the probability of enrolling successfully in the marine science class on each attempt is  $\mu$  and the probability of enrolling successfully in CS 70 on each attempt is  $\lambda$ . Also, these events are independent.

- Suppose Lydia begins by attempting to enroll in the marine science class everyday and gets enrolled in it on day  $M$ . What is the distribution of  $M$ ?
- Suppose she is not enrolled in the marine science class after attempting each day for the first 5 days. What is the conditional distribution of  $M$  given  $M > 5$ ?
- Once she is enrolled in the marine science class, she starts attempting to enroll in CS 70 from day  $M + 1$  and gets enrolled in it on day  $C$ . Find the expected number of days it takes Lydia to enroll in both the classes, i.e.  $\mathbb{E}[C]$ .

- (d) Suppose instead of attempting one by one, Lydia decides to attempt enrolling in both the classes from day 1. Let  $M$  be the number of days it takes to enroll in the marine science class, and  $C$  be the number of days it takes to enroll in CS 70. What is the distribution of  $M$  and  $C$  now? Are they independent?
- (e) Let  $X$  denote the day she gets enrolled in her first class and let  $Y$  denote the day she gets enrolled in both the classes. What is the distribution of  $X$ ?
- (f) What is the expected number of days it takes Lydia to enroll in both classes now, i.e.  $\mathbb{E}[Y]$ .
- (g) What is the expected number of classes she will be enrolled in by the end of 14 days?

### 3 Short Answer

- (a) Let  $X$  be uniform on the interval  $[0, 2]$ , and define  $Y = 2X + 1$ . Find the PDF, CDF, expectation, and variance of  $Y$ .
- (b) Let  $X$  and  $Y$  have joint distribution

$$f(x,y) = \begin{cases} cxy + 1/4 & x \in [1, 2] \text{ and } y \in [0, 2] \\ 0 & \text{else} \end{cases}$$

Find the constant  $c$ . Are  $X$  and  $Y$  independent?

- (c) Let  $X \sim \text{Exp}(3)$ . What is the probability that  $X \in [0, 1]$ ? If I define a new random variable  $Y = \lfloor X \rfloor$ , for each  $k \in \mathbb{N}$ , what is the probability that  $Y = k$ ? Do you recognize this (discrete) distribution?
- (d) Let  $X_i \sim \text{Exp}(\lambda_i)$  for  $i = 1, \dots, n$  be mutually independent. It is a (very nice) fact that  $\min(X_1, \dots, X_n) \sim \text{Exp}(\mu)$ . Find  $\mu$ .

### 4 Useful Uniforms

Let  $X$  be a continuous random variable whose image is all of  $\mathbb{R}$ ; that is,  $\mathbb{P}[X \in (a, b)] > 0$  for all  $a, b \in \mathbb{R}$  and  $a \neq b$ .

- (a) Give an example of a distribution that  $X$  could have, and one that it could not.
- (b) Show that the CDF  $F$  of  $X$  is strictly increasing. That is,  $F(x + \varepsilon) > F(x)$  for any  $\varepsilon > 0$ . Argue why this implies that  $F : \mathbb{R} \rightarrow (0, 1)$  must be invertible.
- (c) Let  $U$  be a uniform random variable on  $(0, 1)$ . What is the distribution of  $F^{-1}(U)$ ?
- (d) Your work in part (c) shows that in order to sample  $X$ , it is enough to be able to sample  $U$ . If  $X$  was a discrete random variable instead, taking finitely many values, can we still use  $U$  to sample  $X$ ?

## 5 It's Raining Fish

A hurricane just blew across the coast and flung a school of fish onto the road nearby the beach. The road starts at your house and is infinitely long. We will label a point on the road by its distance from your house (in miles). For each  $n \in \mathbb{N}$ , the number of fish that land on the segment of the road  $[n, n + 1]$  is independently  $\text{pois}(\lambda)$  and each fish that is flung into that segment of the road lands uniformly at random within the segment. Keep in mind that you can cite any result from lecture or discussion without proof.

- (a) What is the distribution of the number of fish arriving in segment  $[0, n]$  of the road, for some  $n \in \mathbb{N}$ .
- (b) Let  $[a, b]$  be an interval in  $[0, 1]$ . What is the distribution of the number of fish that lands in the segment  $[a, b]$  of the road?
- (c) Let  $[a, b]$  be any interval such that  $a \geq 0$ . What is the distribution of the number of fish that land in  $[a, b]$ ?
- (d) Suppose you take a stroll down the road. What is the distribution of the distance you walk (in miles) until you encounter the first fish? Justify with proof.
- (e) Suppose you encounter a fish at distance  $x$ . What is the distribution of the distance you walk until you encounter the next fish?

## 6 Exponential Expectation

- (a) Let  $X \sim \text{Exp}(\lambda)$ . Use induction to show that  $\mathbb{E}[X^k] = k!/\lambda^k$  for every  $k \in \mathbb{N}$ .
- (b) For any  $|t| < \lambda$ , compute  $\mathbb{E}[e^{tX}]$  directly from the definition of expectation.
- (c) Using part (a), compute  $\sum_{k=0}^{\infty} \frac{\mathbb{E}[X^k]}{k!} t^k$ .
- (d) Let  $M(t) = \mathbb{E}[e^{tX}]$  be a function defined for all  $t$  such that  $|t| < \lambda$ . What is  $\left. \frac{dM(t)}{dt} \right|_{t=0}$ ? What is  $\left. \frac{d^2M(t)}{dt^2} \right|_{t=0}$ ? How does each of these relate to the mean and variance of an  $\text{Exp}(\lambda)$  distribution?

## 7 Noisy Love

Suppose you have confessed to your love interest on Valentine's Day and you are waiting to hear back. Your love interest is trying to send you a binary message: "0" means that your love interest is not interested in you, while "1" means that your love interest reciprocates your feelings. Let  $X$  be your love interest's message for you. Your current best guess of  $X$  has  $\mathbb{P}(X = 0) = 0.7$  and  $\mathbb{P}(X = 1) = 0.3$ . Unfortunately, your love interest sends you the message through a noisy channel, and instead of receiving the message  $X$ , you receive the message  $Y = X + \varepsilon$ , where  $\varepsilon$  is independent Gaussian noise with mean 0 and variance 0.49.

- (a) First, you decide upon the following rule: if you observe  $Y > 0.5$ , then you will assume that your love interest loves you back, whereas if you observe  $Y \leq 0.5$ , then you will assume that your love interest is not interested in you. What is the probability that you are correct using this rule? (Express your answer in terms of the CDF of the standard Gaussian distribution  $\Phi(z) = \mathbb{P}(\mathcal{N}(0, 1) \leq z)$ , and then evaluate your answer numerically.)
- (b) Suppose you observe  $Y = 0.6$ . What is the probability that your love interest loves you back? [Hint: This problem requires conditioning on an event of probability 0, namely, the event  $\{Y = 0.6\}$ . To tackle this problem, think about conditioning on the event  $\{Y \in [0.6, 0.6 + \delta]\}$ , where  $\delta > 0$  is small, so that  $f_Y(0.6) \cdot \delta \approx \mathbb{P}(Y \in [0.6, 0.6 + \delta])$ , and then apply Bayes Rule.]
- (c) Suppose you observe  $Y = y$ . For what values is it more likely than not that your love interest loves you back? [Hint: As before, instead of considering  $\{Y = y\}$ , you can consider the event  $\{Y \in [y, y + \delta]\}$  for small  $\delta > 0$ . So, when is  $\mathbb{P}(X = 1 \mid Y \in [y, y + \delta]) \geq \mathbb{P}(X = 0 \mid Y \in [y, y + \delta])$ ?]
- (d) Your new rule is to assume that your love interest loves you back if (based on the value of  $Y$  that you observe) it is more likely than not that your love interest loves you back. Under this new rule, what is the probability that you are correct?

## 8 Sum of Independent Gaussians

In this question, we will introduce an important property of the Gaussian distribution: the sum of independent Gaussians is also a Gaussian.

Let  $X$  and  $Y$  be independent standard Gaussian random variables. Recall that the density of the standard Gaussian is

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

- (a) What is the joint density of  $X$  and  $Y$ ?
- (b) Observe that the joint density of  $X$  and  $Y$ ,  $f_{X,Y}(x,y)$ , only depends on the quantity  $x^2 + y^2$ , which is the distance from the origin. In other words, the Gaussian is *rotationally symmetric*. Next, we will try to find the density of  $X + Y$ . To do this, draw a picture of the Cartesian plane and draw the region  $x + y \leq c$ , where  $c$  is a real number of your choice.
- (c) Now, rotate your picture clockwise by  $\pi/4$  so that the line  $X + Y = c$  is now vertical. Redraw your figure. Let  $X'$  and  $Y'$  denote the random variables which correspond to the  $\pi/4$  clockwise rotation of  $(X, Y)$  and express the new shaded region in terms of  $X'$  and  $Y'$ .
- (d) By rotational symmetry of the Gaussian,  $(X', Y')$  has the same distribution as  $(X, Y)$ . Argue that  $X + Y$  has the same distribution as  $\sqrt{2}Z$ , where  $Z$  is a standard Gaussian. This proves the following important fact: *the sum of independent Gaussians is also a Gaussian*. Notice that  $X \sim \mathcal{N}(0, 1)$ ,  $Y \sim \mathcal{N}(0, 1)$  and  $X + Y \sim \mathcal{N}(0, 2)$ . *In general, if  $X$  and  $Y$  are independent Gaussians, then  $X + Y$  is a Gaussian with mean  $\mu_X + \mu_Y$  and variance  $\sigma_X^2 + \sigma_Y^2$ .*

(e) Recall the CLT:

If  $\{X_i\}_{i \in \mathbb{N}}$  is a sequence of i.i.d. random variables with mean  $\mu \in \mathbb{R}$  and variance  $\sigma^2 < \infty$ , then:

$$\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\text{in distribution}} \mathcal{N}(0, 1) \quad \text{as } n \rightarrow \infty.$$

Prove that the CLT holds for the special case when the  $X_i$  are i.i.d.  $\mathcal{N}(0, 1)$ .

## 9 Homework Process and Study Group

Citing sources and collaborators are an important part of life, including being a student! We also want to understand what resources you find helpful and how much time homework is taking, so we can change things in the future if possible.

1. **What sources (if any) did you use as you worked through the homework?**
2. **If you worked with someone on this homework, who did you work with?** List names and student ID's. (In case of homework party, you can also just describe the group.)
3. **How did you work on this homework?** (For example, *I first worked by myself for 2 hours, but got stuck on problem 3, so I went to office hours. Then I went to homework party for a few hours, where I finished the homework.*)
4. **Roughly how many total hours did you work on this homework?**