

Section #4 Solutions

1. Are we due for an earthquake?: After the class where we talked about the probability of Earthquakes at Stanford, a student asked a question: “Doesn’t the probability of an earthquake happening change based on the fact that we haven’t had one for a while?” Let’s explore! Recall the USGS rate of earthquakes of magnitude 8+ is $\lambda = 0.002$ earthquakes per year.

- a. What is the probability of no 8+ earthquakes in four years after the 1908 earthquake (recall that earthquakes are exponentially distributed)?

Let X be the time until an earthquake. $X \sim \text{Exp}(\lambda = 0.002)$.

$$\begin{aligned} P(X \geq 4) &= 1 - P(X < 4) \\ &= 1 - F_X(4) \\ &= 1 - [1 - e^{-0.002 \cdot 4}] \\ &= e^{-0.008} \approx 0.992 \end{aligned}$$

- b. What is the probability of no 8+ earthquakes in the 113 years between the 1908 earthquake and four years from now?

$$\begin{aligned} P(X \geq 113) &= 1 - P(X < 113) \\ &= 1 - F_X(113) \\ &= 1 - [1 - e^{-0.002 \cdot 113}] \\ &= e^{-0.226} \approx 0.798 \end{aligned}$$

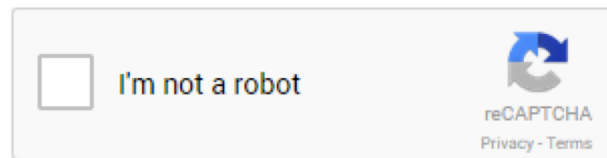
- c. What is the probability of no 8+ earthquakes in the 113 years between the 1908 earthquake and four years from now *given* that there have been no earthquakes in the last 109 years?

$$\begin{aligned} P(X > 113 | X > 109) &= \frac{P(X > 113, X > 109)}{P(X > 109)} \\ &= \frac{P(X > 113)}{P(X > 109)} = \frac{1 - F_X(113)}{1 - F_X(109)} \\ &= \frac{e^{-0.002 \cdot 113}}{e^{-0.002 \cdot 109}} = e^{-0.008} \approx 0.992 \end{aligned}$$

d. Did you notice anything interesting? Would this work for any value of λ ?

It turns out that exponentials are what we call a “memoryless distribution.” If X is an exponential random variable, it holds that $P(X > s + t | X > t) = P(X > s)$.

2. ReCaptcha: Based on browser history, Google believes that there is a 0.2 probability that a particular visitor to a website is a robot. They decide to give the visitor a recaptcha:



Google presents the visitor with a box, 10mm by 10mm. The visitor must click inside the box to show that they are not a robot. You have observed that robots click uniformly in the box. However, the distance location of a human click has X location (mm from the left) and the Y location (mm from the right) distributed as independent normals both with mean $\mu = 5$ and $\sigma^2 = 4$:

a. What the the probability density function of a robot clicking $X = x$ mm from the left of the box and $Y = y$ mm from the top of the box?

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{100} & \text{if } 0 < x, y < 10 \\ 0 & \text{else} \end{cases}$$

b. What the the probability density function of a human clicking $X = x$ mm from the left of the box and $Y = y$ mm from the top of the box?

$$\begin{aligned} f_{X,Y}(x, y) &= f_X(x)f_Y(y) && \text{independence} \\ &= \frac{1}{(2\sqrt{2\pi})^2} e^{-\frac{(x-5)^2}{8}} e^{-\frac{(y-5)^2}{8}} && \text{normal PDF} \\ &= \frac{1}{8\pi} e^{-\frac{(x-5)^2}{8}} e^{-\frac{(y-5)^2}{8}} && \text{normal PDF} \end{aligned}$$

- c. The visitor clicks in the box at $(x = 6 \text{ mm}, y = 6 \text{ mm})$. What is Google’s new belief that the visitor is a robot?

Let Click to be the event that the user clicked at location $X = 6, Y = 6$. We can then use Bayes Rule (with law of total probability in the denominator):

$$\begin{aligned} P(\text{Robot}|\text{Click}) &= \frac{f(\text{Click}|\text{Robot})P(\text{Robot})}{f(\text{Click})} \\ &= \frac{f(\text{Click}|\text{Robot})P(\text{Robot})}{f(\text{Click}|\text{Robot})P(\text{Robot}) + f(\text{Click}|\text{Human})P(\text{Human})} \\ &= \frac{\frac{1}{100} \cdot 0.2}{\frac{1}{100} \cdot 0.2 + \frac{1}{8\pi} e^{-\frac{(1)^2}{8}} e^{-\frac{(1)^2}{8}} \cdot 0.8} \approx 0.075 \end{aligned}$$

3. It’s Complicated

This probability table shows the joint distribution between two random variables: the year of the student at Stanford (Y) and their relationship status (R). The data was volunteered last year by over 200 anonymous students:

	Single	In a Relationship	It’s Complicated
Freshman	0.12	0.07	0.02
Sophomore	0.17	0.12	0.02
Junior	0.10	0.11	0.02
Senior	0.01	0.07	0.00
5+	0.04	0.10	0.03

- a. What is the marginal probability distribution for relationship status at Stanford (R)? Provide your result as a mapping between the values that R can take on and the corresponding probabilities.

For each assignment to R , sum over all the values that S can take on that are consistent with that assignment.

$$\begin{aligned} P(\text{Single}) &= 0.12 + 0.17 + 0.10 + 0.01 + 0.04 = 0.44 \\ P(\text{Relationship}) &= 0.07 + 0.12 + 0.11 + 0.07 + 0.10 = 0.47 \\ P(\text{Complicated}) &= 0.02 + 0.02 + 0.02 + 0.00 + 0.03 = 0.09 \end{aligned}$$

- b. What is the conditional probability of relationship status (R) given that a student is a Senior ($Y = \text{Senior}$)? Provide your result as a mapping between the values that R can take on and the corresponding probabilities.

$$P(\text{Single}|\text{Senior}) = \frac{P(\text{Single, Senior})}{P(\text{Senior})} = \frac{0.01}{0.08} = 0.125$$

$P(\text{Relationship}|\text{Senior}) = 0.88$ and $P(\text{Complicated}|\text{Senior}) = 0$ (same approach).

- c. What is the conditional probability that someone is “In a Relationship” given their year in school, $P(R = \text{In a Relationship}|Y)$? Give your answer as a mapping between the values that Y can take on and the corresponding probabilities.

$$P(\text{Relationship}|\text{Freshman}) = \frac{P(\text{Relationship, Freshman})}{P(\text{Freshman})} = \frac{0.07}{0.21} = 0.33$$

Same approach yields $P(\text{Relationship}|\text{Sophomore}) = 0.39$,
 $P(\text{Relationship}|\text{Junior}) = 0.48$, $P(\text{Relationship}|\text{Senior}) = 0.875$, and
 $P(\text{Relationship}|\text{5+}) = 0.59$

Midterm Prep Guiding Questions The midterm exam is coming up. Below are a few broad, guiding questions you might use to help solidify your thinking, prepare a study guide, etc. This is not meant to be a comprehensive list of exam topics; everything we’ve seen through Friday of week 5 is fair game for the exam. Please come to office hours if you’d like to discuss further!

1. **Counting** What is probability and what are the 3 axioms? What are event and sample spaces? What’s the significance of equally likely events in probability problem-solving? How do we reason differently about distinct vs. indistinct events? What’s the difference between combinations and permutations? What are the sum rule, product rule, inclusion-exclusion and pigeonhole principles, and when do we use them?
2. **Probability Rules** When do we use the definition of conditional probability, the chain rule, the law of total probability, Bayes’ theorem, the Complement Rule, DeMorgan’s law etc.? What are independence and conditional independence?
3. **Random Variables** What does randomness mean? What are expectation and variance, generally? What’s the difference between continuous and discrete random variables? We’ve seen lots of random variables - in which situations would each of them be appropriate? Which ones can be used to approximate others, and in which cases? What’s the difference between PMF, PDF, and CDF? What are multivariate distributions and conditional distributions, and how do we reason about them in both discrete and continuous situations?