Chris Piech
CS109

Section #4
October 24, 2018

Section #4: Joint Random Variables

---

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 λ =0.002 earthquakes per year.

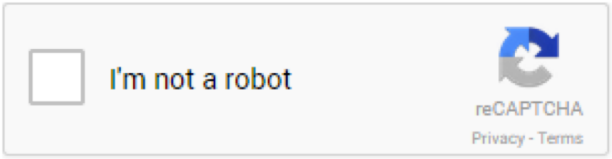
   a. What is the probability of no 8+ earthquakes in four years after the 1908 earthquake (recall that earthquakes are exponentially distributed)?

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

   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?

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

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 and the Y location distributed as normals with mean $\mu = 5$ and $\sigma^2 = 4$:

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

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

   c. The visitor clicks in the box at pixel ($x = 6$, $y = 6$). What is Google's new belief that the visitor is a robot?

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.

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

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.

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?