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 \( \lambda = 0.002 \) 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)?

   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 \( \lambda \)?

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, 10 pixels wide by 10 pixels tall. 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 \( D \) of a human click from the center of the box, in pixels, is distributed by a Rayleigh Distribution with parameter \( \theta = 2 \). A Rayleigh random variable is parameterized by a single scale parameter \( \theta \) and has the following probability density function and cumulative density function:

\[
\begin{align*}
  f_X(x) &= \begin{cases} 
    \frac{x}{\theta} e^{-x^2/2\theta} & x \geq 0 \\
    0 & else 
  \end{cases} \\
  F_X(x) &= \begin{cases} 
    1 - e^{-x^2/2\theta} & x \geq 0 \\
    0 & else 
  \end{cases}
\end{align*}
\]

\(^1\text{ReCaptcha uses more sophisticated statistics of natural human house gestures and clicks, but this problem covers the central idea behind the new click based reCaptcha. It was also a midterm question last Spring.}\)
a. What is 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 is the probability that a human clicks on a pixel that has a distance from the center of the box which is greater than or equal to 1.2 pixels?

c. The visitor clicks in the box at pixel $(x = 1.414, y = 1.414)$ which has a distance of 2 pixels from the center. 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:

<table>
<thead>
<tr>
<th></th>
<th>Single</th>
<th>In a Relationship</th>
<th>It’s Complicated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freshman</td>
<td>0.12</td>
<td>0.07</td>
<td>0.02</td>
</tr>
<tr>
<td>Sophomore</td>
<td>0.17</td>
<td>0.12</td>
<td>0.02</td>
</tr>
<tr>
<td>Junior</td>
<td>0.10</td>
<td>0.11</td>
<td>0.02</td>
</tr>
<tr>
<td>Senior</td>
<td>0.01</td>
<td>0.07</td>
<td>0.00</td>
</tr>
<tr>
<td>5+</td>
<td>0.04</td>
<td>0.10</td>
<td>0.03</td>
</tr>
</tbody>
</table>

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 = $In a Relationship|$Y$)? Give your answer as a mapping between the values that $Y$ can take on and the corresponding probabilities.