Chris Piech
CS 109

Section #7

Section 7 Solution

1. **Binary Tree**: Consider the following function for constructing binary trees:

```
def random_binary_tree(p):
   """
   Returns a dictionary representing a random binary tree structure.
   The dictionary can have two keys, "left" and "right".
   """
   if random_bernoulli(p):   # returns true with probability p
      new_node = {}
      new_node["left"] = random_binary_tree(p)
      new_node["right"] = random_binary_tree(p)
      return random_binary_tree
    else:
      return None
```

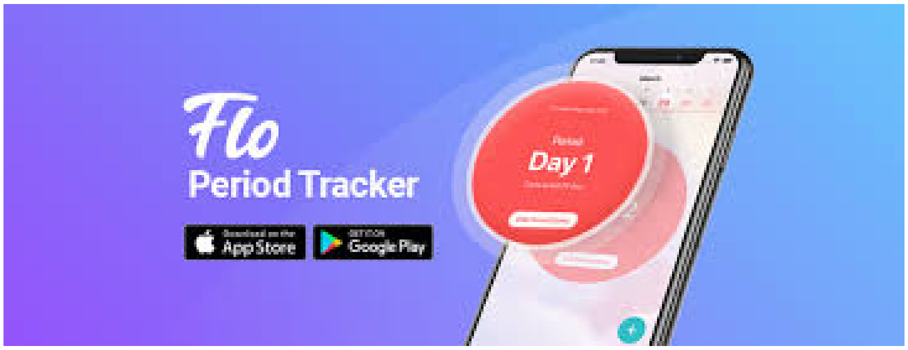
The `if` branch is taken with probability p (and the `else` branch with probability $1 - p$). A tree with no nodes is represented by `nullptr`; so a tree node with no left child has `nullptr` for the `left` field (and the same for the right child).

Let X be the number of nodes in a tree returned by `randomTree`. You can assume $0 < p < 0.5$. What is $E[X]$, in terms of p?

Let X_1 and X_2 be number of nodes the left and right calls to `randomTree`.
$E[X_1] = E[X_2] = E[X]$.

$$
\begin{aligned}
E[X] &= p \cdot E[X \mid \texttt{if}] + (1-p)E[X \mid \texttt{else}] \\
&= p \cdot E[1 + X_1 + X_2] + (1-p) \cdot 0 \\
&= p \cdot (1 + E[X] + E[X]) \\
&= p + 2pE[X] \\
(1-2p)E[X] &= p \\
E[X] &= \frac{p}{1-2p}
\end{aligned}
$$

2. **Flo. Tracking Menstrual Cycles**



Let X represent the length of a menstrual cycle: the number of days, as a continuous value, between the first moment of one period to the first moment of the next, for a given person. X is parameterized by α and β with probability density function:

$$f(X = x) = \beta \cdot (x - \alpha)^{\beta-1} \cdot e^{-(x-\alpha)^2}$$

a. For a particular person, $\alpha = 27$ and $\beta = 2$. Write a simplified version of the PDF of X.

$f(X = x) = 2 * (x - 27) * e^{-(x-27)^2}$

b. For a particular person, $\alpha = 27$ and $\beta = 2$. Write an expression for the probability that they have their period on day 29. In other words, what is the $P(29.0 < X < 30.0)$?

$$P(29.0 < X < 30.0) = \int_{29.0}^{30.0} 2 * (x - 27) * e^{-(x-27)^2}$$

Okay if expression inside integral is incorrect, as long as it's the same answer as part (a).

c. For a particular person, $\alpha = 27$ and $\beta = 2$. How many times more likely is their cycle to last **exactly** 28.0 days than exactly 29.0 days? You do not need to give a numeric answer. Simplify your expression.

$$\frac{f(X = 28)}{f(X = 29)} = \frac{2 * (28 - 27) * e^{-(28-27)^2}}{2 * (29 - 27) * e^{-(29-27)^2}} = \frac{e^3}{2}$$

d. A person has recorded their cycle length for 12 cycles stored in a list: $m = [29.0, 28.5, \ldots, 30.1]$ where m_i is the recorded cycle length for cycle i. Use MLE to estimate the parameter values α and β. Assume that cycle lengths are IID.

You don't need a closed form solution. Derive any necessary partial derivatives and write up to three sentences describing how a program can use the derivatives in order to chose the most likely parameter values.

Define our likelihood function:

$$L(\alpha, \beta) = \prod_{i=1}^{12} f(m_i)$$

Now log likelihood to make the math easier later:

$$LL(\alpha, \beta) = \sum_{i=1}^{12} \log f(m_i)$$

$$\alpha = \arg\max_{\alpha} LL(\alpha, \beta)$$

$$\beta = \arg\max_{\beta} LL(\alpha, \beta)$$

Log of the pdf simplifies:

$$\log f(m) = \log \beta + (\beta - 1)\log(m - \alpha) - (m - \alpha)^2$$

Now take partial derivative w.r.t α and β:

$$\frac{\partial}{\partial \alpha} LL(\alpha, \beta) = \sum_{i=1}^{12} 2(m_i - \alpha) - \frac{\beta - 1}{m_i - \alpha}$$

$$\frac{\partial}{\partial \beta} LL(\alpha, \beta) = \sum_{i=1}^{12} \frac{1}{\beta} + \log(m_i - \alpha)$$

we can use gradient ascent to maximize LL. This computes gradient w.r.t each parameter α, β then moves the parameters a small step in the direction of the gradient.

We also accept valid closed-form solutions. For example, can perform gradient descent on α, then update β by computing closed-form optimal value (given some value of α:

$$\beta = -\frac{12}{\sum_{i=1}^{12} \log(m_i - \alpha)}$$

Note: Flo is a real "AI based" app that helps people track their period lengths. The real world distribution of periods is thought to be a mixture distribution between a normal and a weibell distribution [1]. This problem only has you estimate parameters for a simplified Weibull [2].