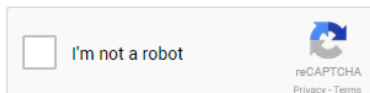


Section 5

Adapted for Winter 2020 by Alex Tsun

Congratulations on making it through the midterm! :)

1. **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 10 mm by 10 mm box. 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 top) distributed as independent normals both with mean $\mu = 5$ and variance $\sigma^2 = 4$.

- a. What is 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?
 - b. What is 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?
 - c. The visitor clicks in the box at $(x = 6$ mm, $y = 6$ mm). What is Google's new belief that the visitor is a robot?
2. **Binary Tree:** Consider the following function for constructing binary trees:

```

struct Node {
    Node *left;
    Node *right;
};

Node *randomTree(float p) {
    if (randomBool(p)) { // returns true with probability p
        Node *newNode = new Node;
        newNode->left = randomTree(p);
        newNode->right = randomTree(p);
        return newNode;
    } else {
        return nullptr;
    }
}

```

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 ?

3. **Hat-Check Again??** Recall the hat-check problem from section 2: n people go to a party and drop off their hats to a hat-check person. When the party is over, a different hat-check person is on duty, and returns the n hats randomly back to each person. Let X be the random variable representing the number of people who get their own hat back. We showed last time that $E[X] = 1$ for any n . What is $Var(X)$? Hint: Be careful when taking the variance of a sum of random variables.

4. **Timing Attack:**

In this problem we are going to show you how to crack a password in linear time, by measuring how long the password check takes to execute (see code below). Assume that our server takes T ms to execute any line in the code where $T \sim N(\mu = 5, \sigma^2 = 0.5)$ seconds. The amount of time taken to execute a line is always independent of other values of T .

```
# An insecure string comparison
def stringEquals(guess, password):
    nGuess = len(guess)
    nPassword = len(password)
    if nGuess != nPassword:
        return False # 4 lines executed to get here
    for i in range(nGuess):
        if guess[i] != password[i]:
            return False # 6 + 2i lines executed to get here
    return True # 5 + 2n lines executed to get here
```

On our site all passwords are length 5 through 10 (inclusive) and are composed of lower case letters only. A hacker is trying to crack the root password which is “gobayes” by carefully measuring how long we take to tell them that her guesses are incorrect.

- a. What is the distribution of time that it takes our server to execute k lines of code? Recall that each line independently takes $T \sim N(\mu = 5, \sigma^2 = 0.5)$ ms.
- b. First the hacker needs to find out the length of the password. What is the probability that the time taken to test a guess of correct length (server executes 6 lines) is longer than the time taken to test a guess of an incorrect length (server executes 4 lines)? Assume that the first letter of the guess does not match the first letter of the password. Hint: $P(A > B)$ is the same as $P(A - B > 0)$.

- c. Now that our hacker knows the length of the password, to get the actual string she is going to try and figure out each letter one at a time, starting with the first letter. The hacker tries the string "aaaaaaa" and it takes 27s. Based on this timing, how much more probable is it that first character did not match (server executes 6 lines) than the first character did match (server executes 8 lines)? Assume that all letters in the alphabet are equally likely to be the first letter.
- d. If it takes the hacker 6 guesses to find the length of the password, and 26 guesses per letter to crack the password string, how many attempts does she need to crack our password, "gobayes"? Yikes!