Chris Piech
CS 109

Section #6
Nov 8, 2017

Section 6: Samples

Thanks to Will Monroe for the Binary Tree question

1. **Warmup**:

   What is the difference between the population variance, σ^2, and sample variance, S^2? What is the difference between sample variance, S^2, and variance of the sample mean, $\text{Var}(\bar{X})$?

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. **Beta Sum**:

   What is the distribution of the sum of 100 IID Betas? Let X be the sum

$$X = \sum_{i=0}^{100} X_i \qquad \text{Where each } X_i \sim \text{Beta}(a = 3, b = 4)$$

   Either simulate the summation 10,000 times or use theory. Note the variance of a Beta:

$$\text{Var}(X_i) = \frac{ab}{(a+b)^2(a+b+1)} \qquad \text{Where } X_i \sim \text{Beta}(a, b)$$

4. **Variance of Height among Island Corgis**:

   A colleague has collected samples of heights of corgis that live on two different islands. The colleague collects 50 samples from both islands.

   The colleague notes that the sample mean is the same between the two groups: both are around 10 inches. However, island B has a sample **variance** that is 3 in^2 greater than island A. The colleague wants to make a scientific claim that corgis on island A have a significantly higher spread of heights than corgis on island B. You are skeptical. It is possible that heights are identically distributed across both islands and that the observed difference in variance was a result of chance and a small sample size (the null hypothesis).

   Calculate the probability of the null hypothesis using bootstrapping. Here is the data. Each number is the height, in inches, of an independently sampled corgi:

   Island A Corgi Heights (S^2 = 6.0):
   13, 12, 7, 16, 9, 11, 7, 10, 9, 8, 9, 7, 16, 7, 9, 8, 13, 10, 11, 9, 13, 13, 10, 10, 9, 7, 7, 6, 7, 8, 12, 13, 9, 6, 9, 11, 10, 8, 12, 10, 9, 10, 8, 14, 13, 13, 10, 11, 12, 9

   Island B Corgi Heights (S^2 = 9.1):
   8, 8, 16, 16, 9, 13, 14, 13, 10, 12, 10, 6, 14, 8, 13, 14, 7, 13, 7, 8, 4, 11, 7, 12, 8, 9, 12, 8, 11, 10, 12, 6, 10, 15, 11, 12, 3, 8, 11, 10, 10, 8, 12, 8, 11, 6, 7, 10, 8, 5

   Discuss: How would this calculation be different if you were interested in looking at the statistical significance of: sample mean? 95th percentile?