Chris Piech
CS 109

Section #6
Feb 20, 2025

Section 6

With questions from Will Monroe and Julia Daniel

1 Beta Sum Warmup

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

$$X = \sum_{i=1}^{100} X_i \quad \text{where each } X_i \sim \text{Beta}(a = 3, b = 4)$$

Note the expectation and variance of a Beta:

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

By the Central Limit Theorem, the sum of equally weighted IID random variables will be Normally distributed. We calculate the expectation and variance of X_i using the Beta formulas:

$$\begin{aligned}
E(X_i) &= \frac{a}{a+b} && \text{Expectation of a Beta} \\
&= \frac{3}{7} \approx 0.43 \\
\text{Var}(X_i) &= \frac{ab}{(a+b)^2(a+b+1)} && \text{Variance of a Beta} \\
&= \frac{3 \cdot 4}{(3+4)^2(3+4+1)} \\
&= \frac{12}{49 \cdot 8} \approx 0.03
\end{aligned}$$

$$\begin{aligned}
X &\sim N(\mu = n \cdot E[X_i], \sigma^2 = n \cdot \text{Var}(X_i)) \\
&\sim N(\mu = 43, \sigma^2 = 3)
\end{aligned}$$

2 Food for Thought

Karel the dog eats an unpredictable amount of food. Every day, the dog is equally likely to eat an amount in the continuous range between 100g and 300g. How much Karel eats one day is independent of all other days.

You only have 6500g of food. What is the probability that 6500g will be enough for the next 30 days?

The distribution of the sum is given by the central limit theorem. Let $X_i \sim \text{Uni}(100, 300)$ where $E[X_i] = 200$ and $Var(X_i) = \frac{1}{12}(200)^2 \approx 3333$.

$$Y = \sum_i X_i$$

Let's approximate Y with a Normal:

$$Y \sim \mathcal{N}(6000, 316.212^2)$$

$$P(Y < 6500)$$

$$P\left(\frac{Y - 6000}{316.212} < \frac{6500 - 6000}{316.212}\right)$$

Let $\frac{Y-6000}{316.212} = Z \sim \mathcal{N}(0, 1)$

$$P\left(Z < \frac{6500 - 6000}{316.212}\right)$$

$$P\,(Z < 1.58)$$

$$\Phi\,(1.58)$$

3 Sorted Random Values

We want to reason about the values produced by the following python code, which samples 10 random uniform values in the range [0, 1] and then sorts them from low to high:

```
# generate 10 random uniform values
values = []
for i in range(10):
   value_i = random_uniform(0,1)  # sample from standard uniform
   values.append(value_i)

# sort all of the values ascending from low to high
sorted_values = sorted(values)
print(sorted_values)
```

Here is what the list could look like when printed. For clarity, each value is rounded to two decimal places, but these values really have near-infinite precision:

```
sorted_values = [0.03,0.13,0.45,0.51,0.52,0.63,0.69,0.82,0.88,0.91]
        index :   0    1    2    3    4    5    6    7    8    9
```

a. The first value produced by `random_uniform(0,1)` is 0.4. What is the probability it ends up at index 4 in the sorted list? *In other words: What is the probability that exactly 4 out of the 9 other values are less than 0.4?*

Let $X \sim \text{Uni}(0, 1)$ be any value produced by `random_uniform`. $P(X < 0.4) = 0.4$, by the CDF of the Uniform (intuitively, it is the fraction of the range that satisfies the event - for a range from 0 to 1, 40% of that range is less than 0.4).

Then, we need to determine the probability that out of 9 values, exactly 4 are less than 0.4. Since each of the 9 values has a consistent probability 0.4 of "succeeding" at being less than 0.4, we can apply the Binomial.

Let Y be the number of values less than 0.4. $Y \sim \text{Bin}(n = 9, p = 0.4)$.

$$P(Y = 4) = \binom{9}{4} 0.4^4 (1 - 0.4)^5$$

b. Let X_4 be the value at index 4 in the sorted list. What is the probability density of $X_4 = x$?
In other words: What is the probability density that $X_4 = x$, given that you know exactly 4 out of 9 numbers are less than x?

We can set up this problem using inference, since we want a probability distribution, given that we have an observation to condition on. Let Y again be the number of values less than x. Using Bayes' Theorem:

$$f(X = x | Y = 4) = \frac{P(Y = 4 | X = x) f(X = x)}{P(Y = 4)}$$

We can generalize part A to write a general expression for $P(Y = 4 | X = x)$.

We also already know that $f(X = x) = 1$ since $X \sim \text{Uni}(0, 1)$.

If we recall that the denominator will be constant that does not change for different values of X, we can just represent it as $\frac{1}{K}$ and know that we would need to normalize the distribution to sum to 1 later.

Plugging everything in:

$$f(X = x|Y = 4) = K \cdot \binom{9}{4} x^4 (1 - x)^5 \cdot 1 = K \cdot x^4 (1 - x)^5$$

This is recognizable as the PDF of the Beta distribution! Specifically, $\text{Beta}(a = 4+1, b = 5+1)$.