

## Section 5

---

### 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:

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

### 2 Microcontroller Precision Error

A low-powered microcontroller can only represent numbers to 3 decimal places.

Let  $X_i$  be an i.i.d. sample from Uniform(0, 1) with infinite precision. Let  $Y_i$  be the representation of  $X_i$  on the microcontroller.  $Y_i$  is equal to  $X_i$  truncated to 3 decimal places. Here are three examples:

$X_i$	$Y_i$
0.12340809 ...	0.123
0.28374110 ...	0.283
0.55555555 ...	0.555

- a (5 points) What is the distribution of  $(X_i - Y_i)$ ? In other words, what is the distribution for the error when a single uniformly sampled value is represented in the microcontroller?
- b (10 points) We sample 1000 values, where  $X_i$  is the precise value and  $Y_i$  is the truncated value of  $X_i$ :

$$X = \sum_{i=1}^{1000} X_i \quad Y = \sum_{i=1}^{1000} Y_i$$

What is the probability that the difference between  $X$  and  $Y$  is more than 0.51?

### 3 Song of the Quarter

This quarter in CS109 there were 167 songs that were voted on. For each song, we have a list of votes where each vote is an integer in the set  $\{1, 2, 3, 4, 5\}$ . We assume all votes for a song are IID samples from the “true” distribution of CS109 opinion on the song.

For each song  $i$  we have  $m_i$  votes stored in a list  $\text{votes}[i] = [x_1, x_2, \dots, x_{m_i}]$ . We have already calculated:

$$\mu_i = \frac{1}{m_i} \sum_{j=1}^{m_i} x_j \quad \text{using } \text{np.mean}(\text{votes}[i])$$

$$\text{var}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} (x_j - \mu_i)^2 \quad \text{using } \text{np.var}(\text{votes}[i])$$

$$\text{svar}_i = \frac{1}{m_i - 1} \sum_{j=1}^{m_i} (x_j - \mu_i)^2 \quad \text{using } \text{np.var}(\text{votes}[i], \text{ddof}=1)$$

a (7 points) Song 1 has  $m_1 = 45$  votes. We have calculated:

$$\mu_1 = 3.82 \quad \text{var}_1 = 1.4 \quad \text{svar}_1 = 1.5$$

Estimate the probability that the true average rating for song 1 is less than 3.

b (8 points) Song 1 has  $m_1 = 45$  votes. Song 2 has  $m_2 = 36$  votes. We have calculated:

$$\begin{aligned} \text{Song 1: } & \mu_1 = 3.82 \quad \text{var}_1 = 1.4 \quad \text{svar}_1 = 1.5 \\ \text{Song 2: } & \mu_2 = 3.79 \quad \text{var}_2 = 1.7 \quad \text{svar}_2 = 1.8 \end{aligned}$$

What is the probability that the true average of Song 1 is greater than the true average for Song 2?