Jerry Cain
CS 109 May 18, 2022

Section 7 Solutions

Based on the work of many CS109 instructors and course staff members.

1 Warmups

1.1 Sample and Population Mean

Computing the sample mean is similar to the population mean: sum all available points and divide by the number of points. However, sample variance is slightly different from population variance.

1. Consider the equation for population variance, and an analogous equation for sample variance.

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \quad (1)$$

$$S^2_{biased} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \quad (2)$$

   S^2_{biased} is a random variable to estimate the constant σ^2. Because it is biased, $E[S^2_{biased}] \neq \sigma^2$. Is $E[S^2_{biased}]$ greater or less than σ^2?

2. Consider an alternative Random Variable, $S^2_{unbiased}$ (known simply as S^2 in class). The technique of unbiasing variance is known as *Bessel's correction*. Write the $S^2_{unbiased}$ equation.

---

1. $E[S^2_{biased}] < \sigma^2$. The intuition is that the spread of a sample of points is generally smaller than the spread of all the points considered together. This becomes more clear when we consider the unbiased version and how it makes the expression evaluate to a larger number.
2. $S^2_{unbiased} = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

1.2 MLE

Suppose $x_1, \ldots, x_n$ are i.i.d. (independent and identically distributed) values sampled from some distribution with density function $f(x|\theta)$, where θ is unknown. Recall that the likelihood of the data is

$$L(\theta) = f(x_1, x_2, \ldots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

Recall we solve an optimization problem to find $\hat{\theta}$ which maximizes $L(\theta)$, i.e., $\hat{\theta} = \arg\max_\theta L(\theta)$.

1. Write an expression for the log-likelihood, $LL(\theta) = \log L(\theta)$.

2. Why *can* we optimize $LL(\theta)$ rather than $L(\theta)$?

3. Why *ought* we optimize $LL(\theta)$ rather than $L(\theta)$?

1. $\sum_{i=1}^{n} \log f(x_i|\theta)$

2. Logarithms are monotonic. For any monotonic function f and any function g, the following holds: $\arg\max g(x) = \arg\max f(g(x))$

3. Finding the max of a function requires taking derivatives. The original expression consists of a product of many functions of θ. This would lead to a difficult derivation because of the chain rule of calculus.

   By taking the log, we convert the product into sums, and the derivative of the sum of many functions of θ is much easier to compute.

   That is, $\frac{\partial}{\partial\theta} \sum_i f_i(\theta) = \sum_i \frac{\partial}{\partial\theta} f_i(\theta)$ which is easy.

   However, $\frac{\partial}{\partial\theta} \prod_i f_i(\theta)$ = a complicated expression

1.3 Beta

1. Suppose you have a coin where you have no prior belief on its true probability of heads p. How can you model this belief as a Beta distribution?

2. Suppose you have a coin which you believe is fair, with "strength" α. That is, pretend you've seen α heads and α tails. How can you model this belief as a Beta distribution?

3. Now suppose you take the coin from the previous part and flip it 10 times. You see 8 heads and 2 tails. How can you model your posterior belief of the coin's probability of heads?

1. $\text{Beta}(1, 1)$ is a uniform prior, meaning that prior to seeing the experiment, all probabilities of heads are equally likely.

2. $\text{Beta}(\alpha + 1, \alpha + 1)$. This is our prior belief about the distribution.

3. $\text{Beta}(\alpha + 9, \alpha + 3)$

2 Problems

2.1 Variance of Hemoglobin Levels

A medical researcher treats patients with dangerously low hemoglobin levels. She has formulated two slightly different drugs and is now testing them on patients. First, she administered drug A to one group of 50 patients and drug B to a separate group of 50 patients. Then, she measured all the patients' hemoglobin levels post-treatment. For simplicity, assume that all variation in the patient outcomes is due to their different reactions to treatment.

The researcher notes that the sample mean is similar between the two groups: both have mean hemoglobin levels around 10g/dL. However, drug B's group has a **sample variance** that is 3 (g/dL)^2 **greater** than drug A's group. The researcher thinks that patients respond to drugs A and B differently. Specifically, she wants to make the scientific claim that drug A's patients will end up with a significantly different spread of hemoglobin levels compared to drug B's.

You are skeptical. It is possible that the two drugs have practically identical effects and that the observed different in variance was a result of chance and a small sample size, i.e. the **null hypothesis**. Calculate the probability of the null hypothesis using bootstrapping. Here is the data. Each number is the level of an independently sampled patient:

Hemoglobin Levels of Drug A's Group ($S^2 = 6.0$): 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

Hemoglobin Levels of Drug B's Group ($S^2 = 9.1$): 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

Complete the prompts in the following Colab notebook to investigate this question using bootstrapping.

`https://colab.research.google.com/drive/1l4Dokxjt9Y0_9_I8ZjYPuExXx7fsc1Tw?usp=sharing`

```
https://colab.research.google.com/drive/15W4ZtyY_nEwiMqlsBODND7GQcjItZHRL?
usp=sharing
```

2.2 Beta Sum

What is the distribution of the sum of 100 i.i.d. Betas? Let X be the sum

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

Note the variance of a Beta:

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

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:

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

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