CS109 May 16, 2024

Continuous Joint Distributions, Central Limit Theorem

Before you leave lab, make sure you click here so that you're marked as having attended this week's section. The CA leading your discussion section can enter the password needed once you've submitted.

1 Warmups

1.1 Food for Thought

Karel the dog eats an unpredictable amount of food. Every day, the dog is equally likely to eat between a continuous amount in the range 100 to 300g. How much Karel eats is independent of all other days. You only have 6.5kg of food for the next 30 days. What is the probability that 6.5kg will be enough for the next 30 days?

1.2 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.

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

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

   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 un-biasing variance is known as *Bessel's correction*. Write the $S^2_{unbiased}$ equation.

2 Problems

2.1 Sum of Two Exponentials

Consider two independent random variables X and Y, each Exponentials with different parameters—specifically, let $X \sim Exp(\frac{1}{2})$ and $Y \sim Exp(\frac{1}{3})$. Assuming $T = X + Y$, derive and present the probability density function $f_T(t)$ by evaluating the relevant convolution. Once you arrive at your $f_T(t)$, verify your answer by calculating $f_T(2)$ out to three decimal places.

2.2 *Grading Exams*

Jacob and Kathleen are planning to grade Problem 1 on your Week 7 exam, and they'll each grade their half independently of the other. Jacob takes $X \sim Exp(\frac{1}{3})$ hours to finish his half while Kathleen takes $Y \sim Exp(\frac{1}{4})$ hours to finish his half.

a. Find the CDF of X/Y, which is the ratio of their grading completion times.

b. What is the probability that Kathleen finishes before Jacob does?

2.3 *Central Limit Theorem and Sampling Calisthenics*

a. Let X_1, X_2, X_3, ..., X_{1000} be iid—that is, independent and identically distributed—such that $X_i \sim \text{NegBin}(r = 10, p = 0.5)$, and let $W = X_1 + X_2 + \ldots + X_{1000}$. According to the Central Limit Theorem, what distribution does W assume, and what are its parameters?

b. Define $\bar{X} = \frac{1}{1000} \sum_{i=1}^{1000} X_i$ to be the sample mean of our 1000 iid samples. What is the standard deviation of the random variable $\bar{X}$?

c. You compute the variance of your 1000 samples, X_1, X_2, X_3, ..., X_{1000} according to the traditional definition of variance—i.e. $\frac{1}{1000} \sum_{i=1}^{1000} (X_i - \bar{X})^2$. Do you expect this variance to, more often than not, be larger, equal to, or smaller than the variance of $\text{NegBin}(10, 0.5)$. Explain your answer.

d. The number of samples needed for the Central Limit Theorem to apply is generally understood to be 30 or more. However, the Central Limit Theorem works well for an even smaller number of samples when $X_i \sim \text{Bin}(10, 0.5)$ than is does when $X_i \sim \text{NegBin}(10, 0.5)$. Briefly explain why.

e. Recall that sampling theory allows a reasonably large sample to stand in for the true population distribution. When resampling from the sample for bootstrapping purposes, we generally do so **with** replacement. Why **with** replacement instead of **without**?