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?

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 R.V.

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

$$P(Y < 6500)$$

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

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

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

$$P\left(Z < 1.58\right)$$

$$\Phi\left(1.58\right)$$

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.

a. $E[S^2_{\text{biased}}] < \sigma^2$. 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.

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

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.

If we let X and Y be continuous random variables with probability density functions $f_X(t)$ and $f_Y(t)$, then the probability density function and $f_T(t)$ of $T = X + Y$ is the convolution of $f_X(t)$ and $f_Y(t)$—that is:

$$f_T(t) = \int_{-\infty}^{\infty} f_X(x)f_Y(t-x)dx$$

In the case of this problem, both X and Y are Exponentials with supports of all nonnegative real numbers, so the bounds of the integral can be compressed to include just those values where both x and $t - x$ are greater than 0 (which is $[0, t]$). Of course, the density function for a general Exponential is $f_X(x) = \lambda e^{-\lambda x}$, so the convolution to be evaluated should be:

$$\begin{aligned}
f_T(t) &= \int_0^t f_X(x) f_Y(t-x)dx \\
&= \int_0^t \frac{1}{2} e^{-\frac{1}{2}x} \cdot \frac{1}{3} e^{-\frac{1}{3}(t-x)} dx \\
&= \frac{1}{6} e^{-\frac{1}{3}t} \int_0^t e^{-\frac{1}{6}x} dx \\
&= -e^{-\frac{1}{3}t} e^{-\frac{1}{6}x} \Big|_0^t \\
&= e^{-\frac{1}{3}t}(1 - e^{-\frac{1}{6}t}) \\
&= e^{-\frac{1}{3}t} - e^{-\frac{1}{2}t}
\end{aligned}$$

That's the probability density function of interest, and its value at $t = 2$ is $f_T(2) = 0.145537$.

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.

The random variable of interest is the ratio X/Y, so the CDF, $F(r)$, in this case would be $P(\frac{X}{Y} < r)$, where r stands for **r**atio and ranges from 0 to ∞. Rearranging, we are interested in computing $P(X < rY)$, which can be computed in terms of the PDFs for X and Y:

$$
\begin{aligned}
F(r) = P(\frac{X}{Y} < r) &= P(X < rY) \\
&= \int_0^\infty \int_0^{ry} \frac{1}{12} e^{-\frac{1}{3}x} e^{-\frac{1}{4}y} dxdy \\
&= \frac{1}{12} \int_0^\infty e^{-\frac{1}{4}y} \int_0^{ry} e^{-\frac{1}{3}x} dxdy \\
&= -\frac{1}{4} \int_0^\infty e^{-\frac{1}{4}y} \left(e^{-\frac{1}{3}x} \right) \Big|_0^{ry} dy \\
&= -\frac{1}{4} \int_0^\infty e^{-\frac{1}{4}y} \left(e^{-\frac{1}{3}ry} - 1 \right) dy \\
&= \frac{1}{4} \int_0^\infty e^{-\frac{1}{4}y} dy - \frac{1}{4} \int_0^\infty e^{-(\frac{1}{3}r + \frac{1}{4})y} dy \\
&= 1 + \frac{\frac{1}{4}}{\frac{1}{3}r + \frac{1}{4}} e^{-(\frac{1}{3}r+\frac{1}{4})y} \Big|_0^\infty \\
&= 1 - \frac{\frac{1}{4}}{\frac{1}{3}r + \frac{1}{4}} = \frac{\frac{1}{3}r + \frac{1}{4}}{\frac{1}{3}r + \frac{1}{4}} - \frac{\frac{1}{4}}{\frac{1}{3}r + \frac{1}{4}} \\
&= \frac{\frac{1}{3}r}{\frac{1}{3}r + \frac{1}{4}}
\end{aligned}
$$

For those question why that first of two integrals vanished to 1, note that the integrand is just the PDF of Expo$(\lambda = \frac{1}{4})$!

Incidentally, we can compute the probability density function from the CDF by differentiating with respect to r:

$$
\begin{aligned}
f(r) &= \frac{dF(r)}{dr} \\
&= \frac{d}{dr} \frac{\frac{1}{3}r}{\frac{1}{3}r + \frac{1}{4}} \\
&= \frac{1}{12(\frac{1}{3}r + \frac{1}{4})^2}
\end{aligned}
$$

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

In comparison, that is delightfully straightforward, because we get to plug $r = 1$ into our result from part a. $P(X < Y) = \frac{1}{3} \cdot \frac{12}{7} = \frac{4}{7}$. That, however, is the probability that Jacob finishing before Kathleen, and we want to opposite. Therefore, the probability of interest is really $\frac{3}{7}$. Given the expected completion times of 3 and 4 hours for Jacob and Kathleen, respectively, this seems right.

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?

This is classic Central Limit Theorem where the distribution of the sum is a Gaussian with mean $1000\,E[X_i]$ and variance $1000\,Var(X_i)$. The formulas for a Negative Binomial's mean and variance are well-defined and are computed as:

$$E[X_i] = \frac{r}{p} = \frac{10}{0.5} = 20$$
$$Var(X_i) = \frac{r(1-p)}{p^2} = \frac{10 \cdot 0.5}{0.5^2} = 20$$

How neat is it that the mean and variance are the same? This all means that $W \sim \mathcal{N}(20000, 20000)$.

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 $\bar{X}$?

The Central Limit Theorem has a lot to say about the distribution of sample means as well. In particular, for this problem, $\bar{X} \sim \mathcal{N}(E[X_i], \frac{Var[X_i]}{1000})$. That's more than we're asking—all I need from you is that $Var(X_i) = \frac{20}{1000} = 0.02$.

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.

Here the population variance is 20, since we know the population distribution is the Negative Binomial. Recall that the unbiased sample variance divides the sum of the differences squared by $n - 1$, or 999. The traditionally computed variance divides by a slightly larger number of 1000, so we expect the traditionally computed variance to, more often than not, be a little too low.

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.

The simple answer is that $X_i \sim \text{Bin}(10, 0.5)$ is symmetric, so there are no asymmetries to overcome as you add samples together. In fact, you can view, say, 3 $\text{Bin}(10, 0.5)$ as 30 $\text{Ber}(0.5)$.

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**?

We sample with replacement because we treat the original set of samples as a probability mass function. If we were to sample without replacement, we're incapable of creating resamples of a size larger than the original sample, and when the size of the resample is close to the size of the original, each resample would essentially be a replica of the original.