

EEL 4930 Stats – Lecture 27

- **Review of in-class assignment from last lecture**

Recall that if x_i are samples drawn from a random variable X , then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n x_i = E[X].$$

Create a Uniform random variable object using `scipy.stats`. Draw 10,000 sample values from it, and use the sample values to estimate $(E[U])^2$ and $E[U^2]$.

Analytical Solution:

- Find the value c that minimizes the expected mean-square error to a random variable X , $E[(X - c)^2]$

THE SAMPLE MEAN ESTIMATOR

- Let $X_i, i = 0, 1, \dots, N - 1$ be a finite sequence of independent random variables drawn from the same distribution, which has some finite mean μ . If we wish to estimate the mean of the underlying distribution, we can do it as in our last simulation:

$$\hat{\mu} = \frac{1}{N} \sum_{i=0}^{N-1} X_i$$

- This is called the **sample mean estimator**
- Note that the sample mean estimator is a *linear combination* of random variables. Thus,
 1. the sample mean estimator is a random variables
 2. we can calculate it's expected value!

MOMENTS

- Moments of a random variable are expected values of the random variable raised to some power
- For a central moment, the mean is subtracted from the random variable before it is raised to a power
- Because different powers spread the values of the random variable in different ways, moments can provide additional information about a random variable than the mean
 - Variance is the second central moment and provides a measure of how much the probability density or mass of random variable is spread away from the mean

- Some common moments (expected values):

- n th moment of X :

$$E[X^n] =$$

- n th central moment of X :

$$E[(X - \mu_X)^n] =$$

where $\mu_X = E[X]$.

- Variance of X is 2nd central moment:

$$\text{Var}[X] = E[(X - \mu_X)^2]$$

$$=$$

$$=$$

$$=$$

(this latter formula is usually a more convenient way to find the variance.)

- The variance of a Gaussian random variable is the parameter σ^2 (you can get it through integration by parts or some clever manipulation)

PROPERTIES OF VARIANCE:

1. $\text{Var}[c] =$

2. $\text{Var}[X + c] =$

3. $\text{Var}[cX] =$

HYPOTHESIS TESTS INVOLVING SAMPLE MEAN WITH KNOWN VARIANCE

- Suppose we have two data sets that we think can be modeled as coming from Gaussian distributions
- We compute the sample means by averaging the data sets
- If we observe a difference in the sample means for the two data sets, how can we determine analytically if the ensemble means are different?
- Let's assume that distributions for the two data sets have a common variance **and we know the variance**, σ^2
- We need to know a few more facts about sums of independent Gaussian random variables:
 1. If X and Y are independent RVs, then

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$$

2. If X and Y are independent Gaussian random variables with

$$X \sim \text{Gaussian}(\mu_X, \sigma_X^2)$$

and

$$Y \sim \text{Gaussian}(\mu_Y, \sigma_Y^2),$$

then

$$Z = X + Y \sim \text{Gaussian}(\mu_Z, \sigma_Z^2).$$

By linearity $\mu_Z = \mu_X + \mu_Y$, and by the previous property $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$.

3. If Z is a random variable, $aZ + b$ is also a Gaussian random variable¹
- Let's start by considering the statistics of a single sample mean,

$$\hat{\mu}_X = \frac{1}{m} \sum_{i=1}^m X_i$$

¹All these things are proved in EEE 5544.

- By the properties above, we can see:

1. $\hat{\mu}_X$ is a Gaussian random variable
2. The mean of $\hat{\mu}_X$ is $E[\hat{\mu}_X] = \mu_X$
3. The variance of $\hat{\mu}_X$ is

$$\begin{aligned}\text{Var}\left[\frac{1}{m}\sum_{i=0}^{m-1}X_i\right] &= \frac{1}{m^2}\text{Var}\left[\sum_{i=0}^{m-1}X_i\right] \\ &= \frac{1}{m^2}\sum_{i=0}^{m-1}\text{Var}[X_i] \\ &= \frac{1}{m^2}\sum_{i=0}^{m-1}\sigma_X^2 = \frac{\sigma_X^2}{m}\end{aligned}$$

(Note that the variance of the sample mean decreases linearly with the number of samples. This can be used to show that the sample mean converges to the true mean if the variance of the original random variable is finite.)



EX

The city of Gainesville claims the mean commute time on SW 24th Ave from I-75 to UF is 23 minutes with a variance of 50. I traveled that route 10 times over the last two weeks and had an average commute time of 27 minutes. Conduct a hypothesis test to determine whether the City of Gainesville's model is reasonable. Reject the null hypothesis if $p < 0.01$.

