Lecture 19

niceguy

April 15, 2023

1 Sampling

We cannot measure the whole population, so we will have to make do with a subset. So we take samples randomly.

Definition 1.1. Sample refers to the data $x_1, \ldots, x_n$ collected, with random variables $X_1, \ldots, X_n$.

Usually, these data are **independent identically distributed** (IID). Suppose each X has mean μ and variance σ^2.

1.1 Sample Mean

The empirical mean is

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

And similarly, the random variable $\overline{X}$ is defined as

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and obviously

$$E[X] = \mu$$

1.2 Sample Variance

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

We want that $E[S^2] = \sigma^2$.

$$\begin{aligned}
E[S^2] &= E\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i^2 - 2X_i\overline{X} + \overline{X}^2\right)\right] \\
&= \frac{1}{n-1}E\left[\left(\sum_{i=1}^{n}X_i^2\right) - n\overline{X}^2\right] \\
&= \frac{1}{n-1}\left(n\mu^2 + n\sigma^2 - n\mu^2 - n\mathrm{var}(\overline{X})\right) \\
&= \frac{1}{n-1}\left(n\sigma^2 - n\mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)\right) \\
&= \frac{1}{n-1}\left(n\sigma^2 - \frac{1}{n}\sum_{i=1}^{n}\mathrm{var}(X_i)\right) \\
&= \frac{1}{n-1}(n\sigma^2 - \sigma^2) \\
&= \sigma^2
\end{aligned}$$

Note that we used

$$E(X_i - \mu)^2 = E(X_i^2) - \mu^2 \Rightarrow E(X_i^2) = E(X_i - \mu)^2 + \mu^2 = \mathrm{var}(X_i) + \mu^2$$

and

$$E(X_i - \mu)^2 = \sigma^2$$

The latter is because the standard deviation is the average value of $(X_i - \mu)^2$. Finally,

$$\mathrm{var}(X + Y) = \mathrm{var}(X) + \mathrm{var}(Y)$$

if X and Y are independent. This can be proven using algebra, but it is annoying.