ECE 3530/CS 3130: Engineering Probability & Stats
Lecture Notes Spring 2015

Lecture #13

1 Expectation

We've been talking about expectation as an extension of the concept of averaging. For example, let's say you have n samples of a random variable X, called $x_1, x_2, \ldots, x_n$. You can average them and compute

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

As shown in Figure 1(a).

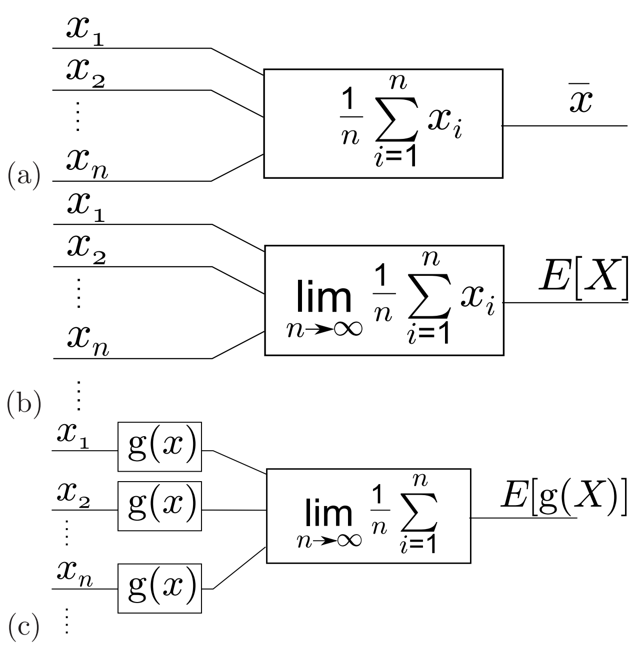

Figure 1: Block diagram representation of (a) averaging, (b) expectation, and (c) expectation of a function.

Intuitively, we want to find a way to come up with the average we'd have if we had an infinite number of samples of the random variable. We will call this expectation rather than the mean, because we will never compute exactly this number with a finite number of samples. In a sense, it is a mathematical limit of the average as $n \to \infty$ (although we are not providing enough theory to back this statement up).

See a problem? We don't ever as engineers have an infinite amount of time or resources to be able to make an arbitrary number of measurements in order to find the expected value this way.

We write the expected value of X as $E[X]$. The point of lecture 12 was that one may compute this from the distribution functions. For a continuous r.v.:

$$E[X] = \int_{x \in S_X} x f_X(x) dx$$

and for a discrete random variable:

$$E[X] = \sum_{x \in S_X} x f_X(x)$$

We also denote the expected value of X as μ_X.

Six notes:

1. The expected value is non-random! More specifically, once you take the expected value w.r.t. X, you will no longer have a function of X.

2. By analogy to statics, the expected value is where on the x-axis you'd need to put your finger to "balance" the mass or density on the pmf or pdf plot.

3. Note we sometimes write $E[X]$ as $E_X[X]$, where the subscript X indicates which r.v. we're taking the expectation of. When there's only one thing in the expectation operator, there's no ambiguity, so no need to write the X in the subscript. When there's any ambiguity, I try to write the subscript. The Walpole book never writes the subscript (as far as I've seen).

4. The Walpole book typically writes $\mathrm{E}(\cdot)$, that is, using parentheses, but sometimes they write $\mathrm{E}[\cdot]$ or $\mathrm{E}\{\cdot\}$. I try to be consistent and use square brackets every time.

5. **The expected value is a linear operator.** Consider $g(X) = aX + b$ for $a, b \in \mathbb{R}$. Then

$$\begin{aligned}
E_X[g(X)] &= E_X[aX + b] \\
&= \sum_{x \in S_X} (ax + b) f_X(x) \\
&= \sum_{x \in S_X} [ax f_X(x) + b f_X(x)] \\
&= \sum_{x \in S_X} ax f_X(x) + \sum_{x \in S_X} b f_X(x) \\
&= a \sum_{x \in S_X} x f_X(x) + b \sum_{x \in S_X} f_X(x) \\
&= a E_X[X] + b
\end{aligned}$$

6. The expected value of a constant is a constant: $E_X[b] = b$.

1.1 Expectation of a Function

In general, we can average any function of a random variable. For example, if X is a voltage, then X^2/R is the power dissipated across resistance R. Perhaps we are interested in the power, but are making a measurement of voltage. No problem, we can average power:

$$\bar{g(x)} = \frac{1}{n}\sum_{i=1}^{n} g(x_i)$$

where

$$g(x) = x^2/R$$

In this case, if we have the distribution function for X, we could calculate the expected value of power. We'd denote this $E[g(x)]$, which is defined as follows. For a continuous r.v.:

$$E[g(X)] = \int_{x \in S_X} g(x) f_X(x) dx$$

If this had been an example where X was a discrete random variable:

$$E[g(X)] = \sum_{x \in S_X} g(x) f_X(x)$$

Now, we used $g(x) = x^2/R$ here, but any function may be used. Let's consider some common functions $g(X)$:

1. Let $g(X) = X$. We've already done this! The mean $\mu_X = E_X[X]$.
2. Let $g(X) = X^2$. The value $E_X\left[X^2\right]$ is called the *second moment*.
3. Let $g(X) = X^n$. The value $E_X[X^n]$ is called the nth moment.
4. Let $g(X) = (X - \mu_X)^2$. This is the *second central moment*. This is also called the variance, or σ_X^2. What are the units of the variance?
5. Let $g(X) = (X - \mu_X)^n$. This is the nth central moment.

Some notes:

1. "Moment" is used by analogy to the moment of inertia of a mass. Moment of inertia describes how difficult it is to get a mass rotating about its center of mass, and is given by:

$$I \triangleq \int\int\int_V r^2 \rho dx\, dy\, dz$$

where ρ is the mass density, and r is the distance from the center.

2. Standard deviation $\sigma = \sqrt{\sigma_X^2}$.

3. Variance in terms of 1st and 2nd moments.

$$E\left[(X-\mu_X)^2\right] = E\left[X^2 - 2X\mu_X + \mu_X^2\right] = E\left[X^2\right] - 2E\left[X\right]\mu_X + \mu_X^2 = E\left[X^2\right] - (E\left[X\right])^2.$$

4. Note $E_X\left[g(X)\right] \neq g\left(E_X\left[X\right]\right)$!

A summary:

Expression	X **is a discrete r.v.**	X **is a continuous r.v.**
$E\left[X\right]$	$= \sum_{x\in S_X} xf(x)$	$= \int_{x\in S_X} xf(x)dx$
$E\left[g(x)\right]$	$= \sum_{x\in S_X} g(x)f(x)$	$= \int_{x\in S_X} g(x)f(x)dx$
$E\left[aX+b\right]$	$= aE\left[X\right] + b$	$= aE\left[X\right] + b$
$E\left[X^2\right]$ 2nd moment	$= \sum_{x\in S_X} x^2 f(x)$	$= \int_{x\in S_X} x^2 f(x)dx$
$\sigma_X^2 = E\left[(X-\mu_X)^2\right]$, $\mu_X = E\left[X\right]$	$= \sum_{x\in S_X}(x-\mu_X)^2 f_X(x)$	$= \int_{x\in S_X}(x-\mu_X)^2 f_X(x)dx$

1.2 Examples

Example: Variance of Uniform r.v.
Let X be a continuous uniform r.v. on (a,b), with $a, b > 0$.

1. What is $E\left[X\right]$? It is

$$\int_a^b \frac{x}{b-a}dx = \frac{1}{2(b-a)}\left. x^2\right|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}.$$

2. What is $E\left[\frac{1}{X}\right]$?

$$\int_a^b \frac{1}{b-a}\frac{1}{x}dx = \frac{1}{b-a}(\ln b - \ln a) = \frac{1}{b-a}\ln\frac{b}{a}.$$

3. What is $E\left[X^2\right]$? It is

$$\int_a^b \frac{x^2}{b-a}dx = \frac{1}{3(b-a)}\left. x^3\right|_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{b^2+ab+a^2}{3}.$$

4. What is σ_X^2? It is

$$\sigma_X^2 = E_{X^2}\left[-\right](E\left[X\right])^2 = \frac{b^2+ab+a^2}{3} - \frac{b^2+2ab+a^2}{4} = \frac{b^2-2ab+a^2}{12} = \frac{(b-a)^2}{12}$$

Example: Bernoulli Moments

What are the mean, 2nd moment and variance of X, a Bernoulli r.v.?
Solution: Recall a Bernoulli r.v. X has two possible outcomes, 0 and 1. The probability that $X = 1$ is denoted p and thus $P\left[X=0\right] = 1-p$. The mean is:

$$\mu_X = E\left[X\right] = \sum_{x\in S_X} xf(x) = 0(1-p) + 1(p) = p.$$

The 2nd moment is:

$$E\left[X^2\right] = \sum_{x \in S_X} x^2 f(x) = 0^2(1-p) + 1^2(p) = p.$$

The variance could be calculated from the definition as:

$$\sigma_x^2 = E\left[(X-\mu_X)^2\right] = \sum_{x \in S_X} (x-p)^2 f(x) = p^2(1-p)+(1-p)^2(p) = p(1-p)(p+1-p) = p(1-p).$$

Alternatively, we could use the relationship shown earlier that $\sigma_X^2 = E\left[X^2\right] - (E\left[X\right])^2$:

$$\sigma_X^2 = p - p^2 = p(1-p).$$