Lecture 12: Expected values

COMS10014 Mathematics for Computer Science A

`cs-uob.github.io/COMS10014/ and github.com/coms10011/2020_21`

November 2020

Expected values

If $g(x)$ is a function we define the **expected value** of $g(X)$ as

$$\langle g(X) \rangle = \sum_{x} p(x) g(x)$$

Expected values

If $g(x)$ is a function we define the **expected value** of $g(X)$ as

$$\langle g(X) \rangle = \sum_x p(x) g(x)$$

or, in another common notation

$$E[g(X)] = \langle g(X) \rangle$$

Expected values

If $g(x) = x$ we get the **expected value** of X which is often just called the **expected value**:

$$\langle X \rangle = \sum_x x p(x)$$

If $p(x)$ is representing the frequencies, then this is the **mean**, often called μ.

More names for the same thing

The expect value of X is also referred to as the **first moment**; the 'first' bit is because it is the expectation value for the first power of X .

Example

X is the number of heads if a coin is flipped three times.

	0	1	2	3
$p_X(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

$$\langle X \rangle = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = \frac{3+6+3}{8} = \frac{3}{2}$$

Sample mean

If we sample multiple times for the sample space and get

$$\{x_1, x_2, \ldots, x_n\}$$

as values, when the probabilities are given by $p_X(x)$ then the sample mean approaches the expected value:

$$\frac{1}{n}\sum_i x_i \to \langle X \rangle$$

as n goes to infinity.

Sample values

If we sample multiple times for the sample space and get

$$\{x_1, x_2, \ldots, x_n\}$$

as values, when the probabilities are given by $p_X(x)$ then the sample mean approaches the expected value:

$$\frac{1}{n}\sum_i g(x_i) \to \langle g(X)\rangle$$

as n goes to infinity.

Variance

The **variance** is

$$V(X) = \langle (X-\mu)^2 \rangle$$

When $p(x)$ represents frequencies this is the square of the **standard deviation**:

$$V(X) = \sigma^2$$

Variance measures spread

	0	1	2	3
p_X	1/8	3/8	3/8	1/8

has expected value is 1.5 and

$$V(X) = 0.75$$

Variance measures spread

	0	1	2	3
p_Y	1/16	7/16	7/16	1/16

has expected value is 1.5 and

$$V(Y) = 0.5$$

More names

$\langle X^2 \rangle$ is called the **second moment**, the variance is called the **second central moment**; the 'central' indicates that it is the second moment you get if you take away the mean first.

Other moments

There are other moments use to describe distributions such as **skewness** based on the third central moment:

$$s = \frac{1}{\sigma^3} \langle (X - \mu)^3 \rangle$$

and the **kurtosis** based on the fourth

$$\kappa = \frac{1}{\sigma^4} \langle (X - \mu)^4 \rangle$$

Nice properties

Scalar multiplication

$$\langle cg(X)\rangle = \sum_x cg(x)p(x) = c\sum_x g(x)p(x) = c\langle g(X)\rangle$$

Also, trivially

$$\langle 1\rangle = \sum_x p(x) = 1$$

Additive

$$\begin{aligned}\langle g_1(X) + g_2(X)\rangle &= \sum_x [g_1(x) + g_2(x)]p(x)\\ &= \sum_x g_1(x)p(x) + \sum_x g_2(x)p(x)\\ &= \langle g_1(X)\rangle + \langle g_2(X)\rangle\end{aligned}$$

Another formula for variance

$$V(X) = \langle (X-\mu)^2 \rangle = \langle (X^2 - 2\mu X + \mu^2) \rangle$$

Now, using the additive property

$$V(X) = \langle X^2 \rangle - 2\mu \langle X \rangle + \langle \mu^2 \rangle$$

Finally, noting $\mu = \langle X \rangle$,

$$V(X) = \langle X^2 \rangle - \mu^2$$