Lecture 4

Math 50051, Topics in Probability Theory and Stochastic Processes

Expectation:

Question: Toss a die. If you toss an even number you loose a dollar, if you toss an odd number you gain a dollar. What is the best approximation for your outcome, if you do not know anything else? Answer: Since the two outcomes are equally likely, the answer will be the average of the two numbers.

In general:

Let X be a discrete random variable with distribution $f(k)$. The **expected value** (**mean**) of X is denoted by $E(X)$ and defined by

$$E(X) = \sum_k k f(k),$$

provided that $\sum_k |k| f(k)$ is finite. (If this condition fails to hold then X has no finite mean.)

You may, if you wish, think of this as a weighted average of the possible values of X, where the weights are $f(k)$.

But if X is a continuous rv the weighted average transforms in the corresponding integral, ie:

The *expected value* (or *mean*) of a continuous rv X is denoted, like in the discrete case, by $E(X)$ and defined by

$$E(X) = \int_{\mathbb{R}} x f(x) dx$$

provided that $\int_{\mathbb{R}} |x| f(x) dx$ is finite.

Remark 1: The mean of a rv is the center of gravity of the distribution of the rv.

Remark 2: If the above conditions are verified the rv are called **integrable**. In particular, this means that the expectation of $|X|$ is finite. Observe that if a rv is integrable then its expectation is finite, but if its expectation is finite that does not mean that it is integrable! If $E(X^2)$ is finite then we say the rv is square integrable.

Exercise: Compute the mean of a Binomial, Poisson and Normal rv.

Example: We denote by I_A the indicator function of the set A, defined by

$$I_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases} \tag{1}$$

Then for any Borel set A, the indicator function is an integrable random variable and its expectation is $P(A)$. Why?

Tail theorem. Let X be non-negative with distribution function $F_X(x)$. Then

$$E(X) = \int_0^\infty P(X > x) dx = \int_0^\infty \{1 - F_X(x)\} dx.$$

Why?

Properties:

1) **Expectation of a composition** Let the random variables X and Y satisfy $Y = g(X)$, where $g(\cdot)$ is a real-valued function on $\mathbb{R}$.

(a) If X is discrete, then

$$E(Y) = \sum_x g(x) f_X(x),$$

provided that $\sum_x |g(x)|\, f_X(x) < \infty$

(b) If X is continuous, then

$$E(Y) = \int_{-\infty}^{\infty} g(x) f_X(x) dx,$$

provided that $\int_{\mathbb{R}} |g(x)|\, f_X(x) < \infty$.

2) **Linearity of E.** If $E(X)$ and $E(Y)$ exist, then for constants a and b

$$E(aX + bY) = aE(X) + bE(Y).$$

3) **Independence case.** If X and Y are independent, then for functions g and h

$$E\{g(X)h(Y)\} = E[g(X)]E[h(Y)],$$

whenever both sides exist.

Moments

(a) The k**th moment** of X is $\mu_k = E(X^k)$.

(b) The k**th central moment** of X is $\sigma_k = E(X - E(X))^k$.

In particular μ_1 is the mean $\mu = E(X)$, and σ_2 is called the *variance* and denoted by σ^2 or var X. Thus

$$\sigma^2 = E(X - \mu)^2 = varX$$

and for the second moment

$$E(X^2) = varX + (E(X))^2 = \sigma^2 + \mu^2.$$

The square root of the central moment is the **standard deviation**. It is a measure of the average deviation of observations from the mean. In financial markets the standard deviation of a price change is called the volatility.

Moment generating function

The **moment generating function** (mgf) of a random variable X is

$$M_X(t) = E(e^{tX})$$

for all real t where the expected value exists. The reason is called mgf is because

$$E(X^r) = M_X^{(r)}(0).$$

Why?

Properties: 1) **Uniqueness.** If $M_X(t) < \infty$ then there is a unique $F_X(x)$ having $M_X(t)$ as its mgf.
2) **Factorization.** If X and Y are independent then

$$M(s,t) = E(e^{sX+tY}) = M_X(s)M_Y(t)$$

3) **Continuity.** If M_n is a sequence of mgf such that $\lim_{n\to\infty} M_n(t) = M(t)$, then if M is the mgf of the distribution F, and M_n are the mgfs of the distributions F_n we have

$$\lim_{n\to\infty} F_n = F.$$

Conditional expectation

Suppose, as before, that X is a r.v measuring the outcome of some random experiment. If we do not know anything about the outcome, we said that the best guess for X is $E(X)$. If, on the other hand, we know completely the outcome of the experiment then we know the exact value of X. The notion of conditional expectation deals with making the best guess for X given some, but not all information.

The discrete case Suppose that X and Y are both discrete rv with joint probability mass function

$$f_{X,Y}(x,y) = P(X = x, Y = y)$$

and marginal probability mass function $f_X(x)$ and $f_Y(y)$ taking values in V_X and V_Y respectively. Then

$$\begin{aligned} E(X|Y = y) &= \sum_{x\in V_X} xP(X = x|Y = y) = \sum_{x\in V_X} x\frac{P(X = x, Y = y)}{P(Y = y)} \\ &= \sum_{x\in V_X} x\frac{f_{X,Y}(x,y)}{f_Y(y)} \end{aligned} \tag{2}$$

Notation: $P(X = x|Y = y) = f_{X|Y}(x|y)$ and it is called **conditional mass function of X given Y.**

The continuous case Assume X and Y are two rv jointly continuous, taking values in V_X and V_Y respectively, and with joint density function $f_{X,Y}(x,y)$ for $x \in V_X$ and $y \in V_y$. Then the **conditional density** of $X = x$ given $Y = y$, $f_{X|Y}(x|y)$ is given by

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

when $f_Y(y) > 0$. And the **conditional expectation** of X given $Y = y$ is given by

$$E(X|Y = y) = \int_{x\in V_X} xf_{X|Y}(x|y)dx,$$

when $f_Y(y) > 0$.

We observe that $E(X|Y = y)$ is a function of y, a rv on the $\sigma-$field generated by Y, denoted by $E(X|Y)$

Example: 3 coins are tossed: 1c, 5c, 10c. The rv X gives the sum of the values of the coins that land heads up. What is $E(X|2$ coins have landed heads up)? What is $E(X|Y)$ if Y gives the total amount shown by the 5c and 10c only?