CmpE 343 Lecture Notes
4: Expectation

Ethem Alpaydın

October 12, 2014

1 Introduction

A probability distribution is a complex object; in the discrete case, it is a table of values and their probabilities, and in the continuous case, it is a function. There are certain values that can be calculated from a distribution relatively easily and they give us a lot of information about the distribution. The mean and variance are such measures. Similarly in the case of a joint distribution, we have covariance and correlation which summarize the relationship between two random variables.

2 Expected Value

The *expected value* or *mean* is the average of the values of the random variable weighted by their probabilities. In the discrete case, we have

$$E[X] = \sum_x xP\{X = x\} = \sum_x xf(x) \tag{1}$$

where $f(x)$ is the probability mass function. In the continuous case, we have

$$E[X] = \int xf(x)dx \tag{2}$$

If we repeat the experiment N times, sum the observed value of X in each, and divide by N, this average converges to the expected value as N increases. For example, let us say we toss a fair coin three times and X is the number of heads, the expected value is

$$E[X] = 0\frac{1}{8} + 1\frac{3}{8} + 2\frac{3}{8} + 3\frac{1}{8} = \frac{12}{8} = 1.5$$

It is also denoted by μ_X, where we omit the subscript if there is a single random variable.

Expected value is useful because although we do not know what the outcome will be for a single experiment, the expected value tells us the expected behavior if the experiment is to be repeated many times. Expected value calculations are done in all applications where there is randomness. For example in buying a lottery ticket, let us say the win is w and the probability of getting the lucky number is p and that the lottery ticket costs c. Then

$$\text{Expected earning } = p * (w - c) - (1 - p) * c$$

It makes sense to buy a ticket for such a lottery if this expected gain is greater than 0—generally the expected earning is less than 0 and that is why gamblers may earn or lose but gambling houses always earn money. You can do a similar calculation about whether it makes sense or not to buy insurance. Though such expected value calculations are generally useful in making decisions, for example, in choosing between actions—whether to buy a ticket or not—, it is known that people do not always make rational decisions[1]; there may also be ethical constraints in that the result of certain actions cannot (and should not) be measured in monetary terms[2].

Consider a function $g(X)$ of the random variable X with probability distribution $f(x)$. Because it is a function of a random variable, it is also a random variable, let us say, $Y \equiv g(X)$. Then

$$E[Y] = E[g(X)] = \sum_x g(x)f(x) \tag{3}$$

As usual, we replace the discrete sum by integration if X is continuous.

3 Variance

The expected value gives us the "center of gravity" of the random variable; *variance* which we define now gives us the "spread" of X around that center.

$$\text{Var}(X) = E[(X-\mu)^2] \tag{4}$$

also denoted by σ_X^2. An observation for X can be different from the mean and it can be smaller or greater; we define a "distance" from the mean $(X-\mu)^2$ and the variance is the expected distance averaged over all possible values of X weighted by their probabilities of their occurrence. Using Equation (3) with $g(X) = (X-\mu)^2$, we get

$$\text{Var}(X) = \sum_x (x-\mu)^2 f(x) \tag{5}$$

where again we replace the discrete sum by integration if X is continuous. The positive square root of the variance is the *standard deviation* and is denoted by σ. The standard deviation is more interpretable because it is in the same scale with X. In scientific literature, frequently we see results reported as, for example, 5.2 ± 0.8; when we see this we understand that 5.2 is the mean and 0.8 is the standard deviation.

It can be shown (quite easily by just expanding Equation (5) that

$$\text{Var}(X) = E[X^2] - (E[X])^2 \tag{6}$$

$E[X^k]$, for integer $k \geq 1$ is called the *kth moment* of X, and hence we see that mean is the first moment and that variance is the second moment minus the square of the first moment.

Let us say we want to calculate the variance of some function $g(X)$:

$$\text{Var}(g(X)) = E[(g(X) - E[g(X)])^2] = \sum_x (g(x) - E[g(X)])^2 f(x) \tag{7}$$

Note that $E[g(X)]$ is not a random variable but a constant which we calculate and plug in the formula. Or using equation (6), we can calculate the variance as

$$\text{Var}(g(X)) = E[(g(X)^2] - (E[g(X)])^2 = \sum_x g(x)^2 f(x) - \left(\sum_x g(x)f(x)\right)^2$$

4 Two (or More) Random Variables

If we have two random variables X and Y with joint probability distribution $f(x,y)$, the expected value of the random variable $Z \equiv g(X,Y)$ is written as

$$E[g(X,Y)] = \sum_x \sum_y g(x,y)f(x,y) \tag{8}$$

[1] Kahneman, Daniel, and Amos Tversky. (1979). "Prospect Theory: An Analysis of Decision Under Risk". *Econometrica* XLVII: 263-291.

[2] Sandel, Micheal J. (2012). *What Money Can't Buy: The Moral Limits of Markets*, Barnes and Noble.

We can calculate the variance of X or Y over the marginal distributions, and we also have the measure of *covariance* defined as

$$\text{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])] \tag{9}$$

Covariance takes a value depending on the arbitrary scales of X and Y; *correlation* is its normalized version and is always between -1 and $+1$:

$$\text{Corr}(X,Y) = \frac{cov(X,Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}} \tag{10}$$

If when X is larger than its mean Y is also larger than its mean, and when X is smaller than its mean Y is also smaller than its mean, the product in Equation (9) is positive and so is the covariance. For example, people who are taller than the average height are generally also heavier than the average weight and people who are shorter than the average height are also generally lighter than the average weight; there is positive covariance between height and weight, or we say they are *positively correlated*. People who are older tend to run slower and people who are younger tend to run faster (valid when age is over 25); in such a case, age and speed are negatively correlated. If the two are independent, covariance and correlation are around 0.

It can be shown that

$$\text{Cov}(X,Y) = E[XY] - E[X]E[Y]$$

If X and Y are independent, $E[XY] = E[X]E[Y]$.

The following hold (X, Y are random variables and a, b are constants):

$$\begin{aligned}
E[aX + b] &= aE[X] + b \\
\text{Var}(aX + b) &= a^2\text{Var}(X) \\
E[aX + bY] &= aE[X] + bE[Y] \\
\text{Var}(aX + bY) &= a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X,Y)
\end{aligned}$$

5 Chebyshev's Inequality

Chebyshev is a Russian mathematician who lived in late 19th century. The inequality we discuss now is named after him and is very useful in that it defines a bound that holds for *any* probability distribution.

The probability that any random variable X takes a value within k standard deviations to the mean is at least $1 - 1/k^2$, or

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2} \tag{11}$$

Let us prove it. We start with the definition of the variance:

$$\begin{aligned}
\sigma^2 &= E[(X-\mu)^2] \\
&= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \\
&= \int_{-\infty}^{\mu - k\sigma} (x-\mu)^2 f(x)dx + \int_{\mu - k\sigma}^{\mu + k\sigma} (x-\mu)^2 f(x)dx + \int_{\mu + k\sigma}^{\infty} (x-\mu)^2 f(x)dx
\end{aligned}$$

We divide $(-\infty, \infty)$ into three regions $(-\infty, \mu - k\sigma)$, $(\mu - k\sigma, \mu + k\sigma)$, and $(\mu + k\sigma, \infty)$. The second term in always 0 or greater, it can never be negative, so we can write:

$$\sigma^2 \geq \int_{-\infty}^{\mu - k\sigma} (x-\mu)^2 f(x)dx + \int_{\mu + k\sigma}^{\infty} (x-\mu)^2 f(x)dx$$

Now in the first region $X < \mu - k\sigma$ and in the second, $X > \mu + k\sigma$, so for both, we have $|X - \mu| > k\sigma$, or $(X - \mu)^2 > k^2\sigma^2$. Hence we get an even smaller sum on the right:

$$\sigma^2 \geq \int_{-\infty}^{\mu - k\sigma} k^2\sigma^2 f(x)dx + \int_{\mu + k\sigma}^{\infty} k^2\sigma^2 f(x)dx$$

$$
\begin{aligned}
&= k^2\sigma^2 \left(\int_{-\infty}^{\mu-k\sigma} f(x)dx + \int_{\mu+k\sigma}^{\infty} f(x)dx \right) \\
&= k^2\sigma^2 \left(1 - \int_{\mu-k\sigma}^{\mu+k\sigma} f(x)dx \right) \\
1 &\geq k^2 \left(1 - \int_{\mu-k\sigma}^{\mu+k\sigma} f(x)dx \right) \\
1 - \frac{1}{k^2} &\leq \int_{\mu-k\sigma}^{\mu+k\sigma} f(x)dx
\end{aligned}
$$

and so we prove that

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$