# Lecture 10-11

Andrei Antonenko

Feb 14-16, 2005

## 1 Conditional expectation: Properties, Examples

One of the most interesting properties of the conditional expectation is a **total expectation theorem:**

$$\mathbf{E}[X] = \sum_y p_Y(y)\mathbf{E}[X|Y=y] \tag{1}$$

or in a different form:

$$\mathbf{E}[X] = \sum_y P(A_i)\mathbf{E}[X|A_i], \tag{2}$$

where events $A_i$ form a partition of the sample space $\Omega$.

*Proof.*

$$\begin{aligned}
\mathbf{E}[X] &= \sum_x x p_X(x) = \\
&= \sum_x x \sum_y p_Y(y) p_{X|Y}(x|y) = \\
&= \sum_y p_Y(y) \sum_x x p_{X|Y}(x|y) = \\
&= \sum_y \mathbf{E}[X|Y=y].
\end{aligned}$$

□

**Example 1.1.** Messages are transmitted from the computer in the New York City to Boston with probability 0.5, to Chicago with probability 0.3, and to San Francisco with probability 0.2. The average transmission time of the message to Boston is 0.05 sec, to Chicago is 0.1 sec, and to San Francisco 0.3 sec. What is the average transmission time of the message?

By total expectation theorem,

$$\mathbf{E}[X] = 0.5 \cdot 0.05 + 0.3 \cdot 0.1 + 0.2 \cdot 0.3 = 0.115 \text{ sec.}$$

**Example 1.2** (Mean and variance of the geometric random variable)**.** The PMF of the geometric random variable is $p_X(k) = p(1-p)^{k-1}$, $k = 1, 2, \ldots$. It describes the number of the experiment on which the first success occurred in the series of experiment. To compute the expectation we must compute the following sum: $\sum_k kp(1-p)^{k-1}$, but it is pretty tedious. We will use the different method. We will use the total expectation theorem with events $A_0 = \{X = 1\}$ and $A_1 = \{X > 1\}$.

If the event $A_0$ happened, we have:

$$\mathbf{E}[X|A_0] = \mathbf{E}[X|X = 1] = 1$$

If the event $A_1$ happened, i.e. the first experiment was unsuccessful. Therefore, we just waisted the first experiment, and we start everything from the beginning. Therefore,

$$\mathbf{E}[X|A_1] = \mathbf{E}[X|X > 1] = \mathbf{E}[X] + 1.$$

Thus, by total expectation theorem, we have:

$$\begin{aligned}\mathbf{E}[X] &= p\mathbf{E}[X|A_0] + (1-p)\mathbf{E}[X|A_1] = \\ &= p \cdot 1 + (1-p)(1 + \mathbf{E}[X]),\end{aligned}$$

from where one can find

$$\mathbf{E}[X] = \frac{1}{p}. \tag{3}$$

By similar reasoning, we get:

$$\mathbf{E}\left[X^2|X = 1\right] = 1,$$

and

$$\mathbf{E}\left[X^2|X > 1\right] = \mathbf{E}\left[(X+1)^2\right] = \mathbf{E}\left[X^2 + 2X + 1\right] = \mathbf{E}\left[X^2\right] + 2\mathbf{E}[X] + 1.$$

By total expectation theorem,

$$\mathbf{E}\left[X^2\right] = p\mathbf{E}\left[X^2|X = 1\right] + (1-p)\mathbf{E}\left[X^2|X > 1\right],$$

from where we can obtain the value for $\mathbf{E}[X^2]$:

$$\mathbf{E}\left[X^2\right] = \frac{2}{p^2} - \frac{1}{p}. \tag{4}$$

Therefore,

$$\mathbf{var}(X) = \mathbf{E}\left[X^2\right] - (\mathbf{E}[X])^2 = \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}. \tag{5}$$

# 2 Independence

As for events, we can define independence for random variables.

**Definition 2.1.** *Random variable $X$ is called **independent** of the event $A$ if*

$$P(X = x \text{ and } A) = P(X = x)P(A) = p_X(x)P(A). \tag{6}$$

Since $p_{X|A} = P(X = x \text{ and } A)/P(A)$, from where we get

$$p_X(x) = p_{X|A}(x), \tag{7}$$

i.e. the occurrence of $A$ does not affect the distribution of the variable $X$.

**Example 2.2.** Consider two independent tosses of the fair coin. Let $X$ be the number of heads. Then $X$ has the following distribution:

$$p_X(x) = \begin{cases} 1/4, & x = 0 \\ 1/2, & x = 1 \\ 1/3, & x = 2 \end{cases}$$

Let the event $A$ be the event that the number of heads is even. Using formula $p_{X|A}(x) = P(X = x \text{ and } A)/P(A)$, we have:

$$p_{X|A}(x) = \begin{cases} 1/2, & x = 0 \\ 0, & x = 1 \\ 1/2, & x = 2 \end{cases}$$

and therefore $X$ and $A$ are (obviously) not independent.

**Definition 2.3.** *Two random variables $X$ and $Y$ are called **independent** if*

$$p_{X,Y}(x,y) = p_X(x)p_Y(y). \tag{8}$$

Formula $p_{X,Y}(x,y) = p_Y(y)p_{X|Y}(x|y)$ implies that $p_X(x) = p_{X|Y}(x|y)$, i.e. value of $Y$ does not tell us anything about $X$.

Also we can introduce a notion of conditional independence.

**Definition 2.4.** *Two random variable $X$ and $Y$ are called conditionally independent given an event $A$ (with positive probability), if*

$$P(X = x, Y = y|A) = P(X = x|A)P(Y = y|A), \tag{9}$$

*or in different notation,*

$$p_{X,Y|A}(x,y) = p_{X|A}(x)p_{Y|A}(y). \tag{10}$$

Now let's look at the expectation of the product of two random variables. In fact, if $X$ and $Y$ are independent, than

$$\mathbf{E}[XY] = \mathbf{E}[X]\,\mathbf{E}[Y]. \tag{11}$$

We will prove this fact:

$$\begin{aligned} \mathbf{E}[XY] &= \sum_{x,y} xyp_{X,Y}(x,y) = \\ &= \sum_x \sum_y xyp_X(x)p_Y(y) = \\ &= \sum_x xp_X(x) \sum_y yp_Y(y) = \\ &= \mathbf{E}[X]\,\mathbf{E}[Y]. \end{aligned}$$

Similar calculations show that

$$\mathbf{E}[g(X)h(Y)] = \mathbf{E}[g(X)]\,\mathbf{E}[h(Y)]. \tag{12}$$

Now, assume $Z = X + Y$. We know that $\mathbf{E}[Z] = \mathbf{E}[X] + \mathbf{E}[Y]$. Now we will compute the variance of $Z$, assuming that $X$ and $Y$ are independent.

$$\begin{aligned}\mathbf{var}(Z) &= \mathbf{E}\left[(X + Y - \mathbf{E}[X+Y])^2\right] = \\ &= \mathbf{E}\left[((X - \mathbf{E}[X]) + (Y - \mathbf{E}[Y]))^2\right] = \\ &= \mathbf{E}\left[(X - \mathbf{E}[X])^2\right] + \mathbf{E}\left[(Y - \mathbf{E}[Y])^2\right] + 2\mathbf{E}\left[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])\right].\end{aligned}$$

Since $X$ and $Y$ are independent, we have:

$$\mathbf{E}\left[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])\right] = \mathbf{E}\left[(X - \mathbf{E}[X])\right]\mathbf{E}\left[(Y - \mathbf{E}[Y])\right] = 0$$

, and therefore

$$\mathbf{var}(Z) = \mathbf{var}(X) + \mathbf{var}(Y). \tag{13}$$

We can generalize the notion of independence to the case of more than 2 random variables. For example, three random variables $X$, $Y$, and $Z$ are called independent if

$$p_{X,Y,Z}(x, y, z) = p_X(x)p_Y(y)p_Z(z). \tag{14}$$

Also, if three or more variables $X_i$ are independent, then

$$\mathbf{var}(X_1 + \cdots + X_n) = \mathbf{var}(X_1) + \cdots + \mathbf{var}(X_n). \tag{15}$$

**Example 2.5** (Variance of binomial random variable)**.** As we saw before, binomial random variable $X$ with parameters $n$ and $p$ can be represented as a sum of $n$ Bernoulli random variables $X_i$ with parameter $p$. Since the variance of Bernoulli r.v. is equal to $\mathbf{var}(X_i) = p(1-p)$, we have

$$\mathbf{var}(X) = np(1-p). \tag{16}$$

**Example 2.6** (Mean and variance of the sample mean)**.** Assume we would like to estimate the approval rating of the president. We ask $n$ people, and let $X_i$ be the random variable which denotes the choice of $i$'th person:

$$X_i = \begin{cases} 1, & \text{if person approves the president's performance} \\ 0, & \text{if person does not approve the president's performance}\end{cases}$$

We model variables $X_i$ as independent Bernoulli random variable with common mean $p$ and variance $p(1-p)$. We view $p$ as a real approval rating of the president. We average the responses and compute the **sample mean**

$$S_n = \frac{X_1 + X_2 + \cdots + X_n}{n}. \tag{17}$$

We have:

$$\mathbf{E}[S_n] = \frac{1}{n}(\mathbf{E}[X_1] + \cdots + \mathbf{E}[X_n]) = \frac{1}{n} \cdot np = p, \tag{18}$$

and since $X_i$'s are independent,

$$\mathbf{var}(S_n) = \mathbf{var}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n^2}(\mathbf{var}(X_1) + \cdots + \mathbf{var}(X_n)) = \frac{p(1-p)}{n}. \tag{19}$$

Therefore, $S_n$ is a good estimate for $p$, since it's average is equal to $p$, and it's variance (accuracy) improves as sample size $n$ increases.