Lecture 11

1. Some properties of expectations

Suppose X is a random variable with mass function f. If $Y = g(X)$ for some function g, then what is the expectation of Y? One way to address this is to first find the mass function f_Y of Y, and then to compute $\mathrm{E}Y$ as $\sum_a a f_Y(a)$ [provided that the sum makes sense, of course]. But there is a more efficient method.

Theorem 11.1. *If X has mass function f and g is some function, then*

$$\mathrm{E}\left[g(X)\right] = \sum_x g(x) f(x),$$

provided that either $g(x) \geq 0$ for all x, or $\sum_x |g(x)| f(x) < \infty$.

Proof. Let $y_1, y_2, \ldots$ denote the possible values of $g(X)$. Consider the set $A_j = \{x : g(x) = y_j\}$ for all $j \geq 1$. Because the y_j's are distinct, it follows that the A_j's are disjoint. Moreover,

$$\begin{aligned}\mathrm{E}\left[g(X)\right] &= \sum_{j=1}^{\infty} y_j \mathrm{P}\{g(X) = y_j\} = \sum_{j=1}^{\infty} y_j \mathrm{P}\{X \in A_j\} \\ &= \sum_{j=1}^{\infty} y_j \sum_{x \in A_j} f(x) = \sum_{j=1}^{\infty} \sum_{x \in A_j} g(x) f(x).\end{aligned}$$

Because the A_j's are disjoint,

$$\sum_{j=1}^{\infty} \sum_{x \in A_j} g(x) f(x) = \sum_{x \in \cup_{j=1}^{\infty} A_j} g(x) f(x).$$

The theorem follows from making one final observation: $\cup_{j=1}^{\infty} A_j$ is the collection of all possible values of X. □

One can derive other properties of expectations by applying similar arguments. Here are some useful properties. For proof see the text.

Theorem 11.2. *Let X be a discrete random variable with mass function f and finite expectation $\mathrm{E}X$. Then:*

(1) $\mathrm{E}(aX + b) = a\mathrm{E}(X) + b$ *for all constants* a, b*;*

(2) $\mathrm{E}a = a$ *for all nonrandom (constant) variables* a*;*

(3) *If* $\mathrm{P}\{a \le X \le b\} = 1$, *then* $a \le \mathrm{E}X \le b$*;*

(4) *If* $g(X)$ *and* $h(X)$ *have finite expectations, then*

$$\mathrm{E}[g(X) + h(X)] = \mathrm{E}[g(X)] + \mathrm{E}[h(X)].$$

This is called linearity.

2. A first example

Suppose X has mass function

$$f(x) = \begin{cases} 1/4 & \text{if } x = 0, \\ 3/4 & \text{if } x = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Recall: $\mathrm{E}X = (\frac{1}{4} \times 0) + (\frac{3}{4} \times 1) = \frac{3}{4}$. Now let us compute $\mathrm{E}(X^2)$ using Theorem 11.1:

$$\mathrm{E}(X^2) = \left(\frac{1}{4} \times 0^2\right) + \left(\frac{3}{4} \times 1^2\right) = \frac{3}{4}.$$

Two observations:

(1) This is obvious because $X = X^2$ *in this particular example*; and

(2) $\mathrm{E}(X^2) \ne (\mathrm{E}X)^2$. In fact, the difference between $\mathrm{E}(X^2)$ and $(\mathrm{E}X)^2$ is an important quantity, called the *variance of* X. We will return to this topic later.

3. A second example

If $X = \mathrm{Bin}(n\,, p)$, then what is $\mathrm{E}(X^2)$? It may help to recall that $\mathrm{E}X = np$. By Theorem 11.1,

$$\mathrm{E}(X^2) = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k q^{n-k} = \sum_{k=1}^{n} k \, \frac{n!}{(k-1)!(n-k)!} \, p^k q^{n-k}.$$

The question is, "how do we reduce the factor k further"? If we had $k-1$ instead of k, then this would be easy to answer. So let us first solve a related problem.

$$\begin{aligned} \mathrm{E}\big[X(X-1)\big] &= \sum_{k=0}^{n} k(k-1)\binom{n}{k}p^k q^{n-k} = \sum_{k=2}^{n} k(k-1)\ \frac{n!}{k!(n-k)!}\ p^k q^{n-k} \\ &= n(n-1)\sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!\big([n-2]-[k-2]\big)!}\ p^k q^{n-k} \\ &= n(n-1)\sum_{k=2}^{n}\binom{n-2}{k-2}p^k q^{n-k} \\ &= n(n-1)p^2\sum_{k=2}^{n}\binom{n-2}{k-2}p^{k-2} q^{[n-2]-[k-2]} \\ &= n(n-1)p^2\sum_{\ell=0}^{n-2}\binom{n-2}{\ell}p^{\ell} q^{[n-2]-\ell}. \end{aligned}$$

The summand is the probability that $\mathrm{Bin}(n-2\,,p)$ is equal to ℓ. Since that probability is added over all of its possible values, the sum is one. Thus, we obtain $\mathrm{E}[X(X-1)] = n(n-1)p^2$. But $X(X-1) = X^2 - X$. Therefore, we can apply Theorem 11.2 to find that

$$\begin{aligned} \mathrm{E}(X^2) &= \mathrm{E}[X(X-1)] + \mathrm{E}X = n(n-1)p^2 + np \\ &= (np)^2 + npq. \end{aligned}$$

4. Expectation inequalities

Theorem 11.3 (The triangle inequality)**.** *If X has a finite expectation, then*

$$\big|\mathrm{E}X\big| \le \mathrm{E}(|X|).$$

Proof. Let $g(x) = |x| - x$. This is a positive function, and $\mathrm{E}[g(X)] = \mathrm{E}(|X|) - \mathrm{E}X$. But $\mathrm{P}\{g(X) \ge 0\} = 1$. Therefore, $\mathrm{E}[g(X)] \ge 0$ by Theorem 11.2. This proves that $\mathrm{E}X \le \mathrm{E}(|X|)$. Apply the same argument to $-X$ to find that $-\mathrm{E}X = \mathrm{E}(-X) \le \mathrm{E}(|-X|) = \mathrm{E}(|X|)$. This proves that $\mathrm{E}X$ and $-\mathrm{E}X$ are both bounded above by $\mathrm{E}(|X|)$, which is the desired result. □

Theorem 11.4 (The Cauchy–Schwarz inequality)**.** *If $\mathrm{E}(|X|) < \infty$, then*

$$\mathrm{E}(|X|) \le \sqrt{\mathrm{E}(X^2)}.$$

Proof. Expand the trivial bound $(|X| - \mathrm{E}(|X|))^2 \geq 0$ to obtain:

$$X^2 - 2|X|\mathrm{E}(|X|) + \left|\mathrm{E}(|X|)\right|^2 \geq 0.$$

Take expectations, and note that $b = \mathrm{E}(|X|)$ is nonrandom. This proves that

$$\mathrm{E}(X^2) - 2\mathrm{E}(|X|)\mathrm{E}(|X|) + \left|\mathrm{E}(|X|)\right|^2 \geq 0.$$

The left-hand side is manifestly equal to $\mathrm{E}(X^2) - |\mathrm{E}(|X|)|^2$, whence follows the theorem. □

One can use more advanced methods to prove the following:

$$\mathrm{E}(|X|) \leq \sqrt{\mathrm{E}(X^2)} \quad \text{for all random variables } X.$$

Note that $|X|$ and X^2 are nonnegative. So the expectations are defined, though possibly infinite. The preceding form of the Cauchy–Schwarz inequality implies that if $\mathrm{E}(X^2)$ is finite, then so is $\mathrm{E}(|X|)$.