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 EY as $\sum_a af_Y(a)$ [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

$$E[g(X)] = \sum_{x} g(x)f(x),$$

provided that either $g(x) \ge 0$ for all x, or $\sum_{x} |g(x)| f(x) < \infty$.

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

$$E[g(X)] = \sum_{j=1}^{\infty} y_j P\{g(X) = y_j\} = \sum_{j=1}^{\infty} y_j P\{X \in A_j\}$$
$$= \sum_{j=1}^{\infty} y_j \sum_{x \in A_j} f(x) = \sum_{j=1}^{\infty} \sum_{x \in A_j} g(x)f(x).$$

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: $\bigcup_{j=1}^{\infty} A_j$ 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* EX. *Then:*

- (1) E(aX + b) = aE(X) + b for all constants a, b;
- (2) Ea = a for all nonrandom (constant) variables a;
- (3) If $P{a \leq X \leq b} = 1$, then $a \leq EX \leq b$;
- (4) If g(X) and h(X) have finite expectations, then

$$E[g(X) + h(X)] = E[g(X)] + E[h(X)].$$

This is called linearity.

2. A first example

Suppose X has mass function

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

Recall: EX = $(\frac{1}{4} \times 0) + (\frac{3}{4} \times 1) = \frac{3}{4}$. Now let us compute E(X²) using Theorem 11.1:

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

Two observations:

- (1) This is obvious because $X = X^2$ in this particular example; and
- (2) $E(X^2) \neq (EX)^2$. In fact, the difference between $E(X^2)$ and $(EX)^2$ is an important quantity, called the *variance of* X. We will return to this topic later.

3. A second example

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

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

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.

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

The summand is the probability that 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 $E[X(X - 1)] = n(n - 1)p^2$. But $X(X - 1) = X^2 - X$. Therefore, we can apply Theorem 11.2 to find that

$$E(X^2) = E[X(X-1)] + EX = n(n-1)p^2 + np$$

= $(np)^2 + npq$.

4. Expectation inequalities

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

$$|\mathbf{E}\mathbf{X}| \leq \mathbf{E}(|\mathbf{X}|).$$

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

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

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

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

$$X^{2} - 2|X|E(|X|) + |E(|X|)|^{2} \ge 0.$$

Take expectations, and note that b = E(|X|) is nonrandom. This proves that

$$E(X^{2}) - 2E(|X|)E(|X|) + |E(|X|)|^{2} \ge 0.$$

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

One can use more advanced methods to prove the following:

 $E(|X|) \leqslant \sqrt{E(X^2)}$ 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 $E(X^2)$ is finite, then so is E(|X|).