Lecture 13

1. Inequalities

Let us start with an inequality.

Lemma 13.1. *If h is a nonnegative function, then for all $\lambda > 0$,*

$$\mathrm{P}\{h(X) \geq \lambda\} \leq \frac{\mathrm{E}[h(X)]}{\lambda}.$$

Proof. We know already that

$$\mathrm{E}[h(X)] = \sum_x h(x)f(x) \geq \sum_{x:\ h(x) \geq \lambda} h(x)f(x).$$

If x is such that $h(x) \geq \lambda$, then $h(x)f(x) \geq \lambda f(x)$, obviously. Therefore,

$$\mathrm{E}[h(X)] \geq \lambda \sum_{x:\ h(x) \geq \lambda} f(x) = \lambda \mathrm{P}\{h(X) \geq \lambda\}.$$

Divide by λ to finish. □

Thus, for example,

$$\mathrm{P}\{|X| \geq \lambda\} \leq \frac{\mathrm{E}(|X|)}{\lambda} \qquad \text{“Markov's inequality.”}$$

$$\mathrm{P}\{|X - \mathrm{E}X| \geq \lambda\} \leq \frac{\mathrm{Var}(X)}{\lambda^2} \qquad \text{“Chebyshev's inequality.”}$$

To get Markov's inequality, apply Lemma 13.1 with $h(x) = |x|$. To get Chebyshev's inequality, first note that $|X - \mathrm{E}X| \geq \lambda$ if and only if $|X - \mathrm{E}X|^2 \geq \lambda^2$. Then, apply Lemma 13.1 to find that

$$\mathrm{P}\{|X - \mathrm{E}X| \geq \lambda\} \leq \frac{\mathrm{E}\left(|X - \mathrm{E}X|^2\right)}{\lambda^2}.$$

Then, recall that the numerator is $\mathrm{Var}(X)$.

In words:

- If $\mathrm{E}(|X|) < \infty$, then the probability that $|X|$ is large is small.
- If $\mathrm{Var}(X)$ is small, then with high probability $X \approx \mathrm{E}X$.

2. Conditional distributions

If X is a random variable with mass function f, then $\{X = x\}$ is an event. Therefore, if B is also an event, and if $\mathrm{P}(B) > 0$, then

$$\mathrm{P}(X = x \mid B) = \frac{\mathrm{P}(\{X = x\} \cap B)}{\mathrm{P}(B)}.$$

As we vary the variable x, we note that $\{X = x\} \cap B$ are disjoint. Therefore,

$$\sum_x \mathrm{P}(X = x \mid B) = \frac{\sum \mathrm{P}(\{X = x\} \cap B)}{\mathrm{P}(B)} = \frac{\mathrm{P}\left(\cup_x \{X = x\} \cap B\right)}{\mathrm{P}(B)} = 1.$$

Thus,

$$f(x \mid B) = \mathrm{P}(X = x \mid B)$$

defines a mass function also. This is called the *conditional mass function of X given B*.

Example 13.2. Let X be distributed uniformly on $\{1, \ldots, n\}$, where n is a fixed positive integer. Recall that this means that

$$f(x) = \begin{cases} \dfrac{1}{n} & \text{if } x = 1, \ldots, n, \\ 0 & \text{otherwise.} \end{cases}$$

Choose and fix two integers a and b such that $1 \le a \le b \le n$. Then,

$$\mathrm{P}\{a \le X \le b\} = \sum_{x=a}^{b} \frac{1}{n} = \frac{b - a + 1}{n}.$$

Therefore,

$$f(x \mid a \le X \le b) = \begin{cases} \dfrac{1}{b - a + 1} & \text{if } x = a, \ldots, b, \\ 0 & \text{otherwise.} \end{cases}$$

3. Conditional expectations

Once we have a (conditional) mass function, we have also a conditional expectation at no cost. Thus,

$$\mathrm{E}(X \mid B) = \sum_x x f(x \mid B).$$

Example 13.3 (Example 13.2, continued)**.** In Example 13.2,

$$\mathrm{E}(X \mid a \le X \le b) = \sum_{k=a}^{b} \frac{k}{b-a+1}.$$

Now,

$$\begin{aligned}\sum_{k=a}^{b} k &= \sum_{k=1}^{b} k - \sum_{k=1}^{a-1} k \\ &= \frac{b(b+1)}{2} - \frac{(a-1)a}{2} \\ &= \frac{b^2 + b - a^2 + a}{2}.\end{aligned}$$

Write $b^2 - a^2 = (b-a)(b+a)$ and factor $b+a$ to get

$$\sum_{k=a}^{b} k = \frac{b+a}{2}(b-a+1).$$

Therefore,

$$\mathrm{E}(X \mid a \le X \le b) = \frac{b+a}{2}.$$

This should not come as a surprise. Example 13.2 actually shows that given $B = \{a \le X \le b\}$, the conditional distribution of X given B is uniform on $\{a, \ldots, b\}$. Therefore, the conditional expectation is the expectation of a uniform random variable on $\{a, \ldots, b\}$.

Theorem 13.4 (Bayes's formula for conditional expectations)**.** *If* $\mathrm{P}(B) > 0$*, then*

$$\mathrm{E}X = \mathrm{E}(X \mid B)\mathrm{P}(B) + \mathrm{E}(X \mid B^c)\mathrm{P}(B^c).$$

Proof. We know from the ordinary Bayes's formula that

$$f(x) = f(x \mid B)\mathrm{P}(B) + f(x \mid B^c)\mathrm{P}(B^c).$$

Multiply both sides by x and add over all x to finish. □

Remark 13.5. The more general version of Bayes's formula works too here: Suppose $B_1, B_2, \dots$ are disjoint and $\cup_{i=1}^{\infty} B_i = \Omega$; i.e., "one of the B_i's happens." Then,

$$\mathrm{E}X = \sum_{i=1}^{\infty} \mathrm{E}(X \mid B_i)\mathrm{P}(B_i).$$

Example 13.6. Suppose you play a fair game repeatedly. At time 0, before you start playing the game, your fortune is zero. In each play, you win or lose with probability 1/2. Let T_1 be the first time your fortune becomes +1. Compute $\mathrm{E}(T_1)$.

More generally, let T_x denote the first time to win x dollars, where $T_0 = 0$.

Let W denote the event that you win the first round. Then, $\mathrm{P}(W) = \mathrm{P}(W^c) = 1/2$, and so

$$\mathrm{E}(T_x) = \frac{1}{2}\mathrm{E}(T_x \mid W) + \frac{1}{2}\mathrm{E}(T_x \mid W^c). \tag{11}$$

Suppose $x \neq 0$. Given W, T_x is one plus the first time to make $x - 1$ more dollars. Given W^c, T_x is one plus the first time to make $x + 1$ more dollars. Therefore,

$$\begin{aligned}\mathrm{E}(T_x) &= \frac{1}{2}\Big[1 + \mathrm{E}(T_{x-1})\Big] + \frac{1}{2}\Big[1 + \mathrm{E}(T_{x+1})\Big] \\ &= 1 + \frac{\mathrm{E}(T_{x-1}) + \mathrm{E}(T_{x+1})}{2}.\end{aligned}$$

Also $\mathrm{E}(T_0) = 0$.

Let $g(x) = \mathrm{E}(T_x)$. This shows that $g(0) = 0$ and

$$g(x) = 1 + \frac{g(x+1) + g(x-1)}{2} \qquad \text{for } x = \pm 1, \pm 2, \dots .$$

Because $g(x) = (g(x) + g(x))/2$,

$$g(x) + g(x) = 2 + g(x+1) + g(x-1) \qquad \text{for } x = \pm 1, \pm 2, \dots .$$

Solve to find that for all integers $x \geq 1$,

$$g(x+1) - g(x) = -2 + g(x) - g(x-1).$$

Lecture 14

Example 14.1 (St.-Petersbourg paradox, continued)**.** We continued with our discussion of the St.-Petersbourg paradox, and note that for all integers $N \geq 1$,

$$\begin{aligned} g(N) &= g(1) + \sum_{k=2}^{N} \Big(g(k) - g(k-1) \Big) \\ &= g(1) + \sum_{k=1}^{N-1} \Big(g(k+1) - g(k) \Big) \\ &= g(1) + \sum_{k=1}^{N-1} \Big(-2 + g(k) - g(k-1) \Big) \\ &= g(1) - 2(N-1) + \sum_{k=1}^{N} \Big(g(k) - g(k-1) \Big) \\ &= g(1) - 2(N-1) + g(N). \end{aligned}$$

If $g(1) < \infty$, then $g(1) = 2(N-1)$. But N is arbitrary. Therefore, $g(1)$ cannot be finite; i.e.,

$$\mathrm{E}(T_1) = \infty.$$

This shows also that $\mathrm{E}(T_x) = \infty$ for all $x \geq 1$, because for example $T_2 \geq 1 + T_1$! By symmetry, $\mathrm{E}(T_x) = \infty$ if x is a negative integer as well.