Expectation of Random Variables

September 17 and 22, 2009

1 Discrete Random Variables

Let $x_1, x_2, \cdots x_n$ be observation, the **empirical mean**,

$$\bar{x} = \frac{1}{n}(x_1 + x_2 \cdots + x_n).$$

So, for the observations, $0, 1, 3, 2, 4, 1, 2, 4, 1, 1, 2, 0, 3$,

$$\bar{x} = \frac{24}{13}.$$

We could also organize these observations and taking advantage of the distributive property of the real numbers, compute $\bar{x}$ as follows

x	$n(x)$	$xn(x)$
0	2	0
1	4	4
2	3	6
3	2	6
4	2	8
	13	24

For $g(x) = x^2$, we can perform a similar compuation:

x	$g(x) = x^2$	$n(x)$	$g(x)n(x)$
0	0	2	0
1	1	4	4
2	4	3	12
3	9	2	18
4	16	2	32
		13	66

Then,

$$\overline{g(x)} = \frac{1}{n}\sum_x g(x)n(x) = \frac{66}{13}.$$

One further notational simplification is to write

$$p(x) = \frac{n(x)}{n} \text{ for the proportion of observations equal to } x, \text{ then } \overline{g(x)} = \sum_x g(x)p(x) = \frac{66}{13}.$$

So, for a finite sample space $S = \{s_1, s_2, \ldots, s_N\}$, we can define the **expectation** or the **expected value** of a random variable X by

$$EX = \sum_{j=1}^{N} X(s_j)P\{s_j\}. \tag{1}$$

In this case, two properties of expectation are immediate:

1. If $X(s) \geq 0$ for every $s \in S$, then $EX \geq 0$
2. Let X_1 and X_2 be two random variables and c_1, c_2 be two real numbers, then

$$E[c_1X_1 + c_2X_2] = c_1EX_1 + c_2EX_2.$$

Taking these two properties, we say that expectation is a **positive linear functional**.
We can generalize the identity in (1) to transformations of X.

$$Eg(X) = \sum_{j=1}^{N} g(X(s_j))P\{s_j\}.$$

Again, we can simplify

$$\begin{aligned} Eg(X) &= \sum_x \sum_{s;X(s)=x} g(X(s))P\{s\} = \sum_x \sum_{s;X(s)=x} g(x)P\{s\} \\ &= \sum_x g(x) \sum_{s;X(s)=x} P\{s\} = \sum_x g(x)P\{X=x\} = \sum_x g(x)f_X(x) \end{aligned}$$

where f_X is the probability density function for X.

Example 1. *Flip a biased coin twice and let X be the number of heads. Then,*

x	$f_X(x)$	$xf_X(x)$	$x^2f_X(x)$
0	$(1-p)^2$	0	0
1	$2p(1-p)$	$2p(1-p)$	$2p(1-p)$
2	p^2	$2p^2$	$4p^2$
		$2p$	$2p+2p^2$

Thus, $EX = 2p$ and $EX^2 = 2p + 2p^2$.

Example 2 (Bernoulli trials)**.** *Random variables $X_1, X_2, \ldots, X_n$ are called a sequence of* **Bernoulli trials** *provided that:*

1. *Each X_i takes on two values 0 and 1. We call the value 1 a* **success** *and the value 0 a* **failure***.*

2. $P\{X_i = 1\} = p$ *for each* i.

3. The outcomes on each of the trials is independent.

For each i,

$$EX_i = 0 \cdot P\{X_i = 0\} + 1 \cdot P\{X_i = 1\} = 0 \cdot (1-p) + 1 \cdot p = p.$$

Let $S = X_1 + X_2 \cdots + X_n$ *be the total number of successes. A sequence having* x *successes has probability*

$$p^x (1-p)^{n-x}.$$

In addition, we have

$$\binom{n}{x}$$

mutually exclusive sequences that have x *successes. Thus, we have the mass function*

$$f_S(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots$$

The fact that $\sum_x f_S(x) = 1$ *follows from the binomial theorem. Consequently,* S *is called a* **binomial random variable**.

Using the linearity of expectation

$$ES = E[X_1 + X_2 \cdots + X_n] = p + p + \cdots + p = np.$$

1.1 Discrete Calculus

Let h be a function on whose domain and range are integers. The **(positive) difference operator**

$$\Delta_+ h(x) = h(x+1) - h(x).$$

If we take, $h(x) = (x)_k = x(x-1)\cdots(x-k+1)$, then

$$\begin{aligned} \Delta_+(x)_k &= (x+1)x\cdots(x-k+2) - x(x-1)\cdots(x-k+1) \\ &= ((x+1)-(x-k+1))(x(x-1)\cdots(x-k+1)) = k(x)_{k-1}. \end{aligned}$$

Thus, the falling powers and the difference operator plays a role similar to the power function and the derivative,

$$\frac{d}{dx} x^k = k x^{k-1}.$$

To find the analog to the integral, note that

$$\begin{aligned} h(n+1) - h(0) &= (h(n+1) - h(x)) + (h(x) - h(x-1)) + \cdots (h(2) - h(1)) + (h(1) - h(0)) \\ &= \sum_{x=0}^{n} \Delta_+ h(x). \end{aligned}$$

For example,

$$\sum_{x=1}^{n} (x)_k = \frac{1}{k+1} \sum_{x=1}^{n} \Delta_+(x)_{k+1} = \frac{1}{k+1}(n+1)_{k+1}.$$

Exercise 3. *Use the ideas above to find the sum to a geometric series.*

Example 4. *Let X be a discrete uniform random variable on $S = \{1, 2, \dots n\}$, then*

$$EX = \sum_{x=1}^{n} x\frac{1}{n} = \frac{1}{n}\sum_{x=1}^{n}(x)_1 = \frac{1}{n}\cdot\frac{(n+1)_2}{2} = \frac{1}{n}\cdot\frac{(n+1)n}{2} = \frac{n+1}{2}.$$

Using the difference operator, $EX(X-1)$ is usually easier to determine for integer valued random variables than EX^2. In this example,

$$\begin{aligned} EX(X-1) &= \sum_{j=1}^{n} x(x-1)\frac{1}{n} = \frac{1}{n}\sum_{j=2}^{n}(x)_2 \\ &= \frac{1}{3n}\sum_{j=2}^{n}\Delta(x)_3 = \frac{1}{3n}(n+1)_3 = \frac{1}{3n}(n+1)n(n-1) = \frac{n^2-1}{3} \end{aligned}$$

We can find $EX^2 = EX(X-1) + EX$ by using the linearity of expectation.

1.2 Geometric Interpretation of Expectation

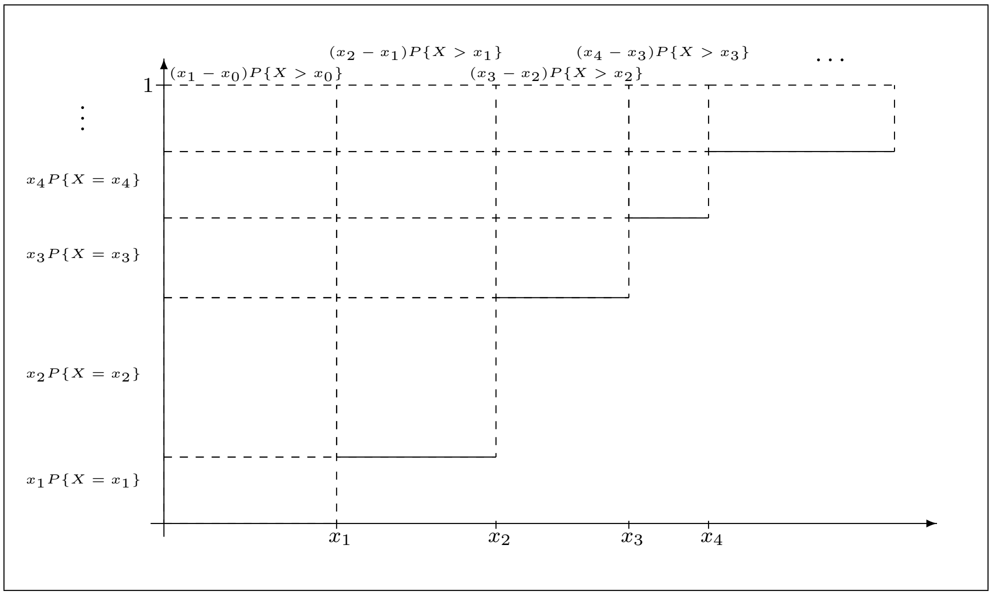

Figure 1: Graphical illustration of EX, the expected value of X, as the area above the cumulative distribution function and below the line $y = 1$ computed two ways.

We can realize the computation of expectation for a nonnegative random variable

$$EX = x_1P\{X = x_1\} + x_2P\{X = x_2\} + x_3P\{X = x_3\} + x_4P\{X = x_4\} + \cdots$$

as the area illustrated in Figure 1. Each term in this sum can be seen as a horizontal rectangle of width x_j and height $P\{X = x_j\}$. This **summation by parts** is the analog in calculus to integration by parts.

We can also compute this area by looking at the vertical rectangle. The j-th rectangle has width $x_{j+1}-x_j$ and height $P\{X > x_j\}$. Thus,

$$EX = \sum_{j=0}^{\infty}(x_{j+1} - x_j)P\{X > x_j\}.$$

if X take values in the nonnegative integers, then $x_j = j$ and

$$EX = \sum_{j=0}^{\infty} P\{X > j\}.$$

Example 5 (geometric random variable). *For a geometric random variable based on the first heads resulting from successive flips of a biased coin, we have that $\{X > j\}$ precisely when the first j coin tosses results in tails*

$$P\{X > j\} = (1-p)^j$$

and thus

$$EX = \sum_{j=0}^{\infty} P\{X \geq j\} = \sum_{j=0}^{\infty}(1-p)^j = \frac{1}{1-(1-p)} = \frac{1}{p}.$$

Exercise 6. *Choose $x_j = (j)_k$ to see that*

$$E(X)_{k+1} = \sum_{j=0}^{\infty}(j)_k P\{X \geq j\}.$$

2 Continuous Random Variables

For X a continuous random variable with density f_X, consider the discrete random variable $\tilde{X}$ obtained from X by rounding down to the nearest multiple of Δx. (Δ has a different meaning here than in the previous section). Denoting the mass function of $\tilde{X}$ by $f_{\tilde{X}}(\tilde{x}) = P\{\tilde{x} \leq X < \tilde{x} + \Delta x\}$, we have

$$\begin{aligned} Eg(\tilde{X}) &= \sum_{\tilde{x}} g(\tilde{x})f_{\tilde{X}}(\tilde{x}) = \sum_{\tilde{x}} g(\tilde{x})P\{\tilde{x} \leq X < \tilde{x} + \Delta x\} \\ &\approx \sum_{\tilde{x}} g(\tilde{x})f_x(\tilde{x})\Delta x \approx \int_{-\infty}^{\infty} g(x)f_X(x)\,dx. \end{aligned}$$

Taking limits as $\Delta x \to 0$ yields the identity

$$Eg(X) = \int_{-\infty}^{\infty} g(x)f_X(x)\,dx. \tag{2}$$

For the case $g(x) = x$, then $\tilde{X}$ is a discrete random variable and so the area above the distribution function and below 1 is equal to $E\tilde{X}$. As $\Delta x \to 0$, the distribution function moves up and in the limit the area is equal to EX.

Example 7. *One solution to finding $Eg(X)$ is to finding f_y, the density of $Y = g(X)$ and evaluating the integral*

$$EY = \int_{-\infty}^{\infty} y f_Y(y)\, dy.$$

However, the direct solution is to evaluate the integral in (2). For $y = g(x) = x^p$ and X, a uniform random variable on $[0, 1]$, we have for $p > -1$,

$$EX^p = \int_0^1 x^p\, dx = \frac{1}{p+1} x^{p+1}\Big|_0^1 = \frac{1}{p+1}.$$

Integration by parts give an alternative to computing expectation. Let X be a positive random variable and g an increasing function.

$$\begin{array}{ll} u(x) = g(x) & v(x) = -(1 - F_X(x)) \\ u'(x) = g'(x) & v(x) = f_X(x) = F_X'(x). \end{array}$$

Then,

$$\int_0^b g(x) f_X(x)\, dx = -g(x)(1 - F_X(x))\Big|_0^b + \int_0^b g'(x)(1 - F_X(x))\, dx$$

Now, substitute $F_X(0) = 0$, then the first term,

$$g(x)(1 - F_X(x))\Big|_0^b = g(b)(1 - F_X(b)) = \int_b^\infty g(b) f_X(x)\, dx \leq \int_b^\infty g(x) f_X(x)\, dx$$

Because, $\int_0^\infty g(x) f_X(x)\, dx < \infty$

$$\int_b^\infty g(x) f_X(x)\, dx \to 0 \text{ as } b \to \infty.$$

Thus,

$$Eg(X) = \int_0^\infty g'(x) P\{X > x\}\, dx.$$

For the case $g(x) = x$, we obtain

$$EX = \int_0^\infty P\{X > x\}\, dx.$$

Exercise 8. *For the identity above, show that it is sufficient to have $|g(x)| < h(x)$ for some increasing h with $Eh(X)$ finite.*

Example 9. *Let T be an exponential random variable, then for some β, $P\{T > t\} = \exp -(t/\beta)$. Then*

$$ET = \int_0^\infty P\{T > t\}\, dt = \int_0^\infty \exp -(t/\beta)\, dt = -\beta \exp -(t/\beta)\Big|_0^\infty = 0 - (-\beta) = \beta.$$

Example 10. *For a normal random variable*

$$EZ = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z \exp(-\frac{z^2}{2})\, dz = 0$$

because the integrand is an odd function.

$$EZ^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 \exp(-\frac{z^2}{2})\, dz$$

To evaluate this integral, integrate by parts

$$\begin{array}{ll} u(z) = z & v(z) = -\exp(-\frac{z^2}{2}) \\ u'(z) = 1 & v'(z) = z\exp(-\frac{z^2}{2}) \end{array}$$

Thus,

$$EZ^2 = \frac{1}{\sqrt{2\pi}} \left(-z\exp(-\frac{z^2}{2})\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \exp(-\frac{z^2}{2})\, dz \right).$$

Use l'Hôpital's rule to see that the first term is 0 and the fact that the integral of a probability density function is 1 to see that the second term is 1.

Using the **Riemann-Stielitjes integral** we can write the expectation in a unified manner.

$$Eg(X) = \int_{-\infty}^{\infty} g(x)\, dF_X(x).$$

This uses limits of Riemann -Stieltjes sums

$$R(g, F) = \sum_{i=1}^{n} g(x_i)\Delta F_X(x_i))$$

For discrete random variables, $\Delta F_X(x_i) = F_X(x_{i+1}) - F_X(x_i) = 0$ if the i-th intreval does not contain a possible value for the random variable X. Thus, the Riemann-Stieltjes sum converges to

$$\sum_x g(x) f_X(x)$$

for X having mass function f_X.

For continuous random variables, $\Delta F_X(x_i) \approx f_X(x_i)\Delta x$. Thus, the Riemann-Stieltjes sum is approximately a Riemann sum for the product $g \cdot f_X$ and converges to

$$\int_{-\infty}^{\infty} g(x) f_X(x)\, dx$$

for X having mass function f_X.