Chapter 4 – Expected Values

4.1 Expected value

For a random variable X, the expected value of X, denoted $E(X)$, is a number (not random) that is an average of all possible experimental outcomes $x \in R_X$ weighted by the probability of seeing the individual outcomes (i.e. weighted by either pmf $p(x)$ or pdf $f(x)$). It is a "typical outcome" of the experiment.

def'n Let X be continuous with pdf $f(x)$. The mean, or *expected value* of X is given by

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

as long as the integral doesn't blow up.

Example: Let $X \sim \exp(\lambda)$.

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x) dx \\ &= \int_{0}^{\infty} x\lambda e^{-\lambda x} dx \\ &= \int_{0}^{\infty} [-e^{-\lambda x}(1+\lambda x)/\lambda]' dx \\ &= [-e^{-\lambda x}(1+\lambda x)/\lambda]_0^{\infty} \\ &= \left[\lim_{x\to\infty} -e^{-\lambda x}(1+\lambda x)/\lambda\right] - -\frac{1}{\lambda} \\ &= 0 + \frac{1}{\lambda} = \frac{1}{\lambda}. \end{aligned}$$

So if X is lifetime of a lightbulb in hours and $X \sim \exp(0.001)$, $E(X) = 1/0.001 = 1000$ hours.

Example: Let $X \sim U(a, b)$.

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x) dx \\ &= \int_a^b \frac{x}{b-a} dx \\ &= \frac{1}{b-a} \int_a^b \left[\frac{1}{2}x^2\right]' dx \\ &= \frac{b^2 - a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{a+b}{2}, \end{aligned}$$

the midpoint between a and b. That makes intuitive sense.

If the waiting time for the bus is $U \sim U(0, 10)$ minutes, on average I'll wait $E(U) = \frac{0+10}{2} = 5$ minutes.

Other expected values:

Distribution	$E(X)$
$X \sim N(\mu, \sigma^2)$	$E(X) = \mu$
$X \sim \text{gamma}(\alpha, \beta)$	$E(X) = \alpha/\beta$
$X \sim \text{beta}(\alpha, \beta)$	$E(X) = \frac{\alpha}{\alpha+\beta}$

See pp. 119-120 for $X \sim$ Cauchy. Here $E(X)$ doesn't exist.

def'n Let X be discrete with pmf $p(x)$. The mean, or *expected value* of X is given by

$$E(X) = \sum_{x \in R_X} xp(x)$$

as long as the sum doesn't diverge.

Example: Let $X \sim \text{Bern}(\pi)$.

$$E(X) = \sum_{x \in R_X} xp(x) = 0p(0) + 1p(1) = 0(1 - \pi) + 1\pi = \pi.$$

Example: Let $X \sim \text{bin}(n, \pi)$.

$$
\begin{aligned}
E(X) &= \sum_{x \in R_X} x p(x) = \sum_{x=0}^{n} x \begin{pmatrix} n \\ x \end{pmatrix} \pi^x (1-\pi)^{n-x} \\
&= 0 + \sum_{x=1}^{n} x \begin{pmatrix} n \\ x \end{pmatrix} \pi^x (1-\pi)^{n-x} \\
&= \sum_{x=1}^{n} \frac{x n!}{(n-x)! x!} \pi^x (1-\pi)^{n-x} \\
&= \sum_{x=1}^{n} n\pi \frac{(n-1)!}{(n-x)!(x-1)!} \pi^{x-1} (1-\pi)^{n-x} \\
&= n\pi \sum_{x=0}^{n-1} \begin{pmatrix} n-1 \\ x \end{pmatrix} \pi^x (1-\pi)^{n-1-x} \\
&= n\pi \times 1 = n\pi.
\end{aligned}
$$

Other expected values:

Distribution	$E(X)$
$X \sim \text{Pois}(\lambda)$	$E(X) = \lambda$
$X \sim \text{Geom}(\pi)$	$E(X) = 1/\pi$
$X \sim \text{hypergeom}(r, n, m)$	$E(X) = m(r/n)$
$X \sim \text{NegBin}(r, \pi)$	$E(X) = r/\pi$

Example: Let X be discrete with $R_X = \{1, 2, 3, 4, \dots\}$ and $p(x) = \frac{6}{(\pi x)^2}$ where here $\pi = 3.14159...$ not a probability. We know that

$$\sum_{x=1}^{\infty} p(x) = \sum_{x=1}^{\infty} \frac{6}{(\pi x)^2} = 1.$$

However,

$$E(X) = \sum_{x=1}^{\infty} x p(x) = \frac{6}{\pi^2} \sum_{x=1}^{\infty} \frac{1}{x} = \infty.$$

The sum diverges and we say that $E(X)$ does not exist.

4.1.1 Expectations of functions of random variables

We've already discussed obtaining the distribution of $Y = g(X)$. A very useful proposition tells us that if alls we care about is $E(Y)$, we can work directly with the distribution of X.

Prop: Let X be continuous with pdf $f(x)$. Then

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx.$$

Prop: Let X be discrete with pmf $p(x)$. Then

$$E(g(X)) = \sum_{x \in R_X} g(x)p(x).$$

Note that in general $E(g(X)) \neq g(E(X))$.

These generalize to a function $Y = g(\mathbf{X}) = g(X_1, \ldots, X_n)$ for jointly distributed $\mathbf{X} = (X_1, \ldots, X_n)$.

Prop: Let $\mathbf{X}$ be continuous with pdf $f(\mathbf{x})$. Then

$$E(g(\mathbf{X})) = \int_{\mathbb{R}^n} g(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}.$$

Prop: Let $\mathbf{X}$ be discrete with pmf $p(\mathbf{x})$. Then

$$E(g(\mathbf{X})) = \sum_{\mathbf{x} \in R_X} g(\mathbf{x}) p(\mathbf{x}).$$

These formulae look intuitive, and in fact they are, but proving them requires some work; see your textbook. These results are collectively referred to as "the law of the unconscious statistician."

Let the pdf of X be $f(x) = 3x^2$ on $R_X = (0, 1)$ (X is beta$(3, 1)$).
Then

$$
\begin{aligned}
E(X) &= \int_0^1 x[3x^2]dx = 3/4. \\
E(X^2) &= \int_0^1 x^2[3x^2]dx = 3/5. \\
E(\sqrt{X}) &= \int_0^1 \sqrt{x}[3x^2]dx = 6/7. \\
E(\sin(X)) &= \int_0^1 \sin(x)[3x^2]dx \approx 0.670. \\
E(e^X) &= \int_0^1 e^x[3x^2]dx = 3(e-2) \approx 2.14. \\
E(a+bX) &= \int_0^1 (a+bx)[3x^2]dx = a + bE(X).
\end{aligned}
$$

Prop: If X and Y are independent,

$$E\{g(X)h(Y)\} = E\{g(X)\}E\{h(Y)\},$$

for any functions $g(x)$ and $h(y)$.

In particular, if X and Y are independent $E(XY) = E(X)E(Y)$.

Let X and Y be independent with the same marginal distribution $f_X(t) = f_Y(t) = 3t^2$ on $R_X = R_Y = (0, 1)$. Then

$$\begin{aligned} E(XY) &= E(X)E(Y) = (3/4)(3/4) = 9/16. \\ E(X^2\sqrt{Y}) &= E(X^2)E(\sqrt{Y}) = (3/5)(6/7) = 18/35. \\ E(X^Y) &= \text{??? (this rule doesn't apply!)} \end{aligned}$$

4.1.2 Expectation of linear combinations of random variables

Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be jointly distributed.

Prop: For any constants $a, b_1, \ldots, b_n$,

$$E\left(a + b_1 X_1 + \cdots + b_n X_n\right) = a + b_1 E(X_1) + b_2 E(X_2) + \cdots + b_n E(X_n),$$

regardless of the joint distribution on $(X_1, \ldots, X_n)$.

Example: Let (X, Y) be jointly distributed with expectations $E(X) = 10$ and $E(Y) = 20$.

$$E(5 + 2X - 0.5Y) = 5 + 2E(X) - 0.5E(Y) = 5 + 2(10) - 0.5(20) = 15.$$

Example: Let $X_1, X_2, \ldots, X_n$ be independent Bern(π) random variables. Then $Y = X_1 + X_2 + \cdots + X_n$ is distributed $Y \sim \text{bin}(n, \pi)$ *by definition.*

$$E(Y) = E(X_1 + \cdots X_n) = E(X_1) + \cdots + E(X_n) = \pi + \cdots + \pi = n\pi.$$

Example: Let $X_1, X_2, \ldots, X_n$ be independent geom(π) random variables. Then $Y = X_1 + X_2 + \cdots + X_r$ is distributed $Y \sim \text{NegBin}(r, \pi)$ *by definition.*

$$E(Y) = E(X_1 + \cdots X_r) = E(X_1) + \cdots + E(X_r) = 1/\pi + \cdots + 1/\pi = r/\pi.$$

Example C, pp. 128-129: group testing

k blood samples are combined into a pool and tested for a disease. If the test is negative, everyone in the group is disease free and only one test need be performed. If the test is positive, each sample must be retested resulting in $k+1$ tests overall. This process is repeated m times, so there are mk individuals total being tested.

What is the expected number of tests assuming the probability that any individual is diseased is p?

p is the disease prevalence in the population.

Let X_i be the number of tests in the i^{th} group. $X_i = 1$ if all k individuals in the i^th group are negative, otherwise $X_i = k + 1$. Let Y_i be the number of people in group i that are infected; $Y_i \sim \text{bin}(k, p)$. Then

$$P(X_i = 1) = P(Y_i = 0) = (1 - p)^k.$$

$$P(X_i = k + 1) = P(Y_i \geq 1) = 1 - (1 - p)^k.$$

So we have for each i,

$$E(X_i) = 1 \times (1 - p)^k + (k + 1) \times (1 - (1 - p)^k).$$

Let $N = \sum_{i=1}^{m} X_i$ be the total number of tests performed.

$$\begin{aligned} E(N) &= E(X_1 + \cdots + X_m) \\ &= E(X_1) + \cdots + E(X_m) = mE(X_i) \\ &= m[(1 - p)^k + (k + 1)(1 - (1 - p)^k)]. \end{aligned}$$

The non-pooling approach simply uses one test for each individual resulting in mk tests. The idea is to save money by pooling blood and only using one test, but the savings will be affected by the number being pooled k, and the prevalence p.

Often tests are performed on pools of size $k = 6$ or $k = 12$. Let's look at $m = 10$ groups of $k = 6$. Say the disease is rare, $p = 0.01$. Then

$$E(N) = 10[(0.99)^6 + (6 + 1)(1 - (0.99)^6)] = 13.51,$$

a substantial savings over $mk = 60$ tests. However, if the disease is common, say $p = 0.3$, then

$$E(N) = 10[(0.6)^6 + (6 + 1)(1 - (0.6)^6)] = 67.2,$$

and on average it's cheaper just to give everyone their own test.

4.2 Variance and standard deviation

We have talked about two measures of "central tendency" for a random variable X, the median $M(X)$, and the mean $E(X)$.

For continuous random variables with densities symmetric about m they are the same $E(X) = M(X) = m$,

$$F^{-1}(0.5) = \int_{-\infty}^{\infty} x f(x) dx = m,$$

as long as $E(X)$ exists. $M(X) = F^{-1}(0.5)$ always exists.

Example: For $X \sim U(a, b)$, the density $f(x) = (b-a)^{-1}$ on $R = (a, b)$ is symmetric about $(a + b)/2$ so $M(X) = E(X) = (a + b)/2$.

Example: For $X \sim N(\mu, \sigma^2)$, the density $f(x) = \exp\{-0.5(x-\mu)^2/\sigma^2\}/\sqrt{2\pi\sigma^2}$ is symmetric about about μ so $M(X) = E(X) = \mu$.

Besides a number summarizing a "typical value" of X, we are often interested in how "spread out" X can be. The length of the range $|R|$ gives some idea, but often $|R| = \infty$ (e.g. normal, gamma, Weibull, exponential) or remains constant for wildly different members within the same family (e.g. beta$(1, 1)$, beta$(0.1, 0.1)$, beta$(10, 10)$ all have range length $|R| = 1$.)

One measure of spread based on quantiles is the *interquartile range* (IQR), the distance between $x_{0.25}$ and $x_{0.75}$. For X with cdf F, $\text{IQR}(X) = F^{-1}(0.75) - F^{-1}(0.25)$. The $\text{IQR}(X)$ is never infinite and gives the length of the interval that encompasses the middle 50% of experimental outcomes X.

The most common measure of spread is the *standard deviation* of X, obtained from the *variance*.

def'n: The variance of X, written $\text{Var}(X)$ a weighted average of how far X is from $E(X)$ squared:

$$\text{Var}(X) = E\{(X - E(X))^2\}.$$

$\text{Var}(X)$ has the same units as X but squared. Taking the square root of the variance gives a measure of spread in the same units as X:

def'n: The standard deviation of X, written $\text{sd}(X)$ is $\text{sd}(X) = \sqrt{\text{Var}(X)}$.

Often, for any X, we denote the mean and variance of X using the Greek letters μ and σ^2

$$\mu = E(X) \text{ and } \sigma^2 = \text{Var}(X),$$

and so the standard deviation is just σ.

Note that for $X \sim N(\mu, \sigma^2)$, the distribution of X is parameterized in terms of its mean and variance right off the bat. That is, for $X \sim N(\mu, \sigma^2)$, $E(X) = \mu$ and $\text{Var}(X) = \sigma^2$.

Another measure of spread is the average absolute deviation of X about the mean $E\{|X - E(X)|\}$. This is probably a more natural thing to look at than σ, but is in general hard to compute and (μ, σ^2) have very nice mathematical properties we'll explore later on.

Can you think of other measures of spread? How about $E\{|X - M(X)|\}$, or better yet $M\{|X - M(X)|\}$? Another: $x_{0.9} - x_{0.1}$.

Recall that if $X \sim \text{Weibull}(\alpha, \lambda)$ then

$$\begin{aligned} f(x) &= \alpha\lambda^{-\alpha}x^{\alpha-1}e^{-\left(\frac{x}{\lambda}\right)^{\alpha}}, \\ F(x) &= 1 - e^{-\left(\frac{x}{\lambda}\right)^{\alpha}}, \\ F^{-1}(p) &= \lambda\{-\log(1-p)\}^{1/\alpha}, \end{aligned}$$

with range $R = (0, \infty)$.

Let $X \sim \text{Weibull}(2, 4)$ months. Then to two decimal places $E(X) = 3.54$ months and $M(X) = F^{-1}(0.5) = x_{0.5} = 3.33$ months. Also,

$$\text{IQR} = x_{0.75} - x_{0.25} = F^{-1}(0.75) - F^{-1}(0.25) = 4.71 - 2.15 = 2.56$$

months and $\text{sd}(X) = 1.85$ months.

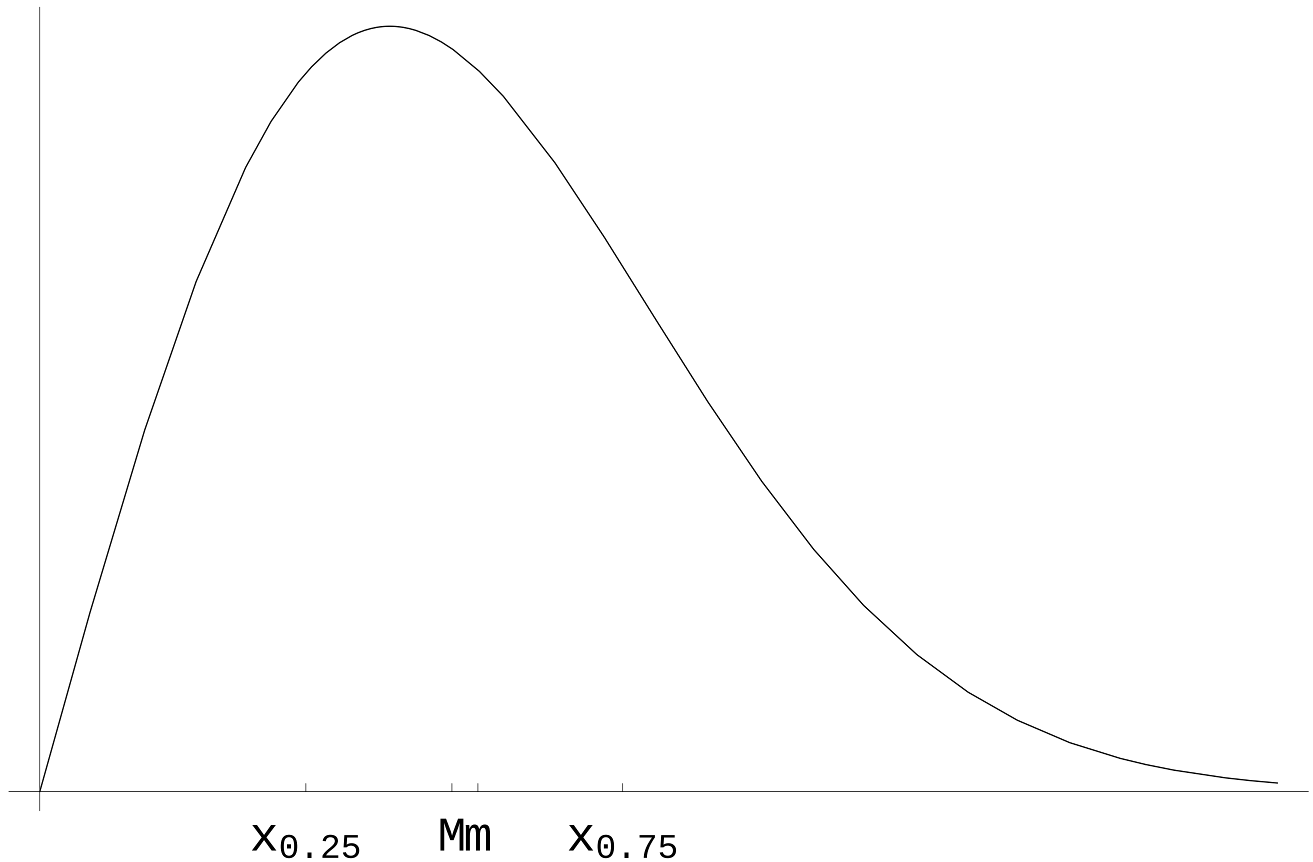

Figure 1: Weibull$(2, 4)$ density with $x_{0.25}$, $x_{0.75}$, $M = x_{0.5}$ and μ.

For any random variable X, let $\mu = E(X)$ and $\sigma^2 = \text{Var}(X)$. If we're talking about more than one random variable, say (X, Y), we can add subscripts to denote the marginal mean and variance μ_X and σ_X^2.

Prop: Let a and b be fixed numbers and X a random variable. Then the variance of the new random variable $a + bX$ is

$$\text{Var}(a + bX) = b^2 \text{Var} X = (b\ \sigma)^2.$$

So $\text{sd}(a + bX) = b\ \sigma$.

Similarly, $\text{IQR}(a + bX) = b\ \text{IQR}(X)$.

Prop: $\text{Var}(X) = E(X^2) - \{E(X)\}^2 = E(X^2) - \mu^2$.

The proofs of these last two propositions are in your book (p. 131 and p. 133).

Example: Let $X \sim \text{beta}(2,3)$. Then $f(x) = 12x(1-x)^2$ on $R = (0,1)$.

$$\begin{aligned}
\mu = E(X) &= \int_0^1 x[12x(1-x)^2]dx = \frac{2}{5} \\
E(X^2) &= \int_0^1 x^2[12x(1-x)^2]dx = \frac{1}{5} \\
\sigma^2 = E\left\{\left(X - \frac{2}{5}\right)^2\right\} &= \int_0^1 \left(x - \frac{2}{5}\right)^2 [12x(1-x)^2]dx = \frac{1}{25}.
\end{aligned}$$

From the definition, we see $\sigma^2 = 1/25$. We can use the proposition to get this another way:

$$\sigma^2 = E(X^2) - \mu^2 = \frac{1}{5} - \left(\frac{2}{5}\right)^2 = \frac{5-4}{25} = \frac{1}{25}, \quad \text{so } \sigma = \frac{1}{5}.$$

Variances from common distributions

$X \sim \exp(\lambda)$	$\text{Var}(X) = 1/\lambda^2$
$X \sim U(a, b)$	$\text{Var}(X) = (b-a)^2/12$
$X \sim N(\mu, \sigma^2)$	$\text{Var}(X) = \sigma^2$
$X \sim \text{gamma}(\alpha, \beta)$	$\text{Var}(X) = \alpha/\beta^2$
$X \sim \text{beta}(\alpha, \beta)$	$\text{Var}(X) = \alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$
$X \sim \text{bin}(n, \pi)$	$\text{Var}(X) = n\pi(1-\pi)$
$X \sim \text{Pois}(\lambda)$	$\text{Var}(X) = \lambda$
$X \sim \text{Geom}(\pi)$	$\text{Var}(X) = (1-\pi)/\pi^2$

There are expressions for $\text{Var}(X)$ for hypergeometric, Weibull, and negative binomial distributions as well if needed.