Expectation

Chapter 7

Definitions and Examples

Properties

- Transformations
- Linearity
- Monotonicity
- Expectation and Independence

Expectation

The Definition. The mean of a random variable X, say, is also called its *expectation* and denoted by $E(X)$.

The Discrete Case. If X is discrete with PMF f, then

$$E(X) = \sum_{x \in \mathcal{X}} x f(x), \tag{1}$$

where $\mathcal{X}$ is the range of X, provided that the sum converges absolutely.

The Continuous Case. If X has density f, then

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx, \tag{2}$$

provided that the integral converges absolutely.

Notation: Combine (1) and (2) by writing,

$$E(X) = \int_{-\infty}^{\infty} x dF(x).$$

Notes:
- Shorthand notation.
- Riemann-Stieltjes Integral.

Examples

Poisson Distributions. If $X \sim \text{Poisson}(\lambda)$, then

$$f(x) = \frac{1}{x!}\lambda^x e^{-\lambda}$$

for $x = 0, 1, 2 \cdots$, and

$$\begin{aligned} E(X) &= \sum_{x=0}^{\infty} x \frac{1}{x!}\lambda^x e^{-\lambda} \\ &= \lambda \sum_{x=1}^{\infty} \frac{1}{(x-1)!}\lambda^{x-1} e^{-\lambda} \\ &= \lambda \sum_{x=0}^{\infty} \frac{1}{x!}\lambda^x e^{-\lambda} = \lambda. \end{aligned}$$

Exponential Distributions. If $X \sim \text{Exp}(\lambda)$, then $f(x) = \lambda e^{-\lambda x}$ for $0 \le x < \infty$ and

$$\begin{aligned} E(X) &= \int_0^{\infty} x \lambda e^{-\lambda x} dx \\ &= \frac{1}{\lambda}\int_0^{\infty} x e^{-x} dx \\ &= \frac{1}{\lambda}\Gamma(2) = 1. \end{aligned}$$

Symmetric Distributions

If X has density f for which

$$f(-x) = f(x),$$
$$\int_{-\infty}^{\infty} |x| f(x) dx < \infty,$$

then

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = 0.$$

Example: The Standard Normal Density.

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

Example
The Standard Cauchy

If

$$f(x) = \frac{1}{\pi(1+x^2)},$$

then

$$\begin{aligned}\int_{-c}^{c} |x| f(x) dx &= 2\int_0^c \frac{x dx}{\pi(1+x^2)} \\ &= \frac{1}{\pi}\log(1+c^2) \\ &\to \infty.\end{aligned}$$

So,

$$\int_{-\infty}^{\infty} |x| f(x) dx = \infty,$$

and the expectation is not defined.

Transformations

Suppose

$$X_1, \cdots, X_m \sim f \text{ joint density or PMF.}$$

Let

$$Y = w(X_1, \cdots, X_m).$$

Theorem. If f is a PMF, then

$$E(Y) = \sum_{\mathbf{x} \in \mathcal{X}} w(\mathbf{x}) f(\mathbf{x}),$$

provided that the sum converges absolutely. If f is a density, then

$$E(Y) = \int_{I\!R^m} w(\mathbf{x}) f(\mathbf{x}) dx,$$

provided that the integral converges absolutely.

Note: By definition,

$$E(Y) = \int_{-\infty}^{\infty} y dF_Y(y).$$

An Example
Distance Between Two Points

Suppose

$$X, Y \sim^{ind} \text{Unif}[0,1].$$

Then

$$f(x,y) = \begin{cases} 1 & if\ 0 \le x, y \le 1 \\ 0 & if\ otherwise \end{cases}$$

Let

$$D = |Y - X|.$$

Then

$$E(D) = \int_0^1 \big[\int_0^1 |y-x| dy\big] dx.$$

Here

$$\begin{aligned}\int_0^1 |y-x| dy &= \int_0^x (x-y) dy \\ &\quad + \int_x^1 (y-x) dy \\ &= -\frac{1}{2}(y-x)^2|_{y=0}^{x} + \frac{1}{2}(y-x)^2|_{y=x}^{1} \\ &= \frac{1}{2}[x^2 + (1-x)^2].\end{aligned}$$

So,

$$\begin{aligned}E(D) &= \frac{1}{2}\int_0^1 [x^2 + (1-x)^2] dx \\ &= \frac{1}{6}[x^3 - (1-x)^3]|_{x=0}^{1} \\ &= \frac{1}{3}.\end{aligned}$$

Example

Towards a Derivation

Coins

PQPNDPQQDP

$$1+1+5+10+1+25+25+10+1=104$$

Or

P	4	$= 4$
N	1	$= 5$
D	2	$= 20$
Q	3	$= 75$
		$= 104$

Transformations

Derivation: The Discrete Case

The PMF of $Y = w(X)$

$$f_Y(y) = P[Y=y] = \sum_{x:w(x)=y} f_X(x).$$

So,

$$\begin{aligned} E(Y) &= \sum_{y\in\mathcal{Y}} y f_Y(y) \\ &= \sum_{y\in\mathcal{Y}} y \left[\sum_{x:w(x)=y} f_X(x)\right] \\ &= \sum_{y\in\mathcal{Y}} \left[\sum_{x:w(x)=y} w(x) f_X(x)\right] \\ &= \sum_{x\in\mathcal{X}} w(x) f_X(x). \end{aligned}$$

Properties of Expectation

Linearity

Theorem. If X and Y are JDRVs with finite expectations and if $a, b \in I\!R$, then

$$E(aX+bY) = aE(X) + bE(Y).$$

Proof-In the Discrete Case. Let f be the joint PMF. Then

$$E(aX+bY) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} (ax+by) f(x,y)$$

by the Transformation Theorem applied to

$$w(x,y) = ax+by.$$

Next,

$$\begin{aligned} &\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} (ax+by) f(x,y) \\ &\qquad = a\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} x f(x,y) \\ &\qquad = b\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} y f(x,y) \\ &\qquad = aE(X) + bE(Y), \end{aligned}$$

by the Transformation Theorem again, applied to $w(x,y) = x$ and $w(x,y) = y$.

Corollary. If $X_1, \cdots, X_n$ are JDRVs with expectations and $c_1, \cdots, c_n \in I\!R$, then

$$\begin{aligned} &E(c_1X_1 + \cdots + c_nX_n) \\ &\qquad = c_1E(X_1) + \cdots + c_nE(X_n). \end{aligned}$$

Counting Variables

If $A_1, \cdots, A_n$ are events and

$$X = \mathbf{1}_{A_1} + \cdots + \mathbf{1}_{A_n},$$

then

$$E(X) = P(A_1) + \cdots + P(A_n).$$

Example: Binomial. If $A_1, \cdots, A_n$ are independent with $P(A_i) = p$, then

$$X \sim \text{Binomial}(n, p),$$
$$E(X) = p + \cdots + p = np.$$

Example

Hypergeometric

Suppose that a simple random sample of size n is drawn *w.o.r.* from R red and $N - R$ white tickets. Let

$$A_k = \{k^{th} \text{ ticket drawn is red}\},$$
$$X = \mathbf{1}_{A_1} + \cdots + \mathbf{1}_{A_n}.$$

Then

$$P(A_k) = \frac{R}{N},$$

and

$$E(X) = n\frac{R}{N},$$

even though $A_1, \cdots, A_n$ are not independent.

Note: $P[X = r] = \binom{R}{r}\binom{N-R}{n-r}/\binom{N}{n}$; and

$$\sum_{r=0}^{n} rP[X = r] = n\frac{R}{N}.$$

The Mean and Variance

If $X \sim F$ is a RV with a finite expectation,

$$\mu = E(X) = \int_{-\infty}^{\infty} x dF(x)$$

depends only on F. It is called the *mean of X and/or F*. Then

$$D^2(X) = E[(X - \mu)^2]$$

is called the *variance of X*. Here

$$E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 dF(x)$$

also depends only on F. It is called the *variance of F and/or X* and denoted by

$$\sigma^2 = D^2(X).$$

Also,

$$\sigma = \sqrt{\sigma^2}$$

is called the *standard deviation of X and/or F*.

Properties of the Variance

Theorem. Let $\mu = E(X)$. Then, for any c,

$$E[(X - c)^2] = \sigma^2 + (c - \mu)^2.$$

Proof. We have

$$\begin{aligned}(X - c)^2 = (X - \mu)^2 &+ 2(\mu - c)(X - \mu) \\ &+ (c - \mu)^2.\end{aligned}$$

So

$$\begin{aligned}E[(X - c)^2] &= E[(X - \mu)^2] \\ &\quad + 2(\mu - c)E(X - \mu) + E[(c - \mu)^2] \\ &= \sigma^2 + 2(\mu - c) \times 0 + (c - \mu)^2.\end{aligned}$$

Corollary. $E[(X - c)^2]$ is minimized by $c = \mu$, and the minimum value is σ^2.

Corollary.

$$\sigma^2 = E(X^2) - \mu^2.$$

Moments

If $X \sim F$, then

$$\mu_k = E(X^k) = \int_{-\infty}^{\infty} x^k dF(x)$$

are called the *moments of X and/or F*, (provided that they exist). Thus,

$$\mu = \mu_1,$$
$$\sigma^2 = \mu_2 - \mu_1^2.$$

Example: Uniform. If $X \sim \text{Unif}[0, 1]$, then

$$\mu_k = \int_0^1 x^k dx = \frac{1}{k+1},$$
$$\mu = \frac{1}{2},$$
$$\sigma^2 = \frac{1}{3} - (\frac{1}{2})^2 = \frac{1}{12}.$$

Properties of Expectation
Monotonicity

If X and Y are JDRVs with expectations for which

$$P[X \le Y] = 1,$$

then

$$E(X) \le E(Y).$$

Proof. First, if $P[Y \ge 0] = 1$, then $P[Y < 0] = 0$ and, therefore,

$$E(Y) = \int_0^{\infty} y dF_Y(y) \ge 0.$$

For the general case,

$$E(Y - X) \ge 0$$

and, therefore,

$$E(Y) \ge E(X).$$

A Simple Identity

$$\left[\sum_{i=1}^{m} a_i\right] \times \left[\sum_{j=1}^{n} b_j\right] = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j.$$

Example

$$\begin{aligned}&(a_1 + a_2) \times (b_1 + b_2)\\ &\qquad = a_1b_1 + a_1b_2 + a2b_1 + b_2a_2.\end{aligned}$$

Proof. Use induction.

Properties of Expectation
Expectation and Independence

Theorem. If X and Y are indpendent RVs with expectations, then

$$E(XY) = E(X)E(Y).$$

Proof-In the Discrete Case. If X and Y are independent, then

$$f(x, y) = f_X(x)f_Y(y).$$

So,

$$\begin{aligned}E(XY) &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} xyf(x, y)\\ &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} xyf_X(x)f_Y(y)\\ &= \Big[\sum_{x\in\mathcal{X}} xf_X(x)\Big]\Big[\sum_{y\in\mathcal{Y}} yf_Y(y)\Big]\\ &= E(X)E(Y).\end{aligned}$$

Independent Random Variables

If $X_1, \cdots, X_n$ are independent with finite means and variances, then

$$E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n) \quad (1)$$

and

$$D^2(X_1 + \cdots + X_n) = D^2(X_1) + \cdots + D^2(X_n). \quad (2)$$

Proof. (1) follows from linearity and does not require independence. For (2) suppose that $n = 2$ and let $E(X_i)$, so that $D^2(X_i) = E(X_i^2)$. Then

$$\begin{aligned} D^2(X_1 + X_2) &= E[(X_1 + X_2)^2] \\ &= E(X_1^2 + 2X_1X_2 + X_2^2) \\ &= E(X_1^2) + 2E(X_1X_2) + E(X_2^2) \\ &= E(X_1^2) + 2E(X_1)E(X_2) + E(X_2^2) \\ &= E(X_1^2) + E(X_2^2) \\ &= D^2(X_1) + D^2(X_2). \end{aligned}$$