Expected value

The *expected value or mean* of a discrete random variable X is defined to be

$$\mathbb{E}[X] = \sum_i x_i \, P(X = x_i)$$

weighted sum of possible values of X.

where the sum runs over the possible values of X.

The operational meaning is that $\mathbb{E}[X]$ is the long-run average value of repeated measurements of the random variable X. That is, suppose that we measure the random variable X in N independent trials, and record the results as $X_1, X_2, \ldots, X_N$. Then the long-run average value is

$$\overline{X}_N = N^{-1} \, (X_1 + X_2 + \cdots + X_N)$$

We will shortly see the Law of Large Numbers which implies that

$$\lim_{N \to \infty} \overline{X}_N = \mathbb{E}[X]$$

For example, if X is the outcome of rolling a die, then the numbers $\{1, 2, \ldots, 6\}$ all occur with probability $1/6$. So the expected value is

$$\mathbb{E}[X] = 1(1/6) + 2(1/6) + \cdots + 6(1/6) = 7/2$$

Example 12 *Roll three fair dice. Let X be the number of different faces that appear. Find the pdf of X and compute $\mathbb{E}[X]$.*

Example 13 *A biased coin has probability p of coming up Heads. Toss the coin until Heads first appears, let N be the number of tosses needed. The pdf of N is a geometric distribution. Compute* $\mathbb{E}[N]$.

$p = \mathbb{P}(H) =$ Prob. of Heads

$1-p = \mathbb{P}(T) =$ Prob. of Tails.

T T T T H $\qquad N = 5$.

$\mathbb{P}(N=5) = \mathbb{P}(TTTTH)$ $\qquad$ geometric r.v.

(pdf is in table).

$= \mathbb{P}(T)\,\mathbb{P}(T)\,\mathbb{P}(T)\,\mathbb{P}(T)\,\mathbb{P}(H)$ $\quad$ (independent)

$= (1-p)^4\, p$

pdf $\quad \mathbb{P}(N=k) = (1-p)^{k-1}\, p \qquad (k=1,2,3\ldots)$

Compute

$$\mathbb{E}[N] = \sum_{k=1}^{\infty} k\, \mathbb{P}(N=k)$$

$$= \sum_{k=1}^{\infty} k\,(1-p)^{k-1} \cdot p$$

Basic idea for infinite series:

geometric series:
$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$$

for $-1 < x < 1$

Cheat:
$$\frac{d}{dx} \sum_{k=0}^{\infty} x^k = \frac{d}{dx}\left(\frac{1}{1-x}\right)$$

$$\sum_{k=0}^{\infty} k x^{k-1} = \frac{1}{(1-x)^2}$$

Apply $x = 1-p$.

$$\sum_{k=1}^{\infty} k(1-p)^{k-1} = \frac{1}{p^2}$$

$$\Rightarrow E[N] = \sum_{k=1}^{\infty} k(1-p)^{k-1} \cdot p$$

$$= \left(\frac{1}{p^2}\right) p$$

$$= \frac{1}{p}.$$

e.g. fair coin $p = \frac{1}{2}$

$\Rightarrow E[N] = 2.$

[Comment: for HW #1 you need

Ex. 4.

pdf $\quad P(X=n) = (1-p)^2 n p^{n-1}$

$(n = 1, 2, 3, \dots)$

$$E[X] = \sum_{n=1}^{\infty} n (1-p)^2 n p^{n-1}$$

$$\sum_{n=1}^{\infty} n^2 p^{n-1}$$

Know $\quad \displaystyle\sum_{n=1}^{\infty} n p^{n-1} = \frac{1}{(1-p)^2}$

Apply $\frac{d}{dp}$ to both sides ....

$$\sum_{n=1}^{\infty} \underbrace{n(n-1)} p^{n-2} = \frac{d}{dp} \frac{1}{(1-p)^2}$$

want $n^2 = n(n-1) + \ldots$

Example 14 *A function of a random variable: $Y = g(X)$. Then*

e.g. $Y = X^2$

$Y = e^{tX}$

⋮

$$P(Y = y) = \sum_{k:y=g(x_k)} P(X = x_k)$$

The expected value of Y can be computed like this:

$$\begin{aligned}
\mathbb{E}[Y] &= \sum_y y\, P(Y = y) \\
&= \sum_y \sum_{k:g(x_k)=y} g(x_k)\, P(X = x_k) \\
&= \sum_y \sum_k 1_{g(x_k)=y}\, g(x_k)\, P(X = x_k) \\
&= \sum_k g(x_k) P(X = x_k) \sum_y 1_{g(x_k)=y} \\
&= \sum_k g(x_k) P(X = x_k)
\end{aligned}$$

So the result is

$$\boxed{\mathbb{E}[g(X)] = \sum_k g(x_k) P(X = x_k)}$$

weighted sum of possible values of $g(X)$.

Important! $\mathbb{E}[g(X)] \neq g(\mathbb{E}[X])$.

Ex. Roll 3 dice

X = # different faces a dice.

X	1	2	3
prob.	$\frac{1}{36}$	$\frac{15}{36}$	$\frac{20}{36}$

$$\mathbb{E}[X] = (1)\left(\frac{1}{36}\right) + (2)\left(\frac{15}{36}\right) + 3\left(\frac{20}{36}\right).$$

$$\mathbb{E}[X^2] = (1)^2 \frac{1}{36} + (2)^2\left(\frac{15}{36}\right) + (3)^2\left(\frac{20}{36}\right).$$

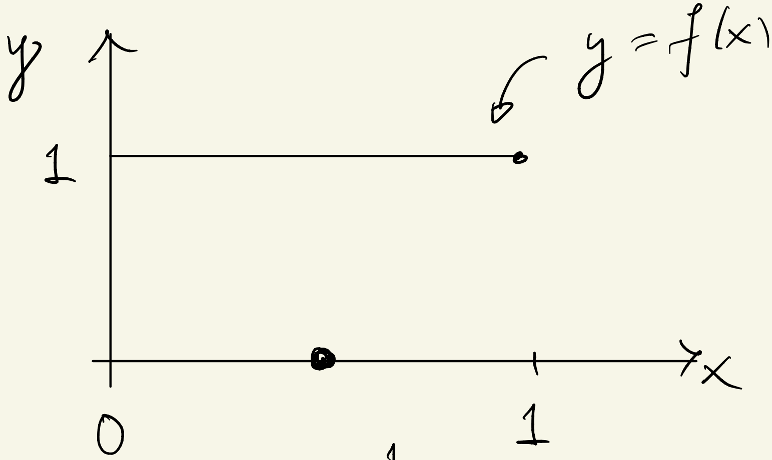
$$\mathbb{E}[2X-1] = [2(1)-1] \cdot \frac{1}{36} + [2(2)-1]\left(\frac{15}{36}\right)$$

$$+ [2(3)-1]\left(\frac{20}{36}\right).$$

__Ex.__ $U \sim$ uniform on $[0,1]$.

pdf. $$f(x) = \begin{cases} 1 & \text{for } 0 \le x \le 1 \\ 0 & \text{else} \end{cases}$$

Continuous r.v.



$$\mathbb{E}[U] = \int_0^1 \underbrace{x \, f(x)}\, dx$$

weighted sum of possible values of U

$$= \int_0^1 x \cdot 1 \cdot dx = \frac{1}{2}.$$

$$\mathbb{E}[U^2] = \int_0^1 x^2 \, f(x) \, dx$$

apply same function

$$= \int_0^1 x^2 \cdot 1 \cdot dx$$

$$= \frac{1}{3}.$$

One very important property of the expected value is linearity, meaning that for any random variables X, Y and any real numbers a, b,

$$\mathbb{E}[aX + bY] = a\,\mathbb{E}[X] + b\,\mathbb{E}[Y]$$

This extends to a sum over any finite collection of r.v.'s:

$$\mathbb{E}[a_1X_1 + \cdots + a_nX_n] = a_1\mathbb{E}[X_1] + \cdots + a_n\mathbb{E}[X_n]$$

Always true! Don't need independence

you always have

mean of sum = sum of the means.

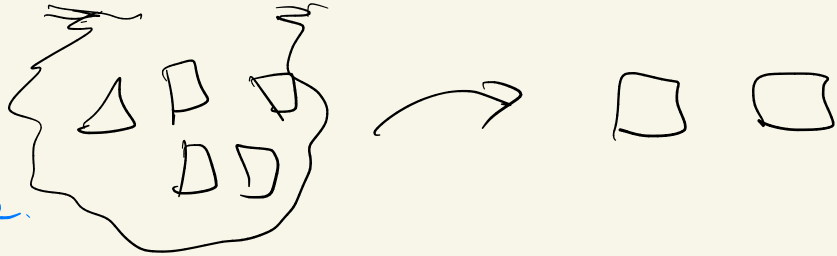
Ex. Suppose have 5 blocks.

Two blocks are Red, three blocks are Blue.

Put them into a bag and randomly draw out 2 blocks.

2 Red

3 Blue



X = number of Red blocks which we draw.

$\mathrm{Ran}(X) = \{0, 1, 2\}$.

$\mathbb{E}[X] = \, ?$

X	0	1	2
prob.	$\frac{6}{20}$	$\frac{12}{20}$	$\frac{2}{20}$

$\mathbb{P}(X=2) =$ Prob. (first draw is Red, second draw is Red)

$$= \left(\frac{2}{5}\right)\left(\frac{1}{4}\right).$$

$$\mathbb{E}[X] = 0 \cdot \left(\frac{6}{20}\right) + 1 \cdot \left(\frac{12}{20}\right) + 2 \cdot \left(\frac{2}{20}\right)$$

Different way:

$$R_1 = \begin{cases} 1 & \text{if first block Red} \\ 0 & \text{if not} \end{cases}$$

$$R_2 = \begin{cases} 1 & \text{if second block Red} \\ 0 & \text{if not} \end{cases}$$

$$X = R_1 + R_2$$

$\boxed{R}$ $\boxed{B}$

$R_1 = 1$ $R_2 = 0$ $X = 1$

Use linearity of $\mathbb{E}[\cdot]$:

$$\mathbb{E}[X] = \mathbb{E}[R_1 + R_2]$$

$$= \mathbb{E}[R_1] + \mathbb{E}[R_2]$$

R_1	0	1
prob.	$\frac{3}{5}$	$\frac{2}{5}$

$$\mathbb{E}[R_1] = \frac{2}{5}$$

$$\mathbb{E}[R_2] = \frac{2}{5}$$

we will come back to this !!

Variance

The mean $\mathbb{E}[X]$ is called the first order statistic of X. The second order statistic is the *variance*, defined by

$$\mathrm{VAR}[X] = \mathbb{E}\left[X - \mathbb{E}[X]\right]^2 = \sum_i (x_i - \mathbb{E}[X])^2 \, P(X = x_i)$$

Some elementary algebra yields

$$\mathrm{VAR}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

The standard deviation is defined as the square root of the variance:

$$\mathrm{STD}[X] = \sqrt{\mathrm{VAR}[X]}$$

$$\mathrm{VAR}[X] = \mathbb{E}\left[\left(X - \mathbb{E}[X]\right)^2\right]$$

$$= \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

measure the spread of the pdf around the mean

Standard deviation $= \sqrt{\mathrm{VAR}[X]}$

same units as X.

Example 15 *Geometric: toss biased coin until first appearance of Heads, N is number of tosses needed. Saw that*

$$\mathbb{E}[N] = \frac{1}{p}$$

where p is the bias. Compute

$$\mathbb{E}[N(N-1)] = \frac{2(1-p)}{p^2}$$

and deduce VAR$[N]$.

$$\mathbb{P}(N=k) = (1-p)^{k-1} \cdot p \qquad (k=1,2,3,\ldots)$$

$$\mathbb{E}[N] = \sum_{k=1}^{\infty} k\,(1-p)^{k-1} \cdot p$$

$$\mathbb{E}[N^2] = \sum_{k=1}^{\infty} k^2\,(1-p)^{k-1} \cdot p$$

Then $\mathrm{VAR}[N] = \mathbb{E}[N^2] - (\mathbb{E}[N])^2$

Joint pdf: Second order statistics

In addition to the means $\mathbb{E}[X]$ and $\mathbb{E}[Y]$ and the variances, there is another statistic which measures the relation between X and Y: the *Covariance* is

$$\mathrm{COV}[X,Y] = \mathbb{E}(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

The units depend on X and Y, so convenient to define the dimensionless *correlation coefficient*:

$$\mathrm{CORR}[X,Y] = \frac{\mathrm{COV}[X,Y]}{\mathrm{STD}[X]\,\mathrm{STD}[Y]}$$

Can show that $-1 \leq \mathrm{CORR}[X,Y] \leq 1$.

$$\mathrm{COV}[X,Y] = \underbrace{\mathbb{E}[XY]}_{} - \mathbb{E}[X]\,\mathbb{E}[Y]$$

use joint pdf to compute this.

For X, Y discrete r.v.'s, the joint pdf is the list of probs for all possible pairs of values.

$$P(X = x_1, Y = y_1),\ P(X = x_2, Y = y_1), \ldots$$

$$\mathbb{E}[XY] = \sum_{x_i, y_j} \underline{x_i y_j}\ \underline{P(X = x_i, Y = y_j)}$$

sum over all possible values for pair (X, Y)

XY

$\rightarrow x_i y_j$

weighting by joint pdf.

Joint pdf: Independence

Recall that two events A, B are *independent* if

$$\mathbb{P}(A \,\text{and}\, B) = \mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$$

The discrete random variables X and Y are *independent* if the events $\{X = x_i\}$ and $\{Y = y_j\}$ are independent for every pair of possible values (x_i, y_j). Equivalently, the joint pdf is the product of the marginal pmf's:

$$P(X = x_i, Y = y_j) = P(X = x_i)\,P(Y = y_j) \quad \text{for all } (x_i, y_j)$$

This definition extends immediately to the continuous case: the two continuous r.v.'s X and Y are independent if the joint pdf is the product of the marginals, that is if

$$f_{X,Y}(x, y) = f_X(x)\,f_Y(y) \quad \text{for all } x, y$$

Similarly, the r.v.'s $X_1, X_2, \ldots, X_n$ are independent if

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = P(X_1 = x_1)\,P(X_2 = x_2) \ldots P(X_n = x_n)$$

for all possible values $x_1, \ldots, x_n$.

joint pdf of independent r.v.'s.

= product of the marginal pdf's.

Bottom Line:

independent r.v.'s naturally arise in many problems, and we can exploit independence to simplify calculations.

Example 16 *Coin tosses are independent. Suppose we toss a coin twice and*

$$X = \begin{cases} 1 & \textit{if first toss is Heads} \\ 0 & \textit{first toss is Tails} \end{cases}$$

indicator r.v.

$$Y = \begin{cases} 1 & \textit{if second toss is Heads} \\ 0 & \textit{second toss is Tails} \end{cases}$$

Suppose the coin is biased and p is the probability of Heads, so $1-p$ is the probability of Tails. Then for example

$$\mathbb{P}(X=1, Y=0) = \mathbb{P}(X=1)\,\mathbb{P}(Y=0) = p(1-p)$$

H T

$X=1$ $Y=0$

Pair $(X,Y) = (1,0)$.

Outcome.

Prob. to get these values:

$$\mathbb{P}(X=1, Y=0) = \mathbb{P}(X=1)\,\mathbb{P}(Y=0)$$

independence

$$= p(1-p).$$

Ex. Two coins, one is fair, the other is biased $\mathbb{P}(H) = \frac{2}{3}$.

Coin 1: fair $p = \frac{1}{2}$

Coin 2: biased $p = \frac{2}{3}$

H T $X=1$

T T $Y=0$

Toss both coins twice:

$X =$ number of Heads for Coin 1

$Y =$ ——— " ——— Coin 2.

$$\mathbb{P}(X > Y) = \mathbb{P}(X=2, Y=0) + \mathbb{P}(X=2, Y=1)$$

$$+ \mathbb{P}(X=1, Y=0)$$

$$= \mathbb{P}(X=2)\,\mathbb{P}(Y=0)$$

$$+ \mathbb{P}(X=2)\,\mathbb{P}(Y=1)$$

$$+ \mathbb{P}(X=1)\,\mathbb{P}(Y=0)$$

(fair)

X	0	1	2
prob	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

(biased) Y	0	1	2
prob	$\frac{1}{9}$	$\frac{4}{9}$	$\frac{4}{9}$

$\mathbb{P}(T\ T) = \left(\frac{1}{3}\right)^2$

$\mathbb{P}(H\ H) = \left(\frac{2}{3}\right)^2$

$$\Rightarrow \mathbb{P}(X > Y) = \left(\frac{1}{4}\right)\left(\frac{1}{9}\right) + \left(\frac{1}{4}\right)\left(\frac{4}{9}\right) + \left(\frac{1}{2}\right)\left(\frac{1}{9}\right)$$

Example 17 *Consider again our example above where X is the value on the first die, and Y is the maximum of the two values. It's easy to check that X and Y are dependent. [For example, there cannot be a row or column with some entries zero and others non-zero].*

$\boxed{3}$ $\boxed{5}$ $X = 3, \ Y = 5.$

Die 1 Die 2

$X =$ number on Die 1.

$Y =$ max. of two Dice.

Claim: (X, Y) are dependent.

To show this, it is sufficient to find one pair of values (k, n) so that

$$\mathbb{P}(X = k, Y = n) \neq \mathbb{P}(X = k)\, \mathbb{P}(Y = n)$$

$k, n = 1, 2, \ldots, 6.$

e.g. $X = 6, \ Y = 6.$

Joint pdf: $\mathbb{P}(X=6, Y=6) = \mathbb{P}(\boxed{6}\ \boxed{x})$

$$= \frac{1}{6}$$

$$\mathbb{P}(X=6) = \frac{1}{6}$$

$$\mathbb{P}(Y=6) = \frac{11}{36}$$

Check: $\frac{1}{6} \stackrel{?}{=} \left(\frac{1}{6}\right)\left(\frac{11}{36}\right)$ NO!

$\Rightarrow$ X, Y are dependent.

Joint pdf: Mean of an independent product

We noted that the expected value is a *linear operator* on random variables:

$$\begin{aligned}\mathbb{E}[X+Y] &= \mathbb{E}[X]+\mathbb{E}[Y] \\ \mathbb{E}[aX+bY] &= a\,\mathbb{E}[X]+b\,\mathbb{E}[Y]\end{aligned}$$

However the expected value of a product is generally *not* the product of expected values, except in one important case: *if X, Y are independent* then

$$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$$

This can be seen as follows:

$$\begin{aligned}\mathbb{E}[XY] &= \sum_{i,j} x_i y_j\, P(X=x_i, Y=y_j) \\ &= \sum_{i,j} x_i y_j\, P(X=x_i)P(Y=y_j) \\ &= \sum_i x_i P(X=x_i) \sum_j y_j P(Y=y_j) \\ &= \mathbb{E}[X]\,\mathbb{E}[Y]\end{aligned}$$

$$\mathrm{COV}[X,Y] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] = 0$$

Similarly for any functions of independent random variables we get

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

[Note that if X and Y are independent then $\mathrm{CORR}[X,Y] = 0$. In general the converse is false, but in many cases it is 'mostly true' ...]

so independent r.v.'s are uncorrelated.