Mean of a Geometric($\frac{1}{2}$).

If X is the trial of the first head, then

$$P(X=x)=\left(\tfrac{1}{2}\right)^x \quad \text{for } x=1,2,3,\ldots$$

Let's compute $E(X)$. directly using the formula.

$$(*) \quad E(X)=\sum_{x=1}^{\infty} x\cdot\left(\tfrac{1}{2}\right)^x = \tfrac{1}{2}+2\left(\tfrac{1}{2}\right)^2+3\left(\tfrac{1}{2}\right)^3+4\left(\tfrac{1}{2}\right)^4+\cdots$$

Also,

$$\tfrac{1}{2}E(X)= \quad \left(\tfrac{1}{2}\right)^2+2\left(\tfrac{1}{2}\right)^3+3\left(\tfrac{1}{4}\right)^4+4\left(\tfrac{1}{2}\right)^5+\cdots$$

And subtracting this from (*) we get

a geometric series

$$\underbrace{E(X)-\tfrac{1}{2}E(X)}_{=\frac{1}{2}E(X)} = \tfrac{1}{2}+\left(\tfrac{1}{2}\right)^2+\left(\tfrac{1}{2}\right)^3+\left(\tfrac{1}{2}\right)^4+\cdots = \frac{\frac{1}{2}}{1-\frac{1}{2}} = 1$$

Therefore, $\tfrac{1}{2}E(X)=1 \Longrightarrow \boxed{E(X)=2.}$

The Expected value of a function of a r.v.

(★) $$E(g(X)) = \sum_x g(x)\, p_X(x)$$

when X is a discrete r.v.

The reason this is true is because: if $Y = g(X)$

$$E(g(X)) = E(Y) = \sum_y y\, p_Y(y)$$

$$= \sum_y \Big(y \sum_{\{x:\, g(x)=y\}} p_X(x) \Big)$$

$$= \sum_y \sum_{x:\, g(x)=y} y\, p_X(x)$$

$$= \sum_y \sum_{x:\, g(x)=y} g(x)\, p_X(x)$$

$$= \sum_x g(x)\, p_X(x).$$

(★) is a <u>very</u> useful formula because it allows us to compute the mean value of $g(X)$ <u>without</u> having to find the pmf of $g(X)$.

Intuitively (★) say $E(g(X))$ is the weighted average the values $g(x)$ against the prob mass that $X = x$.

Suppose X is a discrete r.v. and X has mean μ

i.e.,

$$\mu = E(X) = \sum_x x \, p_X(x) \quad \text{exists and is finite.}$$

Let's consider the function $g(X) = (X-\mu)^2$.

We define the <u>Variance of X</u> by

$$Var(X) = E\left[(X-\mu)^2\right] = \sum_x (x-\mu)^2 p_X(x).$$

<u>Remark</u>. $X-\mu$ means the deviation that X makes from its mean μ, so $(X-\mu)^2$ is measuring the <u>Squared deviation</u>, and $Var(X)$ is the mean Squared deviation or the <u>expected Squared deviation from μ</u> (Nice intuitive meaning)

The Variance formula above is also useful because from it we can immediately see that $Var(X) \geq 0$ (when it exists)

this is because $(X-\mu)^2 \geq 0$ always and since variance is just the weighted average of $(x-\mu)^2$ against $p_X(x) \geq 0$ we must have $Var(X) \geq 0$.

Example Suppose X is the discrete uniform on $\{1,2,3,4,5,6\}$:

x	1	2	3	4	5	6
$p_X(x)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

We saw last time that $E(X) = \frac{7}{2} = \mu$.

Let's compute $Var(X) = \sigma^2$

$$\sigma^2 = \left(1-\tfrac{7}{2}\right)^2 \cdot \tfrac{1}{6} + \left(2-\tfrac{7}{2}\right)^2 \cdot \tfrac{1}{6} + \left(3-\tfrac{7}{2}\right)^2 \cdot \tfrac{1}{6} + \left(4-\tfrac{7}{2}\right)^2 \cdot \tfrac{1}{6} + \left(5-\tfrac{7}{2}\right)^2 \cdot \tfrac{1}{6} + \left(6-\tfrac{7}{2}\right)^2 \cdot \tfrac{1}{6}$$

$$= \frac{25}{4}\cdot\frac{1}{6} + \frac{9}{4}\cdot\frac{1}{6} + \frac{1}{4}\cdot\frac{1}{6} + \frac{1}{4}\cdot\frac{1}{6} + \frac{9}{4}\cdot\frac{1}{6} + \frac{25}{4}\cdot\frac{1}{6}$$

$$= \frac{70}{24} = \frac{35}{12} = 2.916\overline{6}$$

The positive square root σ of the Variance σ^2 is called the standard deviation of X.

$$\sigma = \sqrt{2.916\overline{6}} \doteq 1.707825\ldots$$

It turns out that it is sometimes easier to use a different formula to actually compute Var(X).

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

or, equivalently if $\mu = E(X)$,

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

Let's see why this is true ---

$$Var(X) = E\left[(X-\mu)^2\right] = \sum_x (x-\mu)^2 p_X(x)$$

$$= \sum_x (x^2 - 2\mu x + \mu^2) p_X(x)$$

$$= \sum_x \left\{ x^2 p_X(x) - 2\mu x p_X(x) + \mu^2 p_X(x) \right\}$$

$$= \underbrace{\sum_x x^2 p_X(x)}_{E(X^2)} - 2\mu \underbrace{\sum_x x p_X(x)}_{=\mu} + \mu^2 \underbrace{\sum_x p_X(x)}_{=1}$$

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

Either formula

$$Var(X) = E((X-\mu)^2)$$

or

$$Var(X) = E(X^2) - \mu^2$$

is valid when the expected values on the right exist and are finite. The second formula is sometimes called the Moments expression since

$$E(X^k) = k^{th} \text{ moment of } X.$$

So $\mu = E(X)$ is the first moment.
$E(X^2)$ is the second moment, and so on.

Remark. Since we learned that $\mu = E(X)$ is the "center" of the distribution of X, we say $X-\mu$ is centered and $E((X-\mu)^2)$ is called the 2^{nd} central moment.

Remark. As a by product of the work on the previous page we see that for any constants

$$E(aX+b) = \sum_x (ax+b)p_X(x) = aE(X)+b.$$

In particular $E(X+b) = \mu+b$: if we add b to every possible value of X then the mean μ shifts by b.

Moreover,

$$Var(aX+b) = E\left[\left(aX+b-E(aX+b)\right)^2\right]$$

$$= E\left[\left(aX+b-\{a\mu+b\}\right)^2\right]$$

$$= E\left[\left(aX-a\mu\right)^2\right] = E\left(a^2[X-\mu]^2\right)$$

$$= a^2\, Var(X).$$

<u>Remark</u> In particular $Var(X+b) = Var(X)$, i.e. adding b to every value of X doesn't change the variance.... makes sense since we aren't changing the way the values of X are dispersed about the (new) mean.

Joint probability mass functions

In many situations we can have several r.v.s. defined on the same sample space.

For example, we can have an experiment of rolling two balanced (6-sided) dice and set

X_1 = the number on the first roll

X_2 = maximum of the two up-faces.

For each possible value of X_1 and X_2 we can ask for the probability that X_1 and X_2 take these values. I.e,

$P(X_1 = x_1, X_2 = x_2)$. The function

$$p_{X_1, X_2}(x_1, x_2) = P(X_1 = x_1, X_2 = x_2)$$

is called the joint pmf of X_1, X_2.

Here is the function in tabular form for the above r.v.s.

	$X_1 = 1$	$X_1 = 2$	$X_1 = 3$	$X_1 = 4$	$X_1 = 5$	$X_1 = 6$
$X_2 = 1$	$\frac{1}{36}$	0	0	0	0	0
$X_2 = 2$	$\frac{1}{36}$	$\frac{2}{36}$	0	0	0	0
$X_2 = 3$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{3}{36}$	0	0	0
$X_2 = 4$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{4}{36}$	0	0
$X_2 = 5$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{5}{36}$	0
$X_2 = 6$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{6}{36}$

Computing probabilities involving two discrete rvs is straight-forward once the joint pmf has been specified.

To compute

$$P[(X,Y) \in A] = \sum_{(x,y)\in A} p_{X,Y}(x,y)$$

i.e. add all the probability masses ~~for~~ at each (x,y) that belongs to the set A.

In the above example

$P[X_1 + X_2 \geq 7]$ = sum of all masses within the green outlined region

$$= \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + 0 + \frac{4}{36} + \frac{1}{36} + \frac{1}{36} + \frac{5}{36} + \frac{1}{36} + \frac{6}{36} = \frac{24}{36}$$

Also, we can compute

$$P[X_1 = 1] = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{1}{6}$$

$$P[X_1 = 2] = \frac{2}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{1}{6}$$

$$\vdots$$

$$P[X_1 = 6] = \frac{6}{36} = \frac{1}{6}$$

That is, from the joint pmf of X_1 and X_2 we can "recover" the pmf of X_1 (alone).

In this content we call it the <u>marginal pmf</u> of X_1:

$$p_{X_1}(x_1) = \sum_{x_2} p_{X_1, X_2}(x_1, x_2)$$

The marginal pmf of X_2 is

$$p_{X_2}(x_2) = \sum_{x_1} p_{X_1, X_2}(x_1, x_2)$$

and in our example it is (the row sums...)

$P[X_2 = 1] = 1/36 \quad P[X_2 = 2] = 3/36 \quad P(X_2 = 3) = 5/36$

$P(X_2 = 4) = 7/36 \quad P(X_2 = 5) = 9/36 \quad P(X_2 = 6) = 11/36.$

The concept of expected value for jointly discrete r.v. extends...

If $g = g(x,y)$ is a real-valued function then

$$E(g(X,Y)) = \sum_x \sum_y g(x,y)\, p_{X,Y}(x,y)$$

<u>Example</u> If the joint pmf of X, Y is given as follows:

	$x=-1$	$x=0$	$x=1$	$x=2$	
$y=1$	.2	.1	.1	.1	.5
$y=2$	.1	.1	.1	.2	.5
	.3	.2	.2	.3	

pmf of Y.

then pmf of X

$$E(XY) = (-1)(1)(.2) + (0)(1)(.1) + (1)(1)(.1) + (2)(1)(.1)$$
$$+ (-1)(2)(.1) + 0(2)(.1) + (1)(2)(.1) + (2)(2)(.2)$$

$$= -.2 + 0 + .1 + .2 - .2 + 0 + .2 + .8$$

$$= .9.$$

How about $E(X+Y)$?

$$E(X+Y) = (-1+1)(.2) + (0+1)(.1) + (1+1)(.1) + (2+1)(.1)$$
$$+ (-1+2)(.1) + (0+2)(.1) + (1+2)(.1) + (2+2)(.2)$$

$$= 0 + .1 + .2 + .3 + .1 + .2 + .3 + .8$$

$$= 2.$$

Notice that in this example

$$E(X) = -1(.3) + 0(.2) + 1(.2) + 2(.3) =$$
$$= -.3 + 0 + .2 + .6 = .5$$

$$E(Y) = 1(.5) + 2(.5) = 1.5$$

and

$$E(X) + E(Y) = .5 + 1.5 = 2.$$

That is,

$$E(X+Y) = E(X) + E(Y)$$

<u>This is not coincidence</u>.

Theorem (the linearity of expectation)

Suppose X_1, X_2 are discrete r.v.s and a_1 and a_2 are any constants. Then

$$E(a_1 X_1 + a_2 X_2) = a_1 E(X_1) + a_2 E(X_2).$$

Proof

$$E(a_1 X_1 + a_2 X_2) = \sum_{x_1} \sum_{x_2} (a_1 x_1 + a_2 x_2)\, p_{X_1, X_2}(x_1, x_2)$$

$$= \sum_{x_1} \sum_{x_2} a_1 x_1\, p_{X_1 X_2}(x_1, x_2) + \sum_{x_1} \sum_{x_2} a_2 x_2\, p_{X_1 X_2}(x_1, x_2)$$

$$= a_1 \sum_{x_1} x_1 \sum_{x_2} p_{X_1, X_2}(x_1, x_2) + a_2 \sum_{x_2} \sum_{x_1} x_2\, p_{X_1, X_2}(x_1, x_2)$$

$$= a_1 \sum_{x_1} x_1\, p_{X_1}(x_1) + a_2 \sum_{x_2} x_2 \sum_{x_1} p_{X_1, X_2}(x_1, x_2)$$

$$= a_1 E(X_1) + a_2 \sum_{x_2} x_2\, p_{X_2}(x_2)$$

$$= a_1 E(X_1) + a_2 E(X_2).$$

$\blacksquare$

Remark This theorem remains true even for continuous random variables.

Remark

Once we know that $E(a_1 X_1 + a_2 X_2) = a_1 E(X_1) + a_2 E(X_2)$

we know it holds for any finite number of r.v.s $X_1, X_2, \ldots, X_n$

and constants $a_1, a_2, \ldots, a_n$ as well.

To see why, I will illustrate with $n=3$:

$$E(a_1 X_1 + a_2 X_2 + a_3 X_3)$$

$$= E(\{a_1 X_1 + a_2 X_2\} + a_3 X_3)$$

$$= E(a_1 X_1 + a_2 X_2) + a_3 E(X_3)$$

using the theorem with $n=2$ since $a_1 X_1 + a_2 X_2$ is another r.v.

$$= a_1 E(X_1) + a_2 E(X_2) + a_3 E(X_3).$$

In general,

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$$

Remark this nice property of expectations allows us to Compute Expected values of r.v.s when we recognize the r.v. as a sum of "simpler" r.v.s. as in the next example...

Example We toss a coin with success probability p n times (independently). Let X be the numbers of successes tossed.

Compute $E(X)$.

On one hand, we know $X \sim \text{binomial}(n,p)$

and $p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$ for $x = 0, 1, 2, \ldots, n$

So
$$E(X) = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x}.$$

But we did this earlier and it was involved.

Here is an easy way <u>if</u> we recognize that

$$X = \sum_{i=1}^{n} X_i \quad \text{where}$$

$$X_i = \begin{cases} 1 & \text{if } i^{th} \text{ toss is a success} \\ 0 & \text{if } i^{th} \text{ toss is a failure} \end{cases}$$

(Notice that each X_i is a Bernoulli(p) r.v. whose expected value is p)

Then
$$E(X) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p = np.$$

Example (Men with hats problem - revisited).

n men wearing hats are in a room. They throw their hats into a room — the hats get mixed — each man then takes turns randomly selecting a hat. We say a "MATCH" occurs at i if the i^{th} man selects his own hat.

Let X be the r.v. the counts the total number of matches.

Again, if we let $X_i = \begin{cases} 1 & \text{if } i^{th} \text{ man selects hat } i \\ 0 & \text{if otherwise.} \end{cases}$

then $E(X_i) = \frac{1}{n}$ for all $i = 1, 2, \ldots, n$.

Furthermore, $X = \sum_{i=1}^{n} X_i$ so

$$E(X) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \frac{1}{n} = 1.$$

We expects only 1 man to select their own hat.

Another example where linearity is useful is the following:

<u>Example</u> If we deal a person a 13-card hand from a standard deck of 52 cards, what is the expected number of suits they receive?

Let $X_i = \begin{cases} 1 & \text{if suit } i \text{ is in hand} \\ 0 & \text{if otherwise.} \end{cases}$

each is Bernoulli(p) with $p = P(\text{suit } i \text{ in hand})$

Then

$$X = \text{total \# of suits in hand}$$
$$= \sum_{i=1}^{4} X_i$$

Now

$$E(X) = E\left(\sum_{i=1}^{4} X_i\right) = \sum_{i=1}^{4} E(X_i) = \sum_{i=1}^{4} p = 4p$$

But $p = P(\text{suit } i \text{ is in hand}) = 1 - P(\text{suit } i \text{ is } \underline{\text{not}} \text{ in hand})$

$$= 1 - \frac{\binom{39}{13}}{\binom{52}{13}}.$$

thus,

$$E(X) = 4\left(1 - \frac{\binom{39}{13}}{\binom{52}{13}}\right) \approx 3.9488 \text{ suits.}$$

Remark

In the last example if we had dealt the player only 5 cards

then $E(X) = 4\left(1 - \frac{\binom{39}{5}}{\binom{52}{5}}\right) \approx 3.114$ suits.

Conditional probability Mass functions

X, Y are jointly discrete having joint pmf $p_{X,Y}(x,y)$

and marginal pmfs $p_X(x)$ and $p_Y(y)$.

We define the Conditional pmf of X given $Y=y$

as

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

(here we think of y as fixed and treat this as a function of x)

and the Conditional pmf of Y given $X=x$

as

$$p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)}$$

(similarly, we think of x as being fixed and treat this as a function of y)

The motivation behind these formulas is very straight-forward as it follows directly from Conditional probabilities taught earlier:

Suppose we have two r.v.s X, Y and the events

$$\{X = x\} \quad \text{and} \quad \{Y = y\}$$

Then (assuming $P(\{Y=y\}) = P(Y=y) > 0$)

$$\underbrace{P(\{X=x\} \mid \{Y=y\})}_{p_{X|Y}(x|y)} = \frac{P(\{X=x\} \cap \{Y=y\})}{P(\{Y=y\})} = \frac{P(X=x, Y=y)}{P(Y=y)}$$

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}.$$

and a similar derivation explains the definition of the other conditional pmf.

<u>Example</u> Let X_1 = result of 1^{st} die, X_2 = result of 2^{nd} die.

defined
$$D = |X_1 - X_2|, \quad W = \max\{X_1, X_2\}.$$

Check that the joint pmf of D, W is

$w \backslash d$	$D=0$	$D=1$	$D=2$	$D=3$	$D=4$	$D=5$	row sums
$W=1$	$\frac{1}{36}$	0	0	0	0	0	$1/36$
$W=2$	$\frac{1}{36}$	$\frac{2}{36}$	0	0	0	0	$3/36$
$W=3$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	0	0	0	$5/36$
$W=4$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	0	0	$7/36$
$W=5$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	0	$9/36$
$W=6$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$11/36$
column sums →	$6/36$	$10/36$	$8/36$	$6/36$	$4/36$	$2/36$	

- Find the Conditional pmf of D given $W=4$:

(in tabular form) we have

d	0	1	2	3
$p_{D\|W}(d\|4)$	$\frac{1}{7}$	$\frac{2}{7}$	$\frac{2}{7}$	$\frac{2}{7}$

So, for example,

$$P(D \leq 1 \,|\, W=4) = p_{D|W}(0|4) + p_{D|W}(1|4)$$

$$= \frac{1}{7} + \frac{2}{7} = \frac{3}{7}.$$

- Find the conditional pmf of W given $D=1$:

w	2	3	4	5	6
$p_{W\|D}(w\|1)$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

Notice this conditional pmf is that of a Discrete Uniform distribution on the points 2, 3, 4, 5, 6

$\frac{1}{5}$ prob. each.