Expected Value

The expected value of a random variable indicates its weighted average.

Ex. How many heads would you expect if you flipped a coin twice?

X = number of heads = {0,1,2}
p(0)=1/4, p(1)=1/2, p(2)=1/4
Weighted average = 0*1/4 + 1*1/2 + 2*1/4 = 1

Draw PDF

Definition: Let X be a random variable assuming the values x_1, x_2, x_3, ... with corresponding probabilities $p(x_1)$, $p(x_2)$, $p(x_3)$,..... The mean or expected value of X is defined by E(X) = sum x_k $p(x_k)$.

Interpretations:
(i) The expected value measures the center of the probability distribution - center of mass.
(ii) Long term frequency (law of large numbers… we'll get to this soon)

Expectations can be used to describe the potential gains and losses from games.

Ex. Roll a die. If the side that comes up is odd, you win the $ equivalent of that side. If it is even, you lose $4.

Let X = your earnings

$X=1$	$P(X=1) = P(\{1\}) = 1/6$
$X=3$	$P(X=1) = P(\{3\}) = 1/6$
$X=5$	$P(X=1) = P(\{5\}) = 1/6$
$X=-4$	$P(X=1) = P(\{2,4,6\}) = 3/6$

$E(X) = 1*1/6 + 3*1/6 + 5*1/6 + (-4)*1/2 = 1/6 + 3/6 + 5/6 – 2 = -1/2$

Ex. Lottery – You pick 3 different numbers between 1 and 12. If you pick all the numbers correctly you win $100. What are your expected earnings if it costs $1 to play?

Let X = your earnings

$X = 100-1 = 99$

$X = -1$

$P(X=99) = 1/(12\ 3) = 1/220$

$P(X=-1) = 1-1/220 = 219/220$

$E(X) = 100*1/220 + (-1)*219/220 = -119/220 = -0.54$

Expectation of a function of a random variable

Let X be a random variable assuming the values x_1, x_2, x_3, ... with corresponding probabilities $p(x_1)$, $p(x_2)$, $p(x_3)$,..... For any function g, the mean or expected value of g(X) is defined by $E(g(X)) = \text{sum } g(x_k)\, p(x_k)$.

Ex. Roll a fair die. Let X = number of dots on the side that comes up.

Calculate $E(X^2)$.

$E(X^2)$ = sum_{i=1}^{6} i^2 p(i) = 1^2 p(1) + 2^2 p(2) + 3^2 p(3) + 4^2 p(4) + 5^2 p(5) + 6^2 p(6) = 1/6*(1+4+9+16+25+36) = 91/6

E(X) is the expected value or 1st moment of X.
$E(X^n)$ is called the nth moment of X.

Calculate E(sqrt(X)) = sum_{i=1}^{6} sqrt(i) p(i)
Calculate $E(e^X)$ = sum_{i=1}^{6} e^i p(i)
(Do at home)

Ex. An indicator variable for the event *A* is defined as the random variable that takes on the value 1 when event A happens and 0 otherwise.

I_A = 1 if A occurs
0 if A^C occurs

$P(I_A=1) = P(A)$ and $P(I_A=0) = P(A^C)$
The expectation of this indicator (noted I_A) is $E(I_A)=1*P(A) + 0*P(A^C) = P(A)$.

One-to-one correspondence between expectations and probabilities.

If a and b are constants, then E(aX+b) = aE(X) + b
Proof: E(aX+b) = sum [(ax_k+b) $p(x_k)$] = a sum{$x_k p(x_k)$} + b sum{$p(x_k)$} = aE(X) + b

Variance

We often seek to summarize the essential properties of a random variable in as simple terms as possible.

The mean is one such property.

Let $X = 0$ with probability 1

Let $Y =$ -2 with prob. 1/3
-1 with prob. 1/6
1 with prob. 1/6
2 with prob. 1/3

Both X and Y have the same expected value, but are quite different in other respects. One such respect is in their spread. We would like a measure of spread.

Definition: If X is a random variable with mean E(X), then the variance of X, denoted by Var(X), is defined by $Var(X) = E((X-E(X))^2)$.

A small variance indicates a small spread.

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

$$
\begin{aligned}
Var(X) &= E((X-E(X))^2) \\
&= \text{sum } (x- E(X))^2\, p(x) \\
&= \text{sum } (x^2-2x\, E(X)+ E(X)^2)\, p(x) \\
&= \text{sum } x^2\, p(x) -2\, E(X) \text{ sum } xp(x) + E(X)^2 \text{ sum } p(x) \\
&= E(X^2) -2\, E(X)^2 + E(X)^2 = E(X^2) - E(X)^2
\end{aligned}
$$

Ex. Roll a fair die. Let X = number of dots on the side that comes up.

$Var(X) = E(X^2) - (E(X))^2$
$E(X^2) = 91/6$
$E(X) = 1/6(1+2+3+4+5+6) = 21/6 = 7/2$
$Var(X) = 91/6 – (7/2)^2 = 91/6 – 49/4 = (182-147)/12 = 35/12$

If a and b are constants then $Var(aX+b) = a^2Var(X)$

$E(aX+b) = a E(X) + b$
$Var(aX+b) = E[(aX+b -(a E(X)+b))^2]= E(a^2(X- E(X))^2) = a^2E((X- E(X))^2)= a^2Var(X)$

The square root of Var(X) is called the standard deviation of X.
SD(X) = sqrt(Var(X)): measures scale of X.

Means, modes, and medians

Best estimate under squared loss: mean

i.e., the number m that minimizes E[(X-m)^2] is m=E(X). Proof: expand and differentiate with respect to m.

Best estimate under absolute loss: median.
i.e., m=median minimizes E[|X-m|]. Proof in book. Note that median is nonunique in general.

Best estimate under 1-1(X=x) loss: mode. Ie, choosing mode maximizes probability of being exactly right. Proof easy for discrete r.v.'s; a limiting argument is required for continuous r.v.'s, since P(X=x)=0 for any x.

Moment Generating Functions

The moment generating function of the random variable X, denoted $M_X(t)$, is defined for all real values of t by,

$$M_X(t) = E(e^{tX}) = \begin{cases} \sum_x e^{tx} p(x) & \text{if X is discrete with pmf p(x)} \\ \int_{-\infty}^{\infty} e^{tx} f(x)dx & \text{if X is continuous with pdf f(x)} \end{cases}$$

The reason $M_X(t)$ is called a moment generating function is because all the moments of X can be obtained by successively differentiating $M_X(t)$ and evaluating the result at t=0.

First Moment:

$$\frac{d}{dt}M_X(t) = \frac{d}{dt}E(e^{tX}) = E(\frac{d}{dt}e^{tX}) = E(Xe^{tX})$$

$$M'_X(0) = E(X)$$

(For any of the distributions we will use we can move the derivative inside the expectation).

Second moment:

$$M''_X(t) = \frac{d}{dt}M'(t) = \frac{d}{dt}E(Xe^{tX}) = E(\frac{d}{dt}(Xe^{tX})) = E(X^2e^{tX})$$

$$M''_X(0) = E(X^2)$$

kth moment:

$$M^k{}_X(t) = E(X^k e^{tX})$$

$$M^k{}_X(0) = E(X^k)$$

Ex. Binomial random variable with parameters n and p.

Calculate $M_X(t)$:

$$M_X(t) = E(e^{tX}) = \sum_{k=0}^{n} e^{tk}\binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{n}\binom{n}{k}\left(pe^t\right)^k (1-p)^{n-k} = \left(pe^t + 1 - p\right)^n$$

$$M_X'(t) = n\left(pe^t + 1 - p\right)^{n-1} pe^t$$

$$M_X''(t) = n(n-1)\left(pe^t + 1 - p\right)^{n-2} (pe^t)^2 + n\left(pe^t + 1 - p\right)^{n-1} pe^t$$

$$E(X) = M_X'(0) = n\left(pe^0 + 1 - p\right)^{n-1} pe^0 = np$$

$$E(X^2) = M_X''(t) = n(n-1)\left(pe^0 + 1 - p\right)^{n-2} (pe^0)^2 + n\left(pe^0 + 1 - p\right)^{n-1} pe^0$$

$$= n(n-1)p^2 + np$$

$$Var(X) = E(X^2) - E(X)^2 = n(n-1)p^2 + np - (np)^2 = np(1-p)$$

Later we'll see an even easier way to calculate these moments, by using the fact that a binomial X is the sum of N i.i.d. simpler (Bernoulli) r.v.'s.

<u>Fact:</u> Suppose that for two random variables X and Y, moment generating functions exist and are given by $M_X(t)$ and $M_Y(t)$, respectively. If $M_X(t) = M_Y(t)$ for all values of t, then X and Y have the same probability distribution.

If the moment generating function of X exists and is finite in some region about t=0, then the distribution is uniquely determined.

The proof (which we'll skip) is based on ideas from Fourier analysis. A nice hand-wavey argument for why the theorem "should" be true is as follows (read along with, e.g., HMC pp. 60-61). Let's say someone hands you

$$M(s) = ae^s + be^{-3s} + ce^{6s},$$

say, for positive constants a, b, c. Now, we know, by the definition of mgf, that

$$M(s) = \int e^{su} p(u) du.$$

How many distributions $p(u)$ can you think of that satisfy

$$\int e^{su} p(u) du = ae^s + be^{-3s} + ce^{6s},$$

for all s? It's hard to think of any others besides the one that assigns mass a to the point $u = 1$, b at $u = -3$, and c on $u = 6$.

Note that not every distribution has an mgf, unfortunately: for many distributions (e.g., those with "heavy tails") the relevant integrals will be infinite. (This inconvenient fact inspired people to try complex exponentials — Fourier transforms — which turn out to work for all distributions[9]. But the humble mgf will suit our needs adequately here.)

[9] See a more advanced book on probability theory, e.g., Breiman '68, for more information.

Properties of Expectation

Proposition:

If X and Y have a joint probability mass function $p_{XY}(x,y)$, then

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

If X and Y have a joint probability density function $f_{XY}(x,y)$, then

$$E(g(X,Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{XY}(x,y)$$

It is important to note that if the function g(x,y) is only dependent on either x or y the formula above reverts to the 1-dimensional case.

Ex. Suppose X and Y have a joint pdf $f_{XY}(x,y)$. Calculate E(X).

$$E(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{XY}(x,y)dydx = \int_{-\infty}^{\infty} x\left(\int_{-\infty}^{\infty} f_{XY}(x,y)dy\right)dx = \int_{-\infty}^{\infty} x f_X(x)dx$$

Ex. An accident occurs at a point X that is uniformly distributed on a road of length L. At the time of the accident an ambulance is at location Y that is also uniformly distributed on the road. Assuming that X and Y are independent, find the expected distance between the ambulance and the point of the accident.

Compute E(|X-Y|).

Both X and Y are uniform on the interval (0,L).

The joint pdf is $f_{XY}(x,y) = \frac{1}{L^2}$, $0<x<L$, $0<y<L$.

$$E(|X-Y|) = \int_0^L\int_0^L |x-y| \frac{1}{L^2} dydx = \frac{1}{L^2}\int_0^L\int_0^L |x-y|\, dydx$$

$$\int_0^L |x-y|\, dy = \int_0^x (x-y)dy + \int_x^L (y-x)dy = \left[xy - \frac{y^2}{2}\right]_o^x + \left[\frac{y^2}{2} - yx\right]_x^L$$

$$= (x^2 - \frac{x^2}{2}) + \frac{L^2}{2} - xL - (\frac{x^2}{2} - x^2) = \frac{L^2}{2} + x^2 - xL$$

$$E(|X-Y|) = \frac{1}{L^2}\int_0^L \left(\frac{L^2}{2} + x^2 - xL\right)dx = \frac{1}{L^2}\left[\frac{xL^2}{2} + \frac{x^3}{3} - \frac{x^2}{2}L\right]_0^L = \frac{1}{L^2}\left(\frac{L^3}{2} + \frac{L^3}{3} - \frac{L^3}{2}\right) = \frac{L}{3}$$

Expectation of sums of random variables

Ex. Let X and Y be continuous random variables with joint pdf $f_{XY}(x,y)$. Assume that E(X) and E(Y) are finite. Calculate E(X+Y).

$$E(X+Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)f_{XY}(x,y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xf_{XY}(x,y)dxdy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}yf_{XY}(x,y)dxdy$$

$$= \int_{-\infty}^{\infty}xf_X(x)dx + \int_{-\infty}^{\infty}yf_Y(y)dy = E(X)+E(Y)$$

Same result holds in discrete case.

Proposition: In general if $E(X_i)$ are finite for all i=1,….n, then
$E(X_1+\ldots+X_n) = E(X_1)+\ldots+E(X_n)$.

Proof: Use the example above and prove by induction.

Let X_1, ….. X_n be independent and identically distributed random variables having distribution function F_X and expected value μ. Such a sequence of random variables is said to constitute a sample from the distribution F_X. The quantity $\overline{X}$, defined by
$\overline{X} = \sum_{i=1}^{n}\frac{X_i}{n}$ is called the sample mean.

Calculate $E(\overline{X})$.

We know that $E(X_i) = \mu$.

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

When the mean of a distribution is unknown, the sample mean is often used in statistics to estimate it. (Unbiased estimate)

Ex. Let X be a binomial random variable with parameters n and p. X represents the number of successes in n trials. We can write X as follows:

$$X = X_1 + X_2 + \ldots + X_n$$

where

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

The X_i's are Bernoulli random variables with parameter p.

$$E(X_i) = 1 * p + 0 * (1 - p) = p$$
$$E(X) = E(X_1) + E(X_2) + \ldots + E(X_n) = np$$

Ex. A group of N people throw their hats into the center of a room. The hats are mixed, and each person randomly selects one. Find the expected number of people that select their own hat.

Let X = the number of people who select their own hat.

Number the people from 1 to N. Let

$$X_i = \begin{cases} 1 & \text{if the person i choses his own hat} \\ 0 & \text{otherwise} \end{cases}$$

then $X = X_1 + X_2 + \ldots + X_N$

Each person is equally likely to select any of the N hats, so $P(X_i = 1) = \frac{1}{N}$.

$$E(X_i) = 1\frac{1}{N} + 0(1 - \frac{1}{N}) = \frac{1}{N}.$$

Hence, $E(X) = E(X_1) + E(X_2) + \ldots + E(X_N) = N\frac{1}{N} = 1$

Ex. Twenty people, consisting of 10 married couples, are to be seated at five different tables, with four people at each table. If the seating is done at random, what is the expected number of married couples that are seated at the same table?

Let X = the number of married couples at the same table.

Number then couples from 1 to 10 and let,

$$X_i = \begin{cases} 1 & \text{if couple i is seated at the same table.} \\ 0 & \text{otherwise} \end{cases}$$

Then $X = X_1 + X_2 + \ldots + X_{10}$

To calculate E(X) we need to know $E(X_i)$.

Consider the table where husband *i* is sitting. There is room for three other people at his table. There are a total of 19 possible people which could be seated at his table.

$$P(X_i = 1) = \frac{\binom{1}{1}\binom{18}{2}}{\binom{19}{3}} = \frac{3}{19}.$$

$$E(X_i) = 1\frac{3}{19} + 0\frac{16}{19} = \frac{3}{19}.$$

Hence, $E(X) = E(X_1) + E(X_2) + \ldots + E(X_n) = 10\frac{3}{19} = \frac{30}{19}$

Proposition: If X and Y are independent, then for any functions h and g,

$$E(g(X)h(Y)) = E(g(X))E(h(Y)).$$

Proof:

$$E(g(X)h(Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_{XY}(x,y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)dxdy$$

$$= \int_{-\infty}^{\infty} g(x)f_X(x)dx \int_{-\infty}^{\infty} h(y)f_Y(y)dy = E(g(X))E(h(Y))$$

In fact, this is an equivalent way to characterize independence: if $E(g(X)h(Y)) = E(g(X))E(h(Y))$ for any functions g(X) and h(Y) (but not any function f(X,Y)), then X and Y are independent. To see this, just use indicator functions.

Fact: The moment generating function of the sum of independent random variables equals the product of the individual moment generating functions.

Proof: $M_{X+Y}(t) = E(e^{t(X+Y)}) = E(e^{tX}e^{tY}) = E(e^{tX})E(e^{tY}) = M_X(t)M_Y(t)$

Covariance and correlation

Previously, we have discussed the absence or presence of a relationship between two random variables, i.e. independence or dependence. But if there is in fact a relationship, the relationship may be either weak or strong.

Ex. (a) Let X = weight of a sample of water
Y = volume of the same sample of water

There is an extremely strong relationship between X and Y.

(b) Let X = a persons weight
Y = same persons weight

There is a relationship between X and Y, but not as strong as in (a).

We would like a measure that can quantify this difference in the strength of a relationship between two random variables.

Definition: The covariance between X and Y, denoted by Cov(X,Y), is defined by

$$Cov(X,Y) = E[(X - E(X))(Y - E(Y))].$$

Similarly as with the variance, we can rewrite this equation,

$$Cov(X,Y) = E[(X - E(X))(Y - E(Y))] = E[(XY - E(X)Y - XE(Y) + E(X)E(Y))]$$
$$= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) = E(XY) - E(X)E(Y)$$

Note that if X and Y are independent,
$Cov(X,Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0$.

The converse is however NOT true.

Counter-Example: Define X and Y so that,

P(X=0) = P(X=1) = P(X=-1) = 1/3

and

$$Y = \begin{cases} 0 & \text{if } X \neq 0 \\ 1 & \text{if } X = 0 \end{cases}$$

X and Y are clearly dependent.

XY=0 so we have that E(XY)=E(X)=0, so $Cov(X,Y) = E(XY) - E(X)E(Y) = 0$.

Proposition:

(i) $Cov(X,Y) = Cov(Y,X)$
(ii) $Cov(X,X) = Var(X)$
(iii) $Cov(aX,Y) = aCov(X,Y)$
(iv) $Cov(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_i) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j)$

Proof: (i) – (iii) Verify yourselves.

(iv). Let $\mu_i = E(X_i)$ and $\eta_j = E(Y_j)$

Then $E(\sum_{i=1}^{m} X_i) = \sum_{i=1}^{m} \mu_i$ and $E(\sum_{i=1}^{m} Y_i) = \sum_{i=1}^{m} \eta_i$

$$Cov(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_i) = E\left[(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i)(\sum_{j=1}^{m} Y_i - \sum_{i=1}^{n} \eta_i)\right] = E\left[\sum_{i=1}^{n} (X_i - \mu_i)\sum_{j=1}^{m} (Y_i - \eta_i)\right]$$
$$= E\left[\sum_{i=1}^{n}\sum_{j=1}^{m} (X_i - \mu_i)(Y_i - \eta_i)\right] = \sum_{i=1}^{n}\sum_{j=1}^{m} E((X_i - \mu_i)(Y_i - \eta_i)) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j)$$

Proposition: $Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i<j}\sum Cov(X_i, X_j)$. In particular, V(X+Y)=V(X)+V(Y)+2C(X,Y).

Proof:

$$Var(\sum_{i=1}^{n} X_i) = Cov(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i) = \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, X_j) = \sum_{i=1}^{n} Var(X_i) + \sum_{i\neq j}\sum Cov(X_i, X_j)$$
$$= \sum_{i=1}^{n} Var(X_i) + 2\sum_{i<j}\sum Cov(X_i, X_j)$$

If X_1, …… X_n are pairwise independent for i≠j, then $Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$.

Ex. Let X_1, ….. X_n be independent and identically distributed random variables having expected value μ and variance σ^2. Let $\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$ be the sample mean. The random variable $S^2 = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n-1}$ is called the sample variance.

Calculate (a) Var($\overline{X}$) and (b) E(S^2).

(a) We know that $Var(X_i) = \sigma^2$.

$$Var(\overline{X}) = Var(\sum_{i=1}^{n} \frac{X_i}{n}) = \left(\frac{1}{n}\right)^2 Var(\sum_{i=1}^{n} X_i) = \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n} Var(X_i) = \frac{\sigma^2}{n}.$$

(b) Rewrite the sum portion of the sample variance:

$$\sum_{i=1}^{n} (X_i - \overline{X})^2 = \sum_{i=1}^{n} (X_i - \mu + \mu - \overline{X})^2 = \sum_{i=1}^{n} ((X_i - \mu) - (\overline{X} - \mu))^2$$

$$= \sum_{i=1}^{n} (X_i - \mu)^2 + (\overline{X} - \mu)^2 - 2(X_i - \mu)(\overline{X} - \mu) = \sum_{i=1}^{n} (X_i - \mu)^2 + \sum_{i=1}^{n} (\overline{X} - \mu)^2 - \sum_{i=1}^{n} 2(X_i - \mu)(\overline{X} - \mu)$$

$$= \sum_{i=1}^{n} (X_i - \mu)^2 + n(\overline{X} - \mu)^2 - 2(\overline{X} - \mu)\sum_{i=1}^{n} (X_i - \mu) = \sum_{i=1}^{n} (X_i - \mu)^2 + n(\overline{X} - \mu)^2 - 2(\overline{X} - \mu)n(\overline{X} - \mu)$$

$$= \sum_{i=1}^{n} (X_i - \mu)^2 - n(\overline{X} - \mu)^2$$

$$E(S^2) = \frac{1}{n-1}\left[\sum_{i=1}^{n} E[(X_i - \mu)^2] - nE((\overline{X} - \mu)^2)\right] = \frac{1}{n-1}\left[n\sigma^2 - nVar(\overline{X})\right] = \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2$$

(The sample variance is an unbiased estimate of the variance)

Ex. A group of N people throw their hats into the center of a room. The hats are mixed, and each person randomly selects one.

Let X = the number of people who select their own hat.

Number the people from 1 to N. Let

$$X_i = \begin{cases} 1 & \text{if the person i choses his own hat} \\ 0 & \text{otherwise} \end{cases}$$

then $X = X_1 + X_2 + \ldots + X_n$

We showed last time that E(X)=1.

Calculate Var(X).

$$Var(X) = Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i<j}\sum Cov(X_i, X_j)$$

Recall that since each person is equally likely to select any of the N hats $P(X_i = 1) = \frac{1}{N}$.

Hence,

$$E(X_i) = 1\frac{1}{N} + 0(1 - \frac{1}{N}) = \frac{1}{N}$$

and

$$E(X_i^2) = 1^2\frac{1}{N} + 0(1 - \frac{1}{N}) = \frac{1}{N} \ .$$

$$Var(X_i) = E(X_i^2) - (E(X_i))^2 = \frac{1}{N} - \left(\frac{1}{N}\right)^2 = \frac{N-1}{N^2}.$$

$$Cov(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j)$$

$$X_i X_j = \begin{cases} 1 & \text{if both persons i and j choses their own hat} \\ 0 & \text{otherwise} \end{cases}$$

$$P(X_i = 1, X_j = 1) = P(X_i = 1 \mid X_j = 1)P(X_j = 1) = \frac{1}{N}\frac{1}{N-1}$$

$$E(X_i X_j) = 1\frac{1}{N}\frac{1}{N-1} + 0(1 - \frac{1}{N}\frac{1}{N-1}) = \frac{1}{N}\frac{1}{N-1}$$

$$Cov(X_i, X_j) = \frac{1}{N}\frac{1}{N-1} - \left(\frac{1}{N}\right)^2 = \frac{1}{N^2(N-1)}$$

Hence, $Var(X) = \frac{N-1}{N} + 2\binom{N}{2}\frac{1}{N^2(N-1)} = \frac{N-1}{N} + \frac{1}{N} = 1$

Definition: The correlation between X and Y, denoted by $\rho(X,Y)$, is defined, as long as Var(X) and Var(Y) are positive, by

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}.$$

It can be shown that $-1 \le \rho(X,Y) \le 1$, with equality only if Y=aX+b (assuming E(X^2) and E(Y^2) are both finite). This is called the “Cauchy-Schwarz” inequality.

Proof: It suffices to prove (E(XY))^2<=E(X^2)E(Y^2). The basic idea is to look at the expectations E[(aX+bY)^2] and E[(aX-bY)^2]. We use the usual rules for adding and subtracting variance:

0<= E[(aX+bY)^2] = a^2E(X^2)+b^2E(Y^2)+2abE(XY)
0<= E[(aX-bY)^2] = a^2E(X^2)+b^2E(Y^2)-2abE(XY)

Now let a^2=E(Y^2) and b^2=E(X^2). Then the above two inequalities read

0<=2a^2b^2+2abE(XY)
0<=2a^2b^2-2abE(XY);

dividing by 2ab gives

E(XY)>=-sqrt[E(X^2) E(Y^2)]
E(XY)<=sqrt[E(X^2) E(Y^2)],

and this is equivalent to the inequality $-1 \le \rho(X,Y) \le 1$. For equality to hold, either E[(aX+bY)^2]=0 or E[(aX-bY)^2]=0, i.e., X and Y are linearly related with a negative or positive slope, respectively.

The correlation coefficient is therefore a measure of the degree of linearity between X and Y. If $\rho(X,Y)=0$ then this indicates no linearity, and X and Y are said to be uncorrelated.

Conditional Expectation

Recall that if X and Y are discrete random variables, the conditional mass function of X, given Y=y, is defined for all y such that P(Y=y)>0, by

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

Definition: If X and Y are discrete random variables, the conditional expectation of X, given Y=y, is defined for all y such that P(Y=y)>0, by

$$E(X \mid Y = y) = \sum_x xP(X = x \mid Y = y) = \sum_x xp_{X|Y}(x \mid y).$$

Similarly, if X and Y are continuous random variables, the conditional pdf of X given Y=y, is defined for all y such that $f_Y(y) > 0$, by

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x, y)}{f_Y(y)}.$$

Definition: If X and Y are continuous random variables, the conditional expectation of X, given Y=y, is defined for all y such that $f_Y(y) > 0$, by

$$E(X \mid Y = y) = \int_{-\infty}^{\infty} xP(X = x \mid Y = y) = \int_{-\infty}^{\infty} xf_{X|Y}(x \mid y)dx.$$

Conditional expectations are themselves random variables. The conditional expectation of X given Y=y, is just the expected value on a reduced sample space consisting only of outcomes where Y=y.

E(X|Y=y) is a function of y.

It is important to note that conditional expectations satisfy all the properties of regular expectations:

1. $E[g(X) \mid Y = y] = \sum_x g(x) p_{X|Y}(x \mid y)$ if X and Y discrete.

2. $E[g(X) \mid Y = y] = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x \mid y)dx$ if X and Y continuous.

3. $E\left[\sum_{i=1}^{n} X_i \mid Y = y\right] = \sum_{i=1}^{n} E\left[X_i \mid Y = y\right]$

<u>Proposition:</u> $E(X) = E(E(X \mid Y))$

If Y is discrete $E(X) = E(E(X \mid Y)) = \sum_{y} E(X \mid Y) p_Y(y)$

If Y is continuous $E(X) = E(E(X \mid Y)) = \int E(X \mid Y) f_Y(y)$

<u>Proof:</u> (Discrete case)

$$E(E(X \mid Y)) = \sum_{y} E(X \mid Y) p_Y(y) = \sum_{y}\left[\sum_{x} x p_{X|Y}(x \mid y)\right] p_Y(y)$$

$$= \sum_{y}\sum_{x} x \frac{p_{XY}(x,y)}{p_Y(y)} p_Y(y) = \sum_{x} x \sum_{y} p_{XY}(x,y) = \sum_{x} x p_X(x) = E(X)$$