Today: expectation & variance

Defn. Let X be a r.v. The expectation (aka expected value) of X is

$$E[X] := \sum_x x P(X=x) = \sum_x x p_X(x)$$

whenever this sum is well-defined, i.e. when at most one of

$$\sum_{x>0} x p_X(x) \quad \text{and} \quad \sum_{x<0} x p_X(x)$$

is infinite.

i.e. if the sum is absolutely convergent or if it is unambiguously $+\infty$ or $-\infty$.

The expectation of a r.v. is the typical average value of the r.v. if we perform the experiment many times independently. (law of large numbers - stay tuned)

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

$$E[X] = \sum_{x=0}^{\infty} x \cdot \frac{1}{x!} \lambda^x e^{-\lambda}.$$

$$= e^{-\lambda} \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = e^{-\lambda} \lambda \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = \lambda.$$

Example. (St. Petersburg Paradox).

Consider r.v. X s.t.

$$p_X(2^k) = P(X=2^k) = 2^{-k} \quad \text{for } k = 1, 2, 3, \ldots$$

Then

$$E[X] = \sum_{k=1}^{\infty} 2^k P(X=2^k) = \sum_{k=1}^{\infty} 2^k \cdot 2^{-k}$$

$$= \sum_{k=1}^{\infty} 1 = +\infty.$$

Prop. Suppose that X is a r.v. and $g: \mathbb{R} \to \mathbb{R}$. Then

$$E[g(X)] = \sum_x g(x) P(X=x),$$

(whenever the last sum is well-defined).

Proof. write

$$E[g(X)] = \sum_y y P(g(X)=y)$$

$$= \sum_y \sum_{x: g(x)=y} y P[X=x].$$

$$= \sum_y \sum_{x: g(x)=y} g(x) P(X=x)$$

$$= \sum_x g(x) P(X=x) \quad \square.$$

Corollary. if $c \in \mathbb{R}$ is deterministic, then $E[cX] = cE[X]$.

Theorem. (Linearity of expectation). Suppose that X and Y are r.v.s on the same probability space & $E[X]$ and $E[Y]$ are not both infinite w/ opposite signs, then

$$E[X+Y] = E[X] + E[Y].$$

Note: there is **no condition** on the independence or lack thereof of X and Y.

Proof. Let's first assume that $E[X]$ & $E[Y]$ are absolutely convergent

$$E[Y] = \sum_j j P(Y=j)$$

$$= \sum_j j \sum_i P(Y=j, X=i)$$

$$= \sum_i \sum_j j P(Y=j, X=i).$$

so $E[X] + E[Y]$

$$= \sum_i \left(i P(X=i) + \sum_j j P(Y=j, X=i) \right)$$

$$= \sum_i \left(i \sum_j P(X=i, Y=j) + \sum_j j P(Y=j, X=i) \right)$$

$$= \sum_{i,j} (i+j) P(X=i, Y=j).$$

$$= \sum_z \sum_j z \, P(X=z-j, Y=j)$$

"$z = i+j$"

$$= \sum_z \sum_j z \, P(X+Y=z, Y=j)$$

$$= \sum_z z P(X+Y=z)$$

$$= E[X+Y].$$

For now I will skip the proof in the case when one of $E[X]$ and $E[Y]$ is infinite.

<u>Example</u>. tennis tournament.

Question: we have a single-elimination tournament with 2^n players. What is the probability that two given players play each other, assuming that the players are seeded randomly?

Given two players i, j,
let E_{ij} be the event that i & j play each other, and
let $X_{ij} = 1\!\!1_{E_{ij}} = \begin{cases} 1 & \text{if } i \text{ plays } j \\ 0 & \text{o/w.} \end{cases}$

Also, let X be the total number of games played.

Note: $\sum_{\{i,j\} \in \text{pairs}} X_{ij} = X.$

(since each pair can play at most once)

On the other hand, $P(X = 2^n - 1) = 1$

since in a single-elimination tournament there are $2^n - 1$ games (so that all but 1 player is eliminated).

So:
$$\sum_{\{i,j\} \in \text{pairs}} E[X_{ij}] = E[X] = 2^n - 1.$$
$$\|$$
$$\sum_{\{i,j\} \in \text{pairs}} P(E_{ij}).$$

By symmetry, $P(E_{ij})$ does not depend on i and j.

so $\binom{2^n}{2} P(E_{ij}) = 2^n - 1.$

so $\frac{2^n(2^n - 1)}{2} P(E_{ij}) = 2^n - 1.$

so $P(E_{ij}) = \frac{1}{2^{n-1}}.$

Note: this does not use anything about the structure of the tournament.

$$\binom{m}{2} = \frac{m!}{(m-2)!\,2!} = \frac{m(m-1)}{2}.$$

Theorem. If X & Y are independent, then $E[XY] = E[X]E[Y]$ assuming that $E[X]$ & $E[Y]$ are finite.

Proof.

$$E[XY] = \sum_{\ell} \ell P(XY=\ell)$$

$$= \sum_{\ell} \sum_{j,k:\, jk=\ell} \ell P(X=j \;\&\; Y=k)$$

here we used independence!

$$= \sum_{\ell} \sum_{j,k:\, jk=\ell} jk P(X=j)P(Y=k)$$

$$= \sum_{j,k} jk P(X=j)P(Y=k)$$

$$= \left(\sum_j j P(X=j)\right) \cdot \left(\sum_k k P(Y=k)\right)$$

$$= E[X]E[Y]. \qquad \square$$

Definition. X and Y are called uncorrelated if
$$E[XY] = E[X]E[Y].$$

So we just proved that independent $\Rightarrow$ uncorrelated.

However, there are examples of r.v.s that are uncorrelated but not independent.

Example. Let
$$p_X(x) = \begin{cases} 1/4 & x \in \{-2,-1,1,2\} \\ 0 & \text{o/w}. \end{cases}$$

Let $Z \sim$ Rademacher indep. of X, and let $Y = ZX$.

Then $E[XY] = E[Z \cdot X^2]$
$$= E[Z]E[X^2] = 0$$

but $P(X=1 \text{ and } Y=2) = 0$ since $|Y| = |X|$ always – so X & Y not indep.

Defn. Suppose that $\mu = E[X]$ is finite. Then we define the variance of X to be

$$\mathrm{Var}(X) = E[(X-\mu)^2]$$
$$= E[X^2 - 2\mu X + \mu^2]$$
$$= E[X^2] - 2\mu E[X] + \mu^2$$
$$= E[X^2] - (E[X])^2.$$

Example. If $X \sim \mathrm{Ber}(p)$, then

$E[X] = p$, $E[X^2] = p$, $\mathrm{Var}(X) = p - p^2$
$= p(1-p)$.

the variance is a measure of the spread of a random variable.

the standard deviation of X is $\sqrt{\mathrm{Var}(X)}$.

↳ std. dev. has the same units as X.

given two r.v.s X and Y on the same probability space, we define the covariance

$$\mathrm{Cov}(X,Y) = E[(X-EX)(Y-EY)]$$
$$= E[XY] - E[X]E[Y].$$

and the correlation

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

note: X and Y are uncorrelated
$\iff \mathrm{corr}(X,Y) = 0$
$\iff \mathrm{Cov}(X,Y) = 0$.

Prop: if a and b are deterministic, then

(1) $\text{Var}(aX+b) = a^2\,\text{Var}(X)$.

and

(2) $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\,\text{Cov}(X,Y)$.

Proof. (1).

$$\text{Var}(aX+b) = E[(aX+b - E[aX+b])^2]$$
$$= E[(aX - aE[X])^2]$$
$$= a^2 E[(X - E[X])^2]$$
$$= a^2\,\text{Var}(X).$$

(2). $\text{Var}(X+Y) = E[(X+Y - E[X+Y])^2]$

$$= E[((X - E[X]) + (Y - E[Y]))^2]$$
$$= E[(X - E[X])^2] + E[(Y - E[Y])^2]$$
$$+ 2\,E[(X-EX)(Y-EY)] = \text{Var}(X) + \text{Var}(Y) + 2\,\text{Cov}(X,Y).$$

□.

note: if X and Y are uncorrelated (in particular if they are independent) then:

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y).$$

"Pythagorean theorem of statistics".

covariance is like the dot product of two vectors X and Y and then (2) is the law of cosines what if X and Y are not independent but we still want to at least bound $E[XY]$?

Theorem. (Cauchy-Schwarz) For any r.v.s X & Y for which $E[XY]$ is well-defined, we have

$$|E[XY]| \leq \sqrt{E[X^2]E[Y^2]}.$$

Proof. If $E[X^2]$ or $E[Y^2]$ is 0 then $P(X=0)=1$ or $P(Y=0)=1$ respectively, so the ineq. is trivial. So assume $E[Y^2], E[X^2] \in (0,\infty)$. Then for $\lambda \in \mathbb{R}$ we define $Z = \lambda X - Y$. so

$$0 \leq E[Z^2] = \lambda^2 E[X^2] - 2\lambda E[XY] + E[Y^2].$$

now the r.h.s is quadratic in λ, and it's always non-negative so the discriminant must be non-positive.

the discriminant is

$$0 \geq (2E[XY])^2 - 4E[X^2]E[Y^2]$$

so $(E[XY])^2 \leq E[X^2]E[Y^2]$.

$\square$

Again, can think of X and Y as vectors.

$$\|X\|^2 = E[X^2]$$

$$X \cdot Y = E[XY].$$

so C-S says that

$$|X \cdot Y| \leq \|X\| \cdot \|Y\|.$$

Markov's inequality:

Let X be a r.v. s.t. $P(X \geq 0) = 1$. i.e $X \geq 0$ a.s.

we say that an event E happens $\underbrace{\text{"almost surely"}}_{\text{a.s.}}$ if $P(E) = 1$

Then for any $a > 0$ we have
$$P(X \geq a) \leq \frac{1}{a} E[X].$$

Proof.
$$E[X] = \sum_x x P(X = x)$$
$$\geq \sum_{x \geq a} x P(X = x)$$

b.c $X \geq 0$ a.s

$$\geq a \sum_{x \geq a} P(X = x)$$
$$= a P(X \geq a). \qquad \square$$