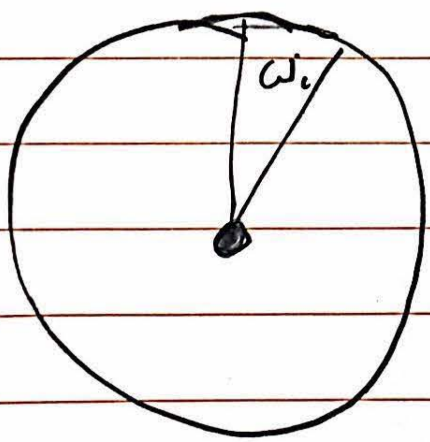
Lesson 8

Moments of Discrete Random Variables

Expectation:

We start our discussion on expectation with an example.

Example (Wheel of fortune)



Assume that the wheel has K slices. Slice i wins w_i dollars.

Also, assume that we spin the wheel K times, and slice i appears n_i times.

Obviously

$$\sum_{i=1}^{K} n_i =$$

The total amount of money that we receive is:

The average/per spin amount of money that we receive is

$$\boxed{}$$

which, when $n \to \infty$, becomes

$=$

This example is a motivation for <u>defining</u> the expectation of a discrete random variable with pmf p_X, $\mathbb{E}[X]$:

Def:

$\mathbb{E}[X] =$

Example: Bernoulli R.V.

$P_X(0) =$ $P_X(1) =$

$\mathbb{E}[X] =$

Example: Zipf distribution

Assume

$$P_X(x) = \frac{C}{x^{3/2}}, \quad x \in \mathbb{N}$$

$C =$

$\mathbb{E}[X] =$

Therefore, the expectation of $X \sim \mathrm{Zipf}(3/2)$ is infinity.

Example: $p_X(x) = \frac{c}{x^2+1}$, $x \in \mathbb{Z}$

$c =$

$\mathbb{E}\{X\}$

Expected Value of A Function of A Random Variable, or the Law of The Unconscious Statistician (LOTUS)

We previously realized that any real function of a discrete r.v. is a discrete random variable, i.e. for any $g: \mathbb{R} \to \mathbb{R}$, and X, $Y = g(X)$ is a R.V.

The question is how to calculate $\mathbb{E}[Y] = \mathbb{E}[g(X)]$. It seems that one has to calculate the pmf of Y to calculate $\mathbb{E}[Y]$.

However, $\mathbb{E}[g(X)]$ can be calculated using the pdf of X as:

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

More generally, if $\underline{X} = (X_1, X_2, \dots, X_n)$ is a vector of discrete r.v.'s with joint pmf $p_{\underline{X}} = p_{X_1, X_2, \ldots, X_n}$

and g is a multivariable

function: $g: \mathbb{R}^n \to \mathbb{R}$,

$$\mathbb{E}[g(\underline{X})] = \sum_{\underline{x}}$$

where $\underline{x} = (x_1, x_2, \ldots, x_n)$

Proof of LOTUS

We saw that if $Y = g(X)$,

$$P_Y(y) = \mathbb{P}(Y = y) = \sum_{\{x \mid g(x) = y\}} \mathbb{P}(X = x)$$

$$= \sum_{x \mid g(x) = y} P_X(x)$$

Then:

$\mathbb{E}[Y] =$

Properties of Expectation

Lemma: Let X, Y be discrete r.v.'s on $(\Omega, \mathcal{F}, \mathbb{P})$.

a) If $X \geq 0$ a.s. $\Rightarrow \mathbb{E}[X] \geq 0$

Remark: $X \geq 0$ a.s. means:

(b) If $X = c$ a.s. for some $c \in \mathbb{R}$ then $\mathbb{E}[X] = c$

(c) $a, b \in \mathbb{R}$, $\mathbb{E}[aX+bY]$

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

provided that expected values are well-defined on both sides (Linearity of Expectation)

Remark: The above equality doesn't hold if some of the expected values are not well-defined.

Example: $P_X(x) = \frac{c}{x^2+1}$

$Y = 2X \quad , \quad Z = aX + bY$

$a = -2 \; , \; b = 1$

(d) If $X^{\to}(\Omega) = \mathbb{Z}^{\geq 0}$ (i.e. X is a r.v. that can take non-negative integers) , then

$$\mathbb{E}[X] = \sum_{n \geq 0} \mathbb{P}(X > n)$$

(e) If X and Y are independent

$\Rightarrow \mathbb{E}[XY] = \boxed{}$

(f) If $X_1, X_2, \dots, X_n$ are independent discrete r.v.'s

then

$$\mathbb{E}\left[\prod_{i=1}^{n} X_i\right] =$$

provided that all expectations are well-defined

Proof: (a), (b) are direct consequences of definitions

(c) $\mathbb{E}[aX+bY]=$

(e) X, Y indep $\Rightarrow E[XY] = E[X]E[Y]$

(d) The proof is left as an exercise. Note that

$$\sum_{n \geq 0} \mathbb{P}(X > n) = \sum_{n \geq 0} \sum_{x > n} p_X(x)$$

The relationship between probability and expectation

Assume that X is a random variable (not necessarily discrete) and B is a Borel subset of $\mathbb{R}$.

Assume that I_B is the indicator function of B defined as

$$I_B : \mathbb{R} \to \{0, 1\}$$

$$I_B(x) = \begin{cases} 0 & x \notin B \\ 1 & x \in B \end{cases}$$

Then $Y = I_B(X)$ is a random variable (why?) and is discrete (why?) and:

$\mathbb{E}\,[I_B(X)] =$

Expectation of products of functions of random Variables.

Theorem: Let X, Y be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. X, Y are independent <u>if and</u> only if for all functions $g: \mathbb{R} \to \mathbb{R}$ and $h: \mathbb{R} \to \mathbb{R}$,

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

Proof:

Moments of Discrete Random Variables

The second moment of X:

Def: If X is a discrete r.v. and $g: \mathbb{R} \to \mathbb{R}$, $g(x) = x^2$,

$\mathbb{E}[g(X)] = \mathbb{E}[X^2] =$

is called the second moment of X.

The r^{th} moment:

Def: More generally, for all $r \in \mathbb{N}$, $\mathbb{E}[X^r] = \boxed{}$

is called the r^{th} moment of X.

Central Moments

Def: $\tilde{X} =$ ______________ is called the centralized version of X.

Def: $\mathbb{E}[\tilde{X}^r] = \mathbb{E}$

is called the r^{th} central moment of X, for all $r \in \mathbb{N}$.

$r=2$; The second central moment is called the variance of X, $Var(X)$.

Also, the standard deviation of X, σ_X, is defined as

$$\sigma_X =$$

Remark: If r is even, the r^{th} central moment is non-negative (why?)

Therefore, if the variance exists,

the standard deviation is well-defined.

Lemma) Let X be a discrete r.v. on $(\Omega, \mathcal{F}, \mathbb{P})$.

a) $\mathbb{E}[X] < \infty \Rightarrow Var(X) =$

(b) $a \in \mathbb{R}$, $Var(aX) =$

(c) If $X_1, X_2, \dots, X_n$ are independent discrete r.v.'s then

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

provided that all expectations are well-defined.

(d) $\mathrm{Var}(X)=0$ if and only if $X=c$ a.s.

Proof:

Skewness and Kurtosis

Def: Let X be a r.v. on $(\Omega, \mathcal{F}, \mathbb{P})$. Then

$$\mathbb{E}\left[\frac{(X - \mathbb{E}[X])^3}{\sigma_X^3}\right]$$

$= \mathbb{E}[\tilde{X}^3/\sigma_X^3]$ is called the skewness of X.

Def. Let X be a r.v. on $(\Omega, \mathcal{F}, \mathbb{P})$. Then

$$\mathbb{E}\left[\frac{(X - \mathbb{E}[X])^4}{\sigma_X^4}\right] =$$

$E\left[\tilde{X}^4 / (\text{var}(X))^2\right]$ is called the kurtosis of X.

Expectation and Variance of Important Discrete R.V.'s:

(a) Bernoulli

$E[X] =$

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

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

(b) $X \sim Bin(n, p)$

X can be seen as the sum of n independent Bernoulli r.v.'s, i.e

$X =$

$\mathbb{E}[X] =$

Also, $Var(X) = Var(\quad)$

$=$

(c) Geometric r.v.

$X \sim \text{Geo}_q(p)$

$p_X(x) = (1-p)^{n-1} p$

$\mathbb{E}[X] =$

Note that $\sum_{x=0}^{\infty} (1-p)^x =$

take derivatives with respect to p:

Second Method

$$\mathbb{E}[X] = \sum_{n=0}^{\infty} \mathbb{P}(X>n)$$

To calculate the variance of $X \sim Geo_1(p)$, one can calculate

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

The calculation of $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$ for $X \sim Geo_0(p)$ are left as an exercise.

(d) $X \sim Pois(\lambda)$

$$P_X(x) = e^{-\lambda} \frac{\lambda^x}{x!} \qquad x = 0, 1, \ldots$$

$$\mathbb{E}[X] = \sum_{x=0}^{\infty} x \, e^{-\lambda} \frac{\lambda^x}{x!}$$

To calculate $Var(X)$, we first calculate $E[X(X-1)]$:

Exercise: A negative binomial random variable $X \sim NegB(k,p)$ can be seen as the sum of k $Geo_1(p)$ independent r.v.'s. Use this to calculate the expectation and variance of X.

Example: Let X and Y be independent r.v.'s with $X \sim \text{Pois}(\lambda)$ and $Y \sim \text{Geo}_1(p)$.

Find $\mathbb{E}[X^2 Y + XY^2]$

(Functions of Indep. r.v.'s)

Exercise: Stein' Characterization for Poisson Random Variables

Assume that $X \sim Pois(\lambda)$ and $f: \mathbb{R} \to \mathbb{R}$. Then $\mathbb{E}[X \cdot f(X)] = \lambda \mathbb{E}[f(X+1)]$ provided that all expectations

exist.

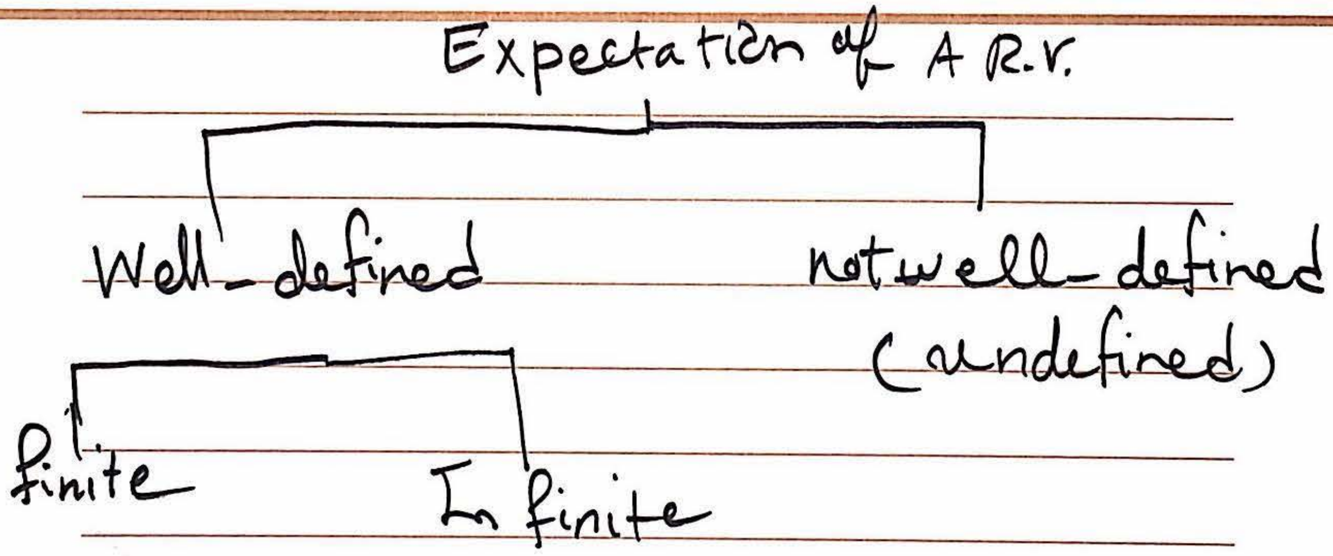
(e) $X \sim Pow(\alpha)$

$$P_X(x) = \frac{1}{x^{\alpha}} - \frac{1}{(x+1)^{\alpha}} \qquad \alpha > 0$$

$$\mathbb{E}[X] = \sum_{n=0}^{\infty} \mathbb{P}(X > n)$$

Existence of Expected Values

The following diagram shows all different possibilities for existence of expectation of a random variable



– Lemma: For every random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$.

(a) $\mathbb{E}[X]$ is undefined if and only if $\sum_{x<0} x p_X(x)$ and $\sum_{x>0} x p_X(x) > 0$ are both infinite.

(b) $\mathbb{E}[X]$ is well-defined and finite if and only if both of the sums $\sum_{x<0} x p_X(x)$ and $\sum_{x>0} x p_X(x)$ are both finite

(c) $\mathbb{E}[X]$ is well-defined and finite if and only if $\mathbb{E}[|X|] = \sum_{x} |x| p_X(x)$ is finite.

Remark: Random variables that have this property are called absolutely integrable (summable)

(d') : $\mathbb{E}[X]$ is well-defined and infinite iff exactly one of $\sum_{x>0} x p_X(x)$ and $\sum_{x<0} x p_X(x)$ is infinite

(e) $\mathbb{E}[X^2]$ is always well defined (finite or infinite)

Remark: If $\mathbb{E}[X^2] < \infty$, we call X square integrable (summable).

Proof: We only show (c) and (e).

(c) Remember if $\sum_n |a_n|$ is finite, then $\sum_n a_n$ is finite (because $a_n \leq |a_n|$)

Therefore $\mathbb{E}[|X|]<\infty \Rightarrow \mathbb{E}[X]<\infty$

because

$$\mathbb{E}[|X|]=\sum_x |x| P_X(x) = \sum_x |x P_X(x)|$$

when it is finite,

$\sum_x x P_X(x)$ is also finite,

which is $\mathbb{E}[X]$.

On the other hand, if

$\mathbb{E}[X]$ is finite, both $\sum_{x<0} x P_X(x)$

and $\sum_{x\geq 0} x P_X(x)$ are finite (See part b)

So $\sum_{x<0} x P_X(x)$ finite $\Rightarrow -\sum_{x<0} x P_X(x)$ finite.

$\Rightarrow \sum_{x<0} -x P_X(x)$ finite.

$$\sum_{x \geq 0} x \, p_X(x) < \infty \quad \Rightarrow \quad \sum_x |x| \, p_X(x) < \infty$$

$$\sum_{x < 0} -x \, p_X(x) < \infty$$

(e) $\sum_{x > 0} x^2 p_X(x)$ is finite or infinite $(+\infty)$

$\sum_{x < 0} x^2 p_X(x)$ is finite or infinite

but cannot be $-\infty$, so

$\mathbb{E}[X^2]$ is <u>always well</u>

<u>defined.</u>

Def: : A random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ is called square integrable iff $\mathbb{E}[X^2] < \infty$.

Lemma: A square integrable random variable is absolute integrable; therefore, has a finite expectation.

Proof: $|x| \leq 1 + x^2$

Lemma (Existence and Finiteness of Variance) $Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

a) If X is square integrable then X has ______ variance

(b) If X is integrable but not square integrable then X has ______ variance

(c) If X is not integrable then the variance of X is ______

Proof: Obvious.

Exercise:

$$p_X(x) = \frac{c}{x^{\alpha}+1} \qquad x \geq 0 \qquad \alpha > 1$$

When does X have finite, infinite, or undefined variance? Repeat the exercise with $x \in \mathbb{Z}$.

Covariance and Correlation

Def:

Assume that X, Y are square integrable r.v.'s on $(\Omega, \mathcal{F}, \mathbb{P})$.

The Covariance of X and Y is defined as:

$$\mathrm{Cov}(X,Y) = \mathbb{E}[\qquad\qquad]$$

(centralized X,Y): $= \mathbb{E}[\qquad\qquad]$

When $\mathrm{Cov}(X,Y)=0$, X,Y are called <u>uncorrelated</u>

Lemma (Properties of Covariance)

a) $Cov(X,X) = Var(X)$

b) $\forall a \in \mathbb{R}$ $Cov(X, Y+a) = Cov(X,Y)$

c) $\forall a,b \in \mathbb{R}$ $Cov(X, aY+bZ)$
$$= a\,Cov(X,Y) + b\,Cov(X,Z)$$

d) $Cov(X,Y) = Cov(Y,X)$

Remark: Any function $f(x,y)$ that is linear with respect to both of its variables is called a <u>bilinear function</u>,

$f(x, ay+bz) =$

$f(ax+by, z)$

(e) (Alternative formula for covariance)

$Cov(X,Y) =$

(f) (Variance of Sum of R.V.'s)

If X_1, X_2 have finite variance

$Var(X_1 + X_2) =$

(g) If $X_1, X_2, \ldots, X_n$ have finite variance:

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} Cov(X_i, X_j)$$

Example:

$Var(X_1 + X_2 + X_3) =$

Proof: (f, g):

Important Note:

If X, Y are independent then they are uncorrelated, because

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

The reverse is NOT TRUE in general.

Example: Assume that

$$P_{X,Y}(1,0) = P_{X,Y}(0,1) = P_{X,Y}(-1,0) = P_{X,Y}(0,-1)$$
$$= P_{X,Y}(2,0) = P_{X,Y}(0,3) = P_{X,Y}(-2,0) = P_{X,Y}(0,-3)$$
$$= 1/8$$

$\mathbb{E}[X] =$

$\mathbb{E}[Y] =$

$\mathbb{E}[XY] =$

$Cor(X, Y) =$

$\Rightarrow X, Y$ are

Are X, Y independent?

$\mathbb{P}(X=0) = P_X(0) =$

$\mathbb{P}(Y=0) = P_Y(0) =$

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

Lemma: (The Cauchy–Schwartz–Bunyakorsky Inequality for Expectation)

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

Proof: See Supplement.

Correlation Coefficient

Def: Assume that X and Y are two random variables with non-zero and finite variances. The correlation coefficient of X and Y is defined as:

$\rho(X,Y) =$

Theorem: Let X and Y be random variables with positive variances $Var(X)>0$, $Var(Y)>0$

(a) $-1 \leq \rho(X,Y) \leq 1$

(b) $\rho(X,Y)=1$ if and only if $\tilde{Y}=a\tilde{X}$ and $a>0$.

(c) $\rho(X,Y)=-1$ if and only if $\tilde{Y}=a\tilde{X}$ and $a<0$.

Exercise: Assume that

$X \sim Bin(n, p)$ and

$Y = n - X$. Find $\rho(X, Y)$.