CS70: Alex Psomas: Lecture 19.

1. Random Variables: Brief Review
2. Some details on distributions: Geometric. Poisson.
3. Joint distributions.
4. Linearity of Expectation.

Random Variables: Definitions

Is a random variable random?
NO!
Is a random variable a variable?
NO!
Great name!

Random Variables: Definitions

Definition

A random variable, X, for a random experiment with sample space Ω is a function $X : \Omega \to \Re$.

Thus, $X(\cdot)$ assigns a real number $X(\omega)$ to each $\omega \in \Omega$.

Definitions

(a) For $a \in \Re$, one defines

$$X^{-1}(a) := \{\omega \in \Omega \mid X(\omega) = a\}.$$

(b) For $A \subset \Re$, one defines

$$X^{-1}(A) := \{\omega \in \Omega \mid X(\omega) \in A\}.$$

(c) The probability that $X = a$ is defined as

$$Pr[X = a] = Pr[X^{-1}(a)].$$

(d) The probability that $X \in A$ is defined as

$$Pr[X \in A] = Pr[X^{-1}(A)].$$

(e) The distribution of a random variable X, is

$$\{(a, Pr[X = a]) : a \in \mathscr{A}\},$$

where $\mathscr{A}$ is the *range* of X. That is, $\mathscr{A} = \{X(\omega), \omega \in \Omega\}$.

Expectation - Definition

Definition: The **expected value** (or mean, or expectation) of a random variable X is

$$E[X] = \sum_{a \in \mathbb{R}} a \times Pr[X = a].$$

Theorem:

$$E[X] = \sum_{\omega \in \Omega} X(\omega) \times Pr[\omega].$$

An Example

Flip a fair coin three times.

$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$.

$X =$ number of H's: $\{3, 2, 2, 2, 1, 1, 1, 0\}$.

- Range of X? $\{0, 1, 2, 3\}$. All the values X can take.
- $X^{-1}(2)$? $X^{-1}(2) = \{HHT, HTH, THH\}$. All the **outcomes** ω such that $X(\omega) = 2$.
- Is $X^{-1}(1)$ an event? **YES**. It's a subset of the outcomes.
- $Pr[X]$? This doesn't make any sense bro....
- $Pr[X = 2]$?

$$Pr[X = 2] = Pr[X^{-1}(2)] = Pr[\{HHT, HTH, THH\}]$$

$$= Pr[\{HHT\}] + Pr[\{HTH\}] + Pr[\{THH\}] = \frac{3}{8}$$

An Example

Flip a fair coin three times.

$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$.

X = number of H's: $\{3, 2, 2, 2, 1, 1, 1, 0\}$.

Thus,

$$E[X] = \sum_{\omega \in \Omega} X(\omega) Pr[\omega] = \frac{3}{8} + \frac{2}{8} + \frac{2}{8} + \frac{2}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + 0 = \frac{12}{8}$$

Also,

$$E[X] = \sum_{a \in \mathbb{R}} a \times Pr[X = a] = 3 \times \frac{1}{8} + 2 \times \frac{3}{8} + 1 \times \frac{3}{8} + 0 \times \frac{1}{8}.$$

Win or Lose.

Expected winnings for heads/tails games, with 3 flips?
Recall the definition of the random variable X:
$\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\} \rightarrow \{3, 1, 1, -1, 1, -1, -1, -3\}$.

$$E[X] = 3 \times \frac{1}{8} + 1 \times \frac{3}{8} - 1 \times \frac{3}{8} - 3 \times \frac{1}{8} = 0.$$

Can you ever win 0?

Apparently: expected value is not a common value, by any means. It doesn't have to be in the range of X.

The expected value of X is not the value that you expect!
Great name once again!
It is the average value per experiment, if you perform the experiment many times:

$$\frac{X_1 + \cdots + X_n}{n}, \text{ when } n \gg 1.$$

The fact that this average converges to $E[X]$ is a theorem:
the Law of Large Numbers. (See later.)

Geometric Distribution

Let's flip a coin with $Pr[H] = p$ until we get H.

For instance:

$$\omega_1 = H, \text{ or}$$
$$\omega_2 = T\ H, \text{ or}$$
$$\omega_3 = T\ T\ H, \text{ or}$$
$$\omega_n = T\ T\ T\ T\ \cdots\ T\ H.$$

Note that $\Omega = \{\omega_n, n = 1, 2, \ldots\}$. (Notice: no distribution yet!)

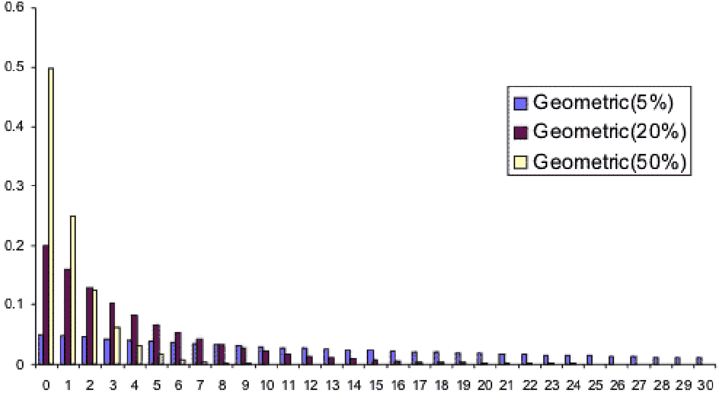
Let X be the number of flips until the first H. Then, $X(\omega_n) = n$.

Also,

$$Pr[X = n] = (1-p)^{n-1}p, \; n \geq 1.$$

Geometric Distribution

$$Pr[X = n] = (1-p)^{n-1}p, n \geq 1.$$



Geometric Distribution: A weird trick

Recall the Geometric Distribution.

$$Pr[X = n] = (1-p)^{n-1}p, n \geq 1.$$

Note that

$$\sum_{n=1}^{\infty} Pr[X = n] = \sum_{n=1}^{\infty} (1-p)^{n-1}p = p\sum_{n=1}^{\infty}(1-p)^{n-1} = p\sum_{n=0}^{\infty}(1-p)^n.$$

We want to analyze $S := \sum_{n=0}^{\infty} a^n$ for $|a| < 1$. $S = \frac{1}{1-a}$. Indeed,

$$\begin{aligned} S &= 1 + a + a^2 + a^3 + \cdots \\ aS &= \quad\; a + a^2 + a^3 + a^4 + \cdots \\ (1-a)S &= 1 + a - a + a^2 - a^2 + \cdots = 1. \end{aligned}$$

Hence,

$$\sum_{n=1}^{\infty} Pr[X = n] = p\,\frac{1}{1-(1-p)} = 1.$$

Geometric Distribution: Expectation

$$X =_D G(p), \text{ i.e., } Pr[X = n] = (1-p)^{n-1}p, n \geq 1.$$

One has

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

Thus,

$$\begin{aligned}
E[X] &= p + 2(1-p)p + 3(1-p)^2p + 4(1-p)^3p + \cdots \\
(1-p)E[X] &= (1-p)p + 2(1-p)^2p + 3(1-p)^3p + \cdots \\
pE[X] &= p + (1-p)p + (1-p)^2p + (1-p)^3p + \cdots \\
&\qquad \text{by subtracting the previous two identities} \\
&= p\sum_{n=0}^{\infty}(1-p)^n = 1.
\end{aligned}$$

Hence,

$$E[X] = \frac{1}{p}.$$

Geometric Distribution: Memoryless

I flip a coin (probability of H is p) until I get H.

What's the probability that I flip it exactly 100 times? $(1-p)^{99}p$

What's the probability that I flip it exactly 100 times if (given that) the first 20 were T?

Same as flipping it exactly 80 times!

$(1-p)^{79}p$.

Geometric Distribution: Memoryless

Let X be $G(p)$. Then, for $n \geq 0$,

$$Pr[X > n] = Pr[\text{ first } n \text{ flips are } T] = (1-p)^n.$$

Theorem

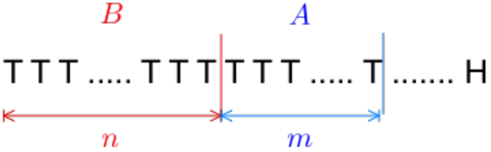
$$Pr[X > n+m | X > n] = Pr[X > m], m, n \geq 0.$$

Proof:

$$\begin{aligned} Pr[X > n+m | X > n] &= \frac{Pr[X > n+m \text{ and } X > n]}{Pr[X > n]} \\ &= \frac{Pr[X > n+m]}{Pr[X > n]} \\ &= \frac{(1-p)^{n+m}}{(1-p)^n} = (1-p)^m \\ &= Pr[X > m]. \end{aligned}$$

Geometric Distribution: Memoryless - Interpretation

$$Pr[X > n+m | X > n] = Pr[X > m], m, n \geq 0.$$



$$Pr[X > n+m | X > n] = Pr[A|B] = Pr[A] = Pr[X > m].$$

The coin is memoryless, therefore, so is X.

Geometric Distribution: Yet another look

Theorem: For a r.v. X that takes the values $\{0,1,2,\ldots\}$, one has

$$E[X] = \sum_{i=1}^{\infty} Pr[X \geq i].$$

[See later for a proof.]

If $X = G(p)$, then $Pr[X \geq i] = Pr[X > i-1] = (1-p)^{i-1}$.

Hence,

$$E[X] = \sum_{i=1}^{\infty} (1-p)^{i-1} = \sum_{i=0}^{\infty} (1-p)^{i} = \frac{1}{1-(1-p)} = \frac{1}{p}.$$

Expected Value of Integer RV

Theorem: For a r.v. X that takes values in $\{0, 1, 2, \ldots\}$, one has

$$E[X] = \sum_{i=1}^{\infty} Pr[X \geq i].$$

Proof: One has

$$\begin{aligned}
E[X] &= \sum_{i=1}^{\infty} i \times Pr[X = i] \\
&= \sum_{i=1}^{\infty} i\,(Pr[X \geq i] - Pr[X \geq i+1]) \\
&= \sum_{i=1}^{\infty} (i \times Pr[X \geq i] - i \times Pr[X \geq i+1]) \\
&= \sum_{i=1}^{\infty} (i \times Pr[X \geq i] - (i-1) \times Pr[X \geq i]) \\
&= \sum_{i=1}^{\infty} Pr[X \geq i].
\end{aligned}$$

□

Poisson Distribution: Definition and Mean

Definition Poisson Distribution with parameter $\lambda > 0$

$$X = P(\lambda) \Leftrightarrow Pr[X = m] = \frac{\lambda^m}{m!} e^{-\lambda}, m \geq 0.$$

Fact: $E[X] = \lambda$.

Proof:

$$\begin{aligned} E[X] &= \sum_{m=1}^{\infty} m \times \frac{\lambda^m}{m!} e^{-\lambda} = e^{-\lambda} \sum_{m=1}^{\infty} \frac{\lambda^m}{(m-1)!} \\ &= e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^{m+1}}{m!} = e^{-\lambda} \lambda \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} \\ &= e^{-\lambda} \lambda e^{\lambda} = \lambda. \end{aligned}$$

□

Used Taylor expansion of e^x at 0 : $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$.

Simeon Poisson

The Poisson distribution is named after:

Siméon Poisson

Siméon Denis Poisson (1781–1840)

Indicators

Definition

Let A be an event. The random variable X defined by

$$X(\omega) = \begin{cases} 1, & \text{if } \omega \in A \\ 0, & \text{if } \omega \notin A \end{cases}$$

is called the indicator of the event A.

Note that $Pr[X = 1] = Pr[A]$ and $Pr[X = 0] = 1 - Pr[A]$.

Hence,

$$E[X] = 1 \times Pr[X = 1] + 0 \times Pr[X = 0] = Pr[A].$$

This random variable $X(\omega)$ is sometimes written as

$$1\{\omega \in A\} \text{ or } 1_A(\omega).$$

Thus, we will write $X = 1_A$.

Review: Distributions

- $U[1,\ldots,n] : Pr[X = m] = \frac{1}{n}, m = 1,\ldots,n;$
  $E[X] = \frac{n+1}{2};$
- $B(n,p) : Pr[X = m] = \binom{n}{m} p^m (1-p)^{n-m}, m = 0,\ldots,n;$
  $E[X] = np;$ (TODO)
- $G(p) : Pr[X = n] = (1-p)^{n-1} p, n = 1,2,\ldots;$
  $E[X] = \frac{1}{p};$
- $P(\lambda) : Pr[X = n] = \frac{\lambda^n}{n!} e^{-\lambda}, n \geq 0;$
  $E[X] = \lambda.$

Joint distribution.

Two random variables, X and Y, in prob space: $(\Omega, P(\cdot))$.

What is $\sum_x Pr[X = x]$? 1. What $\sum_x Pr[Y = y]$? 1.

Let's think about: $Pr[X = x, Y = y]$.
What is $\sum_{x,y} Pr[X = x, Y = y]$?

Are the events "$X = x, \; Y = y$" disjoint?
Yes! Y and X are functions on Ω
Do they cover the entire sample space?
Yes! X and Y are functions on Ω.

So, $\sum_{x,y} Pr[X = x, Y = y] = 1$.

Joint Distribution: $Pr[X = x, Y = y]$.
Marginal Distributions: $Pr[X = x]$ and $Pr[Y = y]$.
Important for inference.

Two random variables, same outcome space.

Experiment: pick a random person.
X = number of episodes of Games of Thrones they have seen.
Y = number of episodes of Westworld they have seen.

X =	0	1	2	3	5	40	All
Pr	0.3	0.05	0.05	0.05	0.05	0.1	0.4

Is this a distribution?

Yes! All the probabilities are non-negative and add up to 1.

Y =	0	1	5	10
Pr	0.3	0.1	0.1	0.5

Joint distribution: Example.

The **joint distribution** of X and Y is:

Y/X	0	1	2	3	5	40	All	
0	0.15	0	0	0	0	0.1	0.05	=0.3
1	0	0.05	0.05	0	0	0	0	=0.1
5	0	0	0	0.05	0.05	0	0	=0.1
10	0.15	0	0	0	0	0	0.35	=0.5
	=0.3	=0.05	=0.05	=0.05	=0.05	=0.1	=0.4	

Is this a valid distribution? Yes!
Notice that $Pr[X = a]$ and $Pr[Y = b]$ are (marginal) distributions!
But now we have more information!

For example, if I tell you someone watched 5 episodes of Westworld, they definitely didn't watch all the episodes of GoT.

Combining Random Variables

Definition
Let X, Y, Z be random variables on Ω and $g : \Re^3 \to \Re$ a function. Then $g(X, Y, Z)$ is the random variable that assigns the value $g(X(\omega), Y(\omega), Z(\omega))$ to ω.

Thus, if $V = g(X, Y, Z)$, then $V(\omega) := g(X(\omega), Y(\omega), Z(\omega))$.

Examples:

- X^k
- $(X - a)^2$
- $a + bX + cX^2 + (Y - Z)^2$
- $(X - Y)^2$
- $X \cos(2\pi Y + Z)$.

Linearity of Expectation

Theorem: Expectation is linear

$$E[a_1X_1+\cdots+a_nX_n]=a_1E[X_1]+\cdots+a_nE[X_n].$$

Proof:

$$\begin{aligned} & E[a_1X_1+\cdots+a_nX_n] \\ &= \sum_{\omega}(a_1X_1+\cdots+a_nX_n)(\omega)Pr[\omega] \\ &= \sum_{\omega}(a_1X_1(\omega)+\cdots+a_nX_n(\omega))Pr[\omega] \\ &= a_1\sum_{\omega}X_1(\omega)Pr[\omega]+\cdots+a_n\sum_{\omega}X_n(\omega)Pr[\omega] \\ &= a_1E[X_1]+\cdots+a_nE[X_n]. \end{aligned}$$

□

Note: If we had defined $Y=a_1X_1+\cdots+a_nX_n$ and had tried to compute $E[Y]=\sum_y yPr[Y=y]$, we would have been in trouble!

Using Linearity - 1: Pips (dots) on dice

Roll a die n times.

X_m = number of pips on roll m.

$X = X_1 + \cdots + X_n$ = total number of pips in n rolls.

$$\begin{aligned} E[X] &= E[X_1 + \cdots + X_n] \\ &= E[X_1] + \cdots + E[X_n], \text{ by linearity} \\ &= nE[X_1], \text{ because the } X_m \text{ have the same distribution} \end{aligned}$$

Now,

$$E[X_1] = 1 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = (1 + 2 + \cdots + 6) \times \frac{1}{6} = \frac{7}{2}.$$

Hence,

$$E[X] = \frac{7n}{2}.$$

Note: Computing $\sum_x xPr[X = x]$ directly is not easy!

Using Linearity - 2: Fixed point.

Hand out assignments at random to n students.

$X =$ number of students that get their own assignment back.

$X = X_1 + \cdots + X_n$ where
$X_m = 1\{$student m gets his/her own assignment back$\}$.

One has

$$
\begin{aligned}
E[X] &= E[X_1 + \cdots + X_n] \\
&= E[X_1] + \cdots + E[X_n], \text{ by linearity} \\
&= nE[X_1], \text{ because all the } X_m \text{ have the same distribution} \\
&= nPr[X_1 = 1], \text{ because } X_1 \text{ is an indicator} \\
&= n(1/n), \text{ because student 1 is equally likely} \\
&\qquad\qquad \text{to get any one of the } n \text{ assignments} \\
&= 1.
\end{aligned}
$$

Note that linearity holds even though the X_m are not independent (whatever that means).

Note: What is $Pr[X = m]$? Tricky ....

Using Linearity - 3: Binomial Distribution.

Flip n coins with heads probability p. X - number of heads

Binomial Distibution: $Pr[X = i]$, for each i.

$$Pr[X = i] = \binom{n}{i} p^i (1-p)^{n-i}.$$

$$E[X] = \sum_i i \times Pr[X = i] = \sum_i i \times \binom{n}{i} p^i (1-p)^{n-i}.$$

No no no no no. NO ... Or... a better approach: Let

$$X_i = \begin{cases} 1 & \text{if } i\text{th flip is heads} \\ 0 & \text{otherwise} \end{cases}$$

$E[X_i] = 1 \times Pr[\text{“}heads''] + 0 \times Pr[\text{“}tails''] = p.$

Moreover $X = X_1 + \cdots X_n$ and

$E[X] = E[X_1] + E[X_2] + \cdots E[X_n] = n \times E[X_i] = np.$

Using Linearity - 4: Expected number of times a word appears.

Alex is typing a document randomly: Each letter has a probability of $\frac{1}{26}$ of being typed. The document will be $100,000,000$ letters long. What is the expected number of times that the word "pizza" will appear?

Let X be a random variable that counts the number of times the word "pizza" appears. We want $E(X)$.

$$E(X) = \sum_{\omega} X(\omega) Pr[\omega].$$

Better approach: Let X_i be the indicator variable that takes value 1 if "pizza" starts on the i-th letter, and 0 otherwise. i takes values from 1 to $100,000,000 - 4 = 99,999,996$.

hpizzafgnpizzadjgbidgne....

$X_2 = 1$, $X_{10} = 1$,...

Using Linearity - 4: Expected number of times a word appears.

$$E(X_i) = (\frac{1}{26})^5$$

Therefore,

$$E(X) = E(\sum_i X_i) = \sum_i E(X_i) = 99,999,996(\frac{1}{26})^5 \approx 8.4$$

Calculating $E[g(X)]$

Let $Y = g(X)$. Assume that we know the distribution of X.

We want to calculate $E[Y]$.

Method 1: We calculate the distribution of Y:

$$Pr[Y = y] = Pr[X \in g^{-1}(y)] \text{ where } g^{-1}(x) = \{x \in \Re : g(x) = y\}.$$

This is typically rather tedious!

Method 2: We use the following result.

Theorem:

$$E[g(X)] = \sum_v g(v) Pr[X = v].$$

Proof:

$$\begin{aligned}
E[g(X)] &= \sum_{\omega} g(X(\omega)) Pr[\omega] = \sum_v \sum_{\omega \in X^{-1}(v)} g(X(\omega)) Pr[\omega] \\
&= \sum_v \sum_{\omega \in X^{-1}(v)} g(v) Pr[\omega] = \sum_v g(v) \sum_{\omega \in X^{-1}(v)} Pr[\omega] \\
&= \sum_v g(v) Pr[X = v].
\end{aligned}$$

$\square$

An Example

Let X be uniform in $\{-2,-1,0,1,2,3\}$.

Let also $g(X) = X^2$. Then (method 2)

$$\begin{aligned} E[g(X)] &= \sum_{x=-2}^{3} x^2 \frac{1}{6} \\ &= \{4+1+0+1+4+9\}\frac{1}{6} = \frac{19}{6}. \end{aligned}$$

Method 1 - We find the distribution of $Y = X^2$:

$$Y = \begin{cases} 4, & \text{w.p. } \frac{2}{6} \\ 1, & \text{w.p. } \frac{2}{6} \\ 0, & \text{w.p. } \frac{1}{6} \\ 9, & \text{w.p. } \frac{1}{6}. \end{cases}$$

Thus,

$$E[Y] = 4\frac{2}{6} + 1\frac{2}{6} + 0\frac{1}{6} + 9\frac{1}{6} = \frac{19}{6}.$$

Summary

Random Variables

- A random variable X is a function $X : \Omega \to \Re$.
- $Pr[X = a] := Pr[X^{-1}(a)] = Pr[\{\omega \mid X(\omega) = a\}]$.
- $Pr[X \in A] := Pr[X^{-1}(A)]$.
- The distribution of X is the list of possible values and their probability: $\{(a, Pr[X = a]), a \in \mathscr{A}\}$.
- Joint distributions.
- $g(X, Y, Z)$ assigns the value .... .
- $E[X] := \sum_a a Pr[X = a]$.
- Expectation is Linear.