CS70: Jean Walrand: Lecture 27.

Expectation; Conditional Expectation; B(n, p); G(p)

1. Review of Expectation
2. Linearity of Expectation
3. Conditional Expectation
4. Independence of RVs
5. Applications
6. Important Distributions and Expectations.

Expectation

Recall: $X : \Omega \to \Re; Pr[X = a]; = Pr[X^{-1}(a)]$;

Definition: The **expectation** of a random variable X is

$$E[X] = \sum_a a \times Pr[X = a].$$

Indicator:

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,

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

The random variable X is sometimes written as

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

Linearity of Expectation

Theorem:

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

Theorem: Expectation is linear

$$E[a_1 X_1 + \cdots + a_n X_n] = a_1 E[X_1] + \cdots + a_n E[X_n].$$

Proof:

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

□

Using Linearity - 1: Pips 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,

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

Hence,

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

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
$X_m = 1\{\text{student } m \text{ gets his/her own assignment back}\}$.

One has

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

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

Using Linearity - 3: Binomial Distribution.

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

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

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

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

Uh oh. ... Or... a better approach: Let

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

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

Moreover $X = X_1 + \cdots X_n$ and

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

Conditional Expectation

How do observations affect expectation?

Example 1:
Roll one die. You are told that the outcome X is at least 3. What is the expected value of X given that information?

Given that $X \geq 3$, we know that X is uniform in $\{3,4,5,6\}$. Hence, the mean value is 4.5. We write

$$E[X|X \geq 3] = 4.5.$$

Similarly, we have

$$E[X|X < 3] = 1.5$$

because, given that $X < 3$, X is uniform in $\{1,2\}$.

Note that

$$\begin{aligned} &E[X|X \geq 3] \times Pr[X \geq 3] + E[X|X < 2] \times Pr[X < 2] \\ &= 4.5 \times \frac{4}{6} + 1.5 \times \frac{2}{6} = 3 + 0.5 = 3.5 = E[X]. \end{aligned}$$

Is this a coincidence?

Conditional Expectation

How do observations affect expectation?
Example 2:
Roll two dice. You are told that the total number X of pips is at least 8. What is the expected value of X given that information?

Recall the distribution of X: $Pr[X=2] = Pr[X=12] = 1/36, Pr[X=3] = Pr[X=11] = 2/36, \ldots$.

Given that $X \geq 8$, the distribution of X becomes

$$\{(8,5/15),(9,4/15),(10,3/15),(11,2/15),(12,1/15)\}.$$

For instance,

$$Pr[X=8|X \geq 8] = \frac{Pr[X=8]}{Pr[X \geq 8]} = \frac{5/36}{15/36} = \frac{5}{15}.$$

Hence,

$$E[X|X \geq 8] = 8\frac{5}{15} + 9\frac{4}{15} + 10\frac{3}{15} + 11\frac{2}{15} + 12\frac{1}{15} = \frac{140}{15} \approx 9.33.$$

Conditional Expectation

How do observations affect expectation?
Example 2: continued
Roll two dice. You are told that the total number X of pips is less than 8. What is the expected value of X given that information?

We find that

$$E[X|X < 8] = 2\frac{1}{21} + 3\frac{3}{21} + \cdots + 7\frac{6}{21} = \frac{112}{21} \approx 5.33.$$

Observe that

$$\begin{aligned} &E[X|X \geq 8]Pr[X \geq 8] + E[X|X < 8]Pr[X < 8] \\ &= 9.33 \times \frac{15}{36} + 5.33\frac{21}{36} \\ &= 7 = E[X]. \end{aligned}$$

Coincidence? Probably not.

Conditional Probability

Definition

Let X be a RV and A an event. Then

$$E[X|A] := \sum_a a \times Pr[X = a|A].$$

It is easy (really) to see that

$$E[X|A] = \sum_\omega X(\omega) Pr[\omega|A] = \frac{1}{Pr[A]} \sum_{\omega \in A} X(\omega) Pr[\omega].$$

Theorem Conditional Expectation is linear

$$E[a_1 X_1 + \cdots + a_n X_n|A] = a_1 E[X_1|A] + \cdots + a_n E[X_n|A].$$

Proof:

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

□

Conditional Probability

Theorem

$$E[X] = E[X|A]Pr[A] + E[X|\bar{A}]Pr[\bar{A}].$$

Proof
The law of total probability says that

$$Pr[\omega] = Pr[\omega|A]Pr[A] + Pr[\omega|\bar{A}]Pr[\bar{A}].$$

Hence,

$$\begin{aligned} E[X] &= \sum_\omega X(\omega) Pr[\omega] \\ &= \sum_\omega X(\omega) Pr[\omega|A] Pr[A] + \sum_\omega X(\omega) Pr[\omega|\bar{A}] Pr[\bar{A}] \\ &= E[X|A]Pr[A] + E[X|\bar{A}]Pr[\bar{A}]. \end{aligned}$$

□

Geometric Distribution

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

For instance:

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

Note that $\Omega = \{\omega_n, n = 1,2,\ldots\}$.

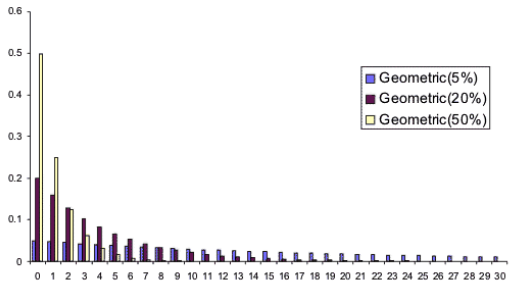
Let X be the number of flips until the first H. Then, $X(\omega_n) = n$.

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

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

Note that

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

Now, if $|a|<1$, then $S:=\sum_{n=0}^{\infty}a^n=\frac{1}{1-a}$. 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,

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

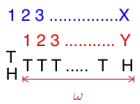
Hence,

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

Geometric Distribution: Renewal Trick

A different look at the algebra.

We flip the coin once, and, if we get T, let ω be the following flips.



Note that

$$X(H\omega)=1 \text{ and } X(T\omega)=1+Y(\omega).$$

Hence,

$$\begin{aligned} E[X] &= \sum_{\omega} 1\times Pr[H\omega]+\sum_{\omega}(1+Y(\omega))Pr[T\omega] \\ &= \sum_{\omega} pPr[\omega]+\sum_{\omega}(1+Y(\omega))(1-p)Pr[\omega] \\ &= p+(1-p)(1+E[Y])=1+(1-p)E[Y]. \end{aligned}$$

But, $E[X]=E[Y]$. Thus, $E[X]=1+(1-p)E[X]$, so that $E[X]=1/p$.

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

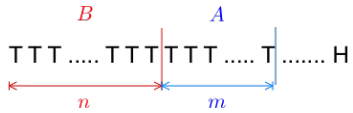
$$Pr[X>n+m|X>n]=Pr[X>m], m,n\geq 0.$$

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

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

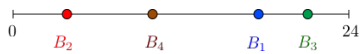
Proof: One has

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

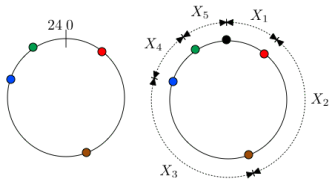
□

Riding the bus.

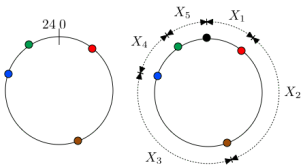
n buses arrive uniformly at random throughout a 24 hour day. What is the time between buses? What is the time to wait for a bus? Here are typical arrival times, independent and uniform in $[0,24]$.



Here is an alternative picture (left)



Riding the bus.



Add the black dot uniformly at random and pretend that it represents 0/24.

This is legitimate, because given the black dot, the other dots are uniform at random. Then,

$$\begin{aligned} 24 &= E[X_1 + \cdots + X_5] \\ &= 5E[X_1], \text{ by linearity and symmetry} = 5E(X_1]. \end{aligned}$$

Hence,

$$E[X_1] = E[X_m] = \frac{24}{5} = \frac{24}{n+1} \text{ for } n \text{ busses.}$$

Summary.

Expectation; Conditional Expectation; B(n, p); G(p)

Expectation: $E[X] = \sum_a a \times Pr[X = a] = \sum_\omega X(\omega) Pr[\omega]$.

Linearity: $E[a_1 X_1 + \cdots + a_n X_n] = a_1 E[X_1] + \cdots + a_n E[X_n]$.

Binomial: $Pr[X = i] = \binom{n}{k} p^i (1-p)^{(n-i)}; E(X) = pn$.

Geometric: $Pr[X = i] = (1-p)^{(i-1)} p; E(X) = \frac{1}{p}$; memoryless.

Condition Expectation: $E[X|A]$. Linear and

$$E[X] = E[X|A]Pr[A] + E[X|\bar{A}]Pr[\bar{A}].$$