# CS70: Jean Walrand: Lecture 26.

Expectation; Geometric & Poisson

1. Random Variables: Brief Review
2. Expectation
3. Linearity of Expectation
4. Geometric Distribution
5. Poisson Distribution

# Random Variables: Review

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

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

**Definitions**
For $a \in \Re$, one defines $X^{-1}(a) := \{\omega \in \Omega \mid X(\omega) = a\}$.

The probability that $X = a$ is defined as $Pr[X = a] = Pr[X^{-1}(a)]$.

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, $\mathscr{A} = \{X(\omega), \omega \in \Omega\}$.

Let $X, Y, Z$ be random variables on $\Omega$ 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 $\omega$.

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

# Expectation

**Definition:** The **expectation** (mean, expected value) 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:**

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

$\square$

# Using Linearity - 1: Dots on dice

Roll a die $n$ times.

$X_m$ = number of dots on roll $m$.

$X = X_1 + \cdots + X_n$ = total number of dots 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,

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

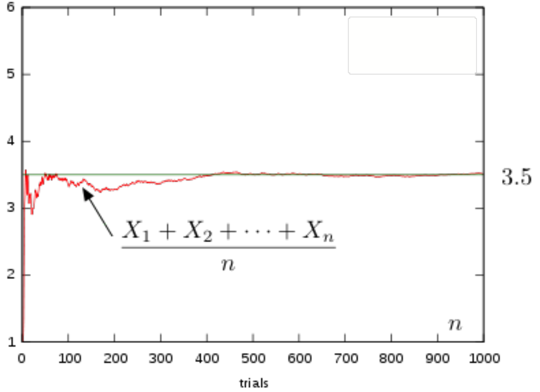
Hence,

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

# Strong Law of Large Numbers: An Example

**Rolling Dice.** $X_m$ = number of dots on roll $m$.

Theorem: $$\frac{X_1 + X_2 + \cdots + X_n}{n} \to E[X_1] = 3.5 \text{ as } n \to \infty.$$



# 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).

# 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

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

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

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

Assume $A$ and $B$ are disjoint events. Then $1_{A\cup B}(\omega) = 1_A(\omega) + 1_B(\omega)$.

Taking expectation, we get

$$Pr[A\cup B] = E[1_{A\cup B}] = E[1_A + 1_B] = E[1_A] + E[1_B] = Pr[A] + Pr[B].$$

In general, $1_{A\cup B}(\omega) = 1_A(\omega) + 1_B(\omega) - 1_{A\cap B}(\omega)$.

Taking expectation, we get $Pr[A\cup B] = Pr[A] + Pr[B] - Pr[A\cap B]$.

Observe that if $Y(\omega) = b$ for all $\omega$, then $E[Y] = b$.

Thus, $E[X + b] = E[X] + b$.

# 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:**

$$E[g(X)] = \sum_x g(x) Pr[X = x].$$

**Proof:**

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

$\square$

# 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,

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

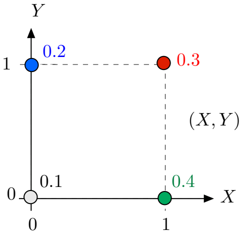
# Calculating $E[g(X,Y,Z)]$

We have seen that $E[g(X)] = \sum_x g(x)Pr[X=x]$.

Using a similar derivation, one can show that

$$E[g(X,Y,Z)] = \sum_{x,y,z} g(x,y,z)Pr[X=x,Y=y,Z=z].$$

**An Example.** Let $X, Y$ be as shown below:

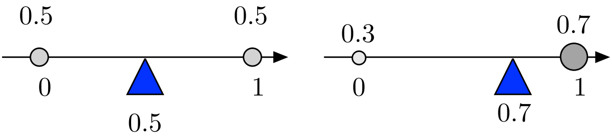


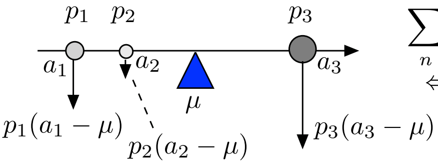

$$(X,Y) = \begin{cases} (0,0), & \text{w.p. } 0.1 \\ (1,0), & \text{w.p. } 0.4 \\ (0,1), & \text{w.p. } 0.2 \\ (1,1), & \text{w.p. } 0.3 \end{cases}$$

$$\begin{aligned} E[\cos(2\pi X + \pi Y)] &= 0.1\cos(0) + 0.4\cos(2\pi) + 0.2\cos(\pi) + 0.3\cos(3\pi) \\ &= 0.1 \times 1 + 0.4 \times 1 + 0.2 \times (-1) + 0.3 \times (-1) = 0. \end{aligned}$$

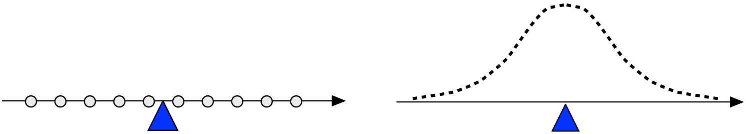
# Center of Mass

The expected value has a *center of mass* interpretation:




$$\sum_n p_n(a_n-\mu)=0$$
$$\Leftrightarrow \mu = \sum_n a_n p_n = E[X]$$



# Monotonicity

**Definition**
Let $X, Y$ be two random variables on $\Omega$. We write $X \leq Y$ if $X(\omega) \leq Y(\omega)$ for all $\omega \in \Omega$, and similarly for $X \geq Y$ and $X \geq a$ for some constant $a$.

**Facts**
(a) If $X \geq 0$, then $E[X] \geq 0$.
(b) If $X \leq Y$, then $E[X] \leq E[Y]$.

**Proof**
(a) If $X \geq 0$, every value $a$ of $X$ is nonnegative. Hence,

$$E[X] = \sum_a aPr[X = a] \geq 0.$$

(b) $X \leq Y \Rightarrow Y - X \geq 0 \Rightarrow E[Y] - E[X] = E[Y - X] \geq 0$.

$\square$

Example:

$$B = \cup_m A_m \Rightarrow 1_B(\omega) \leq \sum_m 1_{A_m}(\omega) \Rightarrow Pr[\cup_m A_m] \leq \sum_m Pr[A_m].$$

# Uniform Distribution

Roll a six-sided balanced die. Let $X$ be the number of pips (dots). Then $X$ is equally likely to take any of the values $\{1,2,\ldots,6\}$. We say that $X$ is *uniformly distributed* in $\{1,2,\ldots,6\}$.

More generally, we say that $X$ is uniformly distributed in $\{1,2,\ldots,n\}$ if $Pr[X=m]=1/n$ for $m=1,2,\ldots,n$.
In that case,

$$E[X]=\sum_{m=1}^{n} mPr[X=m]=\sum_{m=1}^{n} m\times\frac{1}{n}=\frac{1}{n}\frac{n(n+1)}{2}=\frac{n+1}{2}.$$

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

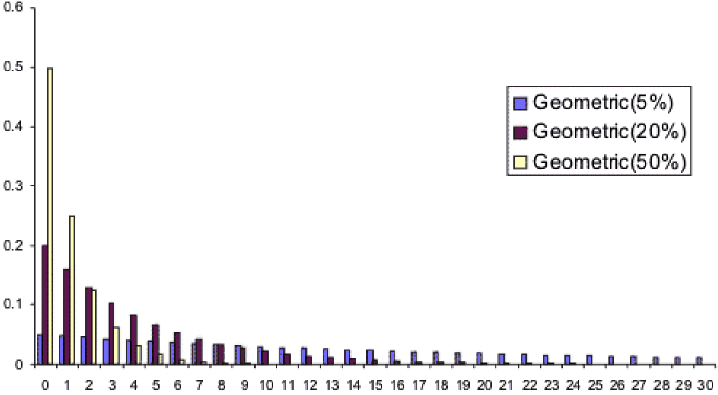
Let $X$ be the number of flips until the first $H$. Then, $X(\omega_n) = n$.

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,

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

Hence,

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

# Geometric Distribution: 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**

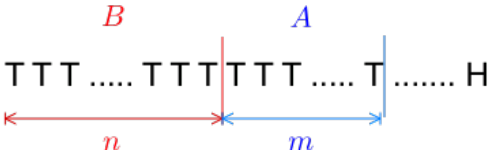
$$Pr[X > n+m | X > n] = Pr[X > m], m, n \geq 0.$$

**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

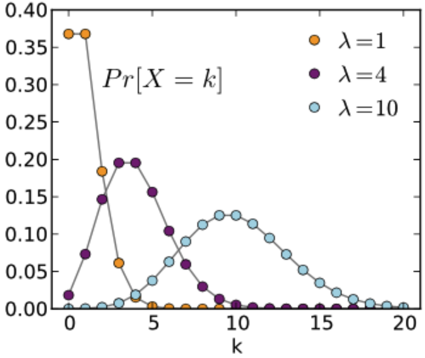
$$\begin{aligned}
E[X] &= \sum_{i=1}^{\infty} i\times Pr[X=i] \\
&= \sum_{i=1}^{\infty} i\{Pr[X\geq i]-Pr[X\geq i+1]\} \\
&= \sum_{i=1}^{\infty} \{i\times Pr[X\geq i]-i\times Pr[X\geq i+1]\} \\
&= \sum_{i=1}^{\infty} \{i\times Pr[X\geq i]-(i-1)\times Pr[X\geq i]\} \\
&= \sum_{i=1}^{\infty} Pr[X\geq i].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

$\square$

# Poisson

Experiment: flip a coin $n$ times. The coin is such that $Pr[H] = \lambda/n$.
Random Variable: $X$ - number of heads. Thus, $X = B(n, \lambda/n)$.

**Poisson Distribution** is distribution of $X$ "for large $n$."



# Poisson

Experiment: flip a coin $n$ times. The coin is such that $Pr[H] = \lambda/n$.
Random Variable: $X$ - number of heads. Thus, $X = B(n, \lambda/n)$.
**Poisson Distribution** is distribution of $X$ "for large $n$."
We expect $X \ll n$. For $m \ll n$ one has

$$\begin{aligned}
Pr[X = m] &= \binom{n}{m} p^m (1-p)^{n-m}, \text{ with } p = \lambda/n \\
&= \frac{n(n-1)\cdots(n-m+1)}{m!} \left(\frac{\lambda}{n}\right)^m \left(1 - \frac{\lambda}{n}\right)^{n-m} \\
&= \frac{n(n-1)\cdots(n-m+1)}{n^m} \frac{\lambda^m}{m!} \left(1 - \frac{\lambda}{n}\right)^{n-m} \\
&\approx^{(1)} \frac{\lambda^m}{m!} \left(1 - \frac{\lambda}{n}\right)^{n-m} \approx^{(2)} \frac{\lambda^m}{m!} \left(1 - \frac{\lambda}{n}\right)^{n} \approx \frac{\lambda^m}{m!} e^{-\lambda}.
\end{aligned}$$

For (1) we used $m \ll n$; for (2) we used $(1 - a/n)^n \approx e^{-a}$.

# 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:**

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

$\square$

# Simeon Poisson

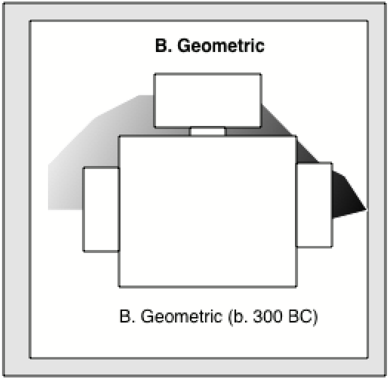
The Poisson distribution is named after:

Siméon Poisson

Siméon Denis Poisson (1781–1840)

# Equal Time: B. Geometric

The geometric distribution is named after:




I could not find a picture of D. Binomial, sorry.

# Summary

$\boxed{\text{Expectation; Geometric \& Poisson}}$

- $E[X] := \sum_a aPr[X = a]$.
- Expectation is Linear.
- $B(n,p), U[1:n], G(p), P(\lambda)$.