

# CS70 @ UC Berkeley, Spring 2026

## Lecture 22

### Geometric and Poisson Distributions

April 14, 2026

# Covariance

**Definition (Covariance):** The covariance of random variables  $X$  and  $Y$  is defined as

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

## Bilinearity of Covariance

Let  $X_1, \dots, X_n$  and  $Y_1, \dots, Y_m$  be random variables on the same probability space, and let  $a_1, \dots, a_n$  and  $b_1, \dots, b_m$  be arbitrary constants. Then,

$$\text{Cov}\left(\sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{Cov}(X_i, Y_j).$$

# Covariance

**Definition (Covariance):** The covariance of random variables  $X$  and  $Y$  is defined as

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

## Interpretation of the sign of $\text{Cov}(X, Y)$

- ▶ If  $\text{Cov}(X, Y) > 0$ :
  - ▶  $X > \mathbb{E}[X]$  and  $Y > \mathbb{E}[Y]$  tend to co-occur, and/or
  - ▶  $X < \mathbb{E}[X]$  and  $Y < \mathbb{E}[Y]$  tend to co-occur.
- ▶ If  $\text{Cov}(X, Y) < 0$ :
  - ▶  $X > \mathbb{E}[X]$  and  $Y < \mathbb{E}[Y]$  tend to co-occur, and/or
  - ▶  $X < \mathbb{E}[X]$  and  $Y > \mathbb{E}[Y]$  tend to co-occur.
- ▶ If  $\text{Cov}(X, Y) = 0$ :
  - ▶ No such association.

The **magnitude** of  $\text{Cov}(X, Y)$  is more difficult to interpret.

# Correlation

**Definition (Pearson's correlation):** Let  $X$  and  $Y$  be random variables on the same probability space with positive standard deviations (i.e.,  $\sigma(X) > 0$  and  $\sigma(Y) > 0$ ). Then, the correlation between  $X$  and  $Y$  is defined as

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$

**Theorem:** For any pair of random variables  $X$  and  $Y$  on the same probability space with  $\sigma(X) > 0$  and  $\sigma(Y) > 0$ ,

$$-1 \leq \text{Corr}(X, Y) \leq 1.$$

See Note 17 for proof.

# Example (Bivariate Bernoulli)

For  $i = 1, \dots, n$ , let  $X_i, Y_i$  be **Bernoulli** random variables such that

$$\mathbb{P}[X_i = 1] = p, \quad \mathbb{P}[X_i = 0] = 1 - p$$

$$\mathbb{P}[Y_i = 1] = q, \quad \mathbb{P}[Y_i = 0] = 1 - q$$

$$\mathbb{P}[(X_i, Y_i) = (1, 1)] = r$$

$$\mathbb{P}[(X_i, Y_i) = (1, 0)] = p - r$$

$$\mathbb{P}[(X_i, Y_i) = (0, 1)] = q - r$$

$$\mathbb{P}[(X_i, Y_i) = (0, 0)] = 1 + r - p - q$$

- ▶  $\text{Var}[X_i] = p(1 - p)$  for all  $i$ .
- ▶  $\text{Var}[Y_i] = q(1 - q)$  for all  $i$ .
- ▶  $\text{Cov}(X_i, Y_i) = \mathbb{E}[X_i Y_i] - \mathbb{E}[X_i]\mathbb{E}[Y_i] = r - pq$
- ▶  $\text{Corr}(X_i, Y_i) = (r - pq) / \sqrt{p(1 - p)q(1 - q)}$
- ▶ In this example,  $X_i \perp\!\!\!\perp Y_i \iff \text{Corr}(X_i, Y_i) = 0$ .
- ▶ In general,  $X_i \perp\!\!\!\perp Y_i \implies \text{Corr}(X_i, Y_i) = 0$ , but **the converse is not true**.

# Example (Bivariate Bernoulli)

For  $i = 1, \dots, n$ , let  $X_i, Y_i$  be **Bernoulli** random variables such that

$$\mathbb{P}[X_i = 1] = p, \quad \mathbb{P}[X_i = 0] = 1 - p$$

$$\mathbb{P}[Y_i = 1] = q, \quad \mathbb{P}[Y_i = 0] = 1 - q$$

$$\mathbb{P}[(X_i, Y_i) = (1, 1)] = r$$

$$\mathbb{P}[(X_i, Y_i) = (1, 0)] = p - r$$

$$\mathbb{P}[(X_i, Y_i) = (0, 1)] = q - r$$

$$\mathbb{P}[(X_i, Y_i) = (0, 0)] = 1 + r - p - q$$

► Suppose  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$  are **i.i.d. draws** from the distribution described above:

- $X_i \perp\!\!\!\perp Y_i$  iff  $r = pq$ .
- $X_i \perp\!\!\!\perp X_j$  for all  $i \neq j$ .
- $Y_i \perp\!\!\!\perp Y_j$  for all  $i \neq j$ .
- $X_i \perp\!\!\!\perp Y_j$  for all  $i \neq j$ .

► Define  $(X, Y) = \sum_{i=1}^n (X_i, Y_i)$ .

How are  $X$  and  $Y$  distributed?

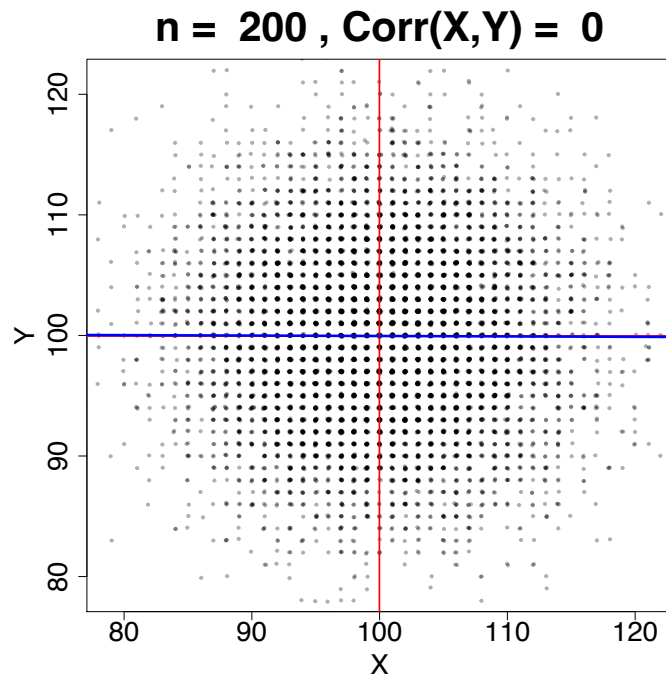
$$X = \sum_{i=1}^n X_i \sim \text{Binomial}(n, p)$$

$$Y = \sum_{i=1}^n Y_i \sim \text{Binomial}(n, q)$$

# Example (Bivariate Bernoulli)

Red lines: marginal means

Blue lines: least-square fit



$$(X, Y) = \sum_{i=1}^n (X_i, Y_i)$$

where  $(X_1, Y_1), \dots, (X_n, Y_n)$

are i.i.d. draws from:

$$\mathbb{P}[(X_i, Y_i) = (1, 1)] = 0.25$$

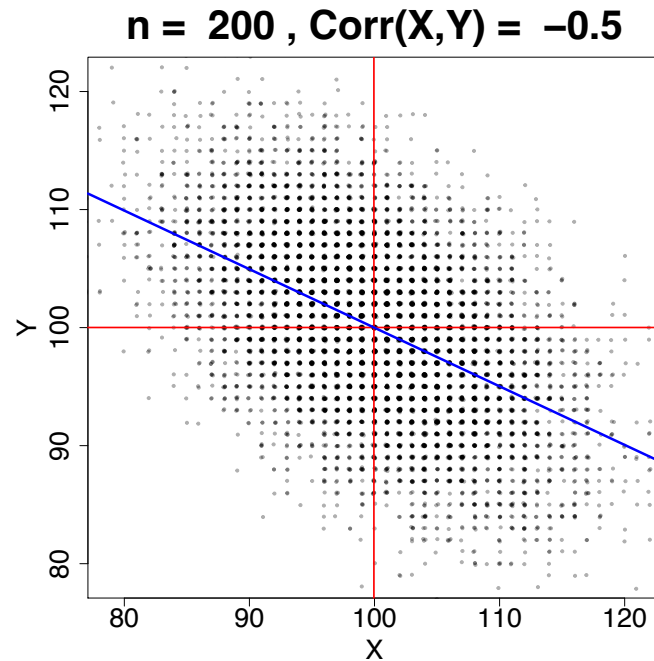
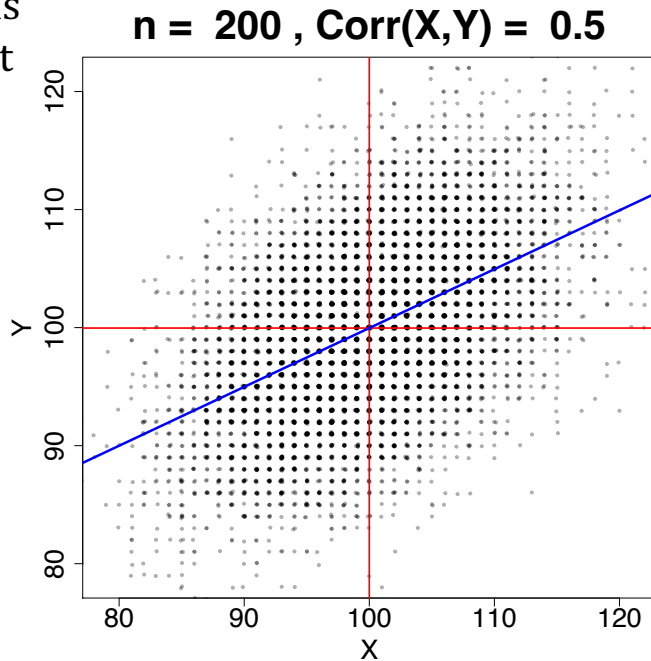
$$\mathbb{P}[(X_i, Y_i) = (1, 0)] = 0.25$$

$$\mathbb{P}[(X_i, Y_i) = (0, 1)] = 0.25$$

$$\mathbb{P}[(X_i, Y_i) = (0, 0)] = 0.25$$

# Example (Bivariate Bernoulli)

Red lines: marginal means  
Blue lines: least-square fit



$$(X, Y) = \sum_{i=1}^n (X_i, Y_i)$$

where  $(X_1, Y_1), \dots, (X_n, Y_n)$   
are i.i.d. draws from:

$$\mathbb{P}[(X_i, Y_i) = (1, 1)] = 0.375$$

$$\mathbb{P}[(X_i, Y_i) = (1, 0)] = 0.125$$

$$\mathbb{P}[(X_i, Y_i) = (0, 1)] = 0.125$$

$$\mathbb{P}[(X_i, Y_i) = (0, 0)] = 0.375$$

$$\mathbb{P}[(X_i, Y_i) = (1, 1)] = 0.125$$

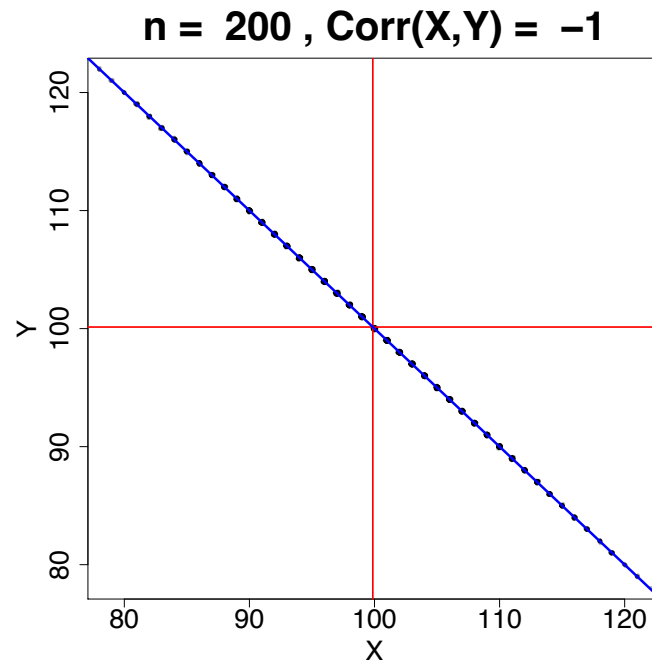
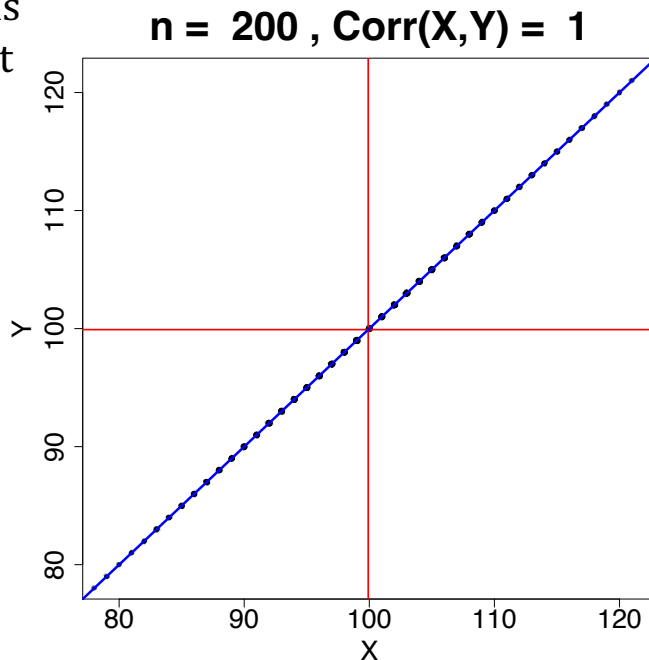
$$\mathbb{P}[(X_i, Y_i) = (1, 0)] = 0.375$$

$$\mathbb{P}[(X_i, Y_i) = (0, 1)] = 0.375$$

$$\mathbb{P}[(X_i, Y_i) = (0, 0)] = 0.125$$

# Example (Bivariate Bernoulli)

Red lines: marginal means  
Blue lines: least-square fit



$(X, Y) = \sum_{i=1}^n (X_i, Y_i)$   
where  $(X_1, Y_1), \dots, (X_n, Y_n)$   
are i.i.d. draws from:

$$\mathbb{P}[(X_i, Y_i) = (1, 1)] = 0.5$$

$$\mathbb{P}[(X_i, Y_i) = (1, 0)] = 0.0$$

$$\mathbb{P}[(X_i, Y_i) = (0, 1)] = 0.0$$

$$\mathbb{P}[(X_i, Y_i) = (0, 0)] = 0.5$$

$$\mathbb{P}[(X_i, Y_i) = (1, 1)] = 0.0$$

$$\mathbb{P}[(X_i, Y_i) = (1, 0)] = 0.5$$

$$\mathbb{P}[(X_i, Y_i) = (0, 1)] = 0.5$$

$$\mathbb{P}[(X_i, Y_i) = (0, 0)] = 0.0$$

# Tail Sum Formula

**Theorem (Tail Sum Formula):** Let  $X$  be a random variable that takes values in  $0, 1, \dots, n$ . Then,

$$\mathbb{E}[X] = \sum_{a=1}^n \mathbb{P}(X \geq a)$$

Why is this useful?

← Important

**Proof:**  $\mathbb{E}[X] = \sum_{a=0}^n a \mathbb{P}(X = a)$

$$\begin{aligned} &= \mathbb{P}(X = 1) \\ &+ \mathbb{P}(X = 2) + \mathbb{P}(X = 2) \\ &+ \mathbb{P}(X = 3) + \mathbb{P}(X = 3) + \mathbb{P}(X = 3) \\ &\vdots \\ &+ \mathbb{P}(X = n) + \mathbb{P}(X = n) + \mathbb{P}(X = n) + \dots + \mathbb{P}(X = n) \\ &= \mathbb{P}(X \geq 1) + \mathbb{P}(X \geq 2) + \mathbb{P}(X \geq 3) + \dots + \mathbb{P}(X \geq n) = \sum_{a=1}^n \mathbb{P}(X \geq a). \end{aligned}$$

**Remark:** For a random variable  $X$  with range  $\mathbb{N}$ ,  $\mathbb{E}[X] = \sum_{a=1}^{\infty} \mathbb{P}(X \geq a)$ .

# Tail Sum Formula (Application)

**Theorem (Tail Sum Formula):** Let  $X$  be a random variable that takes values in  $0, 1, \dots, n$ . Then,

$$\mathbb{E}[X] = \sum_{a=1}^n \mathbb{P}(X \geq a) \quad \leftarrow \text{Important}$$

- Consider a deck of 100 cards numbered  $1, \dots, 100$ .
- Shuffle it well and take a card from the top.
- Repeat this experiment  $n$  times and let  $M = \text{minimum number observed}$ .
- What is  $\mathbb{E}[M]$ ?

Let  $X_i = \text{number observed on the } i\text{th roll}$  and assume independence. Then,

$$\begin{aligned} \mathbb{P}(M \geq a) &= \mathbb{P}(X_1 \geq a, \dots, X_n \geq a) = \mathbb{P}(X_1 \geq a) \cdots \mathbb{P}(X_n \geq a) \quad \text{By independence} \\ &= [\mathbb{P}(X_1 \geq a)]^n = \left[ \frac{100 - (a-1)}{100} \right]^n \end{aligned}$$

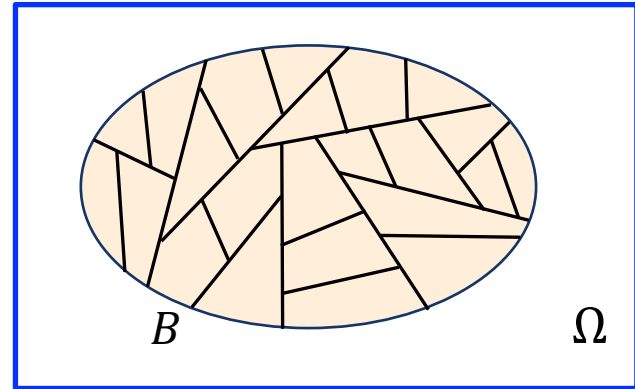
Using the Tail Sum Formula, we obtain  $\mathbb{E}[X] = \sum_{a=1}^n \left[ \frac{100 - (a-1)}{100} \right]^n$

# Rules of Probability (Lecture 17)

**Definition (Partition):** An event  $B \subseteq \Omega$  is said to be partitioned into **infinitely many** events

$B_1, B_2, B_3, \dots$  if

1.  $B = \bigcup_{k=1}^{\infty} B_k$ ,
2.  $B_i \cap B_j = \emptyset$ , for all  $i \neq j$  (that is,  $B_1, \dots, B_n$  are **mutually exclusive**).



1. **(Non-negativity)**  $\mathbb{P}(A) \geq 0$ , for all  $A \subseteq \Omega$ .
2. **(Countable Additivity)** If  $B_1, B_2, B_3, \dots$  is a partition of  $B$ , then


$$\mathbb{P}(B) = \sum_{k=1}^{\infty} \mathbb{P}(B_k)$$

3. **(Normalization)**  $\mathbb{P}(\Omega) = 1$ .

# Geometric Distribution

i.i.d. Bernoulli( $p$ ) trials, where  $p$  = success probability

$X$  = Waiting time to the first success



Trials	1	2	3	4	5	6
$\omega \in \Omega$	$F$	$F$	$F$	$F$	$F$	$S$
Prob	$1 - p$	$1 - p$	$1 - p$	$1 - p$	$1 - p$	$p$

$X(\omega) = 6$  for this outcome

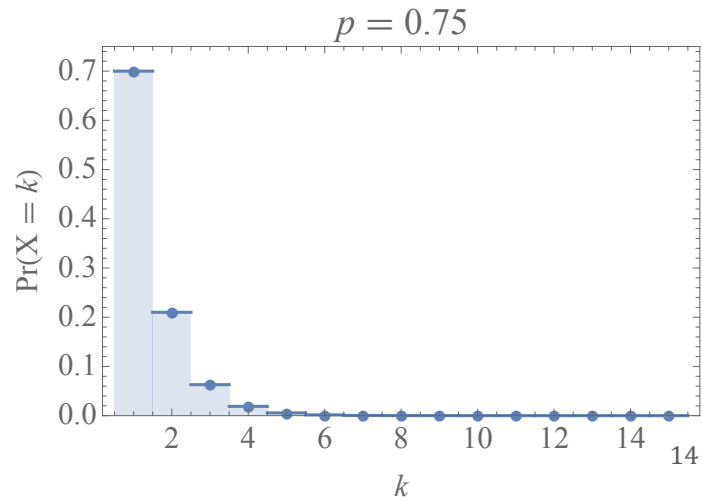
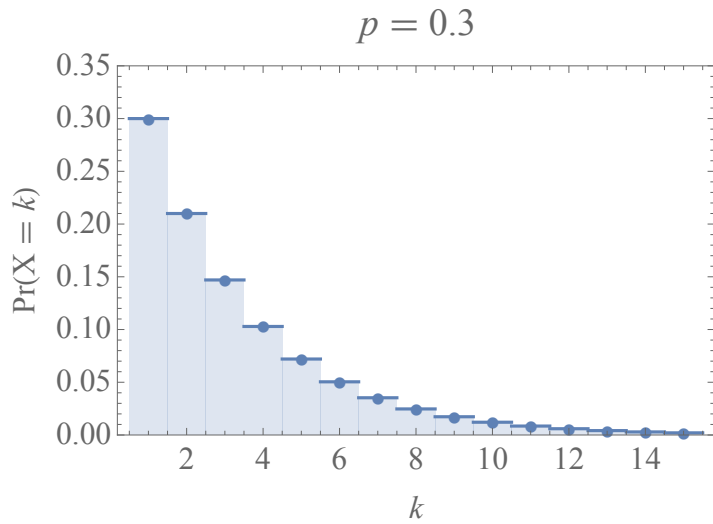
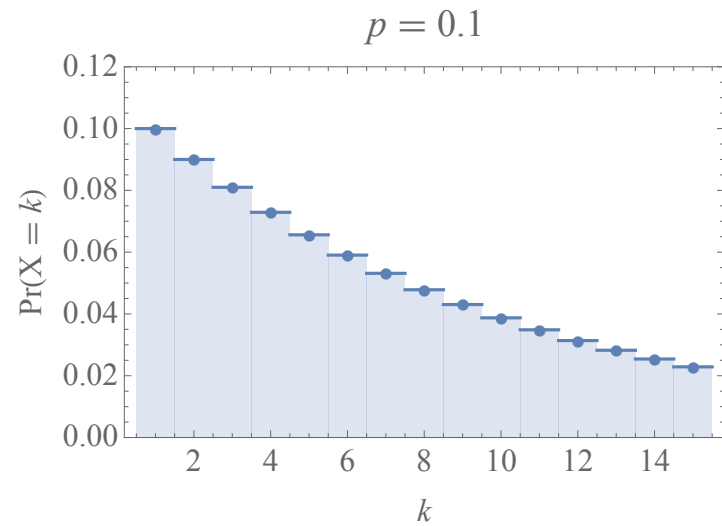
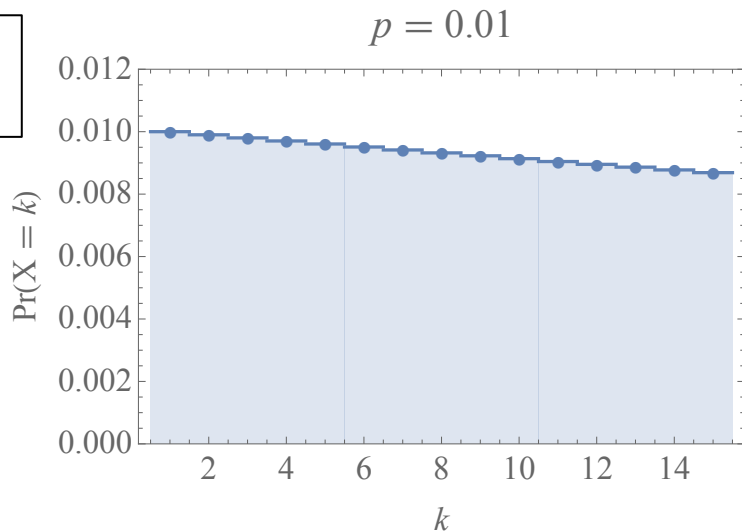
$$\mathbb{P}(X = k) = (1 - p)^{k-1}p, \text{ for } k \in \mathbb{Z}_+$$

$$\mathbb{P}(X > k) = (1 - p)^k$$

We write  $X \sim \text{Geometric}(p)$ ,  $0 < p < 1$ .

As  $p$  decreases,  $\mathbb{E}[X]$  should increase.

$X \sim \text{Geometric}(p)$ ,  
 $0 < p < 1$ .



# Memoryless Property

For  $m, n \in \mathbb{N}$ ,

For  $\omega \in \Omega$ ,  $(X(\omega) > n + m) \Rightarrow (X(\omega) > m)$ ,  
so  $(X > n + m) \subset (X > m)$  and  
 $(X > n + m) \cap (X > m) = (X > n + m)$ .

$$\begin{aligned}\mathbb{P}[X > n + m \mid X > m] &= \frac{\mathbb{P}[X > n + m]}{\mathbb{P}[X > m]} \\ &= \frac{(1 - p)^{n+m}}{(1 - p)^m} \\ &= (1 - p)^n \\ &= \mathbb{P}[X > n].\end{aligned}$$

# Moments of Geometric Random Variable

In Note 18, the Tail Sum Formula is used to compute  $\mathbb{E}[X]$ . We will present an alternate approach here:

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k=1}^{\infty} k \mathbb{P}(X = k) = \sum_{k=1}^{\infty} k (1-p)^{k-1} p \\ &= -p \frac{d}{dp} \sum_{k=1}^{\infty} (1-p)^k \\ &= -p \frac{d}{dp} \left[ \frac{1}{1-(1-p)} - 1 \right] \\ &= -p \frac{d}{dp} \left[ \frac{1}{p} - 1 \right] \\ &= -p \left[ \frac{1}{p^2} \right] = \frac{1}{p}.\end{aligned}$$

**Geometric Series:**

$$\sum_{k=0}^{\infty} a^k = \frac{1}{1-a} \text{ if } |a| < 1.$$

$$\begin{aligned}\mathbb{E}[X^2] &= \sum_{k=1}^{\infty} k^2 \mathbb{P}(X = k) \\ &= \sum_{k=1}^{\infty} k^2 (1-p)^{k-1} p\end{aligned}$$

$$\frac{d}{dp} \left[ (1-p) \frac{d}{dp} (1-p)^k \right] = k^2 (1-p)^{k-1}$$

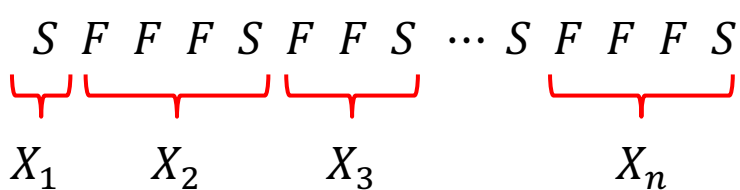
$$\mathbb{E}[X^2] = \frac{d}{dp} \left[ (1-p) \left( -\frac{1}{p^2} \right) \right] = \frac{2-p}{p^2}$$

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{1-p}{p^2}$$

# Coupon Collection (Lecture 19)

- $n$  distinct coupons, **1** coupon per cereal box
- All coupon types are equally likely to be in any given cereal box
- Assume infinitely many cereal boxes so independence of sampling holds
- **Collect all  $n$  distinct coupons to win**

$S$  = found a coupon not obtained before



- $X_1, \dots, X_n$  independent
- $X_k \sim \text{Geometric}(p_k)$ , where  $p_k = \frac{n-(k-1)}{n}$

Already got  $k - 1$   
distinct coupons

$$\mathbb{E}[X_k] = \frac{1}{p_k}$$

$$\begin{aligned} \mathbb{E}[X_1 + \dots + X_n] &= \sum_{k=1}^n \frac{1}{p_k} = \sum_{k=1}^n \frac{n}{n - (k - 1)} \\ &= n \sum_{k=1}^n \frac{1}{k} = n \left[ \ln n + \gamma_E + \frac{1}{2n} \right] \end{aligned}$$

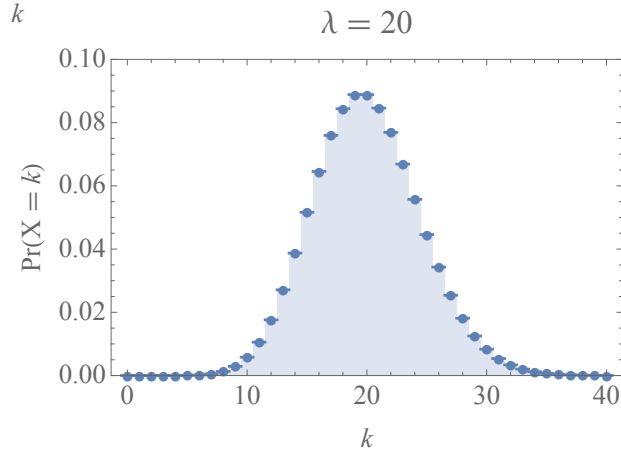
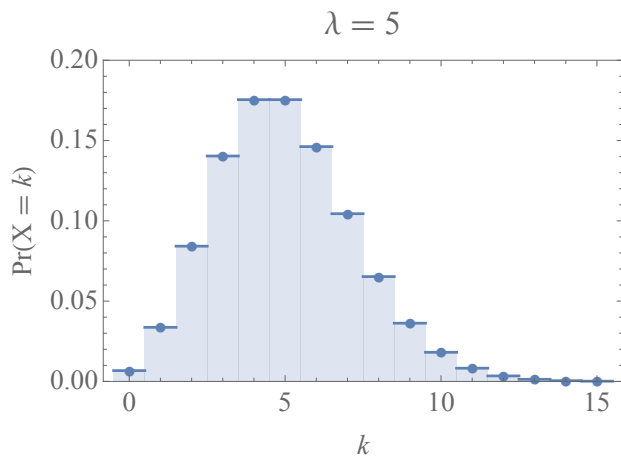
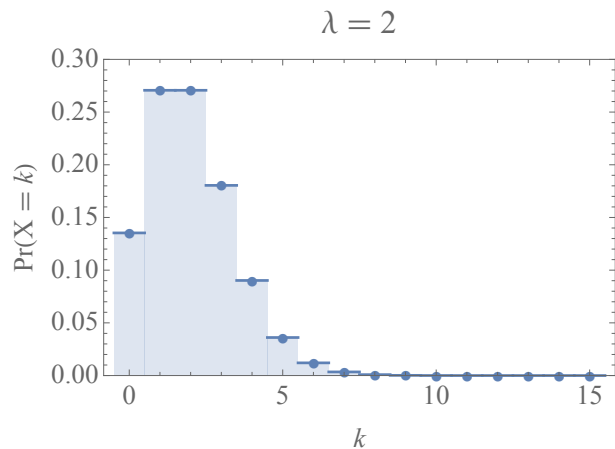
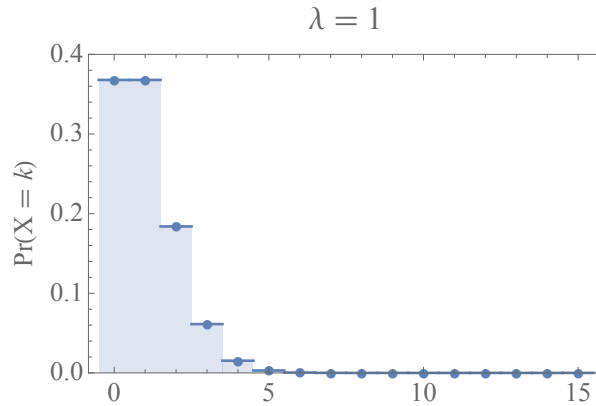
Euler's constant 0.57721

# Poisson Distribution

$N \sim \text{Poisson}(\lambda)$ , where intensity  $\lambda > 0$ .

$$\mathbb{P}(N = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \text{ for } k \in \mathbb{N}.$$

- # rain drops hitting a surface per second
- # radioactive particles emitted by radioactive material during an interval of time



Looks more and more like a “bell curve” as  $\lambda$  gets large. We will see why this happens.

# Poisson Approximation of Binomial( $n, p$ )

- (\*) Limit as  $n \rightarrow \infty$  and  $p \rightarrow 0$ , while  $np = \lambda$  is held fixed.
- Define  $Q_k = \binom{n}{k} p^k (1-p)^{n-k}$
- $Q_0 = (1-p)^n = \left(1 - \frac{\lambda}{n}\right)^n \rightarrow e^{-\lambda}$  in the limit (\*).
- $\frac{Q_k}{Q_{k-1}} = \frac{\binom{n}{k} p^k (1-p)^{n-k}}{\binom{n}{k-1} p^{k-1} (1-p)^{n-k+1}} = \frac{n-k+1}{k} \left(\frac{p}{1-p}\right) = \frac{np}{k} \left[\frac{1 - \frac{k-1}{n}}{1-p}\right] \rightarrow \frac{\lambda}{k}$  in the limit (\*).
- $Q_k = Q_0 \frac{Q_1}{Q_0} \frac{Q_2}{Q_1} \dots \frac{Q_k}{Q_{k-1}} \rightarrow e^{-\lambda} \frac{\lambda}{1} \frac{\lambda}{2} \dots \frac{\lambda}{k} = e^{-\lambda} \frac{\lambda^k}{k!}$  in the limit (\*).

$X \sim \text{Binomial}(n, p)$

$\mathbb{E}[X] = np \rightarrow \lambda$  in the limit (\*).

$\text{Var}(X) = np(1-p) \rightarrow \lambda$  in the limit (\*).

# Independent Poisson Random Variables

**Theorem:** Suppose  $X \sim \text{Poisson}(\lambda)$  and  $Y \sim \text{Poisson}(\mu)$  are independent random variables. Then,  $X + Y \sim \text{Poisson}(\lambda + \mu)$ .

**Proof:** 
$$\begin{aligned}\mathbb{P}(X + Y = n) &= \sum_{k=0}^n \mathbb{P}(X = k, Y = n - k) \\ &= \sum_{k=0}^n \mathbb{P}(X = k)\mathbb{P}(Y = n - k) \\ &= \sum_{k=0}^n \frac{e^{-\lambda}\lambda^k}{k!} \cdot \frac{e^{-\mu}\mu^{n-k}}{(n-k)!} \\ &= \frac{e^{-(\lambda+\mu)}}{n!} \sum_{k=0}^n \frac{n!}{k!(n-k)!} \lambda^k \mu^{n-k}\end{aligned}$$

By additivity.

$(X = 0) \cap (Y = n), \dots, (X = n) \cap (Y = 0)$   
partition  $X + Y = n$

By Independence

Binomial Theorem (Lecture 15)

$$\downarrow \frac{e^{-(\lambda+\mu)}(\lambda + \mu)^n}{n!}$$

□

**Corollary:** Suppose  $X_1, \dots, X_n$  are independent RVs with  $X_i \sim \text{Poisson}(\lambda_i)$ , for  $i = 1, \dots, n$ . Then,  $X_1 + \dots + X_n \sim \text{Poisson}(\lambda_1 + \dots + \lambda_n)$ .