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

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

Bilinearity of Covariance

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

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

Covariance

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

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

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

- If $\mathrm{Cov}(X,Y) > 0$:
  - $X > \mathbb{E}[X]$ 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 $\mathrm{Cov}(X,Y) < 0$:
  - $X > \mathbb{E}[X]$ 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 $\mathrm{Cov}(X,Y) = 0$:
  - No such association.

The magnitude of 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

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

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

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

See Note 19 for proof.

Example (Bivariate Bernoulli)

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

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

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

Example (Bivariate Bernoulli)

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

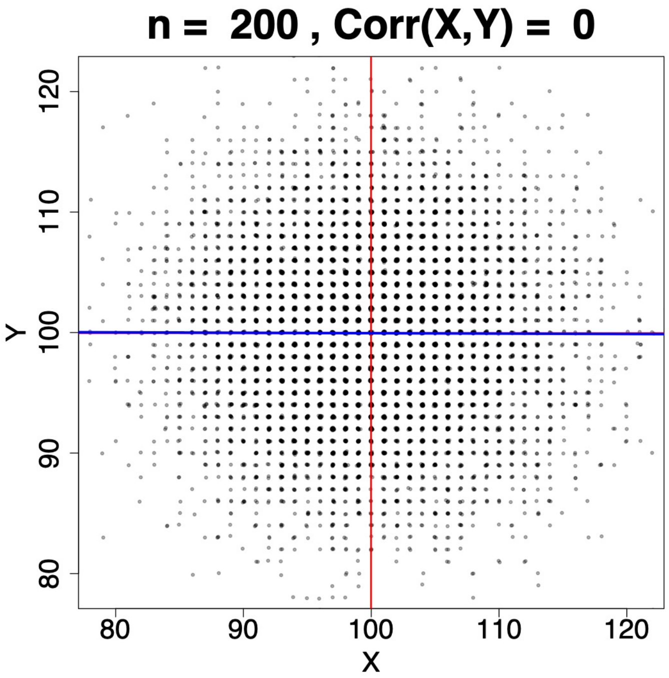
$$\mathbb{P}[X_i = 1] = p, \quad \mathbb{P}[X_i = 0] = 1 - p$$
$$\mathbb{P}[Y_i = 1] = q, \quad \mathbb{P}[Y_i = 0] = 1 - q$$
$$\mathbb{P}[(X_i, Y_i) = (1, 1)] = r$$
$$\mathbb{P}[(X_i, Y_i) = (1, 0)] = p - r$$
$$\mathbb{P}[(X_i, Y_i) = (0, 1)] = q - r$$
$$\mathbb{P}[(X_i, Y_i) = (0, 0)] = 1 + r - p -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

- Suppose $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ 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?

Example (Bivariate Bernoulli)

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

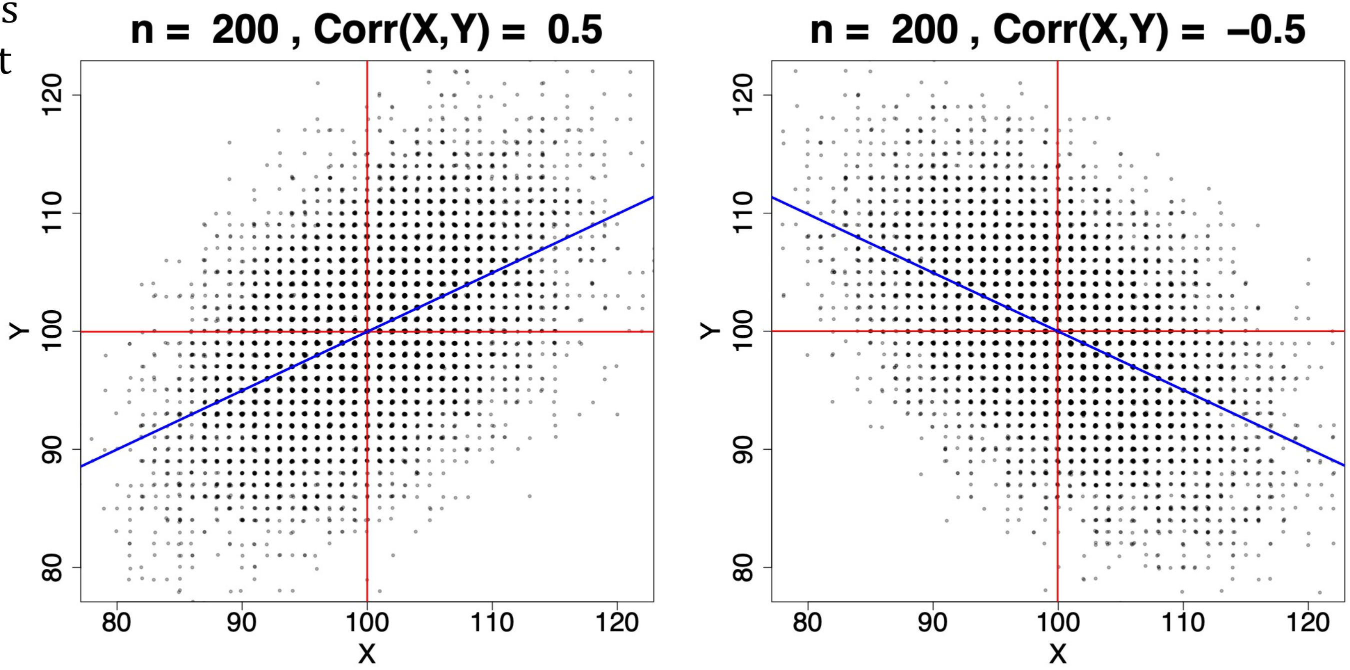


$(X, Y) = \sum_{i=1}^{n} (X_i, Y_i)$
where $(X_1, Y_1), \dots, (X_n, Y_n)$
are i.i.d. draws from:

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

Example (Bivariate Bernoulli)

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



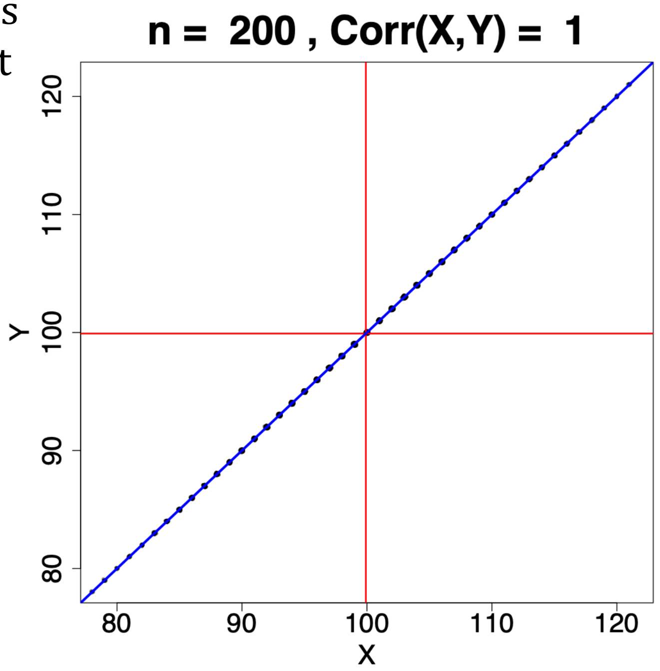
$(X, Y) = \sum_{i=1}^{n} (X_i, Y_i)$
where $(X_1, Y_1), \ldots, (X_n, Y_n)$
are i.i.d. draws from:

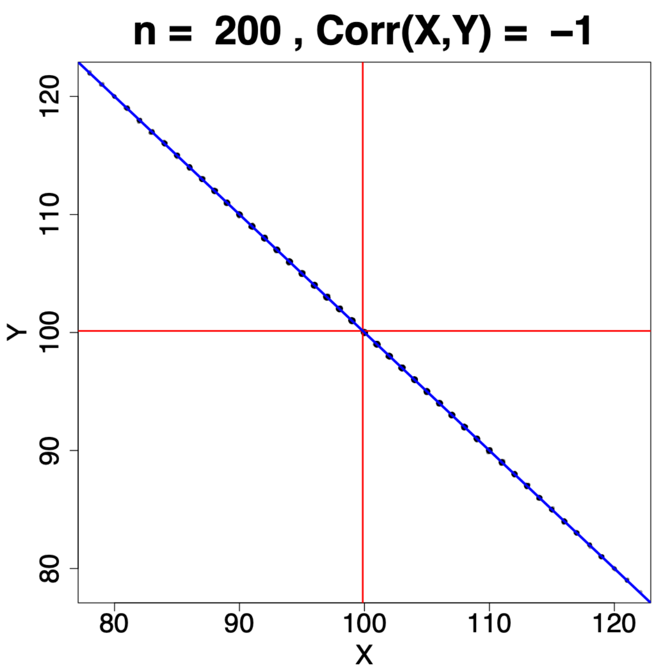
$$\mathbb{P}[(X_i, Y_i) = (1,1)] = 0.375$$
$$\mathbb{P}[(X_i, Y_i) = (1,0)] = 0.125$$
$$\mathbb{P}[(X_i, Y_i) = (0,1)] = 0.125$$
$$\mathbb{P}[(X_i, Y_i) = (0,0)] = 0.375$$

$$\mathbb{P}[(X_i, Y_i) = (1,1)] = 0.125$$
$$\mathbb{P}[(X_i, Y_i) = (1,0)] = 0.375$$
$$\mathbb{P}[(X_i, Y_i) = (0,1)] = 0.375$$
$$\mathbb{P}[(X_i, Y_i) = (0,0)] = 0.125$$

Example (Bivariate Bernoulli)

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





$(X, Y) = \sum_{i=1}^{n} (X_i, Y_i)$
where $(X_1, Y_1), \ldots, (X_n, Y_n)$
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 $0,1,\ldots,n$. Then,

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

Why is this useful?

Important

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

$$\begin{aligned}
&= \mathbb{P}(X=1) \\
&\quad + \mathbb{P}(X=2) + \mathbb{P}(X=2) \\
&\quad + \mathbb{P}(X=3) + \mathbb{P}(X=3) + \mathbb{P}(X=3) \\
&\quad \vdots \\
&\quad + \mathbb{P}(X=n) + \mathbb{P}(X=n) + \mathbb{P}(X=n) + \cdots + \mathbb{P}(X=n)
\end{aligned}$$

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,

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

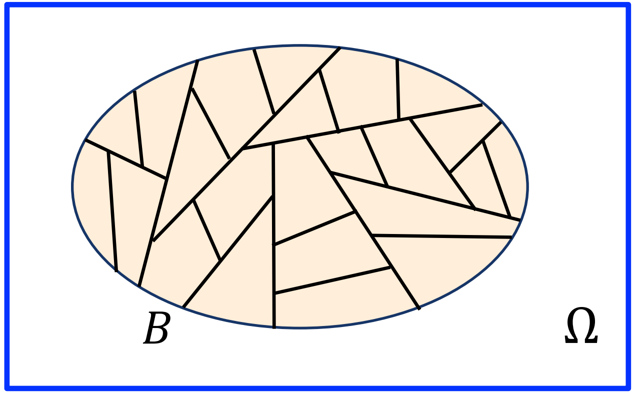
Important

- 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 = minimum number observed.
- What is $\mathbb{E}[M]$?

Rules of Probability (Lecture 17)

Definition (Partition): An event $B \subseteq \Omega$ is said to be partitioned into infinitely many events $B_1, B_2, B_3, \ldots$ if

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



1. **(Non-negativity)** $\mathbb{P}(A) \geq 0$, for all $A \subseteq \Omega$.
2. **(Countable Additivity)** If $B_1, B_2, B_3, \ldots$ 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

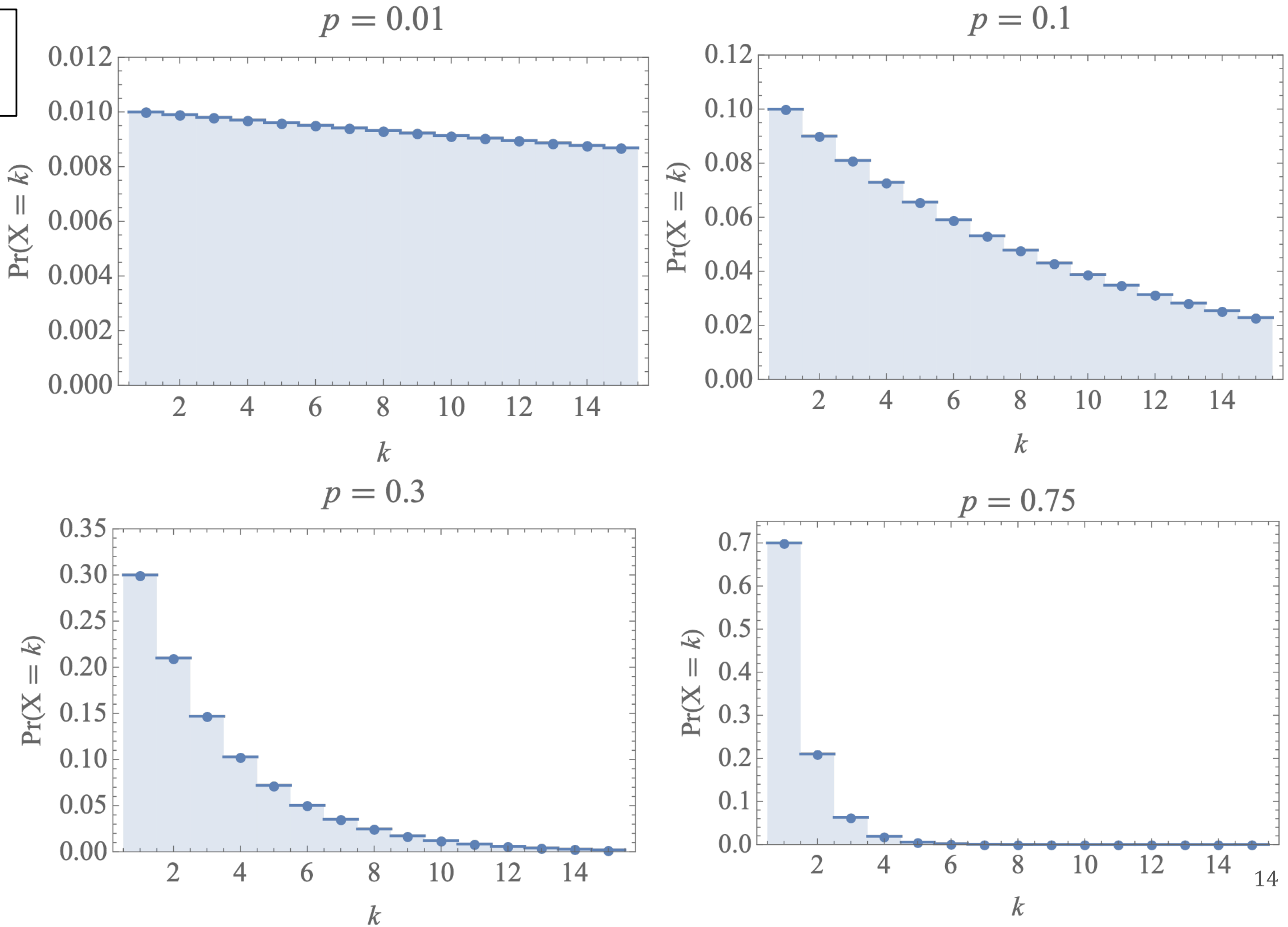
i.i.d. Bernoulli(p) trials, where p = success probability

X = Waiting time to the first success

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

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

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



Memoryless Property

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

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

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

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:

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} k\, \mathbb{P}(X = k) = \sum_{k=1}^{\infty} k\,(1-p)^{k-1}p$$

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

$$\mathbb{E}[X^2] = \sum_{k=1}^{\infty} k^2 \mathbb{P}(X = k)$$

$$= \sum_{k=1}^{\infty} k^2 (1-p)^{k-1} p$$

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

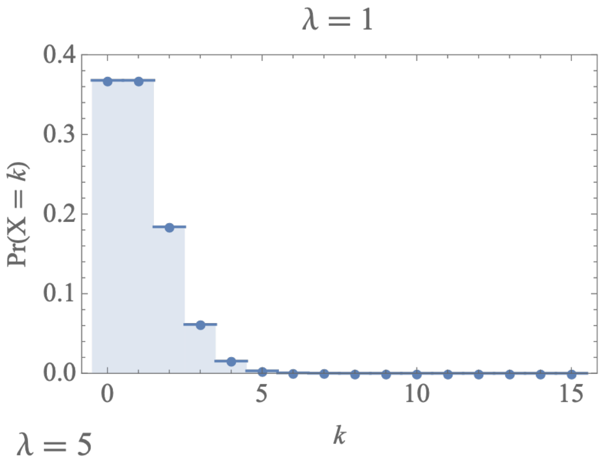
$$\underbrace{S}_{X_1}\ \underbrace{F\ F\ F\ S}_{X_2}\ \underbrace{F\ F\ S}_{X_3}\ \cdots\ S\ \underbrace{F\ F\ F\ S}_{X_n}$$

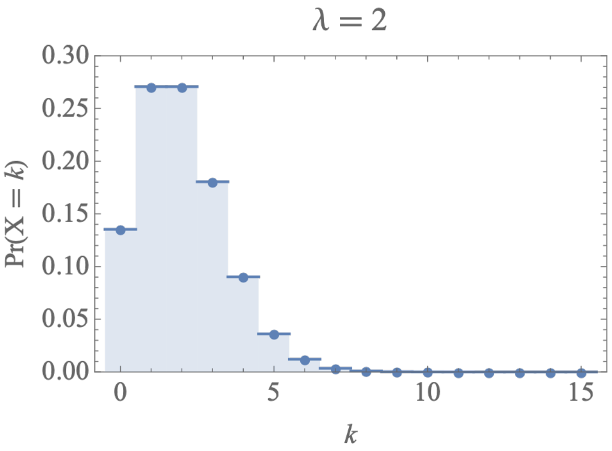
- $X_1, \ldots, X_n$ independent

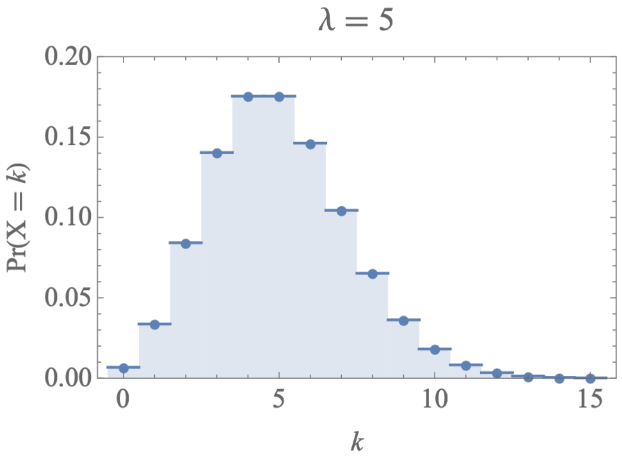
$N \sim \text{Poisson}(\lambda)$, where intensity $\lambda > 0$.

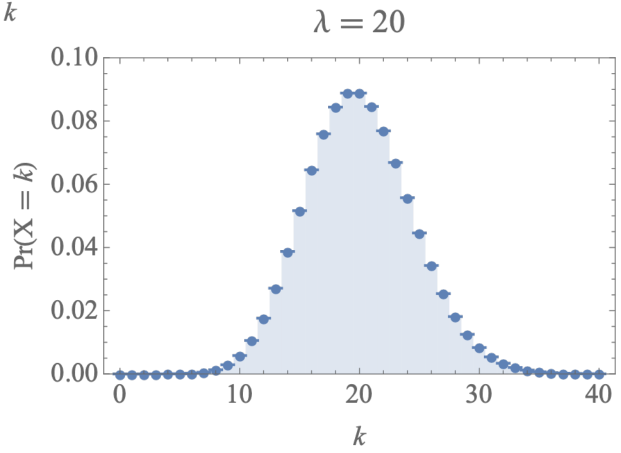
$\mathbb{P}(N = k) = e^{-\lambda}\frac{\lambda^k}{k!}$, for $k \in \mathbb{N}$.

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









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

Poisson Approximation of Binomial(n, p)

- $(*)$ Limit as $n \to \infty$ and $p \to 0$, while $np = \lambda$ is held fixed.
- Define $Q_k = \binom{n}{k} p^k (1-p)^{n-k}$
- $Q_0 = (1-p)^n$
- $\frac{Q_k}{Q_{k-1}} =$
- $Q_k =$

$X \sim \text{Binomial}(n, p)$
$\mathbb{E}[X] = np$
$\text{Var}(X) = np(1-p)$

Independent Poisson Random Variables

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

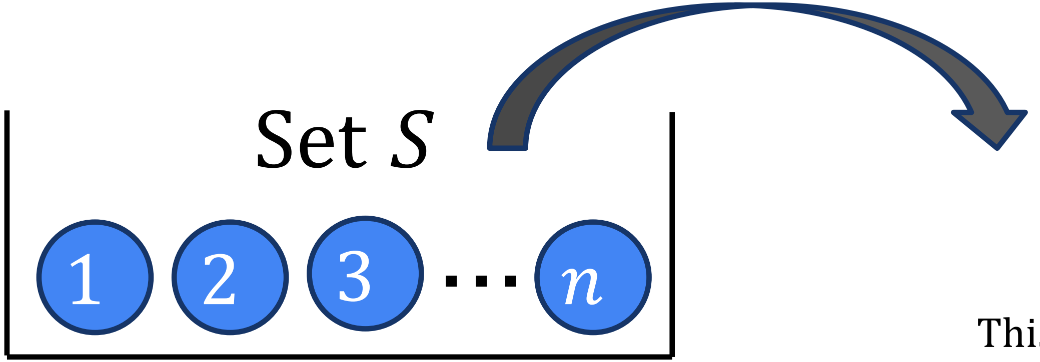
Proof:

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

$\square$

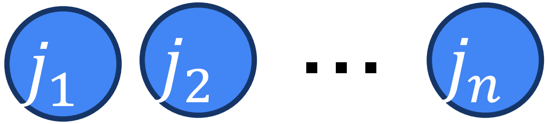
Corollary: Suppose $X_1, \dots, X_n$ are independent RVs with $X_i \sim \text{Poisson}(\lambda_i)$, for $i = 1, \dots, n$. Then, $X_1 + \cdots + X_n \sim \text{Poisson}(\lambda_1 + \cdots + \lambda_n)$.

Permutations and Derangements (Lecture 15)



Sample without replacement

n-element ordered object



This generates a permutation of $\{1,2,\ldots n\}$.

Total number of permutations $=$ $n!$

Iteration i:	1	2	$\boxed{3}$	4	$\boxed{5}$	...	n
Sampled Label π_i:	7	1	$\boxed{3}$	8	$\boxed{5}$	...	2

Fixed points

Bijection $\pi: S \longrightarrow S$

$\pi: i \longmapsto \pi_i$

> **Definition:** A derangement is a permutation with no fixed points.

Let D_n denote the number of derangements of $\{1,2,\ldots,n\}$.

Number of Fixed Points in a Random Permutation

$$D_n = n! \sum_{k=0}^{n} \frac{(-1)^k}{k!}$$

$$\lim_{n\to\infty} \frac{D_n}{n!} = \lim_{n\to\infty} \sum_{k=0}^{n} \frac{(-1)^k}{k!} = e^{-1}$$

- Define $F_{k,n}$ = # permutations of $\{1,2,\ldots,n\}$ with exactly k fixed points.
- $F_{0,n} = D_n$
- $F_{k,n} =$
- $\lim_{n\to\infty} \frac{F_{k,n}}{n!} =$

Proportion of permutations
with k fixed points