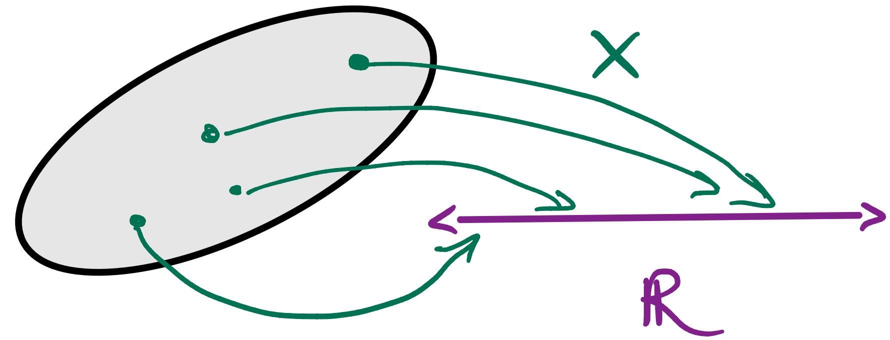
CS70 – Spring 2024

Lecture 19 – March 21

Summary of Last Lecture

- **Random variable** = function $X : \Omega \to \mathbb{R}$



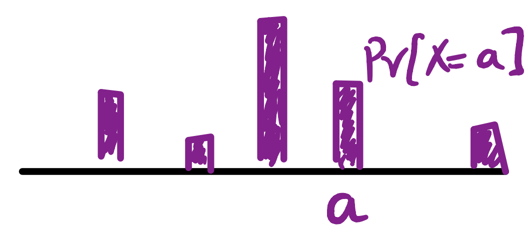
Examples:

Ω = seq. of coin tosses
$X(\omega)$ = # Heads in ω

Ω = two dice rolls
$X(\omega)$ = sum of numbers on the dice

- **Distribution** of a r.v. X:
  $\Pr[X=a]$ for each possible value a of X

Can think of this as a histogram:



$$\sum_a \Pr[X=a] = 1$$

Summary (continued)

- **Expectation** (= mean)

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

  Measures the "center of mass" of the distribution

- **Linearity of expectation**:

  For any r.v.'s X, Y and constants a, b

  $$\boxed{E[aX+bY] = aE[X] + bE[Y]}$$

- Use with indicator r.v.'s to do counting

  E.g. X = no. of fixed points in a random permutation

  $$\boxed{X = \sum_{i=1}^{n} X_i} \quad \text{where } X_i = \begin{cases} 1 & \text{if } i \text{ a fixed point} \\ 0 & \text{otherwise} \end{cases}$$

Summary (continued)

- **Binomial Distribution** Bin(n, p)

  X = # Heads in n tosses of a biased coin (Heads prob. p)

  $$\boxed{\Pr[X=k] = \binom{n}{k} p^k (1-p)^{n-k}} \qquad k = 0, 1, \ldots, n$$

- **Hypergeometric Distribution** HyperGeom(N, n, B)

  X = # black balls in a sample of size n drawn from a box containing N balls, B of which are black

  $$\boxed{\Pr[X=k] = \frac{\binom{B}{k}\binom{N-B}{n-k}}{\binom{N}{n}}}$$

Today

- Joint distributions & independence of random variables
- Two more important distributions:
  - Geometric distribution
  - Poisson distribution

Joint Distributions

<u>Defn</u>: The <u>joint distribution</u> of two r.v.'s X, Y on the same prob. space is the set

$$\boxed{\{(a, b, \Pr[X=a, Y=b]) : a \in \mathcal{A}, b \in \mathcal{B}\}}$$

where $\mathcal{A}, \mathcal{B}$ are the possible values of X, Y resp.

The <u>marginal distribution</u> of X is given by

$$\Pr[X=a] = \sum_{b \in \mathcal{B}} \Pr[X=a, Y=b]$$

X, Y are <u>independent</u> if

$$\Pr[X=a, Y=b] = \Pr[X=a] \times \Pr[Y=b] \quad \forall a, b$$

Joint Distributions

Example: Throw two fair dice

Random variables:

X = score on first die
Y = — " — second — " —
Z = sum of scores

$\Pr[X=3, Y=5] = 1/36$

$\Pr[X=3, Z=9] = 1/36$

X, Y independent?

X, Z independent?

Geometric distribution

Toss a biased coin (Heads prob. p) until you see the first Head

Random variable $X :=$ number of tosses

What is the distribution of X ?

Note : X takes values in $\{1, 2, 3, \ldots\}$

$$\boxed{\begin{aligned} \Pr[X=1] &= p \\ \Pr[X=2] &= (1-p)p \\ \Pr[X=3] &= (1-p)^2 p \\ &\vdots \\ \Pr[X=k] &= (1-p)^{k-1} p \end{aligned}}$$

We say X has the Geometric distribn. with parameter p

$X \sim \mathrm{Geom}(p)$

$$\Pr[X=k] = (1-p)^{k-1}\,p \qquad k = 1, 2, 3, \ldots$$

Check that $\sum_{k=1}^{\infty} \Pr[X=k] = 1$!

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

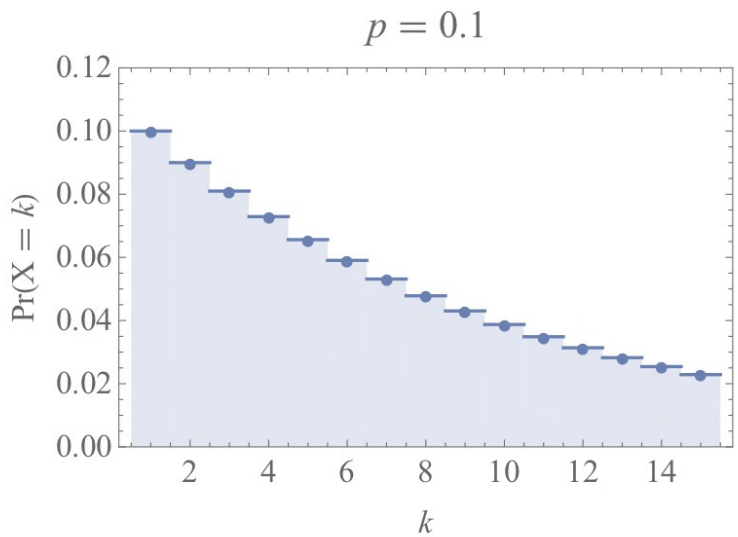
$$= p \sum_{k=0}^{\infty} (1-p)^{k}$$

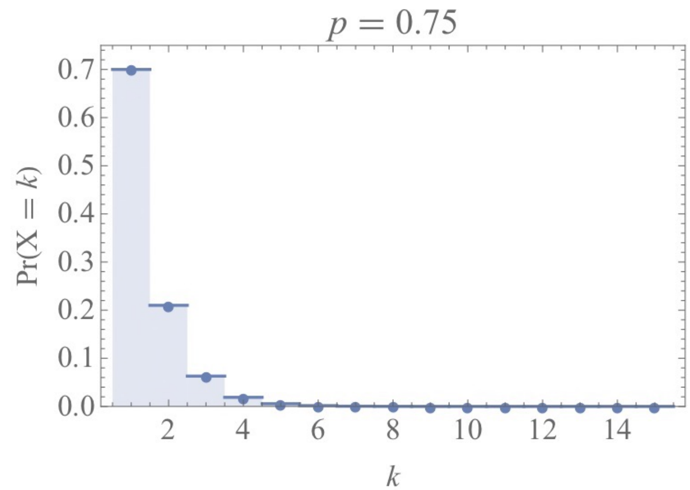
$$= p \times \frac{1}{1-(1-p)}$$

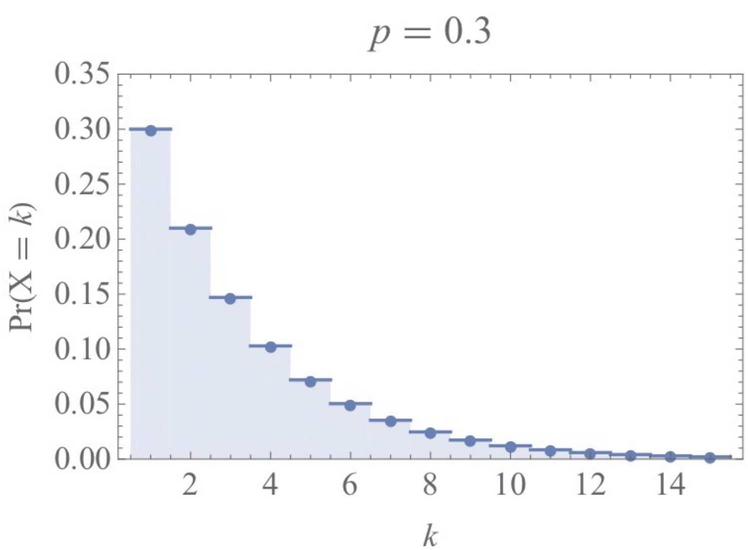
[sum of geometric series]

$$= 1 \quad \checkmark$$

What does the Geometric distribution look like?







Note: Always decreases geometrically (for any p)

Expectation of Geom(p)

Compute $E[X]$ two ways:

(i) Calculus

$$E[X] = \sum_{k=1}^{\infty} k \times \Pr[X=k]$$

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

$$= p \sum_{k=1}^{\infty} k(1-p)^{k-1} \qquad = -\frac{d}{dp}\left(\sum_{k=0}^{\infty}(1-p)^k\right)$$

$$= -\frac{d}{dp}\left(\frac{1}{p}\right)$$

$$= p \times \frac{1}{p^2} \qquad = \frac{1}{p^2}$$

$$= \boxed{\frac{1}{p}}$$

Expectation of Geom(p)

Compute $E[X]$ two ways :

(ii) Tail Sum Formula

Fact: For any r.v. that takes values in $\{0, 1, 2, \ldots\}$ we have

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

Proof: Write $p_i = \Pr[X = i]$ $\quad i = 0, 1, 2, \ldots$

Then $E[X] = (0 \times p_0) + (1 \times p_1) + (2 \times p_2) + (3 \times p_3) + \cdots$

$$= p_1 + (p_2 + p_2) + (p_3 + p_3 + p_3) + \cdots$$

$$= (p_1 + p_2 + p_3 + p_4 + \cdots) + (p_2 + p_3 + p_4 + \cdots) + (p_3 + p_4 + \cdots) + \cdots$$

$$= \Pr[X \geq 1] + \Pr[X \geq 2] + \Pr[X \geq 3] + \cdots$$

Fact: For any r.v. that takes values in $\{0, 1, 2, \ldots\}$ we have

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

Apply to $X \sim \mathrm{Geom}(p)$

Note that $\Pr[X \geq i] = \Pr[\text{first } (i-1) \text{ tosses are Tails}]$

$$= (1-p)^{i-1}$$

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

Bottom line: Expected no. of trials (tosses) until we see first Head is $1/p$

(= 2 for fair coin)

Geometric distribution is $\boxed{\text{Memoryless}}$

Claim: Time until next Head is independent of how long we've been waiting — i.e.

$$\Pr[X > m+k \mid X > m] = \Pr[X > k]$$

Proof: $\forall k$, $\Pr[X > k] = (1-p)^k$

Therefore:

$$\Pr[X > m+k \mid X > m] = \frac{\Pr[X > m+k]}{\Pr[X > m]}$$

$$= \frac{(1-p)^{m+k}}{(1-p)^m}$$

$$= (1-p)^k$$

$$= \Pr[X > k]$$

✓

Coupon collecting revisited

Recall: - n different coupons
- sequence of uniform random samples
- X = # samples until we get at least one of each

Write $X = X_1 + X_2 + \cdots + X_n$

where X_i = no. of samples until we get the ith new coupon, starting after we got the $(i-1)$th

Claim: $\boxed{X_i \sim \text{Geom}\left(\frac{n-i+1}{n}\right)}$

Hence $E[X_i] = \dfrac{n}{n-i+1}$

$\sim \ln n + \gamma$

Linearity: $E[X] = \sum_{i=1}^{n} \dfrac{n}{n-i+1} = n \times \left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right)$

$\boxed{\sim n \ln n}$

Poisson Distribution

Suppose some event (e.g., a radioactive emission, a disconnected phone call etc.) occurs randomly at a certain average density λ per unit time, and occurrences are <u>independent</u>. Then the no. of occurrences in a unit of time is modeled by a <u>Poisson</u> r.v.

$$\boxed{\Pr[X=k] = e^{-\lambda}\frac{\lambda^k}{k!}} \qquad k = 0, 1, 2, \ldots$$

<u>Check</u> :

$$\sum_{k=0}^{\infty} \Pr[X=k] = \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!}$$

$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}$$

$$= 1$$

$$\boxed{X \sim \mathrm{Pois}(\lambda)}$$

$$\boxed{\Pr[X=k] = e^{-\lambda}\frac{\lambda^k}{k!}}$$

E.g., # goals in a World Cup soccer match

$\lambda = 2.5$

$$\Pr[0 \text{ goals}] = e^{-2.5}\frac{(2.5)^0}{0!} = e^{-2.5} \approx 0.082$$

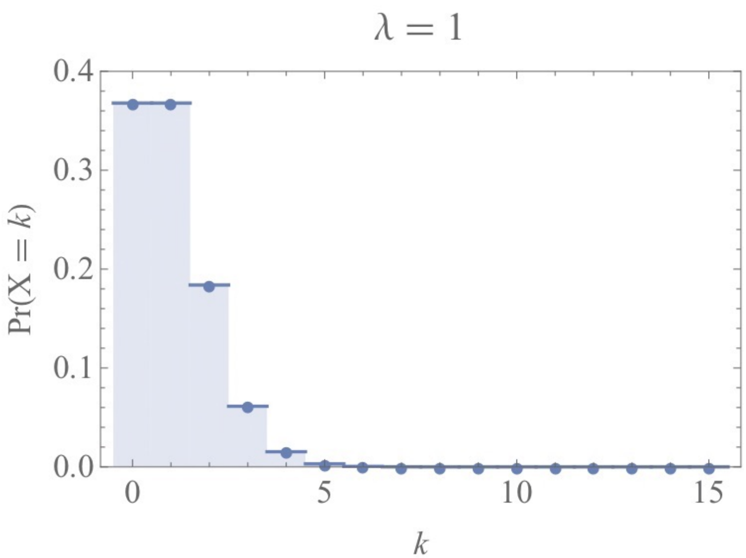
$$\Pr[1 \text{ goal}] = e^{-2.5}\frac{2.5}{1!} \approx 0.205$$

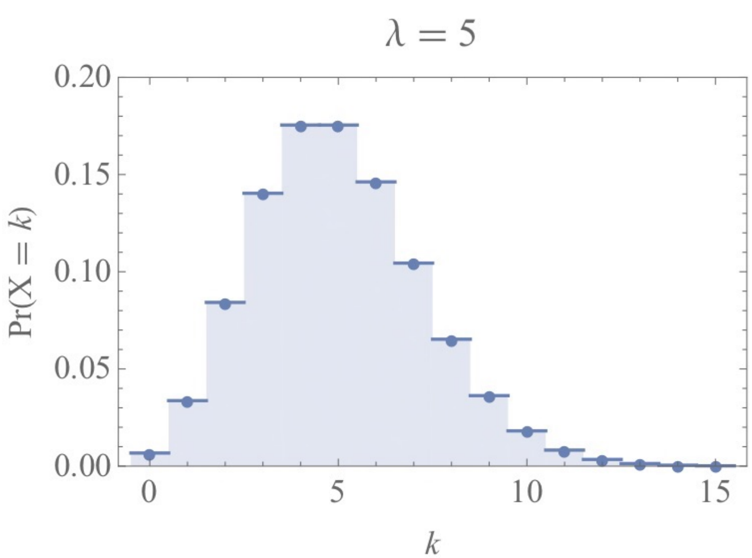
$$\Pr[2 \text{ goals}] = e^{-2.5}\frac{(2.5)^2}{2!} \approx 0.257$$

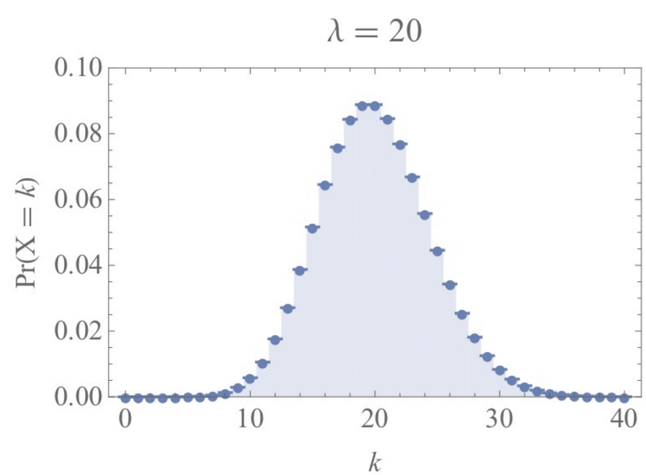
$$\Pr[3 \text{ goals}] = e^{-2.5}\frac{(2.5)^3}{3!} \approx 0.214$$

$$\Pr[>3 \text{ goals}] \approx 0.242$$

Histograms of Pois(λ)







Note : The distribution is unimodal, peaks at $\lfloor \lambda \rfloor$

Expectation of Pois(λ)

$$\Pr[X=k] = e^{-\lambda}\frac{\lambda^k}{k!}$$

$$E[X] = \sum_{k=0}^{\infty} k \times \Pr[X=k]$$

$$= \sum_{k=1}^{\infty} k \times e^{-\lambda}\frac{\lambda^k}{k!}$$

$$= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{\lambda}$$

$$= \lambda e^{-\lambda} e^{\lambda}$$

$$= \boxed{\lambda}$$

Sum of Independent Poisson R.V.'s

Thm: Suppose $X \sim \text{Pois}(\lambda)$ and $Y \sim \text{Pois}(\mu)$ are <u>independent</u>. Then $X+Y \sim \text{Pois}(\lambda+\mu)$

Proof:

$$\Pr[X+Y=k] = \sum_{j=0}^{k} \Pr[X=j,\, Y=k-j]$$

$$= \sum_{j=0}^{k} \Pr[X=j]\Pr[Y=k-j] \qquad \text{(indep.)}$$

$$= \sum_{j=0}^{k} e^{-\lambda}\frac{\lambda^j}{j!} \times e^{-\mu}\frac{\mu^{k-j}}{(k-j)!}$$

$$= e^{-(\lambda+\mu)} \cdot \frac{1}{k!} \sum_{j=0}^{k} \frac{k!}{j!\,(k-j)!}\lambda^j \mu^{k-j}$$

$$= e^{-(\lambda+\mu)} \cdot \frac{1}{k!}(\lambda+\mu)^k \; \checkmark \qquad \text{(binomial theorem)}$$

Poisson vs. Binomial

Example: Balls & bins with n balls, n bins

R.v. $X = $ # balls in bin 1

Then $\boxed{X \sim \text{Bin}(\qquad)}$ So $E[X] =$

So: $\Pr[X=k] = \binom{n}{k}\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{n-k}$ $k = 0, 1, 2, \ldots$

Now fix k and let $n \to \infty$

$$\Pr[X=k] = \binom{n}{k}\frac{1}{n^k}\left(1-\frac{1}{n}\right)^{n-k}$$

$$\frac{1}{n^k}\binom{n}{k} = \frac{1}{k!}\,\frac{n(n-1)\cdots(n-k+1)}{n^k}$$

$$\to \frac{1}{k!} \quad \text{as } n\to\infty$$

$$\xrightarrow[n\to\infty]{} \frac{1}{k!}e^{-1}$$

$$\left(1-\frac{1}{n}\right)^{n-k} \sim e^{-1-\frac{k}{n}} \to e^{-1}$$

as $n \to \infty$

So as $n \to \infty$, $\boxed{X \sim \text{Pois}(1)}$

E.g. $\Pr[X=0] \to e^{-1}$ $\Pr[X=1] \to e^{-1}$

More generally, for any constant λ,

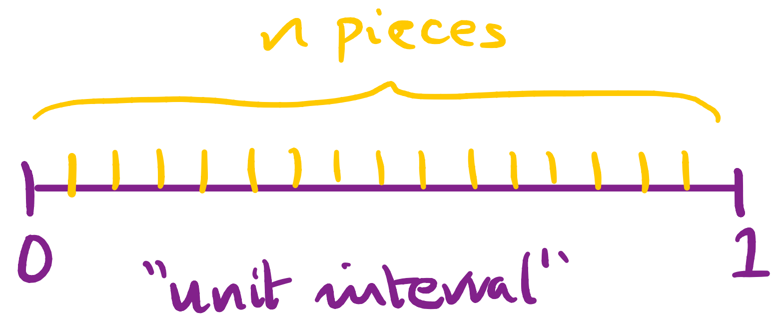
$$\mathrm{Bin}\left(n, \frac{\lambda}{n}\right) \xrightarrow[n \to \infty]{} \mathrm{Pois}(\lambda)$$

[in sense that $\forall k$ $\Pr[\mathrm{Bin}(n, \frac{\lambda}{n}) = k] \to \Pr[\mathrm{Pois}(\lambda) = k]$]

Connection with "rare events"

Assume

- expect λ events per unit interval
- events are "independent"



Divide interval into n equal-sized pieces

$\Pr[\text{event happens in one piece}] = \frac{\lambda}{n}$ (and at most one event per piece as $n \to \infty$)

Events in different pieces mutually independent

$X = \#$ events in interval : $X \sim \mathrm{Bin}(n, \frac{\lambda}{n}) \to \mathrm{Pois}(\lambda)$