BBM 205 Discrete Mathematics
Hacettepe University
http://web.cs.hacettepe.edu.tr/~bbm205

Lecture 10a: Random Variables

Resources:
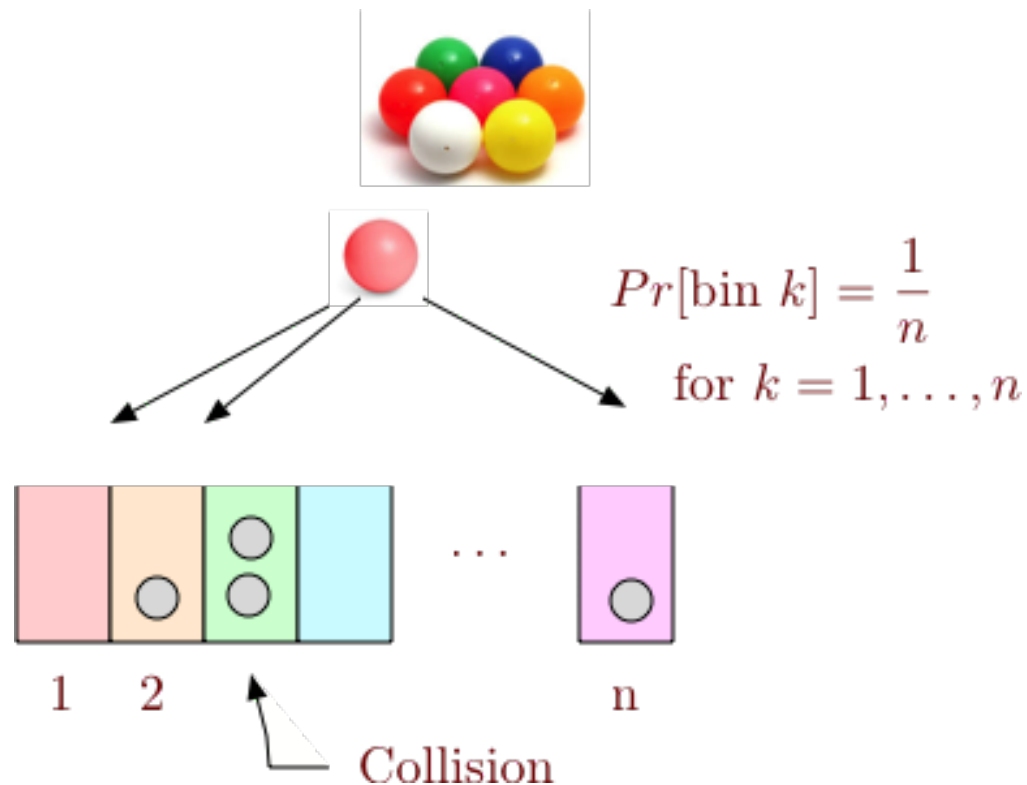**Kenneth Rosen, "Discrete Mathematics and App."**
http://www.eecs70.org/

Balls in bins

One throws m balls into $n > m$ bins.

Balls in bins

One throws m balls into $n > m$ bins.

$$Pr[\text{bin } k] = \frac{1}{n}$$
for $k = 1, \ldots, n$

Collision

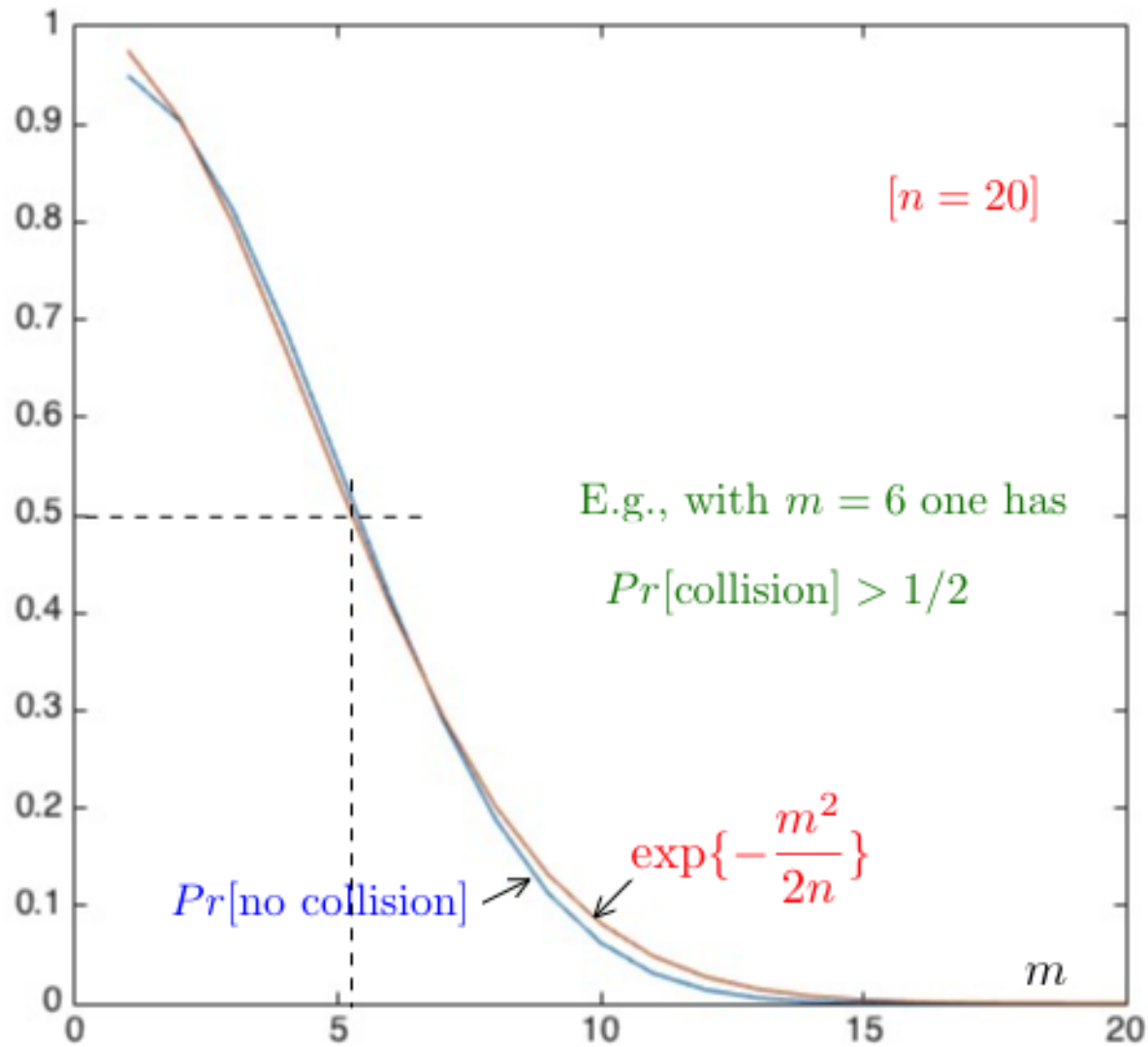**Theorem:**

$Pr[\text{no collision}] \approx \exp\{-\frac{m^2}{2n}\}$, for large enough n.

Balls in bins

Theorem:
$Pr[\text{no collision}] \approx \exp\{-\frac{m^2}{2n}\}$, for large enough n.

$[n = 20]$

E.g., with $m = 6$ one has

$Pr[\text{collision}] > 1/2$

$\exp\{-\dfrac{m^2}{2n}\}$

$Pr[\text{no collision}]$

m

Balls in bins

Theorem:
$Pr[\text{no collision}] \approx \exp\{-\frac{m^2}{2n}\}$, for large enough n.

In particular, $Pr[\text{no collision}] \approx 1/2$ for $m^2/(2n) \approx \ln(2)$, i.e.,

$$m \approx \sqrt{2\ln(2)n} \approx 1.2\sqrt{n}.$$

E.g., $1.2\sqrt{20} \approx 5.4$.

Roughly, $Pr[\text{collision}] \approx 1/2$ for $m = \sqrt{n}$. ($e^{-0.5} \approx 0.6$.)

The Calculation.

A_i = no collision when ith ball is placed in a bin.

$Pr[A_i | A_{i-1} \cap \cdots \cap A_1] = (1 - \frac{i-1}{n})$.

no collision = $A_1 \cap \cdots \cap A_m$.

Product rule:

$Pr[A_1 \cap \cdots \cap A_m] = Pr[A_1] Pr[A_2 | A_1] \cdots Pr[A_m | A_1 \cap \cdots \cap A_{m-1}]$

$$\Rightarrow Pr[\text{no collision}] = \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{m-1}{n}\right).$$
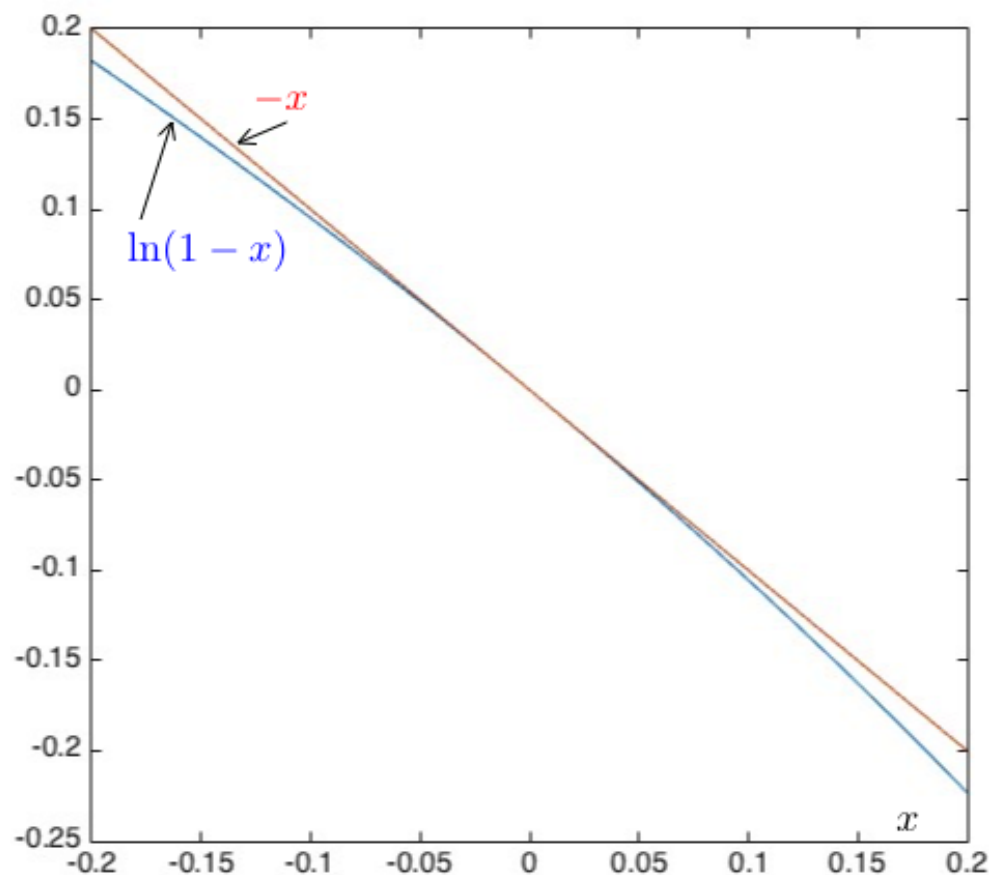
Hence,

$$
\begin{aligned}
\ln(Pr[\text{no collision})) &= \sum_{k=1}^{m-1} \ln(1 - \frac{k}{n}) \approx \sum_{k=1}^{m-1} (-\frac{k}{n})\ ^{(*)} \\
&= -\frac{1}{n} \frac{m(m-1)}{2}\ ^{(\dagger)} \approx -\frac{m^2}{2n}
\end{aligned}
$$

$(*)$ We used $\ln(1 - \varepsilon) \approx -\varepsilon$ for $|\varepsilon| \ll 1$.

$(\dagger)$ $1 + 2 + \cdots + m - 1 = (m-1)m/2$.

Approximation

$$\exp\{-x\} = 1 - x + \frac{1}{2!}x^2 + \cdots \approx 1 - x, \text{ for } |x| \ll 1.$$

Hence, $-x \approx \ln(1-x)$ for $|x| \ll 1$.

Today's your birthday, it's my birthday too..

Probability that m people all have different birthdays?
With $n = 365$, one finds

$Pr[\text{collision}] \approx 1/2$ if $m \approx 1.2\sqrt{365} \approx 23$.

If $m = 60$, we find that

$$Pr[\text{no collision}] \approx \exp\left\{-\frac{m^2}{2n}\right\} = \exp\left\{-\frac{60^2}{2 \times 365}\right\} \approx 0.007.$$

If $m = 366$, then $Pr[\text{no collision}] = 0$. (No approximation here!)

Checksums!

Consider a set of m files.
Each file has a checksum of b bits.
How large should b be for $Pr[\text{share a checksum}] \leq 10^{-3}$?

Claim: $b \geq 2.9\ln(m) + 9$.

Proof:

Let $n = 2^b$ be the number of checksums.
We know $Pr[\text{no collision}] \approx \exp\{-m^2/(2n)\} \approx 1 - m^2/(2n)$.
Hence,

$$Pr[\text{no collision}] \approx 1 - 10^{-3} \Leftrightarrow m^2/(2n) \approx 10^{-3}$$
$$\Leftrightarrow 2n \approx m^2 10^3 \Leftrightarrow 2^{b+1} \approx m^2 2^{10}$$
$$\Leftrightarrow b + 1 \approx 10 + 2\log_2(m) \approx 10 + 2.9\ln(m).$$

Note: $\log_2(x) = \log_2(e)\ln(x) \approx 1.44\ln(x)$.

Coupon Collector Problem.

There are n different baseball cards.
(Brian Wilson, Jackie Robinson, Roger Hornsby, ...)

One random baseball card in each cereal box.

Theorem: If you buy m boxes,

(a) $Pr[\text{miss one specific item}] \approx e^{-\frac{m}{n}}$

(b) $Pr[\text{miss any one of the items}] \leq ne^{-\frac{m}{n}}$.

Coupon Collector Problem: Analysis.

Event A_m = 'fail to get Brian Wilson in m cereal boxes'

Fail the first time: $(1 - \frac{1}{n})$
Fail the second time: $(1 - \frac{1}{n})$
And so on ... for m times. Hence,

$$
\begin{aligned}
Pr[A_m] &= (1 - \frac{1}{n}) \times \cdots \times (1 - \frac{1}{n}) \\
&= (1 - \frac{1}{n})^m \\
\ln(Pr[A_m]) &= m\ln(1 - \frac{1}{n}) \approx m \times (-\frac{1}{n}) \\
Pr[A_m] &\approx \exp\{-\frac{m}{n}\}.
\end{aligned}
$$

For $p_m = \frac{1}{2}$, we need around $n\ln 2 \approx 0.69n$ boxes.

Collect all cards?

Experiment: Choose m cards at random with replacement.

Events: E_k = 'fail to get player k' , for k = 1, ..., n

Probability of failing to get at least one of these n players:

$$p := Pr[E_1 \cup E_2 \cdots \cup E_n]$$

How does one estimate p? Union Bound:

$$p = Pr[E_1 \cup E_2 \cdots \cup E_n] \leq Pr[E_1] + Pr[E_2] \cdots Pr[E_n].$$

$$Pr[E_k] \approx e^{-\frac{m}{n}}, k = 1, \ldots, n.$$

Plug in and get

$$p \leq ne^{-\frac{m}{n}}.$$

Collect all cards?

Thus,

$$Pr[\text{missing at least one card}] \le ne^{-\frac{m}{n}}.$$

Hence,

$$Pr[\text{missing at least one card}] \le p \text{ when } m \ge n\ln(\frac{n}{p}).$$

To get $p = 1/2$, set $m = n\ln(2n)$.

$\quad (p \le ne^{-\frac{m}{n}} \le ne^{-\ln(n/p)} \le n(\frac{p}{n}) \le p.)$

E.g., $n = 10^2 \Rightarrow m = 530; n = 10^3 \Rightarrow m = 7600$.

Quick Review.

Main results:

- Bayes' Rule: $Pr[A_m|B] = p_m q_m/(p_1 q_1 + \cdots + p_M q_M)$.
- Product Rule:
  $Pr[A_1 \cap \cdots \cap A_n] = Pr[A_1]Pr[A_2|A_1] \cdots Pr[A_n|A_1 \cap \cdots \cap A_{n-1}]$.
- Balls in bins: m balls into $n > m$ bins.

$$Pr[\text{no collisions}] \approx \exp\{-\frac{m^2}{2n}\}$$

- Coupon Collection: n items. Buy m cereal boxes.

  $Pr[\text{miss one specific item}] \approx e^{-\frac{m}{n}}$; $Pr[\text{miss any one of the items}] \leq n e^{-\frac{m}{n}}$.

Key Mathematical Fact: $\ln(1 - \varepsilon) \approx -\varepsilon$.

Random Variables

Random Variables

1. Random Variables.
2. Expectation
3. Distributions.

Questions about outcomes ...

Experiment: roll two dice.
Sample Space: $\{(1,1),(1,2),\ldots,(6,6)\} = \{1,\ldots,6\}^2$
How many pips?

Experiment: flip 100 coins.
Sample Space: $\{HHH\cdots H, THH\cdots H, \ldots, TTT\cdots T\}$
How many heads in 100 coin tosses?

Experiment: choose a random student in cs70.
Sample Space: $\{Adam, Jin, Bing, \ldots, Angeline\}$
What midterm score?

Experiment: hand back assignments to 3 students at random.
Sample Space: $\{123, 132, 213, 231, 312, 321\}$
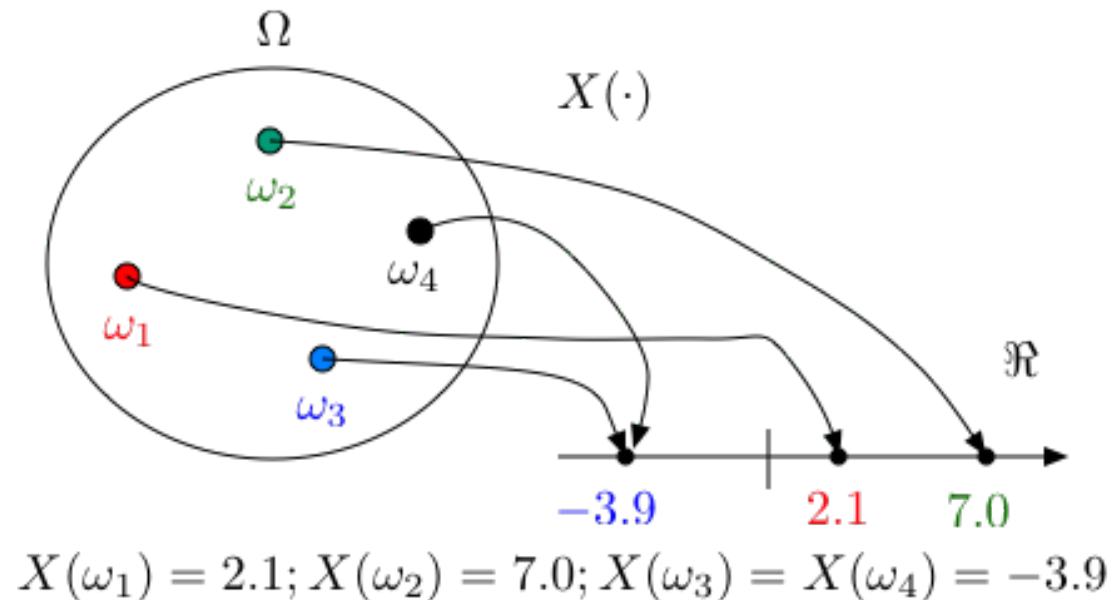How many students get back their own assignment?

In each scenario, each outcome gives a number.

The number is a (known) function of the outcome.

Random Variables.

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

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

$$X(\omega_1) = 2.1; X(\omega_2) = 7.0; X(\omega_3) = X(\omega_4) = -3.9$$

The function $X(\cdot)$ is defined on the outcomes Ω.

The function $X(\cdot)$ is not random, not a variable!

What varies at random (from experiment to experiment)? The outcome!

Example 1 of Random Variable

Experiment: roll two dice.

Sample Space: $\{(1,1),(1,2),\ldots,(6,6)\} = \{1,\ldots,6\}^2$

Random Variable X: number of pips.

$X(1,1) = 2$

$X(1,2) = 3,$

$\vdots$

$X(6,6) = 12,$

$X(a,b) = a+b, (a,b) \in \Omega.$

Example 2 of Random Variable

Experiment: flip three coins

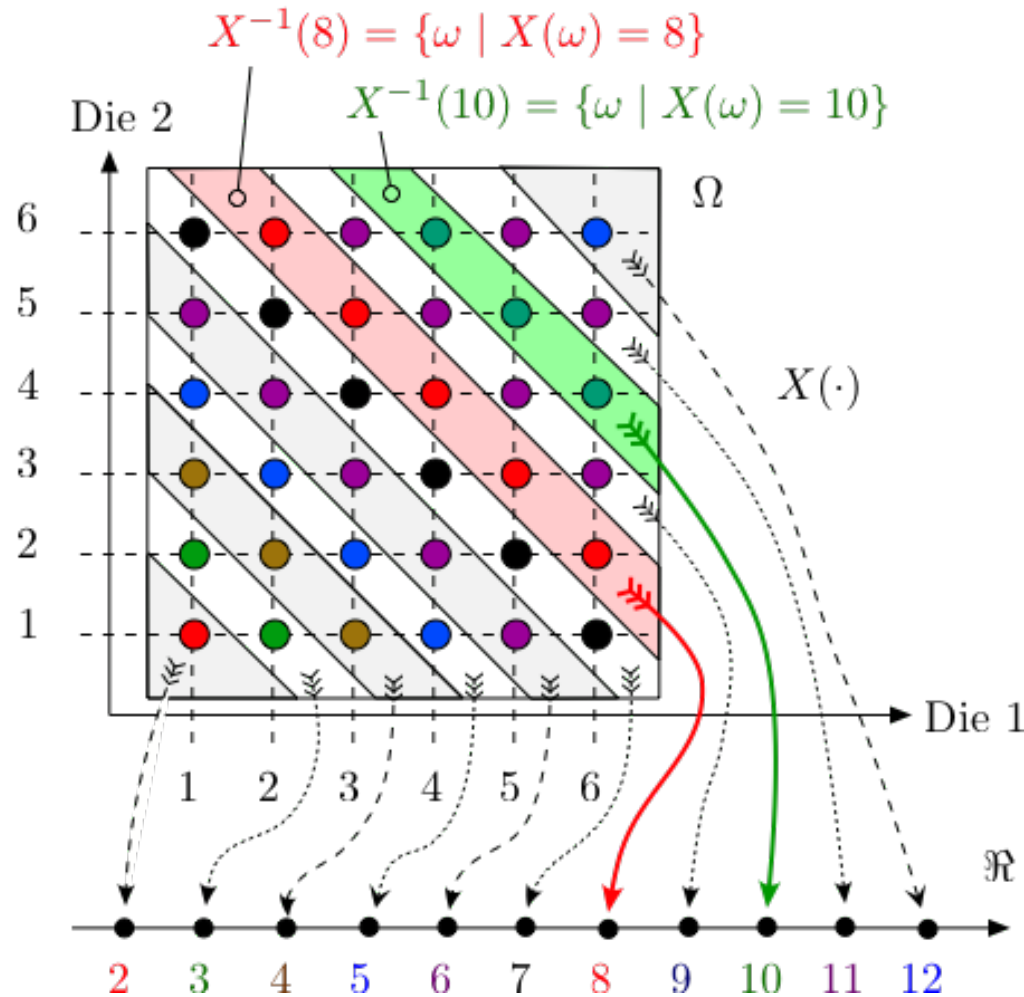Sample Space: $\{HHH, THH, HTH, TTH, HHT, THT, HTT, TTT\}$

Winnings: if win 1 on heads, lose 1 on tails: X

$X(HHH) = 3 \qquad X(THH) = 1 \qquad X(HTH) = 1 \qquad X(TTH) = -1$

$X(HHT) = 1 \qquad X(THT) = -1 \qquad X(HTT) = -1 \qquad X(TTT) = -3$

Number of pips in two dice.

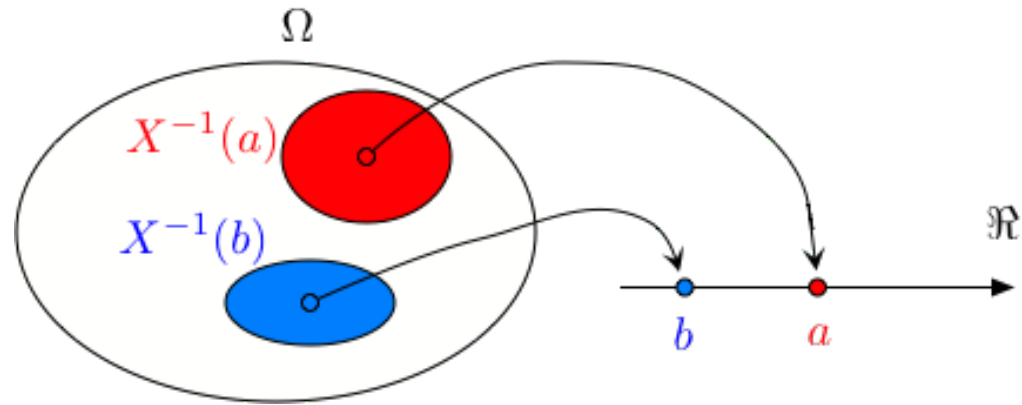"What is the likelihood of getting *n* pips?"

$$Pr[X = 10] = 3/36 = Pr[X^{-1}(10)]; Pr[X = 8] = 5/36 = Pr[X^{-1}(8)].$$

Distribution

The probability of X taking on a value a.

Definition: The **distribution** of a random variable X, is $\{(a, Pr[X = a]) : a \in \mathscr{A}\}$, where $\mathscr{A}$ is the range of X.

$$Pr[X = a] := Pr[X^{-1}(a)] \text{ where } X^{-1}(a) := \{\omega \mid X(\omega) = a\}.$$

Handing back assignments

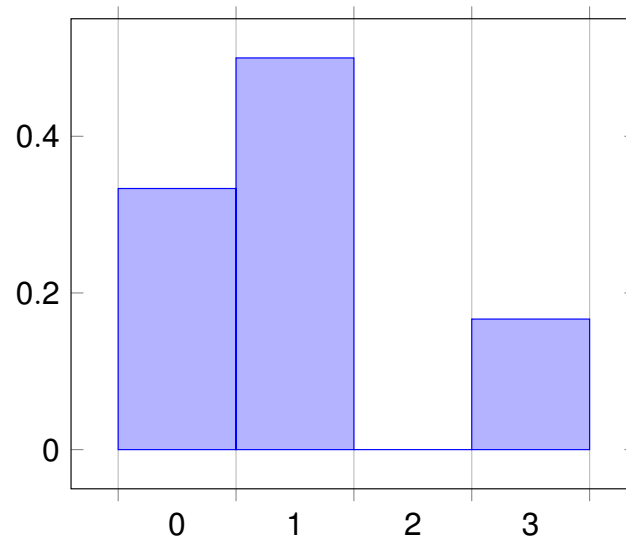Experiment: hand back assignments to 3 students at random.
Sample Space: $\Omega = \{123, 132, 213, 231, 312, 321\}$
How many students get back their own assignment?
Random Variable: values of $X(\omega) : \{3, 1, 1, 0, 0, 1\}$

Distribution:

$$X = \begin{cases} 0, & \text{w.p. } 1/3 \\ 1, & \text{w.p. } 1/2 \\ 3, & \text{w.p. } 1/6 \end{cases}$$

Flip three coins

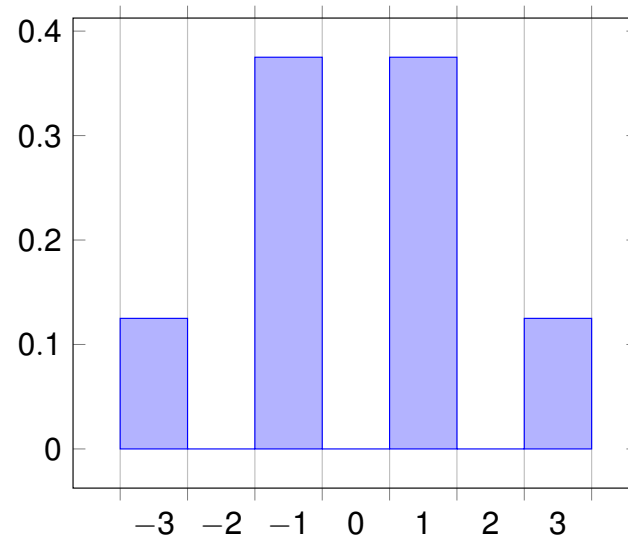Experiment: flip three coins

Sample Space: $\{HHH, THH, HTH, TTH, HHT, THT, HTT, TTT\}$

Winnings: if win 1 on heads, lose 1 on tails. X

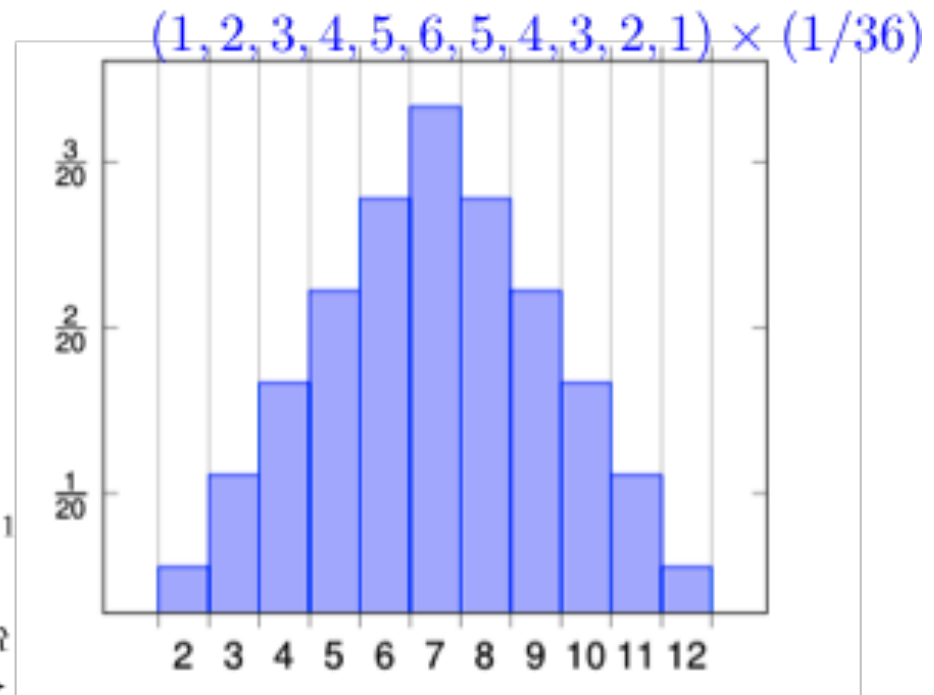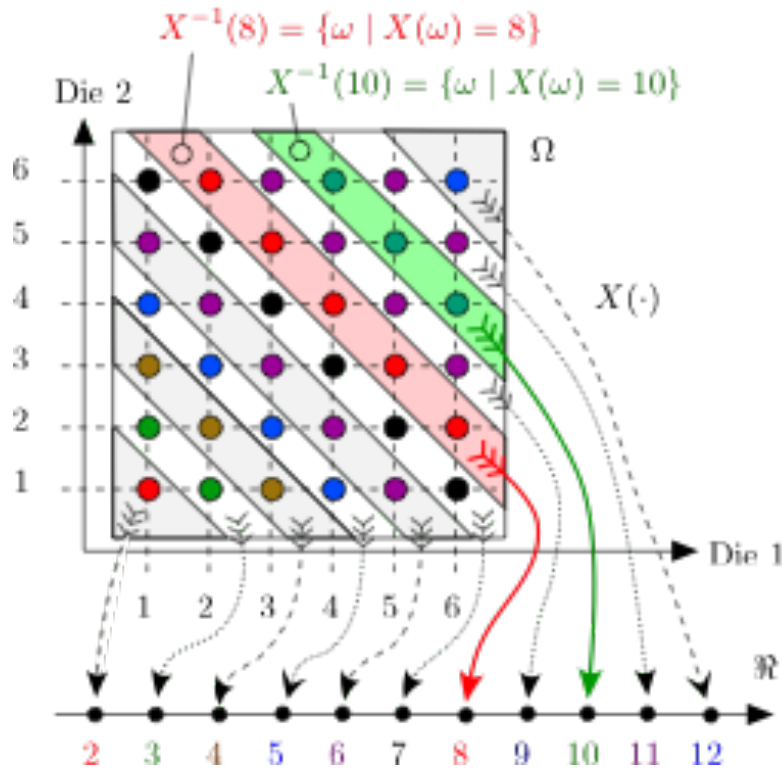Random Variable: $\{3, 1, 1, -1, 1, -1, -1, -3\}$

Distribution:

$$X = \begin{cases} -3, & \text{w. p. } 1/8 \\ -1, & \text{w. p. } 3/8 \\ 1, & \text{w. p. } 3/8 \\ 3 & \text{w. p. } 1/8 \end{cases}$$

Number of pips.

Experiment: roll two dice.

Expectation.

How did people do on the midterm?

Distribution.

Summary of distribution?

Average!

$$E(X) = \sum_{i=1}^{n} x_i p(x_i)$$

GREAT EXPECTATIONS

Expectation - Definition

Definition: The **expected value** of a random variable X is

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

The expected value is also called the mean.

According to our intuition, we expect that if we repeat an experiment a large number N of times and if $X_1, \ldots, X_N$ are the successive values of the random variable, then

$$\frac{X_1 + \cdots + X_N}{N} \approx E[X].$$

That is indeed the case, in the same way that the fraction of times that $X = x$ approaches $Pr[X = x]$.

This (nontrivial) result is called the Law of Large Numbers.

The subjectivist(bayesian) interpretation of $E[X]$ is less obvious.

Expectation: A Useful Fact

Theorem:

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

Proof:

$$
\begin{aligned}
E[X] &= \sum_a a \times Pr[X = a] \\
&= \sum_a a \times \sum_{\omega:X(\omega)=a} Pr[\omega] \\
&= \sum_a \sum_{\omega:X(\omega)=a} X(\omega)Pr[\omega] \\
&= \sum_\omega X(\omega)Pr[\omega]
\end{aligned}
$$

Distributive property of multiplication over addition.

An Example

Flip a fair coin three times.

$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$.

$X =$ number of H's: $\{3, 2, 2, 2, 1, 1, 1, 0\}$.

Thus,

$$\sum_\omega X(\omega)Pr[\omega] = \{3 + 2 + 2 + 2 + 1 + 1 + 1 + 0\} \times \frac{1}{8}.$$

Also,

$$\sum_a a \times Pr[X = a] = 3 \times \frac{1}{8} + 2 \times \frac{3}{8} + 1 \times \frac{3}{8} + 0 \times \frac{1}{8}.$$

What's the answer? Uh.... $\frac{3}{2}$

Expectation and Average.

There are n students in the class;

$X(m)$ = score of student m, for $m = 1, 2, \ldots, n$.

"Average score" of the n students: add scores and divide by n:

$$\text{Average} = \frac{X(1) + X(1) + \cdots + X(n)}{n}.$$

Experiment: choose a student uniformly at random.
Uniform sample space: $\Omega = \{1, 2, \cdots, n\}, Pr[\omega] = 1/n$, for all ω.
Random Variable: midterm score: $X(\omega)$.
Expectation:

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

Hence,

$$\text{Average} = E(X).$$

This holds for a uniform probability space.

Named Distributions.

Some distributions come up over and over again.

...like "choose" or "stars and bars"....

Let's cover some.

The binomial distribution.

Flip *n* coins with heads probability *p*.

Random variable: number of heads.

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

How many sample points in event "$X = i$"?
 i heads out of *n* coin flips $\implies \binom{n}{i}$

What is the probability of ω if ω has *i* heads?
 Probability of heads in any position is *p*.
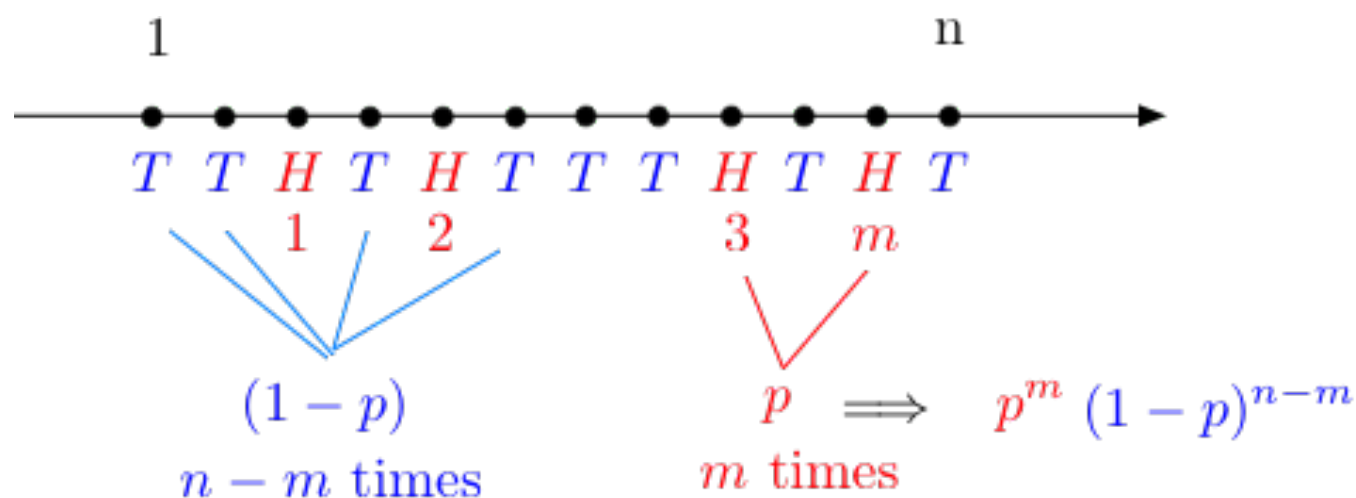 Probability of tails in any position is $(1 - p)$.
So, we get

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

Probability of "$X = i$" is sum of $Pr[\omega]$, $\omega \in$ "$X = i$".

$$Pr[X = i] = \binom{n}{i} p^i(1 - p)^{n-i}, i = 0, 1, \ldots, n : B(n, p) \text{ distribution}$$

The binomial distribution.

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

Error channel and...

A packet is corrupted with probability p.

Send $n+2k$ packets.

Probability of at most k corruptions.

$$\sum_{i \leq k} \binom{n+2k}{i} p^i (1-p)^{n+2k-i}.$$

Also distribution in polling, experiments, etc.

Expectation of Binomial Distribution

Parameter p and n. What is expectation?

$$Pr[X = i] = \binom{n}{i} p^i (1-p)^{n-i}, i = 0, 1, \ldots, n : B(n,p) \text{ distribution}$$

$$E[X] = \sum_i i \times Pr[X = i].$$

Uh oh? Well... It is pn.

Proof? After linearity of expectation this is easy.

Waiting is good.

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:

$$\omega_1 = H, \text{ or}$$
$$\omega_2 = T\,H, \text{ or}$$
$$\omega_3 = T\,T\,H, \text{ or}$$
$$\omega_n = T\,T\,T\,T\,\cdots\,T\,H.$$
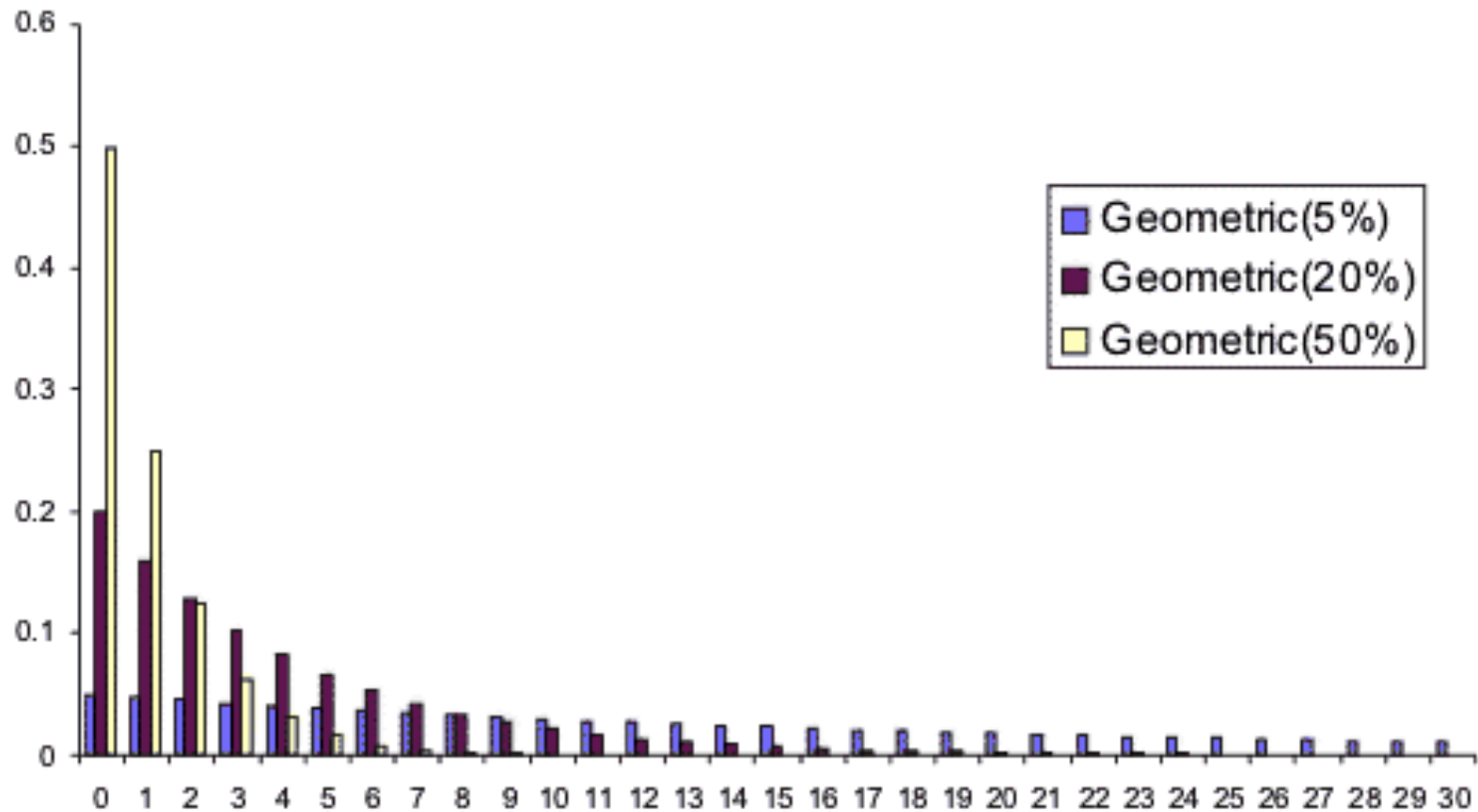
Note that $\Omega = \{\omega_n, n = 1, 2, \ldots\}$.

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,

$$
\begin{aligned}
E[X] &= p + 2(1-p)p + 3(1-p)^2p + 4(1-p)^3p + \cdots \\
(1-p)E[X] &= (1-p)p + 2(1-p)^2p + 3(1-p)^3p + \cdots \\
pE[X] &= p + (1-p)p + (1-p)^2p + (1-p)^3p + \cdots
\end{aligned}
$$

by subtracting the previous two identities

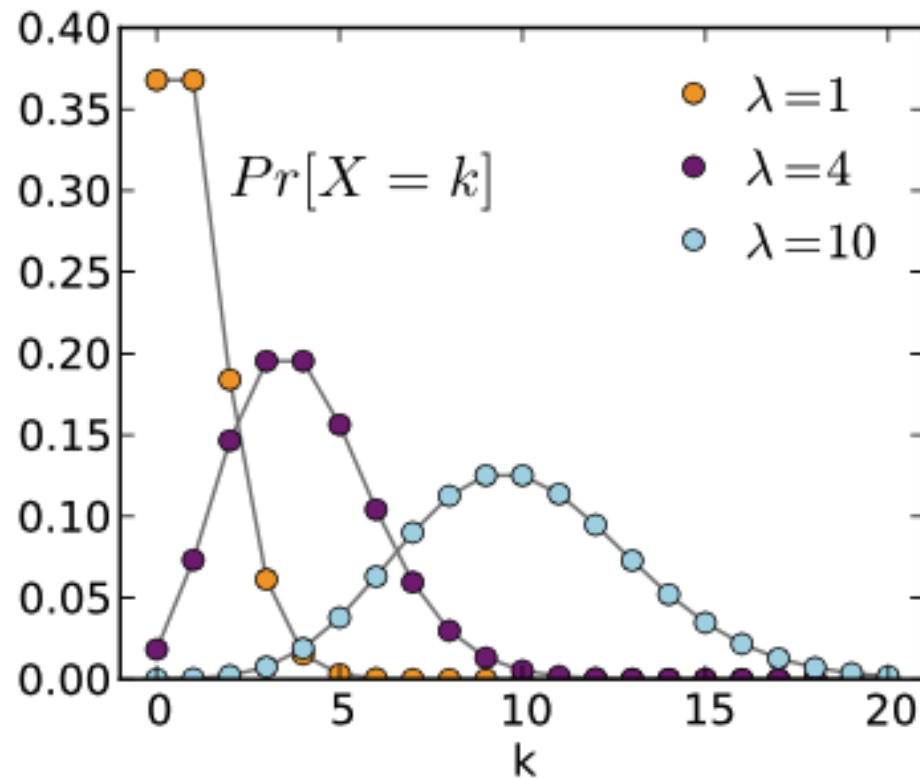$$= \sum_{n=1}^{\infty} Pr[X = n] = 1.$$

Hence,

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

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

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

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:

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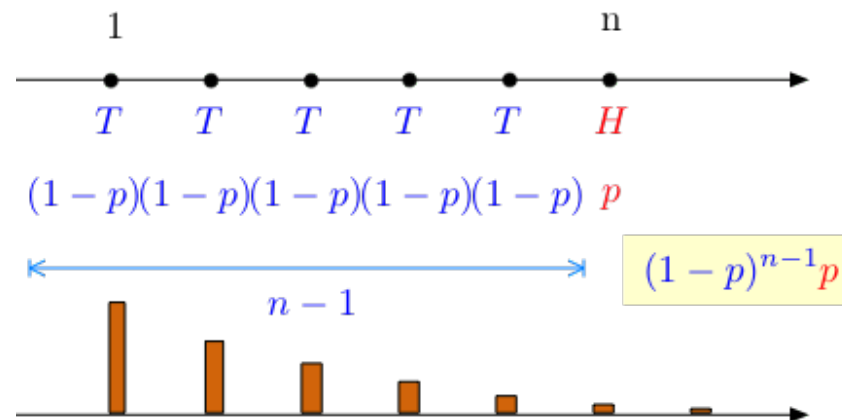&= e^{-\lambda} \lambda e^{\lambda} = \lambda.
\end{aligned}
$$

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

- A random variable X is a function $X : \Omega \to \mathfrak{R}$.
- $Pr[X = a] := Pr[X^{-1}(a)] = Pr[\{\omega \mid X(\omega) = a\}]$.
- $Pr[X \in A] := Pr[X^{-1}(A)]$.
- The distribution of X is the list of possible values and their probability: $\{(a, Pr[X = a]), a \in \mathscr{A}\}$.
- $E[X] := \sum_a a Pr[X = a]$.
- Expectation is Linear.
- $B(n, p), U[1 : n], G(p), P(\lambda)$.