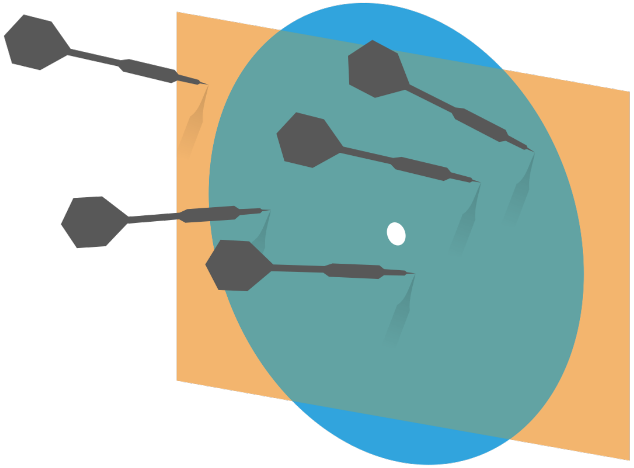
Probability and Statistics for Computer Science



Credit: wikipedia

Who discovered this?

$$e = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n$$

Hongye Liu, Teaching Assistant Prof, CS361, UIUC, 09.24.2020

what is the number?

$$(p+q)^n = \sum_{k=0}^{n} a_k p^k q^{n-k}$$

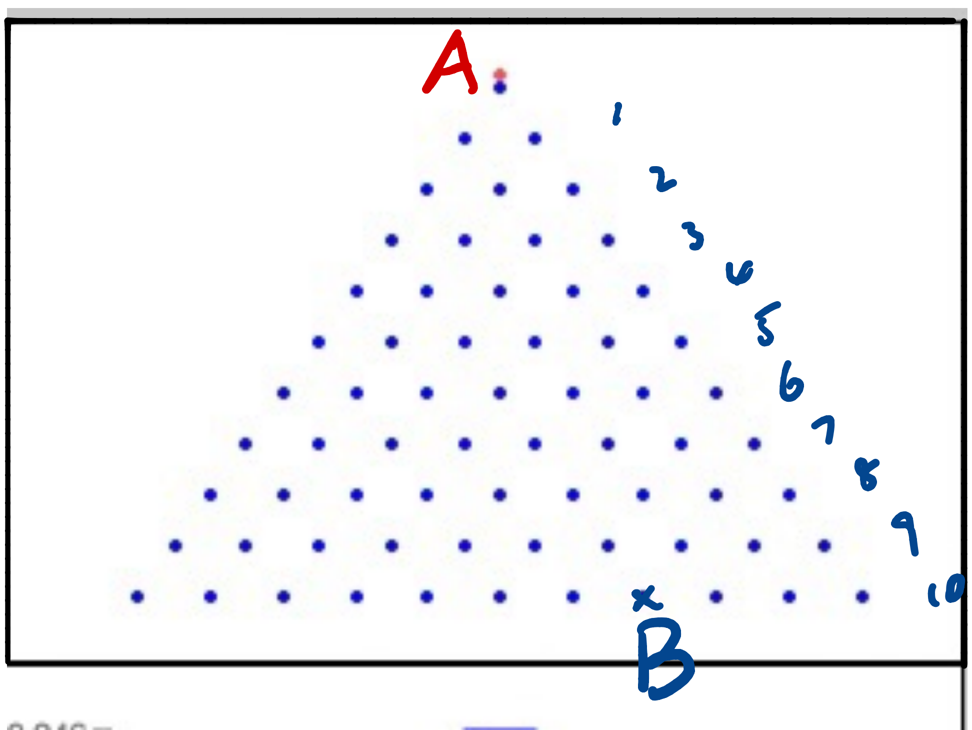
$a_k = ?$ $\binom{n}{k}$

$a_k = \binom{n}{k}$

$p+q=1$ then

$$\Longrightarrow \underbrace{\sum_{k=0}^{n} a_k p^k q^{n-k} = 1}$$

How many paths that lead $A \to B$?



$$\binom{10}{7} = \binom{10}{3}$$

Last time

Random variable (R.V.)

* Review of expectations
* Markov's Inequality
  Chebyshev's Inequality
* The weak law of large numbers

Proof of Weak law of large numbers

- Apply Chebyshev's inequality

$$P(|\overline{\mathbf{X}} - E[\overline{\mathbf{X}}]| \geq \epsilon) \leq \frac{var[\overline{\mathbf{X}}]}{\epsilon^2}$$

- Substitute $E[\overline{\mathbf{X}}] = E[X]$ and $var[\overline{\mathbf{X}}] = \frac{var[X]}{N}$

$$P(|\overline{\mathbf{X}} - E[\mathbf{X}]| \geq \epsilon) \leq \frac{var[\mathbf{X}]}{N\epsilon^2} \xrightarrow[N \to \infty]{} 0$$

$$\boxed{\lim_{N \to \infty} P(|\overline{\mathbf{X}} - E[X]| \geq \epsilon) = 0}$$

Applications of the Weak law of large numbers

- The law of large numbers ***justifies using simulations*** (instead of calculation) to estimate the expected values of random variables

$$\lim_{N\to\infty} P(|\overline{\mathbf{X}} - E[X]| \geq \epsilon) = 0$$

- The law of large numbers also ***justifies using histogram*** of large random samples to approximate the probability distribution function $P(x)$, see proof on

Pg. 353 of the textbook by DeGroot, et al.

Read off-line

Histogram of large random IID samples approximates the probability distribution

- The law of large numbers justifies using histograms to approximate the probability distribution. Given ***N*** IID random variables X_1, …, X_N
  - According to the law of large numbers

$$\overline{\mathbf{Y}} = \frac{\sum_{i=1}^{N} Y_i}{N} \xrightarrow{N \to \infty} E[Y_i]$$

read off line.

  - As we know for $\boxed{\text{indicator function}}$ → text book Pg 100-101

$$E[Y_i] = P(c_1 \leq X_i < c_2) = P(c_1 \leq X < c_2)$$

Objectives

Bernoulli Distribution

Binomial Distribution

Geometric Distribution

Bernoulli trials

Discrete Uniform Distribution

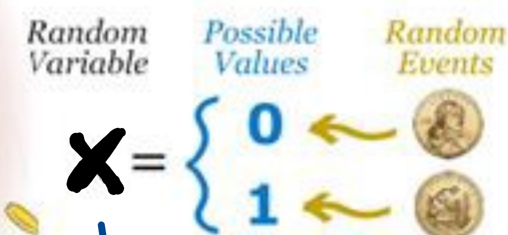
Continuous Random Variable

Random variables

A random variable maps all outcomes (ω) to Numbers (X), so it's a function!!

Bernoulli:

$p(\text{tail}) = \frac{1}{2}$



ω is tail

ω is head

$X(\omega)$

Bernoulli Random Variable

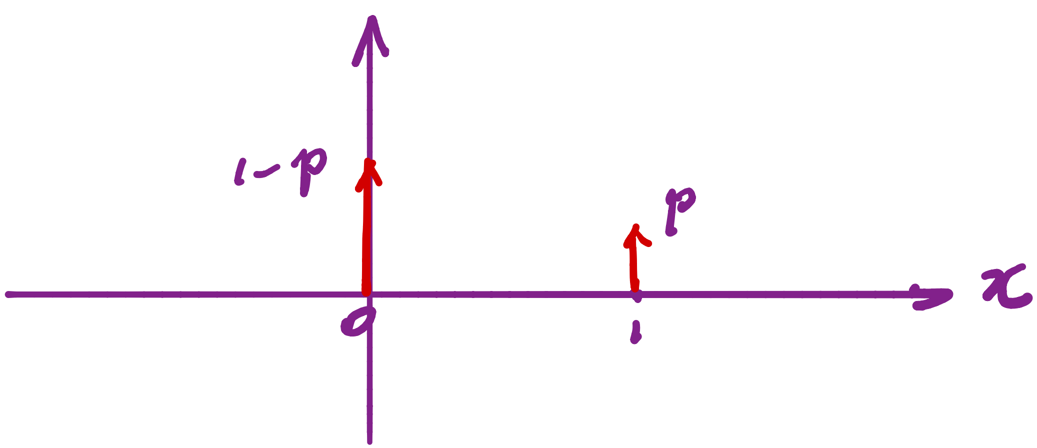
$$X(\omega) = \begin{cases} 1 & \omega = \text{event } A \\ 0 & \omega = \text{otherwise} \end{cases}$$

→ Head

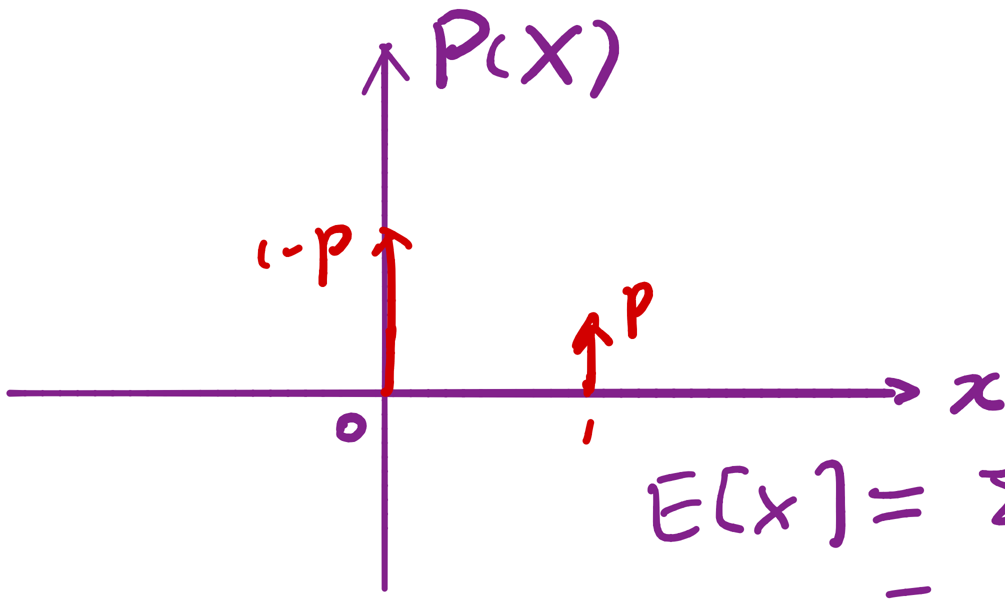
→ Tail

$P(A) = ?\ p$

$P(X=x)$



Bernoulli Distribution



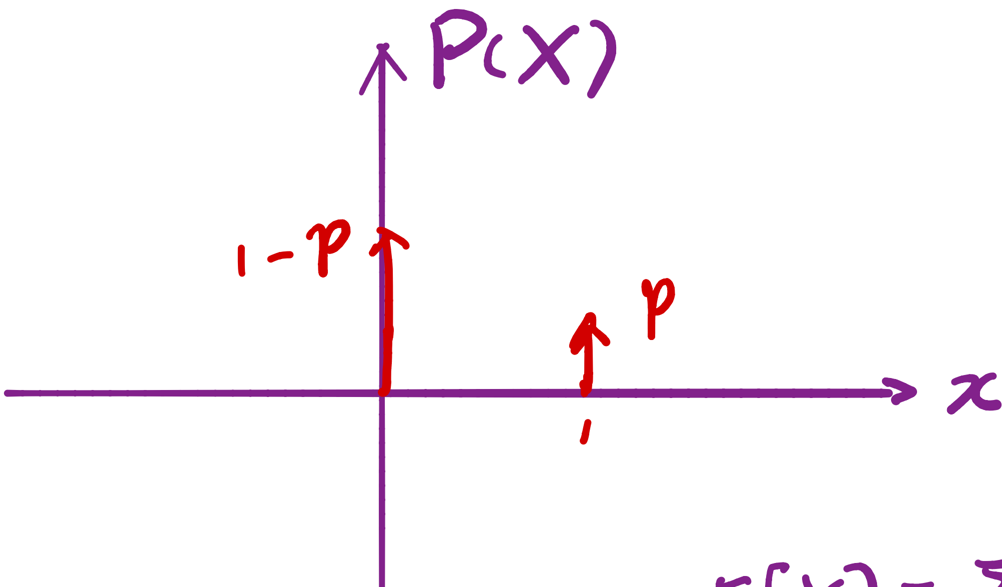
$E[X] = ?$

$var[X] = ?$

$$E[X] = \sum x P(x)$$
$$= 1 \cdot p + 0 \cdot (1-p) = p$$

$$var[X] = E[X^2] - E^2[X]$$
$$= \sum x^2 p(x) - p^2$$
$$= 1^2 \cdot p + 0^2 \cdot (1-p) - p^2$$
$$= p - p^2 = p(1-p)$$

Bernoulli Distribution



$E[X] = ?$

$var[X] = ?$

$$E[X] = \sum x\,P(x) = 1 \times p + 0 \cdot (1-p)$$
$$= p$$

$$var[X] = E[X^2] - E^2[X]$$
$$= 1^2 \cdot p + 0^2 \cdot (1-p) - p^2 = p - p^2$$
$$= p(1-p)$$

Bernoulli distribution

- A random variable X is **Bernoulli** if it takes on two values 0 and 1 such that

$$P(X=x) = \begin{cases} p & x=1 \\ 1-p & x=0 \\ 0 & \text{otherwise} \end{cases}$$

$$E[X] = p$$

$$var[X] = p(1-p)$$

Jacob Bernoulli (1654-1705)

Credit: wikipedia

Bernoulli distribution

- Examples
  - Tossing a biased (or fair) coin
  - Making a free throw
  - Rolling a six-sided die and checking if it shows 6
  - **Any indicator function** of a random variable

$$\mathbb{I}_A = \begin{cases} 1 & \text{Event A happens} \\ 0 & \text{otherwise} \end{cases}$$

$P(A) = p$

$$E[\mathbb{I}_A] = 1 \times P(A) + 0 \times (1 - P(A))$$
$$= P(\text{Event } A)$$

Binomial Distribution

Binomial RV X_s is the sum of N independent Bernoulli RVs

$$X_s = \sum_{i=1}^{N} X_i$$

$$X_i(\omega) = \begin{cases} 1 & \omega = \text{event } A \\ 0 & \omega = \text{other} \end{cases}$$

Range of X_s is ?

$[0, N]$

Binomial Distribution

Binomial RV X_s is the sum of N

independent Bernoulli RVs

→ Toss N times a biased coin,

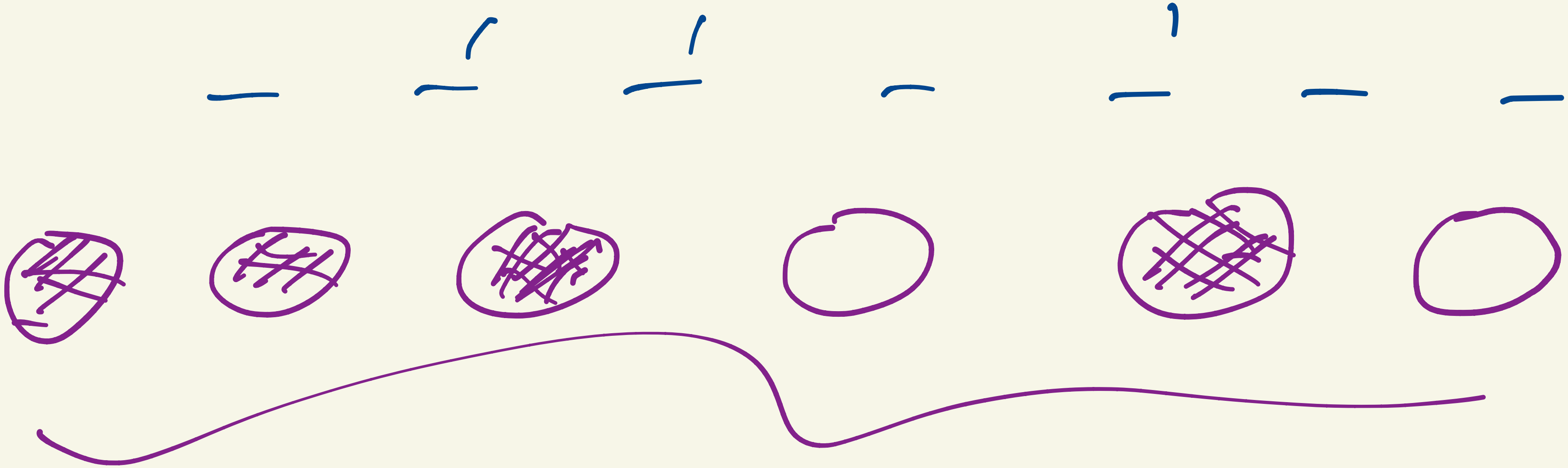
how many heads ?
k

$P(X_s = k) = ?$

$= \binom{N}{k} \cdot p^k (1-p)^{N-k}$

$\binom{N}{K} \underbrace{\overbrace{p \cdot p \cdots p}^{K} \; \overbrace{(1-p) \cdots (1-p)}^{N-K}}$

$k \in [0, N]$



6 positions

$$\binom{6}{4}$$

$$\binom{N}{K}$$

N positi

k heads

Expectations of Binomial distribution

- A discrete random variable X is binomial if

$$P(X=k) = \binom{N}{k} p^k (1-p)^{N-k} \quad for\ integer\ 0 \le k \le N$$

with $E[X] = Np \quad \& \quad var[X] = Np(1-p)$

$$E[X] = \sum x\, p(x)$$

$$= \sum k\, p(k)$$

$$= \sum k \binom{N}{k} p^k (1-p)^{N-k}$$

when $N = 1$

$\rightarrow$ Bernoulli:

$$var[X] = E[(X - E[X])^2] = \sum (x - \cdots)^2 \cdot p(x=k)$$

$$X_s = \sum_{i=1}^{N} X_i$$

$$E[X_s] = E\left(\sum_{i=1}^{N} X_i\right)$$

$$= \sum_{i=1}^{N} E[X_i]$$

$$= \sum_{i=1}^{N} p$$

$$= N \cdot p$$

X_i iid.

$E[X_i] = E[X]$

Bernoulli RV

$$\left\{\begin{array}{l} var[X+Y] = var[X] + var[Y] \\ corr(X,Y) = 0 \\ E[XY] = E[X]E[Y] \end{array}\right\} \text{ if } X, Y \text{ indpt} \quad \text{uncorr.}$$

$$E[(X-E[X])^2]$$

$$var[X_S]$$

$$= var[\sum_i X_i]$$

$$= \sum_i var[X_i]$$

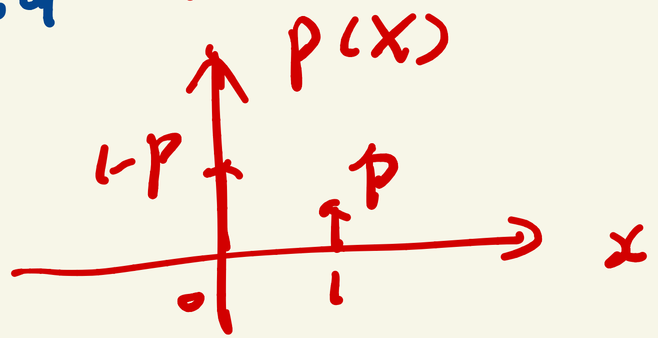
$$= N \cdot p(1-p)$$

X_i are iid R.V.

identical

independent

Bernoulli:



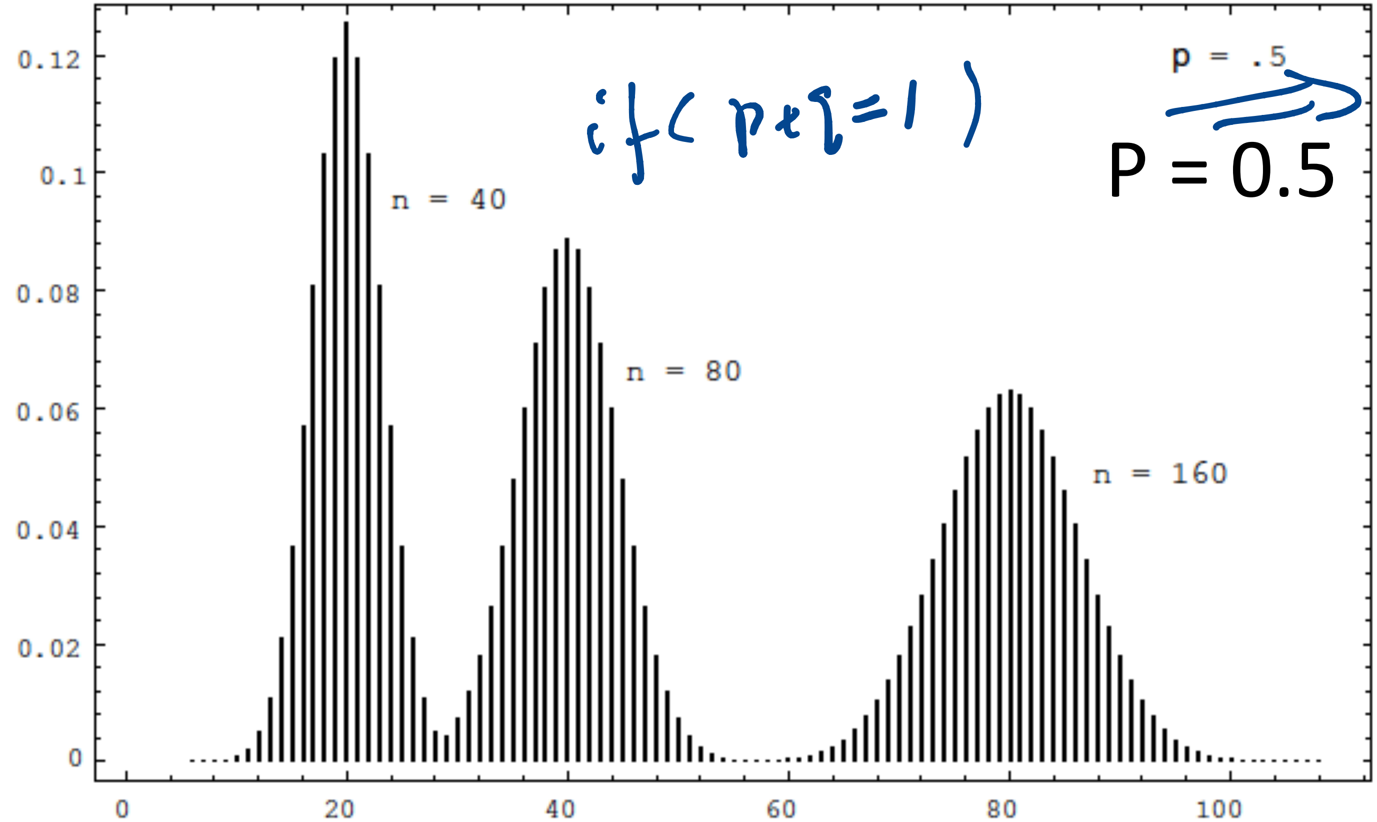
$$var[X_i] = p(1-p)$$

indep^t $\Longrightarrow$ uncorrelated

Binomial distribution

$$(p+q)^n = \sum_{k=0}^{N} \binom{n}{k} p^k q^{(n-k)}$$

if $(p+q=1)$ $\Rightarrow$ RHS $=1$



P = 0.5

Credit: Prof. Grinstead

Binomial distribution: die example

- Let X be the number of sixes in 36 rolls of a fair six-sided die. What is P(X=k) for k =5, 6, 7

$N=36 \quad p=\frac{1}{6}$

$P(X=k)=\binom{36}{k}\cdot p^k\left(1-\frac{1}{6}\right)^{36-k}$

- Calculate E[X] and var[X]

$E[X]=N\cdot p=36\times\frac{1}{6}=6$

$var[X]=N\cdot p(1-p)=36\times\frac{1}{6}\times\frac{5}{6}=5$

Geometric Distribution

H

T H

T T H

$\underbrace{T\ T\ -\ -\ -\ T\ H}_{k \text{ time to see a "H"}}$

$P(\text{toss 1 time}) = p$

$p(\ 2 \text{ times}) = (1-p) \cdot p$

$p(\quad) = (1-p)^2 p$ no N

$\vdots$

$P(k \text{ times}) = (1-p)^{k-1} p$

$k \geq 1$

Geometric distribution

- A discrete random variable X is geometric if

$$P(X = k) = (1-p)^{k-1}p \qquad k \geq 1$$

H, TH, TTH, TTTH, TTTTH, TTTTTH,...

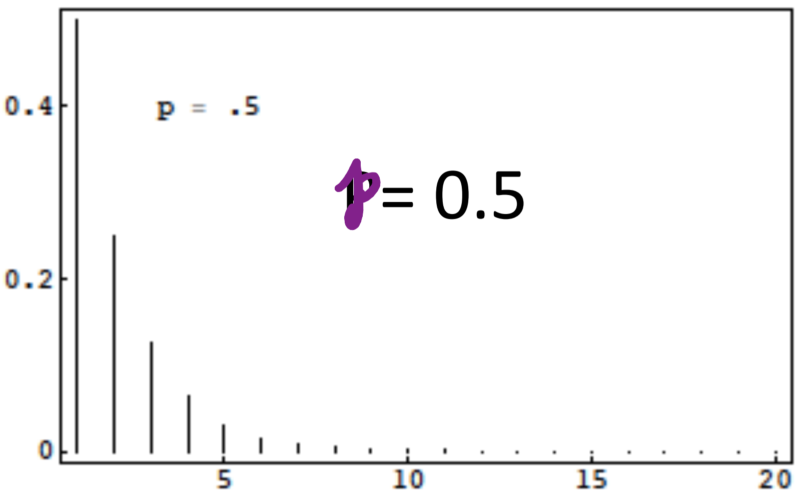
- Expected value and variance

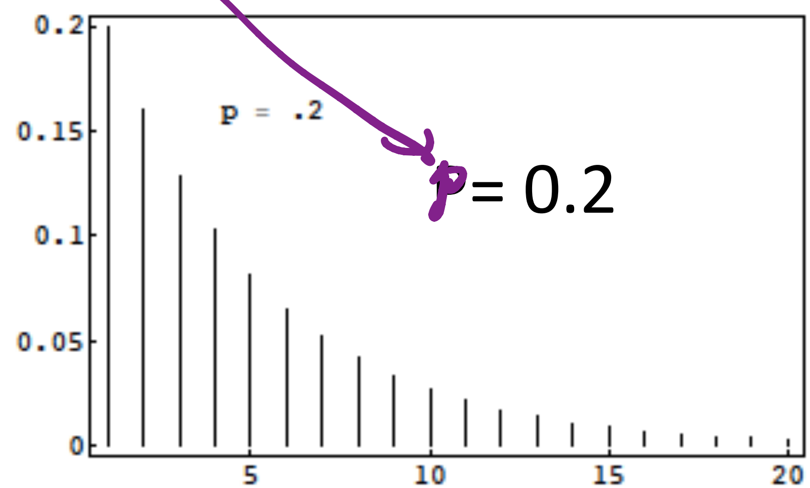
$$E[X] = \frac{1}{p} \quad \& \quad var[X] = \frac{1-p}{p^2}$$

$E(X^2)$

Geometric distribution

$$P(X = k) = (1-p)^{k-1}p \qquad k \geq 1$$





Credit: Prof. Grinstead

Geometric distribution

- Examples:
  - How many rolls of a six-sided die will it take to see the first 6?
  - How many Bernoulli trials must be done before the first 1?
  - How many experiments needed to have the first success?
  - Plays an important role in the theory of **theory of queues**

Derivation of geometric expected value

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

$E[X] = \sum x p(x)$

$k \; p(k)$

$p(k)$

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

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

$$\sum_{n=1} n x^n = \frac{x}{(1-x)^2}$$

Derivation of geometric expected value

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

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

Derivation of geometric expected value

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

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

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

Derivation of geometric expected value

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

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

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

- For we have

this power series:

Derivation of geometric expected value

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

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

$$1-p = x$$

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

- For we have this power series:

$$\sum_{n=1}^{\infty} n x^n = \frac{x}{(1-x)^2}; \quad , \quad |x| < 1$$

Derivation of the power series

$$S(x) = \sum_{n=1}^{\infty} n x^n = \frac{x}{(1-x)^2}; \quad |x| < 1$$

Proof:

$$\frac{S(x)}{x} = \sum_{n=1}^{\infty} n x^{n-1} \; ; \; \sum_{n=0}^{\infty} x^n = \frac{1}{1-x}; \quad |x| < 1$$

$$\int_0^x \frac{S(t)}{t} = \sum_{n=1}^{\infty} x^n = x \cdot \frac{1}{1-x} = \frac{x}{1-x}$$

$$\frac{S(x)}{x} = \left(\frac{x}{1-x}\right)'$$

$$S(x) = \frac{x}{(1-x)^2}$$

Read at home

Geometric distribution: die example

- Let X be the number of rolls of a fair six-sided die needed to see the first 6. What is $P(X = k)$ for k = 1, 2?

$p = \frac{1}{6}$

$P(X=1) = p^1 = \frac{1}{6}$

$P(X=2) = (1-p)p = \frac{5}{6} \times \frac{1}{6} = \frac{5}{36}$

- Calculate E[X] and var[X]

$$E[X] = \frac{1}{p} \quad \& \quad var[X] = \frac{1-p}{p^2}$$

$E[X] = \frac{1}{p} = \frac{1}{\frac{1}{6}} = 6$

$var[X] = \frac{1-p}{p^2} =$

Betting brainteaser

- What would you rather bet on?
  - How many rolls of a fair six-sided die will it take to see the first 6?
  - How many sixes will appear in 36 rolls of a fair six-sided die?
- Why?

Multinomial distribution

- A discrete random variable X is Multinomial if

$$P(X_1 = n_1, X_2 = n_2, ..., X_k = n_k) = \frac{N!}{n_1!n_2!...n_k!} p_1^{n_1} p_2^{n_2} ... p_k^{n_k}$$

$$where \;\; N = n_1 + n_2 + ... + n_k$$

- The event of throwing *N* times the k-sided die to see the probability of getting n_1 X_1, n_2 X_2, n_3 X_3...n_k X_k

Read off-line

Multinomial distribution

- A discrete random variable X is Multinomial if

$$P(X_1 = n_1, X_2 = n_2, ..., X_k = n_k) = \frac{N!}{n_1!n_2!...n_k!} p_1^{n_1} p_2^{n_2} ... p_k^{n_k}$$

$$where \ \ N = n_1 + n_2 + ... + n_k$$

- The event of throwing k-sided die to see the probability of getting n_1 X_1, n_2 X_2, n_3 X_3...

ILLINOIS?

$$\frac{8!}{\underset{I}{3!}\underset{L}{2!}1!1!1!}$$

Read off-line

Multinomial distribution

- Examples
  - If we roll a six-sided die N times, how many of each value will we see?
  - What are the counts of N independent and identical distributed trials?
  - This is very widely used in genetics

Read off-line

Multinomial distribution: die example

- What is the probability of seeing 1 one, 2 twos, 3 threes, 4 fours, 5 fives and 0 sixes in 15 rolls of a fair six-sided die?

solve off-line

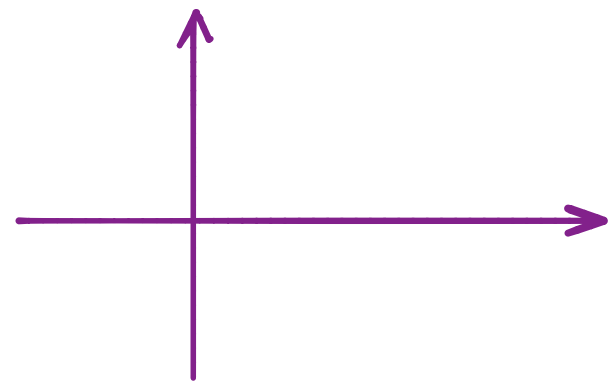
Discrete uniform distribution

- A discrete random variable X is uniform if it takes k different values and

  $X(\omega) = \begin{cases} x_1 \\ x_2 \\ \vdots \\ x_k \end{cases}$

$$P(X = x_i) = \frac{1}{k}$$

For all x_i that X can take



- For example:
  - Rolling a fair k-sided die
  - Tossing a fair coin (k=2)

Discrete uniform distribution

- Expectation of a discrete random variable X that takes k different values uniformly

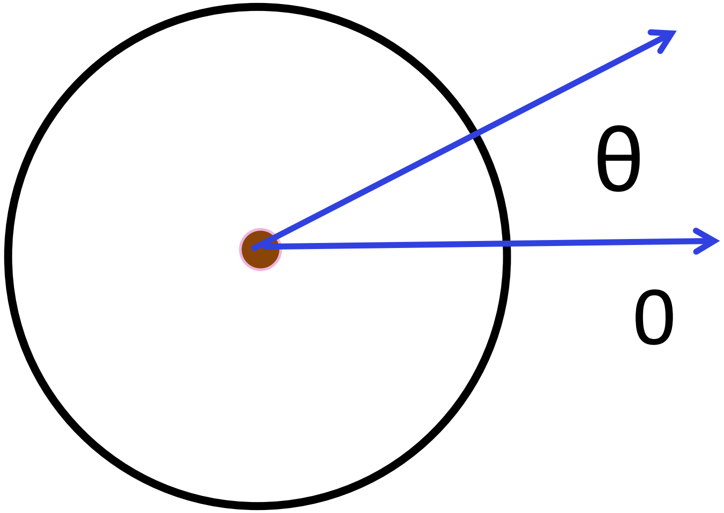
$$E[X] = \frac{1}{k}\sum_{i=1}^{k} x_i$$

- Variance of a uniformly distributed random variable X .

$$var[X] = \frac{1}{k}\sum_{i=1}^{k}(x_i - E[X])^2$$

Example of a continuous random variable

- The spinner

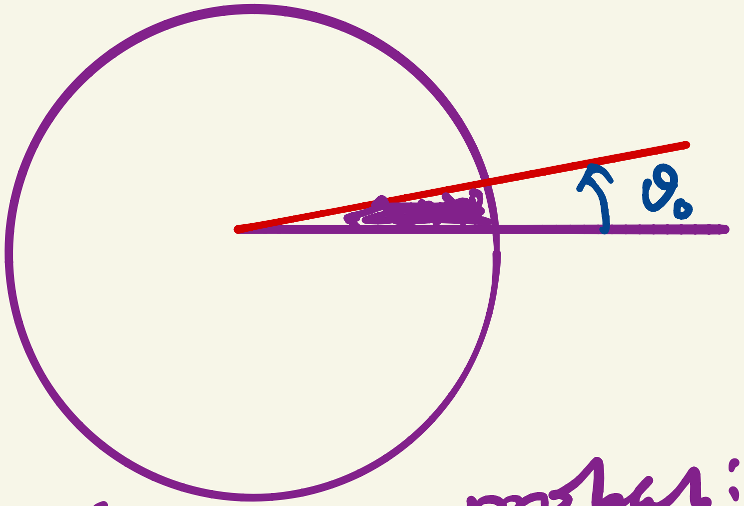


$$\theta \in (0, 2\pi]$$

- The sample space for all outcomes is not countable

* What is the probability of

$P(\theta = \theta_0)$? θ_0 is a constant in $(0, 2\pi]$



* what is the probability of

$P(\underline{0 < \theta \leq \theta_0})$?

$$\frac{\theta_0}{2\pi}$$

Probability density function (pdf)

- For a continuous random variable X, the probability that X=x is essentially zero for all (or most) x, so we can't define $P(X = x)$
- Instead, we define the **probability density function** (pdf) over an infinitesimally small interval dx, $p(x)dx = P(X \in [x, x + dx])$
- For $a < b$ $$\int_a^b p(x)dx = P(X \in [a, b])$$

Properties of the probability density function

- $p(x)$ **resembles** the probability function of discrete random variables in that
  - $p(x) \geq 0$ for all x
  - The probability of X taking all possible values is 1.

$$\int_{-\infty}^{\infty} p(x)dx = 1$$

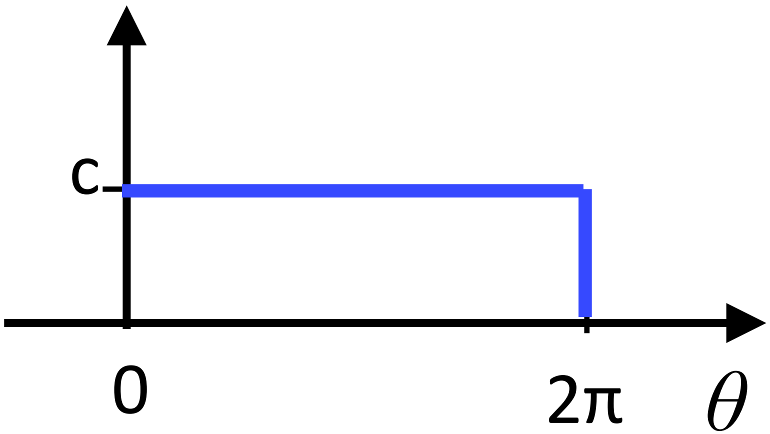
Properties of the probability density function

- $p(x)$ **differs** from the probability distribution function for a discrete random variable in that
  - $p(x)$ is not the probability that $X = x$
  - $p(x)$ can exceed 1

Probability density function: spinner

- Suppose the spinner has equal chance stopping at any position. What's the pdf of the angle θ of the spin position?

$$p(\theta) = \begin{cases} c & if\ \theta \in (0, 2\pi] \\ 0 & otherwise \end{cases}$$



- For this function to be a pdf,

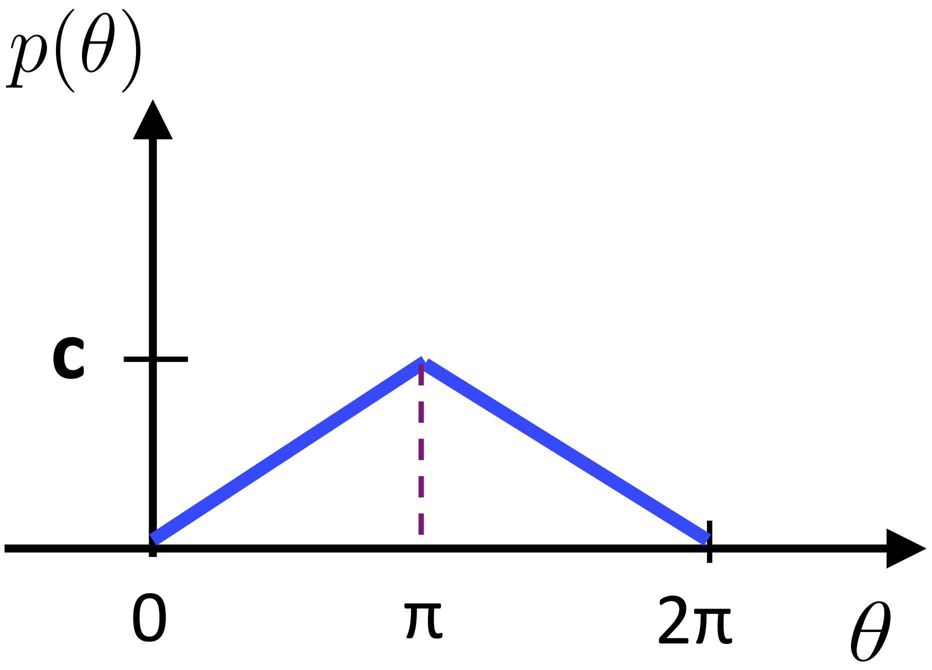
Then $$\int_{-\infty}^{\infty} p(\theta) d\theta = 1$$

Probability density function: spinner

- What the probability that the spin angle θ is within [$\frac{\pi}{12}, \frac{\pi}{7}$]?

Q: Probability density function: spinner

- What is the constant **c** given the spin angle θ has the following pdf?



A. 1
B. 1/π
C. 2/π
D. 4/π
E. 1/2π

Expectation of continuous variables

- Expected value of a continuous random variable X

weight

$$E[X] = \int_{-\infty}^{\infty} x\boxed{p(x)dx}$$

- Expected value of function of continuous random variable $Y = f(X)$

$$E[Y] = E[f(X)] = \int_{-\infty}^{\infty} f(x)p(x)dx$$

Probability density function: spinner

- Given the probability density of the spin angle θ

$$p(\theta) = \begin{cases} \frac{1}{2\pi} & if\ \theta \in (0, 2\pi] \\ 0 & otherwise \end{cases}$$

- The expected value of spin angle is

$$E[\theta] = \int_{-\infty}^{\infty} \theta p(\theta) d\theta$$

Properties of expectation of continuous random variables

- The linearity of expected value is true for continuous random variables.

$$\sum \longrightarrow \int$$

- And the other properties that we derived for variance and covariance also hold for continuous random variable

Q.

- Suppose a continuous variable has pdf

$$p(x) = \begin{cases} 2(1-x) & x \in [0,1] \\ 0 & otherwise \end{cases}$$

What is E[X]?

A. 1/2 B. 1/3 C. 1/4

D. 1 E. 2/3

$$E[X] = \int_{-\infty}^{\infty} xp(x)dx$$

Variance of a continuous variable

Assignments

- Read Chapter 5 of the textbook
- Next time: more classic known probability distributions

Additional References

- Charles M. Grinstead and J. Laurie Snell "Introduction to Probability"
- Morris H. Degroot and Mark J. Schervish "Probability and Statistics"

See you next time

See You!