Distribution Families

Many distribution families

Most natural and important

Theoretical and Practical significance

Distributions

Discrete

Bernoulli

Binomial

Poisson

Geometric

Continuous

Uniform

Exponential

Normal

Discuss

Motivation

Applications

Formulate

Visualize

Examples

Properties

μ

V

σ

Python

Plot and experiment

Show Distribution

Nonnegative

Sum to 1

Blendtec

Tom Dickson

Bernoulli Distribution

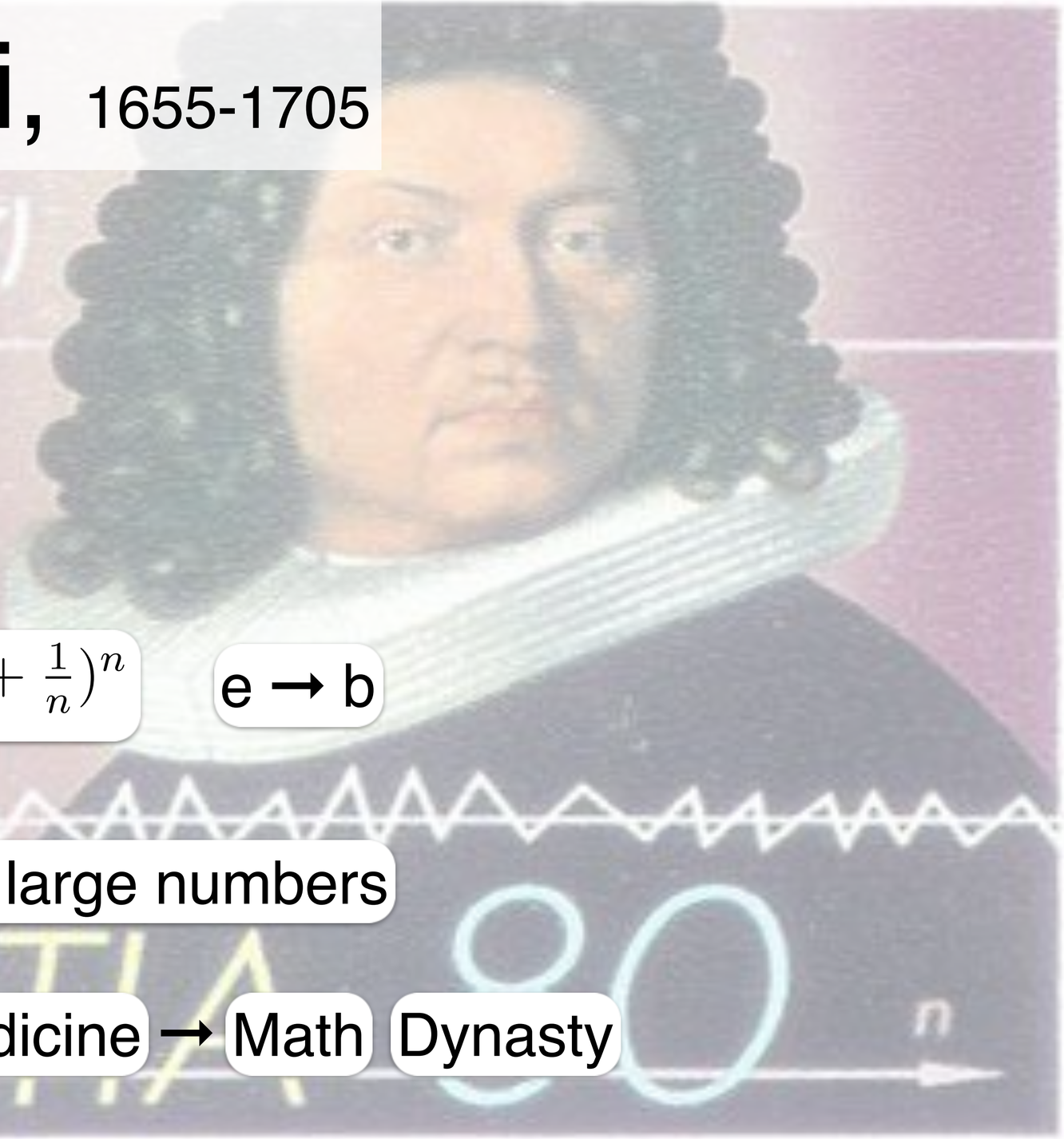
Simplest non-constant distribution

Foundation of many others

μ V σ

Repeated experiments

Jacob Bernoulli, 1655-1705



Theology → mathematics

Calculus Integrals

"Euler" number $e = \lim_{n\to\infty}(1+\frac{1}{n})^n$ e → b

Ars Conjectandi First law of large numbers

Mentored brother Johann Medicine → Math Dynasty

The simplest Distribution

Simplest

One value | 5 | Constant | Always same | Trivial

Simplest non-trivial

Two values

Simplest values

0 and 1

Bernoulli Coin!

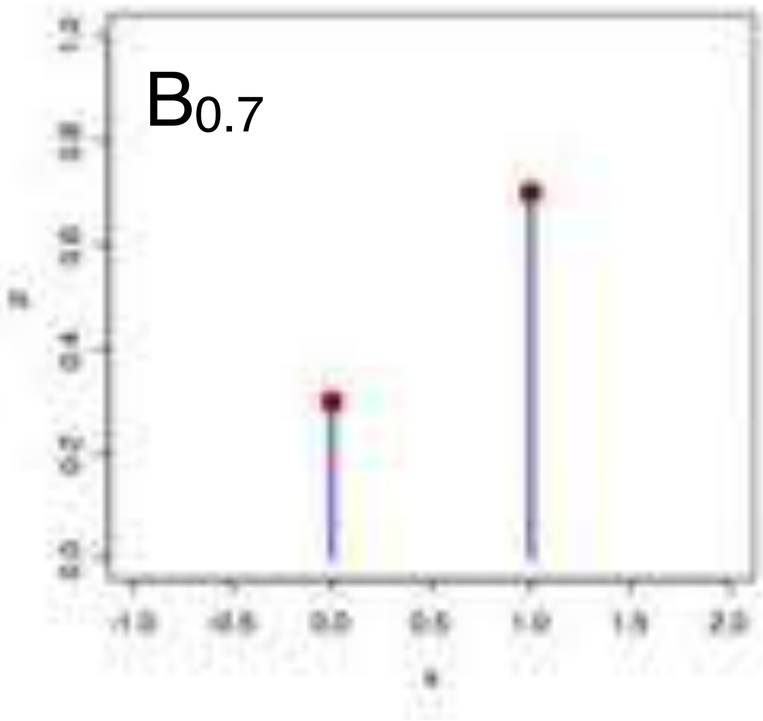
Bernoulli Distribution

B_p $0 \le p \le 1$

failure success

Two values	0	1
Probability	1-p	p

$\overline{p}$ q



$p(0) + p(1) = (1-p)+p = 1$

$X \sim B_p$ Bernoulli random variable, coin, experiment, trial

Who Cares About Two Values?

Everyone!

Binary version of complex events

- Products: 80 good, 20 defective
- Select one, good or not
- $\sim B_{.8}$
- Next child will be a boy
- $\sim B_{.5}$

Generalizes to more complex variables

- Patient has one of three diseases

Repeated trials yield # successes

- Many important distributions
- Binomial, Geometric, Poisson, Normal

Mean

$X \sim B_p$ $\quad$ $p(0) = 1-p$ $\quad$ $p(1) = p$

$EX = \sum p(x) \cdot x = (1-p) \cdot 0 + p \cdot 1 = p$

$X \sim B_{0.8}$ $\quad$ $EX = 0.8$

$EX = P(X=1)$

Fraction of times expect to see 1

Variance

$X \sim B_p$

$EX = p$

Variance

Easy route

$0^2 = 0$

$1^2 = 1$

$X^2 = X$

$E(X^2) = EX = p$

$V(X) = E(X^2) - (EX)^2 = p - p^2 = p(1-p) = pq$

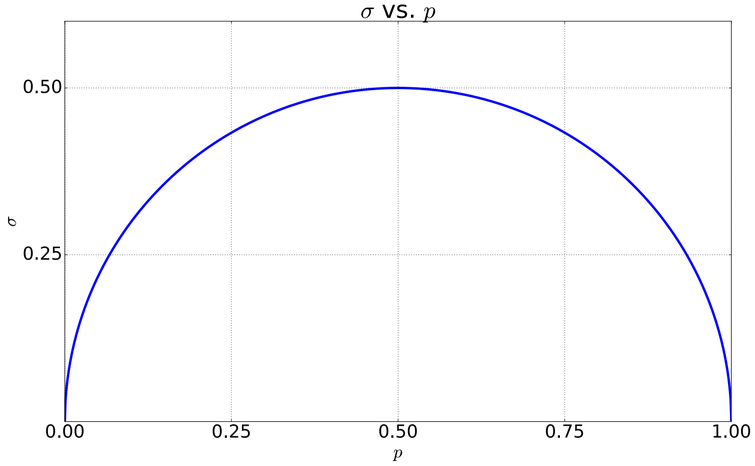
Standard Deviation

$\sigma = \sqrt{pq}$

B_p varies most when $p = ½$

E, V, σ for various p

p	EX	V(X)	σ
0	0	0	0
1	1	0	0
½	½	¼	½



Independent Trials

Much of B_p importance stems from multiple trials

Most common Independent $\perp\!\!\!\perp$

$0 \le p \le 1$ $X_1, X_2, X_3 \sim B_p$ $\perp\!\!\!\perp$

$q \stackrel{\text{def}}{=} 1-p$ $P(110) = p^2 q = P(101) = P(011)$

Generally $X_1, X_2, \ldots, X_n \sim B_p$ $\perp\!\!\!\perp$

$x^n = x_1, x_2, \ldots, x_n \in \{0,1\}^n$ n_0 0's and n_1 1's

$$P(x_1, \ldots, x_n) = p^{n_1} q^{n_0} \qquad P(10101) = p^{n_1} q^{n_0} = p^3 q^2$$

Typical Samples

Distribution	Typical seq.	Description	Probability
B_0	0000000000	constant 0	$1^{10} = 1$
B_1	1111111111	constant 1	$1^{10} = 1$
$B_{0.8}$	1110111011	80% 1's	$0.8^8 \cdot 0.2^2$
$B_{0.5}$	1011010010	50% 1's	0.5^{10}

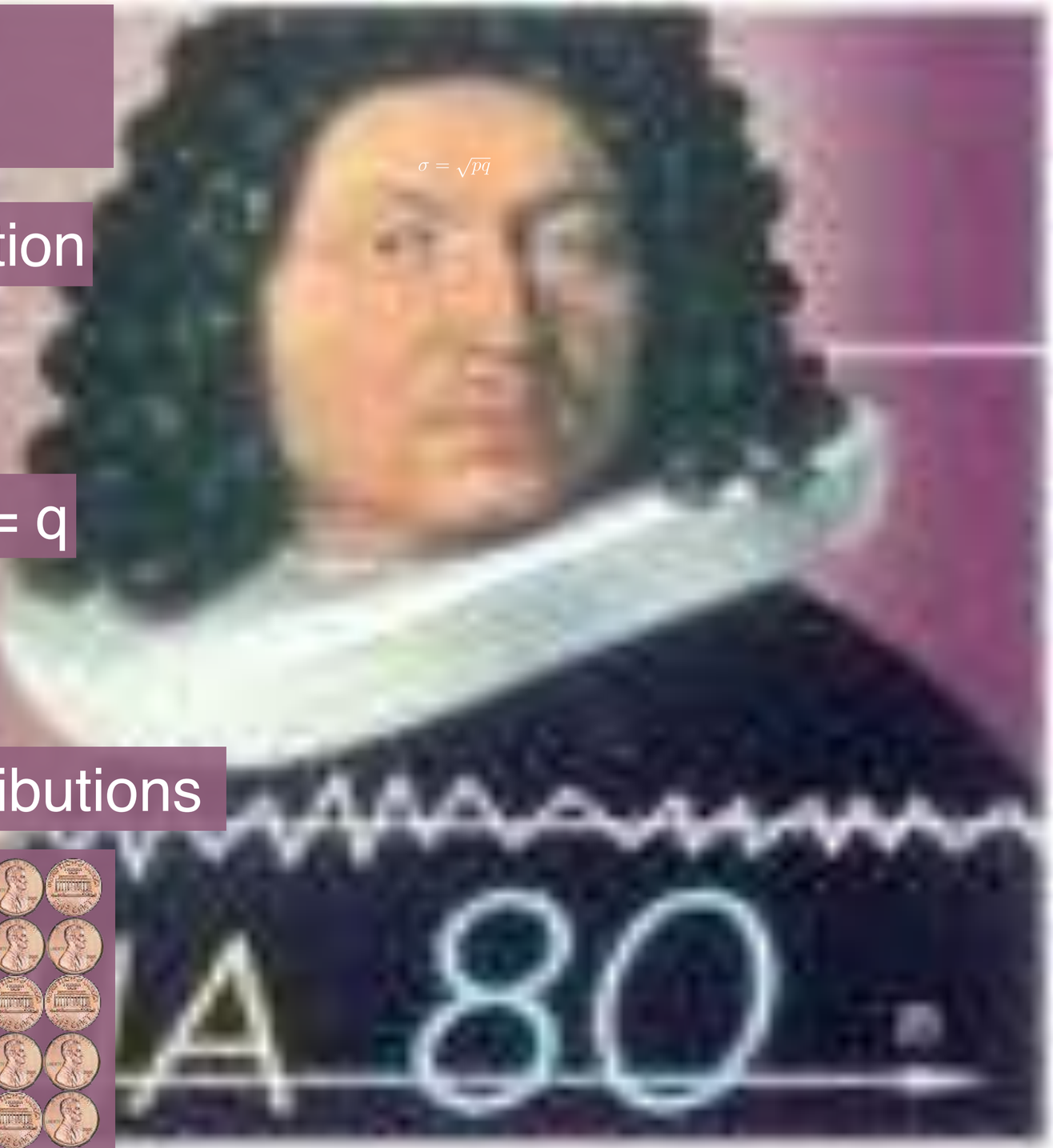
Fair coin flip

Not most probable

Most probable: 1…1

Unlikely to be seen

Bernoulli Distribution



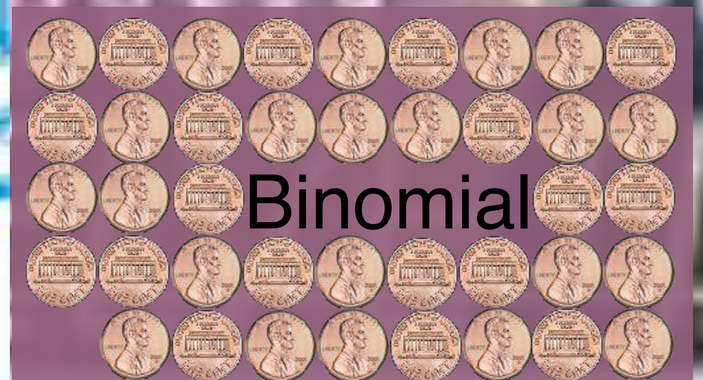
Simplest non-constant distribution

B_p $0 \le p \le 1$

0 and 1 $p(1) = p$ $p(0) = 1-p = q$

$\mu = p$ $V = pq$ $\sigma = \sqrt{pq}$

Foundation of many other distributions



Number of successes in n Bernoulli trials

Useful in many applications

μ, V, σ

Binomial distribution

21 heads

19 tails

The Binomial Distribution

n independent Bernoulli experiments

$\overline{p}$, q

Each “success” with same probability p, “failure" with probability 1 - p

$B_{n,p}$ or $B_{p,n}$ - distribution of # successes

$B_{n,p}$ more common,
use $B_{p,n}$ as generalizes B_p,
natural for Poisson Binomial

n independent coin flips

p(heads) = p

$B_{p,n}$ - distribution of # heads

Applications

Positive responses to a treatment

Faulty components

Rainy days in a month

Delayed flights

Small n

n independent experiments

Success probability p

Failure probability $q = 1 - p$

$b_{p,n}(k)$ - probability of k successes

n = 0

k	$b_{p,0}(k)$
0	1

n = 1

k	$b_{p,1}(k)$
0	q
1	p

$p+q=1$

n = 2

k	seq's	$b_{p,1}(k)$
0	00	q^2
1	01,10	2pq
2	11	p^2

$p^2+2pq+q^2=(p+q)^2=1^2=1$

General n and k

$n \perp\!\!\!\perp B_p$ experiments

successes $0 \le k \le n$

$b_{p,n}(k) = p(k \text{ successes})$

Every k-success sequence: n-k failures, probability $p^k \cdot q^{n-k}$

$\binom{n}{k}$ such sequences

$= \binom{n}{k} p^k q^{n-k}$

Distribution over n+1 values

ΣWILL IT ADD?

$0 \le k \le n$

$p(X = k) = b_{p,n}(k) = \binom{n}{k} p^k q^{n-k}$

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

$$= (p+q)^n$$

$$= 1^n = 1$$

Binomial Theorem

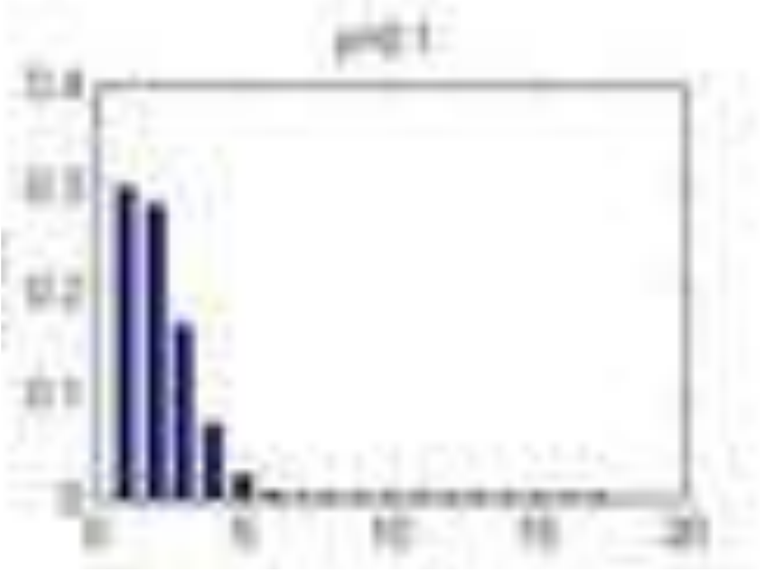
$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

YES IT ADDS!

Sample Distributions

n=18

$b_{p,18}(k)$



Notebook: experiment with different p and n

Multiple Choice

Exam has 6 multiple-choice questions, each with 4 possible answers

Each question, student selects one of the 4 answers randomly

X = # correct answers ~ $B_{1/4,6}$

Passing: ≥ 4 correct answers P(passing) = ?

$$P(4) = \binom{6}{4} \cdot (\tfrac{1}{4})^4 \cdot (\tfrac{3}{4})^2 \approx 0.0329$$

$$P(5) = \binom{6}{5} \cdot (\tfrac{1}{4})^5 \cdot (\tfrac{3}{4})^1 \approx 0.00439$$

$$P(6) = \binom{6}{6} \cdot (\tfrac{1}{4})^6 \cdot (\tfrac{3}{4})^0 \approx 0.000244$$

$$P(\geq 4) = P(4) + P(5) + P(6) \approx 0.03759$$

Interpretation as a Sum

$B_{p,n}$ a sum of n B_p

$X_1, \ldots, X_n \sim B_p \perp\!\!\!\perp$

$$X \stackrel{\text{def}}{=} \sum_{i=1}^{n} X_i$$

P(X = k) = P(exactly k of $X_1, \ldots, X_n$ are 1) $= \binom{n}{k} p^k q^{n-k} = b_{p,n}(k)$

$X \sim B_{p,n}$

Apply to mean and variance

Mean and Variance

$X \sim B_{p,n}$

$$X = \sum_{i=1}^{n} X_i \qquad X_1, \ldots, X_n \sim B_p \perp\!\!\!\perp$$

$$E(X) = E(\sum_{i=1}^{n} X_i) \overset{LE}{=} \sum EX_i \overset{B_p}{=} \sum p = np$$

$$V(X) = V(\sum_{i=1}^{n} X_i) \overset{\perp\!\!\!\perp}{=} \sum V(X_i) \overset{B_p}{=} \sum pq = npq$$

$$\sigma = \sqrt{npq}$$

Multiple Choice

Exam has 6 multiple-choice questions, each with 4 possible answers

For each question, student selects one of the 4 answers randomly

X = # correct answers $\sim B_{1/4,6}$

Mean $EX = np = 6 \cdot 1/4 = 1.5$

Standard deviation $\sigma = \sqrt{npq} = \sqrt{6 \cdot \frac{1}{4} \cdot \frac{3}{4}} = \frac{\sqrt{18}}{4}$

Why Vote

For simplicity odd # voters: 2n + 1

Each equally likely D or R

P(voter makes a difference) = P(other 2n voters equally split)

$$b_{p,n}(k) = \binom{n}{k} p^k q^{n-k}$$

$$= \binom{2n}{n} \frac{1}{2^n} \cdot \frac{1}{2^n}$$

$$= \frac{(2n)!}{n! \cdot n! \cdot 2^n \cdot 2^n}$$

$$\approx \frac{\sqrt{2\pi \cdot 2n}(\frac{2n}{e})^{2n}}{(\sqrt{2\pi n}(\frac{n}{e})^n)^2 2^{2n}}$$

$$= \frac{1}{\sqrt{\pi n}}$$

Stirling's Approximation

$$n! \sim \sqrt{2\pi n}(\tfrac{n}{e})^n$$

Poisson Binomial

Generalizes the binomial distribution

$n \geq 1$	Binomial	$B_{p,n}$	For $1 \leq i \leq n$	$X_i \sim B(p)$	$\perp\!\!\!\perp$	$X = \sum_{i=1}^{n} X_i$
	Poisson Binomial	$PB_{p_1, \ldots, p_n}$		$X_i \sim B(p_i)$		

$PB(1/4, 2/3)$ $X_1 \sim B(1/4)$ $X_2 \sim B(2/3)$ $\perp\!\!\!\perp$

X_1	X_2	P	X
0	0	$3/4 \cdot 1/3 = 1/4$	0
0	1	$3/4 \cdot 2/3 = 1/2$	1
1	0	$1/4 \cdot 1/3 = 1/12$	1
1	1	$1/4 \cdot 2/3 = 1/6$	2

X	P(x)
0	$1/4$
1	$7/12$
2	$1/6$

Expectation and Variance

$$X \sim PB_{p_1,p_2,\ldots,p_n} \qquad X = \sum_{i=1}^{n} X_i \qquad X_i \sim B_{p_i}$$

$$E(X) = E(\sum_{i=1}^{n} X_i) \overset{\text{LE}}{=} \sum_{i=1}^{n} EX_i \overset{\text{Bpi}}{=} \sum_{i=1}^{n} p_i$$

$$V(X) = V(\sum_{i=1}^{n} X_i) \overset{\perp\!\!\!\perp}{=} \sum_{i=1}^{n} V(X_i) \overset{\text{Bpi}}{=} \sum_{i=1}^{n} p_i(1-p_i)$$

p(k) No closed form Computationally Homework

$B_{p,n}$ Number of successes in n B_p trials

$$b_{p,n}(k) = \binom{n}{k} p^k \overline{p}^{n-k}$$

$\mu = np$ $V = np\bar{p}$ $\sigma = \sqrt{np\overline{p}}$

Why vote

Binomial distribution

Coin Flips

Most basic convergence to average is B(p)

Flip n B(p) coins, average # 1's will approach np

Probability of a sequence with k 1's and n-k 0's is $p^k q^{n-k}$

Wolog assume p>0.5, then most likely is 1^n

Yet by WLLN with probability →1 we see roughly pn 1's and qn 0's

Why do we observe these sequences and not the most likely ones?

Strength in #s. # sequences of a given composition increases near 1/2

pn balances # x probability.

Definition

Applications

Motivation

Derivation

μ, σ

Example

Poisson distribution

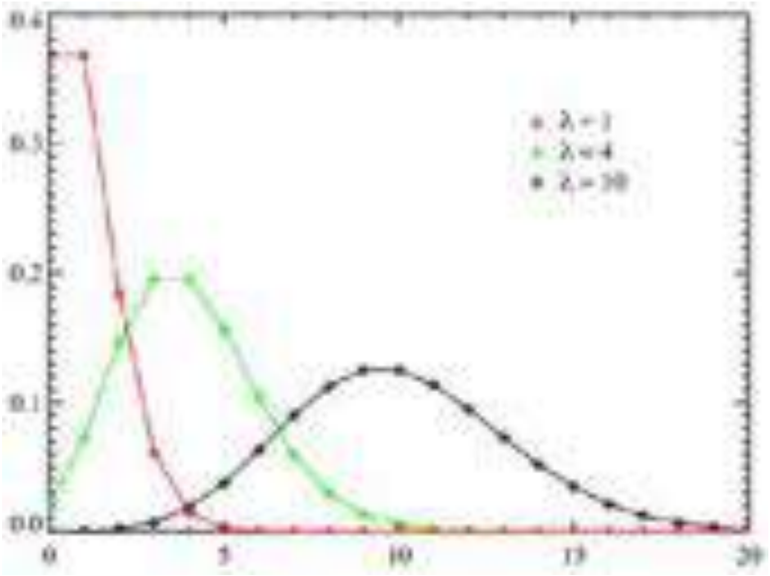
The Poisson Distribution

Parameter $\lambda \geq 0$ (some > 0)

Support $\mathbb{N}$

PMF

$$P_\lambda(k) = e^{-\lambda}\frac{\lambda^k}{k!}$$



More in notebook

Significance Approximates $B_{p,n}$ for large n and small p so that $np = \lambda$ is moderate

We are Poisson

P_λ approximates $B_{p,n}$ for small p, large n

Numerous applications

$450M

- # People clicking ad
- Responses to spam
- Rare-disease infections
- Daily 911 calls
- Daily store customers
- Gallery purchasing customers
- Flight no shows
- Typos in a page

Small k

k

	λ	$P_\lambda(k)$	0	1	2	3
$\frac{\lambda^k}{e^\lambda}$	General	$e^{-\lambda}\frac{\lambda^k}{k!}$	$\frac{1}{e^\lambda}$	$\frac{\lambda}{e^\lambda}$	$\frac{\lambda^2}{2e^\lambda}$	$\frac{\lambda^3}{6e^\lambda}$
$\downarrow$ $\frac{1}{e}$	1	$\frac{1}{ek!}$	$\frac{1}{e}$	$\frac{1}{e}$	$\frac{1}{2e}$	$\frac{1}{6e}$
$\frac{2^k}{e^2}$	2	$\frac{2^k}{e^2k!}$	$\frac{1}{e^2}$	$\frac{2}{e^2}$	$\frac{2}{e^2}$	$\frac{4}{3e^2}$
0^k	0	$\frac{0^k}{k!}$	1	0	0	0

Binomial Approximation

P_λ approximates $B_{p,n}$ for $\lambda = pn$, when $n \gg 1 \gg p$

$$B_{p,n}(k) = \binom{n}{k} p^k q^{n-k} \qquad q = 1 - p$$

$$p = \frac{\lambda}{n}$$

$$= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{n^{\underline{k}}}{k!} \cdot \frac{\lambda^k}{n^k} \cdot \frac{(1 - \frac{\lambda}{n})^n}{(1 - \frac{\lambda}{n})^k}$$

Fix k, fixe λ while n ↗ and p ↘

Derive Poisson

Limit of Binomial

$$B_{p,n}(k) = \frac{\lambda^k}{k!} \cdot \underbrace{\frac{n^{\underline{k}}}{n^k}}_{1} \cdot \frac{\underbrace{(1-\frac{\lambda}{n})^n}_{e^{-\lambda}}}{\underbrace{(1-\frac{\lambda}{n})^k}_{1}} \to e^{-\lambda}\,\frac{\lambda^k}{k!}$$

λ and k fixed, n $\to \infty$

$$\lambda = p \cdot n$$

① $\frac{n^{\underline{k}}}{n^k} = \underbrace{\frac{n}{n}}_{1} \cdot \underbrace{\frac{(n-1)}{n}}_{1} \cdot \ldots \cdot \underbrace{\frac{(n-k+1)}{n}}_{1} \to 1$ fixed # (k) terms, each → 1

② $(1-\frac{\lambda}{n})^k \to 1$ fixed # (k) terms, each → 1

③ $(1-\frac{\lambda}{n})^n = ((1-\frac{\lambda}{n})^{\frac{n}{\lambda}})^\lambda \to (e^{-1})^\lambda = e^{-\lambda}$

increasing # terms, each → 1 $(1-\frac{1}{m})^m \to e^{-1}$

ΣWILL IT ADD?

$$P_\lambda(k) = e^{-\lambda}\frac{\lambda^k}{k!}$$ k≥0

≥ 0 ✅

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

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

YES IT ADDS!

Mean and Variance

P_λ approximates $B_{p,n}$ for $\lambda = np$ when $n \gg 1 \gg p$

	μ	V
$B_{p,n}$	np	npq
P_λ	λ	λ

$\longleftarrow$ Expect

Calculate next

Observation

$$\frac{d}{d\lambda}\lambda^k = k\lambda^{k-1} = \frac{k}{\lambda}\lambda^k$$

$$\frac{d^2}{d\lambda^2}\lambda^k = k^{\underline{2}}\lambda^{k-2} = \frac{k^{\underline{2}}}{\lambda^2}\lambda^k$$

$$\frac{d^r}{d\lambda^r}\lambda^k = k^{\underline{r}}\lambda^{k-r} = \frac{k^{\underline{r}}}{\lambda^r}\lambda^k$$

$$k^{\underline{r}}\lambda^k = \lambda^r \frac{d^r}{d\lambda^r}\lambda^k$$

Falling Moments

$$X \sim P_\lambda$$

$$k^{\underline{r}}\lambda^k = \lambda^r \frac{d^r}{d\lambda^r}\lambda^k$$

$$E(X^{\underline{r}}) = \sum_{k=0}^{\infty} k^{\underline{r}} P_\lambda(k) = \sum_k k^{\underline{r}} e^{-\lambda}\frac{\lambda^k}{k!}$$

$$= e^{-\lambda}\sum_k k^{\underline{r}}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_k \frac{\lambda^r}{k!}\frac{d^r}{d\lambda^r}\lambda^k$$

$$= e^{-\lambda}\lambda^r\frac{d^r}{d\lambda^r}\sum_k \frac{\lambda^k}{k!} = e^{-\lambda}\lambda^r\frac{d^r}{d\lambda^r}e^{\lambda}$$

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

$$EX = EX^{\underline{1}} = \lambda \qquad EX(X-1) = EX^{\underline{2}} = \lambda^2$$

Mean and Variance

$$EX = EX^{\underline{1}} = \lambda$$ ✅

$$EX(X-1) = EX^{\underline{2}} = \lambda^2$$

$$E(X^2) = E(X(X-1)+X) = E(X(X-1)) + E(X) = \lambda^2 + \lambda$$

$$V(X) = E(X^2) - (EX)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$ ✅

$$\sigma = \sqrt{\lambda}$$

Small relative to the mean

Approximation Example

Factory produces 200 items, each defective with probability 1%

P(3 defective)?

Binomial (precise) $B_{0.01,200}(3) = \binom{200}{3}(0.01)^3(0.99)^{197} \approx 0.181$

Poisson (approximation) $\lambda = 200 \cdot 0.01 = 2$ $P_2(3) = e^{-2}\frac{2^3}{3!} \approx 0.18$

P(some defective)?

$B_{0.01,200}(0) = \binom{200}{0}(0.99)^{200} \approx 0.134$ $P_2(0) = e^{-2}\frac{2^0}{0!} = e^{-2} \approx 0.135$

$B_{0.01,200}(\geq 1) = 1 - 0.134 \approx 0.866$ $P_2(\geq 1) = 1 - 0.135 \approx 0.865$

$$P_\lambda(k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

$\lambda \geq 0$ $k \geq 0$

Approximates $B_{p,n}$ for $\lambda = np$, when $n \gg 1 \gg p$

# of ad clicks, rare diseases, production defects

$\mu = \lambda$ $V = \lambda$ $\sigma = \sqrt{\lambda}$

Geometric Distribution

Poisson

distribution

Geometric distributions

Back to Basics

Independent B_p coin flips

$p(1) = p$

$p(0) = 1-p \stackrel{\text{def}}{=} q$

Two Derived Distributions

Binomial	$B_{p,n}$	n flips, # 1's
Geometric	G_p	# flips till first 1

Flips	X
10101	1
01011	2
00011	4

n	$X_1,\ldots,X_n$	p(n)
1	$X_1 = 1$	p
2	$X_1 = 0 \quad X_2 = 1$	qp
3	$X_1 = X_2 = 0 \quad X_3 = 1$	q^2p
n	$X_1 = \ldots = X_{n-1} = 0 \quad X_n = 1$	$q^{n-1}p$

Geometric Distribution

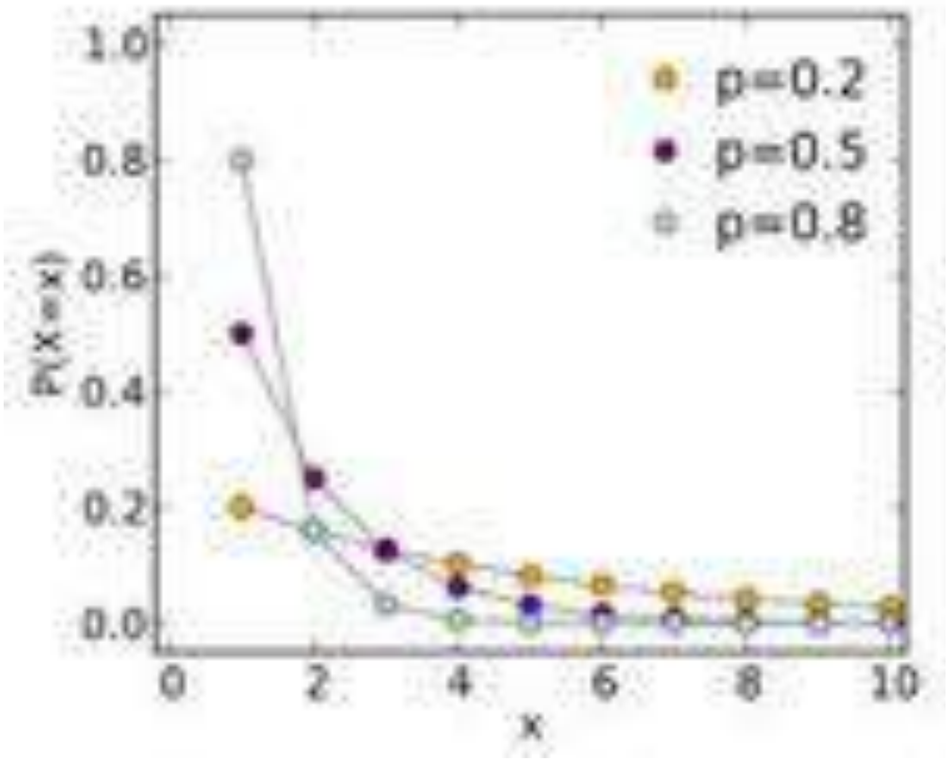
G_p $0 < p \leq 1$

$p(n) = q^{n-1}p \stackrel{\text{def}}{=} g_p(n)$ $n \geq 1$

Notebook Plot different p

Observations $p \neq 0$

n can be arbitrarily high



Who’s Geometric

30 years ago

Thief trying to find a matching key

Trials to hit a target

Attempts till success

Till failure

Nowadays

Σ Will it ADD?

$$P(n) = pq^{n-1} \qquad n \geq 1 \qquad q = 1 - p$$

$$(1 + q + q^2 + \ldots)(1 - q) = 1 + \cancel{q} + \cancel{q^2} + \ldots$$
$$- \cancel{q} - \cancel{q^2} - \ldots$$

$$\sum_{i=0}^{\infty} q^i = \frac{1}{1 - q}$$

YES IT ADDS!

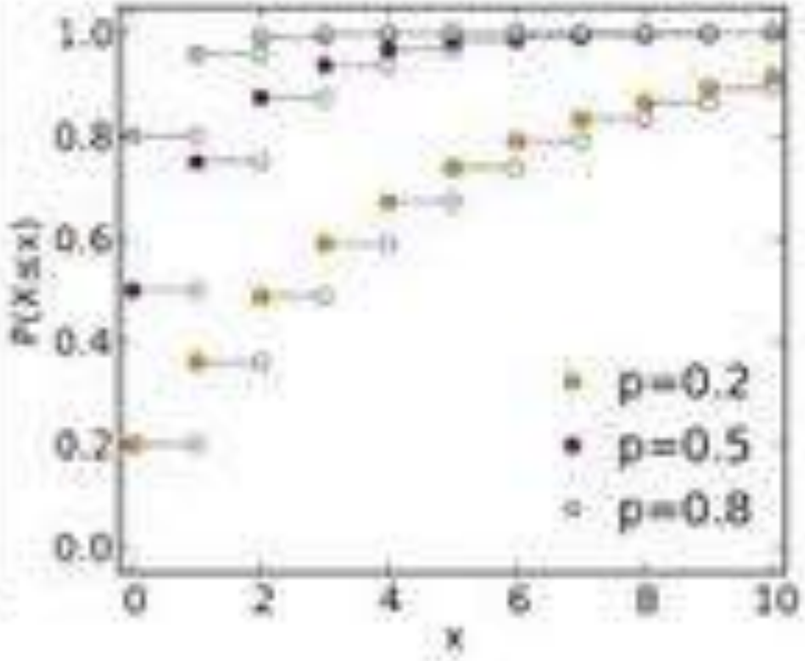
$$\sum_{n=1}^{\infty} p(1-p)^{n-1} = p \sum_{i=0}^{\infty} (1-p)^i = p \cdot \frac{1}{1 - (1-p)} = \frac{p}{p} = 1$$

CDF

$n \in \mathbb{N}$

$P(X>n) = P(X_1 = \ldots = X_n = 0) = q^n$

$$F(n) = P(X \leq n) = 1 - P(X > n) = 1 - q^n$$



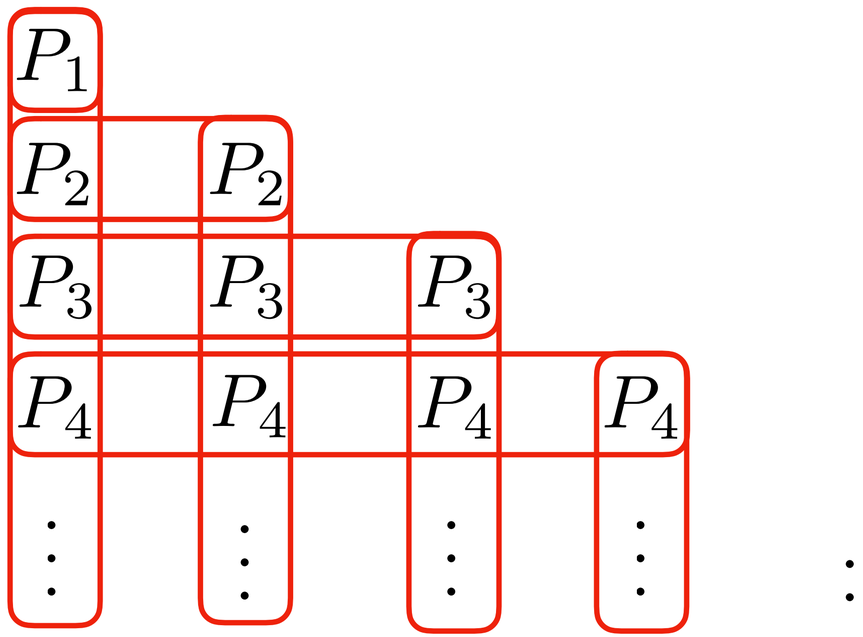
Expectation via “Right” CDF

$x \in \mathbb{N} \qquad P_k = P(X = k)$

$$E(X) = \sum_{k=0}^{\infty} k P_k$$

$$= P_1 + 2P_2 + 3P_3 + \dots$$

$$= P(X \geq 1) + P(X \geq 2) + P(X \geq 3) + \dots$$



Geometric distribution

$$EX = \sum_{k=1}^{\infty} P(X \geq k) = \sum_{i=0}^{\infty} P(X > i) = \sum_{i=0}^{\infty} q^i = \frac{1}{1-q} = \frac{1}{p}$$

Variance

$$EX(X-1) = \sum_{n=1}^{\infty} n(n-1) \cdot P(X=n) = p \sum_{n=2}^{\infty} n(n-1)q^{n-1}$$

$$= pq \sum_{n=2}^{\infty} \frac{d^2}{dq^2} q^n = pq \frac{d^2}{dq^2} \sum_{n=2}^{\infty} q^n$$

$$= pq \frac{d^2}{dq^2} \left(\frac{1}{1-q} - 1 - q \right)$$

$$= pq \frac{2}{(1-q)^3} = \frac{2q}{p^2}$$

$$\left(\frac{1}{1-q} \right)' = \frac{1}{(1-q)^2}$$

$$\left(\frac{1}{(1-q)^2} \right)' = \frac{2}{(1-q)^3}$$

$$EX^2 = EX(X-1) + EX = \frac{2q}{p^2} + \frac{1}{p} = \frac{2q+p}{p^2} = \frac{1+q}{p^2}$$

$$V(X) = EX^2 - (EX)^2 = \frac{1+q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2} \qquad \sigma = \frac{\sqrt{q}}{p}$$

Fair Coin

$$X \sim G_{\frac{1}{2}}$$

$$P(X = k) = g_{0.5}(k) = (\tfrac{1}{2})^{k-1} \cdot \tfrac{1}{2} = \frac{1}{2^k}$$

$$EX = \frac{1}{p} = 2$$

$$VX = \frac{q}{p^2} = 2$$

Memoryless

A distribution over $\mathbb{P} = \{1,2,..\}$ is memoryless if for all $n \geq 0$, $m > 1$

$$P(X=n+m \mid X>n) = P(X=m)$$

$P(X=12 \mid X>10) = P(X=2)$

Geometric → Memoryless

$$P(X = n + m | X > n) = \frac{P(X{=}n{+}m, X{>}n)}{P(X{>}n)}$$

$$= \frac{P(X{=}n{+}m)}{P(X{>}n)}$$

$$= \frac{p \cdot q^{n+m-1}}{q^n}$$

$$= p \cdot q^{m-1}$$

$$= P(X = m)$$

All geometric distributions are memoryless

Memoryless → Geometric

Any discrete memoryless distribution over $\mathbb{P}$ is geometric

Let $p \stackrel{\text{def}}{=} P(X=1), \qquad q \stackrel{\text{def}}{=} 1-p = P(X>1)$

$$\forall\, n \geq 1, \quad P(X=n+1) = P(X>1 \;\wedge\; X=n+1)$$

$$= P(X>1) \cdot P(X=n+1 | X>1)$$

$$= q \cdot P(X=n)$$

Hence $P(X=2)=qp$, $P(X=3)=q^2p$, ….

$P(X=n)=q^{n-1}p$ Geometric

r Successes

$$\text{Geometric } P(X = n) = P(\text{first success at n'th trial})$$

$$P(\text{r'th success at n'th trial}) = P(\text{r - 1 successes in n - 1 trials}) \cdot P(\text{n'th trial is success})$$

$$n \geq r$$

$$b_{n-1,p}(r-1) \qquad \text{p}$$

$$= \binom{n-1}{r-1} p^{r-1} q^{n-r} p$$

$$= \binom{n-1}{r-1} p^{r} q^{n-r}$$

$$r = 1 \rightarrow pq^{n-1} = g_p(n)$$

Negative binomial distribution

Geometric Distribution

$$P(n) = pq^{n-1} \qquad n \geq 1 \qquad q = 1 - p$$

Memoryless

$$E(X) = \frac{1}{p}$$

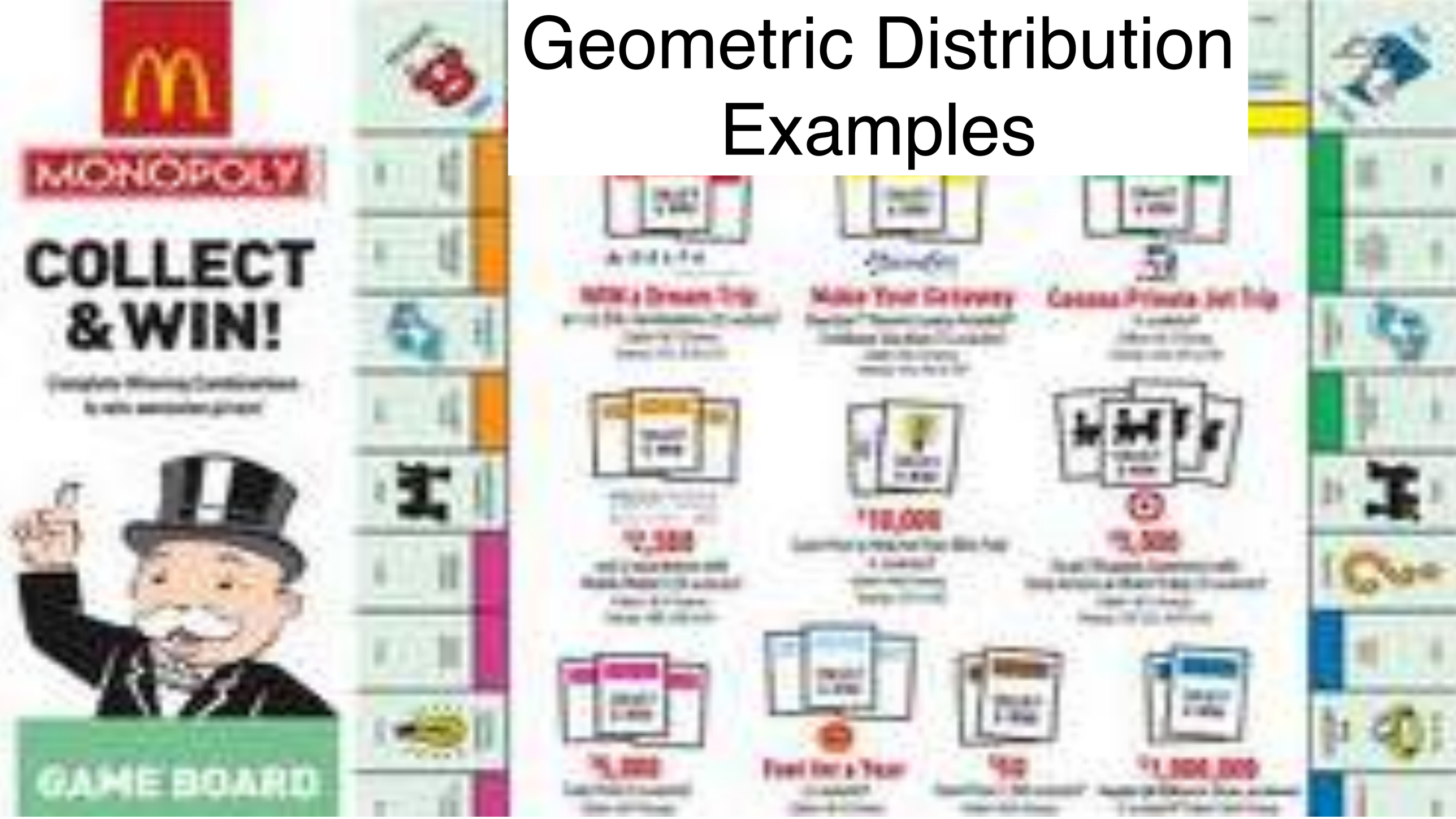
$$V(X) = \frac{q}{p^2} \qquad \sigma = \frac{\sqrt{q}}{p}$$

$$P(\text{r'th success at n'th trial}) = \binom{n-1}{r-1} p^r q^{n-r}$$

Next:

Examples

Geometric Distribution Examples



Startup Statistics

P(startup success) = 20%, independent of previous attempts

Expected # startups till first success

$X \sim G_{0.2}$ $\qquad$ $E(X) = \frac{1}{.2} = 5$

Home-Grown Entrepreneur

One of first three startups succeeds

Dad will fund up to three startups $\quad$ P(success)?

$$P(X \leq 3) = F(3) = 1 - (0.8)^3 \approx 0.49$$

Cry Uncle

Even wealthier uncle funds next three startups (4,5,6)

P(success with uncle if dad's help did not suffice)?

$$P(X \in \{4,5,6\}|X > 3) = P(4|X > 3) + P(5|X > 3) + P(6|X > 3)$$
$$= P(1) + P(2) + P(3) = P(X \leq 3) \approx 49\%$$

P(success with uncle)? 1,2,3 failed but one of 4, 5, 6 succeeded

$$P(3 < X \leq 6) = P(X > 3 \cap X \leq 6) = P(X > 3) \cdot P(x \leq 6|x > 3)$$
$$= (0.8)^3 \cdot 0.49 \approx 25\%$$

P(X_1,X_2,X_3 failed) = q^3

$$P(3 < X \leq 6) = F(6) - F(3) = (1 - 0.8^6) - (1 - 0.8^3)$$
$$= 0.8^3 - 0.8^6 \approx 25\%$$

Foreign-Born Entrepreneur

X - time to first success p=0.2

r^X - fraction of company you keep r=0.5

$$E(r^X) = \sum_{k=1}^{\infty} r^k P(X=k) = \sum_{k=1}^{\infty} pq^{k-1} r^k = pr \sum_{i=0}^{\infty} (qr)^i$$

$$= \frac{pr}{1-qr} = \frac{0.2 \cdot 0.5}{1-0.8 \cdot 0.5} = \frac{0.1}{1-0.4} = \frac{0.1}{0.6} \approx 16.67\%$$

Coupon Collector Problem

Pre GROUPON

n coupons

Each item contains one coupon selected uniformly

Collect all coupons, get a prize

How many items need to buy to collect all?

Expectation

X - # items to collect all coupons

n = 3 Items 1 2 3 4 5 6 7 X = 7 EX?

Coupon $\underbrace{2}\ \underbrace{2\ \ 3}\ \underbrace{2\ \ 3\ \ 3\ \ 1}$

$X_1 = 1$ $X_2 = 3 - 1 = 2$ $X_3 = 7 - 3 = 4$

X_i - # items to get i^{th} coupon after getting i - 1 coupons

$X = X_1 + X_2 + X_3$

$7 = 1 + 2 + 4$

$$\left.\begin{array}{l} X_1 = 1 \\ X_2 \sim G_{2/3} \\ X_3 \sim G_{1/3} \end{array}\right\} \perp\!\!\!\perp$$

General n

$$X_i \sim G(\tfrac{n-(i-1)}{n}) = G(\tfrac{n-i+1}{n})$$

$$EX_i = \tfrac{n}{n-i+1}$$

$$X = \sum_{i=1}^{n} X_i$$

$$EX = \sum_{i=1}^{n} EX_i = \sum_{i=1}^{n} \tfrac{n}{n-i+1} = \tfrac{n}{n} + \tfrac{n}{n-1} + \tfrac{n}{n-2} + \ldots + \tfrac{n}{1}$$

$$= n(\tfrac{1}{1} + \tfrac{1}{2} + \ldots + \tfrac{1}{n}) = nH_n \approx n \ln n + 0.577n$$

Harmonic Sum $\quad H_n = \tfrac{1}{1} + \tfrac{1}{2} + \ldots + \tfrac{1}{n} \to \ln n + 0.577\ldots$

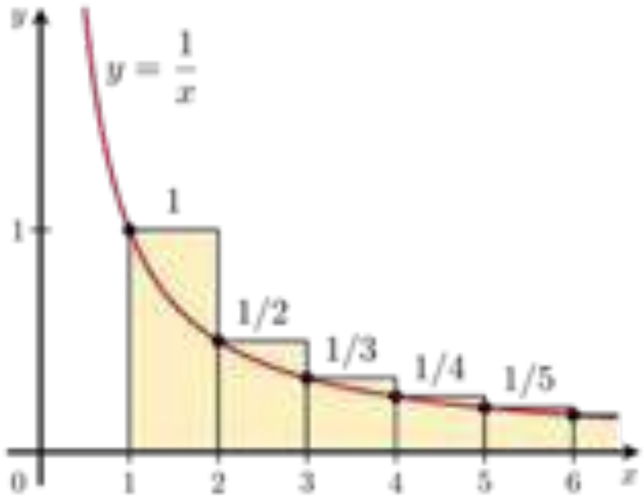
Harmonic Sum

$$H_n > \int_1^{n+1} \frac{1}{x} dx$$

$$= \ln x \Big|_1^{n+1} = \ln(n+1)$$

$$H_n \leq 1 + \int_1^n \frac{1}{x} dx = 1 + \ln x \,\Big|_1^n = 1 + \ln n$$

$$H_n \to \ln n + 0.577\ldots$$



Variance

$X \sim G(P)$

$V(X) = \frac{1-p}{p^2} \leq \frac{1}{p^2}$

$$V(X) = V\left(\sum_{i=1}^{n} X_i\right)$$

$$\overset{\perp\!\!\!\perp}{=} \sum_{i=1}^{n} V(X_i)$$

$$\leq \sum_{i=1}^{n} \frac{1}{(\frac{n-i+1}{n})^2}$$

$$= n^2\left(\frac{1}{n^2} + \frac{1}{(n-1)^2} + \ldots + \frac{1}{1^2}\right)$$

$$\leq \frac{\pi^2}{6} n^2$$

$\sigma \leq \frac{\pi}{\sqrt{6}} n$

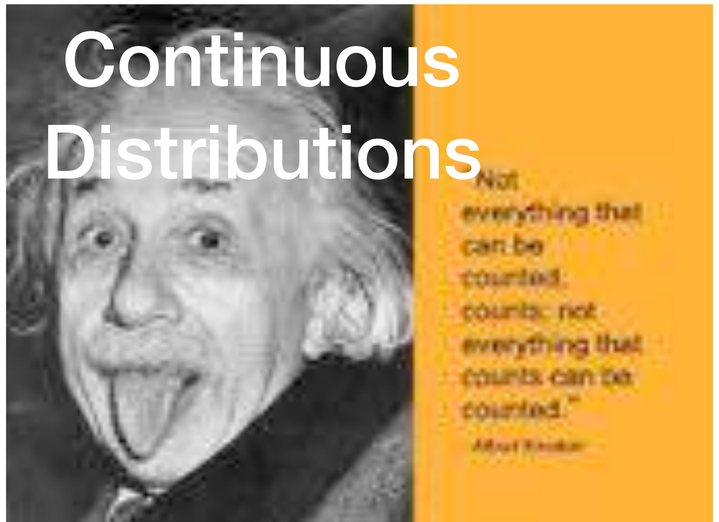
Summary

Geometric-distribution examples

Coupon collector problem

Discrete distribution families

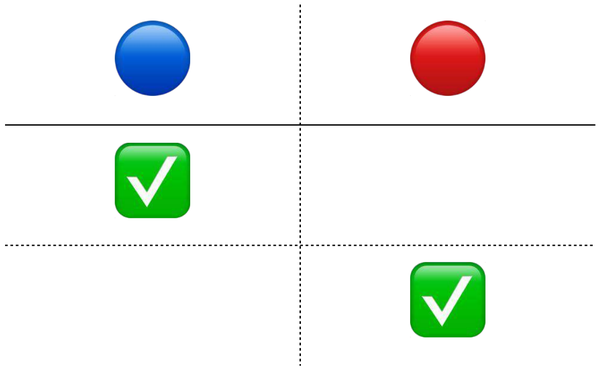
Bernoulli, Binomial, Poisson, Geometric



Hypergeometric Distribution

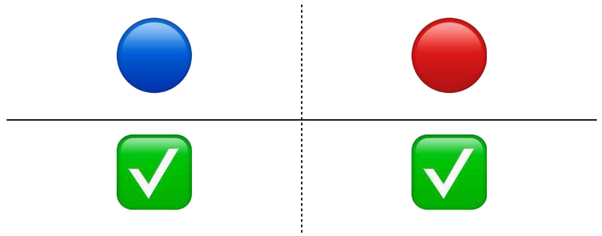
b = r = 1 n = 1

k	P(k)
0	½
1	½



b = r = 1 n = 2

k	P(k)
0	0
1	1
2	0



$$P(0) = \frac{\binom{1}{0}\cdot\overset{0}{\cancel{\binom{1}{2}}}}{\binom{2}{2}} = 0$$

$$P(1) = \frac{\binom{1}{1}\cdot\binom{1}{1}}{\binom{2}{2}} = \frac{1\cdot 1}{1} = 1$$

$$P(2) = 0$$

b = 1 r = 2 n = 2

k	P(k)
0	⅓
1	⅔
2	0

🔵	🔴	🔴
✅	✅	
✅		✅
	✅	✅

$$P(\text{k blue balls}) = \frac{\binom{b}{k} \cdot \binom{r}{n-k}}{\binom{b+r}{n}}$$

Hypergeometric Distribution

b blue balls r red balls

Pick n ≥ 0 balls

$$|\Omega| = \binom{b+r}{n}$$

If k blue balls, then n-k red balls

$$P(\text{k blue balls}) = \frac{\binom{b}{k} \cdot \binom{r}{n-k}}{\binom{b+r}{n}}$$

Recall

$$\sum_{k=0}^{n} \binom{b}{k}\binom{r}{n-k} = \binom{b+r}{n}$$

$$\sum_{k=0}^{n} P(k) = \frac{\sum_{k=0}^{n} \binom{b}{k}\binom{r}{n-k}}{\binom{b+r}{n}} = \frac{\binom{b+r}{n}}{\binom{b+r}{b}} = 1$$

Socks

Box contains 4 white, 3 black socks Choose 5

Find probability of 0,1,2,3,4,5 white

Total # choices $\binom{7}{5} = \binom{7}{2} = \frac{7 \cdot 6}{2} = 21$

0 white $\rightarrow$ 5 black $\rightarrow$ impossible

1 white $\rightarrow$ 4 black $\rightarrow$ impossible

2 white $\rightarrow$ 3 black $\rightarrow$ $\dfrac{\binom{3}{3}\binom{4}{2}}{21} = \dfrac{6}{21}$

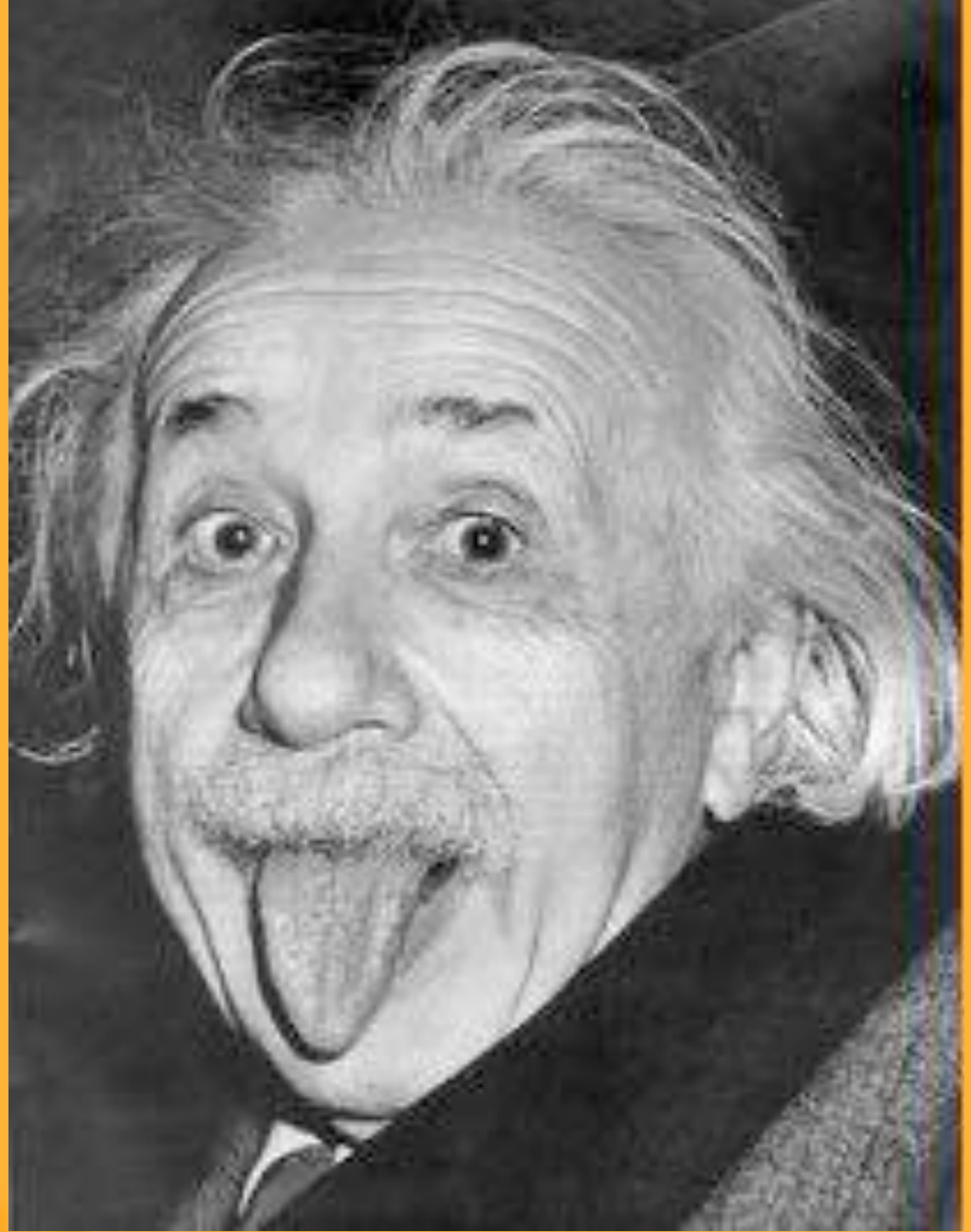
3 white $\rightarrow$ 2 black $\rightarrow$ $\dfrac{\binom{3}{2}\binom{4}{3}}{21} = \dfrac{12}{21}$

4 white $\rightarrow$ 1 black $\rightarrow$ $\dfrac{\binom{3}{1}\binom{4}{4}}{21} = \dfrac{3}{21}$

5 white $\rightarrow$ impossible

Continuous

Distributions



"Not everything that can be counted, counts; not everything that counts can be counted."

- *Albert Einstein*

Discrete to Continuous

Discrete distributions: Countable # values (finite or countably-infinite)

Continuous distributions: Uncountable # values, intervals

Why Continuous

Anything physics

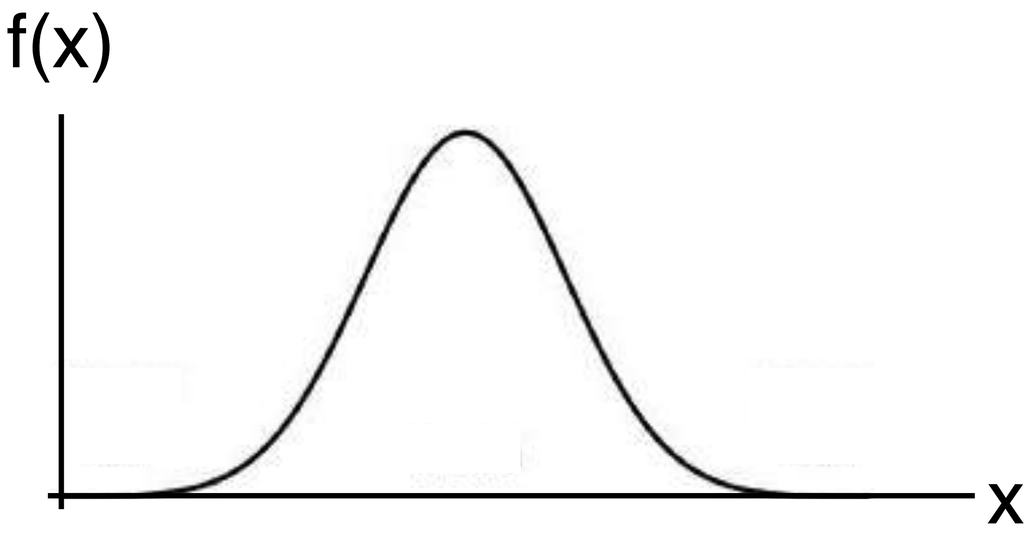
- **Time** — flight, delivery, disease, life
- **Space** — height, storm area
- **Mass** — pet, cookie
- **Temperature** — air, body

Nearly continuous variables

- **Cost** — stock, house, pork bellies
- **Rates** — interest, exchange, unemployment

Probability Density Function

Replaces the discrete pmf



$f(x) \geq 0$ relative likelihood of x

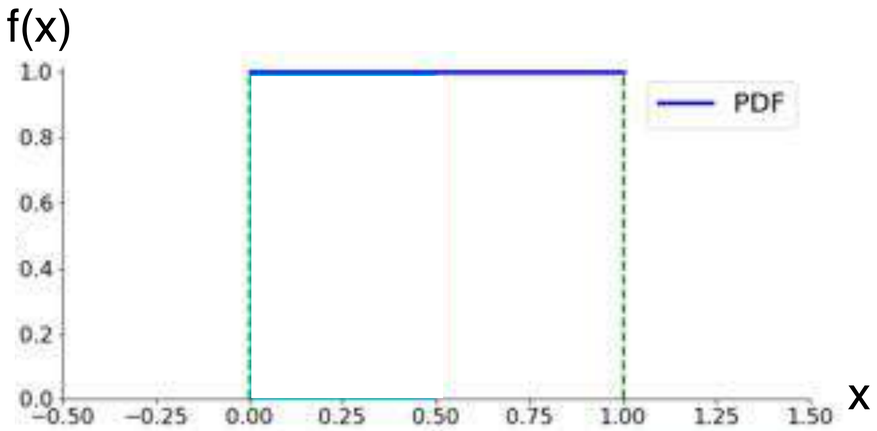
$\int_{-\infty}^{\infty} f(x)dx = 1$ area under curve (area)

Comparison to Discrete

	Discrete	Continuous
Probability function	mass (pmf)	density (pdf)
≥ 0	$p(x) \geq 0$	$f(x) \geq 0$
$\sum = 1$	$\sum_x p(x) = 1$	$\int_{-\infty}^{\infty} f(x)dx = 1$

Uniform

$$f(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

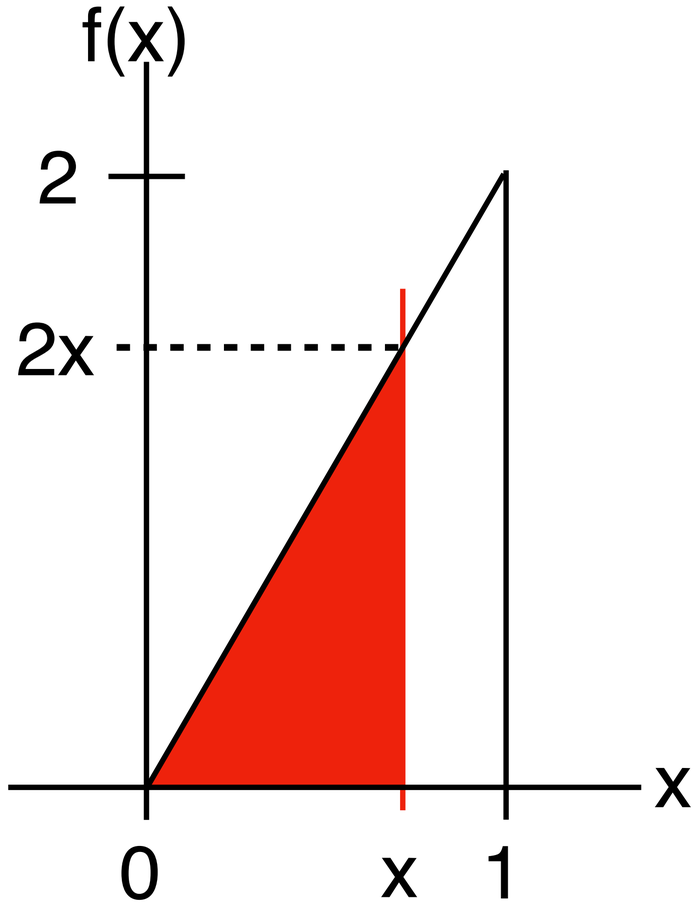


Will it ∫? Area = 1 · 1 = 1

$$\int_{-\infty}^{\infty} f(x)dx = \int_0^1 1dx = x\Big|_0^1 = 1$$

Triangle

$$f(x) = \begin{cases} 2x & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$



Will it $\int$?

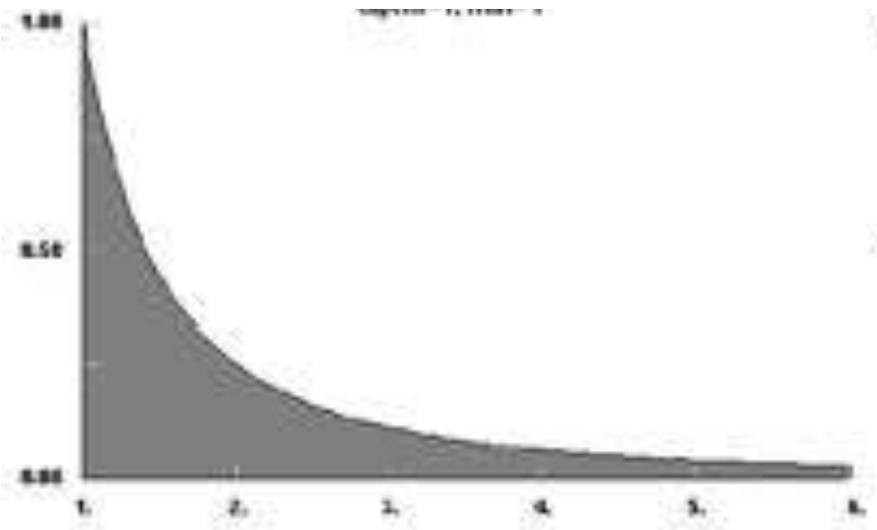
Area $= 2 \cdot 1 \cdot \frac{1}{2} = 1$

$$\int_{-\infty}^{\infty} f(x)dx = \int_0^1 2x dx = x^2\Big|_0^1 = 1 - 0 = 1$$

Infinite Support

Power law

$$f(x) = \begin{cases} \frac{1}{x^2} & x \geq 1 \\ 0 & x < 1 \end{cases}$$



Will it $\sum$?

$$\int_{-\infty}^{\infty} f(x)dx = \int_{1}^{\infty} \frac{1}{u^2} du = \frac{-1}{u}\Big|_{1}^{\infty} = 1$$

Event Probability

	Discrete	Continuous
P(A)	$\sum_{x \in A} p(x)$	$\int_{x \in A} f(x)dx$

Typically interested in interval probability P(a ≤ X ≤ b)

Area between a and b

P(X ≤ b) - P(X ≤ a)

Cumulative distribution function

Cumulative Distribution Function (CDF)

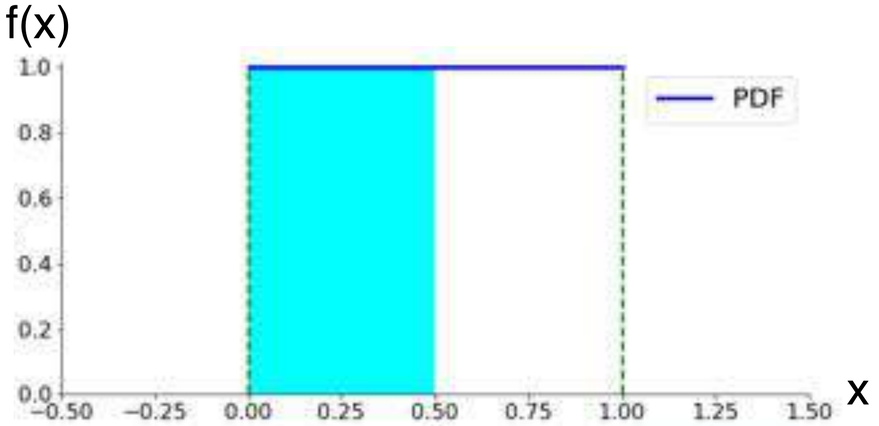
$$F(x) \triangleq P(X \leq x)$$

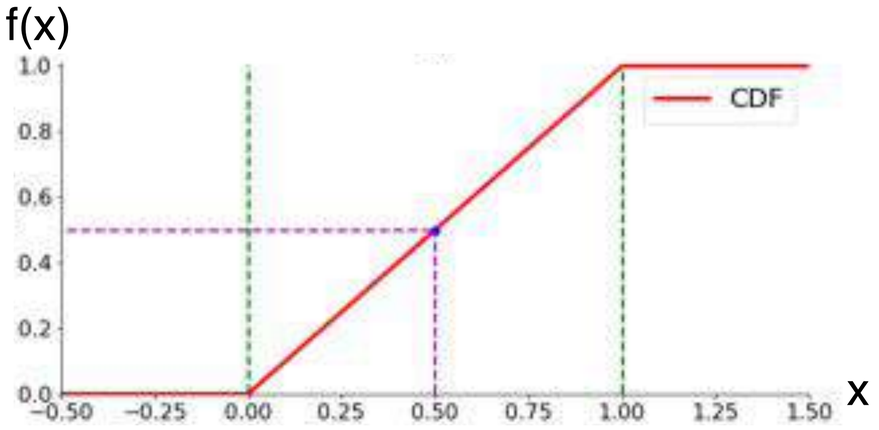
	Discrete	Continuous
PF → CDF	$\sum_{u \leq x} p(u)$	$\int_{-\infty}^{x} f(u)du$
CDF → PF	$p(x) = F(x) - F(x^*)$	$f(x) = F'(x)$

x^* - element preceding x

Uniform

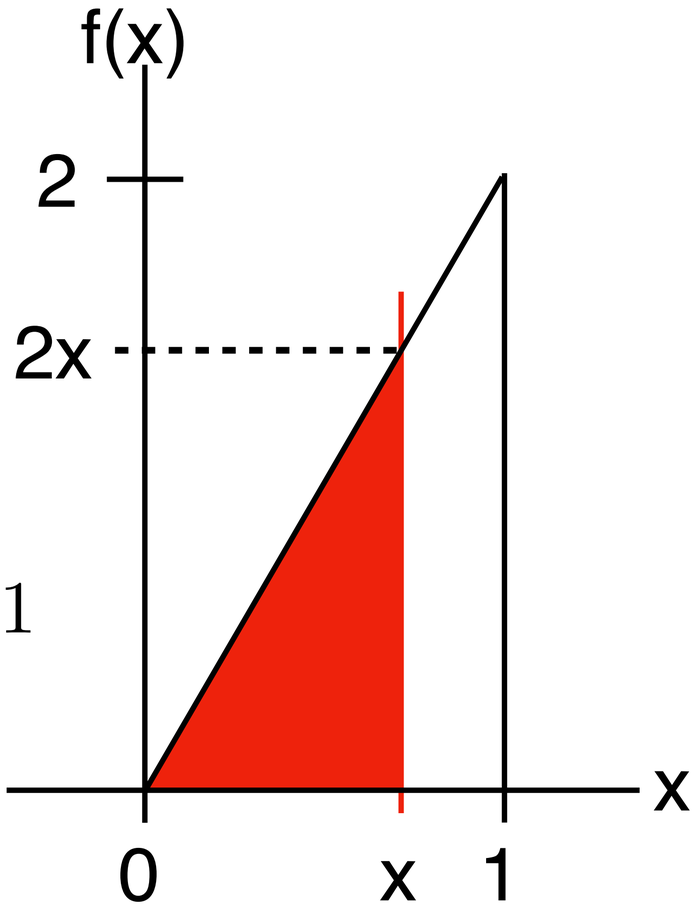
$$F(x) = \int_{-\infty}^{x} f(u)du = \begin{cases} 0 & x \leq 0 \\ \int_0^x 1du = u\Big|_0^x = x & 0 \leq x \leq 1 \\ 1 & 1 \leq x \end{cases}$$





$$F'(x) = \begin{cases} (0)' = 0 & x \leq 0 \\ (x)' = 1 & 0 < x < 1 \\ (1)' = 0 & 1 < x \end{cases}$$

Triangle

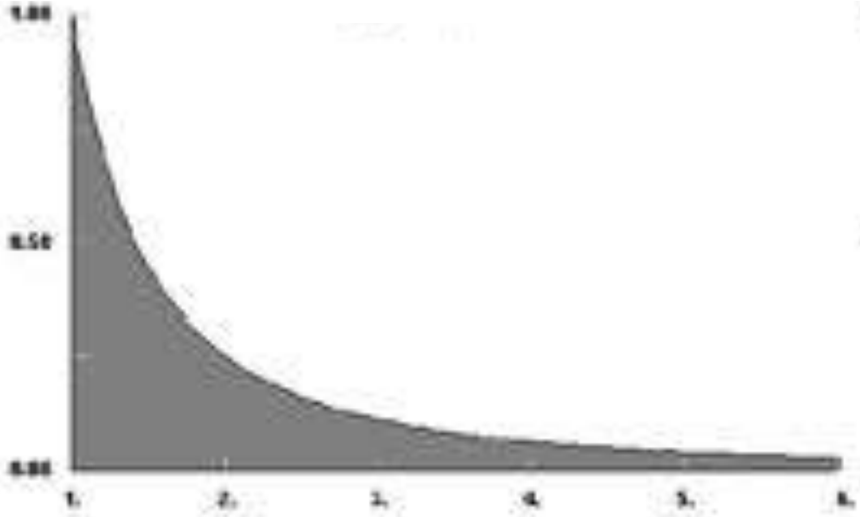


$$F(x) = \int_{-\infty}^{x} f(u)du = \begin{cases} 0 & x \leq 0 \\ \int_0^x 2udu = u^2\big|_0^x = x^2 & 0 \leq x \leq 1 \\ 1 & 1 \leq x \end{cases}$$

$$F'(x) = \begin{cases} (0)' = 0 & x < 0 \\ (x^2)' = 2x & 0 < x < 1 \\ (1)' = 0 & 1 < x \end{cases}$$

Infinite Support

$$F(x) = \begin{cases} 0 & x \leq 1 \\ \int_1^x \frac{1}{u^2} du = \frac{-1}{u}\Big|_1^x = 1 - \frac{1}{x} & x \geq 1 \end{cases}$$



$$F'(x) = \begin{cases} (0)' = 0 & x < 1 \\ (1 - \frac{1}{x})' = \frac{1}{x^2} & x > 1 \end{cases}$$

Properties of the CDF

$F(x)$ = integral

Nondecreasing

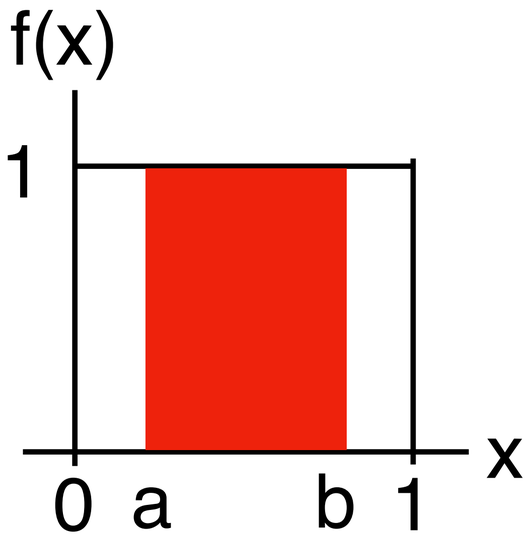
$F(-\infty)=0$

$F(\infty)=1$

Continuous

Interval Probabilities

Examples



Uniform

$0 \le a \le b \le 1$

$$P(a \le X \le b) = \begin{cases} \text{Area} = (b - a) \cdot 1 = b - a \\ \int_a^b f(x)dx = \int_a^b 1dx = x\Big|_a^b = b - a \\ F(b) - F(a) = b - a \end{cases}$$

$P(0.6 \le X \le 1.3) = P(0.6 \le X \le 1) = 0.4$

$= F(1.3) - F(0.6) = 1 - 0.6 = 0.4$

Power law

$1 \le a \le b$ $\quad P(a \le X \le b) = F(b) - F(a) = (1 - \frac{1}{b}) - (1 - \frac{1}{a}) = \frac{1}{a} - \frac{1}{b}$

Differences

Discrete	Continuous
$p(x) \leq 1$	$f(x)$ can be >1
Generally $p(x) \neq 0$	$p(x)=0$
Generally $P(X \leq a) \neq P(X < a)$	$P(X \leq a) = P(X < a) = F(a)$
	$P(X \geq a) = P(X > a) = 1 - F(a)$
	$P(a \leq X \leq b) = P(a < X < b) = F(b) - F(a)$

Expectation

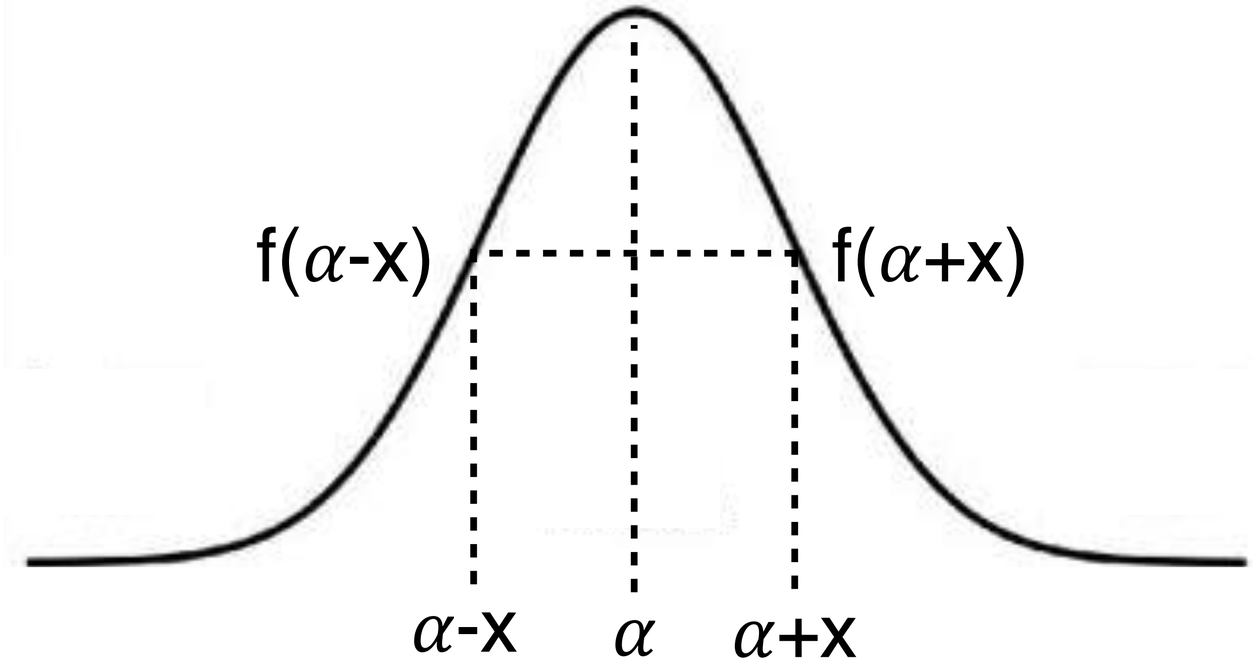
	Discrete	Continuous
EX	$\sum x \cdot p(x)$	$\int_{-\infty}^{\infty} x f(x) dx$

As discrete: Average of many samples

Properties

Support set = [a,b] $a \le EX \le b$

Symmetry If for some α, $f(\alpha+x)=f(\alpha-x)$ for all x then $EX = \alpha$



Examples

Uniform $EX = \int_{-\infty}^{\infty} x f(x) dx = \int_0^1 x\ 1\ dx = \frac{x^2}{2}\Big|_0^1 = \frac{1}{2}$

Triangle $EX = \int_0^1 x\ 2x\ dx = \frac{2x^3}{3}\Big|_0^1 = \frac{2}{3}$

Power law $EX = \int_1^{\infty} x\ \frac{1}{x^2}\ dx = \int_1^{\infty} \frac{1}{x}\ dx = \ln x\Big|_1^{\infty} = \infty$

Later: power laws with finite expectation

Variance

	Discrete	Continuous
$V(X) \triangleq E(X-\mu)^2$	$\sum_x p(x)(x-\mu)^2$	$\int_{-\infty}^{\infty} f(x)(x-\mu)^2 dx$

$$E(X^2) - \mu^2$$

As for discrete

$$\begin{aligned} E(X-\mu)^2 &= \int (x-\mu)^2 f(x) dx \\ &= \int (x^2 - 2x\mu + \mu^2) f(x) dx \\ &= \int x^2 f(x)\ dx - 2\mu \int x f(x)\ dx + \mu^2 \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2 \end{aligned}$$

$$\sigma = \sqrt{V(X)}$$

Examples

Uniform

$$EX = \frac{1}{2}$$

$$E(X^2) = \int_0^1 x^2 \; 1 \; dx = \frac{x^3}{3}\Big|_0^1 = \frac{1}{3}$$

$$V(X) = E(X^2) - (EX)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

$$\sigma = \frac{1}{\sqrt{12}} = \frac{1}{2\sqrt{3}}$$

Triangle

$$EX = \frac{2}{3}$$

$$EX^2 = \int_0^1 x^2 \; 2x \; dx = \frac{2}{4}x^4\Big|_0^1 = \frac{1}{2}$$

$$V(X) = E(X^2) - (EX^2) = \frac{1}{2} - (\frac{2}{3})^2 = \frac{9-8}{18} = \frac{1}{18}$$

$$\sigma = \frac{1}{\sqrt{18}} = \frac{1}{3\sqrt{2}}$$

Discrete vs. Continuous

	Discrete	Continuous
Prob. Fun.	pmf - p	pdf - f
≥ 0	$p(x) \geq 0$	$f(x) \geq 0$
∑ = 1	$\sum p(x) = 1$	$\int f(x)dx = 1$
P(A)	$\sum_{x \in A} p(x)$	$\int_{x \in A} f(x)dx$
F(X)	$\sum_{u \leq x} p(u)$	$\int_{-\infty}^{x} f(u)du$
μ = E(X)	$\sum xp(x)$	$\int xf(x)dx$
V(X)	$\sum (x-\mu)^2 p(x)$	$\int (x-\mu)^2 f(x)dx$

p(x) ≤ 1 f can be larger

f(x) = F'(x)

P(a ≤ X ≤ b) = P(a < X < b) = F(b) - F(a)

Functions of Continuous Random Variables

Functions of Random Variables

Functions of Random Variables

X - income, Y = 0.3X - income tax

X - driving speed, Y= 2^x - speeding ticket

$X \sim f_X$ $\quad$ f_X known distribution

$Y = g(X)$ $\quad$ g known deterministic function

$f_Y \quad ?$

What is the distribution of Y?

Power-law pdf's

$a > -1$

$$\int_0^1 \underbrace{(a+1)x^a}_{\geq 0} dx = x^{a+1}\Big|_0^1 = 1$$

$a < -1$ later

$$f(x) = \begin{cases} (a+1)x^a & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

pdf!

$$F(x) = \int_0^x (a+1)u^a du = u^{a+1}\Big|_0^x = x^{a+1}$$

$a \leq x \leq 1$

Use $f(x) = 3x^2$

$F(x) = x^3$

$0 \leq x \leq 1$

$$X \sim f_X(x) = 3x^2$$

$$F_X(x) = x^3$$

$$0 \le x \le 1$$

g ↗

$$Y = g(X) \triangleq X^{\frac{3}{2}}$$

$$F_Y(y) = P(Y \le y) \qquad 0 \le y \le 1$$

$$= P(X^{\frac{3}{2}} \le y)$$

$$= P(X \le y^{\frac{2}{3}})$$

$$= F_X(y^{\frac{2}{3}})$$

$$= y^2$$

$$F_Y(0) = 0$$

$$F_Y(1) = 1 \checkmark$$

$$f_Y(y) = F'_Y(y) = 2y$$

g ↘

$$Y = g(X) \triangleq X^{-3}$$

$$F_Y(y) = P(Y \le y) \qquad y \ge 1$$

$$= P(X^{-3} \le y)$$

$$= P(X \ge y^{-\frac{1}{3}})$$

$$= 1 - P(X \le y^{-\frac{1}{3}})$$

$$= 1 - F_X(y^{-\frac{1}{3}})$$

$$= 1 - y^{-1}$$

$$F_Y(1) = 0$$

$$\checkmark\ F_Y(\infty) = 1$$

$$f_Y(y) = F'_Y(y) = y^{-2}$$

$g \nearrow$

$f_X(x) = 3x^2$ $\quad F_X(x) = x^3$ $\quad g(x) = x^{\frac{3}{2}}$ $\quad h(y) = y^{\frac{2}{3}}$ $\quad f_Y(y) = 2y$

$F_Y(y) \triangleq P(Y \leq y)$

$f_Y(y) = F'_Y(y)$

$= P(g(X) \leq y)$ $\quad P(X^{\frac{3}{2}} \leq y)$ $\quad = [F_X(h(y))]'$ $\quad [(y^{\frac{2}{3}})^3]'$

$= P(X \leq g^{-1}(y))$ $\quad P(X \leq y^{\frac{2}{3}})$ $\quad = F'_X(h(y)) \cdot h'(y)$ $\quad 3(y^{\frac{2}{3}})^2 \cdot \frac{2}{3} y^{-\frac{1}{3}}$

$= F_X(g^{-1}(y))$ $\quad F_X(y^{\frac{2}{3}})$ $\quad = f_X(h(y)) \cdot h'(y)$ $\quad 2y$

$= F_X(h(y))$ $\quad h(y) \stackrel{\text{def}}{=} g^{-1}(y)$

g ↘

$$F_Y(y) \triangleq P(Y \leq y)$$

$$= P(g(X) \leq y)$$

$$= P(X \geq g^{-1}(y))$$

$$= 1 - P(X \leq g^{-1}(y))$$

$$= 1 - F_X(g^{-1}(y))$$

$$= 1 - F_X(h(y))$$

$$f_Y(y) = F'_Y(y)$$

$$= [1 - [F_X(h(y))]'$$

$$= -F'_X(h(y)) \cdot h'(y)$$

$$= -f_X(h(y)) \cdot h'(y)$$

Combining

$g \nearrow$ $$f_Y(y) = f_X(h(y)) \cdot h'(y)$$

$g \searrow$ $$f_Y(y) = -f_X(h(y)) \cdot h'(y)$$

For both $$f_Y(y) = f_X(h(y)) \cdot |h'(y)|$$

Alternative formulation $$f_Y(y) = \frac{1}{|g'(x)|} f_X(x)\Big|_{y=g(x)}$$

Functions of Random Variables

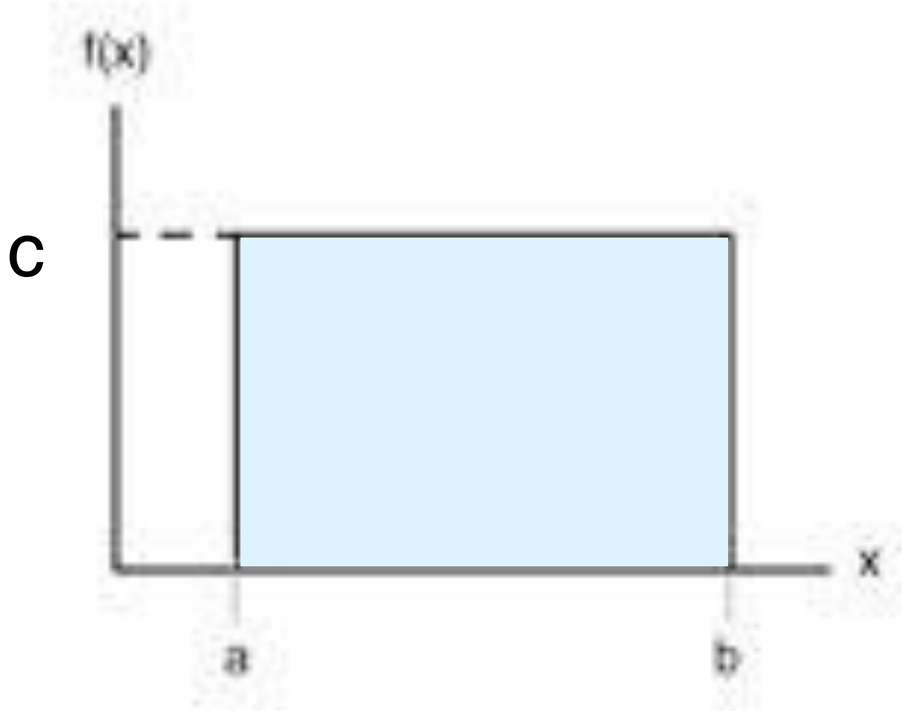
Uniform Distributions

Uniform distribution

Definintuition

For a<b, the ***uniform*** distribution $U_{[a,b]}$ is constant inside [a,b] and 0 outside

$$f(x) = \begin{cases} c & x \in [a,b] \quad \text{equally likely} \\ 0 & x \notin [a,b] \quad \text{never happen} \end{cases}$$



For $(\alpha, \beta) \subseteq [a,b]$

Probability determined by, and $\propto$, length β-α

Area under curve is always a rectangle

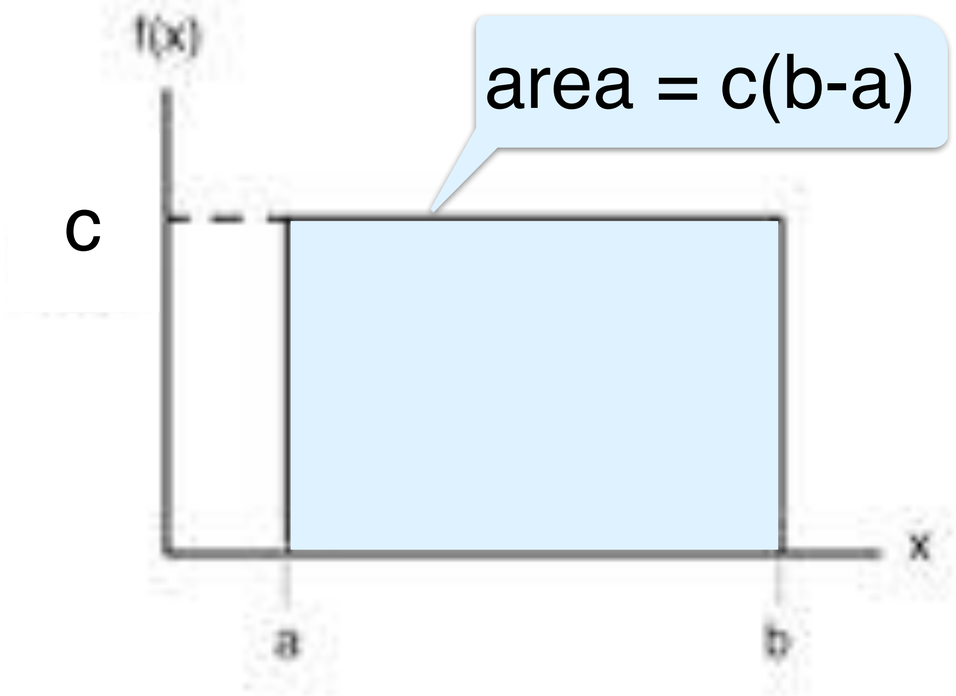
Integrals are just height x width

c?

$$f(x) = \begin{cases} c = \frac{1}{b-a} & x \in [a,b] \\ 0 & x \notin [a,b] \end{cases}$$

$$1 = \int_{-\infty}^{\infty} f(x)dx = c(b-a)$$

$$c = \frac{1}{b-a}$$



Who’s Uniform

Departure times

Wait time for a bus

Location of chip defect

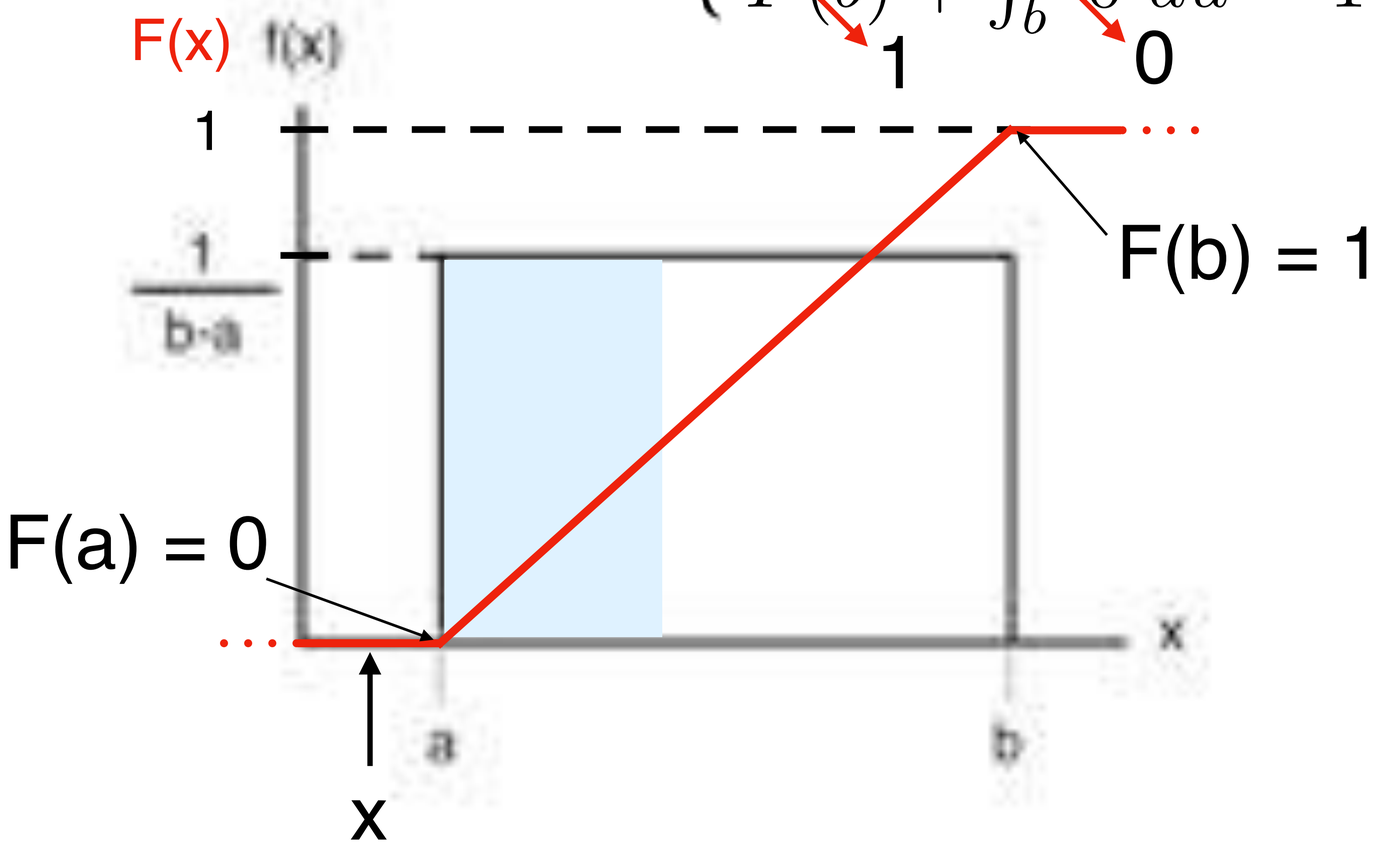
Location of a molecule in space

Considering a small area in time or space

Not so many…

CDF

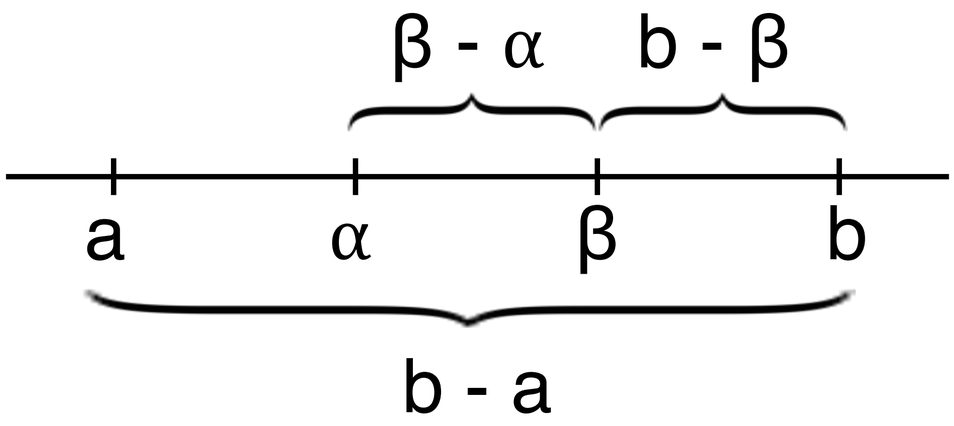
$$F(x) = \int_{-\infty}^{x} f(u)du = \begin{cases} \int_{-\infty}^{x} 0 \; du = 0 & x \leq a \\ \underbrace{\cancel{F(a)}}_{0} + \int_{a}^{x} \frac{1}{b-a} du = \frac{x-a}{b-a} & a \leq x \leq b \\ \underbrace{\cancel{F(b)}}_{1} + \underbrace{\cancel{\int_{b}^{x} 0 \; du}}_{0} = 1 & x \geq b \end{cases}$$



Interval Probabilities

For a ≤ α ≤ β ≤ b

Interval	Probability
$(\alpha, \beta]$	$F(\beta) - F(\alpha) = \frac{\beta - a}{b - a} - \frac{\alpha - a}{b - a} = \frac{\beta - \alpha}{b - a}$
$[\beta, \infty)$	$F(\infty) - F(\beta) = 1 - \frac{\beta - a}{b - a} = \frac{b - \beta}{b - a}$
$\{\alpha\}$	$F(\alpha) - F(\alpha) = 0$



μ & σ

X ~ $U_{[0,\,1]}$ first

$$f(x) = 1, \quad 0 \le x \le 1$$

$$EX = \int_0^1 x \; dx = \frac{x^2}{2}\Big|_0^1 = \frac{1}{2}$$

$$EX^2 = \int_0^1 x^2 \; dx = \frac{x^3}{3}\Big|_0^1 = \frac{1}{3}$$

$$V(X) = EX^2 - (EX)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

$$\sigma = \sqrt{\frac{1}{12}} = \frac{1}{2\sqrt{3}} \approx 0.29$$

Translation and Scaling

Uniformity preserved under translation and scaling

$X \sim U_{[0,\,1]}$ wolog $(a \neq 0)$

For any constants $a > 0$ and b, $Y \stackrel{\text{def}}{=} aX + b \sim U_{[b,\,a+b]}$

Range: $0 \rightarrow b$ $1 \rightarrow a+b$

PDF $Y = aX + b \stackrel{\text{def}}{=} g(X)$

$$f_Y(y) = \left.\frac{f_X(x)}{|g'(x)|}\right|_{x=g^{-1}(y)} = \frac{1}{a}$$

Or: equal-length interval map to =-length intervals

General μ & σ

$Y \sim U_{[a, b]}$

$Y = (b - a)X + a$

$$EX = \frac{1}{2}$$

$$EY = (b-a)EX + a = \frac{b-a}{2} + a = \frac{a+b}{2}$$

$$V(X) = \frac{1}{12}$$

$$V(Y) = V[(b-a)X + a] = (b-a)^2 V(X) = \frac{(b-a)^2}{12}$$

$$\sigma = \frac{b-a}{2\sqrt{3}} \approx 0.29(b-a)$$

Uniform Distributions

$U_{[a,b]}$ $a<b$

PDF $f(x) = \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & x \notin [a,b] \end{cases}$

CDF $F(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x \geq b \end{cases}$

$\mu = \frac{a+b}{2}$ $V = \frac{(b-a)^2}{12}$ $\sigma = \frac{b-a}{2\sqrt{3}}$

Exponential Distribution

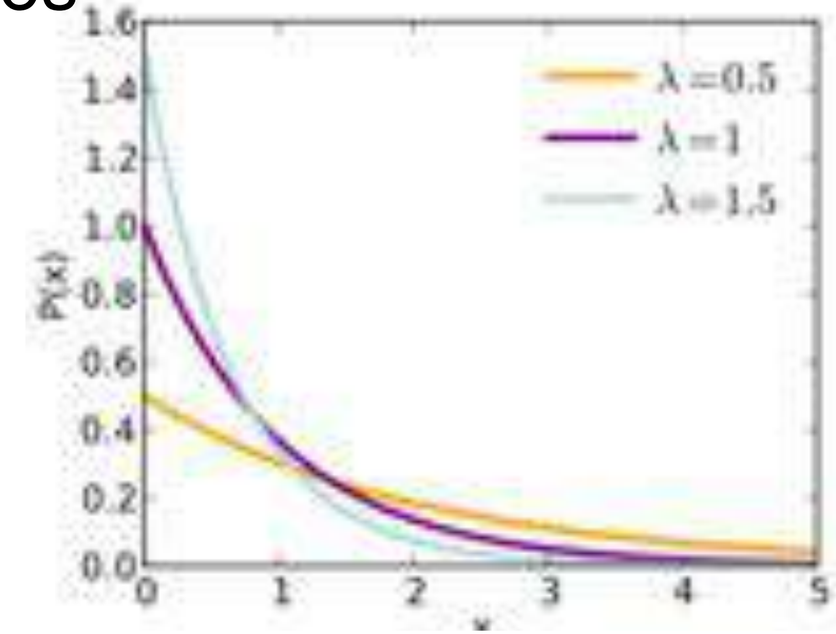
Definition

Extends geometric distribution to continuous values

$\lambda > 0$

$$f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

≥ 0 ✅



$$\int f(x)dx = \int_0^\infty \lambda e^{-\lambda x}\, dx$$
$$= -\, e^{-\lambda x}\Big|_0^\infty$$
$$= 0 - (-1)$$
$$= 1$$

Who’s Exponential

Duration of a phone call

Wait time when you call an airline

Lifetime of a car

Time between accidents

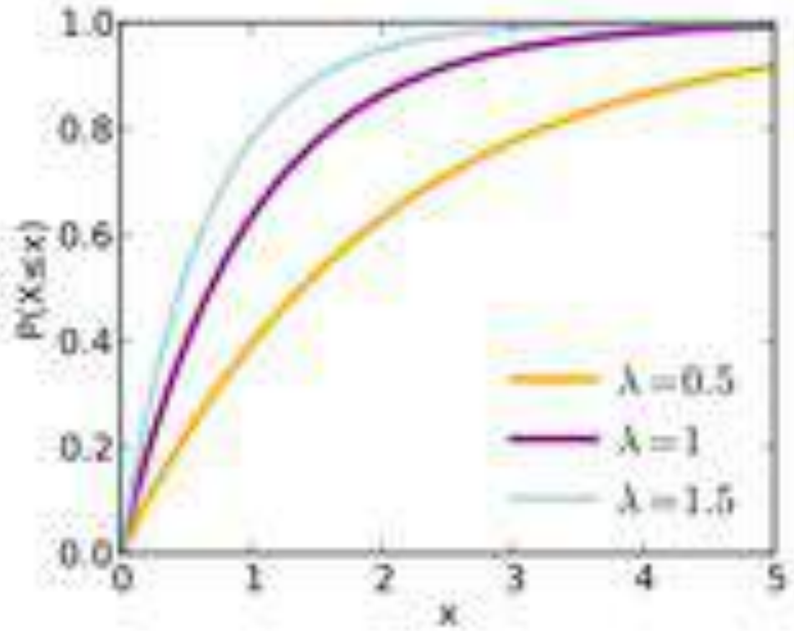
CDF

$x \geq 0$

$$P(X > x) = \int_x^\infty \lambda e^{-\lambda u} du$$

$$= -e^{-\lambda u}\Big|_x^\infty$$

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



$$F(x) = P(X \leq x) \quad = 1 - P(X > x) \; = 1 - e^{-\lambda x}$$

$x \geq 0$

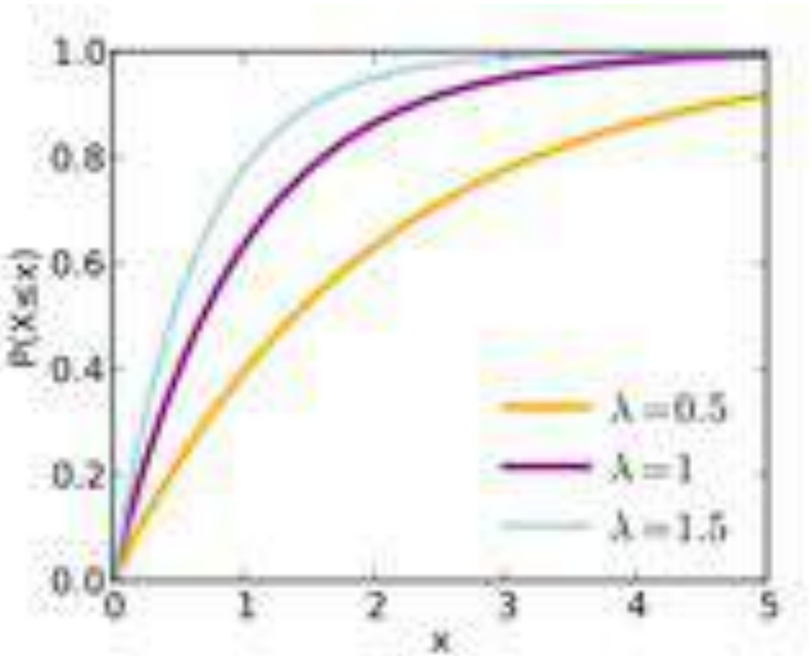
$$F(x) = 0$$

$x \leq 0$

CDF

$$P(X > x) = \begin{cases} \int_x^\infty \lambda e^{-\lambda u} du \ = -e^{-\lambda u}\Big|_x^\infty \ = e^{-\lambda x} & x \geq 0 \\ 1 & x \leq 0 \end{cases}$$

$$F(x) = P(X \leq x) = \begin{cases} 1 - P(X > x) \ = 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x \leq 0 \end{cases}$$



Example

$0 \leq a \leq b$

$$P(a \leq X \leq b) = P(a < X < b)$$

$$= F(b) - F(a)$$

$$= (1 - e^{-\lambda b}) - (1 - e^{-\lambda a})$$

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

Expectation

$$EX = \int_0^\infty x\lambda e^{-\lambda x} dx$$

$$u = x \qquad dv = \lambda e^{-\lambda x} dx$$

$$\int u \; dv = uv - \int v \; du$$

$$du = 1 \qquad v = -e^{-\lambda x}$$

$$= -xe^{-\lambda x}\Big|_0^\infty + \int_0^\infty e^{-\lambda x} dx$$

$$= 0 - \frac{1}{\lambda} e^{-\lambda x}\Big|_0^\infty$$

$$= \frac{1}{\lambda}$$

Variance

$$EX^2 = \int_0^\infty x^2 \lambda e^{-\lambda x} dx$$

$$u = x^2 \qquad dv = \lambda e^{-\lambda x} dx$$

$$\int u \; dv = uv - \int v \; du$$

$$du = 2x \; dx \qquad v = -e^{-\lambda x}$$

$$= -x^2 e^{-\lambda x}\Big|_0^\infty + \int_0^\infty 2x e^{-\lambda x} dx$$

$$EX = \int_0^\infty x \lambda e^{-\lambda x} dx = \frac{1}{\lambda}$$

$$= 0 + \frac{2}{\lambda} EX = \frac{2}{\lambda^2}$$

$$V(X) = EX^2 - (EX)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

$$\sigma = \frac{1}{\lambda}$$

Memoryless

$X \sim f_\lambda$ $\quad$ $a, b \geq 0$

$$P(X \geq a+b | X \geq a) = \frac{P(X \geq a+b, X \geq a)}{P(X \geq a)}$$

$$= \frac{P(X \geq a+b)}{P(X \geq a)}$$

$$= \frac{e^{-\lambda(a+b)}}{e^{-\lambda a}}$$

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

$$= P(X \geq b)$$

$$P(X < a+b | x \geq a) = 1 - P(X \geq a+b | X \geq a) = 1 - P(X \geq b) = P(X < b)$$

$$f(X = a+b | X \geq a) = f(X = b)$$

While Waiting in Line

DMV has 2 clerks, each with exponential service time

When you arrive, one person is in line 🙁

While you wait, someone cuts in front of you 😡

At some point a clerk becomes available and starts serving the first person

Before first person finishes, other clerk starts serving second person

If all three of you served randomly, P(you finish last) = ⅓

P(you finish last now)?

Evaluation

A - time first person finishes

B - time second person finishes

C - time you finish

Service	$P(A < B < C)$
Fixed	1
Exponential	?

Orders	Probability
$A < B < C$	¼
$A < C < B$	¼
$B < A < C$	¼
$B < C < A$	¼
$C < A < B$	0
$C < B < A$	0

?

$$P(A<B<C) = \underbrace{P(A < B)}_{½} \cdot \underbrace{P(B < C \mid A < B)}_{½} = ¼$$

$$P(B<C<A) = \overbrace{P(B < A)}^{½} \cdot \overbrace{P(C < A \mid B < A)}^{½} = ¼$$

Conclusion

All three of you served randomly, P(you finish last) = ⅓

Fixed service time, P(you finish last) = 1

Exponential (memoryless) service time

- You won't finish first
- All 4 other orders equally likely
- P(you finish last) = ¼ + ¼ = ½
- Only slightly larger than ⅓

Orders	P
A < B < C	¼
A < C < B	¼
B < A < C	¼
B < C < A	¼
C < A < B	0
C < B < A	0

Summary

Exponential

PDF $f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x \leq 0 \end{cases}$

CDF $F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x \leq 0 \end{cases}$

$EX = \frac{1}{\lambda}$ $\quad V(X) = \frac{1}{\lambda^2}$ $\quad \sigma = \frac{1}{\lambda}$

Memoryless

Normal Distribution

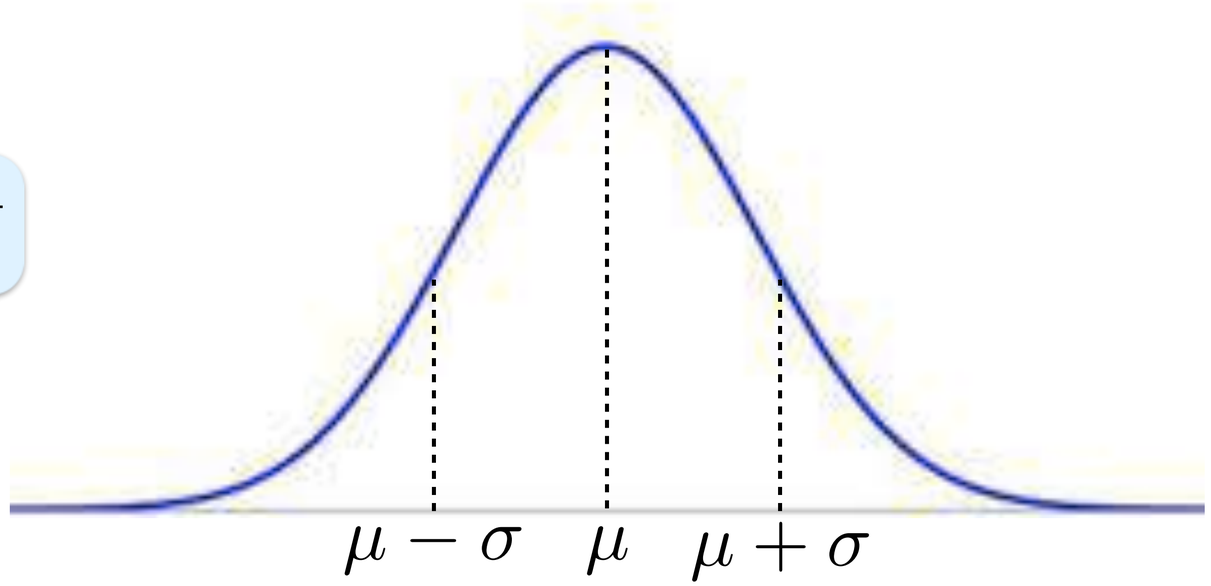
Normal (Gaussian) Distributions

Definition

Mean Variance

$$X \sim N(\mu, \sigma^2)$$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Bell Curve

Very common

Occurs whenever adding many independent factors

height, weight, rainfall, salaries, ….

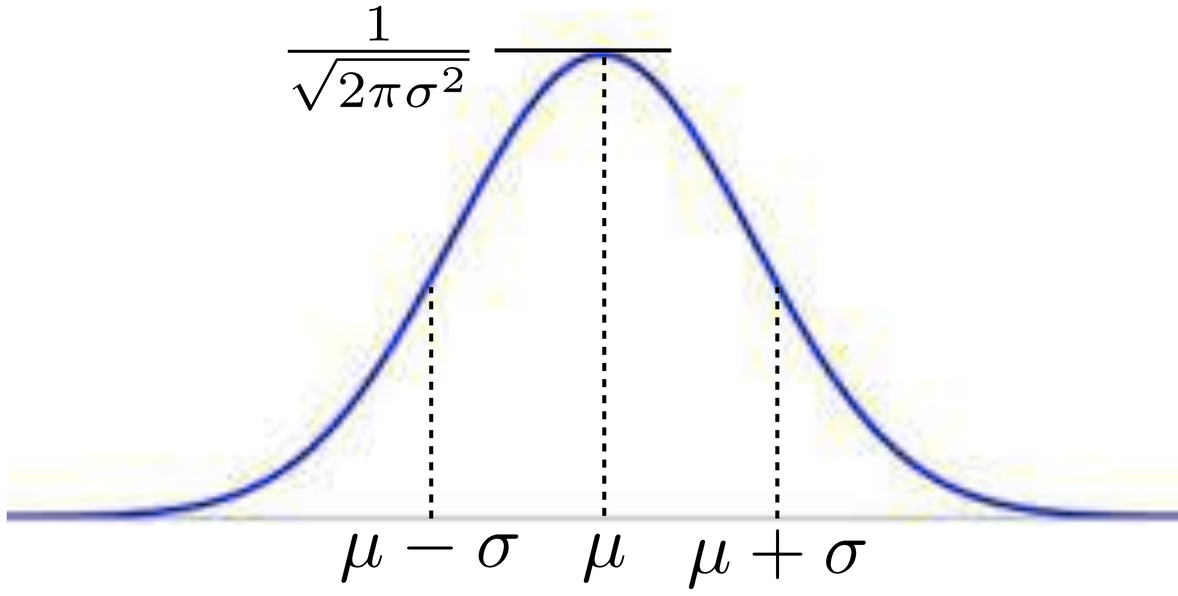
Approximates binomial distribution

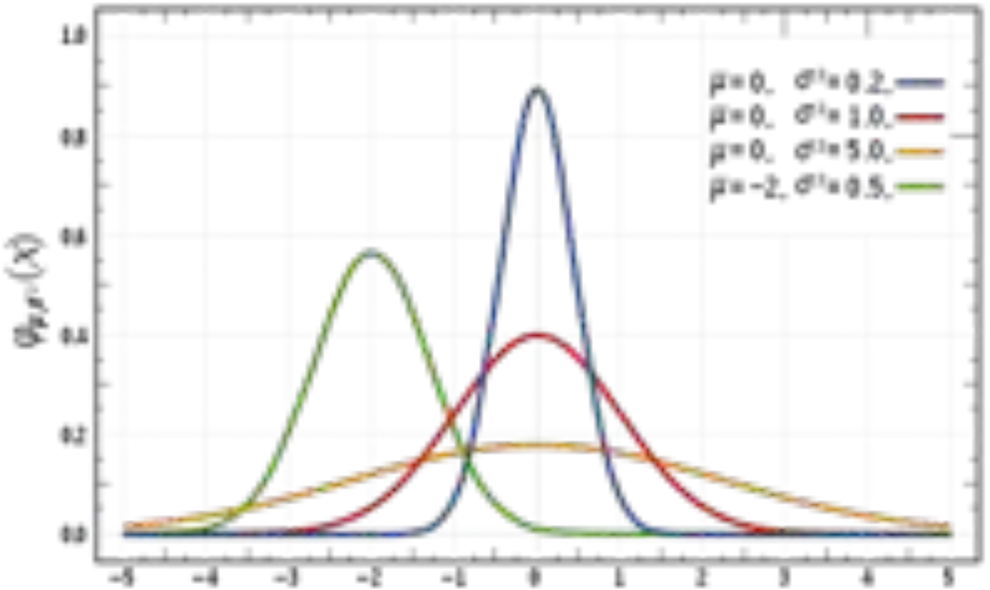
Observations

Symmetric around mean μ

μ most likely $f(\mu) = \frac{1}{\sqrt{2\pi\sigma^2}}$

As σ grows, distribution gets more flat





Linear Transformations

Linear transformations of normal distributions are normal

$$X \sim N(\mu, \sigma)$$

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$Y = aX + b$$

As for all r.v. $\mu_Y = a\mu_X + b$ $\sigma_Y = a\sigma_X$

Show normal

Variable transformation

$$f_Y(y) = \frac{1}{(ax+b)'} f_X(x)\Big|_{y=ax+b} \longleftrightarrow x = \frac{y-b}{a}$$

$$= \frac{1}{a} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\frac{y-b}{a}-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\underbrace{(a\sigma)^2}_{\sigma_Y}}} e^{-\frac{(y-\overbrace{(a\mu+b)}^{\mu_Y})^2}{2\underbrace{(a\sigma)^2}_{\sigma_Y}}}$$

$$Y \sim N(a\mu + b, (a\sigma)^2)$$

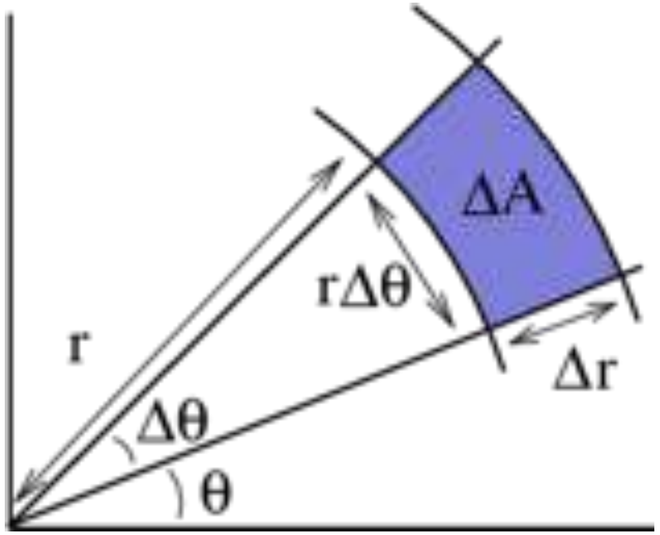
Standard Normal Distribution

Without loss of generality consider X ~ N(0,1)

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$$

Helpful integral $\int x e^{-\frac{x^2}{2}} dx = -e^{-\frac{x^2}{2}}$

ΣWILL IT ADD?



$$I = \int_{-\infty}^{\infty} e^{\frac{-x^2}{2}}\, dx$$

$$I^2 = \left(\int_{-\infty}^{\infty} e^{\frac{-x^2}{2}}\, dx\right)\left(\int_{-\infty}^{\infty} e^{\frac{-y^2}{2}}\, dy\right)$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{\frac{-(x^2+y^2)}{2}}\, dx\ dy$$

$$= \int_{0}^{\infty}\int_{0}^{2\pi} e^{\frac{-r^2}{2}}\, r\ d\theta\ dr$$

$$x = r\cos\theta \qquad y = r\sin\theta$$

$$dx\ dy = r\ dr\ d\theta$$

$$= \int_{0}^{\infty} r\ e^{\frac{-r^2}{2}} \int_{0}^{2\pi} d\theta\ dr$$

$$= 2\pi \int_{0}^{\infty} r\ e^{\frac{-r^2}{2}}\, dr$$

$$\int r e^{\frac{-r^2}{2}}\, dr = -e^{\frac{-r^2}{2}}$$

$$= -2\pi e^{\frac{-r^2}{2}}\Big|_0^{\infty} = 2\pi$$

$$I = \sqrt{2\pi}$$

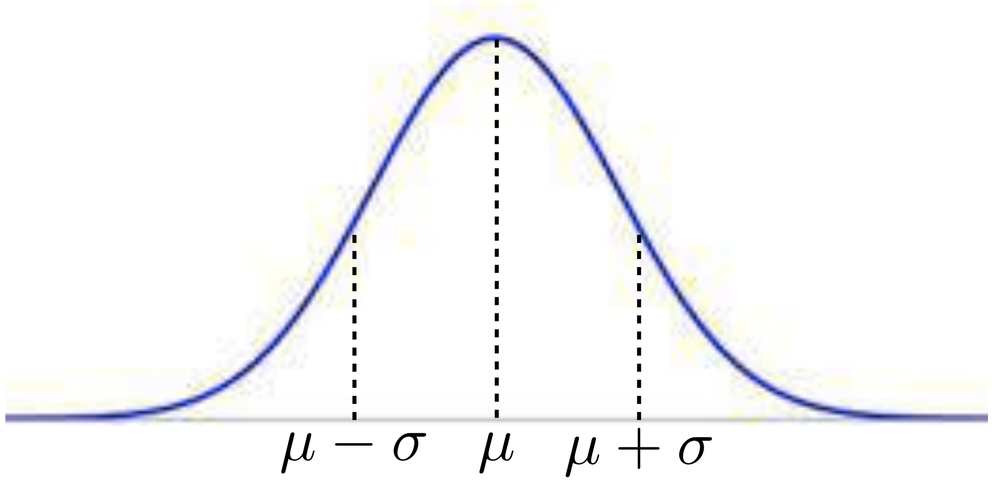
$$\int \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}\, dx = 1$$

YES IT ADDS!

Expectation

Symmetry

$$E(X) = 0$$



Calculation

$$E(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{\frac{-x^2}{2}} dx$$

$$= \frac{1}{\sqrt{2\pi}} [-e^{\frac{-x^2}{2}} \Big|_{-\infty}^{\infty} = 0$$

Variance

$E(X^2)$ $= \frac{1}{\sqrt{2\pi}} \int x^2 e^{\frac{-x^2}{2}} dx$

$u = x$ $\quad dv = x\ e^{\frac{-x^2}{2}} dx$

$\int u\ dv = uv - \int v\ du$

$du = dx$ $\quad v = -e^{\frac{-x^2}{2}}$

$= \frac{-1}{\sqrt{2\pi}} x e^{\frac{-x^2}{2}} \Big|_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} dx = 0 + 1 = 1$

$V(X)$ $= E(X^2) - (EX)^2 = 1 - 0 = 1$

Normal (Gaussian) Distributions

$\mathbb{N}(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad -\infty < x < \infty$$

$$EX = \mu$$

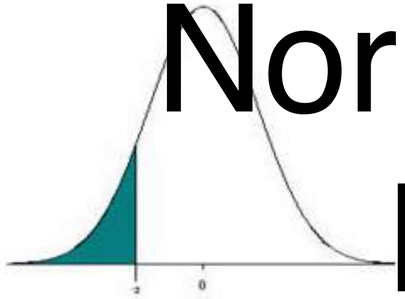
$$V = \sigma^2$$

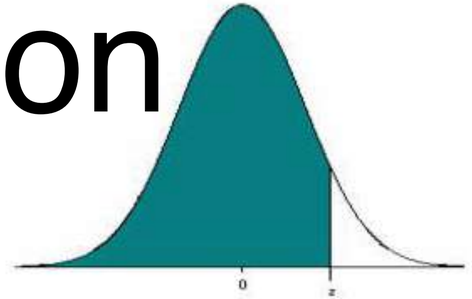
$$\sigma = \sigma$$

Very common in nature

Next Probabilities

Normal Distribution Probabilities





Interval probabilities

Z table and scores

Standard deviation

Approximating Bernoulli

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0011	0.0011	0.0010	0.0010
-2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0015	0.0015	0.0014	0.0014
-2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0094	0.0091	0.0089	0.0087	0.0084
-2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0122	0.0119	0.0116	0.0113	0.0110
-2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0606	0.0594	0.0582	0.0571	0.0559
-1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0735	0.0721	0.0708	0.0694	0.0681
-1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.1922	0.1894	0.1867
-0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.2206	0.2177	0.2148
-0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.2514	0.2483	0.2451
-0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

CDF

$$X \sim N(0,1)$$

$$\Phi(x) \triangleq F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-y^2}{2}} dy$$

No known formula

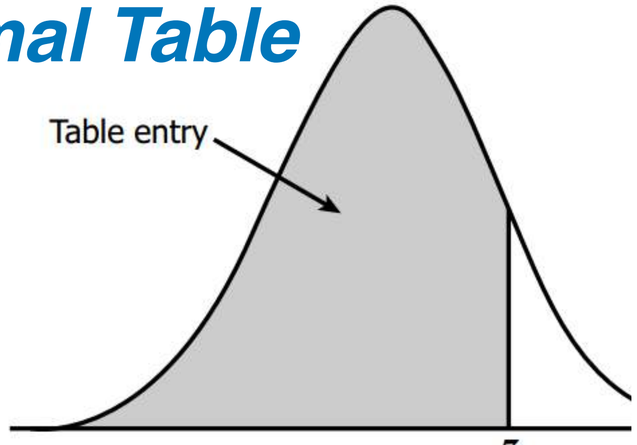
Instead use table or computer

Table for each μ, σ?

1 suffices!

Standard Normal Table

Z Table

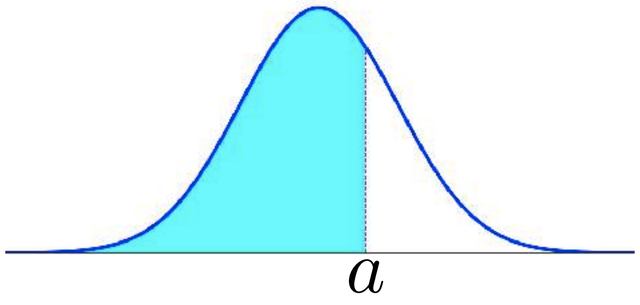


z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

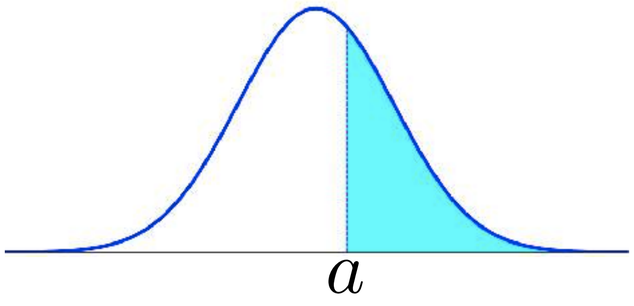
Using the Z Table

$$X \sim N(0,1)$$

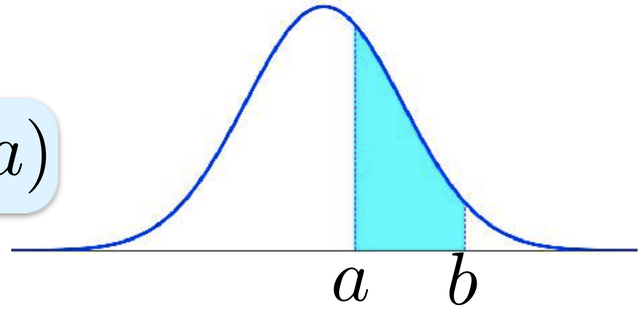
$$P(X \leq a) = \Phi(a)$$



$$P(X \geq a) = 1 - \Phi(a)$$



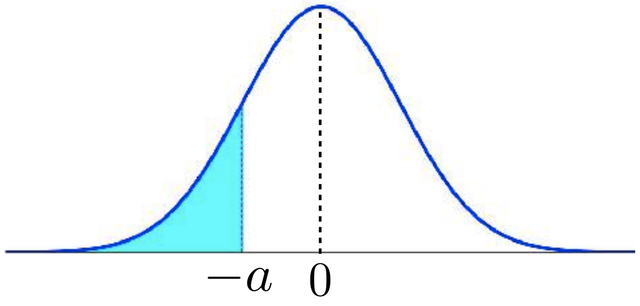
$$P(a \leq X \leq b) = \Phi(b) - \Phi(a)$$

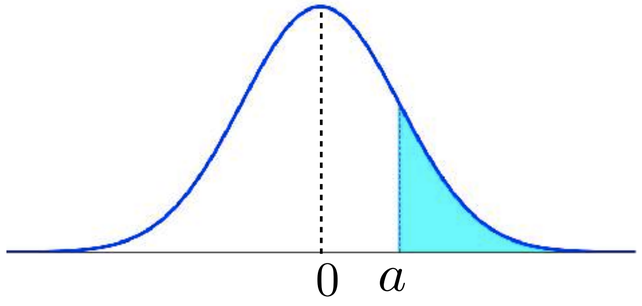


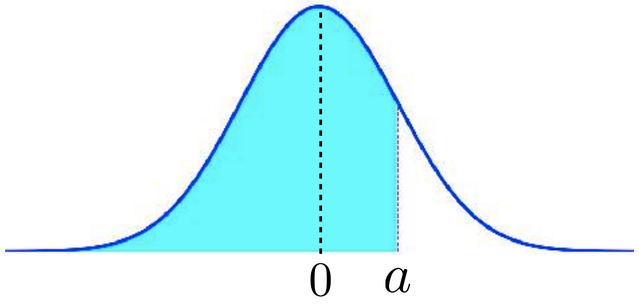
Negative Values

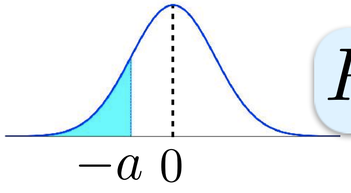
a > 0

$$\Phi(-a) = P(X \leq -a) = P(X \geq a) = 1 - P(X \leq a) = 1 - \Phi(a)$$

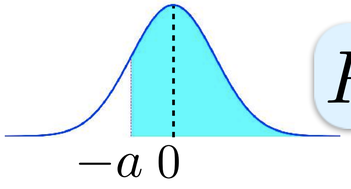




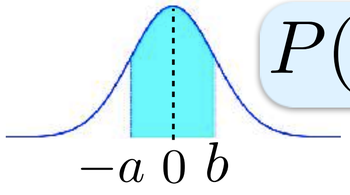




$$P(X \leq -a) = \Phi(-a) = 1 - \Phi(a)$$



$$P(X \geq -a) = 1 - \Phi(-a) = \Phi(a)$$



$$P(-a \leq X \leq b) = \Phi(b) - \Phi(-a) = \Phi(b) - (1 - \Phi(a))$$

$$= \Phi(a) + \Phi(b) - 1$$

General Normal

$$X \sim N(\mu, \sigma^2)$$

Variable	E(X)	V(X)
X	μ	σ^2
$X - \mu$	0	σ^2
$\frac{X-\mu}{\sigma}$	0	1

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

Standardized Version of X

$$P(a \leq X \leq b)$$

$$= P\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right)$$

$$= P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right)$$

Example

$X \sim N(15, 4)$ $\mu = 15$ $\sigma = 2$

Z score $Z = \frac{X-15}{2}$

$$P(11 \leq X \leq 17) = P(\frac{11-15}{2} \leq Z \leq \frac{17-15}{2})$$

$$= P(-2 \leq Z \leq 1)$$

$$= \Phi(1) + \Phi(2) - 1$$

$$\approx 0.8413 + 0.9772 - 1$$

$$= 0.8185$$

z	.00	.01	.02	.03	.04
0.0	.5000	.5040	.5080	.5120	.5160
0.1	.5398	.5438	.5478	.5517	.5557
0.2	.5793	.5832	.5871	.5910	.5948
0.3	.6179	.6217	.6255	.6293	.6331
0.4	.6554	.6591	.6628	.6664	.6700
0.5	.6915	.6950	.6985	.7019	.7054
0.6	.7257	.7291	.7324	.7357	.7389
0.7	.7580	.7611	.7642	.7673	.7704
0.8	.7881	.7910	.7939	.7967	.7995
0.9	.8159	.8186	.8212	.8238	.8264
1.0	.8413	.8438	.8461	.8485	.8508
1.1	.8643	.8665	.8686	.8708	.8729
1.2	.8849	.8869	.8888	.8907	.8925
1.3	.9032	.9049	.9066	.9082	.9099
1.4	.9192	.9207	.9222	.9236	.9251
1.5	.9332	.9345	.9357	.9370	.9382
1.6	.9452	.9463	.9474	.9484	.9495
1.7	.9554	.9564	.9573	.9582	.9591
1.8	.9641	.9649	.9656	.9664	.9671
1.9	.9713	.9719	.9726	.9732	.9738
2.0	.9772	.9778	.9783	.9788	.9793
2.1	.9821	.9826	.9830	.9834	.9838
2.2	.9861	.9864	.9868	.9871	.9875
2.3	.9893	.9896	.9898	.9901	.9904
2.4	.9918	.9920	.9922	.9925	.9927
2.5	.9938	.9940	.9941	.9943	.9945
2.6	.9953	.9955	.9956	.9957	.9959

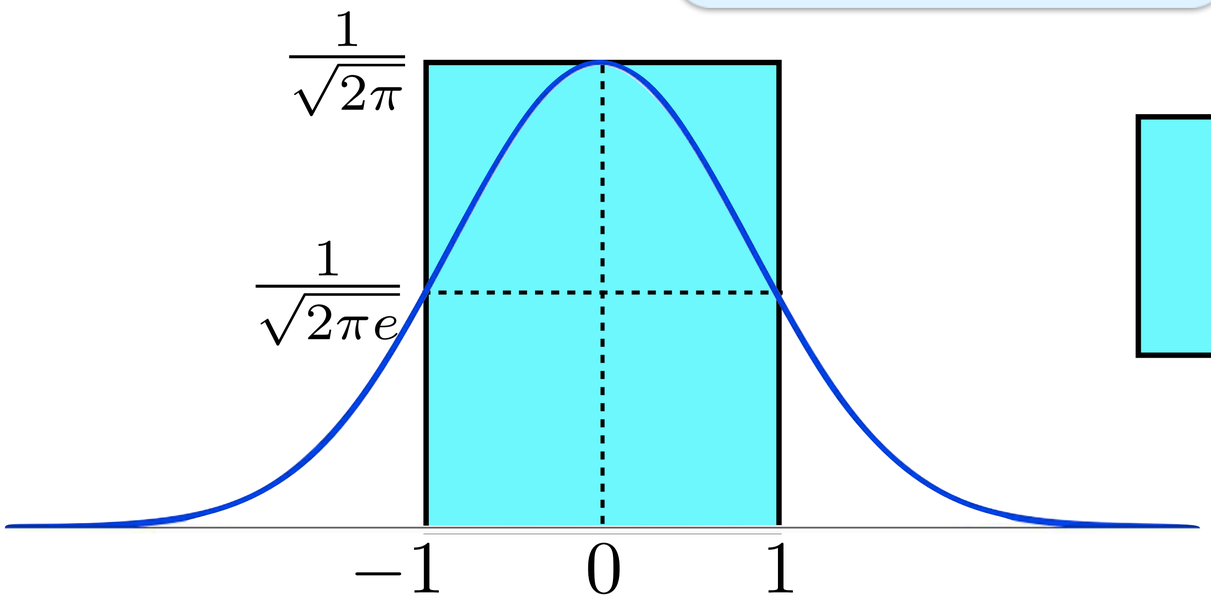
Standard Deviation and Probability

Probability of within σ

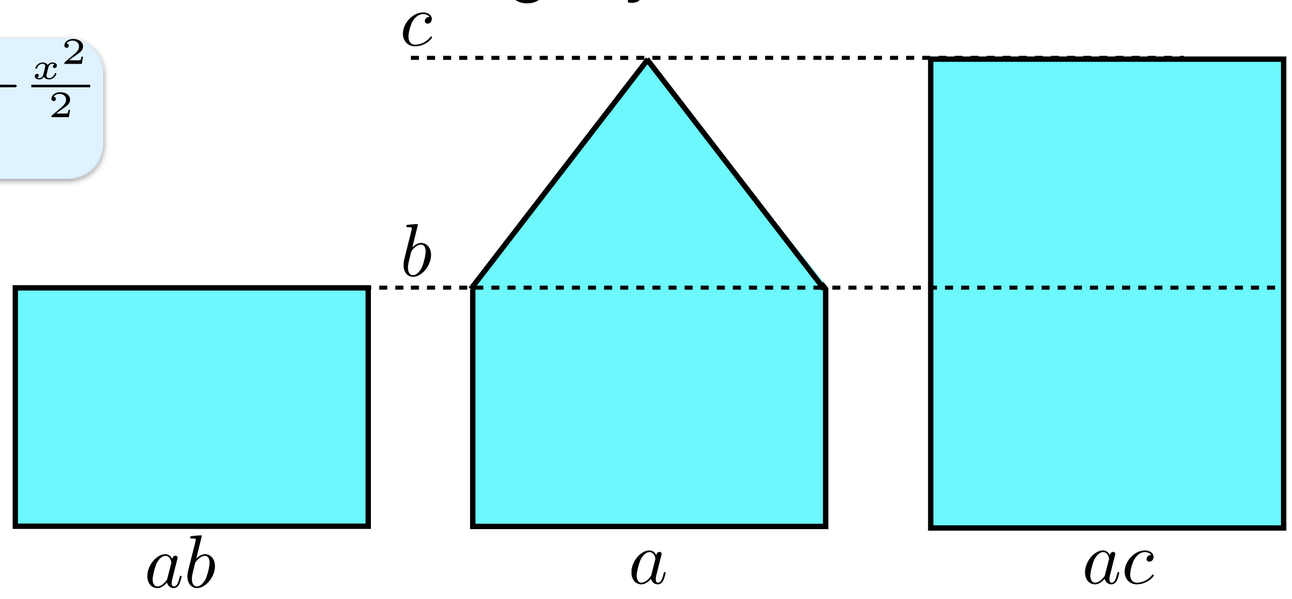
$P(|X-\mu| \le \sigma)$?

Standard normal

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$



How big's your house?



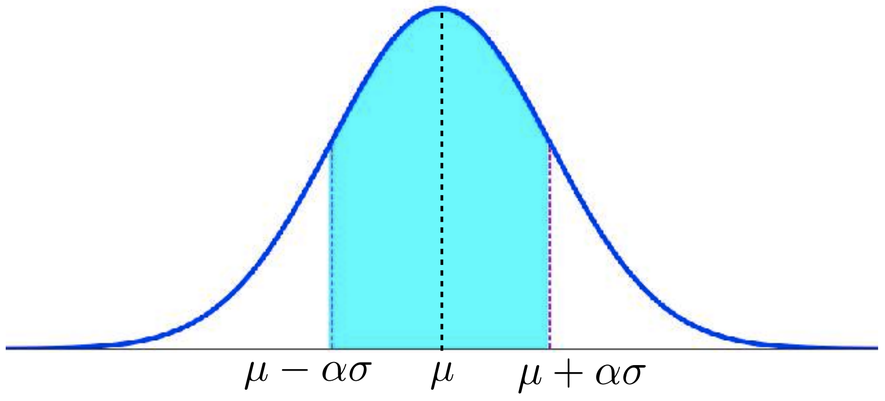
$$ab + \frac{a(c-b)}{2} = \frac{ab+ac}{2}$$

$$0.48 \approx \sqrt{\frac{2}{\pi e}} \le P(|X-\mu| \le \sigma) \le \sqrt{\frac{2}{\pi}} \approx 0.8$$

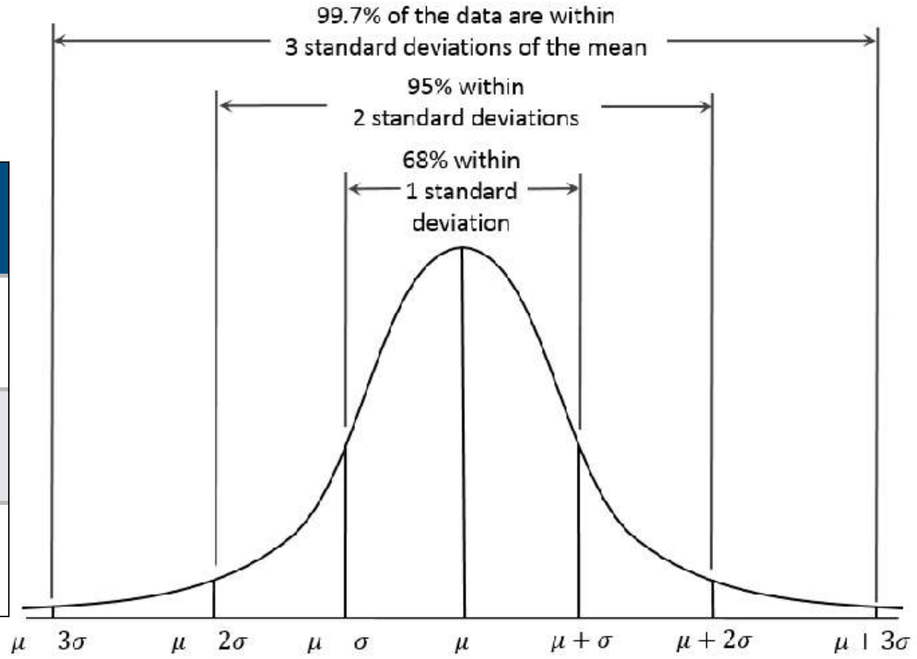
$$\frac{\sqrt{\frac{2}{\pi e}} + \sqrt{\frac{2}{\pi}}}{2} \approx 0.64$$

68 - 95 - 99.7 Rule

$$P(\mu - \alpha\sigma \leq X \leq \mu + \alpha\sigma) \;= P(-\alpha \leq Z \leq \alpha) \;\; = 2\Phi(\alpha) - 1$$



α	P(|X-μ| ≤ $\alpha\sigma$)
1	2(0.8413) - 1 = 0.682
2	2(0.9772) - 1 = 0.9544
3	2(0.9987) - 1 = 0.9974



Rare Events

α	$P(\lvert X - \mu \rvert \leq \alpha\sigma)$	$P(\lvert X - \mu \rvert > \alpha\sigma)$	Daily event ~
1	68.2%	31.8%	3 days
2	95.4%	4.6%	3 weeks
3	99.7%	0.3%	year
4	99.99%	0.01%	43 years
5	99.9999%	10^{-6}	4776 years
6	99.9999998%	$2 \cdot 10^{-9}$	1.3 millions years

ab-

6 σ philosophy

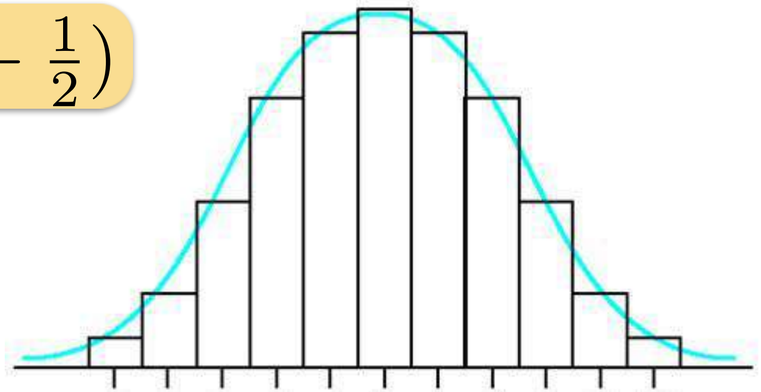
Life is normal

Normal Approximation of Binomial Distribution

Normal Approximation of Binomial

$X \sim B_{n,p}$ $\mu = np$ $\sigma = \sqrt{npq}$

$Y \sim N(np, npq)$ $P(X = k) \approx P(k - \frac{1}{2} \leq Y \leq k + \frac{1}{2})$



$X \sim B_{100,0.5}$ $np = 50$ $\sqrt{npq} = 5$

$Y \sim N(50, 25)$

$Z = \frac{Y-50}{5}$

$P(X = 60) \approx P(59.5 \leq Y \leq 60.5)$

$= P(1.9 = \frac{59.5-50}{5} \leq Z \leq \frac{60.5-50}{5} = 2.1)$

$= \Phi(2.1) - \Phi(1.9) \approx 0.9821 - 0.9713 = 0.0108$

$P(X = 60) = \binom{100}{60} 0.5^{100} \approx 0.0108$

Interval Probabilities

$X \sim B_{100,0.5}$

$np = 50$

$\sqrt{npq} = 5$

$Y \sim N(50, 25)$

$$P(42 \leq X \leq 53) = P(41.5 \leq Y \leq 53.5)$$

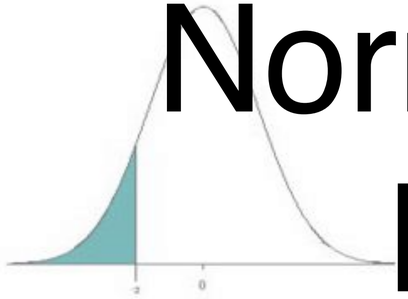
$$= P(-1.7 = \tfrac{41.5-50}{5} \leq Z \leq \tfrac{53.5-50}{5} \leq 0.7)$$

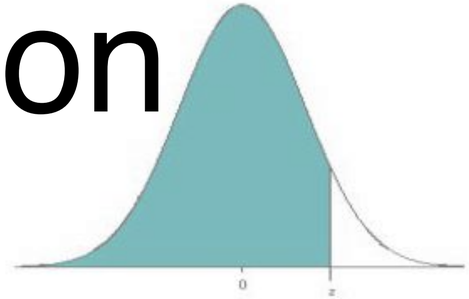
$$= \Phi(0.7) + \Phi(1.7) - 1$$

$$\approx 0.7580 + 0.9554 - 1 = 0.7134$$

$$\text{Actual} = \sum_{k=42}^{53} \binom{100}{k} \left(\tfrac{1}{2}\right)^{100} \approx 0.7136$$

Normal Distribution Probabilities





Interval probabilities

Z table and scores

Standard deviation

Approximating Bernoulli

Next

Inequalities and limits

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0004	0.0004	0.0004	0.0003
-3.2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0006	0.0005	0.0005	0.0005
-3.1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0008	0.0008	0.0007	0.0007
-3.0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0853	0.0838	0.0823
-1.2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.1020	0.1003	0.0985
-1.1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.1210	0.1190	0.1170
-1.0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	[illegible]	0.3050	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.2776
-0.4	[illegible]	0.3409	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.3121
-0.3	[illegible]	0.3783	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.3483
-0.2	[illegible]	0.4168	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998