MATH 5010 –Foundations of Statistical Theory and Probability

❖ Common Families of Distributions

Instructor: He Wang
Department of Mathematics
Northeastern University

❖ **Outline:**

1. **Common Discrete Distributions**
2. **Common Continuous Distributions**

Introduction:

We usually deal with a family of distributions indexed by parameters, which allows us to vary certain characteristics of the distribution while staying with one functional form.

Example: Random Variable X $\backsim$ $\text{Bernouli}\ (\phi)$

pmf function $p_X(k) = \phi^k(1-\phi)^{1-k} = \begin{cases} \phi & if\ k = 1 \\ 1-\phi & if\ k = 0 \end{cases}$

Example: Uniform Distribution on interval [a, b].

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{for } a \le x \le b. \\ 0 & \text{for } x < a \text{ or } x > b \end{cases}$$

➢ Common Discrete Distributions:

1. Binomial distribution is a generalization of Bernoulli distribution.

Given a series of n independent trials with two outcomes (T or F) with constant probability p and $1 - p$.

Let X be the number of T appears in the n trials. Then $\mathrm{X} \backsim Binomial(n, p)$

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Example: flip a coin n times.

Example: an airline knows that 5% of people will not show up for a flight, so they overbook 52 people on a plane with 50 seats. What is the probability that nobody is bumped off the flight?

2. Multinomial is a generalization of Categorical distribution.

Given a series of n independent trials with m outcomes $(O_1, \dots O_m)$ with constant probability $(\phi_1, \dots , \phi_m)$.

Let $\vec{X}$ be the number of O_i appears in the n trials.

$$\text{Then } \vec{X} \backsim Multinomial(n, \phi_1, \dots , \phi_m)$$

$$P(X_i = n_i) = \frac{n!}{n_1! \cdots n_m!} \phi_1^{n_1} \cdots \phi_m^{n_m}$$

for each $i = 1, \dots, m$, and each $n_1 + \cdots + n_m = n$

For example, Toss a m-side die n times.

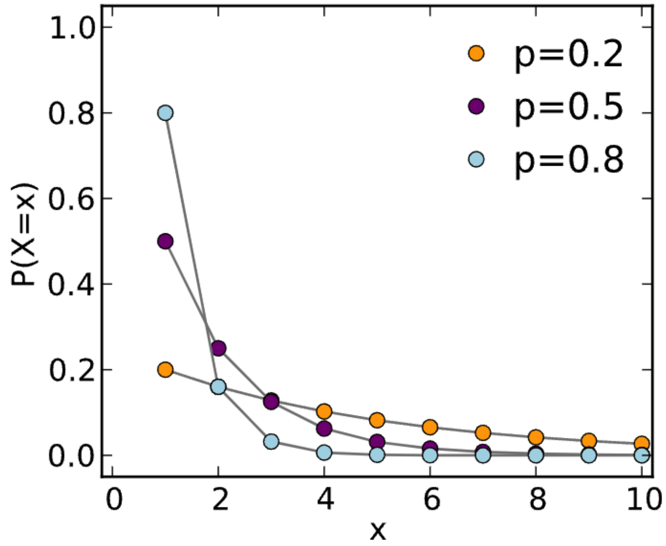
3. **Geometric distribution**.

If X is a geometric random variable with parameter p, then

$$P(X = n) = (1 - p)^{n-1}p \text{ for } n = 1, 2, 3, \dots \ .$$

Geometric random variable can be constructed using 'the number of trials of the first success occurs'.

Example: Flipping coin with a probability p that we gets a head.



The CDF of geometric random variable is

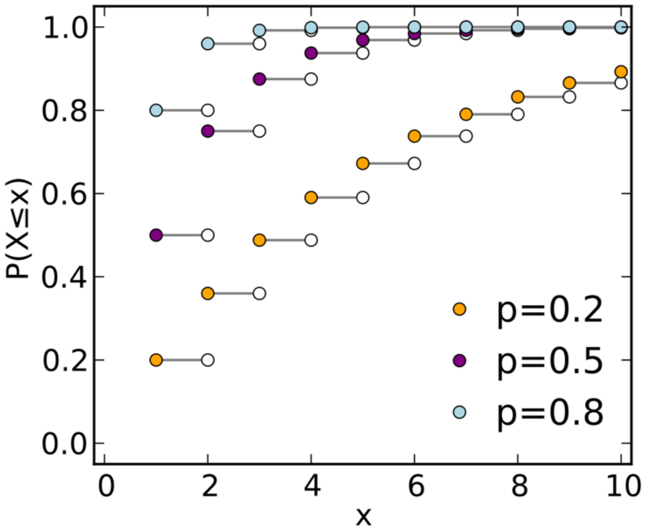
$$F(x) = P(X \leq x) = \sum_{i=1}^{x} P(X = i)$$

$$= \sum_{i=1}^{x} (1 - p)^{i-1} p$$

Geometric sum

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

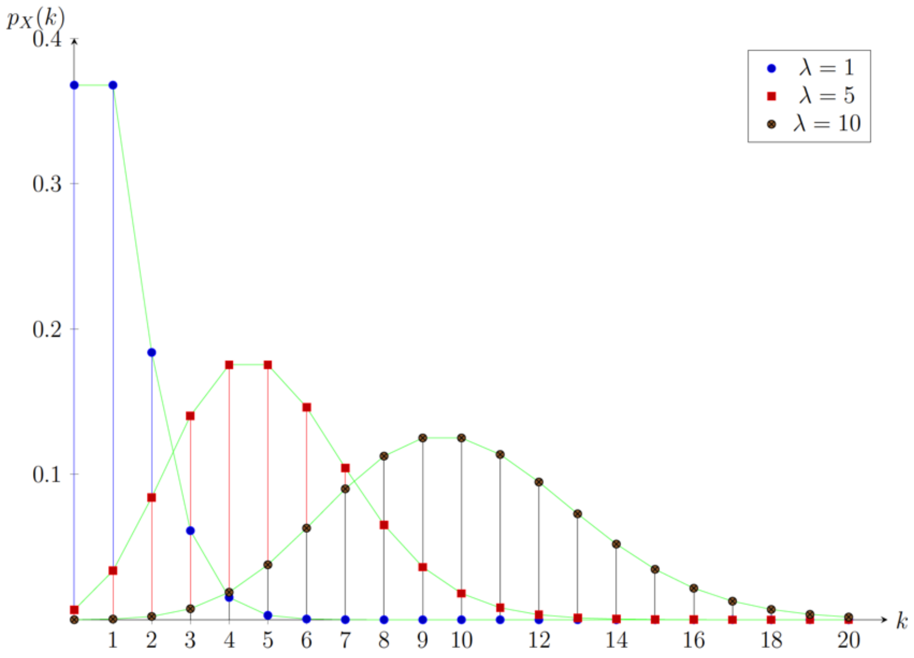
$$= 1 - (1 - p)^x$$



4. **Poisson Distribution**

Definition. The **Poisson Distribution** $Poisson(\lambda)$ is a discrete **pdf** function defined as

$$p_X(k) = P(X = k) := \frac{\lambda^k e^{-\lambda}}{k!}$$



for $k = 0,1,2,3, \dots$.
Here, λ is a positive constant.

Applications:

1.) Poisson approximation for binomial distribution

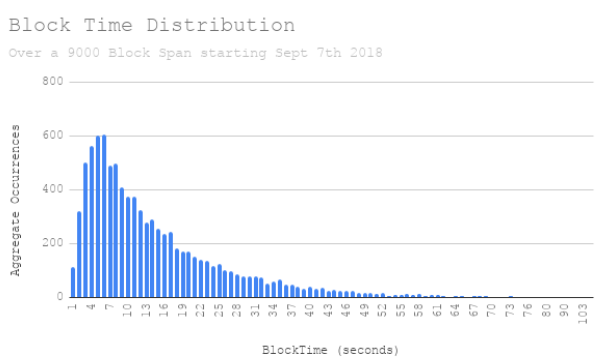
As the number of small intervals n increases (i.e., the intervals become infinitesimally small), the binomial distribution converges to the Poisson distribution.

Taking the **limit of the binomial distribution** as $n \to \infty$ and $p \to 0$, but keeping $np = \lambda$ constant, we arrive at the **Poisson distribution**:

$$\lim_{\substack{n\to\infty \\ p\to 0 \\ np=\lambda}} \binom{n}{k} p^k (1-p)^{n-k} = \frac{\lambda^k e^{-\lambda}}{k!}$$

2.) **Poisson Model**. The number of occurrences in a time interval with a given rate.

A waiting-for-occurrence application: consider a telephone operator who, on the average, handles 5 calls every 3 minutes. What is the probability that there will be no calls in the next minute? At least two calls?



➢ Common Continuous Distributions:

1. Normal Distribution

If X is a **normal random variable** with parameters $\mu\ and\ \sigma^2$

$$X \sim \mathrm{Normal}(\mu, \sigma^2)$$

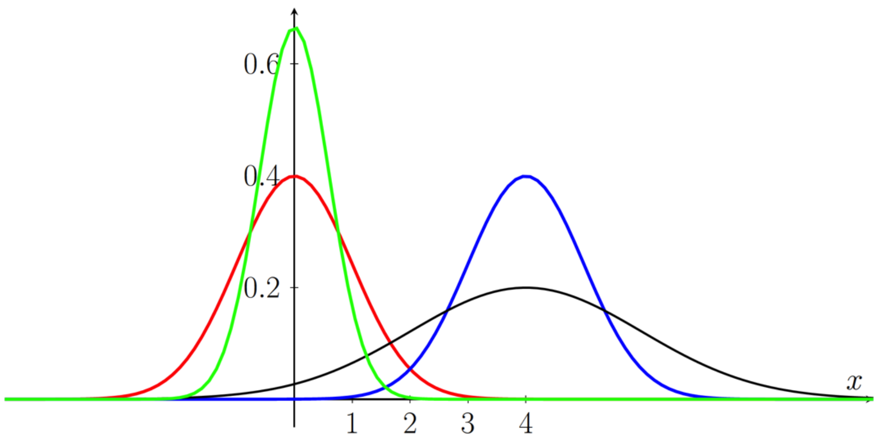
The pdf of X is

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

for $x \in \mathbb{R}$

The normal distribution is by far the most important probability distribution.

One of the main reasons for that is the Central Limit Theorem (CLT)



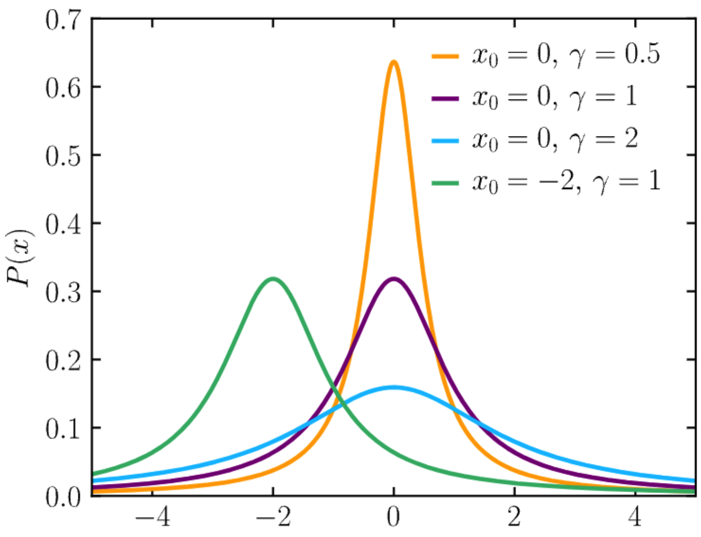
Red: $\mu = 0, \sigma = 1$. Green: $\mu = 0, \sigma = 0.6$. Blue: $\mu = 4, \sigma = 1$. Black: $\mu = 4, \sigma = 2$.

2. Cauchy Distribution

The Cauchy random variable with parameters μ, σ^2 has pdf:

$$p(x) = \frac{1}{\pi\sigma\left[1 + \frac{(x-\mu)^2}{\sigma^2}\right]} = \frac{1}{\pi\left[\sigma^2 + (x-\mu)^2\right]}$$

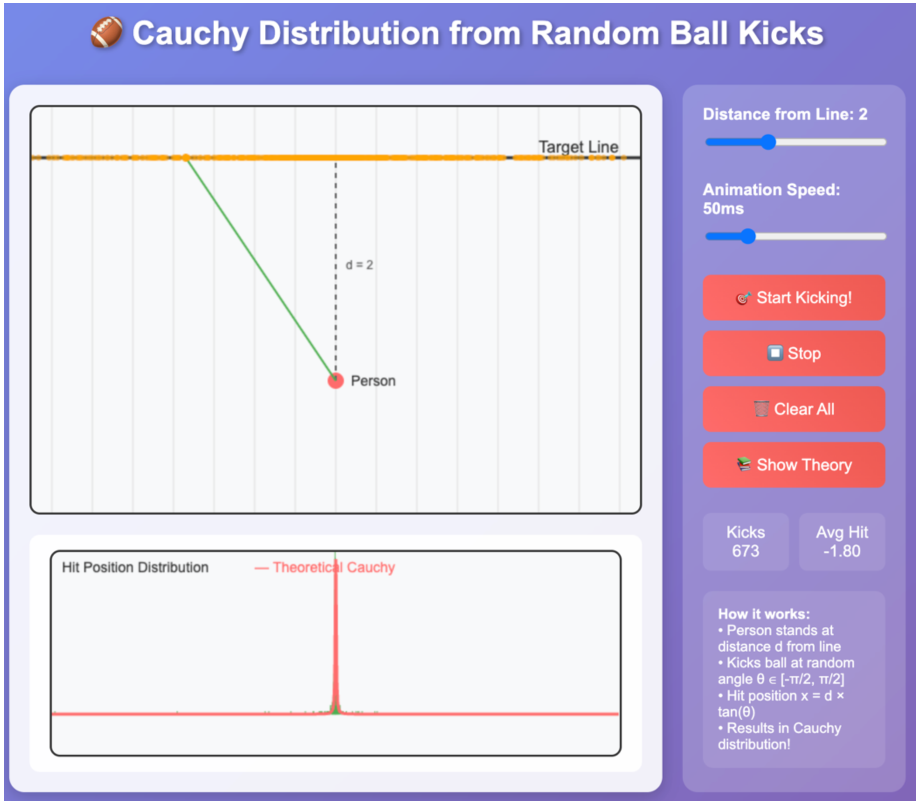
the Cauchy distribution has no mean; the parameter μ is the median.



Example: Kicking a Ball at a Random Angle

If one stands in front of a line and kicks a ball with a direction (more precisely, an angle) uniformly at random towards the line, then the distribution of the point where the ball hits the line is a Cauchy distribution.

I created the following activities by claude.ai



https://drive.google.com/file/d/1q0RZbHXGxJPdIWo38iSz2r-EbPPL9Lgt/view?usp=sharing

Reason: (later in the section of transformation)

The connection between the angle $\theta \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ and the hitting point X

$$\tan(\theta) = \frac{X}{d} \qquad\qquad X = d \cdot \tan(\theta)$$

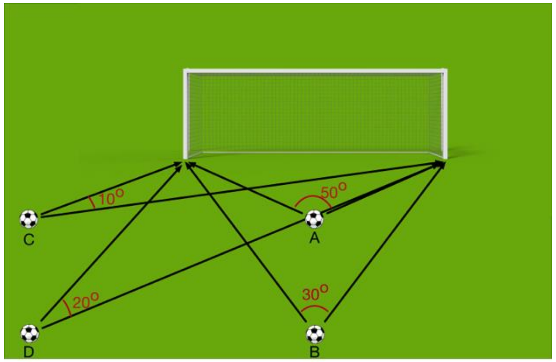
The probability density function (PDF) for a uniform distribution of θ

$$f(\theta) = \frac{1}{\pi} \text{ for } \theta \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$$

Use the method of transformation of variables we can find

$$f_X(x) = \frac{1}{\pi}\frac{d}{d^2 + x^2}$$

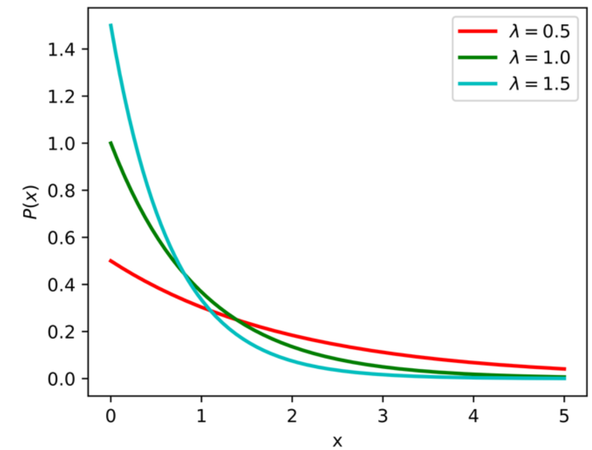
Question: Restrictions on angle and on kick point.



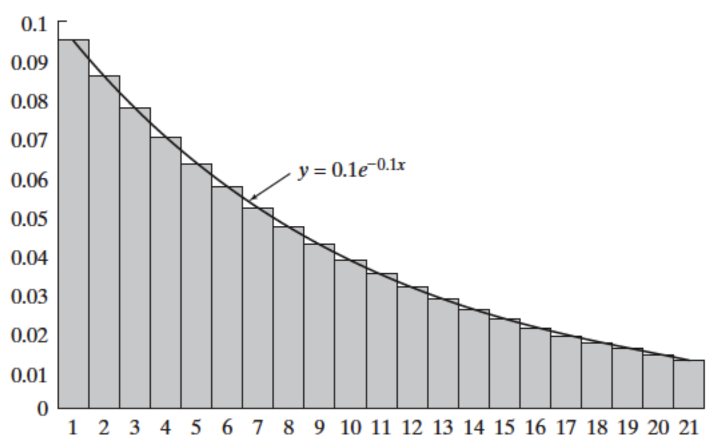
3. Exponential random variable is a continuous random variable with pdf given by

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

where λ is a fixed positive number.



- Exponential distribution is a contiguous analogue of the geometric distribution.



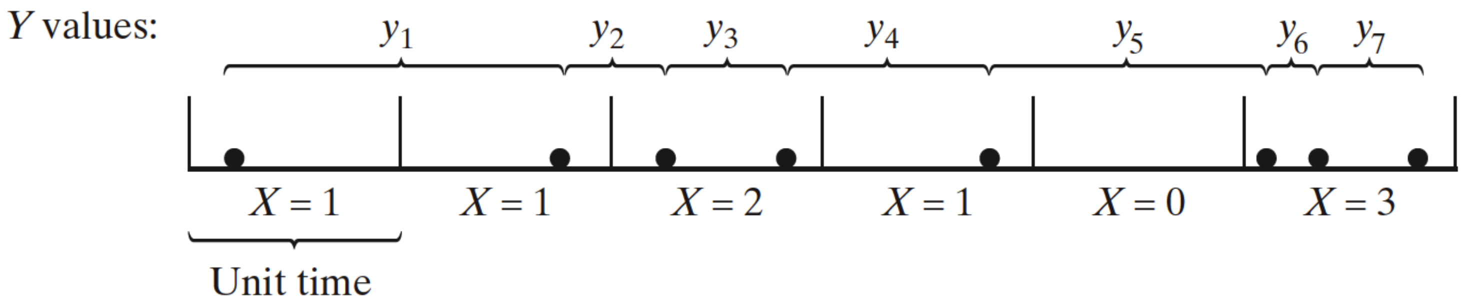
$$P_X(k) = (1 - p)^{k-1} p$$

$$p = 0.095$$

Exponential v.s. Poisson

- Exponential distribution models the time between occurrences in a time interval.

Example: Exponential distribution is often used to model the time until failure of a device.

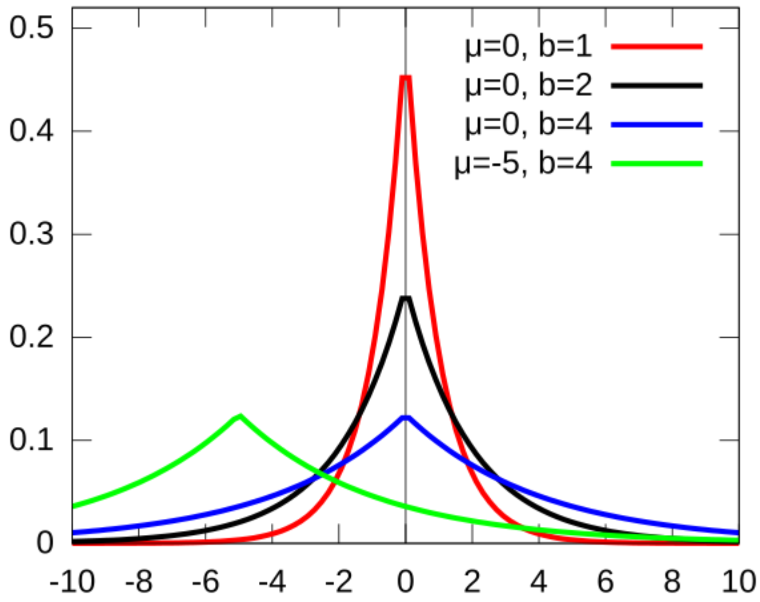


- A double exponential random variable X is that $|X|$ is an exponential random variable. (also called **Laplace distribution**)

$$f_X(x) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right)$$

CDF:

$$F(x) = \begin{cases} \frac{1}{2} \exp\left(\frac{x-\mu}{b}\right) & \text{for } x \geq \mu \\ 1 - \frac{1}{2} \exp\left(-\frac{x-\mu}{b}\right) & \text{for } x < \mu \end{cases}$$



Remark: the maximum likelihood (MLE) estimator of μ is the sample median.

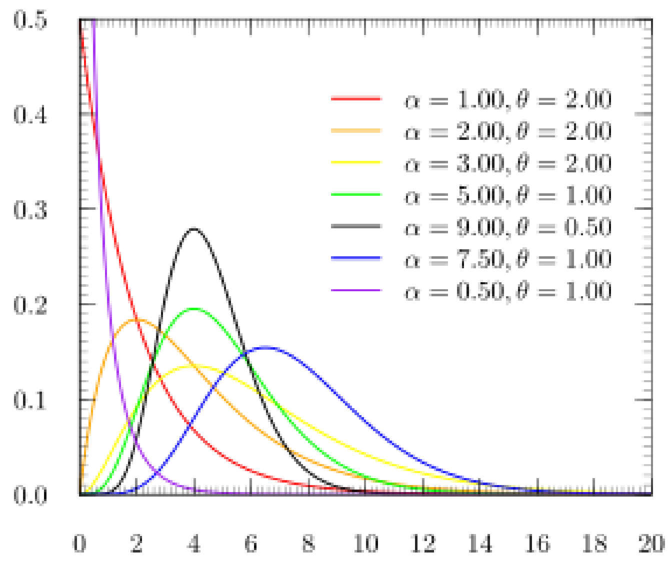
4. Gamma Distribution

Gamma Distribution $(Gamma(\alpha, \theta))$

$$p(x; \alpha, \theta) = \frac{x^{\alpha-1} e^{-\frac{x}{\theta}}}{\theta^{\alpha} \Gamma(\alpha)} \qquad \text{for } x \geq 0$$

Here, $\Gamma(\alpha)$ is the gamma function.

$$\Gamma(\alpha) := \int_0^{\infty} t^{\alpha-1} e^{-t} dt \ .$$



Applications:

Gamma Distribution is popular as a prior on coefficients. Obtained from integral over waiting times in Poisson distribution.

Erlang distribution is the distribution of the time until the k-th event of a Poisson process with a rate of λ.

Example: Waiting times in *Queueing Systems*: Erlang distribution models number of telephone calls which might be made at the same time to the operators of the switching stations.

Another popular parameterizations $Gamma(\alpha, \beta)$ with $\beta = \frac{1}{\theta}$

There is an interesting relationship between the gamma and Poisson distributions. If $X \sim Gamma(\alpha, \beta)$, where α is an integer, and $Y \sim Poisson\left(\frac{x}{\beta}\right)$ then for any x,

$$P(X \leq x) = P(Y \geq \alpha)$$

Special cases:

- $Gamma(1, \lambda) = \mathrm{Exp}(\lambda)$ the **exponential distribution**.

- If $\alpha = k$ is a positive integer, $\Gamma(k) = (k-1)!$ the distribution $Gamma(k, \theta)$ is called **Erlang distribution**.

- **Weibull distribution**

  If $X \sim \text{exponential}(\beta)$ then $Y = X^{1/\gamma}$ has a $\text{Weibull}(\gamma, \beta)$ distribution

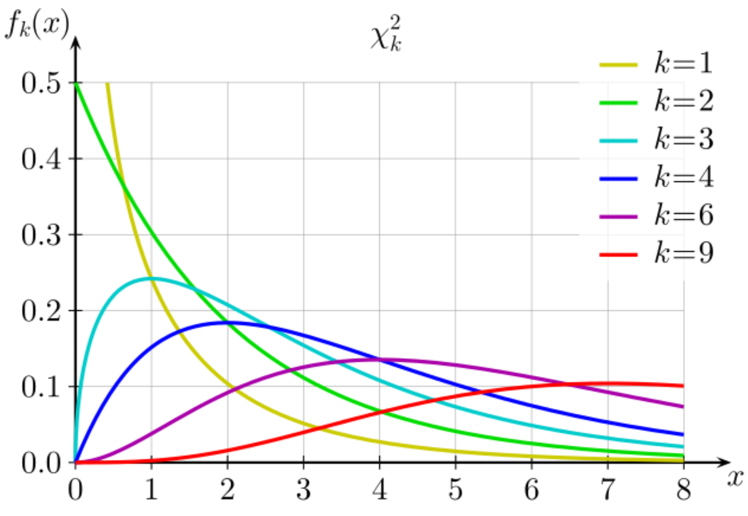
$$f_Y(y|\gamma, \beta) = \frac{\gamma}{\beta} y^{\gamma-1} e^{-y^\gamma/\beta}, \quad 0 < y < \infty, \quad \gamma > 0, \quad \beta > 0.$$

- **Chi-squared distribution**

Chi-squared distribution χ_k^2 with degree freedom k is a special case of $Gamma$ distribution

$$\chi_k^2 \sim Gamma\left(\alpha = \frac{k}{2}, \theta = 2\right)$$

Chi-squared distribution χ_k^2 is the distribution of a sum of the squares of k independent standard normal random variables $Normal(0,1)$



5. Beta Distribution

Beta Distribution ($Beta(\alpha, \beta)$**)** is often used as **prior** on Binomial distributions (it is a conjugate prior).

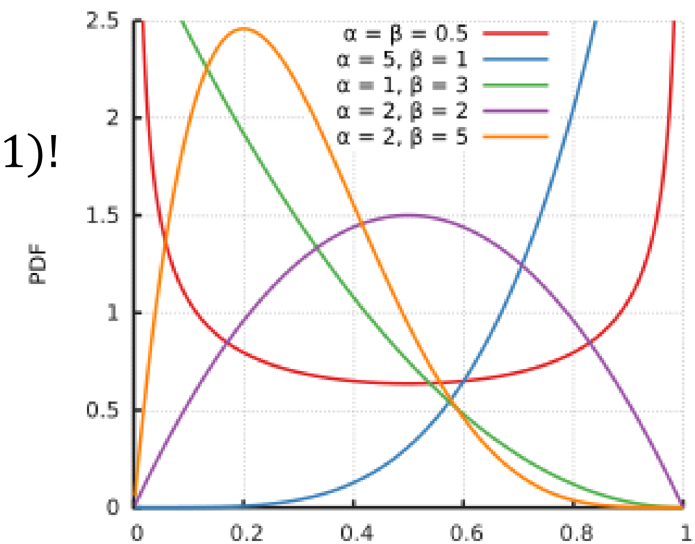
$$f(x; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha)} x^{\alpha-1}(1-x)^{\beta-1}$$

where $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$ is the Gamma function.

Usually denote $\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha)} = \frac{1}{B(\alpha,\beta)}$

If $z = k$ is an positive integer, then $\Gamma(k) = (k-1)!$

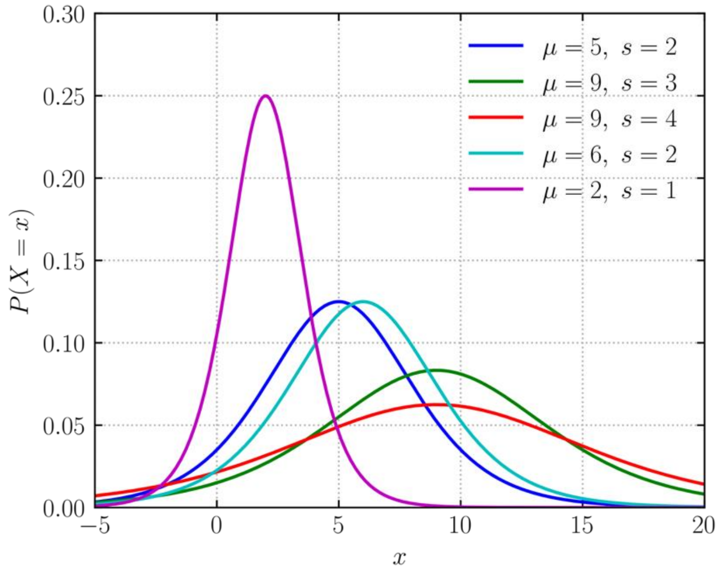
Example: The beta distribution is a suitable model for the random behavior of percentages and proportions.



6. **Logistic Distribution**

The logistic distribution is a random variable with pdf

$$p(x) = \frac{\beta e^{-\alpha-\beta x}}{(1+e^{-\alpha-\beta x})^2} = \frac{\beta e^{\alpha+\beta x}}{(1+e^{\alpha+\beta x})^2}$$



It has has a CDF

$$F(x) = \frac{1}{1+e^{-\alpha-\beta x}}$$

$$= \frac{1}{2} + \frac{1}{2}\tanh\left(\frac{x-\mu}{2s}\right)$$

- One of the most common applications is in logistic regression, which is used for modeling categorical dependent variables.
- The United States Chess Federation and FIDE have switched its formula for calculating chess ratings from the normal distribution to the logistic distribution.

Summary of distributions:

Random variables and a useful `story' that describes how they arise.

Name	Story
$Bernoulli(p)$	Toss a coin with probability p of turning up heads. X = Number of heads in one toss.
$Binomial(n, p)$	Toss a coin with probability p of turning up heads. X = Number of heads in n tosses. Binomial(n, p) is the sum of n independent Bernoulli(p).
$Geometric(p)$	Toss a coin with probability p of turning up heads. X = Number of tosses until the first Head.
$Poisson(\lambda)$	Random calls arrive with rate λ . X = Number of calls that arrive in one time unit.
$Exponential(\lambda)$	Random calls arrive with rate λ . X = Time until the first arrival.
$Gamma(n, \lambda)$	Random calls arrive with rate λ . X = Time until the n-th arrival.
$Uniform(a, b)$	Pick a random number X between a and b .
$Normal(\mu, \sigma^2)$	Pick an individual in a large population. X = Height of the individual.
$Beta(\alpha, \beta)$	Pick a random number X representing the probability of success in α successes and β failures.

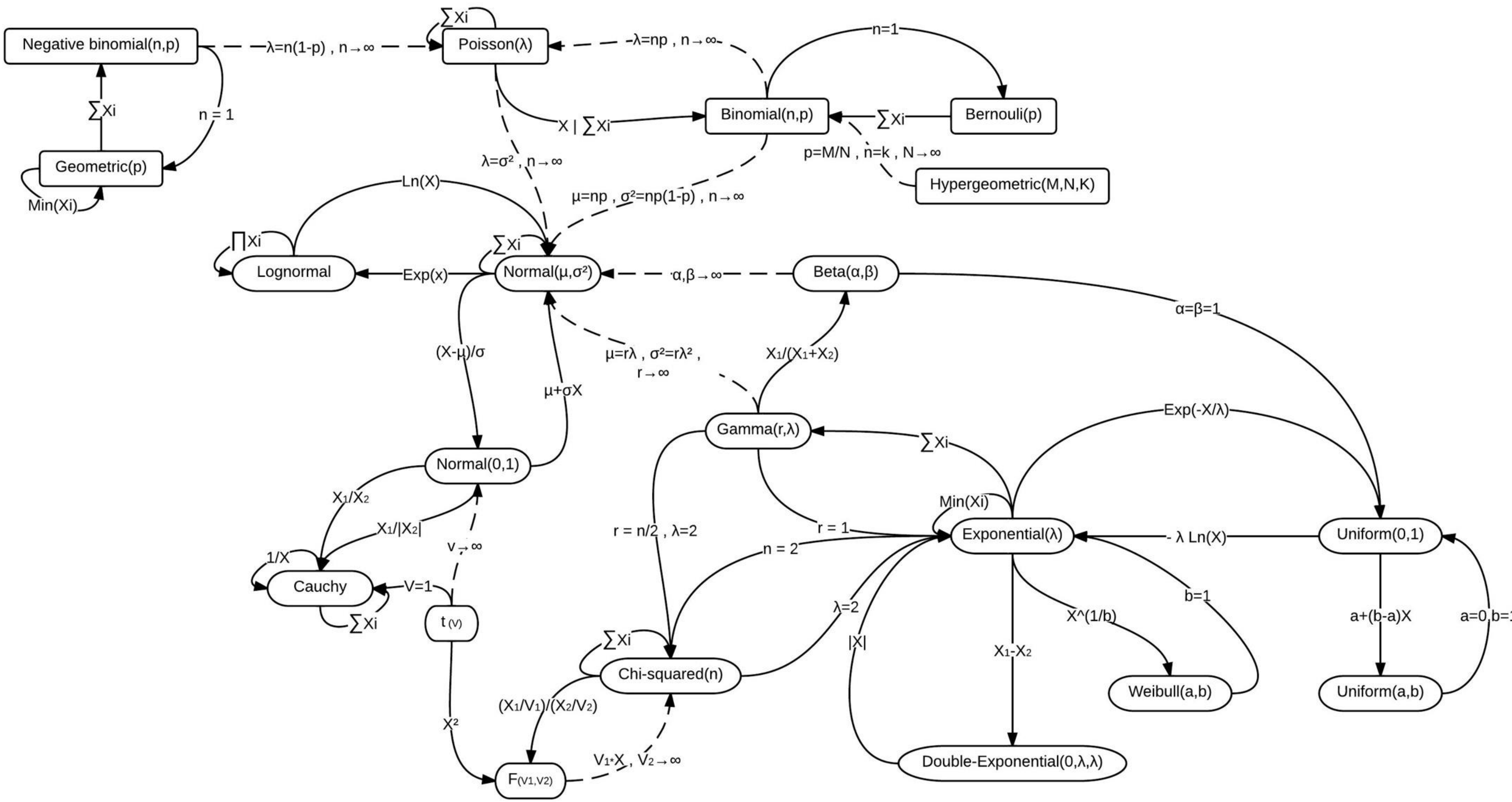
Lists the **pdf's** of these random variables.

Name	Pmf/pdf
$Bernoulli(p)$	$p_X(k) = p^k(1-p)^{1-k}$
$Binomial(n, p)$	$p_X(k) = \binom{n}{k} p^k(1-p)^{n-k}$
$Geometric(p)$	$p_X(k) = (1-p)^{k-1}p$
$Poisson(\lambda)$	$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}$
$Exponential(\lambda)$	$f_X(k) = \lambda e^{-\lambda x}$
$Gamma(n, \lambda)$	$f(x; k, \theta) = \frac{x^{k-1}e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)}$
$Uniform(a, b)$	$f_X(x) = \frac{1}{b-a}$
$Normal(\mu, \sigma^2)$	$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$
$Beta(\alpha, \beta)$	$f(x; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha)} x^{\alpha-1}(1-x)^{\beta-1}$

Lists the means, variances these random variables.

Name	**Mean**	**Variance**
$Bernoulli(p)$	p	$p(1-p)$
$Binomial(n,p)$	np	$np(1-p)$
$Geometric(p)$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
$Poisson(\lambda)$	λ	λ
$Exponential(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$Gamma(n,\lambda)$	$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$
$Uniform(a,b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$Normal(\mu,\sigma^2)$	μ	σ^2
$Beta(\alpha,\beta)$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Relationships among probability distributions



https://en.wikipedia.org/wiki/Relationships_among_probability_distributions

References:

- **Book 1. [CB] Statistical Inference**, by Casella, George, Berger, Roger L, 2nd edition
- **Book 2. [W]: All of Statistics: Larry Wasserman**
- **Book 3. Introduction to Probability**. C.M. Grinstead and J.L. Snell. American Mathematical Society, 2012
- **Book 4. Introduction to Probability Models**, S. Ross, 12th edition (published by Academic Press).

Online books:

https://www.probabilitycourse.com/

Extra Reading:

Baby Measure Theory: https://www.stat.umn.edu/geyer/8501/measure.pdf

YouTube video about coin flips by a famous statistician,
YouTube video about dice rolls (Part I and Part II).