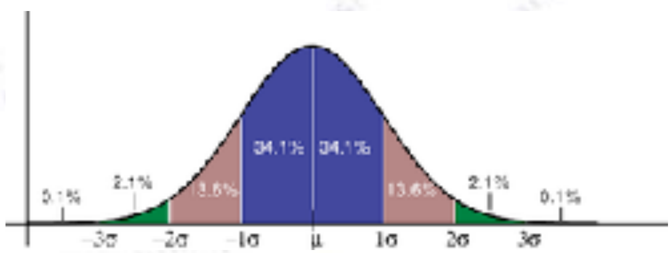
Applied Statistics

Probability Density Functions (PDFs)



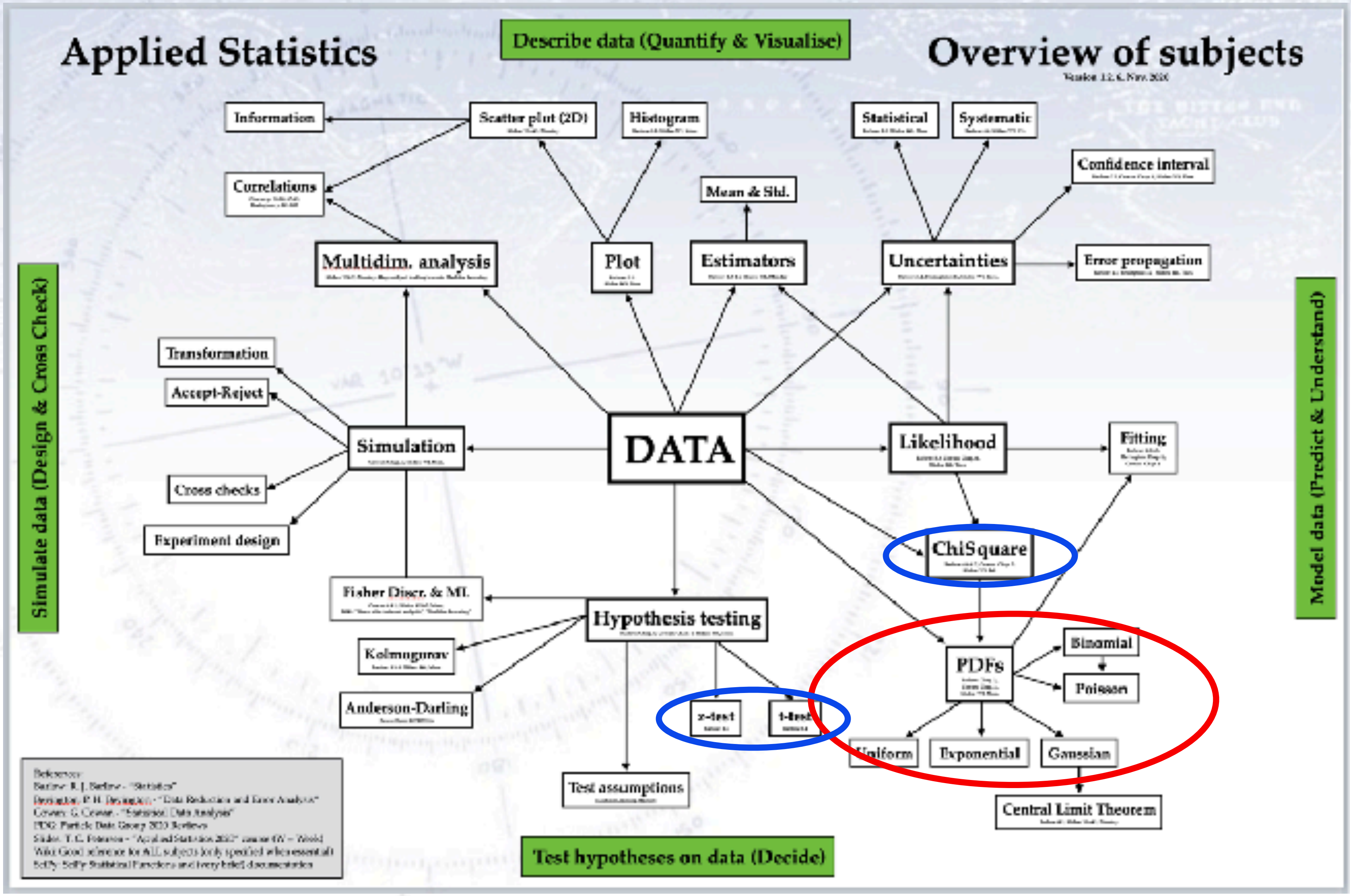



Troels C. Petersen



"Statistics is merely a quantisation of common sense"

Probability Density Functions



Probability Density Functions

A Probability Density Function (PDF) f(x) describes the probability of an outcome x:

probability to observe x in the interval [x, x+dx] = f(x) dx

PDFs are required to be normalised:

$$\int_S f(x)dx = 1$$

The expectation value (aka. mean) and the variance (i.e. standard deviation squared) are then defined as follows:

$$\mu = \int_{-\infty}^{\infty} x f(x) dx$$

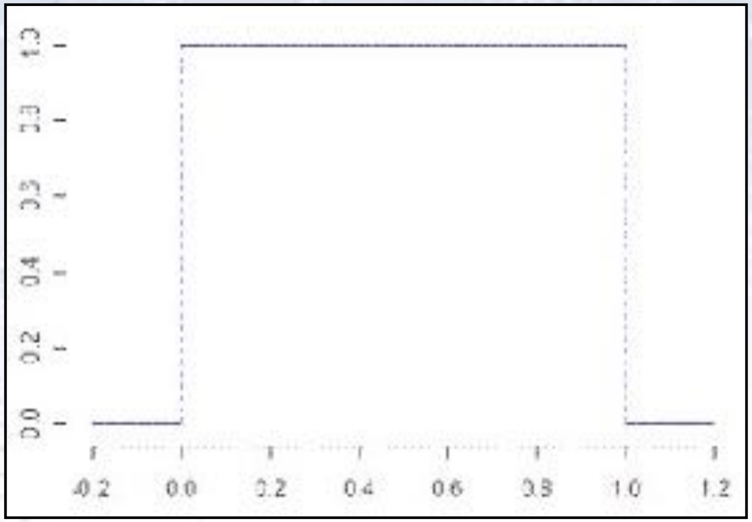
$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

Probability Density Functions

Example:

Consider a uniform distribution:

$$f(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & else \end{cases}$$



Calculating the mean and variance:

$$\mu = \int_{-\infty}^{\infty} x f(x) dx = \int_0^1 x dx = [\frac{1}{2}x^2]_0^1 = \frac{1}{2}$$

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx = \int_0^1 (x - \frac{1}{2})^2 dx =$$

$$[\frac{1}{3}x^3 - \frac{1}{2}x^2 + \frac{1}{4}x]_0^1 = \frac{1}{3} - \frac{1}{2} + \frac{1}{4} = \frac{1}{12}$$

Cumulative distributions functions

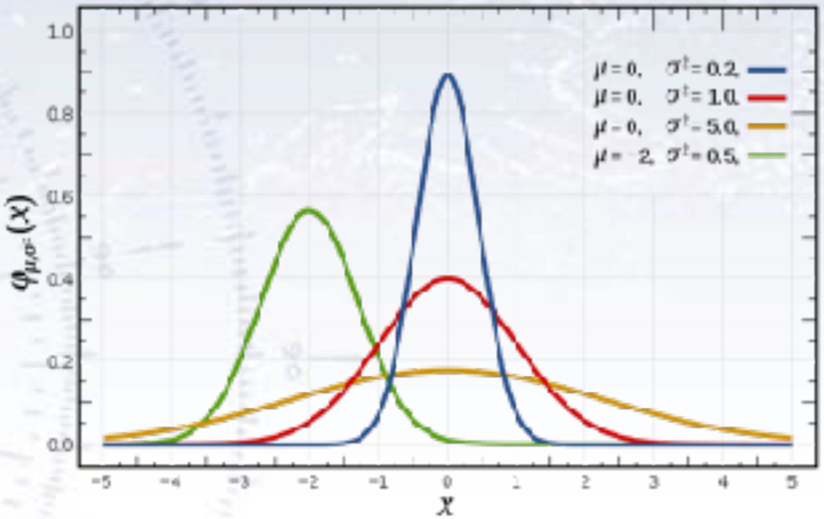
Completely basic to every PDF is the **cumulative distribution function,** CDF, defined as:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt.$$

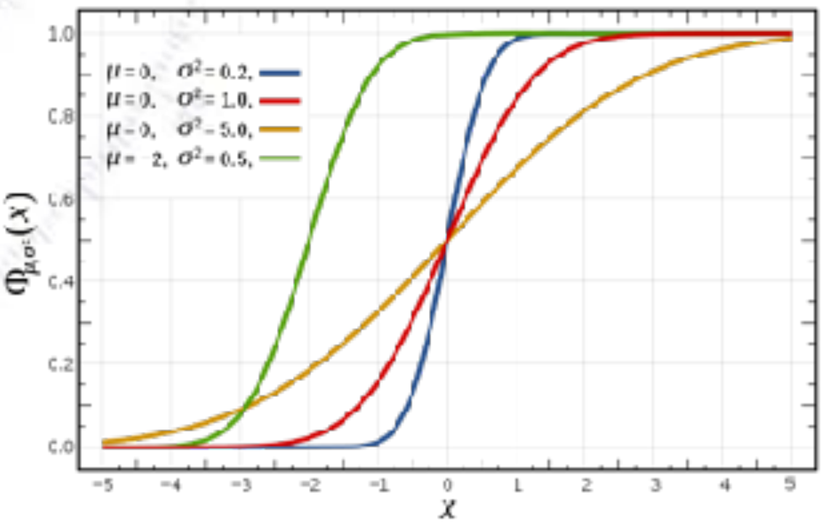
In words, this means that it is the probability of getting x, or something below that value.

The CDF is used in many ways, and we will meet it again soon, when we discuss hypothesis testing.

Gaussian PDF



Gaussian CDF



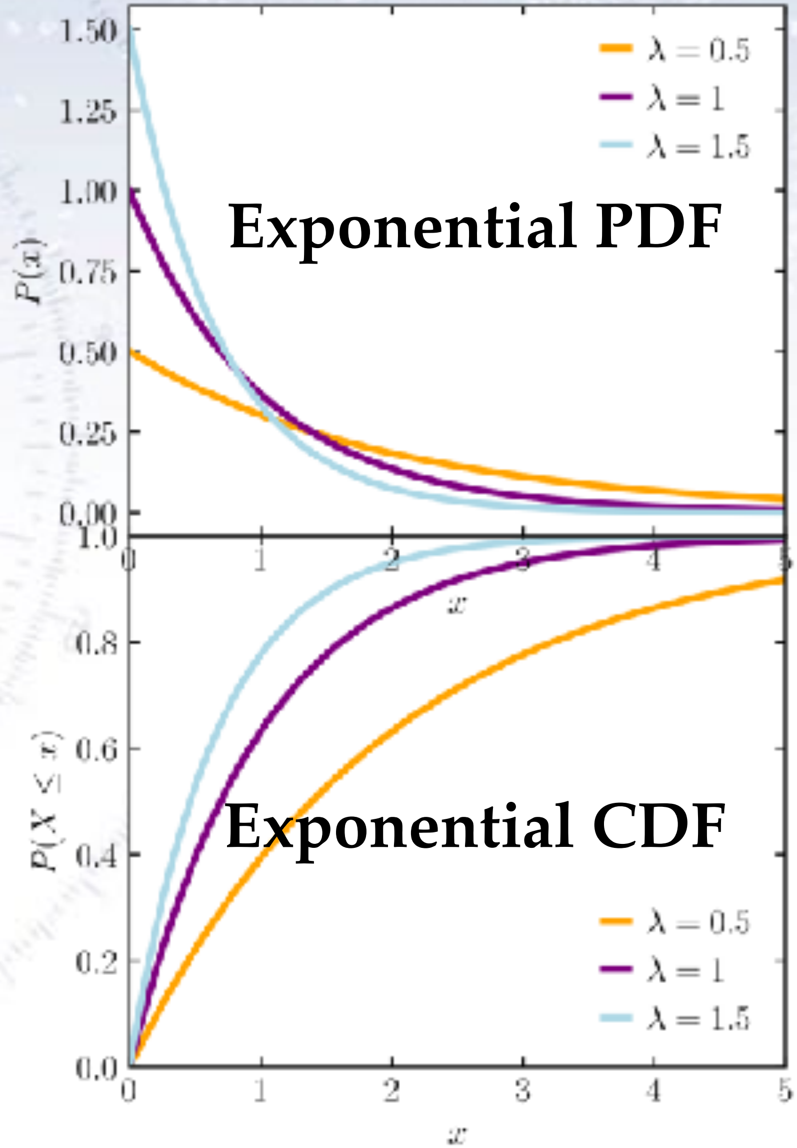
Cumulative distributions functions

Completely basic to every PDF is the **cumulative distribution function,** CDF, defined as:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt.$$

In words, this means that it is the probability of getting x, or something below that value.

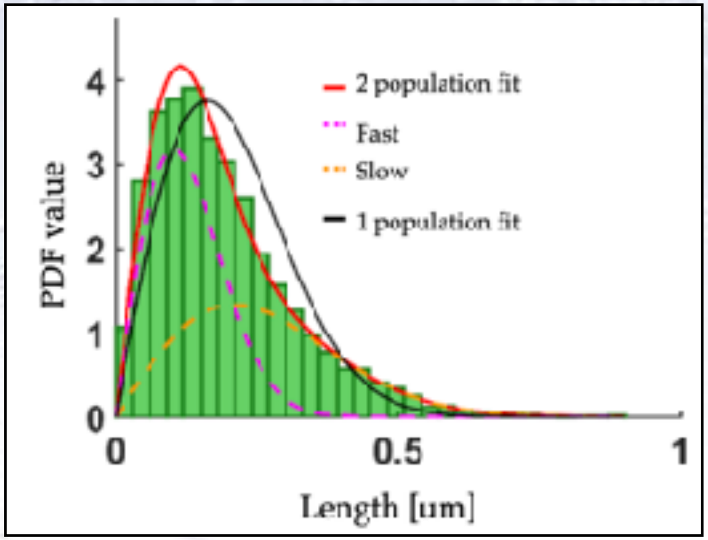
The CDF is used in many ways, and we will meet it again soon, when we discuss hypothesis testing.



Why PDFs

Could we not just use mean and variance and call it a day?

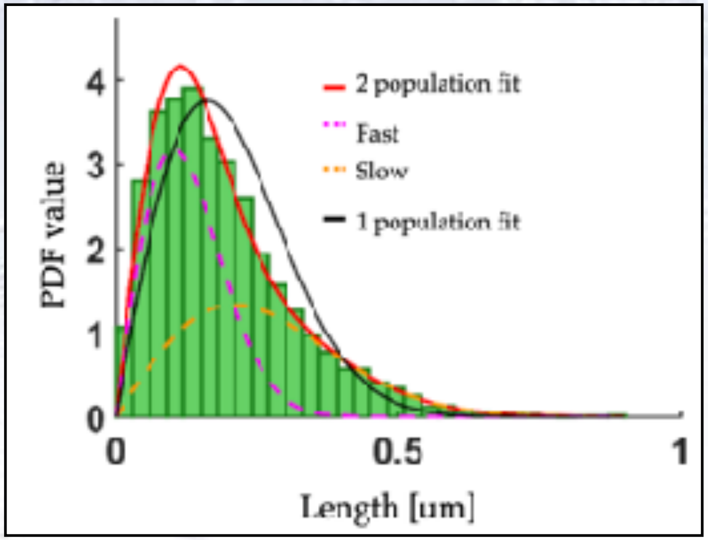
Well, PDFs makes us able to ask what the *probability* of a certain event given the underlying model (i.e. PDF), and this allows for new discoveries!



Why PDFs

Could we not just use mean and variance and call it a day?

Well, PDFs makes us able to ask what the *probability* of a certain event given the underlying model (i.e. PDF), and this allows for new discoveries!

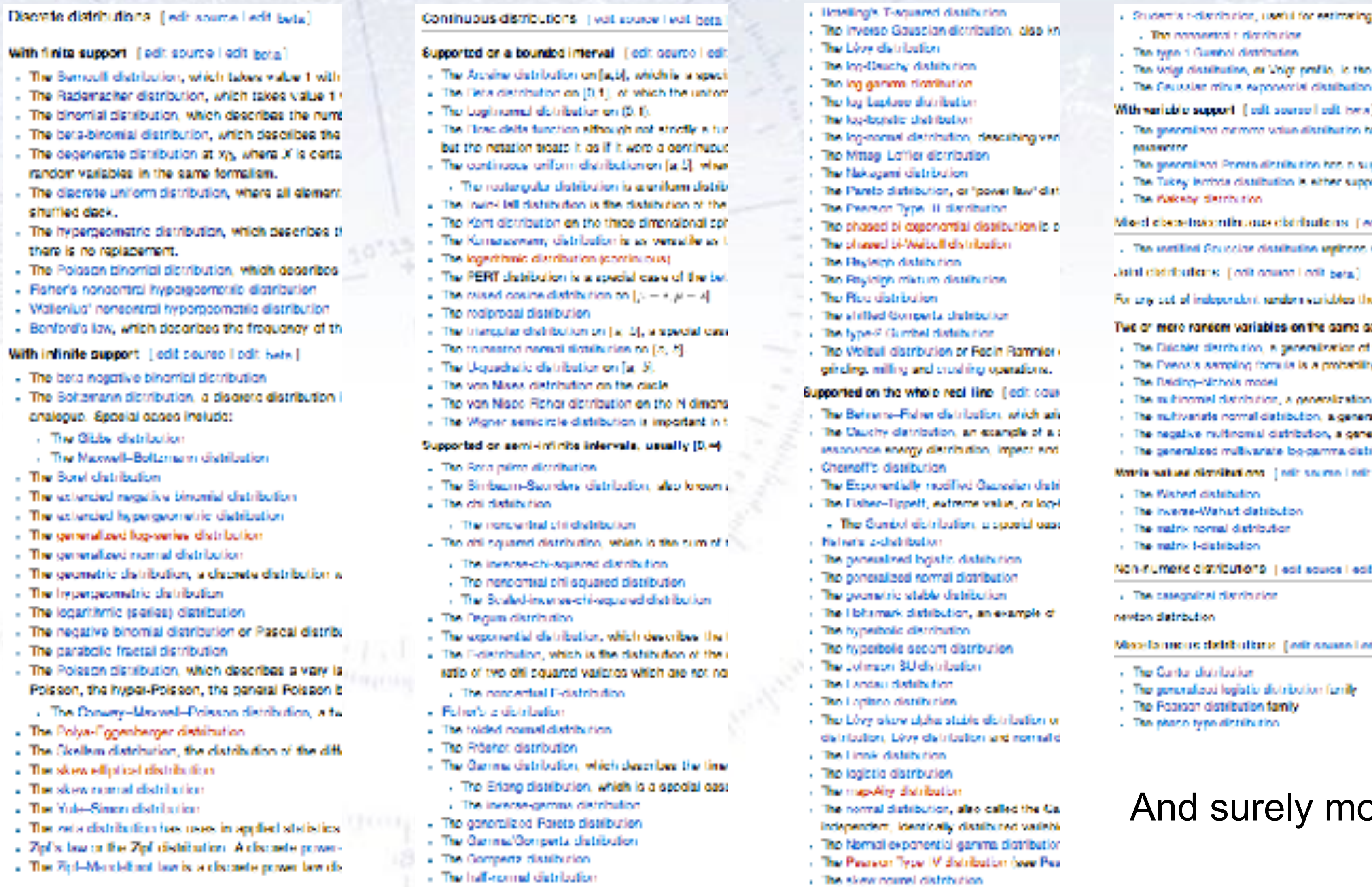


On notation:
In the literature it is often we use large letters for a random variable X. This means an *outcome* for an event! If I roll a die, we say that X takes on values in {1,2,3,4,5,6}, which is a *discrete* case.

Small letters are typically real numbers. So we could write: P(X < c), which translated means that we calculate the probability P that in one event X, we obtain a value of X smaller than the real value c.

Probability Density Functions

The number of PDFs is infinite, and nearly so is the list of known ones:



Discrete distributions

With finite support

- The Bernoulli distribution
- The Rademacher distribution
- The binomial distribution
- The beta-binomial distribution
- The degenerate distribution
- The discrete uniform distribution
- The hypergeometric distribution
- The Poisson binomial distribution
- Fisher's noncentral hypergeometric distribution
- Wallenius' noncentral hypergeometric distribution
- Benford's law

With infinite support

- The beta negative binomial distribution
- The Boltzmann distribution
  - The Gibbs distribution
  - The Maxwell–Boltzmann distribution
- The Borel distribution
- The extended negative binomial distribution
- The extended hypergeometric distribution
- The generalized log-series distribution
- The geometric distribution
- The logarithmic (series) distribution
- The negative binomial distribution or Pascal distribution
- The parabolic fractal distribution
- The Poisson distribution
  - The Conway–Maxwell–Poisson distribution
- The Polya–Eggenberger distribution
- The skew elliptical distribution
- The Yule–Simon distribution
- Zipf's law or the Zipf distribution
- The Zipf–Mandelbrot law

Continuous distributions

Supported on a bounded interval

- The Arcsine distribution
- The Beta distribution
- The Logitnormal distribution
- The Dirac delta function
- The continuous uniform distribution
  - The rectangular distribution
- The Kent distribution
- The Kumaraswamy distribution
- The logarithmic distribution (continuous)
- The PERT distribution
- The raised cosine distribution
- The reciprocal distribution
- The triangular distribution
- The truncated normal distribution
- The U-quadratic distribution
- The von Mises distribution
- The von Mises–Fisher distribution
- The Wigner semicircle distribution

Supported on semi-infinite intervals, usually [0,∞)

- The Beta prime distribution
- The Birnbaum–Saunders distribution
- The chi distribution
  - The noncentral chi distribution
- The chi-squared distribution
  - The inverse-chi-squared distribution
  - The noncentral chi-squared distribution
  - The Scaled-inverse-chi-squared distribution
- The Dagum distribution
- The exponential distribution
- The F-distribution
  - The noncentral F-distribution
- Fisher's z-distribution
- The folded normal distribution
- The Fréchet distribution
- The Gamma distribution
  - The Erlang distribution
  - The inverse-gamma distribution
- The generalized Pareto distribution
- The Gamma/Gompertz distribution
- The Gompertz distribution
- The half-normal distribution
- Hotelling's T-squared distribution
- The inverse Gaussian distribution
- The Lévy distribution
- The log-Cauchy distribution
- The log gamma distribution
- The log-Laplace distribution
- The log-logistic distribution
- The log-normal distribution
- The Mittag-Leffler distribution
- The Nakagami distribution
- The Pareto distribution
- The Pearson Type III distribution
- The phased bi-exponential distribution
- The phased bi-Weibull distribution
- The Rayleigh distribution
- The Rayleigh mixture distribution
- The Rice distribution
- The shifted Gompertz distribution
- The type-2 Gumbel distribution
- The Weibull distribution or Rosin Rammler distribution

Supported on the whole real line

- The Behrens–Fisher distribution
- The Cauchy distribution
- Chernoff's distribution
- The Exponentially modified Gaussian distribution
- The Fisher–Tippett, extreme value, or log-Weibull distribution
  - The Gumbel distribution
- Fisher's z-distribution
- The generalized logistic distribution
- The generalized normal distribution
- The geometric stable distribution
- The Holtsmark distribution
- The hyperbolic distribution
- The hyperbolic secant distribution
- The Johnson SU distribution
- The Landau distribution
- The Laplace distribution
- The Lévy skew alpha-stable distribution
- The Linnik distribution
- The logistic distribution
- The map-Airy distribution
- The normal distribution
- The Normal-exponential-gamma distribution
- The Pearson Type IV distribution
- The skew normal distribution
- Student's t-distribution
  - The noncentral t-distribution
- The type-1 Gumbel distribution
- The Voigt distribution
- The Gaussian minus exponential distribution

With variable support

- The generalized extreme value distribution
- The generalized Pareto distribution
- The Tukey lambda distribution
- The Wakeby distribution

Mixed discrete/continuous distributions

- The rectified Gaussian distribution

Joint distributions

Two or more random variables on the same sample space

- The Dirichlet distribution
- The Ewens's sampling formula
- The Balding–Nichols model
- The multinomial distribution
- The multivariate normal distribution
- The negative multinomial distribution
- The generalized multivariate log-gamma distribution

Matrix-valued distributions

- The Wishart distribution
- The inverse-Wishart distribution
- The matrix normal distribution
- The matrix t-distribution

Non-numeric distributions

- The categorical distribution

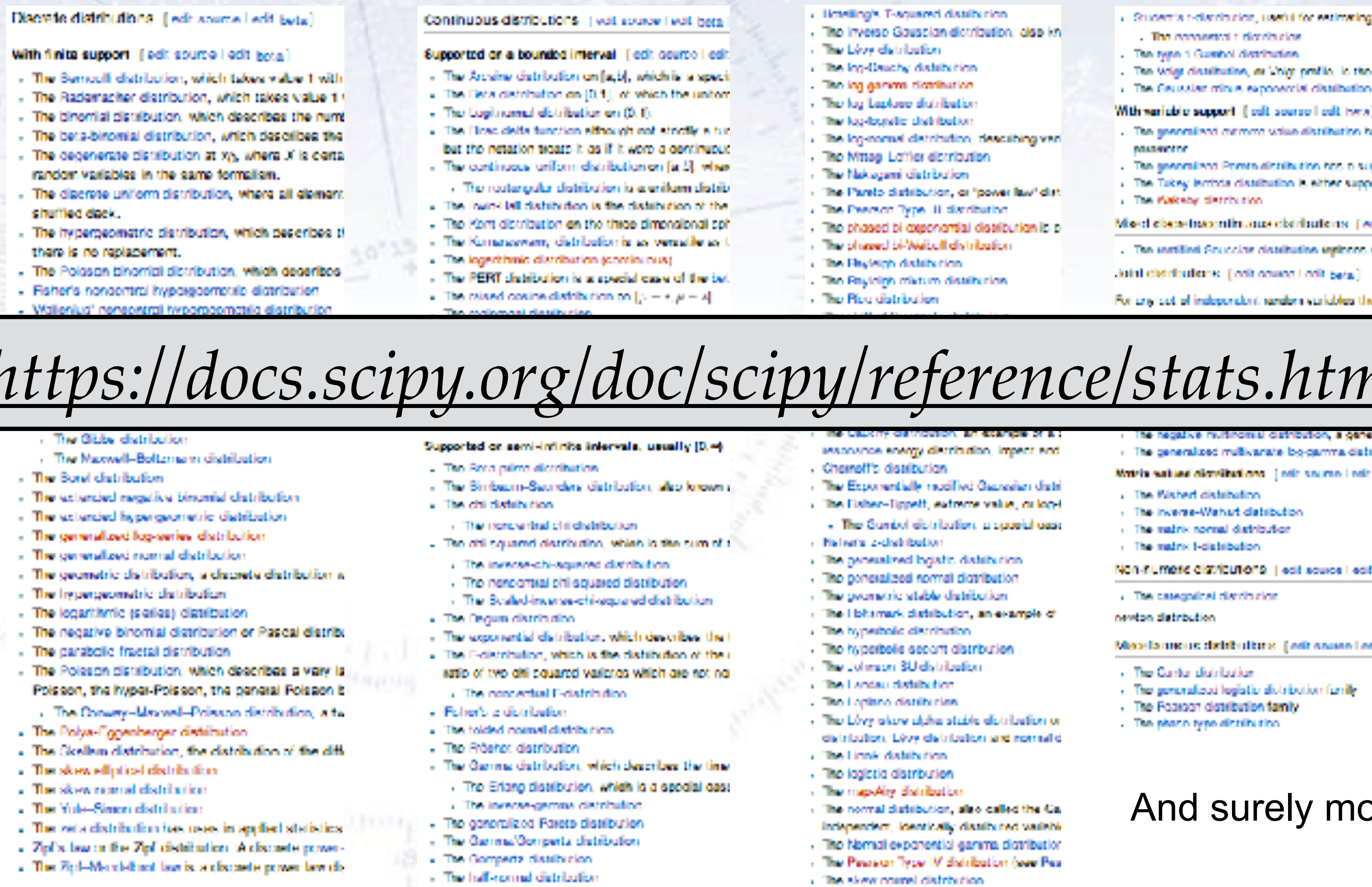
Miscellaneous distributions

- The Cantor distribution
- The generalized logistic distribution family
- The Poisson distribution family
- The phase-type distribution

And surely more!

Probability Density Functions

The number of PDFs is infinite, and nearly so is the list of known ones:

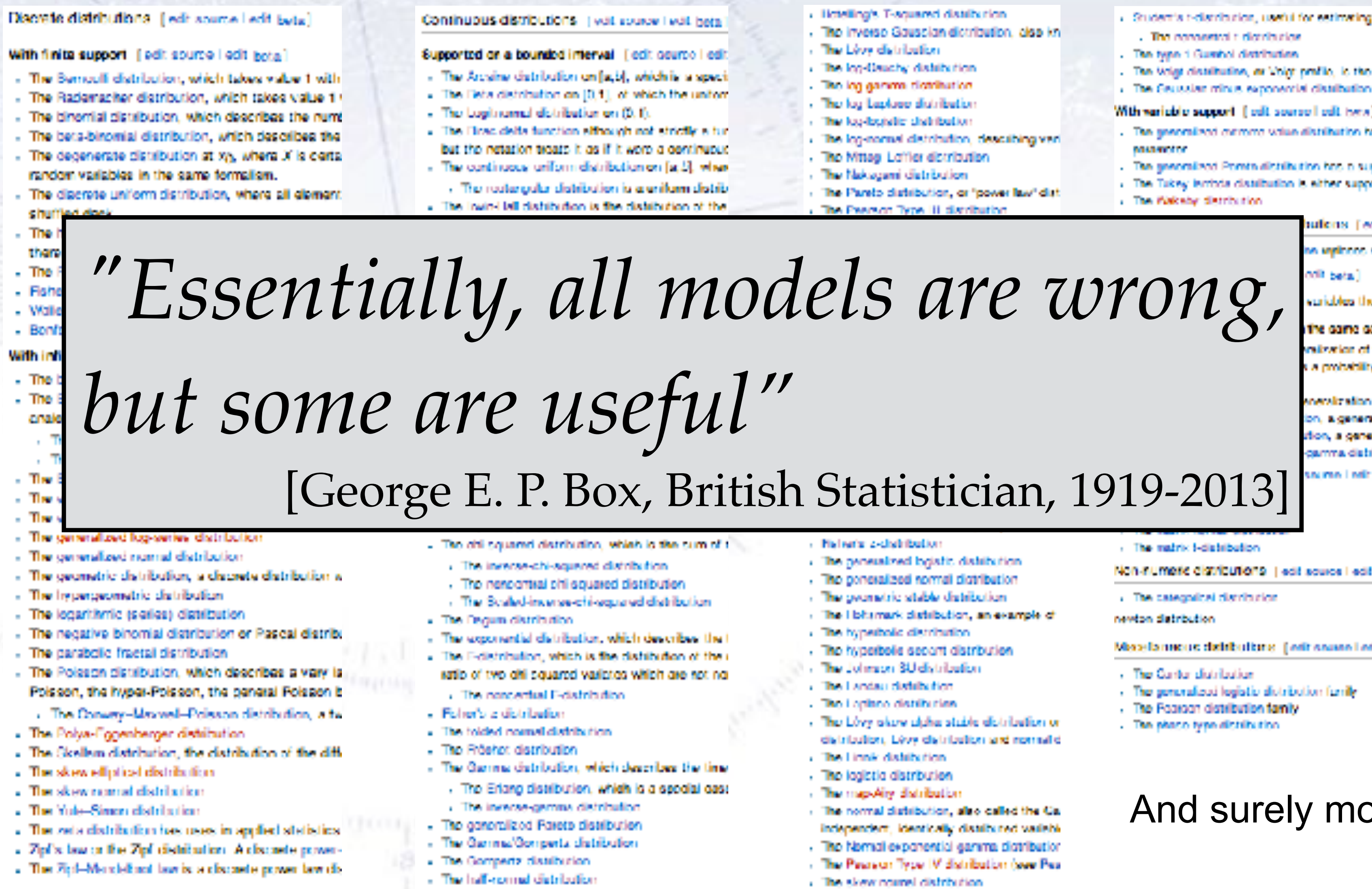


<u>https://docs.scipy.org/doc/scipy/reference/stats.html</u>

And surely more!

Probability Density Functions

The number of PDFs is infinite, and nearly so is the list of known ones:



"Essentially, all models are wrong, but some are useful"

[George E. P. Box, British Statistician, 1919-2013]

And surely more!

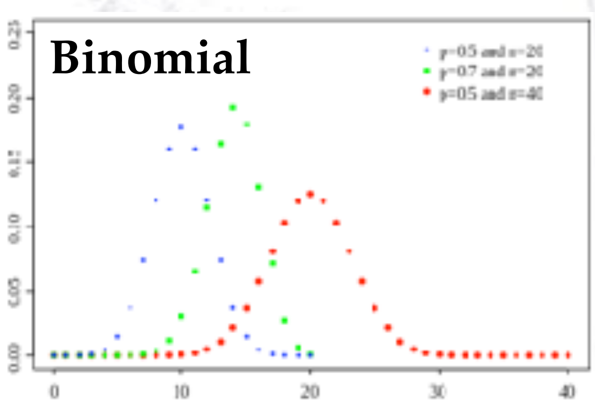
Probability Density Functions

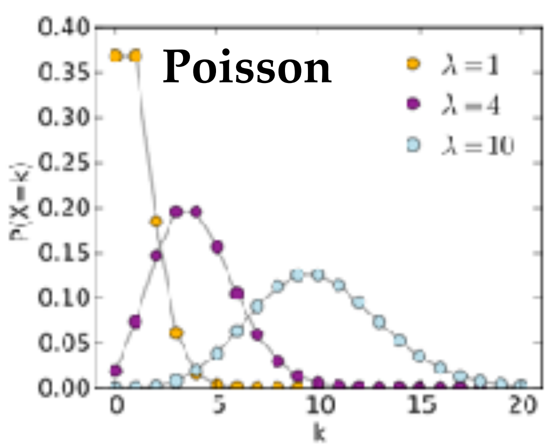
An almost complete list of those we will deal with in this course is:

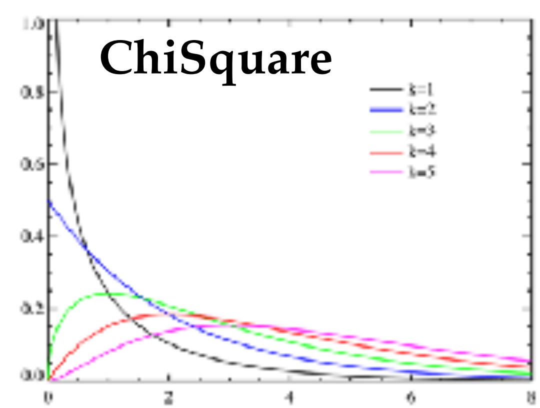
- **Gaussian** (aka. Normal)
- **Poisson**
- **Binomial** (and also Multinomial)
- Uniform
- Exponential
- ChiSquare
- Students t-distribution

See Barlow chap.3 and Cowan chap.2

You should already know most of these, and the rest will be explained.







Binomial, Poisson, Gaussian

$$f(n;N,p) = \frac{N!}{n!(N-n)!}p^n(1-p)^{N-n}$$

Given **N trials** each with **p chance of success**, how many **successes n** should you expect in total?

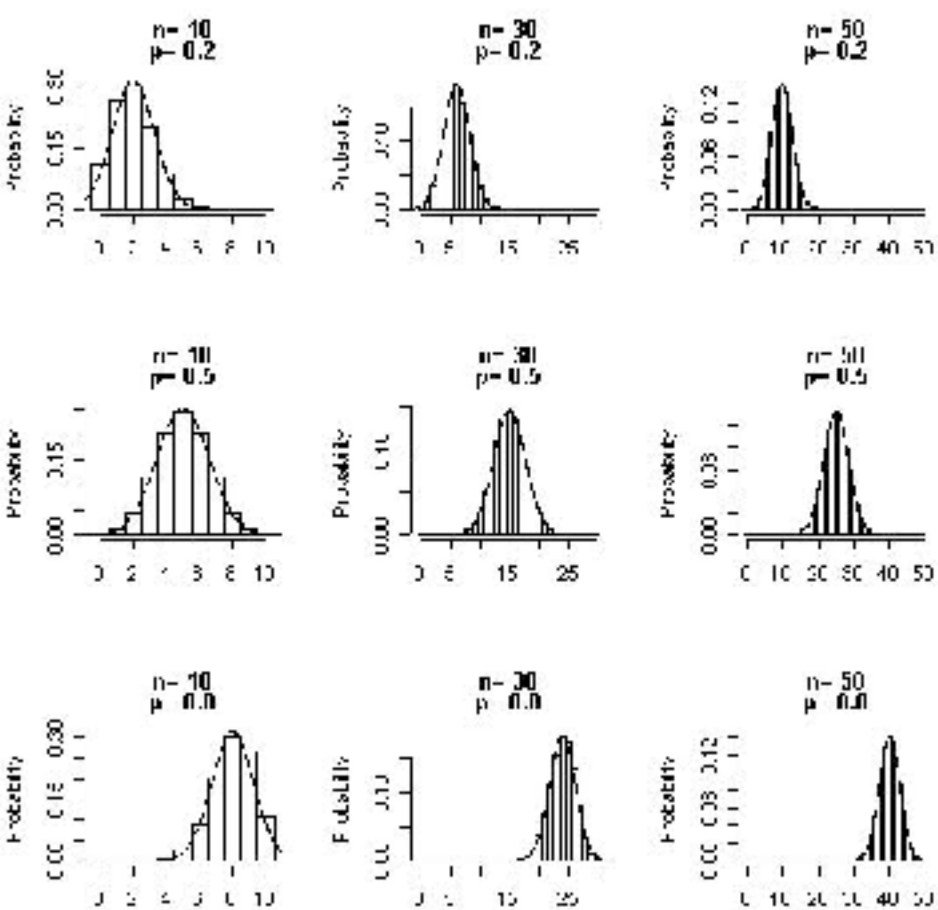
This distribution is… **Binomial**, with

Mean = Np

Variance = Np(1-p)

This means, that the error on a fraction f = n/N is:

$$\sigma(f) = \sqrt{\frac{f(1-f)}{N}}$$



Binomial, Poisson, Gaussian

$$f(n;N,p) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$$

The binomial distribution was first introduced by Jacob Bernoulli in 1713 (posthumously).

The binomial distribution basically consists of two elements: The binomial coefficient (green) and the probabilities of exactly n such events (blue).

Even though a system has many outcomes, it is typically possible to refer to either “success” of “failure”.

$$(x+y)^4 = x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4.$$

0:	1							
1:	1	1						
2:	1	2	1					
3:	1	3	3	1				
4:	1	4	6	4	1			
5:	1	5	10	10	5	1		
6:	1	6	15	20	15	6	1	
7:	1	7	21	35	35	21	7	1
8: 1	8	28	56	70	56	28	8	1

Assume the probability to have COVID19 is 1%. In a sample of 50 people the chance to have 1 or more infected is: 1-p(0) = 1 - 0.99^{50} = 0.60

Binomial, Poisson, Gaussian

You count 100 cars, and see 15 red cars. What is your estimate of the fraction (i.e. probability) of red cars and its uncertainty?

a) 0.150 ± 0.050
b) 0.150 ± 0.026
c) 0.150 ± 0.036
d) 0.125 ± 0.030
e) 0.150 ± 0.081

From previous page: $$\sigma(f) = \sqrt{\frac{f(1-f)}{N}}$$

A friend tells you, that 8% of the cars on Blegdamsvej are red. What is the chance of that? Could he be right?

Binomial, Poisson, Gaussian

You count 100 cars, and see 15 red cars. What is your estimate of the fraction (i.e. probability) of red cars and its uncertainty?

a) 0.150 ± 0.050
b) 0.150 ± 0.026
c) 0.150 ± 0.036 ← (0.150 - 0.080) / 0.036 = 1.9 σ
d) 0.125 ± 0.030
e) 0.150 ± 0.081

From previous page: $$\sigma(f) = \sqrt{\frac{f(1-f)}{N}}$$

A friend tells you, that 8% of the cars on Blegdamsvej are red. What is the chance of that? Could he be right?

Binomial, Poisson, Gaussian

You count 100 cars, and see 15 red cars. What is your estimate of the fraction (i.e. probability) of red cars and its uncertainty?

a) 0.150 ± 0.050
b) 0.150 ± 0.026
c) 0.150 ± 0.036 ← **(0.150 - 0.080) / 0.036 = 1.9 σ**
d) 0.125 ± 0.030
e) 0.150 ± 0.081

Notice - this was actually a hypothesis test!

From previous page: $$\sigma(f) = \sqrt{\frac{f(1-f)}{N}}$$

A friend tells you, that 8% of the cars on Blegdamsvej are red. What is the chance of that? Could he be right?

Binomial, Poisson, Gaussian

Requirements to be Binomial:

- Fixed number of trials, N
- Independent trials.
- Only two outcomes (success/failure).
- Constant probability of success/failure.

If number of possible outcomes is more than two ⇒ **Multinomial distribution**.

Examples of Binomial experiments:

- Tossing a coin 20 times counting number of tails.
- Asking 200 people if they watch sports on TV.
- Rolling a die to see if a 6 appear (Multinomial for all outcomes!).
- Asking 100 die-hards from Enhedslisten, if they would vote for Konservative at the next election!

Examples which aren't Binomial experiments:

- Rolling a die until a 6 appears (not fixed number of trials).
- Drawing 5 cards for a poker hand (no replacement ⇒ not independent)

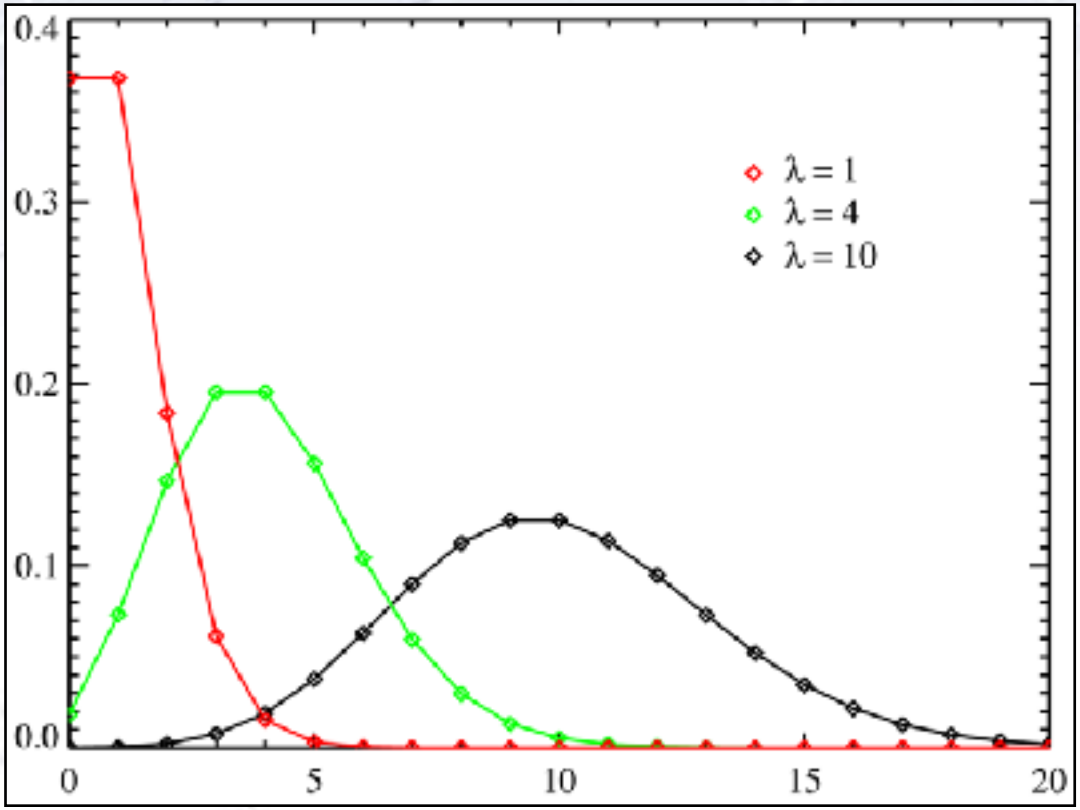
Binomial, **Poisson**, Gaussian

If $N \to \infty$ and $p \to 0$, but $Np \to \lambda$ then a Binomial

Binomial, **Poisson**, Gaussian

If N → ∞ and p → 0, but Np → λ then a Binomial approaches a Poisson: (see Barlow 3.3.1)

$$f(n,\lambda) = \frac{\lambda^n}{n!}e^{-\lambda}$$



In reality, the approximation is already quite good at e.g. N=50 and p=0.1.

The Poisson distribution only has one parameter, namely λ.
Mean = λ
Variance = λ

So the error on a number is...

...the square root of that number!

Binomial, **Poisson**, Gaussian

The error on a (Poisson) number... is **the square root** of that number!!!

Binomial, **Poisson**, Gaussian

The error on a (Poisson) number.

A very useful case of this is the error to assign a bin in a histogram, if there is reasonable statistics (N_i > 5-20) in each bin.

is **the square root** of that number!!!

The error on a (Poisson) number... is **the square root** of that number!!!

Note: The sum of two Poissons with λ_a and λ_b is a new Poisson with $\lambda = \lambda_a + \lambda_b$.
(See Barlow pages 33-34 for proof)

Binomial, **Poisson**, Gaussian

The Poisson distribution has the advantage that **neither the number of trials N nor the probability of succes p has to be known** - just their product.

A typical use is when dealing with **rates** in a given interval of time, distance, area, volume, etc.

Binomial, **Poisson**, Gaussian

The Poisson distribution has the advantage that **neither the number of trials N nor the probability of succes p has to be known** - just their product.

A typical use is when dealing with **rates** in a given interval of time, distance, area, volume, etc.

Example (real from 1898):
There were 122 deaths by horse kicks over 10 different regiments, over 20 years. What is the predicted number of deaths in a specific regiment and year?

First we estimate the mean value:

$$\mu = \frac{122}{20 * 10} = 0.61$$

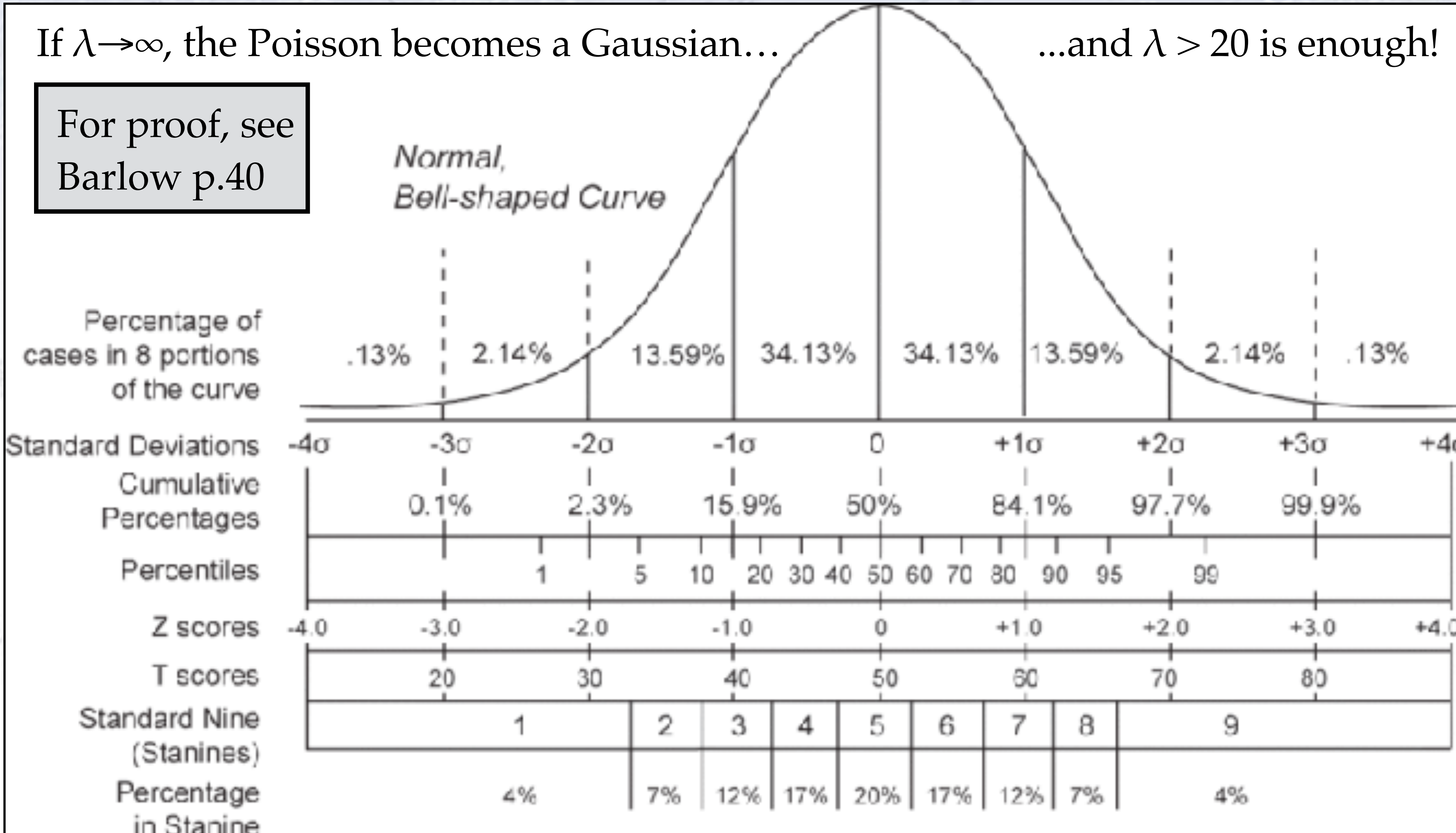
This means that the probability that 0 will die is given by:

$$P(0) = e^0 \frac{0.61^0}{0!} = 0.54$$

Binomial, Poisson, **Gaussian**

If $\lambda \rightarrow \infty$, the Poisson becomes a Gaussian… …and $\lambda > 20$ is enough!

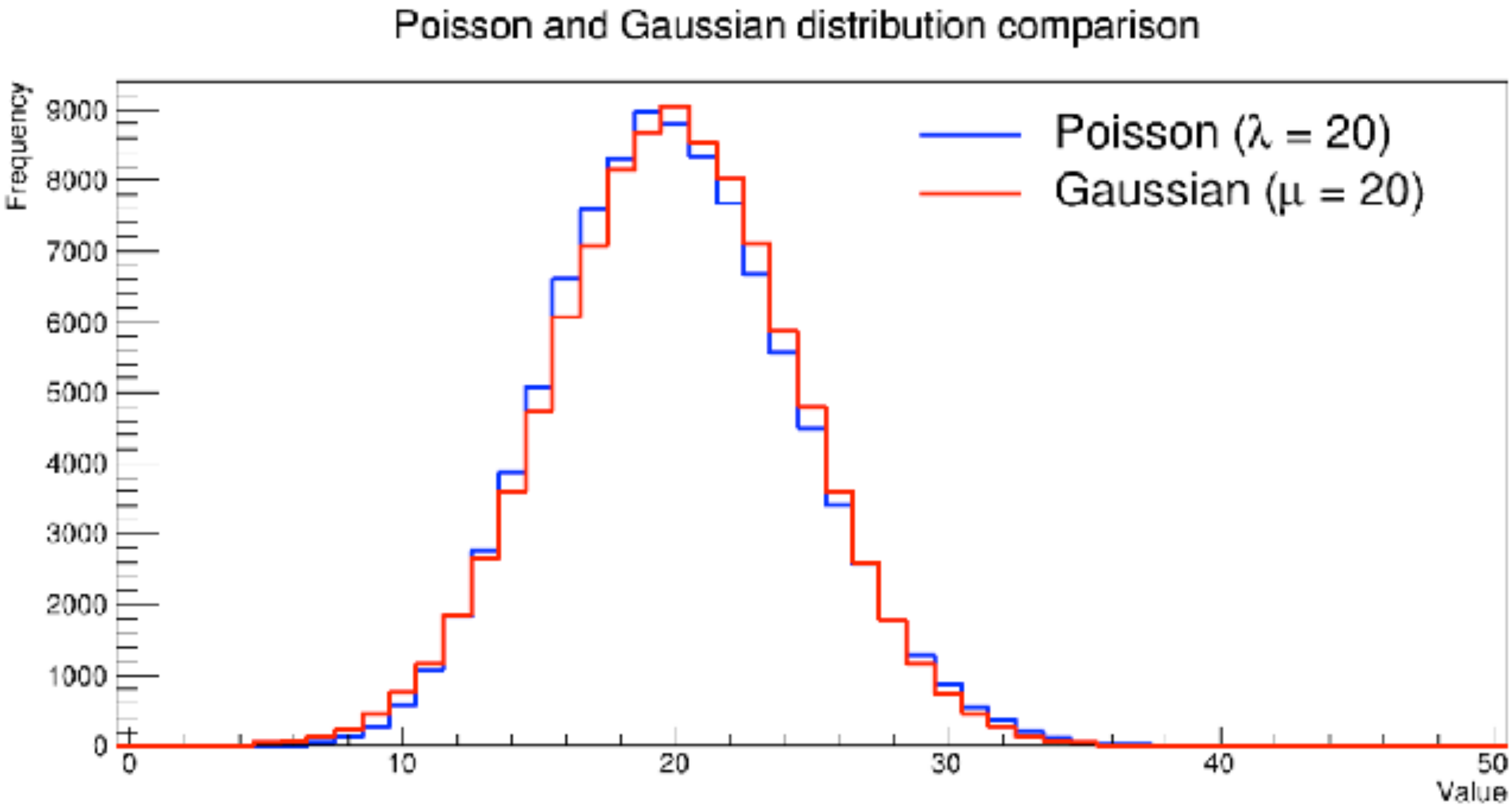
For proof, see Barlow p.40



All fields encounter the Gaussian, and for this reason, its scale has many names!

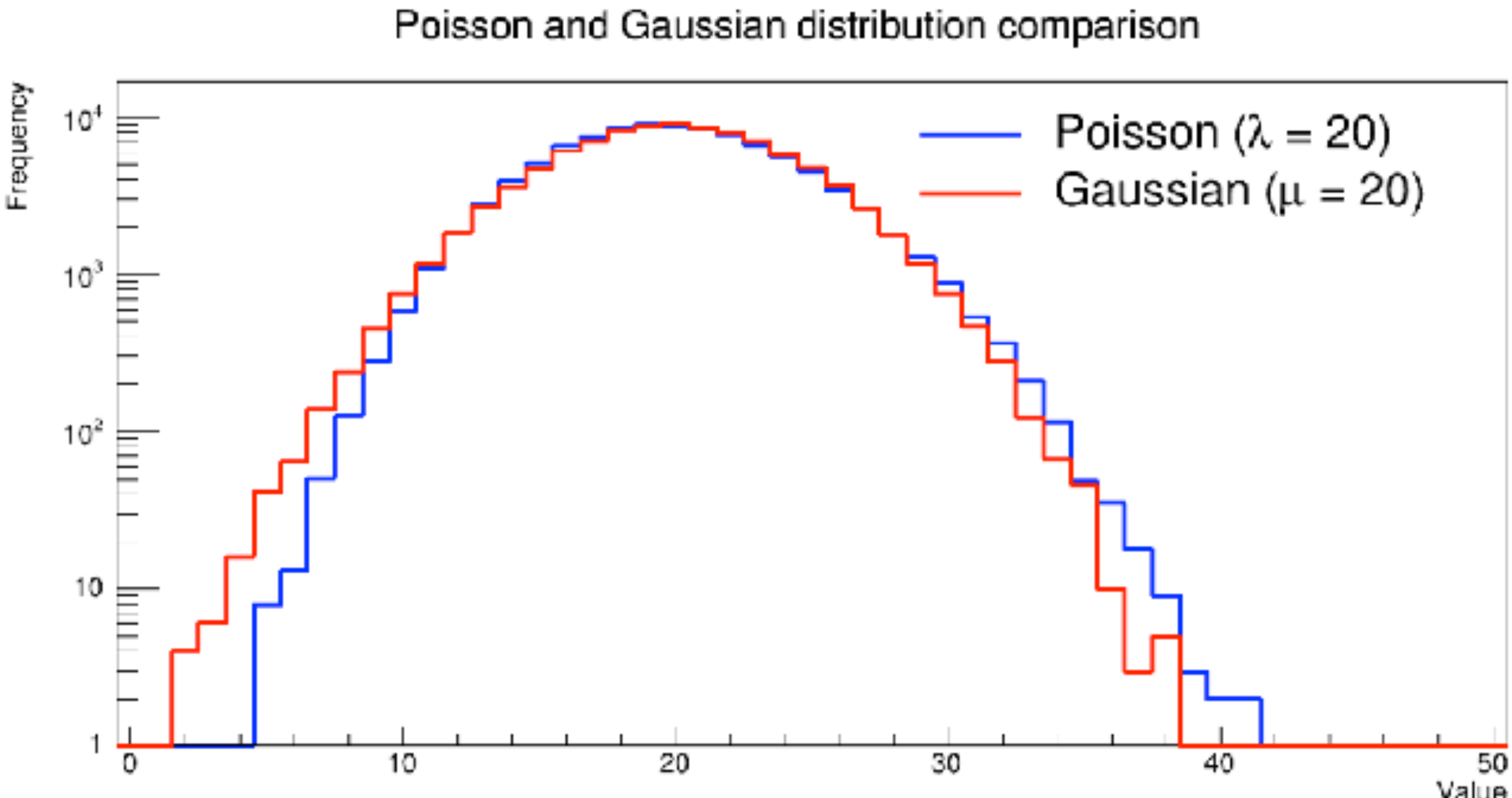
Binomial, Poisson, **Gaussian**

If $\lambda \to \infty$, the Poisson becomes a Gaussian… ...and $\lambda > 20$ is enough!



Binomial, Poisson, **Gaussian**

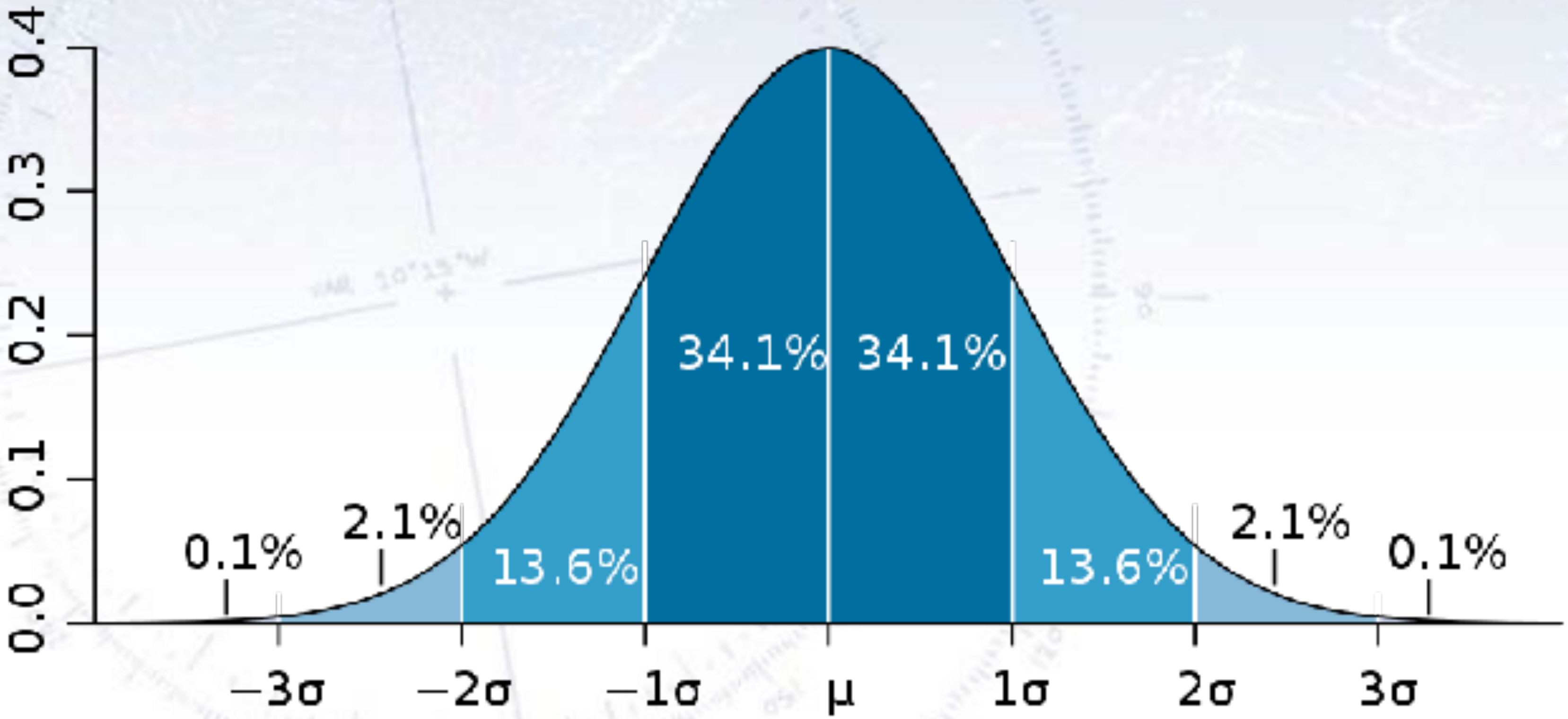
If $\lambda \rightarrow \infty$, the Poisson becomes a Gaussian… ...and $\lambda > 20$ is enough!



However, note that the TAILS are not quite the same!!!
This is the very reason for the difference between Chi2 and (binned) likelihood!

Binomial, Poisson, **Gaussian**

"If the Greeks had known it, they would have deified it."

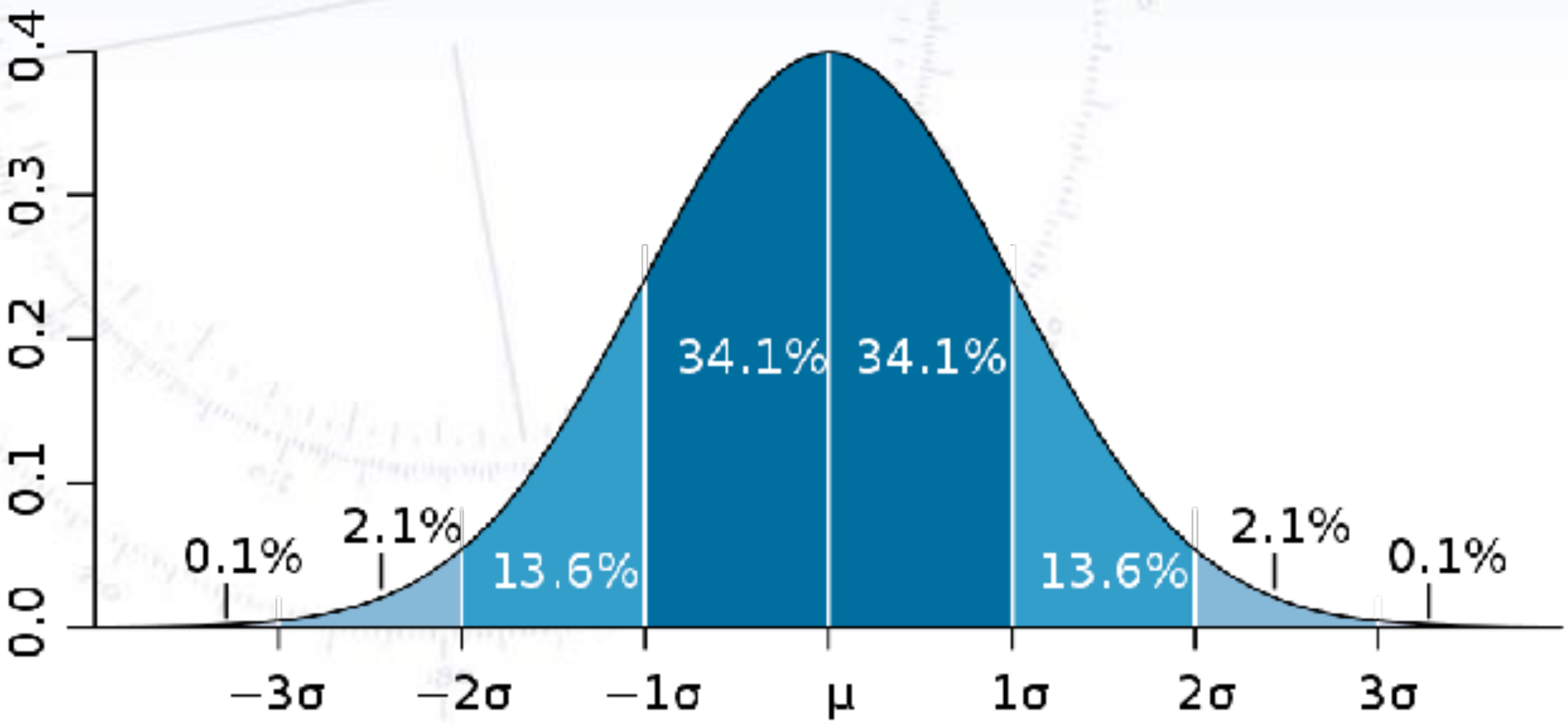


"If the Greeks had known it, they would have deified it. It reigns with serenity and in complete self-effacement amids the wildest confusion. ***The more huge the mob and the greater the apparent anarchy, the more perfect is its sway****. It is the supreme Law of Unreason. Whenever a large sample of chaotic elements are taken in hand and marshalled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to be latent all along." [Karl Pearson]*

Binomial, Poisson, **Gaussian**

The Gaussian **defines** the way we consider uncertainties.

Range	Inside	Outside
$\pm\ 1\sigma$	**68** %	32 %
$\pm\ 2\sigma$	**95** %	5 %
$\pm\ 3\sigma$	**99.7** %	0.3 %
$\pm\ 5\sigma$	99.99995 %	0.00005 %



Student's t-distribution

Given only a small (n obs.) sample (still assumed Gaussian), we don't know the mean μ and width σ well - we only know estimates of them! This changes the PDF to:

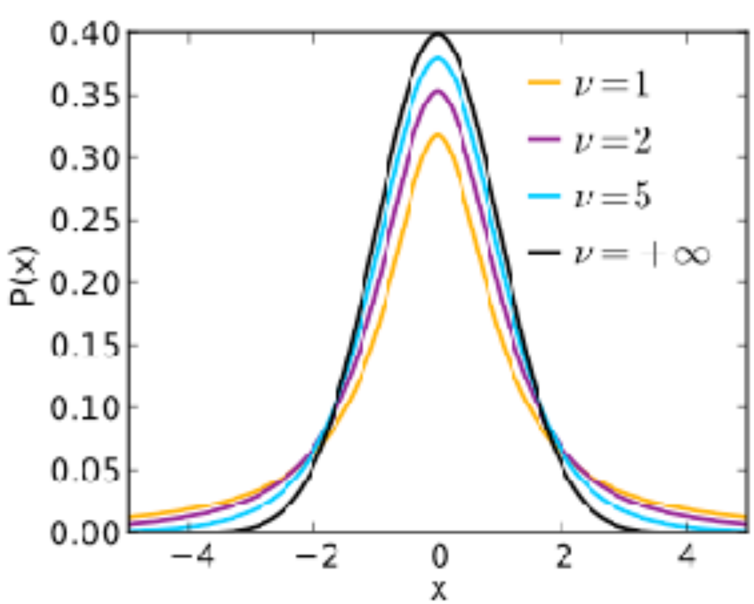
$$p(x \,|\, \nu, \hat{\mu}, \hat{\sigma}^2) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\,\sqrt{\pi\nu\hat{\sigma}^2}} \left(1 + \frac{1}{\nu}\left(\frac{x-\hat{\mu}}{\hat{\sigma}}\right)^2\right)^{-\frac{\nu+1}{2}} \qquad \nu = N_{\mathrm{DoF}} = n - 1$$

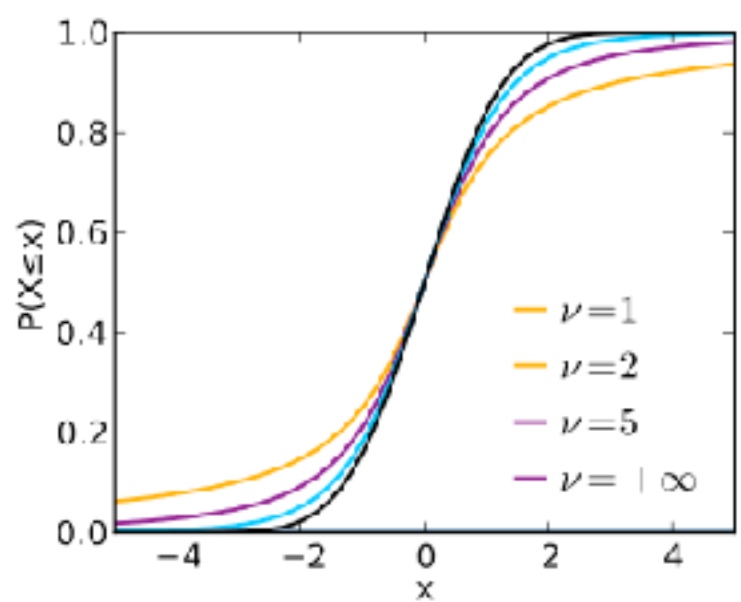
Student's t-distribution

Given only a small (n obs.) sample (still assumed Gaussian), we don't know the mean μ and width σ well - we only know estimates of them! This changes the PDF to:

$$p(x \,|\, \nu, \hat{\mu}, \hat{\sigma}^2) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\,\sqrt{\pi\nu\hat{\sigma}^2}} \left(1 + \frac{1}{\nu}\left(\frac{x - \hat{\mu}}{\hat{\sigma}}\right)^2\right)^{-\frac{\nu+1}{2}} \qquad \nu = N_{\mathrm{DoF}} = n - 1$$

"Discovered" by William Gosset, student's t-distribution takes into account the **lacking knowledge of the mean and variance** (as is the case for small samples).



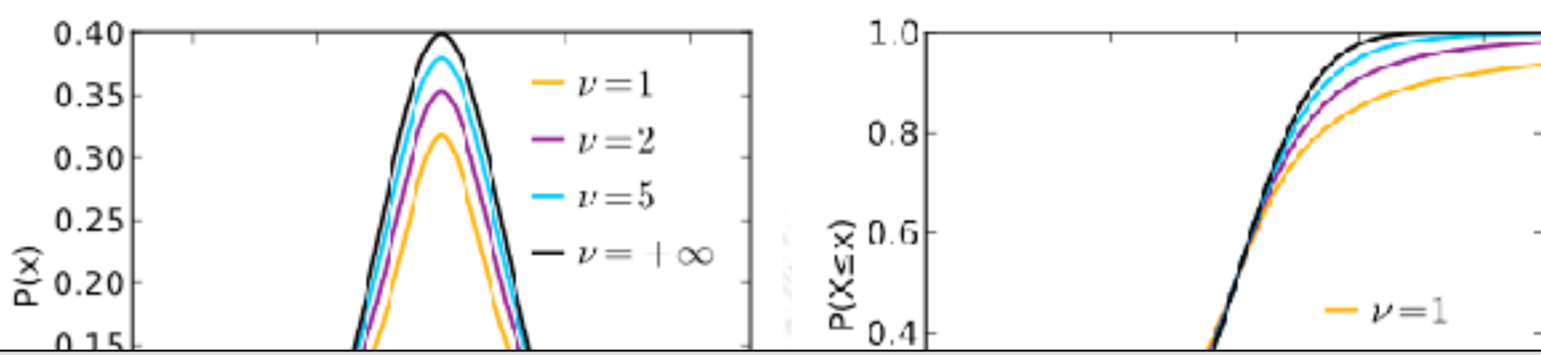


Student's t-distribution

Given only a small (n obs.) sample (still assumed Gaussian), we don't know the mean μ and width σ well - we only know estimates of them! This changes the PDF to:

$$p(x\,|\,\nu,\hat{\mu},\hat{\sigma}^2) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu\hat{\sigma}^2}}\left(1+\frac{1}{\nu}\left(\frac{x-\hat{\mu}}{\hat{\sigma}}\right)^2\right)^{-\frac{\nu+1}{2}} \qquad \nu = N_{\mathrm{DoF}} = n-1$$

"Discovered" by William Gosset, student's t-distribution takes into account the **lacking knowledge of the mean and variance** (as is the case for small samples).



When mean and width are poorly known, estimating it from sample gives:

Gaussian: $z = \frac{x-\mu}{\sigma}$ **Student's:** $t = \frac{x-\hat{\mu}}{\hat{\sigma}}$

Exponential distribution

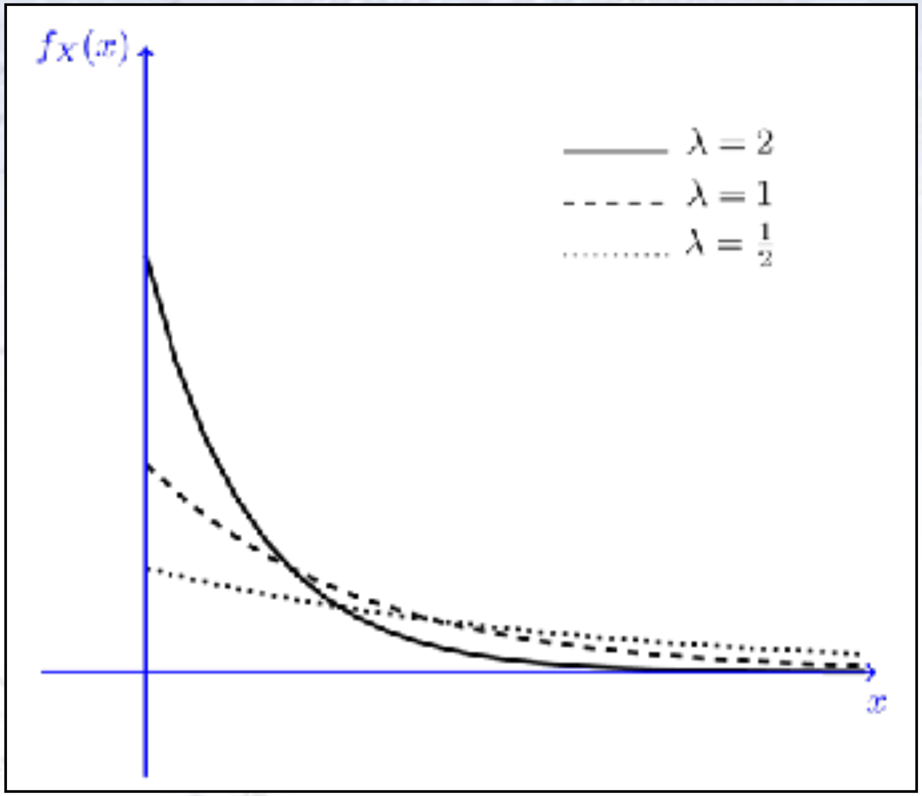
One particularly important PDF is the exponential distribution:

$$f(x) = \lambda e^{-\lambda x}$$

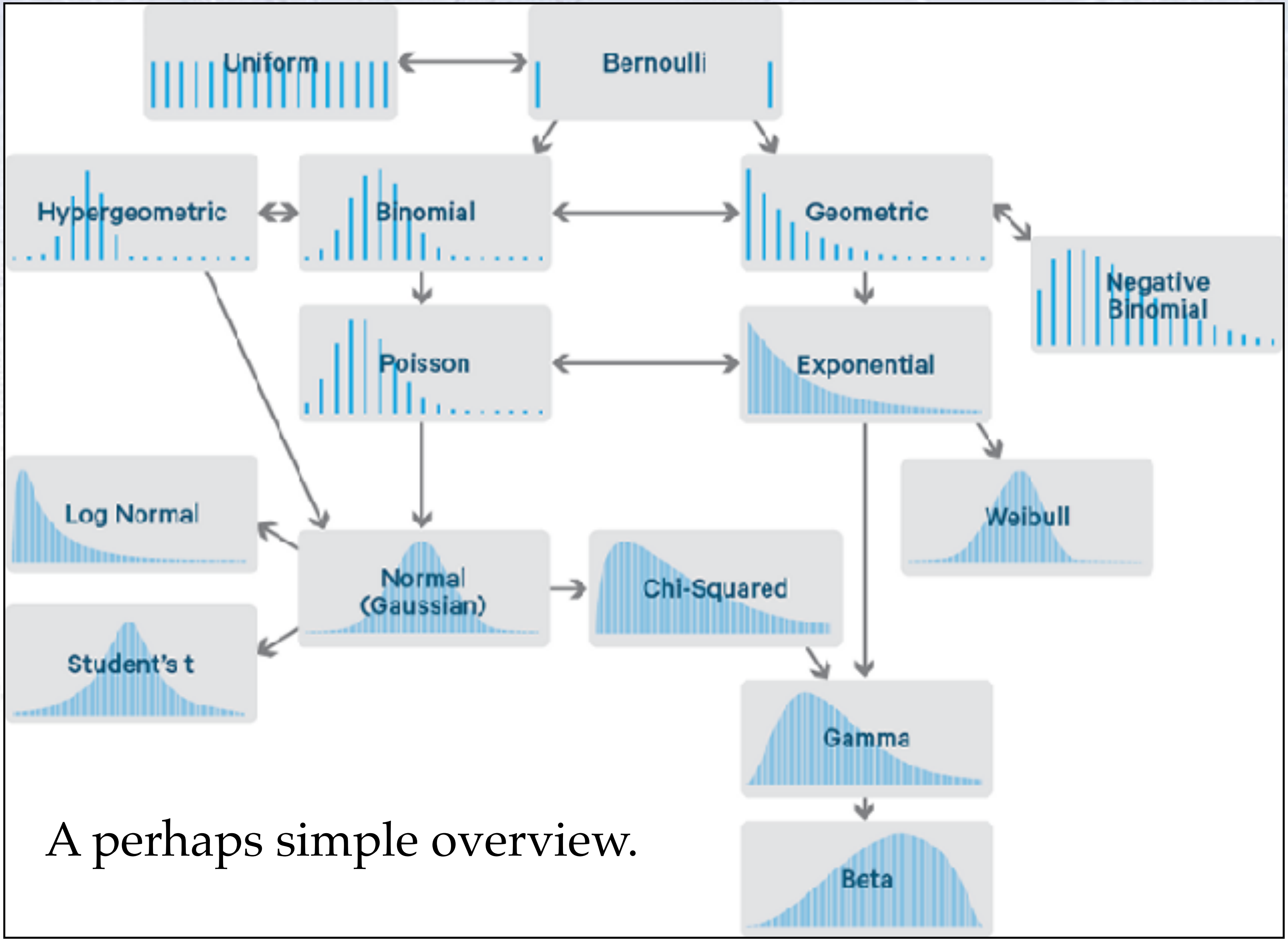
It has mean value $1/\lambda$ and width $1/\lambda$.

Its importance comes from the fact that:
If a process occurs with a constant rate, the distribution of time before the event happens will be exponentially distributed

This is really the case for many systems. Of course the most prominent example is the decay of particles/cells/etc.

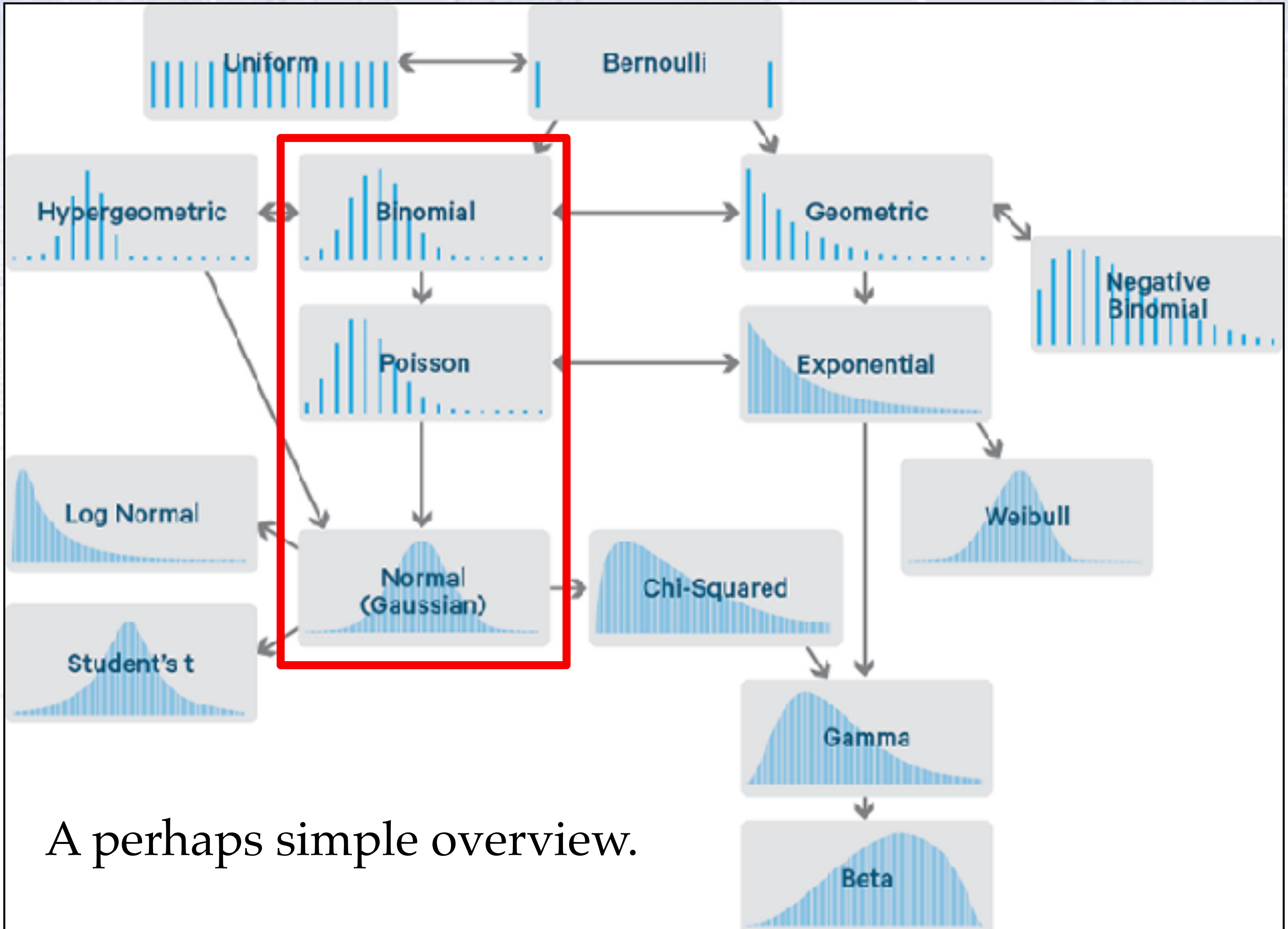


Distribution Overview



A perhaps simple overview.

Distribution Overview

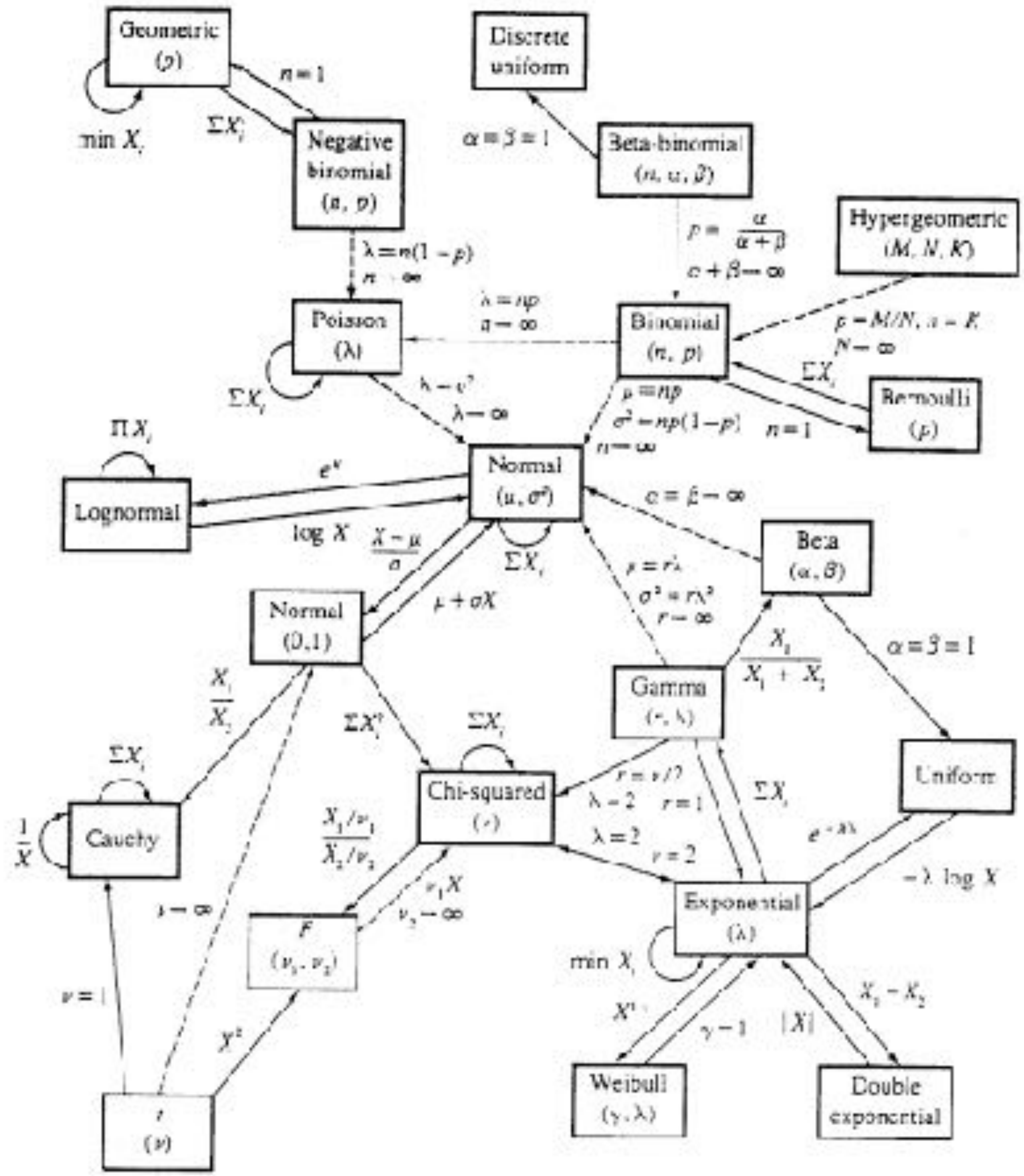


A perhaps simple overview.

Distribution Relationship

The different PDFs are related.

As can be seen, essentially all PDFs "converges" towards the Gaussian (normal) distribution.



Relationships among common distributions. Solid lines represent transformations and special cases, dashed lines represent limits. Adapted from Leemis (1986).

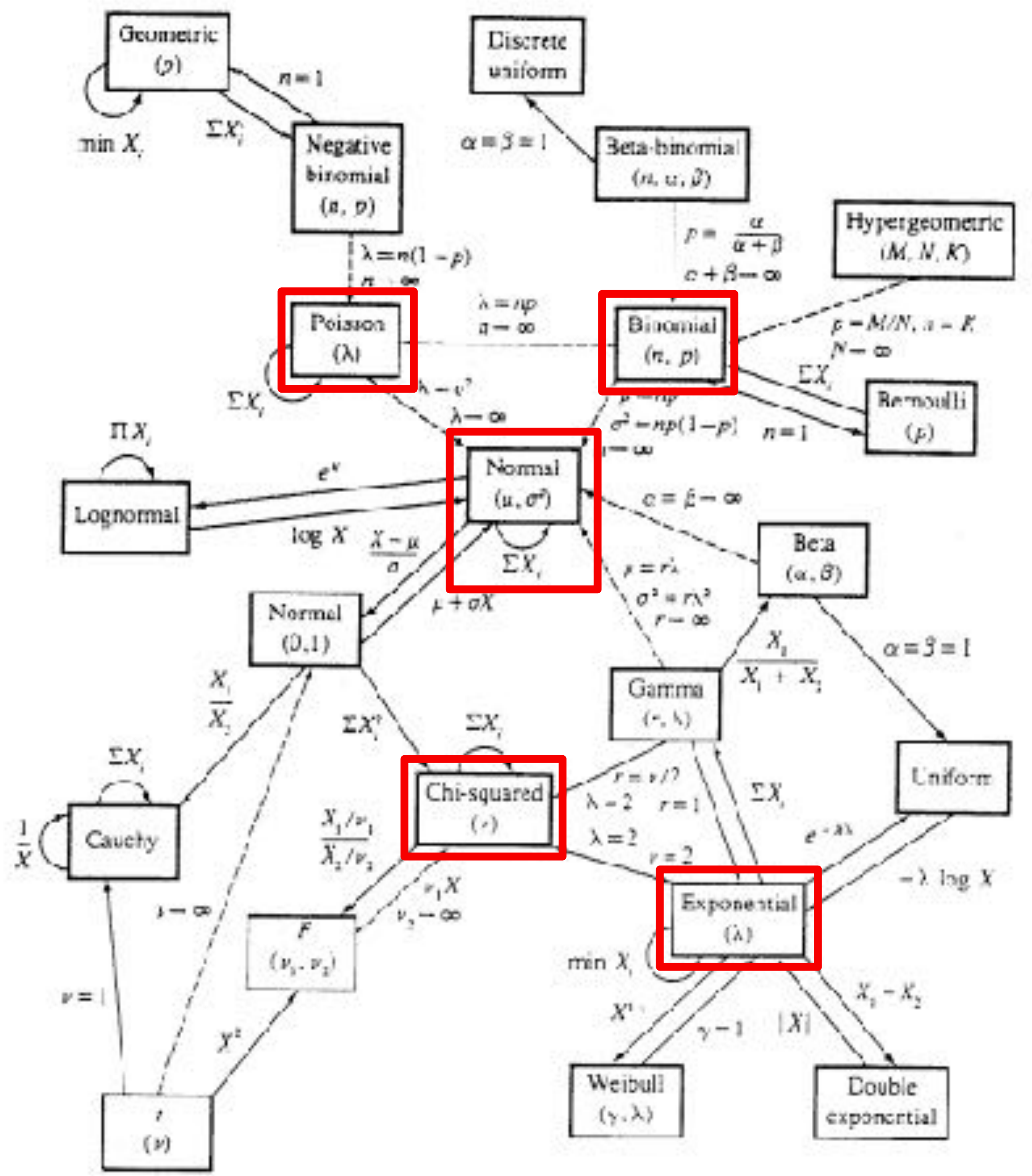
Distribution Relationship

The different PDFs are related.

As can be seen, essentially all PDFs "converges" towards the Gaussian (normal) distribution.

Don't worry about knowing them all…. Through a long life in statistics, I have still yet to encounter all of these in use!

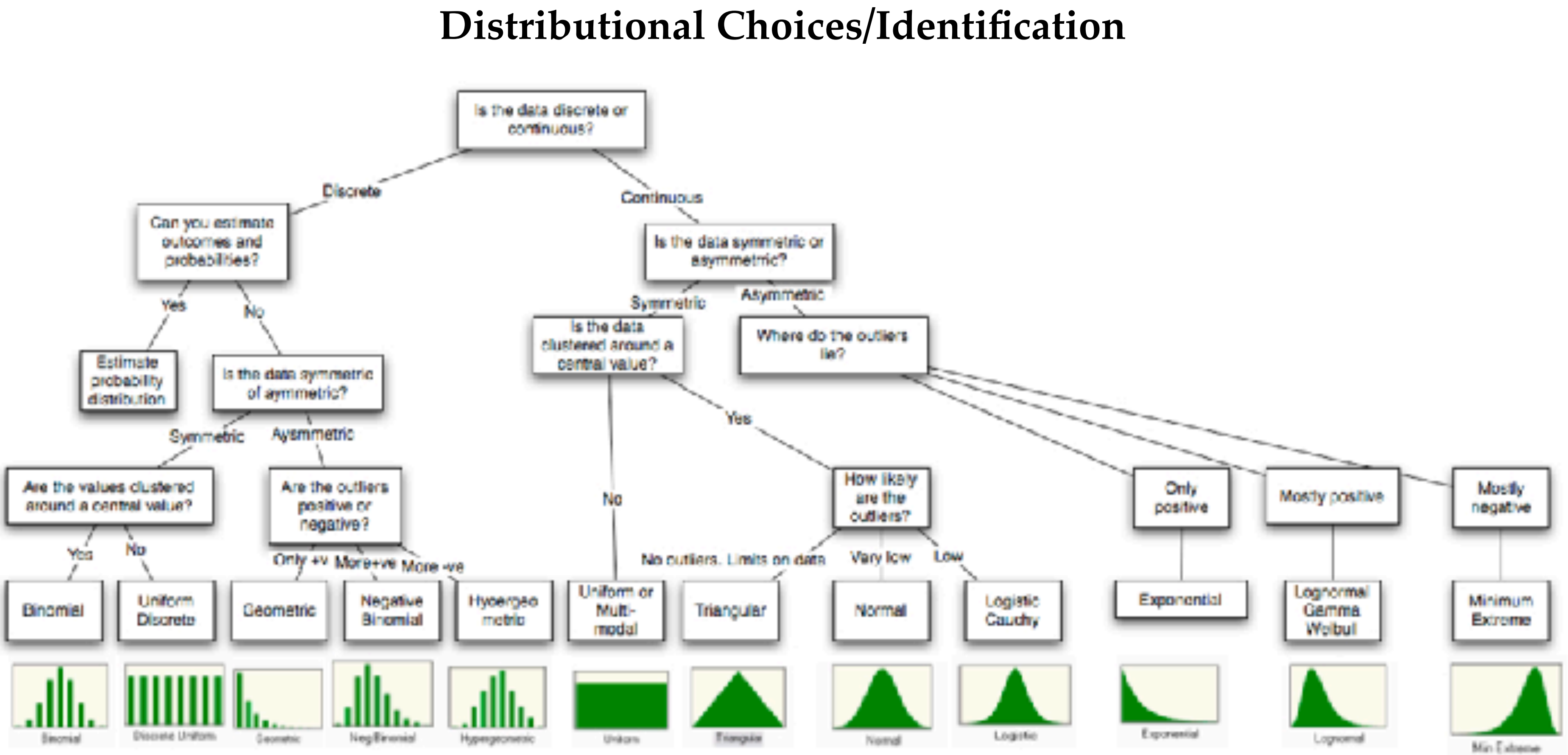
630 Table of Common Distributions



Relationships among common distributions. Solid lines represent transformations and special cases, dashed lines represent limits. Adapted from Leemis (1986).

Distribution Overview

I like the following overview of the most common PDFs, though it is far from perfect. However, it shows what makes the essential differences between PDFs.



From: A. Damodaran

Distribution Overview

Summary of lecture

All PDFs are normalized functions, that describe the probability of getting a certain value/outcome from evaluating the PDF function.

Among the most fundamental PDFs are the Binomial, Poisson and Gaussian.

Remember that the error on a (Poisson) number is **the square root** of that number!

Remember that the Gaussian distribution defines the uncertainties, that we report on experiments.