Data Analysis in Astronomy and Physics

Lecture 5: Probability Distributions

M. Röllig

Introduction

A probability distribution $f(x)$ describes the expectation of occurrence of event x.

Example: Toss 4 coins simultaneously.

The probability of	no heads is	$(1/2)^4 = 1/16$
	1 head	$4 \times (1/2)^4 = 4/16$
	two heads	$6 \times (1/2)^4 = 6/16$ ($\binom{4}{2}$=6)
	three heads	$4 \times (1/2)^4 = 4/16$
	and four heads	$(1/2)^4 = 1/16$.

The probability sum of all possibilities is 1.

If x is the number of heads $x = \{0, 1, 2, 3, 4\}$,
the corresponding probabilities are $\mathcal{P}(x) = (1/16,\ 4/16,\ 6/16,\ 4/16, 1/16\}$, a distribution of probabilities describing our expectation of the outcome of the experiment.

Introduction

In the example we have a mapping between a set of integer numbers and the outcome of the experiment.

If the outcome does not correspond to discrete numbers, we have a mapping of a set of outcomes to real numbers. If so, we discretize the real numbers into small sub-ranges, in which we assume that the probability does not change.

The real number x indexing an outcome is associated with a probability density $f(x)$, such that the probability of getting a number near x, say in the range δx, is $\mathcal{P}(x)\,\delta$x.

Definition

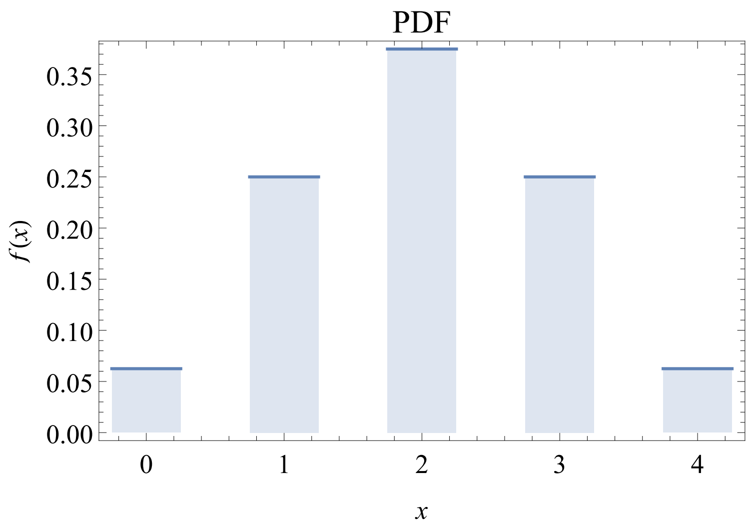
If x is a continuous random variable, then $f(x)$ is its **probability density function (PDF)**, commonly called **probability distribution**, when,

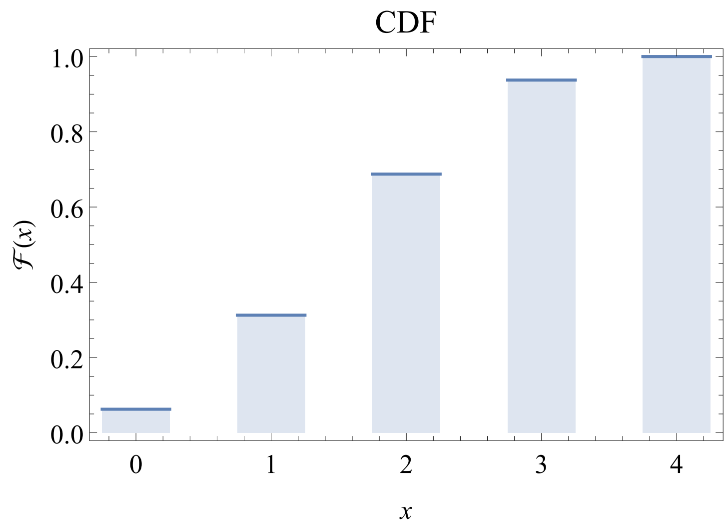
1. $\mathcal{P}(a < x < b) = \int_a^b f(x)\, dx$
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$, and
3. $f(x)$ is a single-valued non-negative number for all real x.

Definition

The corresponding **cumulative probability distribution function (CDF)** is $\mathcal{F}(x) = \int_{-\infty}^{x} f(y)\, dy$.

Out[•]=





If distributions are functions of more than one variable they are called **multivariate**.

Distribution Measures

To quantify a distribution, one can describe its location (“Where is the center?”) and its dispersion (“What is its spread/width?”).

These quantifiers can be given by the first two **moments of the distributions**:

mean: $\mu_1 = \mu = \int_{-\infty}^{\infty} x\, f(x)\, dx$

variance: $\mu_2 = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu_1)^2\, f(x)\, dx$

The square root of the variance is called **standard deviation** σ.

Higher moments exist as well (third moment ~ “skewness”). There are probability distributions we can calculate resulting from ideal experiments, outcomes or combinations of these. The best-known are the **UNIFORM**, **BINOMIAL**, **POISSON** and **GAUSSIAN** (or **NORMAL**) distributions.

Expectation values

Parameters such as μ and σ of the Poisson or Gaussian distribution define these distribution functions. They are not statistics! (Statistics are a function of data!). But we may expect or anticipate that our data follows a certain distribution and we may wish to relate statistics from the data to parameters describing the distribution.

This is done through expectation values. The expectation $E[g(x)]$ of some function $g(x)$ of a random variable x, with distribution function $f(x)$, is defined as

Out[ ]//TraditionalForm=

$$\boxed{\mathrm{E}(g(x)) = \int g(x)\, f(x)\, dx}$$

i.e. the sum of all possible values of f **weighted by the probability of their occurrence**. Think of the expectation as being the result of repeating an experiment many times, and averaging the results.

Expectation values

For example, compute an average value of $\langle X \rangle$. If we repeat the experiment many times, we will find that the average of $\langle X \rangle$ will converge to the true mean value, the expectation of the function $f(x) = x$.

Out[ ]//TraditionalForm=

$$\boxed{\mathrm{E}(x) = \int x\, f(x)\, dx}$$

Likewise, the statistics S^2 should converge to the variance defined by

Out[ ]//TraditionalForm=

$$\boxed{\mathrm{var}\left((x-\mu)^2\right) = \mathrm{E}\left((x-\mu)^2\right) = \int (x-\mu)^2 f(x)\, dx}$$

In fact ($\langle X \rangle$,S^2) (functions of data alone) converge to (μ, σ^2) when we have plenty of data.

Convention

We will in the following use the convention that if applicable, Greek letters describe the whole population while Latin letters describe sample statistics, e.g. a population might be normally distributed with mean μ and variance σ^2, while a sample from this sample will be distributed around the sample mean $\bar{x}$ and the sample variance S^2.

Moments of a distribution

Expectations of a distribution can also be expressed in terms of moments. For the list $\{x_1, x_2, \ldots, x_n\}$, the r^{th} central moment is given by $\frac{1}{n}\sum_i (x_i - \overline{x})^r$, where $\overline{x}$ is the mean of the list.

Out[ ]//TraditionalForm=

$$\boxed{\mu_r = \int (x-\mu)^r f(x)\, dx}$$

Moments of a distribution

The zero-th moment is just the area of $f(x)$.

Out[∘]//TraditionalForm=

$$\boxed{\mu_0 = \int f(x)\, dx}$$

The second moment is the variance:

Out[∘]//TraditionalForm=

$$\boxed{\mu_2 = \sigma^2 = \int (x-\mu)^2 f(x)\, dx}$$

Moments of a distribution

The skewness is

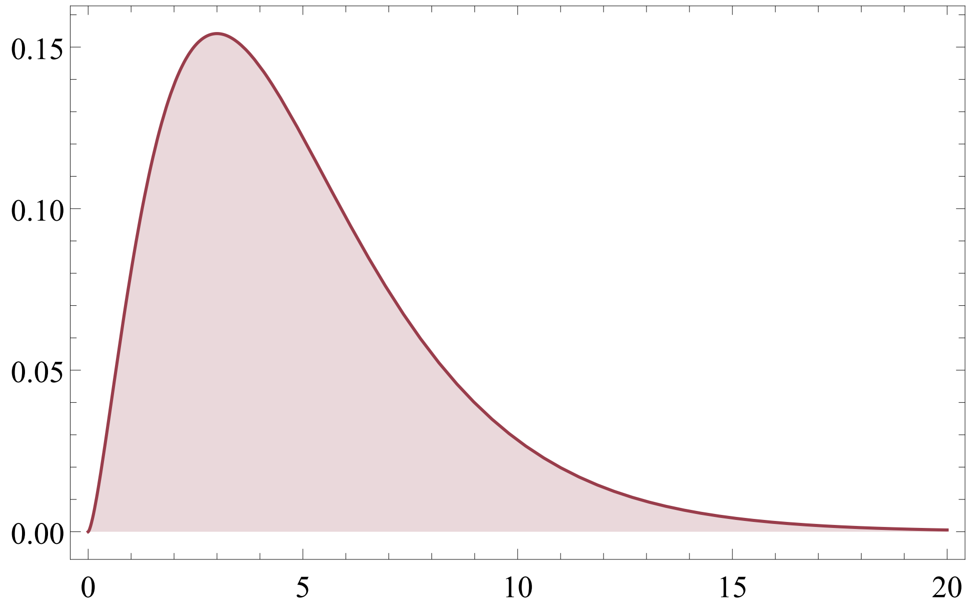
Out[∘]//TraditionalForm=

$$\boxed{\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\int (x-\mu)^3 f(x)\, dx}{\left(\int (x-\mu)^2 f(x)\, dx\right)^{3/2}}}$$

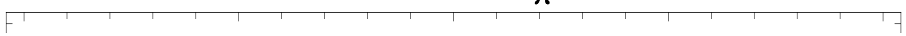
Skewness measures the asymmetry of a distribution/data. A positive skewness indicates a distribution with a long right tail, a negative skewness a distribution with a long left tail.

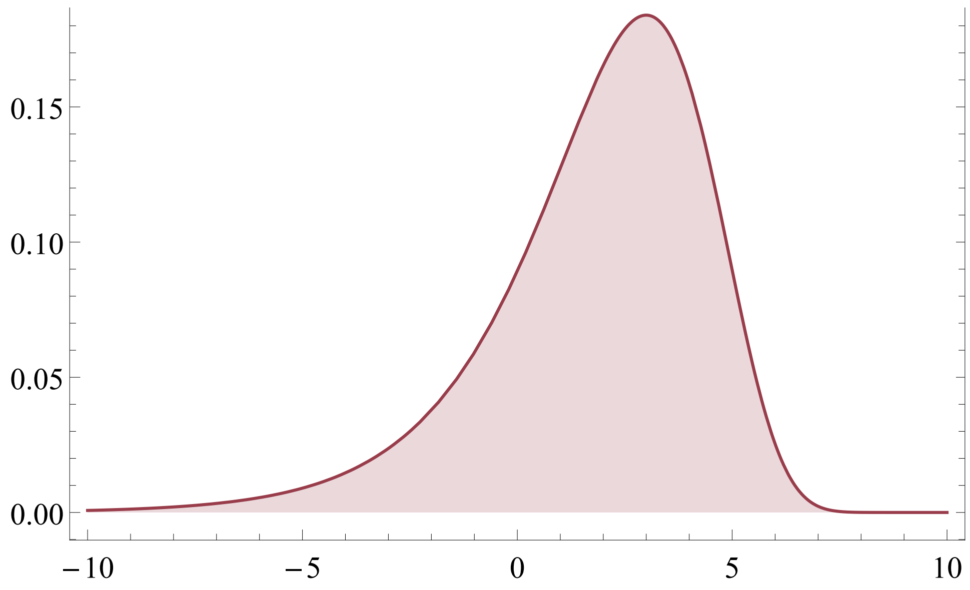
Out[∘]=

$$\text{Skewness}=2\sqrt{\frac{2}{5}}$$



$$\text{Skewness}=-\frac{12\sqrt{6}\,\zeta(3)}{\pi^3}$$




0.15
0.10
0.05
0.00
−10
−5
0
5
10

Moments of a distribution

ATTENTION: Abramowitz and Stegun (1972, p. 928) also confusingly refer to both γ_1 and $\beta = \gamma_1{}^2$ as "skewness."

Several types of skewness are defined, the terminology and notation of which are unfortunately rather confusing.

For a summary see :http://mathworld.wolfram.com/Skewness.html

Moments of a distribution

Kurtosis is defined as a normalized form of the fourth central moment μ_4 of a distribution and is a measure of the degree of peakiness(e.g.: 3 for Gaussian):

Out[∘]//TraditionalForm=

$$\boxed{\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\int (x-\mu)^4 f(x)\, dx}{\left(\int (x-\mu)^2 f(x)\, dx\right)^2}}$$

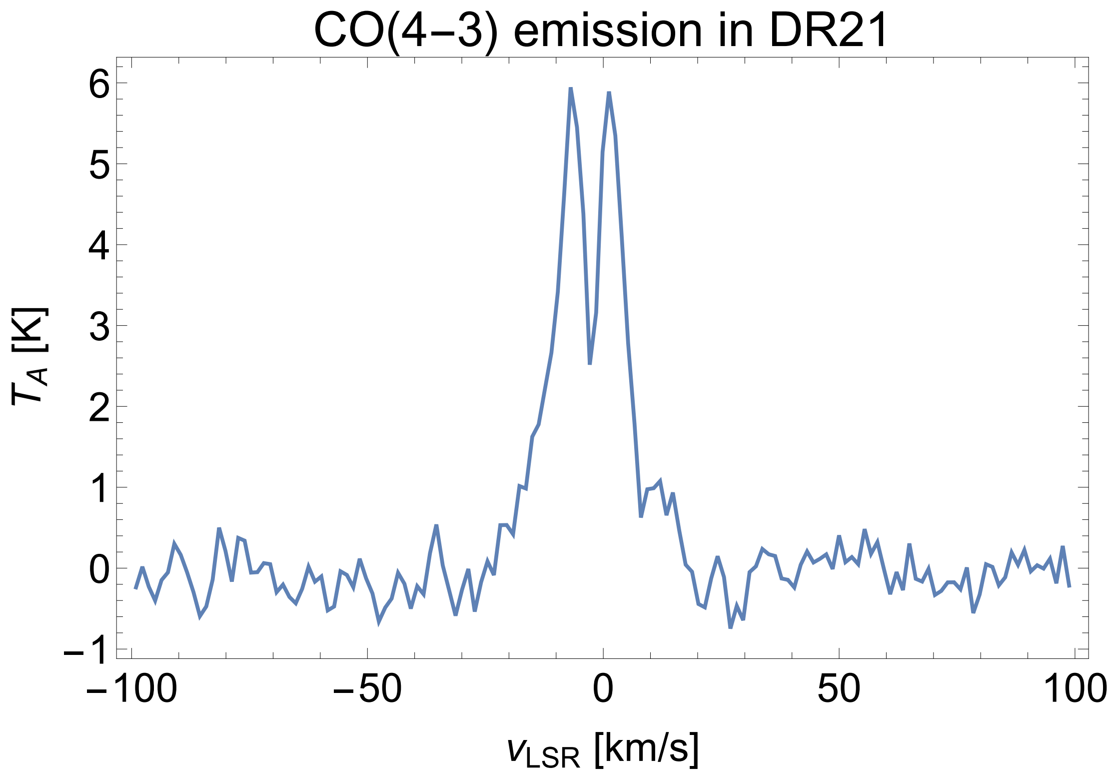
In[∘]:=

```
Integrate[(x - 0)^4 PDF[NormalDistribution[0, 2], x], {x, -∞, +∞}] / Integrate[(x - 0)^2 PDF[NormalDistribution[0, 2], x], {x, -∞, +∞}]^2
```

Out[∘]= 3

Moments of spectra

Out[ ]=



In[ ]:=
```
spectrum[[100 ;; 102]]
```

Out[ ]=
```
{{35.0739, 0.171445}, {36.429, 0.150318}, {37.784, -0.127574}}
```

Imagine we want to characterize the observed spectral emission line. Interesting statistics on are, e.g. the peak position, the integrated line intensity,

Moments of spectra

The line integrated intensity is given by the 0th moment, $\int I(V)\,dv$, or in case of the discrete data points: $\Delta v \sum_i I_i$ (if the velocity channels are equally wide).

Out[ ]//TraditionalForm=

$$\boxed{M_0 = \Delta \mathrm{v} \sum_i \mathrm{I}_i}$$

The first moment defines the intensity-weighted velocity of the spectral line. It can be taken as a measure for the mean velocity of the gas. The first moment is defined by

Out[ ]//TraditionalForm=

$$\boxed{M_1 = \frac{\sum_i v_i \, \mathrm{I}_i}{\sum_i \mathrm{I}_i}}$$

Moments of spectra

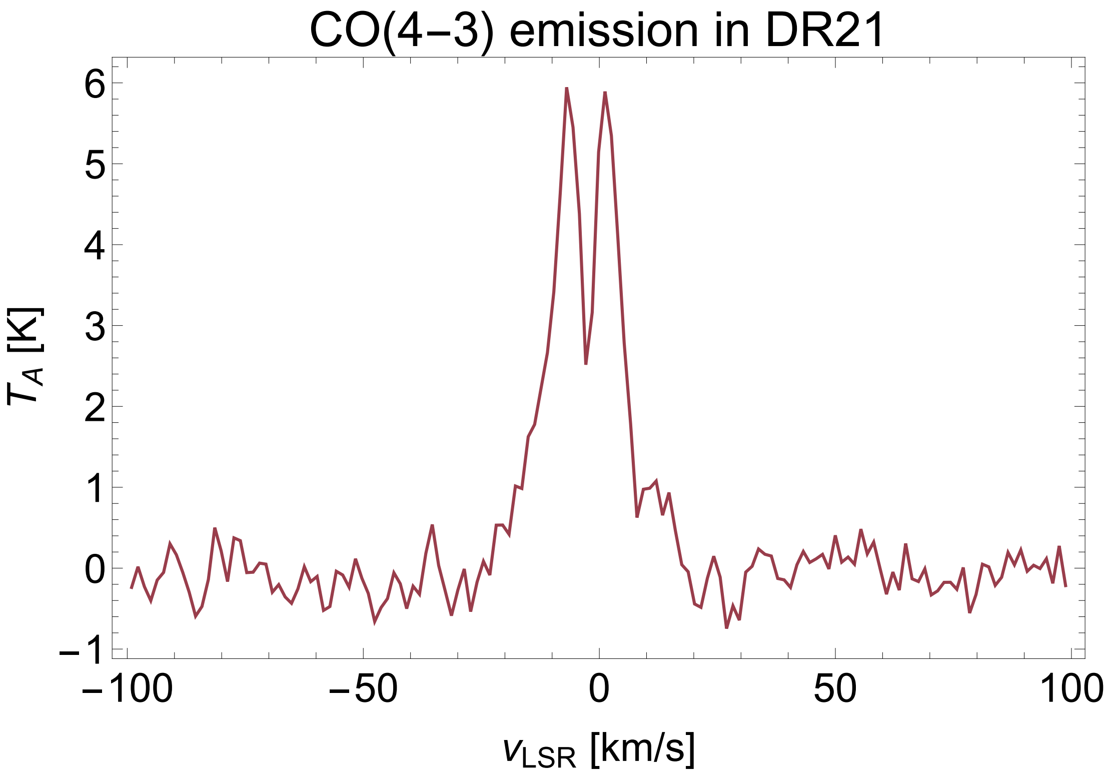
The second moment is a measure for the velocity dispersion, σ, of the gas along the line of sight, i.e. the width of the spectral line. It is defined by the intensity-weighted square of the velocity:

Out[ ]//TraditionalForm=

$$\boxed{M_2 = \sqrt{\frac{\sum_i (v_i - M_1)^2 \, \mathrm{I}_i}{\sum_i \mathrm{I}_i}}}$$

Moments of spectra

Out[ ]=



In[ ]:=
```
deltaV = Mean@Differences@spectrum[[All, 1]]
M0 = deltaV * Total[spectrum[[All, 2]]]
```

Out[ ]= 1.35505

Out[ ]= 82.1292

Moments of spectra

Assuming a proper baseline, the integrate line intensity is 82 K km/s

Where is the center (of mass) of the spectrum? $\frac{\int_{-\infty}^{\infty} v\, I\, dv}{\int_{-\infty}^{\infty} I\, dv} == \langle v \rangle$

In[ ]:=
```
M1 = Total[spectrum[[All, 1]] × spectrum[[All, 2]]] / Total[spectrum[[All, 2]]]
```

Out[ ]=
```
1.59657
```

And what is the FWHM (full width half maximum):

In[ ]:=
```
M2 = Sqrt[Abs@(Total[(spectrum[[All, 1]] - M1)^2 spectrum[[All, 2]]] / Total[spectrum[[All, 2]]])]
```

Out[ ]=
```
23.1812
```

Overview table

Uniform	f(x;a,b)=	$\begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{True} \end{cases}$	$\frac{\text{a+b}}{2}$	$\frac{1}{12}(-\text{a}+\text{b})^2$	In the study of rounding errors; as a tool in studies of other continuous distributions
Binomial	f(x;n,p)=	$\begin{cases} p^x \binom{n}{x} (1-p)^{n-x} & 0 \leq x \leq n \\ 0 & \text{True} \end{cases}$	n p	n (1 − p) p	x is the number of successes in an experiment (with n trials) with two possible outcomes, one ("success") of probability p, and the other ("failure") with probability q=1−p.

					Becomes a Normal distribution when $n \to \infty$.
Poisson	$f(x;\mu)=$	$\begin{cases} \frac{e^{-\mu}\mu^x}{x!} & x \geq 0 \\ 0 & \text{True} \end{cases}$	μ	μ	The limit for the Binomial distribution as $p \ll 1$, setting $\mu = n\,p$. It is the count−rate distribution, e.g. take a star from which an average of μ photons are received per hour (out of a total of n emitted; hence $p \ll 1$); the probability of receiving xphotons in Δt is $f(x;\mu)$. Tends

Out[ ]=

					to the Normal distribution as $\mu \to \infty$.
Normal (Gaussian)	$f(\mathrm{x};\mu,\sigma)=$	$\frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\,\sigma}$	μ	σ^2	The essential distribution; The central limit theorem ensures that the majority of "scattering things" are dispersed according to $f(\mathrm{x};\mu,\sigma)$.
Chi−square	$f(\chi^2;\nu)=$	$\begin{cases} \left(2^{-\nu/2}\, e^{-x/2}\, x^{\frac{\nu}{2}-1}\right)/\Gamma\left(\frac{\nu}{2}\right) & x>0 \\ 0 & \text{True} \end{cases}$	ν	$2\,\nu$	Vital in the comparison of samples, model testing; characterizes the dispersion of observed samples from the expected

					dispersion because if x_i is a sample of ν variables normally and independently distributed with means μ_i and variances σ_i^2, then $\chi^2=\sum_{i=1}^{N}(x_i-\mu_i)^2/$ obeys $f(\chi^2,\nu)$. Invariably tabulated and used in integral form. Tends to Normal distribution as $\nu \to \infty$.
Student t	$f(t;\nu)=$	$\frac{\left(\frac{\nu}{\nu+x^2}\right)^{\frac{\nu+1}{2}}}{\sqrt{\nu}\,\mathrm{B}\left(\frac{\nu}{2},\frac{1}{2}\right)}$	$\begin{cases} 0 & \nu > 1 \\ \text{Indeterminate} & \text{True} \end{cases}$	$\begin{cases} \frac{\nu}{-2+\nu} & \nu > 2 \\ \text{Indeterminate} & \text{True} \end{cases}$	For comparison of means, Normally

					distributed populations; if n x_i's are taken from a Normal population (μ,σ), and if x_s and σ_s are determined, then $t=\sqrt{n(\overline{x_s}-\mu)}/\sigma_s$ is distributed as f(t,ν) where \"degrees of freedom\" $\nu=n-$. Statistic t can also be formulated to compare means for samples from Normal populations with the same σ, different μ. Tends to Normal

					as $\nu \to \infty$.

Binomial Distribution

There are two outcomes - `success' or `failure'. This common distribution gives the chance of *n* successes in *N* trials, with the probability of a success at each trial *ρ*, and successive trials are independent. This probability is

Out[∘]//TraditionalForm=

$$\boxed{f(n, \mathrm{N}, \rho) = \binom{\mathrm{N}}{n} \rho^n (1-\rho)^{\mathrm{N}-n}}$$

The first term is the Binomial coefficient (or combinatorial coefficient) and gives the number of distinctive ways of choosing *n* items out of *N*.

Out[∘]//TraditionalForm=

$$\boxed{\binom{\mathrm{N}}{n} = \frac{\mathrm{N}!}{n!\,(\mathrm{N}-n)!}}$$

Binomial Distribution

Motivation

Imagine $N = 5$ people want to sit in only $n = 3$ chairs. How many possible arrangements are there (permutations)?

The first chair (C1) can be occupied by 5 different persons. The second chair (C2) can be occupied by 4 different people (one is already sitting). On the last chair (C3), three different people can sit. In total we have

$5 \times 4 \times 3 = 60$ possible ways to arrange 5 people on 3 chairs

Another way of writing this is

$\frac{5\times4\times3\times2\times1}{2\times1} = 5 \times 4 \times 3 = 60$, or generally $\frac{N!}{(N-n)!}$

$\frac{N!}{(N-n)!}$ is the number of permutations, i.e. the number of different ways to choose n items from a set of N, **when we care about the order!**

Binomial Distribution

Motivation

When the order is irrelevant, i.e. we just wish to know which combination of people end up sitting on a chair, we have to divide by the different possible combinations of 3 people sitting on 3 chairs, which is 3!.

To get the number of combinations, when choosing n items from a set of N, we find

$\frac{N!}{n!\,(N-n)!}$ which is written shortly as $\binom{N}{n}$

Binomial Distribution

Properties

The Binomial distribution has a mean value given by

$$\sum_{n=0}^{N} n\, f(n) = N \times p$$

and a variance of

$$\sum_{n=0}^{N} (n - N \times p)^2 f(n) = N \times p \times (1 - p)$$

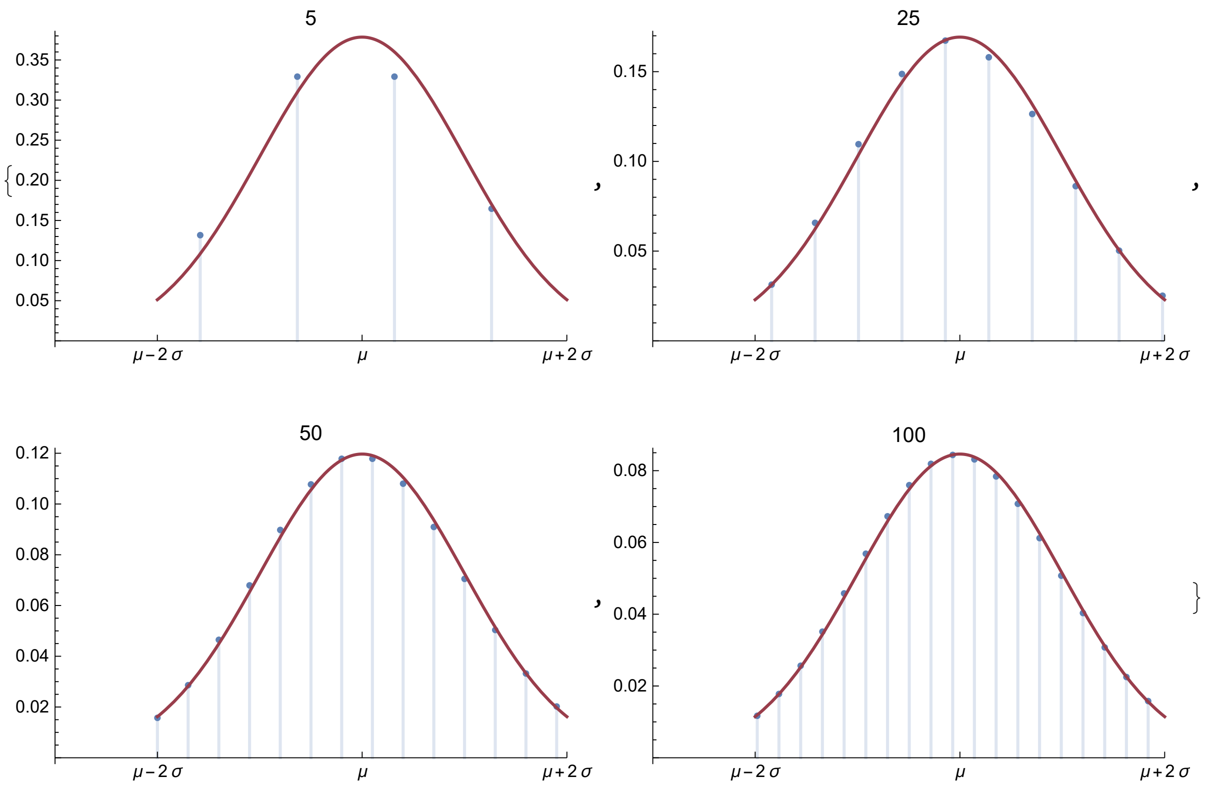
The Binomial distribution converges to a normal distribution as $n \to \infty$:

Binomial Distribution

Properties

In[•]:=
```
{μ, σ} = {Mean[BinomialDistribution[n, p]], StandardDeviation[BinomialDistribution[n, p]]}
```

Out[•]= $\{n\,p, \sqrt{n\,(1-p)\,p}\}$

Out[•]=


Binomial Distribution

Properties

Skewness measures the asymmetry in a distribution. A positive skewness indicates a distribution with a long right tail. A negative skewness indicates a distribution with a long left tail.

```
Skewness[BinomialDistribution[n, p]]
```

$$\frac{1-2p}{\sqrt{n(1-p)p}}$$

The distribution is symmetric for $p = 0.5$:

```
Solve[Skewness[BinomialDistribution[n, p]] == 0, p]
```

$$\left\{\left\{p \to \frac{1}{2}\right\}\right\}$$

The distribution becomes symmetric for large n:

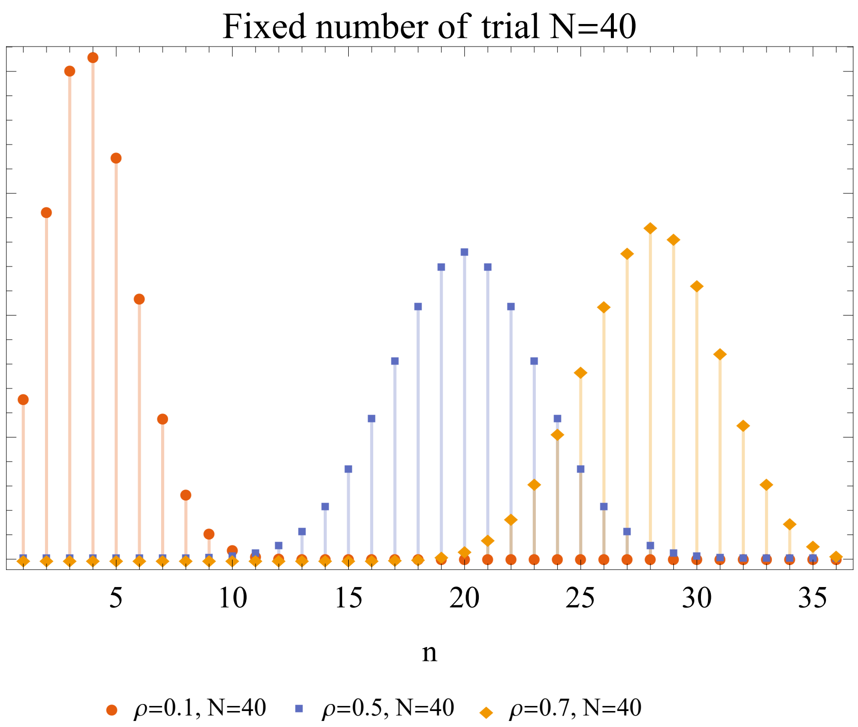
```
Limit[Skewness[BinomialDistribution[n, p]], n → ∞]
```

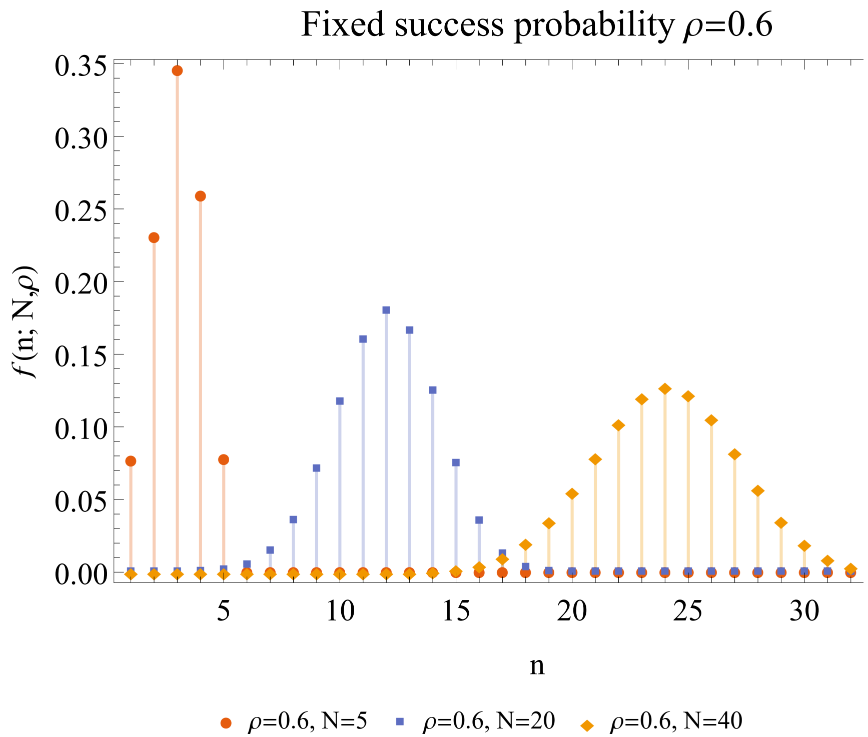
0

Binomial Distribution

P.M.F. - probability mass function

$$f(k;\ N, \rho) = \mathcal{P}(X = n) = \binom{N}{n} \rho^n (1 - \rho)^{N-n}$$





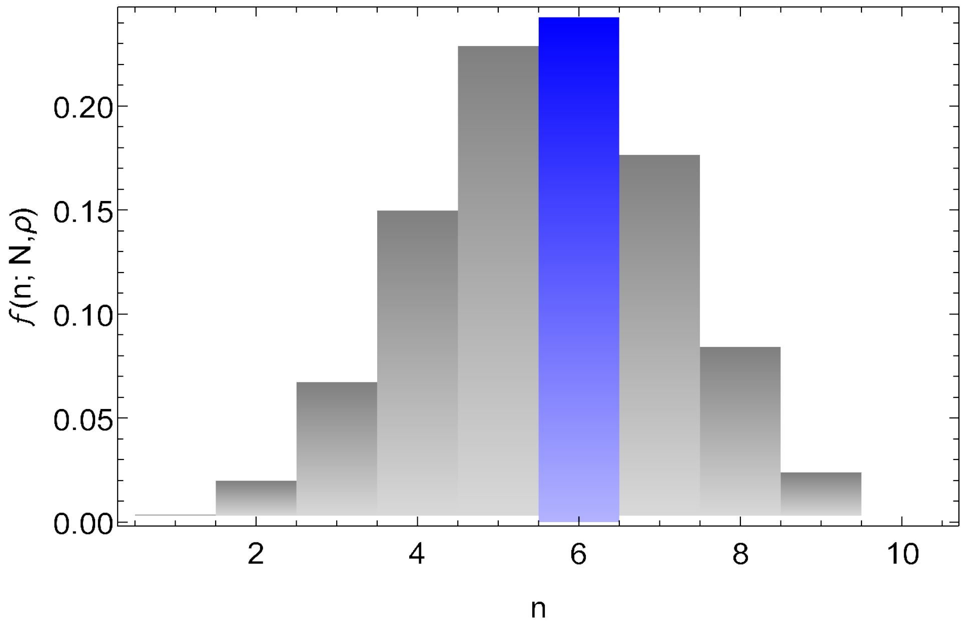
The term probability **mass** function is used for discrete random variables.

Binomial Distribution

Example

According to the latest exoplanet data (numbers made up!), a stellar system has a 56% chance to host at least one planet. What is the probability that in a random sample of 10 stars exactly 6 will host at least one planet?

$\mathcal{P}(X=6)=\binom{10}{6}0.56^{6}\,(1-0.56)^{10-6}=210\times0.56^{6}\times0.44^{4}=0.243$



Binomial Distribution

Success-failure rule:

A binomial distribution with at least 10 expected successes and 10 expected failures closely follows a normal distribution.

$$n \times p \geq 10$$
$$n \times (1-p) \geq 10$$

Normal approximation to the binomial: If the success- failure condition holds,

$$\text{Binomial}(n,p) \approx \text{Normal}(\mu,\sigma) \quad , \text{ where } \mu = n \times p, \text{ and } \sigma = \sqrt{n \times p \times (1-p)}$$

Quiz

What is is the minimum required n for a binomial distribution with p = 0.21 to closely follow a normal distribution?

Binomial Distribution

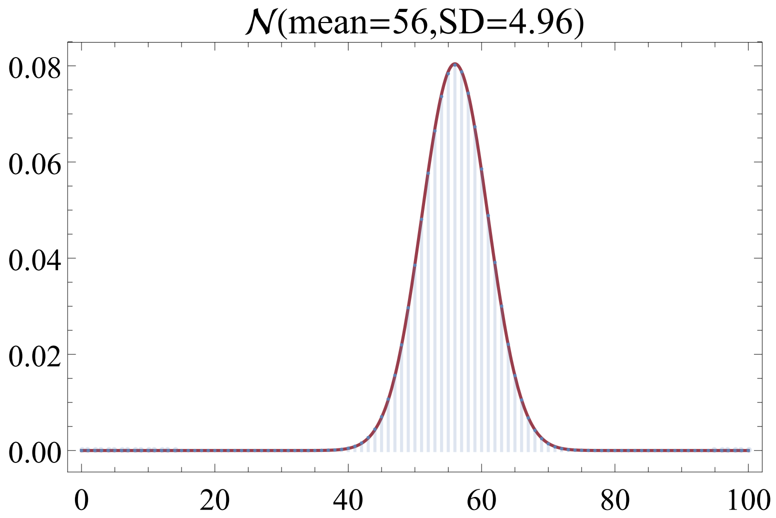
Example

Describe the probability distribution of stellar systems hosting at least one planet among a random sample of 100.

$p = 0.56,\ n = 100$

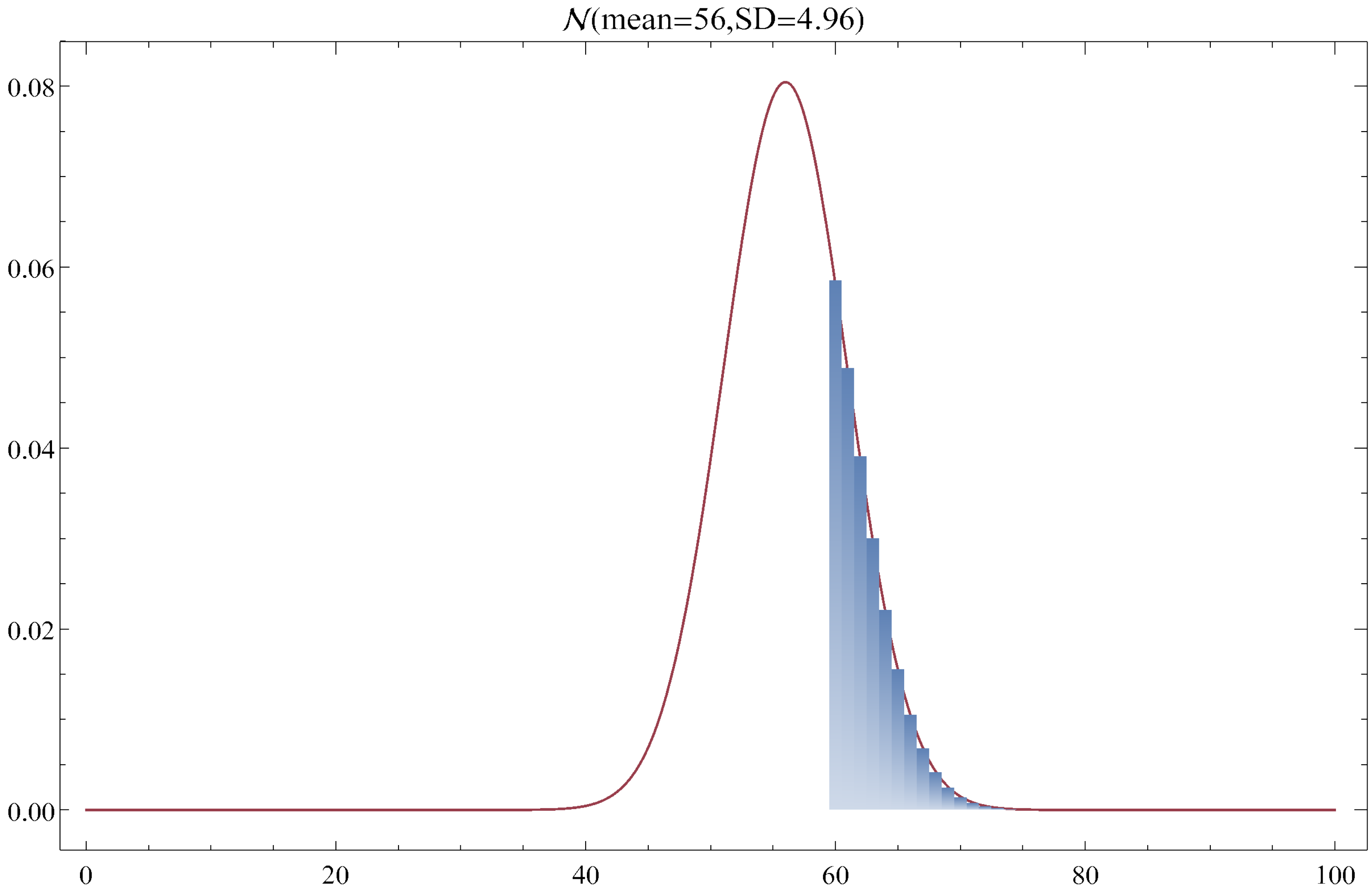
expected successes: $p \times n = 56 > 10$ expected fails: $(1 - p) \times n = 44 > 10$

μ=56 $\sigma = \sqrt{100 \times 0.56 \times 0.44}$=4.96



Example

What is the probability, that at least 60 out of a random sample of 100 stellar systems host a planet?

Out[∘]=


Example

What is the probability, that at least 60 out of a random sample of 100 stellar systems host a planet?

```
Total[Table[PDF[BinomialDistribution[100, 0.56], x], {x, 60, 100}]]
```

```
0.241106
```

Using the normal distribution as approximation for the Binomial distribution gives slightly different result!

```
1 - CDF[NormalDistribution[56, 4.96], 60]
```

```
0.209991
```

We have to account for the discreteness of the Binomial distribution.

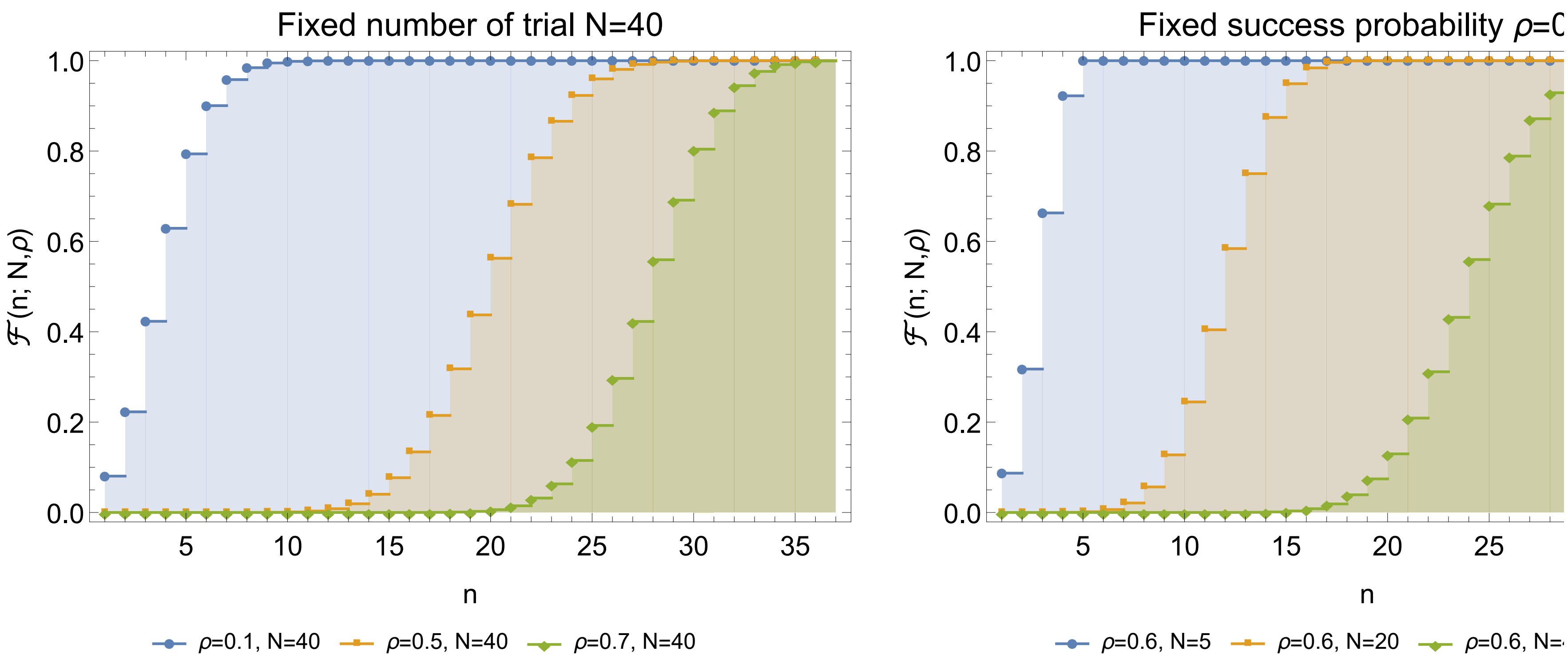
```
1 - CDF[NormalDistribution[56, 4.96], 59.5]
```

```
0.240204
```

Binomial Distribution

CDF - cumulative distribution function

$$\mathcal{F}(k;\ N, \rho) = \mathcal{P}(X \leq k) = \sum_{n=0}^{k} \binom{N}{n} \rho^n (1-\rho)^{N-n}$$



Example

In a sample of 100 galaxy clusters selected by automatic techniques,10 contain a dominant central galaxy. We plan to check a different sample of 30 clusters, now selected by X-ray emission. How many of these clusters do we expect to have a dominant central galaxy?

If we assume that the 10 per cent probability holds for the X-ray sample, then the chance of getting n dominant central galaxies is

$$\mathcal{P}(n) = \binom{30}{n} 0.1^{n} \times 0.9^{30-n}$$

Out[∘]//TableForm=

galaxies found	probability	cumulative probability
1	0.141304	0.183695
2	0.227656	0.411351
3	0.236088	0.647439
5	0.102305	0.92681
8	0.00576379	0.99798
10	0.000365277	0.999911
15	3.19373×10^{-8}	1.
20	1.0476×10^{-13}	1.

Example

$$\mathcal{P}(n) = \binom{30}{n} 0.1^n \times 0.9^{30-n}$$

galaxies found	probability	cumulative probability
1	0.141304	0.183695
2	0.227656	0.411351
3	0.236088	0.647439
5	0.102305	0.92681
8	0.00576379	0.99798
10	0.000365277	0.999911
15	3.19373×10^{-8}	1.
20	1.0476×10^{-13}	1.

For example, the chance of getting 10 is < 1%; if we found this many we would be suspicious that the X-ray cluster population differed from the general population.

Example

Suppose we made these observations and did find 10 centrally-dominated clusters. What can we do with this information?

A Bayesian calculation that parallels the supernova example! Assuming the X-ray galaxies are a homogeneous set, we can deduce the probability distribution for the fraction of these galaxies that have a dominant central galaxy. A relevant prior would be the results for the original larger survey.

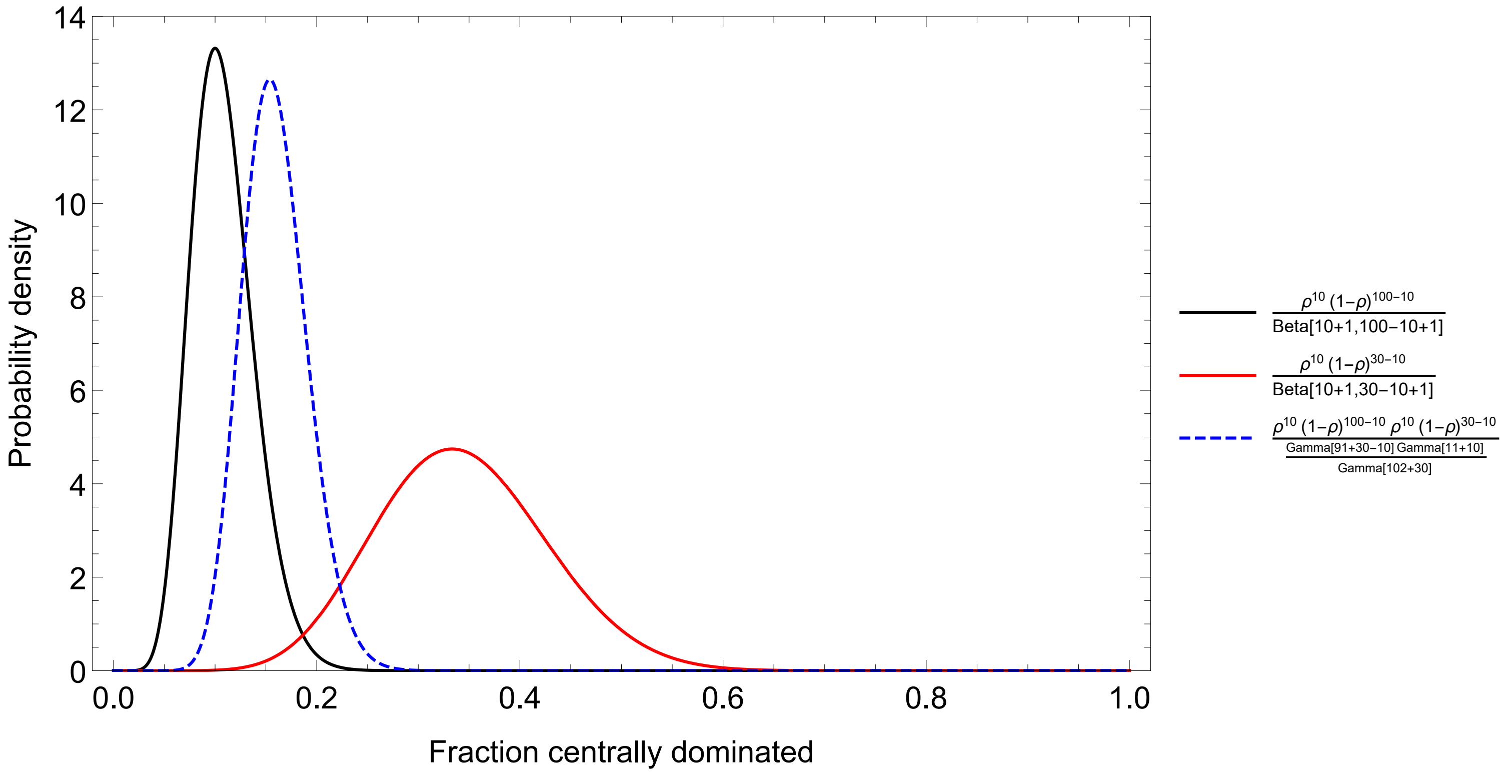
This is the corresponding normalizing constant:

```
Integrate[ρ^10 (1 - ρ)^(100-10) ρ^n (1 - ρ)^(m-n), {ρ, 0, 1}, Assumptions :→ {n > 0, m - n > 0}] // FunctionExpand
```

$$\frac{\text{Gamma}[91+m-n]\ \text{Gamma}[11+n]}{\text{Gamma}[102+m]}$$

Example

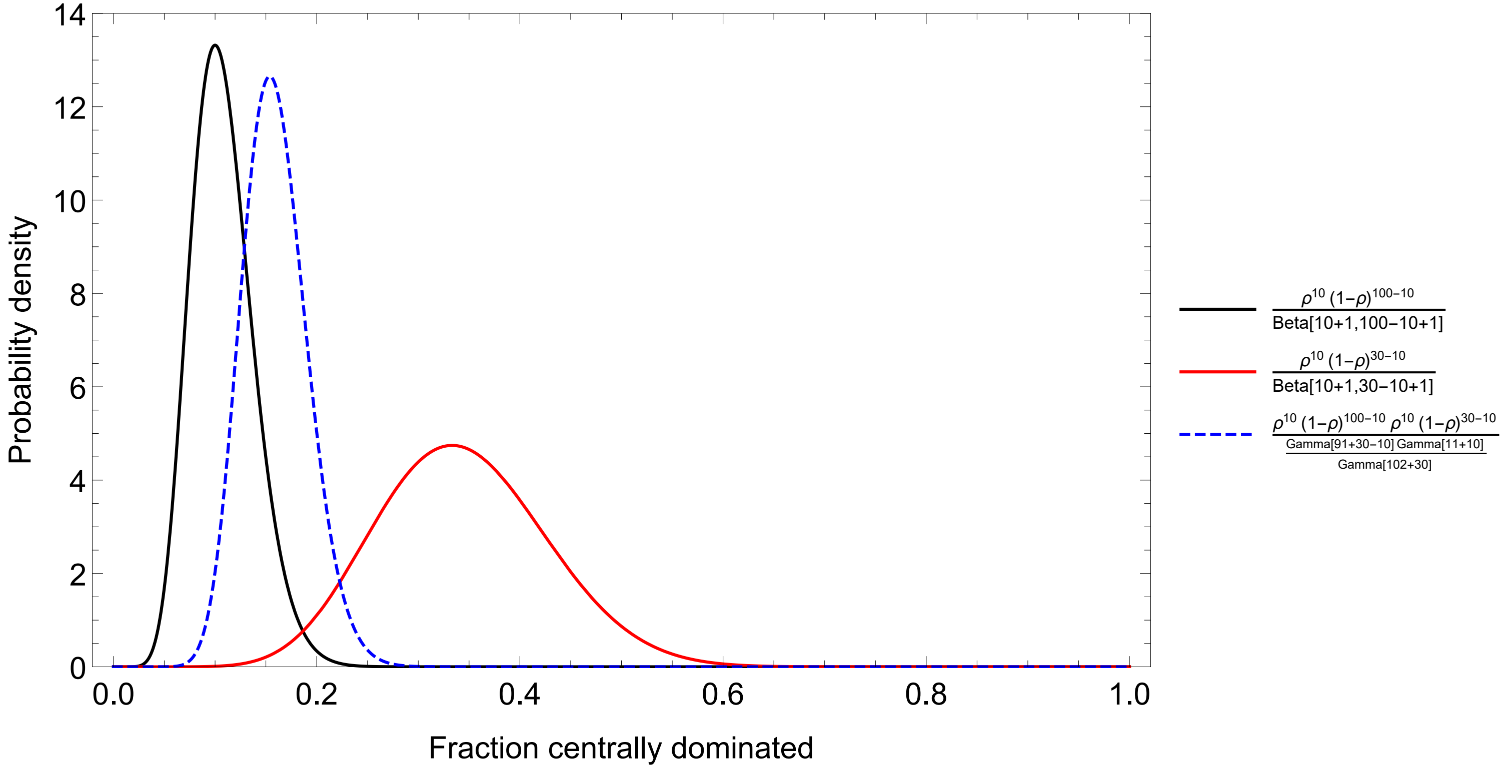
Out[]=



The posterior probability distribution for the observation that 10/30 X-ray selected clusters are centrally-dominated. The red line uses a uniform prior distribution for this fraction; the dashed line uses the prior derived from an assumed previous sample in which 10 out of 100 clusters had dominant central members. The black curve shows the distribution for this earlier sample.

Example

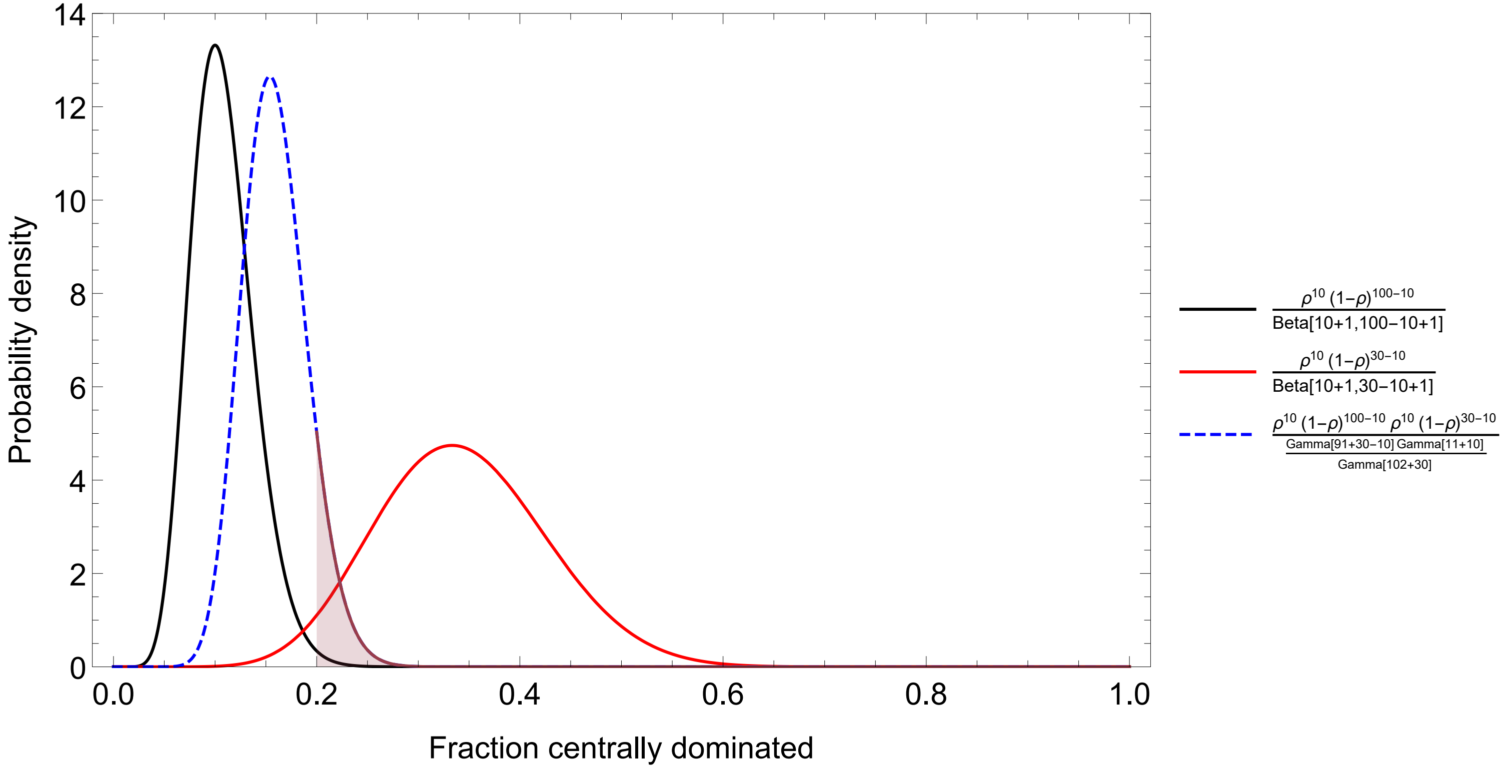
Out[ ]=



The figure makes clear that the data are not really sufficient to alter our prior very much. For example, there is only a 10~per cent chance that the centrally-dominant fraction exceeds even 0.2; and indeed we see that the possibility of it being as high as 33% is completely negligible. Our X-ray clusters differ markedly from the general population.

Example

Out[•]=




```
NIntegrate[(ρ^10 (1 - ρ)^(100-10) ρ^10 (1 - ρ)^(30-10)) / (Gamma[91 + 30 - 10] Gamma[11 + 10]) / Gamma[102 + 30], {ρ, 0.2, 1}]
```

0.103948

Geometric Distribution

The geometric distribution gives the probability that the first occurrence of success requires k independent trials, each with success probability p. If the probability of success on each trial is p, then the probability that the k-th trial (out of k trials) is the first success is for $k = 1, 2, 3, \ldots$

Out[∘]//TraditionalForm=

$$\boxed{f(X = k,\ p) = p\,(1 - p)^{k-1}}$$

The mean (expectation value) is

Out[∘]//TraditionalForm=

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

The variance is

Out[∘]//TraditionalForm=

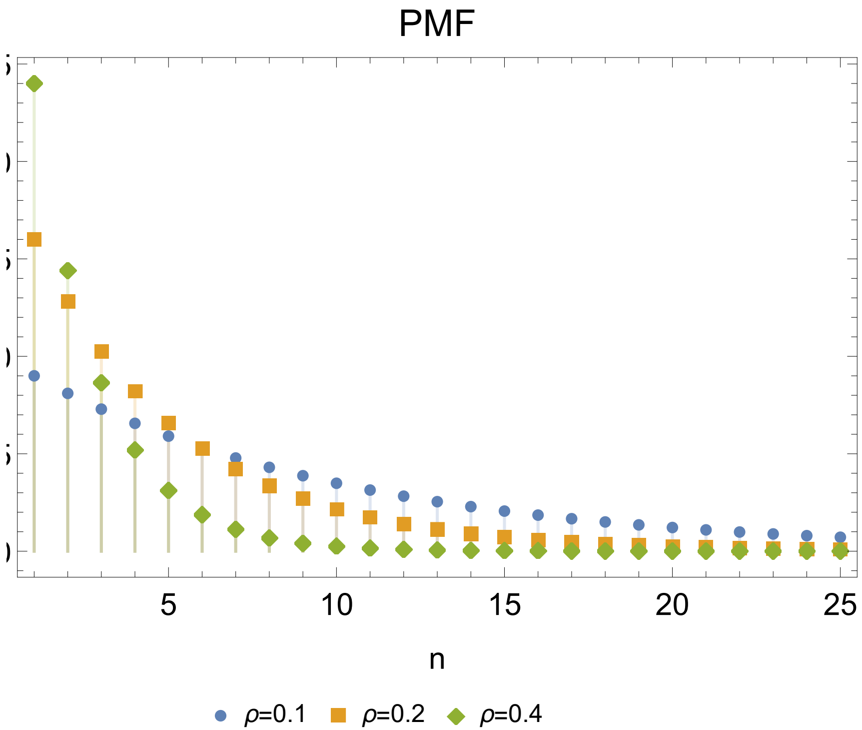
$$\boxed{\mathrm{var}(X) = \frac{1 - p}{p^2}}$$

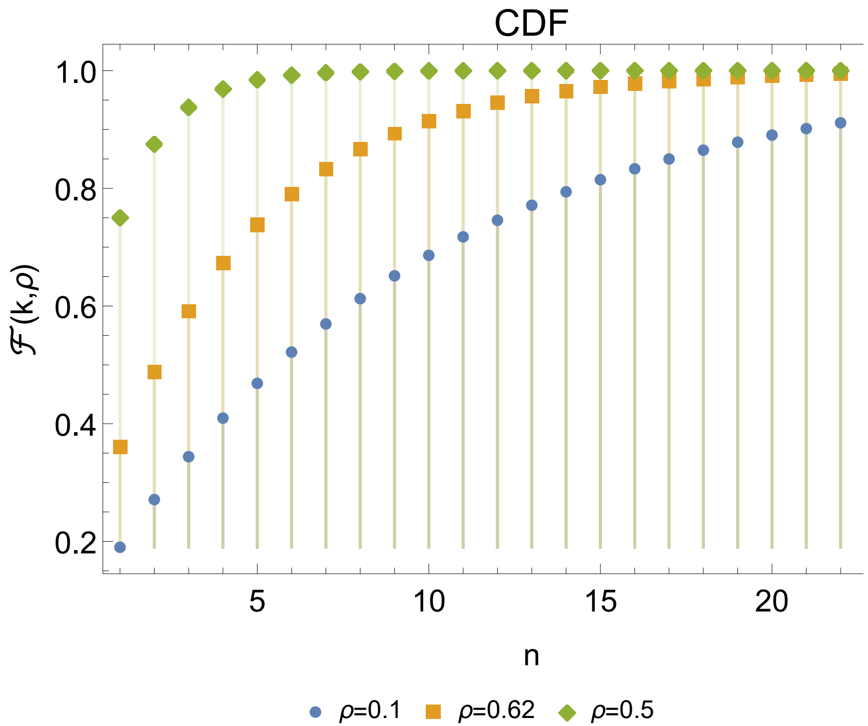
P.M.F. - probability <u>mass</u> function and C.D.F. Cumulative distribution function

$$f(k;\ \rho) = \mathcal{P}(X = k) = \rho(1 - \rho)^{k-1}$$

$$F(k;\ \rho) = \mathcal{P}(X = k) = 1 - (1 - \rho)^{k}$$

Out[ ]=





The term probability **<u>mass</u>** function is used for discrete random variables.

Beta Distribution

the beta distribution is the conjugate prior probability distribution for the Bernoulli, binomial, negative binomial and geometric distributions

Out[ ]=

$$\boxed{f(X = x, \alpha, \beta) = \frac{x^{\alpha-1} (1-x)^{\beta-1}}{\mathrm{B}(\alpha, \beta)}} \quad \text{with} \quad \boxed{\mathrm{B}(\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}}$$

where Γ is the gamma function and *B* is the Beta function, which is the normalizing constant to ensure that *f* is a probability density function

Out[ ]//TraditionalForm=

$$\int_0^1 x^{\alpha-1} (1-x)^{\beta-1} \, dx = \mathrm{B}(\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}$$

The mean value is

Out[ ]//TraditionalForm=

$$\boxed{\mathrm{E}(X) = \frac{\alpha}{\alpha+\beta}}$$

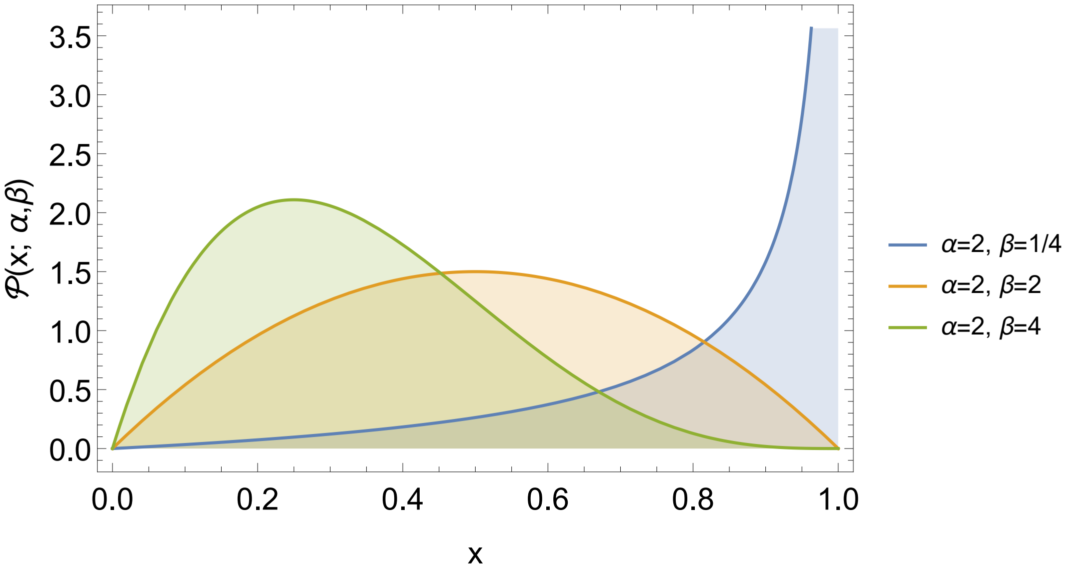
The variance is

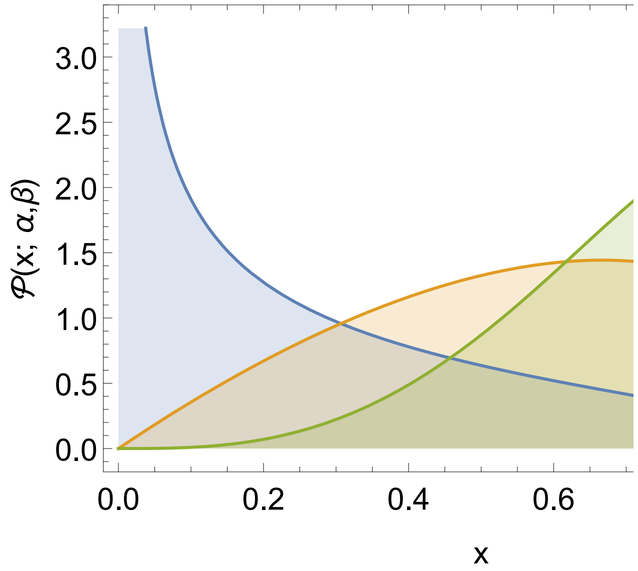
Out[ ]//TraditionalForm=

$$\boxed{\mathrm{var}(X) = \frac{\alpha\,\beta}{(\alpha+\beta)^2\,(\alpha+\beta+1)}}$$

PDF - probability density function

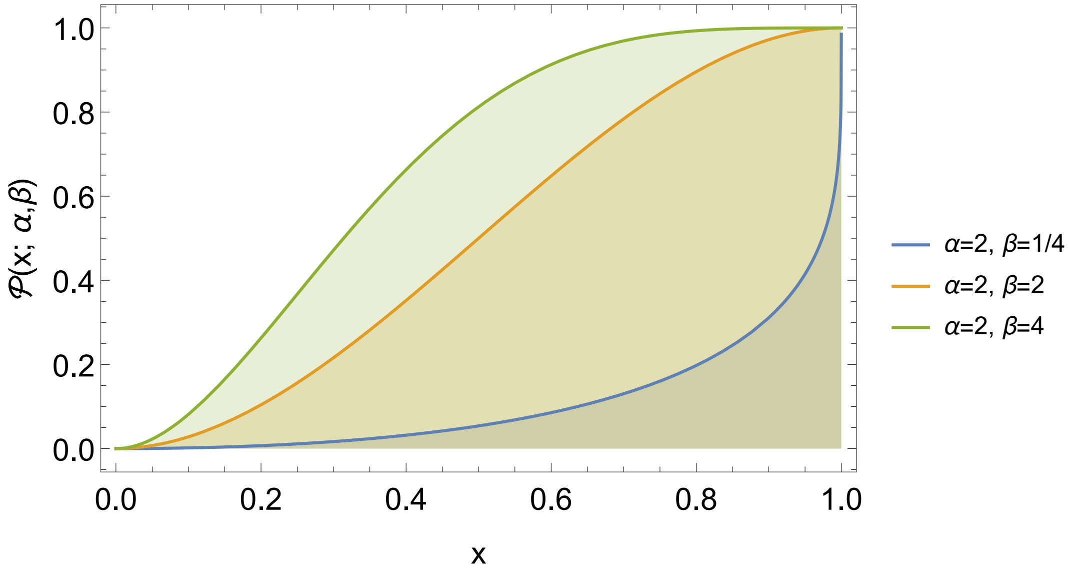
Out[◦]=

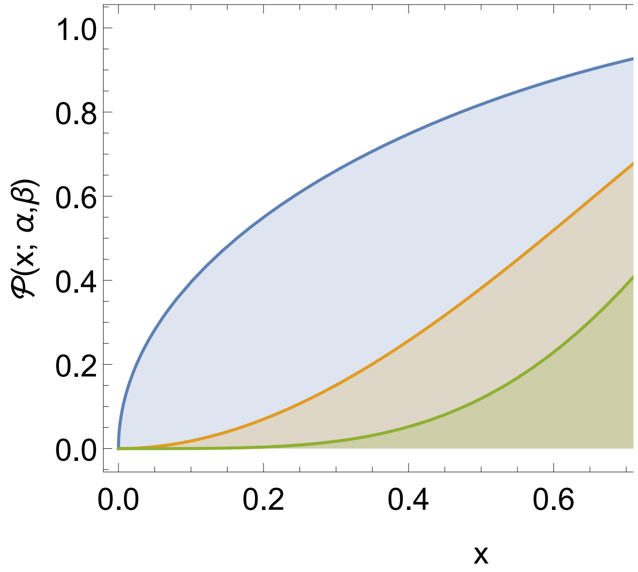




CDF - cumulative distribution function

Out[ ]=





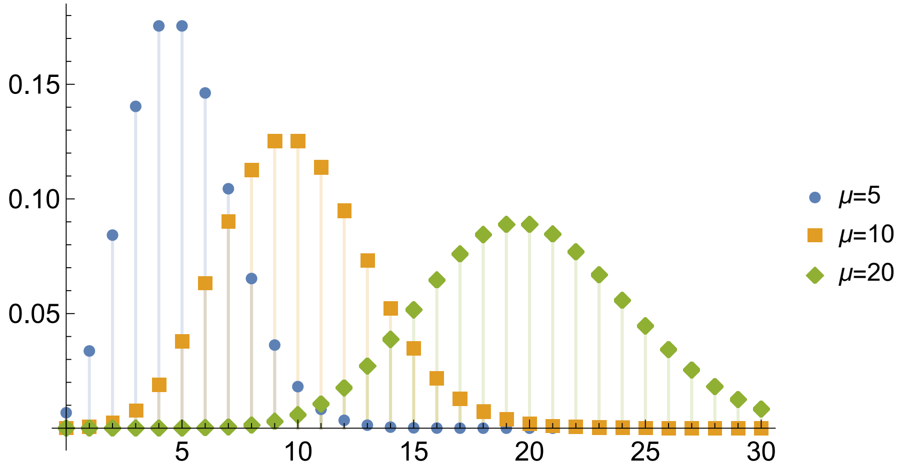
The Poisson Distribution

The Poisson distribution derives from the binomial in the limiting case of very rare events and a large number of trials, so that although $\rho \to 0$, $N \times \rho \to$ (a finite value). Calling this finite **mean value** μ, the Poisson distribution is

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

The variance of the Poisson distribution is also μ.

Out[ ]=



For higher values of μ the distribution becomes more symmetric.

The Poisson Distribution

In[ ]:=
```
StandardDeviation[PoissonDistribution[μ]]
```

Out[ ]=
$$\sqrt{\mu}$$

```
Skewness[PoissonDistribution[μ]]
```

$$\frac{1}{\sqrt{\mu}}$$

```
Limit[Skewness[PoissonDistribution[μ]], μ → Infinity]
```

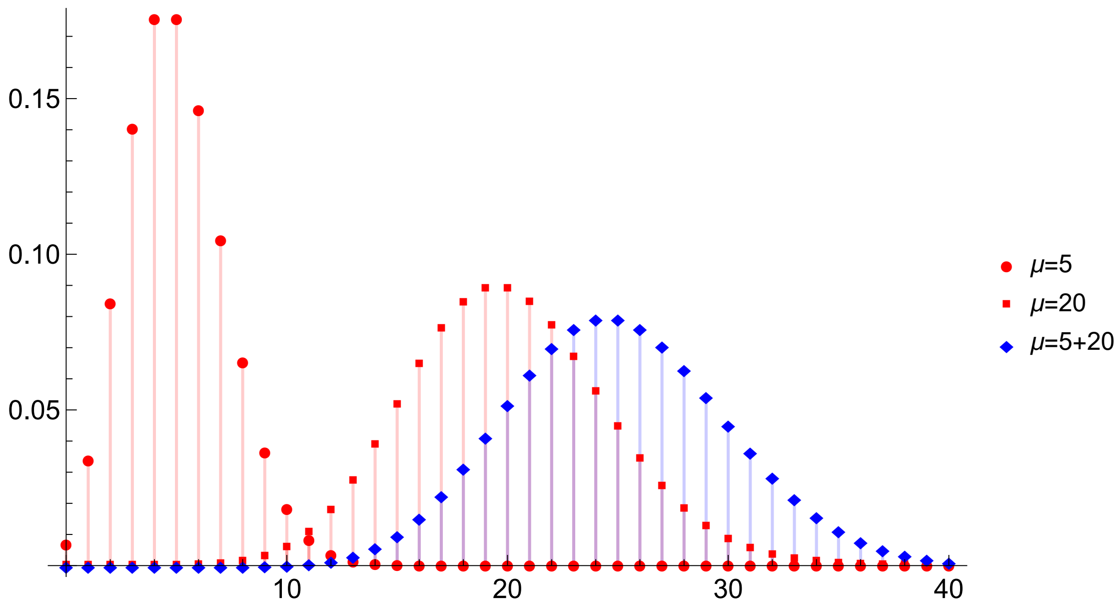
0

The Poisson Distribution

The sum of Poisson variables is Poisson distributed:

```
TransformedDistribution[u + v, {u ≈ PoissonDistribution[λ1], v ≈ PoissonDistribution[λ2]}]
```

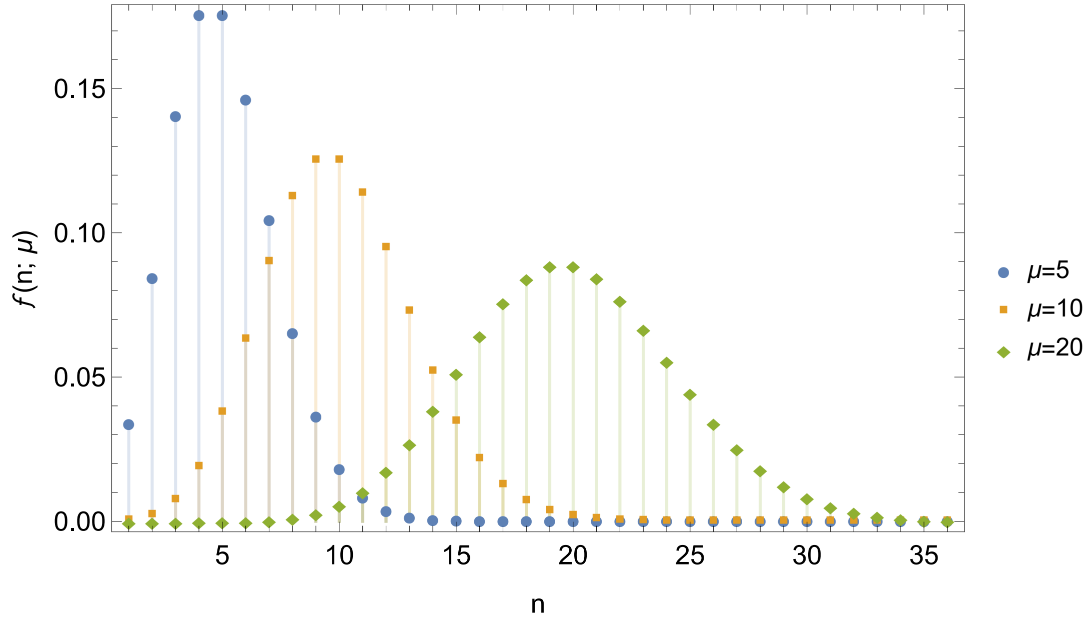
```
PoissonDistribution[λ1 + λ2]
```



The Poisson Distribution

P.M.F. - probability mass function

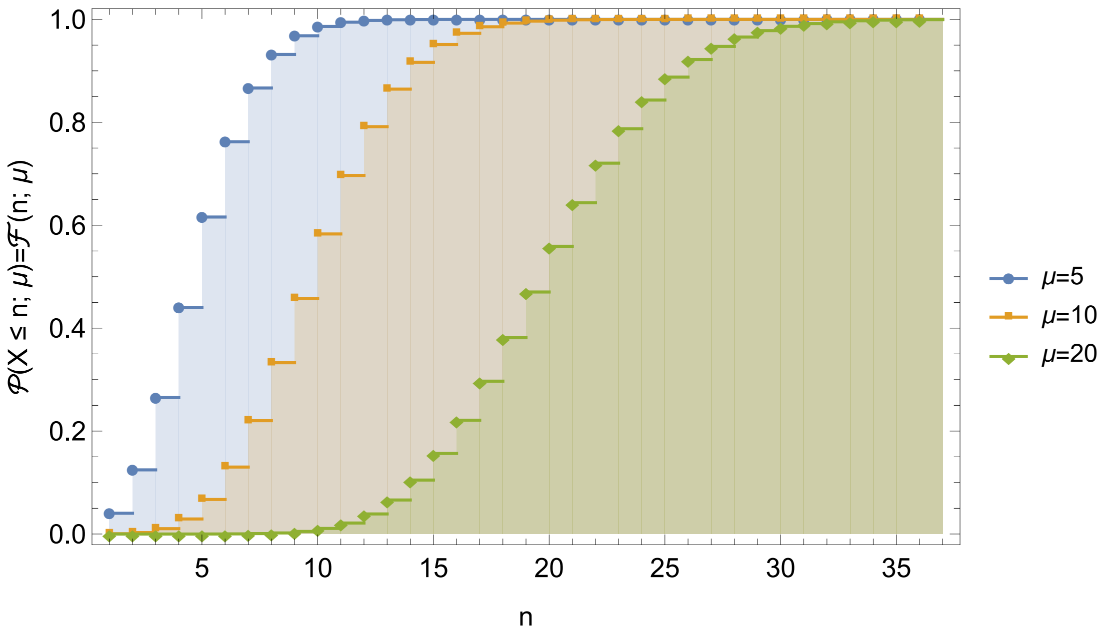
$$f(n;\ \mu) = \mathcal{P}(X = n) = \sum_{n=0}^{k} \frac{\mu^n}{n!} e^{-\mu}$$



The Poisson Distribution

CDF - cumulative distribution function

$\mathcal{F}(k;\ \mu) = \mathcal{P}(X \leq k) = \sum_{n=0}^{k} \frac{\mu^n}{n!} e^{-\mu}$



Example - Typographical errors

Typographical errors in a book are occurring randomly according to a Poisson process. On 384 pages, 158 errors are counted. Find the distribution of errors per page:

The mean numbers of errors on a page is $\mu = \frac{158}{384} = 0.411458$. The probability to find n errors on a single page is $\mathcal{P}(n) = \frac{0.412^n}{n!}\,e^{-0.412}$.

errors on one page	probability
0	0.662683
1	0.272666
2	0.0560955
3	0.00769365
4	0.000791404
5	0.0000651259

Example - Typographical errors

errors on one page	probability
0	0.662683
1	0.272666
2	0.0560955
3	0.00769365
4	0.000791404
5	0.0000651259

The probability of fewer than 2 errors per page is: $\mathcal{P}(0) + \mathcal{P}(1) = 0.663 + 0.273$. This is the cumulative distribution

```
CDF[PoissonDistribution[158/384], 1] // N
```

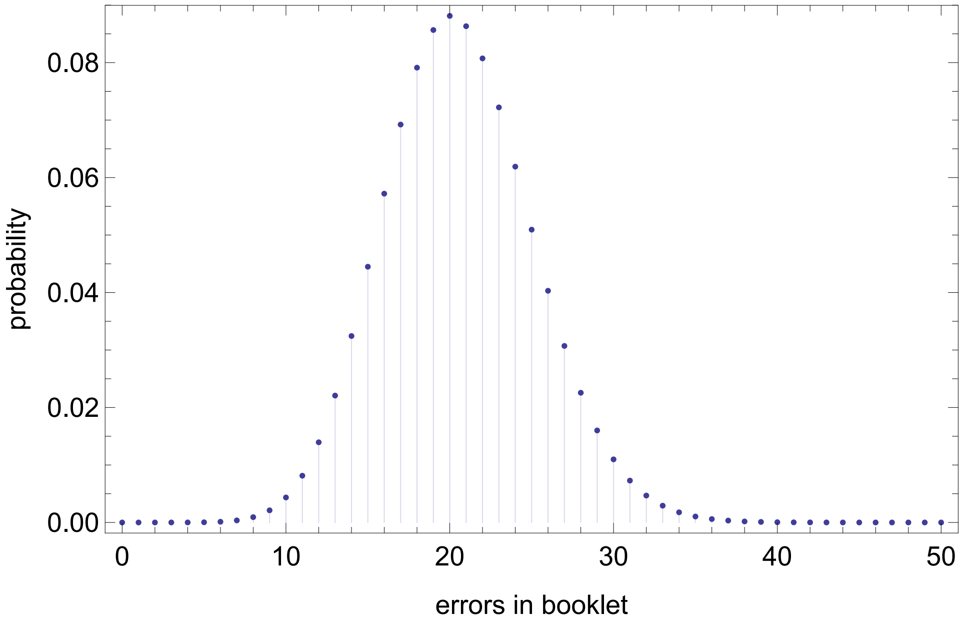
```
0.93535
```

The probability to find 1 or more errors per page is $\mathcal{P}(n \geq 1) = (1 - \mathcal{P}(0)) = 1 - 0.663 = 0.337$. This sometimes also called the **survival function**.

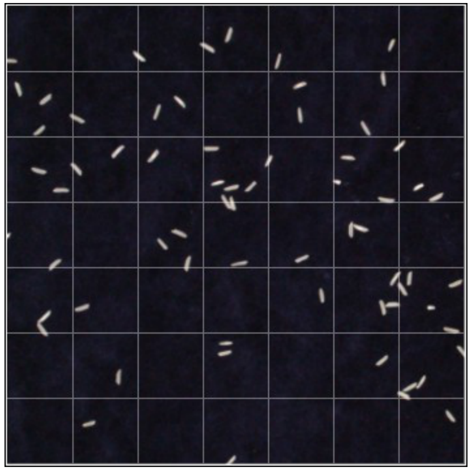
Example - Typographical errors

If we are not interested in the number of errors per page, but, e.g. the number of errors in a 50 page booklet, we have to adjust the distribution mean: $\mu = p \times \frac{158}{384} = 50 \times \frac{158}{384} = 20.5729$.

The probability to find n errors in the booklet is $\mathcal{P}(n) = \frac{20.5729^n}{n!}\, e^{-20.5729}$.



Example - Scattered rice grains



Randomly scattered rice on ground: 66 rice grains in 49 squares.

In[ ]:=
```
dat = {{0, 16}, {1, 14}, {2, 10}, {6, 3}, {4, 1}, {5, 2}, {6, 0}};
```

66 rice grains on 49 squares gives $\lambda = N/n$ =66/49=1.35 grains/square.

$\mathcal{P}(n) = \frac{\mu^n}{n!} e^{-\mu} \longrightarrow \mathcal{P}(n) = \frac{1.35^n}{n!} e^{-1.35}$

Out[ ]//TableForm=

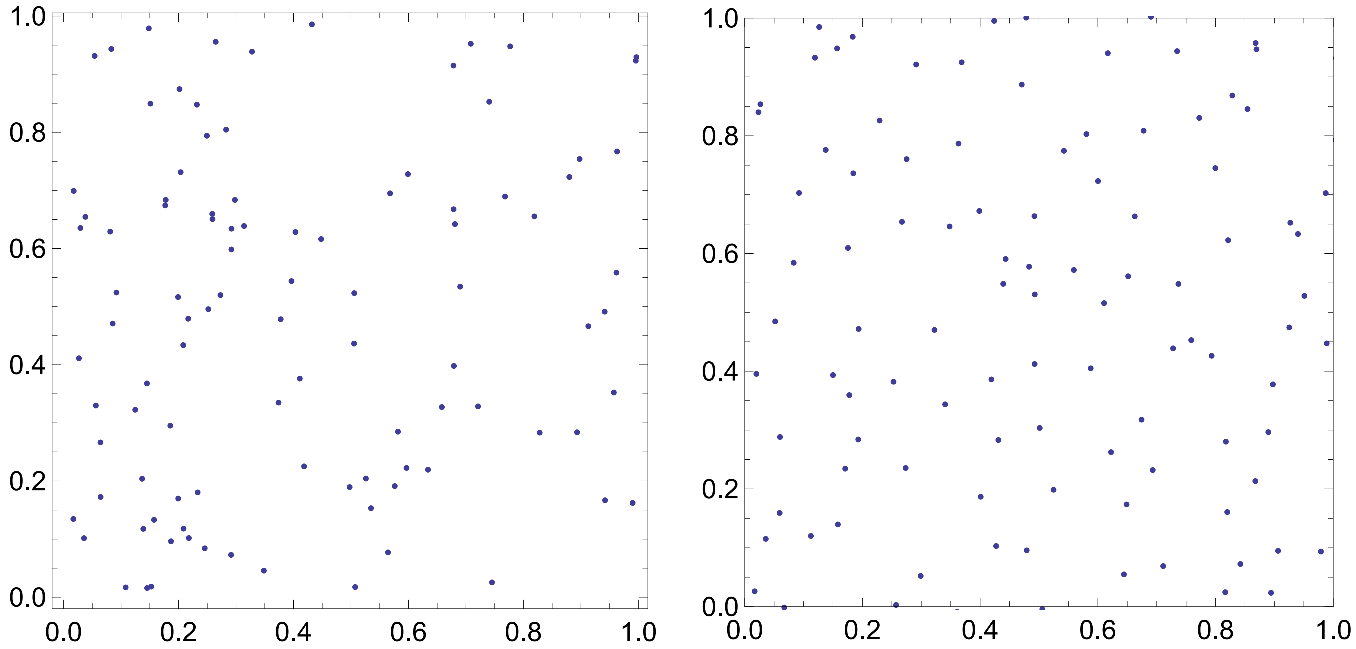
grains / square	expectation	counted
0	12.7417	16
1	17.1623	14
2	11.5583	10
3	5.18944	3
4	1.74746	1
5	0.470745	2
6	0.105678	0

Example - Scattered rice grains

grains/square	expectation	counted
0	12.7417	16
1	17.1623	14
2	11.5583	10
3	5.18944	3
4	1.74746	1
5	0.470745	2
6	0.105678	0

The probability that a square remains empty is $\mathcal{P}(0) = \frac{1.35^0}{0!}\, e^{-1.35} = e^{-1.35} = 0.25924$.

Example - Random distribution



In the figure, we placed 100 dots in a unit square. The left panel shows random placement (actually there are no true random number generators on a computer). In the right panel, we first created a regular grid of 10×10 dots and then slightly shifted them randomly in x and y direction. The left panel shows a much more irregular pattern compared to the right panel. We see prominent ‘holes’ and ‘clumping’ in the left panel (Poisson clumping).

Example - Photon statistics

A familiar example of a process obeying Poisson statistics is the number of photons arriving during an integration. The probability of a photon arriving in a fixed interval of time is (often) small (at least at wavelengths shorter than infrared (IR)). The arrivals of successive photons are **independent**. Thus the conditions necessary for the Poisson distribution are met.

Hence, if the integration over time t of photons arriving at a rate λ has a mean of $\mu = \lambda t$ photons, then the fluctuation on this number will be $\sigma = \sqrt{\mu}$, because we know that the variance is μ (In practice we usually only know the number of photons in a single exposure, rather than the mean number; obviously we can then only estimate the μ.). There are the following limiting cases:

Example - Photon statistics

Suppose we are detecting our objects with no effective background (either from the sky or the instrument). The observed photons are solely from the objects measured. In this idealized situation, with $\mu = \lambda \times t$, the scatter on μ is (Poisson) $\sigma = \sqrt{\lambda \times t}$. If we integrate more, by simply waiting for more photons, the **photon-limited** case,

$$\sigma \propto \sqrt{t}, \text{ while signal} \propto t.$$

Thus,

$$\text{signal}/\text{noise} \propto \sqrt{t}$$

Example - Photon statistics

Now suppose our object is barely visible against the sky background. Our signal is still $\mu = \lambda \times t$, but our noise is $\sigma = \sqrt{\lambda_{\text{sky}} \times t}$. The net result is

$$S/N \propto \frac{\lambda \times t}{\sqrt{\lambda_{\text{sky}} \times t}}$$

Again,

$$S/N \propto \sqrt{t}$$

but the sky emission $\lambda_{\text{sky}} \times t$ will be much stronger than the $\lambda \times t$ (**sky-limited** case). So, in order to achieve the same S/N as in the first case, we need to integrate much longer.

Also note, that the fact that $S/N \propto \sqrt{t}$ poses a practical limit on how good an observation can be; half the S/N requires 4 times longer observations, etc..

Example - Photon statistics

Imagine, we have very bright, photon-limited objects, for which we require very short exposures only, e.g. observation of bright stars or hot dust. Thus, with strong signal and short exposure, the **readout noise** of the detector (fixed, time-independent, added on top of the signal) may dominate the errors. Calling this fixed error σ_{readout}:

$$S/N \propto \frac{\lambda \times t}{\sigma_{\text{readout}}}, \text{ or } \propto t$$

Example - Photon statistics

At the long-wavelength end of the spectrum (sub-millimeter & radio), the photon flux from objects is so strong, that the Poisson statistics of rare events no longer applies! We are in the **receiver-limited** case; S/N is governed by the receiver sensitivity. We have now such a quantity of photons from the object (photon flux S), that it is the receiver noise which requires the integration:

$$S/N \propto \frac{S}{\sigma_{\text{rec}}/\sqrt{t}}, \text{ or } \propto \sqrt{t}$$

for a receiver of thermal noise σ_{rec}.

Power law distribution (Pareto distribution)

Take N objects (or events) with measured property (e.g. luminosity) either greater than a value L or within a bin dL centered on L. With γ as the power-law exponent

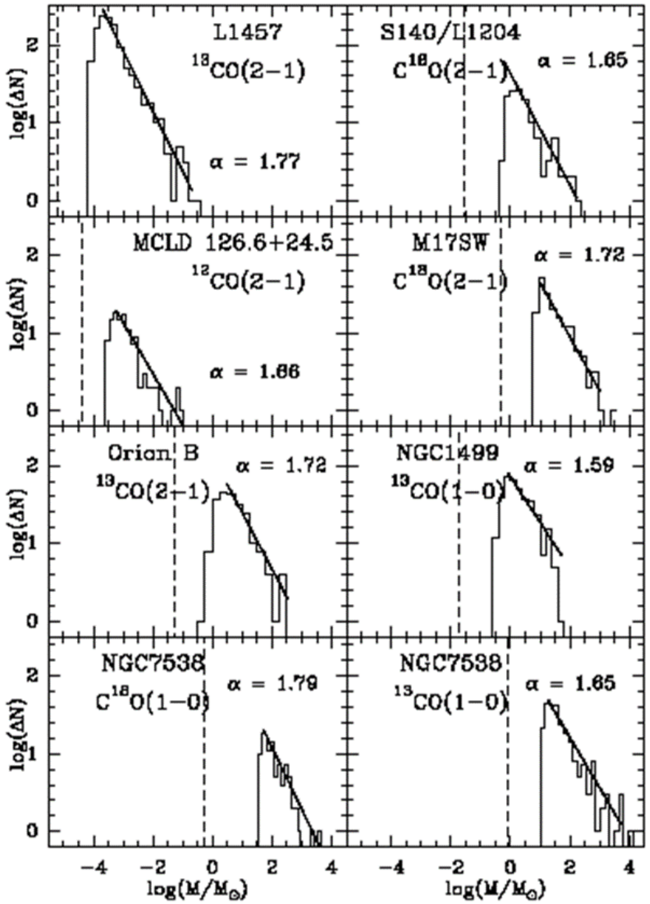
$N(>L) = K \times L^{\gamma+1}$ integral form, $N > L$, or

$dN = (\gamma + 1) \times K \times L^{\gamma} \times dL$, differential form, dN objects in dL.

- This is a scale-free or scale-independent distribution, because if
$$f(x) = x^{\gamma}, \text{ then } f(a\,x) = a^{\gamma}\,x^{\gamma} = \text{const} \times x^{\gamma} = \text{const} \times f(x)$$
(definition of scale independence).
- Not formally a probability distribution because $\int dN \to \infty$
- formally, mean and variance $\to \infty$
- normally there are physical bounds so that it works
- Steep power laws -4 < γ < 0, extending over decades, occur in astronomy frequently

Power law distribution (Pareto distribution)

Example



Observations of molecular clouds show an almost universal clump – mass distribution with

$dN / dM \propto M^{-1.6...-1.8}$

(Kramer et al. 1998)

Power law distribution (Pareto distribution)

Describing results for objects selected from power-law distributions in terms of means and sigmas becomes very misleading (or wrong). Generally, the slopes of the power laws are steep and inverse, i.e. there are many more objects with small L than large ⇒ **strong bias when objects are drawn from such a parent population**.

- Astronomical examples:
  - Salpeter Mass function (IMF)
  - Clump-mass distribution and mass-size relation for molecular clouds
  - magnitude or source counts (surface density of objects on the sky)
  - luminosity functions
  - primordial fluctuations spectrum
- There are always more faint objects, more low-mass or low-luminosity objects than high ⟶ **γ is invariably negative.**
- **Power-law distributions indicate that something interesting is going on.**

Power law distribution (Pareto distribution)

- **Pitfalls:**
  - Characterizing by means or variances completely misleading.The Central Limit Theorem fails us badly.
  - The index!
    - differential or integral form?
    - binning: uniform or on a $\Delta\log(L)$ scale $\left(d(\log(L)) = \text{const} \times L^{-1}\, dL\right) \rightarrow$ index reduced by 1!

Given a fixed range of power law between a and b, the **mean** can be calculated:

$$\mu = \left(\frac{\gamma+1}{\gamma+2}\right)\left[\frac{b^{\gamma+2}-a^{\gamma+2}}{b^{\gamma+1}-a^{\gamma+1}}\right]$$

The expression breaks down for $\gamma = -1$ or -2;

Power law distribution (Pareto distribution)

PDF - probability density function

In a looser sense, a power-law probability distribution is a distribution whose density function (or mass function in the discrete case) has the form

$f(x) \propto L(x)\, x^{-\alpha}$ **BE CAREFUL - DIFFERENT DEFINITON!**

For instance, if $L(x)$ is the constant function, then we have a power law that holds for all values of x. In many cases, it is convenient to assume a lower bound $x_{\min}$ from which the law holds. Combining these two cases, and where x is a continuous variable, the power law has the form

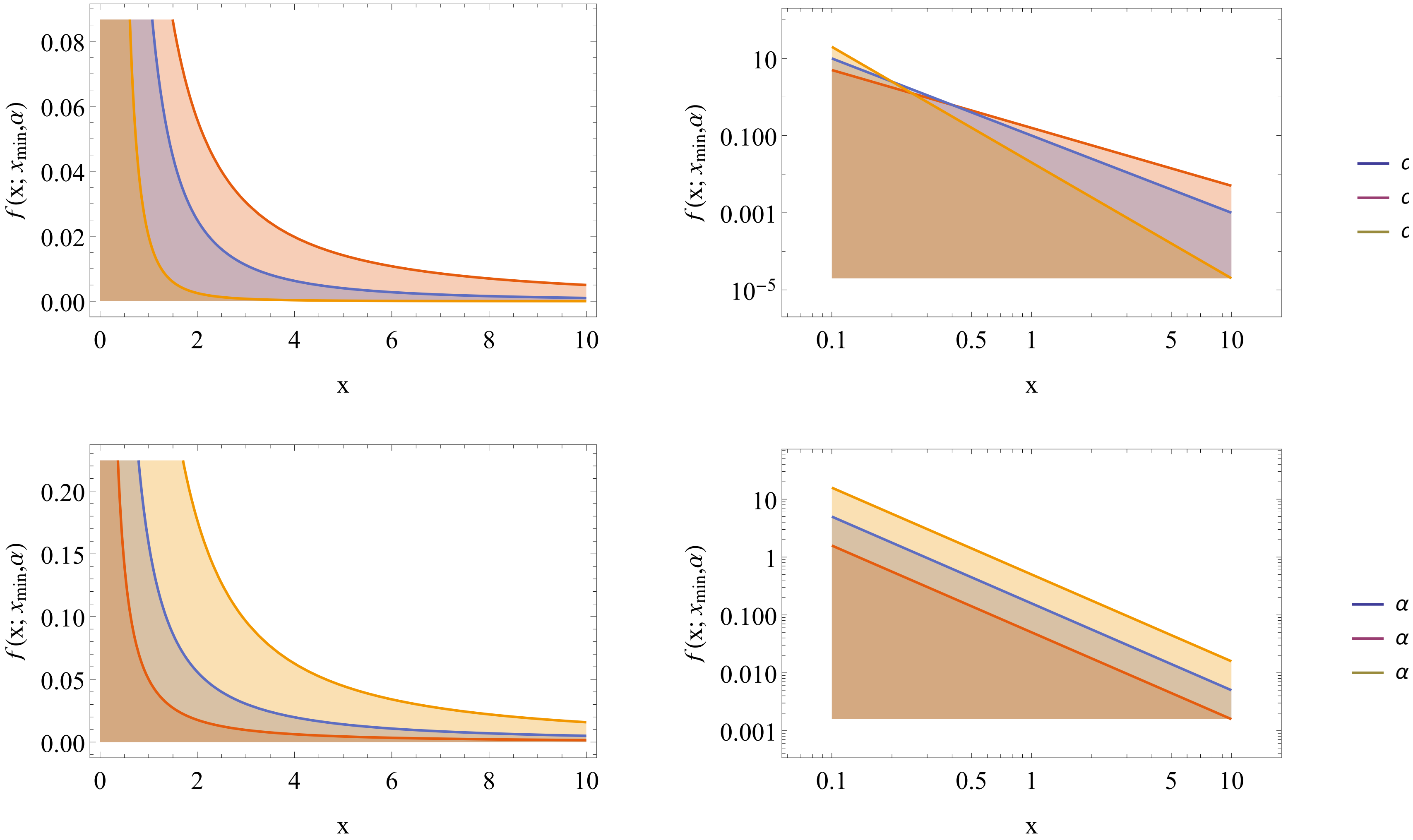
$$f(x) = \frac{\alpha-1}{x_{\min}} \left(\frac{x}{x_{\min}}\right)^{-\alpha}$$

The pre-factor is the normalizing factor. With this distribution, we can now define, e.g. the mean value

$$\langle f(x) \rangle = \int_{x_{\min}}^{\infty} x\, f(x)\, dx = \frac{\alpha-1}{\alpha-2} x_{\min}$$

Power law distribution (Pareto distribution)

PDF - probability density function



Power law distribution (Pareto distribution)

CDF - cumulative distribution function

In general, power - law distributions are plotted on doubly logarithmic axes, which emphasizes the upper tail region.The most convenient way to do this is via the **(complementary) cumulative distribution** (cdf),

Complementary cumulative distribution function

$$\overline{\mathcal{F}}(x) = \mathcal{P}(X > x) = C \int_x^\infty f(x)\, dx = \frac{\alpha-1}{x_{\min}{}^{-\alpha+1}} \int_x^\infty X^{-\alpha}\, dX = \left(\frac{x}{x_{\min}}\right)^{-\alpha+1}$$

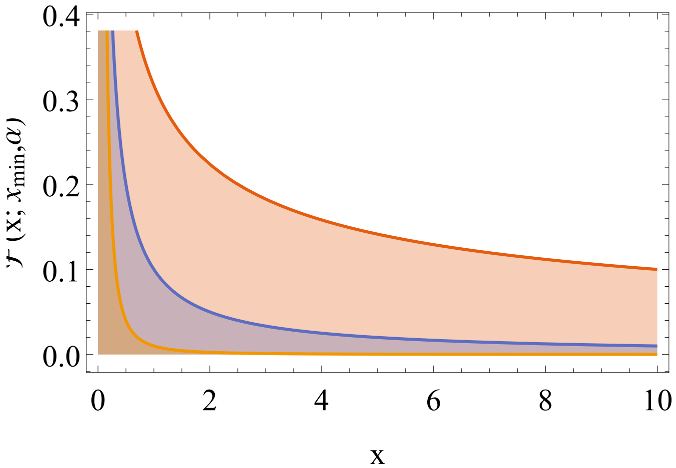
The regular CDF is then

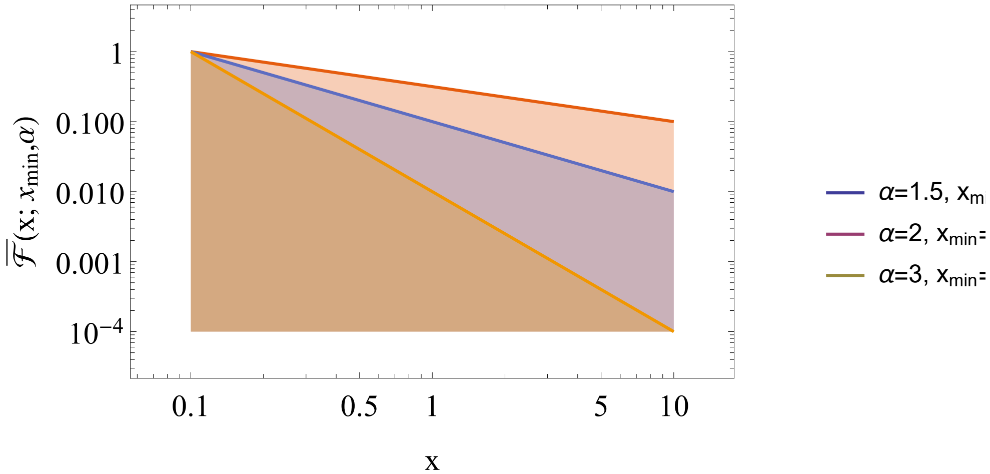
$$\mathcal{F}(x) = \mathcal{P}(X \leq x) = 1 - \overline{\mathcal{F}}(x) = 1 - \left(\frac{x}{x_{\min}}\right)^{-\alpha+1}$$

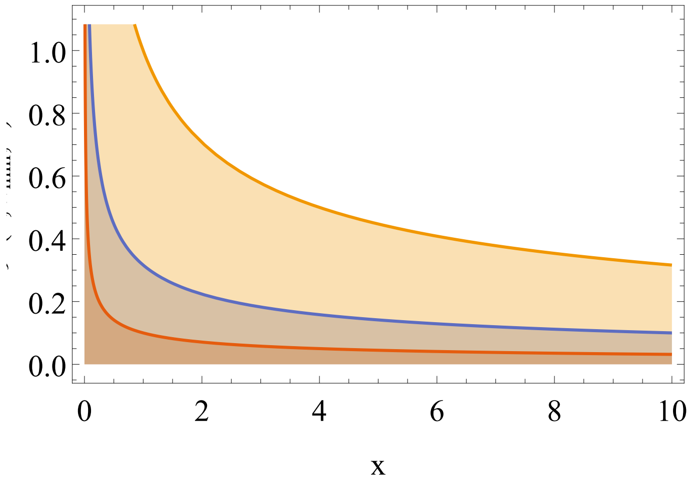
Note that the cdf is also a power-law function, but with a smaller scaling exponent.

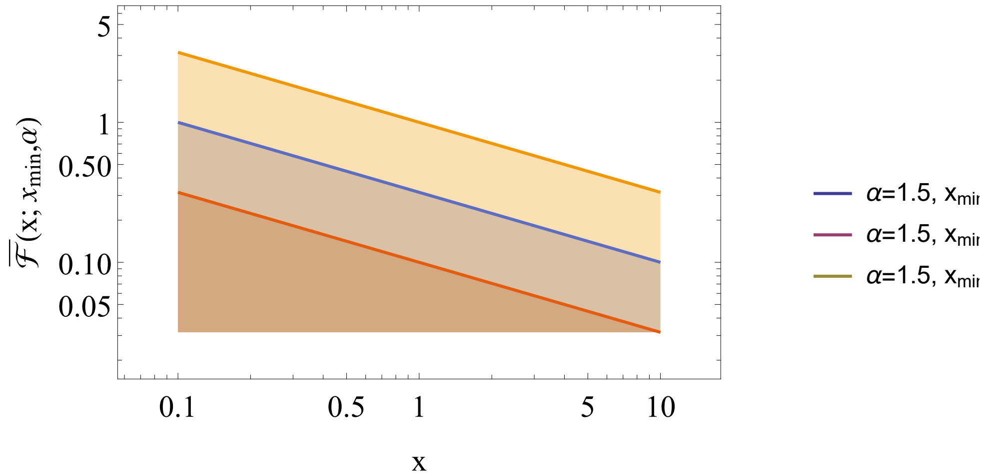
Power law distribution (Pareto distribution)

CDF - cumulative distribution function







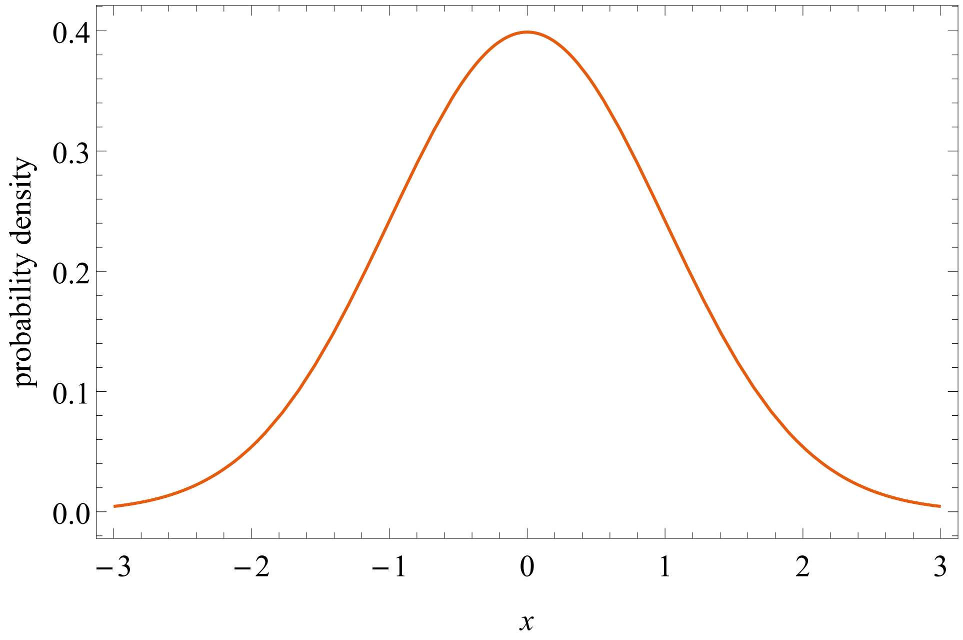


Gaussian (Normal) distribution

The Binomial and the Poisson distribution both tend to the Gaussian distribution (large *N* in case of Binomial, large μ in case of Poisson). The univariate Gaussian (Normal) distribution is:

Out[∘]//TraditionalForm=

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



Gaussian (Normal) distribution

Properties

- It is symmetric around the point $x = \mu$ which is at the same time the mode, the median and the mean of the distribution.
- It has two inflection points (where the second derivative of $\mathcal{P}$ is zero and changes sign), located one standard deviation away from the mean, namely at $x = \mu - \sigma$ and $x = \mu + \sigma$.
- The area under the curve is unity (because of the normalizing factor $1/(\sigma\sqrt{2\pi})$)

```
Integrate[Exp[-(x - μ)^2/(2 σ^2)], {x, -∞, ∞}, Assumptions :→ σ > 0]
```

$\sqrt{2\pi}\ \sigma$

- The Full Width at Half Maximum (FWHM) is the width of a spectrum curve measured between those points on the y-axis which are half the maximum amplitude (log is the natural logarithm!)

Out[ ]//TraditionalForm=

$$\text{FWHM} = 2\sqrt{2\log(2)}\ \sigma \approx 2.335\sigma$$

In[ ]:=
```
Solve[1/2 == Exp[-(x - μ)^2/(2 σ^2)], x, Reals][[All, 1, 2]] // Differences
```

Out[ ]= $\{2\sqrt{\sigma^2}\ \sqrt{2\,\text{Log}[2]}\}$

Gaussian (Normal) distribution

Properties

The **mean value** of the Normal distribution **is** μ.

The **variance is** σ^2**.**

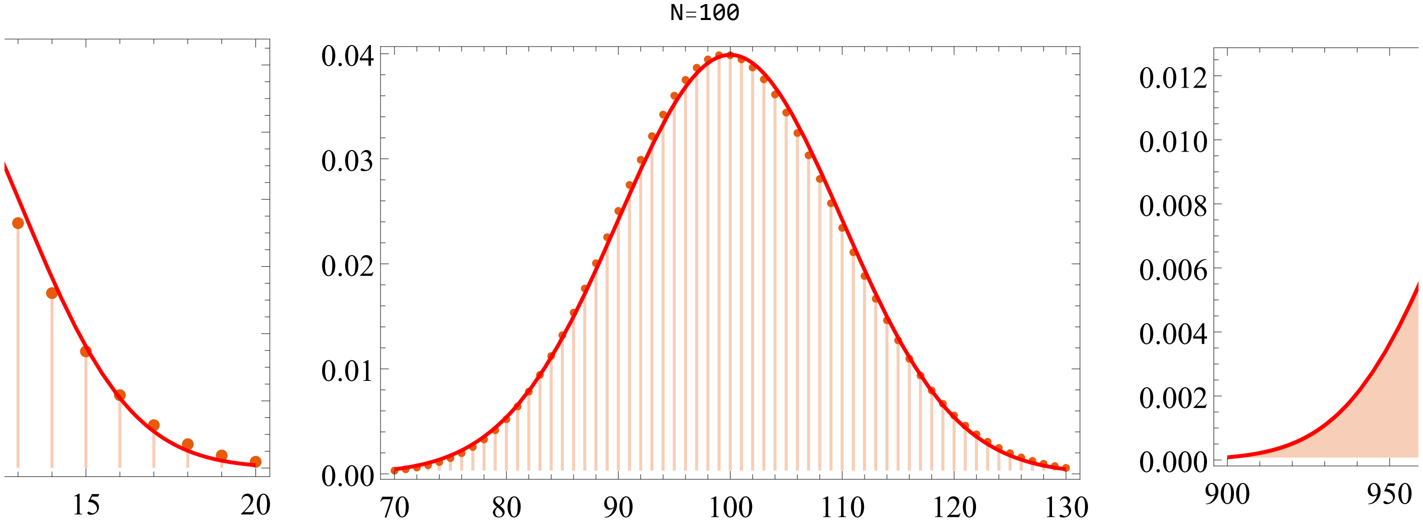
Gaussian (Normal) distribution

Properties

For a very large sample size N, the **discrete** Binomial distribution tends to the **continuous** Normal distribution with mean $\mu = N \times p$ and variance $\sigma^2 = N \times p \times (1 - p)$.

Out[ ]//TraditionalForm=

$$\boxed{\mathcal{P}_r(n) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(n-\mu)^2}{2\sigma^2}\right)}$$

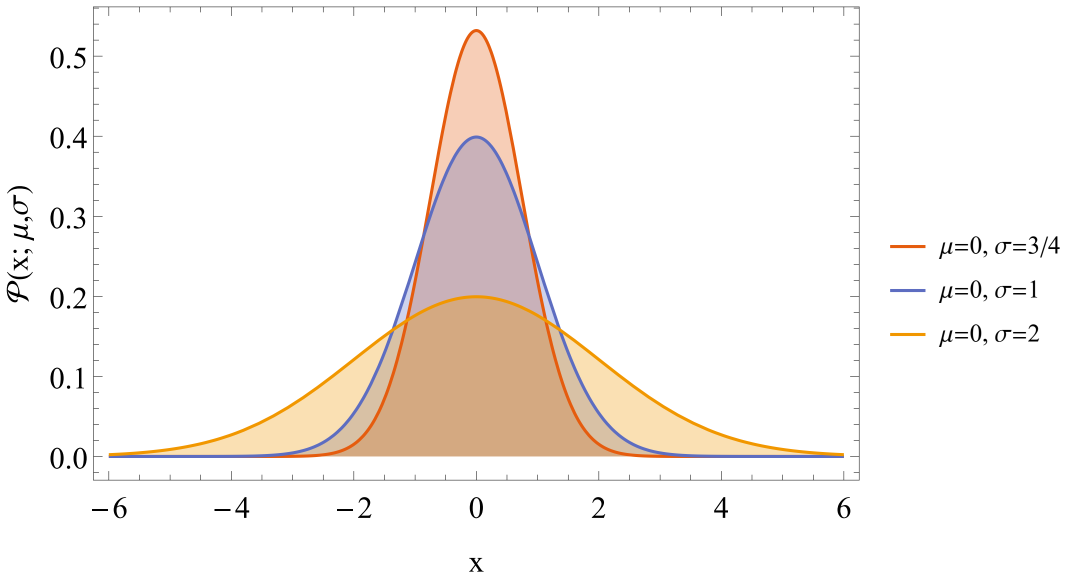


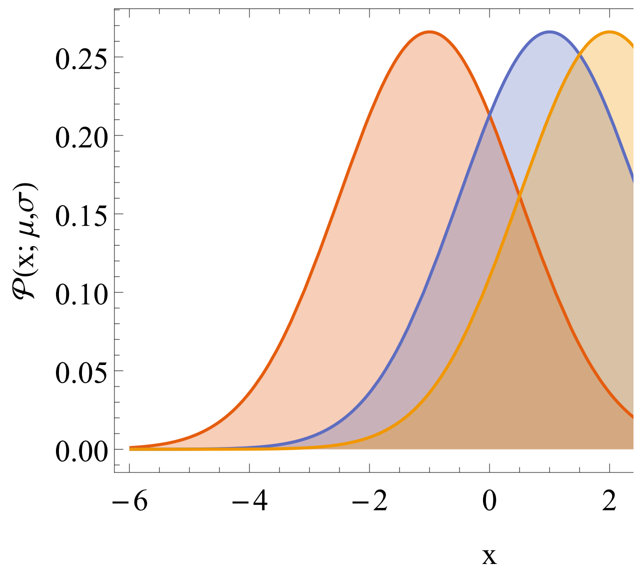
Gaussian (Normal) distribution

PDF - probability density function

Out[ ]//TraditionalForm=

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



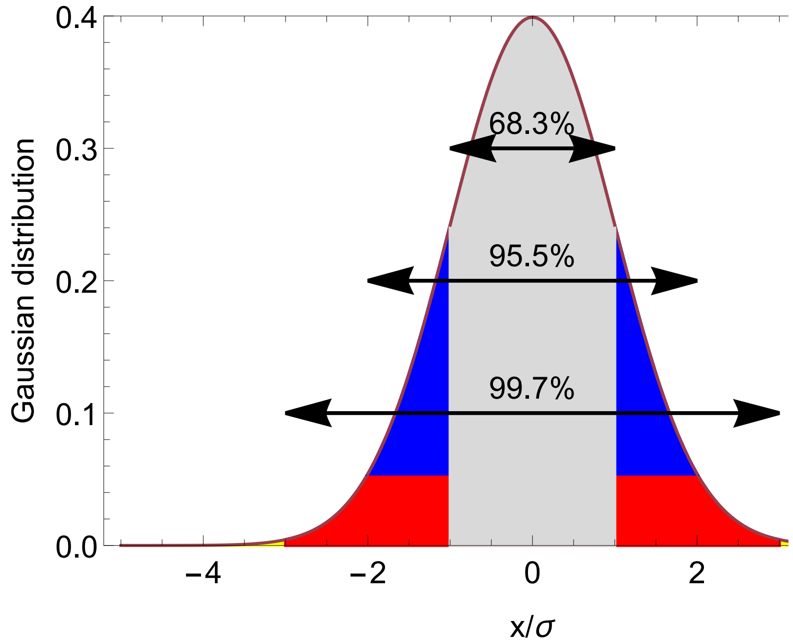


Gaussian (Normal) distribution

Out[•]=

Percentage area under the Gaussian curve

x/σ	one tail	both tails	between tails
0.	50.	100.	0.
0.5	30.8538	61.7075	38.2925
1.	15.8655	31.7311	68.2689
1.5	6.68072	13.3614	86.6386
2.	2.27501	4.55003	95.45
2.5	0.620967	1.24193	98.7581
3.	0.13499	0.26998	99.73
3.5	0.0232629	0.0465258	99.9535
4.	0.00316712	0.00633425	99.9937
4.5	0.000339767	0.000679535	99.9993
5.	0.0000286652	0.0000573303	99.9999



Gaussian (Normal) distribution

CDF - cumulative distribution function

Out[ ]//TraditionalForm=

$$\boxed{\mathcal{F}(k\,;\mu,\,\sigma)=\mathcal{P}_r(X\leq k)=\int_{-\infty}^{k}\frac{1}{\sigma\,\sqrt{2\,\pi}}\exp\left(-\frac{(x-\mu)^2}{2\,\sigma^2}\right)dx=\frac{1}{2}\left(1+\operatorname{erf}\left(\frac{k-\mu}{\sqrt{2}\,\sigma}\right)\right)}$$

Gaussian (Normal) distribution

CDF - cumulative distribution function

```
Assuming[σ > 0, ∫_{-∞}^{k} e^(-(x-μ)^2/(2 σ^2))/(Sqrt[2 π] σ) dx]
```

$$\frac{1}{2}\left(\operatorname{erf}\left(\frac{k-\mu}{\sqrt{2}\,\sigma}\right)+1\right)$$

```
CDF[NormalDistribution[μ, σ], x]
```

$$\frac{1}{2}\operatorname{erfc}\left(\frac{\mu-x}{\sqrt{2}\,\sigma}\right)$$

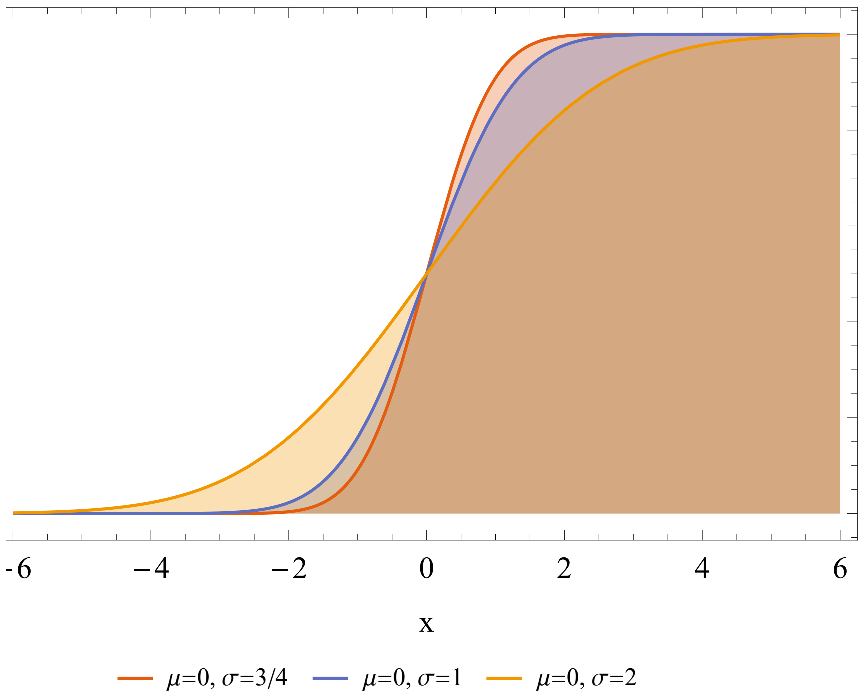
The complementary error function erfc(z) is given by erfc(z)=1-erf(z). The error function erf(z) is the integral of the (normalized) Gaussian distribution, given by $\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2}\, dt$.

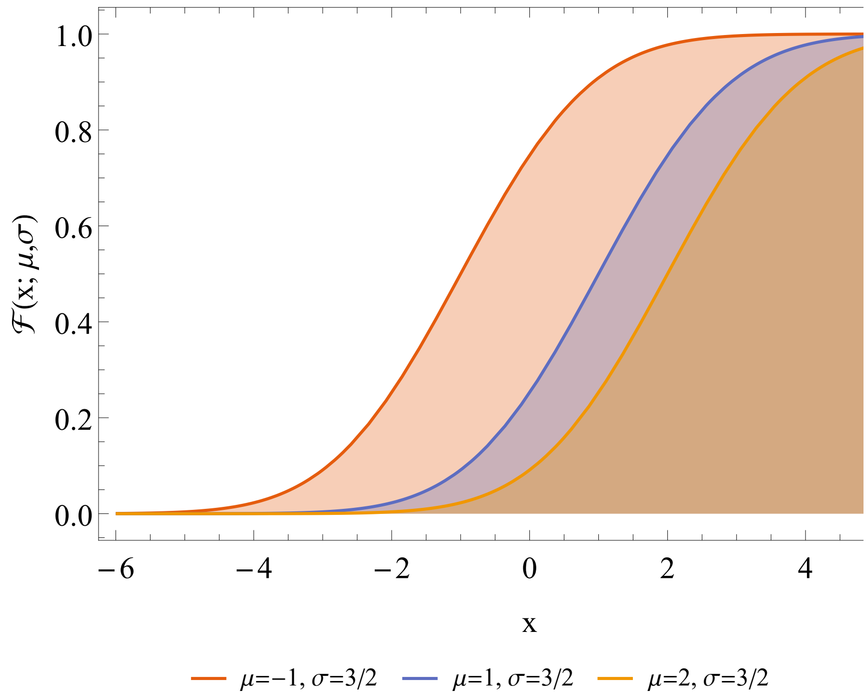
Gaussian (Normal) distribution

CDF - cumulative distribution function

Out[ ]//TraditionalForm=

$$\mathcal{F}(k;\mu,\sigma)=\mathcal{P}r(X\le k)=\int_{-\infty}^{k}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx=\frac{1}{2}\left(1+\operatorname{erf}\left(\frac{k-\mu}{\sqrt{2}\,\sigma}\right)\right)$$



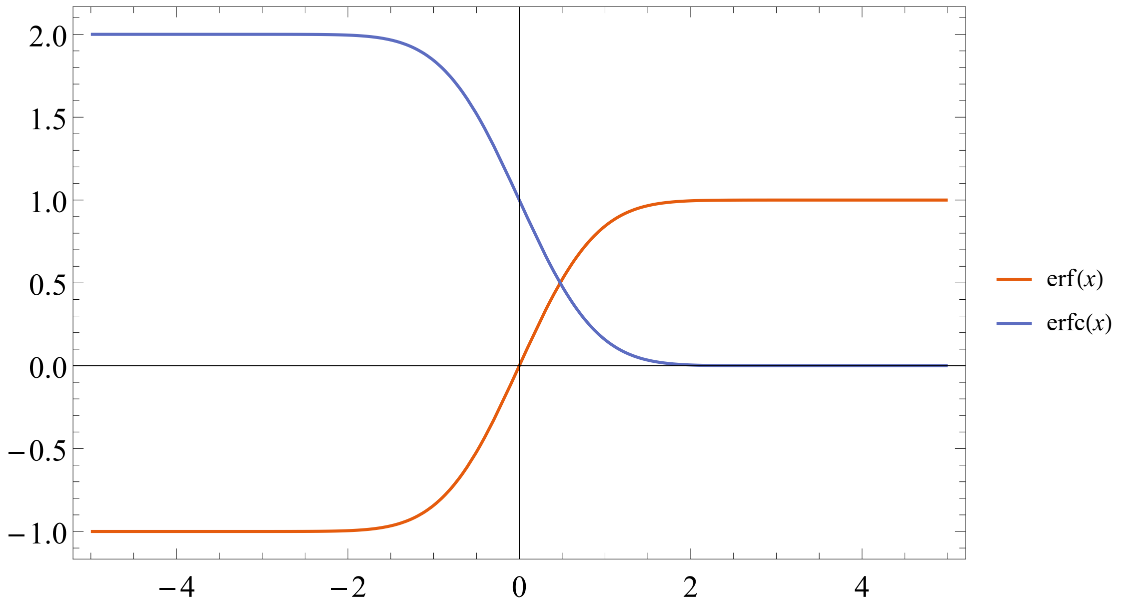


Gaussian (Normal) distribution

Error function

Out[ ]=

$$\mathrm{erf}(z) = \frac{2\int_0^z \exp(-t^2)\,dt}{\sqrt{\pi}}, \qquad \mathrm{erfc}(z) = 1 - \mathrm{erf}(z)$$

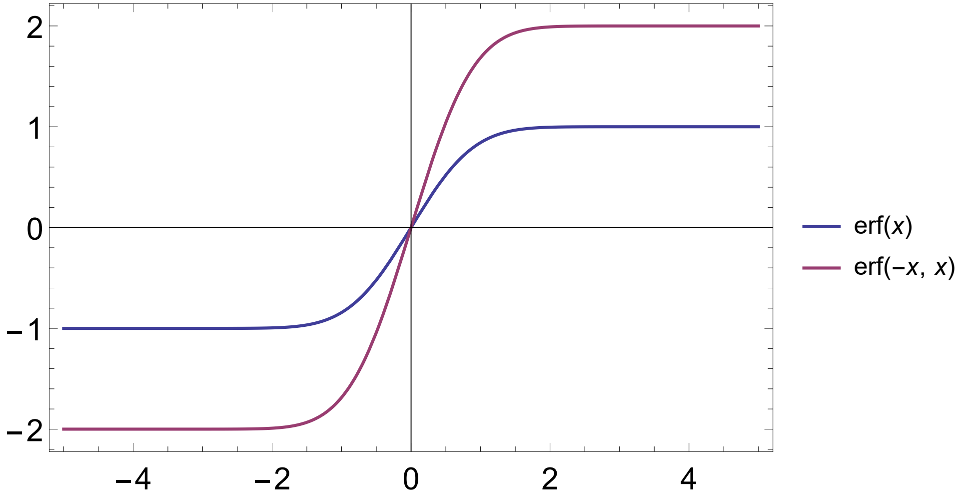


Gaussian (Normal) distribution

Error function

Be careful what definition is actually implied, sometimes erf(μ) is interpreted as $\frac{2}{\sqrt{\pi}} \int_{-\mu}^{\mu} e^{-t^2}\, dt$.

Out[ ]=



The error function expressed in terms of a series expansion:

```
Series[Erf[x], {x, 0, 10}]
```

$$\frac{2\,x}{\sqrt{\pi}} - \frac{2\,x^3}{3\,\sqrt{\pi}} + \frac{x^5}{5\,\sqrt{\pi}} - \frac{x^7}{21\,\sqrt{\pi}} + \frac{x^9}{108\,\sqrt{\pi}} + O[x]^{11}$$

Gaussian (Normal) distribution

Comparing draws from different (normal) distributions

Suppose a student sits 2 exams, getting 55 in a verbal test and 60 in a numerical reasoning test. The class scores for each exam are normally distributed. For the verbal test, the mean is 50 and standard deviation 5; for the numerical test, the mean is 50 and standard deviation is 12.
Now it is plain to see that the student did above average for each test, and did better at numerical reasoning. How did this student perform relative to everyone else? We can answer this by calculating the z-score.

Out[∘]//TraditionalForm=

$$\boxed{Z = \frac{\text{observation} - \text{mean}}{\text{SD}}}$$

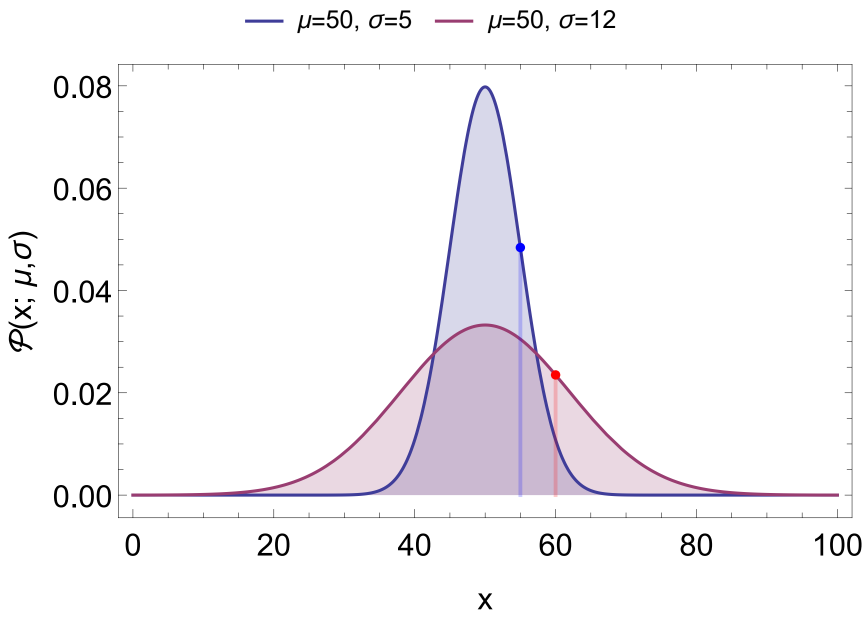
Gaussian (Normal) distribution

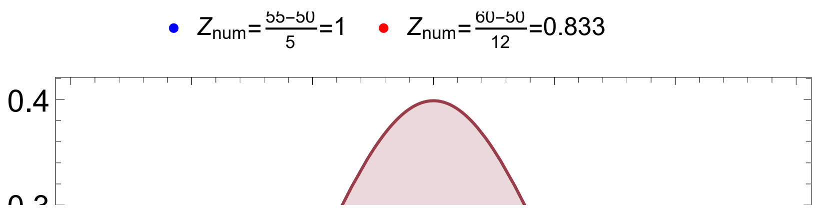
Comparing draws from different (normal) distributions

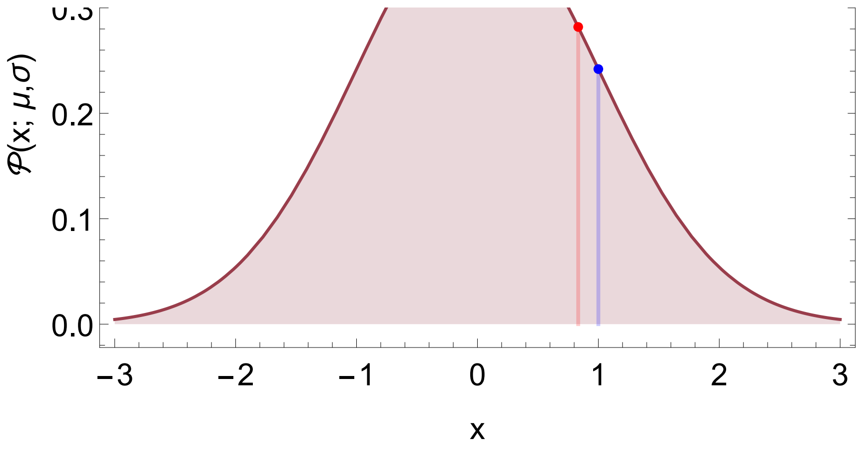
Out[ ]//TraditionalForm=

$$Z = \frac{\text{observation} - \text{mean}}{\text{SD}}$$

Out[ ]=







0.3
0.2
0.1
0.0
$\mathcal{P}(x;\,\mu,\sigma)$
−3
−2
−1
0
1
2
3
x

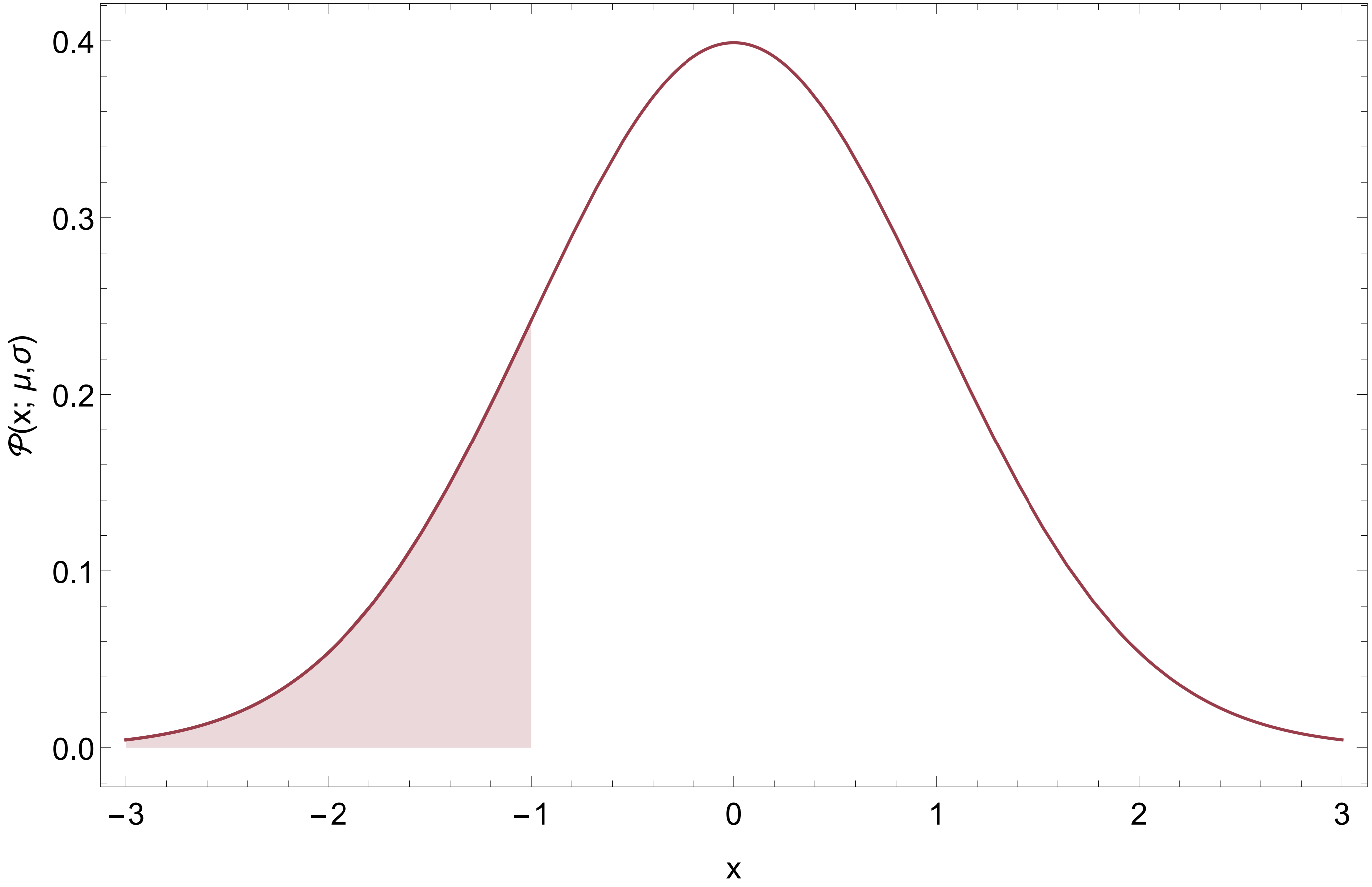
Standardization with Z scores

Out[ ]//TraditionalForm=

$$\boxed{Z = \frac{\text{observation} - \text{mean}}{\text{SD}}}$$

- standardized (Z) score of an observation is the number of standard deviations it falls above or below the mean
- Z score of mean = 0
- unusual observation: $|Z| > 2$
- defined for distributions of any shape

Standardization with Z scores

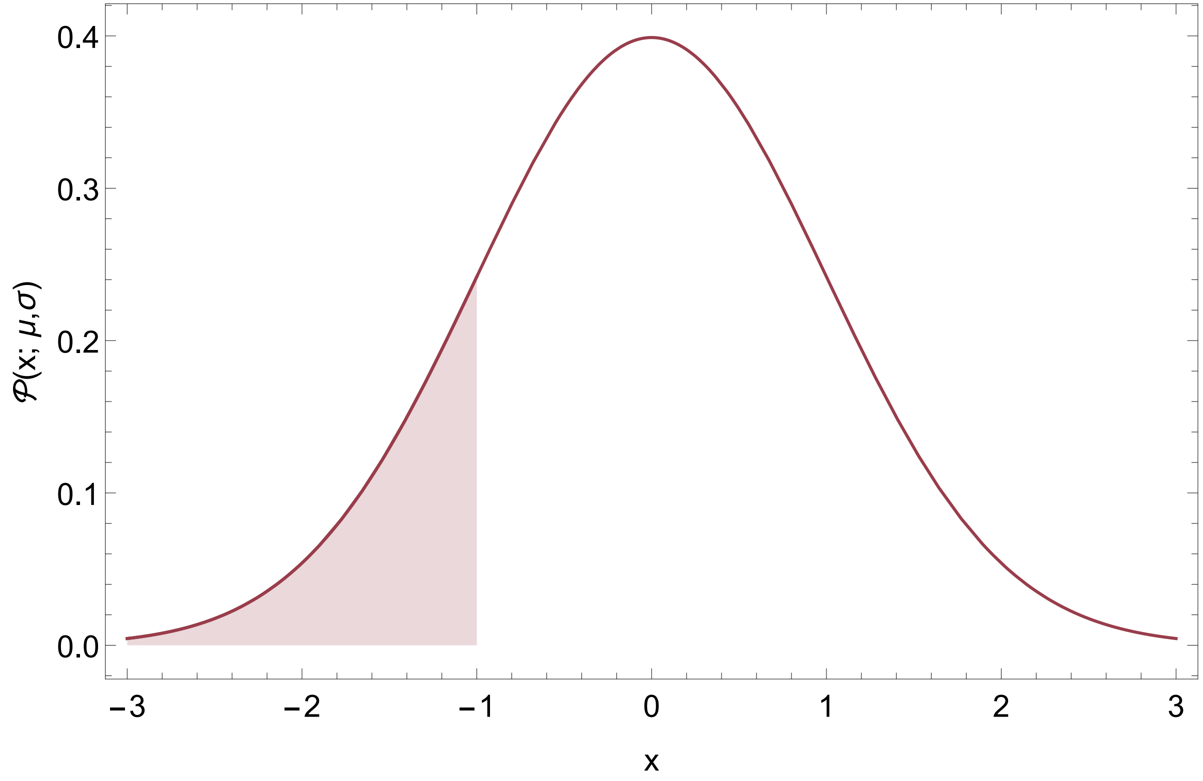


1. when the distribution is normal, Z

2. percentile is the percentage of ar
point

3. graphically, percentile is the area
left of that observation (sometimes

Standardization with Z scores



Computing Z scores with R

```
> pnorm(-1, mean = 0,
[1] 0.1586553
```

Computing Z scores with tabulated values

Second decimal place of Z					
0.04	0.03	0.02	0.01	0.00	Z
0.0003	0.0003	0.0003	0.0003	0.0003	−3.4
0.0004	0.0004	0.0005	0.0005	0.0005	−3.3
0.0006	0.0006	0.0006	0.0007	0.0007	−3.2
0.0505	0.0516	0.0526	0.0537	0.0548	−1.6
0.0618	0.0630	0.0643	0.0655	0.0668	−1.5
0.0749	0.0764	0.0778	0.0793	0.0808	−1.4
0.0901	0.0918	0.0934	0.0951	0.0968	−1.3
0.1075	0.1093	0.1112	0.1131	0.1151	−1.2
0.1271	0.1292	0.1314	0.1335	0.1357	−1.1
0.1492	0.1515	0.1539	0.1562	0.1587	−1.0

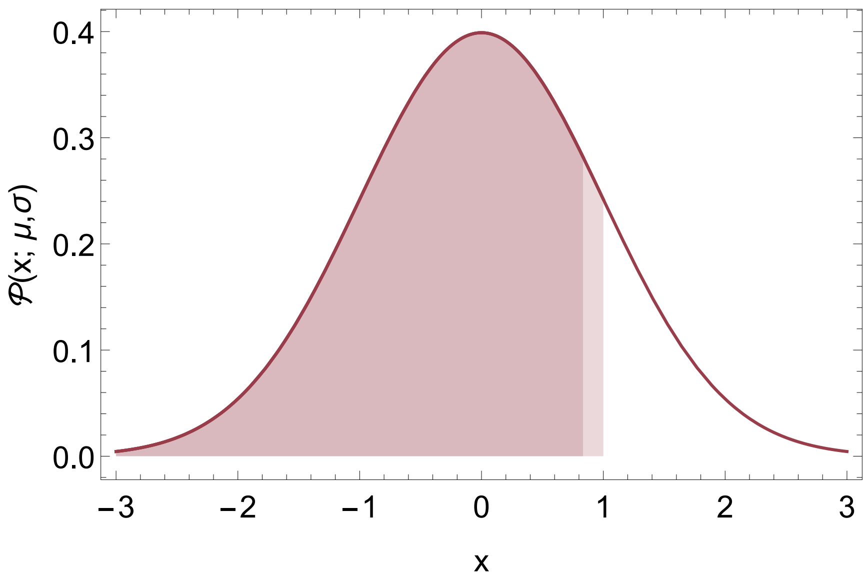
Computing Z scores with *Mathematica*

```
CDF[NormalDistribution[0, 1], -1.]
```

```
0.158655
```

Alternatively, use a web applet: e.g. http://bitly.com/dist_calc or http://vassarstats.net/tabs.html

Standardization with Z scores



Given our earlier example, the exam results expressed in percentiles are

```
CDF[NormalDistribution[0, 1], 0.833]
```

```
0.797578
```

```
CDF[NormalDistribution[0, 1], 1.]
```

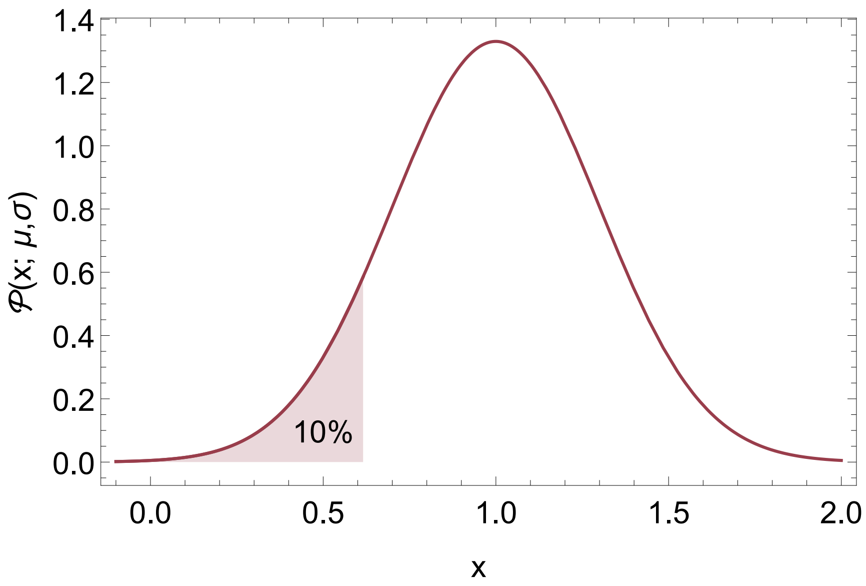
```
0.841345
```

The student performed better than 84% of all the students in the verbal exam and better than 79% in the numerical exam.

Standardization with Z scores

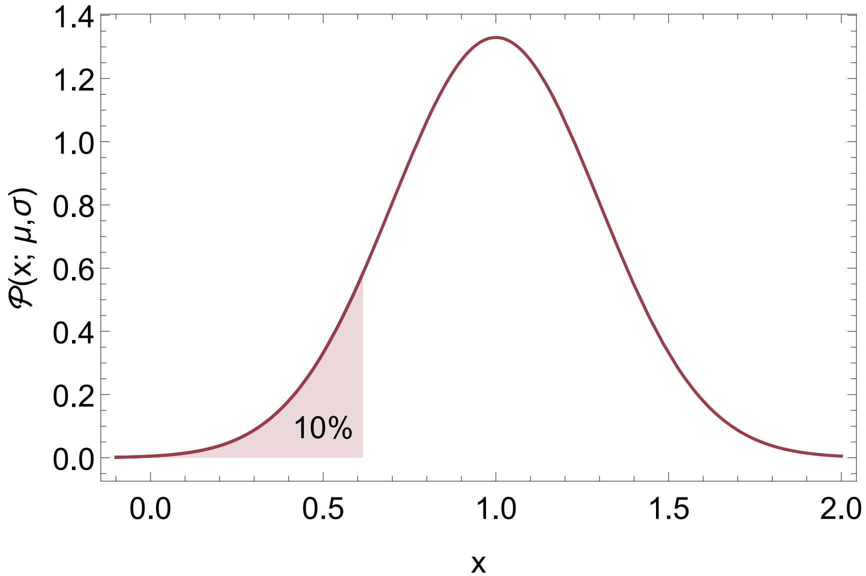
Example: Weather conditions

Radio observations require dry atmospheric conditions. Suppose the average perceptible water vapor (PWV) at an astronomical observatory site is 1 mm with a standard deviation of 0.3. Suppose the PWV statistics follow closely a Normal distribution. How dry is the atmosphere at the best 10% of observing days?



Standardization with Z scores

Example: Weather conditions



In other words, at which value of x is the c.d.f. equal to 0.1?

```
InverseCDF[NormalDistribution[1, 0.3], 0.1]
```

```
0.615535
```

10% of the time, the atmospheric conditions will be such that pwv<0.62 mm.

Example: Weather conditions

Another way to solve this problem is by using tabulated values of the normalized c.d.f.. In the table below,we show a part of such a table, where each column and row represents the CDF for one Z score (row: 1. decimal, col: 2nd decimal)

	0.	0.01	0.02	0.03	0.04
-1.6	0.0547993	0.0559174	0.0570534	0.0582076	0.0593799
-1.5	0.0668072	0.0681121	0.0694366	0.0707809	0.072145
-1.4	0.0807567	0.0822644	0.0837933	0.0853435	0.086915
-1.3	0.0968005	**0.0985253**	**0.100273**	0.102042	0.103835
-1.2	0.11507	0.117023	0.119	0.121	0.123024
-1.1	0.135666	0.137857	0.140071	0.14231	0.144572
-1.	0.158655	0.161087	0.163543	0.166023	0.168528

Since we don’t know the Z score we use the table to solve the inverse problem:

$$Z = ? = \frac{X-\mu}{\sigma} = \frac{X-1}{0.3}$$

The given CDF value is 0.1. The corresponding Z score is between -1.31 and -1.32 (highlighted values in table). Therefore

$$Z = -1.315 = \frac{X-\mu}{\sigma} = \frac{X-1}{0.3} \;\Rightarrow\; -1.315 = \frac{X-1}{0.3} \Rightarrow X = 0.6055$$

The Central Limit Theorem

The true importance of the Gaussian distribution and its dominant position in experimental science, stems from the **Central Limit Theorem**. A non-rigorous statement of this is as follows.

Form averages M_n from repeatedly drawing n samples from a population x_i with finite mean μ, variance σ^2. Then the distribution of

$$\frac{(M_n-\mu)}{\sigma/\sqrt{n}} \longrightarrow \text{Gaussian distribution}$$

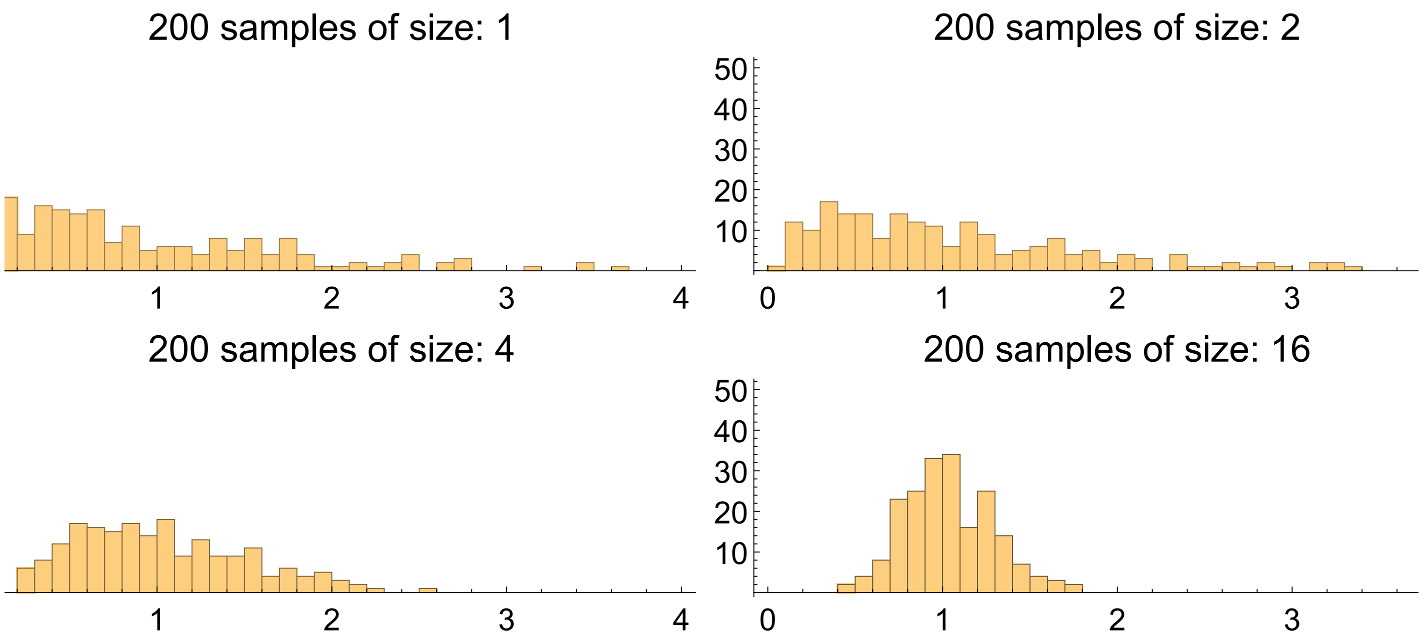
with mean 0, variance 1, as $n \to \infty$.

This is remarkable!

The Central Limit Theorem

- It says that averaging will produce a Gaussian distribution of results - no matter the shape of distribution from which the sample is drawn.
- Errors on averaged samples will always look `Gaussian'.
- The Central Limit Theorem shapes our entire view of experimentation ⇒ error language of sigmas, describing tails of Gaussian distributions.

Out[•]=

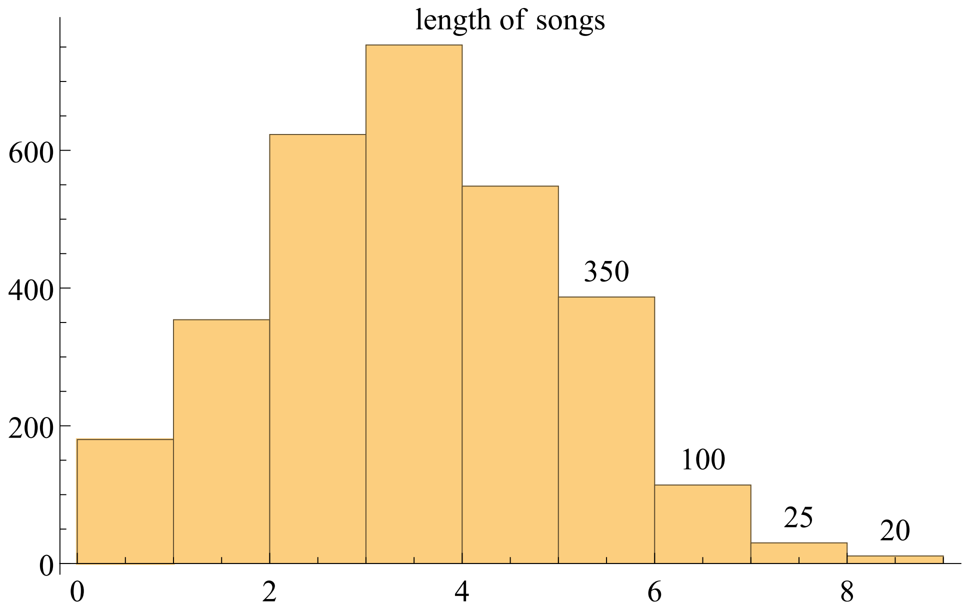


Histograms of 200 values drawn from an exponential distribution: averages of 1,2,4, 16.

Example - CLT

Credits: Dr. Mine Çetinkaya-Rundel, Duke University

Suppose my iPod has 3,000 songs. The histogram below shows the distribution of the lengths of these songs. We also know that, for this iPod, the mean length is 3.45 minutes and the standard deviation is 1.63 minutes. Calculate the probability that a randomly selected song lasts more than 5 minutes.



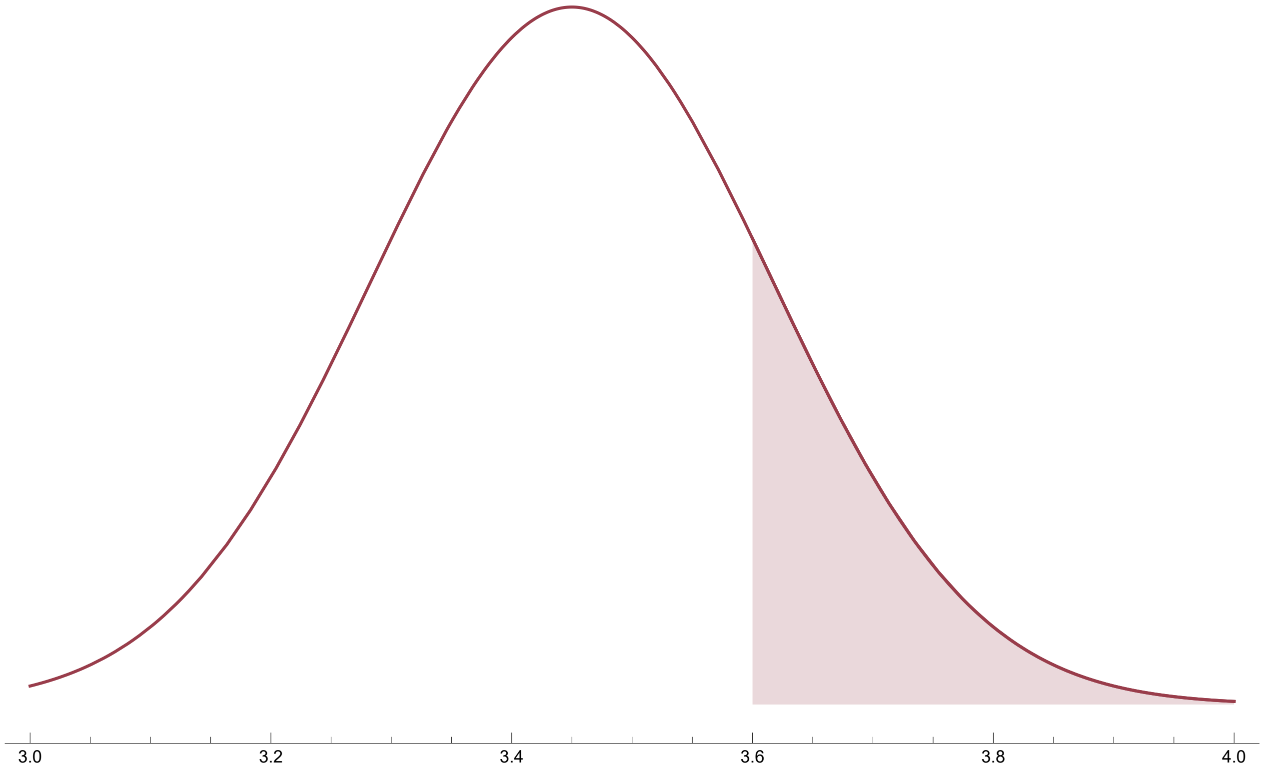
X = length of one song

$\mathcal{P}_r(X > s) = (350 + 100 + 25 + 20 + 5)/3000 = \frac{500}{3000} = 0.166667$

Example - CLT 2

Credits: Dr. Mine Çetinkaya-Rundel, Duke University

I'm about to take a trip to visit my parents and the drive is 6 hours. I make a random playlist of 100 songs. What is the probability that my playlist lasts the entire drive?



6 hours = 360 minutes

$\mathcal{P}_r(X_1 + X_2 + \ldots + X_{100} > 360 \text{ min}) = ?$

$\mathcal{P}_r(\overline{X} > 3.6 \text{ min}) = ?$

$\overline{X} \sim \mathcal{N}\left(\text{mean} = \mu = 3.45,\ \text{SE} = \frac{\sigma}{\sqrt{n}} = \frac{1.63}{\sqrt{100}} = 0.16\right.$

Z–score: $Z = \frac{3.6-3.45}{0.163} = 0.92$

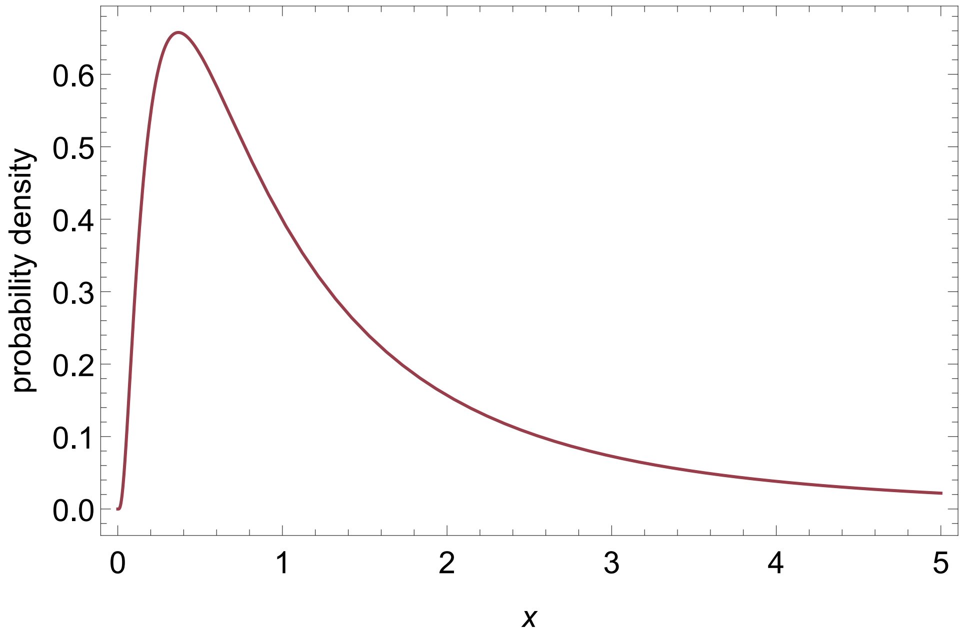
$\mathcal{P}_r(Z > 0.92) = 0.179$

Log-Normal Distribution

A log-normal distribution is a distribution where the logarithm of the random variable *x* is normally distributed:

Out[ ]//TraditionalForm=

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



Log-Normal Distribution

The PDF takes on its maximum value

Out[ ]//TraditionalForm=

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

at the position $x_{\max} = \exp(\mu - \sigma^2)$. The mean value (expectation value) and variance are:

Out[ ]//TraditionalForm=

$$\mathrm{E(X)} = \frac{1}{\sigma\,\sqrt{2\pi}} \int_0^{+\infty} \frac{x\left(\exp\left(-\frac{(\ln \mathrm{x}-\mu)^2}{2\,\sigma^2}\right)\right)}{x}\,dx = \exp\left(\mu + \frac{\sigma^2}{2}\right)$$

Out[ ]//TraditionalForm=

$$\mathrm{Var(X)} = \frac{1}{\sigma\,\sqrt{2\pi}} \int_0^{+\infty} \frac{\left(x - \mathrm{e}^{\mu+\frac{\sigma^2}{2}}\right)^2 \left(\exp\left(-\frac{(\ln \mathrm{x}-\mu)^2}{2\,\sigma^2}\right)\right)}{x}\,dx = \exp\left(2\,\mu + \sigma^2\right)\left(\mathrm{e}^{\sigma^2} - 1\right)$$

Log-Normal Distribution

Given a random variable $Y \approx \mathcal{N}(\mu, \sigma)$, then the random variable $X = e^Y$ is log-normally distributed. Given a particular target expectation value E and variance Var, it follows:

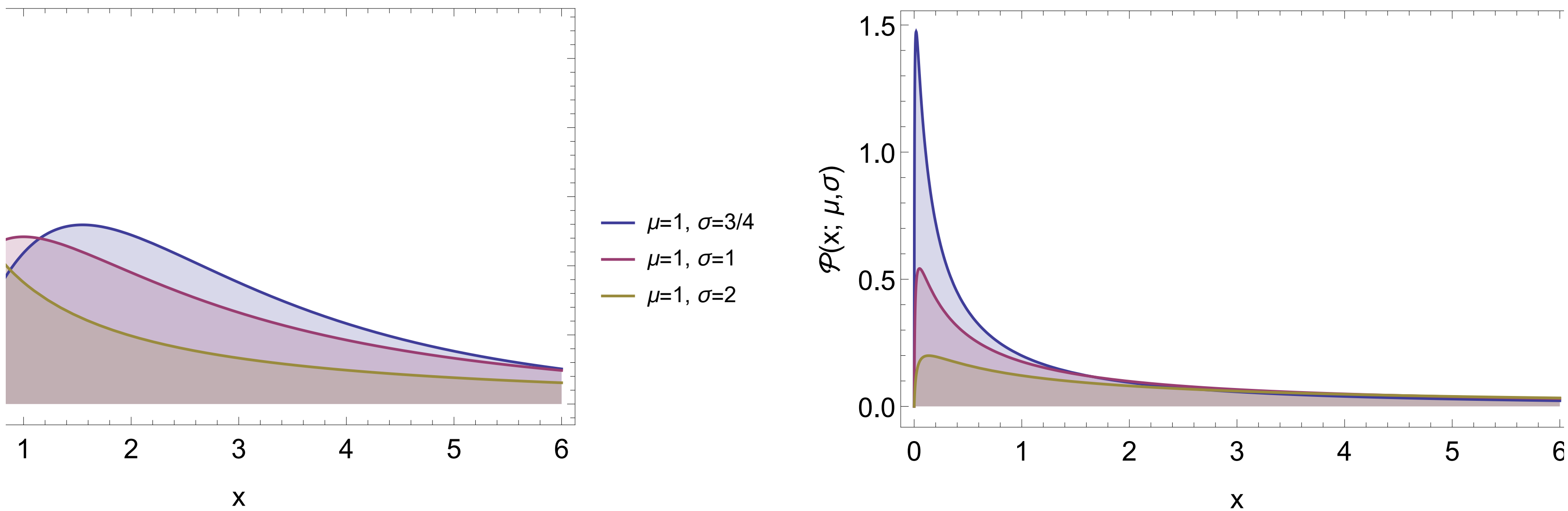
Out[ ]= $\sigma^2 = \ln\left(\frac{\text{Var}}{E^2} + 1\right) \quad \text{and} \quad \mu = \ln(E) - \frac{\sigma^2}{2} = \ln\left(E^2 \sqrt{\frac{1}{\text{Var} + E^2}}\right)$

Log-Normal Distribution

P.M.F. - probability mass function

Out[∘]//TraditionalForm=

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

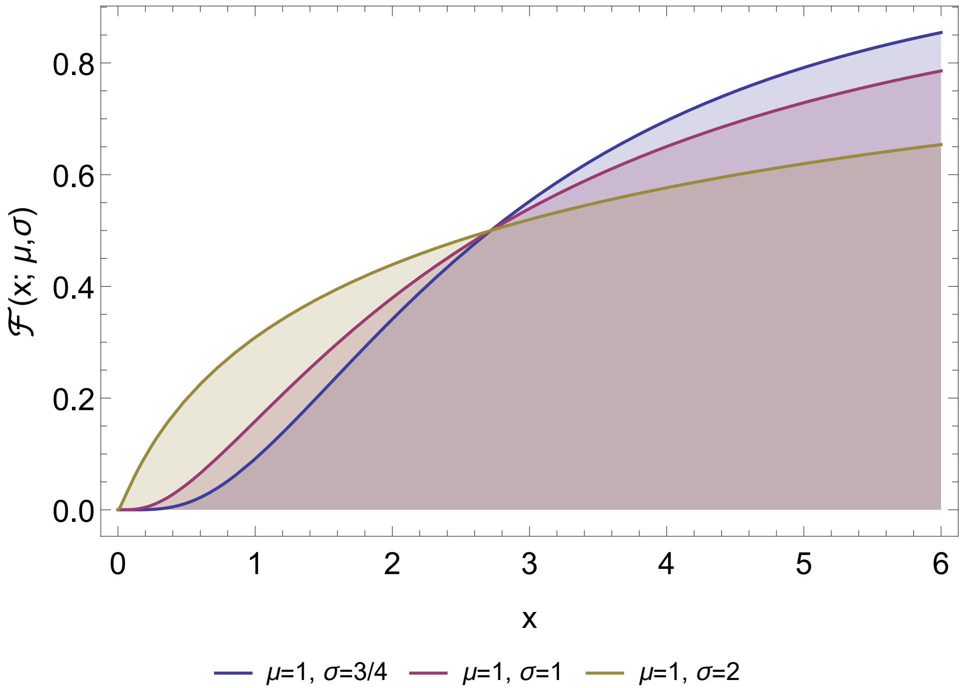


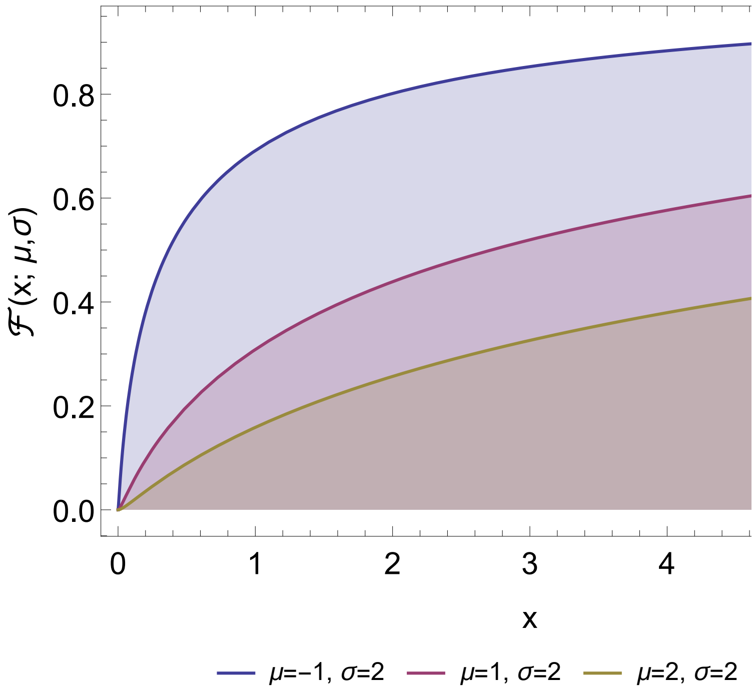
Log-Normal Distribution

CDF - cumulative distribution function

Out[∘]//TraditionalForm=

$$\mathcal{F}(k\,;\mu,\,\sigma)=\mathcal{P}(X\leq k)=\int_{-\infty}^{k}\frac{1}{x\,\sigma\,\sqrt{2\,\pi}}\exp\left(-\frac{(\ln x-\mu)^2}{2\,\sigma^2}\right)dx=\frac{1}{2}\left(1+\operatorname{erf}\left(\frac{\ln k-\mu}{\sqrt{2}\,\sigma}\right)\right)=\frac{1}{2}\left(\operatorname{erfc}\left(-\frac{\ln k-\mu}{\sqrt{2}\,\sigma}\right)\right)$$





Chebyshev's Inequality

In probability theory, Chebyshev's inequality guarantees that **in any probability distribution**, *"nearly all"* values are close to the mean — the precise statement being that **no more than $1/k^2$** of the distribution's values can be more than *k* standard deviations away from the mean (or equivalently, at least $1 - 1/k^2$ of the distribution's values are within *k* standard deviations of the mean).

Out[ ]= $\boxed{\mathcal{P}\left(\left|X-\mu\right| \geq k\,\sigma\right) \leq \frac{1}{k^2}}$ $\boxed{\mathcal{P}\left(\left|X-\mu\right| \geq k\right) \leq \frac{\sigma^2}{k^2}}$ $\boxed{\mathcal{P}\left(\left|X-\mu\right| < k\right) \geq 1-\frac{\sigma^2}{k^2}}$

Only the case $k > 1$ provides useful information. When $k < 1$ the right-hand side is greater than one, so the inequality becomes vacuous, as the probability of any event cannot be greater than one. When $k = 1$ it just says the probability is less than or equal to one, which is always true for probabilities.

Random numbers from probability distributions

Most number generators generate their output uniformly distributed (between 0 and 1). In general, we require numbers to be distributed according to a particular distribution. Let us define uniform deviates $p(r)$ drawn from a standard probability density distributions that is uniform between $r = 0$ and $r = 1$.

Out[∘]//TraditionalForm=

$$p(r) = \begin{cases} 1 & 0 \leq r < 1 \\ 0 & \text{True} \end{cases}$$

The distribution is normalized so that

Out[∘]//TraditionalForm=

$$\int_{-\infty}^{\infty} p(r)\, dr = \int_0^1 1\, dr = 1$$

Random numbers from probability distributions

We will refer to $p(r)$ as the *uniform distribution*. Suppose that we require random deviates from a different normalized probability density distribution $\mathcal{P}(x)$ which is defined to be uniform between $x = -1$ and $x = 1$; i.e. the distribution

Out[ ]//TraditionalForm=

$$\mathcal{P}(x) = \left(\begin{cases} \frac{1}{2} & -1 \leq x < 1 \\ 0 & \text{True} \end{cases} \right)$$

Random numbers from probability distributions

How to calculate x? If we choose a random deviate r between 0 and 1 from the uniform distribution, it is obvious that we can write

Out[ ]//TraditionalForm=

$$x = f(r) = 2\,r - 1$$

which will be uniformly distributed between -1 and +1. This is an example of a simple linear transformation.

Let us generalize the approach. The task is to find a general relation for obtaining a random deviate x drawn from any probability density distribution $\mathcal{P}(x)$, in terms of the random deviate r drawn from the uniform probability distribution $p(r)$.

Conservation of probability requires

Out[ ]//TraditionalForm=

$$\boxed{|p(r)\, dr| = |\mathcal{P}(x)\, dx|}$$

Random numbers from probability distributions

Therefore we can write

Out[ ]=
$$\int_{-\infty}^{r} p(r)\,dr = \int_{-\infty}^{x} \mathcal{P}(x)\,dx \quad \text{or} \quad \int_{0}^{r} 1\,dr = \int_{-\infty}^{x} \mathcal{P}(x)\,dx$$

which gives the general result

Out[ ]//TraditionalForm=
$$\boxed{r = \int_{-\infty}^{x} \mathcal{P}(x)\,dx}$$

Thus, to find x, selected randomly from the probability distribution $\mathcal{P}(x)$, we generate a random number r from the uniform distribution and find the value of the limit x that satisfies the general equation above.

Random numbers from probability distributions

Example

Consider the distribution

Out[∘]//TraditionalForm=

$$p(r) = \left(\begin{cases} A\left(1 + a\,x^2\right) & -1 \leq r < 1 \\ 0 & \text{True} \end{cases} \right)$$

where $\mathcal{P}(x)$ is positive or zero everywhere within the specified range and the normalization constant A is chosen so that

Out[∘]//TraditionalForm=

$$\int_{-1}^{1} \mathcal{P}(x)\, dx = 1$$

Random numbers from probability distributions

Example

We have

Out[ ]//TraditionalForm=

$$\boxed{r = \int_{-\infty}^{\mathtt{x}} \mathcal{P}(x)\, dx = \int_{-1}^{\mathtt{x}} A\left(1 + a\, x^2\right) dx = A\left(\mathtt{x} + \frac{a\, \mathtt{x}^3}{3} + 1 + \frac{a}{3}\right)}$$

Random numbers from probability distributions

Example

And therefore to find x we have to solve the third-degree equation

In[∘]:=
```
Solve[r == A (x + a x^3/3 + 1 + a/3), x] // FullSimplify[#, Reals] & // TraditionalForm
```

Out[∘]//TraditionalForm=

$$\left\{\left\{x \to \frac{2\sqrt[3]{2}\, a A^2 - 2^{2/3}\left(\sqrt{a^3 A^4\left((a+1)^2 (a+4) A^2 - 6 a (a+3) A r + 9 a r^2\right)} + a^2 A^2 ((a+3) A - 3 r)\right)^{2/3}}{2 a A \sqrt[3]{\sqrt{a^3 A^4\left((a+1)^2 (a+4) A^2 - 6 a (a+3) A r + 9 a r^2\right)} + a^2 A^2 ((a+3) A - 3 r)}}\right\},\right.$$

$$\left\{x \to -\frac{\sqrt[3]{-2}\, A}{\sqrt[3]{\sqrt{a^3 A^4\left((a+1)^2 (a+4) A^2 - 6 a (a+3) A r + 9 a r^2\right)} + a^2 A^2 ((a+3) A - 3 r)}} + \frac{\left(1 - i\sqrt{3}\right)\sqrt[3]{\sqrt{a^3 A^4\left((a+1)^2 (a+4) A^2 - 6 a (a+3) A r + 9 a r^2\right)} + a^2 A^2 ((a+3) A - 3 r)}}{2\sqrt[3]{2}\, a A}\right\},$$

$$\left.\left\{x \to \frac{A \boxed{-0.630\ldots + 1.09\ldots i}}{\sqrt[3]{\sqrt{a^3 A^4\left((a+1)^2 (a+4) A^2 - 6 a (a+3) A r + 9 a r^2\right)} + a^2 A^2 ((a+3) A - 3 r)}} + \frac{\left(1 + i\sqrt{3}\right)\sqrt[3]{\sqrt{a^3 A^4\left((a+1)^2 (a+4) A^2 - 6 a (a+3) A r + 9 a r^2\right)} + a^2 A^2 ((a+3) A - 3 r)}}{2\sqrt[3]{2}\, a A}\right\}\right\}$$

Random numbers from probability distributions

Example

This is called the *transformation method of generating random deviates* from probability distributions. In general, neither the integral equation, nor its solution can be obtained analytically, so numerical solutions are necessary. The following steps are required:

1. Decide on the range of x. Some probability density functions are defined in a finite range, others, such as the Gaussian extend to infinity. For numerical calculations, reasonable finite limits must be set.
2. Normalize the probability function. If it is necessary to impose limits on the range of the variable x, then the function must be *renormalized* to assure that the integral is unity over the newly defined range. *The normalization integral should be calculated numerically by the same routine that is used to find x.*
3. Generate a random variable r drawn from the uniform distribution.
4. Integrate the normalized probability function $\mathcal{P}(x)$ from negative infinity (or its defined lower limit) to the value $x = \mathrm{x}$, where x satisfies the general result from above.

Rejection Method

Often another method, the rejection method, is easier to use.

Suppose we wish to obtain random deviates between $x = -1$ and $x = +1$, drawn from the distribution function

Out[]//TraditionalForm=

$$\mathcal{P}(x) = 1 + a\,x^2$$

We begin by generating a random deviate x' uniformly distributed between -1 and +1, corresponding to the allowed range of x, and a second random deviate y' uniformly distributed between 0 and $(1 + a)$, corresponding to the allowed range of $\mathcal{P}(x)$. We can see that x' and y' must be given by

Out[]//TraditionalForm=

$$x' = -1 + 2\,r_i \qquad \text{and} \qquad y' = (1 + a)\,r_{i+1}$$

where r_i and r_{i+1} are successively generated random numbers drawn from the uniform distribution.

Rejection Method

We count an event as

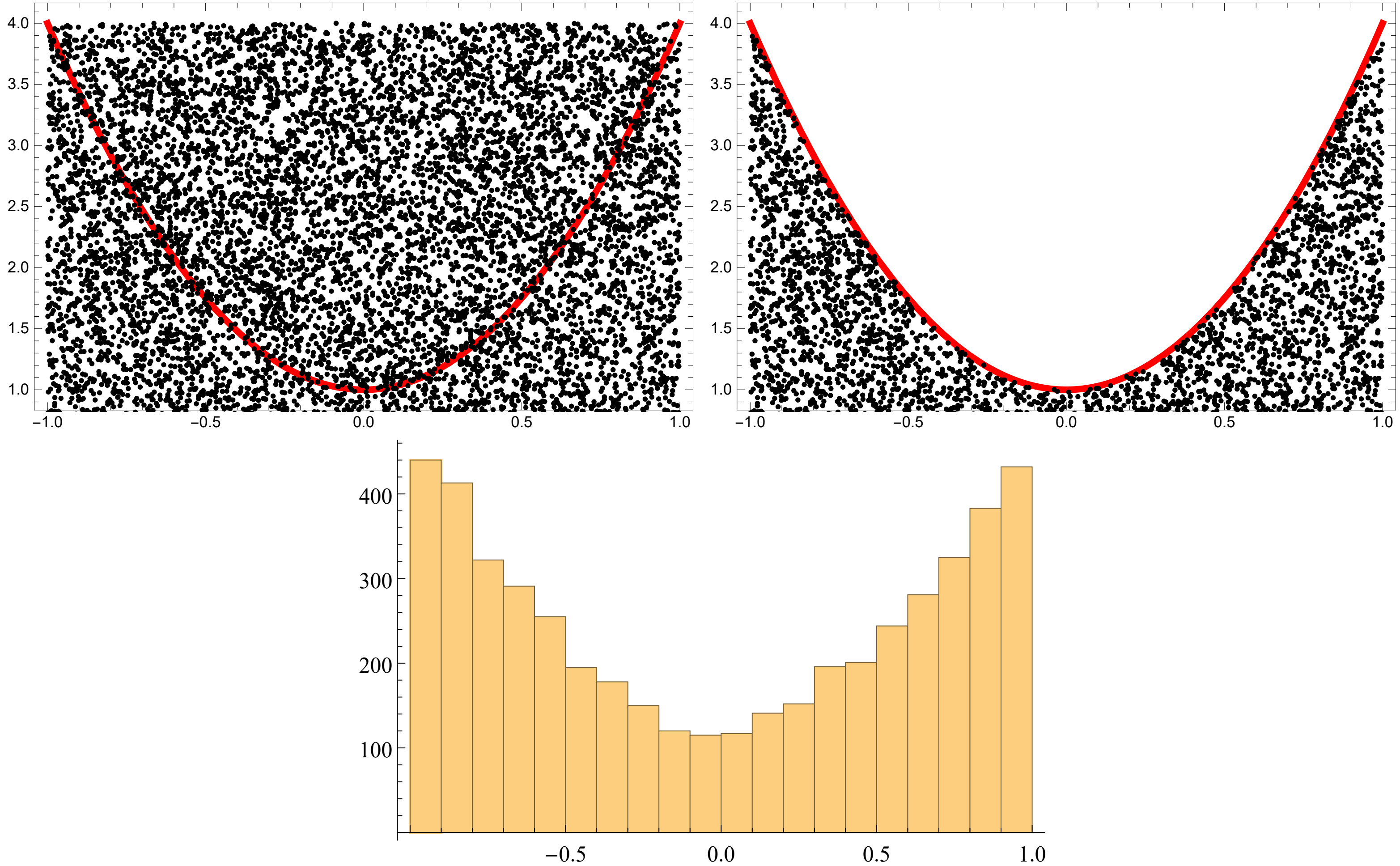
- a “hit” if the point $\{x', y'\}$ falls between the curve defined by by $\mathcal{P}(x)$ and the x axis, that is, if $y' < \mathcal{P}(x')$, and
- a “miss” if it falls above the curve.

In the limit of a large number of trials, the entire plot, including the area between the curve and the x axis will be uniformly populated by this operation and **our selected samples will be the x coordinates of the “hits”**, or the values of x' drawn randomly from the distribution $\mathcal{P}(x)$.

Note, that with this method **it is not necessary to normalize the distribution to form a true probability function**. It is sufficient that the distribution be positive and well behaved within its allowed range.

Rejection Method

Out[ ]=



The advantage of this method is its simplicity. The disadvantage is its inefficiency. In a complex Monte Carlo program only a small fraction of the events may survive, especially for higher dimensional cases.

Init

In[ ]:=
```
SetDirectory[NotebookDirectory[]];
```

In[ ]:=
```
SetOptions[$FrontEndSession, UnderoverscriptBoxOptions → {LimitsPositioning → False}]
```

In[ ]:=
```
ClearAll[equation]
Attributes[equation] = {HoldAll, HoldAllComplete};
equation[eq___] := TraditionalForm@Framed[HoldForm@Defer[eq], FrameStyle → Directive[Gray], FrameMargins → 10]
ClearAll[equationNoBox]
Attributes[equationNoBox] = {HoldAll, HoldAllComplete};
equationNoBox[eq___] := TraditionalForm[HoldForm@Defer[eq]]
```

In[ ]:=
```
equation[ℱ[k, "N", ρ] == 𝒫[X ≤ k] == Sum[Binomial["N", m] ρ^n (1 - ρ)^("N"-n), {n, 0, k}]]
```

Out[ ]//TraditionalForm=

$$\boxed{\mathcal{F}(k, \mathrm{N}, \rho) = \mathcal{P}(X \leq k) = \sum_{n=0}^{k} \binom{\mathrm{N}}{m} \rho^{n} (1-\rho)^{\mathrm{N}-n}}$$

In[ ]:=
```
equationNoBox[ℱ[k, "N", ρ] == 𝒫[X ≤ k] == Sum[Binomial["N", m] ρ^n (1 - ρ)^("N"-n), {n, 0, k}]]
```

Out[ ]//TraditionalForm=

$$\mathcal{F}(k, \mathrm{N}, \rho) = \mathcal{P}(X \leq k) = \sum_{n=0}^{k} \binom{\mathrm{N}}{m} \rho^{n} (1-\rho)^{\mathrm{N}-n}$$

In[ ]:=
```
Attributes[equationInvBox] = {HoldAll, HoldAllComplete};
equationInvBox[eq___] := TraditionalForm@Framed[HoldForm@Defer[eq], FrameStyle → Transparent, FrameMargins → 10]
```

In[∘]:=
```
Format[parens[e_]] := Style[DisplayForm@RowBox[{"(", MakeBoxes@e, ")"}], SpanMaxSize → Infinity]
Format[brackets[e_]] := Style[DisplayForm@RowBox[{"[", MakeBoxes@e, "]"}], SpanMaxSize → Infinity]
Format[parens[e_, i_]] := Style[DisplayForm@RowBox[{"(", MakeBoxes@e, ")"}], SpanMaxSize → Infinity, SpanMinSize → i]
Format[brackets[e_, i_]] := Style[DisplayForm@RowBox[{"[", MakeBoxes@e, "]"}], SpanMaxSize → Infinity, SpanMinSize → i]
 equationNoBox[Subscript[OverBar[parens[X / Y]], "geom"] == foo]
equation[Subscript[OverBar[brackets[X / Y]], "geom"] == foo]
equationNoBox[Subscript[OverBar[brackets[X / Y, 3]], "geom"] == foo]
 equation[Subscript[OverBar[parens[X / Y]], "geom"] == foo]
equation[Subscript[OverBar[brackets[X / Y]], "geom"] == foo]
equation[Subscript[OverBar[parens[X / Y, BoxMargins → {{0, 0}, {0, 0.5}}]], "geom"] == foo]
```

Out[∘]//TraditionalForm=

$$\overline{\left(\frac{X}{Y}\right)}_{\text{geom}} = \text{foo}$$

Out[∘]//TraditionalForm=

$$\boxed{\overline{\left[\frac{X}{Y}\right]}_{\text{geom}} = \text{foo}}$$

Out[∘]//TraditionalForm=

$$\overline{\left[\frac{X}{Y}\right]}_{\text{geom}} = \text{foo}$$

Out[∘]//TraditionalForm=

$$\boxed{\overline{\left(\frac{X}{Y}\right)}_{\text{geom}} = \text{foo}}$$

Out[∘]//TraditionalForm=

$$\boxed{\overline{\left[\frac{X}{Y}\right]}_{\text{geom}} = \text{foo}}$$

Out[•]//TraditionalForm=

$$\overline{\left(\frac{X}{Y}\right)}_{\text{geom}} = \text{foo}$$

Special cells

```
In[•]:= ToPictureCell[expr_] := CellPrint[ExpressionCell[expr, "Picture", ShowStringCharacters → False]]
```

```
In[•]:= PicturePrint[expr_, opts : OptionsPattern[Cell]] := CellPrint[TextCell[expr, "Picture", opts, ShowStringCharacters → False]]
```

Set options

```
In[•]:= imgsize = 500;
fontfam = "Times";
fontsize = 13;
sty = Directive[fontsize, FontFamily → fontfam];
```

```
In[•]:= SetOptions[#, BaseStyle → Thick, PlotStyle → ColorData[1, "ColorList"], ImageSize → imgsize, PlotRange → All,
      FillingStyle → Automatic, TicksStyle → Directive[fontsize, FontFamily → fontfam], Frame → True, Axes → False] & /@
    {Plot, ListPlot, LogPlot, LogLogPlot, ListLogLogPlot, LogLinearPlot, ListLogLinearPlot};
```

```
In[•]:= SetOptions[Histogram, {ImageSize → Large, BaseStyle → {Directive[fontsize, FontFamily → fontfam]}}];
```

```
In[•]:= SetOptions[{Histogram3D, DensityHistogram}, {ImageSize → Medium, BaseStyle → {Directive[fontsize, FontFamily → fontfam]}}];
```

```
In[•]:= OverviewMouseover[header_, subs_] :=
  CellPrint@Cell[BoxData[ToBoxes@Mouseover[Framed[Style[header, "OverviewSection"], FrameMargins → {{15, 15}, {5, 5}},
        BoxFrame → 2, FrameStyle → None, RoundingRadius → 5],

       Evaluate@If[subs == "", Framed[Style[header, "OverviewSection", White],
          FrameMargins → {{15, 15}, {5, 5}}, BoxFrame → 2, FrameStyle → None, RoundingRadius → 5, Background → m8red[1]],
         Row[{Framed[Style[header, "OverviewSection", White], FrameMargins → {{15, 15}, {5, 5}}, FrameStyle → None,
            RoundingRadius → 5, Background → m8red[1]], " » ", Framed[Style[subs, "OverviewSection"], FrameMargins → {{15, 15},
              {5, 5}}, FrameStyle → m8red[1], BoxFrame → 2, RoundingRadius → 5, Background → White]}]]]], "OverviewSection"]
```

```
In[]:= LargeOverviewMouseover[header_, subs_] :=
  CellPrint@Cell[BoxData[ToBoxes@Mouseover[Framed[Style[header, "LargeOverviewSection"],
        FrameMargins → {{15, 15}, {5, 5}}, BoxFrame → 2, FrameStyle → None, RoundingRadius → 5],

       Evaluate@If[subs == "", Framed[Style[header, "LargeOverviewSection", White], FrameMargins → {{15, 15}, {5, 5}}, BoxFrame → 2,
          FrameStyle → None, RoundingRadius → 5, Background → m8red[1]], Row[{Framed[Style[header, "LargeOverviewSection", White],
            FrameMargins → {{15, 15}, {5, 5}}, FrameStyle → None, RoundingRadius → 5, Background → m8red[1]], " » ",
           Framed[Style[subs, "LargeOverviewSection"], FrameMargins → {{15, 15}, {5, 5}}, FrameStyle → m8red[1],
            BoxFrame → 2, RoundingRadius → 5, Background → White]}]]]], "LargeOverviewSection"]

In[]:= MapThread[LargeOverviewMouseover, Transpose[{{"What is Mathematica?", "Uses, community, prebuilt materials"},
    {"A Look at Workflow", "syntax, palettes, free-form input"}, {"Mathematica in the Classroom",
     "teaching and learning applications"}, {"Work with Data", "built-in curated data, your own data"}}]];
```

What is *Mathematica*?

A Look at Workflow

Mathematica in the Classroom

Work with Data

```
In[ ]:= Clear[numberLine];
numberLine[ssize_] := {SeedRandom[1234];
  Block[{data = RandomReal[{-1, 1}, 20], mean, dev, sample, pos, smean, sdev},
   mean = Mean[data];
   dev = 1/Length[data] Total[(data - mean)^2];
   SeedRandom[];
   sample = RandomChoice[data, ssize];
   pos = Position[data, #] & /@ sample;
   smean = Mean[sample]; sdev = 1/Length[sample] Total[(sample - smean)^2];
   Sow[{smean, sdev}];
   Graphics[{
     PointSize[Large], Red, Point[{#, 0.05}] & /@ data,
     Black, Thick, Arrowheads[{-.03, .03}], Arrow[{{-1, 0}, {1, 0}}],
     Line[{{mean, 0.05}, {mean, -0.05}}],
     Line[{{mean + dev, 0.0}, {mean + dev, -0.05}}], Line[{{mean - dev, 0.0}, {mean - dev, -0.05}}],
     Circle[{#, 0.05}, 0.03] & /@ sample,
     Orange,
     Line[{{smean, 0.05}, {smean, -0.05}}],
     Line[{{smean + sdev, 0.0}, {smean + sdev, -0.05}}], Line[{{smean - sdev, 0.0}, {smean - sdev, -0.05}}],}, ImageSize → 700,
    PlotLabel → Row[{"μ=", mean, "   σ²=", dev}]]]}
```