Data Analysis in Astronomy and Physics

Lecture 9: Simple Error Analysis

M. Röllig

Error propagation

Suppose we wish to find the volume V of a box of length L, width W, and height H. We can measure each of the three dimensions to be L_0, W_0, and H_0 and combine these measurements to yield a value for the volume

$$V_0 = L_0\, W_0\, H_0.$$

How do the uncertainties in the estimates L_0, W_0, and H_0, affect the resulting uncertainties in the final result V_0?

If we knew the actual errors $\Delta L = L - L_0$, and so forth in each dimension, we could obtain an estimate of the error in the final result V_0 by expanding V about the point $\{L_0, W_0, H_0\}$ in a Taylor series. The first term in the Taylor expansion gives:

$$V \simeq V_0 + \Delta L\left(\frac{\delta V}{\delta L}\right)_{W_0\, H_0} + \Delta H\left(\frac{\delta V}{\delta H}\right)_{L_0\, W_0} + \Delta W\left(\frac{\delta V}{\delta W}\right)_{L_0\, H_0}$$

from which we can find $\Delta V = V - V_0$.

Error propagation

The term in the parentheses are the partial derivates of V, with respect to each of the dimensions L, W, and H, evaluated at the point $\{L_0, W_0, H_0\}$. They are the proportionality constants between changes in V and infinitesimally small changes in the corresponding dimensions.

The partial derivate of V, with respect to L, for example, is evaluated with the other variables W and H held fixed at the values W_0 and H_0 as indicated by the subscript.

This neglects higher order terms in the Taylor series. For very large errors, we need to include them.

For our example above, we find an error ΔV of

$$\Delta V \simeq \Delta L\, W_0\, H_0 + \Delta H\, L_0\, W_0 + \Delta W\, L_0\, H_0$$

which we could evaluate if we knew the uncertainties ΔL, ΔW, and ΔH.

Uncertainties

In general, we do not know the actual errors in the determination of the dependent variables. Instead we may be able to estimate the error in each measured quantity, or to estimate some characteristics, such as the standard deviation σ or the probability distribution of the measured quantities.

How can we combine the standard deviation of the individual measurements to estimate the uncertainty in the result?

Uncertainties

Suppose, we want to determine a quantity x that is a function of at least two measured variables u and v. We want to determine the characteristics of x from those of u and v and from the fundamental dependence

Out[∘]//TraditionalForm=

$$x = f(u, v, \ldots)$$

We assume (not necessarily exact) that

Out[∘]//TraditionalForm=

$$\overline{x} = f(\overline{u}, \overline{v}, \ldots)$$

The uncertainty in the resulting value for x can be found by considering the spread of the values x_i resulting from combining the individual measurements u_i, v_i, … into individual results

Out[∘]//TraditionalForm=

$$x_i = f(u_i, v_i, \ldots)$$

Uncertainties

In the limit of in infinite number of measurements, the mean of the distribution will coincide with $\overline{x}$ and we find the variance $\sigma_x{}^2$:

Out[∘]//TraditionalForm=

$$\sigma_x^2 = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{x}_i)^2$$

Similar to the example above, we can express the deviations $x_i - \overline{x}$ in terms of the deviations $u_i - \overline{u}$, $v_i - \overline{v}$, … of the observed parameters

Out[∘]//TraditionalForm=

$$x_i - \overline{x} = (u_i - \overline{u}) \frac{\partial x}{\partial u} + (v_i - \overline{v}) \frac{\partial x}{\partial v} + \ldots$$

Variance and Covariance

Combining the last two equations we get

Out[∘]//TraditionalForm=

$$\sigma_x^2 \simeq \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left((u_i - \overline{u}) \frac{\partial x}{\partial u} + (v_i - \overline{v}) \frac{\partial x}{\partial v} + \ldots \right)^2$$

Out[∘]//TraditionalForm=

$$\sigma_x^2 \simeq \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left((u_i - \overline{u})^2 \left(\frac{\partial x}{\partial u}\right)^2 + (v_i - \overline{v})^2 \left(\frac{\partial x}{\partial v}\right)^2 + 2\,(u_i - \overline{u})\,(v_i - \overline{v}) \frac{\partial x}{\partial u} \frac{\partial x}{\partial v} + \ldots \right)$$

The first two terms can be expressed in terms of variances $\sigma_u{}^2$ and $\sigma_v{}^2$

Out[∘]=

$$\boxed{\sigma_u^2 = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} (u_i - \overline{u}_i)^2} \qquad \boxed{\sigma_v^2 = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} (v_i - \overline{v}_i)^2}$$

Variance and Covariance

In order to express the third term in a similar form, we (re-)introduce the covariance $\sigma_{u\,v}{}^2$ between the variables u and v

Out[ ]//TraditionalForm=

$$\boxed{\sigma_{u\,v}^2 = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} (u_i - \overline{u}_i)(v_i - \overline{v}_i)}$$

Variance and Covariance

With these replacements the approximation for the standard deviation σ_x becomes

Out[∘]//TraditionalForm=

$$\boxed{\sigma_x^2 \simeq \sigma_u^2 \left(\frac{\partial x}{\partial u}\right)^2 + \sigma_v^2 \left(\frac{\partial x}{\partial v}\right)^2 + \ldots + 2\,\sigma_{u\,v}^2 \frac{\partial x}{\partial u} \frac{\partial x}{\partial v} + \ldots}$$

This is the **error propagation equation**.

If the fluctuations in the measured quantities u and v are uncorrelated we should expect the last term to approach zero. This is often a reasonable approximation and the error propagation equation then reduces to

Out[∘]//TraditionalForm=

$$\boxed{\sigma_x^2 \simeq \sigma_u^2 \left(\frac{\partial x}{\partial u}\right)^2 + \sigma_v^2 \left(\frac{\partial x}{\partial v}\right)^2 + \ldots}$$

Specific Error Formulas

Simple sums and differences

Given a constant parameter a, the dependent variable be

Out[∘]//TraditionalForm=

$$x = u \pm a$$

then, $\partial x / \partial u = 1$ and the (absolute) uncertainty in x is just

Out[∘]//TraditionalForm=

$$\sigma_x = \sigma_u$$

and the relative uncertainty is given by

Out[∘]//TraditionalForm=

$$\frac{\sigma_x}{x} = \frac{\sigma_u}{x} = \frac{\sigma_u}{u \pm a}$$

Example

In an experiment to count particles emitted by a decaying radioactive source, we count N_1 = 723 counts in a 15 s time interval at the beginning of the experiment and N_2 = 19 counts in a 15 s time interval later in the experiment. The events are random and obey Poisson statistics so that we know that the uncertainties in N_1 and N_2 are just their square roots. Assume that we have made a very careful measurement of the background radiation in the absence of the radioactive source and obtained a value B = 14.2 counts with negligible error for the same time interval Δt. Because we have averaged over a long time period, the mean number of background counts in the 15 s interval is not an integral number.

Example

For the first time interval, corrected number of counts is

Out[ ]//TraditionalForm=

$$x_i = \mathrm{N}_1 - B = 723 - 14.2 = 708.8 \text{ counts}$$

The uncertainty in x_1 is given by

Out[ ]//TraditionalForm=

$$\sigma_{x_1} = \sigma_{\mathrm{N}_1} = \sqrt{723} \simeq 26.9 \text{ counts}$$

And the relative uncertainty is

Out[ ]//TraditionalForm=

$$\frac{\sigma_{x_1}}{x_1} = \frac{26.9}{708.8} = 0.038$$

Example

For the second time interval we find

Out[]= $x_2 = \mathrm{N}_2 - B = 19 - 14.2 = 4.8$ counts, $\sigma_{x_1} = \sigma_{\mathrm{N}_1} = \sqrt{19} \simeq 4.4$ counts, $\dfrac{\sigma_{x_1}}{x_1} = \dfrac{4.4}{4.8} = 0.91$

Specific Error Formulas

Weighted Sums and Differences

If *x* is the weighted sum of *u* and *v*

Out[∘]//TraditionalForm=

$$x = a\,u \pm b\,v$$

The partial derivatives are simply the constants

Out[∘]=

$$\left(\frac{\partial x}{\partial u}\right) = a \qquad \left(\frac{\partial x}{\partial v}\right) = \pm b$$

And we obtain

Out[∘]//TraditionalForm=

$$\sigma_x^2 = a^2\,\sigma_u^2 + b^2\,\sigma_v^2 \pm 2\,a\,b\,\sigma_{u\,v}^2$$

Example

Suppose, that, in the previous example, the background radiation B had not been averaged over a long time period, but was simply measured for 15 s to give $B = 14$ with standard deviation $\sigma_B = \sqrt{14} = 3.7$ counts. Then the uncertainty in x would be given by

Out[ ]//TraditionalForm=

$$\sigma_x^2 = \sigma_{\mathrm{N}}^2 + \sigma_B^2 = \mathrm{N} + B$$

because the uncertainties in N and B are equal to their square roots. For the first time interval we would calculate

Out[ ]//TraditionalForm=

$$x_i = (723 - 14) \pm \sqrt{723 + 14} = 709 \pm 27.1 \text{ counts}$$

Example

And the relative uncertainty would be

Out[ ]//TraditionalForm=

$$\left(\frac{\sigma_{x_1}}{x_1} = \frac{27.1}{709}\right) \simeq 0.038$$

For the second time interval we find

Out[ ]=

$$x_2 = (19 - 14) \pm \sqrt{19 + 14} = 5 \pm 5.7 \text{ counts}, \qquad \frac{\sigma_{x_2}}{x_2} = \frac{5.7}{5} = 1.1$$

Specific Error Formulas

Multiplication and Division

If x is the weighted product of u and v

Out[ ]//TraditionalForm=

$$x = \pm a\, u\, v$$

The partial derivatives of each variable are functions of the other variable

Out[ ]=

$$\left(\frac{\partial x}{\partial u}\right) = \pm a\, v \qquad \left(\frac{\partial x}{\partial v}\right) = \pm a\, u$$

And the standard deviation becomes

Out[ ]//TraditionalForm=

$$\sigma_x^2 = (a\, v\, \sigma_u)^2 + (a\, u\, \sigma_v)^2 + 2\, a^2\, u\, v\, \sigma_{u\,v}^2$$

Out[ ]//TraditionalForm=

$$\frac{\sigma_x^2}{x^2} = \frac{\sigma_u^2}{u^2} + \frac{\sigma_v^2}{v^2} + \frac{2\, \sigma_{u\,v}^2}{u\, v}$$

Specific Error Formulas

Multiplication and Division

Similarly, if *x* is obtained through division

Out[ ]//TraditionalForm=

$$x = \pm \frac{a\,u}{v}$$

The variance for x is given by

Out[ ]//TraditionalForm=

$$\frac{\sigma_x^2}{x^2} = \frac{\sigma_u^2}{u^2} + \frac{\sigma_v^2}{v^2} - \frac{2\,\sigma_{u\,v}^2}{u\,v}$$

Example

The area of a triangle is equal to half the product of the base times the height $A = b\,h/2$. If the base and height have values $b = 5.0 \pm 0.1$ cm and $h = 10.0 \pm 0.3$ cm, the area is $A = 25.0\ \text{cm}^2$ and the uncertainty in the area is given by

Out[ ]//TraditionalForm=

$$\frac{\sigma_A^2}{A^2} = \frac{\sigma_b^2}{b^2} + \frac{\sigma_h^2}{h^2}$$

Out[ ]//TraditionalForm=

$$\sigma_A^2 = A^2\left(\frac{\sigma_b^2}{b^2} + \frac{\sigma_h^2}{h^2}\right) = 25^2\ \text{cm}^4\left(\frac{0.1^2}{5^2} + \frac{0.3^2}{10^2}\right)\left(\frac{\text{cm}^2}{\text{cm}^2}\right) = 0.81\ \text{cm}^2$$

Although the absolute uncertainty in the height is 3 times the absolute uncertainty in the base, the relative uncertainty σ_h is only 3/2 as large and its contribution to the variance of the area is only $(3/2)^2$ as large.

Specific Error Formulas

Powers

If x is obtained by raising the variable u to power

Out[ ]//TraditionalForm=

$$x = a\, u^{\pm b}$$

the derivative of x with respect to u is

Out[ ]//TraditionalForm=

$$\left(\frac{\partial x}{\partial u}\right) = \pm a\, b\, u^{\pm b-1} = \pm \frac{b\, x}{u}$$

Specific Error Formulas

Powers

and the relative error in *x* becomes

Out[ ]//TraditionalForm=

$$\frac{\sigma_x}{x} = \frac{\pm b\,\sigma_u}{u}$$

For the special case of $b = \pm 1$ we have

Out[ ]//TraditionalForm=

$$\frac{\sigma_x}{x} = \pm \frac{\sigma_u}{u}$$

The negative sign indicates that, in division, a positive error in u will produce a corresponding negative error in x.

Example

Example: The area of a circle is proportional to the square of the radius $A = \pi r^2$. If the radius is determined to be $r = 10.0 \pm 0.3$ cm, the area is $A = 100\,\pi\,\text{cm}^2$ with an uncertainty given by

Out[ ]//TraditionalForm=

$$\frac{\sigma_A}{A} = \frac{2\,\sigma_r}{r} \Rightarrow \sigma_A = \frac{2\,A\,\sigma_r}{r} = \frac{2\,\pi\left(10.\,\text{cm}^2\right)(0.3\,\text{cm})}{10.\,\text{cm}} = 6\,\pi\,\text{cm}^2$$

Specific Error Formulas

Exponentials

If x is obtained by raising the natural base to a power proportional to u

Out[∘]//TraditionalForm=

$$x = a\, e^{\pm b\, u}$$

the derivative of x with respect to u is

Out[∘]//TraditionalForm=

$$\left(\frac{\partial x}{\partial u}\right) = \pm a\, b\, e^{\pm b\, u} = \pm b\, x$$

Specific Error Formulas

Exponentials

and the relative error in x becomes

Out[∘]//TraditionalForm=

$$\frac{\sigma_x}{x} = \pm b\, \sigma_u$$

If the constant that is raised to the power is not equal to e, the expression can be rewritten as

Out[∘]//TraditionalForm=

$$x = a^{\pm b\, u} = \left(e^{\ln[a]}\right)^{\pm b\, u} = e^{\pm(b \ln(a)\, u)} = e^{\pm c\, u} \text{ with } c = b \ln(a)$$

Out[∘]//TraditionalForm=

$$\frac{\sigma_x}{x} = \pm c\, \sigma_u = \pm(b \ln[a])\, \sigma_u$$

Specific Error Formulas

Logarithm

If x is obtained by taking the logarithm of u

Out[∘]//TraditionalForm=

$$x = a \ln(b\, u)$$

The derivative with respect to u is

Out[∘]//TraditionalForm=

$$\left(\frac{\partial x}{\partial u}\right) = \frac{a}{u}$$

so

Out[∘]//TraditionalForm=

$$\sigma_x = \frac{a\, \sigma_u}{u}$$

Random Error or Systematic Error?

The average is a very common statistic; it is what we are doing all the time, for example, in `integrating' on a faint object. The variance on the the average is:

Out[ ]//TraditionalForm=

$$\boxed{S_m^2 = \mathrm{E}\left(\left(\frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}} X_i - \mu\right)^2\right)}$$

Which can be written as:

Out[ ]//TraditionalForm=

$$\boxed{S_m^2 = \frac{\sigma^2}{\mathrm{N}} + \frac{1}{\mathrm{N}^2}\sum_{i\neq j}\mathrm{E}((X_i - \mu)(X_j - \mu))}$$

Random Error or Systematic Error?

The first term expresses generally-held belief : the **error on the mean of some data decreases like $\sqrt{N}$**, as the amount of data is increased. **This is one of the most important tenets of observational astronomy**.

Out[ ]=

$$\boxed{S_m = \frac{\sigma}{\sqrt{N}} = \sqrt{\frac{\sum_{i=1}^{N} \left(X_i - \overline{X}\right)^2}{N\,(N-1)}}}$$

for vanishing covariance

Random Error or Systematic Error?

But apart from infinite variances (e.g. the Cauchy distribution), the $\sqrt{N}$ result holds only when the last term is zero. The term contains the covariance, defined as

Out[ ]//TraditionalForm=

$$\boxed{\operatorname{cov}(X_i,\ X_J) = \mathrm{E}\big((X_i - \mu_i)\,(X_j - \mu_j)\big)}$$

it is closely related to the correlation coefficient between X_i and X_j.

In the simplest cases, the data are independent and identically distributed (probability of $\mathcal{P}(X_i \text{ and } X_j) = \mathcal{P}(X_i) \times \mathcal{P}(X_j)$). ⇒ **covariance is zero.**

This is a condition (probably the likeliest) for the $\sqrt{N}$ averaging away of noise.

If so, errors are called '**random**' If not –'**systematic**'–but there's a continuum.

Combining Distributions

Often, we want to know more details of the probability distribution of a derived quantity, not only one or two measures. The simplest case is a transformation from the measured x, with probability distribution g, to some derived quantity $f(x)$ with probability distribution h. Since probability is conserved, we have the requirement (from the conservation of probability) that

Out[ ]//TraditionalForm=

$$h(f)\, df = g(x)\, dx$$

so that h involves the derivative df/dx.

Example

Suppose we are taking the logarithm of some exponentially-distributed data. Here $g(x) = \exp(-x)$ for positive x, and $f(x) = \log(x)$. Applying our rule gives

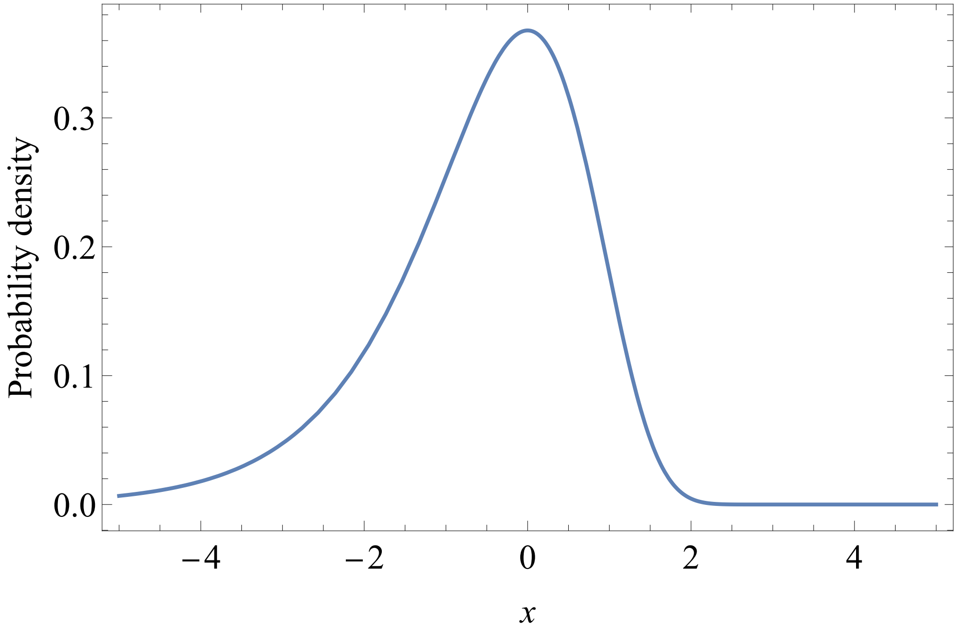
Out[∘]//TraditionalForm=

$$f(x) = \ln(x)\text{ , therefore}\quad x = \exp(f)$$

Out[∘]//TraditionalForm= $$h(f) = g(x)\,\frac{dx}{df} = g(x)\left(\frac{1}{x}\right)^{-1}$$

Out[∘]//TraditionalForm= $$\boxed{h(f) = \int \exp(-\exp(f))\,\exp(f)\,dx}$$

Out[∘]=



Example

The Figure shows a pronounced tail to negative values and is correctly normalized to unity. Our simpler methods would give us $\delta h = \delta x/x$, which cannot give a good representation of the asymmetry of h. Quoting ``h ± δh'' is clearly not very informative.

Beware if f is not monotonic! This technique rapidly becomes difficult to apply for more than one variable. Results for some useful cases:

1. Suppose we have two identically-distributed independent variables x and y, both with distribution function g. What is the distribution of their sum $z = x + y$? For each x, we have to add up the probabilities of the all the numbers $y = z - x$ that yield the z we are interested in. The probability distribution $h(z)$ is therefore

Out[ ]//TraditionalForm=

$$\boxed{h(z) = \int g(z - x)\, g(x)\, dx}$$

where the probabilities are simply multiplied because of the assumption of independence. h is the **autocorrelation** of g. The result generalizes to the sum of many variables, and is often best calculated using the Fourier transform of the distribution g. This transform is called the **characteristic function**.

Example

2. We often need the distribution of the product or quotient of two variables. Without details, the results are as follows:

   For $z = x\,y$, the distribution of z is

Out[ ]//TraditionalForm=

$$\boxed{h(z) = \int \frac{1}{|x|}\, g(x)\, g\!\left(\frac{z}{x}\right) dx}$$

   For $z = x / y$, the distribution of z is

Out[ ]//TraditionalForm=

$$\boxed{h(z) = \int |x|\, g(x)\, g(z\,x)\, dx}$$

   In almost any case of interest, these integrals are too hard to do analytically.

Example

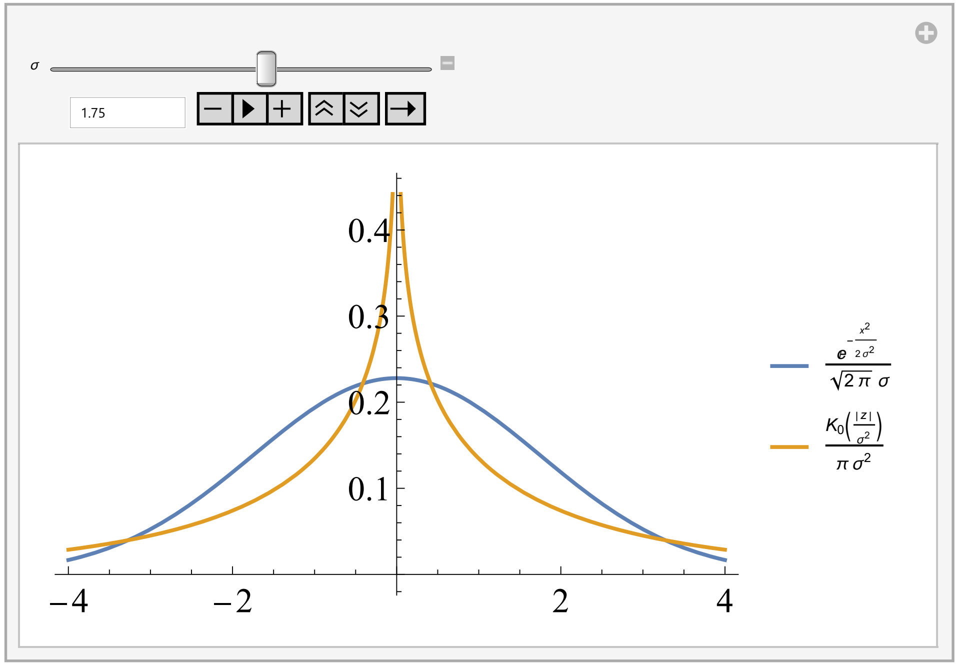
One exception is the product of two Gaussian variables of zero mean; this has applicability for radio-astronomical correlator, for instance.

Out[ ]//TraditionalForm=

$$\int \frac{1}{|x|} \mathcal{N}(0,\sigma;x)\,\mathcal{N}\left(0,\sigma;\frac{z}{x}\right) dx = \int_{-\infty}^{\infty} \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi}\,\sigma} \frac{e^{-\frac{z^2}{2x^2\sigma^2}}}{\sqrt{2\pi}\,\sigma} \frac{1}{|x|}\, dx = \frac{K_0\left(\frac{|z|}{\sigma^2}\right)}{\pi\,\sigma^2}$$

Leaving out the mathematical details, the result emerges in the form of a standard modified Bessel function. The input Gaussians are of zero mean and variance σ^2.

Example



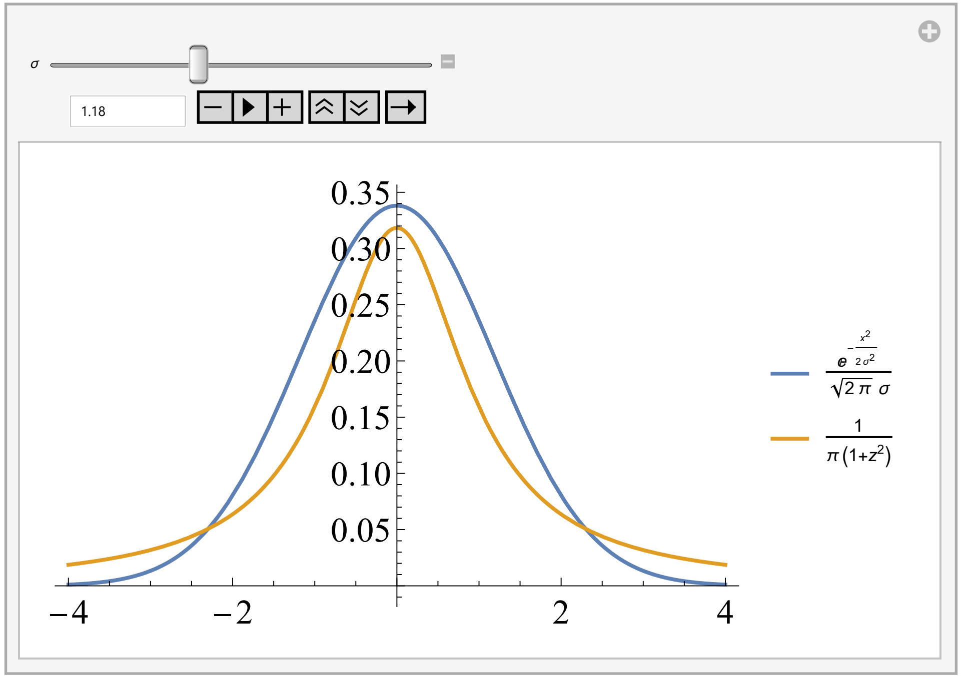
The case of the ratio is also interesting. Here we get:

Out[ ]//TraditionalForm=

$$\int |x|\, \mathcal{N}(0,\sigma;x)\, \mathcal{N}(0,\sigma;(z\,x))\, dx = \int_{-\infty}^{\infty} \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi}\,\sigma} \frac{e^{-\frac{z^2 x^2}{2\sigma^2}}}{\sqrt{2\pi}\,\sigma} |x|\, dx = \frac{1}{\pi\left(1+z^2\right)}$$

Example

This is a Cauchy distribution. It has infinite variance and, as seen from the equation, the σ from the original Gaussian does not influence the form of the resulting distribution!
The Cauchy distribution has the property that the expectation of the average of N data is again exactly the same Cauchy distribution! It does not tend to a Normal distribution!
This is a unrealistic case. It corresponds to forming the ratio of data of zero signal-to-noise ratio (same mean values!). But it illustrates, that ratios involving low signal-to-noises are likely to have very broad wings.



Some statistics and their distributions

For N data X_i, some useful statistics are the average, the sample variance, and the order statistics. If the X_i are independent and identically distributed Gaussian variables, where the original Gaussian has mean μ and variance σ^2, then:

1. the average $\overline{X}$ obeys a Gaussian distribution around μ, with variance σ^2/N.
2. the sample variance $\sigma_s{}^2$ is distributed like $\sigma^2\,\chi^2/(N-1)$, where the chi-square variable has $N-1$ degrees of freedom.
3. the ratio

   $$\frac{\sqrt{N}\,\left(\overline{X}-\mu\right)}{\sigma_s{}^2}$$

   is distributed like the t-statistic, with $N-1$ degrees of freedom. This ratio tells us how far our average might be from the true mean.
4. if we have two independent samples (sizes N and M) drawn from the same Gaussian distribution, then the ratio of sample variances $\sigma_{s_1}{}^2$ and $\sigma_{s_2}{}^2$ follows an F-distribution. This allows us to check if the data were indeed drawn from Gaussians of the same width.

Order statistics

The order statistics are simply the result of rearranging the data X_i in order of size, relabeled as $Y_1, Y_2, +\ldots$ with Y_1 the smallest value of X and Y_N the largest. Maximum values (Y_N) are often of interest, but also the median $Y_{N/2}$ (N even) is a useful robust indicator of location.

Suppose the distribution of x is $f(x)$ with cumulative distribution $\mathcal{F}(x)$. The the distribution g_n of the n-th order statistic is

Out[ ]//TraditionalForm=

$$g_n(y) = \frac{\mathrm{N}!}{(n-1)!\,(\mathrm{N}-n)!}\,\mathcal{F}(y)^{n-1}\,(1-\mathcal{F}(y))^{\mathrm{N}-n}\,f(y)$$

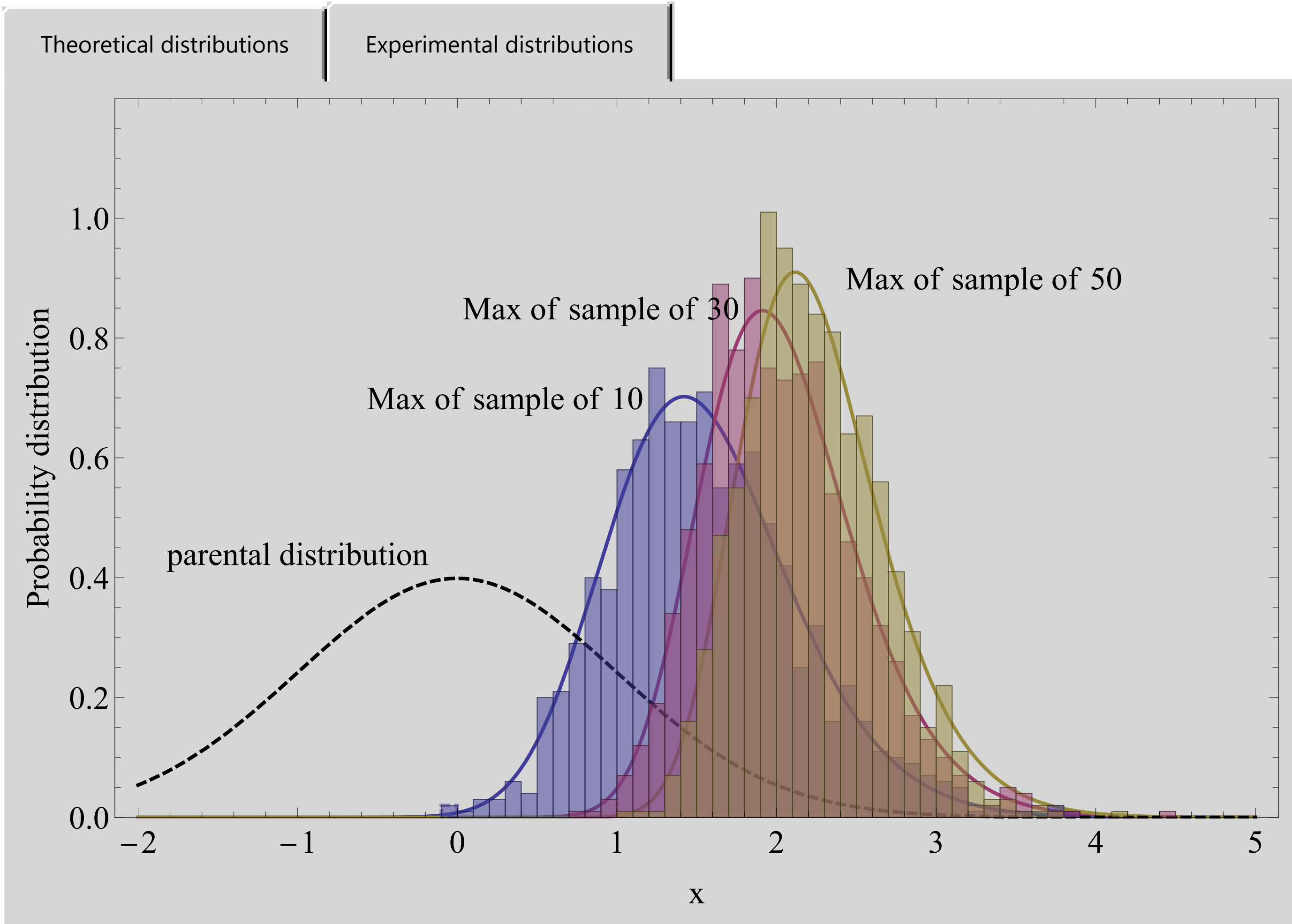
and the cumulative distribution is

Out[ ]//TraditionalForm=

$$\mathcal{G}_n(y) = \sum_{j=n}^{\mathrm{N}} \binom{\mathrm{N}}{j} \mathcal{F}(y)^j\,(1-\mathcal{F}(y))^{\mathrm{N}-j}$$

Example

Draw samples of $N = 10, 30,$ and 50 from a normally distributed population with $\mu = 0$ and $\sigma = 1$. Take the maximum value of the sample and repeat. The maximal values are distributed according to the n-th order distribution $g_n(x)$ from above.



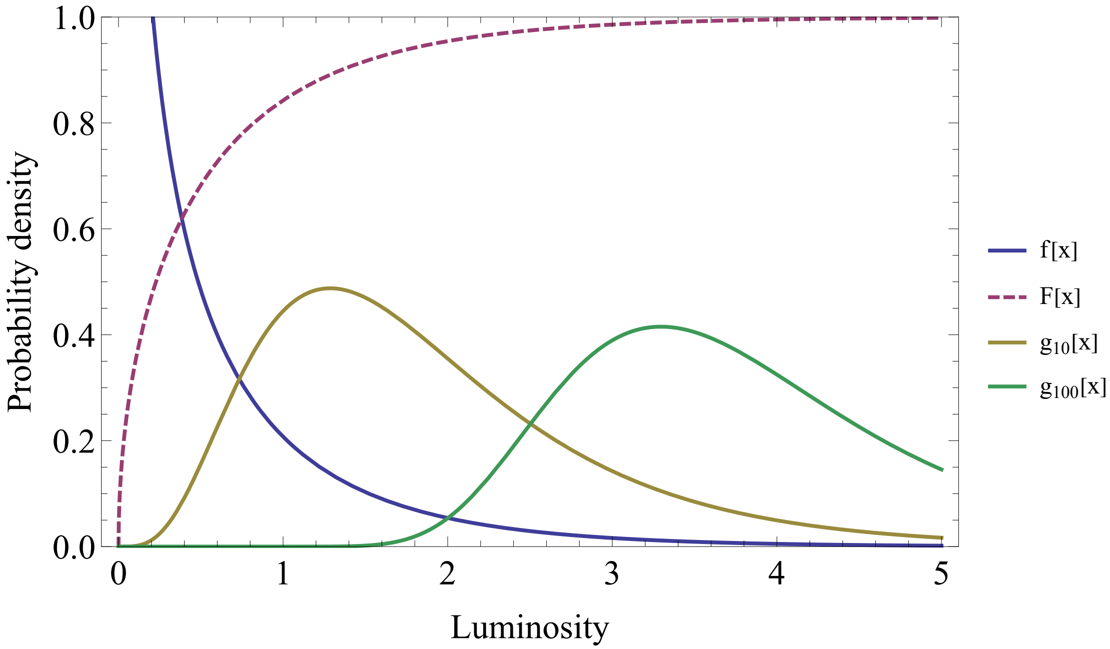
Example: Luminosity function of galaxies

The Schechter luminosity function

Out[ ]//TraditionalForm=

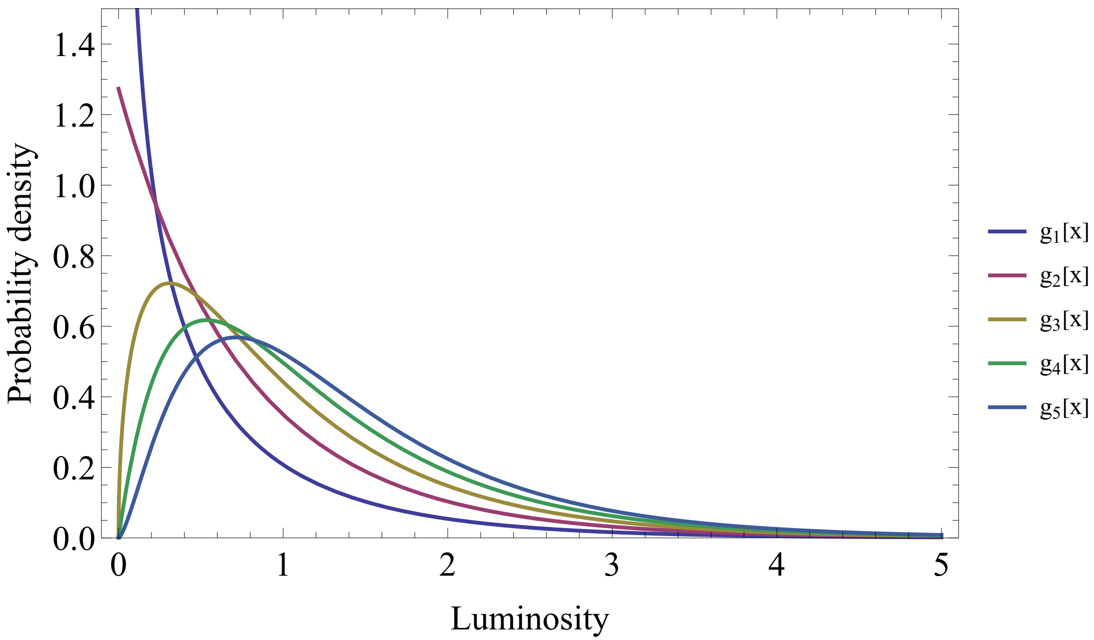
$$f(x) = \left(\frac{x^*}{x}\right)^{\gamma} \exp\left(-\frac{x}{x^*}\right)$$

is a useful model for the luminosity function of field galaxies. The observed value for γ is close to unity, but we will take $\gamma = 0.5$ (so that the distribution function can be normalized in the range 0 to infinity. We also take $x^* = 1$. If we select 10 galaxies from the distribution (we observe 10 galaxies) the maximum of the 10 will follow the 10-th order distribution $g_{10}(x)$ as given above.



Example: Luminosity function of galaxies

In the figure we see, that if we instead take the maximum of 100 galaxies, the distribution will look different. Of course, in the large sample, we will more likely find a brighter maximum than in the sample of 10 galaxies, accordingly, the distribution is shifter to higher luminosities.



Init