Physics 403

Propagation of Uncertainties

Segev BenZvi

Department of Physics and Astronomy
University of Rochester

Table of Contents

Reading

- Sivia: Ch. 3.6
- Cowan: Ch. 7.6

Maximum Likelihood and Method of Least Squares

- Suppose we measure data $\boldsymbol{x}$ and we want to find the posterior of the model parameters $\boldsymbol{\theta}$. If our priors on the parameters are uniform then

$$p(\boldsymbol{\theta}|\boldsymbol{x},I) \propto p(\boldsymbol{x}|\boldsymbol{\theta},I)\,p(\boldsymbol{\theta}|I) = p(\boldsymbol{x}|\boldsymbol{\theta},I) = \mathcal{L}(\boldsymbol{x}|\boldsymbol{\theta})$$

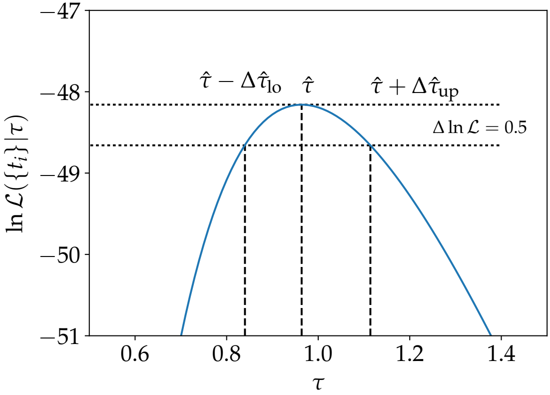
- In this case finding the best estimate $\hat{\boldsymbol{\theta}}$ is equivalent to maximizing the likelihood $\mathcal{L}$
- If $\{x_i\}$ are independent measurements with Gaussian errors then

$$p(\boldsymbol{x}|\boldsymbol{\theta},I) = \mathcal{L}(\boldsymbol{x}|\boldsymbol{\theta}) = \frac{1}{(2\pi\boldsymbol{\Sigma})^{N/2}}\exp\left(-\sum_{i=1}^{N}\frac{(f(x_i)-x_i)^2}{2\sigma_i^2}\right)$$

- Least Squares: equivalent to maximizing $\ln\mathcal{L}$, except you minimize

$$\chi^2 = \sum_{i=1}^{N}\frac{(f(x_i)-x_i)^2}{\sigma_i^2}$$

Obtaining Uncertainty Intervals from $\Delta \ln \mathcal{L}$ and $\Delta\chi^2$



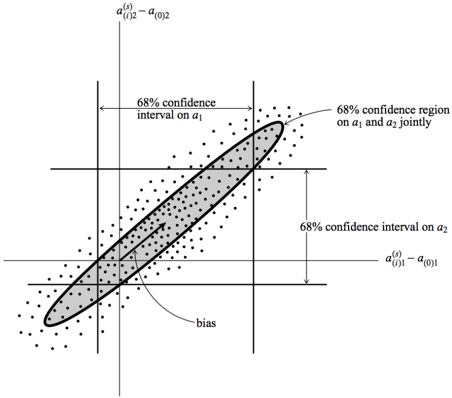
For Gaussian uncertainties we can obtain 1σ, 2σ, and 3σ intervals using the rules

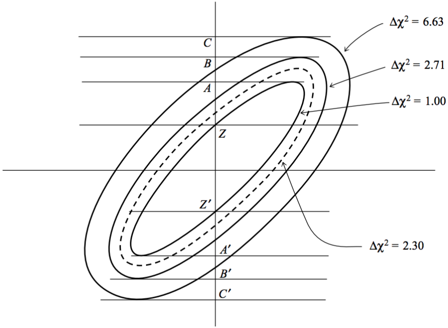
Error	$\Delta \ln \mathcal{L}$	$\Delta\chi^2$
1σ	0.5	1
2σ	2	4
3σ	4.5	9

Even without Gaussian errors this can work reasonably well. But, a safe alternative is simulation of $\ln \mathcal{L}$ with Monte Carlo

Marginal and Joint Confidence Regions

The curves $\Delta\chi^2 = 1.00, 2.71, 6.63$ project onto 1D intervals containing 68.3%, 90%, and 99% of normally distributed data





Note that it's the intervals, not the ellipses themselves, that contain 68.3%. The ellipse that contains 68% of the 2D space is $\Delta\chi^2 = 2.30$ [1]

Joint Confidence Intervals

If we want multi-dimensional error ellipses that contain 68.3%, 95.4%, and 99.7% of the data, we use these contours in $\Delta \ln \mathcal{L}$:

		joint parameters					
Range	p	1	2	3	4	5	6
1σ	68.3%	0.50	1.15	1.76	2.36	2.95	3.52
2σ	95.4%	2.00	3.09	4.01	4.85	5.65	6.4
3σ	99.7%	4.50	5.90	7.10	8.15	9.10	10.05

Or these in $\Delta\chi^2$ [1]:

		joint parameters					
Range	p	1	2	3	4	5	6
1σ	68.3%	1.00	2.30	3.53	4.72	5.89	7.04
2σ	95.4%	4.00	6.17	8.02	9.70	11.3	12.8
3σ	99.7%	9.00	11.8	14.2	16.3	18.2	20.1

Table of Contents

Propagation of Uncertainties

- We know that measurements (or fit parameters) $\boldsymbol{x}$ have uncertainties, and these uncertainties need to be propagated when you calculate functions of measured quantities $f(\boldsymbol{x})$
- From undergraduate lab courses you know the formula [2]

$$\sigma_f^2 \approx \sum_{i=1}^{N} \left(\frac{\partial f}{\partial x_i}\right)^2 \sigma_{x_i}^2$$

- **Question**: what does this formula assume about the uncertainties on $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$?
- **Question**: what does this formula assume about the PDFs of the $\{x_i\}$ (if anything)?
- **Question**: what does this formula assume about f?

Propagation of Uncertainties

- Let's start with a set of N random variables $\boldsymbol{x}$. E.g., the $\{x_i\}$ could be parameters from a fit
- We want to calculate a function $f(\boldsymbol{x})$, but suppose we don't know the PDFs of the $\{x_i\}$, just best estimates of their means $\hat{\boldsymbol{x}}$ and the covariance matrix $\boldsymbol{V}$
- Linearize the problem: expand $f(\boldsymbol{x})$ to first order about the means of the x_i:

$$f(\boldsymbol{x}) \approx f(\hat{\boldsymbol{x}}) + \sum_{i=1}^{N} \left.\frac{\partial f}{\partial x_i}\right|_{\boldsymbol{x}=\hat{\boldsymbol{x}}} (x_i - \hat{x}_i)$$

- The name of the game: calculate the expectation and variance of $f(\boldsymbol{x})$ to derive the error propagation formula. To first order,

$$\mathrm{E}\left[f(\boldsymbol{x})\right] \approx f(\hat{\boldsymbol{x}})$$

Error Propagation Formula

- Get the variance by calculating the expectation of f^2:

$$
\begin{aligned}
\mathrm{E}\left[f^2(\boldsymbol{x})\right] &\approx f^2(\hat{\boldsymbol{x}}) + 2f(\hat{\boldsymbol{x}}) \sum_{i=1}^{N} \left.\frac{\partial f}{\partial x_i}\right|_{\boldsymbol{x}=\hat{\boldsymbol{x}}} \mathrm{E}\left(x_i - \hat{x}_i\right) \\
&\quad + \mathrm{E}\left[\left(\sum_{i=1}^{N} \left.\frac{\partial f}{\partial x_i}\right|_{\boldsymbol{x}=\hat{\boldsymbol{x}}} (x_i - \hat{x}_i)\right)\left(\sum_{j=1}^{N} \left.\frac{\partial f}{\partial x_j}\right|_{\boldsymbol{x}=\hat{\boldsymbol{x}}} (x_j - \hat{x}_j)\right)\right] \\
&= f^2(\hat{\boldsymbol{x}}) + \sum_{i,j=1}^{N} \left.\frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j}\right|_{\boldsymbol{x}=\hat{\boldsymbol{x}}} V_{ij}
\end{aligned}
$$

- Since $\mathrm{var}\,(f) = \sigma_f^2 = \mathrm{E}\left(f^2\right) - \mathrm{E}\left(f\right)^2$, we find that

$$
\sigma_f^2 \approx \sum_{i,j=1}^{N} \left.\frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j}\right|_{\boldsymbol{x}=\hat{\boldsymbol{x}}} V_{ij}
$$

Error Propagation Formula

- For a set of m functions $f_1(\boldsymbol{x}), \ldots, f_m(\boldsymbol{x})$, we have a covariance matrix
$$\text{cov}\,(f_k, f_l) = U_{kl} \approx \sum_{i,j=1}^{N} \frac{\partial f_k}{\partial x_i} \frac{\partial f_l}{\partial x_j}\bigg|_{\boldsymbol{x}=\hat{\boldsymbol{x}}} V_{ij}$$
- Writing the matrix of derivatives as $A_{ij} = \partial f_i / \partial x_j$, the covariance matrix can be written
$$\boldsymbol{U} = \boldsymbol{A}\boldsymbol{V}\boldsymbol{A}^\top$$
- For uncorrelated x_i, $\boldsymbol{V}$ is diagonal and so
$$\sigma_f^2 \approx \sum_{i=1}^{N} \frac{\partial f}{\partial x_i}\bigg|_{\boldsymbol{x}=\hat{\boldsymbol{x}}} \sigma_i^2$$

This is the form you're used to from elementary courses.

Propagation of Uncertainties for Two Variables

- Let $\boldsymbol{x} = (x, y)$. The general form of σ_f^2 is

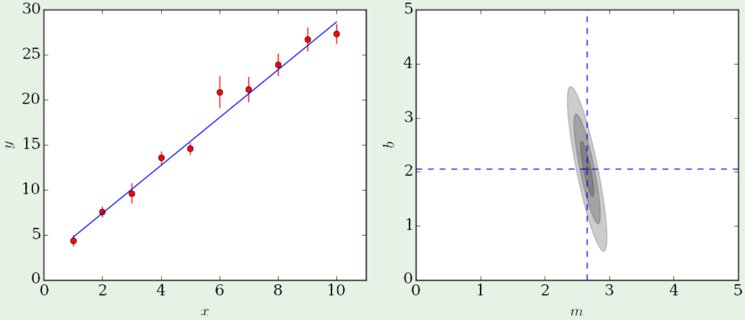
$$\sigma_f^2 = \left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2 + 2\frac{\partial f}{\partial x}\frac{\partial f}{\partial y}\rho\sigma_x\sigma_y$$

- The final cross term is often in lab courses, but it's important! Since the correlation between x and y can be negative, you can overestimate the uncertainty in f by failing to include it
- Don't forget the assumptions underlying this expression:
  1. Gaussian uncertainties with known covariance matrix
  2. f is approximately linear in the range $(x \pm \sigma_x, y \pm \sigma_y)$
- If the assumptions are violated, the error propagation formula breaks down

Interpolation of Linear Fit

Example

Example LS fit: best estimators $\hat{m} = 2.66 \pm 0.10$, $\hat{b} = 2.05 \pm 0.51$, $\text{cov}\,(m, b) = -0.10 \implies \rho = -0.94$



$y(5.5) = 16.68 \pm 0.75$ without using the correlation. With the correlation, $y(5.5) = 16.68 \pm 0.19$.

Breakdown of Error Propagation

Example

Imagine two independent variables x and y with $\hat{x} = 10 \pm 1$ and $\hat{y} = 10 \pm 1$. The variance in the ratio $f = x^2/y$ is

$$\sigma_f^2 = \left[4 \left(\frac{x}{y} \right)^2 \sigma_x^2 + \left(\frac{x}{y} \right)^4 \sigma_y^2 \right]_{x=\hat{x}}$$

For $\hat{x} = \hat{y} = 10$ and $\sigma_x^2 = \sigma_y^2 = 1$,

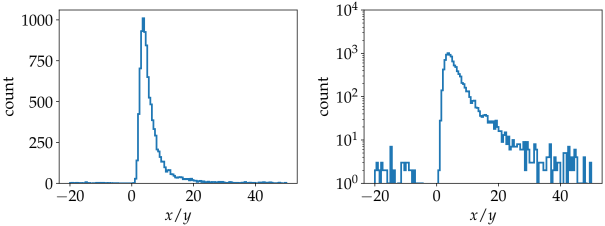
$$\sigma_f^2 = 4 \left(\frac{10}{10} \right)^2 (1)^2 + \left(\frac{10}{10} \right)^4 (1)^2 = 5$$

But, suppose $\hat{y} = 1$. Then the uncertainty blows up

$$\sigma_f^2 = 4 \left(\frac{10}{1} \right)^2 (1)^2 + \left(\frac{10}{1} \right)^4 (1)^2 = 10400$$

Breakdown of Error Propagation

- What happened? If $\hat{y} = 1$, then y can be very close to zero when $f(x, y)$ is expanded about the mean, so f can blow up and become non-linear
- **Note**: be careful even when the error propagation assumptions of small uncertainties and linearity apply; the resulting distribution could still be non-Gaussian. Example: x/y, with $\hat{x} = 5 \pm 1$ and $\hat{y} = 1 \pm 0.5$:



- In this case, reporting a central value and RMS for $f = x/y$ is clearly inadequate

Case Study: Polarization Asymmetry

Example

- Early evidence supporting the Standard Model of particle physics came from observing the difference in cross sections σ_R and σ_L for inelastic scattering of right- and left-handed polarized electrons on a deuterium target [3]
- The experiment studied the polarization asymmetry defined by

$$\alpha = \frac{\sigma_R - \sigma_L}{\sigma_R + \sigma_L}$$

- Must be careful about using the error on α to conclude whether or not α is consistent with zero
- More robust approach: check whether or not $\sigma_R - \sigma_L$ alone is consistent with zero

Averaging Correlated Measurements using Least Squares

- Imagine we have a set of measurements $x_i \pm \sigma_i$ of some "true value" λ. Since λ is the same for all measurements, we can minimize

$$\chi^2 = \sum_{i=1}^{N} \frac{(x_i - \lambda)^2}{\sigma_i^2}$$

- The LS estimator for λ is the weighted average

$$\hat{\lambda} = \frac{\sum y_i/\sigma_i^2}{\sum 1/\sigma_i^2}, \qquad \mathrm{var}\,(\hat{\lambda}) = \frac{1}{\sum 1/\sigma_i^2}$$

- For correlated measurements, we can write

$$\chi^2 = \sum_{i,j=1}^{N} (x_i - \lambda)(V^{-1})_{ij}(x_j - \lambda)$$

$$\therefore \hat{\lambda} = \sum_{i=1}^{N} w_i x_i, \qquad w_i = \frac{\sum_{j=1}^{N}(V^{-1})_{ij}}{\sum_{k,l=1}^{N}(V^{-1})_{kl}}, \qquad \mathrm{var}\,(\hat{\lambda}) = \sum_{i,j=1}^{N} w_i V_{ij} w_j$$

Example: Averaging Correlated Measurements

See Cowan Ch. 7.6.1

Example

We measure a length with two rulers made of different materials (and different coefficients of thermal expansion). Both are calibrated to be accurate at $T = T_0$ but otherwise have a temperature dependence

$$y_i = L_i + c_i(T - T_0)$$

We know the c_i and the uncertainties, T, and L_1 and L_2 from the calibration. We want to combine measurements and get $\hat{y}$. The variances and covariance are

$$\text{var}\,(y_i) = \sigma_i^2 = \sigma_{L_i}^2 + c_i^2\sigma_T^2$$

$$\text{cov}\,(y_1, y_2) = \text{E}\,(y_1 y_2) - \hat{y}^2 = c_1 c_2 \sigma_T^2$$

Solve for $\hat{y}$ with the weighted mean derived using least squares

Example: Averaging Correlated Measurements

Example

Plug in the following values: $T_0 = 25$, $T = 23 \pm 2$, and

Ruler	c_i	L_i	y_i
1	0.1	2.0 ± 0.1	1.80 ± 0.22
2	0.2	2.3 ± 0.1	1.90 ± 0.41

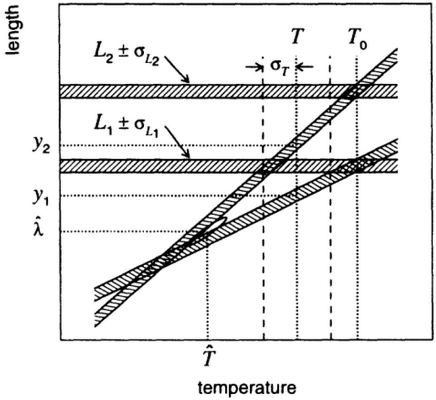
Solving, we find the weighted average is

$$\hat{y} = \frac{(\sigma_{L_2}^2 + (c_2^2 - c_1 c_2)\sigma_T^2)y_1 + \sigma_{L_1}^2 + (c_1^2 - c_1 c_2)\sigma_T^2)y_2}{\sigma_{L_1}^2 + \sigma_{L_2}^2 + (c_1 - c_2)^2\sigma_T^2} = 1.75 \pm 0.19$$

So the effect of the correlation is that the weighted average is less than either of the two individual measurements. Moreover, if $\sigma_L \to$ small and $\sigma_T \to$ large, $\sigma_y \to 0$.

Does that make sense?

Averaging Correlated Measurements



- Horizontal bands: lengths L_i from two rulers
- Slanted: lengths y_i corrected for T
- If L_1 and L_2 are known accurately, but y_1 and y_2 differ, then the true temperature must be different than the measured value of T
- The χ^2 favors reducing $\hat{y}$ until $y_1(T)$ and $y_2(T)$ intersect
- If the correction $\Delta T \gg \sigma_T$, some assumption is probably wrong. This would be reflected as a large value of χ^2 and a small p-value

Asymmetric Uncertainties

- You will often encounter published data with asymmetric error bars σ_+ and σ_-, e.g., if the author found an error interval with the maximimum likelihood method
- What do you do if you have no further information about the form of the likelihood, which is almost never published?
- Suggestion due to Barlow [4, 5]: parameterize the likelihood as

$$\ln \mathcal{L} = -\frac{1}{2}\frac{(\hat{x}-x)^2}{\sigma(x)^2}$$

where $\sigma(x) = \sigma + \sigma'(x-\hat{x})$. Requiring it to go through the $-1/2$ points gives

$$\ln \mathcal{L} = -\frac{1}{2}\left(\frac{(\hat{x}-x)(\sigma_+ + \sigma_-)}{2\sigma_+\sigma_- + (\sigma_+ - \sigma_-)(x-\hat{x})}\right)$$

- When $\sigma_+ = \sigma_-$ this reduces to an expression that gives the usual $\Delta \ln \mathcal{L} = 1/2$ rule

Table of Contents

Full Bayesian Approach

Transformation of Variables

- In the Bayesian universe, you would ideally know the complete PDF and use that to propagate uncertainties
- In this case, if we have some $p(x|I)$ and we define $y = f(x)$, then we need to map $p(x|I)$ to $p(y|I)$
- Consider a small interval δx around x' such that

$$p(x' + \delta x/2 \leq x < \delta x/2|I) \approx p(x = x'|I)\ \delta x$$

- $y = f(x)$ maps x' to $y' = f(x')$ and δx to δy. The range of y values in $y' \pm \delta y/2$ is equivalent to a variation in x between $x' \pm \delta x/2$, and so

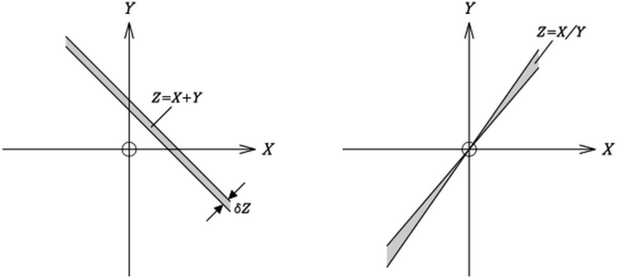
$$p(x = x'|I)\ \delta x = p(y = y'|I)\ \delta y$$

In the limit $\delta x \to 0$, this yields the PDF transformation rule

$$p(x|I) = p(y|I)\ \left|\frac{dy}{dx}\right|$$

Application to Simple Problems

- If we want to estimate a sum like $z = x + y$ or a ratio $z = x/y$, we integrate the joint PDF $p(x,y|I)$ along the shaded strips defined by $\delta(z - f(x,y))$:



- The explicit marginalization is

$$
\begin{aligned}
p(z|I) &= \iint dx\,dy\, p(z|x,y,I)\, p(x,y|I) \\
&= \iint dx\,dy\, \delta(z - f(x,y))\, p(x,y|I)
\end{aligned}
$$

Sum of Two Random Variables

- The sum $z = x + y$ requires that we marginalize

$$p(z|I) = \iint dx\, dy\, \delta(z - (x + y))\, p(x, y|I)$$

- If we are given $x = \hat{x} \pm \sigma_x$ and $y = \hat{y} \pm \sigma_y$, then we can assume x and y are independent and factor the joint PDF into separate PDFs by the product rule:

$$\begin{aligned} p(z|I) &= \int dx\, p(x|I) \int dy\, p(y|I)\, \delta(z - x - y) \\ &= \int dx\, p(x|I)\, p(y = z - x|I) \end{aligned}$$

- Assuming Gaussian PDFs for x and y,

$$p(z|I) = \frac{1}{2\pi\sigma_x\sigma_y} \int dx\; \exp\left\{-\frac{(x - \hat{x})^2}{2\sigma_x^2}\right\}\, \exp\left\{-\frac{(z - x - \hat{y})^2}{2\sigma_y^2}\right\}$$

Sum of Two Random Variables

After some rearranging of terms and changes of variables, we can express

$$p(z|I) = \frac{1}{2\pi\sigma_x\sigma_y} \int dx \, \exp\left\{-\frac{(x-\hat{x})^2}{2\sigma_x^2}\right\} \exp\left\{-\frac{(z-x-\hat{y})^2}{2\sigma_y^2}\right\}$$

as

$$p(z|I) = \frac{1}{\sqrt{2\pi}\sigma_z} \exp\left\{-\frac{(z-\hat{z})^2}{2\sigma_z^2}\right\}$$

where

$$\hat{z} = \hat{x} + \hat{y} \qquad \text{and} \qquad \sigma_z^2 = \sigma_x^2 + \sigma_y^2$$

Hence, we see how the quadrature sum rule for adding uncertainties derives directly from the assumption of Gaussian errors. Note that for a difference $z = x - y$, the uncertainties still add in quadrature but $\hat{z} = \hat{x} - \hat{y}$, as you'd expect

Case Study: Amplitude of a Bragg Peak in Crystallography

Isn't this serious overkill given that we have the error propagation formula? Unfortunately, recall that the formula can break down

Example

- In crystallography, one measures a Bragg peak $A = \hat{A} \pm \sigma_A$
- The peak is related to the structure factor $A = |F|^2$
- We want to estimate $f = |F| = \sqrt{A}$. From the propagation formula,

$$f = \sqrt{\hat{A}} \pm \frac{\sigma_A}{2\sqrt{\hat{A}}}$$

- Problem: suppose $\hat{A} < 0$, which is an allowed measurement due to reflections
- Now we're in trouble, because the error propagation formula requires us to take the square root of a negative number

Solution with Full PDF

- Let's write down the full posterior PDF

$$p(A|\{\text{data}\}, I) \propto p(\{\text{data}\}|A, I)\, p(A|I)$$

- By applying the error propagation formula, we assumed A is distributed like a Gaussian, so

$$p(\{\text{data}\}|A, I) \propto \exp\left\{-\frac{(A - \hat{A})^2}{2\sigma_A^2}\right\}$$

- Since $A < 0$ is a problem, let's define the prior to force A into a physical region:

$$p(A|I) = \begin{cases} \text{constant} & A \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

When $\hat{A} < 0$, the prior will truncate the Gaussian likelihood

Solution with Full PDF

- Truncating the PDF violates the error propagation formula, because it depends on a Taylor expansion about a central maximum
- There is no such restriction on the formal change of variables to f:

$$p(f|\{\text{data}\},I) = p(A|\{\text{data}\},I) \cdot \left|\frac{dA}{df}\right|$$

- The Jacobian is $|dA/df| = 2f$, with $f = |F| \geq 0$, so

$$p(f|\{\text{data}\},I) \propto f \cdot \exp\left\{-\frac{(f^2 - \hat{A})^2}{2\sigma_A^2}\right\} \quad \text{for } f \geq 0$$

- Find $\hat{f}$ by maximizing $\ln p$, and σ_f^2 from $\sigma_f^2 = (-\partial^2 \ln p/\partial f^2)^{-1}$:

$$2\hat{f}^2 = \hat{A} + \sqrt{\hat{A}^2 + 2\sigma_A^2}, \qquad \sigma_f^2 = \left[\frac{1}{\hat{f}^2} + \frac{2(3\hat{f}^2 - \hat{A})}{\sigma_A^2}\right]^{-1}$$

Asymptotic Agreement of PDF and Error Propagation

- When $\hat{A} > 0$ and $\hat{A} \gg \sigma_A$, the expression for f is

$$
\begin{aligned}
2\hat{f}^2 &= \hat{A} + \left(\hat{A}^2 + 2\sigma_A^2\right)^{1/2} \\
&= \hat{A}\left[1 + \left(1 + 2\left(\frac{\sigma_A}{\hat{A}}\right)^2\right)^{1/2}\right] \\
&\approx \hat{A}\left[1 + \left(1 + \left(\frac{\sigma_A}{\hat{A}}\right)^2\right)\right] \approx 2\hat{A} \\
\therefore \hat{f} &\approx \sqrt{\hat{A}}
\end{aligned}
$$

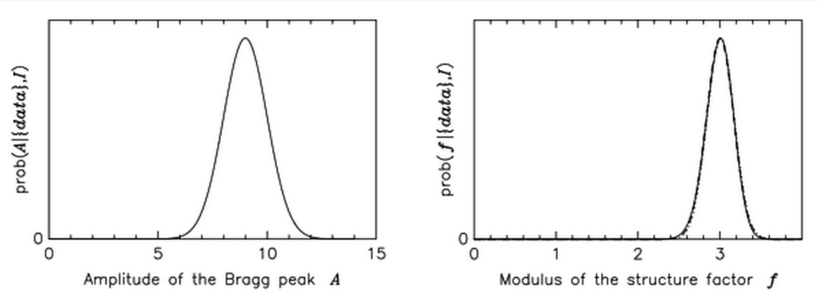
- Similarly, the expression for σ_f reduces to

$$
\sigma_f^2 = \left[\frac{1}{\hat{f}^2} + \frac{2(3\hat{f}^2 - \hat{A})}{\sigma_A^2}\right]^{-1} \rightarrow \frac{\sigma_A^2}{4\hat{A}}
$$

Asymptotic Agreement of PDF and Error Propagation

Example

For example, if $A = 9 \pm 1$, the posterior PDFs of A and f look very similar to the Gaussian PDF implied by the error propagation formula:

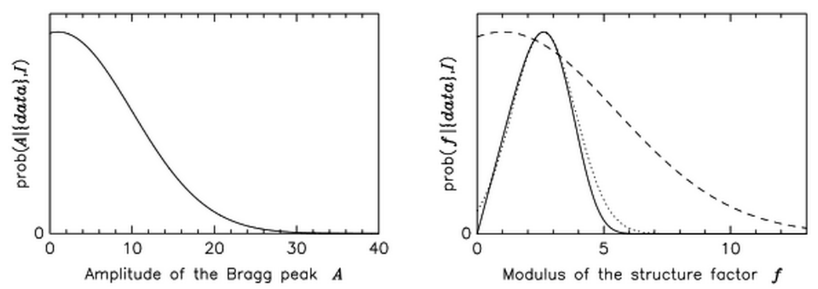


The transformed PDF is shown as a solid line, and the propagated Gaussian PDF is a dashed line.

Asymptotic Agreement of PDF and Error Propagation

Example

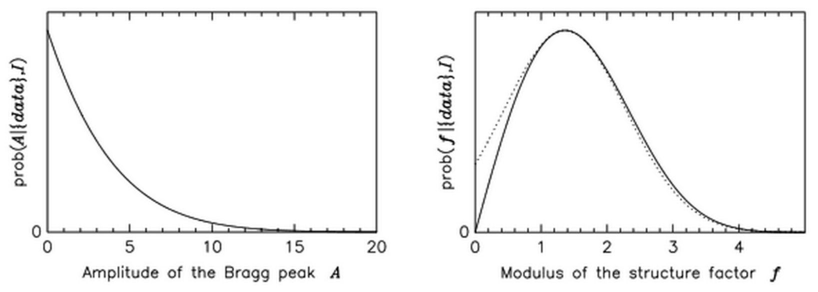
If $A = 1 \pm 9$, the error propagation formula (dashed) begins to blow up compared to the full PDF:



Asymptotic Agreement of PDF and Error Propagation

Example

If $A = -20 \pm 9$, the error propagation formula can't even be applied. The posterior PDF looks like a Rayleigh distribution:



The dotted line shows $2\hat{f}^2 = \hat{A} + \left(\hat{A}^2 + 2\sigma_A^2\right)^{1/2}$.

Summary

- The standard error propagation formula applies when uncertainties are Gaussian and $f(x)$ can be approximated by a first-order Taylor expansion (linearized)
- Most undergraduate courses emphasize only uncorrelated uncertainties, but you need to account for correlations
- Often authors will report asymmetric error bars, implying non-Gaussian uncertainties, without giving the form of the PDF. In this case there are some approximations to the likelihood that you can try to use
- Standard error propagation breaks down when the errors are asymmetric or $f(x)$ can't be linearized
- The general case is to use the full PDF to construct a new uncertainty interval on your best estimator. It's a pain (and often overkill) but it is always correct and can help you when standard error propagation fails

References I

[1] W. Press et al. *Numerical Recipes in C*. New York: Cambridge University Press, 1992. URL: http://www.nr.com.

[2] John R. Taylor. *An Introduction to Error Analysis*. Sausalito, CA: University Science Books, 1997.

[3] C.Y. Prescott et al. "Parity Nonconservation in Inelastic Electron Scattering". In: *Phys.Lett.* B77 (1978), pp. 347–352.

[4] Roger Barlow. "Asymmetric Errors". In: *PHYSTAT 2003*. SLAC, 2003, pp. 250–255.

[5] Roger Barlow. "Asymmetric Statistical Errors". In: (2004), pp. 56–59. arXiv: physics/0406120 [physics].