Analysis of Uncertainties **Fall 2014**

Lecture 1: 20 October 2014

Lecturer: Prof. Kate Scholberg *Notes by: Douglas Davis*

1.1 Introduction

Think about the value for Newton's gravitational constant G; the current accepted value in literature is something around (call it G_{m1}, this doesn't have to be correct we're just making a point).

$$G_{\mathrm{m1}} = (6.674215 \pm 0.000092) \times 10^{11}\ \mathrm{m}^3\ \mathrm{kg}^{-1}\ \mathrm{s}^{-2}.$$

Now imagine we perform some experiments where we measure G. Imagine we measure a a new value, call it G_{m2},

$$G_{\mathrm{m2}} = (6.674400 \pm 0.000010) \times 10^{11}\ \mathrm{m}^3\ \mathrm{kg}^{-1}\ \mathrm{s}^{-2}.$$

And we measure another value, G_{m3},

$$G_{\mathrm{m3}} = (6.6741 \pm 0.0221) \times 10^{11}\ \mathrm{m}^3\ \mathrm{kg}^{-1}\ \mathrm{s}^{-2}.$$

If we look closely at the values and the uncertainties, we see that the value G_{m2} is in disagreement with G_{m1}, because the values do not lie within the other value's uncertainty. This means some further investigation is needed! If we look at G_{m3}, we see that it is in agreement with both G_{m1} and G_{m1}, but the uncertainty is so large that this measurement really did nothing for the advancement of science. This can be seen graphically in Figure 1.1. The point of this example is to show that uncertainties seriously matter. When interpreting

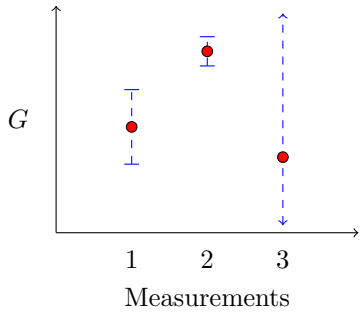

Figure 1.1: Three measurements for G, the values are red bullets, and the uncertainties are shown with blue dashed error bars

results and planning new experiments, a scientist must always be thinking about uncertainties. Why build an experiment if the expected uncertainty in the measurement will yield nothing better than what is already known? Or, what does this new measurement mean that is it so much more constrained than the previous measurement? Is this method so great we should expand it elsewhere, and why is it so much better?

1.2 Important Definitions

Our first two important definitions are *statistical* and *systematic uncertainties*. A *statistical uncertainty* is a random uncertainty that arises simply from random fluctuations in the data. As the sample size of the data approaches infinity ($N \to \infty$), the statistical uncertainty approaches 0. A *systematic uncertainty* comes from a lack of knowledge about the experimental apparatus/measurement. An example would be the possibly of an inexact measurement from an imperfect ruler. Systematic uncertainties can also have some randomness, if an apparatus is affected by temperature change, the random fluctuations of temperature in the room cause systematic uncertainties because of the effect on the apparatus. To quote the systematic and statistical uncertainties, results are often written in the form:

$$x \pm \Delta x\,(\text{sys.}) \pm \Delta x\,(\text{stat.}), \tag{1.1}$$

with the two values separate. As the amount of data increases, the statistical uncertainty always decreases. The same cannot be said for systematic. In theory, studying systematic uncertainties more thoroughly may lead to a decrease, but it is possible that a better understanding of the systematics may cause an increase in the uncertainty. Our next definition is a *sample*. The *sample* is the collection of measurements. If we want to measure some value x, then the sample is simply:

$$x_1,\, x_2,\, x_3,\, \ldots,\, x_N,$$

where x_i is just the ith measurement. An infinite number of measurements would yield the sample's *parent distribution*, which we denote with $P(x)$. Figure 1.2 shows an example of a histogrammed sample with it's parent distribution. As $N \to \infty$, the shape of the histogram would approach the shape of $P(x)$.

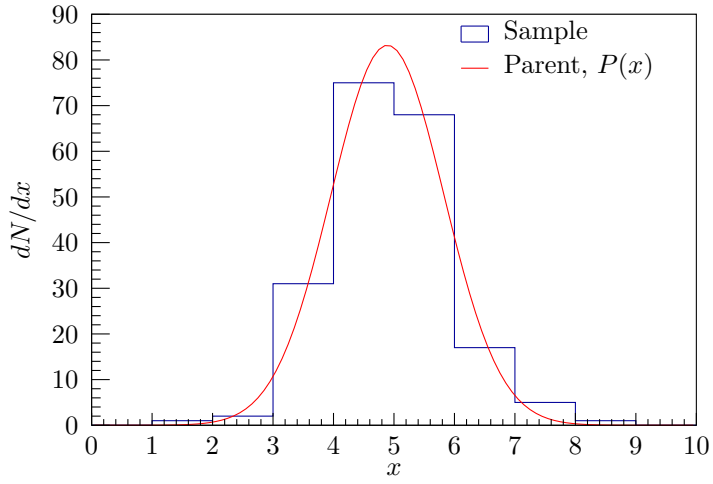

Figure 1.2: A sample distribution in histogram form overlayed by its parent distribution.

Now we will define some characteristics of the sample and the parent distribution. First, the *mean of the sample*, which we denote with $\overline{x}$:

$$\overline{x} = \frac{1}{N}\sum_{i}^{N} x_i. \tag{1.2}$$

The *mean of the parent distribution*, denoted by μ:

$$\mu = \lim_{N\to\infty} \frac{1}{N} \sum_{i}^{N} x_i. \tag{1.3}$$

To quantify the spread we may first think to use the *deviation*, denoted by d,

$$d_i = x_i - \mu, \tag{1.4}$$

and take the limit:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i}^{N} d_i = \lim_{N\to\infty} \frac{1}{N} \sum_{i}^{N} (x_i - \mu), \tag{1.5}$$

but because d_i can be positive or negative, this limit tends to zero. We may also think to take the absolute value of the deviation, $|d_i|$, but this math turns out to be pretty bad. Instead we go with something called the *variance*, which we denote with σ^2:

$$\sigma^2 = \lim_{N\to\infty} \frac{1}{N} \sum_{i}^{N} (x_i - \mu)^2. \tag{1.6}$$

Which can also be written as:

$$\sigma^2 = \lim_{N\to\infty} \left[\frac{1}{N} \sum_{i}^{N} x_i^2 \right] - \mu^2 \tag{1.7}$$

In real world experiments, we probably do not know the parent mean, so we take $\mu \to \overline{x}$. Now we define the *sample variance*, denoted by s^2:

$$s^2 = \frac{1}{N-1} \sum_{i}^{N} (x_i - \overline{x})^2 \tag{1.8}$$

Notice the difference between the fraction in equation 1.6 and equation 1.8, we change from $1/N$ to $1/(N+1)$. The reason for this can be worked out mathematically (ref), but we will just explain the following reason. Imagine taking only 1 measurement ($N = 1$). Equation 1.6 would yield a variance of only 1. This does not make sense for a single measurement, the outcome using equation 1.8 would be an undefined variance, which does make sense. There is no variance in a single measurement. The sample variance defined in equation 1.8 is known as an *unbiased estimator*. Closing side note: for practical purposes many scientists use the notation of σ^2 for the sample variance and μ for the sample mean.

1.3 Propagation of uncertainties

Let's suppose the value we are attempting to measure, x, is a function of u:

$$x = f(u) = u^2 \tag{1.9}$$

We know that our measurement, x, will have some uncertainty (which we will denote as δx or σ_x), and our independent variable must also have some uncertainty: δu, σ_u. We can (to first order, assuming δx is small) write:

$$\delta x = f(u + \delta u) - f(u). \tag{1.10}$$

We can then write:

$$\frac{\delta x}{\delta u} = \frac{f(u+\delta u) - f(u)}{\delta u} = \frac{df}{du} \rightarrow \delta x = \left|\frac{df}{du}\right| \delta u, \tag{1.11}$$

or:

$$\sigma_x = \left|\frac{df}{du}\right| \sigma_u \tag{1.12}$$

Therefore, the measurement related to equation 1.9 would be of the form:

$$\text{measured value} = x \pm |2u|\sigma_u \tag{1.13}$$

An example of this would be the area of a square with side length u, and an uncertainty in the ruler used to measure the side which is σ_u.

Now let's expand this to an arbitrary number of independent variables. Let's take x in the form of f again:

$$x = f(u, v, w, ...), \tag{1.14}$$

it follows that:

$$\overline{x} = f(\overline{u}, \overline{v}, \overline{w}, ...) \qquad x_i = f(u_i, v_i, w_i, ...) \tag{1.15}$$

We can write:

$$x_i - \overline{x} \cong (u_i - \overline{u}) \left(\frac{\partial x}{\partial u}\right)_{v,w,...} + (v_i - \overline{v}) \left(\frac{\partial x}{\partial v}\right)_{u,w,...} + ... \tag{1.16}$$

Now recall equation 1.6, we can write:

$$\sigma_x^2 = \lim_{N\to\infty} \frac{1}{N} \sum_i^N \Bigg[(u_i - \overline{u})^2 \left(\frac{\partial x}{\partial u}\right)^2 + (v_i - \overline{v})^2 \left(\frac{\partial x}{\partial v}\right)^2 + ... + \underbrace{2(u_i - \overline{u})(v_i - \overline{v}) \left(\frac{\partial x}{\partial u}\right) \left(\frac{\partial x}{\partial v}\right) + ...}_{\text{cross terms from squaring eq. 1.16}} \Bigg] \tag{1.17}$$

Now using our standard variance equation with the form:

$$\sigma_u^2 = \lim_{N\to\infty} \frac{1}{N} \sum_i^N (u_i - \overline{u})^2, \tag{1.18}$$

and now introducing the covariance between u and v:

$$\sigma_{uv}^2 = \lim_{N\to\infty} \frac{1}{N} \sum_i^N (u_i - \overline{u})(v_i - \overline{v}). \tag{1.19}$$

We can now rewrite equation 1.17 in the form:

$$\sigma_x^2 \cong \sigma_u^2 \left(\frac{\partial x}{\partial u}\right) + \sigma_v^2 \left(\frac{\partial x}{\partial v}\right) + ... + 2\sigma_{uv}^2 \left(\frac{\partial x}{\partial u}\right) \left(\frac{\partial x}{\partial v}\right) + ... \tag{1.20}$$

If u and v are independent of eachother, in the large N limit, $\sigma_{uv}^2 \rightarrow 0$ and all of the "cross-terms" vanish, yielding a variance of:

$$\sigma_x^2 \cong \sigma_u^2 \left(\frac{\partial x}{\partial u}\right) + \sigma_v^2 \left(\frac{\partial x}{\partial v}\right) + ... \tag{1.21}$$

Analysis of Uncertainties **Fall 2014**

Lecture 2: 24 October 2014

Lecturer: Prof. Kate Scholberg *Notes by: Douglas Davis*

2.1 Probability

We define probability with

$$P = \frac{\text{number of occasions some particular result occurs}}{\text{total number of occurrences}} \tag{2.1}$$

The probability can only be between 0 and 1 (normalized). We write a discrete normalized distribution as:

$$\sum_i P_i = 1 \tag{2.2}$$

An example would be a the probability of rolling a dice and getting a certain number. If the dice is fair, we have 6 total outcomes with each having a probability of 1/6

$$\sum_{i=1}^{6} P_i = 6 \times \frac{1}{6} = 1$$

An example of continuous probabilities would be the isotropic emission of say a photon (let's take this to be 2-dimensional for simplicity). The probability that the photon comes off at a set angle θ is just

$$P(\theta) = \frac{1}{2\pi},$$

it's just constant for any angle. The probability that the photon is emitted in some range of angles, $\Delta\theta$, would be $\Delta\theta/2\pi$. Integrating over all of 2π angles, we show:

$$\int_0^{2\pi} P(\theta) d\theta = \frac{1}{2\pi} \int_0^{2\pi} d\theta = 2\pi/2\pi = 1.$$

The familiar normalization condition for wave functions in quantum mechanics is just a property of probability. The probability that the state represented by the wave function ψ exists from $-\infty$ to ∞ is 1.

Let's develop some definitions and notation. If A and B are possible outcomes, then:

$$P(A + B) \equiv \text{probability of } A \text{ or } B, \tag{2.3}$$

When A and B are independent (exclusive) the $P(A + B)$ is just equal to the sum of $P(A)$ and $P(B)$, but in general:

$$P(A + B) \leq P(A) + P(B).$$

Another definition:

$$P(AB) \equiv \text{probability of } A \text{ and } B = P(A|B)P(B). \tag{2.4}$$

The new form we've introduced is the *conditional probability*, $P(A|B)$, which is the probability of A given B. Let's jump into a couple examples using the conditional probability. Let's say that A represents that it's a Friday in October, and B represents we have a 771 class day during this day in October. There are 4 class days for 771 in October, and there are 5 Fridays in the current October (2 are on Friday, 2 are not). First let's determine if the A and B are independent (if they are, $P(A)P(B) = P(AB)$).

$$P(A)P(B) = \frac{5}{31} \times \frac{4}{31} \approx .02$$
$$P(AB) = P(B|A)P(A) = \frac{2}{4} \times \frac{4}{31} = \frac{2}{31} \approx .065.$$

Therefore, A and B are definitely not independent (as expected). We can also calculate $P(AB)$ using a different ordering of equation 2.4:

$$P(AB) = P(B|A)P(A) = \frac{2}{5} \times \frac{5}{31} = \frac{2}{31},$$

as expected.

We've now reached a point where we can introduce Bayes' Theorem:

$$P(A|B)P(B) = P(B|A)P(A) \tag{2.5}$$
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2.6}$$

Taking an example that works perfectly for current events, let A be the probability that you have ebola and let B be the probability that you test positive. Let's say a test has a 99% detection rate and a 1% false detection rate.

$$P(B|A) = \text{testing positive if you have ebola } (99\%)$$
$$P(A|B) = \text{probability of having ebola if you test positive } (\text{unknown})$$

We now need a value for $P(A)$, let's call the probability of having ebola 0.001. $P(B)$ is our false positive rate of 1%. Now we can calculate:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{0.99 \times 0.001}{0.01} \approx 10\%.$$

With this we something a little counter intuitive – if you test positive for ebola, there's really only a 10% chance you have it under these circumstances. Let's look at a better physics example, proton decay. Let A be the probability of proton decay, B is the probability of any event passing selection cuts.

$$P(B|A) = \text{effeciency (probability the event passes cuts given proton decay)}$$
$$P(A|B) = \text{probability of proton decay given an event passing selection cuts}$$

Using Bayes':

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \tag{2.7}$$

where $P(B)$ is the probability of a background event – therefore the experiment needs high efficiency and low background to make a serious measurement (increase the value of $P(A|B)$).

Let's now consider a parameter estimation scenario. Take $\frac{1}{2}(1 + \alpha \cos\theta)$. Let A be the experimental result probability and B be the model parameterized by α. $P(B|A) = P(A|B)P(B)/P(A)$ is the probability of α given the data. In this case $P(B)$ is very important, (this term is called the prior). If the experimenter does not know $P(B)$ well (if one does not have a reasonable distribution for it) then one cannot make much sense of $P(B|A)$ with much confidence. This is where Bayes' theorem has a weakness.

2.2 Probability Distributions

2.2.1 Binomial

Let's imagine we have three dice – how often will we get 3 ones? This is pretty simple:

$$(1/6)^3 \approx 0.0046,$$

about half a percent. How about 2 ones? Now we have 3 different scenarios where this will happen, The first a second dice will be 1, the first and third dice will be 1, and the second and third dice will be 1. Therefore wehave:

$$3 \times (1/6)^2(5/6) \approx 6.9\%$$

In this case – a "success" is getting a 1. Now we define a probability distribution using the following parameters:

$$P(x; n, p) \tag{2.8}$$

where x is the number of successes, n is the number of trials, and p is the probability of success in one trial. In the above case, our distribution would be $P(x; 3, 1/6)$, and it would be a **Binomial Distribution**. The probability distribution function for a binomial distribution is:

$$P(x; n, p) = \binom{n}{x} p^x q^{n-x}, \tag{2.9}$$

where

$$\binom{n}{x} \equiv n \text{ "choose" } x = \frac{n!}{x!\,(n-x)!} \text{ and } q \equiv 1 - p$$

Now we define the mean and the variance of the binomial distribution:

$$\langle x \rangle = \mu = \sum_{x=0}^{n} x P(x; n, p) = np \tag{2.10}$$

$$\sigma^2 = np(1-p) \tag{2.11}$$

2.2.2 Poisson

The limit of the binomial distribution for low rate processes ($\mu \ll n, p \ll 1$) yields the **Poisson distribution**:

$$P(x; \mu) = \frac{\mu^x}{x!} e^{-\mu}. \tag{2.12}$$

Poisson is unique in that the expected value and variance are the same:

$$\langle x \rangle = \mu \tag{2.13}$$

$$\sigma^2 = \mu \tag{2.14}$$

Poisson is continuous in μ and discrete in x.

2.2.3 Gaussian

Now we take $np \gg 1$ and arrive at the **Gaussian distribution**:

$$P(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]. \tag{2.15}$$

The Gaussian is defined for all x and is continuous. Some properties of the distribution:

$$\begin{aligned}
1 &= \int_{-\infty}^{\infty} dx\, P(x;\mu,\sigma) \\
\langle x\rangle &= \int_{-\infty}^{\infty} dx\, x\, P(x;\mu,\sigma) \\
\langle (x-\mu)^2\rangle &= \int_{-\infty}^{\infty} dx\, (x-\mu)^2\, P(x;\mu,\sigma) \\
P(\mu\pm\sigma;\mu,\sigma) &= \frac{1}{e} P(\mu;\mu,\sigma)
\end{aligned}$$

The "full width at half max", Γ, is 2.354σ. And finally:

$$\begin{aligned}
1\sigma \text{ out from mean } &\approx 68\% \text{ of area} \\
2\sigma &\approx 95\% \text{ of area} \\
3\sigma &\approx 99\% \text{ of area}
\end{aligned}$$

Analysis of Uncertainties **Fall 2014**

Lecture 3: 28 October 2014

Lecturer: Prof. Kate Scholberg *Notes by: Douglas Davis*

3.1 Estimating means & errors from data

In lecture one, we stated the following connection:

sample quantities	*truth*/parent quantities
$\overline{x}$	μ
s^2	σ^2

In many common situations the parent is a Gaussian or Poisson distribution (Poisson for low mean, low rate). For now let's assume we have a parent which is Gaussian. The probability dQ of measuring between possibility x and $x + dx$ is given by

$$dQ = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx. \tag{3.1}$$

Let's say we our particular measurements are x_i and the probability of measuring P_i is $P(x_i)$. We'll say that each x_i has the same μ and σ. Our set is

$$\{x_i\} : \{x_1, x_2, ..., s_N\}.$$

Using the notation μ = truth distribution and μ' = best estimate, let's work out μ' assuming that $\sigma' = \sigma$ and is constant. We start by defining:

$$P_i(\mu_i) = P(x_i|\mu_i), \text{ probability of getting } x_i \text{ given } \mu_i, \tag{3.2}$$

$$P_i(\mu') = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x_i-\mu'}{\sigma}\right)^2\right]. \tag{3.3}$$

We assume all probabilities are independent, therefore all probabilities can be multiplied to gather the probability of getting our entire set, recall that when A and B are independent:

$$P(AB) = P(A)P(B).$$

Therefore, given μ' the probability of getting $x_1 ... x_N$ (our collection) is:

$$P(\mu') = \prod_i^N P_i(\mu') \equiv \text{“Likelihood”}. \tag{3.4}$$

For our current scenario:

$$P(\mu') = \prod_i^N P_i(\mu') = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N \exp\left[-\frac{1}{2}\sum_i^N \left(\frac{x_i-\mu'}{\sigma}\right)^2\right]. \tag{3.5}$$

To get the best estimate for the mean we need to "maximize the likelihood." This is called the method of maximum likelihood. To do this, it is clear that we need to minimize the sum in equation 3.5. We'll minimize as a function of μ':

$$\frac{d}{d\mu'}\left[\frac{1}{2}\sum_i^N\left(\frac{x-\mu'}{\sigma}\right)^2\right] = 0 = \sum_i^N\left(\frac{x_i-\mu'}{\sigma^2}\right) \quad \rightarrow \quad \sum_i^N x_i = \sum_i^N \mu' = N\mu'.$$

This yields the familiar mean of the sample:

$$\mu' = \frac{1}{N}\sum_i^N x_i. \tag{3.6}$$

Therefore, using the general maximum likelihood method, we arrive at the simple mean of the sample maximizes the likelihood and is therefore the best estimate of the mean. Next, we will work out the uncertainty on μ'. It's a common mistake to take σ as the uncertainty on the mean, but this is actually incorrect – we have more knowledge about the mean than that (the uncertainty is less than σ). We'll take our error propagation from lecture one. For some variable y as a function of $u, v, w... = \{\alpha_i\}$:

$$\sigma_y^2 = \sum_\alpha \left|\frac{\partial f}{\partial x_\alpha}\right|^2 \sigma_\alpha^2 + \text{ covariant terms.} \tag{3.7}$$

Here we are assuming no correlation so no covariant terms. Now we identify μ' as y and the x_i's as the $u, v, w...$ set.

$$\sigma_{\mu'}^2 = \sum_i^N \sigma_i^2 \left|\frac{\partial \mu'}{\partial x_i}\right|^2 = \sigma^2 \sum_i^N \left|\frac{\partial \mu'}{\partial x_i}\right|^2,$$

for constant $\sigma_i = \sigma$ (all x_i drawn from same sample).

$$\frac{\partial \mu'}{\partial x_i} = \frac{1}{N},$$

therefore:

$$\sigma_{\mu'}^2 = \sigma^2 \sum_i^N \frac{1}{N^2} = \sigma^2 N \frac{1}{N^2} = \frac{\sigma^2}{N}.$$

And now we've arrived at the uncertainty on the mean:

$$\sigma_{\mu'} = \frac{\sigma}{\sqrt{N}}. \tag{3.8}$$

Now we suppose that σ_i's are not all constant σ. We begin by just taking equation 3.5 and using σ_i for σ. Now it is iterated over in the product.

$$P(\mu') = \prod_i^N P_i(\mu') = \left(\frac{1}{\sigma_i\sqrt{2\pi}}\right)^N \exp\left[-\frac{1}{2}\sum_i^N\left(\frac{x_i-\mu'}{\sigma_i}\right)^2\right]. \tag{3.9}$$

We again maximize the likelihood by minimizing the sum in the exponential, now with unique σ_i:

$$\frac{d}{d\mu'}\frac{1}{2}\sum_i^N\left(\frac{x_i-\mu'}{\sigma_i}\right)^2 = 0 = \sum_i^N \frac{x_i-\mu'}{\sigma^2} \quad \rightarrow \quad \sum_i^N \frac{x_i}{\sigma_i^2} - \mu' \sum_i^N \frac{1}{\sigma^2} = 0.$$

We arrive at:

$$\mu' = \frac{\sum_i^N \frac{x_i}{\sigma_i^2}}{\sum_i^N \frac{1}{\sigma_i^2}}. \tag{3.10}$$

And the variance of the estimate of the mean (uncertainty on the mean):

$$\sigma_{\mu'}^2 = \frac{1}{\sum_i^N \frac{1}{\sigma_i^2}}. \tag{3.11}$$

Notice that if $\sigma_i \to$ constant σ equations 3.10 & 3.11 just go to equations 3.6 & 3.8, respectively.

Let's take a minute to talk about Poisson. For Poisson distributions $\sigma \to \sqrt{\mu}$. When we think Poisson we think low rate counting, for example: a low number of events in some Δt. Let μ_t be the mean number of events in Δt. The Poisson distribution:

$$P(x;t) \to P(x;\mu) = \frac{\mu^x e^{-\mu}}{x!}, \tag{3.12}$$

for $x = 0, 1, 2, \ldots$. Imagine N measurements (for $\mu \geq 10$ this would $\sim$ Gaussian).

$$\mu_t \sim \overline{x}_t, \tag{3.13}$$

$\overline{x}_t$ is the best estimate of μ_t, the best estimate of the variance is

$$\sigma_t = \sqrt{\mu_t} \sim \sqrt{\overline{x}_t}. \tag{3.14}$$

This is the spread. The error on the mean is

$$\sigma_{t\mu} = \frac{\sigma_t}{\sqrt{N}} \sim \sqrt{\frac{\overline{x}}{N}}. \tag{3.15}$$

We'll look at a quick example: Take $N = 4$, with measurements 3, 4, 7, and 2.

$$\begin{aligned} \mu_t = \overline{x}_t &= \frac{1}{4} \times (3 + 4 + 7 + 2) = 4 \\ \sigma_t &= 2 \\ \sigma_{t\mu} &= \sqrt{4/4} = 1 \\ \text{therefore, } \mu_t &= 4 \pm 1 \end{aligned}$$

3.2 Error Matrices Introduction

Error matrices are helpful for understanding correlations. consider two variables u, v – each is a Gaussian with σ_u, σ_v and $\mu_u = \mu_v = 0$ (binormal).

$$P(u) = \frac{1}{\sigma_u\sqrt{2\pi}} e^{-u^2/2\sigma_u^2} \qquad P(v) = \frac{1}{\sigma_v\sqrt{2\pi}} e^{-v^2/2\sigma_v^2}.$$

The uncorrelated combination (probability of u and v):

$$P(uv) = P(u)P(v) = \frac{1}{2\pi\sigma_u\sigma_v} \exp\left[-\frac{1}{2}\left(\frac{u^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2}\right)\right]. \tag{3.16}$$

Let's focus on the $\left(u^2/\sigma_u^2 + v^2/\sigma_v^2\right)$ term in the exponential. If the sum of this term is equal to one, then the probability is goes down by a factor of $e^{-1/2}$. The term equal to 1 is also the equation of an ellipse.

$$\frac{u^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} = 1 \tag{3.17}$$

Figure 3.1a shows the ellipse and Figure 3.1b shows the ellipse overlayed with a chosen value of u. Equation 3.17 can be written in matrix form:

$$\begin{pmatrix} u & v \end{pmatrix} \underbrace{\begin{pmatrix} 1/\sigma_u^2 & 0 \\ 0 & 1/\sigma_v^2 \end{pmatrix}}_{M^{-1}} \begin{pmatrix} u \\ v \end{pmatrix} = 1,$$

where

$$M = \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_v^2 \end{pmatrix}. \tag{3.18}$$

M above represents uncorrelated u and v. In general the matrix elements of M are represented as:

$$M_{ij} = \langle (w_i - \overline{w}_i)\,(w_j - \overline{w}_j) \rangle\,, \tag{3.19}$$

where w_i, w_j in our matrix above is either u or v. Therefore:

$$\begin{aligned} \text{for } i = j \rightarrow \left\langle (u - \overline{u})^2 \right\rangle &= \sigma_u^2 \\ \rightarrow \left\langle (v - \overline{v})^2 \right\rangle &= \sigma_v^2 \\ \text{for } i \neq j \rightarrow \langle (u - \overline{u})(v - \overline{v}) \rangle &= \sigma_{uv}^2 = \lim_{N \to \infty} \frac{1}{N} \sum_k^N (u_k - \overline{u})\,(v_k - \overline{v}) \end{aligned}$$

Now we'll wrap up with a quick example: suppose we have uncorrelated u and v where:

$$\begin{aligned} u' &= u\cos\theta - v\sin\theta \\ v' &= u\sin\theta + v\cos\theta \end{aligned}$$

Which allows us to write an equation similar in form to equation 3.17, with a, b, c some constants:

$$\frac{u'^2}{a} + \frac{v'^2}{b} + \frac{u'v'}{c} = 1$$

Now this equation is no longer a standard ellipse with a major and minor axis, it has some tilt to it, which we show in Figure 3.1c. The most likely values for u and v are no longer the easy find at the 0 for the partner variable. This will continue in the next lecture.

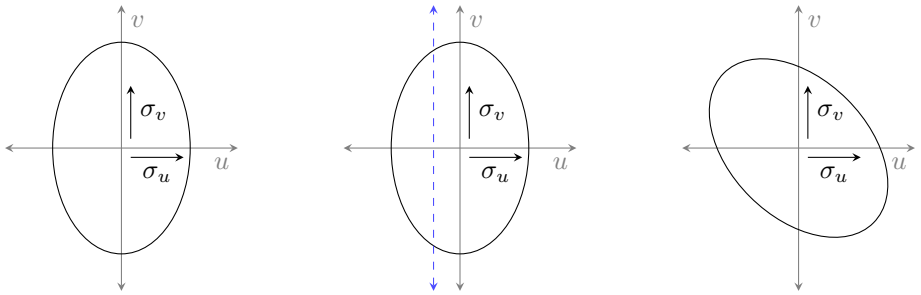

Figure 3.1: For (a), (b), and (c), the black ellipse is called the *contour of equal probability*. In (b), the blue dash represents a chosen u, and here the maximum probability would be at $v = 0$. In (c), the ellipse has been rotated by some angle θ, and at some chosen v or u, the maximum probability is not longer at $u = 0$ and $v = 0$, respectively.

Analysis of Uncertainties **Fall 2014**

Lecture 4: 31 October 2014

Lecturer: Prof. Kate Scholberg *Notes by: Douglas Davis*

4.1 Error matrices continued

Last lecture we introduced error matrices. Let's briefly remind ourselves of error propagation for some measurement x which is a function of many variables $u, v, w, ...$

$$\sigma_x^2 = \sigma_u^2 \left(\frac{\partial x}{\partial u}\right) + \sigma_v^2 \left(\frac{\partial x}{\partial v}\right) + ... + 2\sigma_{uv}^2 \left(\frac{\partial x}{\partial u}\right)\left(\frac{\partial x}{\partial v}\right) + ... \tag{4.1}$$

where

$$\sigma_{uv}^2 = \lim_{N\to\infty} \frac{1}{N} \sum_i^N (u_i - \overline{u})(v_i - \overline{v}) = \langle (u - \overline{u})(v - \overline{v}) \rangle . \tag{4.2}$$

If u and v are uncorrelated, σ_{uv} is of course zero. If u and v are correlated – we have two possible situations, *positive correlation* (or simply *correlated* variables) and *negative correlation* (or simply *anticorrelated* variables). For two variables u and v, positive correlation would correspond to u and v growing or decreasing together – an example would be height and weight (with a large sample, the majority of taller people would weight more than those shorter than them). Negative correlation between u and v is a situation where as one grows the other decreases. Take the height of a horse racing jockey and the speed the horse will run. Figure 4.1 shows samples with their respective correlations. The covariance σ_{uv}^2 has dimensions – it's often

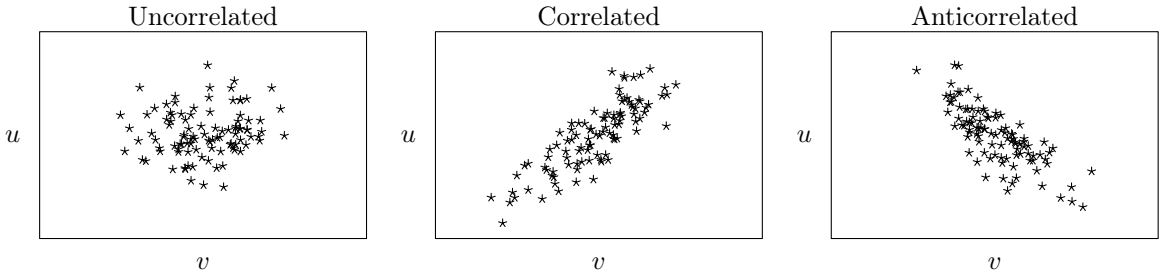

Figure 4.1: Three samples which correspond to uncorrelated variables, correlated variables, and uncorrelated variables.

desired to have a dimensionless quantity. For this we will define the correlation coefficient:

$$\rho = \frac{\sigma_{uv}^2}{\sigma_u \sigma_v} = \frac{\mathrm{cov}(u,v)}{\sigma_u \sigma_v} \tag{4.3}$$

It's clear that ρ is bounded: $-1 \leq \rho \leq 1$. We can define three levels of correlation in table 4.1.

Now let's dive back into error ellipses. Take two uncorrelated outcomes u and v with Gaussian probabilities:

$$P(u) = \frac{1}{\sqrt{2\pi}\sigma_u} e^{-u^2/2\sigma_u^2} \quad P(v) = \frac{1}{\sqrt{2\pi}\sigma_v} e^{-v^2/2\sigma_v^2}. \tag{4.4}$$

Table 4.1: ρ values table

$\rho = 0$	uncorrelated
$\rho = 1$	complete correlation
$\rho = -1$	complete anticorrelation

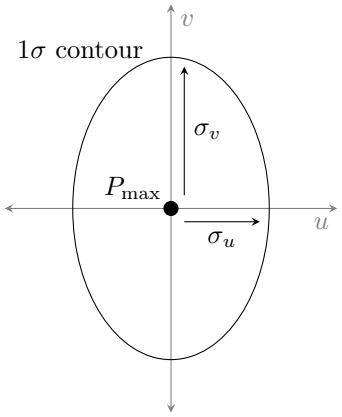

Figure 4.2: The 1σ contour displayed in an error ellipse corresponding to a pair of uncorrelated variables

It follows that

$$P(u)P(v) = \frac{1}{2\pi\sigma_u\sigma_v} \exp\left[-\frac{1}{2}\left(\frac{u^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2}\right)\right]. \tag{4.5}$$

We can write that $P = P_{\max}e^{-1/2}$ when

$$\frac{u^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} = 1.$$

This is the equation of an ellipse where the major and minor axes are given by the the square root of the variance of each quantity. Figure 4.2 shows this in graphical form. The volume inside the 1σ contour holds approximately 68% of the totla probability. We can write equation 4.5 in a new form using vectors and matrices:

$$P(u,v) = \frac{1}{2\pi\left|\mathbf{M}\right|^{1/2}} \exp\left[-\frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{x}\right]. \tag{4.6}$$

The term inside the exponential ($\mathbf{x}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{x}$) can be written:

$$\begin{pmatrix} u & v \end{pmatrix} \begin{pmatrix} 1/\sigma_u^2 & 0 \\ 0 & 1/\sigma_v^2 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}.$$

From here we define the error matrix $\mathbf{M}$ (for uncorrelated variables):

$$\mathbf{M} = \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_v^2 \end{pmatrix}. \tag{4.7}$$

In generaal (or say for possibly correlated variables) the matrix takes the general form:

$$M_{ij} = \left\langle (u_i - \overline{u}_i)\,(u_j - \overline{u}_j) \right\rangle, \tag{4.8}$$

where u_i corresponds to some variable set $\{u, v, w, \dots\}$. Using the numbered indices again, we can write an arbitrary error matrix in the form:

$$\mathbf{M} = \begin{pmatrix} \sigma_1^2 & \text{cov}(1,2) & \text{cov}(1,3) & \dots \\ \text{cov}(1,2) & \sigma_2^2 & \text{cov}(2,3) & \dots \\ \text{cov}(1,3) & \text{cov}(2,3) & \sigma_3^2 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{4.9}$$

Therefore if we go back to our standard two variable correlated form (using u and v again) we would write the error matrix in the form:

$$\mathbf{M} = \begin{pmatrix} \sigma_u^2 & \text{cov}(u,v) \\ \text{cov}(u,v) & \sigma_v^2 \end{pmatrix}. \tag{4.10}$$

Using the definition of the correlation coefficient, we can now write the probability equation (with the ellipse equation in the exponential) in the following form:

$$P(u,v) = \frac{1}{2\pi\sigma_u\sigma_v}\frac{1}{\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2}\left[\frac{1}{(1-\rho)^2}\left(\frac{u^2}{\sigma_u^2}+\frac{v^2}{\sigma_v^2}-\frac{2\rho uv}{\sigma_u\sigma_v}\right)\right]\right\}. \tag{4.11}$$

This is assuming a mean of zero for u and v, for non zero means we simply change all u and v terms to $(u-\overline{u})$ and $(v-\overline{v})$, respectively.

Now we'll generalize to an arbitrary number of variables (here we say k variables):

$$P(x_1, x_2, \dots, x_k) = \frac{1}{(2\pi)^{k/2}}\frac{1}{|\mathbf{M}|^{1/2}}\exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\mathbf{M}(\mathbf{x}-\boldsymbol{\mu})\right], \tag{4.12}$$

where

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix} \qquad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{pmatrix} \quad \text{or} \quad \boldsymbol{\mu} = \begin{pmatrix} \overline{x}_1 \\ \overline{x}_2 \\ \vdots \\ \overline{x}_k \end{pmatrix}. \tag{4.13}$$

We note here that the eigenvalues of the error matrix are the semi axes of the error ellipse.

Make diagram showing error ellipse

Now that we've done a pretty in depth discussion of the error matrix let's talk about using it. Let's say we have a variable which is a function of measurable quantities with known uncertainties – we can use the error matrix for those values to determine the uncertainties in our desired variable. Our desired uncertainty is the uncertainty in some parameter ζ, and it is a function of N variables $\{x_1, \dots, x_N\}$:

$$\zeta = f(x_1, x_2, \dots, x_N).$$

We can define vector $\mathbf{D}$ of derivates of f w.r.t. all independent variables:

$$\mathbf{D} = \begin{pmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \\ \vdots \\ \partial f/\partial x_N \end{pmatrix} \tag{4.14}$$

Then the variance in ζ is defined as:

$$\sigma_\zeta^2 = \mathbf{D}^{\mathrm{T}}\mathbf{M}\mathbf{D} \tag{4.15}$$

For variable transformations – for example if x is a function of r and θ and y is a function of r and θ, then we develop a Jacobi style matrix to determine an error matrix for x and y systems:

$$\mathbf{A} = \begin{pmatrix} \partial x/\partial r & \partial y/\partial r \\ \partial x/\partial \theta & \partial y/\partial \theta \end{pmatrix}. \tag{4.16}$$

We then define the new error matrix (for x and y) as:

$$\mathbf{M}_{xy} = \mathbf{A}^{\mathrm{T}}\mathbf{M}_{r\theta}\mathbf{A} \tag{4.17}$$

This of course can be generalized to many variables. For example if $\{\alpha_1, \ldots, \alpha_N\}$ are dependent on $\{\beta_1, \ldots, \beta_M\}$, then $\mathbf{A}$ would have the form:

$$\mathbf{A} = \begin{pmatrix} \partial\alpha_1/\partial\beta_1 & \partial\alpha_2/\partial\beta_1 & \ldots & \partial\alpha_N/\partial\beta_1 \\ \partial\alpha_1/\partial\beta_2 & \partial\alpha_2/\partial\beta_2 & \ldots & \partial\alpha_N/\partial\beta_2 \\ \vdots & \vdots & \ddots & \vdots \\ \partial\alpha_1/\partial\beta_M & \partial\alpha_2/\partial\beta_M & \ldots & \partial\alpha_N/\partial\beta_M \end{pmatrix} \tag{4.18}$$

Now if we go back to the (x, y) and (r, θ) example, we can define some variable z which is a function of x and y:

$$z = f(x, y)$$

And now to determine the uncertainty on a variable in some transformed coordinates we develop another derivative vector in the new system (call it the prime system):

$$\mathbf{D}' = \begin{pmatrix} \partial z/\partial x \\ \partial z/\partial y \end{pmatrix} \tag{4.19}$$

And we calculate the variance of z as:

$$\sigma_z^2 = \mathbf{D}'^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{M}_{r\theta}\mathbf{A}\mathbf{D}', \tag{4.20}$$

which would just be (as intuitively expected based on equation 4.15):

$$\sigma_z^2 = \mathbf{D}'^{\mathrm{T}}\mathbf{M}_{xy}\mathbf{D}' \tag{4.21}$$

And of course this can be generalized back to our α and β system if we define some ξ as a function of all of the α's. Then:

$$\mathbf{D}' = \begin{pmatrix} \partial\xi/\partial\alpha_1 \\ \partial\xi/\partial\alpha_2 \\ \vdots \\ \partial\xi/\partial\alpha_N \end{pmatrix} \tag{4.22}$$

Analysis of Uncertainties **Fall 2014**

Lecture 5: 12 November 2014

Lecturer: Prof. Kate Scholberg *Notes by: Douglas Davis*

5.1 Synopsis of error matrix equations

Recall from the previous lecture the error matrix (for correlated variables $x_1, \ldots, x_n$):

$$\mathbf{M} = \begin{pmatrix} \sigma_{x_1}^2 & \mathrm{cov}(x_1, x_2) & \ldots & \mathrm{cov}(x_1, x_n) \\ \mathrm{cov}(x_1, x_2) & \sigma_{x_2}^2 & \ldots & \mathrm{cov}(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(x_1, x_n) & \mathrm{cov}(x_2, x_n) & \ldots & \sigma_{x_n}^2 \end{pmatrix} \tag{5.1}$$

And this came from:

$$P(x_1, x_2, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}} \frac{1}{|\mathbf{M}|^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^{\mathrm{T}} \mathbf{M}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right] \tag{5.2}$$

We can use the error matrix for three scenarios:

1. Error propagation of a variable as a function of independent variables $x_1, \ldots, x_n$.

$$r = f(x_1, \ldots, x_n) \quad \rightarrow \quad \sigma_r^2 = \mathbf{D}^{\mathrm{T}} \mathbf{M} \mathbf{D} \quad \text{where} \quad \mathbf{D} = \begin{pmatrix} \partial r / \partial x_1 \\ \vdots \\ \partial r / \partial x_n \end{pmatrix} \tag{5.3}$$

2. Changing variables. If $\{y_m\}$ are all functions of $\{x_n\}$.

$$\mathbf{A} = \begin{pmatrix} \partial y_1 / \partial x_1 & \ldots & \partial y_m / \partial x_1 \\ \vdots & \ddots & \vdots \\ \partial y_1 / \partial x_n & \ldots & \partial y_m / \partial x_n \end{pmatrix} \quad \rightarrow \quad \mathbf{M}_{\{y_m\}} = \mathbf{A}^{\mathrm{T}} \mathbf{M}_{\{x_n\}} \mathbf{A} \tag{5.4}$$

3. And uncertainties in values dependent on changed variables. If z is a function of all $\{y_m\}$:

$$z = f(y_1, \ldots, y_m) \quad \rightarrow \quad \sigma_z^2 = \mathbf{D'}^{\mathrm{T}} \mathbf{A}^{\mathrm{T}} \mathbf{M}_{\{x_n\}} \mathbf{A} \mathbf{D'} = \mathbf{D'}^{\mathrm{T}} \mathbf{M}_{\{y_m\}} \mathbf{D'} \quad \text{where} \quad \mathbf{D'} = \begin{pmatrix} \partial z / \partial y_1 \\ \vdots \\ \partial z / \partial y_m \end{pmatrix} \tag{5.5}$$

5.2 Error matrices example uses

Analysis of Uncertainties **Fall 2014**

Lecture 5: 12 November 2014

Lecturer: Prof. Kate Scholberg *Notes by: Douglas Davis*

5.1 Synopsis of error matrix equations

Recall from the previous lecture the error matrix (for correlated variables $x_1, \ldots, x_n$):

$$\mathbf{M} = \begin{pmatrix} \sigma_{x_1}^2 & \mathrm{cov}(x_1, x_2) & \ldots & \mathrm{cov}(x_1, x_n) \\ \mathrm{cov}(x_1, x_2) & \sigma_{x_2}^2 & \ldots & \mathrm{cov}(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(x_1, x_n) & \mathrm{cov}(x_2, x_n) & \ldots & \sigma_{x_n}^2 \end{pmatrix} \tag{5.1}$$

And this came from:

$$P(x_1, x_2, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}} \frac{1}{|\mathbf{M}|^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^{\mathrm{T}} \mathbf{M}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right] \tag{5.2}$$

We can use the error matrix for three scenarios:

1. Error propagation of a variable as a function of independent variables $x_1, \ldots, x_n$.

$$r = f(x_1, \ldots, x_n) \quad \rightarrow \quad \sigma_r^2 = \mathbf{D}^{\mathrm{T}} \mathbf{M} \mathbf{D} \quad \text{where} \quad \mathbf{D} = \begin{pmatrix} \partial r / \partial x_1 \\ \vdots \\ \partial r / \partial x_n \end{pmatrix} \tag{5.3}$$

2. Changing variables. If $\{y_m\}$ are all functions of $\{x_n\}$.

$$\mathbf{A} = \begin{pmatrix} \partial y_1 / \partial x_1 & \ldots & \partial y_m / \partial x_1 \\ \vdots & \ddots & \vdots \\ \partial y_1 / \partial x_n & \ldots & \partial y_m / \partial x_n \end{pmatrix} \quad \rightarrow \quad \mathbf{M}_{\{y_m\}} = \mathbf{A}^{\mathrm{T}} \mathbf{M}_{\{x_n\}} \mathbf{A} \tag{5.4}$$

3. And uncertainties in values dependent on changed variables. If z is a function of all $\{y_m\}$:

$$z = f(y_1, \ldots, y_m) \quad \rightarrow \quad \sigma_z^2 = \mathbf{D'}^{\mathrm{T}} \mathbf{A}^{\mathrm{T}} \mathbf{M}_{\{x_n\}} \mathbf{A} \mathbf{D'} = \mathbf{D'}^{\mathrm{T}} \mathbf{M}_{\{y_m\}} \mathbf{D'} \quad \text{where} \quad \mathbf{D'} = \begin{pmatrix} \partial z / \partial y_1 \\ \vdots \\ \partial z / \partial y_m \end{pmatrix} \tag{5.5}$$

5.2 Error matrices example uses

Analysis of Uncertainties **Fall 2014**

Lecture 6: 18 November 2014

Lecturer: Prof. Kate Scholberg *Notes by: Douglas Davis*

6.1 TBA