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 $\mu = $ truth distribution and $\mu' = $ 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 $x_1 \ldots x_N$ (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.

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

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:

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

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

$$u' = u\cos\theta - v\sin\theta$$
$$v' = u\sin\theta + v\cos\theta$$

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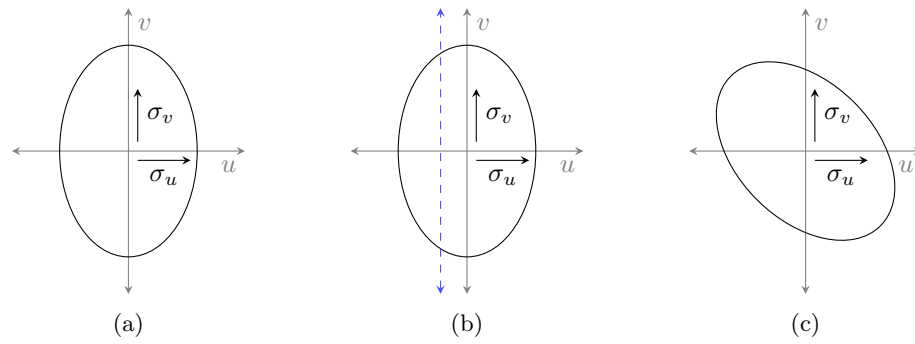
Figure 3.1: For (a), (b), and (c), the black ellipse is called the *contour of equal probability.* In (b), the blue dash represents a chosen u, and here the maximum probability would be at $v = 0$. In (c), the ellipse has been rotated by some angle θ, and at some chosen v or u, the maximum probability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