2.6 Inverse chi-squared statistic

Our starting distribution is the Gaussian

$$\text{prob}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{(x - \mu)^2}{2\sigma^2} \right]$$

and so for $n$ data $X_i$ we have the posterior distribution for $\mu$ and $\sigma$

$$\text{prob}(\mu, \sigma \mid \text{data}) \propto \left( \frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left[ -\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{2\sigma^2} \right] \text{prob}(\mu, \sigma).$$

There are reasonably strong theoretical arguments for taking the combined prior for the location and scale parameters to be the Jeffreys prior (see Lee, 1997)

$$\text{prob}(\mu, \sigma) \propto \frac{1}{\sigma}$$

with obvious caveats about the divergent nature of the distribution.

Given this, if we wish to know the posterior distribution of the mean without regard to the possible values of $\sigma$, we may simply marginalize it out:

$$\text{prob}(\mu \mid \text{data}) = \int \text{prob}(\mu, \sigma \mid \text{data}) d\sigma.$$

A neat trick to tidy up the expression is to use

$$\sum_{i=1}^{n}(X_i - \mu)^2 = S + n(<X> - \mu)^2$$

where

$$S = \sum_{i=1}^{n}(X_i - <X>)^2$$

and the angle brackets denote the simple arithmetic mean of the data. The marginalization is now a standard integration (change variables to $u = 1/\sigma$) and we find that

$$\text{prob}(\mu \mid \text{data}) \propto \frac{1}{(S + n(<X> - \mu)^2)^{n/2}}$$

which shows clearly how the most probable value of the mean concentrates near the arithmetic mean; the scatter in the data, which must diminish this, appears in the number $S$ — and the parameter $\sigma$ has of course vanished. This is very close to a standard t-distribution, which as we will see in Chapter 5, turns up naturally when we examine differences of means.
By the same methods we can eliminate the parameter $\mu$, marginalizing it out to obtain

$$\text{prob}(\sigma | \text{data}) \propto \frac{1}{\sigma^{n+1}} \exp \left[ -\frac{S}{2\sigma^2} \right].$$

This has the form of an “inverse chi-square distribution”, meaning that changing variables to $u = 1/\sigma$ will give a standard chi-square distribution. As in the previous case, we see that the “nuisance parameter” ($\mu$ this time) has conveniently vanished, its effect being mediated through the number $S$ again.

In fact the two important numbers $S$ and $< X >$ tell us everything we need to know about the data. They are called sufficient statistics, and we will meet them several times in the next Chapter for this reason.