CASE STUDY: WHAT IS THE MEAN OF A POPULATION?

ALAN HEAVENS

Suppose we have N independent samples $\{x_i\}$, each drawn from a gaussian with mean $\mu \equiv \langle x \rangle$ and (for simplicity the same) standard deviation σ. What is μ? We can do this in a Bayesian and Frequentist way, and find very similar-looking answers, but the interpretations are very different.

Bayesian: What we want is the posterior pdf for μ, given the data $\{x_i\}$ and any prior information $p(\mu)$. i.e. we want $p(\mu|\{x_i\})$. To do this, we use Bayes' theorem,

$$p(\mu|\{x_i\}) = \frac{p(\{x_i\}|\mu)p(\mu)}{p(\{x_i\})}.$$

We assume a flat prior, $p(\mu)$ =constant, and so

$$p(\mu|\{x_i\}) \propto p(\{x_i\}|\mu) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right].$$

This is clearly a gaussian in μ:

$$p(\mu|\{x_i\}) \propto \exp\left[-\frac{N\mu^2}{2\sigma^2} + \frac{\mu}{\sigma^2}\sum_{i=1}^{N} x_i\right] \propto \exp\left[-\frac{(\mu-\bar{x})^2}{2\sigma^2/N}\right]$$

where

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

is the data average. So we see that the posterior pdf for μ, assuming a uniform prior, is a gaussian, centred on the *average of the data*, with a variance given by σ^2/N. It is easily to generalise to the case where σ_i varies with i.

Frequentist: The concept of the distribution of μ makes no sense to a frequentist; it has a value, and we want to estimate it. Here we might as well let σ depend on i. Let us consider an *estimator* $\hat{\mu}$ for it, and it makes sense to consider

$$\hat{\mu} = \sum_{i=1}^{N} w_i x_i$$

for some weights w_i. Now imagine doing the experiment many times. The expected value of $\hat{\mu}$ will be

$$\langle \hat{\mu} \rangle = \mu \sum_{i=1}^{N} w_i$$

so we can make the estimator *unbiased* (i.e. $\langle \hat{\mu} \rangle = \mu$) if we choose

$$\sum_{i=1}^{N} w_i = 1.$$

Next, we want to minimise the scatter of the estimator from experiment to experiment, so we minimise the variance of $\hat{\mu}$. Since the samples are independent the variances add, and the variance of $w_i x_i$ is $w_i^2 \sigma_i^2$, hence

$$\sigma_{\hat{\mu}}^2 = \sum_{i=1}^{N} w_i^2 \sigma_i^2.$$

We minimise this subject to the constraint $\sum w_i = 1$, so introduce a Lagrange multiplier λ:

$$\frac{\partial}{\partial w_j} \left[\sum_{i=1}^{N} w_i^2 \sigma_i^2 - \lambda \sum_{i=1}^{N} w_i \right] = 0$$

Hence

$$2 w_j \sigma_j^2 - \lambda = 0$$

i.e.

$$w_j \propto \frac{1}{\sigma_j^2}$$

and we see that the *minimum-variance estimate* has inverse-variance weighting. Clearly, for the case $\sigma_j = \sigma$constant, we obtain $w_j = 1/N$. The distribution of $\hat{\mu}$ is also a gaussian (not shown here), with a variance given by the above calculations as

$$\sigma_{\hat{\mu}}^2 = \frac{\sigma^2}{N}.$$

Interpretation: Both methods give a gaussian of the same width. In the Bayesian interpretation, the gaussian is centred on the data average, and is the *posterior pdf for* μ, given the data and a flat prior. In the frequentist approach, the gaussian is the *pdf of the estimator*, from repeated trials, given (and centred on) the true value μ[1]. You can decide which you think is closer to what we are looking for.

[1]Note that in this simple case, the estimator is unbiased and has a distribution which is independent of μ except for being shifted, but one can envisage more complicated cases where, for example, the distribution of $\hat{\mu}$ is always close to $\mu = 0$, even if the true value is far from 0. This would be much more problematic.