WIKIPEDIA

Effective sample size

In statistics, **effective sample size** is a notion defined for a sample from a distribution when the observations in the sample are correlated or weighted.[1]

Contents

Correlated observations

Suppose a sample of several observations y_i is drawn from a distribution with mean μ and standard deviation σ. Then the mean of this distribution is estimated by the mean of the sample:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

In that case, the variance of $\hat{\mu}$ is given by

$$\operatorname{Var}(\hat{\mu}) = \frac{\sigma^2}{n}$$

However, if the observations in the sample are correlated, then $\operatorname{Var}(\hat{\mu})$ is somewhat higher. For instance, if all observations in the sample are completely correlated ($\rho_{(i,j)} = 1$), then $\operatorname{Var}(\hat{\mu}) = \sigma^2$ regardless of n.

The effective sample size n_{eff} is the unique value (not necessarily an integer) such that

$$\operatorname{Var}(\hat{\mu}) = \frac{\sigma^2}{n_{\text{eff}}}$$

n_{eff} is a function of the correlation between observations in the sample. Suppose that all the correlations are the same and nonnegative, i.e. if $i \neq j$, then $\rho_{(i,j)} = \rho \geq 0$. In that case, if $\rho = 0$, then $n_{\text{eff}} = n$. Similarly, if $\rho = 1$ then $n_{\text{eff}} = 1$. More generally,

$$n_{\text{eff}} = \frac{n}{1 + (n-1)\rho}$$

The case where the correlations are not uniform is somewhat more complicated. Note that if the correlation is negative, the effective sample size may be larger than the actual sample size. Similarly, it is possible to construct correlation matrices that have an $n_{\text{eff}} > n$ even when all correlations are positive. Intuitively, n_{eff} may be thought of as the information content of the observed data.

Weighted samples

If the data has been weighted, then several observations composing a sample have been pulled from the distribution with effectively 100% correlation with some previous sample. In this case, the effect is known as Kish's Effective Sample Size[2]

$$n_{\text{eff}} = \frac{(\sum_{i=1}^{n} w_i)^2}{\sum_{i=1}^{n} w_i^2}$$

References

1. Tom Leinster (December 18, 2014). "Effective Sample Size" (https://golem.ph.utexas.edu/category/2014/12/effective_sample_size.html) (html).
2. "Design Effects and Effective Sample Size" (http://docs.displayr.com/wiki/Design_Effects_and_Effective_Sample_Size) (html).

Further reading

- M. B., Priestley (1981), *Spectral Analysis and Time Series 1*, Academic Press, §5.3.