WIKIPEDIA
The Free Encyclopedia

Inverse-variance weighting

In statistics, **inverse-variance weighting** is a method of aggregating two or more random variables to minimize the variance of the weighted average. Each random variable is weighted in inverse proportion to its variance, i.e., proportional to its precision.

Given a sequence of independent observations y_i with variances σ_i^2, the inverse-variance weighted average is given by[1]

$$\hat{y} = \frac{\sum_i y_i/\sigma_i^2}{\sum_i 1/\sigma_i^2}.$$

The inverse-variance weighted average has the least variance among all weighted averages, which can be calculated as

$$Var(\hat{y}) = \frac{1}{\sum_i 1/\sigma_i^2}.$$

If the variances of the measurements are all equal, then the inverse-variance weighted average becomes the simple average.

Inverse-variance weighting is typically used in statistical meta-analysis or sensor fusion to combine the results from independent measurements.

Context

Suppose an experimenter wishes to measure the value of a quantity, say the acceleration due to gravity of Earth, whose true value happens to be μ. A careful experimenter makes multiple measurements, which we denote with n random variables $X_1, X_2, \ldots, X_n$. If they are all noisy but unbiased, i.e., the measuring device does not systematically overestimate or underestimate the true value and the errors are scattered symmetrically, then the expectation value $E[X_i] = \mu \ \forall i$. The scatter in the measurement is then characterised by the variance of the random variables $Var(X_i) := \sigma_i^2$, and if the measurements are performed under identical scenarios, then all the σ_i are the same, which we shall refer to by σ. Given the n measurements, a typical estimator for μ, denoted as $\hat{\mu}$, is given by the simple average $\overline{X} = \frac{1}{n}\sum_i X_i$. Note that this empirical average is also a random variable, whose expectation value $E[\overline{X}]$ is μ but also has a scatter. If the individual measurements are uncorrelated, the square of the error in the estimate is given by $Var(\overline{X}) = \frac{1}{n^2}\sum_i \sigma_i^2 = \left(\frac{\sigma}{\sqrt{n}}\right)^2$. Hence, if all the σ_i are equal, then the error in the estimate decreases with increase in n as $1/\sqrt{n}$, thus making more observations preferred.

Instead of n repeated measurements with one instrument, if the experimenter makes n of the same quantity with n different instruments with varying quality of measurements, then there is no reason to expect the different σ_i to be the same. Some instruments could be noisier than others. In the example of measuring the acceleration due to gravity, the different "instruments" could be measuring g from a simple pendulum, from analysing a projectile motion etc. The simple average is no longer an optimal estimator, since the error in $\overline{X}$ might actually exceed the error in the least noisy measurement if different measurements have very different errors. Instead of discarding the noisy measurements that increase the final error, the experimenter can combine all the measurements with appropriate weights so as to give more importance to the least noisy measurements and vice versa. Given the knowledge of $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$, an optimal estimator to measure μ would be a weighted mean of the measurements $\hat{\mu} = \frac{\sum_i w_i X_i}{\sum_i w_i}$, for the particular choice of the weights $w_i = 1/\sigma_i^2$. The variance of the estimator $Var(\hat{\mu}) = \frac{\sum_i w_i^2 \sigma_i^2}{(\sum_i w_i)^2}$, which for the optimal choice of the weights become $Var(\hat{\mu}_{\text{opt}}) = \left(\sum_i \sigma_i^{-2} \right)^{-1}$.

Note that since $Var(\hat{\mu}_{\text{opt}}) < \min_j \sigma_j^2$, the estimator has a scatter smaller than the scatter in any individual measurement. Furthermore, the scatter in $\hat{\mu}_{\text{opt}}$ decreases with adding more measurements, however noisier those measurements may be.

Derivation

Consider a generic weighted sum $Y = \sum_i w_i X_i$, where the weights w_i are normalised such that $\sum_i w_i = 1$. If the X_i are all independent, the variance of Y is given by

$$Var(Y) = \sum_i w_i^2 \sigma_i^2.$$

For optimality, we wish to minimise $Var(Y)$ which can be done by equating the gradient with respect to the weights of $Var(Y)$ to zero, while maintaining the constraint that $\sum_i w_i = 1$. Using a Lagrange multiplier w_0 to enforce the constraint, we express the variance:

$$Var(Y) = \sum_i w_i^2 \sigma_i^2 - w_0 (\sum_i w_i - 1).$$

For $k > 0$,

$$0 = \frac{\partial}{\partial w_k} Var(Y) = 2 w_k \sigma_k^2 - w_0,$$

which implies that:

$$w_k = \frac{w_0/2}{\sigma_k^2}.$$

The main takeaway here is that $w_k \propto 1/\sigma_k^2$. Since $\sum_i w_i = 1$,

$$\frac{2}{w_0} = \sum_i \frac{1}{\sigma_i^2} := \frac{1}{\sigma_0^2}.$$

The individual normalised weights are:

$$w_k = \frac{1}{\sigma_k^2}\left(\sum_i \frac{1}{\sigma_i^2}\right)^{-1}.$$

It is easy to see that this extremum solution corresponds to the minimum from the second partial derivative test by noting that the variance is a quadratic function of the weights. Thus, the minimum variance of the estimator is then given by:

$$Var(Y) = \sum_i \frac{\sigma_0^4}{\sigma_i^4}\sigma_i^2 = \sigma_0^4 \sum_i \frac{1}{\sigma_i^2} = \sigma_0^4 \frac{1}{\sigma_0^2} = \sigma_0^2 = \frac{1}{\sum_i 1/\sigma_i^2}.$$

Normal distributions

For normally distributed random variables inverse-variance weighted averages can also be derived as the maximum likelihood estimate for the true value. Furthermore, from a Bayesian perspective the posterior distribution for the true value given normally distributed observations y_i and a flat prior is a normal distribution with the inverse-variance weighted average as a mean and variance $Var(Y)$

Multivariate case

For multivariate distributions an equivalent argument leads to an optimal weighting based on the covariance matrices $\mathbf{C}_i$ of the individual vector-valued estimates $\mathbf{x}_i$:

$$\hat{\mathbf{x}} = \left(\sum_i \mathbf{C}_i^{-1}\right)^{-1} \sum_i \mathbf{C}_i^{-1}\mathbf{x}_i$$

$$\hat{\mathbf{C}} = \left(\sum_i \mathbf{C}_i^{-1}\right)^{-1}$$

For multivariate distributions the term "precision-weighted" average is more commonly used.

See also

- Weighted least squares
- Portfolio theory
- Cramér-Rao bound

References

1. Joachim Hartung; Guido Knapp; Bimal K. Sinha (2008). *Statistical meta-analysis with applications* (https://archive.org/details/statisticalmetaa0000hart). John Wiley & Sons. ISBN 978-0-470-29089-7.