Weighted estimators

Subhaneil Lahiri
Harvard University

November 9, 2011

Abstract

We compute estimators and standard errors for samples that are independent but not identically distributed, assuming that cumulants scale in a particular way.

1 Grouping data points

Suppose that we start with an independent, identically distributed (iid) sample, i.e. a set of random variables $\{y_a\}$, $a = 1...M$

$$\mathrm{cum}_r(y_a) = c_r. \tag{1}$$

Now, say we put them into groups of size $\{n_i\}$, $i = 1...N$, and only recorded the means of each group

$$x_i = \frac{\sum_{b=1}^{n_i} y_{n_{i-1}+b}}{n_i}. \tag{2}$$

Then the $\{n_i\}$ would be independent but not identically distributed. In fact, their cumulants would be

$$\mathrm{cum}_r(x_i) = \frac{c_r}{(n_i)^{r-1}}. \tag{3}$$

In practice, we often do not have any concept of "number of samples per group". Instead we have some quantity that scales the same way, i.e.

$$\mathrm{cum}_r(x_i) = \frac{\tilde{c}_r}{(w_i)^{r-1}}. \tag{4}$$

We will refer to this as independent, differently scaled (ids). One can think of these quantities as being related to (3) by

$$w_i = \alpha n_i, \qquad \tilde{c}_r = \alpha^{r-1} c_r. \tag{5}$$

Note the w_i and $\tilde{c}_r$ are not separately well defined, as we can redefine them (changing α in (5)) with,

$$w_i \to \lambda w_i, \qquad \tilde{c}_r \to \lambda^{r-1} \tilde{c}_r. \tag{6}$$

Only quantities that are invariant under this redefinition are well defined. In particular, $\tilde{c}_1$ is the only one of the $\tilde{c}_r$ that is really meaningful.

2 Estimators

First let us define

$$\mu = \tilde{c}_1, \quad \sigma^2 = \tilde{c}_2, \quad \gamma = \tilde{c}_3, \quad \kappa = \tilde{c}_4. \tag{7}$$

As the x_i are independent, we have

$$\begin{aligned}
\mathbb{E}(x_i) =& \mu, \\
\mathbb{E}(x_i x_j) =& \mu^2 + \sigma^2 \frac{\delta_{ij}}{w_i}, \\
\mathbb{E}(x_i x_j x_k) =& \mu^3 + \mu\sigma^2 \left(\frac{\delta_{ij}}{w_i} + \frac{\delta_{jk}}{w_j} + \frac{\delta_{ki}}{w_k} \right) + \gamma \frac{\delta_{ijk}}{w_i^2}, \\
\mathbb{E}(x_i x_j x_k x_l) =& \mu^4 + \mu^2\sigma^2 \left(\frac{\delta_{ij}}{w_i} + \frac{\delta_{ik}}{w_i} + \frac{\delta_{il}}{w_i} + \frac{\delta_{jk}}{w_j} + \frac{\delta_{jl}}{w_j} + \frac{\delta_{kl}}{w_k} \right) \\
&+ \sigma^4 \left(\frac{\delta_{ij}}{w_i}\frac{\delta_{kl}}{w_k} + \frac{\delta_{ik}}{w_i}\frac{\delta_{jl}}{w_j} + \frac{\delta_{il}}{w_i}\frac{\delta_{jk}}{w_j} \right) \\
&+ \mu\gamma \left(\frac{\delta_{ijk}}{w_i^2} + \frac{\delta_{ijl}}{w_j^2} + \frac{\delta_{ikl}}{w_k^2} + \frac{\delta_{jkl}}{w_l^2} \right) + \kappa \frac{\delta_{ijkl}}{w_i^3},
\end{aligned} \tag{8}$$

where the generalised Kronecker-delta symbols are defined by

$$\delta_{ijkl\ldots} = \begin{cases} 1 & \text{if} \quad i = j = k = l \ldots \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

We will now go through the construction of an estimator for the mean, μ, in detail. Let

$$\hat{\mu} = \sum_i g_i x_i. \tag{10}$$

We have

$$\mathrm{Bias}(\hat{\mu}) = \left(\sum_i g_i - 1 \right) \mu, \qquad \mathrm{Var}(\hat{\mu}) = \left(\sum_i \frac{g_i^2}{w_i} \right) \sigma^2. \tag{11}$$

If we minimise the variance subject to the constraint that the bias is zero,

$$\frac{\partial\, \mathrm{Var}(\hat{\mu})}{\partial g_i} - \beta \frac{\partial\, \mathrm{Bias}(\hat{\mu})}{\partial g_i} = \frac{2g_i}{w_i}\sigma^2 - \beta\mu = 0, \tag{12}$$

where β is a Lagrange multiplier. Therefore, we must set

$$g_i = \frac{w_i}{\sum_j w_j}. \tag{13}$$

This leaves us with

$$\hat{\mu} = \frac{\sum_i w_i x_i}{\sum_i w_i}, \qquad \mathrm{Bias}(\hat{\mu}) = 0, \qquad \mathrm{Var}(\hat{\mu}) = \frac{\sigma^2}{\sum_i w_i}. \tag{14}$$

Therefore, this estimator is unbiased and consistent, assuming that the w_i/σ^2 do not decrease faster than the sample size increases.

We will construct an estimator for the variance in much less detail, starting with a good guess and improving it. Consider the quantity

$$\hat{s}^2 = \frac{\sum_i w_i (x_i - \hat{\mu})^2}{\sum_i w_i} . \tag{15}$$

Using (8), one can show that

$$\begin{aligned} \mathbb{E}(\hat{s}^2) &= \frac{N-1}{\sum_i w_i}\sigma^2, \\ \mathrm{Var}(\hat{s}^2) &= \frac{2(N-1)}{\left(\sum_i w_i\right)^2}\sigma^2 + \frac{\left(\sum_i w_i^{-1}\right)\left(\sum_i w_i\right) - 2N + 1}{\left(\sum_i w_i\right)^3}\kappa. \end{aligned} \tag{16}$$

So we can define an estimator for the variance

$$\begin{aligned} \hat{\sigma}^2 = \frac{\sum_i w_i (x_i - \hat{\mu})^2}{N-1}, \qquad &\mathrm{Bias}(\hat{\sigma}^2) = 0, \\ &\mathrm{Var}(\hat{\sigma}^2) = \frac{2\sigma^4}{N-1} + \frac{\left[\left(\sum_{ij} w_i/w_j\right) - 2N + 1\right]\kappa}{\left(\sum_i w_i\right)(N-1)^2} . \end{aligned} \tag{17}$$

This estimator is unbiased and consistent, as above. However it is not invariant under the rescaling (6). This is not surprising, given that σ^2 is itself not invariant under a rescaling of all the weights. Nevertheless, it can be used to compute a standard error in the mean:

$$\delta\mu^2 = \frac{\sum_i w_i (x_i - \hat{\mu})^2}{\left(\sum_i w_i\right)(N-1)}, \qquad \mathbb{E}(\delta\mu^2) = \mathrm{Var}(\hat{\mu}). \tag{18}$$

Not that both the estimators for the mean (14) and its standard error (18) are invariant under the rescaling (6). These are the only two fully meaningful formulae in the section.

3 One sample T-test

In this section we will assume that all cumulants except for the first two vanish, i.e. that the distributions are normal

$$x_i \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{w_i}\right) . \tag{19}$$

We wish to test the hypothesis that the mean is equal to some value, μ_0. Let

$$z_i \equiv \frac{\sqrt{w_i}}{\sigma}(x_i - \mu_0), \qquad Z = \sum_i z_i^2 . \tag{20}$$

Under the null hypothesis that $\mu = \mu_0$,

$$z_i \sim \mathrm{N}(0,1), \qquad Z \sim \chi^2_N. \tag{21}$$

We can rewrite Z as

$$\begin{aligned}
Z &= \frac{1}{\sigma^2}\sum_i w_i (x_i - \mu_0)^2 \\
&= \frac{1}{\sigma^2}\sum_i w_i (x_i - \hat\mu + \hat\mu - \mu_0)^2 \\
&= \frac{1}{\sigma^2}\sum_i w_i \left[(x_i - \hat\mu)^2 + (\hat\mu - \mu_0)^2\right] \\
&= \left(\frac{\sum_i w_i}{\sigma^2}\right)\left[(N-1)\delta\hat\mu^2 + (\hat\mu - \mu_0)^2\right].
\end{aligned} \tag{22}$$

This can be rewritten back in terms of the z_i as

$$\begin{aligned}
\left(\frac{\sum_i w_i}{\sigma^2}\right)(N-1)\delta\hat\mu^2 &= \sum_{ij} S_{ij} z_i z_j, \qquad & S_{ij} &= \delta_{ij} - \frac{\sqrt{w_i w_j}}{\sum_k w_k}, \\
\left(\frac{\sum_i w_i}{\sigma^2}\right)(\hat\mu - \mu_0)^2 &= \sum_{ij} M_{ij} z_i z_j, \qquad & M_{ij} &= \frac{\sqrt{w_i w_j}}{\sum_k w_k}.
\end{aligned} \tag{23}$$

We can see that M_{ij} is the projection operator onto a vector of the square root of the weights and S_{ij} is the perpendicular projection operator, therefore

$$\mathrm{rank}(S_{ij}) = N-1, \qquad \mathrm{rank}(M_{ij}) = 1. \tag{24}$$

By Cochran's theorem [1],

$$\begin{aligned}
\left(\frac{\sum_i w_i}{\sigma^2}\right)(N-1)\delta\hat\mu^2 &\sim \chi^2_{N-1}, \\
\left(\frac{\sum_i w_i}{\sigma^2}\right)(\hat\mu - \mu_0)^2 &\sim \chi^2_1,
\end{aligned} \tag{25}$$

and these two quantities are independent. This allows us to conclude that

$$\begin{aligned}
T^2 \equiv \frac{(\hat\mu - \mu_0)^2}{\delta\hat\mu^2} \quad &\sim \mathrm{F}_{1,N-1}, \\
T \equiv \frac{\hat\mu - \mu_0}{\delta\hat\mu} \quad &\sim \mathrm{T}_{N-1}.
\end{aligned} \tag{26}$$

Therefore, we can do one sample T-tests in the same way as usual.

References

[1] W. G. Cochran, "The distribution of quadratic forms in a normal system, with applications to the analysis of covariance," *Mathematical Proceedings of the Cambridge Philosophical Society* **30** (1934) no. 02, 178–191.