BRIEF ARTICLE

THE AUTHOR

1. Optimality of the AHL Estimator

Consider the problem of estimating the mean of a random variable $X \sim F_X(\mu; \sigma^2)$ using two observations X_1 and X_2. With probability π, X_1 is drawn from the true distribution F_X and X_2 is noise drawn from the distribution $F_Y(\kappa, \omega^2)$. With probability $1 - \pi$, X_2 is drawn from the correct distribution and X_1 is noise. Under this specification, exactly one of X_1 or X_2 is drawn from the distribution of interest at all times.

Observe that if π is known, we can construct an unbiased estimator using only X_1,

$$\hat{\mu}_1 = \frac{X_1}{\pi} - \frac{1-\pi}{\pi}\kappa \tag{1}$$

Similarly, we can construct an unbiased estimator using only X_2,

$$\hat{\mu_2} = \frac{X_2}{1-\pi} - \frac{\pi}{1-\pi}\kappa \tag{2}$$

Compare these to an estimator that uses both X_1 and X_2,

$$\hat{\mu} = a_1 X_1 + a_2 X_2 - a_3 \kappa \tag{3}$$

which has the following expectation,

$$E[\hat{\mu}] = (a_1 \pi + a_2(1-\pi))\mu + (a_1(1-\pi) + a_2\pi - a_3)\kappa$$

so that unbiased, requires

$$a_1\pi + a_2(1-\pi) = 1 \implies a_2(a_1) = \frac{1}{1-\pi} - \frac{a_1\pi}{1-\pi} \tag{4}$$

$$a_1(1-\pi) + a_2\pi = a_3 \implies a_3(a_1) = \frac{\pi}{1-\pi} + \frac{a_1 - 2a_1\pi}{1-\pi} \tag{5}$$

Any unbiased estimator of the form CITE $\hat{\mu}$ can be written as a linear combination of $\hat{\mu}_1$ and $\hat{\mu}_2$:

$$\hat{\mu} = d_1\hat{\mu}_1 + d_2\hat{\mu}_2$$

The variance of $\hat{\mu}$ depends on:

$$\begin{aligned}
\mathrm{Var}\,(\hat{\mu}_1) &= \frac{\mathrm{Var}(X_1)}{\pi^2} = \frac{1}{\pi^2}\left(\pi\sigma^2 + (1-\pi)\omega^2 + \pi(1-\pi)(\mu-\kappa)^2\right) \\
\mathrm{Var}\,(\hat{\mu}_2) &= \frac{\mathrm{Var}\,(X_2)}{(1-\pi)^2} = \frac{1}{(1-\pi)^2}\left((1-\pi)\sigma^2 + \pi\omega^2 + \pi(1-\pi)(\mu-\kappa)^2\right)
\end{aligned}$$

This follows from the law of total variance, with the random variable $D = 1$ if X_1 is drawn from the correct distribution and $D = 0$ otherwise.

$$\begin{aligned}
\mathrm{Var}\,(X_1) &= E[\mathrm{Var}\,(X_1|D)] + \mathrm{Var}\,(E[X_1|D]) \\
&= P(D=1)\sigma^2 + P(D=0)\omega^2 + \mathrm{Var}\,(()\,\mu D + \kappa(1-D)) \\
&= \pi\sigma^2 + (1-\pi)\omega^2 + \pi(1-\pi)(\mu-\kappa)^2
\end{aligned}$$

and similarly

$$\mathrm{Var}\,(X_2) = (1-\pi)\sigma^2 + \pi\omega^2 + \pi(1-\pi)(\mu-\kappa)^2$$

The important observation is that the variances are functions parameterized by $(\sigma^2, \omega^2, (\mu-\kappa)^2)$, and we can write:

$$\text{Var}\,(\hat{\mu}_1) = g(\sigma^2, \omega^2, (\mu - \kappa)^2, \pi)$$
$$\text{Var}\,(\hat{\mu}_2) = g(\sigma^2, \omega^2, (\mu - \kappa)^2, 1 - \pi)$$

They are exactly equal when $\pi = 1 - \pi$; otherwise they are quadratic as in the plot below. But they are exactly symmetric, so that the optimal estimator will always assign $d_1 = 1, d_2 = 0$ or $d_1 = 0, d_2 = 1$ depending on whether $\pi > 1 - \pi$. However, if π is not estimated precisely, the resulting variance explodes. The most conservative way is by treating $\pi = 0.5$, which gives the estimate of the mean! Which is the AHL estimator. .

Also $\frac{\partial g}{\partial \pi} < 0$ on $\pi \in [0, 1]$ so that there is a single crossing at $\pi = 0.5$!!! So unless $\pi = 0.5$ you always weight the more likely match and give it the proper weights. Could do this for every single i in the sample? Divide by π_i and you get the Horwitz Thompson.