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

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

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

Hence we can write $\hat{\mu}$ as a function of a_1,

$$\hat{\mu}(a_1) = a_1X_1 + \left(\frac{1}{1-\pi} - \frac{a_1\pi}{1-\pi}\right)X_2 - \left(\frac{\pi}{1-\pi} + \frac{a_1-2a_1\pi}{1-\pi}\right)\kappa$$

When $a_1 = \frac{1}{\pi}$, then $\hat{\mu} = \hat{\mu}_1$; and if $a_1 = 0$ then $\hat{\mu} = \hat{\mu_2}$.

We can write:

$$\begin{aligned}\hat{\mu} &= (a_1\pi)\hat{\mu}_1 + (1-a_1\pi)\hat{\mu}_2 + a_1(1-\pi)\kappa + \frac{\pi}{1-\pi}(1-a_1)\kappa - \left(\frac{\pi}{1-\pi} + \frac{a_1-2a_1\pi}{1-\pi}\right)\kappa \\ &= (a_1\pi)\hat{\mu}_1 + (1-a_1\pi)\hat{\mu}_2 - (a_1\pi)\kappa\end{aligned}$$

Hence any unbiased estimator $\hat{\mu}$ that uses X_1 and X_2 can be written as a linear combination of estimators using only X_1 or X_2. The problem of finding the minimum variance, unbiased estimator $\hat{\mu}$ reduces to finding d^* that solves

$$\min_d \ \mathrm{Var}\left(d\hat{\mu}_1 + (1-d)\hat{\mu}_2\right)$$

which is solved by $d^* = 0$ or $d^* = 1$ depending on whether $\mathrm{Var}(\hat{\mu}_1)$ or $\mathrm{Var}(\hat{\mu}_2)$ is smaller.

I now show that $\mathrm{Var}(\hat{\mu}_1) > \mathrm{Var}(\hat{\mu}_2)$, without loss of generality, except when $\pi = 0.5$, or $\sigma^2 = \omega^2 = (\mu-\kappa)^2$. Observe that,

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

This follows from the law of total variance, with the random variable $D = 1$ if X_1 is drawn from the correct distribution (and X_2 is drawn from the incorrect distribution), and

$D = 0$ otherwise.

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

Similarly,

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

Thus, $\text{Var}\,(\hat{\mu}_1)$ and $\text{Var}\,(\hat{\mu}_2)$ can be written as functions of σ^2, ω^2, and $(\mu-\kappa)^2$,

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

Importantly, $\frac{\partial g}{\partial x} < 0$ for $x \in [0,1]$; so there is a unique crossing $\text{Var}\,(\hat{\mu}_1) = \text{Var}\,(\hat{\mu}_2)$ when $\pi = 0.5$. The variance is decreasing in the probability that the observation is correct; hence the minimum variance unbiased estimator assigns $d = 1$ to the observation that has the highest probability of being correct, and applies inverse probability weighting accordingly.

This result holds for $L > 2$. Pick the $\pi_{i\ell}$ that is maximal and give it full weight. However, there are problems with weighting, especially if no observation has particularly high $\pi_{i\ell}$. Apply threshold

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.

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