Using information about match quality

Rachel Anderson

August 21, 2019

1 Estimating the mean

1.1 Setup

The goal is to estimate the mean the mean of a random variable X with mean μ and variance σ^2. The econometrician observes $\{X_{i1}, X_{i2}\}_{i=1}^n$, where only one observation is drawn from the true distribution of X. The other observation is noise, drawn from a distribution with mean κ and variance ω^2.

Suppose the econometrician knows that the first observation X_{1i} is drawn from the correct distribution with probability π, and drawn from the noisy distribution with probability $1 - \pi$. The two observations are perfectly dependent, so that if X_{i1} is drawn from the correct distribution, then X_{i2} is drawn from the incorrect distribution.

Suppose we want to estimate a mean, $\mu = E[X]$. For each i, we have two observations, X_{1i} and X_{2i}. One is drawn from the correct distribution which has mean μ and variance σ^2 and one is drawn from a known incorrect distribution with *known* mean κ and variance ω^2.

Suppose that the probability that the first is drawn from the correct correct dis-

tribution is π. Then

$$\begin{aligned} E[X_{1i}] &= \pi\mu + (1-\pi)\kappa \\ E[X_{2i}] &= \pi\kappa + (1-\pi)\mu \end{aligned}$$

so

$$E[X_{1i} + X_{2i}] - \kappa = \pi(\mu+\kappa) + (1-\pi)(\kappa+\mu) - \kappa = \mu$$

It therefore follows that

$$\frac{1}{n}\sum_{i=1}^{n}(X_{1i} + X_{2i}) - \kappa$$

is a consistent estimator of μ.

More generally consider an estimator of the form

$$\widehat{\mu} = \frac{a_1}{n}\sum_{i=1}^{n} X_{1i} + \frac{a_2}{n}\sum_{i=1}^{n} X_{2i} - a_3\kappa \tag{1}$$

Its mean would be

$$\begin{aligned} & E[\widehat{\mu}] = \pi(a_1\mu + a_2\kappa) + (1-\pi)(a_1\kappa + a_2\mu) - a_3\kappa \qquad (2) \\ = \; & (\pi a_1 + (1-\pi)a_2)\mu + (\pi a_2 + (1-\pi)a_1 - a_3)\kappa \end{aligned}$$

For unbiasedness, we then need

$$(\pi a_1 + (1-\pi)a_2) = 1 \tag{3}$$

or

$$a_2 = \frac{1-\pi a_1}{1-\pi} = \frac{1}{1-\pi} - \frac{\pi}{1-\pi}a_1$$

and

$$(\pi a_2 + (1-\pi)a_1 - a_3) = 0 \tag{4}$$

The only way to do this without knowing π is to set $a_1 = a_2$. But in that case (4) inplies that $a_1 = a_2 = 1$.

If we know π then (3 and (4) can be solved for a_2 and a_3 as a function of a_1:

[here insert numerical example]

True p indicates the probability that X_1 is drawn from the correct distribution.

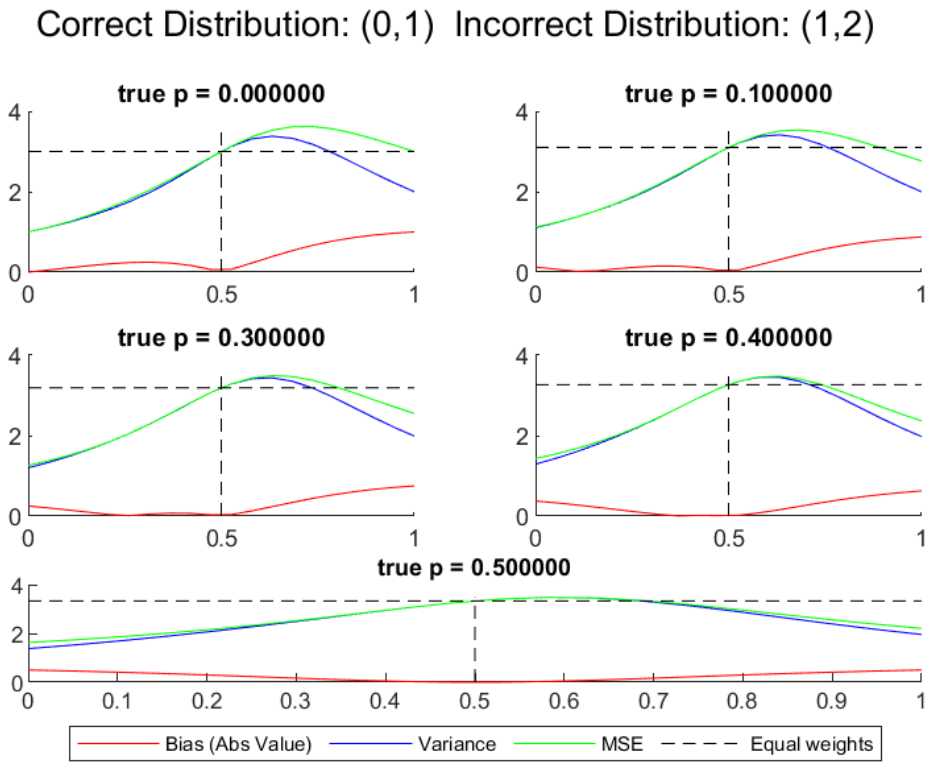

Figure 1: Bias-variance tradeoff implied by different beliefs about π

1.2 Results

The minimum variance estimator is achieved by placing all weight on the $X_{i\ell}$ with lowest variance. The bias function is quadratic in π, with zeros achieved at $\hat{\pi} = \pi_0$ and $\hat{\pi} = 0.5$. The minimum MSE estimator depends on the choice of correct and incorrect distributions.

With two observations, everything is symmetric so we only need to look at values between 0 and 0.5.