

ECON 6200
Problem Set 2

Gabe Sekeres

March 5, 2025

1. We have that $Y \sim U[a, b]$

(a) Recall that the OLS estimator in the simple mean case is defined as

$$\hat{\mu}_{OLS} \in \underset{m}{\operatorname{argmin}} (Y - m)^2$$

Since Y is scalar-valued and i.i.d., we can consider a single scalar case in which

$$\hat{\mu}_{OLS} \in \underset{m}{\operatorname{argmin}} (Y_i - m)^2$$

which admits the first order condition on m

$$2(Y_i - m) = 0 \implies \frac{1}{n} \sum_{i=1}^n m = \frac{1}{n} \sum_{i=1}^n Y_i \implies \hat{\mu}_{OLS} = \bar{Y}$$

(b) First, to show that $\hat{\mu}_{\text{order}}$ is unbiased, it suffices to show that $\mathbb{E}[\hat{\mu}_{\text{order}}] = \mathbb{E}[Y] = \mu$. Notice that from the moments of order statistics,

$$\mathbb{E}[Y_1^*] = a + \frac{b-a}{n+1} \quad \text{and} \quad \mathbb{E}[Y_n^*] = b - \frac{b-a}{n+1}$$

Thus, we have that

$$\mathbb{E}[\hat{\mu}_{\text{order}}] = \frac{\mathbb{E}[Y_1^* + Y_n^*]}{2} = \frac{1}{2} \mathbb{E} \left[a + \frac{b-a}{n+1} + b - \frac{b-a}{n+1} \right] = \frac{b+a}{2} = \mathbb{E}[Y]$$

from the properties of uniform distributions. Next, recall that the variance of a Y is

$$\sigma^2 = \frac{(a-b)^2}{12}$$

This means that

$$\operatorname{Var}(\bar{Y}) = \operatorname{Var} \left(\frac{1}{n} \sum_{i=1}^n Y_i \right) \stackrel{\text{i.i.d.}}{=} \frac{1}{n^2} n \sigma^2 = \frac{(a-b)^2}{12n}$$

Noting that the sample minimum and maximum are each order statistics, and using the properties of the 1st and n th order statistics, we have that

$$\begin{aligned} \operatorname{Var}(Y_1^*) &= \frac{n(b-a)^2}{(n+1)^2(n+2)} \\ \operatorname{Var}(Y_n^*) &= \frac{n(b-a)^2}{(n+1)^2(n+2)} \\ \operatorname{cov}(Y_1^*, Y_n^*) &= \frac{(b-a)^2}{(n+1)^2(n+2)} \end{aligned}$$

Then, we have directly that

$$\text{Var}(\hat{\mu}_{\text{order}}) = \frac{\text{Var}(Y_1^*) + \text{Var}(Y_n^*) + 2 \text{cov}(Y_1^*, Y_n^*)}{4} = \frac{(b-a)^2}{2(n+1)(n+2)}$$

and for all positive integers n ,

$$2(n+1)(n+2) = 2n^2 + 6n + 4 \geq 12n \iff (n-2)(n-1) \geq 0$$

so this estimator has weakly lower variance, and has strictly lower variance for any $n \geq 3$.

- (c) The Gauss-Markov Theorem *does* apply to \bar{Y} , as the assumptions (linearity, unbiasedness, strong exogeneity, and the rank condition) hold. However, though it would seem that we had found an unbiased linear estimator with lower variance than the BLUE, we actually haven't, since $\hat{\mu}_{\text{order}}$ is *not* linear in Y . This can be clearly seen by the fact that the covariance term of its variance is positive, meaning that as Y_1^* increases, Y_n^* increases, and vice versa, so a linear transformation of the terms of $\hat{\mu}_{\text{order}}$ will have an oversized effect on $\hat{\mu}_{\text{order}}$, so it is not linear.

2. Consider simple linear regression both without a constant

$$Y = \beta \cdot X + \varepsilon$$

and with a constant

$$Y = \alpha + \beta \cdot X + \varepsilon$$

- (a) Recall that $\tilde{\beta}$ solves the least-squares minimization problem:

$$\tilde{\beta} \in \underset{b}{\text{argmin}} (Y - Xb)^2$$

which admits the first order condition

$$-2X'Y + 2(X'X)b = 0$$

And in finite sample, this becomes

$$\tilde{\beta} = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}$$

- (b) Recall from lecture the analogous closed form of $\hat{\beta}$:

$$\hat{\beta} = \frac{\sum_{i=1}^n (X_i - \bar{X}) Y_i}{\sum_{i=1}^n (X_i - \bar{X}) X_i}$$

So, as always, we have that (from exogeneity)

$$\mathbb{E}[\hat{\beta}] = \mathbb{E} \left[\frac{\sum_{i=1}^n (X_i - \bar{X})(X_i \beta + \varepsilon)}{\sum_{i=1}^n (X_i - \bar{X}) X_i} \right] = \beta + \underbrace{\mathbb{E} \left[\frac{\sum_{i=1}^n X_i \varepsilon}{\sum_{i=1}^n (X_i - \bar{X})^2} \right]}_{=0} = \beta$$

and

$$\text{Var}(\hat{\beta}) = \text{Var} \left(\beta + \frac{\sum_{i=1}^n X_i \varepsilon}{\sum_{i=1}^n (X_i - \bar{X}) X_i} \right) = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

Using our closed form for $\tilde{\beta}$, we get that (from exogeneity)

$$\mathbb{E}[\tilde{\beta}] = \mathbb{E} \left[\frac{\sum_{i=1}^n X_i (X_i \beta + \varepsilon)}{\sum_{i=1}^n X_i^2} \right] = \beta + \underbrace{\mathbb{E} \left[\frac{\sum_{i=1}^n X_i \varepsilon}{\sum_{i=1}^n X_i^2} \right]}_{=0} = \beta$$

and

$$\text{Var}(\tilde{\beta}) = \text{Var}\left(\frac{\sum_{i=1}^n X_i \varepsilon}{\sum_{i=1}^n X_i^2}\right) = \frac{\sigma^2}{\sum_{i=1}^n X_i^2}$$

So the variance is weakly smaller, strictly as long as $\bar{X} \neq 0$.

(c) First, assume that the true model is

$$Y = \alpha + \beta X + \varepsilon$$

but that we use the regression

$$Y = \beta X + \varepsilon$$

As we saw above, our slope coefficient is

$$\tilde{\beta} = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2} = \frac{\sum_{i=1}^n X_i (\alpha + \beta X_i + \varepsilon)}{\sum_{i=1}^n X_i^2} = \beta + \frac{\alpha \sum_{i=1}^n X_i}{\sum_{i=1}^n X_i^2} + \frac{\sum_{i=1}^n X_i \varepsilon}{\sum_{i=1}^n X_i^2}$$

so we have that

$$\mathbb{E}[\tilde{\beta} \mid X] = \mathbb{E}\left[\beta + \frac{\alpha \sum_{i=1}^n X_i}{\sum_{i=1}^n X_i^2} + \frac{\sum_{i=1}^n X_i \varepsilon}{\sum_{i=1}^n X_i^2}\right] = \beta + \alpha \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n X_i^2}$$

so $\tilde{\beta}$ is unbiased only if $\sum_{i=1}^n X_i = 0$, the class of models in which X is already demeaned. This allows us to state a modification of Gauss-Markov:

Theorem 1. *If the previous assumptions (linearity, strong exogeneity, rank condition, and spherical errors) hold, and $\tilde{\beta}$ is a linear unbiased estimator for a model where $\alpha = 0$, then*

$$\text{Var}(\tilde{\beta} \mid X) \geq \sigma^2 (X'X)^{-1}$$

Proof. Assume that $\tilde{\beta} = CY$ for some matrix C , and since $\tilde{\beta}$ is unbiased it must be that $CX = I$. Thus, we write

$$C = (X'X)^{-1}X' + D \quad \text{where } D = C - (X'X)^{-1}X'$$

As in class, unbiasedness requires that

$$\mathbb{E}[(X'X)^{-1}X' + D)Y \mid X] = \beta \implies \mathbb{E}[DX\beta \mid X] = 0 \implies DX = 0$$

Thus, we have that

$$\begin{aligned} \text{Var}(\tilde{\beta} \mid X) &= \text{Var}(((X'X)^{-1}X' + D)Y \mid X) \\ &= \text{Var}(((X'X)^{-1}X' + D)(X\beta + \varepsilon) \mid X) \\ &= \text{Var}(((X'X)^{-1}X' + D)\varepsilon \mid X) \\ &= \sigma^2((X'X)^{-1}X' + D)((X'X)^{-1}X' + D)' \\ &= \sigma^2((X'X)^{-1}X'X(X'X)^{-1} + \underbrace{DX(X'X)^{-1}}_{=0} + \underbrace{(X'X)^{-1}D'X'}_{=0} + D'D) \\ &\geq \sigma^2(X'X)^{-1} \end{aligned}$$

where conclusion follows from the fact that $D'D$ is positive semi-definite. □

3. Estimate the efficacy of a vaccine.

- (a) Our estimator is the simple sample average:

$$\hat{\pi} = \frac{9}{178} \approx 0.05056$$

A Wald 95% confidence interval is constructed with the bounds

$$\hat{\pi} \pm z_{0.975} \sqrt{\frac{\hat{\pi}(1 - \hat{\pi})}{n}}$$

so we have the approximate confidence interval

$$\pi \in (\hat{\pi} - 0.0322, \hat{\pi} + 0.0322) \approx (0.0184, 0.0827)$$

We can alternatively construct a Poisson approximation by taking $\lambda = 178\pi$ and using test inversion on the Poisson distribution where $k = X = 9$. Then, the formula for our (Poisson) confidence interval is

$$\lambda \in \left(\frac{1}{2} \chi_{18,0.025}^2, \frac{1}{2} \chi_{20,0.975}^2 \right) \approx (4.12, 17.09)$$

and since $\lambda = 178\pi$, we have the confidence interval

$$\pi \in (0.0231, 0.0960)$$

Finally, using the relationship between the F distribution and the binomial distribution (which I googled), we can construct an exact inversion of the binomial distribution, and get a confidence interval defined by, for $k = 9$, $n = 178$, and $\alpha = 0.05$,

$$\left(\frac{k}{k + (n - k + 1)F_{1-\alpha/2;2(n-k+1),2k}}, \frac{(k + 1)F_{1-\alpha/2;2(k+1),2(n-k)}}{(n - k) + (k + 1)F_{1-\alpha/2;2(k+1),2(n-k)}} \right)$$

We have that, from a distribution table,

$$F_{0.975;340,18} \approx 2.2117$$

$$F_{0.975;20,338} \approx 1.7495$$

so this becomes

$$\left(\frac{9}{9 + 170 \cdot 2.2117}, \frac{10 \cdot 1.7495}{169 + 10 \cdot 1.7495} \right) \approx (0.0234, 0.0938)$$

In the last two, we are essentially finding $\underline{\lambda}, \bar{\lambda}$ such that when the Poisson distribution takes a value $k = 9$, we are sure with 95% confidence that the true λ is within those values. We are doing the same with the binomial inversion – finding $\underline{\pi}, \bar{\pi}$ such that since we observed 9 successes of 178 trials, we are sure with 95% confidence that π is within those bounds.

- (b) **Proof.** We have that $[\underline{\theta}, \bar{\theta}]$ is a valid confidence interval for θ and that $f(\cdot)$ is a known monotonic function. WLOG, assume that f is monotonically increasing. We know that since this confidence interval is valid,

$$\mathbb{P}\{\theta \in [\underline{\theta}, \bar{\theta}]\} = 1 - \alpha \implies \mathbb{P}\{\underline{\theta} \leq \theta \leq \bar{\theta}\} = 1 - \alpha$$

Since f is monotonically increasing, $f(x) \leq f(y) \implies x \leq y$. Thus,

$$\underline{\theta} \leq \theta \leq \bar{\theta} \iff f(\underline{\theta}) \leq f(\theta) \leq f(\bar{\theta})$$

so we have that

$$\mathbb{P}\{f(\theta) \in [f(\underline{\theta}), f(\bar{\theta})]\} = 1 - \alpha$$

so $[f(\underline{\theta}), f(\bar{\theta})]$ is a $1 - \alpha$ confidence interval for $f(\theta)$. The same logic applies, reversing the inequalities, if f is monotonically decreasing. \square

(c) If we know $\rho = \mathbb{P}\{\text{vaccinated}\}$, then by Bayes' Theorem, we have that

$$VE = 1 - \frac{\frac{\pi \cdot \mathbb{P}\{\text{infected}\}}{\rho}}{\frac{(1-\pi) \cdot \mathbb{P}\{\text{infected}\}}{1-\rho}} = 1 - \frac{\pi(1-\rho)}{(1-\pi)\rho}$$

so

$$\frac{\partial}{\partial \pi} VE = \frac{\rho-1}{\rho} \frac{1}{(1-\pi)^2}$$

and VE is strictly monotonic since for $\pi, \rho \in (0, 1)$ the partial with respect to π is strictly negative.

(d) We have that

$$\rho = \frac{18,559}{18,559 + 18,708} \approx 0.498$$

Thus,

$$VE = 1 - \frac{0.502 \cdot \pi}{0.498 \cdot (1-\pi)}$$

which is a strictly monotonic (decreasing) function of π . Thus, we have that our estimator is

$$\hat{VE} = 1 - \frac{0.502 \cdot \hat{\pi}}{0.498 \cdot (1-\hat{\pi})} \approx 0.946$$

From part (b), we have that we can simply transform our confidence intervals using the monotonically decreasing function

$$f(x) = 1 - \frac{0.502 \cdot x}{0.498 \cdot (1-x)}$$

We get that the Wald confidence interval is

$$(f(0.0827), f(0.0184)) \approx (0.9091, 0.9811)$$

The Poisson approximation confidence interval is

$$(f(0.0960), f(0.0231)) \approx (0.8930, 0.9762)$$

Finally, the binomial confidence interval is

$$(f(0.0938), f(0.0234)) \approx (0.8957, 0.9758)$$

(e) Yes! We can see the binomial confidence interval in the overall vaccine efficacy line, and their estimator has the same mean as ours. They cite the exact same Clopper and Pearson method I used!