

ECON 6200
Problem Set 5

Gabe Sekeres

April 30, 2025

1. Extremum Estimation

- (a) Here, we do not assume that θ_0 is a well-separated optimum, and we do not assume uniform convergence. We do assume (above the class assumptions about consistency) that m is strictly convex. Strict convexity does imply the above, as combined with continuity we will have directly that the minimum is well-separated and that the sample m_n converge uniformly to m as long as they converge pointwise.
- (b) **Proof.** It suffices to show that

$$\left| \frac{1}{n} \sum_i m(W_i, \theta) - \mathbb{E}m(W, \theta) \right| \xrightarrow{p} 0 \quad \forall \theta \in \mathbb{R}$$

Fix some $\theta \in \mathbb{R}$, and recall that we are assuming that $\mathbb{E}m(W, \theta) \in \mathbb{R}$. If m is continuous in some open neighborhood of θ , conclusion follows immediately from the continuous mapping theorem. Therefore, assume that m has a discontinuity at exactly θ . Since m is strictly convex, it must be that $m(W, \theta) = \lim_{W_n \searrow W} m(W_n, \theta)$, i.e. as W is approached from above. Conclusion then directly follows from Jensen's Inequality, since $\frac{1}{n} \sum m(W_i, \theta) \geq \mathbb{E}m(W, \theta)$ almost surely as n increases, and so is approaching from above where the discontinuity will have no bite. \square

2. Worked Example with Geometric Distribution

- (a) The random variable X will equal x in precisely the case where there are $x - 1$ failures followed by one success. Since the trials are i.i.d., we have immediately that

$$\mathbb{P}(X = x) = (1 - \pi)^{x-1} \pi$$

Since the geometric distribution is memoryless, we have that

$$\begin{aligned} \mathbb{E}X &= \pi \cdot \mathbb{E}[X \mid X = 1] + (1 - \pi) \cdot \mathbb{E}[X \mid X > 1] = \pi + (1 - \pi) \cdot (\mathbb{E}X + 1) \\ \implies \mathbb{E}X &= \frac{1}{\pi} \end{aligned}$$

We have that

$$\begin{aligned} \mathbb{E}X^2 &= \pi + (1 - \pi)(\mathbb{E}(X + 1)^2) = \pi + (1 - \pi)(\mathbb{E}X^2 + 2\mathbb{E}X + 1) \\ \implies \mathbb{E}X^2 &= \frac{2(1 - \pi)}{\pi^2} + \frac{1}{\pi} \end{aligned}$$

So

$$\text{Var}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{2(1 - \pi)}{\pi^2} + \frac{1}{\pi} - \frac{1}{\pi^2} = \frac{1 - \pi}{\pi^2}$$

- (b) Given data $\{x_i\}$, the likelihood function is

$$L_n(\pi) = \prod_{i=1}^n (1 - \pi)^{x_i-1} \pi$$

so the log likelihood function is

$$\ell_n(\pi) = n \log \pi + \sum_{i=1}^n (x_i - 1) \log(1 - \pi)$$

Thus,

$$\hat{\pi}_{MLE} = \operatorname{argmax}_{\pi \in [0,1]} n \log \pi + \sum_{i=1}^n (x_i - 1) \log(1 - \pi)$$

Since this is concave by inspection, the maximum is attained and precisely characterized by the first order condition:

$$\frac{n}{\pi} - \frac{\sum_{i=1}^n (x_i - 1)}{1 - \pi} = 0 \implies \hat{\pi}_{MLE} = \frac{1}{\bar{X}}$$

We can use the delta method which will apply as long as $\pi \neq 0$. We have that

$$\sqrt{n}(\bar{X} - \mathbb{E}X) \xrightarrow{d} \mathcal{N}(0, \operatorname{Var}(X)) = \mathcal{N}\left(0, \frac{1 - \pi}{\pi^2}\right)$$

So using $g(x) = \frac{1}{x}$, which admits $g'(x) = -x^{-2}$ we have that

$$\sqrt{n}(\hat{\pi}_{MLE} - \pi) = \sqrt{n}(g(\bar{X}) - g(\mathbb{E}X)) \xrightarrow{d} \mathcal{N}\left(0, \frac{1 - \pi}{\pi^2} \cdot (-\pi^2)^2\right) = \mathcal{N}(0, \pi^2(1 - \pi))$$

- (c) This estimator is not unbiased. The function $\frac{1}{x}$ is convex for all positive x , so from Jensen's Inequality we have that $\mathbb{E}[\hat{\pi}_{MLE}] \geq \pi$, where strict inequality holds except when $\pi = 1$. Thus, the estimator is biased upwards.

3. Uniform Distribution

- (a) We will use the simple moment function $g(x, \mu) = x - \mu$. Then, we have that $\mathbb{E}g(X, \mu) = \mathbb{E}X - \mu$, which equals zero only when $\mu = \mathbb{E}X = \mu_0$. The sample moment function is directly $g_n(\mu) = \frac{1}{n} \sum_{i=1}^n g(X_i, \mu) = \bar{x}_n - \mu$ so the GMM estimator is clearly the sample average. We can characterize the asymptotics using the central limit theorem:

$$\sqrt{n}(\hat{\mu} - \mu_0) \xrightarrow{d} \mathcal{N}(0, \operatorname{Var}(X)) = \mathcal{N}\left(0, \frac{(\beta - \alpha)^2}{12}\right)$$

- (b) We first verify the assumptions. (1) follows from the law of large numbers, (2) we have that $W = \hat{W} = 1$, which is trivially symmetric and positive definite, (3) we have that $\mu_0 \in [\alpha, \beta] \subset \operatorname{int}(\mathbb{R})$, (4) we have that since $g(\cdot)$ is linear in μ , $g(x, \mu) \in \mathbf{C}^\infty$ for any $\mu \in \mathbb{R}$, including μ_0 . For (5) we have from the central limit theorem that $S = \frac{(\beta - \alpha)^2}{12}$, which is positive. For (6), take a sequence $\mu_n \rightarrow \mu_0$, and we have that $\frac{\partial g(x, \mu_n)}{\partial \mu} = -1 \xrightarrow{p} -1 = G(\mu_0) = \mathbb{E} \frac{\partial g(x, \mu_0)}{\partial \mu}$, and (7) we know that singletons always have full column rank. Thus, the assumptions hold, and we have that $W = 1$, $S = \frac{(\beta - \alpha)^2}{12}$, and $G = -1$.
- (c) Recall that the density of a uniform distribution is $\mathbb{1}_{x \in [\alpha, \beta]} \frac{1}{\beta - \alpha}$. The likelihood function with n i.i.d. observations is therefore

$$L_n(\alpha, \beta) = \prod_{i=1}^n \frac{1}{\beta - \alpha} \mathbb{1}\{x_i \in [\alpha, \beta]\} = (\beta - \alpha)^{-n} \mathbb{1}\{\alpha \leq x_{[1]} \leq x_{[n]} \leq \beta\}$$

Observe that maximizing the likelihood is equivalent to minimizing the length of $\beta - \alpha$ subject to $\beta \geq x_{[n]}$ and $\alpha \leq x_{[1]}$. Thus, clearly the likelihood is maximized at equality, so

$$\hat{\alpha}_{ML} = x_{[1]} \quad ; \quad \hat{\beta}_{ML} = x_{[n]}$$

- (d) We know from above that $\mathbb{E}\hat{\mu} = \mathbb{E}\frac{1}{n}\sum_{i=1}^n x_i = \frac{n}{n}\mathbb{E}\bar{x}_n = \mu_0$, so $\hat{\mu}$ is unbiased. We also have that

$$\text{Var}(\hat{\mu}) = \frac{1}{n^2}n \text{Var}(X) = \frac{(\beta - \alpha)^2}{12n}$$

So the mean squared error is:

$$\text{MSE}(\hat{\mu}) = \text{bias}(\hat{\mu})^2 + \text{Var}(\hat{\mu}) = \frac{(\beta - \alpha)^2}{12n}$$

We have that

$$\mathbb{E}\tilde{\mu} = \frac{\mathbb{E}x_{[1]} + \mathbb{E}x_{[n]}}{2} = \frac{\alpha + \beta}{2} = \mu_0$$

so $\tilde{\mu}$ is unbiased. We also have that

$$\text{Var}(\tilde{\mu}) = \frac{1}{4} (\text{Var}(x_{[1]}) + \text{Var}(x_{[n]})) = \frac{n(\beta - \alpha)^2}{2(n+1)^2(n+2)}$$

So the mean squared error is:

$$\text{MSE}(\tilde{\mu}) = \text{bias}(\tilde{\mu})^2 + \text{Var}(\tilde{\mu}) = \frac{n(\beta - \alpha)^2}{2(n+1)^2(n+2)}$$

- (e) The Gauss-Markov Theorem does apply to $\hat{\mu}$, and $\hat{\mu}$ does attain the BLUE variance. However, $\tilde{\mu}$ is not a linear estimator, since there is correlation between $\hat{\alpha}_{ML}$ and $\hat{\beta}_{ML}$. It is unbiased and (asymptotically) attains a lower variance, but is not linear.
- (f) The corresponding population criterion is

$$Q(\mu) = (\beta - \mu)^2 + (\mu - \alpha)^2$$

The consistency theorem for convex Q applies here, since this function is strictly convex in the argument, μ_0 is in the interior and uniquely minimizes Q , the sample criterion functions are convex, and Q_n converges pointwise to Q for any μ .

- (g) Note that

$$\tilde{\mu} - \mu_0 = \frac{\hat{\alpha}_{ML} - \alpha + \hat{\beta}_{ML} - \beta}{2}$$

Thus, since $\sqrt{n}(\hat{\alpha}_{ML} - \alpha)$ and $\sqrt{n}(\hat{\beta}_{ML} - \beta)$ are both asymptotically normal, $\sqrt{n}(\tilde{\mu} - \mu_0)$ is asymptotically the composition of normals, which since we have correlation is not asymptotically normal. I can go no further than this.

- (h) The theorem is not contradicted because the Hessian is not nonsingular everywhere, and we can improve asymptotically when $\mu_0 = 0$.

4. GMM versus ML in a Classic Application

- (a) They say that GMM is advantaged over ML because the moment functions are functions of true preference parameters α and β . ML would require those to be functions of policy, while Lucas says that such parameters are policy-invariant.
- (b) The moment conditions are that there is no improvement to be made in expectation, basically that markets will clear and there is no arbitrage or other free lunch.
- (c) b_T^* is optimal in the sense of all estimators based on the specific moment conditions in this problem, but does not attain Cramér-Rao efficiency or other general bounds.

- (d) We could match the efficiency of ML, as long as we made all of the distributional assumptions that the authors don't want to. In that case, we would be numerically identical to the ML estimator. However, we would inherit exactly ML's disadvantages from part (a).
- (e) α is the coefficient of relative risk aversion and β is the discount factor. The estimates seem relatively plausible for α and β both – people are relatively risk-averse and relatively patient. The variation between specifications makes sense to me – α is very sensitive to specification, but β barely changes at all. That intuitively holds.