ARE 212 - Assignment 1

Thor Larson

April 4, 2023

1 Admin

This document and affiliated code can be found here.

2 Exercises

2.1

Define a 2x1 Z with

$$Z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} x + y \\ y \end{bmatrix} = g \begin{pmatrix} x \\ y \end{pmatrix} \tag{1}$$

Then we can define g^{-1} as follows:

$$g^{-1}(Z) = \begin{bmatrix} z_1 - z_2 \\ z_2 \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} \tag{2}$$

Observe the Jacobian of g^{-1} is $\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$ which has determinant 1.

So $f_Z(Z) = f(x, y) \cdot 1 = f_X(x) f_Y(y)$ since $X \perp\!\!\!\perp Y$.

Then

$$f_z(z) = \int_{-\infty}^{\infty} f(z_1, z_2) dz_2 = \int_{-\infty}^{\infty} f_X(x) f_Y(y) \cdot \mathbb{1}(x + y = z) dy = \int_{-\infty}^{\infty} f_X(z - y) f_Y(y) dy.$$

But this is just a convolution; $\boxed{f_Z = f_X * f_Y}$.

2.2

Let s have finite support $\{s_1, s_2 \ldots s_n\}$ with probability weights $\{p(s_i)\}_1^n$. We can rewrite the density of s as

$$f_s(s) = \sum_{i=1}^{n} p(s_i) \delta(s - s_i)$$

where δ is Dirac's delta function.

Then,

$$f_x * f_s = \sum_{i=1}^{n} p(s_i) f_x * \delta(s - s_i) = \sum_{i=1}^{n} p(s_i) f_x(x - s_i)$$

Since f_x is continuous, a finite sum of shifted f_x is also continuous. Thus $x + s$ has a continuous distribution.

2.3

The generalized inverse condition is

$$\mathbf{0} A^- \mathbf{0} = \mathbf{0}$$

It is clear that *any* $n \ x \ m$ matrix will satisfy this.

2.4

(a) Denote random variables as y, X

(b) Denote r.v. realizations as y_i, X_i

(c) Denote arrays of realizations in bold as $\boldsymbol{y}, \boldsymbol{X}$

2.5

2.5.1

$$\boldsymbol{A}^+ = \boldsymbol{A}^+\boldsymbol{0}\boldsymbol{A}^+ = \boldsymbol{A}^+(\boldsymbol{0}\boldsymbol{A}^+) = \boldsymbol{A}^+\boldsymbol{0} = \boxed{\boldsymbol{0}}.$$

2.5.2

$X^+X = (X^\top X)^{-1}(X^\top X) = I$ since X has full column rank.

- $XX^+X = XI = X.$ ✓

- $X^+XX^+ = IX^+ = X^+.$ ✓

- $(X^+X)^\top = I^\top = I = (X^+X).$ ✓

- $(XX^+)^\top = (X(X^\top X)^{-1}X^\top)^\top = (X^\top)^\top(X^\top(X^\top)^\top)^{-1}X^\top = X(X^\top X)^{-1}X^\top = XX^+.$ ✓

So X^+ is a Moore-Penrose inverse for X. Thus, it is also the unique inverse.

2.5.3

Since $X^\top u = 0, X^+u = 0.$
So $X^+y = X^+Xb + X^+u = b.$
$\boxed{X^+y = b.}$

3 Convolutions

3.1

See Section 2.2

3.2

Jupyter Notebook

3.3

We can't approximate the convolution of two continuous distributions exactly unless we solve the equation mathematically. This is because both distributions have infinitely many points in their support to iterate over, making a for-loop impossible.

4 General Weighted Linear Regressions

- OLS: $T = I$

- GLS: $T = C^\top$, where $C^\top C = \Omega^{-1}(X).$

- IV: $T = Z$, where Z is a vector of instruments for X.

- 2SLS: $T = I$ for both stages of the regression.

- We cannot express logit/probit as weighted linear regressions (although they can be expressed as GLM models)

5 Simultaneous Equations

5.1

- $X : N \times m$

- $\beta : m \times k$

- $T : N \times k$

- $u : N \times k$

5.2

Yes! We can think of this procedure as running k independent regressions of y_i^{Nx1} on X_i^{Nx1} to generate k unique $\{\beta_i\}_{i=1}^k$. Matrix operations enable us to solve these k equations simultaneously, so the estimator will still work. However, we will need to take care when calculating standard errors, since we assume the y_i are independent.

5.3

See the workbook here for the case where $m = 2$.

5.4

- $N \geq m$ so that $T^\top X$ is full column rank

- $E(T'u) = \mathbf{0}_{kx1}$

- Each of the y_i could be estimated independently - i.e., variance matrix $cov(u|X)$ is diagonal. This means that error terms u_i are uncorrelated.

6 SUR

6.1

Any dependence between y_1 and y_2 would stem from their unobserved errors being correlated.

6.2

See the workbook here for the case where $m = 2$.

6.3

The estimates from this procedure will be the same as before; however, the standard errors will differ. This is because we've used a feasible GLS procedure to correct for sample heteroskedasticity.
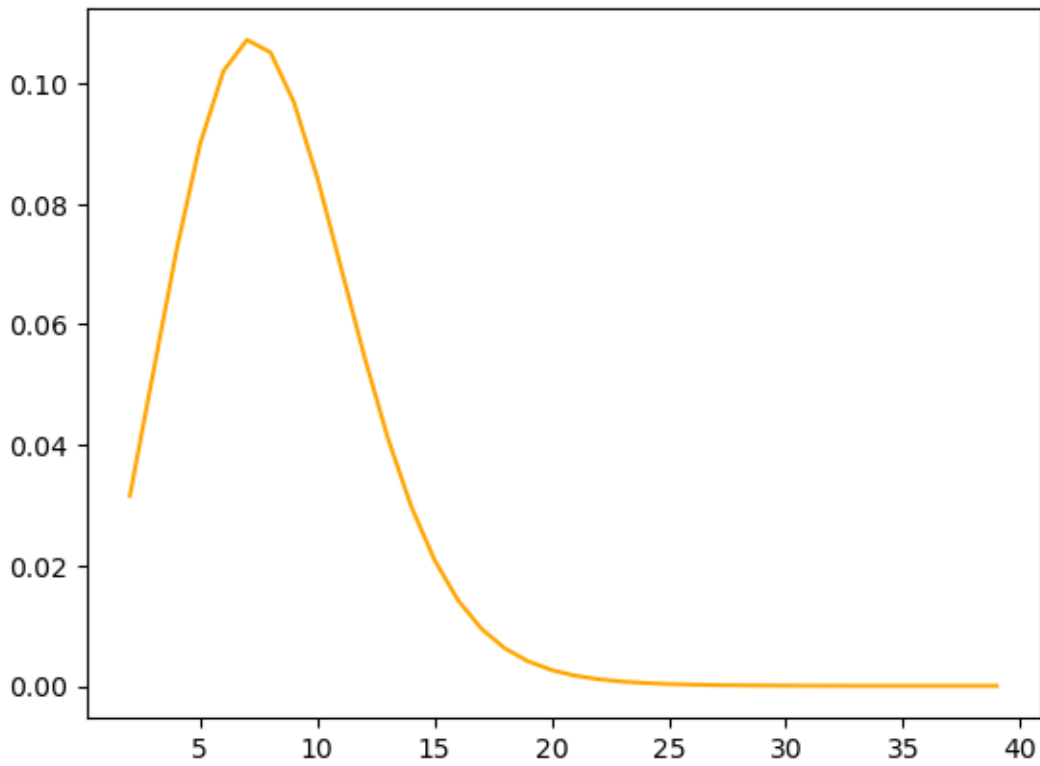
7

7.1

See the workbook here. A Gaussian kernel with $h = 4$ seems to best fit the data - any smaller, and the the peakiness at low-N high household sizes skews the distribution. As h increases, the distribution becomes too smooth at low household sizes.
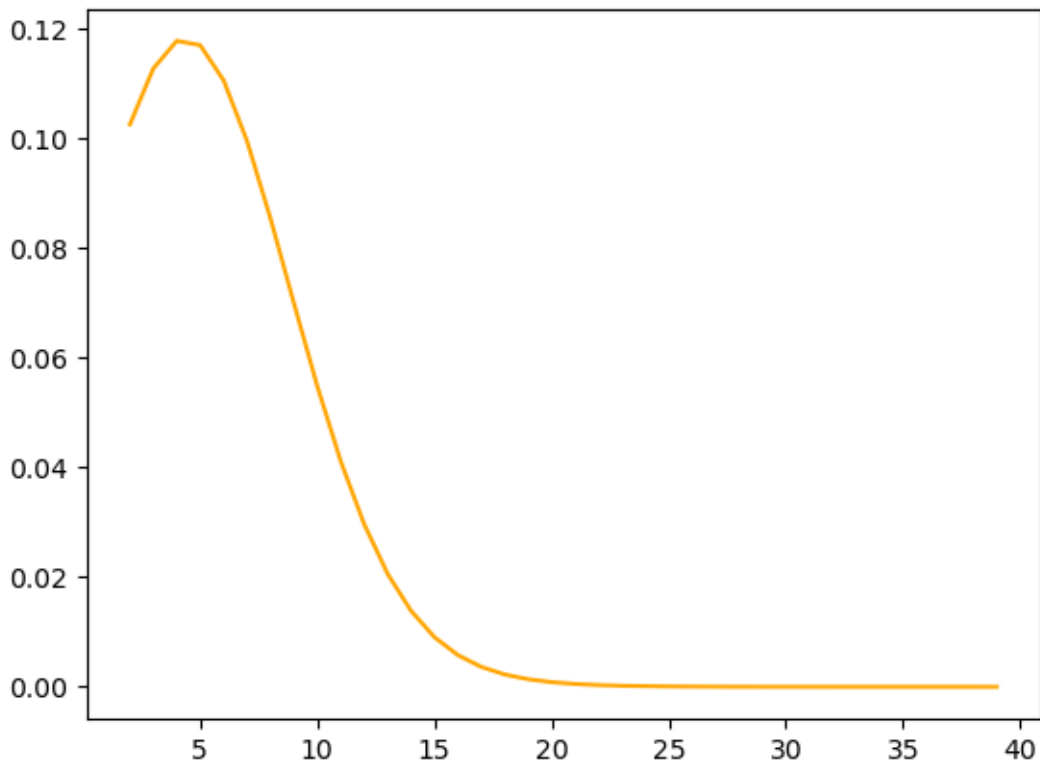
7.2

For $Y = \log(X)$, change of variables yields:

$$f_Y(y) = f_X(x) \cdot e^y = f_X(x) \cdot x$$

7.3

We observe the center of mass is shifted further left when we take the log first. This approach means the kernel is stretched over a larger (and nonlinear) range of x when we calculate the logarithm.

8

The expression for bias would become

$$Bias \equiv E(\hat{f}) - \hat{f}_n = E\left(\frac{1}{h}k\left(\frac{X-x}{h}\right)\right) - \frac{1}{nh}\sum_{1}^{n}k\left(\frac{X-x}{h}\right)$$

As $n \to \infty$, $\hat{f}_n \to E(\hat{f})$ and bias necessarily goes to 0.

Thus, the expression is not inherently useful for large samples since it tells us nothing about whether our sample is close to the original distribution.

However, a strongly non-zero bias *is* informative: it tells us that our choice of n is insufficiently small to account for the selected h. Thus, we would need to decrease our bandwidth or increase sample size.