Basic Stats/Probability

Some definitions for everything below:

- X, Y, and Z are random variables
- a, b, c, d are constants
- Whenever you see μ_X or σ_X^2, know that this is the mean and variance (respectively) of the random variable X
- $\overline{X}_N$ is a sample mean from a sample of random variables X_n drawn from the same distribution.
- s_X^2 is the sample variance calculated from a sample of X.

Most of the rules follow almost immediately from the basic definitions, so you should test yourself by trying to prove each of them. It will also help you a lot if you know these rules like the back of your hand.

- $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$
- $\mathbb{C}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbb{C}(Y, X)$
- $\mathbb{C}(a + bX, Y) = b\mathbb{C}(X, Y)$

  $$\Rightarrow \mathbb{C}(a + bX, c + dY) = bd\mathbb{C}(X, Y).$$
- $\mathbb{C}(aX + bY, Z) = a\mathbb{C}(X, Z) + b\mathbb{C}(Y, Z)$
- $X \perp\!\!\!\perp Y \Rightarrow \mathbb{C}(X, Y) = 0$
- $\mathbb{V}[X + Y] = X + Y + 2\mathbb{C}(X, Y)$
- $\Rightarrow\ \mathbb{V}\left[\sum_n X_n\right] = \sum_n \left[X_n\right]$ if $X_1 \perp\!\!\!\perp X_2 ... \perp\!\!\!\perp X_N$
- $\mathbb{E}[\overline{X}_N] = \mu_X$
- $\mathbb{V}[\overline{X}_N] = \frac{1}{N}\mathbb{V}[X]$ if $X_1 \perp\!\!\!\perp X_2 ... \perp\!\!\!\perp X_N$ (i.e. iid sample)
- $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$ (Law of Iterated Expectations)
- $\mathbb{E}[X|Y] = 0\ \Rightarrow \mathbb{E}[XY] = 0$
- $\mathbb{E}[X|Y] = a\ \Rightarrow \mathbb{C}(X, Y) = 0$
- If $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ and $Z = X + Y$, then $Z \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Z^2 + 2\sigma_{XY})$ where $\sigma_{XY} = \mathbb{C}(X, Y)$.
- When $\overline{X}_N$ comes from an iid sample and $\mathbb{V}[X] < \infty$, $\lim_{N\to\infty} P\left[\frac{\overline{X}_N - \mu_X}{s_X} \leq z_\alpha\right] = 1 - \alpha$ where z_α solves $P[Z \leq z_\alpha] = 1 - \alpha$ where $Z \sim \mathcal{N}(0, 1)$.

Now, let X be a random column vector, and let $X_1, X_2, ..., X_N$ be an iid sample of size N. Let $\mu_X = \mathbb{E}[X]$ and $\Sigma_X = \mathbb{V}[X] = \mathbb{E}[(X - \mu_X)(X - \mu_X)^T]$. Let C be a constant matrix.

- $\mathbb{V}[CX] = C\Sigma_X C^T$.

- $\overline{X} \to_p \mu_X$ (law of large numbers)
- $\sqrt{N}(\overline{X} - \mu_X) \to_d \mathcal{N}(0, \Sigma_X)$ (central limit theorem)
- If $\sqrt{N}(\hat{\beta} - \beta) \to_d \mathcal{N}(0, \Sigma)$ and $H_0 : R\beta - c = 0$ then we can test H_0 with the Wald or Chi squared test statistic which follows under the null:

$$(R\hat{\beta} - c)^T (R\hat{\Sigma}R^T)^{-1}(R\hat{\beta} - c) \sim \chi^2_L$$

  where L is the "degrees of freedom", the number of rows in R. We reject the null if this statistic exceeds a suitably defined critical value.

OLS

Let $\mathbb{E}[Y_n|\boldsymbol{x}_n] = \boldsymbol{x}_n\beta$ where $\boldsymbol{x}_n$ is a row vector. Let $\boldsymbol{X}$ be N iid observations of $\boldsymbol{x}_n$ stacked vertically and the same for $\boldsymbol{Y}$.

- OLS estimator: $\hat{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$
- Variance for $\hat{\beta}$:

$$\mathbb{V}[\hat{\beta}] = \frac{1}{N}\mathbb{E}[\boldsymbol{x}_n^T\boldsymbol{x}_n]^{-1}\mathbb{E}[\boldsymbol{x}_n^T\boldsymbol{x}_n\epsilon_n^2]\mathbb{E}[\boldsymbol{x}_n^T\boldsymbol{x}_n]^{-1}$$

- And if $\mathbb{E}[\epsilon_n^2|\boldsymbol{x}_n] = \sigma^2$,

$$\mathbb{V}[\hat{\beta}] = \frac{1}{N}\mathbb{E}[\boldsymbol{x}_n^T\boldsymbol{x}_n]^{-1}\sigma^2$$

- To estimate $\mathbb{V}[\hat{\beta}]$ in general:

$$\hat{V}_\beta = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\sum_n \boldsymbol{x}_n^T\boldsymbol{x}\hat{\epsilon}_n^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$$

- And if $\mathbb{E}[\epsilon_n^2|\boldsymbol{x}_n] = \sigma^2$:

$$\hat{V}_{\hat{\beta}} = \frac{1}{N}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\sum_n \hat{\epsilon}_n^2$$

Instrumental Variables

Let the first stage be

$$\boldsymbol{x}_n = \boldsymbol{z}_n\pi + \epsilon_n$$

with a second stage

$$Y_n = \boldsymbol{x}_n\gamma + \eta_n$$

Let $\boldsymbol{X}$ be all of the $\boldsymbol{x}_n$ stacked in a vector and $\boldsymbol{Z}$ the same for $\boldsymbol{z}$.

- The 2SLS estimator is

$$\hat{\gamma} = (\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}})^{-1}\hat{\boldsymbol{X}}'\boldsymbol{Y}$$

  and

$$\hat{\boldsymbol{X}} = \boldsymbol{Z}(\boldsymbol{Z}'\hat{\boldsymbol{Z}})^{-1}\boldsymbol{Z}'\boldsymbol{Y}$$

- The variance of the 2SLS estimator is **approximately**:[1]

$$V_{\hat{\gamma}} = \sigma_\eta^2 \left(\boldsymbol{Q}_{XZ} \boldsymbol{Q}_{ZZ}^{-1} \boldsymbol{Q}_{XZ}' \right)^{-1}$$

  where $\sigma_\eta^2 = \mathbb{E}[\eta^2 | \boldsymbol{z}]$, $\boldsymbol{Q}_{XZ} = \mathbb{E}[\boldsymbol{x}' \boldsymbol{z}]$ and $\boldsymbol{Q}_{ZZ} = \mathbb{E}[\boldsymbol{z}' \boldsymbol{z}]$. The variance can be estimated as

$$\hat{V}_{\hat{\gamma}} = s^2 (\boldsymbol{X}' \boldsymbol{Z} (\boldsymbol{Z}' \boldsymbol{Z})^{-1} \boldsymbol{Z}' \boldsymbol{X})^{-1}$$

  where

$$s^2 = \frac{1}{N} \sum \hat{\eta}_n^2, \qquad \hat{\eta}_n = Y_n - \boldsymbol{x}_n \hat{\gamma}$$

[1]This approximation is derived from the asymptotic variance implied by the central limit theorem.