ECON 251

Discussion Section

Week 2 Solutions

1. Review some important theoretical concepts

- Law of iterated expectations (LIE)

  $E(Y) = E_X(E[Y|X])$

- Populations, parameters, samples, statistics… and their distributions!

  $Y \sim f(y;\theta)$ $\quad \theta = (\mu_Y, \sigma_Y^2)$ $\quad \{Y_1, Y_2, \dots, Y_N\}$ $\quad T(Y_1, \dots, Y_N)$

- Estimators (aka statistics) vs estimates

  $T(Y_1, \dots, Y_N)$ $\quad T(y_1, \dots, y_N)$

- Properties: finite sample (bias, variance, efficiency) vs "large" sample or asymptotic (consistency, asymptotic variance)

  $T(Y_1, \dots, Y_N)$ is unbiased for θ whenever $E[T(Y_1, \dots, Y_N)] = \theta$

  $T(Y_1, \dots, Y_N)$ is consistent for θ whenever $\operatorname*{plim}_{N\to\infty} T(Y_1, \dots, Y_N) = \theta$

  Note that plim is equivalent to 2 regular limits: $\lim_{N\to\infty} Bias[T(Y_1, \dots, Y_N)] = 0$

  $$\lim_{N\to\infty} Var[T(Y_1, \dots, Y_N)] = 0$$

- Law of large numbers (LLN)

  If $\{Y_1, \dots, Y_N\}$ is a random sample from population $Y \sim f(y; \mu_Y, \sigma_Y^2)$, then

  $$\operatorname*{p\,lim}_{N\to\infty} \overline{Y} = \mu_Y$$

  In other words, the mean of a random sample $\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$ is always consistent for the population mean $\mu_Y \coloneqq E[Y]$ of any random variable Y!

- Central limit theorem (CLT)

  If $\{Y_1, \dots, Y_N\}$ is a random sample from population $Y \sim f(y; \mu_Y, \sigma_Y^2)$ with finite variance $\sigma_Y^2 < \infty$, then the "standardized" random variable

  $$Z \coloneqq \frac{\overline{Y} - \mu_Y}{\sigma_Y}$$

  has an asymptotic distribution that is standard normal: $Z \sim N(0,1)$ as $N \to \infty$.

- Regression as conditional expectation

2. Prove that the method of moments estimator $\hat{\sigma}_{MOM}^2 \coloneqq \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2$ is consistent for population variance parameter $\sigma_X^2 \coloneqq E[(X-\mu_X)^2]$ of random variable $X \sim f_X(\mu_X, \sigma_X^2)$.

$$\operatorname*{plim}_{N\to\infty} \hat{\sigma}_{MOM}^2 = \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2 = \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\left(X_i^2 - 2X_i\overline{X} + \overline{X}^2\right)$$

$$= \operatorname*{plim}_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N} X_i^2 - \overline{X}^2\right] = \operatorname*{plim}_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N} X_i^2\right] - \operatorname*{plim}_{N\to\infty}\overline{X}^2$$

$$= E[X^2] - \left[\operatorname*{plim}_{N\to\infty}\overline{X}\right]^2 = E[X^2] - E[X]^2$$

$$= E[(X-\mu_X)^2] = \sigma_X^2$$

3. Let $\{X_1, \dots, X_N\}$ denote a random sample of size N from $X \sim f_X(\mu_X, \sigma_X^2)$. Consider the following candidate estimators of the population mean:

   $\hat{\mu}_1 \coloneqq \overline{X} + \frac{1}{N}$ $\qquad\qquad$ $\hat{\mu}_2 \coloneqq 0.9 \cdot \overline{X}$ $\qquad\qquad$ $\hat{\mu}_3 \coloneqq \frac{X_1 + X_N}{2}$

   Which estimators are unbiased for μ_X? Which are consistent for μ_X?

$E[\hat{\mu}_1] \coloneqq E\left[\overline{X} + \frac{1}{N}\right] = E\left[\frac{1}{N}\sum_{i=1}^{N} X_i\right] + E\left[\frac{1}{N}\right] = \frac{1}{N}E\left[\sum_{i=1}^{N} X_i\right] + \frac{1}{N}$

$\quad = \frac{1}{N}\sum_{i=1}^{N} E[X_i] + \frac{1}{N} = \frac{1}{N}\sum_{i=1}^{N}\mu_X + \frac{1}{N} = \frac{1}{N}N\mu_X + \frac{1}{N} = \mu_X + \frac{1}{N} \neq \mu_X$ -> biased!

$E[\hat{\mu}_2] \coloneqq E\left[0.9 \cdot \overline{X}\right] = 0.9 \cdot E\left[\overline{X}\right] = 0.9 \cdot \mu_X \neq \mu_x$ -> biased!

$E[\hat{\mu}_3] \coloneqq E\left[\frac{X_1 + X_N}{2}\right] = \frac{1}{2} \cdot E[X_1 + X_N] = \frac{1}{2}(\mu_X + \mu_X) = \mu_X$ -> unbiased!

$\underset{N\to\infty}{\text{plim}}\, \hat{\mu}_1 = \underset{N\to\infty}{\text{plim}}\left[\overline{X} + \frac{1}{N}\right] = \underset{N\to\infty}{\text{plim}}\left[\overline{X}\right] + \underset{N\to\infty}{\text{plim}}\left[\frac{1}{N}\right] = \mu_X$ -> consistent!

$\underset{N\to\infty}{\text{plim}}\, \hat{\mu}_2 = \underset{N\to\infty}{\text{plim}}\left[0.9 \cdot \overline{X}\right] = 0.9 \cdot \underset{N\to\infty}{\text{plim}}\left[\overline{X}\right] = 0.9 \cdot \mu_X \neq \mu_X$ -> not consistent!

Since $Var(\hat{\mu}_3) = \frac{1}{2}\sigma_X^2$, we have $\lim_{N\to\infty} \frac{1}{2}\sigma_X^2 \neq 0$ -> not consistent!

4. Let $Y = \beta_0 + \beta_1 X + e$ denote a linear population regression function. Prove that whenever $Cov(X, e) = 0$ we can write the values of $\{\beta_0, \beta_1\}$ in terms of $E(X)$, $E(Y)$, $Var(Y)$, and $Cov(X, Y)$. What is the economic meaning behind the assertion that the value of the parameter $Cov(X, e)$ must be $= 0$ in the population? (Bonus: if it fails, why would an infinite sample, or the whole population, be useless for establishing causality?)

$Cov(X, e) = 0$

$Cov(X, Y - \beta_0 - \beta_1 X) = 0$

$Cov(X, Y) - Cov(X, \beta_0) - Cov(X, \beta_1 X) = 0$

$Cov(X, Y) - 0 - \beta_1 Cov(X, X) = 0$

$Cov(X, Y) - \beta_1 Var(X) = 0$

$\beta_1 = \frac{Cov(X,Y)}{Var(X)} = \frac{E[XY] - E[X]E[Y]}{E[X^2] - E[X]^2}$

$\Rightarrow \hat{\beta}_1^{MOM} := \frac{\widehat{Cov(X,Y)}}{\widehat{Var(X)}} = \hat{\beta}_1^{OLS}$

$E[Y] = E[\beta_0 + \beta_1 X + e]$

$E[Y] = \beta_0 + \beta_1 E[X] + E[e]$

$E[Y] = \beta_0 + \beta_1 E[X] + 0$

$\beta_0 = E[Y] - \beta_1 E[X]$

$\Rightarrow \hat{\beta}_0^{MOM} := \widehat{E[Y]} - \hat{\beta}_1^{MOM} \cdot \widehat{E[X]} = \overline{Y} - \hat{\beta}_1^{MOM}\overline{X} = \hat{\beta}_1^{OLS}$