

Empirical Analysis I - Problem Set 3

Timothy Schwieg

Paulo Henrique Ramos

Samuel Barker

Rafeh

Question 1

Notice that, because we are interested in the limit distribution of the variance estimator, we can assume that the observations are centered (that is, consider that the mean of the distribution was already subtracted from the X_i ; i.e. $X_i = Y_i - \mu$ for some Y_i with the distribution we are interested in). This gives us that $\mathbb{E}[X_i] = 0$ (which we can interpret as the first centered moment, and similarly with higher moments, all being interpreted as centered), and it does not change the variance or its estimator.

Also, notice that $S_n^2 = \frac{1}{n-1} \sum (X_i - \bar{X}_n)^2 = \frac{n}{n-1} \left[\frac{1}{n} \sum X_i^2 - \bar{X}_n^2 \right]$. Thus we can write $S_n^2 = g(\frac{n}{n-1}, \bar{X}_n^2, \bar{X}_n)$, where $g(a, b, c) = a[b - c^2]$ is a continuous and differentiable function on the relevant domain.

Because $\frac{n}{n-1} \xrightarrow{p} 1$, $\bar{X}_n^2 \xrightarrow{p} \mathbb{E}[X_i^2]$ and $\bar{X}_n \xrightarrow{p} \mathbb{E}[X_i]$ due to the WLLN (which we can apply, because $\mathbb{E}[X_i^4] \leq \infty$), and marginal convergence in probability implies joint convergence in probability, we have $(\frac{n}{n-1}, \bar{X}_n^2, \bar{X}_n) \xrightarrow{p} (1, \mathbb{E}[X_i^2], \mathbb{E}[X_i])$.

Also, when applied to the limits, $g(1, \mathbb{E}[X_i^2], \mathbb{E}[X_i]) = [\mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2] = \sigma^2$, the true variance of X_i .

Also, because $\mathbb{E}[X_i^4] \leq \infty$, we can apply the CLT to this vector:

$$\begin{aligned} \sqrt{n} \left(\begin{bmatrix} \frac{n}{n-1} \\ \bar{X}_n^2 \\ \bar{X}_n \end{bmatrix} - \begin{bmatrix} 1 \\ \mathbb{E}[X_i^2] \\ \mathbb{E}[X_i] \end{bmatrix} \right) &\xrightarrow{d} N \left(0, \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbb{E}[X_i^4] - \mathbb{E}[X_i^2]^2 & \mathbb{E}[X_i^3] - \mathbb{E}[X_i^2] \mathbb{E}[X_i] \\ 0 & \mathbb{E}[X_i^3] - \mathbb{E}[X_i^2] \mathbb{E}[X_i] & \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 \end{bmatrix} \right) \\ &\text{(because } \mathbb{E}[X_i] = 0 \text{)} = N \left(0, \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbb{E}[X_i^4] - \mathbb{E}[X_i^2]^2 & \mathbb{E}[X_i^3] \\ 0 & \mathbb{E}[X_i^3] & \mathbb{E}[X_i^2] \end{bmatrix} \right) \end{aligned} \quad (1)$$

Now we can use the delta method with $g(\cdot)$ (its derivative on (a, b, c) is $(1, 1, -2c)$) and

conclude:

$$\begin{aligned} \sqrt{n}(S_n^2 - \sigma^2) &\xrightarrow{d} N(0, [1 \quad 1 \quad -2\mathbb{E}[X_i] (=0)] \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbb{E}[X_i^4] - \mathbb{E}[X_i^2]^2 & \mathbb{E}[X_i^3] \\ 0 & \mathbb{E}[X_i^3] & \mathbb{E}[X_i^2] \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -2\mathbb{E}[X_i] (=0) \end{bmatrix}) \\ &= N(0, \mathbb{E}[X_i^4] - \mathbb{E}[X_i^2]^2) \end{aligned} \quad (2)$$

Thus, recalling that $\mathbb{E}[X_i^4]$ and $\mathbb{E}[X_i^2]$ are centered moments, we have that $\sqrt{n}(S_n^2 - \sigma^2) \xrightarrow{d} N(0, \mathbb{E}[(Y_i - \mu)^4] - \sigma^4)$, if we consider Y_i as the uncentered variable.

Question 5

a

(\implies) If $p(1|1) = p(1|0)$ and $p(0|1) = p(0|0)$, then the ratio $\frac{p(1|1)}{p(1|0)} \bigg/ \frac{p(0|1)}{p(0|0)} = 1/1 = 1 = \rho$.

(\impliedby) If $\rho = 1$, then we have:

$$\begin{aligned} \frac{p(1|1)}{p(1|0)} &= \frac{p(0|1)}{p(0|0)} \\ &\iff \\ \frac{p(0|0)}{p(1|0)} + \frac{p(1|0)}{p(1|0)} &= \frac{p(0|1)}{p(1|1)} + \frac{p(1|1)}{p(1|1)} \\ &\iff \\ \frac{p(0|0) + p(1|0)}{p(1|0)} &= \frac{p(0|1) + p(1|1)}{p(1|1)} \quad (3) \\ &\iff \\ \frac{1}{p(1|0)} &= \frac{1}{p(1|1)} \\ &\iff \\ p(1|0) &= p(1|1) \end{aligned}$$

But, this implies that $1 = \frac{p(0|1)}{p(0|0)}$, which implies $p(0|1) = p(0|0)$.

b

Using bayes rules, we have:

$$\begin{aligned} \rho &= \frac{\frac{p(1,1)}{p(1)}}{\frac{p(1,0)}{p(0)}} \bigg/ \frac{\frac{p(0,1)}{p(1)}}{\frac{p(0,0)}{p(0)}} \\ \rho &= \frac{p(1,1)}{p(1,0)} \bigg/ \frac{p(0,1)}{p(0,0)} \end{aligned} \quad (4)$$

c

Let's first take $\hat{p}_n^{(1,1)} := \frac{1}{n} \sum \mathbb{1}_{\{Y_i=1, X_i=1\}}$. Because $\mathbb{1}_{\{Y_i=1, X_i=1\}}$ is a Bernoulli random variable, it has finite expectation ($p(1,1)$) and variance ($(p(1,1)[1 - p(1,1)])$). Thus, since the sample of students is i.i.d., the random variable $\mathbb{1}_{\{Y_i=1, X_i=1\}}$ is also i.i.d., and the WLLN gives us that $\hat{p}_n^{(1,1)} \xrightarrow{P} p(1,1)$. The same rationale applies to $\hat{p}_n^{(1,0)}$, $\hat{p}_n^{(0,1)}$ and $\hat{p}_n^{(0,0)}$. And because marginal convergence in probability implies joint convergence in probability, we have: $(\hat{p}_n^{(1,1)}, \hat{p}_n^{(1,0)}, \hat{p}_n^{(0,1)}, \hat{p}_n^{(0,0)}) \xrightarrow{P} (p(1,1), p(1,0), p(0,1), p(0,0))$.

Now we know from item (b) that $\rho = \frac{p(1,1)}{p(1,0)} \bigg/ \frac{p(0,1)}{p(0,0)}$, and no terms in this expression are zero. Thus, $f(a, b, c, d) = \frac{a}{b} \bigg/ \frac{c}{d}$ is a continuous function on the relevant domain, and we can apply the continuous mapping theorem to conclude that:

$$\hat{\rho}_n = f(\hat{p}_n^{(1,1)}, \hat{p}_n^{(1,0)}, \hat{p}_n^{(0,1)}, \hat{p}_n^{(0,0)}) \xrightarrow{P} f(p(1,1), p(1,0), p(0,1), p(0,0)) = \rho \quad (5)$$

Thus our estimator is $\hat{\rho}_n := \frac{\hat{p}_n^{(1,1)}}{\hat{p}_n^{(1,0)}} \bigg/ \frac{\hat{p}_n^{(0,1)}}{\hat{p}_n^{(0,0)}}$, and its is consistent.

d

As seen in item (c) above, $\mathbb{1}_{\{Y_i=1, X_i=1\}}$ is bernoulli, with finite expectation and variance (and similarly with the other indicator functions for different outcomes). Therefore, we can apply the multivariate CLT to conclude that:

$$\sqrt{n} \left(\begin{bmatrix} \hat{p}_n^{(1,1)} \\ \hat{p}_n^{(1,0)} \\ \hat{p}_n^{(0,1)} \\ \hat{p}_n^{(0,0)} \end{bmatrix} - \begin{bmatrix} p(1,1) \\ p(1,0) \\ p(0,1) \\ p(0,0) \end{bmatrix} \right) \xrightarrow{d} N(0, \Sigma), \quad (6)$$

where:

$$\Sigma = \begin{bmatrix} p(1,1)[1 - p(1,1)] & -p(1,1)p(1,0) & -p(1,1)p(0,1) & -p(1,1)p(0,0) \\ -p(1,0)p(1,1) & p(1,0)[1 - p(1,0)] & -p(1,0)p(0,1) & -p(1,0)p(0,0) \\ -p(0,1)p(1,1) & -p(0,1)p(1,0) & p(0,1)[1 - p(0,1)] & -p(0,1)p(0,0) \\ -p(0,0)p(1,1) & -p(0,0)p(1,0) & -p(0,0)p(0,1) & p(0,0)[1 - p(0,0)] \end{bmatrix} \quad (7)$$

We can then take the function $g(a, b, c, d) = \ln \frac{a}{b} \bigg/ \frac{c}{d} = \ln a - \ln b - \ln c + \ln d$, which is continuous and differentiable on $(p(1,1), p(1,0), p(0,1), p(0,0))$, since they are greater than zero. Using the delta method, we obtain that:

$$\begin{aligned} \sqrt{n}(g(\hat{p}_n^{(1,1)}, \hat{p}_n^{(1,0)}, \hat{p}_n^{(0,1)}, \hat{p}_n^{(0,0)}) - g(p(1,1), p(1,0), p(0,1), p(0,0))) \\ \xrightarrow{d} Dg(p(1,1), p(1,0), p(0,1), p(0,0))N(0, \Sigma) \\ \sqrt{n}(\ln \hat{\rho}_n - \ln \rho) \xrightarrow{d} Dg(p(1,1), p(1,0), p(0,1), p(0,0))N(0, \Sigma) \end{aligned} \quad (8)$$

Now we can calculate the variance of $Dg(.)N(0, \Sigma)$ and call it τ^2 as we wanted:

$$\begin{aligned} \begin{bmatrix} \frac{1}{p(1,1)} \\ -\frac{1}{p(1,0)} \\ -\frac{1}{p(0,1)} \\ \frac{1}{p(0,0)} \end{bmatrix}^T & \begin{bmatrix} p(1,1)[1 - p(1,1)] & -p(1,1)p(1,0) & -p(1,1)p(0,1) & -p(1,1)p(0,0) \\ -p(1,0)p(1,1) & p(1,0)[1 - p(1,0)] & -p(1,0)p(0,1) & -p(1,0)p(0,0) \\ -p(0,1)p(1,1) & -p(0,1)p(1,0) & p(0,1)[1 - p(0,1)] & -p(0,1)p(0,0) \\ -p(0,0)p(1,1) & -p(0,0)p(1,0) & -p(0,0)p(0,1) & p(0,0)[1 - p(0,0)] \end{bmatrix} \begin{bmatrix} \frac{1}{p(1,1)} \\ -\frac{1}{p(1,0)} \\ -\frac{1}{p(0,1)} \\ \frac{1}{p(0,0)} \end{bmatrix} = \\ & \begin{bmatrix} 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{p(1,1)} \\ -\frac{1}{p(1,0)} \\ -\frac{1}{p(0,1)} \\ \frac{1}{p(0,0)} \end{bmatrix} = \frac{1}{p(1,1)} + \frac{1}{p(1,0)} + \frac{1}{p(0,1)} + \frac{1}{p(0,0)} =: \tau^2 \end{aligned} \quad (9)$$

e

The estimator is: $\hat{\tau}_n^2 := 1/\hat{p}_n^{(1,1)} + 1/\hat{p}_n^{(1,0)} + 1/\hat{p}_n^{(0,1)} + 1/\hat{p}_n^{(0,0)}$. As seen in item (c) above, we have $(\hat{p}_n^{(1,1)}, \hat{p}_n^{(1,0)}, \hat{p}_n^{(0,1)}, \hat{p}_n^{(0,0)}) \xrightarrow{p} (p(1,1), p(1,0), p(0,1), p(0,0))$. Thus, because all terms are greater than zero, the function $f(a, b, c, d) = 1/a + 1/b + 1/c + 1/d$ is continuous on the relevant domain, and we can apply the CMT to conclude that $\hat{\tau}_n^2 \xrightarrow{p} \tau^2$.

f

We know from previous items that $\sqrt{n}(\ln \hat{\rho}_n - \ln \rho) \xrightarrow{d} N(0, \tau^2)$. Because the function $g(x) = \exp x$ is continuous and differentiable at the true ρ , we can apply the delta method to obtain:

$$\sqrt{n}(\hat{\rho}_n - \rho) \xrightarrow{d} \rho N(0, \tau^2) = N(0, \tau^2 \rho^2) \quad (10)$$

g

If gender is independent of supporting Obama, then we have that the conditional probabilities equal the unconditional ones: $p(1|0) = p(1|1)$ and $p(0|0) = p(0|1)$. As seen in item (a), this is equivalent to $\rho = 1$, thus we can define $H_0 : \rho = 1$ and $H_1 : \rho \neq 1$.

Now define $T_n := \left| \frac{\sqrt{n}(\hat{\rho}_n - 1)}{\sqrt{\hat{\tau}_n^2}} \right| = \frac{\sqrt{n}(|\hat{\rho}_n - 1|)}{\sqrt{\hat{\tau}_n^2}}$. We will reject the null if $T_n > z_{1-\frac{\alpha}{2}}$.

Under the null hypothesis and using CMT and Slutsky (because $\hat{\tau}_n^2 \xrightarrow{p} \tau^2 > 0$, $g(x) = 1/\sqrt{x}$ is continuous on the relevant domain, and τ^2 is constant), we know that: $T_n \xrightarrow{d} |z| \sim N(0, 1)$. Therefore, we have:

$$\limsup \Pr(T_n > z_{1-\frac{\alpha}{2}}) \leq \limsup \Pr(T_n \geq z_{1-\frac{\alpha}{2}}) \leq \Pr(|z| \geq z_{1-\frac{\alpha}{2}}) = \alpha. \quad (11)$$

Thus the test is consistent in level α . And we can also calculate the p -value using the

following:

$$\begin{aligned}
 p - \text{value} &= \inf \{ \alpha \in (0, 1) | T_n \geq z_{1-\frac{\alpha}{2}} \} \\
 &= \inf \{ \alpha \in (0, 1) | \phi(T_n) \geq \phi(z_{1-\frac{\alpha}{2}}) \} \\
 &= \inf \{ \alpha \in (0, 1) | \phi(T_n) \geq 1 - \frac{\alpha}{2} \} \\
 &= \inf \{ \alpha \in (0, 1) | \alpha \geq 2(1 - \phi(T_n)) \} = 2(1 - \phi(T_n))
 \end{aligned} \tag{12}$$

where T_n is as defined above and $\phi(\cdot)$ is the cdf of the standard normal distribution.

Question 9

Mean independence means that $\mathbb{E}[Y|X] = c$, a constant, and using the law of iterated expectations, we have that $c = \mathbb{E}[c] = \mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$. Thus we have:

$$\begin{aligned}
 \text{Cov}[Y, X] &= \mathbb{E}[YX] - \mathbb{E}[X] \mathbb{E}[Y] \\
 (\text{By LIE}) &= \mathbb{E}[\mathbb{E}[YX|X]] - \mathbb{E}[X] \mathbb{E}[Y] \\
 (\text{Because } X \text{ is a function of } X) &= \mathbb{E}[X \mathbb{E}[Y|X]] - \mathbb{E}[X] \mathbb{E}[Y] \\
 (\text{Because of mean independence}) &= \mathbb{E}[X] \mathbb{E}[Y] - \mathbb{E}[X] \mathbb{E}[Y] = 0
 \end{aligned} \tag{13}$$

But $\text{Cov}[Y, X] = 0$ does not imply mean independence. Take, for instance, $X \sim N(0, 1)$ and $Y = X^2$. Then we have $\text{Cov}[Y, X] = \mathbb{E}[X^3] - \mathbb{E}[X] \mathbb{E}[X^2] = 0$ (because the mean and third moment of the normal are zero), but $\mathbb{E}[Y|X] = Y = X^2$, because Y is a function of X .

Also, mean independence does not imply independence. Take $Y|X \sim N(0, \sigma^2 X)$, with a non-degenerate random variable. Then we do have $\mathbb{E}[Y|X] = 0$, a constant, but the distributions of Y and X are dependent (by construction, the conditional distribution of Y changes with X , and thus cannot be equal the unconditional distribution for all X).

Question 13

a

i

Yes. Because we are interpreting the regression as the best linear predictor of Y given X , the vector $\beta := (\beta_0, \beta_1)'$ minimizes $\mathbb{E}[(Y - \bar{X}'\beta)^2]$, where $\bar{X} = (1, X)$. This results in the first order conditions:

$$\begin{aligned}
 -2\mathbb{E}[\bar{X}(Y - \bar{X}'\beta)] &= 0 \\
 \mathbb{E}\left[\begin{pmatrix} 1 \\ X \end{pmatrix} (Y - (1 \ X) \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix})\right] &= 0 \\
 \mathbb{E}\left[\begin{pmatrix} Y - \beta_0 - \beta_1 X \\ X(Y - \beta_0 - \beta_1 X) \end{pmatrix}\right] &= 0 \\
 \begin{pmatrix} \mathbb{E}[U] \\ \mathbb{E}[XU] \end{pmatrix} &= 0
 \end{aligned} \tag{14}$$

Since $Cov[X, U] = \mathbb{E}[XU] - \mathbb{E}[X]\mathbb{E}[U] = 0$, we have that U and X are uncorrelated.

ii

From the regression equation $Y = \beta_0 + \beta_1 X + U$ we have:

$$\begin{aligned}\mathbb{E}[Y|X=0] &= \beta_0 + \mathbb{E}[U|X=0] \\ \mathbb{E}[Y|X=1] &= \beta_0 + \beta_1 + \mathbb{E}[U|X=1]\end{aligned}\tag{15}$$

Now we can show that $\mathbb{E}[U|X=0] = \mathbb{E}[U|X=1] = 0$ using the conditions obtained in item (i) above ($\mathbb{E}[U] = 0 = \mathbb{E}[XU]$):

$$\begin{aligned}0 = \mathbb{E}[XU] &= \mathbb{E}[XU|X=0]\Pr(X=0) + \mathbb{E}[XU|X=1]\Pr(X=1) \\ &= 0 + \mathbb{E}[U|X=1]\Pr(X=1) \\ \implies \mathbb{E}[U|X=1] &= 0 \text{ (because } \Pr(X=1) > 0),\end{aligned}\tag{16}$$

and this also implies:

$$\begin{aligned}0 = \mathbb{E}[U] &= \mathbb{E}[U|X=0]\Pr(X=0) + \mathbb{E}[U|X=1]\Pr(X=1) \\ &= \mathbb{E}[U|X=0]\Pr(X=0) + 0 \\ \implies \mathbb{E}[U|X=0] &= 0 \text{ (because } \Pr(X=0) > 0).\end{aligned}\tag{17}$$

Therefore, we have that:

$$\begin{aligned}\beta_0 &= \mathbb{E}[Y|X=0] \\ \beta_1 &= \mathbb{E}[Y|X=1] - \mathbb{E}[Y|X=0]\end{aligned}\tag{18}$$

iii

Yes. As seen in item (ii), $\mathbb{E}[Y|X] = 0$, a constant, for all possible values of X . This happens because, when we have binary variables as conditioners, the conditional expectation takes a linear form, and thus our best linear predictor is exactly equal to the conditional expectation. As a consequence, by definition of U as the difference between the conditional expectation and the linear predictor, it will be constantly zero.

a

i

Not necessarily. The interpretation of causality only assumes that $Y = g(X.U)$, i.e., that Y can be determined as a function of X and U . But it does not assume anything else about the relationship between X and U .

ii

Notice that we can still take conditional expectations on both sides of the regression and get:

$$\begin{aligned}\beta_0 &= \mathbb{E}[Y|X=0] - \mathbb{E}[U|X=0] \\ \beta_1 &= \mathbb{E}[Y|X=1] - \mathbb{E}[Y|X=0] + \mathbb{E}[U|X=0] - \mathbb{E}[U|X=1]\end{aligned}\tag{19}$$

If X and U were in fact uncorrelated, then we could use the same steps in item (ii) of letter (a) above to conclude that $\mathbb{E}[U|X] = 0$ always, and obtain the same values of β . But if this is not case, the values will differ, because $\mathbb{E}[U|X]$ is not constant in X .