Econ 2120: Section 2

Part II - Random Sampling and the Linear Predictor

Ashesh Rambachan

Fall 2018

Outline

Outline

So far

We have been discussing population properties of the best linear predictor by considering a single "draw" of data.

> We modeled that single draw of data as a random vector (Y, X).

In Elie's notation, we considered

$$(Y_1, \dots, Y_M, Z_1, \dots, Z_L) \sim F,$$

where F is some population distribution

So far

Examples:

(1) $M = 1$, $L = 2$: Population of individuals, observe 1 individual randomly drawn and record $Y =$ wage, $Z_1 =$ years of education, $Z_2 =$ parental income

So far

Examples:

(1) $M = 1$, $L = 2$: Population of individuals, observe 1 individual randomly drawn and record Y = wage, Z_1 = years of education, Z_2 = parental income

(2) $M = 2$, $L = 3$: Population of twins, observe 1 pair of twins randomly drawn and record Y_1 = wage of twin 1, Y_2 = wage of twin 2, Z_1 = education of twin 1, Z_2 = education of twin 2, Z_3 = parental income.

So far

Examples:

(1) $M = 1$, $L = 2$: Population of individuals, observe 1 individual randomly drawn and record $Y =$ wage, $Z_1 =$ years of education, $Z_2 =$ parental income

(2) $M = 2$, $L = 3$: Population of twins, observe 1 pair of twins randomly drawn and record $Y_1 =$ wage of twin 1, $Y_2 =$ wage of twin 2, $Z_1 =$ education of twin 1, $Z_2 =$ education of twin 2, $Z_3 =$ parental income.

(3) $M = T$, $L = T + 1$: Population of individuals, observe 1 individual randomly drawn and record $Y_t =$ wage in year t, $Z_{1t} =$ occupation industry, $Z_2 =$ years of education.

The population distribution, F

The distribution F is unknown.

The population distribution, F

The distribution F is unknown.

Assume F belongs to some set of distributions indexed by a parameter $\theta \in \Theta$.

$$\mathcal{P} = \{P_\theta : \theta \in \Theta\}, \quad F = P_{\theta^*}$$

for some $\theta^* \in \Theta$.

The population distribution, F

The distribution F is unknown.

Assume F belongs to some set of distributions indexed by a parameter $\theta \in \Theta$.

$$\mathcal{P} = \{P_\theta : \theta \in \Theta\}, \quad F = P_{\theta^*}$$

for some $\theta^* \in \Theta$.

Jargon:

$\mathcal{P}$ is parametric if Θ has finite dimension.

$\mathcal{P}$ is non-parametric if Θ is an infinite dimension parameter space.

The population distribution, F

Key question: Given data, what statements can we make about F?

The population distribution, F

Key question: Given data, what statements can we make about F?

We will begin to explore this question with the population linear predictor and the least squares estimator?

The population distribution, F

Key question: Given data, what statements can we make about F?

We will begin to explore this question with the population linear predictor and the least squares estimator?

> What can we say about the sampling distribution of the least squares estimator, $\hat{E}^*[Y|X]$?

The population distribution, F

Key question: Given data, what statements can we make about F?

We will begin to explore this question with the population linear predictor and the least squares estimator?

- What can we say about the sampling distribution of the least squares estimator, $\hat{E}^*[Y|X]$?
- How does it relate to the population linear predictor $E^*[Y|X]$?

Outline

Outline

A Random Sample

Random sample of size n:

n independent draws from the population F (with replacement). The i-th draw is

$$(Y_{i1}, \ldots, Y_{iM}, Z_{i1}, \ldots, Z_{iL}).$$

The joint distribution of this random vector is F.

A Random Sample

Random sample of size n:

n independent draws from the population F (with replacement). The i-th draw is

$$(Y_{i1}, \ldots, Y_{iM}, Z_{i1}, \ldots, Z_{iL}).$$

The joint distribution of this random vector is F.

Notation for random sample:

$$D_i = (Y_{i1}, \ldots, Y_{iM}, Z_{i1}, \ldots, Z_{iL}) \overset{i.i.d.}{\sim} F \quad \text{for} \quad i = 1, \ldots, n.$$

A Random Sample

Additional notation:

$$\underset{M\times 1}{Y_i} = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{iM} \end{pmatrix}, \quad \underset{L\times 1}{Z_i} = \begin{pmatrix} Z_{i1} \\ \vdots \\ Z_{iL} \end{pmatrix}$$

May also write random sample of size n as

$$(Y_i, Z_i) \overset{i.i.d.}{\sim} F \quad \text{for} \quad i = 1, \ldots, n.$$

A Random Sample

Additional notation:

$$\underset{M\times 1}{Y_i} = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{iM} \end{pmatrix}, \quad \underset{L\times 1}{Z_i} = \begin{pmatrix} Z_{i1} \\ \vdots \\ Z_{iL} \end{pmatrix}$$

May also write random sample of size n as

$$(Y_i, Z_i) \overset{i.i.d.}{\sim} F \quad \text{for} \quad i = 1, \ldots, n.$$

More notation:

$$\underset{n\times M}{Y} = \begin{pmatrix} Y_1' \\ \vdots \\ Y_M' \end{pmatrix}, \quad \underset{n\times L}{Z} = \begin{pmatrix} Z_1' \\ \vdots \\ Z_n' \end{pmatrix}$$

Outline

Notation

The linear predictors will use transformations of Z variables

$$X_{ik} = g_k(Z_i) \quad \text{for} \quad k = 1, \ldots, K.$$

Each $g_k(\cdot)$ is a known, specified function.

E.g. $Z_i =$ age and we transform this into age, age squared, etc.

Notation

The linear predictors will use transformations of Z variables

$$X_{ik} = g_k(Z_i) \quad \text{for} \quad k = 1, \ldots, K.$$

Each $g_k(\cdot)$ is a known, specified function.

E.g. $Z_i =$ age and we transform this into age, age squared, etc.

As before,

$$\underset{K \times 1}{X_i} = \begin{pmatrix} X_{i1} \\ \vdots \\ X_{iK} \end{pmatrix}, \quad \underset{n \times K}{X} = \begin{pmatrix} X_1' \\ \vdots \\ X_n' \end{pmatrix}$$

Outline

Least-squares estimator

Linear predictor:

$$E^*[Y_i|X_i] = \beta_1 X_{i1} + \ldots + \beta_K X_{iK}.$$

Least-squares estimator

Linear predictor:

$$E^*[Y_i|X_i] = \beta_1 X_{i1} + \ldots + \beta_K X_{iK}.$$

Recall:

$$\beta = E[X_i X_i']^{-1} E[X_i Y_i]$$

β does not depend on i because (Y_i, X_i) are i.i.d.

Least-squares estimator

Linear predictor:

$$E^*[Y_i|X_i] = \beta_1 X_{i1} + \ldots + \beta_K X_{iK}.$$

Recall:

$$\beta = E[X_i X_i']^{-1} E[X_i Y_i]$$

β does not depend on i because (Y_i, X_i) are i.i.d.

F is unknown and so, need to estimate β. How?

Least-squares estimator

Natural estimator:

$$b(Y, Z) = \Big(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\Big)^{-1}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i Y_i\Big) = (X'X)^{-1}X'Y$$

Where's it come from?

(1) Replace expectations with sample averages.

(2) Solve "sample analogue" to minimum mean-square error/minimum norm problem.

Expectation of Least-Squares Estimator

We wish to compute

$$E[b(Y, Z)]$$

Looks hard...

Expectation of Least-Squares Estimator

We wish to compute

$$E[b(Y, Z)]$$

Looks hard...

Iterated expectations to the rescue!

Expectation of Least-Squares Estimator

We wish to compute

$$E[b(Y, Z)]$$

Looks hard...

Iterated expectations to the rescue!

Condition on $Z = z$. Since $X = g(Z)$, this fixes $X = x$.

$$b(Y, Z)|Z = z \sim (x'x)^{-1}x'Y|Z = z,$$

where $x = g(z)$ is a matrix of fixed numbers.

Expectation of Least-Squares Estimator

So,

$$E[b(Y,Z)|Z=z] = (x'x)^{-1}x'E[Y|Z=z].$$

Since each (Y_i, Z_i) independent,

$$E[Y_i|Z=z] = E[Y_i|Z_i=z_i] = r(z_i).$$

So,

$$E[b(Y,Z)|Z=z] = (x'x)^{-1}x' \begin{pmatrix} r(z_1) \\ \vdots \\ r(z_n) \end{pmatrix}$$

Only used random sampling to get here. Can we say anything more without any additional assumptions?

Expectation of Least-Squares Estimator

We'd need to know something about $r(z)$. Do we know anything?

Expectation of Least-Squares Estimator

We'd need to know something about $r(z)$. Do we know anything?

No - $r(z)$ is unrestricted and without more assumptions, least-squares estimator is NOT unbiased for the linear predictor coefficients.

Why?

Expectation of Least-Squares Estimator

We'd need to know something about $r(z)$. Do we know anything?

No - $r(z)$ is unrestricted and without more assumptions, least-squares estimator is NOT unbiased for the linear predictor coefficients.

Why?

We can write

$$Y_i = \beta_1 X_1 + \ldots + \beta_K X_K + U_i,$$

where $U_i \perp X_1, \ldots, X_K$. The orthogonality conditions do NOT imply that

$$E[U_i | X_i = x] = 0.$$

Expectation of Least-Squares Estimator

Claim: $E[U_i|X_i] = 0$ if and only if $E[U_i h(X_i)] = 0$ for all functions X_i.

Proof:

Expectation of Least-Squares Estimator

Claim: $E[U_i|X_i] = 0$ if and only if $E[U_i h(X_i)] = 0$ for all functions X_i.

Proof:

$(\Rightarrow)$ $E[U_i h(X_i)] = E[E[U_i|X_i]h(X_i)] = 0.$

Expectation of Least-Squares Estimator

Claim: $E[U_i|X_i] = 0$ if and only if $E[U_i h(X_i)] = 0$ for all functions X_i.

Proof:

($\Rightarrow$) $E[U_i h(X_i)] = E[E[U_i|X_i]h(X_i)] = 0.$

($\Leftarrow$) By the projection theorem, the unique solution h^* to

$$\min_h E[(U - h(X))^2]$$

satisfies $E[(U - h^*(X))h(X)] = 0$ and we know that $h^*(X) = E[U|X]$ and $h(X) = 0$ satisfies the orthogonality conditions. By the uniquenesss of the solution to the projection problem, $E[U|X] = 0$.

Expectation of Least-Squares Estimator

Claim: $E[U_i|X_i] = 0$ if and only if $E[U_i h(X_i)] = 0$ for all functions X_i.

Proof:

($\Rightarrow$) $E[U_i h(X_i)] = E[E[U_i|X_i]h(X_i)] = 0$.

($\Leftarrow$) By the projection theorem, the unique solution h^* to

$$\min_h E[(U - h(X))^2]$$

satisfies $E[(U - h^*(X))h(X)] = 0$ and we know that $h^*(X) = E[U|X]$ and $h(X) = 0$ satisfies the orthogonality conditions. By the uniquenesss of the solution to the projection problem, $E[U|X] = 0$.

$E[U_i|X_i] = 0$ is only true for the "conditional expectation" projection problem. Not true for the best linear predictor.

Expectation of the Least Squares Estimator

If we make more assumptions, we can get more!

Suppose the linear predictor is equal to the conditional expectation function:

$$r(z_i) = x_i'\beta.$$

Then,

$$E[b|Z=z] = (x'x)^{-1}x' \begin{pmatrix} x_1' \\ \vdots \\ x_n' \end{pmatrix} \beta = \beta$$

and so,

$$E[b] = \beta.$$

So if $r(z_i) = x_i'\beta$, then b is an unbiased estimator for β.

Outline

Covariance

Recall: Let $S = (S_1, \ldots, S_T)'$ be a vector of random variables. Then,

$$\underset{T \times T}{Cov(S)} = \begin{pmatrix} Cov(S_1, S_1) & \ldots & Cov(S_1, S_t) \\ \vdots & & \vdots \\ Cov(S_T, S_1) & \ldots & Cov(S_T, S_T) \end{pmatrix}$$

Property: Let a_1 be an $M \times T$ matrix of scalars. Then,

$$Cov(a_1 S) = a_1 Cov(S) a_1'.$$

Covariance Matrix of Least Squares

Once again, we condition on $Z = z$. What's $Cov(b(Y, Z)|Z = z)$?

Covariance Matrix of Least Squares

Once again, we condition on $Z = z$. What's $Cov(b(Y, Z)|Z = z)$?

Simple:

$$\begin{aligned} Cov(b(Y, Z)|Z = z) &= Cov((X'X)^{-1}X'Y|Z = z) \\ &= Cov((x'x)^{-1}x'Y|Z = z) \\ &= (x'x)^{-1}x'Cov(Y|Z = z)x(x'x)^{-1} \end{aligned}$$

Covariance Matrix of Least Squares

What's $Cov(Y|Z = z)$? We have that

$$Cov(Y|Z = z) = \begin{pmatrix} V(Y_1|Z_1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & V(Y_n|Z_n) \end{pmatrix}$$

If we additionally assume **homoskedasticity**

$$V(Y_i|Z_i = z) = \sigma^2,$$

then

$$Cov(Y|Z = z) = \sigma^2 I_n$$

and

$$Cov(b|Z = z) = \sigma^2 (x'x)^{-1}.$$

Outline

Summarizing

Under random sampling, if $E[Y_i|Z_i = z_i] = x_i'\beta$ and $V(Y_i|Z_i = z_i) = \sigma^2$, then

$$E[b|Z = z] = \beta, \quad Cov(b|Z = z) = \sigma^2(x'x)^{-1}.$$