The Population Regression Function

Gianluca Cubadda
Università di Roma "Tor Vergata"

8th November 2017

Joint, conditional and marginal distributions

Consider the following k-dimensional random variable

$$Z = \left[\begin{array}{c} y \\ X^* \end{array} \right],$$

where y is a scalar and $X^* = [x_2, ..., x_k]'$.
We will focus our attention on the conditional distribution of y given X^*:

$$f(y|X^*) = \frac{f_z(y, X^*)}{f_{X^*}(X^*)},$$

where $f_z(y, X^*)$ is the joint distribution of y and X, and $f_{X^*}(X^*)$ is the marginal distribution of X^*.

The conditional expectation

- **Assumption 1:** the first and the second moment of $y|X^*$ exist.
- **Assumption 2:** the expected value of $y|X^*$ is a linear function of the elements of X^*:

$$\mathrm{E}(y|X^*) = \beta_1 + \sum_{j=2}^{k} \beta_j x_j \equiv \beta' X$$

  where $\beta = [\beta_1, ..., \beta_k]'$, and $X = [1, X^{*\prime}]'$.
- **Remark 1**: Ass. 2 holds if Z is multivariate Gaussian or if y and X^* are both scalar Bernoulli variables.

The error term

Under Ass.'s 1 & 2, we can write the population regression equation:

$$y = \beta' X + \varepsilon,$$

where β is a k-vector of regression coefficients and ε is denoted as the error term.
Notice that by construction we have

$$\mathrm{E}(\varepsilon | X^*) = 0$$

By the law of iterate expectations and in view of the equation above we get

$$\begin{aligned} \mathrm{E}(\varepsilon) &= \mathrm{E}_{X^*}\mathrm{E}(\varepsilon|X^*) = 0 \\ \mathrm{Cov}(X^*, \varepsilon) &= \mathrm{E}(X^*\varepsilon) = \mathrm{E}_{X^*}\mathrm{E}(X^*\varepsilon|X^*) = \mathrm{E}_{X^*}\left(X^*\mathrm{E}(\varepsilon|X^*)\right) = 0 \end{aligned}$$

Hence, the error term ε has zero mean and it is uncorrelated with X^*.

The regression intercept

We can rewrite the regression equation as follows

$$y = \beta_1 + \beta^{*\prime} X^* + \varepsilon,$$

where $\beta^* = [\beta_2, ..., \beta_k]'$.
Keeping in mind that the error term ε has zero mean, we get

$$\beta_1 = \mathrm{E}(y) - \beta^{*\prime}\mathrm{E}(X^*)$$

By inserting the second equation into the first one, we get the demeaned regression equation

$$\overline{y} = \beta^{*\prime}\overline{X}^* + \varepsilon$$

where $\overline{y} = y - \mathrm{E}(y)$ and $\overline{X}^* = X^* - \mathrm{E}(X^*)$.

The regression slopes

Keeping in mind that $\mathrm{Cov}(\varepsilon, X^*) = 0$, if we post-multiply both sides of the demeaned regression equation with $\overline{X}^*$ and take expectations, we get

$$\mathrm{Cov}(y, X^*) = \beta^{*\prime}\mathrm{Var}(X^*)$$

where $\mathrm{Cov}(y, X^*) = \mathrm{E}(\overline{y}\overline{X}^{*\prime})$ and $\mathrm{Var}(X^*) = \mathrm{E}(\overline{X}^*\overline{X}^{*\prime})$.
Since $\mathrm{Var}(X^*)$ is by construction a positive definite matrix, from the previous equation we easily obtain

$$\beta^* = \mathrm{Var}(X^*)^{-1}\mathrm{Cov}(X^*, y)$$

where $\mathrm{Cov}(X^*, y) = \mathrm{E}(\overline{X}^*\overline{y}) = \mathrm{Cov}(y, X^*)'$.

The conditional variance

Assumption 3: the variance of $y|X^*$ is a constant function of X^*: $\mathrm{Var}(y|X^*) = \sigma^2$.
Remark 2: Ass. 3 implies that the conditional and the unconditional variance of the error term ε are both equal to σ^2. Indeed, by taking the conditional variance on both sides of the demeaned regression equation, we get

$$\mathrm{Var}(y|X^*) = \beta^{*\prime}\underbrace{\mathrm{Var}(X^*|X^*)}_{=0}\beta^* + \mathrm{Var}(\varepsilon|X^*) + 2\beta^{*\prime}\underbrace{\mathrm{E}(X^*\varepsilon|X^*)}_{=X^*\mathrm{E}(\varepsilon|X^*)=0}$$

Moreover, by the law of iterated expectations, we get

$$\mathrm{Var}(\varepsilon) = \mathrm{E}(\varepsilon^2) = \mathrm{E}_{X^*}\mathrm{E}(\varepsilon^2|X^*) = \mathrm{E}_{X^*}\mathrm{Var}(\varepsilon|X^*) = \sigma^2$$

The error variance

Keeping in mind that the error term ε and the variables X^* are uncorrelated, it is easy to get

$$\begin{aligned}\mathrm{Var}(\varepsilon) &= \mathrm{Var}(y) - \beta^{*\prime}\mathrm{Var}(X^*)\beta^* \\ &= \mathrm{Var}(y) - \mathrm{Cov}(y, X^*)\mathrm{Var}(X^*)^{-1}\mathrm{Cov}(X^*, y)\end{aligned}$$

Question: Do the following formulae remind you something?

$$\begin{aligned}\mathrm{E}(\overline{y}|X^*) &= \mathrm{Cov}(X^*, y)\mathrm{Var}(X^*)^{-1}\overline{X}^* \\ \mathrm{Var}(y|X^*) &= \mathrm{Var}(y) - \mathrm{Cov}(y, X^*)\mathrm{Var}(X^*)^{-1}\mathrm{Cov}(X^*, y)\end{aligned}$$

The random sample

Simple random sampling from $f(y|X^*)$ provides us with the i.i.d. sequence

$$\{y_i, x_{2i}, ..., x_{ki}; i = 1, ..., n\}$$

Under Ass.'s 1-3, we get the system:

$$y_i = \beta_1 + \sum_{j=2}^{k} \beta_j x_{ij} + \varepsilon_i \quad i = 1, ..., n$$

Remark 3: since the sample is drawn from the conditional distribution $f(y|X^*)$, we can treat the variables

$$\{x_{2i}, ..., x_{ki}; i = 1, ..., n\}$$

as fixed rather than stochastic. This implies that

$$\mathrm{E}(y_i) = \beta_1 + \sum_{j=2}^{k} \beta_j x_{ij}$$

$$\mathrm{Var}(y_i) = \mathrm{Var}(\varepsilon_i) = \sigma^2$$

for $i = 1, ..., n$.