Note on : **EXPECTED VALUE AND CONDITIONAL EXPECTATION**

Pierre Villedieu (Sciences Po, 2019)

1 Expected value

Let X be a random variable. Intuitively, the **expected value** of X (denoted $\mathbb{E}(X)$) is the value you get if you could compute the *average* ($\bar{X}$) over an infinite number of *realizations* of X. It is also referred as the *theoretical mean* or *first order moment*.

We can express the expected value in two different way according to the case :

- Finite case (X can only take a finite number of values) :

$$\boxed{\mathbb{E}(X) = \sum_{k=1}^{K} x_k * p_k} \tag{1}$$

  where the x_k are the finite values taken by X with probabilities p_k.

- Continuous case, with $f(x)$ the density probability of X :

$$\boxed{\mathbb{E}(X) = \int_{\mathbb{R}} x f(x) dx} \tag{2}$$

In both cases, you can see the expected value as weighted average of the different values that X can take, each value being weighted according its probability to occur. In the continuous case these probabilities are the $f(x) * dx$, which is the probability associated to the event $\{X \in [x; x + dx]\}$

Taking a **concrete example**, the following figure is an illustration of a fair coin toss where you win 1 if "heads" (X("heads") = 1) and loose 1 if "tails" (X("tails") = -1) at each draw. Imagine that we repeat the experiment n times and we define X_i as the random variable giving the outcome of the i^{th} draw. Then, the Law of large numbers states that the sample average $\bar{X}_n$ converges[1] from the expected value, which here is 0.

Formalizing we have :

$$\bar{X}_n = \frac{1}{n} * \sum_{i=1}^{n} X_i \underset{n\to\infty}{\longrightarrow} \mathbb{E}(X) = 1/2 * 1 + 1/2 * (-1) = 0$$

[1] We won't go into the details of what "converges" exactly means here, you can just keep in mind that the sample average is becoming closer and closer from the actual expected value as n increases

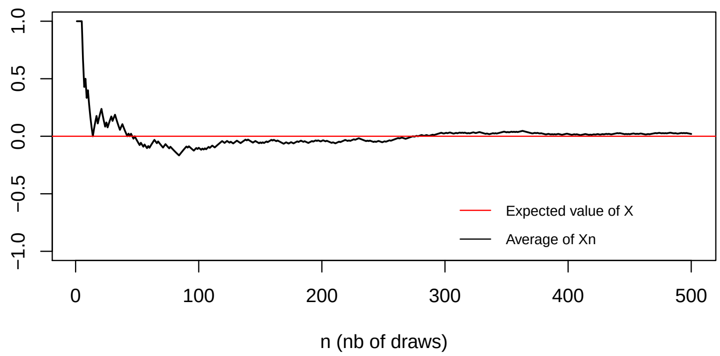

Figure 1: Law of large numbers

2 Conditional expectation

Definition We now focus on the expected value of a particular random variable (Y) but **conditioning on another random variable** (X). For example, you can think about the number of cars' accidents in a day conditioning on the weather, the life expectancy given the gender, ... Conceptually, the **conditional expectation** of Y with respect to X is itself a random variable and it will be denoted $\mathbb{E}(Y/X)$. In other words, it means that $\mathbb{E}(Y/X)$ is a function of X : the **Conditional Expectation Function**. This is actually the function of X that best approximates Y, where "best" means it minimize the mean square error.[2]. In concrete examples, you may be interested in computing the CEF at a given value of X : $X = x$.

As the expected value, the conditional expectation is a theoretical object. In particular, when thinking of common variables of interest in social sciences like income, education, employment status or life expectancy, this is a **population concept**. However most of the time, we only have access to a sample of this population. We thus compute a **conditional mean** which is the **empirical counterpart** of the conditional expected value, and is expected to give a proper approximation of the underlying population value. Conditional mean is a very standard object in every day life : every time you make a breakdown of a given quantitative variable by another variable you compute the associated conditional mean for each category. The figure 2 gives the CEF of the log of income by educational level. Each black dot is equal to the sample analog of $\mathbb{E}(Y/X = x)$. For certain values of education (X = 4, 8, 12, 16), the distribution of income among the people is represented in shaded area.

[2] $\mathbb{E}(Y - f(X))^2$ reaches is minimum for f = $\mathbb{E}(Y/X)$.

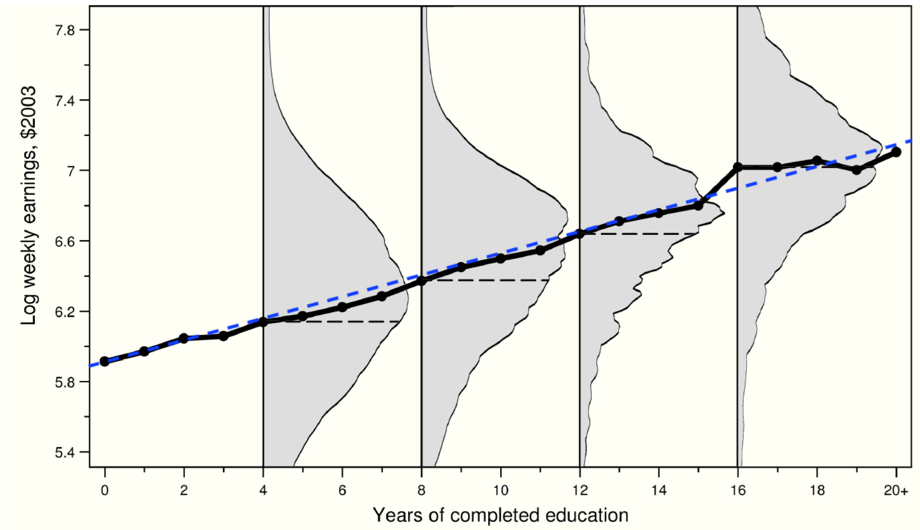

Figure 2: Mean income by educational level
(The sample includes white men aged 40-49 in the 1980 IPUMS 5 percent file)

(Source : Angrist, J., & Pischke, J. S. (2009))

Link with regression framework The conditional expectation function is also closely linked to the **regression framework** in econometrics. Actually, the standard **OLS** (Ordinary Least Squares) predictor provides the best linear approximation of the CEF (this is the blue dotted line added on the figure 2). To sum up, the CEF gives you the best unrestricted approximation of your dependent variable Y depending on X and the OLS estimator will give you the best linear approximation of this CEF.

3 Some useful properties

Ensure you can make an intuitive interpretation of these properties and think to concrete examples.

$\forall a, b \in \mathbb{R}$, X, Y and Z random variables (defined on the same probability space) :

- Expected value of a constant : $\mathbb{E}(a) = a$
- **Linearity** :
  - $\mathbb{E}(aX + bY) = a\,\mathbb{E}(X) + b\,\mathbb{E}(Y)$
  - $\mathbb{E}(aY + bZ/X) = a\,\mathbb{E}(Y/X) + b\,\mathbb{E}(Z/X)$
- **Law of total expectation** : $\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y/X)$
  $\hookrightarrow$ Example (finite case) : Y = Income and X = Gender

$\mathbb{E}(Y) = \mathbb{E}(Y/"X = woman") * \mathbb{P}("X = woman") + \mathbb{E}(Y/"X = man") * \mathbb{P}("X = man") = \mathbb{E}(\mathbb{E}(Y/X)$

- If X and Y are independent : $\mathbb{E}(Y/X) = \mathbb{E}(Y)$. The realization of Y is not affected by the realization of X, so conditioning on X does not change the expected value of Y.

4 References

- For formal definition and concepts :
  - Knight, Keith (2000): Mathematical Statistics. Chapman & Hall/CRC, Boca Raton, FL.
  - Lecture notes of Gordan Žitković
- For application to econometric framework :
  - Angrist, J., & Pischke, J. S. (2009). Mostly harmless econometrics: an empiricist's guide. (see Ch 1.; 2.; 3.1 and 3.2)