

Probability Total Theorems

Anis Rezgui

Mathematics & Computer Science Department
INSAT - Carthage University - Tunis - TUNISIA and
Laboratoire d'Ingénierie Mathématiques LIM - Ecole Polytechnique de
Tunisie.

November 20, 2020

Introduction

Plan

- 1 Introduction
- 2 Total probabilities Theorem
- 3 Total expectation Theorem
- 4 Total variance Theorem
- 5 The determination coefficient
- 6 The linear Correlation Coefficient
- 7 Conclusion

Introduction

After careful study of this chapter you should be able to:

- 1 State and prove the total probabilities Theorem.
- 2 State the total expectation Theorem.
- 3 State the total variance Theorem.
- 4 Be aware of the importance of Total Theorems in applications.
- 5 Understand the determination coefficient and its relation with the correlation coefficient.

Total probabilities Theorem

Plan

- 1 Introduction
- 2 Total probabilities Theorem**
- 3 Total expectation Theorem
- 4 Total variance Theorem
- 5 The determination coefficient
- 6 The linear Correlation Coefficient
- 7 Conclusion

Total probabilities Theorem

Theorem

Let $(\Omega, \Sigma, \mathbb{P})$ be a given probability space and A, B two events then:

$$\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|\overline{B})\mathbb{P}(\overline{B}).$$

Total probabilities Theorem

Theorem

Let $(\Omega, \Sigma, \mathbb{P})$ be a given probability space and A, B two events then:

$$\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|\overline{B})\mathbb{P}(\overline{B}).$$

Proof:

The proof is straightforward, we just need to recall the trivial partition of Ω :

$$\Omega = B \cup \overline{B}$$

and so

$$A = (A \cap B) \cup (A \cap \overline{B})$$

and finally

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B}) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|\overline{B})\mathbb{P}(\overline{B}).$$

Total probabilities Theorem

Note

- ① Although it is easy, the latter result is very important in probability, in fact, it simplifies the computation of $\mathbb{P}(A)$ by conditioning on any given B and its complement \overline{B} .
- ② It can be generalized trivially to any complete system $\{B_1, \dots, B_p\}$:

$$\mathbb{P}(A) = \sum_{i=1}^p \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

where "complete" means that the B_i 's are mutually disjoint and that they cover the whole sample space Ω .

Total expectation Theorem

Plan

- 1 Introduction
- 2 Total probabilities Theorem
- 3 Total expectation Theorem**
- 4 Total variance Theorem
- 5 The determination coefficient
- 6 The linear Correlation Coefficient
- 7 Conclusion

Total expectation Theorem

The Total Expectation Theorem

Theorem

For any random pair (X, Y) we have

$$\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(Y).$$

Total expectation Theorem

Proof (discrete case)

$$\begin{aligned}\mathbb{E}(\mathbb{E}(Y|X)) &= \sum_i \mathbb{E}(Y|X = x_i) \times \mathbb{P}\{X = x_i\} \\&= \sum_i \left\{ \sum_j y_j \mathbb{P}\{Y = y_j|X = x_i\} \right\} \times \mathbb{P}\{X = x_i\} \\&= \sum_j y_j \left\{ \sum_i \mathbb{P}\{Y = y_j|X = x_i\} \times \mathbb{P}\{X = x_i\} \right\} \\&= \sum_j y_j \mathbb{P}\{Y = y_j\} = \mathbb{E}(Y)\end{aligned}$$

Total expectation Theorem

Note

- 1 In certain cases, the direct computation of the expectation of Y may be difficult. To overcome to this difficulty we start by computing the expectation of $\mathbb{E}(Y|X)$ for any given X and then we use the Total expectation expectation Theorem to compute $\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|X))$.
- 2 One can express Y as:

$$Y = \mathbb{E}(Y|X) + Y - \mathbb{E}(Y|X)$$

and since $\mathbb{E}(Y - \mathbb{E}(Y|X)) = 0$ it gives sense to the decomposition of Y as the sum of a function of X and a **residual** part:

$$Y = \mathbb{E}(Y|X) + \text{"residual"}.$$

Total variance Theorem

Plan

- 1 Introduction
- 2 Total probabilities Theorem
- 3 Total expectation Theorem
- 4 Total variance Theorem**
- 5 The determination coefficient
- 6 The linear Correlation Coefficient
- 7 Conclusion

Total variance Theorem

The total variance Theorem

Theorem

Let (X, Y) be a random pair then

$$\mathbb{V}(Y) = \mathbb{V}(\mathbb{E}(Y|X)) + \mathbb{E}(\mathbb{V}(Y|X)).$$

Total variance Theorem

Proof

$$\begin{aligned}\mathbb{V}(Y) &= \mathbb{E}\left([Y - \mathbb{E}(Y)]^2\right) \\&= \mathbb{E}\left([Y - \mathbb{E}(Y|X) + \mathbb{E}(Y|X) - \mathbb{E}(Y)]^2\right) \\&= \mathbb{E}\left([Y - \mathbb{E}(Y|X)]^2\right) + \mathbb{E}\left([\mathbb{E}(Y|X) - \mathbb{E}(Y)]^2\right) + \\&\quad 2\mathbb{E}\left([Y - \mathbb{E}(Y|X)][\mathbb{E}(Y|X) - \mathbb{E}(Y)]\right) \\&= \mathbb{E}\left(\mathbb{V}(Y|X)\right) + \mathbb{V}\left(\mathbb{E}(Y|X)\right) + 2 \times 0\end{aligned}$$

this finishes the proof.

Total variance Theorem

Note

- 1 The total variance Theorem enables to decompose the variance of Y into two parts:

$$\mathbb{V}(Y) = \mathbb{V}(\mathbb{E}(Y|X)) + \mathbb{E}(\mathbb{V}(Y|X))$$

Total Variance = Explained Variance + Residual Variance.

The latter decomposition is known as "Analysis of Variance" and denoted in short ANOVA.

- 2 As much as the "residual variance" is small, compared to the "total variance", as strong as the relationship between Y and X is.
- 3 Equivalently, as much as the "explained variance" is large, compared to the "total variance", as strong as the relationship between Y and X is.

The determination coefficient

Plan

- 1 Introduction
- 2 Total probabilities Theorem
- 3 Total expectation Theorem
- 4 Total variance Theorem
- 5 The determination coefficient**
- 6 The linear Correlation Coefficient
- 7 Conclusion

The determination coefficient

The determination coefficient

As suggested by the last slide, the ratio

$$\frac{\mathbb{V}(\mathbb{E}(Y|X))}{\mathbb{V}(Y)} = \frac{\text{Explained Variance}}{\text{Total Variance}}$$

belongs to $[0, 1]$ and as much as it is close to one as strong as the relationship between X and Y is.

The latter ratio is called "The determination coefficient" and is denoted:

$$R^2_{Y|X}$$

The linear Correlation Coefficient

Plan

- 1 Introduction
- 2 Total probabilities Theorem
- 3 Total expectation Theorem
- 4 Total variance Theorem
- 5 The determination coefficient
- 6 The linear Correlation Coefficient**
- 7 Conclusion

The linear Correlation Coefficient

Covariance

Definition

Let (X, Y) be a random couple the covariance is quantity defined below

$$\begin{aligned}\text{Cov}(X, Y) &= \mathbb{E}\left((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right) \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\end{aligned}$$

- The discrete case

$$\mathbb{E}(XY) = \sum_{i,j} x_i y_j \mathbb{P}\{X = x_i, Y = y_j\}$$

- The continuous case

$$\mathbb{E}(XY) = \int_{\mathbb{R}} \int_{\mathbb{R}} xy f_{XY}(x, y) dx dy$$

The linear Correlation Coefficient

Linear Correlation Coefficient

Definition

Let (X, Y) be a random couple the covariance is quantity defined below

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

- $\rho_{XY} \in [-1, 1]$.
- $\rho_{XY} = \pm 1$ if and only if $Y = \frac{\sigma_Y}{\sigma_X}(X - \mathbb{E}(X)) + \mathbb{E}(Y)$.
- If $\rho_{XY} = 0$ then we can claim that there is no linear relationship between X and Y .

Conclusion

Plan

- 1 Introduction
- 2 Total probabilities Theorem
- 3 Total expectation Theorem
- 4 Total variance Theorem
- 5 The determination coefficient
- 6 The linear Correlation Coefficient
- 7 Conclusion**

Conclusion

Relation with the linear correlation coefficient

Theorem

Let (X, Y) be a random pair, then

$$\rho_{XY}^2 \leq R_{Y|X}^2$$

and there is equality if and only if

$$\mathbb{E}(Y|X) = \mathbb{E}(Y) + \rho_{XY} \frac{\sigma_Y}{\sigma_X} (X - \mathbb{E}(X))$$

Conclusion

So what

- ① If $\rho_{XY} \sim \pm 1$ or equivalently $\rho_{XY}^2 \sim 1$ then $\rho_{XY}^2 \simeq R_{Y|X}^2 \sim 1$ and then

$$Y \simeq \mathbb{E}(Y|X) \simeq \frac{\sigma_Y}{\sigma_X} (X - \mathbb{E}(X)) + \mathbb{E}(Y).$$

- ② If $\rho_{XY}^2 \sim R_{Y|X}^2$ then $\mathbb{E}(Y|X)$ is linear and is expressed as

$$\mathbb{E}(Y|X) = \rho_{XY} \frac{\sigma_Y}{\sigma_X} (X - \mathbb{E}(X)) + \mathbb{E}(Y),$$

but we don't know whether Y is significantly related to X !

- ③ If $\rho_{XY}^2 \ll R_{Y|X}^2$ then $\mathbb{E}(Y|X)$ is definitely non-linear.

- ④ If $R_{Y|X}^2 \sim 1$

$$Y \simeq \mathbb{E}(Y|X),$$

but we don't know whether this relation is linear or not.