

Probabilités, Statistiques, Combinatoire - CM 12 (2ème partie) - statistique

Adrian Tanasă

Université de Bordeaux – Licence Informatique

Plan cours

- ▶ échantillon et échantillonnage
- ▶ moyenne (espérance) et variance empirique
- ▶ échantillonnage et loi forte de grand nombres
- ▶ statistique bidimensionnelle

But de statistiques

Soit X une v.a d'espérance m et de variance σ^2 à priori inconnues.

but des statistiques : retrouver des informations sur X à partir d'informations parcellaires (l'échantillon)

Définition d'un échantillon

Définition.

Soit $n \in \mathbb{N}^*$ et soient X_1, \dots, X_n n v. a. On appelle **échantillon** de taille n toute suite (x_1, \dots, x_n) de réalisations des n variables aléatoires.

La façon d'obtenir cet échantillon se nomme l'échantillonnage.

Exemple.

vouloir connaître la taille de 12000 étudiants adultes. En mesurant seulement 100 étudiants, on obtient un échantillon de taille 100.

Échantillonnage

1. exhaustif
2. non-exhaustif

Moyenne (espérance) empirique

Définition :

Soit X une v.a d'espérance m et de variance σ^2 , soient X_1, \dots, X_n n variables aléatoires indépendantes de même loi que X , et soit (x_1, \dots, x_n) l'échantillon de taille n associé. On appelle moyenne (espérance) empirique de l'échantillon la quantité :

$$\tilde{m} = \frac{1}{n} (x_1 + \dots + x_n).$$

Moyenne (espérance) empirique

Définition :

Soit X une v.a d'espérance m et de variance σ^2 , soient X_1, \dots, X_n n variables aléatoires indépendantes de même loi que X , et soit (x_1, \dots, x_n) l'échantillon de taille n associé. On appelle variance empirique de l'échantillon la quantité :

$$\tilde{\sigma}^2 = \frac{1}{n} \left((x_1 - \tilde{m})^2 + \dots (x_n - \tilde{m})^2 \right).$$

Statistique bidimensionnelle

on relève simultanément un échantillon (x_1, \dots, x_n) et (y_1, \dots, y_n) des deux variables aléatoires X et Y .

Définition :

On appelle covariance empirique des échantillons la quantité

$$\tilde{\sigma}_{XY} = \frac{1}{n} \sum_{i=1}^n (x_i - \tilde{m}_X)(y_i - \tilde{m}_Y)$$

Définition :

On appelle coefficient de corrélation empirique des échantillons la quantité

$$\rho_{XY} = \frac{\tilde{\sigma}_{XY}}{\tilde{\sigma}_X \tilde{\sigma}_Y}$$