Probabilités et statistiques
Université Clermont Auvergne
IUT Info 2A

Chafik Samir
Adrien Wohrer
Année 2017-2018

TP5 : Bases des statistiques

Objectifs

1. *Statistiques descriptives :* Comprendre les propriétés de la **moyenne empirique** et de la **variance empirique**.
2. *Statistiques inférentielles :* Application du principe de base sur un premier exemple (estimer si un jeu de pile ou face est biaisé ou pas).

À récupérer sur l'ENT

- Un fichier de données *patinage_vitesse_2002.sce*, pour l'exercice 2.
- Un squelette de fichier de rendu, *tp5.sce*. Vous devez remplir **ce fichier** (pas un autre !) et le déposer sur l'ENT. Pas de dépôt par mail ! Ce fichier devra s'exécuter **sans erreur** lorsqu'on le lance, et afficher toutes vos réponses.

1 Propriétés de la moyenne empirique

Dans cette première partie, nous vérifions *par la simulation* les propriétés des moyenne et variance empiriques, vues en cours. Rappel : pour un échantillon aléatoire $\{X_1, \ldots, X_N\}$ de taille N, la moyenne m_N et la variance empiriques s_N^2 s'écrivent respectivement :

$$m_N := \frac{X_1 + \ldots + X_N}{N},$$
$$s_N^2 := \frac{X_1^2 + \ldots + X_N^2}{N} - m_N^2.$$

La loi de départ pour cet exercice est la loi exponentielle de paramètre $\lambda = 0.2$.

1. On considère un échantillon de $N = 2$ variables aléatoires X_1 et X_2, tirées suivant Exp(0.2). Tirez 10000 répétitions de cet échantillon, et calculez les 10000 valeurs correspondantes de m_N et s_N.
2. Utilisez ces valeurs pour :
   a) Tracer un histogramme de la distribution de m_N.
   b) Calculer une valeur (approchée) de $E(m_N)$.
   c) Calculer une valeur (approchée) de $E(s_N^2)$.
3. Reprendre les questions 1 et 2 pour les valeurs de N suivantes : $N = 5$, 10, 50, 200.
   - Affichez tous les histogrammes dans la même figure. (Utilisez **histplot** avec l'argument optionnel `style` pour différencier les différentes courbes.)
   - Dans une autre figure, tracez la courbe de $E(m_N)$ en fonction de N.
   - Dans une autre figure, tracez la courbe de $E(s_N^2)$ en fonction de N.
4. Quelles sont la moyenne et la variance théoriques de la distribution Exp(0.2) ? Déduisez-en que m_N est un bon estimateur de la moyenne.

2 Patinage de vitesse

Le patinage de vitesse est une discipline dans laquelle deux patineurs s'affrontent. L'épreuve se déroule sur une patinoire circulaire, chaque patineur ayant son *couloir* réservé.

Aux J.O. de Salt Lake City, en 2002, certains journaux ont critiqué la patinoire, affirmant que le couloir *intérieur* était plus lent que le couloir *extérieur*, si bien que les courses étaient systématiquement biaisées en faveur de l'un des deux patineurs. Le but de cet exercice est d'estimer si ces allégations étaient fondées, avec des outils de statistiques inférentielles.

Statistiques descriptives

Les données sont constituées des 23 courses ayant eu lieu sur la patinoire pendant les J.O. Elles sont fournies sur l'ENT, dans le fichier **patinage_vitesse_2002.sce**. La première colonne donne le temps du patineur à l'intérieur, et la seconde colonne le temps du coureur à l'extérieur.

1. On considère la variable aléatoire $\Delta =$ *temps(extérieur)-temps(intérieur)*. Affichez les 23 valeurs de Δ dans un histogramme.

2. Trouvez le vainqueur de chaque course, et calculez la proportion empirique de victoires à l'extérieur.

Test d'hypothèse

Dans une première étude, nous regardons si le pourcentage de victoires à l'extérieur est suspect.

Si on prend deux patineurs complètement au hasard, appelons p la probabilité que le coureur extérieur remporte la course. Ce nombre p est donc une caractéristique « intrinsèque » de la piste, qu'on cherche à estimer. On considère les deux hypothèses :

- H_0 : les chances de victoire ne sont pas biaisées ($p = 0.5$).
- H_1 : les chances de victoire sont biaisées vers l'extérieur ($p > 0.5$).

On note k_0 le nombre moyen de victoires à l'extérieur sur 23 courses, sous l'hypothèse H_0. On note $k^{(obs)}$ le nombre de victoires à l'extérieur observé en réalité.

3. Calculez k_0 et $k^{(obs)}$. Calculez le nombre $d = |k^{(obs)} - k_0|$, et dites ce qu'il représente.
4. Sous l'hypothèse H_0, quelle loi devrait suivre la v.a. K, nombre de victoires à l'extérieur sur 23 courses ? Calculez et stockez la loi de probabilité correspondante dans un vecteur **prob0**. Représentez-la dans un graphe.
5. Sous l'hypothèse H_0, estimez la probabilité $P(|K - k_0| \geq d)$.
   Indice : Trouvez toutes les valeurs possibles de k vérifiant $|k - k_0| \geq d$, et sommez les probabilités associées.
6. Concluez : au vu des données observées, peut-on rejeter l'hypothèse H_0 ?