3. Covariance, corrélation et estimation ponctuelle

3.1. **Variance de la somme.** Soient X et Y deux variables aléatoires pas nécesseriament indépendantes. Calculer la variance de $X + Y$ en fonction des variances de X et Y et de leur covariance.

3.2. **Covariance pour les dés.** On désigne par $Y^{(1)}$ le nombre de 1 et par $Y^{(6)}$ le nombre de 6 apparaissant à l'issue de n lancers indépendants d'un dé équitable.

(1) Quelle est la loi de probabilité de $Y^{(1)}$? de $Y^{(6)}$? de $Y^{(1)} + Y^{(6)}$?
(2) Quel signe attendez-vous pour la covariance de $Y^{(1)}$ et $Y^{(6)}$?
(3) Utiliser la formule pour la variance de la somme pour calculer la covariance et la corrélation de $Y^{(1)}$ et $Y^{(6)}$.

3.3. **Corrélation pour un geyser [R].** R a un jeu de données intégré nommé *faithful*. Il se compose d'une collection d'observations du geyser Old Faithful dans le parc national américain de Yellowstone. Pour avoir un aperçu du jeu, utiliser la fonction *head*. Le jeu de données contient 272 observations de durées d'éruption et intervalles de temps entre deux éruptions.

(1) Tracer l'histogramme des durées de éruptions.
(2) Tracer l'histogramme des intervalles d'attente.
(3) Tracer le diagramme de dispersion des durées d'éruption et des intervalles d'attente. Est-ce qu'il révèle une relation entre les variables ?
(4) Calculer la corrélation entre les 2 variables.

Suggestions: utiliser les fonctions R *hist*, *cor* et le symbole $ pour extraire une colonne du jeu de données (par exemple *faithful$eruptions*).

3.4. **Estimateur pour la v.a. exponentielle.** Soit $\{x_1, x_2, \ldots x_n\}$ un échantillon provenant d'une loi de densité exponentielle de paramètre λ. Déterminer l'estimateur du maximum de vraisemblance de λ. Est-il biasé ? Est-il convergent ?

3.5. **Estimateur de la probabilité de pile.** Considérez des tirages indépendantes à pile ou face, et soit p la probabilité de pile à chaque tirage.

(1) Fixez n et soit k le nombre de piles observées dans n tirages. Trouver l'estimateur du maximum de vraisemblance de p basé sur k.
(2) Fixez k et soit n le nombre de tirages jusqu'à que la k-ème pile se produise. Trouver l'estimateur du maximum de vraisemblance de p basé sur N.

3.6. **Estimation du paramètre d'une variable aléatoire uniforme.** On nous donne un échantillon de n observations $\{x_1, x_2, \ldots x_n\}$, qui sont tirées d'une distribution uniforme sur $[0, \theta]$. Quel est l'estimateur du maximum de vraisemblance de θ ? Est-il convergent ? Est-il sans biais ou asymptotiquement sans biais ? Pouvez-vous construire des estimateurs alternatifs qui sont sans biais ?

3.7. **Au restaurant.** Soit une certaine population P dont on sait que le pourcentage mangeant x fois au restaurant sur 3 jours est :

- $x = 0; 50\%$
- $x = 1; 20\%$
- $x = 2; 20\%$
- $x = 3; 10\%$

On considère un sous-ensemble R de la population (par exemple les personnes répondant au téléphone l'après-midi), en gros $1/4$ de P. Pour R les pourcentages sont :

- $x = 0; 76\%$
- $x = 1; 12\%$
- $x = 2; 8\%$
- $x = 3; 4\%$

(1) Déterminer le nombre moyen μ_P (resp. μ_R) de fois qu'une personne de P (resp. R) mange au restaurant.
(2) Un premier statisticien fait un sondage par téléphone dans l'après-midi jusqu'à ce qu'il obtienne 50 réponses. Il prend commme estimateur $\hat{\mu}_{P,1} = \frac{x_1+\ldots+x_{50}}{50}$ où x_i est la valeur donnée par la personne i. Quel est le biais de $\hat{\mu}_{P,1}$ $B(\hat{\mu}_{P,1}) := \mathrm{E}\left[\hat{\mu}_{P,1}\right] - \mu_P$? Quelle est l'erreur quadratique moyenne $\mathrm{E}\left[(\hat{\mu}_{P,1} - \mu_P)^2\right]$?
(3) Même question avec 200 réponses.
(4) Un autre statisticien fait un sondage sur un échantillon aléatoire de 20 personnes à qui il téléphone matin et après-midi jusqu'à ce qu'il obtienne une réponse et prend comme estimateur $\hat{\mu}_{P,2} = \frac{y_1+\ldots+y_{20}}{20}$ où y_i est la valeur donnée par la personne i. Calculer le biais $B(\hat{\mu}_{P,2})$ et l'erreur quadratique moyenne $\mathrm{E}\left[(\hat{\mu}_{P,2} - \mu_P)^2\right]$.
(5) Conclusion ?

3.8. **Niveau d'eau.** Un limnimètre peut mesurer le niveau d'un fleuve entre h_1 et h_2. On assume que le niveau suit une loi exponentielle avec espérance θ inconnue. Supposons que n mesures $\{x_1, x_2, \ldots x_n\}$ soient disponibles.

(1) Quelle est la fonction de vraisemblance ?
(2) Si $h_1 = 1$, $h_2 = 20$ and les mesures sont $\{1.5, 2, 3, 4, 5, 12\}$, trouvez approximativement l'estimation du maximum de vraisemblance de θ.