1 TD1

1.1 Intervalles de confiance (1).

Soit $Y_1, \ldots, Y_n$ une suite i.i.d. de loi de Bernoulli de paramètre $p \in (0,1)$.

1. (a) Proposer un estimateur $\widehat{p}_n$ de p de type moyenne empirique.
   (b) En utilisant l'inégalité de Tchébychev, construire un intervalle de confiance pour p de niveau $1-\alpha$, pour $0 < \alpha < 1$.
   (c) Montrer que $\widehat{p}_n$ est asymptotiquement normal. En déduire un intervalle de confiance asymptotique pour p de niveau $1-\alpha$.
2. Soient $Z_1, \ldots, Z_n$ des variables indépendantes avec $0 \leq Z_i \leq 1$ pour tout i. L'inégalité de Hoeffding, que l'on admettra, s'écrit

$$\mathbb{P}[|\overline{Z}_n - E[\overline{Z}_n]| \geq t] \leq 2e^{-2nt^2}, \quad \text{où} \quad \overline{Z}_n = n^{-1}\sum_{i=1}^{n} Z_i.$$

   (a) Déduire de cette inégalité un intervalle de confiance pour p de niveau $1-\alpha$.
   (b) Comparer avec les intervalles contruits précédemment.

1.2 Modèle statistique, identifiabilité

On s'intéresse à des colonies d'insectes pondeurs d'oeufs. On suppose que pour chacun des n sites observés le nombre X_i d'oeufs suit une loi de Poisson de paramètre $\lambda > 0$ et que chaque oeuf pondu donne un insecte avec une probabilité p (le développement des oeufs est supposé indépendant de X_i et les différents sites sont indépendants). On notera Y_i le nombre de naissance au site i.

1. On suppose qu'on observe à la fois X et Y.
   (a) Déterminer la loi conditionnelle de Y_1 sachant $X_1 = x_1$.
   (b) Ecrire le modèle statistique associé.
   (c) Le modèle est-il identifiable ?
   (d) Donner un estimateur de λ et p.
2. On n'observe plus que Y.
   (a) Ecrire le modèle statistique associé.
   (b) Quelle est la loi de Y ?
   (c) Le modèle est-il identifiable ?

1.3 Lois conditionnelles (1)

Soient X, Y deux variables aléatoires réelles de loi jointe de densité sur $\mathbb{R}^2$ donnée par

$$(x,y) \to \frac{1}{2\pi}\exp\left(-x^2 + xy - \frac{y^2}{2}\right).$$

1. Déterminer les quantités suivantes

   (a) les densités marginales f_X et f_Y de X et Y et les lois marginales associées,
   (b) la loi conditionnelle de Y sachant $X = x$ et celle de X sachant $Y = y$.
2. Vérifier par ailleurs que la loi de (X, Y) est celle d'un vecteur gaussien sur $\mathbb{R}^2$, dont on précisera la moyenne et la matrice de variance-covariance.

1.4 Lois conditionnelles (2)

Soient Y, Z des variables aléatoires indépendantes de loi exponentielle de densité $\lambda e^{-\lambda y}\mathbb{1}_{\mathbb{R}^+}(y)$. On pose $X = Y + Z$.

1. Quelle est la densité du couple (Y, Z) ? En déduire la densité du couple (X, Y) sur $\mathbb{R}^2$, puis la densité de X.
2. Déterminer la loi conditionnelle de Y sachant $X = x$, puis l'espérance conditionnelle $E[Y \mid X = x]$. Commenter. Quelle est la loi conditionnelle de Z sachant $X = x$?

1.5 Tirages de pile ou face

On considère une expérience de tirage de pile–ou–face. Soit P_θ la loi de Bernoulli de paramètre $\theta \in [0, 1]$. Le modèle considéré est $\mathcal{P} = \{P_\theta,\ \theta \in \Theta\}$, avec $\Theta \subset [0, 1]$. On dispose par ailleurs d'observations $X_1, \ldots, X_n$ à valeurs dans $\{0, 1\}$.

1. Dans cette question, $\Theta = \{a, b\}$ avec $a = \frac{1}{3}$ et $b = \frac{1}{2}$. On note Π la loi sur l'ensemble Θ définie par
$$\Pi[\{a\}] = \Pi[\{b\}] = \frac{1}{2}.$$
   On regarde d'abord le cas d'une observation, soit $n = 1$. On pose le modèle suivant pour l'observation X_1 et pour θ,
$$\theta \sim \Pi$$
$$X_1 \mid \theta \sim P_\theta.$$
   (a) En quoi s'agit-il d'une approche bayésienne ? Quelle est la loi a priori ?
   (b) Calculer la loi a posteriori $\theta \mid X_1$.
   (c) Donner également la loi jointe de (X_1, θ) et la loi marginale de X_1

   Puis, on pose le modèle suivant pour les observations $X_1, \ldots, X_n$ et pour θ.
$$\theta \sim \Pi$$
$$X_1, \ldots, X_n \mid \theta \sim P_\theta^{\otimes n}.$$
   (d) De même, déterminer la loi a posteriori $\theta \mid X_1, \ldots, X_n$.
   (e) Calculer l'espérance a posteriori $E[\theta \mid X_1, \ldots, X_n] = \int \theta d\Pi(\theta \mid X_1, \ldots, X_n)$.
2. Thomas Bayes (1763) considère le problème suivant. Une boule de billard roule sur une ligne de longueur 1, avec une probabilité uniforme de s'arrêter en un point. Supposons qu'elle s'arrête en p. Une deuxième boule roule n fois dans les mêmes conditions, et on note X le nombre de fois où elle s'est arrêtée *à gauche* de la première boule. Bayes se demande : connaissant X, quelle inférence peut-on mener sur p ?
   (a) Ecrire cette expérience dans un formalisme bayésien (c'est le cas de le dire !), où Θ est cette fois l'intervalle $[0, 1]$. Quelle est le paramètre, la loi a priori ?
   (b) Répondre à la question de Bayes en calculant la densité a posteriori.