Estimation

Exercice 1

On procède à une série de mesures avec un même appareil. On suppose que le résultat d'une mesure est une v.a. X suivant une loi normale de moyenne inconnue m et d'écart-type 0.5. On a observé la série de valeurs suivantes :

12.5 15.8 14.7 15.3 15.2 15.1 15.6 14.9 14.9

1. Construire un intervalle de confiance I_1 de niveau 90 % pour m.
2. Construire un intervalle de confiance I_2 de niveau 98 % pour m.
3. Comment obtenir un intervalle de confiance qui ait même longueur que I_1 et même niveau de confiance que I_2 ?

Exercice 2

Une usine fabrique des câbles. On suppose que la charge maximale supportée par un câble, exprimée en tonnes, est une v.a. normale de moyenne m inconnue et d'écart-type 0.3). Une étude portant sur 50 câbles a donné une moyenne des charges maximales supportées égale à 12.2 tonnes.

1. Déterminer un intervalle de confiance à 99 % de la charge maximale moyenne de tous les câbles fabriqués par l'usine.
2. Quelle doit être la taille minimale de l'échantillon étudié pour que la longueur de l'intervalle de confiance à 99 % soit inférieure ou égale à 0.2 ?

Exercice 3

Une entreprise fabrique une pièce de moteurs industriels. Parfois ces pièces se révèlent immédiatement défaillantes après la vente. Le taux de défaillance doit être limité à 0.04. Sur 500 pièces contrôlées, 28 sont défaillantes.

1. Donner un intervalle de confiance pour le taux en question (avec le risque de 5 %).
2. La norme de qualité de production est-elle respectée ?

Exercice 4

On dit qu'une v.a. Y suit la loi « log-normale » de paramètres (m, σ) si $X = \ln Y$ suit la loi normale paramètres (m, σ^2). On dispose d'un échantillon de neuf observations d'une loi log-normale de paramètres (m, σ) inconnus :

11.3 2.1 1.1 8.9 4.6 5.7 13.5 24.5 16.4

1. Construire un intervalle de confiance de niveau 90 % pour m .
2. Même question pour σ.

Exercice 5

Une chaîne de fabrication produit un pourcentage p d'appareils défectueux. Le paramètre p est inconnu mais on sait que $p \leqslant 0.08$. On voudrait construire un estimateur sans biais de p de précision supérieure à $1/0.002$. Quelle doit être la taille de l'échantillon à prélever ?

Exercice 6

Une machine fabrique des pièces dont la longueur est une v.a. X acceptable si elle est comprise entre 28.18 et 28.22 mm.

1. Dans un premier temps, on estime que X suit la loi normale de moyenne 28.20 et d'écart-type 0.012. Calculer le pourcentage de pièces acceptables auquel il faut s'attendre.
2. Sans mettre en cause l'écart-type estimé $\sigma = 0.012$, on désire vérifier la valeur moyenne de la longueur des pièces produites m par l'analyse d'un échantillon. Quelle doit être la taille de l'échantillon qui placera m dans un intervalle d'amplitude 0.01mm centré sur $\widehat{m}$, moyenne de l'échantillon au risque de 1 % ?
3. On prélève un échantillon de 50 pièces qui donne le résultat suivant : moyenne de l'échantillon 28.195, écart-type de l'échantillon 0.015. Déterminer l'intervalle centré sur $\widehat{m}$ où se trouve m au risque 1 %.
4. On complète l'analyse par une étude des retours de l'atelier de montage pour non-conformité. Sur 250 pièces expédiées, 21 sont retournées. Déterminer la probabilité qu'une pièce soit hors tolérance, dans un intervalle de confiance de niveau 95 %.

Exercice 7

Soit $(X_1, \dots, X_n)$ un échantillon de la loi normale de paramètres (m, σ^2) où m et σ sont inconnus et soit $k > 0$. On pose $\widehat{m}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ et $\widehat{s_n^2} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \widehat{m}_n)^2$.

1. En se rappelant que $n\widehat{s_n^2}/\sigma^2$ suit la loi du Khi-Deux à $(n-1)$ d.d.l., calculer le moment d'ordre k de $\widehat{s_n^2}$, $\mathbb{E}[(\widehat{s_n^2})^k]$.
2. Trouver la constante $\alpha_{k,n}$ telle que $\widehat{\sigma_n^{2k}} = \alpha_{k,n}(\widehat{s_n^2})^k$ soit un estimateur sans biais de σ^{2k}. Examiner les cas particuliers $k = 1$ et $k = \frac{1}{2}$.
3. Montrer que $\widehat{\sigma_n^{2k}}$ est un estimateur correct.
   On donne $\dfrac{\Gamma(x+a)}{\Gamma(x+b)} = x^{a-b}\Big[1 + \dfrac{(a-b)(a+b-1)}{2x} + o\Big(\dfrac{1}{x}\Big)\Big]$ lorsque $x \to +\infty$.

Exercice 8

Le nombre d'appels téléphoniques entre 9 h et 10 h est une v.a. X suivant une loi de Poisson de paramètre inconnu λ. On cherche à évaluer la probabilité $\mathbb{P}(X = 0) = e^{-\lambda} = \mu$ qu'il n'y ait pas d'appel pendant ce laps de temps. On note $X_1, \dots, X_n$ les nombres d'appels (aléatoires) sur n jours durant cet intervalle de temps. $(X_1, \dots, X_n)$ est un échantillon de X. Désignons la moyenne de ces appels par $\widehat{m} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ et le nombre de jours sans appel par $n\hat{F} = \displaystyle\sum_{i=1}^{n} \mathbb{1}_{\{X_i = 0\}}$.

On propose trois estimateurs différents pour μ : $\widehat{\mu}_1 = \hat{F}$; $\widehat{\mu}_2 = e^{-\widehat{m}}$; $\widehat{\mu}_3 = \Big(1 - \dfrac{1}{n}\Big)^{n\widehat{m}}$.

1. Déterminer le biais et la variance de chacun de ces estimateurs.
2. Etudier la convergence de chacun de ces estimateurs, puis comparer leur précision.

Exercice 9

La hauteur maximale H de la crue annuelle d'un fleuve est soigneusement observée, car une crue supérieure à 6 mètres serait catastrophique. On a modélisé la loi de la v.a. H comme étant de Rayleigh, i.e. de densité

$$f_H(x) = \frac{x}{\theta}\, e^{-\frac{x^2}{2\theta}}\, \mathbb{1}_{[0,+\infty[}(x),$$

où θ est un paramètre inconnu. Pendant 8 ans on a observé les hauteurs de crues suivantes en mètres :

2.5	2.9	1.8	0.9	1.7	2.1	2.2	2.8

1. Donner l'estimateur du maximum de vraisemblance de θ.
2. Une compagnie d'assurance estime qu'une catastrophe n'arrive en moyenne qu'au plus une fois tous les mille ans. Cette estimation peut-elle être justifiée par les observations ?

Exercice 10

On considère une v.a. X de densité $f_X(x) = e^{-(x-\theta)}\mathbb{1}_{[\theta,+\infty[}(x)$, où $\theta > 0$ est un paramètre inconnu. Soit $(X_1, \dots, X_n)$ un échantillon de X.

1. (a) Donner les valeurs de $\mathbb{E}(X)$ et $\operatorname{var}(X)$.
   (b) Trouver un estimateur sans biais simple T_n de θ. Etudier sa convergence.
   (c) Donner une loi approchée de T_n pour $n \geqslant 100$. En déduire un intervalle de confiance pour θ de crédibilité 0.9.
2. On se propose maintenant d'étudier l'estimateur du maximum de vraisemblance U_n de θ.
   (a) Exprimer U_n en fonction de $X_1, \dots, X_n$.
   (b) Déterminer la fonction de répartition F_{U_n} ainsi que la densité f_{U_n} de U_n.
   (c) Calculer la précision de U_n.
3. Comparer les estimateurs T_n et U_n. Lequel est préférable ?

Exercice 11

Soient $X_1, \ldots, X_n$ des v.a. indépendantes issues d'une loi normale centrée de variance θ. On note $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ et on définit deux statistiques U et V par :

$$U = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n-1}\Big[\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\Big] \quad \text{et} \quad V = \frac{1}{n(n-1)}\Big[(n-2)\sum_{i=1}^{n} X_i^2 + (n\overline{X})^2\Big].$$

1. Montrer que U et V sont des estimateurs sans biais de θ.
2. Calculer leur variance. (On a $\mathbb{E}\Big[\Big(\sum_{i=1}^{n} X_i^2\Big)^2\Big] = n(n+2)\theta^2$, $\mathbb{E}\Big[\overline{X}^2\sum_{i=1}^{n} X_i^2\Big] = \frac{n+2}{n}\,\theta^2$, $\mathbb{E}(\overline{X}^4) = \frac{3}{n^2}\,\theta^2$.)
3. Soit $T_a = aU + (1-a)V$ avec $a \in \mathbb{R}$. Montrer que T_a est un estimateur sans biais de θ et donner la valeur de a qui minimise $\operatorname{var}(T_a)$.

Exercice 12

L'organisateur d'une exposition picturale s'intéresse au rythme d'arrivée des groupes de visiteurs à partir des observations faites au cours des premières journées. Il constate que le temps séparant l'arrivée de deux groupes successifs peut être assimilé à une v.a. uniforme sur $[0,\theta]$ et que les temps inter-arrivées sont indépendants. Pour l'organisation ultérieure des caisses réservées aux entrées des groupes, il souhaite estimer avec précision le paramètre θ, ayant à sa dispostion un échantillon $(X_1, \ldots, X_n)$ de X.

1. (a) Donner les valeurs de $\mathbb{E}(X)$ et $\operatorname{var}(X)$.
   (b) En déduire la valeur de α telle que $T_n = \alpha\widehat{m}_n$ soit un estimateur sans biais du paramètre θ, où $\widehat{m}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.
   (c) Montrer que T_n est un estimateur correct.
2. On se propose maintenant d'étudier l'estimateur du maximum de vraisemblance U_n de θ.
   (a) Exprimer U_n en fonction de $X_1, \ldots, X_n$.
   (b) Déterminer la fonction de répartition F_{U_n} ainsi que la densité f_{U_n} de U_n, puis calculer $\mathbb{E}(U_n)$ et $\operatorname{var}(U_n)$.
   (c) Déterminer β_n tel que $V_n = \beta_n U_n$ soit un estimateur sans biais de θ.
   (d) Montrer que V_n est correct.
3. (a) Comparer les estimateurs T_n et V_n.
   (b) Calculer les valeurs numériques de T_n et V_n observées pour $n = 75$, $\sum_{i=1}^{n} x_i = 28$ h et $\max(x_1, \ldots, x_n) = 0.71$ h.

Exercice 13

Soit $X_1, \ldots, X_n$ des v.a. indépendantes issues d'une loi uniforme sur $[k\theta, (k+1)\theta]$ où $k > 0$ est connu et $\theta > 0$ est inconnu. On note $X_{(1)} = \min(X_1, \ldots, X_n)$ et $X_{(n)} = \max(X_1, \ldots, X_n)$.

1. (a) Trouver l'estimateur du maximum de vraisemblance de θ que l'on notera U.
   (b) Déterminer la fonction de répartition F_U ainsi que la densité f_U de U.
   (c) Calculer $\mathbb{E}(U)$ et $\operatorname{var}(U)$.
   (d) En déduire le risque $\mathbb{E}[(U-\theta)^2]$.
2. Soit $V = \frac{1}{k}X_{(1)}$ un autre estimateur de θ.
   (a) Déterminer la fonction de répartition F_V ainsi que la densité f_V de V.
   (b) Calculer $\mathbb{E}(V)$ et $\operatorname{var}(V)$.
   (c) En déduire le risque $\mathbb{E}[(V-\theta)^2]$.
3. Soit $T_a = aU + (1-a)V$ avec $\alpha \in \mathbb{R}$ une famille d'estimateurs de θ. Trouver la valeur de a qui minimise le risque $\mathbb{E}[(T_a-\theta)^2]$. On donne $\mathbb{E}[X_{(1)}X_{(n)}] = \Big[k(k+1) + \frac{1}{n+2}\Big]\frac{\theta^2}{k(k+1)}$.

Exercice 14

Prouver le résultat suivant :

Il existe presque-sûrement au plus un estimateur d'un paramètre θ sans biais et de variance minimale.

(Indication : si $\widehat{\theta}_1$ et $\widehat{\theta}_2$ sont deux tel estimateurs, introduire $\widehat{\theta}_3 = \frac{1}{2}(\widehat{\theta}_1 + \widehat{\theta}_2)$. Calculer $\mathbb{E}(\widehat{\theta}_3)$ et $\mathrm{var}\,(\widehat{\theta}_3)$, puis remarquer que l'on doit nécessairement avoir $\mathrm{var}\,(\widehat{\theta}_3) \geqslant V$ où $V = \mathrm{var}\,(\widehat{\theta}_1) = \mathrm{var}\,(\widehat{\theta}_2)$ est la variance minimale. Conclure.)

Exercice 15 (Statistiques exhaustives)

Soit X une v.a. continue (resp. discrète) de densité $x \mapsto f_X(x;\theta)$ (resp. de loi $x \mapsto p_X(x;\theta) = \mathbb{P}(X = x)$) et $(X_1, \dots, X_n)$ un échantillon de X. On introduit la vraisemblance de $(X_1, \dots, X_n)$: $L(x_1, \dots, x_n; \theta) = f_X(x_1;\theta) \dots f_X(x_n;\theta)$ (resp. $p_X(x_1;\theta) \dots p_X(x_n;\theta)$).

On dit qu'une statistique de $(X_1, \dots, X_n)$, $T = \varphi(X_1, \dots, X_n)$, de densité $t \mapsto f_T(t;\theta)$ (resp. de loi discrète $t \mapsto p_T(t;\theta) = \mathbb{P}(T = t)$) est exhaustive pour θ s'il existe une fonction h indépendante de θ telle que :

$$L(x_1, \dots, x_n; \theta) = f_T(t;\theta) h(x_1, \dots, x_n) \quad (\text{resp. } p_T(t;\theta) h(x_1, \dots, x_n))$$

où $t = \varphi(x_1, \dots, x_n)$. La densité conditionnelle de $(X_2, \dots, X_n)$ sachant $T = t$ est alors donnée par

$$f_{(X_2,\dots,X_n)|T=t}(x_1, \dots, x_n) = \frac{h(x_1, \dots, x_n)}{\frac{\partial \varphi}{\partial x_1}(x_1, \dots, x_n)}$$

expression indépendante de θ. Cette remarque signifie, qu'une fois T connue, l'échantillon n'apporte plus d'information concernant θ et donc que T porte toute l'information disponible sur θ.

1. Vérifier que T est exhaustive pour θ dans chacun des cas suivants :
   (a) X suit la loi de Poisson de paramètre θ et $T = \displaystyle\sum_{i=1}^{n} X_i$;
   (b) X suit la loi exponentielle de paramètre θ et $T = \displaystyle\sum_{i=1}^{n} X_i$;
   (c) X suit la loi normale de paramètres (θ, σ^2) où σ est connu et $T = \displaystyle\sum_{i=1}^{n} X_i$;
   (d) X suit la loi normale de paramètres (m, θ) où m est connu et $T = \displaystyle\sum_{i=1}^{n} (X_i - m)^2$;
   (e) X suit la loi uniforme sur $[0, \theta]$ et $T = \max(X_1, \dots, X_n)$.
2. Démontrer le résultat suivant (théorème de Darmois) :
   On suppose les hypothèses suivantes vérifiées :
   le support de f_X, $S_\theta = \{x \in \mathbb{R} : f_X(x;\theta) > 0\}$, est indépendant de θ ;
   la densité f_X est de la forme $f_X(t;\theta) = \exp[a(x)\alpha(\theta) + b(x) + \beta(\theta)]$;
   la fonction $x_1 \mapsto \displaystyle\sum_{i=1}^{n} a(x_i)$ est une bijection de classe C^1 pour tout $(x_2, \dots, x_n)$.
   Dans ces conditions, la statistique $T = \displaystyle\sum_{i=1}^{n} a(X_i)$ est exhaustive.
   Retrouver les résultats précédents (excepté le dernier) à partir de ce théorème.

Exercice 16 (Information de Fisher, inégalité de Cramér-Rao)

Soit X une v.a. de densité $x \mapsto f_X(x;\theta)$. On introduit l'information de Fisher fournie par X sur le paramètre θ :

$$I_X(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial \theta} \ln f_X(X;\theta)\right)^2\right] = \int_{-\infty}^{+\infty} \left[\frac{\partial}{\partial \theta} \ln f_X(x;\theta)\right]^2 f_X(x;\theta) dx.$$

Si le support de f_X, $S_\theta = \{x \in \mathbb{R} : f_X(x;\theta) > 0\}$, est indépendant de θ, alors sous certaines conditions de régularité pour f, on a :

$$I_X(\theta) = \mathrm{var}\left[\frac{\partial}{\partial \theta} \ln f_X(X;\theta)\right] = -\mathbb{E}\left[\frac{\partial^2}{\partial \theta^2} \ln f_X(X;\theta)\right].$$

(Dans le cas où X est une v.a. discrète, remplacer f_X par $x \mapsto p_X(x;\theta) = \mathbb{P}(X = x)$ ainsi que l'intégrale par une somme.)

1. Calculer I_X dans les cas suivants :
   (a) X suit la loi de Poisson de paramètre θ ;
   (b) X suit la loi exponentielle de paramètre θ ;
   (c) X suit la loi normale de paramètres (θ, σ^2) où σ est connu ;
   (d) X suit la loi normale de paramètres (m, θ^2) où m est connu ;
   (e) X suit la loi uniforme sur $[0, \theta]$.
2. On suppose maintenant que S_θ est indépendant de θ. Soit $(X_1, \dots, X_n)$ un échantillon de X, et $I_n = I_{(X_1,\dots,X_n)}$.
   (a) Montrer que $I_n = nI_X$.
   (b) Montrer que pour toute statistique exhaustive T de $(X_1, \dots, X_n)$, on a $I_T = I_n$.
3. On a le résultat suivant (inégalité de Cramér-Rao) :

   On suppose que S_θ est indépendant de θ. Soit $T = \phi(X_1, \dots, X_n)$ une statistique de $X_1, \dots, X_n$. Alors,

   $$\operatorname{var} T \geqslant \frac{\left[\frac{\partial}{\partial\theta}\mathbb{E}(T)\right]^2}{I_n(\theta)}.$$

   En particulier, si T est un estimateur sans biais de θ, on a $\operatorname{var} T \geqslant \dfrac{1}{I_n(\theta)}$, ce qui donne une borne supérieure pour la précision d'un tel estimateur.

   Tester l'efficacité de T dans les cas suivants :
   (a) X suit la loi de Poisson de paramètre θ et $T = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$;
   (b) X suit la loi exponentielle de paramètre θ et $T = (n-1)/\displaystyle\sum_{i=1}^{n} X_i$;
   (c) X suit la loi normale de paramètres (θ, σ^2) où σ est connu et $T = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$;
   (d) X suit la loi normale de paramètres (m, θ) où m est connu et $T = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - m)^2$;
   (e) X suit la loi uniforme sur $[0, \theta]$ et $T = \dfrac{n+1}{n}\max(X_1, \dots, X_n)$.