1 Estimation

Exercice 1

Un fabricant de pneus relève le nombre de kilomètres parcourus par cent pneus neufs d'un modèle donné avant la première crevaison et obtient les résultats suivants :

208	1685	554	1348	1037	226	1405	211	1305	44
1321	1421	141	680	3044	184	102	65	1016	1376
21	757	96	336	684	620	167	150	1265	764
468	438	289	345	577	507	910	1832	2030	208
455	2493	56	1499	1086	83	83	1332	88	1379
138	223	121	2816	288	740	616	53	2408	1610
60	1678	423	656	199	129	168	720	580	192
1112	632	458	697	2053	888	467	721	1865	41
2618	46	530	398	1572	119	364	229	142	632
356	770	136	235	700	504	1453	125	193	717

Il sait par ailleurs que le nombre de kilomètres parcourus par un pneu neuf suit une loi exponentielle. Estimer le paramètre de cette loi.

Exercice 2

Soient Y_1, Y_2 et Y_3 trois variables aléatoires indépendantes suivant toutes une loi géométrique de paramètre $\frac{1}{k}$, où k est un nombre réel positif à déterminer. On note $Y = Y_1 + Y_2 + Y_3$.

1. En déduire que $\frac{1}{3}Y$ est un estimateur sans biais de k.
2. Calculer le risque quadratique de cet estimateur.

Exercice 3

On lance un dé équilibré à n faces numérotées de 1 à n. On note X la variable aléatoire égale au résultat du dé.

1. Quelle est la loi suivie par X ?
2. Soient $(X_i)_{1\leq i\leq 5}$ cinq variables aléatoires indépendantes qui suivent la même loi que X.

   Montrer que $Y = \frac{2}{5}\sum_{i=1}^{5} X_i - 1$ est un estimateur sans biais de n.
3. Calculer le risque quadratique de Y.

Exercice 4

Soit $X \hookrightarrow \mathcal{U}([0;a])$ et $(X_1, \cdots, X_n)$ un n-échantillon de X. On cherche à estimer a. On note $\overline{X_n} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$.

1. Soit $T_n = 2\overline{X_n}$. Montrer que T_n est un estimateur sans biais de a et calculer son risque quadratique.
2. Soit $T'_n = \max(X_1, \cdots, X_n)$.
   Donner la fonction de répartition de X, puis déterminer celle de T'_n.
   En déduire une densité de T'_n puis son biais et son risque quadratique.
3. Soit $T''_n = \dfrac{n+1}{n}T'_n$. Déterminer son biais et son risque quadratique.
4. Pour de grandes valeurs de n, quelle est le meilleur estimateur de a ?

2 Intervalle de confiance

Exercice 5

Une urne contient des boules blanches et des boules noires. On cherche à estimer la proportion p de boules noires. Pour cela on tire 100 fois une boule de l'urne avec remise et on obtient 74 boules noires.
On note X la variable aléatoire égale à 1 si la boule tirée est noire, et à 0 sinon, et X_1, ..., X_{100} un échantillon de variables aléatoires indépendantes de même loi que X. On note $\overline{X} = \dfrac{1}{100}\sum_{i=1}^{100} X_i$.

1. Quelle est la loi de X ?
2. Montrer que $\overline{X}$ est un estimateur sans biais de p.
3. Déterminer son risque quadratique.
4. Montrer que pour tout $p \in [0; 1]$, $p(1-p) \leq \dfrac{1}{4}$.
5. En déduire, en utilisant l'inégalité de Bienaymé-Tchebychev, que $P(|\overline{X} - p| \geq 0,2) \leq \dfrac{1}{16}$.

On note $U = \overline{X} - 0,2$ et $V = \overline{X} + 0,2$.

6. Montrer que l'intervalle $[U; V]$ est un intervalle de confiance de p au niveau de confiance 0,9375.
7. Donner une estimation de cet intervalle de confiance.

Exercice 6

On suppose que le paramètre p, qui exprime la probabilité qu'un individu contagieux transmette un virus donné à un individu sain, est inconnu, et on cherche à l'estimer.
Pour m entier supérieur ou égal à 1, on considère un m-échantillon $(Y_1, Y_2, \ldots, Y_m)$ de variables aléatoires indépendantes, de même loi de Bernoulli de paramètre p.

On pose : $\overline{Y_m} = \dfrac{1}{m}\sum_{i=1}^{m} Y_i$.

1. (a) Montrer que $\overline{Y_m}$ est un estimateur sans biais de p ;
   (b) Déterminer son risque quadratique.
2. À l'aide de l'inégalité de Bienaymé-Tchebycheff, montrer que l'intervalle $\left[\overline{Y_m} - \sqrt{\dfrac{5}{m}}, \overline{Y_m} + \sqrt{\dfrac{5}{m}}\right]$ est un intervalle de confiance de p au niveau de confiance $0,95$.

Exercice 7

Un sériciculteur a pesé 100 cocons provenant de son élevage de vers à soie. Il a obtenu les résultats suivants :

Poids en g	0,64	0,65	0,66	0,67	0,68	0,69	0,70	0,71
Effectif	3	5	2	6	6	10	12	10

Poids en g	0,72	0,73	0,74	0,75	0,76	0,77	0,78	0,79
Effectif	9	8	8	6	5	4	3	3

On suppose que le poids X d'un cocon suit une loi normale de moyenne m et de variance $0,0016$ g^2.

1. Préciser les paramètres de cette loi normale.
2. On choisit au hasard un échantillon de 100 cocons. On appelle X_i le poids du i-ème cocon de cet échantillon. On suppose que $X_1, \cdots, X_{100}$ sont mutuellement indépendantes, de même loi que X. On pose $T = \dfrac{1}{100}\sum_{i=1}^{100} X_i$ et on admet que T suit une loi normale.
   Préciser les paramètres de cette loi normale et déterminer la variable centrée réduite T^* associée à T.
3. Déterminer un réel t tel que $P(-t \leqslant T^* \leqslant t) \geqslant 0,95$.
4. Montrer que $P(T - 0,00784 \leqslant m \leqslant T + 0,00784) \geqslant 0,95$.
5. En déduire un intervalle de confiance pour m au niveau de confiance $0,95$ et donner une estimation de cet intervalle.

Exercice 8

On cherche à estimer la proportion p de personnes qui voteront pour le candidat A au sein d'une population. Lors d'un sondage sur 100 personnes interrogées, 60 pensent voter pour A.

On modélise le choix par un échantillon $(X_1, \cdots, X_{100})$ de variables indépendantes de même loi de Bernoulli de paramètre p.

On cherche à déterminer un intervalle de confiance pour p au niveau de confiance 99 %.

1. Déterminer l'espérance et la variance de la moyenne empirique $F = \dfrac{1}{100}\displaystyle\sum_{i=1}^{100} X_i$.
2. On note F^* la variable centrée réduite associée à F.
   On approche désormais F^* par la loi $\mathcal{N}(0,1)$.
   Déterminer t tel que $P(-t \leqslant F^* \leqslant t) \geqslant 0,99$ et en déduire que
   $$P\left(F - t\frac{\sqrt{p(1-p)}}{10} \leqslant p \leqslant F + t\frac{\sqrt{p(1-p)}}{10}\right) \geqslant 0,99$$
3. Montrer que pour tout $p \in [0;1]$, $p(1-p) \leqslant \dfrac{1}{4}$ et en déduire un intervalle de confiance pour p au niveau de confiance $0,99$ puis en donner une estimation.

Exercice 9

Soit X une variable aléatoire à valeurs dans $[0,\theta]$ où θ est un paramètre réel strictement positif à déterminer. Une densité f de X est donnée par : $f(x) = \begin{cases} \dfrac{2x}{\theta^2} & \text{si } x \in]0,\theta]. \\ 0 & \text{sinon} \end{cases}$

1. Calculer l'espérance et la variance de X.
2. (a) Déterminer la fonction de répartition F de X.
   (b) Tracer dans un repère orthogonal, l'allure de la courbe représentative de F.

Pour n entier de $\mathbb{N}^*$, soit $(X_1, X_2, \dots, X_n)$ un n-échantillon de variables aléatoires indépendantes et de même loi que X. On pose pour tout $n \in \mathbb{N}^*$: $\overline{X}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$

3. (a) Déterminer un estimateur T_n de θ, sans biais et de la forme $c\overline{X}_n$, où c est un réel que l'on précisera.
   (b) Quels sont les risques quadratiques respectifs associés aux estimateurs $\overline{X}_n$ et T_n de θ ?
4. On pose pour tout $n \in \mathbb{N}^*$: $M_n = max(X_1, X_2, \dots, X_n)$.
   (a) Déterminer la fonction de répartition G_n et une densité g_n de M_n.
   (b) Calculer l'espérance de M_n . En déduire un estimateur sans biais W_n de θ.
   (c) Entre T_n et W_n , quel estimateur doit-on préférer pour estimer θ ?
5. Soit α un réel donné vérifiant $0 < \alpha < 1$
   (a) Établir l'existence de deux réels a et b tels que $0 < a < 1$ et $0 < b < 1$, vérifiant $P(M_n \leq a\theta) = \alpha/2$ et $P(b\theta \leq M_n \leq \theta) = \alpha/2$.
   (b) En déduire un intervalle de confiance pour le paramètre θ au niveau de confiance $1 - \alpha$.