Travaux Dirigés de Statistique SY02

T. Denœux et G. Govaert

Automne 2011

Table des matières

Chapitre 1

Enoncés

1.1 Statistiques descriptives

1. Une étude sur le taux de glucose dans le sang de 16 femmes au début de leur grossesse a fourni les valeurs suivantes :

60	56	80	55	62	74	64	73
60	70	66	83	68	78	103	77

   Tracer le diagramme par tige et feuilles puis donner l'expression de la fonction de répartition empirique de cet échantillon. Calculer la moyenne empirique, la moyenne tronquée d'ordre 1, la variance empirique corrigée, la médiane et les quartiles, l'étendue et l'étendue interquartiles.

2. Le tableau suivant indique les nombres de voitures particulières vendues au Etats-Unis en 1990, par classes de poids. La signification des notations est la suivante :
   - c_k : centre de la classe k,
   - $[a_k, a_{k+1}[$: intervalle définissant la classe k,
   - n_k : nombre de véhicules vendus en 1990 (en milliers) pour la classe k.

c_k	$[a_k, a_{k+1}[$	n_k
1750	[1500,2000[	510
2250	[2000,2500[	1290
2750	[2500,3000[	4070
3500	[3000,4000[	4094
5000	[4000,6000[	720

   (a) Calculer les fréquences relatives et les fréquences relatives cumulées.

   (b) Tracer un histogramme de la distribution des ventes de véhicules en 1990.

   (c) Donner la moyenne, l'écart-type et la classe modale de cette distribution.

3. Les données ci-dessous sont le résultat d'une étude effectuées entre le 25/09/86 et le 18/10/87 dans la forêt équatoriale de Nouvelle-Guinée. Elles concernent les quantités de poisson (en kg) pêchées par chaque habitant de sexe masculin d'un village. Sont indiqués, pour chaque individu :

l'âge, la situation familiale (M = marié, W = veuf, B = célibataire, BM = marié au cours de la période d'étude, Y = jeune homme, C = enfant), le nombre de nuits passées dans les zones de chasse et de pêche du village, et les quantités en kg de poissons pêchées par deux procédés différents.

Nom	Age	Sit.	Nuits	javelot	hameçon
Bisaeo	45	M	327	8.1	2.5
Gugwi	45	M	363	54.1	13.6
Wodai	45	W	43	0	12.1
Mamo	40	M	310	11.7	3.4
Simo	35	M	295	3.3	21.8
Gwase	28	BM	346	10.3	12.0
Tufa	25	BM	155	2.2	2.7
Gwuho	25	B	136	2.0	7.1
Filifi	25	B	274	30.7	0.9
Sinio	22	M	267	52.2	8.0
Maubo	20	Y	362	16.2	27.9
Dogo	15	Y	314	11.8	14.9
Hegogwa	15	Y	122	3.8	15.9
Gawua	10	C	263	0	2.0
Okre	3	C	355	0	0

(a) Indiquer la nature (qualitative ou numérique) de chacune des variables du tableau, et résumer les distributions des variables « situation familiale » et « nombre de nuits » sous forme de tableaux de fréquences.

(b) A l'aide d'un diagramme en boîtes, comparer les quantités de poissons obtenues par les deux modes de pêche.

1.2 Probabilités

1. La durée de vie en heures d'un composant électronique est modélisée par une variable aléatoire X de densité de probabilité

$$f(x) = \begin{cases} \alpha^2 x e^{-\alpha x} & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases}$$

avec $\alpha > 0$.

   (a) Vérifiez que f est bien une densité de probabilité.

   (b) Calculer l'espérance mathématique μ et la variance σ^2 de X.

   (c) Calculer la probabilité p qu'un composant pris au hasard ait une durée de vie supérieure ou égale à μ.

   Rappel : $\int_0^\infty x^n e^{-\alpha x} dx = \frac{n!}{\alpha^{n+1}}$.

2. Trouver les constantes a en fonction de b pour que les fonctions suivantes soient des densités de probabilité. Donner ensuite les expressions de l'espérance mathématique et de la variance (si elles existent) des v.a. correspondantes :

   (a)

$$f(x) = \begin{cases} ax & \text{si } 0 \leq x \leq b \\ 0 & \text{sinon} \end{cases}$$

   (b)

$$f(x) = \begin{cases} \frac{a}{x^2} & \text{si } x \geq b \\ 0 & \text{sinon} \end{cases}$$

   où $b > 0$.

   Tracer les fonctions de densité et les fonctions de répartition correspondantes.

3. Une machine à embouteiller peut tomber en panne. La probabilité d'une panne est de 0.01 à chaque emploi de la machine. La machine doit être utilisée 100 fois.

   (a) Le nombre de pannes obtenues est une variable aléatoire X. Calculer les probabilités d'obtenir :

      i. $X = 1$,

      ii. $X = 2$,

      iii. $X = 3$,

      iv. $X \geq 4$.

   (b) On estime le coût d'une réparation à 500 euros. La dépense (exprimée en euros) pour les réparations de la machine est une variable aléatoire Y. Calculer l'espérance mathématique de Y, et son écart-type.

4. Une machine outil débite des plaques carrées dont le côté, mesuré en centimètres, est une variable aléatoire X suivant une loi normale d'espérance $\mu = 10$ cm et d'écart-type $\sigma = 0,4$ cm.

   (a) Quelle est la probabilité p_0 pour une plaque d'avoir une surface au moins égale à 110 centimètres carrés ?

(b) On fabrique 30 plaques. Quelle est la probabilité p_1 pour qu'au plus 5 plaques aient cette propriété ?

5. On a constaté que la répartition du taux de cholestérol dans une population de grande taille est la suivante :
   – taux inférieur à 165 cg : 58 % ;
   – taux compris entre 165 et 180 cg : 38 % ;
   – taux supérieur à 180 cg : 4 %.

   (a) Sachant que le taux de cholestérol est distribué selon une loi normale, calculer la valeur moyenne et l'écart-type du taux de cholestérol dans la population.

   (b) On admet que les personnes dont le taux est supérieur à 183 cg doivent subir un traitement. Quel est le nombre de personnes à soigner dans une population d'un million de personnes ?

1.3 Échantillonnage. Théorème de la limite centrale

1. Une boîte de céréales d'une certaine marque contient en moyenne 500 grammes de céréales avec un écart-type de 10 grammes. Si l'on prélève 100 boîtes au hasard, quelle est la probabilité que leur poids moyen soit :
   (a) au moins égal à 498 grammes ;
   (b) au plus égal à 503 grammes ;
   (c) compris entre 499 et 506 grammes.
2. Un avion de ligne a une capacité de 350 passagers. On suppose que le poids d'un passager est une v.a. de moyenne 70 kg et d'écart-type 10 kg, et le poids des bagages d'un passager est une v.a. de moyenne 45 kg et d'écart-type 5 kg. Si le poids total de la cargaison de l'avion (passagers et bagages) ne doit pas excéder 45 tonnes, et en admettant que le taux de remplissage est de 100 %, quelle est la probabilité :
   (a) que la limite de poids soit dépassée ;
   (b) que le poids total de la cargaison soit compris entre 39.5 et 40.5 tonnes.
3. Un dé est lancé 1000 fois. Soit X le nombre de 6 obtenus au cours des 1000 lancés. Calculer :
   (a) $\mathbb{P}(X > 200)$;
   (b) $\mathbb{P}(X \leq 175)$.
4. On cherche à comparer les ampoules de type A et celles de type B dont la durée de vie suit une loi normale avec les paramètres suivants :

Marque	μ	σ
A	1400 h	200 h
B	1200 h	100 h

   On prélève 125 ampoules de chaque type. Calculer la probabilité des événements suivants :
   (a) la durée de vie moyenne des ampoules de type B est supérieure ou égale à celle des ampoules du type A.
   (b) la durée de vie moyenne des ampoules de type A est supérieure de plus de 160 heures à celle des ampoules du type B ;
5. Un coureur à pied couvre en moyenne 140 cm à chaque foulée, avec un écart-type de 5 cm. Calculer approximativement la probabilité que le coureur parcoure 100 m en moins de 70 foulées, puis en moins de 75 foulées.
6. Soient n nombres aléatoires $X_1, \ldots, X_n$. Chaque nombre X_i est approché par l'entier le plus proche, noté J_i. Soit $U_i = X_i - J_i$ l'erreur d'arrondi. En supposant que les variables $U_1, \ldots, U_n$ suivent indépendamment une loi uniforme sur l'intervalle $[-0.5, 0.5]$, calculer de manière approchée la probabilité que l'erreur d'arrondi sur la somme $X_1 + \ldots + X_n$ soit en valeur absolue supérieure à une valeur a quelconque.

7. Soit $X_1, X_2, \ldots$ une suite de v.a. indépendantes et identiquement distribuées suivant chacune une loi continue uniforme sur $[0,1]$. On considère la moyenne géométrique
$$G_n = (X_1 X_2 \ldots X_n)^{1/n}.$$
Montrer que $G_n \xrightarrow{P} c$ pour une constante c que l'on précisera. [On utilisera le résultat suivant : si pour une suite (Z_n) de v.a. et une constante a on a $Z_n \xrightarrow{P} a$, alors pour toute fonction continue φ on a $\varphi(Z_n) \xrightarrow{P} \varphi(a)$].

1.4 Estimation, méthode des moments

1. Soit $X_1, \ldots, X_n$ un échantillon i.i.d. de v.a. parente X admettant une espérance μ et une variance σ^2. Montrer que
$$T_1 = \frac{2}{n(n+1)} \sum_{i=1}^{n} iX_i$$
et
$$T_2 = \frac{6}{n(n+1)(2n+1)} \sum_{i=1}^{n} i^2 X_i$$
sont des estimateurs sans biais et convergents de μ. Rappels :
$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6} \qquad \sum_{i=1}^{n} i^4 = \frac{n(n+1)(2n+1)(3n^2+3n-1)}{30}.$$

2. Suite à l'avarie d'un navire acheminant des déchets toxiques, un conteneur renfermant un produit dangereux pour l'environnement a disparu dans l'océan. Le temps X (en heures) nécessaire à la dégradation naturelle du produit suit une loi exponentielle, définie par la fonction de densité de probabilité f_X suivante :
$$f_X(x) = \begin{cases} \frac{1}{\sigma}\exp\left(-\frac{x-\theta}{\sigma}\right), & \text{si } x \geq \theta; \\ 0 & \text{sinon.} \end{cases}$$
Ici, le paramètre θ représente la durée pendant laquelle le conteneur restera hermétiquement clos, et le paramètre σ caractérise la vitesse à laquelle le produit disparaîtra. On sait en outre que $\mathbb{E}(X) = \theta + \sigma$ et $\mathrm{Var}(X) = \sigma^2$. L'analyse des données collectées lors de 10 catastrophes écologiques a fourni l'échantillon suivant de durées de dégradation :
$$95 \quad 122 \quad 104 \quad 116 \quad 196 \quad 141 \quad 172 \quad 153 \quad 141 \quad 114.$$
   (a) On suppose dans un premier temps que la durée de résistance du conteneur est connue : $\theta = 72$h. Proposer un estimateur $\widehat{\sigma_1}$ du paramètre σ par la méthode des moments, et calculer sa réalisation pour l'échantillon de valeurs disponibles. Est-il sans biais ? Est-il convergent ?
   (b) On considère à présent le cas le plus général où l'on ne dispose d'aucune information sur la durée de résistance θ du conteneur. Toujours à l'aide de la méthode des moments, proposer deux estimateurs $\widehat{\sigma_2}$ et $\widehat{\theta}$ des paramètres σ et θ, et calculer leur réalisation.
   (c) Les estimateurs $\widehat{\sigma_2}$ et $\widehat{\theta}$ sont-ils convergents ?

3. On dispose d'un échantillon $X_1, \ldots, X_n$ de n variables aléatoires indépendantes issues d'une loi continue uniforme définie sur l'intervalle $[0, \theta]$ et on se propose d'estimer le paramètre θ.
   (a) Soit $\widehat{\theta}_1$ l'estimateur de θ obtenu par la méthode des moments en utilisant les moments d'ordre 1. Donner l'expression de $\widehat{\theta}_1$.

(b) Montrer que $\widehat{\theta}_1$ est un estimateur sans biais et convergent de θ.

(c) Soit $\widehat{\theta}_2 = X_{(n)}$. Montrer que $\widehat{\theta}_2$ est un estimateur asymptotiquement sans biais et convergent (Indication : pour calculer l'espérance de cet estimateur, déterminer d'abord sa fonction de répartition).

(d) Proposer un estimateur $\widehat{\theta}_3$ sans biais et convergent, fonction de $\widehat{\theta}_2$.

1.5 Méthode du maximum de vraisemblance

1. On suppose que le nombre K d'absences par mois dans une entreprise suit une loi de Poisson de paramètre λ. Sachant que l'on dispose d'un échantillon iid de taille n issu de cette v.a.,
   (a) déterminer $\widehat{\lambda}_{MV}$ l'estimateur du maximum de vraisemblance de λ ;
   (b) déterminer l'information de Fisher apportée au paramètre λ par l'échantillon ;
   (c) et en déduire la loi asymptotique de $\widehat{\lambda}_{MV}$.
2. On suppose que le salaire annuel brut d'un jeune diplômé de l'UTC pris au hasard est une variable aléatoire continue R dont la densité s'exprime en fonction de deux paramètres r_0 et θ de la façon suivante :

$$\begin{aligned} f_R(r) &= \frac{1}{\sqrt{\pi}(r-r_0)} \exp(-(\ln(r-r_0)-\theta)^2) \text{ si } r > r_0 \\ &= 0 \text{ sinon} \end{aligned}$$

   Soit $R_1, \ldots, R_n$ un échantillon iid de variable parente R.
   (a) On suppose r_0 connu. Trouver l'estimateur du maximum de vraisemblance $\widehat{\theta}_{MV}$ de θ.
   (b) Soit la variable aléatoire $Z = \ln(R - r_0)$. Calculer F_Z, fonction de répartition de Z en fonction de F_R, fonction de répartition de R. En déduire la densité f_Z de Z en fonction de la densité f_R de R, puis la loi de $\widehat{\theta}_{MV}$.
3. Soit $(X_1, \ldots, X_n)$ un échantillon extrait de la loi uniforme sur $[\theta, \theta+1]$ avec $\theta > 0$. On pose

$$\begin{aligned} S_n &= \max_{1 \le i \le n} X_i \\ I_n &= \min_{1 \le i \le n} X_i. \end{aligned}$$

   (a) Montrer qu'il existe, sauf cas particulier que l'on précisera, une infinité d'estimateurs du maximum de vraisemblance de θ, de forme générale :

$$\widehat{\theta}_n(\alpha) = \alpha(S_n - 1) + (1-\alpha)I_n \qquad (0 \le \alpha \le 1)$$

   (b) Calculer la fonction de répartition, puis la fonction de densité des v. a. S_n et I_n.
   (c) En déduire que

$$\begin{aligned} \mathbb{E}(I_n) &= \theta + \frac{1}{n+1} \\ \mathbb{E}(S_n) &= \theta + \frac{n}{n+1} \end{aligned}$$

   (On pourra au besoin admettre ce résultat et passer directement à la question suivante).
   (d) Quelle est l'unique valeur α^* telle que $\widehat{\theta}_n = \widehat{\theta}_n(\alpha^*)$ soit un estimateur sans biais de θ ?
   (e) En utilisant la méthode des moments, proposer un autre estimateur sans biais de θ. Quelle est sa variance ?

1.6 Estimation par intervalle de confiance

1. Les lectures de tension artérielle systolique (en mm Hg) sur un individu à la même heure pendant 7 jours consécutifs ont été :

jour	1	2	3	4	5	6	7
x_i	161	155	142	157	150	192	156

   (a) Calculer la moyenne empirique et la médiane de cet échantillon.

   (b) En faisant l'hypothèse que la mesure de tension X suit une loi normale d'espérance μ et de variance $\sigma^2 = 100$, quel intervalle de confiance bilatéral à 95% peut on donner sur μ ?

   (c) Combien de jours faut-il observer la tension artérielle systolique pour que la longueur de l'intervalle de confiance à 95% n'excède pas 5 mm de Hg ?

   (d) Que devient l'intervalle calculé pour répondre à la question (b) si l'on suppose maintenant que la variance est inconnue ?

   (e) Donner un intervalle de confiance unilatéral à 95% sur σ^2, de la forme $[0, U]$, en supposant $\mu = 160$, puis en supposant μ inconnu.

2. Le nombre X de plaintes enregistrées au commissariat après une soirée étudiante suit une loi de Poisson de paramètre λ. On a relevé les nombres de plaintes $X_1, \ldots, X_n$ consécutives à n soirées.

   (a) Quel est l'estimateur du maximum de vraisemblance $\widehat{\lambda}$ de λ ?

   (b) En utilisant le théorème de la limite centrale, proposer une fonction asymptotiquement pivotale pour λ.

   (c) En utilisant le théorème de Slutsky, montrer que $\frac{\widehat{\lambda}-\lambda}{\sqrt{\widehat{\lambda}/n}}$ est une fonction asymptotiquement pivotale pour λ.

   (d) En déduire l'expression d'un intervalle de confiance bilatéral approché au niveau $1 - \alpha$.

   (e) Application numérique : aux mois de mai et juin, 20 soirées (officielles) ont été organisées ; au total, 73 plaintes ont été déposées. Quelle est la réalisation de l'intervalle proposé à la question (d), au niveau de confiance de 95% ?

3. Soit X une v. a. suivant une loi de densité $f(x) = \theta^2 x e^{-\theta x} 1_{[0,+\infty[}(x)$ avec $\theta > 0$ et $X_1, \ldots, X_n$ un échantillon i.i.d. de variable parente X. Une réalisation de cet échantillon pour $n = 100$ a donné des valeurs x_i vérifiant $\sum_{i=1}^n x_i = 398$.

   (a) Déterminer l'estimateur du maximum de vraisemblance de θ, que l'on notera $\widehat{\theta}$.

   (b) Quelle est l'information de Fisher associée à θ ? En déduire une fonction asymptotiquement pivotale pour θ et un intervalle de confiance bilatéral approché sur θ au niveau de confiance $1 - \alpha$.

   (c) A.N. Calculer l'intervalle de confiance pour $1 - \alpha = 95\%$.

4. Une enquête effectuée dans une école primaire a mis en évidence un problème de dentition chez 15 enfants parmi 50. Donner, en utilisant les abaques puis par le calcul :

(a) un intervalle de confiance bilatéral à 95 % sur la proportion p d'enfants ayant un problème de dentition dans la population totale ;

(b) deux intervalles unilatéraux à 95 % sur p.

1.7 Estimation optimale

1. On dispose d'un échantillon de n variables aléatoires indépendantes tiré d'une loi de Laplace de densité

$$f(x) = \frac{\theta}{2} e^{-\theta|x|} \ \forall x \in \mathbb{R} \text{ et } \theta > 0.$$

   (a) Existe-t-il un estimateur efficace de θ ou d'une fonction $u(\theta)$ de θ ?

   (b) Calculer sa variance.

2. On considère une urne contenant des boules blanches et noires en proportions p et $1-p$, le paramètre $p \in]0,1[$ étant inconnu. On effectue une suite de tirages avec remise et on s'arrête à la première boule blanche tirée. Soit K le nombre de tirages effectués. On sait que K suit une loi géométrique définie par :

$$\mathbb{P}(K = k) = (1-p)^{k-1} p, \quad k \in \{1, 2, \dots, \infty\}.$$

   (a) Existe-t-il un estimateur efficace de p ou d'une fonction de p ? Si oui, donner son espérance et sa variance.

   (b) Quel est l'estimateur $\widehat{p}_{MV}$ du maximum de vraisemblance de p ?

   (c) Soit T la variable aléatoire définie par :

$$T = \begin{cases} 1 & \text{si } K = 1 \\ 0 & \text{sinon} \end{cases}$$

   Montrer que T est un estimateur sans biais de p. Calculer sa variance.

   (d) Calculer l'information de Fisher $I_K(p)$ apportée par K sur p.

   (e) Comparer la variance de T à la borne de Fréchet pour les estimateurs sans biais de p. Pouvait-on prévoir ce résultat ?

3. Soit X une v.a. normale de paramètres μ et σ^2 inconnus et $X_1, \dots, X_n$ un échantillon i.i.d. de variable parente X. On se propose d'étudier la famille des estimateurs de σ^2 de la forme :

$$\widehat{\sigma}_a^2 = a \sum_{i=1}^{n} (X_i - \bar{X})^2, \quad a \in \mathbb{R}.$$

   On supposera dans tout cet exercice que $n > 1$.

   (a) Calculer $\mathbb{E}(\widehat{\sigma}_a^2)$ et en déduire la valeur de a pour laquelle $\widehat{\sigma}_a^2$ est sans biais.

   (b) En utilisant le théorème de Fisher, calculer la variance de l'estimateur $\widehat{\sigma}_a^2$.

   (c) En déduire l'expression du risque quadratique de $\widehat{\sigma}_a^2$, fonction du paramètre a, défini par :

$$\text{EQM}(\widehat{\sigma}_a^2, \sigma^2) = \mathbb{E}[(\widehat{\sigma}_a^2 - \sigma^2)^2].$$

   (d) En déduire la valeur de a pour laquelle l'estimateur $\widehat{\sigma}_a^2$ est le plus précis.

1.8 Principe des tests d'hypothèses ; théorème de Neyman-Pearson et test UPP

1. On cherche à déterminer si un analyseur en continu de SO_2 en sortie d'une usine d'incinération de déchets est bien étalonné. Pour cela, on a comparé les valeurs X fournies par l'analyseur aux valeurs Y déterminées en laboratoire pour 10 prélèvements :

analyseur x_i	33	24	18	41	37	28	21	43	30	8
laboratoire y_i	33	25	19	43	39	27	21	44	29	9
différence z_i	0	-1	-1	-2	-2	1	0	-1	1	-1

   (a) On suppose que $Z = X - Y \sim \mathcal{N}(\mu, \sigma^2)$ avec $\sigma = 1$. On désigne par H_0 l'hypothèse selon laquelle l'analyseur est bien étalonné, et par H_1 l'hypothèse contraire. Formuler H_0 et H_1 en fonction de μ.

   (b) On considère le test de région critique $W = \{|\overline{Z}| > k\}$. Calculer le seuil critique k pour $\alpha^* = 0.1$. Peut-on rejeter H_0 à ce niveau de signification ?

   (c) Calculer le degré de signification du test pour les données de l'exercice.

2. On a relevé pendant 30 jours ouvrables consécutifs les nombres quotidiens d'actes de délinquance commis dans un centre commercial. Les résultats sont reportés dans le tableau suivant (n_k désigne le nombre de jours où k actes de délinquance ont été commis) :

k	0	1	2	3	4	5
n_k	4	10	11	1	1	3

On suppose que le nombre X d'actes de délinquance commis chaque jour dans le centre commercial suit une loi de Poisson de paramètre λ inconnu. D'après les statistiques du Ministère de l'Intérieur, il se produit en moyenne 1.5 acte de délinquance par jour dans chacun des centres commerciaux de la région parisienne. Ce chiffre est considéré comme sous-estimé par une association des commerçants. Nous allons utiliser la théorie des tests pour tenter de trancher cette question.

   (a) On considère le problème de test suivant :

$$\begin{array}{ll} H_0 : & \lambda = \lambda_0 \\ H_1 : & \lambda = \lambda_1 \end{array}$$

   avec $\lambda_1 > \lambda_0$. En utilisant le théorème de Neyman-Pearson, montrer que le test optimal au niveau α^* pour ce problème s'exprime en fonction de la statistique $\overline{X}$. Donner la forme de la région critique de ce test.

   (b) En supposant n grand, déterminer l'expression littérale de la région critique. Quelle hypothèse accepte-t-on avec $\lambda_0 = 1.5$, $\alpha^* = 0.05$?

   (c) Calculer la puissance du test pour $\lambda_1 = 2$.

   (d) On considère cette fois les hypothèses suivantes :

$$\begin{array}{ll} H_0 : & \lambda = \lambda_0 \\ H_1 : & \lambda > \lambda_0 \end{array}$$

Montrer qu'il existe un test UPP pour ce problème. Comparer la région critique de ce test avec celle du test précédent. Y a-t-il lieu de remettre en cause la validité des statistiques du Ministère de l'Intérieur ?

3. Soit $X_1, \ldots, X_n$ un échantillon i.i.d. de variable parente $X \sim \mathcal{E}(\theta)$, de densité

$$f_X(x) = \theta e^{-\theta x} 1_{[0,+\infty[}(x),$$

θ étant un paramètre positif.

   (a) Donner l'expression de $\widehat{\theta}$, estimateur du maximum de vraisemblance de θ.

   (b) Calculer l'information de Fisher $I_n(\theta)$ relative au paramètre θ. En déduire une fonction asymptotiquement pivotale pour θ.

   (c) On considère le test suivant

$$\begin{array}{lcl} H_0 & : & \theta = \theta_0 \\ H_1 & : & \theta = \theta_1 \end{array}$$

   avec $\theta_1 > \theta_0$. Montrer que la région critique W du test le plus puissant pour ce problème au niveau α^* s'exprime en fonction de $\widehat{\theta}$, puis donner une approximation de W en supposant n grand.

   (d) On considère maintenant le test suivant

$$\begin{array}{lcl} H_0 & : & \theta = \theta_0 \\ H_1 & : & \theta > \theta_0 \end{array}$$

   Existe-t-il un test UPP pour ce problème ?

1.9 Test du rapport de vraisemblance

1. Un lion peut se trouver dans trois états : très actif (θ_1), peu actif (θ_2) et léthargique (θ_3). Le nombre X de personnes mangées en une nuit par un lion est une v.a. dont la loi dépend de l'état du lion $\theta \in \Theta = \{\theta_1, \theta_2, \theta_3\}$, selon le tableau suivant :

x	0	1	2	3	4
$p(x;\theta_1)$	.00	.05	.05	.80	.10
$p(x;\theta_2)$	.05	.05	.80	.10	.00
$p(x;\theta_3)$	.90	.08	.02	.00	.00

   Ayant appris que x personnes ont été mangées par un lion au cours d'une nuit, on cherche à en déduire l'état du lion. Plus précisément, on considère les hypothèses $H_0 = \{\theta_3\}$ contre $H_1 = \{\theta_1, \theta_2\}$.

   (a) Donner dans un tableau les valeurs de la statistique $\lambda(x)$ du rapport de vraisemblance, pour les différentes valeurs possibles de x.

   (b) Donner la loi de $\lambda(X)$ sous l'hypothèse H_0.

   (c) En déduire la région critique du test du rapport de vraisemblance au niveau $\alpha^* = 2$ %.

2. Ayant lancé 1000 fois une pièce de monnaie, on a obtenu 450 fois « face ». L'objet de cet exercice est de tester l'hypothèse selon laquelle la pièce est équilibrée.

   (a) Soit X le nombre de « faces » obtenu au cours de n lancers. On a observé une seule réalisation de X. Calculer l'estimateur du maximum de vraisemblance de p.

   (b) On considère le problème de test suivant : $H_0 : p = p_0$ contre $H_1 : p \neq p_0$. Calculer l'expression littérale de la statistique λ du test du rapport de vraisemblance, en fonction de X, p_0 et n.

   (c) En utilisant l'approximation asymptotique $-2 \ln \lambda \sim \chi^2_1$, donner une expression littérale de la région critique du test du rapport de vraisemblance, au niveau α^*.

   (d) Donner le résultat du test du rapport de vraisemblance avec les données de l'exercice, $p_0 = 1/2$, et $\alpha^* = 5\%$.

3. Un cultivateur désire comparer les performances de deux types d'engrais. Pour cela, il observe les rendements $X_1, \ldots, X_n$ de n parcelles traitées par l'engrais I, et ceux $Y_1, \ldots, Y_m$ de m parcelles traitées par l'engrais II. On suppose que les rendements de chaque type suivent des lois normales d'espérances respectives μ_1 et μ_2, et de variance commune σ^2 connue.

   (a) On considère le problème de test :

$$\begin{array}{ll} H_0 : & \mu_1 = 2,\ \mu_2 = 4 \\ H_1 : & \mu_1 \neq 2 \text{ ou } \mu_2 \neq 4 \end{array}$$

   En utilisant le test du RV, proposer une solution en supposant n et m grands.

   (b) A.N. : $\overline{x} = 2.1$, $\overline{y} = 3.8$, $\sigma = 1$, $n = 40$, $m = 60$, $\alpha^* = 0.05$.

1.10 Tests de conformité

1. Chez un fabriquant de joints en caoutchouc, le département d'ingénierie de la qualité a mis en œuvre un plan d'échantillonnage pour vérifier le poids d'un joint d'étanchéité, poids qui est affecté par les variations d'écoulement du caoutchouc provenant de l'extrudeuse. La valeur cible du poids du joint est de 270 g.

   On considère que le poids X est distribué normalement avec une espérance μ et un écart-type $\sigma = 4,5$ g. Pour maîtriser le procédé, on prélève régulièrement $n = 5$ pièces de caoutchouc de l'extrudeuse. Chaque pièce est pesée et le poids moyen est calculé.

   (a) Donner sans démonstration l'expression de la région critique W du test de l'hypothèse $H_0 : \mu = 270$ g contre $H_1 : \mu \neq 270$ g au niveau de signification de 5 %.

   (b) Lors d'un récent contrôle, on a obtenu, pour un échantillon de cinq pièces, un poids moyen de 265,5 g. Doit-on poursuivre ou arrêter la production ?

   (c) Avec ce plan de contrôle, quel est la probabilité β d'accepter l'hypothèse selon laquelle l'extrudeuse opère à 270 g alors qu'en réalité le procédé est centré à 264 g ?

2. Un industriel affirme que sa production a moins de 10% de défectueux. Après un contrôle de 50 pièces prises au hasard, X n'ont pas fonctionné. On notera p la proportion de défectueux.

   (a) On considère le problème de test suivant :

   $$\begin{cases} H_0 : & p = p_0 \\ H_1 : & p > p_0 \end{cases}$$

   avec $p_0 = 0.1$. Montrer qu'il existe un test UPP et donner la forme de la région critique.

   (b) En utilisant l'approximation normale de la loi binomiale, calculer le seuil critique au niveau $\alpha^* = 0.05$.

   (c) Quelle décision prend-on si il y a 9 pièces défectueuses ?

   (d) Calculer le degré de signification $\widehat{\alpha}$ associé.

1.11 Tests de comparaison

1. Pour comparer la fiabilité de deux types T_1 et T_2 de machines à laver le linge, on choisit comme critère la durée de fonctionnement sans panne sur le programme « blanc », représentée comme une v.a. X pour T_1 et Y pour T_2. On suppose que X et Y suivent des lois normales de moyennes respectives μ_1 et μ_2 et de variances σ_1^2 et σ_2^2. Les observations, exprimées en jours de fonctionnement, sur 10 machines de type T_1 et 13 machines de type T_2 ont donné les résultats suivants :

$$\sum_{i=1}^{10} x_i = 136.6 \quad \sum_{i=1}^{10} x_i^2 = 1923.2$$

$$\sum_{i=1}^{13} y_i = 187.0 \quad \sum_{i=1}^{13} y_i^2 = 2773.3$$

   (a) Tester l'égalité des variances, au niveau de signification $\alpha^* = 0.10$.

   (b) Les durées moyennes de fonctionnement sont-elles significativement différentes au seuil $\alpha^* = 0.05$?

   (c) Donner un intervalle de confiance à 95 % sur la différence $\mu_1 - \mu_2$ des durées de vie moyennes des deux types de machines.

   (d) En quoi cet intervalle de confiance est-il cohérent avec les résultats de la question b ?

2. Une année, le taux de réussite nationale au baccalauréat dans une série donnée a été de 67 %.

   (a) Dans un centre d'examens A, il y a eu 216 reçus sur 300 candidats présentés. Les résultats de ce centre sont-ils conformes aux résultats nationaux ?

   (b) Dans un autre centre d'examen B, il y a eu 128 reçus sur 200 candidats. Les résultats des centres A et B sont-ils significativement différents ?

   (Les tests seront realisés au niveau de signification $\alpha^* = 5\%$).

1.12 Tests d'adéquation

1. L'examen de 320 familles ayant 5 enfants s'est traduit par les résultats suivants :

Nb de garçons	5	4	3	2	1	0	Total
Nb de filles	0	1	2	3	4	5	
Nb de familles	18	56	110	88	40	8	320

   (a) Sous l'hypothèse que la naissance d'un garçon et la naissance d'une fille sont des événements équiprobables, calculer les probabilités de chacun des 6 types de familles.

   (b) Peut-on admettre, au niveau de signification de 5% que les données obtenues sont compatibles avec cette hypothèse d'équiprobabilité ?

   (c) Calculer le degré de signification.

2. On a interrogé 200 élèves d'un lycée sur le type d'études supérieures qu'ils désiraient entreprendre. Les résultats de l'enquête figurent dans le tableau ci-dessous. Au vu de ces résultats, semble-t-il exister une relation entre le choix des études et le sexe ?

Etudes \ Sexe	Garçons	Filles	Total
Littéraires	60	60	120
Scientifiques	42	18	60
Techniques	18	2	20
Total	120	80	200

3. On s'intéresse à la durée de vie d'un produit que l'on modélise par une variable aléatoire X. On considère la réalisation suivante d'un échantillon iid de X :

$$6.1 \quad 1.3 \quad 11.05 \quad 4.6 \quad 9.5 \quad 1.75.$$

   Au vu de ces données, peut-on admettre, au niveau de signification $\alpha^* = 0.05$, que X obéit à

   (a) une loi exponentielle de moyenne $\mu = 5$?

   (b) une loi normale ?

1.13 Analyse de la variance

1. La composition en lipide de la myéline du système nerveux d'un échantillon provenant de quatre espèces différentes a donné les taux de cholestérol (en μ mol/mg lipide) suivants :

	Homme	boeuf	rat	grenouille
	0.70	0.58	0.68	0.64
	0.73	0.62	0.64	0.66
	0.68	0.65	0.71	0.69
	0.72	0.56	0.67	0.67
	0.69	0.59	0.65	0.65
$\overline{x}_k$	0.704	0.600	0.670	0.662
s_k^{*2}	$4.3\ 10^{-4}$	$12.5\ 10^{-4}$	$7.5\ 10^{-4}$	$3.7\ 10^{-4}$

On désire comparer, à l'aide d'une analyse de la variance, les taux de cholestérol entre les espèces.

(a) Vérifier les conditions d'application de l'analyse de la variance, c'est-à-dire :

i. l'hypothèse de normalité de la distribution correspondant à l'espèce grenouille. On supposera pour la suite que les autres distributions vérifient la même propriété.

ii. l'hypothèse concernant les variances des 4 populations.

On prendra $\alpha^* = 5\%$.

(b) Tester l'effet du facteur espèce en prenant un risque de première espèce de 5%.

(c) Dans le cas d'un effet du facteur, préciser pour quelles espèces il existe des différences significatives.

2. Les données suivantes représentent le nombre de problèmes arithmétiques simples (sur 85) résolus (de manière correcte ou non) en une heure par des sujets ayant reçu un médicament dépresseur, un stimulant et un placebo :
– dépresseur : 55, 0, 1 , 40
– stimulant : 75, 85, 51, 63
– placebo : 61, 54, 80, 47
Au vu des ces résultats, peut-on admettre que ces trois médicaments induisent des taux de performance différents ? (faire une test de Kruskal-Wallis avec $\alpha^* = 0.10$).

1.14 Régression linéaire

1. On a relevé dans le tableau suivant les moyennes x au baccalauréat de 10 élèves, et leurs scores Y à un test de QI :

x	8.8	9.6	11.2	10.4	12.8	15.2	12.0	16.0	8.0	9.2
Y	108	112	115	118	121	125	122	130	96	113

   On suppose que les Y_i sont des v.a. indépendantes avec $Y_i \sim \mathcal{N}(a+bx_i, \sigma^2)$, les x_i étant des nombres fixés.

   (a) Déterminer les estimateurs du maximum de vraisemblance des paramètres a, b et σ^2.

   (b) Construire un intervalle de confiance bilatéral sur a, puis une borne inférieure, au niveau de confiance 95 %.

   (c) Tester l'hypothèse $H_0 : b = 0$ contre $H_1 : b > 0$ au niveau de signification $0,01$.

   (d) Construire un intervalle de confiance bilatéral sur b, puis une borne supérieure, au niveau de confiance 95 %.

2. La différence de potentiel mesurée aux bornes d'une résistance r traversée par un courant d'intensité x_i $(i = 1, \ldots, n)$ est modélisée par une variable aléatoire

$$U_i = rx_i + \epsilon_i$$

   où ϵ_i est un bruit de mesure supposé suivre une loi normale d'espérance nulle et de variance σ^2. On considère un échantillon indépendant $U_1, \ldots, U_n$ de n mesures réalisées pour des intensités $x_1, \ldots, x_n$.

   (a) Expliciter les estimateurs du maximum de vraisemblance des paramètres r et σ^2, que l'on notera respectivement $\widehat{r}$ et $\widehat{\sigma^2}$.

   (b) Montrer que $\widehat{r}$ est sans biais. Calculer sa variance. Quelle est la loi de $\widehat{r}$?

   (c) En supposant que $\sigma^2 = 1$, donner l'expression d'un intervalle de confiance de niveau $1 - \alpha$ pour r.

   (d) Application numérique. On a obtenu les résultats suivants :

x_i	0,5	1,5	3	5	8
u_i	1,73	2,41	8,18	9,86	16,11

   Calculer $\widehat{r}$ ainsi qu'un intervalle de confiance de niveau 95 % pour r (en supposant $\sigma^2 = 1$).

   (e) Toujours en supposant $\sigma^2 = 1$, tester l'hypothèse $H_0 : r = 1,9$ contre l'hypothèse alternative $H_1 : r \neq 1,9$, au niveau de signification $0,01$.

Chapitre 2

Problèmes

2.1 Estimation

2.1.1 Comparaison d'intervalles de confiance

On considère dans ce problème une variable aléatoire X de fonction de densité

$$f_X(x) = \frac{kx}{\sigma^2} \exp\left(-\frac{kx^2}{2\sigma^2}\right) 1_{[0,+\infty[}(x)$$

où k est la constante $2 - \pi/2$. On admettra que la variance de X est égale à σ^2. On dispose d'un échantillon de n variables aléatoires indépendantes $X_1, \ldots, X_n$ de même loi que X, et l'on cherche à estimer le paramètre σ.

1. Première méthode.
   (a) Montrer que l'estimateur de σ^2 par la méthode du maximum de vraisemblance est :
   $$\widehat{\sigma}^2_{MV} = \frac{k}{2n} \sum_{i=1}^{n} X_i^2$$
   (b) Cet estimateur est-il efficace ? Calculer sa variance.
   (c) En utilisant l'estimateur $\widehat{\sigma}^2_{MV}$, déterminer un intervalle de confiance bilatéral symétrique pour σ^2 au niveau $1 - \alpha$.
   (d) On a observé un échantillon de taille 200 et on obtient
   $$\sum_{i=1}^{200} x_i = 771.4 \quad \text{et} \quad \sum_{i=1}^{200} x_i^2 = 3793$$
   Calculer numériquement les réalisations de $\widehat{\sigma}^2_{MV}$ et de l'intervalle de confiance bilatéral symétrique pour σ^2 au niveau 95%.
2. Deuxième méthode.
   (a) Montrer la relation
   $$\mathbb{E}(X^2) = \frac{2\sigma^2}{k}$$
   (Pour cela, on exprimera l'espérance $\mathbb{E}(\widehat{\sigma}^2_{MV})$.)

(b) En déduire que l'espérance $\mathbb{E}(X)$ est égale à $\sigma\sqrt{\frac{2-k}{k}}$ et proposer un estimateur $\widehat{\sigma}_m$ de σ par la méthode des moments.

(c) Montrer que cet estimateur est sans biais et calculer sa variance.

(d) En utilisant l'estimateur $\widehat{\sigma}_m$, déterminer un intervalle de confiance bilatéral symétrique pour σ au niveau $1-\alpha$. En déduire un intervalle de confiance bilatéral symétrique pour σ^2 au même niveau de confiance.

(e) Avec les mêmes données numériques que dans la question 1(d), calculer numériquement les réalisations de $\widehat{\sigma}_m^2$ et de l'intervalle de confiance bilatéral symétrique pour σ^2 au niveau 95 %.

2.1.2 Loi exponentielle

Soit X une variable aléatoire suivant une loi de densité

$$f(x) = \begin{cases} e^{\theta - x} & \text{si } x \geq \theta \\ 0 & \text{sinon} \end{cases}$$

On admettra que $\mathbb{E}(X) = \theta + 1$ et $\mathrm{Var}(X) = 1$ et on supposera disposer d'un échantillon i.i.d de grande taille.

1. Première partie

   (a) Déterminer un estimateur $\widehat{\theta}_1$ de θ par la méthode des moments. Montrer que cet estimateur est sans biais. Déterminer sa variance.

   (b) Déterminer un intervalle de confiance bilatéral approché au niveau de confiance $1-\alpha$ en utilisant $\widehat{\theta}_1$.

   (c) Application numérique : on dispose d'un échantillon de taille 2000 dont les principales caractéristiques sont : $\sum x_i = 8019$, $\sum x_i^2 = 34410$, $\min(x_i) = 3.0001$ et $\max(x_i) = 12.3751$. Calculer l'intervalle de confiance pour $\alpha = 5\%$.

2. Seconde partie

   (a) Calculer la fonction de vraisemblance $L(\theta; x_1, \ldots, x_n)$ et tracer son graphe.

   (b) En déduire l'estimateur $\widehat{\theta}_2$ du maximum de vraisemblance. On admettra pour la suite que $\mathbb{E}(\widehat{\theta}_2) = \theta + \frac{1}{n}$ et $\mathrm{Var}(\widehat{\theta}_2) = \frac{1}{n^2}$.

   (c) Déterminer la fonction de répartition de $\widehat{\theta}_2$.

   (d) Déterminer un intervalle de confiance bilatéral au niveau de confiance α en utilisant $\widehat{\theta}_2$.

   (e) En prenant les mêmes données numériques que dans la première partie, calculer l'intervalle de confiance ainsi obtenu. Le comparer à celui obtenu avec l'estimateur $\widehat{\theta}_1$.

2.1.3 Estimateur le plus précis

Soit X une v.a. normale de paramètres μ et σ^2 inconnus et $X_1, \ldots, X_n$ un échantillon i.i.d. de variable parente X.

1. Première partie
   On se propose tout d'abord d'étudier la famille des estimateurs de σ^2 de la forme :
$$\widehat{\sigma}_a^2 = a\sum_{i=1}^{n}(X_i - \overline{X})^2, \quad a \in \mathbb{R}.$$
   (a) Montrer que la méthode du maximum de vraisemblance conduit à choisir $a = 1/n$.
   (b) Calculer $\mathbb{E}(\widehat{\sigma}_a^2)$ et en déduire la valeur de a pour laquelle $\widehat{\sigma}_a^2$ est sans biais.
   (c) On s'intéresse maintenant à la précision de $\widehat{\sigma}_a^2$ définie par la quantité :
$$R(\widehat{\sigma}_a^2, \sigma^2) = E[(\widehat{\sigma}_a^2 - \sigma^2)^2].$$
      i. Montrer que
$$R(\widehat{\sigma}_a^2, \sigma^2) = [a^2(n^2 - 1) - 2a(n-1) + 1]\sigma^4.$$
      (On rappelle que $\frac{\widehat{\sigma}_a^2}{a\sigma^2} \sim \chi^2_{n-1}$).
      ii. En déduire la valeur de a pour laquelle l'estimateur $\widehat{\sigma}_a^2$ est le plus précis.
2. Seconde partie
   On considère maintenant les estimateurs de μ de la forme :
$$\widehat{\mu}_b = b\sum_{i=1}^{n} X_i, \quad b \in \mathbb{R}$$
   (a) Calculer la valeur de b obtenue par la méthode du maximum de vraisemblance et montrer que l'estimateur correspondant est sans biais.
   (b) Montrer que la précision de $\widehat{\mu}_b$ est égale à :
$$R(\widehat{\mu}_b, \mu) = b^2 n(\sigma^2 + n\mu^2) - 2bn\mu^2 + \mu^2$$
   et en déduire la valeur de b minimisant $R(\widehat{\mu}_b, \mu)$.
   (c) Est-il possible en pratique de déterminer une valeur de b telle que l'estimateur $\widehat{\mu}_b$ soit le plus précis ? Pourquoi ?
   Quelle approximation peut-on faire pour n grand et σ^2 petit ?

2.2 Tests

2.2.1 Loi $\gamma(t, \lambda)$

Soit X une v.a. suivant une loi gamma de paramètres t et λ, de densité :

$$f(x) = \begin{cases} 1/\Gamma(t)\lambda^t x^{t-1} e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

On considère la réalisation suivante d'un échantillon i.i.d. de X :

$$1.8 \quad 6.8 \quad 0.5 \quad 1.0 \quad 6.6 \quad 4.5 \quad 4.8 \quad 2.8$$

1. Montrer que :
$$\Gamma(t) = \int_0^{+\infty} x^{t-1} e^{-x}\, dx.$$
Montrer que $\Gamma(1) = \Gamma(2) = 1$.
2. En admettant que $t = 2$, donner un estimateur efficace d'une fonction de λ. Calculer son espérance et sa variance. En déduire l'espérance et la variance de X.
3. Toujours en admettant que $t = 2$, tester l'hypothèse $H_0 : \lambda = 1/2$ contre l'hypothèse $H_1 : \lambda > 1/2$, pour un risque de première espèce $\alpha = 5\%$. On précisera notamment :
   (a) la forme de la région critique en la justifiant,
   (b) la règle de décision (faire l'approximation normale),
   (c) l'application numérique avec les données de l'exercice.
   (d) le risque de seconde espèce pour $\lambda = 1$.
4. En fixant $\lambda = 1/2$, donner la *forme* de la région critique optimale pour les hypothèses $H_0 : t = 2$ contre $H_1 : t = 1$.
5. Tester l'hypothèse H_0 : l'échantillon est issu d'une loi gamma de paramètres $t = 2$ et $\lambda = 1/2$, contre l'hypothèse H_1 : l'échantillon est issu d'une autre loi.

2.2.2 Temps de fonctionnement sans panne d'un appareil

L'instant T de panne d'un appareil est une variable aléatoire obéissant à une loi exponentielle de paramètre $\mu > 0$:

$$f(t) = \frac{1}{\mu} e^{-t/\mu} 1_{]0,+\infty[}(t)$$

On réalise deux expériences.

– 1ère expérience
  On met $n = 225$ appareils en service à la même date t_0 et on note T_i l'instant de panne de l'appareil numéro i.
  1. Calculer l'espérance mathématique de T.
  2. Donner l'estimateur de μ par la méthode du maximum de vraisemblance. Est-il efficace ? Si oui, donner sa variance.

3. On suppose que le temps moyen de fonctionnement sans panne de ce type d'appareil est égal à 750 heures. Tester cette hypothèse avec $\alpha^* = 0.05$. Application numérique : $\bar{t} = 812$ heures.
4. Calculer la puissance du test si le temps moyen de bon fonctionnement est de 900 heures.

– 2ème expérience
On met n appareils en service durant un temps $\tau = 500$ heures et on compte le nombre d'appareils en panne à l'issue de cette période τ.
   1. Calculer la probabilité p pour qu'un appareil tombe en panne entre les instants 0 et τ.
   2. Soit X, le nombre d'appareils en panne avant τ, sur les n qui avaient été mis en service. Donner la loi de X.
   3. Reformuler les hypothèses du test précédent comme des hypothèses sur le paramètre de la loi de X et tester ces hypothèses.
   4. Calculer la puissance du test sous la même hypothèse que précédemment.
   5. Quelle taille devra avoir l'échantillon pour avoir la même puissance que dans la première expérience ?

2.2.3 Test du rapport de vraisemblance

Soit $X_1, \ldots, X_n$ un échantillon i.i.d. de variable parente $X \sim \mathcal{E}(\theta)$, de densité

$$f_X(x) = \theta e^{-\theta x} 1_{[0,+\infty[}(x),$$

θ étant un paramètre positif.

1. Donner l'expression de $\widehat{\theta}$, estimateur du maximum de vraisemblance de θ.
2. Calculer l'information de Fisher $I_n(\theta)$ relative au paramètre θ. En déduire une fonction asymptotiquement pivotale pour θ.
3. On considère le problème de test suivant

$$\begin{array}{lcl} H_0 & : & \theta = \theta_0 \\ H_1 & : & \theta = \theta_1 \end{array}$$

   avec $\theta_1 > \theta_0$. Montrer que la région critique W du test le plus puissant pour ce problème au niveau α^* s'exprime en fonction de $\widehat{\theta}$, puis donner une approximation de W en supposant n grand.
4. On considère maintenant le problème de test suivant

$$\begin{array}{lcl} H_0 & : & \theta = \theta_0 \\ H_1 & : & \theta \neq \theta_0 \end{array}$$

   Existe-t-il un test UPP pour ce problème ?
5. Calculer la statistique du rapport de vraisemblance λ pour le problème de test de la question précédente.
6. Etudier les variations de $\ln(\lambda)$ en fonction de $\widehat{\theta}$. En déduire la forme de la région critique W' du test du rapport de vraisemblance pour le problème de la question 4, puis une approximation de W' en supposant n grand.

2.2.4 Test randomisé

On prélève 20 pièces dans un lot et on compte le nombre X de pièces défectueuses. On admet que $X \sim \mathcal{B}(20, p)$, p étant la proportion inconnue de pièces défectueuses dans le lot. On souhaite tester les hypothèses $H_0 : p = 0.5$ contre $H_1 : p > 0.5$, au niveau de signification $\alpha^* = 0.05$.

1. Montrer qu'il existe un test UPP pour ce problème, de la forme

$$X > A.$$

   Déterminer A pour que le risque de première espèce α ait la plus grande valeur possible tout en respectant la contrainte $\alpha \leq 0.05$.

2. L'inconvénient de la procédure précédente est qu'elle a un risque de première espèce strictement inférieur à 5 %. On propose donc la procédure plus complexe suivante :
   - si $X > A$, on accepte H_1 ;
   - si $X < A$, on accepte H_0 ;
   - si $X = A$, on fait un tirage au sort : on accepte H_1 avec une probabilité γ, et H_0 avec une probabilité $1 - \gamma$.

   Exprimer le risque de première espèce α de ce test en fonction de γ (A ayant la valeur déterminée dans la question 1), puis déterminer γ pour avoir $\alpha = 0.05$.

3. Calculer la puissance du test défini dans la question 2, pour $p = 0.8$.

Chapitre 3

Éléments de correction

3.1 Statistiques descriptives

1. Résumés numériques : $\overline{x} = 70.56$, $M_1 = 69.36$, $s^{*2} = 148.80$, $s^* = 12.20$, $\widehat{f}_{0.5} = 68$, $\widehat{f}_{0.25} = 60$, $\widehat{f}_{0.75} = 77$, $W = 48$, $H = 17$.
2. (c) $\overline{x} = 3080.9$, $s = 724.65$.
3. (b) Javelot : $\widehat{f}_{0.25} = 2$, $\widehat{f}_{0.5} = 8.1$, $\widehat{f}_{0.75} = 16.2$; 2 pts extrêmes $> \widehat{f}_{0.75} + 1.5(\widehat{f}_{0.75} - \widehat{f}_{0.25})$. Hameçon : $\widehat{f}_{0.25} = 2.5$, $\widehat{f}_{0.5} = 8$, $\widehat{f}_{0.75} = 14.9$.

3.2 Probabilités

1. (b) $\mathbb{E}[X] = \frac{2}{\alpha}$ et $\mathrm{Var}(X) = \frac{2}{\alpha^2}$. (c) $\mathbb{P}(X \geq \mu) = 3e^{-2} = 0.406$.
2. (a) $a = \frac{2}{b^2}$, $\mathbb{E}[X] = \frac{2b}{3}$ et $\mathrm{Var}[X] = \frac{b^2}{18}$. (b) $a = b$, pas d'espérance ni de variance.
3. (a)
   i. $\mathbb{P}(X = 1) \simeq 0.370$,
   ii. $\mathbb{P}(X = 2) \simeq 0.185$,
   iii. $\mathbb{P}(X = 3) \simeq 0.061$,
   iv. $\mathbb{P}(X \geq 4) \simeq 0.018$.

   (b) $\mathbb{E}[Y] = 500$, $\mathrm{Var}(Y) = 247500 \Rightarrow \sigma_Y \simeq 497.5$.
4. (a) $p_0 \approx 0.11$; (b) $p_1 \approx 0.93$.
5. (a) $\mu = 163$ cg, $\sigma = 9,7$ cg. (b) Environ 20000 personnes.

3.3 Échantillonnage. Théorème de la limite centrale

1. (a) ≈ 0.9772 (b) ≈ 0.9987 (c) ≈ 0.8413.
2. (a) $\approx 1 - \Phi(22.71) \approx 0$ (b) $\Phi(1.20) - 1 + \Phi(3.55) = 0.8847$.
3. (a) $1 - \Phi(2.87) = 0.0021$ (b) $\Phi(0.75) = 0.7734$.
4. (a) $\Phi(-10) \approx 0.000$ (b) $1 - \Phi(-2) = \Phi(2) = 0.977$.
5. $p_{70} = 1 - \Phi(4.8) \approx 0$ et $p_{75} = 1 - \Phi(-11) \approx 1$.
6. $2 - 2\Phi(\frac{a}{\sqrt{n/12}})$.

3.4 Estimation, méthode des moments

1. Calculer l'espérance et la variance de chacun des deux estimateurs, et appliquer la condition suffisante de convergence en probabilité.
2. (a) $\widehat{\sigma_1} = \overline{X} - \theta$: pour $\theta = 72$, on obtient donc $\overline{x} - \theta = 63.4$. C'est un estimateur sans biais et convergent.
   (b) $\widehat{\sigma_2} = S$, $\widehat{\theta} = \overline{X} - S$: on obtient $s = 30.03$ et $\overline{x} - s = 105.37$.
   (c) Ces estimateurs sont convergents.
3. (a) $\widehat{\theta}_1 = 2\overline{X}$.
   (b) $\mathbb{E}(\widehat{\theta}_1) = \theta$ et $\mathrm{Var}(\widehat{\theta}_1) = \frac{\theta^2}{3n}$.
   (c) $\mathbb{E}(\widehat{\theta}_2) = \frac{n}{n+1}\theta$ et $\mathrm{Var}(\widehat{\theta}_2) = \frac{n}{(n+2)(n+1)^2}\theta^2$.
   (d) $\widehat{\theta}_3 = \frac{n+1}{n}\widehat{\theta}_2$: $\mathbb{E}(\widehat{\theta}_3) = \theta$ et $\mathrm{Var}(\widehat{\theta}_3) = \frac{1}{n(n+2)}\theta^2$.

3.5 Méthode du maximum de vraisemblance

1. (a) $\hat{\lambda}_{MV} = \overline{K}$.
   (b) $I_n(\lambda) = \frac{n}{\lambda}$
   (c) $\mathcal{N}(\lambda, \lambda/n)$
2. (a) $\widehat{\theta}_{MV} = \frac{\sum_i \ln(R_i - r_0)}{n}$
   (b) $F_Z(z) = F_R(e^z + r_0)$, $f_Z(z) = \frac{1}{\sqrt{\pi}}e^{-(z-\theta)^2}$, $Z \sim \mathcal{N}(\theta, \frac{1}{2})$ et $\widehat{\theta}_{MV} \sim \mathcal{N}(\theta, \frac{1}{2n})$.
3. (a) $L(\theta; x_1, ...x_n) = 1_{[I_n, S_n]}(\theta)$: toutes les valeurs de θ comprises entre I_n et S_n maximisent la vraisemblance $\Rightarrow \widehat{\theta}_n = \alpha(S_n - 1) + (1 - \alpha)I_n \quad (0 \leq \alpha \leq 1)$.
   (b) $F_S(x) = 0$ si $x \leq \theta$, $(x-\theta)^n$ si $\theta \leq x \leq \theta + 1$ et 1 sinon ; $f_S(x) = n(x-\theta)^{n-1}1_{[\theta,\theta+1]}(x)$
   $F_I(x) = 0$ si $x \leq \theta$, $1 - (1 - x + \theta)^n$ si $\theta \leq x \leq \theta + 1$ et 1 sinon ; $f_I(x) = n(1 - x + \theta)^{n-1}1_{[\theta,\theta+1]}(x)$.
   (c) $\mathbb{E}(I_n) = \theta + \frac{1}{n+1}$ et $\mathbb{E}(S_n) = \theta + \frac{n}{n+1}$.
   (d) $\mathbb{E}(\widehat{\theta}_n) = \theta + \frac{1-2\alpha}{n+1} \Rightarrow \alpha^* = 1/2$.
   (e) $\widehat{\theta}_m = \overline{X} - \frac{1}{2}$, $\mathrm{Var}(\widehat{\theta}_m) = \frac{1}{12n}$.

3.6 Estimation par intervalle de confiance

1. (a) $\overline{x} = 159$, $M = x_{(4)} = 156$.
   (b) $\overline{X} \pm \frac{\sigma}{\sqrt{n}}u_{1-\alpha/2}$, a.n. : [151.59,166.41].
   (c) $n \geq \frac{4\sigma^2 u_{1-\alpha/2}^2}{\varepsilon^2}$ où ε est la précision attendue, a.n. : $n \geq 62$.
   (d) $\overline{X} \pm \frac{S^*}{\sqrt{n}}t_{n-1,1-\alpha/2}$, a.n. : [144.41,173.59].
   (e) μ connue : $\sigma^2 < \frac{n\widehat{\sigma}^2}{\chi^2_{n,\alpha}}$, a.n. : [0,690.8].
   μ inconnue : $\sigma^2 < \frac{(n-1)S^{*2}}{\chi^2_{n-1,\alpha}}$., a.n. : [0,909.8].

2. (a) $\widehat{\lambda} = \overline{X}$
   (b) $\frac{\widehat{\lambda}-\lambda}{\sqrt{\lambda/n}} \xrightarrow{L} \mathcal{N}(0,1)$.
   (c)
   (d) $\left[\widehat{\lambda} - u_{1-\alpha/2}\sqrt{\frac{\widehat{\lambda}}{n}}; \widehat{\lambda} + u_{1-\alpha/2}\sqrt{\frac{\widehat{\lambda}}{n}}\right]$.
   (e) $[2.81; 4.49]$.
3. (a) $\widehat{\theta}_{MV} = \frac{2}{\overline{X}}$.
   (b) $I_n(\theta) = \frac{2n}{\theta^2}$, $\sqrt{2n}\left(\frac{\widehat{\theta}}{\theta} - 1\right) \sim \mathcal{N}(0,1)$, $IC = \left[\frac{\widehat{\theta}}{1+u_{1-\frac{\alpha^*}{2}}/\sqrt{2n}}, \frac{\widehat{\theta}}{1-u_{1-\frac{\alpha^*}{2}}/\sqrt{2n}}\right]$
   (c) A.N. : $[0.44, 0.58]$.

3.7 Estimation optimale

1. (a) Abaque : $[0.17, 0.44]$, calcul : $[0.17, 0.43]$.
   (b) Abaque : $[0, 0.42]$ et $[0.19, 1]$, calcul : $[0, 0.41]$ et $[0.19, 1]$
2. (a) $\widehat{u} = \frac{\sum |X_i|}{n}$ est un estimateur efficace de $1/\theta$.
   (b) $\mathrm{Var}(\widehat{u}) = \frac{1}{n\theta^2}$.
3. (a) Il n'existe pas d'estimateur efficace de p. K est un estimateur efficace de $1/p$. $\mathbb{E}(K) = 1/p$. $\mathrm{Var}(K) = \frac{1-p}{p^2}$.
   (b) $\hat{p}_{MV} = 1/K$.
   (c) $\mathbb{E}(T) = p$ (estimateur sans biais) et $\mathrm{Var}(T) = p(1-p)$.
   (d) $I_K(p) = \frac{1}{p^2(1-p)}$.
   (e) $B_F(p) = p^2(1-p) < \mathrm{Var}(T)$.
4. (a) $\widehat{\sigma}^2_a$ est sans biais ssi $a = 1/(n-1)$.
   (b) $\mathrm{Var}(\widehat{\sigma}^2_a) = 2(n-1)a^2\sigma^4$.
   (c) $R(\widehat{\sigma}^2_a, \sigma^2) = \sigma^4[a^2(n^2-1) - 2a(n-1) + 1]$.
   (d) L'estimateur le plus précis est celui qui minimise le risque quadratique. Pour minimiser cette fonction, il suffit d'annuler sa dérivée. On obtient $a = \frac{1}{n+1}$.

3.8 Principe des tests d'hypothèses ; théorème de Neyman-Pearson et test UPP

1. (a) $H_0 : \mu = 0$, $H_1 : \mu \neq 0$.
   (b) $k = 0.52$. On rejette H_0.
   (c) $\widehat{\alpha} = 0.058$.
2. (a) $W = \{\widehat{\lambda} > A\}$
   (b) $A \approx \lambda_0 + u_{1-\alpha^*}\sqrt{\frac{\lambda_0}{n}}$, a.n. $A \approx 1.87$ et $\widehat{\lambda} = 1.8$: H_0 acceptée.
   (c) $\pi \approx 1 - \phi\left(\frac{A-\lambda_1}{\sqrt{\lambda_1/n}}\right)$, a.n. $\pi \approx 0.69$.

(d) Pour tout $\lambda_1 > \lambda_0$, la RC optimale est celle de la question (a) et ne dépend donc pas de λ_1 : test UPP ; pas de remise en cause de l'affirmation du Ministère de l'Intérieur.

3.

3.9 Test du rapport de vraisemblance

1. (a) Pour $x = 0, 1, 2, 3, 4$ on a resp. $\lambda(x) = 1, 1, 0.025, 0.00, 0.00$
   (b) $\mathbb{P}_{H_0}(\lambda(X) = 1) = 0.98$, $\mathbb{P}_{H_0}(\lambda(X) = 0.025) = 0.02$ et $\mathbb{P}_{H_0}(\lambda(X) = 0.00) = 0$.
   (c) $W = \{\lambda(x) < 1\}$.
2. (a) $\mathbb{E}_{MV} = \widehat{p} = X/n$
   (b) $\lambda = \left(\frac{np_0}{X}\right)^X \left(\frac{n(1-p_0)}{n-X}\right)^{n-X}$
   (c) $W = \left\{-2\left[X \ln\left(\frac{np_0}{X}\right) + (n-X)\ln\left(\frac{n(1-p_0)}{n-X}\right)\right] > \chi^2_{1,1-\alpha^*}\right\}$
   (d) $-2\ln\lambda = 10.0167$, $\chi^2_{1,0.95} = 3.84$, rejet de H_0
3. (a) $W = \{-2\log\lambda > \chi^2_{2,1-\alpha^*}\}$ avec $-2\log\lambda = \frac{n(\overline{x}-2)^2 + m(\overline{y}-4)^2}{\sigma^2}$
   (b) A.N. : $-2\log\lambda = 2.8$ et $\chi^2_{2;0.95} = 5.99$: on ne rejette pas H_0.

3.10 Tests de conformité

1. (a) $W = \{|\overline{x} - 270| > 3.94\}$
   (b) On doit donc arrêter la production
   (c) $\beta \approx 0.15$
2. (a) $\widehat{p} = X/n$. Fonction asymptotiquement pivotale : $\frac{X-np}{\sqrt{np(1-p)}} \xrightarrow{L} \mathcal{N}(0,1)$.
   (b) Forme de la région critique : $W = \{X/n > k\} = \{X > k'\}$.
   (c) Seuil critique $k' = np_0 + u_{1-\alpha^*}\sqrt{np_0(1-p_0)}$, A.N. : $W = \{X \geq 9)$.
   (d) Décision : On rejette H_0.
   (e) $\widehat{\alpha} = 0.03$.

3.11 Tests de comparaison

1. (a) Test de Fisher : $s^{*2}_X/s^{*2}_Y < F_{n_1-1;n_2-1;\alpha/2}$ ou $> F_{n_1-1;n_2-1;1-\alpha^*/2}$, a.n. $s^{*2}_X/s^{*2}_Y < 0.33$ ou > 2.80, $s^{*2}_X = 6.36$, $s^{*2}_Y = 6.95$, $s^{*2}_X/sY*2 = 0.92$: égalité des variances.
   (b) Test de Student : $\frac{|\overline{x}-\overline{y}|}{s^*\sqrt{1/n_1+1/n_2}} > t_{n_1+n_2-2;1-\alpha^*/2}$, a.n. $t_{21;0.975} = 2.080$, $s^{*2} = 6.70$, $\frac{|\overline{x}-\overline{y}|}{S^*\sqrt{1/n_1+1/n_2}} = 0.66$: égalité des moyennes.
   (c) $\overline{X} - \overline{Y} \pm s^* t_{n_1+n_2-2}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$, a.n. $[-2.98, 1.54]$.
2. (a) Test de l'hypothèse $H_0 : p_A = p_0 = 0.67$ contre $H_1 : p_A \neq p_0$. Région critique du test : $W = \{|\widehat{p}_A - p_0|\,(p_0(1-p_0)/n_A)^{-1/2} > u_{1-\alpha^*/2}\}$ avec $\widehat{p}_A = X_A/n_A$. A.N. : $\widehat{p}_A = 216/300 = 0.72$, $|\widehat{p} -$

$p_0|\left(p_0(1-p_0)/n_A\right)^{-1/2} = 1.84$ et $u_{0.975} = 1.96$. On ne rejette donc pas H_0 au niveau de 5 %.

(b) Test de l'hypothèse $H_0 : p_A = p_B$ contre $H_1 : p_A \neq p_B$. Région critique du test : $W = \{|\widehat{p}_A - \widehat{p}_B|\left(\widehat{p}(1-\widehat{p})\left(1/n_A + 1/n_B\right)\right)^{-1/2} > u_{1-\alpha^*/2}\}$ avec $\widehat{p}_A = X_A/n_A$, $\widehat{p}_B = X_B/n_B$, $\widehat{p} = (X_A + X_B)/(n_A + n_B)$. A.N. : $1.89 < 1.96$: la différence entre les deux centres n'est pas significative au niveau de 5 %.

3.12 Tests d'adéquation

1. (a) On obtient donc les probabilités suivantes :

	$5-0$	$4-1$	$3-2$	$2-3$	$1-4$	$0-5$
probabilités	1/32	5/32	10/32	10/32	5/32	1/32

   (b) $D^2 = 11.96 > \chi^2_{5;0.95} = 11.1$. L'hypothèse d'équiprobabilité est rejetée.

2. $D^2 = 15 > \chi^2_{2;0.95} = 5.99$: hypothèse H_0 d'indépendance rejetée.

3. (a) Test de Kolmogorov : $D_n = 0.27 < d_{n,1-\alpha^*} = 0.521$: on ne rejette pas H_0.

   (b) Test de normalité de Stephens : $D_n^* = 0.17$ et donc $(\sqrt{6} + \frac{0.85}{\sqrt{6}} - 0.01)D_n^* = 0.48 < 0.895$: on ne rejette pas H_0.

3.13 Analyse de la variance

1. (a) $D_n = 0.14$, $(\sqrt{5} + \frac{0.85}{\sqrt{5}} - 0.01) \times 0.14 = 0.36 < 0.895$. On ne rejette pas l'hypothèse de normalité.
   Test de Bartlett : $W : B = (N-K)\ln(MSW) - \sum_{k=1}^{K}(n_k-1)\ln(S_k^{*2}) > \chi^2_{K-1,1-\alpha}$, $MSW = 7\times 10^{-4}$, $b = 1.90$ et $\chi^2_{3,0.95} = 7.81$. On ne rejette pas l'hypothèse d'égalité des variances.

   (b) $F = \frac{MSB}{MSW} = 13.419$ et $F_{K-1,N-K,1-\alpha} = F_{3,16,0.95} = 3.24$: effet espèce significatif.

   (c) bœuf, homme : <u>6.22</u>, rat, homme : 2.03, grenouille, homme : <u>2.51</u>, rat, bœuf : <u>4.18</u>, grenouille, bœuf : <u>3.71</u>, grenouille, rat : 0.48.

2. $H = 5.35 > \chi^2_{2;0.9} = 4.61$. Donc, les effets induits par les trois médicaments sont significativement différents au niveau 10%.

3.14 Régression Linéaire

1. (a) $\widehat{b} = 3.22$, $\widehat{a} = 79.59$, $\widehat{\sigma}^2_{MV} = S_{res} = 15.66$ et $\widehat{\sigma}^2 = \frac{n}{n-2}S_{res} = 19.58$.

   (b) Intervalle bilatéral : $\widehat{a} \pm t_{n-2;1-\alpha/2}\frac{\widehat{\sigma}}{\sqrt{n}}\sqrt{1 + \frac{\overline{x}^2}{s_x^2}}$, a.n. $[64.94, 94.25]$.
   Intervalle unilatéral : $a > \widehat{a} - t_{n-2;1-\alpha}\frac{\widehat{\sigma}}{\sqrt{n}}\sqrt{1 + \frac{\overline{x}^2}{s_x^2}} = 67.78$.

   (c) $\frac{\widehat{b}}{\widehat{\sigma}/\sqrt{ns_x^2}} = 5.87 > t_{8;0.99} = 2.90$: H_0 rejetée.

(d) Intervalle bilatéral $\widehat{b} \pm t_{n-2;1-\alpha/2}\frac{\widehat{\sigma}}{\sqrt{ns_x^2}}$, a.n. $[1.95, 4.48]$.

Intervalle unilatéral : $b < \widehat{b} + t_{n-2;1-\alpha}\frac{\widehat{\sigma}}{\sqrt{ns_x^2}} = 4.23$.

2. (a) $\widehat{r} = \frac{\sum_{i=1}^n x_i U_i}{\sum_{i=1} x_i^2}$ et $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (U_i - \widehat{r}x_i)^2$.
   (b) $\mathbb{E}(\widehat{r}) = r$, $\mathrm{Var}(\widehat{r}) = \sigma^2 / \sum_{i=1}^n x_i^2$ et $\widehat{r} \sim \mathcal{N}\left(r, \frac{\sigma^2}{\sum_{i=1}^n x_i^2}\right)$.
   (c) $\widehat{r} \pm \frac{u_{1-\alpha/2}}{\sqrt{\sum_{i=1}^n x_i^2}}$
   (d) A.n. : $\sum x_i^2 = 100.5, \sum x_i u_i = 207.2, \widehat{r} = 2.0617, IC = [1.8662, 2.2572]$.
   (e) Test sur la moyenne d'une v.a. gaussienne de variance connue : $W = \{\frac{|\widehat{r} - r_0|}{\sqrt{1/\sum_{i=1}^n x_i^2}} > u_{1-\alpha^*/2}\}$, a.n. : $1.6210 < 2.5758$: H_0 acceptée.

3.15 Problèmes

3.15.1 Estimation

Comparaison d'intervalles de confiance

1. Première méthode
   (a) Equation de vraisemblance : $\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{\sigma^2} + \frac{k}{2\sigma^4}\sum_{i=1}^n x_i^2 = 0$.
   (b) $\widehat{\sigma}^2_{MV}$ est efficace et $\mathrm{Var}(\widehat{\sigma}^2_{MV}) = \frac{\sigma^4}{n}$.
   (c) $1 - \alpha = P\left(\frac{\widehat{\sigma}^2_{MV}}{1+\frac{u_{1-\alpha/2}}{\sqrt{n}}} < \sigma^2 < \frac{\widehat{\sigma}^2_{MV}}{1-\frac{u_{1-\alpha/2}}{\sqrt{n}}}\right)$
   (d) $\widehat{\sigma}^2_{MV} = 4.07$. $IC = [3.57; 4.72]$.
2. Deuxième méthode
   (a) $\mathbb{E}(\widehat{\sigma}^2_{MV}) = \frac{k}{2n}n\mathbb{E}(X^2) = \sigma^2$.
   (b) $\widehat{\sigma}_m = \overline{X}\sqrt{\frac{k}{2-k}}$.
   (c) $\mathbb{E}(\widehat{\sigma}_m) = \mathbb{E}(X)\sqrt{\frac{k}{2-k}} = \sigma$. $\mathrm{Var}(\widehat{\sigma}_m) = \frac{k}{2-k}\frac{\sigma^2}{n}$.
   (d) $1 - \alpha = P\left(\frac{\widehat{\sigma}^2_m}{\left(1+u_{1-\alpha/2}\sqrt{\frac{k}{n(2-k)}}\right)^2} < \sigma^2 < \frac{\widehat{\sigma}^2_m}{\left(1-u_{1-\alpha/2}\sqrt{\frac{k}{n(2-k)}}\right)^2}\right)$.
   (e) $\widehat{\sigma}^2_m = 4.06$. $IC' = [3.53; 4.72]$.

Loi exponentielle

1. Premiére partie
   (a) $\widehat{\theta}_1 = \overline{X} - 1$. $\mathbb{E}(\widehat{\theta}_1) = \theta$. $\mathrm{Var}(\widehat{\theta}_1) = 1/n$.
   (b) $I = [\overline{X} - 1 - \frac{u_{1-\alpha/2}}{\sqrt{n}}, \overline{X} - 1 + \frac{u_{1-\alpha/2}}{\sqrt{n}}]$.
   (c) A.N. $\overline{x} = 4.0095$, $I = [2.9657, 30533]$.
2. Deuxième partie
   (a) $L(\theta; x_1, \ldots, x_n) = e^{n(\theta - \overline{x})}$ si $\theta \leq \min x_i$, 0 sinon.
   (b) $\widehat{\theta}_2 = \min(X_i)$.

(c) $F_{\widehat{\theta}_2}(x) = \begin{cases} 1 - e^{n(\theta - x)} & \text{si } x > \theta \\ 0 & \text{sinon} \end{cases}$

(d) $I' = [\min(X_i) + \log(\alpha/2)/n, \min(X_i) + \log(1-\alpha/2)/n]$.

(e) $I' = [2.99826, 3.00008]$.

Estimateur le plus précis

1. Première partie

   (a) Estimateurs du maximum de vraisemblance de μ et σ^2 : $\widehat{\mu} = \overline{X}$ et $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = S^2$.

   (b) $\mathbb{E}(\widehat{\sigma}_a^2) = a(n-1)\sigma^2$. Donc, $\widehat{\sigma}_a^2$ est sans biais ssi $a = \frac{1}{n-1}$.

   (c) i. On utilise la décomposition du risque en variance + biais au carré. Ici, $\mathrm{Var}(\widehat{\sigma}_a^2) = a^2\sigma^4(2n-2)$, d'où le résultat.

   ii. En annulant la dérivée du risque par rapport à a, on obtient $a = \frac{1}{n+1}$.

2. Seconde partie

   (a) $b = 1/n$.

   (b) On applique à nouveau la formule de décomposition risque = variance + biais au carré.

   (c) Le risque dépend de μ et σ^2. Lorsque n tend vers l'infini, b^* tend vers $1/n$, valeur qui constitue donc une bonne approximation de b^* pour n grand.

3.15.2 Tests

Loi $\gamma(t, \lambda)$

1. Utiliser le fait que l'intégrale de la fonction de densité est égale à 1, et faire un changement de variable.
2. $\overline{X}$ est l'estimateur efficace de $\frac{2}{\lambda}$. $\mathbb{E}(\overline{X}) = \frac{2}{\lambda}$. $\mathrm{Var}(\overline{X}) = \frac{2}{n\lambda^2}$. D'où $\mathbb{E}(X) = \frac{2}{\lambda}$ et $\mathrm{Var}(X) = \frac{2}{\lambda^2}$.
3. (a) Test UPP de région critique : $W = \{\overline{x} < A\}$.

   (b) $A \approx \frac{2}{\lambda_0}(1 - \frac{u_{1-\alpha^*}}{\sqrt{2n}})$.

   (c) $\overline{x} = 3.6$ et $A = 2.36$: on accepte l'hypothèse H_0.

   (d) $\beta(\lambda_1) = 0.25$.
4. Test de Neyman-Pearson de la forme $W = \{\prod_i x_i < cste\}$.
5. Test de Kolmogorov-Smirnov. $D_n = 01.598 > d_{8,0.95} = 0.457$: on accepte H_0.

Temps de fonctionnement sans panne d'un appareil

1. Première expérience

   (a) $\mathbb{E}(T) = \mu$.

   (b) $\widehat{\mu} = \overline{T}$, estimateur efficace de μ, de variance μ^2/n.

   (c) Problème de test $H_0 : \mu = \mu_0 (= 750)$ contre $H_1 : \mu > \mu_0$. Test UPP de région critique $W = \{\overline{T} > A\}$, avec $A = \mu_0(1 + u_{1-\alpha^*}/\sqrt{n})$. A ; N. : $A = 832$; on accepte H_0.
   (d) $\pi = \Phi(\sqrt{n} - \frac{\mu_0}{\mu_1}(\sqrt{n} + u_{1-\alpha^*}))$. A.N. : $\pi = 0.87$.
2. Deuxième expérience
   (a) $p = 1 - e^{-\tau/\mu}$.
   (b) $X \sim \mathcal{B}(n, p)$.
   (c) Hypothèses : $H_0 : p = p_0$ contre $H_1 : p < p_0$, avec $p_0 = 1 - e^{-\tau/\mu_0} = 0.486$. Test UPP : $W = \{X < A\}$ avec $A \approx np_0 - \sqrt{np_0(1-p_0)}u_{1-\alpha^*}$. A. N. : $A = 98$.
   (d) $\pi \approx \Phi(\frac{\sqrt{n}(p_0-p_1) - \sqrt{p_0(1-p_0)}u_{1-\alpha^*}}{\sqrt{p_1(1-p_1)}})$. A.N. : $p_1 = 0.4262$, $\pi = 0.57$.
   (e) $n = \left(\frac{\sqrt{p_1(1-p_1)}u_{1-\beta} + \sqrt{p_0(1-p_0)}u_{1-\alpha^*}}{p_0 - p_1}\right)^2$. A.N. : $n = 523$.

Test du rapport de vraisemblance

1. $\widehat{\theta}_{MV} = \frac{1}{\overline{X}}$.
2. $I_n(\theta) = \frac{n}{\theta^2}$. D'après la propriété de l'estimateur du maximum de vraisemblance : $\frac{\widehat{\theta}-\theta}{\theta/\sqrt{n}} \xrightarrow{L} \mathcal{N}(0, 1)$.
3. Test de Neyman-Pearson : $W = \{\widehat{\theta} > k'\}$ pour une certaine constante k', avec $k' \approx \theta_0\left(1 + \frac{u_{1-\alpha^*}}{\sqrt{n}}\right)$.
4. Il n'existe pas de test UPP.
5. $\lambda = \left(\frac{\theta_0}{\widehat{\theta}}\right)^n \exp\left[n\left(1 - \frac{\theta_0}{\widehat{\theta}}\right)\right]$.
6. La fonction $\ln \lambda = g(\widehat{\theta})$ est strictement croissante sur $[0, \theta_0[$, puis strictement décroissante sur $]\theta_0, +\infty[$. La RC du test du rapport de vraisemblance est donc de la forme $W' = \{\widehat{\theta} < k_1$ ou $\widehat{\theta} > k_2\}$, avec $k_1 \approx \theta_0\left(1 - \frac{u_{1-\alpha^*/2}}{\sqrt{n}}\right)$ et $k_2 \approx \theta_0\left(1 + \frac{u_{1-\alpha^*/2}}{\sqrt{n}}\right)$.

Test randomisé

1. $A = 14 \Rightarrow \alpha = 0.0207$.
2. $\alpha = \mathbb{P}_{H_0}(X > A) + \gamma\mathbb{P}_{H_0}(X = A) = 0.0207 + 0.037\gamma$. Donc $\alpha = 0.05 \Leftrightarrow \gamma = 0.79$.
3. $\pi = \mathbb{P}_{H_1}(X > A) + \gamma\mathbb{P}_{H_1}(X = A) = \mathbb{P}_{H_1}(n - X \leq n - A) + \gamma\mathbb{P}_{H_1}(n - X = n - A) = 0.9133 + 0.79(0.9133 - 0.8042) = 0.999$.