Ecole Polytechnique
Année 1998-1999

Petites classes de Statistique

Christian Robert

0. Modélisation

Le tableau ci-dessous reproduit les résultats d'une expérience médicale sur les durées de survie de malades sous un traitement donné.

mois	≤ 5	$5-10$	$10-15$	$15-20$	$20-25$	$25-30$	$30-35$	$35-40$	$40-45$	$45-50$	≥ 50
survivants	40	35	28	22	18	13	9	5	5	3	2
décés	5	7	6	4	5	4	4	0	2	1	2

1. Donner la loi de l'échantillon si l'on suppose les observations indépendantes et exponentielles $\mathcal{E}xp(\lambda)$. Déduire de la vraisemblance la valeur de λ la plus "probable".
2. On associe à chaque intervalle l'indice entier correspondant et on suppose que la probabilité de tomber dans l'intervalle i est donnée par la loi de Poisson, $\mu^i \exp(-\mu)/i!$. Ecrire la vraisemblance correspondant aux observations et en déduire la valeur de μ la plus "probable".

1. – Capture-recapture.

Soit une population de taille N inconnue. On sélectionne (*capture*) n_1 individus (*animaux*) dans cette population et on les identifie (*marquage*) avant de les relâcher. On sélectionne à nouveau n_2 individus et on observe n_{11} individus marqués.

1. Donner la loi de n_{11} en fonction de n_1, n_2 et N.
2. Donner l'espérance de n_{11} conditionnelle-ment à n_1 et n_2 et en déduire un estimateur intuitif de N.
3. On installe successivement r trappes. Soit q la probabilité qu'un animal soit capturé par une trappe et N la taille de la population avant l'expérience.
   a. Supposant que les animaux capturés ne sont pas relachés, donner la probabilité que les effectifs de capture soient $n_1, \ldots, n_r$.
   b. Si les animaux capturés sont relachés après *marquage*, donner la probabilité que les effectifs des captures soient $n_1, \ldots, n_r$ et que les effectifs d'animaux *recapturés* soient $m_1, \ldots, m_{r-1}$.

1. – Mélanges de distributions.

Etant donnés k distributions de densités $f_1, \ldots, f_k$ et des réels $p_1, \ldots, p_k$ tels que $0 \leq p_i$ et $p_1 + \ldots + p_k = 1$, la distribution de densité

$$p_1 f_1 + \ldots + p_k f_k \tag{1}$$

est dite *mélange des k distributions.*

1. Soient f_1 et f_2 les densités de $\mathcal{N}(\theta_1, \sigma_1^2)$ et de $\mathcal{N}(\theta_2, \sigma_2^2)$. Pour $k = 2$, montrer que, si $(\theta_1, \sigma_1) \neq (\theta_2, \sigma_2)$ et $0 < p_2 < 1$, le modèle ne peut être normal.

2. Montrer que, dans le cas particulier $\sigma_1 = \sigma_2$, $p = 1/2$, le mélange est bimodal si $|\theta_1 - \theta_2| > 2\sigma_1$.
3. Montrer que les moments non centrés de (1) s'écrivent comme combinaison linéaire des moments correspondants des f_i. En déduire une méthode d'estimation des paramètres éventuels du modèle.
4. Montrer que le modèle (1) peut aussi s'écrire comme un modèle conditionnel où $z \sim \mathcal{M}_1(p_1, \ldots, p_k)$ et $x|z \sim f_z(x)$.

3. – Invariance du rapport de Fisher.

Soit $x = (y^t, z^t)^t \in \mathbb{R}^p$ avec $y \in \mathbb{R}^q$ $(q < p)$. Montrer que, si x suit une loi de densité $f(||y-\theta||^2 + ||z||^2)$ par rapport à la mesure de Lebesgue, la loi de

$$T(x) = \frac{||y-\theta||^2}{||z||^2}\frac{p-q}{q}$$

est la loi de Fisher $\mathcal{F}_{q,p-q}$, quel que soit f.

4. – Moments de modèles exponentiels.

Soit

$$x \sim f(x|\theta) = e^{\theta.x - \psi(\theta)}$$

par rapport à une certaine mesure ν de $\mathbb{R}^n$.

1. Montrer que $\mathbb{E}_\theta[x] = \nabla\psi(\theta)$ et obtenir des relations similaires sur les moments d'ordre supérieur. Comment exprime-t-on la fonction génératrice?
2. Montrer que, pour tout échantillon $x_1, \ldots, x_m$, il existe une statistique exhaustive de dimension n.
3. Si $\theta \sim \pi(\theta) \propto \exp(\theta.x_0 - \lambda\psi(\theta))$, exprimer $\pi(\theta|x)$ et montrer que $\mathbb{E}^\pi[\nabla\psi(\theta)] = x_0/\lambda$.

5. – Exemples de familles exponentielles.

Montrer que les lois suivantes sont dans des familles exponentielles:

(i) Négative binomiale $\mathcal{N}eg(n,p)$;
(ii) Wishart $\mathcal{W}_p(n,\Sigma)$;
(iii) Pareto $\mathcal{P}a(\alpha,1)$ (qu'en est-il de $\mathcal{P}(\alpha,\theta)$ pour α fixé?);
(iv) Multinomiale $\mathcal{M}_n(k; p_1, \ldots, p_k)$;
(v) Khi-deux décentré $\chi^2_n(\lambda)$;
(vi) Weibull(α,β).

Pour chaque loi, identifier l'espace naturel des paramètres et l'éventuelle reparamétrisation des observations.

6. – Simulation de lois par acceptation–rejet.

Soient f et g des densités de probabilité telles qu'il existe M tel que $f(x) \le Mg(x)$. Montrer que l'algorithme

(i) Générer x suivant $g(x)$ et u suivant $\mathcal{U}_{[0,1]}$;

(ii) Prendre x si $u \leq f(x)/Mg(x)$ et repartir en (i) sinon.

permet de générer une observation suivant la loi de densité g.

0. Modélisation

On dispose de données de terrain sur l'habitat de lézards.

		Hauteur	
		< 9	≥ 9
Diamètre	≤ 2	54	44
	> 2	25	18

		Période		
		matin	a.m.	soir
Ensoleillement	soleil	47	20	22
	ombre	94	205	43

1. Montrer que ces tables sont des statistiques exhaustives pour les paramètres du modèle si l'on suppose les observations distribuées suivant des lois de Poisson $\mathcal{P}oi(\lambda_{ij})$ pour chaque modalité des catégories ((hauteur, diamètre) ou (période,ensoleillement)). En est-il de même si l'on fait une hypothèse de loi multinomiale $\mathcal{M}_k(n; p_{ij})$ avec $k = 4, 6$ et $n = 151, 429$? On pourra comparer les fonctions de vraisemblance dans les deux cas. Donner les maxima des fonctions de vraisemblance.
2. Un modèle logistique représente les probabilités sous la forme

$$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \mu + \alpha_i + \beta_j \qquad (1)$$

   (a) Montrer qu'avec les contraintes d'identifiabilité $\alpha_1 = \beta_1 = 0$, le nouveau modèle est identifiable.
   (b) Donner des statistiques exhaustives pour ce modèle.
   (c) Modéliser par (1) la probabilité d'exposition au soleil, décrite par la table suivante:

		matinée	a.m.	soir
soleil	$H < 5$	28	12	9
	$H \geq 5$	19	8	13
ombre	$H < 5$	51	129	26
	$H \geq 5$	43	76	17

1. – Exhaustivité.

Soit $x \sim f(x|\theta)$. Une statistique L est dite *libre* si la distribution de $L(x)$ ne dépend pas de θ.

1. Montrer que la statistique d'ordre $(X_{(1)}, \ldots, X_{(n)})$ de l'échantillon $X_1, \ldots, X_n$ est toujours exhaustive. Montrer que, si $f(x|\theta)$ est la loi de Cauchy de paramètre de position θ, cette statistique d'ordre est aussi exhaustive minimale. Montrer que le vecteur $(X_{(1)} - X_{(2)}, \ldots, X_{(1)} - X_{(n)})$ est libre pour toute famille à paramètre de position.

2. Soient $x_1, \ldots, x_n$ i.i.d. $\mathcal{U}([\theta - 1/2, \theta + 1/2])$ avec $\theta \in \mathbb{R}$.
   (a) Montrer que $(X_{(1)}, X_{(n)})$ est exhaustive minimale.
   (b) Montrer que $\Delta = X_{(n)} - X_{(1)}$ est *libre* et en déduire que $(X_{(1)}, X_{(n)})$ n'est pas *totale*. Pour $\delta = (X_{(1)} + X_{(n)})/2$, utiliser la majoration $|\delta - \theta| \leq \Delta/2$ pour montrer que les statistiques libres sont parfois utiles.
3. Soient $(x_1, y_1), \ldots, (x_n, y_n)$ i.i.d. et distribués suivant $f(x, y) = \exp\{-\theta x - y/\theta\}$.
   (a) Montrer que $(\sum_i x_i, \sum_i y_i)$ forme une statistique exhaustive minimale.
   (b) Montrer que, si

$$T = \sqrt{\sum_i y_i / \sum_i x_i}, \qquad U = \sqrt{\sum_i y_i \sum_i x_i},$$

   T est l'estimateur du maximum de vraisemblance de θ et U est libre.
4. (*Théorème de Basu*) Montrer que, si S est exhaustive et totale et si T est libre, T et S sont indépendantes. Utiliser les cas ci-dessus pour montrer que l'exhaustivité minimale n'est pas suffisante.
5. Soient X_1, X_2 iid suivant $\alpha x^{\alpha-1} \exp\{-x^\alpha\}$. Montrer que $\log(X_1)/\log(X_2)$ est libre.
6. Soient $X_1, \ldots, X_n$ iid suivant $\exp\{-(x - \mu)\}\mathbb{I}\{\mu \leq x\}$. Montrer que $X_{(1)}$ est totale et exhaustive et déduire que S^2 et $X_{(1)}$ sont indépendantes.
7. Montrer que T totale et exhaustive est nécessairement exhaustive minimale.
8. Montrer que, si la loi des observations est inconnue, la statistique d'ordre $(X_{(1)}, X_{(n)})$ est totale.

2. – Le phénomène de Stein.

1. Soit $x \sim \mathcal{N}(\theta, 1)$ une observation normale. Si $L(\theta, d) = (\theta - d)^2$ est le coût sous lequel on évalue les estimations d de θ, montrer que le *risque* de $\delta_0(x) = x$ est $R(\theta, \delta_0) = 1$.
2. On prend une loi *a priori* $\mathcal{N}(0, n)$ sur θ. Montrer que l'espérance *a posteriori* de θ sachant x est $\delta_n(x) = (n/n + 1)x$ et donner le risque de δ_n, $R(\theta, \delta_n)$.
3. Montrer que

$$\int \{R(\theta, \delta_0) - R(\theta, \delta_n)\} e^{-\theta^2/2} d\theta$$

   tend vers 0 quand n tend vers $+\infty$ et en déduire que δ_0 est admissible.
4. Soit $x \sim \mathcal{N}_p(\theta, I_p)$. Montrer le *lemme de Stein*: si ψ est une fonction de $\mathbb{R}^p$ dans $\mathbb{R}^p$ différentiable, alors

$$\mathbb{E}_\theta[(x - \theta)^t \psi(x)] = \mathbb{E}_\theta[\nabla^t \psi(x)].$$

   En déduire que, si $\delta(x) = x + \psi(x)$ et si ψ satisfait

$$||\psi(x)||^2 + 2\nabla^t \psi(x) \leq 0,$$

   l'estimateur δ domine $\delta_0(x) = x$ sous le coût $L(\theta, d) = ||\theta - d||^2$.

5. Examiner les cas particuliers $p = 1$ et $\psi(x) = (p-2)x/||x||^2$ $(p > 2)$.

3. – Coûts quadratiques.

1. Soit $x \sim \mathcal{N}_p(\theta, I_p)$. On cherche à estimer $\eta = ||\theta||^2$ et à évaluer les estimations δ de η sous le coût quadratique $(\delta - \eta)^2$.
   a) Comparer les estimateurs de la forme $||x||^2 + c$ et montrer que $\delta^*(x) = ||x||^2 - p$ domine uniformément tous les autres estimateurs dans cette classe.
   b) Montrer que δ^* est sans biais. Proposer un estimateur améliorant δ^*.
2. Soient les deux fonctions de coût "intrinsèques":

$$L_E(\theta, \delta) = \mathbb{E}_\theta \left[\log(f(x|\theta)/f(x|\delta))\right], \qquad L_H(\theta, \delta) = \mathbb{E}_\theta \left[\left(\sqrt{f(x|\theta)/f(x|\delta)} - 1\right)^2\right],$$

   a) Montrer que L_H est toujours défini, au contraire de L_E, et borné.
   b) Montrer que les deux coûts admettent comme approximation du second ordre $(\theta - \delta)^2 I(\theta)$, où $I(\theta)$ désigne l'information de Fisher.
   c) Si $f(x|\theta) = h(x)\exp\{\theta \cdot x - \psi(\theta)\}$, donner la forme générale de L_E et L_H.

I. – Moindres carrés contraints.

Soit le modèle linéaire

$$y = X\beta + \epsilon$$

avec $y, \epsilon \in \mathbb{R}^n$, $\beta \in \mathbb{R}^p$ et X une matrice (n,p).

1. On suppose que $R\beta = 0$ où R est une matrice (q,p). Donner l'estimateur des moindres carrés de β sous cette contrainte. Si on suppose que l'erreur ϵ a une matrice de variance Σ, quelle est la forme de l'estimateur des moindres carrés généralisés? Donner la forme de la matrice de variance de cet estimateur restreint.
2. On suppose à présent que $\beta_1 \leq \beta_2 \leq \ldots \leq \beta_p$. Montrer que l'algorithme suivant conduit à l'estimateur des moindres carrés satisfaisant cette contrainte (on prendra $X^tX = I_p$) :
   (a) Calculer l'estimateur des moindres carrés, $\hat\beta$;
   (b) Moyenner les groupes de composantes mal ordonnées, c'est-à-dire $\hat\beta_i, \hat\beta_{i+1}, \ldots, \hat\beta_j$, si $\hat\beta_{j+1} \geq \hat\beta_i > \hat\beta_j \geq \hat\beta_{i-1}$ et prendre $\tilde\beta_i = \ldots = \tilde\beta_j = (\hat\beta_i + \ldots \hat\beta_j)/(j-i+1)$.

II. – Paradoxe de prévision.

On observe le modèle linéaire $y = X\beta + \epsilon$, défini comme en I. On observe $x^0 = (x_1, \ldots, x_p)$ et on cherche à prédire $y^0 \sim \mathcal{N}(x_1\beta_1 + \ldots + x_p\beta_p, \sigma^2)$ sachant que $\epsilon \sim \mathcal{N}_n(0, \sigma^2 I_n)$.

1. Montrer que chercher à minimiser $\mathbb{E}^{y,y_0}_{\beta,\sigma}[(\hat y(y) - y_0)^2]$ revient à minimiser
$$\mathbb{E}^{y,y_0}_{\beta,\sigma}[(x_0\beta - \hat y(y))^2]$$
en travaillant conditionnellement à X et x_0.
2. En déduire que la meilleure prédiction linéaire de y_0 est $x_0\hat\beta$ où $\hat\beta$ est l'EMC de β.
3. On considère à présent x^0 comme un vecteur aléatoire de moyenne 0 et de variance A. Montrer que le problème revient à considérer le risque
$$\mathbb{E}_{\beta,\sigma}[(\beta - \hat\beta)^t A(\beta - \hat\beta)], \tag{1}$$
donc à travailler sous un coût quadratique associé à A.
4. En déduire que, si le rang de A est au moins égal à 3, il peut exister des estimateurs préférables à $\hat\beta$ suivant ce critère (voir PC(2)). Le critère (1) est-il défendable?

III. – Moyenne commune.

On considère $y_1 \sim \mathcal{N}_p(\beta_1, \sigma^2 I_p)$ et $y_2 \sim \mathcal{N}_p(\beta_1, \lambda\sigma^2 I_p)$. Montrer que le modèle n'est pas identifiable si λ n'est pas connu. Si λ est connu, montrer que l'estimateur

$$\frac{\lambda y_1}{1-\lambda} + \frac{y_2}{1-\lambda}$$

est l'estimateur du maximum de vraisemblance de β_1 et le meilleur estimateur de la forme $\alpha y_1 + (1-\alpha) y_2$ sous coût quadratique. Etendre au cas où la variance de y_i est Σ_i $(i = 1, 2)$.

IV. – Modèles à erreurs dans les variables.

Soit un modèle linéaire $y_i = \alpha + \beta x_i + \epsilon_i$ avec $x_i \sim \mathcal{N}_1(\xi_i, \tau^2)$ et $\epsilon_i \sim \mathcal{N}_1(0, \sigma^2)$. Ce modèle est dit *à erreurs dans les variables* car on observe ξ_i avec une erreur $(x_i - \xi_i)$.

1. On considère l'extension au cas où les ξ_i sont aussi aléatoires, avec $\xi_i \sim \mathcal{N}_1(\xi, \sigma_\xi^2)$. (Cette extension est dite *structurelle*.) Exprimer la vraisemblance du modèle initial (ou *fonctionnel*) et montrer que l'extension structurelle se ramène à un traitement bayésien du modèle fonctionnel. Donner les lois marginales de (y_i, x_i) et de $(x_1, \ldots, x_n, y_1, \ldots, y_n)$ en fonction de (α, β).
2. Expliquer pourquoi il convient de remplacer le critère usuel des moindres carrés par celui des *moindres carrés orthogonaux*,
$$\min_{(\hat{x}_i, \hat{y}_i)} \sum_{i=1}^n \left\{ (x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right\} , \tag{2}$$
les $(\hat{x}_i, \hat{y}_i)$ appartenant à la droite $y = \hat{\alpha} + \hat{\beta} x$.
3. Montrer que les solutions $(\hat{x}_i, \hat{y}_i)$ de (2) satisfont
$$\hat{x}_i = \frac{\hat{\beta} y_i + x_i - \hat{\alpha}\hat{\beta}}{1 + \hat{\beta}^2}, \qquad \hat{y}_i = \hat{\alpha} + \frac{\hat{\beta}}{1 + \hat{\beta}^2} (\hat{\beta} y_i + x_i - \hat{\alpha}\hat{\beta}).$$
En déduire que (2) peut s'écrire
$$\frac{1}{1 + \hat{\beta}^2} \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta} x_i)^2 = \frac{1}{1 + \hat{\beta}^2} (S_y + \hat{\beta}^2 S_x - 2\hat{\beta} S_{xy})$$
avec $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$, $S_y = \sum_i (y_i - \bar{y})^2$, $S_x = \sum_i (x_i - \bar{x})^2$ et $S_{xy} = \sum_i (y_i - \bar{y})(x_i - \bar{x})$.
4. Montrer que les estimateurs des moindres carrés sont alors
$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}, \quad \hat{\beta} = \frac{-(S_x - S_y) + \sqrt{(S_x - S_y)^2 + 4 S_{xy}}}{2 S_{xy}}.$$
5. On suppose $y_i \sim \mathcal{N}(\alpha + \beta \xi_i, \sigma^2)$ et $x_i \sim \mathcal{N}(\xi_i, \tau^2)$ avec $\tau^2 = \lambda \sigma^2$ et $\lambda > 0$ connu. Montrer que cette hypothèse est nécessaire pour obtenir une vraisemblance bornée et en déduire que les estimateurs du maximum de vraisemblance satisfont
$$\xi_i^* = \frac{x_i + \lambda \beta^* (y_i - \alpha^*)}{1 + \lambda {\beta^*}^2}.$$
6. Définissant $\tilde{\alpha} = \sqrt{\lambda}\alpha$, $\tilde{\beta} = \sqrt{\lambda}\beta$ et $\tilde{y}_i = \sqrt{\lambda} y_i$, montrer que la maximisation de la vraisemblance revient à minimiser
$$\frac{1}{1 + \tilde{\beta}^2} \sum_{i=1}^n (\tilde{y}_i - \tilde{\alpha} - \tilde{\beta} x_i)^2$$

et en déduire α^*, β^* et σ^*.

V. – Estimateurs ridges.

Soit $y \sim \mathcal{N}_p(X\beta, \sigma^2 I_p)$, $\beta \in \mathbb{R}^q$.

1. Montrer que le modèle n'est pas identifiable si $rg(X) < p$. Comment cette condition s'interprète-t-elle en termes de valeurs propres de $X^t X$?
2. Montrer que, si $\lambda_{\min}(X^t X)$ tend vers 0, la variance de certaines projections sur $\mathbb{R}$ de l'estimateur du maximum de vraisemblance $\hat{\beta}$ tend vers ∞.
3. Afin de corriger cet effet, on introduit l'estimateur *ridge*

$$\hat{\beta}_\xi = (X^t X + \xi I_p)^{-1} X^t y.$$

Déterminer moyenne et variance de $\hat{\beta}_\xi$. On montrera en particulier que cette variance est inférieure ou égale à

$$(X^t X + \xi I_p)^{-1}$$

au sens des matrices définies positives. L'estimateur β_ξ domine-t-il uniformément $\hat{\beta}$?

0. Modélisation

A. Dans des contextes comme celui de la capture-recapture (voir PC(1)), il est possible d'observer des lois binomiales dont le paramètre de taille est inconnu. Soit donc $x_1, \ldots, x_n$ un échantillon iid de $\mathcal{B}(k,p)$ où p est supposé connu et k est le paramètre inconnu.

1. Montrer que le rapport des vraisemblances est
$$\frac{[k(1-p)]^n}{\prod_{i=1}^n (k-x_i)} \mathbb{I}_{k \geq x_{(n)}}$$
et en déduire que $x_{(n)}$ n'est pas exhaustive. (Ce modèle est-il encore exponentiel ?)
2. Montrer que l'équation
$$(1-p)^n = \prod_{i=1}^n (1-x_i z)$$
admet une unique solution z^* dans $[0, 1/x_{(n)}]$ et en déduire que l'estimateur du maximum de vraisemblance $\hat{k}$ est la partie entière de $1/z^*$.
3. Appliquer aux deux échantillons $(16, 18, 22, 25, 27)$ et $(16, 18, 22, 25, 28)$ pour $p = 0.3$.

B. La *distribution de Weibull* a pour densité
$$f(t) = \lambda^\gamma \gamma (t-\mu)^{\gamma-1} \exp[-\lambda^\gamma (t-\mu)^\gamma] \, \mathbb{I}_{\mu < t} \, .$$

1. Montrer que la fonction de répartition associée est explicite et que γ détermine son évolution.
2. Montrer que cette distribution n'est pas dans une famille exponentielle si, et seulement si, (γ, μ) est inconnu. Donner les équations de vraisemblance en (γ, δ) à μ fixé. L'estimateur du maximum de vraisemblance de μ est-il toujours $x_{(1)}$?
3. Construire les estimateurs du maximum de vraisemblance de (μ, γ, λ) pour des données de survie à un carcinogène :
$143, 164, 188, 190, 206, 209, 213, 216, 220, 227, 230, 234, 246, 265, 304$

1. Exercices

1. Dans le contexte d'un mélange de deux distributions normales $\mathcal{N}(\theta_i, \sigma_i^2)$, montrer que l'estimateur du maximum de vraisemblance n'est pas défini si les cinq paramètres $\theta_1, \theta_2, \sigma_1, \sigma_2, p$ sont inconnus. Donner une interprétation de ce phénomène en termes d'identifiablité statistique.
2. Soit $y \sim \mathcal{B}(1, (1+e^\theta)^{-1})$. Montrer qu'il n'existe pas d'estimateur du maximum de vraisemblance de θ et que le même phénomène se produit si $y \sim \mathcal{B}(n, (1+e^\theta)^{-1})$ et $y = 0, n$.

3. Soient x_1, x_2 i.i.d. $\mathcal{C}(\theta, 1)$. Montrer que la fonction de vraisemblance est bimodale quand $|x_1 - x_2| > 2$ et conclure sur l'existence de l'estimateur du maximum de vraisemblance. Montrer que la multimodalité disparait pour 3 observations et plus.
4. Dans une urne contenant 1000 tickets, 20 sont marqués θ et 980 sont marqués 10θ. Un ticket est tiré au hasard et son numéro, x, lu. Donner l'estimateur du maximum de vraisemblance de $\theta, \delta(x)$, et montrer que $P(\delta(X) = \theta) = 0.98$. On renumérote les tickets 10θ par $a_i\theta$ ($i \leq 980$), avec $a_i \in [10, 10.1]$ et $a_i \neq a_j$ ($i \neq j$). Donner le nouvel estimateur du maximum de vraisemblance, δ', et montrer que $P(\delta'(x) \; < \; 10\;\theta) \; = \; 0.02$.
5. Pour
$$f(x) \; = \; \frac{1}{x} \; \exp\left[-50\left(\frac{1}{x} \; - \; 1\right)^2\right], \qquad (x \; > \; 0)$$
montrer que f est intégrable et qu'il existe $a, b > 0$ tels que
$$\int_0^b af(x)dx = 1 \quad \text{et} \quad \int_1^b af(x)dx = \; 0.99.$$
Ayant ainsi construit la densité
$$p(y \mid \theta) \; = \; a\theta^{-1}f(y\theta^{-1})\mathbb{I}_{[0,b\theta]}(y),$$
donner l'estimateur du maximum de vraisemblance de θ, $\delta(y)$ et montrer que $P(\delta(y) \; > \; 10\theta) = 0.99$.

2. Paramètre de non centralité

Soit un vecteur normal $x \sim \mathcal{N}_p(\theta, I_p)$. Le paramètre d'intérêt est $\lambda = ||\theta||^2$.

1. Montrer que l'estimateur du maximum de vraisemblance de λ est $\delta_0(x) = ||x||^2$.
2. Donner l'espérance de δ_0, en déduire le biais de cet estimateur et un estimateur dominant δ_0.
3. Montrer que $\delta_1(x) = ||x||^2 - p$ est le meilleur estimateur de la forme $||x||^2 - c$ ($c \in \mathbb{R}$).
4. Donner la loi de $\delta_0(x)$, montrer qu'elle ne dépend que de λ et en déduire l'estimateur du maximum de vraisemblance de λ fondé sur $\delta_0(x)$, $\delta_2(x)$. Le fait que δ_2 est différent de δ_0 est-il paradoxal ? Ce phénomène peut-il se produire quand δ_0 est exhaustive ?

3. Maximiser la vraisemblance.

Soit $\ell(\theta|x_1, \ldots, x_n)$ vraisemblance d'un échantillon i.i.d., fonction continue de θ telle qu'il existe p_0 avec
$$\int_\Theta \ell^p(\theta|x_1, \ldots, x_n)d\theta < \infty \tag{1}$$
pour $p \geq p_0$.

1. Montrer que, si $\ell^p(\theta|x_1, \ldots, x_n)$ admet un unique maximum global $\hat{\theta}$, il vient
$$\lim_{p\to\infty} \frac{\int f(\theta)\ell^p(\theta|x_1, \ldots, x_n)d\theta}{\int \ell^p(\theta|x_1, \ldots, x_n)d\theta} = f(\hat{\theta}).$$

Donner une interprétation bayésienne à ce résultat et en déduire une méthode pratique de maximisation de ℓ.

2. Soit $f(x_i|\theta) = h(x_i)\exp(\theta \cdot x_i - \psi(\theta))$. Montrer que (1) est vérifiée pour tout $p > 0$ et en déduire que, si $\pi(\theta)$ est la loi a priori “conjuguée”, $\pi(\theta|\lambda, x_0) \propto \exp(\theta \cdot x_0 - \lambda\psi(\theta))$, l'espérance a priori $\mathbb{E}^\pi[f(\theta)]$ tend vers $f(\hat{\theta})$ quand $x_0 = \sum x_i/n$ et λ tend vers ∞.
3. Appliquer aux cas où $P(x_i = 1) = 1 - P(x_i = 0) = \exp(\alpha^t z_i)/(1 + \exp(\alpha^t z_i))$ (régression *logistique*) et où $x_i \sim \mathcal{N}(\beta^t z_i, \sigma^2)$ sous la contrainte que $x_i \geq \underline{\alpha}$. On montrera en particulier qu'il existe un unique maximum global de la vraisemblance.

4. – Familles exponentielles abruptes.

On dit qu'une famille exponentielle naturelle,

$$f(x|\theta) = \exp\{\theta \cdot x - \psi(\theta)\} \quad (\nu)$$

est *abrupte* si, définissant

$$N = \{\theta; \int e^{\theta x} d\nu(x) < \infty\},$$

on a, pour tout $\theta^0 \in N^0$, $\theta_1 \in N - N^0$ et $\theta_\varrho = \varrho\theta_1 + (1-\varrho)\theta_0$,

$$\lim_{\varrho \to 1} \frac{\partial \psi}{\partial \varrho}(\theta_\varrho) = +\infty.$$

1. Montrer qu'une famille exponentielle est abrupte si, et seulement si, $\mathbb{E}[||\theta||] = +\infty$ pour $\theta \in N - N^0$.
2. Montrer que la famille *gaussienne inverse*, de densité

$$(1/\sqrt{\pi})z^{-3/2}\exp\{\theta_1 z - \theta_2(1/z) - \sqrt{2\theta_1\theta_2} - \log(2\theta_2)/2\}$$

sur $\mathbb{R}_+$, est abrupte sans être régulière.
3. Montrer qu'une famille abrupte peut se paramétriser en fonction de $\xi(\theta) = \nabla\psi(\theta)$, qui définit une transformation bijective de N^0 dans $\xi(N^0)$.
4. Montrer que, pour les familles exponentielles abruptes, l'estimateur du maximum de vraisemblance de θ, $\hat{\theta}$, satisfait $\nabla\psi(\hat{\theta}) = x$.

0. Modélisation

La table ci-dessous décrit un ensemble d'observations reliant des concentrations d'acide ascorbique pour des haricots à la température et au temps de stockage.

	Sem.			
Temp.	2	4	6	8
0	45	47	46	46
10	45	43	41	37
20	34	28	21	16

1. Etant donnée la nature de ces observations, positives et comprises entre 0 et 100, proposer des modèles de pseudo-vraisemblance convergents.
2. Adoptant une modélisation normale avec la contrainte
$$\mathbb{E}[Y|T,t] = \exp\{\alpha - \beta_T t\},$$
donner la vraisemblance et la loi limite de l'estimateur du maximum de vraisemblance. (On supposera la variance de Y constante.)
3. Comparer l'adéquation de la pseudo-vraisemblance normale à celles des pseudo-vraisemblances gamma et beta en termes de *déviance*, c'est-à-dire de différence entre la log-vraisemblance sans contrainte et la log-vraisemblance correspondant à la contrainte sur la moyenne.

1. Meilleurs estimateurs sans biais.

1. Si $f(x|\theta) = 1/\theta$ est la densité de $\mathcal{U}_{[0,\theta]}$, déduire que la borne de Cramer-Rao est θ^2/n pour un échantillon de taille n, $X_{(1)} \leq \ldots \leq X_{(n)}$. En considérant l'estimateur sans biais $Y = (n+1)X_{(n)}/n$, montrer que $\mathrm{Var}(Y) = \theta^2/n(n+2)$ et conclure.
2. Soient $x_1, \ldots, x_n$ des réalisations indépendantes de $\mathcal{N}(\mu, \sigma^2)$.
   a. Montrer que l'estimateur de μ, $\bar{x} = \sum_i x_i/n$, atteint la borne de Cramer-Rao.
   b. Montrer que $s^2 = \sum_i (x_i - \bar{x})^2/(n-1)$ est un estimateur sans biais de σ^2 mais n'atteint pas la borne de Cramer-Rao.
   c. Déduire de la complétude de s^2 qu'il est le meilleur estimateur sans biais de σ^2.
3. Soit $x_1, \ldots, x_n$ un échantillon de $\mathcal{P}(\lambda)$.
   a. Montrer que $x_1, \bar{x}$ et s^2 sont des estimateurs sans biais de λ.
   b. Montrer que $\bar{x}$ atteint la borne de Cramer-Rao.
   c. Déduire de $\mathrm{var}(s^2|\bar{x}) = \bar{x}$ que $\mathrm{var}(s^2) > \mathrm{var}(\bar{x})$.
4. Montrer qu'un estimateur sans biais δ est le meilleur estimateur sans biais si et seulement si il est non corrélé avec tout estimateur sans biais de 0.

5. Soit $x \sim \mathcal{U}_{[\theta,\theta+1]}$. Montrer que les estimateurs sans biais de 0 sont les fonctions périodiques de période 1 telles que

$$\int_0^1 h(x)dx = 0.$$

En déduire que $x - 0.5$ n'est pas le meilleur estimateur sans biais.
6. Si $x_1, \ldots, x_n$ sont $\mathcal{B}(k, \theta)$ et $\tau(\theta) = P_\theta(X = 1)$, trouver le meilleur estimateur sans biais de $\tau(\theta)$. (Utiliser $h(x_1, \ldots, x_n) = \mathbb{I}(x_1 = 1)$.)
7. Soit $x_1, \ldots, x_n$ i.i.d. $\mathcal{U}_{[\theta,2\theta]}$. Donner l'estimateur du maximum de vraisemblance et l'estimateur des moments. Lequel peut on améliorer par le théorème de Rao-Blackwell?

2. Echantillons censurés.

On observe des durées de vie $x_1, \ldots, x_n$ indépendantes, pour un même phénomène (composants, épidémie, chômage, etc.), que l'on modélise par des lois exponentielles, $\mathcal{E}xp(\lambda)$. Pour des raisons endogènes ou exogènes, on n'observe pas toutes les fins de fonctionnement: l'échantillon est alors *censuré*. On distingue trois formes de censure:

(i) Type I: Un temps d'observation, T, est fixé et on observe $x_1 \wedge T, \ldots, x_n \wedge T$.
(ii) Type II: Un nombre d'observations menées à terme, $N < n$, est donné et on observe seulement $x_{(1)}, \ldots, x_{(N)}$.
(iii) Type III: On associe aux x_i des temps aléatoires de censure $y_1, \ldots, y_n$ de loi indépendante de λ et on observe $x_1 \wedge y_1, \ldots, x_n \wedge y_n$.

1. Donner les fonctions de vraisemblance correspondantes aux trois types.
2. Determiner les situations où les trois modèles donnent des fonctions de vraisemblance proportionnelles et calculer l'estimateur du maximum de vraisemblance de λ.
3. Donner l'information de Fisher, $I(\lambda)$.
4. Dans ce contexte, on introduit des variables explicatives $y_i \in \mathbb{R}^p$ en les reliant à la distribution des x_i par $x_i \sim \mathcal{P}(\lambda_i)$ avec $\lambda_i = \exp(y_i^t\beta)$. Discuter des conditions d'identifiabilité de β et des équations de vraisemblance.

Un cas extrême de *troncation* est celui où seule l'appartenance à une classe est connue. On observe, à a et b fixés, $\mathbb{I}_{[a,b]}(y_i)$, $\mathbb{I}_{[0,a]}(y_i)$ et $\mathbb{I}_{[b,+\infty[}(y_i)$ $(1 \leq i \leq n)$.

5. Donner la vraisemblance du modèle en fonction de λ et discuter de l'identifiabilité de λ. Montrer que le modèle se présente comme un modèle *trinomial* contraint.
6. Soit $v_1 = \sum_i \mathbb{I}_{[0,a]}(y_i)$ le nombre d'observations inférieures à a. Montrer que

$$\sqrt{n}(v_1/n - p_1) \xrightarrow{\mathcal{L}} \mathcal{N}(0, p_1(1 - p_1)),$$

où $1 - p_1 = \exp(-\lambda a)$.
7. En déduire que

$$\tilde{v}_n = -\log(1 - v_1/n))/a$$

est un estimateur convergent de λ.

3. L'algorithme EM.

Dans le cadre d'un mélange de deux distributions normales

$$p\,\mathcal{N}_1(\mu_1, \sigma_1^2) + (1-p)\,\mathcal{N}_1(\mu_0, \sigma_0^2),$$

on observe un échantillon $y_1, \ldots, y_n$ afin d'estimer le paramètre $\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2, p)$.

1. On introduit $z_i \in \{0, 1\}$, variable indicatrice de la composante associée à y_i, c'est à dire que $y_i | z_i \sim \mathcal{N}(\mu_{z_i}, \sigma_{z_i}^2)$. Montrer que $z_i \sim \mathcal{B}(p)$ et donner la loi de (y_i, z_i). (Noter que z_i n'est pas observé.)
2. Soit $\mathrm{L}(\theta | y_i, z_i)$ la log-vraisemblance associée au couple (y_i, z_i), dite *vraisemblance complétée*. Donner $\sum_i \mathbb{E}_{\theta_0}[\mathrm{L}(\theta | y_i, z_i) | y_i]$, espérance de la vraisemblance complétée conditionnellement aux observations, et déterminer la valeur de θ, θ_1, qui maximise cette quantité.
3. Montrer que la *vraisemblance observée*, $\sum_i \mathrm{L}(\theta | y_i)$, vérifie

$$\sum_{i=1}^n \mathrm{L}(\theta_1 | y_i) \geq \sum_{i=1}^n \mathrm{L}(\theta_0 | y_i).$$

(Utiliser le fait que

$$\int \log \frac{f(x|\theta_0)}{f(x|\theta_1)} f(x|\theta_0) dx \geq 0.)$$

En déduire l'intérêt de cette méthode, dit *algorithme EM*, pour le calcul de maxima locaux de la vraisemblance.

1. – Tests asymptotiques.

On dispose de mesures de vitesses $x_i \sim \mathcal{N}(\mu, \sigma^2)$ $(1 \leq i \leq n)$ et on cherche à tester si l'erreur de mesure est proportionnelle à μ. Par une calibration appropriée, on se ramène au test de $H_0 : \sigma^2 = \mu^2$. (On rappellera la matrice d'information de Fisher pour ce modèle.)

(a) Calculer les emv contraint et non-contraint. Montrer que le meilleur estimateur pour l'erreur quadratique de μ de la forme $a\overline{x}_n$ est associé à $a^* = n/(n+1)$, où $\overline{x}_n$ désigne la moyenne empirique de l'échantillon.

(b) Construire les statistiques de Wald, du score et du rapport de vraisemblance pour ce problème.

(c) Comparer les performances des trois tests au test du χ^2 pour les valeurs suivantes : $n = 53$, $\overline{x}_n = 1.8$, $s_n^2 = 2.51$ et pour les classes suivantes :

A_i	$]-\infty, 0.58[$	$[0.58, 1.32[$	$[1.32, 1.99[$	$[1.99, 2.74[$	$[2.74, +\infty[$
n_i	10	9	11	13	9

On rappelle que $\Phi(-0.65) = 0.26$ et $\Phi(-0.20) = 0.42$. Par ailleurs, $\chi^2_{0.95}(3) = 7.8$ et $\chi^2_{.99}(3) = 11.25$.

2. – Le problème de Behrens-Fisher.

Soient $x_1, \ldots, x_n$ un échantillon de $\mathcal{N}(\theta, \sigma^2)$ et $y_1, \ldots, y_m$ un échantillon de $\mathcal{N}(\mu, \tau^2)$. On cherche à tester l'égalité $H_0 :\ \mu = \theta$ dans le cas général.

1. On suppose $\tau^2 = c\sigma^2$, avec c donné. Montrer qu'une statistique exhaustive de ce modèle est donnée par

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \bar{y} = \frac{1}{m}\sum_{j=1}^{m} y_j \quad \text{et} \quad S_c^2 = c\sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{j=1}^{m}(y_j - \bar{y})^2.$$

En déduire qu'on peut fonder le test sur

$$T = \frac{\bar{x} - \bar{y}}{(\frac{c}{m} + \frac{1}{n})^{1/2} S_c / \sqrt{m+n-2}} \sim \mathcal{T}_{m+n-2} \tag{1}$$

et que la région critique construite à partir de (1) dépend de c.

2. Si $m/(m+n)$ tend vers $0 < \varrho < 1$ quand m, n tendent vers $+\infty$, montrer que

$$t = \left(\frac{1}{m} + \frac{1}{n}\right)^{-1/2} (\bar{y} - \bar{x}) D^{-1} \approx \mathcal{N}\left(0, \frac{(1-\varrho)\sigma^2 + \varrho\tau^2}{\varrho\sigma^2 + (1-\varrho)\tau^2}\right),$$

si $D^2 = \{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{j=1}^m (y_j - \bar{y})^2\}/(m+n-2)$. En déduire que l'utilisation de la statistique T en remplaçant S_c^2 par S_1^2 et c par 1 n'est acceptable que si les deux échantillons sont en proportions similaires ($\varrho \simeq 1/2$).

3. – Robustesse du test de Student.

La *statistique de Student* sur laquelle s'appuie le test de $\mathbb{E}[X_i] = \theta_0$ est

$$T(x_1, \ldots, x_n) = \frac{\sqrt{n}(\bar{x} - \theta_0)}{s} \underset{H_0}{\sim} \mathcal{T}_{n-1},$$

où $s^2 = \sum (x_i - \bar{x})^2/(n-1)$.

1. On suppose les x_i indépendants de moyenne θ_0. Montrer que le test correspondant est *robuste* (ou *asymptotiquement convergent*) au sens où $T(x_1, \ldots, x_n)$ converge en loi vers la loi normale $\mathcal{N}(0,1)$.
2. Montrer qu'il n'en va pas de même pour le test de Fisher sur l'égalité des variances de deux échantillons $x_1, \ldots, x_n$ et $y_1, \ldots, y_m$,

$$F(x_1, \ldots, x_n, y_1, \ldots, y_m) = \frac{s_x^2/n-1}{s_y^2/m-1} \sim \mathcal{F}_{n-1,m-1}.$$

Pour montrer ce manque de robustesse, on supposera que $m/(m+n) = \varrho$ et on établira que

$$\sqrt{m+n}[\log s_x^2 - \log s_y^2] \underset{H_0}{\approx} \mathcal{N}(0, \gamma^2/\varrho(1-\varrho)\sigma^4),$$

avec $\gamma^2 = \text{var}(x_i^2) = \text{var}(y_i^2)$.

3. On suppose les x_i normaux $\mathcal{N}(\theta, \sigma^2)$ et *corrélés*, avec $cov(x_i, x_j) = \varrho_{ij}$. Sous les conditions ($n \to +\infty$)

$$\frac{1}{n^2} \sum_{i,j} \varrho_{ij} \to 0, \qquad \text{var}\left(\frac{1}{n} \sum_i x_i^2\right) \to 0, \qquad \frac{1}{n} \sum_{i \neq j} \varrho_{ij} \to \gamma^2,$$

montrer que le test de Student au niveau nominal α est au niveau réel

$$1 - \Phi(u_\alpha/\sqrt{1+\gamma^2}) > 1 - \Phi(u_\alpha) = \alpha.$$

4. Appliquer les résultats précédents aux modèles suivants pour discuter de la validité du test de Student:
   (i) *Aggrégation:* on observe les x_i par paquets de taille m et la corrélation entre les membres de ce paquet est ϱ;
   (ii) *AR(1):* les x_i sont reliés par $x_i = \theta + \varrho(x_{i-1} - \theta) + \epsilon_i$, avec $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

4. – Lois de Poisson et modèles linéaires généralisés.

1. On considère des observations y_{ij} de lois de Poisson $\mathcal{P}(\lambda_i)$ $(1 \leq i \leq I, 1 \leq j \leq J)$. Proposer un test asymptotique de l'hypothèse $\lambda_i = \lambda_1$ $(1 \leq i \leq I)$.
2. On teste l'effet différent de quatre vaccins contre la tuberculose sur 30 vaches pour chaque vaccin. Les réponses par site d'innoculation et par vaccin sont

vaccin/site	1	2	3	4
A	454	322	411	290
B	408	249	266	285
C	523	283	349	290
D	364	268	312	249

   Pouvez vous détecter une différence entre les vaccins? entre les sites d'innoculation?

3. On s'intéresse à l'influence sur y_{ij} d'une variable explicative x_{ij}, sous la forme $\lambda_i = \exp(\alpha_i + \beta_i x_{ij})$. Donner log-vraisemblance, information de Fisher et test du score sur l'influence de x_{ij}. Tester cette influence dans le cas où

i	1	2	3	4
$\sum_j x_{ij}(y_{ij} - \bar{y}_i)$	28.5	32.3	22.8	25.2
$\bar{y}_i$	11.2	9.4	13.7	11.9
$\sum_j (x_{ij} - \bar{x}_i)^2$	5.3	4.7	6.6	4.3

0. – Modélisation

On observe les temps de survie suivants, déjà traités en PC (1).

Temps de survie	Patients entrant dans l'intervalle	Décés dans l'intervalle
$0-5$	40	5
$5-10$	35	7
$10-15$	28	6
$15-20$	22	4
$20-25$	18	5
$25-30$	13	4
$30-35$	9	4
$35-40$	5	0
$40-45$	5	2
$45-50$	3	1
≥ 50	2	2

Comparer les modélisations exponentielle, Poisson et normale par un test du χ^2.

1. – Tests asymptotiques (suite).

1. Soit $y_1, \ldots, y_n$ un échantillon i.i.d. de $\mathcal{P}(\lambda)$. Etudier la forme des tests du score, de Wald et du rapport de vraisemblance pour $H_0 : \lambda = \lambda_0$ et discuter de la valeur de l'approximation asymptotique.
2. Soit $y_1, \ldots, y_n$ un échantillon de la loi multinomiale $\mathcal{M}_m(1; p_1, \ldots, p_m)$. On cherche à tester $H_0 : (p_1, \ldots, p_m) \in \mathcal{P} = \{(p_i) = (p_i(\gamma)), \gamma \in \Gamma\}$, où $\dim \Gamma = q$ et $(\partial p_i / \partial \gamma_j)$ est de rang q. On note $\theta = (p_1, \ldots, p_{m-1})$.
   a) Donner la matrice d'information de Fisher, $I(\theta)$ et en déduire la statistique de Hausman.
   b) Montrer qu'un test équivalent est le test du χ^2.
   c) Trouver les tests du score et du rapport de vraisemblance dans le cas particulier où $\mathcal{P} = \{(p_1^0, \ldots, p_m^0)\}$.
3. Soient $x_1, \ldots, x_n$ iid $\mathcal{P}(\lambda)$ et $y_1, \ldots, y_n$ iid $\mathcal{P}(\mu)$. Comparer les expressions du test de Wald pour $H_0 : \lambda - \mu = 0$ et $H_0' : \lambda/(\lambda + \mu) = 1/2$.

2. – Tests séquentiels.

On dispose d'une suite infinie d'observations potentielles $x_1, x_2, \ldots$

1. On cherche à tester si $H_0 : x_i \sim f_0$ ou si $H_1 : x_i \sim f_1$. Le *test séquentiel de Wald* est fondé sur la statistique

$$R_n = \frac{f_1(x_1) \ldots f_1(x_n)}{f_0(x_1) \ldots f_0(x_n)}$$

et accepte H_0 si $R_n \leq B$, rejette H_0 si $R_n \geq A$ et continue les observations sinon. On note α et β les erreurs de première et seconde espèces liées à cette procédure.

a) Montrer que, si l'échantillonnage s'arrête avec probabilité 1, on a $A \leq (1-\beta)/\alpha$ et $B \geq \beta/(1-\alpha)$.

b) On modifie le test en prenant $A' = (1-\beta)/\alpha$ et $B' = \beta/(1-\alpha)$. Montrer que les nouvelles erreurs vérifient

$$\alpha' \leq \frac{\alpha}{1-\beta}, \qquad \beta' \leq \frac{\beta}{1-\alpha}.$$

c) Soit $z(x) = \log f_1(x)/f_0(x)$. Montrer que le test s'arrête avec probabilité 1 dès que z n'est pas identiquement nul.

d) On définit N comme la taille de l'échantillon conduisant à la décision et on note $S_N = \log R_N$. Montrer que

$$\mathbb{E}_{f_0}[N] \simeq \frac{b(1-\alpha)+a\alpha}{\mathbb{E}_{f_0}[z(X)]}, \quad \mathbb{E}_{f_1}[N] \simeq \frac{b\beta+a(1-\beta)}{\mathbb{E}_{f_1}[z(X)]},$$

où $a = \log A$, $b = \log B$.

2. Appliquer le test séquentiel de Wald au cas où f_0 est $\mathcal{N}(\theta_0, 1)$ et f_1 est $\mathcal{N}(\theta_1, 1)$. En déduire une approximation de la taille moyenne, $\mathbb{E}_{f_0}[N]$, et comparer avec un test à taille fixe d'erreurs α et β. (On montrera que la taille fixe s'écrit

$$N_{\alpha,\beta} = \frac{(d_\alpha + d_\beta)^2}{(\theta_0-\theta_1)^2}.\Big)$$

3. – Régions de confiance.

1. Soit $x \sim \exp\{\theta \cdot U(x) + \eta \cdot T(x) - \psi(\theta,\eta)\}$, famille exponentielle avec paramètre de nuisance $\eta \in \mathbb{R}^q$ et paramètre d'intérêt $\theta \in \mathbb{R}^p$.

a) Montrer que T est exhaustive pour le paramètre η (à θ fixé) et en déduire que la loi de U conditionnellement à T ne dépend pas de η.

b) Montrer qu'une région de confiance sur θ est de la forme

$$\{\theta;\ \theta \cdot U(x) - \psi_t(\theta) \geq c_t\}$$

quand $T(x) = t$.

c) Appliquer au cas où $x \sim \mathcal{P}(\lambda)$, $y \sim \mathcal{P}(\mu)$ et où $\theta = \ln(\lambda/\mu)$.

2. Soient $x_1, \ldots, x_n$ des observations i.i.d. de $\mathcal{N}(0, \sigma^2)$. Déterminer les intervalles de confiance au niveau $1-\alpha$, $[as^2, bs^2]$, sous les contraintes:

(i) $b - a$ est minimale;

(ii) $P(\sigma^2 < as^2) = P(\sigma^2 > bs^2)$;

(i) $[as^2, bs^2]$ est la région *uniformément plus précise*, i.e. la région duale de la région critique du test UPP sans biais.

3. (Région de Fieller) Soient $(x_1, y_1), \ldots, (x_n, y_n)$ des observations i.i.d. normales de moyenne (μ_1, μ_2), de variances σ_1^2, σ_2^2 et de covariance $\varrho\sigma_1\sigma_2$. On recherche une région de confiance sur le rapport $\theta = \mu_2/\mu_1$.
   a) Donner moyenne et variance de $\bar{z}_\theta = \bar{y} - \theta\bar{x}$ et montrer que cette variance peut s'estimer par
   $$\hat{\tau}^2 = \frac{1}{n}(s_2^2 - 2\theta s_{12} + \theta^2 s_1^2),$$
   où s_1^2, s_2^2 et s_{12} sont les estimateurs usuels de σ_1^2, σ_2^2 et $\varrho\sigma_1\sigma_2$.
   b) Donner la loi de $\hat{\tau}^2$ et déduire que
   $$\{\theta;\ \bar{z}_\theta^2/\hat{\tau}^2 \leq t_{n-1,\alpha/2}^2\}$$
   définit une région de confiance pour θ au niveau $1 - \alpha$.
   c) Exprimer cette région et déduire qu'elle peut s'écrire comme un intervalle, le complémentaire d'un intervalle ou même la droite réelle toute entière.

4. – Miscellanées.

1. On considère $y = X\beta + \epsilon$, avec $y \in \mathbb{R}^p$, $\beta \in \mathbb{R}^q$ et $\epsilon \sim \mathcal{N}(0, \sigma^2 I_p)$, où σ est connu. Montrer que les tests du score, de Wald et du rapport de vraisemblance pour H_0 : $R\beta = 0$, avec R matrice $r \times q$, sont identiques. Etendre au cas où σ est inconnu.

2. Soient $(x_1, y_1), \ldots, (x_n, y_n)$ des observations i.i.d. de loi (p_{ij}) $(1 \leq i \leq I, 1 \leq j \leq J)$. Exprimer les tests du score, de Wald et du rapport de vraisemblance pour l'hypothèse d'indépendance entre x et y et retrouver le test du χ^2.

3. On suppose que les x_i suivent une loi normale $\mathcal{N}(\theta, 1)$. On décide de recueillir des observations jusqu'à ce que $|\bar{x}| > 1.96\sqrt{n}$. Montrer que l'échantillon ainsi obtenu est p.s. fini et qu'il ne peut pas être utilisé comme un échantillon classique.

4. En contrôle de la qualité, on cherche des tests qui détectent très rapidement les défauts mais qui n'arrêtent pas la production en l'absence de défauts. On observe $x_1, \ldots$ variables $\mathcal{B}(p)$ et $H_0 : p < 0.01$. On rejette H_0 si $R_n > 0.01n + a\sqrt{n}$.
   a) Montrer que, pour tout a, la procédure s'arrête p.s. sous H_1.
   b) Montrer que a peut être choisi de manière à ce que la procédure s'arrête avec probabilité α sous H_0 si $p = p_0 < 0.01$.

5. Soit $y = X\beta + \epsilon$. On note $\hat{\beta}$ l'estimateur des moindres carrés de β. Montrer que la région de confiance
   $$\{\beta; ||X\beta - X\hat{\beta}||^2 \leq f_\alpha ||y - X\hat{\beta}||^2\}$$
   est au même niveau α quelque soit la loi de $\epsilon \sim g(||\epsilon||)$.

6. (Intervalles simultanés de Scheffé) Soient $x_1, \dots, x_n$ des observations indépendantes de $\mathcal{N}(\xi_i, \sigma^2)$. On cherche des régions de confiance simultanées pour toutes les combinaisons linéaires $\sum_i u_i \xi_i$, de la forme $\{L(u,x) \leq \sum_i u_i \xi_i \leq M(u,x)\}$, qui vérifient

$$P\left(L(u,x) \leq \sum_i u_i \xi_i \leq M(u,x), \forall u\right) = \alpha.$$

   a) Montrer que l'encadrement de la *droite de régression* se ramène à ce modèle.
   b) Justifier les contraintes de (i) *symétrie*, (ii) d'*orthogonalité* et (iii) d'*homogénéité*,

   $$(i)\ L(u,x) = -M(-u,x), \quad (ii)\ L(u,x) = L(Q^{-1}u, x), \quad (iii)\ L(cu,x) = cL(u,x)$$

   où Q varie dans l'ensemble des matrices othogonales, et déduire que

   $$L(u,x) = ||u|| h(u^t x, x^t x).$$

   c) Montrer que, si L doit de plus satisfaire, $L(u, x+x_0) = L(u,x) + \sum_i u_i x_{0i}$, on a

   $$L(u,x) = c||u|| + u^t x$$

   et déduire que la région de confiance associée est

   $$\{\xi;\ \left|\sum u_i(x_i - \xi_i)\right| < c||u||, \forall u\},$$

   où c est le α-quantile de χ^2_n.
   d) Montrer que l'ensemble ci-dessus est en fait la boule

   $$\{\xi;\ \sum (x_i - \xi_i)^2 < c^2\}.$$

I. – Analyse de la variance.

On observe des $y_{ij} \sim \mathcal{N}(\theta_i, \sigma^2)$ en nombres respectifs n_i. On cherche à tester $\theta_1 = \ldots = \theta_k$.

1. Montrer que cette hypothèse est équivalente à

$$\sum_j a_j \theta_j = 0 \quad \text{pour tout } a \text{ tel que} \sum_j a_j = 0.$$

On supposera dans la suite de l'exercice que $\sum_j a_j = 0$.

2. Donner les tests de Wald, du score et du rapport de vraisemblance de l'hypothèse nulle

$$H_0 : \sum_j a_j \theta_j = 0$$

à a fixé. Sont-ils identiques au test standard fondé sur

$$T_a = \frac{\sum_j a_j \bar{y}_{i\cdot}}{\sqrt{\frac{1}{n-k} \sum_{ij} (y_{ij} - \bar{y}_{i\cdot})^2 \sum_i a_i^2 / n_i}} \ ?$$

3. Justifier le test de $H_0 : \theta_1 = \ldots = \theta_k$ au moyen de la statistique $Z = \sup_a T_a$.

4. Montrer que

$$\sup_a \frac{(\sum_i a_i v_i)^2}{\sum_i a_i^2 / c_i} = \sum_i c_i (v_i - \bar{v}_c)^2$$

où $\bar{v}_c = \sum_i c_i v_i / \sum_i c_i$ et en déduire que Z correspond à

$$Z = \frac{\sum_i n_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2}{\sum_{ij} (y_{ij} - y_{i\cdot})^2} = \frac{\sum_{ij} (y_{ij} - \bar{y}_{\cdot\cdot})^2 - \sum_{ij} (y_{ij} - y_{i\cdot})^2}{\sum_{ij} (y_{ij} - y_{i\cdot})^2}.$$

Donner la loi de Z et déduire un test exact de H_0.

4. On examine l'effet d'une toxine sur des truites en mesurant le nombre de toxines ingérées par les truites :

Toxine	1	2	3	4
	28	33	18	11
	23	36	21	14
	14	34	20	11
	27	29	22	16
		31	24	
		34		

Tester la différence des toxines par le test ci-dessus.

II. — Capture-recapture (suite).

On considère n animaux se déplaçant de manière indépendante sur un territoire subdivisé en 3 zones géographiques A,B,C, tels que, entre deux instants d'observation, les déplacements de chaque animal sont régis par la matrice de transition

$$P = \begin{pmatrix} p_{AA} & p_{AB} & 0 \\ 0 & p_{BB} & p_{BC} \\ p_{CA} & 0 & p_{CC} \end{pmatrix}$$

avec $p_{AB} = 1 - p_{AA}$, $p_{BC} = 1 - p_{BB}$ et $p_{CA} = 1 - p_{CC}$. On fixe à n_A^1, n_B^1 et n_C^1 $(n = n_A^1 + n_B^1 + n_C^1)$ le nombre d'animaux observés dans chaque zone au premier instant $(t = 1)$ et on note n_{AB}^t, n_{BC}^t et n_{CA}^t le nombre d'animaux ayant migré de A en B, B en C et C en A, respectivement, entre $t-1$ et t $\ (t \geq 2)$ et n_{AA}^t, n_{BB}^t et n_{CC}^t le nombre d'animaux demeurant en A, B et C entre $t-1$ et t.

(a) Ecrire les relations entre $(n_{AA}^t, \cdots, n_{CA}^t)$ et $(n_{AA}^{t+1}, \cdots, n_{CA}^{t+1})$ traduisant la conservation du nombre total d'animaux. On notera n_A^t, n_B^t et n_C^t le nombre d'animaux en A, B et C à l'instant t. Donner la vraisemblance du modèle.

(b) Donner les emv des probabilités de migration pour T périodes d'observation si les observations correspondantes sont $n_{AB}^2, \cdots, n_{CA}^T$.

(c) On cherche à tester si les probabilités de migration p_{AB}, p_{BC} et p_{CA} sont identiques. Exprimer l'hypothèse nulle en fonction des probabilités de migration et donner l'emv de P sous cette contrainte. En déduire la forme du test du rapport de vraisemblance.

(d) Si $P^{(t)}$ désigne la t-ième puissance de P et

$$Q = \operatorname{diag}\left(\frac{1}{p_{AB}(1-p_{AB})}, \frac{1}{p_{BC}(1-p_{BC})}, \frac{1}{p_{CA}(1-p_{CA})}\right),$$

$e' = (n_1, n_2, n_3)$, montrer que

$$I(p_{AB}, p_{BC}, p_{CA}) = \sum_{t=1}^{T} Q \operatorname{diag}\left(P^{(t)} e\right)$$

et en déduire que les tests de Wald et du score ne sont pas utilisables dans ce cas, à une exception près portant sur une valeur particulière de e. On détaillera, si possible, ce cas particulier. (*Indication :* Déterminer le vecteur propre de somme 1 associé à la valeur propre 1 pour P.)

(e) Par un développement limité approprié, comparer le test du rapport de vraisemblance avec le test du χ^2 pour ce problème.