Corrigé de l'examen janvier 2006
Exercice 1

Partie A

1) $L_p(x_1, ..., x_n) = (1-p)^{\sum_{i=1}^n x_i}\ p^n \prod_{i=1}^{n} \mathbf{1}_{\ x_i \in N}$

2) L_p est dérivable par rapport à p.
Si tous les $x_i \in N$, on peut écrire

$$ln(L_p(x_1, ..., x_n)) = \sum_{i=1}^{n} x_i(ln(1-p)) + n\ ln(p)$$

$$\frac{\delta ln(L_p)}{\delta p} = -\frac{\sum_{i=1}^n x_i}{1-p} + \frac{n}{p}$$

.

Cherchons $\hat{p}$ tel que $-\frac{\sum_{i=1}^n x_i}{1-\hat{p}} + \frac{n}{\hat{p}} = 0.$

D'où $\hat{p} = \frac{n}{\sum_{i=1}^n x_i + n}$ est un extremum. Vérifions que c'est un maximum.

$$\frac{\delta^2 ln(L_p)}{\delta^2 p} = -\frac{\sum_{i=1}^n x_i}{(1-p)^2} - \frac{n}{p^2}$$

qui est négative en $\hat{p}$.

On a donc montré que $\hat{P} = \frac{1}{1+\bar{X}_n}$ est l'emv de p.

3) $\hat{p} = 0,08$.

Partie B
Soit $T = 1 + \bar{X}_n$
1) $E(T_n) = 1 + E(X) = p$ donc T_n est un esb de p. Comme T_n est un esb de p, le risque quadratique est donné par $R_p = Var(T_n) = Var(X)/n$ qui converge vers 0 quand n tend vers l'infini, donc T_n est un estimateur convergent de p.

2) On peut appliquer le théorème central limite à l'échantillon $(X_1, ..., X_n)$. Alors $Z = \frac{\bar{X}_n - E(\bar{X}_n)}{var(\bar{X}_n)} = \sqrt{n}\ \frac{T_n - 1/p}{\sqrt{\frac{1-p}{p^2}}}$ suit approximativement une loi $N(0,1)$. Pour construire

l'intervalle de confiance pour $\frac{1}{p}$, on utilise un bon estimateur de $\frac{1}{p}$. Soit donc T_n un estimateur convergent de $\frac{1}{p}$. Cherchons maintenant une fonction pivotale fonction de T_n.Soit $Z = \sqrt{np}\,\frac{T_n - 1/p}{\sqrt{1-p}}$ suit approximativement une loi $N(0,1)$.

Posons $0.95 = P(-t < Z < t)$ oú le quantile $t = 1,96$ est lu dans la table de la loi normale.

D'où $t_n \; - \; 1,96\frac{\sqrt{1-p}}{p\sqrt{n}} \; < \; \frac{1}{p} \; < \; t_n \; + \; 1,96\frac{\sqrt{1-p}}{p\sqrt{n}}$.

3) $\hat{\frac{1}{p}} = t_n = 1 + \frac{608}{50} = 13,16$ et $\hat{p} = \frac{1}{13,16} = 0,076$.

Comme $n > 30$, on peut appliquer le résultat précédent de l'intervalle de confiance en remplaçant dans les bornes p par son estimateur $\hat{p}$.
D'où $9.654 \; < \; \frac{1}{p} \; < \; 16,667$ et $0,06 \; < p \; < \; 0,1036$.

Exercice 2 Test d'homogénéité

On considère deux populations P_1 =société d'assurance A et P_2 =société d'assurance B dont on extrait respectivement un échantillon de taille 200 sur lesquels on étudie un caractère statistique X =type de contrat. On peut représenter ces données dans un tableau de contingence:

X/P	A	B	$n_{i.}$
1	108	91	199
2	64	74	138
3	22	29	51
4	6	6	12
$n_{.j}$	200	200	400

On veut tester au niveau :

H_0 : les deux populations sont homogènes pour le caractère X contre
H_1 : les deux populations ne sont pas homogènes pour le caractère X

On construit le tableau des effectifs théoriques sous H_0:

X/P	A	B	$n_{i.}$
1	99,5	99.5	199
2	69	69	138
3	25,5	25,5	51
4	6	6	12
$n_{.j}$	200	200	400

Sous H_0, comme n_1 et $n_2 \geq 30$ et que les effectifs théoriques sont supérieurs ou égaux à 5.
$T_n = \sum_{i=1}^{k}\sum_{j=1}^{l} \frac{(N_{ij} - \frac{n_{i.}n_{.j}}{n})^2}{\frac{n_{i.}n_{.j}}{n}}$suit approximativement une loi $\chi^2(3)$.

La région de rejet, au niveau α de la forme

$$W = (T_n \geq 7,815)$$

On calcule la valeur de T_n sur l'échantillon et on obtient $t_n = 3,136$. Comme $t_n < 17,815$ n'est pas dans la région de rejet, on ne rejette pas H_0 au niveau 0.05, on accepte l'hypothèse que le comportement est le même pour les clients A ou B.

Exercice 3 Test sur l'espérance d'une loi normale $N(m, \sigma^2)$, σ inconnu
On veut tester les hypothèses suivantes, pour un niveau α donné:

$$\begin{aligned} H_0 :\ m &= 100 \\ \text{contre} & \\ H_1 : m &> 100 \end{aligned}$$

Sous H_O, la v.a $Z = \sqrt{n}\frac{\bar{X}_n - m}{\hat{S}_n}$ suit une loi de Student $St(14)$.

La région de rejet est donnée par $W = (\bar{X}_n > K)$

$$0,1 = P_{m_0}(\bar{X}_n > K)$$

D'où, $K = 100 + 1,365\frac{10,351}{\sqrt{15}}$ avec $1,365$ quantile d'ordre $0,9$ de la loi $N(0,1)$ et $\hat{s} = \sqrt{\frac{15}{14}}s = 10,351$.

La région de rejet du test est donc:
$W = (\bar{X}_n > 103,6481)$
$\bar{x}_n = 105 \in W$, on rejette H_0 au risque $0,1$, le prix moyen est en dehors de la norme.

Exercice 4 Comparaison de deux espérances pour des échantillons indépendants de grande taille et de variances inconnues

On veut tester au niveau $0,05$:

$$\begin{aligned} H_0: \ m_X &= m_Y \\ \text{contre} & \\ H_1: \ m_X &< m_Y \end{aligned}$$

Sous H_0, comme $n_X, n_Y \geq 30$, par le théorème de la limite centrale, la v.a. $\dfrac{\bar{X}_n - \bar{Y}_n}{\sqrt{\frac{\hat{S}_X^2}{n_X} + \frac{\hat{S}_Y^2}{n_Y}}}$ suit approximativement une loi $N(0,1)$.
On en déduit la région de rejet $W = (\bar{X}_n - \bar{Y}_n < K)$.

$$0,05 \ = \ P_{H_0}(\bar{X}_n - \bar{Y}_n < K)$$

D'où, $K \ = -1,645\sqrt{\frac{\hat{s}_X^2}{n_X} + \frac{\hat{s}_Y^2}{n_Y}}$ avec $-1,645$ quantile d'ordre $0,05$ de la loi $N(0,1)$ et $\hat{s}_X = 1,589, \hat{s}_X = 1,601$.

La région de rejet du test est donc:
$W \ = \ (\bar{X}_n - \bar{Y}_n < -0.481)$
$\bar{x}_n - \bar{y}_n = -2,8 \in W$, on rejette H_0 au risque $0,05$,les filles affichent en moyenne un plus haut niveau d'attention que les gar cons.