SERIE DE TD N° 1

ENONCES

EXERCICE 1

On considère la variable aléatoire X_n de loi de probabilité uniforme sur $\left\{0, \frac{1}{n}, ..., \frac{n-1}{n}, 1\right\}$.

Montrer que la suite de variables aléatoires $(X_n,\ n < 1)$ converge en loi ver la variable aléatoire X de loi uniforme sur le segment $[0,\ 1]$.

EXERCICE 2

Soit une suite $(X_n,\ n < 1)$ de variables aléatoires réélles définies sur un espace probabilisé $(\Omega,\ \Lambda,\ P)$, la loi de X_n étant donnée par $P\left(X_n = 1 - \frac{1}{n}\right) = \frac{1}{2} = P\left(X_n = 1 + \frac{1}{n}\right)$.

1 Montrer que la suite (X_n) converge en loi vers la variable aléatoire $X = 1$.

2 Est-ce-que pour tout x, $\lim_{n\to\infty} P(X_n = x) = P(X = x)$?.

3- Montrer que la suite (X_n) converge en probabilité vers 1.

4- (X_n) converge-t-elle en moyenne quadratique vers 1?.

5- (X_n) converge-t-elle presque sûrement vers 1?.

EXERCICE 3

Soit une suite $(X_n,\ n < 1)$ de variables aléatoires réélles mutuellement indépendantes de même loi uniforme sur $[0,\ a]$, $a < 0$.

1- Soit $S_n = \frac{X_1 + X_2 + ... + X_n}{n}$. Etudier la suité (S_n) suivant différents modes de convergence.

2- Etudier la limite de la suité $\left(\sqrt{n}\left(S_n - \frac{a}{2}\right)\right)$.

3- Montrer que la suite $(M_n = \sup(X_1, ,X_2, ...,\ X_n))$ converge en loi, converge-t-elle enprobabilité ?.

4- Calculer la distance de Kolmogorove entre la fonction de répartition de la loi de M_n et celle de la variable aléatoire constante égale à a et déterminer sa limite quand $n \to \infty$.

EXERCICE 4 Lemme de Borel-Cantelli

Soit $(\Omega,\ \Lambda,\ P)$ un espace probabilisé et (A_n) une suite d'évènement de Λ.On définit l'évènement B " pour une infinité de n, A_n est réalisé".

1- On pose $B_n = \cup_{n\langle\rangle m} A_n$; montrer que $B = \lim_{n\to\infty} B_n$.

2- Montrer que si la série de terme général $P(A_n)$ est convergente, alors $P(B) = 0$.

Exercice 5

Pour tout entier naturel n non nul, on considère la fonction f_n définie par

$f_n(x) = n^2 x \exp(\frac{-n^2x^2}{2}) 1_{\mathbb{R}_+}(x)$.

Montrer que fnfn est la densité d'une variable aléatoire.

Soit $(X_n)_n$ une suite de variables aléatoires telle que, pour tout entier $n \geq 1$, X_n admet pour densité f_n. Démontrer que la suite (X_n) converge en probabilité vers une variable aléatoire X que l'on précisera.

Exercice 6

Soit (Un) une suite de variables aléatoires indépendantes suivant toutes la loi uniforme sur [0,1]. On note

$M_n = \max(U_1, \ldots, U_n)$ *et* $X_n = n(1 - M_n)$.

Quelle est la fonction de répartition de X_n ?

Etudier la convergence en loi de la suite (X_n).

Indication : $\lim_{n\to\infty} \left(1 - \dfrac{x}{n}\right) = \exp(x)$.

Exercice 7

On dit qu'une variable aléatoire Y suit une loi de Gumbel si elle admet pour densité $f(x) = e^{-x-e^{-x}}$.

Vérifier que f est une densité, et calculer la fonction de répartition de Y.

Soit $(X_n)_n$ une suite de variables aléatoires indépendantes identiquement distribuées de loi exponentielle de paramètre 1. On pose $M_n = max(X_1, \ldots, X_n)$. Démontrer que la suite $(M_n - \ln(n))$ converge en loi vers Y suivant une loi de Gumbel.

Indication : $\int\limits_{-\infty}^{t} e^{-x-e^{-x}} dx = \left[e^{-e^{-x}}\right]_{-\infty}^{t}$.

SERIE DE TD N° 2

ENONCES

Exercice 1 : Familles Exponentielles

On considère les modè les suivants :

Modèle Binomial $\{B(m,p) : p \in [0,1]\}$;

Modèle de Poisson $\{P(\lambda) : \lambda > 0\}$;

Modèle gaussien à variance fixée $\{N(\mu,\sigma^2) : \mu \in \mathbb{R}\}$;

Modèle gaussien à paramètre bi-dimensionnel $\{N(n\mu,\sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$;

Modèle Gamma

$$\{G(\alpha,\beta) : \alpha > 0,\ \beta > 0\} = \{f_{\alpha,\beta}(x) = \frac{\beta}{\alpha}\Gamma(\alpha)x^{\alpha-1}e^{-\beta x}1_{R+}(x) : \alpha > 0,\ \beta > 0\};$$

Modèle uniforme $\{U_{[0,\theta]} : \theta > 0\}$;

Modèle de Cauchy $\{f_\theta(x) = \dfrac{1}{\Pi(1+(x-\theta)^2)} : \theta \in R\}$; • Modèle multinomial $\{M(n,p_1,...,p_k) : 0 < p_i < 1, \forall i = 1,...,k\ et\ \sum_{i=1}^{k} p_i = 1\}$. Pour tous ces modèles, répondre aux questions suivantes.

1) Quelle est l'expression de la densité $f_{theta}(x)$?

2) Le modèle constitue-t-il une famille exponentielle générale ? Naturelle ? Quel est le paramètre canonique du modèle ?

3) Quelle est la vraisemblance d'un échantillon$x = (x_1,...,x_n)$?

Exercice 2 : (Modèles position-échelle)

1) Construire un modèle position-échelle à partir de la loi exponentielle exp(1). Préciser la forme des $f.d.r.$ des lois de ce modèle ainsi que leurs densités.

2) Montrer que le modèle uniforme $\{U_{[a,b]} : -\infty < a < b < +\infty\}$ est un modèle position-échelle.

Exercice 3 (Statistiques d'ordre)

Soit $X_1,...,X_n$ des v.a.r. définies sur un même espace probabilisé (Ω, A, P), indépendantes et demême loi absolument continueparrapport à la mesure de Lebesgue de densité f. Pour tout ω dans Ω, on peut ordonner les réels $X_1(\omega), ..., X_i(\omega), ..., X_n(\omega)$ sous la forme $X_{(1)}(\omega) \leq X_{(2)}(\omega) \leq \cdots \leq X_{(i)}(\omega) \leq \cdots \leq X_{(n)}(\omega)$.

L'application $X_{(i)} : \omega \in \Omega \to X_{(i)}(\omega)$ ainsi définie pour chaque i est une v.a.r. dite ième statistique d'ordre.

1) Calculer la loi de $X_{(n)} = sup\{X_1, ..., X_n\}$ ($f.d.r.$ et densité).

2) Calculer la loi de $X_{(1)} = inf\{X1, ..., X_n\}$ ($f.d.r.$ et densité).

3) Calculer la loi du couple $(X_{(1)}, X_{(n)})$.

4) Soit N_y le nombre de X_i inférieurs à y. Quelle est la loi de N_y ? Que dire des événements $\{N_y \geq k\}$ et $\{X_{(k)} \leq y\}$? En déduire la $f.d.r.$ de $X_{(k)}$.

Solutions

Modèle statistique de la loi Binomiale $(B(m;p) : p \in [0;1])$ La densité, pour tout x dans N, est $f_p(x) = C_m^x p^x(1-p)^{m-x} = exp[xln(\dfrac{p}{1-p})](1-p)^m C_m^x$ En posant C(p)=(1-p)^m, $h(x) = C_m^x, T(x) = xet\eta(p) = \dfrac{p}{1-p})$

on constate que le modèle de la loi Binomiale est une famille exponentielle naturelle dont le paramètre canonique est $\theta = ln(\dfrac{p}{1-p}).Lavraisemblancedel'chantillon$ x$_1$; ...x_n est : $L(x_1; ...; x_n); p) = p^{\sum i=1^m}(1-p)^{nm-\sum i=1^n \Pi_{i=1}^m C_m^{x_i}}$

Modèle Statistique de la loi de Poisson, la densité en tout piont x de $\mathbb{N}, estf_P(\lambda) : \frac{e^{-\lambda}\lambda^x}{x!}$

En posant $C(\lambda) = e^{-\lambda}; h(x) = \dfrac{1}{x!}; T(x) = xet\eta(\lambda) = ln(\lambda)$,

on vérifie que ce modèle est une famille exponentielle naturelle de paramètre canonique $\theta = ln(\lambda)$. La vraisemblance de l'échantillon x$_1$; ; x_nest : L(x$_1$; ...; x_n; λ) $= \frac{e^{-n\lambda}\lambda^{\sum i=1^n}}{\Pi_{i=1}^m x_i!}$ Modèle Statistique de la loi normale à deux paramètres (N$(\mu;\sigma^2) : \mu \in \mathbb{R}; \sigma^2 > 0$)

Dans ce modèle la densité est : $f_{(\mu;\sigma^2)} = \frac{1}{\sigma(\Pi)^{\frac{1}{2}}} exp[-\dfrac{\mu^2}{2\sigma^2}]$, où $\eta(\mu, \sigma^2) = (\frac{\mu}{2\sigma^2}) et T(x) = (x, -x^2)$

En posant $C(\mu; \sigma^2) = \sigma(\Pi)_{\frac{1}{2}} exp[-\dfrac{\mu^2}{2\sigma^2}] et h(x) = 1;$

on constate que ce modèle est une famille exponentielle générale de paramètre canonique $\theta = (\mu/\sigma^2, 1/2\sigma^2)$

Modèle statistique de la loi Gamma $f_{(\alpha;\beta)}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} : avcx > 0, \alpha\beta > 0$

En posant
$C(\alpha; \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}$; $h(x) = I_{\mathbb{R}^+}(x)$; $\eta(\alpha; \beta) = (\alpha - 1; \beta)$ et $T(x) = (lnx; -x)$;
ce modèle s'écrit sous la forme d'une famille exponentielle générale. Modèle statistique de la loi uniforme $(U_{[0,\theta]} : \theta > 0)$
La densité est $\mathrm{f}_\theta(x) = \frac{1}{\theta} I_{[0,\theta]}(x)$

et on constate que l'on ne peut pas l'écrire sous la forme d'une famille exponentielle.

Modèle statistique de la loi de Cauchy $\mathrm{f}_\theta(x) = \dfrac{1}{\Pi(1 + (x - \theta)^2)} : \theta \in R$

que l'on ne peut pas écrire sous la forme exponentielle, ains il ne s'agit pas d'une famille exponentielle.

SERIE DE TD N° 3

ENONCES

Exercice 1 (Statistiques exhaustives)
On considère les modèles suivants :
modèle de Poisson (N ; P(N) ; P(λ) :λ > 0) ;
modèle de la loi de exponentielle $(\mathbb{R}_+; B_{\mathbb{R}_+}, e(\lambda)$: λ >0);

modèle gaussien avec σ^2 positif connu : $(\mathbb{R}; B()\mathbb{R}); \mathbb{N}(\mu; \sigma^2 : \sigma^2$ >0);

modèle gaussien avec μ dans $\mathbb{R}$ connu : (R ; B(R) N($\mu; \sigma^2 : \sigma^2$ >0) :

modèle gaussien général : (R ; B(R) N($\mu; \sigma^2 : \mu \in \mathbb{R}, \sigma^2$ >0) :

1) Pour chacun de ces modèles donner l'expression d'une statistique exhaustive (éventuellement vectorielle).
2) Retrouver le résultat pour le modèle de Poisson en utilisant une autre méthode.

Exercice 2 : (Statistique exhaustive et Famille Exponentielle Générale) On considère une famille exponentielle générale de statistique canonique $T(X)$ où X est la variable générique dans ce modèle.

1) Montrer que $\sum_{i=1}^n X_i$ = T(x$_i$) est une statistique exhaustive pour le modèle d'échantillonnage associé.

2) En utilisant un résultat obtenu dans l'Exercice 1 série de TD N° 2, montrer que la moyenne empirique $\frac{1}{n}\sum_{i=1}^{n} X_i estunestatistiqueexhaustivedansunmodled'chantillonnagedelaloibinomiale.$

Exercice 3 (Modèle Gamma et Méthode des moments)

On considère le Modèle Statistique de la loi Gamma :

(R^+; $B(R^+)$; $G(\alpha,\beta)$: α >0 , β >0) :

On rappelle que la densité d'une v.a. X de loi $G(\alpha;\beta)$ est : $f_{(\alpha;\beta)}(x) = \dfrac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} 1_{\mathbb{R}^+}$.

1) Calculer $E_{(\alpha;\beta)}(X)$ et $\mathrm{var}_{(\alpha;\beta)}(X)$

2) Par la méthode des moments, donner un estimateur du paramètre bidimensionnel $E_{(\alpha;\beta)}(X)$ du modèle, basé sur l'observation d'un échnatillon $X_1; ...; X_n$.

3) Déterminer des estimateurs de α et β en utilisant conjointement des estimateurs empiriques des moments et la méthode de substitution.

Exercice 4 (Modèle de la loi exponentielle et Méthode des moments)
On a vu que la méthode des moments permet d'obtenir un estimateur du paramètre λ dans un modèle de la loi exponentielle :

$\lambda = 1/(\dfrac{1}{n}\sum_{i=1}^n X_i)$ basé sur la relation E(X) = $\frac{1}{\lambda}$. L'intérêt de cet exercice est de montrer que cette méthode permet la construction de plusieurs estimateurs de ce même paramètre λ. 1) On suppose qu'une v.a.r. X suit une loi exponentielle $exp(\lambda)$. Calculer $E(X^2$

2) Soit $t_0 > 0$. Écrire la fiabilité $1 - F(t_0) = P(X > t_0)$ sous forme d'une espérance.

3) On considère le modèle de la loi exponentielle $(\mathbb{R}_+; B_{\mathbb{R}_+}, e(\lambda) : \lambda > 0)$;
En vous inspirant des résultats des deux questions précédentes et en utilisant à chaque fois la méthode des moments, proposer deux autres estimateurs du paramètre λ.

Exercice 5 (Maximum de vraisemblance pour un modèle gaussien)
1) On considère le modèle gaussien : $(\mathbb{R}; B(\mathbb{R})N(\mu;\sigma^2) : \mu \in \mathbb{R})$:
Donner l'estimateur du maximum de vraisemblance du paramètreμ basé sur une observation $x_1; ...; x_n$ d'un échantillon issu de ce modèle.
2) On considère maintenant le modèle gaussien avec paramètre bidimensionnel,i.e.
$(\mathbb{R}; B(\mathbb{R})N(\mu;\sigma^2 : \mu \in \mathbb{R}, \sigma^2$ >0) : Donner l'estimateur du maximum de vraisemblance du paramètre $\theta = (\mu, \sigma^2)$ pour le modèle d'échantillonnage associé.

Exercice 6 (Maximum de vraisemblance pour un modèle de loi uniforme) On considère le modèle uniforme $U_{[0,\theta]} : \theta > 0$

1) Montrer que la vraisemblance associée à un échantillon $x_1, ..., x_n$ observé dans ce modèle est :

L($x_1, ..., x_n;\ \theta$) $= \dfrac{1}{\theta^n} I_{X_{(1)} \geq 0} I_{x_{(n)} \leq \theta}$, où x_1 et $x_{(n)}$ sont respectivement les observations des statistiques d'ordre X(1) et X(n).

2) Donner l'estimateur du maximum de vraisemblance du paramètre θ.

Exercice 7 (Maximum de vraisemblance)

Pour les modèles suivants, donner l'estimateur du maximum de vraisemblance associé à l'observation d'un échantillon $X_1; ...; X_n$. 1) Modèle de la loi exponentielle décalée :
$(\mathbb{R}_+; B_{\mathbb{R}_+}, e_{t_0}(\lambda) : \lambda >0,\ t_0 \in R)$;

On rappelle que la densité de la loi exponentielle décalée $E_{t_0(\lambda)}$ est :

$f_{(\lambda,t_0)}(x) = \lambda exp(-\lambda(x - t_0))I_{[0,\infty])}$

2) Modèle de la loi Bêta à un seul paramètre : $(\mathbb{R}_+; B_{\mathbb{R}_+}, Beta(1, \theta) : \theta > 1)$

On rappelle que la densité de la loi Beta(a ; b) est :

$f_{a,b}(x) = \dfrac{1}{\beta(a, b)} x^{a-1}(1 - x)^{b-1} I_{[0,1]}(x)o$

(a ; b) est la valeur de la fonction Eulérienne Bêta prise en a et b. Ind. On pourra montrer en premier lieu que la densité pour le modèle considéré est : $f_\theta(x) = \theta(1 - x)^{\theta-1} I_{[0,1]}(x)$.

<u>Solutions</u> 1) Modèle de Poisson

La vraisemblance des observations est dans ce modèle : L($x_1; ...; x_n; \lambda$) $=$ $\Pi_{i=1}^{n} \dfrac{\lambda^{x_i}}{x_i!} e^{-\lambda} = \lambda^{\sum i=1^n x_i} e^{-n\lambda} \Pi_{i=1}^{n} \frac{1}{x_i!} = g_\lambda(T(x))h(x)$

en posant $h(x) = \frac{1}{\Pi_{i=1}^{n} x_i!}$; $g_\lambda(\mu) = \lambda^\mu e^{-n\lambda}$ et $T(x) = \sum_{i=1}^{n} x_i$

Par le théorème de factorisation la statistique T(x)=T(x)=$\sum i = 1^n x_i estexhau$

Modèle Exponentiel

L($x_1; ...; x_n; \lambda) = \Pi_{i=1}^{n} \lambda e^{-\lambda x} = \lambda^n exp[-\lambda \sum i = 1^n x_i] = g_\lambda(T(x))h(x)$

en posant $h(x) = 1$; $g_\lambda(t) = \lambda e^{-t\lambda}$ et $T(x) = \sum_{i=1}^{n} x_i$

Par le théorème de factorisation, la statistique T(x)=$\sum_{i=1}^{n} x_i estexhaustivep$

Modèle Gaussien avec σ^2 connu

La vraisemblance de l'échantillon s'écrit :

L($x_1; ...; x_n) = \frac{1}{\sigma}^n (2\Pi)^{\frac{n}{2}} exp[\frac{\mu}{\sigma^2} t - \frac{n\mu^2}{2\sigma^2}] exp[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} x_i^2] = g_{mu}(T(x))h(x)$

h(x)=exp[-$\frac{1}{2\sigma^2} \sum_{i=1}^{n} x_i^2$] $g_{mu}(t) = \frac{1}{\sigma}^n (2\Pi)^{\frac{n}{2}} T(x) = \sum_{i=1}^{n} x_i$

Le théorème de factorisation nous assure que T(x)=$\sum_{i=1}^{n} x_i estexhaustivepo$

Modèle Gaussien avec $\mu connu;$

Le théorème de factorisation nous assure que la statistique T (x)$\sum_{i=1}^{n}(x_i - \mu)^2 estexhaustivepour \sigma^2$

SERIE DE TD N° 4 AVEC SOLUTIONS

ENONCES

EXERCICE 1 :

en désire intérpréter les résultats suivants : le nombre de guérisons du concer de la peau à été de 1712 individus sur 2015 patient pour un traitement A et de 757 individus sur 1010 patients pour un traitement B.

Tester l'hypothèse H_0"un individu à la même probabilté d'être guéri dans les deux traitements" contre l'hypothèse H_1"les deux traitements sont caractérisés par deux probabilités de guérison différentes".

EXERCICE 2 :

Une enquête a été effectué en milieu hospitalier pour déterminer si l'usage du tabac favorise l'apparition du concer broncopulmonaire. Cette enquête a été menée de la manière suivante :

Les individus intérrogés sont répartis en quatre catégories selon leur consommation journalière en cigarette : A (non fumeurs) , B(de 1 à 9), C (de 10 à 19), D(de 20 ou plus) ; il s'agit d'une consommation moyenne évaluée sur les deux dernières années précédant l'enquête.

Un premier échantillon est constitué de concéreux. Un échantillon témoin a ensuite été choisi parmi les accidentés, c'est-à-dire les patients hospitalisés pour des raisons qui n'ont rien à voir avec le tabac, de plus pour éliminer tout autre facteur, àchaque concéreu correspond un témoin de même sexe, de même âge et intérrogé par le même enquiteur.

A partir des résultats ci-dessous, peut-on conclure à l'influence du tabac ?.

Ca	A	B	C	D	TOT
Co	25	66	177	334	602
T	130	136	165	171	602
TOT	155	202	342	505	1204

Ca=Catégorie, Co=Concéreux, T=Témoins.

EXERCICE : On a vacciné contre la grippe 300 personnes réparties en deux groupes A et B en fonction de l'âge :

Le groupe A comporte 120 individus de 55 ans au plus.

Le groupe B comporte 180 individus de plus de 55 ans.

On a constaté que, dans le groupe A, 38 individus ont eu la grippe l'hiver suivant la vaccination, tandis que 73 individus du groupe B ont eu la grippe ce même hiver.

Pet-on, au risque 10%, considérer qu'il existe un liaison entre l'éfficacité du vaccin et l'âge de la personne vaccinée ?.

EXERCICE 4 : On a vacciné contre la grippe 300 personnes réparties en deux groupes A et B en fonction de l'âge :

Le groupe A comporte 120 individus de 55 ans au plus.

Le groupe B comporte 180 individus de plus de 55 ans.

On a constaté que, dans le groupe A, 38 individus ont eu la grippe l'hiver suivant la vaccination, tandis que 73 individus du groupe B ont eu la grippe ce même hiver.

Pet-on, au risque 10%, considérer qu'il existe un liaison entre l'éfficacité du vaccin et l'âge de la personne vaccinée ?.

EXERCICE : On a croisé deux races de plantes différant par deux caractères : la couleur (rouge ou blanche) et la taille (grande ou petite) des fleurs qu'elle produisent.

La première génération est homogène et donne de grandes fleurs rouges. La seconde génération fait apparaître quatre type de plantes en fonction des fleurs qu'elles produisent : grandes fleurs rouges, grandes fleurs blanches, petites fleurs rouges et petites fleurs blanches.

Sur un échantillon de 320 plantes on a observé les résultats suivants :

phénotypes	GR	GB	PR	PB
effectifs	202	59	45	14

Peut-on considérer, au risque 5% , que les deux caractères étudiés se transmettent selon les lois de MENDEL ?.

EXERCICE 6 : Il est admet qu'en Algerie les groupes sanguins sont réparties de la façon suivante : O :40%, A :43% , B :12%, AB :5%.

Un échantillon de 300 étudiants à l'université de jijel a fourni les résultats :

Groupes	O	A	B	AB
effectifs	112	123	44	21

Peut-on affirmer, au risque 5% , que la répartition des groupes sanguins à l' université de jijel ne diffère pas sensiblement de celle de l'Algerie ?.

EXERCICE 7 : Dan une population de 500 personnes (300 hommes et 200 femmes) on a mesuré la tension artérielle dechaque individu, ce qui a donné les résultats suivants :

	Hypert	TN	Hypot
H	72	192	36
F	38	118	44

Peut-on, au risque 5%, émettre l'hypothèse H_0 d'une liaison entre le sexe de l'individu et la tension artérielle ?.

INDICATION : Le nombre de degrés de liberté est le nombre minimum des case du tableau dont il faut connaître l'effectif pour déterminer l'ensemble du tableau où les sommes de chaque ligne et chaque conlonne sont données.

Dans l'exercice précédent le nombre de degrés de liberté est 2.

EXERCICE 8 : Un médicament a été expérimenté sur 200 malades dévisés en deux groupes M_1 et M_2 indépendants :

-le groupe M_1 composé de 110 malades a aborbé le médicament étudié.

-le groupe M_2 composé de 90 malades a aborbé un placebo.

Les résultats sont les suivants : 60 malades guéris dans le groupe M_1, 36 malades guéris dans le groupe M_2.

1°) Calculer le pourcentage de guérisons et l'écart-type de ce pourcentage pour chacun des échantillons M_1 et M_2.

2°) En admettant que le phénomène étudié suit une loi normale, construir un test permettant d'accepter ou de rejeter l'hypothèse de l'éfficacité du médicament au risque de 5%.

EXERCICE 9 : On veut savoir si une maladie M modifie le taux de certaines protéines dans le song. On a mesuré leurs concentrations dans un échantillon de sujets atteints pa M et dans un autre échantillons formé de sujets en bonne santé (sujets témoins). Les résultats (dans une unité convenable) sont les suivants :

	effectifs	moyenne échantillon	variance échantillon
Malades	77	141	40
Témoins	33	131	32

Tester l'hypothèse "taux identiques chez les malades et les témoins" contre l'hypothèse :

a) "taux différent chez les malades et les témoins".

b) "taux supérieur chez les malades".

EXERCICE 10 : On a mesuré les dimensions d'une tumeur chez les souris traitées ou non par une substance anti-tumorale et on a obtenu :

Surface (cm^2)	5	5,5	6	6,5	7	7,5	8
Nombre de témoins	0	0	2	3	8	4	3
Nombre traités	4	4	8	3	0	1	0

La différence observée est-elle significative ?.

EXERCICE 11 : Dans une maternité, on a compré les poids à la naissance des des bébés de mères primipares et multipares. On a obtenu les résultats suivants :

primipares	$n_1 = 100 \quad \overline{x_1} = 3180g \quad \sigma_{e1}^2 = 214400$
multipares	$n_2 = 110 \quad \overline{x_2} = 3400g \quad \sigma_{e2}^2 = 243300$

Peut-on admettre au coefficient de confiance de 99% que les enfants nés de mères multipares sont plus lourds que ceux nés de mères primipares ?.

SOLUTIONS

EXERCICE 3

L'hypothèse à tester est : H_0 " l'efficacité du vaccin ne dépend pas de l'âge e la perssonne vaccinée".

Loi théorique : On désigne respectivement par G et $\overline{G}$ le fait que la personne vaccinée ait eu, ou non la grippe l'hiver suivant la vaccination.

Sous l'hypothèse H_0, l'indépendence des évènements conduit, par exemple à : $P(A \cap G) = P(A)P(G) = \frac{120}{300} * \frac{110}{300} = 0,147$ Par suite l'effectif théorique correspondant est $0,147 * 300 \simeq 44$ En procédant de même pour es autrescouples on obtient le tableau de contingence suivant :

	G	$\overline{G}$	Somme
A	38/44	82/76	120
B	72/66	108/114	180
Somme	110	190	300

La loi du χ^2 est ici à 1 degré de liberté ;on a $\chi^2_{observé} = 2,15$

D'autre part $\alpha = 0,1$ et $v = 1$ la table donne $\chi^2_{lu} = 2,71$

On peut donc admettre au risque 10% , l'hypothèse selon laquelle l'efficacité du vaccin ne dépend pas de l'âge de la personne vaccinée

EXERCICE 5

Le tableau suivant résume les calculs Pour défìnr la loi théorique nous n'utilisé que la relation donnant l'effectif total et nous disposons de quatre classes donc la loi du χ^2 à 3 degrés de liberté. la table donne $\chi^2_{lu} = 7,81$. Nous somme donc amenés à rejeter l'hypothèse suivant laquelle les deux caractères étudiés se transmettent suivant les lois de MENDEL

phénotype	O_i	T_i	O_i-T_i	$\frac{(O_i-T_i)^2}{T_i}$
O	202	180	22	2,69
A	59	60	-1	0,02
B	45	60	-5	3,75
AB	14	20	-6	1,80
Somme	320	320	0	8,26

EXERCICE 6

Le tableau suivant résume les calculs nous avons quatre classes et nous n'avons utilisé que la relation donnant l'effectif total pour déterminer la loi théorique, donc la loi du χ^2 à 3 degrés de liberté. la table donne $\chi^2_{lu} = 7,81$. Nous acceptons l'hypothèse selon laquelle la répartition des groupes sanguins dan la population étudiée est conforme à celle de l'algérie.

phénotype	O_i	T_i	O_i-T_i	$\frac{(O_i-T_i)^2}{T_i}$
GR	112	120	-8	0,53
GB	123	129	-6	0,28
PR	44	36	8	1,78
PB	21	15	6	2,40
Somme	300	300	0	4,99

EXERCICE 7

On obtient le tableau de contingence suivant
On trouve $\chi^2_{observé} = 8,9$. Pour $\alpha = 5\%$ *et* $v = 2$ la table donne $\chi^2_{lu} = 5,99$. Nous rejetons H_0 et concluons à une dépendance des deux paramètres étudiés.

	Hyper	Normale	Hypo	*somme*
Hommes	72∕66	192∕166	36∕48	300
Femmes	38∕44	118∕124	44∕32	200
sommes	110	310	80	500

EXERCICE 8

1°) Dans un échantillon de taille n, extrait d'une population P où la fréquence de guérisons est p, la variable aléatoitre F prenant pour valeurs la fréquence de guérisons suit la loi normale $N(p, \sqrt{\dfrac{p(1-p)}{n}}$ lorsque n ≥ 30.

Si p est inconnu, on l'estime par f, fréquence de guérisons dans l'échantillon.

Dans M_1 celle-ci est $f_1 = \frac{66}{110} = 0,66$ et l'écart-type est $\sigma_1 = \sqrt{\dfrac{0,6(0,4)}{110}} \simeq 4,67 * 10^{-2}$

Dans M_2 celle-ci est $f_2 = \frac{36}{90} = 0,4$ et l'écart-type est $\sigma_2 = \sqrt{\dfrac{0,6(0,4)}{90}} \simeq 5,14 * 10^{-2}$

2°) On costruit un test d'homogénéité permettant de comparer les fréquences f_1 et f_2.

Hypothèse à tester

Soit H_0 l'hypothèse " le médicament est inefficace", c'est-à dire "f_1 et f_2 ne sont pas significativement différents"

Nature du test : c'est un test unilatéral au seuil de risque 5%.

L'hypothèse contraductoire est (H_1) "le médicament est efficace" c'est-à-dire f_1 est significativement supérieur à f_2.

Condition de rejet de H_0 : sous l'hypothèse H_0 la variable aléatoire $F_1 - F_2$ suit approximativement la loi normale $N(0, \sqrt{p(1-p)(\frac{1}{n_1} + \frac{1}{n_2}})$.

Ici p est inconnu

sous H_0 nous réunissons les deux échantillons et nous etimons p par $\widehat{p}$ tel que $\widehat{p} = \dfrac{n_1 f_1 + n_2 f_2}{n_1 + n_2} \simeq 0,51$.

Alors sous l'hypothèse H_0 $F_1 - F_2$ suit approximativement la loi normale $N(0, \sqrt{\widehat{p}(1-\widehat{p})(\frac{1}{n_1} + \frac{1}{n_2}})$).soit $N(0; 0,071)$ la variable aléatoire $T = \dfrac{F_1 - F_2}{\sqrt{\widehat{p}(1-\widehat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$ suit approximativement la loi normale $N(0,1)$.

Au seuil de risque 5%, il existe un réel u_α strictement positif tel que $p(T \geq u_\alpha) = 0,05$ ou ce qui est équivalent à $p(T \leq u_\alpha) = 0,95$; la table donne $\Pi\,(1,64) = 0,95$, la condition de rejet est donc $T > 1,64$.

Mise en oeuvre du test

On a $t = \dfrac{f_1 - f_2}{\sqrt{p(1-p)(\frac{1}{n_1} + \frac{1}{n_2})}} = \dfrac{0,6 - 0,4}{\sqrt{0,51 * 0,49(\frac{1}{110} + \frac{1}{90})}} \simeq 2,82.$

Donc $t > 1,64$ et nous rejetons H_0.

Conclusion : Nous acceptons l'hypothèse de l'efficacité du médicament, avec un risque 5% de se tromper.

EXERCICE 9

On se propose e tester l'hypothèse nulle H_0 suivante :

H_0 :"le taux moyen des protéines est identique chez les malades et les temoins"

Il s'agit de comparer deux moyennes à partir des observations fournies par deux échantillons de taille assez grande pour que les distributillons d'échantillonnage suivent une loi normale. On test cette hypothèse au seuil $\alpha = 5\%$.

a) : Si H_1 est :"les taux moyens sont différent", il s'agit bilatéral est la valeur critique à ne pas dépasser est $u_\alpha = u_{5\%} = 1,96$.

On calcul alors : $Z = \dfrac{|\overline{x_1} - \overline{x_2}|}{\sqrt{\frac{\sigma_{e1}^2}{n_1 - 1} + \frac{\sigma_{e2}^2}{n_2 - 1}}} = \dfrac{|141 - 131|}{\sqrt{\frac{40}{76} + \frac{32}{32}}} \simeq 8,094 >> u_{5\%} = 1,96$; on est conduit à rejeter l'hypothèse H_0 . Il y a tout lieu de penser que les taux moyens sont differents.

b) H_1 est telle que maintenant le test est unilatéral, on compar $Z \simeq 8,094$ avec $v_{5\%} = u_{2,5\%} = 1,64$.

La conclusion rest identique pour H_0. Tout laisse penser que le taux moyen est supérieur chez les individus malades.

Remarque : On peut noter que lorsqu'un test bilatéral est significatif (H_0 rejetée) le test unilatéral correspondant l'est toujours (puisque $u_\alpha > v_\alpha$).

EXERCICE 10.

Il s'agit de tester l'égalité de deux moyennes expérimentales sur deux petits échantillons. On a ici $n_1 = n_2 = 20$ et on calcule :

$\overline{x_1} = 7,075\ cm^2 \quad et \quad \sigma_{e1} = 0,576\ cm^2$

$\overline{x_2} = 5,850\ cm^2 \quad et \quad \sigma_{e2} = 0,614\ cm^2$

Posons l'hypothèse de travail H_0 "les surfaces moyennes sont égales" et teston la contre H_1 "les surfaces moyennes sont différentes" au seuil $\alpha = 5\%$.

Avant de mener ce test, il faut comparer les variances des populations avec un test F à partir des estimations $S_1^2 = (0,576)^2 * \frac{20}{19} = 0,349$ et $S_2^2 = (0,614)^2 * \frac{20}{19} = 0,397$.

On calcule la valeur :$F = \frac{0,397}{0,349} = 1,37$ car $S_2^2 > S_1^2$. Le test est bilatéral, aussi va-t-on comparer cette valeur à $F_{2,5\%;6;6} = 5,82$. La valeur calculée est inferieur à la valeur critique ; on accepte donc l'hypothèse H_0 au risque de 5% et on peut remplacer les deux variences par une variance commune donnée par $\widehat{\sigma}^2 = \frac{n_1\sigma_{e1}^2 + n_2\sigma_{e2}^2}{n_1 + n - 2_2} = \frac{14,175}{38} = 0,373$.

Il est maintenant possible de comparer les moyennes en posant H_0 "les moyennes des deux populations son égales" contre l'alternative d'inégalité. Cel revient à supposer que le traitement n'a pas d'effet.

On calcule alors : $t = \frac{|\overline{x_1} - \overline{x_2}|}{\widehat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{1,225}{0,61 * 0,316} \simeq 0,344$.

On compare cette valeur à celle de la loi de Student pour $\alpha = 5\%$ et $v = n_1 + n_2 - 2 = 12$. Elle est déduite de la table et vaut 2,179 qui est bien inferieur à la valeur calculée. On est conduit à ne pas retenir l'hypothèse nulle et à en conclure que la différence constatée est significative.

EXERCICE 11

Il s'agit de tester l'hypothèse nulle H_0 " le poids moyen des bébés est le même que soit lamère est primipare ou multipare", si m_1 et la moyenne vraie du poids des enfants des mères primipares et m_2 celle des enfants des mères multipares. On test H_0 contre l'hypothèse alternative H_1 "$m_1 < m_2$". On mènera donc un test unilatéral à partir de deux grands échantillons ; la valeur critique à ne pas dépasser est $v_{1\%} = 2,326$.

Pour prendre une décision il faut calculé la quantité : $Z = \frac{|\overline{x_1} - \overline{x_2}|}{\sqrt{\frac{\sigma_{e1}^2}{n_1 - 1} + \frac{\sigma_{e2}^2}{n_2 - 1}}} = \frac{220}{66,316} \simeq 3,317 > v_\alpha = 2,326$.

Il faut donc rejeter H_0 au profit de H_1 au seuil $\alpha = 1\%$. Il y a tout lieu de penser que les enfants nés de mères multipares sont plus lourds que les autres.

Modèle Gaussien (R ; B(R) ; $\mu in R\sigma^2 > 0) Lethormedefactorsationassurequelastatistique$ T (x) $=$ $(\sum_{i=1}^{n}(x_i, \sum_{i=1}^{n}(x_i - \mu)^2)$ est exhaustive pour (μ, σ^2)