Loi khi-deux χ^2

On rappelle que la fonction Γ, définie par $x \longmapsto \displaystyle\int_0^{+\infty} t^{x-1}e^{-t}\,\mathrm{d}t$ sur $\mathbb{R}_+^*$, vérifie les relations

— $\forall x \in \mathbb{R}_+^*$: $\quad \Gamma(x+1) = x\Gamma(x)$

— $\forall n \in \mathbb{N}^*$: $\quad \Gamma(n) = (n-1)!$

— $\Gamma\left(\dfrac{1}{2}\right) = \sqrt{\pi}$

Partie I: Loi Gamma et loi du χ^2

Dans toute cette partie, les variables aléatoires réelles, définies sur le même espace probabilisé (Ω, T, P).

1. (a) Établir que, pour tout couple (x, y) de nombres réels strictement positifs, la fonction $t \mapsto t^{x-1}(1-t)^{y-1}$ est intégrable sur l'intervalle $]0,1[$.
   Pour tout couple (x, y) de nombres réels strictement positifs, on pose :
   $$\beta(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,\mathrm{d}t$$
   (b) Soit $s, x, y \in \mathbb{R}_+^*$, montrer que $\displaystyle\int_0^s t^{x-1}(s-t)^{y-1}\,\mathrm{d}t = \beta(x,y)s^{x+y-1}$.
2. Soit b et p sont deux paramètres réels strictement positifs et X une variable aléatoire suivant la loi $\Gamma\left(p, \dfrac{1}{b}\right)$
   (a) Reconnaitre f la densité de X ;
   (b) Donner l'espérance et la variance de X ;
   (c) Dans le cas particulier où $p = 1$, reconnaître la loi de X et exprimer sa fonction de répartition.
3. Soit p_1, p_2 et λ trois réels strictement positifs. X_1 et X_2 deux variables aléatoires réelles indépendantes, suivant les lois respectives $\Gamma(p_1, \lambda)$ et $\Gamma(p_2, \lambda)$. Posons $S = X_1 + X_2$ et soit f_S sa fonction densité.
   (a) Montrer que :
   $$\forall s \in \mathbb{R}, \quad f_S(s) = \begin{cases} \dfrac{\lambda^{p1+p_2}\beta(p_1,p_2)}{\Gamma(p_1)\Gamma(p_2)} s^{p_1+p_2-1}e^{-\lambda s} & \text{si } s > 0 \\ 0 & \text{si } s \leqslant 0 \end{cases}$$
   (b) En déduire que $\beta(p_1,p_2) = \dfrac{\Gamma(p_1)\Gamma(p_2)}{\Gamma(p_1+p_2)}$, puis déduire $X_1 + X_2 \hookrightarrow \Gamma(p_1+p_2, \lambda)$
4. Soient $n \in \mathbb{N}^*$, $X_1, \cdots, X_n$ variables aléatoires réelles mutuellement indépendantes, suivant les lois respectives $\Gamma(p_i, \lambda)$ pour tout $i \in [\![1, n]\!]$
   Déterminer la loi de la variable aléatoire $\displaystyle\sum_{i=1}^{n} X_i$

Soit s un entier strictement positif; on appelle loi du χ^2 à s degrés de liberté ou loi $\chi^2(s)$ la loi $\Gamma\left(\frac{s}{2}, \frac{1}{2}\right)$

5. (a) Donner l'espérance et la variance d'une variable aléatoire suivant la loi $\chi^2(s)$.
   (b) Soit $n \in \mathbb{N}^*$. Si pour tout $i \in [\![1, n]\!]$, Les X_i sont des variables mutuellement indépendantes, X_i suivant la loi $\chi^2(s)$, donner la loi de $\displaystyle\sum_{i=1}^{n} X_i$
6. Pour $x > 0$ et $n \in \mathbb{N}^*$, Justifier la formule : $e^x = \displaystyle\sum_{k=0}^{n-1} \frac{x^k}{k!} + \int_0^x e^{x-t}\frac{t^{n-1}}{(n-1)!}\,\mathrm{d}t$
7. Soient X_{2n} une variable aléatoire suivant la loi $\chi^2(2n)$ et λ un réel strictement positif.
   Démontrer que $P(X_{2n} > 2\lambda) = P(Y_\lambda < n)$ où Y_λ désigne une variable aléatoire suivant la loi de Poisson de paramètre λ
8. (a) Soit X une variable aléatoire suivant une loi normale centrée réduite. Montrer que X^2 suit la loi du χ^2 à un degré de liberté.

(b) $X_1, \cdots, X_n$ variables aléatoires réelles mutuellement indépendantes, suivant la loi normale centrée réduite. Quelle est la loi de $\sum_{i=1}^{n} X_i^2$

Partie II: De l'urne au χ^2

Une urne contient des boules de couleurs $C_1, \cdots, C_k$ où $k \in \mathbb{N}$ et $k \geqslant 2$. Les boules de couleur C_i sont en proportion p_i non nulle avec $\sum_{i=1}^{n} p_i = 1$. Étant donné n entier naturel non nul, on tire successivement dans l'urne n boules, avec remise après chaque tirage.
Pour $i \in [\![1, k]\!]$, on appelle X_i la variable aléatoire désignant le nombre de boules de couleur C_i obtenues lors du tirage. X_i suit par conséquent une loi binomiale de paramètres n et p_i.

9. (a) Déterminer l'espérance et la variance de X_i
   (b) Pour $i \neq j$, déterminer la loi de $X_i + X_j$ et en déduire que $\operatorname{cov}(X_i, X_j) = -np_ip_j$
10. Soit k entier supérieur ou égal à 2. On pose $Y_i = \dfrac{X_i - np_i}{\sqrt{np_i}}$ et on note M la matrice de covariance des variables aléatoires $Y_1, \cdots, Y_k$ définie par $M = (\operatorname{cov}(Y_i, Y_j))_{1 \leqslant i,j \leqslant k}$
   (a) Montrer que pour $i \neq j$, $\operatorname{cov}(Y_i, Y_j) = -\sqrt{p_ip_j}$ et que $\operatorname{cov}(Y_i, Y_i) = 1 - p_i$.
   En déduire l'expression de la matrice M.
   Dans la suite, $M_{n,p}(\mathbb{R})$ désigne l'ensemble des matrices à n lignes et p colonnes à coefficients réels. On notera I_k la matrice unite de $M_k(\mathbb{R})$
   (b) Préciser la matrice P telle que $M = I_k - P$, puis la matrice $C \in M_{k,1}(\mathbb{R})$ vérifiant $P = C\,{}^tC$, où ${}^tC$ désigne la transposée de la matrice C.
   (c) Déterminer le rang de P.
   (d) Calculer P^2. Préciser les valeurs propres de P et leur multiplicité.
11. Soit $J \in M_k(\mathbb{R})$ dont les coefficients $a_{i,j}$ vérifient : $a_{i,j} = \begin{cases} 1 & \text{si } i = j \leqslant k-1 \\ 0 & \text{sinon} \end{cases}$.
   Justifier l'existence d'une matrice $S \in M_k(\mathbb{R})$ vérifiant : ${}^tSS = I_k$ et ${}^tSMS = J$
12. On définit les variables aléatoires Z_1, $Z_2, \cdots, Z_k$ par la relation :

$$\begin{pmatrix} Z_1 \\ \vdots \\ Z_k \end{pmatrix} = {}^tS \begin{pmatrix} Y_1 \\ \vdots \\ Y_k \end{pmatrix}$$

   et on pose $S = (s_{ij})_{1 \leqslant i,j \leqslant k}$
   (a) Prouver que, pour tout entier i entre 1 et k, Z_i est centrée.
   (b) Déterminer la matrice de covariance de $Z_1, Z_2, \cdots, Z_k$. En déduire que Z_k est la variable certaine égale à zéro.
   (c) On pose $Q = \sum_{i=1}^{k} Y_i^2$. Calculer Q en fonction de $Z_1, Z_2, \cdots, Z_k$.
   (d) On suppose que n est grand, que les variables $Z_1, Z_2, \cdots, Z_{k-1}$ sont mutuellement indépendantes et qu'elles suivent des lois normales centrées réduites. Justifier le fait que Q suit la loi du χ^2 à $(k-1)$ degrés de liberté.

Loi khi-deux χ^2

Partie I: Loi Gamma et loi du χ^2

1. (a) La fonction $t \mapsto t^{x-1}(1-t)^{y-1}$ est continue sur $]0,1[$.

   - Équivalente en 0^+ à t^{x-1}, elle est intégrable sur $\left]0, \frac{1}{2}\right]$ si, et seulement si, $x-1>-1$, c'est-à-dire $x>0$;
   - Équivalente en 1^- à $(1-t)^{y-1}$, elle est intégrable sur $\left[\frac{1}{2}, 1\right[$ si, et seulement si, $y-1>-1$, c'est-à-dire $y>0$.

   Elle est donc intégrable sur $]0,1[$ si, et seulement si, x et y sont strictement positifs.

   (b) L'intégrale $\displaystyle\int_0^1 t^{x-1}(1-t)^{y-1}\,\mathrm{d}t$ converge et l'application $t \longmapsto \dfrac{t}{s}$ est une bijection de classe $\mathcal{C}^1$ de $]0,s[$ vers $]0,1[$, donc par intégration par parties on a :

$$\int_0^1 t^{x-1}(1-t)^{y-1}\,\mathrm{d}t = \int_0^s \left(\frac{t}{s}\right)^{x-1}\left(1-\frac{t}{s}\right)^{y-1}\frac{1}{s}\,\mathrm{d}t = \frac{1}{s^{x+y-1}}\int_0^s t^{x-1}(s-t)^{y-1}\,\mathrm{d}t$$

   Ainsi l'égalité demandée

2. (a) f est définie sur $\mathbb{R}$ par :

$$\forall x\in\mathbb{R}, f(x) = \begin{cases} \dfrac{x^{p-1}e^{-\frac{x}{b}}}{b^p\Gamma(p)} & \text{si } x>0 \\ 0 & \text{si } x\leqslant 0 \end{cases}$$

   (b) $\mathbb{E}(X) = bp$ et $\mathbb{V}(X) = pb^2$

   (c) Lorsque $p=1$, on a $f(x) = \dfrac{1}{b}e^{-\frac{x}{b}}\chi_{]0,+\infty[}$, donc la loi de X est la loi exponentielle de paramètre $\dfrac{1}{b}$. On a alors $F_X(x)=0$ si $x\leqslant 0$ et si $x>0$

$$\begin{aligned} F_X(x) &= \int_{-\infty}^x f(t)\,\mathrm{d}t = \int_0^x \frac{1}{b}e^{-\frac{t}{b}}\,\mathrm{d}t \\ &= 1-e^{-\frac{x}{b}} \end{aligned}$$

3. (a) On sait que X_1+X_2 est une variable à densité de densité h définie par :

$$\forall s\in\mathbb{R},\quad f_S(s) = \int_{-\infty}^{+\infty} f_1(s-t)f_2(t)\,\mathrm{d}t$$

   - Pour $s\leqslant 0$, on a $f_S(s)=0$
   - Pour $s>0$, on a

$$\begin{aligned} f_S(s) &= \int_0^s (s-t)^{p_1-1}\frac{\lambda^{p_1}e^{-\lambda(s-t)}}{\Gamma(p_1)}t^{p_2-1}\frac{\lambda^{p_2}e^{-\lambda t}}{\Gamma(p_2)}\,\mathrm{d}t \\ &= \frac{\lambda^{p_1+p_2}e^{-\lambda s}}{\Gamma(p_1)\Gamma(p_2)}\int_0^s (s-t)^{p_1-1}\,t^{p_2-1}\,\mathrm{d}t \\ &= \frac{\lambda^{p_1+p_2}e^{-\lambda s}}{\Gamma(p_1)\Gamma(p_2)}s^{p_1+p_2-1}\beta(p_1,p_2) \end{aligned}$$

   f_S étant une densité, donc elle est intégrable sur $\mathbb{R}$ d'intégrale 1. Or

$$\begin{aligned} \int_{-\infty}^{+\infty} f_S(s)\,\mathrm{d}s &= \frac{\beta(p_1,p_2)}{\Gamma(p_1)\Gamma(p_2)}\underbrace{\int_0^{+\infty}\lambda^{p_1+p_2}s^{p_1+p_2-1}e^{-\lambda s}\,\mathrm{d}s}_{=\Gamma(p_1+p_2)} \\ &= \frac{\beta(p_1,p_2)}{\Gamma(p_1)\Gamma(p_2)}.\Gamma(p_1+p_2) \end{aligned}$$

   Et $\displaystyle\int_{-\infty}^{+\infty} f_S(s)\,\mathrm{d}s = 1$, donc $\beta(p_1,p_2) = \dfrac{\Gamma(p_1)\Gamma(p_2)}{\Gamma(p_1+p_2)}$

Donc $X_1 + X_2$ est de loi $\Gamma(p_1 + p_2, \lambda)$

4. Par récurrence simple sur $n \in \mathbb{N}^*$, on montre que $\sum_{i=1}^{n} X_i$ est de loi $\Gamma\left(\sum_{i=1}^{n} p_i, \lambda\right)$

   - Pour $n = 1$, rien à démontrer
   - Soit $n \in \mathbb{N}^*$ et $X_1, \cdots, X_{n+1}$ des variables indépendantes suivant des lois respectives $\Gamma(p_i, \lambda)$ pour $i \in [\![1, n+1]\!]$. Par hypothèse de récurrence $\sum_{i=1}^{n} X_i$ est de loi$\Gamma\left(\sum_{i=1}^{n} p_i, \lambda\right)$. Or, par indépendance héritée, les variables $\sum_{i=1}^{n} X_i$ et X_{n+1} sont indépendantes, alors on conclut d'après la question précédente que $\sum_{i=1}^{n+1} X_i = \sum_{i=1}^{n} X_i + X_{n+1}$ est de loi $\Gamma\left(\sum_{i=1}^{n+1} p_i, \lambda\right)$

5. (a) L'espérance vaut s et la variance vaut $2s$

   (b) D'après la question précédente $\sum_{i=1}^{n} X_i$ est de loi $\Gamma\left(\frac{ns}{2}, \frac{1}{2}\right) = \chi^2(ns)$

6. Soit $n \in \mathbb{N}^*$ et $x \in \mathbb{R}_+^*$.
   La fonction exp est de classe $\mathcal{C}^\infty$, alors d'après la formule de Taylor avec reste intégrale à l'ordre $n-1$, on a

$$
\begin{aligned}
e^x &= \sum_{k=0}^{n-1} \frac{x^k}{k!} \exp^{(k)}(0) + \int_0^x \frac{(x-t)^{n-1}}{(n-1)!} \exp^{(n)}(t)\,\mathrm{d}t \\
&= \sum_{k=0}^{n-1} \frac{x^k}{k!} + \int_0^x \frac{(x-t)^{n-1}}{(n-1)!} e^t\,\mathrm{d}t \\
&\underset{u=x-t}{=} \sum_{k=0}^{n-1} \frac{x^k}{k!} + \int_0^x e^{x-u} \frac{u^{n-1}}{(n-1)!}\,\mathrm{d}u
\end{aligned}
$$

7. Par définition de X_{2n}, on a pour tout λ un réel strictement positif :

$$
\begin{aligned}
P\left(X_{2n} > 2\lambda\right) &= 1 - \int_0^{2\lambda} \frac{t^{n-1} e^{-\frac{t}{2}}}{2^n \Gamma(n)}\,\mathrm{d}t \\
&= 1 - \int_0^{\lambda} \frac{s^{n-1} e^{-s}}{\Gamma(n)}\,\mathrm{d}t \\
&= 1 - e^{-\lambda} \int_0^{\lambda} \frac{s^{n-1} e^{\lambda-s}}{(n-1)}\,\mathrm{d}t \\
&= 1 - e^{-\lambda}\left(e^\lambda - \sum_{k=0}^{n-1} \frac{\lambda^k}{k!}\right) \\
&= \sum_{k=0}^{n-1} \frac{\lambda^k}{k!} e^{-\lambda} = P\left(Y_\lambda < n\right)
\end{aligned}
$$

8. (a) On sait que $P\left(X \leqslant x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,\mathrm{d}t$.
   Soit $y \in \mathbb{R}$, on a :

   - Si $y \leqslant 0$, alors $P\left(X^2 \leqslant y\right) = 0$

- Si $y > 0$, alors

$$\begin{aligned} P\left(X^2 \leqslant y\right) &= P\left(|X| \leqslant \sqrt{y}\right) \\ &= P\left(-\sqrt{y} \leqslant X \leqslant \sqrt{y}\right) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{y}}^{\sqrt{y}} e^{-\frac{t^2}{2}}\, \mathrm{d}t \\ &= \frac{2}{\sqrt{2\pi}} \int_{0}^{\sqrt{y}} e^{-\frac{t^2}{2}}\, \mathrm{d}t \\ &= \frac{2}{\sqrt{2\pi}} \int_{0}^{\sqrt{y}} e^{-\frac{t^2}{2}}\, \mathrm{d}t \\ &= \frac{1}{\sqrt{2\pi}} \int_{0}^{y} \frac{e^{-\frac{s}{2}}}{\sqrt{s}}\, \mathrm{d}s \end{aligned}$$

Donc X^2 est de loi $\chi^2(1)$

(b) Les variables $X_1, \cdots, X_n$ sont mutuellement indépendantes, donc Les variables $X_1^2, \cdots, X_n^2$ le sont aussi donc $\displaystyle\sum_{i=1}^{n} X_i^2$ est de loi $\chi^2(n)$

Partie II: De l'urne au χ^2

9. (a) X_i suit la loi binomiale de taille n et de paramètre p_i. Donc $\mathbb{E}(X_i) = np_i$ et $\mathbb{V}(X_i) = np_i(1-p_i)$

(b) Pour $i \neq j$, la variable $X_i + X_j$ suit la loi binomiale de taille n et de paramètre $p_i + p_j$. Donc

$$\mathbb{V}\left(X_i + X_j\right) = n\left(p_i + p_j\right)\left(1 - p_i - p_j\right)$$

On déduit, alors

$$\begin{aligned} \operatorname{cov}\left(X_i, X_j\right) &= \frac{1}{2}\left(\mathbb{V}(X_i + X_j) - \mathbb{V}(X_i) - \mathbb{V}(X_j)\right) \\ &= \frac{1}{2}\left(n\left(p_i + p_j\right)\left(1 - p_i - p_j\right) - np_i(1-p_i) - np_j(1-p_j)\right) \\ &= -np_ip_j \end{aligned}$$

10. (a) La covariance étant bilinéaire, et inchangée quand on ajoute une constante à une des variables, on obtient pour $i \neq j$

$$\operatorname{cov}\left(Y_i, Yj\right) = \frac{1}{\sqrt{np_i}} \frac{1}{\sqrt{np_j}} \operatorname{cov}\left(\tilde{X}_i, \tilde{X}_j\right) = \frac{1}{\sqrt{np_i}} \frac{1}{\sqrt{np_j}} \operatorname{cov}\left(X_i, X_j\right) = -\sqrt{p_ip_j}$$

D'autre part, $\operatorname{cov}\left(Y_i, Y_i\right) = \mathbb{V}\left(Y_i\right) = \dfrac{1}{np_i}\mathbb{V}\left(X_i\right) = 1 - p_i$

(b) Pour $i, j \in [\![1, k]\!]^2$, le coefficient de position (i, j) de P avec $i \neq j$ vaut

$$P_{ij} = -\operatorname{cov}\left(X_i, X_j\right) = \sqrt{p_ip_j}$$

et

$$P_{ii} = 1 - \operatorname{cov}\left(X_i, X_i\right) = p_i$$

Bref $P_{ij} = \sqrt{p_ip_j}$. Ainsi, on déduit $C = \begin{pmatrix} \sqrt{p_1} \\ \vdots \\ \sqrt{p_k} \end{pmatrix}$ et on a $P = C^tC$

(c) Toutes les lignes de P sont deux à deux colinéaires et au moins une est non nulle puisqu'il existe un p_i non nul (la somme fait 1), donc P est de rang 1.

(d) P est du rang 1, donc $P^2 = \operatorname{Tr}(P)\, P = P$. Les valeurs propres de la matrice symétrique P sont réelles , et il y en a k, comptées avec leur multiplicité car elle est diagonalisable. Comme $P^2 - P = 0$ on sait que les valeurs propres vérifient $\lambda^2 - \lambda = 0$, donc c'est 0 ou 1. On sait que 0 est valeur propre de multiplicité $k-1$ car $\dim \operatorname{Ker}(P) + \mathbf{rg}(P) = k$. La seule possibilité est donc que 1 est valeur propre simple.

Loi khi-deux χ^2

11. $M = I_k - P$ donc la matrice M admet 1 pour valeur propre de multiplicité $k-1$ et 0 pour valeur propre simple. Comme elle est symétrique, il existe une base orthonormée de vecteurs propres. Si on note S la matrice de passage de la base canonique à cette base, on obtient $S^tS = I_k$ et ${}^tSMS = J$ où J est la matrice diagonale avec des coefficients diagonaux $\mathbf{diag}\,(1, 1, \cdots, 1, 0)$
12. On définit les variables aléatoires Z_1, $Z_2, \cdots, Z_k$ par la relation :

$$\begin{pmatrix} Z_1 \\ \vdots \\ Z_k \end{pmatrix} = {}^tS \begin{pmatrix} Y_1 \\ \vdots \\ Y_k \end{pmatrix}$$

et on pose $S = (s_{ij})_{1 \leqslant i,j \leqslant k}$

(a) L'espérance étant linéaire, on a

$$\begin{pmatrix} \mathbb{E}(Z_1) \\ \vdots \\ \mathbb{E}(Z_k) \end{pmatrix} = {}^tS \begin{pmatrix} \mathbb{E}(Y_1) \\ \vdots \\ \mathbb{E}(Y_k) \end{pmatrix} = {}^tS \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

Donc pour tout entier i entre 1 et k, Z_i est centrée.

(b) Comme les variables sont centrées, $\operatorname{cov}(Z_i, Z_j) = \mathbb{E}(Z_iZ_j)$ donc la matrice de covariance M' des Z_j est l'espérance de la matrice $(Z_iZ_j)_{1 \leqslant i,j \leqslant k}$. On a

$$(Z_iZ_j)_{1 \leqslant i,j \leqslant k} = {}^tS \begin{pmatrix} Y_1 \\ \vdots \\ Y_k \end{pmatrix} \begin{pmatrix} Y_1 & \cdots & Y_k \end{pmatrix} S$$

et en passant à l'espérance de chaque terme, $M' = {}^tSMS = J$. Il en résulte que $\operatorname{cov}(Z_k, Z_k) = 0 = E\left(Z_k^2\right)$ et que $Z_k = 0$ presque partout

(c) La matrice S est orthogonale donc $Q = \displaystyle\sum_{i=1}^{k} Z_i^2 = \sum_{i=1}^{k-1} Z_i^2$.

(d) Q suit une loi du χ^2 à $k-1$ degrés de liberté.