Exercices, 3

Correction Exo 12

1. On peut prendre
   - $\mathcal{X} = \{1, \ldots d\}^{\mathbb{N}}$ (suites à valeurs dans $\{1, \ldots d\}$).
   - $\mathcal{A} = \sigma\left(\bigcup_{i=1}^{+\infty} \mathcal{P}(\{1, \ldots d\}^i) \times \{1, \ldots, d\}^{\mathbb{N}-i}\right)$ (plus petite tribu rendant observable les réalisations de $(X_1, \ldots, X_N) = \pi_{1:N}(X_1, \ldots, X_\infty)$).
   - Loi de $(X_{1:N}, N)$ donnée par $N \sim \mathcal{P}(n)$, et
$$\mathbb{P}\left(X_{1:N} = x_{1:k} \mid N = j\right) = \begin{cases} \prod_{i=1}^{k} p_{x_i} & \text{si} \quad k = j \\ 0 & \text{sinon.} \end{cases}$$
2. Soit $k \in \mathbb{N}$, $E \mid N = k$ suit une loi multinomiale $\mathcal{M}(k, \mathbf{p})$, c'est à dire
$$\mathbb{P}\left(E = e_{1:d} \mid N = k\right) = \frac{k!}{e_1! \times \ldots \times e_d!} \prod_{j=1}^{d} p_j^{e_j} \mathbb{1}_{\sum_{j=1}^d e_j = k}.$$
   On en déduit
$$\begin{aligned}\mathbb{P}\left(E = e_{1:d}\right) &= \sum_{k=1}^{+\infty} e^{-n} \frac{n^k}{k!} \frac{k!}{e_1! \times \ldots \times e_d!} \prod_{j=1}^{d} p_j^{e_j} \mathbb{1}_{\sum_{j=1}^d e_j = k} \\ &= \sum_{k=0}^{+\infty} \prod_{j=1}^{d} \frac{(np_j)^{e_j}}{e_j!} e^{-np_j} \mathbb{1}_{\sum_{j=1}^d e_j = k} \\ &= \prod_{j=1}^{d} \frac{(np_j)^{e_j}}{e_j!} e^{-np_j}.\end{aligned}$$
   On en déduit que les E_j sont indépendantes, de loi de Poisson $\mathcal{P}(np_j)$.
3. Soit $u \in \mathbb{R}^d$.
$$\Phi_{Z_n}(u) = \mathbb{E}(e^{i\langle u, Z_n\rangle}) = \mathbb{E}\left(\prod_{j=1}^{d} e^{iu_j Z_{n,j}}\right) = \prod_{j=1}^{d} \mathbb{E}\left(e^{iu_j Z_{n,j}}\right) \to \prod_{j=1}^{d} \mathbb{E} e^{iu_j Z_j} = \mathbb{E} e^{i\langle u, Z\rangle},$$
   où $Z = (Z_1, \ldots, Z_d)^T$, avec les Z_j indépendantes.
4. On se place sous H_0. Soit $j \in \{1, \ldots, d\}$, on peut écrire
$$E_j = \sum_{i=1}^{n} Y_i,$$
   pour des Y_i i.i.d. de loi $\mathcal{P}(p_{0,j})$. Le théorème central limite donne alors (on a bien $\mathbb{E}Y_1^2 = p_1 + p_1^2 < +\infty$),
$$\frac{\sqrt{n}}{\sqrt{p_{0,j}}}\left(\frac{E_j}{n} - p_{0,j}\right) \rightsquigarrow \mathcal{N}(0, 1),$$

ce dont on déduit que, pour tout j,

$$Z_{n,j} := \frac{(E_j - np_{0,j})^2}{np_{0,j}} \rightsquigarrow \chi^2(1).$$

Les $(Z_{n,j})_j$ étant indépendantes, de la partie précédente on déduit que

$$S := \sum_{j=1}^{d} \frac{(E_j - np_{0,j})^2}{np_{0,j}} \rightsquigarrow \chi^2(d).$$

Un test de niveau α pour $H_0 : \mathbf{p} = \mathbf{p_0}$ est alors

$$\mathbb{1}_{S \geq q_{1-\alpha}},$$

où $q_{1-\alpha}$ est le $1-\alpha$-quantile d'une $\chi^2(d)$.

5. Dans le cas où n est inconnu (mais supposé tendant vers $+\infty$), on va essayer de remplacer n par N. On se place encore sous H_0, et commence par regarder

$$Z = (E_j - Np_{0,j})_{j=1,\ldots,d} = (I_d - \mathbf{p}_0 \mathbf{1}^T)E,$$

où E est le vecteur des effectifs. En remarquant que $(I_d - \mathbf{p}_0\mathbf{1}^T)\mathbb{E}(E) = (I_d - \mathbf{p}_0\mathbf{1}^T)((np_{0,j})_{j=1,\ldots,d}) = 0$, on en déduit que

$$\frac{1}{\sqrt{n}}Z = \frac{1}{\sqrt{n}}(I_d - \mathbf{p}_0\mathbf{1}^T)(E - \mathbb{E}(E)) \rightsquigarrow (I_d - \mathbf{p}_0\mathbf{1}^T)\mathcal{N}(0, Diag(\mathbf{p}_0)).$$

Un calcul matriciel donne

$$\begin{aligned}
(I_d - \mathbf{p}_0\mathbf{1}^T)Diag(\mathbf{p}_0)(I_d - \mathbf{1}\mathbf{p}_0^T) &= (Diag(\mathbf{p}_0) - \mathbf{p}_0\mathbf{p}_0^T)(I_d - \mathbf{1}\mathbf{p}_0^T) \\
&= Diag(\mathbf{p}_0) - \mathbf{p}_0\mathbf{p}_0^T - \mathbf{p}_0\mathbf{p}_0^T + \mathbf{p}_0\mathbf{p}_0^T\mathbf{1}\mathbf{p}_0^T \\
&= Diag(\mathbf{p}_0) - \mathbf{p}_0\mathbf{p}_0^T.
\end{aligned}$$

Comme pout un test du Chi-deux d'adéquation classique, on en déduit alors que

$$Diag((\mathbf{p}_0)^{-1/2})\frac{1}{\sqrt{n}}Z \rightsquigarrow \mathcal{N}(0, \pi_{\sqrt{\mathbf{p}_0}^\perp}),$$

et donc

$$\frac{1}{n}\left\|Diag((\mathbf{p}_0)^{-1/2})Z\right\|^2 \rightsquigarrow \chi^2(d-1),$$

en utilisant Cochran. Reste un dernier terme en n à faire disparaître. Pour cela on remarque que $\frac{N}{n} \to 1$ en probabilité (loi des grands nombres), et une application du Lemme de Stlusky donne

$$\frac{1}{N}\left\|Diag((\mathbf{p}_0)^{-1/2})Z\right\|^2 \rightsquigarrow \chi^2(d-1).$$

En prenant comme statistique de test

$$\widetilde{S} = \sum_{j=1}^{d} \frac{(E_j - Np_{0,j})^2}{Np_{0,j}},$$

un test de niveau asymptotique α est alors $\mathbb{1}_{\widetilde{S} \geq q_{1-\alpha}}$, où $q_{1-\alpha}$ est le $1-\alpha$ quantile d'une $\chi^2(d-1)$.

Correction Exo 13

Comme $\mathrm{Cov}(AX) = AA^T$, si $A \in O_d(\mathbb{R})$, $AX \sim X$. Il suffit alors d'appliquer le résultat de l'exercice 2 (un peu plus général), en prenant pour A des rotations pour deux coordonnées laissant invariant les autres coordonnées.

Correction Exo 14

On commence par le cas $d = 2$. On prend alors comme vecteurs orthogonaux $a = (1, 1)$ et $b = (1, -1)$, et on peut alors écrire, pour tous $t_1, t_2 \in \mathbb{R}$,

$$\begin{aligned}\mathbb{E}(\exp(i\,(t_1X_1 + t_1X_2 + t_2X_1 - t_2X_2)) &= \mathbb{E}\left(\exp(i\,(t_1X_1 + t_1X_2)\right)\mathbb{E}\left(\exp(i\,(t_2X_1 - t_2X_2)\right)\\ &= \varphi_1(t_1)\varphi_2(t_1)\varphi_1(t_2)\varphi_2(-t_2),\end{aligned}$$

où φ_j est la fonction caractéristique de X_j. Par ailleurs, en exploitant directement l'indépendance des X_j, on a

$$\mathbb{E}(\exp(i\,(t_1X_1 + t_1X_2 + t_2X_1 - t_2X_2)) = \varphi_1(t_1 + t_2)\varphi_2(t_1 - t_2),$$

ce dont on déduit l'équation

$$\varphi_1(t_1 + t_2)\varphi_2(t_1 - t_2) = \varphi_1(t_1)\varphi_2(t_1)\varphi_1(t_2)\varphi_2(-t_2). \tag{1}$$

Les choix $t_1 = t_2 = t$ et $t = t_1 = -t_2$ donnent le système

$$\begin{aligned}\varphi_1(2t) &= \varphi_1(t)^2\varphi_2(t)\varphi_2(-t),\\ \varphi_2(2t) &= \varphi_2(t)^2\varphi_1(t)\varphi_1(-t).\end{aligned}$$

Par une récurrence immédiate, on en déduit que, pour tout $n \geq 1$ et $t \in \mathbb{R}$,

$$\varphi_1(t) = \prod_{j=1}^{4^n} \varphi_{\varepsilon_j}(\sigma_j 2^{-n} t),$$

où ε et σ sont des suites de 4^n éléments à valeurs dans $\{1, 2\}$ et $\{-1, 1\}$. Par continuité de φ_1 en 0, on peut trouver $r > 0$ tel que, pour tout $|t| < r$, $|\varphi(t) - 1| < 1$. Pour de tels t on peut prendre les log (complexes), qu'on note ℓ_j. Cela donne

$$\ell_1(t) = \sum_{j=1}^{4^n} \ell_{\varepsilon_j}(\sigma_j 2^{-n} t).$$

Or, pour $j \in [\![1; 4^n]\!]$,

$$\ell_{\varepsilon_j}(\sigma_j 2^{-n} t) = \log(1 - \frac{t^2}{2}4^{-n} + o_n(4^{-n}) = -\frac{t^2}{2}4^{-n} + o_n(4^{-n}),$$

en utilisant le fait que X est centrée et de variance 1. On en déduit que

$$\ell_1(t) = -\frac{t^2}{2} + o_n(1) = -\frac{t^2}{2}.$$

On a donc montré que pour tout t dans $]-r, r[$, $\varphi_1(t) = \exp(-t^2/2)$. Notons R le rayon maximal pour lequel cela arrive, c'est à dire $R = \sup\{r > 0 \mid \forall t \in]-r; r[\quad |\varphi_1(t) - 1| < 1\}$. Si $R < +\infty$, comme $|\exp(-R^2/2) - 1| < 1$, par continuité de φ_1 autour de $\pm R$ on peut trouver ε tel que $\forall |t| < R + \varepsilon$ $|\varphi_1(t) - 1| < 1$, ce qui contredit la définition de R. On en déduit que $R = +\infty$, et donc que $X_1 \sim \mathcal{N}(0, 1)$. Pareillement on déduit $X_2 \sim \mathcal{N}(0, 1)$.

Pour le cas d quelconque, il suffit de regarder les vecteurs de type $e_i + e_j$ et $e_i - e_j$, où les e_i sont les vecteurs de la base canonique.

Remarque: On peut prouver le Théorème de Darmois Skitovitch comme ceci, en remarquant que $\varphi_1(t)$ se met sous la forme $\varphi_1(t) = \prod_{j=1}^{4} \varphi_{\varepsilon_j}(\sigma_j a_j t)$, avec $\sum_{j=1}^{4} a_j^2 = 1$.