Semestre 4, Année 2011-2012 SESSION 2

UE N1MA6031 : Probabilités et Statistique

Date : 22 Juin 2012, 14h00-17h00 Durée : 3h00

Polycopié de cours autorisé, à l'exclusion de tout autre document.

Exercice 1. *Sur une autoroute, les statistiques officielles révèlent qu'un conducteur sur 1000 ne respecte pas la limitation de vitesse à 130 km/h. Un radar, placé sur cet autoroute, repère l'excès de vitesse chez* 99% *des conducteurs fautifs (en les flashant). Cependant, du fait de défaillances ponctuelles du système,* 0.3% *de conducteurs en règle se trouvent aussi flashés. Calculer la probabilité pour qu'un conducteur se retrouvant flashé le soit effectivement à juste titre.*

On prend comme univers des possibles l'ensemble de tous les conducteurs, avec comme disribution de probabilité la distribution uniforme. La probabilité de l'évènement NR (un conducteur pris au hasard ne respecte pas la limitation de vitesse) est $P(NR) = 10^{-3}$. Si F désigne l'évènement « un conducteur pris au hasard est flashé », on a $P(F \mid NR) = 0.99$. D'autre part, si R désigne l'évènement « un conducteur pris au hasard respecte la limitation de vitesse », on a $P(F \mid R) = 3/100 = 3 \times 10^{-3}$. D'après la formule des causes (Proposition 1.3 du cours), on a

$$P(NR \mid F) = \frac{P(F \mid NR)\, P(NR)}{P(F \mid NR)\, P(F) + P(F \mid R)\, P(R)} = \frac{0.99}{0.99 + 3 \times 0.999} \simeq 1/4.$$

Surprenant, non ?

Exercice 2.

1. *Soient X_1 et X_2 deux VAR indépendantes suivant l'une une loi de Poisson de paramètre $\lambda_1 > 0$, l'autre une loi de Poisson de paramètre $\lambda_2 > 0$. Montrer que $X_1 + X_2$ suit encore une loi de Poisson dont on précisera le paramètre.*

La fonction génératrice de X_1 est $s \mapsto e^{\lambda_1 s}$, celle de X_2 est $s \mapsto e^{\lambda_2 s}$. Comme X_1 et X_2 sont supposées indépendantes, la fonction génératrice de la somme est le produit des fonctions génératrices, soit $s \mapsto e^{(\lambda_1+\lambda_2)s}$. On reconnait là la fonction génératrice d'une loi de Poisson de paramètre $\lambda_1 + \lambda_2$. La loi de $X_1 + X_2$ est donc une loi de Poisson $\mathscr{P}(\lambda_1 + \lambda_2)$.

2. *Soit $(X_k)_{k\in\mathbb{N}^*}$ une suite de VAR mutuellement indépendantes toutes de même loi, à savoir la loi de Poisson de paramètre $\lambda > 0$. Quelle est la loi de la VAR*

$$X_{[n]}^{\text{emp}} : \frac{X_1 + \cdots + X_n}{n} \quad ?$$

Que valent espérance et écart-type de cette VAR $X_{[n]}^{\rm emp}$?
Comme les variables X_j, $j = 1, ..., n$, sont mutuellement indépendantes, la fonction génératrice de la VAR $X_{[n]}^{\rm emp}$ est le produit des fonctions génératrices des VAR X_j/n. Ces VAR sont toutes de même loi, la loi de Poisson $\mathscr{P}(\lambda/n)$, et chacune de ces fonctions génératrices est $s \mapsto e^{\lambda s/n}$. La fonction génératrice de $X_{[n]}^{\rm emp}$ est donc $s \mapsto (e^{\lambda s/n})^n = e^{-\lambda s}$. La VAR $X_{[n]}^{\rm emp}$ suit donc une loi de Poisson $\mathscr{P}(\lambda)$. Son espérance vaut λ, ainsi que sa variance (d'après le cours).
3. *Vers quelle VAR la suite de VAR*

$$\Big(\sqrt{\frac{n}{\lambda}}(X_{[n]}^{\rm emp} - \lambda)\Big)_{n\geq 1}$$

converge t-elle en loi lorsque n tend vers $+\infty$?
D'après le Théorème Limite Central (Théorème 1.1 du cours), cette suite converge en loi vers une VAR suivant une loi $\mathscr{N}(0,1)$.
4. *On suppose N grand. Déterminer un seuil x tel que*

$$P\Big(\Big\{\lambda \in \Big]X_{[N]}^{\rm emp} - x\sqrt{\frac{\lambda}{N}}, X_{[N]}^{\rm emp} + x\sqrt{\frac{\lambda}{N}}\Big[\Big\}\Big) \geq 95\%$$

(on exploitera la table donnée en annexe).
Ce seuil vaut $x = 1.96$ d'après la table donnée en annexe.
5. *Dans une entreprise, on constate au final, sur une durée de deux ans, qu'il y a eu 20 accidents du travail à déplorer au sein du personnel. Pourquoi la loi de Poisson est elle un modèle raisonnable pour modéliser la loi de la VAR X égale au nombre d'accidents susceptibles de survenir par jour ? Si l'on suppose que l'on choisit ce modèle, et donc que X suit une telle loi, déterminer (en utilisant le résultat établi à la question* **(4)***) un intervalle de confiance pour la moyenne λ du nombre d'accidents journaliers, au niveau de confiance* 95%.
Comme la loi de Poisson peut être considérée heuristiquement comme la « loi des évènements rares », c'est un modèle naturel pour modéliser la loi de la VAR X. Sur une durée de $365 \times 2 = 730$ jours, la réalisation de $X_{[730]}^{\rm emp}$ est $2/73$. Les valeurs extrèmes de λ telles que

$$\lambda \in \Big]2/73 - 1.96\sqrt{\frac{\lambda}{730}}, 2/73 + 1.96\sqrt{\frac{\lambda}{730}}\Big[$$

sont données en cherchant les racines positives des deux équations du second degré en $\sqrt{\lambda}$:

$$(2/73 - \lambda) = +1.96\sqrt{\lambda}/\sqrt{730}, \quad (2/73 - \lambda) = -1.96\sqrt{\lambda}/\sqrt{730}.$$

L'unique racine (positive) de la première équation est $\sqrt{\lambda} \simeq 0.1332$, tandis que celle de la seconde équation est $\sqrt{\lambda} \simeq 0.2057$. La réalisation de l'intervalle de confiance proposée est donc

$$](0.133)^2, (0.206)^2[=]0.017, 0.043[.$$

Exercice 3.

Soit $p \in \mathbb{N}^*$, $c > 0$, et $f_{p,c}$ la fonction

$$f_{p,c} \ : t \in \mathbb{R} \longmapsto c\, \chi_{[0,\infty[}(t)\, \exp(-t/2)\, t^{p-1}, \quad c > 0.$$

1. *Comment faut-il choisir la constante $c = c_p$ pour que f_{p,c_p} soit la densité d'une loi de VAR ?*

On a

$$\int_0^\infty t^{p-1} e^{-t/2}\, dt = 2^p \int_0^\infty u^{p-1} e^{-u}\, du = 2^p\, (p-1)!\,.$$

On doit donc, pour que l'intégrale sur $\mathbb{R}$ de f_{p,c_p} vaille 1, choisir

$$c_p = 2^{-p}/(p-1)!\,.$$

2. *Une fois que c_p a été choisie comme au* **(1)**, *que valent espérance et variance de la VAR X_p de densité f_{p,c_p} ?*

On a

$$\int_0^\infty t^p e^{-t/2}\, dt = 2^{p+1} \int_0^\infty u^p e^{-u}\, du = 2^{p+1}\, p!.$$

L'espérance de X_p vaut donc $c_p \times 2^{p+1} \times p! = 2p$. D'autre part

$$\int_0^\infty t^{p+1} e^{-t/2}\, dt = 2^{p+2} \int_0^\infty u^{p+1} e^{-u}\, du = 2^{p+2}\, (p+1)!.$$

Donc $E[X_p^2] = c_p \times 2^{p+2}\, (p+1)^2 = 4\, p(p+1)$. On en déduit donc

$$\mathrm{Var}\,(X_p) = 4p(p+1) - 4p^2 = 4p.$$

3. *Soit $\epsilon > 0$. Donner en fonction de p une majoration de*

$$P\big(\{|X_p - 2p| \geq \epsilon\}\big).$$

Pour quelles valeurs de ϵ une telle inégalité est-elle intéressante ?

Cette majoration est donnée par l'inégalité de Bienaymé-Tchebychev (Proposition 2.1 du cours), qui donne :

$$P\big(\{|X_p - 2p| \geq \epsilon\}\big) \leq \frac{\mathrm{Var}(X_p)}{\epsilon^2} = \frac{4p}{\epsilon^2}.$$

Cette inégalité n'est intéressante que si $\epsilon > \sigma(X_p) = 2\sqrt{p}$.

4. *Soient p_1 et p_2 deux entiers strictement positifs. Vérifier la formule :*

$$\forall t \in \mathbb{R}, \qquad \int_{\mathbb{R}} f_{p_1,c_{p_1}}(s)\, f_{p_2,c_{p_2}}(t-s)\, ds = f_{p_1+p_2,c_{p_1+p_2}}(t).$$

Déduire de ce calcul la densité de la loi de $X_{p_1} + X_{p_2}$ lorsque X_{p_1} et X_{p_2} sont supposées être deux VAR indépendantes, de densités respectives $f_{p_1,c_{p_1}}$ et $f_{p_2,c_{p_2}}$.

Le membre de droite de la formule vaut 0 si $t < 0$ et, si $t \geq 0$:

$$c_{p_1}\, c_{p_2} e^{-t} \int_0^t s^{p_1-1}(t-s)^{p_2-1}\, ds = c_{p_1}\, c_{p_2} t^{p_1+p_2-1} e^{-t} \int_0^1 u^{p_1-1}(1-u)^{p_2-1}\, du. \tag{$*$}$$

En intégrant par parties, on trouve

$$\int_0^1 u^{p_1-1}(1-u)^{p_2-1}\, du = \frac{(p_1-1)!\ (p_2-1)!}{(p_1+p_2-1)!}.$$

En reportant dans $(*)$, on retrouve bien la valeur de $f_{p_1+p_2,c_{p_1+p_2}}(t)$. La densité de la loi de $X_{p_1} + X_{p_2}$ est, puisque les deux variables sont indépendantes, la fonction

$$t \in \mathbb{R} \longmapsto \int_{\mathbb{R}} f_{p_1,c_{p_1}}(s)\, f_{p_2,c_{p_2}}(t-s)\, ds.$$

En effet, au niveau infinitésimal

$$P\Big(\{X_{p_1} \in [s, s+ds]\} \cap \{X_{p_2} \in [t-s, t-s+dt]\}\Big) = f_{p_1}(s) \times f_{p_2}(t-s) ds dt$$

du fait de l'indépendance (les probabilités des deux évènements se multiplient). La densité de la loi de $X_{p_1} + X_{p_2}$ est donc exactement celle de $X_{p_1+p_2}$.

Exercice 4.

Un surfeur pianote sur le net (il y a un nombre fini N de pages répertoriées, numérotées de 1 à N). S'il consulte à l'instant $t = k$ la page i, la probabilité qu'il aille à l'instant $t = k+1$ consulter la page j ne dépend pas de k et vaut un certain nombre $a_{i,j} \in [0,1]$.

1. *On note X_k la VAR désignant le numéro de la page que consulte le surfeur à l'instant k. Vérifier que l'on a :*

$$\forall\, k \in \mathbb{N}, \quad P(X_{k+1} = j) = \sum_{i=1}^{N} a_{i,j} P(X_k = i).$$

On peut appliquer la formule de Bayes (Proposition 1.2 du cours), car $a_{i,j}$ s'interprète comme la probabilité (conditionnelle) que le surfeur aille consulter la page j à l'instant $k+1$, conditionnée par l'évènement : « le surfeur consulte la page i à l'instant k ».

2. *On note A la matrice $[a_{i,j}]_{1 \leq i,j \leq N}$ (i : indice de ligne, j : indice de colonne). Que peut-on dire de la somme des éléments de chaque ligne de la matrice A ? Calculer en fonction de A la loi de probabilité de la VAR X_k (on suppose $P(\{X_1 = 1\}) = 1$).*

La somme des éléments de la matrice A vaut 1 puisque l'on peut interpréter le vecteur ligne des entrées de A figurant sur la ligne i comme une distribution de probabilité (conditionnelle) sur $\{1, ..., n\}$ (à condition toutefois que la probabilité que cette page i soit visitée par X_k ne soit pas nulle pour toutes les valeurs de k). Comme $P(\{X_1 = 1\}) = 1$, on a, en tenant compte des formules inductives établies à la question 1 :

$$L_k = [1, 0, ..., 0] * A^{k-1}$$

si L_k désigne le vecteur ligne correspondant à la distribution de probabilité de la VAR X_k sur $\{1, ..., N\}$.

Annexe

Quantiles de dépassement de l'écart absolu de la loi normale $\mathcal{N}(0,1)$

$$\mathbb{P}(|X| > z_\alpha) = \alpha$$

(par exemple : si $\alpha = 0.13 + 0.005$, alors $z_\alpha = 1.495$)

α	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009
0.00	∞	3.291	3.090	2.968	2.878	2.807	2.748	2.697	2.652	2.612
0.01	2.576	2.543	2.512	2.484	2.457	2.432	2.409	2.387	2.366	2.346
0.02	2.326	2.308	2.290	2.273	2.257	2.241	2.226	2.212	2.197	2.183
0.03	2.170	2.157	2.144	2.132	2.120	2.108	2.097	2.086	2.075	2.064
0.04	2.054	2.044	2.034	2.024	2.014	2.005	1.995	1.986	1.977	1.969
0.05	1.960	1.951	1.943	1.935	1.927	1.919	1.911	1.903	1.896	1.888
0.06	1.881	1.873	1.866	1.859	1.852	1.845	1.838	1.832	1.825	1.818
0.07	1.812	1.805	1.799	1.793	1.787	1.780	1.774	1.768	1.762	1.757
0.08	1.751	1.745	1.739	1.734	1.728	1.722	1.717	1.711	1.706	1.701
0.09	1.695	1.690	1.685	1.680	1.675	1.670	1.665	1.660	1.655	1.650
0.10	1.645	1.640	1.635	1.630	1.626	1.621	1.616	1.612	1.607	1.603
0.11	1.598	1.594	1.589	1.585	1.580	1.576	1.572	1.567	1.563	1.559
0.12	1.555	1.551	1.546	1.542	1.538	1.534	1.530	1.526	1.522	1.518
0.13	1.514	1.510	1.506	1.502	1.499	1.495	1.491	1.487	1.483	1.480
0.14	1.476	1.472	1.468	1.465	1.461	1.457	1.454	1.450	1.447	1.443
0.15	1.440	1.436	1.433	1.429	1.426	1.422	1.419	1.415	1.412	1.408
0.16	1.405	1.402	1.398	1.395	1.392	1.388	1.385	1.382	1.379	1.375
0.17	1.372	1.369	1.366	1.363	1.359	1.356	1.353	1.350	1.347	1.344
0.18	1.341	1.338	1.335	1.332	1.329	1.326	1.323	1.320	1.317	1.314
0.19	1.311	1.308	1.305	1.302	1.299	1.296	1.293	1.290	1.287	1.284
0.20	1.282	1.279	1.276	1.273	1.270	1.267	1.265	1.262	1.259	1.256