RAPPORT SUR L'EPREUVE DE PROBABILITES ET STATISTIQUES

Le problème développait des techniques d'approximation poissonnienne dans la lignée des travaux de Le Cam (1960) et Serfling (1978). Un exemple d'application était joint, concernant l'étude de certaine propriétés des records (Rényi (1962)).

Aucune difficulté sérieuse n'était présente dans l'énoncé, et il aurait été à la rigueur possible à un étudiant n'ayant que des bases sommaires en probabilités d'arriver à obtenir une note honorable au prix d'un effort minimal.

Dans l'ensemble, les résultats ont été malgré tout assez décevants. Il semble surtout que les candidats manquent de "savoir faire" sur des questions, somme toute, d'abord relativement élémentaire .

La leçon à retenir devrait être la suivante. Au niveau de l'agrégation, il n'est pas indispensable de rédiger un problème excessivement sophistiqué pour arriver à classer convenablement les candidats de l'option probabilités et statistique. Dans l'attente hypothétique d'un relèvement massif du niveau général des étudiants, il serait sans doute préférable d'adopter un profil assez bas pour les épreuves à venir.

Il est clair que les probabilités et la statistique sont des disciplines complexes nécessitant dès le départ un excellent niveau en analyse. De ce fait, on ne peut s'attendre à des miracles, et il serait sans doute utile de vérifier si les candidats savent calculer $P(A|B)$ en fonction de $P(A \cap B)$ et $P(B)$, avant de supposer bien connu le fait que l'espérance conditionnelle est un projecteur de L_p (par exemple).

Les commentaires de détail joints au corrigé résumé du problème donneront une idée des résultats de l'épreuve.

Corrigé résumé du problème

Première Partie

1°) $P(U \in D)-P(V \in D) = a+b$, où $a = P(U \in D \cap A)-P(V \in D \cap A) \leq 0$, et
$b = P(U \in D \cap C)-P(V \in D \cap C) \geq 0$.

Par conséquent, $-a$ est minimum pour $D \subset A$, de même que b pour $D \subset C$. De plus,

$$P(U \in C)-P(V \in C) = P(V \in A)-P(U \in A) \geq 0.$$

Donc nécessairement

$$d(L(U),L(V)) = P(U \in C)-P(V \in C) = P(V \in A)-P(U \in A) = \frac{1}{2}\sum_k |P(U=k)-P(V=k)|.$$

De l'identité $|x-y|=x+y-2\min(x,y)$ et de $\sum_k P(U=k) = \sum_k P(V=k) = 1$, on déduit

$$d(L(U),L(V)) = 1 - \sum_k \min(P(U=k),P(V=k)).$$

Commentaire

Déjà dans cette question, on observe un grand nombre de réponses incomplètes, consistant à n'établir que des inégalités en lieu et place des égalités demandées.

2°) Pour tout k, $P(U=k)-P(V=k)=P(\{U\neq V\}\cap\{U=k\})-P(\{U\neq V\}\cap\{V=k\})$. Par suite

$$\sum_k |P(U=k)-P(V=k)| \leq \sum_k (P(\{U\neq V\}\cap\{U=k\})+P(\{U\neq V\}\cap\{V=k\}) = 2P(U\neq V),$$

d'où $d(L(U),L(V)) \leq P(U\neq V)$ par le 1°.

Commentaire

Une quantité considérable de réponses fausses, parmi lesquelles on remarque entr autres, des affirmations telles que $P(\{U\neq V\})=P(U\in A)-P(V\in A)$ (par exemple), qui r vent de la plus haute fantaisie.

3°)a) Il suffit de vérifier que $R_{ij}\geq 0$ et que $\sum_j R_{ij} = P_i$, $\sum_i R_{ij} = Q_j$, et $\sum_{i,j} R_{i.}$

b) En se servant de la question précédente, on peut définir $p_{k\ell}$, $k\in C$, $\ell\in A$ par

$$p_{k\ell} = \frac{(P(U=k)-P(V=k))(P(V=\ell)-P(U=\ell))}{d(L(U),L(V))}\text{ , si } d(L(U),L(V))\neq 0\text{, et } p_{k\ell} = 0 \text{ autremen}$$

c) Pour la construction précédente, $P(U=V=k)=\min(P(U=k),P(V=k))$ pour tout k, et donc $P(U\neq V)= 1 - \sum_k P(U=V=k) = 1 - \sum_k \min(P(U=k),P(V=k)) = d(L(U),L(V))$ par le 1°.

Commentaire

A partir du moment où l'on sait qu'une probabilité sur un ensemble dénombrable se ramène à la donnée d'une série de termes non négatifs de somme totale un, on devra être en mesure de résoudre cette question. Le résultat a été très mauvais dans l'ense ble.

Par ailleurs, certains candidats ont tenu à vérifier au 3°a que les R_{ij} vérifiaie tous les axiomes des probabilités. On ne leur en demandait pas tant.

4°)a) $d(L(U),L(V)) = \frac{1}{2}\sum_{k=0} |P(U=k)-P(V=k)| = \frac{1}{2}\{|1-p-e^{-p}|+|p-pe^{-p}|+(1-e^{-p}-pe^{-p})\}$.

Il suffit alors de remarquer que $e^{-p}-1+p \geq 0$, tandis que $e^{-p} \leq 1$ pour en déduire

$$d(L(U),L(V)) = p(1-e^{-p}) \leq p^2.$$

b) Posons V=UW où U et W sont indépendants. On en déduit par des calculs évider que

$$P(W=0) = \frac{e^{-p}-1+p}{p} \geq 0\text{, tandis que } P(W=k) = \frac{p^{k-1}e^{-p}}{k!}\text{ pour } k\geq 1.$$

On vérifie que $\sum_k P(W=k) = 1$, ce qui rend possible la construction par produits.

Enfin $P(U\neq V) = P(U=1)P(W\neq 1) = p(1-e^{-p}) = d(L(U),L(V))$.

Commentaire

Beaucoup de candidats n'aboutissent pas, ne sachant vérifier que $1-p-e^{-p} \geq 0$, ou se trompant dans l'évaluation du signe de cette expression. Pour le reste, la notion de produit de mesures semble mal comprise ...

5°)a) T_n suit une loi de Poisson d'espérance $\sum_{i=1}^{n} p_i$ (cours, fonctions caractéristiques, ou vérification directe).

b) On répète la construction du 4° pour tout n, de manière à avoir une suite de couples (X_i,Y_i), $i=1,2,\ldots$, indépendants et tels que $d(L(X_i),L(Y_i))=P(X_i \neq Y_i)$. On a alors

$$d(L(S_n),L(T_n)) \leq P(S_n \neq T_n) = 1-P(S_n=T_n) \leq 1-P(\bigcap_{i=1}^{n} \{X_i=Y_i\}) = 1-\prod_{i=1}^{n}(1-P(X_i \neq Y_i)) =$$

$$= 1-\prod_{i=1}^{n}(1-d(L(X_i),L(Y_i)) \leq \sum_{i=1}^{n} d(L(X_i),L(Y_i)) \leq \sum_{i=1}^{n} p_i(1-e^{-p_i}) \leq \sum_{i=1}^{n} p_i^2.$$

Commentaire

Les candidats connaissent mal les propriétés de la loi de Poisson (au programme) et tentent parfois vainement de vérifier le a au prix de pages de calculs faux. Pour le reste, des inégalités usuelles telles

$\prod_{i=1}^{n}(1-u_i) \geq 1 - \sum_{i=1}^{n} u_i$ semblent au delà de la portée du candidat moyen,

malgré les indications transparentes de l'énoncé.

Deuxième Partie

1°) $E(N(m,n)) = \sum_{i=m+1}^{n} \frac{1}{i} = \log(\frac{n}{m}) + o(1)$, uniformément en $\underline{n<m<N}$ lorsque $N \to \infty$.

$$V(N(m,n)) = \sum_{i=m+1}^{n} \frac{1}{i} - \sum_{i=m+1}^{n} \frac{1}{i^2}, \text{ et } \sum_{i=m+1}^{n} \frac{1}{i^2} \leq \sum_{i=N+1}^{\infty} \frac{1}{i^2} \to 0 \text{ pour } \underline{n \geq m \geq} N \to \infty.$$

On se ramène donc au cas précédent.

Commentaire

Question particulièrement aisée, destinée à sauver de la note zéro les candidats désarçonnés par le reste du problème. Malgré tout, nombre de candidats n'ont pas vu cette question, et une énorme proportion s'avère incapable de calculer la variance de $N(m,n)$, écrivant que la variance de X_i est $1/i^2$, au lieu de $(1-1/i)/i$.

2°)a) $E(\exp(iuN(m,n)) = \prod_{j=m+1}^{n} \{1+\frac{1}{j}(e^{iu}-1)\} = \exp(\sum_{j=m+1}^{n} \log(1+\frac{1}{j}(e^{iu}-1)))$.

Lorsque $T \to \infty$, si $m=[Te^s]$ et $n=[Te^t]$, $n/m \to e^{t-s}$ et $\sum_{j=m+1}^{n} \frac{1}{j}(e^{iu}-1) \to (t-s)(e^{iu}-1)$, tandis que $\sum_{j=m+1}^{n} (\frac{1}{j}(e^{iu}-1))^2 \to 0$. La limite de la fonction caractéristique de $M_T(s,t)$ est donc la fonction caractéristique d'une loi de Poisson d'espérance t-s.

b) Les v.a. considérées étant indépendantes, on obtient comme loi limite un produit de lois de Poisson d'espérances s_1, s_2-s_1, ...

Commentaire

Bien qu'un peu plus technique, cette question restait simple dans la mesure où elle ne faisait appel qu'à des propriétés simples du logarithme. A la correction, on est surpris de l'aversion qu'éprouvent les candidats à la mani lation des logarithmes complexes. Certains d'entre eux vont jusqu'à décrire leur raisonnement comme un calcul formel, qui "serait vrai si on avait le droit d'écrire" le logarithme d'un nombre complexe. Les lacunes ne viennent pas ici des probabilités

c) On a

$$E(\exp\{iu \frac{S_n - \log n}{\sqrt{\log n}}\}) = \exp\{\sum_{j=1}^{n} \log\{1+ \frac{1}{j}(\exp(\frac{iu}{\sqrt{\log n}})-1\} - iu\sqrt{\log n}\} =$$

$$= \exp\{\sum_{j=1}^{n} \{\frac{1}{j} (\frac{iu}{\sqrt{\log n}} - \frac{u^2}{2\log n} (1+o(1))) + \frac{O(1)}{\log n} - iu\sqrt{\log n}\} =$$

$$= \exp\{(\log n + \gamma + o(1))(\frac{iu}{\sqrt{\log n}} - \frac{u^2}{2\log n} (1+o(1)) + o(1) - iu\sqrt{\log n}\} =$$

$$= \exp\{-\frac{1}{2} \frac{u^2}{\log n} + o(1)\}$$ pour $n\to\infty$, d'où le résultat.

d) $\lim_{n\to\infty} (E(S_n)-\log n) = \gamma$, $\lim_{n\to\infty} (V(S_n)-\log n) = \gamma - \frac{\pi^2}{6}$.

Commentaire

Tandis que le c n'est traité que par une toute petite élite, le d est fait dans l'ensemble correctement, malgré des erreurs dues à des candidats qui tentent de ramen le d au c.

3°)a) Par les I,4° et 5°, il est possible de construire une suite $\{W_n, n\geq 1\}$ de v.a indépendantes des X_i et mutuellement indépendantes (extension de l'espace par produit de manière à obtenir le résultat cherché en posant $Y_i = W_i X_i$.

b) T_n suit une loi de Poisson d'espérance $\sum_{1\leq i\leq n} (1/i)$, lorsque les Y_i sont choi: indépendants.

Comme $P(X_i \neq Y_i) \leq 1/i^2$, et donc que $\sum_{i=1}^{\infty} P(X_i\neq Y_i) < \infty$, on conclut par Borel-Cante

Commentaire

Très peu de candidats semblent connaître le lemme de Borel-Cantelli (ou être en mesure de l'utiliser).

4°)a) $$P(\xi=m, \eta=n | \zeta=m+n) P(\zeta=m+n) = \frac{(m+n)!}{m!n!} \theta^m (1-\theta)^n \frac{\alpha^{m+n}}{(m+n)!} e^{-\alpha} =$$

$$\{\frac{(\alpha\theta)^m}{m!} e^{-\alpha\theta}\}\{\frac{(\alpha(1-\theta))^n}{n!} e^{-\alpha(1-\theta)}\} .$$

Dans cette dernière expression, on reconnaît un produit de probabilités. Ainsi ξ et η suivent des lois de Poisson indépendantes d'espérances $\alpha\theta$ et $\alpha(1-\theta)$.

b) On procède par récurrence sur n en utilisant le a.

Commentaire

Cette question a été souvent abordée. Bien peu ont réalisé l'indépendance de ξ et η.

c) On applique la loi des grands nombres à $\sum_{i=1}^{L(n)} Z_i \sim L(n)$ p.s., d'où le résultat par les encadrements du b.

Troisième Partie

1°) L'ensemble des $\{\omega_1,\dots,\omega_n\}$ non touts distincts est de mesure de Lebesgue nulle.

Commentaire

Question de cours facile, bonnes réponses en général.

2°)a) $D_n(t)$ suit une loi binômiale $B(t,n)$.

b) $P(\omega_{i,n}\leq t) = P(D_n(t)\geq i) = \sum_{j=i}^{n} \binom{n}{j} t^j(1-t)^{n-j}$, $0<t<1$.

c) Par dérivation (par exemple)

$$P(\omega_{i,n}\leq t) = \frac{1}{\beta(i,n-i+1)} \int_0^t t^{i-1}(1-t)^{n-i}dt.$$

3°)a) Conséquence directe du 1°.

b) La matrice étant de permutation, son jacobien est un. La densité est obtenue comme $\sum f(x_{r_1},\dots,x_{r_n})$, où $(r_1,\dots,r_n)$ varie dans toutes les permutations de $(1,\dots,n)$. On obtient donc $n!$.

c) Soit $r=(r_1,\dots,r_n)$ une permutation donnée de $(1,\dots,n)$ et doit $R=(r_{1,n},\dots,r_{n,n})$. Pour tout borélien A de l'ensemble des $0<x_1<\dots<x_n<1$, on a

$$P(\{(\omega_{1,n},\dots,\omega_{n,n}) \in A\} \cap \{R=r\}) = P((\omega_{s_1},\dots,\omega_{s_n}) \in A).$$

En prenant pour A tout l'espace, on obtient $P(R=r)=1/n!$, notant qu'ici $(s_1,\dots,s_n)$ est une permutation de $(1,\dots,n)$ définie par r. Enfin, la loi conditionnelle de $(\omega_{1,n},\dots,\omega_{n,n})$ sachant $R=r$ ne dépend pas de r, d'où l'indépendance.

Commentaires

Beaucoup de bonnes réponses au 2°, résultats très médiocres au 3°.

4°)a) Evident (cf. 3°)

b) $P(R(n)\leq i|\omega_{i,n-1}=t) = P(\omega_n<t|\omega_{i,n-1}=t) = t$. Ainsi $P(R(n)\leq i) = E(\omega_{i,n-1})$ [illegible] en utilisant le 2°c. D'où $P(R(n)=i) = 1/n$, indépendamment de $1\leq i\leq n$.

Commentaires

Question mal traitée.

c) $r_{1,n},\dots,r_{n,n}$ déterminent $R(1),\dots,R(n)$, tandis que $R(n+1)$ ne dépend que de $_{1,n},\dots,{}_{n,n}$ et de $_{n+1}$, d'où l'indépendance.

5°)a) Convergence vers une loi de Poisson d'espérance log k (cf.II,2°).

b) On a

$$\frac{\theta^N}{N!}e^{-\theta} \leq \sum_{r=N}^{\infty}\frac{\theta^r}{r!}e^{-\theta} \leq \frac{\theta^N}{N!}e^{-\theta}\sum_{i=0}^{\infty}\frac{\theta^i}{(N+1)\dots(N+i)} \leq \frac{\theta^N}{N!}e^{-\theta}\sum_{i=0}^{\infty}\left(\frac{\theta}{N}\right)^i \leq \frac{\theta^N e^{-\theta}}{N!(1-$$

c) Par le II,3°, $N(n,kn)-M(n,kn)$ est p.s. nul à partir d'un certain rang. Il suffit donc de raisonner pour $M(n,kn)$. Or

$$P_n = P\left(M(n,kn)\geq(1+\varepsilon)\frac{\log n}{\log\log n}\right) \leq \frac{\theta^N}{N!}e^{-\theta}\left(\frac{1}{1-\frac{\theta}{N}}\right), \text{ où } \theta = \sum_{i=n+1}^{kn}\frac{1}{i} \to \log k,$$

et où $N = \frac{(1+\varepsilon)\log n}{\log\log n} \to \infty$. Par Stirling, on en déduit que $P_n \leq \exp(-(1+\varepsilon+o(1))\log n)$

pour $n\to\infty$, d'où pour $\varepsilon>0$ la convergence de $\sum P_n$ et le résultat par Borel-Cantelli.

Commentaire

Le a et le b sont souvent traités. Par contre, le c ne semble pas avoir retenu l'attention des candidats.

6°)a) De toute évidence $1 \leq \lim\inf_{n\to\infty}\{L(nk)-L(n)\} \leq \lim\sup_{n\to\infty}\{L(nk)-L(n)\} \leq 1+[\log k$

b,c) τ_n suit une loi de Poisson d'espérance ℓ. Le même démonstration qu'au 5°c montre que

$$\sum_n P\left(\tau_n \geq \frac{(1+\varepsilon)\log n}{\log\log n}\right) < \infty \quad \text{si} \quad \varepsilon>0.$$

Posons $\delta_n = \tau_{n\ell}$. Ces v.a. sont indépendantes et telles que

$$\sum_n P\left(\delta_n > \frac{(1-\varepsilon)\log n}{\log\log n}\right) = \infty \quad \text{si} \quad \varepsilon>0.$$

Le résultat s'en déduit par Borel-Cantelli et $\tau_{L(n)} \leq N(n,kn)$, $k\geq 2$.

d) Il suffit de refaire les mêmes raisonnements, en remplaçant les v.a. Z_n par des v.a. de Poisson d'espérance 1/2 (par exemple).

Commentaire

Cette question n'est généralement pas traitée.

Indications bibliographiques

P. Deheuvels et D. Pfeifer (1986) A semigroup approach to Poisson approximation, Annals of Probability, Vol. 14.

Option Probabilités-Statistiques

Répartition des notes

0 à 4	:	59 copies
5 à 9	:	46 copies
10 à 14	:	53 copies
15 à 19	:	64 copies
20 à 24	:	28 copies
25 à 29	:	17 copies
30 à 34	:	11 copies
35 à 40	:	12 copies
		290