Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2016 – 2017

Statistiques II

Contrôle continu n°1, mars 2017

Examen de 1h30. Tout document ou calculatrice est interdit.

1. On considère une suite $(X_i)_{i\in\mathbb{N}}$ de variables aléatoires indépendantes et identiquement distribuées suivant une loi discrète à valeurs dans $\{-1,0,1\}$ telle que $\Pr(X_0=0)=p_0>0$ and $p_1=\Pr(X_0=1)>\Pr(X_0=-1)>0$. On définit également:
$$Y_n=\sum_{i=1}^n X_i, \quad \text{pour tout } n\in\mathbb{N}^*.$$
   (a) Déterminer $\mathbb{E}(X_0)$ en fonction de p_0 et p_1 et montrer que $\mathbb{E}(X_0)=m>0$ **(1pt)**. Déterminer en fonction de p_0 et p_1, $\sigma^2=\mathrm{var}(X_0)$ **(1pt)**.

   (b) Montrer que Y_n est une variable aléatoire discrète et préciser l'ensemble de ses valeurs possibles **(0.5pts)**.

   (c) Déterminer $\mathbb{E}(Y_n)$ puis $\mathrm{var}(Y_n)$ pour tout $n\in\mathbf{N}^*$ **(1pt)**. (Y_n) est-il un processus stationnaire **(0.5pts)**?

   (d) Montrer que $Y_n \underset{n\to+\infty}{\overset{p.s.}{\longrightarrow}} +\infty$ **(1.5pts)**.

   (e) Montrer que $(Y_n)_{n\in\mathbf{N}}$ peut se mettre sous la forme
$$Y_n=m\,n+\sqrt{n}\,\varepsilon_n \quad \text{pour tout } n\in\mathbb{N}^*,$$
   où $(\varepsilon_n)_{n\in\mathbf{N}^*}$ est une suite de variables aléatoires centrées de même variance (à préciser) **(1pt)**. Quelle est la tendance de $(Y_n)_{n\in\mathbf{N}}$ **(0.5pts)**? Montrer que la loi de ε_n converge vers une loi à préciser lorsque $n\to\infty$ **(1pt)**. En déduire que les variables $(\varepsilon_n)_{n\in\mathbf{N}^*}$ ne forment pas un processus stationnaire **(1.5pts)**.

   (f) Déterminer $\mathrm{cov}(Y_i,Y_j)$ puis $\mathrm{cov}(\varepsilon_i,\varepsilon_j)$ pour $i,j\in\mathbf{N}^*$ **(2pts)**. Que devient cette quantité lorsque $|i-j|$ est "grand" **(1pt)**?

   (g) On suppose maintenant que $(Y_1,\cdots,Y_N)$ est un échantillon observé de $(Y_n)_{n\in\mathbf{N}^*}$. On suppose que tous les paramètres m, p_0, p_1 sont inconnus. Montrer que l'estimateur $\widehat{m}_N$ de m par régression linéaire par moindres carrés est défini par
$$\widehat{m}_N=\frac{6}{N(N+1)(2N+1)}\sum_{i=1}^N i\,Y_i \qquad \textbf{(3pts)}.$$
   Calculer $\mathbb{E}(\widehat{m}_N)$ et $\mathrm{var}(\widehat{m}_N)$ **(2pts)**. L'estimateur $\widehat{m}_N$ est-il un estimateur convergent de m **(1pt)**?

   (h) Montrer que les estimateurs $\widehat{p}_0=\frac{1}{N}\sum_{i=0}^{N-1}\mathbb{I}_{Y_{i+1}-Y_i=0}$ et $\widehat{p}_1=\frac{1}{N}\sum_{i=0}^{N-1}\mathbb{I}_{Y_{i+1}-Y_i=1}$ (on pose par convention $Y_0=0$) sont les estimateurs de p_0 et p_1 par maximum de vraisemblance (on pourra écrire la vraisemblance à l'aide d'un produit de probabilités conditionnelles) **(3pts)**. En justifiant, donner les théorèmes de la limite centrale vérifiés par $\widehat{p}_0$ et $\widehat{p}_1$ **(2pts)**.

   (i) En utilisant l'expression de m en fonction de p_0 et p_1, en déduire un autre estimateur $\widetilde{m}_N$ non biaisé de m **(1pt)**. Converge-t-il plus vite que $\widehat{m}_N$ vers m (au sens du risque quadratique) **(2pts)**?