Examen du 7 janvier 2008 (Durée 2 heure)

Les documents ne sont pas autorisés mais les calculettes et les tables statistiques sont autorisées.
Il sera tenu compte de la rigueur de la rédaction et de la clarté de la présentation

Popularité de pages

Dans le cadre du test d'un serveur web, on souhaite générer une suite de requêtes indépendantes parmi les pages du serveur. Pour cela on suppose que les pages sont numérotées par popularité décroissante et on utilise l'algorithme suivant

Génère_numéro_page() ;
U=random() ; {Les appels successifs à la fonction random renvoient une séquence de nombres réels indépendants et uniformément distribués sur $[0, 1[$.}
X =INT($\frac{1}{U}$) ; {INT est la partie entière inférieure }
Retourner X

Question 1.1 : Loi de la popularité

Calculer la loi de la variable X retournée par l'algorithme.

Question 1.2 : Popularité moyenne

Calculer la popularité moyenne, commenter votre résultat.

Question 1.3 :

Calculer la probabilité que 2 requêtes indépendantes demandent la même page. Qu'en concluez-vous ? [1]

Calcul de surface

La méthode générale d'estimation de la valeur d'une intégrale par simulation est la suivante :
- soit f une fonction définie sur $[0, 1[$ dont on veut calculer l'intégrale I, en général f est très compliquée et sans primitive connue,

$$I = \int_0^1 f(x)dx.$$

- si U est une variable aléatoire de loi uniforme sur $[0, 1[$, on montre facilement que

$$\mathbb{E}f(U) = \int_0^1 f(u)densite(u)du = \int_0^1 f(u)du;$$

car la densité de la loi uniforme est 1 sur $[0, 1[$.

On en déduit que la valeur de I peut être estimée par $\hat{I}_n$

$$\hat{I}_n = \frac{1}{n}\sum_{i=1}^{n} f(U_i);$$

où $\{U_i\}_{i\in\mathbb{N}}$ est une suite de tirages aléatoires indépendants de loi uniforme sur $[0, 1[$.

Question 2.1 : Justification

Par quels théorèmes justifie-t-on cette approche et comment peut-on calculer l'erreur de cette estimation ?

[1] On rappelle que $\sum_{k=1}^{+\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$

On considère maintenant la fonction $f(x) = \sqrt{1-x^2}$.

Question 2.2 : Propriétés de $f(U)$

Pour U de loi uniforme sur $[0,1[$ montrer que $\mathbb{E}f(U) = \frac{\pi}{4}$ (faire un dessin).

Calculer $\mathbb{E}f(U)^2$ et en déduire la valeur de $\mathbb{V}ar f(U)$.

Question 2.3 : Taille de l'échantillon

Quelle est la taille de l'échantillon nécessaire pour que l'erreur d'estimation de I soit inférieure à 10^{-2} avec une confiance de 95%.

On considère maintenant la fonction $g(x) = f(x) + c(x - \frac{1}{2})$ où c est un coefficient que l'on calculera ultérieurement.

Question 2.4 : Propriétés de $g(U)$

Montrer que

$$\mathbb{E}g(U) = \mathbb{E}f(U) = I.$$

En déduire que l'on peut remplacer f par g dans l'algorithme de simulation.

Calculer $\mathbb{E}g(U)^2$ en fonction de c. En déduire $\mathbb{V}ar g(U)$ en fonction de c. Montrer qu'il existe une valeur optimale c^* qui minimise la variance de $g(U)$.

Question 2.5 : Simulation optimale

Donner la valeur exacte de c^* en fonction de π. En déduire $\mathbb{V}ar g^*(U)$. Quelle est, maintenant la taille de l'échantillon nécessaire pour que l'erreur d'estimation de I soit inférieure à 10^{-2} avec une confiance de 95%.

De combien la taille de l'échantillon a-t-elle été réduite ?

Question 2.6 :

Expliquer pourquoi la connaissance approchée de c^* peut suffire.

Roue de la fortune

On considère un jeu de fête foraine constitué d'une roue montée sur pivot décomposée en secteurs de taille égale. Sur chaque secteur on a écrit une lettre, le mot formé par la roue est "BARBAPAPA".

Une observation des lettres tirées par la roue fournit le tableau suivant :

lettre	B	A	R	P
effectif	211	426	112	251

Question 3.1 : Interprétation

Qu'en pensez-vous ?

Question 3.2 : Générateur

A partir d'un générateur de bits de loi uniforme sur $\{0,1\}$, écrire un algorithme simulant la roue foraine.