Eléments du calcul des probabilités, 22 janvier 2021

Toutes les réponses doivent être justifiées. *L'usage de la calculatrice est interdit.*

Question 1. Soit D l'ensemble des couples (a, b) tels que a et b soient strictement compris entre 0 et 1. La fonction *random* renvoie un réel uniformément distribué entre 0 et 1. On l'appelle n fois et on obtient n valeurs dont la plus grande est Y. Construire une fonction f de domaine D, telle que $P(Y \geq a) \geq b$ si et seulement si $n \geq f(a, b)$.

Question 2. Une urne contient 4 boules blanches et 4 boules noires. On choisit 4 boules au hasard. S'il y a 2 boules blanches et 2 boules noires, on s'arrête. Sinon, on remet les boules dans l'urne et on recommence. Quelle est la probabilité que l'arrêt survienne après plus de n essais ?

Question 3. Qu'est-ce que l'inégalité de Tchebyshev ? On donnera son énoncé et sa démonstration, en veillant à préciser si elle s'applique seulement aux v.a. discrètes, ou seulement aux v.a. continues, ou aux deux types de variables.

Question 4. La v.a. X suit une loi binomiale de paramètres n et p. Quelles sont sa moyenne et sa variance ? Quelle est la moyenne de la v.a. $Y = X^2$? Indiquer, sans détailler les calculs, comment on pourrait déterminer $var(Y)$.

Bonus (projet). Dans le cours, on a construit une relation de récurrence pour la fonction de Wilcoxon $(n, p, i) \mapsto w(n, p, i)$. De manière analogue, construire une relation de récurrence permettant un calcul simple et efficace de la fonction $(N, K, I) \mapsto q(N, K, I)$.

Eléments du calcul des probabilités, 22 janvier 2021

Question 1. Si X_i est la $i$ème valeur fournie par *random*, on a $P(Y \geq a) = 1 - P(\wedge_{i=1}^{n} X_i \leq a) \ = \ 1 - a^n$. On a donc $P(Y \geq a) \geq b \longleftrightarrow b \leq 1 - a^n$. On a aussi

$$b \leq 1 - a^n \longleftrightarrow a^n \leq 1 - b \longleftrightarrow n \log(a) \leq \log(1-b) \longleftrightarrow n \geq \log(1-b)/\log(a)\,.$$

Le changement de sens de l'inégalité est dû au fait que $\log(a)$ est négatif puisque $a < 1$.
On définit donc

$$f(a,b) = \frac{\log(1-b)}{\log(a)}.$$

Remarque. La base des logarithmes utilisés peut être n'importe quel réel plus grand que 1, comme 2, e ou 10. [1]

Question 2. La probabilité de tirer 2 boules blanches et 2 boules noires est $p = \mathsf{C}_4^2\mathsf{C}_4^2 \,/\, \mathsf{C}_8^4 = 36/70 = 18/35$ (distribution hypergéométrique, tr. 69). La probabilité que l'arrêt survienne après plus de n essais est de $(1-p)^n = (17/35)^n$ (distribution géométrique, tr. 61).

Question 3. Tout est au transparent 101. L'inégalité s'applique à toute v.a. admettant une moyenne et une variance ; la v.a. peut donc être continue ou discrète ; elle ne doit pas nécessairement être positive.

Question 4. Les transparents 56 et 59 permettent de calculer $E[X^n]$ et $var(X^n)$ pour tout n ; en particulier, on a $E[X] = np$, $var(X) = np(1-p)$ et $E[Y] = E[X^2] = (np)^2 + np(1-p)$. On a aussi $var(Y) = E[Y^2] - (E[Y])^2 = E[X^4] - (E[X^2])^2$.

Question Bonus. Pour la fonction de Wilcoxon, on avait exprimé w pour (n,p,i) en fonction de deux valeurs plus simples des arguments, à savoir $(n, p-1, i-n)$ et $(n-1, p, i)$ (tr. 17). On procède de même ici. Quand on regarde la Kième et dernière vignette du lot, elle peut être nouvelle, auquel cas la situation auparavant était $(N, K-1, I-1)$, ou un doublon, auquel cas la situation auparavant était $(N, K-1, I)$. Notons aussi que, quand on possède déjà $I-1$ vignettes distinctes, la probabilité qu'une vignette supplémentaire soit nouvelle est de $(N-I+1)/N$; quand on possède déjà I vignettes distinctes, la probabilité qu'une vignette supplémentaire soit un doublon est de I/N. On en déduit immédiatement la relation de récurrence demandée :

$$q(N,K,I) \ = \ \frac{N-I+1}{N} q(N,K-1,I-1) + \frac{I}{N} q(N,K-1,I)\,,$$

valable pour $K > 1$ et $I > 1$. [2]

1. Ces bases sont souvent prévues dans les langages de programmation usuels. Rien n'interdirait de choisir la base dans $]0:1[$ (par exemple a) ; le calcul serait alors un peu différent. Cependant, on évite généralement un tel choix, qui conduirait à une fonction log décroissante.
2. Les cas de base $K = 1$ et $I = 1$ sont évidents.