MPA - Ensimag 2A **2012-13**

Séance 2 - Formule de Bayes et exemples d'inférence bayésienne

Exercice 1. Recherche d'un objet. Un objet a disparu dans une zone déterminée divisée en trois régions de même surface pour les recherches. La recherche commence par la fouille de la région 1. Soit q_1 la probabilité de ne pas trouver l'objet dans cette région conditionnellement à l'événement qu'il s'y trouve effectivement. Quelle est la probabilité que l'objet se trouve dans les régions 2 ou 3 sachant que les recherches dans la région 1 ont été infructueuses.

Exercice 2. La loi de l'ouest. A Las Vegas, on rencontre une proportion p de tricheurs, $0 < p < 1$. Lorsque l'on joue contre un tricheur, la probabilité de gagner une partie est nulle, tandis que, lorsque l'on joue contre une personne honnête, cette probabilité est $1/2$. On joue et on perd une partie à Las Vegas. Quelle est la probabilité d'avoir joué contre un tricheur ?

Exercice 3. Tarzan est-il un bon biologiste de terrain ? La loi du célèbre naturaliste Ramon Homstein-Grunberger indique que la taille des girafes pigmées est uniformément répartie entre 0.5 et 1.5m. Adepte du scepticisme, Tarzan fait une sortie dans la jungle avec son mètre mesureur. Au retour il prétend que les girafes pigmées ne dépassent pas 1.25m. Le problème consiste à estimer le nombre de girafes mesurées par Tarzan, et à donner la probabilité qu'il ait mesuré moins de 10 girafes. Pour cela, on considère des réalisations indépendantes, $(U_i)_{i\geq 1}$, de loi uniforme sur $(0,1)$ (les tailles des girafes moins 0.5m). Soit $t \in (0,1)$.

1. Soit $k \geq 1$. On pose $X_k = \max_{i=1,\ldots,k} U_i$. Montrer que
$$\mathrm{P}(X_k \leq t) = t^k\,.$$
2. Soit $n \geq 1$. On observe le résultat du maximum d'un nombre inconnu et aléatoire, N, de variables U_i. On suppose que N est indépendant des (U_i) et de loi uniforme sur $\{1,\ldots,n\}$. Déterminer la fonction de répartition de la variable X_N. (Rappel : $\sum_{\ell=0}^{n} t^\ell = (1-t^{n+1})/(1-t)$.)
3. Tarzan observe $(X_N \leq t)$, $t = 3/4$. Montrer que la loi conditionnelle de N sachant $(X_N \leq t)$ est donnée par
$$\mathrm{P}(N = k | X_N \leq t) = \frac{t^{k-1}(1-t)}{1-t^n}\,, \quad k = 1,\ldots,n.$$
4. Démontrer que, pour tout $k = 1,\ldots,n$,
$$\mathrm{P}(N = k | X_N \leq t) = \mathrm{P}(N_\star = k | N_\star \leq n).$$
où $N_\star$ est une variable de loi géométrique de paramètre $1-t$.
5. Déterminer la limite de la loi conditionnelle de N sachant $(X_N \leq t)$ lorsque n tend vers l'infini. Quelle est l'espérance mathématique de la loi limite ?

6. Donner une estimation du nombre de girafes mesurées par Tarzan s'appuyant sur les questions précédentes. Peut on la comparer à l'estimateur de maximum de vraisemblance ? Calculer la probabilité que Tarzan ait mesuré moins de 10 girafes.

Exercice 4. Estimation de la fréquence revisitée par Laplace. Pierre-Simon observe qu'un événement (de probabilité inconnue) se produit $y = 9$ fois lors de la répétition de $n = 20$ épreuves indépendantes. Il souhaite estimer la probabilité que cet événement se réalise à nouveau lors de la 21[e] épreuve. Il note θ la probabilité inconnue, et la traite comme variable aléatoire de loi uniforme sur $(0, 1)$.

1. Calculer la probabilité conditionnelle $p(y|\theta)$ (vraisemblance).
2. Montrer que la loi conditionnelle $p(\theta|y)$ (loi a posteriori) est la loi beta$(y+1, n+1-y)$, dont la densité est de forme générale

$$p(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}\,, \quad x \in (0,1), \alpha > 0, \beta > 0.$$

3. Calculer l'espérance conditionnelle de θ sachant y. Que se passe-t-il lorsque $n \to \infty$? Interprétation ?
4. Ecrire un algorithme en langage R permettant de simuler la loi conditionnelle $p(\theta|y)$ sans la calculer (Idée : utiliser un algorithme de rejet s'appuyant sur les générateurs `runif` et `rbinom`, indiquer le nombre de rejets).
5. Calculer la probabilité que l'événement se réalise lors d'un épreuve future, d'abord théoriquement, puis donner une estimation numérique à l'aide de l'algorithme de simulation.
6. Calculer la probabilité que θ soit supérieur à $1/2$.