**Cursus HPC-BigData 2023 / Analyse bayésienne :**

**TP1 Modèle Bêta-Binomial : Avalanche à Montroc**

En 1999, une énorme avalanche balayait le hameau de Montroc près de Chamonix, faisant 12 morts et détruisant 14 chalets dans une zone déclarée constructible car considérée comme hors d'atteinte d'après la cartographie des risques établie en 1992.

Imaginez-vous en 1992, le conseil municipal de Chamonix attend vos conclusions issues d'une analyse bayésienne pour déclarer la zone constructible ou pas.

Répondre à la question : « fallait-il construire à Montroc ? » à partir des éléments vus en cours et du TP suivant.

-------------------------------------

On considère que l'occurrence d'une année à avalanche à Montroc suit un processus de Bernoulli de paramètre $\theta$, probabilité d'observer au moins une avalanche une année donnée.

Deux actions sont alors possibles : construire (=décision a1) ou ne pas construire (=décision a2).

On considère la variable X égale au nombre d'années à avalanche sur une période de n années. Soit P(h) la probabilité d'observer au moins une avalanche lors des h prochaines années.

Les coûts, en millions d'euros, associés aux décisions a1 et a2 sont les suivants :

- C1 si on décide de construire les chalets et qu'on observe au moins une avalanche lors des h prochaines années (indemnisation).
- C2 si on refuse le projet de construction et qu'on n'observe pas d'avalanche lors des h prochaines années (manque à gagner).
- Les autres coûts sont ici considérés comme étant nuls. On a de plus C1>C2.

On choisit de travailler avec un modèle Bêta-Binomial de posterior $f(\theta|x)$ (cf.Annexe). On consulte un spécialiste des avalanches qui pense que $\theta$ a une chance sur 10 de dépasser la valeur 0.05, et 5 chances sur 100 d'être inférieur à la valeur 0.01.

D'autre part, la consultation des archives a révélé qu'il y a eu 6 années à avalanches en 150 ans.

-------------------------------------

*Problème tiré de « Pratique du calcul bayésien », Boreux, Parent & Bernier, Springer.*

1) Initialisation :

Installer puis charger le package ***rootSolve***

2) Phase d'élicitation du prior :

On cherche à construire le prior à partir de l'expertise qui vous a été fournie. En utilisant la fonction ***multiroot*** initialisée avec ***start=c(1,10)*** et le code ci-dessous, quelles valeurs des hyperparamètres p et q du prior bêta obtenez-vous ?

***model = function(x,pa,qa,pb,qb){***

***c(F1=pa-pbeta(qa,x[1],x[2]),***

***F2=pb-pbeta(qb,x[1],x[2]))}***

3) Elaboration du posterior :

Calculer les paramètres de la loi a posteriori et représenter sur un même graphe les densités a priori $f(\theta)$ et a posteriori $f(\theta|x)$.

4) Distribution prédictive a posteriori :

Créer une fonction permettant de calculer la loi de probabilité prédictive a posteriori $P(X^*=x^*|X=x)$ Bêta-Binomiale en fonction de x*, h, n, p, q, x, en utilisant les fonctions ***beta*** et ***choose*** de R. Représenter la loi prédictive a posteriori pour les horizons h= 5, 10, 20 et 30 ans.

5) Outil d'aide à la prise de décision :

P(h) étant la probabilité qu'il y ait au moins une avalanche à Montroc dans les h prochaines années, le rapport r = [P(h)/(1-P(h)].[C1/C2] fournit une règle de décision naturelle : on choisit a1 (construction) si r < 1, et a2 (non construction) sinon.

***On choisit en effet la décision qui minimise la perte moyenne sur les h années.*** (cf.Annexe)

Représenter graphiquement les domaines de décision a1 et a2 en fonction de h et du rapport C1/C2 et conclure sachant que C1/C2 était estimé à 6 dans la réalité.

6) Epilogue :

Le promoteur à l'origine du projet immobilier attaque votre étude, jugeant irréaliste l'expertise fournie par votre spécialiste des avalanches.

Refaire alors l'étude avec un prior non informatif (le prior de Jeffreys Be(½,½) relatif à la vraisemblance Binomiale ou le prior uniforme). Conclure de nouveau.

*Problème tiré de « Pratique du calcul bayésien », Boreux, Parent & Bernier, Springer.*

# <u>Annexe</u> :

<u>Modélisation bayésienne</u> :

La variable X, égale au nombre d'années à avalanche sur un horizon de n années suit la loi Binomiale de paramètres (n,θ) et on cale, dans le cadre d'une modélisation bayésienne, le prior conjugué Bêta(p,q) sur la probabilité inconnue θ.

En exploitant le posterior Bêta(p+x,q+n-x) induit par cette modélisation, on obtient la loi dite Bêta-Binomiale (ou loi Polya), loi prédictive a posteriori relative à un horizon de h années :

$$P(X^* = x^* | X = x) = \int_0^1 f(x^*|\theta) f(\theta|x) d\theta = \binom{h}{x^*} \frac{B(x + x^* + p, n + q + h - x - x^*)}{B(x + p, n + q - x)}$$

On définit donc la probabilité P(h) d'avoir au moins une année à avalanche sur l'horizon h :

$$P(h) = P(X^* \geq 1 | X = x) = 1 - P(X^* = 0 | X = x) = 1 - \frac{B(x + p, n + q + h - x)}{B(x + p, n + q - x)}$$

<u>Stratégie d'aide à la décision</u> :

La décision sera prise en couplant l'outil statistique bayésien que constitue la loi prédictive a posteriori avec la fonction de coût L(a,θ) du problème. Le tableau ci-dessous définit cette fonction L ainsi que sa loi a posteriori sur un horizon h, pour chacune des deux décisions :

| **Fonction de coût L(a,θ)** | **Avalanche : P(h)** | **Non Avalanche : 1-P(h)** |
|---|---|---|
| **Décision a1** | C1 | 0 |
| **Décision a2** | 0 | C2 |

La stratégie de décision reposera sur la minimisation du coût moyen sur l'horizon h considéré :
on choisira donc la décision qui mènera au coût moyen a posteriori le plus faible.

Coût moyen a posteriori sur les h futures années en cas de décision a1 :
E[L(a1,θ)|X=x] = P(h).C1

Coût moyen a posteriori sur les h futures années en cas de décision a2 :
E[L(a2,θ)|X=x] = (1-P(h)).C2

**On prendra donc la décision a1 de construire si P(h).C1 < (1-P(h)).C2 ,**
**et donc si le rapport r = [P(h)/(1-P(h)].[C1/C2] < 1.**

*Problème tiré de « Pratique du calcul bayésien », Boreux, Parent & Bernier, Springer.*