

TP 5 : Révisions : simulation et estimation

Ce TP n°5 reprend tles grandes lignes de le simulation et l'estimation. Il peut-être utile de reprendre les techniques utilisées dans les TP 2 à 4.

Exercice 1 : Simulation et estimation pour la Loi Gamma. La loi Gamma de paramètres $\lambda > 0$ et $m \in \mathbb{N}^*$, notée $\Gamma(m, \lambda)$, est la loi à densité sur \mathbb{R}_+ :

$$f_{m,\lambda}(x) := \frac{x^{m-1}}{(m-1)!} \lambda^m e^{-\lambda x}.$$

1. Que dire d'une variable de loi $\Gamma(m, \lambda)$ quand m vaut 1 ?
2. Définir une fonction `f(m, lambda, x)` qui renvoie $f_{m,\lambda}(x)$. Représenter sur un même graphique des densités des Lois Gamma sur l'intervalle $[0, 8]$ quand $\lambda = 2$, pour les cas $m = 1$, $m = 2$ et $m = 5$. *Indication : on pourra utiliser la fonction `factorial`.*
3. Simuler un échantillon de $n = 2000$ variables de loi $\Gamma(5, 1.5)$. Représenter l'histogramme du jeu de données obtenu, et superposer la courbe correspondante à la densité. *Indication : c'est une loi classique sous R...*
4. Reprendre ce code précédent en faisant augmenter m : quelle semble être une approximation naturelle de la loi $\Gamma(m, \lambda)$ quand m tend vers l'infini ? Illustrer ce phénomène par une méthode votre choix.
5. On donne le résultat suivant : si X_1, \dots, X_m sont m variables exponentielles indépendantes de paramètre λ , alors

$$X_1 + \dots + X_m \sim \Gamma(m, \lambda).$$

Illustrer ce résultat à l'aide d'un histogramme pour $m = 4$ et $\lambda = 3$.

6. On dispose de Y_1, \dots, Y_n un échantillon de variables i.i.d. de loi $\Gamma(m, \lambda)$. On cherche à estimer m et λ . Proposer deux estimateurs \hat{m} et $\hat{\lambda}$ construits à partir de la moyenne empirique notée M_n et l'estimateur de la variance sans biais noté V_n .
7. Simuler $N = 2000$ échantillons Y_1, \dots, Y_n de loi $\Gamma(m = 5, \lambda = 5)$ de taille $n = 5000$. Calculer et afficher le biais, la variance, et le risque (estimés) de $\hat{\lambda}$ et \hat{m} . Commenter les résultats obtenus.

Pour ceux qui veulent aller plus loin : Une transition de phase pour l'estimation Dans ce problème, on suppose qu'on dispose d'un vecteur de \mathbb{R}^n dont toutes les entrées sont nulles, sauf une qui vaut 1. On notera e_i ce vecteur si le 1 est à la position i . On tire i uniformément au hasard dans $\{1, \dots, n\}$ et observe ensuite une version bruitée de e_i :

$$X = e_i + \sigma Z,$$

où Z est un vecteur Gaussien standard (toutes les entrées sont des Gaussiennes standard indépendantes), et $\sigma > 0$ est le paramètre de bruit. Le but est de retrouver la valeur de i en observant X .

1. Coder une fonction qui prend en entrée n et σ et qui renvoie une réalisation du couple (X, i) .
2. On considère l'estimateur suivant pour i :

$$\hat{i} = \arg \max_{1 \leq i \leq n} X_i.$$

Pour $n = 500$, pour σ variant entre 0 et 1, simuler $N = 1000$ réalisations de X et calculer \hat{i} . Tracer ainsi la courbe de la probabilité (estimée) $\mathbb{P}(\hat{i} = i)$, en fonction de σ . Qu'observe-t-on ?