Exploration Numérique 5

04/10/2019

1 Modèle de Poisson

Soit une suite de n-échantillons du modèle statistique $(\mathbb{N}, \mathscr{P}(\mathbb{N}), \{p_\theta \cdot \mu, \theta \in \Theta := \mathbb{R}_+^\star\})$ où

$$p_\theta : k \mapsto \frac{\theta^k}{k!} \exp(-\theta),$$

et μ est la mesure de comptage sur $\mathbb{N}$. On pose $\overline{X}_n := n^{-1} \sum_{i=1}^n X_i$.

La suite $\{\overline{X}_n,\ n \in \mathbb{N}\}$ est une suite consistante d'estimateurs de θ.

1. En prenant $\theta = 5$ et $n = 5$, obtenir 10^4 réalisations du modèle statistique et calculer $\overline{X}_n$ pour chaque réalisation. Visualiser l'histogramme de cees valeurs, en rajoutant un trait vertical en $\theta = 5$ pour visualiser la vraie valeur du paramètre.
2. Répéter la question précédente en gardant le nombre de réalisations constant et en augmentant le nombre d'échantillons : $n = 20$, $n = 80$, $n = 320$. Que remarquez-vous ?
3. Montrer que la suite $\{\overline{X}_n,\ n \in \mathbb{N}\}$ est asymptotiquement normale (calculer la variance asymptotique) ? Confirmer ce résultat en superposant aux histogrammes la densité de probabilité de la loi normale théorique.

2 Transformation de stabilisation de la variance

La connaissance de propriétés asymptotiques de suites d'estimateurs nous permet de construire des intervalles de confiance asymptotiques. Toutefois, la variance de la loi normale dépend du paramètre que l'on souhaite estimer, donc que le calcul de l'intervalle de confiance asymptotique nécessite d'estimer la variance asymptotique de la suite d'estimateurs (voir poly, section II-1.5). On utilise la technique de la transformation de stabilisation de la variance pour contourner ce problème.

Soit $\{X_n, n \in \mathbb{N}\}$ un échantillon de Bernoulli

$$(\{0,1\}, \mathscr{P}(\{0,1\}), \{Ber(\theta), \theta \in \Theta =]0,1[\}).$$

La moyenne empirique $\overline{X}_n$, définie comme dans l'exercice précédent, est un estimateur non biaisé de θ. Pour tout $\theta \in \Theta$, l'application du théorème de la limite centrale nous permet de montrer que :

$$\sqrt{n}(\overline{X}_n - \theta) \xrightarrow{\mathbb{P}_\theta} \mathrm{N}(0, \theta(1-\theta))$$

Une transformation de stabilisation de la variance est donnée par $x \mapsto \arcsin(\sqrt{x})$ (cf. poly). Par conséquent, pour tout $\theta \in \Theta$, pour $n \to \infty$:

$$\sqrt{n}(\arcsin(\sqrt{\overline{X}_n}) - \arcsin(\sqrt{\theta})) \xrightarrow{\mathbb{P}_\theta} \mathrm{N}(0, 1/4) \tag{1}$$

1. En prenant $\theta = 0.3$ et $n = 100$, obtenir 10^4 réalisations du modèle statistique et calculer $\arcsin(\sqrt{\overline{X}_n})$ pour chacune. Afficher les valeurs dans un histogramme et vérifier que la moyenne et la variance des valeurs obtenues sont conformes à l'équation 1.
2. De l'équation 1 on déduit un intervalle de confiance asymptotique à 95% de θ :

$$\left[\sin^2(\arcsin(\sqrt{\overline{X}_n}) - z_{97.5}/2\sqrt{n}); \sin^2(\arcsin(\sqrt{\overline{X}_n}) + z_{97.5}/2\sqrt{n})\right],$$

où $z_{97.5}$ est le quantile de la loi $\mathrm{N}(0,1)$. Sur les 10^4 réalisations, quelle est la proportion d'intervalles qui contiennent effectivement la vraie valeur du paramètre ? Cette proportion est appelée *taux de couverture* de l'intervalle.