DM n°3 – à rendre pour le 14 avril 2015

Des points sont prévus pour la qualité de la rédaction.

1) Soit $X_1, X_2, \ldots$ une suite de variables aléatoires indépendantes de loi de Bernoulli de paramètre $p \in]0,1[$. On appelle $S_n = \sum_{k=1}^{n} X_k$.

   a) Donner la moyenne et la variance de S_n.

   b) Énoncer le théorème central limite pour S_n.

   c) À l'aide d'une table de la loi normale centrée réduite (on en trouve sur internet), donner une valeur approchée d'un réel $x_0 > 0$ tel que

$$\mathbb{P}\Bigg(- x_0 < \sqrt{\frac{n}{p(1-p)}} \left(\frac{S_n}{n} - p \right) < x_0 \Bigg) \approx 0.95.$$

2) Après avoir lancé une pièce 10 000 fois de manière indépendante, on obtient 5 800 fois "face".

   a) Est-il raisonable de supposer que la pièce est non biaisée (c'est-à-dire que la probabilité d'avoir "pile" est égale à la probabilité d'avoir "face") ?
   Expliquer pourquoi, en utilisant le théorème central limite pour estimer $\mathbb{P}(S_{10000} \geq 5800)$ dans le cas d'une pièce non biaisée ($p = 1/2$).
   En langage statistique, on vient de "tester l'hypothèse $\{p = 1/2\}$".

   On ne connaît donc pas la valeur du paramètre p de la pièce ($p = \mathbb{P}(\text{"pile"})$). On cherche donc un *intervalle de confiance à 95%* pour la valeur de p, c'est-à-dire un intervalle $[a, b]$ telle que la probabilité que le paramètre $p \in [a, b]$ soit plus grande que 95%.

   b) Nous allons utiliser la question 1c) On obtient

$$\mathbb{P}\left(\frac{S_n}{n} - x_0 \frac{\sqrt{p(1-p)}}{\sqrt{n}} < p < \frac{S_n}{n} + x_0 \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right) \approx 0.95$$

   Cela ressemble presque à un intervalle de confiance, à ceci près que les bornes dépendent du paramètre inconnu p. En admettant l'approximation $\sqrt{p(1-p)} \approx \sqrt{\frac{S_n}{n}(1 - \frac{S_n}{n})}$, donner un intervalle de confiance à 95% pour p.

   c) Répéter la procédure pour avoir un intervalle de confiance à 99% pour p.