L2Mass Statistique - Corrigé de l'interrogation du 6 mars 2019

Q.1 Une expérience aléatoire conduit à un succès avec probabilité 0.3. On répète l'expérience 100 fois ; expliquez (brièvement) pourquoi le nombre total de succès est une variable aléatoire ; quelle est sa loi ?

Quelle est l'espérance de la fréquence de succès observée ? Quel est son écart type ?

Réponse : Le nombre de succès dépend des 100 résultats (succès ou échec) obtenus lors de la répétition de l'expérience, et ces résultats sont aléatoires. Le nombre de succès suit la loi binomiale $\mathcal{B}(100, 0.3)$.

Notons N le nombre de succès après 100 répétitions et f la fréquence de succès. On a

$f = \frac{N}{100}$

$\mathrm{E}(N) = 100 \times 0.3 = 30$ puis $\mathrm{E}(f) = \mathrm{E}(N)/100 = 0.3$

$\mathrm{V}(N) = 100 \times 0.3 \times (1 - 0.3) = 21$ puis $\mathrm{V}(f) = \mathrm{V}(N)/100^2$,
$\sigma(f) = \sqrt{\mathrm{V}(N)}/100 = \sqrt{21}/100 \in [0.04, 0.05]$

En interpolant ou avec une calculatrice on trouve $\sigma(f) \approx 0.046$

Q.2 On répète 100 fois une expérience conduisant à un succès avec probabilité 0.3. Comment s'exprime l'intervalle de fluctuation de la fréquence observée f au seuil 80% après approximation gaussienne de la loi de f et en terme de la fonction de répartition $\mathrm{F}_{\mathcal{N}}$ de la loi normale centrée réduite $\mathcal{N}(0, 1)$? (On ne demande pas l'expression de $\mathrm{F}_{\mathcal{N}}$.)

Quelle valeur donne la table de la loi normale donnée au dos ?

Réponse : L'intervalle de fluctuation de f au seuil 80% est, après approximation de la loi de f par une loi continue, $[q_0, q_1]$ où q_0 est le 1er décile, c'est à dire tel que $\mathrm{F}_f(q_0) = 0.1$ et q_1 le 9ème décile, c'est à dire tel que $\mathrm{F}_f(q_1) = 0.9$. (Si la loi de f est continue on a bien $\mathrm{P}(f \in [q_0, q_1]) = \mathrm{F}_f(q_1) - \mathrm{F}_f(q_0) = 0.9 - 0.1 = 0.8$.)

Réponse courte : après approximation gaussienne de la loi de f, l'intervalle de fluctuation de f est $[\mathrm{E}(f) - \sigma(f)x, \mathrm{E}(f) + \sigma(f)x]$ où $[-x, x]$ est l'intervalle de fluctuation de la loi $\mathcal{N}(0, 1)$, c'est à dire $\mathrm{F}_{\mathcal{N}(0,1)}(x) = \text{milieu}([0.8, 1]) = 0.9$. La table 2 donne $x \approx 1.282$ d'où $[q_0, q_1] \approx [0.3 - 0.046 \times 1.28, 0.3 + 0.046 \times 1.28] \approx [0.24, 0.36]$. (On a reporté les résultats de la question 1.)

Réponse détaillée : On approxime la loi de $\frac{f - \mathrm{E}(f)}{\sigma(f)}$ par $\mathcal{N}(0, 1)$; concrètement on approxime la fonction de répartition de la première par celle de la seconde.

On a

$$\mathrm{F}_f(q) \overset{def}{=} \mathrm{P}(f \leq q) = \mathrm{P}\big(\frac{f - \mathrm{E}(f)}{\sigma(f)} \leq \frac{q - \mathrm{E}(f)}{\sigma(f)}\big) \approx \mathrm{F}_{\mathcal{N}(0,1)}\big(\frac{q - \mathrm{E}(f)}{\sigma(f)}\big)$$

On cherche alors q_0 et q_1 tels que, après approximation gaussienne, $\mathrm{F}_{\mathcal{N}(0,1)}\big(\frac{q_0 - \mathrm{E}(f)}{\sigma(f)}\big) = 0.1$ et $\mathrm{F}_{\mathcal{N}(0,1)}\big(\frac{q_1 - \mathrm{E}(f)}{\sigma(f)}\big) = 0.9$.

La table 1 donne $\frac{q_1 - \mathrm{E}(f)}{\sigma(f)} \approx 1.28$ d'où $q_1 \approx \sigma(f) \times 1.28 + \mathrm{E}(f) \approx 0.046 \times 1.28 + 0.3 \approx 0.36$ et par symétrie de la loi $\mathcal{N}(0, 1)$, $\frac{q_0 - \mathrm{E}(f)}{\sigma(f)} \approx -1.28$ d'où $q_0 \approx \sigma(f) \times (-1.28) + \mathrm{E}(f) \approx 0.24$ (ou bien plus simplement q_0 est le symétrique de q_1 par rapport à $\mathrm{E}(f) = 0.3$).

Ex.2 Une petite compagnie d'assurance assure 100 personnes contre un même risque de probabilité 0.1. Si ce risque se réalise, la compagnie paye à l'assuré la somme 5000 (€). La cotisation de chaque assuré est de la forme $500 + 5000 * c$, où c est un paramètre déterminé par la compagnie.

Comment s'exprime le bénéfice de la compagnie ? En quoi est ce une variable aléatoire ? Quelle est sa loi ?

Pour quelles valeurs de c la compagnie a t-elle une probabilité inférieure à 1/100 de perdre de l'argent ?

Réponse : Notons N le nombre de personnes assurées pour lesquelles le risque s'est réalisé ; N suit la loi binomiale $\mathcal{B}(100, 0.1)$ d'espérance 10.

Le bénéfice X de la compagnie s'écrit $X = 100 \times (500 + 5000c) - 5000N = 100 \times (5000c + 500 - 50N)$. C'est une variable aléatoire puisque N est aléatoire et sa loi se déduit de celle de N. (Précisément N est à valeurs dans $\{0, \ldots, 100\}$ donc X prend les valeurs $100 \times (5000c + 500 - 50k)$ avec probabilité $\binom{n}{k}(0.1)^k(0.9)^{100-k}$, $k \in \{0, \ldots, 100\}$.)

On cherche c tel que $\mathrm{P}(X < 0) \leq 0.01$.

Après approximation gaussienne de N, X suit elle même une loi gaussienne d'espérance et de variance

$$\mathrm{E}(X) = \mathrm{E}(100 \times (5000c + 500 - 50N)) = 100 \times (5000c + 500 - 50\mathrm{E}(N)) = 500\,000\,c$$

$$\mathrm{Var}(X) = \mathrm{Var}(100 \times (500 + 5000c) - 5000N) = 5000^2\mathrm{Var}(N) = 5000^2 \times 100 \times 0.1 \times 0.9 = 5000^2 \times 3^2.$$

On a $\mathrm{P}(X < 0) = \mathrm{P}\left(\frac{X - \mathrm{E}(X)}{\sigma(X)} < -\frac{\mathrm{E}(X)}{\sigma(X)}\right) \approx \mathrm{F}_{\mathcal{N}(0,1)}\left(-\frac{\mathrm{E}(X)}{\sigma(X)}\right) = \mathrm{F}_{\mathcal{N}(0,1)}\left(-\frac{500\,000\,c}{15000}\right) = \mathrm{F}_{\mathcal{N}(0,1)}\left(-\frac{100\,c}{3}\right)$.

On cherche donc c tel que $\mathrm{F}_{\mathcal{N}(0,1)}\left(-\frac{100\,c}{3}\right) = 0.01$ ce qui équivaut par symétrie de $\mathcal{N}(0,1)$ à $\mathrm{F}_{\mathcal{N}(0,1)}\left(\frac{100\,c}{3}\right) = 1 - 0.01 = 0.99$.

On lit sur la table 1 $\frac{100\,c}{3} \approx 2.33$ soit $c \approx 0.07$.