Échantillonnage

Table des matières

I Rappels sur les lois usuelles 2

II Approximations de la loi binomiale 2
II.1 Approximation par la loi de poisson 2
II.2 Approximation par la loi normale 3

III Lois limites 4
III.1 Loi faible des grands nombres 4
III.2 Théorème de la limite centrée 4
III.3 Application : lois d'échantillonnage 5
III.3.1 Loi d'échantillonnage des moyennes 5
III.3.2 Loi d'échantillonnage des fréquences 5

I Rappels sur les lois usuelles

Voici un tableau récapitulatif représentant les principales formules des lois usuelles vues en première année :
(Dans toutes les formules, on a $p + q = 1$ c'est-à-dire $q = 1 - p$).

Loi	Notation	Formule	Espérance	Variance
Loi de Bernoulli	$\mathcal{B}(p)$	$P(X = 1) = p\,;\ P(X = 0) = q$	$E(X) = p$	$V(X) = pq)$
Loi Binomiale	$\mathcal{B}(n;p)$	$P(X = k) = C_n^k \times p^k \times q^{n-k}$	$E(X) = np$	$V(X) = npq)$
Loi de Poisson	$\mathcal{P}(\lambda)$	$P(X = k) = e^{-\lambda}\dfrac{\lambda^k}{k!}$	$E(X) = \lambda$	$V(X) = \lambda$
Loi Normale	$\mathcal{N}(m;\sigma)$	$f(x) = \dfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$	$E(X) = m$	$V(X) = \sigma^2$
Centrée réduite	$\mathcal{N}(0;1)$	$f(x) = \dfrac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$	$E(X) = 0$	$V(X) = 1$

II Approximations de la loi binomiale

II.1 Approximation par la loi de poisson

On admet le résultat suivant :

Propriété 1
Pour n « assez grand » et p « proche » de 0 tels que $np(1-p)$ ne soit « pas trop grand », on peut approcher la loi binomiale $\mathcal{B}(n,p)$ par la loi de poisson $\mathcal{P}(\lambda)$ où $\lambda = np$.

On convient de faire cette approximation pour $n > 30$, $p \leq 0,1$ et $np(1-p) \leq 10$.

Exemple 1
Dans une chaîne de fabrication, 5% des pièces sont défectueuses ; on prélève une pièce, on examine si elle est défectueuse et on la replace parmi les autres. On répète 120 fois cette expérience. On désigne par X la variable aléatoire qui à chaque tirage de 120 pièces associe le nombre de pièces défectueuses.

1. Justifier que X suit une loi binomiale, en préciser les paramètres.
2. Calculer $P(X = 5)$.
3. Montrer qu'une approximation de la loi binomiale par une loi de poisson convient.
4. Calculer $P(X = 5)$ à l'aide de cette approximation.
5. Comparer pour apprécier la qualité de l'approximation.

Solution :

1. Pour chaque tirage, on a deux résultats possibles : ou bien la pièce est défectueuse avec une probabilité de $p = 0,05$; ou bien elle ne l'est pas avec une probabilité de $q = 1 - p = 0,95$.
   On effectue 120 tirages de manière indépendante.
   On peut donc conclure que X suit la loi binomiale $\mathcal{B}(120; 0,05)$.
2. $P(X = 5) = C_{120}^{5} \times 0,05^5 \times 0,95^{115} = 0,1634$.
3. On a $n > 30$; $p < 0,1$ et $np(1 - p) = 5,7 < 10$.
   On peut donc faire une approximation grâce à la loi de poisson $\mathcal{P}(120 \times 0,05) = \mathcal{P}(6)$.
4. On obtient : $P(X = 5) = e^{-6}\dfrac{6^5}{5!} = 0,1606$.
5. La loi de poisson donne la même valeur à 10^{-2} près, ce qui est une bonne approximation.

II.2 Approximation par la loi normale

Propriété 2
Pour n « assez grand » et pour p « ni proche de 0 ni de 1 » tels que $np(1 - p)$ ne soit « pas trop petit », on peut approcher la loi binomiale $\mathcal{B}(n; p)$ par la loi normale $\mathcal{N}(m; \sigma)$ où $m = np$ et $\sigma = \sqrt{np(1 - p)}$.

On convient de faire cette approximation pour $n \geq 50$, $p \leq 50$ et $np(1 - p) > 10$.

Exemple 2
On lance 300 fois une pièce de monnaie truquée ce qui constitue une partie. La probabilité d'obtenir « face » est $\frac{2}{3}$.
On désigne par X la variable aléatoire qui à chaque partie associe le nombre de « face » obtenus.

1. Justifier que X suit une loi binomiale, en préciser les paramètres.
2. Peut-on calculer simplement $P(X > 210)$?
3. Montrer qu'une approximation de la loi binomiale par une loi normale se justifie.
4. Calculer $P(X > 210)$ à l'aide de cette approximation.

Solution :

1. Pour chaque jet, on a deux résultats possibles : ou bien on obtient « face » avec une probabilité de $p = \frac{2}{3}$, ou bien on obtient « pile » avec une probabilité de $q = 1 - p = \frac{1}{3}$.
   On lance 300fois la pièce de manière indépendante.
   On peut donc conclure que X suit la loi Binomiale $\mathcal{B}\left(300; \frac{2}{3}\right)$.
2. $P(X > 210) = \sum\limits_{i=211}^{300} C_{300}^{i} \times \dfrac{2}{3}^{i} \times \dfrac{1}{3}^{300-i}$, la calculatrice ne peut pas toujours effectuer un tel calcul.
3. On a $n \geq 50$, $p = \frac{2}{3}$ et $np(1 - p) = 66,66 > 10$.
   On peut donc faire une approximation par la loi normale $\mathcal{N}\left(300 \times \dfrac{2}{3}; \sqrt{300 \times \dfrac{2}{3} \times \dfrac{1}{3}}\right) = \mathcal{N}(200; 8,16)$.
4. On utilise le changement de variable $T = \dfrac{X - 200}{8,16}$. T suit la loi normale $\mathcal{N}(0,1)$.
   $$\begin{aligned} P(X > 210) &= P(8,16T + 200 > 210) \\ &= P(T > 1,22) \\ &= 1 - P(T \leq 1,22) \\ &= 1 - 0,8888 \\ &= 0,1112. \end{aligned}$$

III Lois limites

III.1 Loi faible des grands nombres

Propriété 3
Soit X_1, X_2, ..., X_n n variables aléatoires indépendantes ayant même espérance m et même écart-type σ et soit $\overline{X_n} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$, alors :

$$\text{Pour tout } \epsilon > 0, \lim_{n \to +\infty} P(|\overline{X_n} - m| < \epsilon) = 1.$$

Concrètement, ce théorème signifie que plus n est grand plus la variable aléatoire $\overline{X_n}$ se rapproche de l'espérance mathématique m.

Exemple 3
On lance un dé. Si on obtient 6, c'est gagné et on marque 1 point. Sinon, c'est perdu et on marque 0 point.
Soit X_i la variable aléatoire correspondant au nombre de point obtenu lors du $i - ième$ lancer.
On a donc : $P(X = 0) = \frac{5}{6}$, $P(X = 1) = \frac{1}{6}$ et $E(X) = \frac{1}{6}$.
On répète n fois cette même expérience, les n variables aléatoires X_1, X_2, ..., X_n ont la même loi de probabilité.
Pour connaitre le nombre de succès, on étudie la variable aléatoire $\overline{X_n}$: « Fréquence des succès »
avec $\overline{X_n} = \dfrac{\text{Nombre de succès}}{\text{Nombre d'expériences aléatoires}} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$.
Normalement, on devrait trouver $\overline{X_n} = \frac{1}{6}$.

- Pour $n = 3$ par exemple, il y a peu de chance pour que l'on trouve $\overline{X_3} = \frac{1}{6}$.
- Pour $n = 30$, la probabilité de trouver $\overline{X_{30}} = \frac{1}{6}$ augmente sans être très forte.
- Pour $n = 1000$, on se rapproche de cette valeur de $\frac{1}{6}$.

Le théorème dit que plus n est grand, plus $\overline{X_n}$ se rapproche de la valeur théorique $\frac{1}{6}$.

III.2 Théorème de la limite centrée

Propriété 4
Soit X_1, X_2, ..., X_n n variables aléatoires indépendantes ayant même espérance m et même écart-type σ et soit $\overline{X_n} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$, alors :
Pour n suffisamment grand, $\overline{X_n}$ suit approximativement la loi normale $\mathcal{N}\left(m, \dfrac{\sigma}{\sqrt{n}}\right)$.

Remarque 1
Dans la plupart des cas, on considère que n est « suffisamment grand » lorsque n atteint quelques dizaines, par exemple lorsque $n \geq 30$, mais cela dépend de la nature, de la population et du contexte de l'étude

III.3 Application : lois d'échantillonnage

En statistiques, il est en général impossible d'étudier un caractère sur toute une population de taille N élevée. La théorie de l'échantillonnage se pose la question suivante :

En supposant connus les paramètres statistiques de la population, que peut-on en déduire sur les échantillons prélevés dans la population ?

On suppose que ces échantillons sont prélevés au hasard et que le tirage de ces échantillons est effectué avec remise. L'ensemble de ces échantillons de taille n est appelé échantillonnage de taille n.

On peut étudier dans ces conditions :
- la loi d'échantillonnage des moyennes,
- la loi d'échantillonnage des fréquences,

III.3.1 Loi d'échantillonnage des moyennes

Étant donné une population de taille N et X une variable aléatoire telle que $E(X) = m$ et $\sigma(X) = \sigma$.

Pour prélever les échantillons de taille n, on a procédé à n épreuves indépendantes de variables aléatoires X_1, X_2, ..., X_n de même loi que X.

La variable aléatoire $\overline{X_n} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$ associe à tout échantillon de taille n sa moyenne.

D'après le théorème de la limite centrée, pour n assez grand, on a :

Propriété 5
La loi d'échantillonnage de taille n de la moyenne $\overline{X_n}$ quand $n \geq 30$, peut être approchée par la loi normale $\mathcal{N}\left(m, \dfrac{\sigma}{\sqrt{n}}\right)$.

Exemple 4
Une machine fabrique des pièces en grande série. A chaque pièce tirée au hasard, on associe sa longueur exprimée en millimètres ; on définit ainsi une variable aléatoire X.
On suppose que X suit la loi normal $\mathcal{N}(28,20; 0,027)$.
Soit M_n la variable aléatoire qui à tout échantillon aléatoire de taille n associe la moyenne des longueurs des n pièces de l'échantillon.
La propriété nous dit alors que pour n assez grand, M_n suit la loi normale $\mathcal{N}\left(m, \dfrac{\sigma}{\sqrt{n}}\right)$ soit $\mathcal{N}\left(28,20; \dfrac{0,027}{\sqrt{n}}\right)$.
Supposons que les échantillons soient de taille 100, alors M_{100} suit la loi $\mathcal{N}(28,20; 0,0027)$.

III.3.2 Loi d'échantillonnage des fréquences

On étudie, dans une population de taille N, un caractère X suivant une loi de bernoulli $\mathcal{B}(p)$, c'est-à-dire que les éléments possèdent une certaine propriété d fréquence p.

Dans un échantillon de taille n, on répète n fois la même épreuve de façon indépendante. On obtient n variables aléatoires X_1, X_2, ..., X_n de même loi que X.

La variable aléatoire $f_n = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$ associe à tout échantillon de taille n la fréquence de succès sur cet échantillon.

Propriété 6
La loi d'échantillonnage de taille n de la fréquence f_n pour n « assez grand » peut être approchée par la loi normale $\mathcal{N}\left(p; \sqrt{\dfrac{p(1-p)}{n}}\right)$.

On convient de dire que n est « assez grand » lorsque $n \geq 50$.

Remarque 2
Ce résultat est un cas particulier du précédent en l'appliquant à $m = p$ et $\sigma = \sqrt{p(1-p)}$

Exemple 5
Une urne contient 100 boules numérotées de 1 à 100, indiscernables au toucher. Lors d'un tirage aléatoire d'une boule, la probabilité d'obtenir un nombre inférieur ou égal à 37 est $p = 0,37$. On appelle succès l'événement qui consiste à tirer une des boules numérotées de 1 à 37.
Un échantillon de taille 50 est obtenu par un tirage aléatoire, avec remise, de 50 boules. On s'intéresse à la fréquence d'apparition d'un succès lors du tirage de ces 50 boules.
Soit f_{50} la variable aléatoire qui à chaque échantillon de taille 50 associe sa fréquence de succès.
X_i est la variable aléatoire qui à chaque échantillon associe 1 si la $i-ième$ boule apporte un succès, O sinon. Les X_i sont des variables aléatoires indépendantes et suivent le même loi de Bernoulli de paramètre $p = 0,37$ d'espérance $E(X_i) = 0,37$ et d'écart-type $\sigma(X) = \sqrt{p(1-p)} = 0,48$.
On a $f_{50} = \dfrac{X_1 + X_2 + \cdots + X_{50}}{50}$ qui a pour espérance mathématique $p = 0,37$ et pour écart-type $\sqrt{\dfrac{0,37 \times 0,63}{50}} = 0,068$.