30

Convergences et approximations en probabilités

Dans tout ce chapitre, on fixe un espace probabilisé $(\Omega, \mathcal{A}, P)$.
Toutes les variables aléatoires introduites seront définies sur cet espace probabilisé.

1 Deux inégalités utiles : Markov & Bienaymé-Tchebychev

Commençons par énoncer deux inégalités utiles pour majorer certaines probabilités.

Proposition 1 (Inégalité de Markov)

Soit X une variable aléatoire discrète prenant des valeurs positives et admettant un espérance.

Alors on a : $\forall \lambda > 0, \ \ P(X \geqslant \lambda) \leqslant \dfrac{E(X)}{\lambda}$.

Preuve :

Pour tout $\lambda > 0$,

$$\begin{aligned}\lambda \times P(X \geqslant \lambda) = \lambda \sum_{\substack{x \in X(\Omega) \\ x \geqslant \lambda}} P(X = x) &= \sum_{\substack{x \in X(\Omega) \\ x \geqslant \lambda}} \underbrace{\lambda}_{\leqslant x} P(X = x) \\ &\leqslant \sum_{\substack{x \in X(\Omega) \\ x \geqslant \lambda}} xP(X = x) \leqslant \sum_{x \in X(\Omega)} xP(X = x) = E(X).\end{aligned}$$

On obtient le résultat voulu en divisant par λ. □

Proposition 2 (Inégalité de Bienaymé-Tchebychev)

Soit X une variable aléatoire discrète, admettant une variance.

Alons on a : $\forall \varepsilon > 0, \ \ P\Big(|X - E(X)| \geqslant \varepsilon\Big) \leqslant \dfrac{V(X)}{\varepsilon^2}$.

Preuve :

Puisque X admet une variance, on sait que $Y = (X - E(X))^2$ admet une espérance
(puisque, par définition, $V(X) = E((X - E(X))^2) = E(Y)$)

Pour tout $\varepsilon > 0$, d'après l'inégalité de Markov appliqué à Y :

$$P(|X - E(X)| \geqslant \varepsilon) = P(|X - E(X)|^2 \geqslant \varepsilon^2) = P(Y \geqslant \varepsilon^2) \leqslant \frac{E(Y)}{\varepsilon^2} = \frac{V(X)}{\varepsilon^2}.$$

□

S P O I L E R . . .

Ces deux résultats (en particulier l'inégalité de Bienaymé-Tchebychev) sont au coeur des **statistiques paramétriques**, dont l'objectif est d'estimer un paramètre inconnu à partir d'un échantillon de réalisations d'une variable aléatoire... Ces problèmes d'estimations seront étudiés en deuxième année !

Donnons tout de suite un exemple basique d'estimation.

Exercice 1

On dispose d'une pièce biaisée, donnant Pile avec probabilité $p \in [0,1]$, Face avec probabilité $1-p$.
La valeur de p nous est inconnue : on souhaite l'estimer en "testant" la pièce...
On lance la pièce n fois d'affilée et on observe :

- S_n le nombre de Pile obtenus
- $Z_n = \dfrac{S_n}{n}$ la fréquence d'apparition de Pile.

Intuitivement, quand le nombre de lancers n est grand, la fréquence Z_n se rapproche de la valeur p...

1. Montrer que pour tout $\varepsilon > 0$ et $n \geqslant 1$, $\quad P(|Z_n - p| \geqslant \varepsilon) \leqslant \dfrac{1}{4n\varepsilon^2}$.
2. (a) On effectue $n = 10$ lancers.
   Quelle valeur de ε peut-on choisir pour avoir au moins 90% de chances que $p \in]Z_n - \varepsilon, Z_n + \varepsilon[$?
   (b) On effectue $n = 100$ lancers.
   Avec quelle certitude peut-on affirmer $p \in]Z_n - 0.1, Z_n + 0.1[$?
   (c) Combien de lancers faut-il effectuer pour pouvoir affirmer
   que l'on a $p \in]Z_n - 0.1, Z_n + 0.1[$ avec une certitude d'au moins 90% ?

1. Soient $\varepsilon > 0$ et $n \geqslant 1$.
On sait que $S_n \hookrightarrow B(n,p)$. On a donc $E(S_n) = np$ et $V(S_n) = np(1-p)$.

$$P(|Z_n - p| \geqslant \varepsilon) = P(|\frac{S_n}{n} - p| \geqslant \varepsilon) = P(|S_n - np| \geqslant n\varepsilon) = P\Big(|S_n - E(S_n)| \geqslant n\varepsilon\Big)$$

Ainsi, d'après l'inégalité de Bienaymé-Tchebychev :

$$P(|Z_n - p| \geqslant \varepsilon) \leqslant \frac{V(S_n)}{(n\varepsilon)^2} = \frac{np(1-p)}{n^2\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2}.$$

Enfin, on montre facilement que $\forall p \in [0,1],\ \ p(1-p) \leqslant \dfrac{1}{4}$. Ainsi : $P(|Z_n - p| \geqslant \varepsilon) \leqslant \dfrac{1}{4n\varepsilon^2}$.

2.(a) Dire que $p \in]Z_n - \varepsilon, Z_n + \varepsilon[$ revient à dire $|Z_n - p| < \varepsilon$.
On veut donc que $P(|Z_n - p| < \varepsilon) \geqslant 90\%$, c'est à dire $P(|Z_n - p| \geqslant \varepsilon) \leqslant 10\%$.
Il suffit donc de choisir $\varepsilon > 0$ pour que $\dfrac{1}{4n\varepsilon^2} \leqslant 10\%$:

$$\frac{1}{4n\varepsilon^2} \leqslant 0.1 \iff 4n\varepsilon^2 \geqslant \frac{1}{0.1} = 10 \iff \varepsilon^2 \geqslant \frac{5}{2n} \iff \varepsilon \geqslant \sqrt{\frac{5}{2n}}.$$

Avec ici $n = 10$ lancers, cela donne $\varepsilon \geqslant \sqrt{\dfrac{5}{20}} = \sqrt{\dfrac{1}{4}} = \dfrac{1}{2}$. Au minimum, on peut choisir $\varepsilon = \dfrac{1}{2}$.

Conclusion : Après 10 lancers, on peut affirmer avec 90% de certitude que $p \in]Z_n - 0.5, Z_n + 0.5[$.
Par exemple, si on observe $Z_n = 0.35$, on est sûr à 90% que $p \in]-0.15, 0.85[$ (pas terrible...)

2.(b) Pour $n = 100$ et $\varepsilon = 0.1$:

$$P(|Z_n - p| \geqslant 0.1) \leqslant \frac{1}{4n\varepsilon^2} = \frac{1}{4 \times 100 \times (0.1)^2} = \frac{1}{4}.$$

Ainsi $P(|Z_n - p| \geqslant 0.1) \leqslant 25\%$ et donc $P(|Z_n - p| < 0.1) \geqslant 75\%$.
Conclusion : Après 100 lancers, on a donc une certitude de (au moins) 75% que $p \in]Z_n - 0.1, Z_n + 0.1[$.

(c) On veut avoir $P(|Z_n - p| < 0.1) \geqslant 90\%$, c'est à dire $P(|Z_n - p| \geqslant 0.1) \leqslant 10\%$.
Il suffit de choisir n assez grand pour que $\dfrac{1}{4n\varepsilon^2} \leqslant 10\%$:

$$\frac{1}{4n \times (0.1)^2} \leqslant 0.1 \iff 4n \times 0.01 \geqslant 10 \iff 4n \geqslant 1000 \iff \underline{n \geqslant 250}.$$ (il faut faire au moins 250 lancers)

Remarque : Si l'on veut 90% de chances d'avoir $p \in]Z_n - 0.01, Z_n + 0.01[$, il faut faire 25000 lancers...

2 Loi des grands nombres

L'exercice précédent suggère que, lorsque le nombre d'expériences répétées indépendamment augmente, la **fréquence empirique** d'apparition d'un évènement converge vers la **probabilité "réelle"** de l'évènement.

On peut énoncer un résultat plus général encore : la **moyenne empirique**, construite à partir d'une succession de variables aléatoires indépendantes de même loi, converge vers la moyenne "réelle" : l'**espérance**.

Théorème 1 (Loi (faible) des grands nombres)

Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires discrètes mutuellement indépendantes et de même loi.

On suppose qu'elles admettent une espérance que l'on note $m = E(X_1) = E(X_2) = \ldots$ etc.
On suppose qu'elles admettent également une variance.

Pour tout $n \in \mathbb{N}^*$, on définit la **moyenne empirique** : $\overline{X_n} = \dfrac{X_1 + X_2 + \ldots + X_n}{n} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i.$

Alors $\overline{X_n}$ converge vers l'espérance m quand n tends vers l'infini, au sens où :

$$\forall \varepsilon > 0, \quad P(|\overline{X_n} - m| \geqslant \varepsilon) \xrightarrow[n\to+\infty]{} 0.$$

Preuve :

Puisque les X_i sont de même loi, elles ont la même espérance et également la même variance.
On note donc $\sigma^2 = V(X_1) = V(X_2) = \ldots$ etc. $\quad$ (σ est l'écart type de $X_1, X_2, \ldots$ etc.)
Par linéarité de l'espérance,

$$E(\overline{X_n}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} m = m.$$

Puisque $X_1, X_2, \ldots, X_n$ sont mutuellement indépendantes, on a $V\left(\displaystyle\sum_{i=1}^{n} X_i\right) = \displaystyle\sum_{i=1}^{n} V(X_i)$, donc :

$$V(\overline{X_n}) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} V\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{\sigma^2}{n}.$$

Pour tout $\varepsilon > 0$, l'inégalité de Bienaymé-Tchebychev pour $\overline{X_n}$ s'écrit :

$$P(|\overline{X_n} - E(\overline{X_n})| \geqslant \varepsilon) \leqslant \frac{V(\overline{X_n})}{\varepsilon^2} \qquad \text{c'est à dire :} \quad P(|\overline{X_n} - m| \geqslant \varepsilon) \leqslant \frac{\sigma^2}{n\varepsilon^2}$$

Puisque $\dfrac{\sigma^2}{n\varepsilon^2} \xrightarrow[n\to+\infty]{} 0$, on en déduit par Théorème des gendarmes que $P(|\overline{X_n} - m| \geqslant \varepsilon) \xrightarrow[n\to+\infty]{} 0.$ $\square$

Pour $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoire mutuellement indépendantes, de loi de Bernoulli $\mathcal{B}(p)$, on obtient en particulier le résultat suivant (similaire à celui étudié dans l'Exercice 1) :

Corollaire 1 (Convergence des fréquences empirique vers la probabilité réelle)

On répète indéfiniment et indépendamment une épreuve de Bernoulli de probabilité de succès $p \in [0,1]$.

Pour tout $n \in \mathbb{N}^*$, on pose : $X_n = \begin{cases} 1 & \text{si succès à la } n\text{-ème épreuve} \\ 0 & \text{si échec à la } n\text{-ème épreuve.} \end{cases}$

Ainsi, $(X_n)_{n\in\mathbb{N}^*}$ est une suite de variables aléatoires mutuellement indépendantes, de loi $\mathcal{B}(p)$.

$\overline{X_n} = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$ s'interprète comme la fréquence de succès sur n épreuves.

Cette fréquence de succès converge vers la probabilité p quand n tends vers l'infini, au sens où :

$$\forall \varepsilon > 0, \quad P(|\overline{X_n} - p| \geqslant \varepsilon) \xrightarrow[n\to+\infty]{} 0.$$

Remarques 1

• La loi des grands nombres (et en particulier le corollaire sur les fréquences empiriques) est un résultat fondamental de la théorie des probabilités. Lorsque le nombre n d'épreuves répétées est très grand, on a $\overline{X_n} \simeq p$: la fréquence empirique des succès (aléatoire) se rapproche de la "probabilité théorique" p d'obtenir un succès (déterministe !)

Ceci vient, *a posteriori*, justifier la notion intuitive que l'on peut avoir d'une probabilité : dans la limite où le nombre d'expériences répétées est grand, on retrouve en quelque sorte la formule d'équi-probabilité :

$$P(A) \simeq \frac{\text{"Nombre de cas favorables"}}{\text{"Nombre de cas total"}}$$

• **Remarque hors programme :**

Il existe des versions plus générales de l'énoncé de la loi des grands nombres. D'abord, l'hypothèse d'existence de variance n'est en fait pas nécessaire (mais la preuve devient plus complexe !). De plus :

Le résultat annoncé ici est la loi "faible" des grands nombres : $\forall \varepsilon > 0,\ P(|\overline{X_n} - m| \geqslant \varepsilon) \underset{n\to+\infty}{\longrightarrow} 0.$

Il existe une version "forte" : $\overline{X_n} \underset{n\to+\infty}{\longrightarrow} m$ presque-sûrement, c'est à dire : $P\Big(\lim\limits_{n\to+\infty} \overline{X_n} = m\Big) = 1.$
(ce n'est pas tout à fait la même chose...)

3 Approximation Binomiale-Poisson (loi des évènements rares)

Évoquons un dernier résultat, plus anecdotique, mais qui prépare le terrain pour la notion de "convergence en loi" qui sera étudiée l'année prochaine :
"Dans la limite où n est grand, la loi binomiale $\mathcal{B}(n, \frac{\lambda}{n})$ peut être approchée par une loi de Poisson $\mathcal{P}(\lambda)$."

Théorème 2 (Loi des évènements rares)

Soit $\lambda > 0$.
Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires telle que : $\forall n \geqslant 1,\ X_n \hookrightarrow \mathcal{B}\left(n, \dfrac{\lambda}{n}\right)$.

(notons que l'on a bien $\frac{\lambda}{n} \in [0,1]$ au moins pour n assez grand...)

Alors, pour tout $k \in \mathbb{N}$: $P(X_n = k) \underset{n\to+\infty}{\longrightarrow} P(X = k)$, où X est une variable aléatoire de loi $\mathcal{P}(\lambda)$.

Preuve :

Soit $k \in \mathbb{N}$ fixé. Pour tout $n \geqslant 1$, comme $X_n \hookrightarrow \mathcal{B}(n, \frac{\lambda}{n})$, on a $P(X_n = k) = \dbinom{n}{k}\left(\dfrac{\lambda}{n}\right)^k\left(1 - \dfrac{\lambda}{n}\right)^{n-k}$.
(avec la convention habituelle $\dbinom{n}{k} = 0$ si $k > n$).

$\boxed{1}$ $\dbinom{n}{k} = \dfrac{\overbrace{n(n-1)\dots(n-k+1)}^{k \text{ termes}}}{k!} \underset{n\to+\infty}{\sim} \dfrac{n^k}{k!}$ donc $\dbinom{n}{k}\left(\dfrac{\lambda}{n}\right)^k \underset{n\to+\infty}{\sim} \dfrac{n^k}{k!}\dfrac{\lambda^k}{n^k} = \dfrac{\lambda^k}{k!}$

c'est à dire : $\dbinom{n}{k}\left(\dfrac{\lambda}{n}\right)^k \underset{n\to+\infty}{\longrightarrow} \dfrac{\lambda^k}{k!}$.

$\boxed{2}$ $\left(1 - \dfrac{\lambda}{n}\right)^{n-k} = e^{(n-k)\ln(1-\frac{\lambda}{n})}$.

On sait que $\ln(1+x) \underset{x\to0}{\sim} x$, donc $\ln\left(1 - \dfrac{\lambda}{n}\right) \underset{n\to+\infty}{\sim} -\dfrac{\lambda}{n}$. De plus $(n-k) \underset{n\to+\infty}{\sim} n$.

Ainsi : $(n-k)\ln\left(1 - \dfrac{\lambda}{n}\right) \underset{n\to+\infty}{\sim} n \times -\dfrac{\lambda}{n} = -\lambda$, c'est à dire $(n-k)\ln\left(1 - \dfrac{\lambda}{n}\right) \underset{n\to+\infty}{\longrightarrow} -\lambda$.

Par composition de limite (exp est continue), $\left(1 - \dfrac{\lambda}{n}\right)^{n-k} \underset{n\to+\infty}{\longrightarrow} e^{-\lambda}$.

<u>Conclusion :</u> $P(X_n = k) \underset{n\to+\infty}{\longrightarrow} \dfrac{\lambda^k}{k!}e^{-\lambda} = P(X = k)$ où $X \hookrightarrow \mathcal{P}(\lambda)$. □

Conséquence : une interprétation de la loi de Poisson

Un bureau de poste accueille en moyenne $\lambda > 0$ clients en une heure.

On souhaite proposer un modèle mathématique pour décrire le nombre (aléatoire) X de clients qui se présentent au guichet au cours d'une heure. Une façon naturelle de procéder est la suivante :

• On segmente l'heure en 4 portions :

Tous les quarts d'heures, un client décide de rentrer (ou pas!) avec probabilité p.

Si X désigne le nombre de clients qui sont rentrés au bout d'une heure, on a : $X \hookrightarrow \mathcal{B}(4, p)$..

Comme on veut que $E(X) = 4p = \lambda$ il faut choisir $p = \frac{\lambda}{4}$.

Pour ce premier modèle : $X \hookrightarrow \mathcal{B}(4, \frac{\lambda}{4})$. Mais on peut être plus précis...

• On segmente l'heure en 60 portions :

Toutes les minutes, un client décide de rentrer (ou pas!) avec probabilité p.

Si X désigne le nombre de clients qui sont rentrés au bout d'une heure, on a : $X \hookrightarrow \mathcal{B}(60, p)$.

Comme on veut que $E(X) = 60p = \lambda$ il faut choisir $p = \frac{\lambda}{60}$.

Pour ce deuxième modèle : $X \hookrightarrow \mathcal{B}(60, \frac{\lambda}{60})$. Mais on peut être plus précis...

• On segmente l'heure en 3600 portions :

Toutes les secondes, un client décide de rentrer (ou pas!) avec probabilité p.

Si X désigne le nombre de clients qui sont rentrés au bout d'une heure, on a : $X \hookrightarrow \mathcal{B}(3600, p)$.

Comme on veut que $E(X) = 3600p = \lambda$ il faut choisir $p = \frac{\lambda}{3600}$.

Pour ce troisième modèle : $X \hookrightarrow \mathcal{B}(3600, \frac{\lambda}{3600})$.

En segmentant une heure en n morceaux, on aboutit donc à : $X \hookrightarrow \mathcal{B}(n, \frac{\lambda}{n})$.

Dans la limite où $n \to +\infty$ (segmentation "infinie" : un client peut débarquer à chaque instant), on choisit naturellement comme modèle $\boxed{X \hookrightarrow \mathcal{P}(\lambda)}$.

On comprend mieux cette appellation de **loi des évènements rares** : à chaque instant fixé, la probabilité de voir arriver un client est très faible ($= \frac{\lambda}{n}$), mais ceci est "compensé" par le fait que le nombre d'instants où une arrivée est possible est très grand ($= n$).

La loi de Poisson est ainsi particulièrement adaptée pour modéliser le nombre d'occurrences d'un évènement au cours d'un intervalle de temps donné.

À savoir faire à l'issue de ce chapitre :

Au minimum

- Retenir l'inégalité de Markov et l'inégalité de Bienaymé-Tchebychev.
- Connaître l'énoncé de la loi faible des grands nombres.
- Connaître la Loi des évènements rares.

Pour suivre

- Bien comprendre le problème d'estimation de l'Exercice 1.
- Se convaincre que la loi des grand nombres est un résultat "naturel".

Pour les ambitieux

$\{ \emptyset$