

Markov, Bienaymé et Tchebychev...

Vous êtes nulle ou nul en probabilités ? Ouf, je me sens moins seul !!! Nous allons donc ensemble, en prenant notre temps, étudier ces trois individus...

I Rappels : espérance, variance et écart-type.

On travaille ici avec X , une variable aléatoire discrète pouvant prendre les valeurs notées x_i (en nombre fini ou infini).

Lorsque X admet une espérance $E(X)$, elle se calcule selon : $E(X) = \sum_i x_i p(X = x_i)$.

Remarque : la « moyenne » représente le côté concret (calculable en vrai) de la notion d'espérance.

L'espérance est linéaire, ce qui se traduit par les relations :

$$E(X + Y) = E(X) + E(Y) \quad E(kX) = k E(X) \quad \text{avec } k \text{ un réel.}$$

Cette propriété permet en particulier ceci :

Soient n variables aléatoires admettant la même espérance μ , alors, la variable aléatoire Y définie par $Y = \frac{X_1 + X_2 + \dots + X_n}{n}$ admet pour espérance μ . En effet, on peut écrire :

$$E(Y) = E\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{1}{n} E(X_1 + \dots + X_n) = \frac{1}{n} (E(X_1) + \dots + E(X_n)) = \mu$$

Lorsque X admet une variance $V(X)$, elle vaut : $V(X) = E\left(\left(X - E(X)\right)^2\right)$. On peut aussi écrire : $V(X) = E(X^2) - \left(E(X)\right)^2$.

La variance permet d'apprécier la façon dont la variable X est « dispersée » par rapport à sa moyenne. L'écart-type σ est la racine carrée de la variance : $\sigma = \sqrt{V(X)}$ ou $\sigma^2 = V(X)$.

Attention, la variance n'est pas linéaire.

On a : $V(X + Y) = V(X) + V(Y) + 2 \text{Cov}(X, Y)$ où $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$

$$V(kX) = k^2 V(X)$$

En revanche, si X et Y sont deux variables *indépendantes*, on a : $V(X + Y) = V(X) + V(Y)$

II Inégalité de Markov

Soit X une variable aléatoire positive. Alors on peut écrire : $\forall a > 0, p(X \geq a) \leq \frac{E(X)}{a}$

X admet bien évidemment une espérance pour cette situation ! Cette relation permet une majoration simple (donc très imparfaite...) de la probabilité $p(X \geq a)$. Petite démo ?

Cette démonstration permet de mettre le nez dans le côté théorique de ce domaine. Elle n'est pas à savoir pas cœur, mais il est intéressant d'essayer de la suivre, elle est abordable...

On a par définition de l'espérance : $E(X) = \sum_i x_i p_i$ où $p_i = p(X = x_i)$

On peut séparer la somme en deux parties selon les $x_i \geq a$ et les $x_i < a$.

Ainsi : $E(X) = \sum_{x_i \geq a} x_i p_i + \sum_{x_i < a} x_i p_i$ or, si $x_i \geq a$ on aura $x_i p_i \geq a p_i$

D'où : $E(X) \geq \sum_{x_i \geq a} a p_i + \sum_{x_i < a} x_i p_i = a \sum_{x_i \geq a} p_i + \sum_{x_i < a} x_i p_i$

Avec $\sum_{x_i \geq a} p_i = p(X \geq a)$

On peut donc remplacer et obtenir : $E(X) \geq a p(X \geq a) + \sum_{x_i < a} x_i p_i$

Le terme $\sum_{x_i < a} x_i p_i$ étant positif, on a donc a plus forte raison : $E(X) \geq a p(X \geq a)$

(en effet, si l'on a $x \geq 3 + 4$, il y a de fortes chances pour que $x \geq 3$!!).

Finalement, avec la condition $a > 0$, on peut retrouver la relation $p(X \geq a) \leq \frac{E(X)}{a}$.

Exemple :

Tirage d'un dé non truqué à six faces. Soit X la variable aléatoire donnant le numéro sorti après un lancer. On peut calculer l'espérance : $E(X) = \frac{1+2+3+4+5+6}{6} = 3,5$.

On a deux possibilités sur six d'obtenir $X \geq 5$. Donc : $p(X \geq 5) = \frac{2}{6} = \frac{1}{3}$ (valeur exacte).

On peut comparer cette valeur avec la majoration de Markov $p(X \geq 5) \leq \frac{3,5}{5}$ qui donne $p(X \geq 5) \leq 0,7$. L'inégalité de Markov dit finalement : $0,333 \leq 0,7$. Vous voyez dans ce cas simple que l'inégalité est assez large !!

Remarque :

Dès que la valeur de a est inférieure à $E(X)$, l'inégalité de Markov est inutile. En effet, on obtient dans un tel cas : $p(X \geq a) \leq \frac{E(X)}{a}$ et comme $a \leq E(X)$, cela donne $1 \leq \frac{E(X)}{a}$.

L'inégalité dit alors que la probabilité $p(X \geq a)$ sera inférieure à une valeur du genre 1,2. Autant dire qu'on le savait déjà !!!

Conclusion :

L'inégalité de Markov sert à majorer de manière large la probabilité $p(X \geq a)$ dans les cas où la valeur a est plus grande que l'espérance de X . Elle est donc plutôt utile pour évaluer la probabilité de dépassement de X du côté de ses plus grandes valeurs.



Andreï Markov (1856-1922) s'est spécialisé à la fin de sa vie dans le domaine des probabilités. On lui doit la notion de chaîne de Markov qui étudie la dynamique de phénomènes aléatoires. Le futur peut être étudié via la connaissance unique de l'état précédent, et non de tous les états précédents. On parle de phénomènes aléatoires à mémoire courte.

III Inégalité de Bienaymé-Tchebychev

Soit X une variable aléatoire admettant une espérance notée μ (histoire de changer un peu...) et une variance notée σ^2 . Alors : $\forall \varepsilon > 0, p(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$

Cette très intéressante relation permet de relier la notion d'écart-type de manière tangible avec le monde des probabilités. On a sous les yeux une majoration de la probabilité que la variable X s'écarte de son espérance d'une quantité plus grande que ε . On voit concrètement que la valeur de l'écart-type σ joue un rôle dans la dispersion des valeurs autour de μ .



Non, Bienaymé n'est pas le prénom de Tchebychev ! Ce sont deux scientifiques différents...

Irénée-Jules Bienaymé (1796-1878) est un mathématicien français qui a contribué à l'essor de la théorie des probabilités ainsi qu'à ses applications. Il a été l'un des membres fondateurs de la Société mathématique de France.

Pafnouti Tchebychev (1821-1894) est un mathématicien russe très connu pour ses ignobles polynômes ! Des sujets de concours étudient les propriétés nombreuses de ces polynômes. Les deux premiers sont $P_0 = 1$ et $P_1 = 2X$ et les suivants répondent à la relation de récurrence : $P_{n+1} = 2XP_n - P_{n-1}$. Il établit une « loi des grands nombres » très générale (d'après Wikipédia...).



Un ou deux petits exemples de Bienaymé-Tchebychev ?

1) Revenons à notre dé à six faces...

On va estimer $p(|X - \mu| \geq 2)$ et la comparer avec la fameuse inégalité.

$|X - \mu| \geq 2$ est équivalent à $X = 1$ ou $X = 6$. Donc : $p(|X - \mu| \geq 2) = \frac{2}{6} = \frac{1}{3}$.

On peut ensuite calculer la variance de X : $V(X) = \sum_i x_i^2 p_i - \mu^2 \approx 2,92$

On obtient alors : $p(|X - \mu| \geq 2) \leq \frac{2,92}{2^2} \approx 0,73$. Autant dire que cette majoration n'est pas très performante... On le comprend au sens où X est une variable uniforme, donc la variance est assez grande ce qui limite la pertinence de notre inégalité.

On pourrait tricher un peu en écrivant : $p(|X - \mu| \geq 2,49) = \frac{1}{3} \leq \frac{2,92}{2,49^2} \approx 0,47$. Cela améliore la qualité de l'inégalité...

2) Soit X une variable aléatoire suivant une loi binomiale $\mathcal{B}(n = 20 ; p = \frac{1}{2})$.

On peut associer X au nombre de fois où une pièce donne « Pile » après 20 lancers.

On peut calculer : $\mu = np = 10$ et $\sigma^2 = np(1-p) = 5$. Ainsi : $p(|X - 10| \geq k) \leq \frac{5}{k^2}$.

Pour $k = 8$, on obtient : $p(|X - 10| \geq 8) \leq 0,08$.

D'un autre côté : $|X - 10| \geq 8$ donne $X \leq 2$ ou $X \geq 18$. Alors : $p(|X - 10| \geq 8) \approx 4 \cdot 10^{-4}$.

Le majorant assez faible (0,08) montre qu'il y a relativement peu de chance que X prenne une valeur très éloignée de μ . Cette information est intéressante. Néanmoins, encore une fois, la comparaison entre la valeur exacte et le majorant montre une grande différence.

Une démonstration de l'inégalité de Bienaymé-Tchebychev (reléguée ici car un peu délicate...)

Soit t un réel positif.

On va démontrer dans un premier temps : $p(\mu - t\sigma \leq X \leq \mu + t\sigma) \geq 1 - \frac{1}{t^2}$

Pour la suite, on va noter :

x_i^{inf} les événements tels que $X \leq \mu - t\sigma$ (*inf* pour « inférieur »)

x_i^{ent} les événements tels que $\mu - t\sigma \leq X \leq \mu + t\sigma$ (*ent* pour « entre »)

x_i^{sup} les événements tels que $\mu + t\sigma \leq X$ (*sup* pour « supérieur »)

Alors :

$$\sigma^2 = \sum_i (x_i - \mu)^2 p_i = \sum_{i_{inf}} (x_i^{inf} - \mu)^2 p_i + \sum_{i_{ent}} (x_i^{ent} - \mu)^2 p_i + \sum_{i_{sup}} (x_i^{sup} - \mu)^2 p_i$$

Comme $\sum_i (x_i^{ent} - \mu)^2 p_i \geq 0$, on a : $\sigma^2 \geq \sum_{i_{inf}} (x_i^{inf} - \mu)^2 p_i + \sum_{i_{sup}} (x_i^{sup} - \mu)^2 p_i$

On utilise ensuite les relations : $(x_i^{inf} - \mu)^2 \geq t^2 \sigma^2$ et de même $(x_i^{sup} - \mu)^2 \geq t^2 \sigma^2$

$$\text{Donc : } \sigma^2 \geq \sum_{i_{inf}} t^2 \sigma^2 p_i + \sum_{i_{sup}} t^2 \sigma^2 p_i$$

$$\text{On obtient : } \sigma^2 \geq t^2 \sigma^2 \left(\sum_{i_{inf}} p_i + \sum_{i_{sup}} p_i \right) = t^2 \sigma^2 (1 - p(\mu - t\sigma \leq X \leq \mu + t\sigma))$$

Ainsi : $1 \geq t^2 (1 - p(\mu - t\sigma \leq X \leq \mu + t\sigma))$ d'où le $p(\mu - t\sigma \leq X \leq \mu + t\sigma) \geq 1 - \frac{1}{t^2}$

En posant $\varepsilon = t\sigma$, on obtient : $p(\mu - \varepsilon \leq X \leq \mu + \varepsilon) = p(|X - \mu| \leq \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}$

Avec l'événement contraire, on obtient : $1 - p(|X - \mu| \geq \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}$

Et donc finalement : $p(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$

IV Loi faible des grands nombres

Soient n variables aléatoires X_i suivant la même loi de probabilité, admettant la même espérance μ . La variable aléatoire définie par $Y = \frac{X_1+X_2+\dots+X_n}{n}$ admet alors pour espérance μ (revoir le I !). Les X_i peuvent être liées à la mesure répétée d'une même expérience.

Si les n variables X_i sont *indépendantes*, alors d'après le I, on peut calculer la variance de Y selon : $V(Y) = V\left(\frac{X_1+X_2+\dots+X_n}{n}\right) = \frac{1}{n^2}V(X_1 + X_2 + \dots + X_n) = \frac{1}{n^2}(V(X_1) + \dots + V(X_n))$

Au final, en notant σ^2 la variance des X_i , on obtient : $V(Y) = \frac{\sigma^2}{n}$.

L'inégalité de Bienaymé-Tchebychev conduit alors à : $p(|Y - \mu| \geq \varepsilon) \leq \frac{V(Y)}{\varepsilon^2}$ que l'on peut réécrire sous : $p\left(\left|\frac{X_1+X_2+\dots+X_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}$ (ceci est une inégalité de concentration)

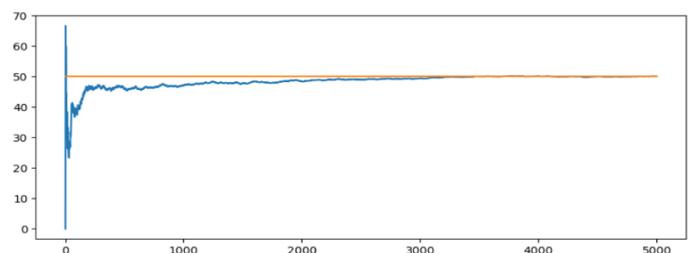
En faisant tendre n vers l'infini : $\forall \varepsilon > 0, \lim_{n \rightarrow +\infty} p\left(\left|\frac{X_1+X_2+\dots+X_n}{n} - \mu\right| \geq \varepsilon\right) = 0$

On dit que la suite de variables $Y_n = \frac{X_1+X_2+\dots+X_n}{n}$ converge en probabilité vers μ .

Interprétation : en répétant un grand nombre de fois une même expérience aléatoire, il est rare que la moyenne des valeurs effectivement mesurées s'écarte de manière significative de l'espérance théorique. Mais, notons que cela peut arriver.

Tout le monde a constaté cette loi en lançant un très grand nombre de fois, noté N , une pièce et en remarquant qu'on obtenait un nombre, noté N_{Pile} , de « Pile » qui est très proche de $\frac{N}{2}$. Plus N est grand, plus cette constatation est remarquable (néanmoins, on peut très bien tomber sur une affreuse série improbable de 100 000 lancers comportant seulement 45 % de « Pile » !).

Ci-contre, j'ai tracé l'évolution du pourcentage de « Pile » durant 5 000 tirages aléatoires (« aléatoires » selon le générateur aléatoire de l'ordinateur...). On se rapproche bien progressivement de la barre des 50 %.



Info : J'ai lancé (en vrai) 1000 fois un dé à 6 faces et ai obtenu une moyenne de 3,4 points (3,5 en théorie).

V Loi forte des grands nombres

On peut aller plus loin avec, dans les mêmes conditions que IV, le théorème suivant :

$$p\left(\lim_{n \rightarrow \infty} \frac{X_1 + X_2 + \dots + X_n}{n} = \mu\right) = 1$$

Il annonce cette fois *avec certitude* que nous pourrions connaître l'espérance μ en faisant un nombre de plus en plus grand de mesures indépendantes d'une variable X .