Feuille d'exercices n^o 4

Inégalités de Markov et de Bienaymé-Tchebychev, loi des grands nombres

Exercice 1. Le nombre de pièces sortant d'une usine en une journée est une variable aléatoire d'espérance 50. On veut estimer la probabilité que la production de demain dépasse 75 pièces.
a) En utilisant l'inégalité de Markov, quelle estimation obtient-on sur cette probabilité ?
b) Que peut-on dire de plus sur cette probabilité si on sait que l'écart-type de la production quotidienne est 5 ?
Solution : a) Inégalité de Markov : $P(X \geq 75) \leq \frac{E(X)}{75} = \frac{2}{3}$.
b) Inégalité de Tchebychev : $P(|X-50| \geq 25) \leq \frac{\mathrm{Var}(X)}{25^2} = \frac{5^2}{25^2} = 0.04$. *Donc* $P(X \geq 75) \leq 0.04$.

Exercice 2. Pour étudier les particules émises par une substance radioactive, on dispose d'un détecteur. On note X la variable aléatoire représentant le nombre de particules qui atteignent le détecteur pendant un intervalle de temps Δt. Le nombre maximal de particules que le détecteur peut compter pendant un intervalle de temps Δt est de 10^3. On suppose que X suit une loi de Poisson de paramètre $\lambda = 10^2$. Donner une majoration de la probabilité que X dépasse 10^3.
Caractéristiques d'une loi de Poisson $\mathcal{P}(\lambda)$ *: son espérance et sa variance sont égales à* λ.
Solution : Inégalité de Tchebychev : $P(|X-10^2| \geq 10^3 - 10^2) \leq \frac{Var(X)}{(10^3-10^2)^2} = \frac{100}{900^2} \simeq 0.00012$. *Donc* $P(X \geq 10^3) \leq 0.00013$.

Exercice 3. On lance n fois un dé à 6 faces et on regarde la fréquence d'obtention de la face "6" (c'est-à-dire le nombre de fois qu'on obtient "6", divisé par n). Que peut-on dire de cette fréquence quand n devient grand ? (si on parle de limite, préciser en quel sens)
Solution : Loi des grands nombres : si S_n *est la v.a. comptant le nombre de "6",* $\frac{S_n}{n}$ *tend vers* $1/6$ *en proba (loi faible) ou p.s. (loi forte).*

Théorème Central Limite, intervalles de confiance

Exercice 4. On considère l'expérience consistant à lancer 100 fois une pièce (équilibrée) et on note S la variable aléatoire comptant le nombre de "pile" obtenu lors d'une expérience. Que vaut $P(40 \leq S \leq 60)$? Quelle est la probabilité pour que S soit supérieur à 60 ?
Solution : TCL (en écrivant S *comme somme de 100 v.a. Bernoulli indépendantes) ou approximation d'une binomiale par une loi normale :* $\frac{S-50}{5}$ *est approximativement de loi* $\mathcal{N}(0,1)$.
$P(40 \leq S \leq 60) = P(-2 \leq \frac{S-50}{5} \leq 2) \simeq 0.954$ *(table). Par symétrie,* $P(S > 60) = P(S < 40)$, *d'où* $P(S > 60) = \frac{1}{2}(1 - P(40 \leq S \leq 60)) \simeq 0.023$.

Exercice 5. Selon une étude, 20% des consommateurs se déclarent influencés par la marque lors d'un achat. Si on interroge 100 consommateurs pris au hasard, quelle est la probabilité pour qu'au moins 28 d'entre eux se déclarent influencés par la marque ?
Solution : Soit F *la v.a. "proportion de consommateurs dans l'échantillon de taille* $n = 100$ *qui se déclarent influencés par la marque". On cherche à calculer* $P(F \geq 0.28)$.
nF *suit la loi binomiale* $\mathcal{B}(n,p)$ *avec* $n = 100$ *et* $p = 0.2$, *donc* $E(F) = 0.2$ *et* $\sigma(F) = \sqrt{\frac{0.2\times 0.8}{100}} = 0.04$. *Soit* $T = \frac{F-0.2}{0.04}$, T *suit à peu près une loi normale* $\mathcal{N}(0,1)$ *(TCL ou approximation d'une binomiale par une loi normale).* $P(F \geq 0.28) = P(F - 0.2 \geq 0.08) = P(T \geq 2)$. *Or, avec la table,* $P(T < 2) \simeq 0.5 + 0.477 = 0.977$, *d'où* $P(F \geq 0.28) \simeq 1 - 0.977 = 0.023$.
Conclusion. : Il y a environ 2,3% de chance pour que plus de 28 consommateurs dans un échantillon de 100 personnes se disent influencés par la marque.

Exercice 6. La firme Comtec vient de développer un nouvel appareil électronique. On veut en estimer la fiabilité en termes de durée de vie. D'après une étude, l'écart-type de la durée de vie d'un appareil serait de l'ordre de 100 heures. On suppose également que la durée de vie suit une loi normale, et que les durées de vie de différents appareils sont indépendantes.
Déterminer le nombre d'essais requis pour estimer, avec un niveau de confiance de 95%, la durée de vie moyenne d'une grande production de sorte que la marge d'erreur dans l'estimation n'excède pas ± 20 heures.
Solution : La variable X = *"durée de vie d'un appareil" suit une loi normale d'espérance* m *(inconnue) et d'écart-type* $\sigma = 100$. *Si on fait* n *essais indépendants donnés par les v.a.* $X_1, \ldots, X_n$, *on note* $\bar{X} = \frac{X_1 + \cdots + X_n}{n}$. *Alors* $T = \sqrt{n}\frac{\bar{X}-m}{\sigma}$ *suit une loi normale* $\mathcal{N}(0,1)$ *(la loi de* T *est exactement* $\mathcal{N}(0,1)$ *en utilisant la propriété qu'une somme de lois normales indépendantes est une loi normale ; si on utilise le TCL, on conclut que la loi de* T *est approximativement* $\mathcal{N}(0,1)$ *si* n *est assez grand). Donc* $p(|T| \leq 1.96) \geq 0.95$ *(table). Au niveau de confiance 95%, on peut affirmer que l'intervalle* $I = [\bar{x} - 1.96\frac{100}{\sqrt{n}}, \bar{x} + 1.96\frac{100}{\sqrt{n}}]$ *contient la durée de vie moyenne cherchée* m, *où* $\bar{x}$ *est la valeur effectivement mesurée de* $\bar{X}$, *c'est-à-dire* $P(m \in I) \geq 0.95$ *(*$\bar{x} = \bar{X}(\omega)$, *où* ω *représente la réalisation d'une série d'expériences donnée).*
La marge d'erreur n'excède pas 20h dès que $1.96\frac{100}{\sqrt{n}} < 20$, *i.e. si* $n \geq 97$.
Remarque : m *est inconnu mais fixé. C'est l'intervalle* $[\bar{X} - 1.96\frac{100}{\sqrt{n}}, \bar{X} + 1.96\frac{100}{\sqrt{n}}]$ *qui varie selon l'expérience car* $\bar{X}$ *dépend de* ω.

Exercice 7. On interroge 1000 électeurs, 521 déclarent vouloir voter pour le candidat A. Indiquer avec une probabilité de 0.95 entre quelles limites se situe la proportion du corps électoral favorable à A au moment du sondage.
Solution : On note p *la proportion (inconnue) d'électeurs favorables à A. Soit* $n = 1000$. *Soit* X_i *la variable qui vaut 1 si le i-ème électeur interrogé déclare vouloir voter pour A, 0 sinon. Les* X_i *suivent une loi de Bernoulli* $b(p)$ *; on suppose que ces v.a. sont indépendantes. Donc le nombre* $S = \sum_{i=1}^{n} X_i$ *suit une loi binomiale* $\mathcal{B}(n,p)$ *;* $E(S) = np$ *et* $\mathrm{Var}(S) = np(1-p)$. *Comme* n *est grand, on peut approcher la loi de* $T = \frac{S-np}{\sqrt{np(1-p)}}$ *par* $\mathcal{N}(0,1)$ *(TCL ou approximation d'une binomiale par une loi normale). Donc* $P(|T| \leq 1.96) \geq 0.95$ *(table). On a*

$$\begin{aligned} P(|T| \leq 1.96) &= P(-1.96 \leq \frac{S-np}{\sqrt{np(1-p)}} \leq 1.96) = P\left(-1.96\sqrt{\frac{p(1-p)}{n}} \leq \frac{S}{n} - p \leq 1.96\sqrt{\frac{p(1-p)}{n}}\right) \\ &= P\left(\frac{S}{n} - 1.96\sqrt{\frac{p(1-p)}{n}} \leq p \leq \frac{S}{n} + 1.96\sqrt{\frac{p(1-p)}{n}}\right) \end{aligned}$$

Comme n *est grand, on peut estimer* p *par la fréquence* $f = 0.521$ *observée dans l'échantillon (loi des grands nombres), et approximer* $p(1-p)$ *par* $f(1-f)$, *ce qui donne*

$$P\left(\frac{S}{n} - 1.96\sqrt{\frac{f(1-f)}{n}} \leq p \leq \frac{S}{n} + 1.96\sqrt{\frac{f(1-p)}{f}}\right) \geq 0.95.$$

Lors du sondage, on trouve $\frac{S}{n} = f$ *(c'est-à-dire* $\frac{S(\omega)}{n} = f$ *où* ω *correspond au sondage réalisé), ce qui donne l'intervalle de confiance* $[f - 1.96\sqrt{\frac{f(1-f)}{n}}, f + 1.96\sqrt{\frac{f(1-f)}{n}}] \simeq [0.49, 0.55]$. *On conclut que* $[0.49, 0.55]$ *(autrement dit* $52 \pm 3\%$*) est un intervalle de confiance relatif à* p *au seuil de 95% (c'est-à-dire* $P(0.49 \leq p \leq 0.55) \geq 0.95$*).*
Variante : plutôt que d'estimer $p(1-p)$ *par* $f(1-f)$, *on peut utiliser* $p(1-p) \leq \frac{1}{4}$ *(inégalité*

vraie pour tout $p \in [0,1]$), ce qui donne

$$P\left(\frac{S}{n} - 1.96\frac{0.5}{\sqrt{n}} \leq p \leq \frac{S}{n} + 1.96\frac{0.5}{\sqrt{n}}\right) \geq P\left(\frac{S}{n} - 1.96\sqrt{\frac{p(1-p)}{n}} \leq p \leq \frac{S}{n} + 1.96\sqrt{\frac{p(1-p)}{n}}\right) \geq 0.95$$

d'où l'intervalle de confiance $[f - 1.96\frac{0.5}{\sqrt{n}}, f + 1.96\frac{0.5}{\sqrt{n}} \simeq [0.49, 0.55]$.
Remarque : en majorant $p(1-p)$ par $\frac{1}{4}$, on trouve a priori un intervalle plus grand que le précédent, avec une différence d'autant plus grande que p est loin de 0.5. Ici, p est proche de 0.5 et on ne voit pas la différence avec 2 décimales.

Exercices supplémentaires

Exercice 8. On considère une marche aléatoire sur $\mathbb{Z}$ définie de la façon suivante : on part de 0 et, à chaque étape, on a une probabilité p de faire un pas vers la droite et une probabilité $1-p$ de faire un pas vers la gauche. Autrement dit, on considère une suite de variables aléatoires $(X_n)_{n\geq 1}$, indépendantes, de même loi donnée par

$$\forall n \in \mathbb{N}^*, \ P(X_n = 1) = p \quad \text{et} P(X_n = -1) = 1 - p,$$

et $S_n = X_1 + X_2 + \cdots + X_n$ représente la position sur $\mathbb{Z}$ de la marche aléatoire à l'étape n.
a) Que vaut $\lim\limits_{n\to+\infty} \dfrac{S_n}{n}$? (on précisera de quel type de limite on parle)
b) On suppose que $p > \frac{1}{2}$. En utilisant la loi forte des grands nombres, montrer que S_n tend presque sûrement vers $+\infty$.
Solution : a) $E(X_i) = p - (1-p) = 2p - 1$. Loi des grands nombres : $\frac{S_n}{n}$ tend vers $E(X_1)$ en proba, c'est-à-dire

$$\forall \varepsilon > 0, \lim_{n\to+\infty} P\left(\left|\frac{S_n}{n} - (2p-1\right| > \varepsilon\right) = 0.$$

Loi forte des grands nombres $\frac{S_n}{n}$ tend vers $E(X_1)$ presque sûrement, c'est-à-dire

$$\exists \Omega' \subset \Omega, P(\Omega') = 1, \forall \omega \in \Omega', \lim_{n\to+\infty} \frac{S_n(\omega)}{n} = 2p - 1.$$

b) $p > \frac{1}{2}$ donc $2p - 1 > 0$. Soit $0 < \delta < 2p - 1$. Par b), $\forall \omega \in \Omega', \exists N, \forall n \geq N, S_n(\omega) \geq n\delta$, donc $S_n(\omega) \to +\infty$. Autrement dit S_n tend presque sûrement. vers $+\infty$.

Exercice 9. On effectue un sondage sur un échantillon de 10000 personnes à la veille d'un référendum : 4903 d'entre elles s'apprêtent à voter oui, et 5097 à voter non. On note p la proportion (inconnue) de personnes dans la population s'apprêtant à voter oui. Donner un intervalle de confiance à 95% pour p.
Solution : Idem que l'exercice 7 avec $n = 10000$ et $f = 0.4903$, ce qui donne l'intervalle de confiance $[f - 1.96\sqrt{\frac{f(1-f)}{n}}, f + 1.96\sqrt{\frac{f(1-f)}{n}}] \simeq [0.4805, 0.5001]$. On conclut que $[0.4805, 0.5001]$ (autrement dit $49 \pm 1\%$) est un intervalle de confiance relatif à p au seuil de 95%.
Si on utilise $p(1-p) \leq \frac{1}{4}$, on trouve le même l'intervalle de confiance si on se limite à 4 décimales ou moins.

Exercice 10. Chaque jour, un train subit un retard aléatoire au départ, évalué en minutes. On modélise la loi du retard par une loi exponentielle $\mathcal{E}(\lambda)$ (mais on ignore la valeur de λ). On suppose également que les retards sont indépendants entre eux. Sur 400 jours, le retard moyen est de 10 minutes. Donner un intervalle de confiance de niveau approximativement 95% pour λ.
Caractéristiques d'une loi exponentielle $\mathcal{E}(\lambda)$: son espérance et son écart-type valent $1/\lambda$.

Solution : Loi estimée $\mathcal{E}(\lambda)$ d'espérance $m = 1/\lambda$ et d'écart-type $\sigma = 1/\lambda$. TCL : $P(\frac{\sqrt{n}|S_n/n-m|}{\sigma} < c) \simeq P(|\mathcal{N}(0,1)| < c)$. Pour avoir un intervalle de confiance à 95%, on prend $c = 1.96$. L'encadrement $-c < \frac{\sqrt{n}(S_n/n-1/\lambda)}{\sigma} < c$ donne

$$\frac{1}{S_n/n + c\sigma/\sqrt{n}} < \lambda < \frac{1}{S_n/n - c\sigma/\sqrt{n}}.$$

Avec les données : $\frac{S_n(\omega)}{n} = 10$. C'est un estimateur de m. Donc on approxime λ par $1/10$ et σ par 10. D'où l'intervalle de confiance $[0.091, 0.111]$.

Exercice 11. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes, suivant toutes la loi de Bernoulli de paramètre p. Pour tout $n \geq 1$, on note $S_n = X_1 + \cdots + X_n$. On fixe $\varepsilon > 0$.
a) En utilisant l'inégalité de Bienaymé-Tchebychev, minorer $P\left(p - \varepsilon < \frac{S_n}{n} < p + \varepsilon\right)$.

b) Déterminer a_n (dépendant de n, p et ε) pour que les encadrements suivants soient équivalents :

$$-a_n < \frac{S_n - np}{\sqrt{np(1-p)}} < a_n \Longleftrightarrow -\varepsilon < \frac{S_n}{n} - p < \varepsilon.$$

On suppose que n est assez grand pour appliquer le théorème central limite. En déduire une approximation de $P\left(p - \varepsilon < \frac{S_n}{n} < p + \varepsilon\right)$ (sous forme d'intégrale faisant intervenir la densité de la loi normale).
c) Application numérique : comparer les estimations obtenues en a) et b) pour $p = \frac{1}{2}$, $n = 400$ et $\varepsilon = 0.05$.
Solution : a) $E(\frac{S_n}{n}) = p$ et $\mathrm{Var}(\frac{S_n}{n}) = \frac{p(1-p)}{n}$. D'où $P(|\frac{S_n}{n} - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2}$ (inégalité de Bienaymé-Tchebychev) et donc $P\left(p - \varepsilon < \frac{S_n}{n} < p + \varepsilon\right) \geq 1 - \frac{p(1-p)}{n\varepsilon^2}$.
b)

$$\begin{aligned}
-a_n < \frac{S_n - np}{\sqrt{np(1-p)}} < a_n \quad &\Longleftrightarrow \quad -a_n\sqrt{np(1-p)} < S_n - np < a_n\sqrt{np(1-p)} \\
&\Longleftrightarrow \quad -a_n\sqrt{\frac{p(1-p)}{n}} < \frac{S_n}{n} - p < a_n\sqrt{\frac{p(1-p)}{n}}
\end{aligned}$$

On veut $a_n\sqrt{\frac{p(1-p)}{n}} = \varepsilon$, il faut donc prendre $a_n = \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}$.
On note f la densité de $\mathcal{N}(0,1)$ ($f(x) = \frac{1}{\sqrt{2\pi}}\exp(-x^2/2)$, mais on n'a pas besoin de connaître l'expression de f). Pour n grand, le TCL donne

$$P\left(p - \varepsilon < \frac{S_n}{n} < p + \varepsilon\right) = P\left(-a_n < \frac{S_n - np}{\sqrt{np(1-p)}} < a_n\right) \simeq \int_{-a_n}^{a_n} f(x)\,dx$$

c) application numérique : $p(1-p) = 1/4$.
Avec a), $P\left(p - \varepsilon < \frac{S_n}{n} < p + \varepsilon\right) \geq 0.75$.
Avec b), $P\left(p - \varepsilon < \frac{S_n}{n} < p + \varepsilon\right) \simeq \int_{-2}^{2} f(x)\,dx \simeq 0.954$ (table).
La 2ème estimation est donc bien plus précise.

Exercice 12. On considère le nombre de garçons parmi n naissances choisies au hasard. On suppose que, pour chaque naissance, la probabilité que ce soit un garçon est $p = 0.514$, et que les naissances sont indépendantes entre elles. A partir de quelle valeur de n y a-t-il une probabilité inférieure à 1% pour que le nombre de filles soit supérieur ou égal au nombre de garçons ?
Solution : Soit X le nombre de garçons. X suit une loi binomiale $B(n,p)$. Le nombre de filles est supérieur ou égal au nombre de garçons si $X \leq \frac{n}{2}$. On cherche n tel que $P(X \leq \frac{n}{2}) \leq 0.01$

ou, de façon équivalente, $P(X > \frac{n}{2}) \geq 0.99$. *On suppose que* n *est assez grand pour approximer* $B(n,p)$ *par une loi normale, c'est-à-dire que* $T = \frac{X-np}{\sqrt{np(1-p)}}$ *suit presque la loi* $\mathcal{N}(0,1)$. *On a*

$$X > \frac{n}{2} \Longleftrightarrow T > \frac{(0.5-p)\sqrt{n}}{p(1-p)}.$$

Vu la table et la symétrie de $\mathcal{N}(0,1)$, $P(T > -2.326) = 0.5 + P(T < 2.326) = 0.99$. *Donc* $P(X > \frac{n}{2})$ *équivaut à*

$$\frac{(0.5-p)\sqrt{n}}{p(1-p)} < -2.326 \Leftrightarrow \frac{(p-0.5)\sqrt{n}}{p(1-p)} > 2.326 \Leftrightarrow n > 2.326^2 \frac{p(1-p)}{(p-0.5)^2} \simeq 6895.5$$

Il faut donc prendre $n \geq 6896$.

Exercice 13. On considère l'intervalle $[0,1]$ muni de la probabilité uniforme. Si $x \in [0,1]$, on note $X_1(x), X_2(x), \ldots, X_n(x), \ldots$ les chiffres du développement décimal de x. On admet que les v.a. $(X_n)_{n\geq 1}$ sont indépendantes de loi uniforme dans l'ensemble des chiffres $\{0,1,\ldots,9\}$.
Un nombre $x \in [0,1]$ est dit normal si, pour tout $k \in \{0,1,\ldots,9\}$, le chiffre k apparaît avec une proportion $\frac{1}{10}$ dans le développement décimal de x, c'est-à-dire $\lim_{n\to+\infty} \frac{N_n(x)}{n} = \frac{1}{10}$, où $N_n(x) = \mathrm{Card}\{1 \leq i \leq n \mid X_i(x) = k\}$. Montrer qu'un nombre est normal avec probabilité 1.

Table (partielle) pour une v.a. X **de loi normale** $\mathcal{N}(0,1)$

t	$P(0 \leq X \leq t)$	$P(-t \leq X \leq t)$
0.6	0.226	0.451
0.68	**0.25**	**0.5**
0.8	0.288	0.576
1.26	0.396	0.792
1.32	0.407	0.813
1.645	0.45	**0.90**
1.96	0.475	**0.95**
2	0.477	0.954
2.326	0.49	0.98