

Probabilité
Fiche TD 19 - réponses possibles
Les punaises préfèrent-elles tomber sur le dos ?
Intervalle de confiance d'une estimation

Lorsqu'on vide une boîte de punaises sur la table, on a l'impression qu'une punaise manifeste une "préférence" p à tomber sur le dos plutôt que sur la pointe.

1. Proposer un modèle probabiliste pour décrire cette situation.

Réponse possible : lorsqu'on écrit "une punaise", on pense bien entendu à une punaise générique, pas à une punaise particulière. Le modèle auquel je pense alors, c'est celui d'une *loi de probabilité*. Comme on veut rendre compte qu'il y a deux états possible (sur le dos ou sur la pointe) et on pense donc à une loi à deux états, qu'il est commode de *coder* par des nombres (et non des lettres, par exemple) ce qui permettra aussi de calculer. Si l'on code par 0 et 1, cela permet en suite de représenter facilement l'activité de comptage. On code l'état "la punaise est sur le dos" par 1, et par 0 l'état "la punaise est sur la pointe". Une loi de Bernoulli $\mathcal{B}(1, p)$, avec p chargé de représenter une (hypothétique) "préférence à tomber sur le dos", semble alors toute désignée pour modéliser la situation envisagée.

2. On vide une boîte de n punaises sur la table et on compte le nombre D_n de punaises tombées sur le dos. Donner un modèle probabiliste pour D_n .

Réponse possible : On pense tout naturellement à modéliser l'état de la i -ème punaise parmi les n punaises par une v.a. $X_i \sim \mathcal{B}(1, p)$. On imagine que les positions adoptées par les punaises sont indépendantes l'une de l'autre ; on suppose donc les X_i *indépendantes* et on pose $D_n = X_1 + \dots + X_n$, qui compte le nombre de X_i codant l'état "la punaise i est tombée sur le dos".

Notons ici que si le modèle se révélait aboutir à des résultats en contradiction avec l'expérience, il y aurait au moins deux points qui semblaient "aller de soit" : l'indépendance des X_i : si les punaises étaient fortement magnétisées, cette indépendance serait sans doute une hypothèse inefficace ; plus subtilement, ce modèle sous-entend qu'on peut individualiser les diverses punaises, puisqu'on les numérote. La physique quantique fournit des exemples où cette hypothèse en apparence anodine serait nuisible.

3. Donner un estimateur P_n (ou $\hat{P}_n$) de p .

Réponse possible : Par estimateur de p , nous entendons une v.a. $P_n = \hat{P}_n = \theta(X_1, \dots, X_n)$ telle que p soit limite (dans un sens à préciser) des P_n . La *loi des grands nombres* (LGN) assure que $p = \mathbb{E}(X_i)$ est *limite en probabilité* des $\frac{1}{n}(X_1 + \dots + X_n)$ pour des X_i i.i.d. . Nous posons donc $P_n = \frac{1}{n}D_n = \frac{1}{n}(X_1 + \dots + X_n)$ pour des $X_i \sim \mathcal{B}(1, p)$ indépendantes

4. Ecrire P_n sous la forme $P_n = aZ_n + b$, où Z_n est proche en loi d'une v.a. $Z \sim \mathcal{N}(0, 1)$; que valent alors a et b (en fonction de p et n) ?

Réponse possible : Le théorème limite central (TLC) assure que, sous les hypothèses ici vérifiées (X_i i.i.d. telles que $\sigma^2 = \text{Var}(X_i)$ existe), il y a *convergence en probabilité* des $Z_n := \frac{1}{\sigma\sqrt{n}}(X_1 + \dots + X_n - n\mathbb{E}(X_1))$ vers $Z \sim \mathcal{N}(0, 1)$. Ici $\mathbb{E}(X_1) = p$ et $\sigma^2 = p(1 - p)$, et $D_n = X_1 + \dots + X_n = nP_n$. On a donc $P_n = p + \frac{\sigma}{\sqrt{n}}Z_n$, et donc $a = \frac{\sigma}{\sqrt{n}}$ et $b = p$. L'application pratique du TLC consiste donc à assimiler Z_n à sa limite Z . On assimile donc P_n à $\hat{P}_n = p + \frac{\sigma}{\sqrt{n}}Z$

5. On a vidé une boîte de 54 punaises et on a trouvé $d_{54} = 30$ punaises sur le dos ; quelle est l'estimation $\hat{p}$ de p issue de cette expérience ?

Réponse possible : Cela conduit à poser $\hat{p} = p_{54} = \frac{d_{54}}{54} = 0,555\dots = 55,5\%$.

6. On cherche un *intervalle de confiance au seuil* $\alpha = 5\%$ du type $[P_n - \Delta, P_n + \Delta]$, c'est-à-dire tel que $\mathbb{P}(\{p \notin [P_n - \Delta, P_n + \Delta]\}) \leq \alpha = 5\%$. Montrer tout d'abord que

$$\{p \in [P_n - \Delta, P_n + \Delta]\} = \{P_n \in [p - \Delta, p + \Delta]\}.$$

Réponse possible : On a $\{p \in [P_n - \Delta, P_n + \Delta]\} = \{P_n - \Delta \leq p, P_n + \Delta \geq p\} = \{P_n \leq p + \Delta, P_n \geq p - \Delta\} = \{P_n \in [p - \Delta, p + \Delta]\}$

7. Trouver z_- et z_+ tels que $\{P_n \in [p - \Delta, p + \Delta]\} = \{Z_n \in [z_-, z_+]\}$.

Réponse possible : On cherche z_- et z_+ tels que $\{Z_n \in [z_-, z_+]\} = \{aZ_n + b \in [p - \Delta, p + \Delta]\}$. La relation $aZ_n + b \geq p - \Delta$ équivaut à $Z_n \geq \frac{1}{a}(p - \Delta - b) = \frac{\sqrt{n}}{\sigma}(p - \Delta + p) = -\frac{\sqrt{n}}{\sqrt{p(1-p)}}\Delta =: z_-$; on obtiendrait de même $z_+ := \frac{\sqrt{n}}{\sqrt{p(1-p)}}\Delta$.

8. En déduire finalement une approximation (majoration) Δ de $\Delta_\alpha = \Delta_{5\%}$.

Réponse possible : Par définition de Δ , il suffit de choisir $\Delta \geq \Delta_\alpha$, avec $0,05 = 5\% = \alpha = \mathbb{P}(\{p \notin [P_n - \Delta_\alpha, P_n + \Delta_\alpha]\}) = \mathbb{P}(\{Z_n \notin [z_-, z_+]\})$. Le TLC nous conduit à assimiler, pour n assez grand ($n \geq 30\dots$), $\mathbb{P}(\{Z_n \notin [z_-, z_+]\})$ à $\mathbb{P}(\{Z \notin [z_-, z_+]\})$. Comme $z_- = -z_+$ et que la densité de Z est paire, ceci équivaut donc à Δ_α tel que $F(z_+) = \mathbb{P}(\{Z \geq z_+\} = 0,025 = F(1,96\dots)$) d'après la table de la loi normale $\mathcal{N}(0,1)$; d'où $\frac{\sqrt{n}}{\sigma}\Delta_\alpha = z_+ = 1,96\dots$, et donc $\Delta_\alpha = \frac{\sqrt{p(1-p)}}{\sqrt{n}}1,96\dots \leq \frac{1}{2\sqrt{n}}1,96\dots =: \Delta'_\alpha$ (puisque $p(1-p) \leq \frac{1}{4}$). Comme $n = 54$, on obtient qu'il suffit de choisir $\Delta \geq \Delta'_\alpha = \frac{1}{2\sqrt{54}}1,96\dots = 0,0133\dots = 1,3\dots\%$

9. Peut-on affirmer qu'au vu de l'expérience, *au seuil $\alpha = 5\%$, les punaises préfèrent tomber sur le dos ?*

Réponse possible : Non, puisque l'intervalle de confiance $55,5\% \pm 13,3\%$ ne comporte pas que des nombres supérieurs à $0,5 = 50\%$. Le choix d'une valeur plus élevée pour n (par exemple $n = 1000$) conduirait à un Δ plus petit, et permettrait peut-être d'avoir un intervalle de confiance entièrement composé de nombres supérieurs à 50% , mais il se peut aussi que p_{1000} soit plus proche de 50% : cela dépendra du résultat de l'expérience !

10. Même question si on avait $n = 540$ et $d_{540} = 299$.

Réponse possible : Ici on obtient $\Delta'_\alpha = \frac{1,96}{2\sqrt{540}} = \frac{1}{\sqrt{10}}0,0133\dots\% = 4,2\dots\%$ et $p_{540} = \frac{299}{540} = 55,37\dots\%$; dans ces conditions on peut affirmer, au seuil de 5% (c'est-à-dire avec un "risque inférieur à 5% " d'être démenti par d'autres expériences) que "les punaises préfèrent tomber sur le dos".



Edwin Moore, inventeur, en 1900, de la punaise :