Exercice 1 Test d'hypothèses
Sur 53680 familles de lapins ayant 8 petits (soit 429440 petits en tout), il y a 221023 mâles et 208417 femelles. On souhaite savoir si le nombre de mâles est significativement plus élevé que celui des femelles.

1. Si l'on suppose que les chances sont égales d'avoir un mâle ou une femelle, quelle loi suit la v.a. "le ième petit est un mâle"
2. Grâce au théorème de la limite centrée, déterminer x_0 tel que la v.a. S représentant le nombre de mâles vérifie (puis conclure)

$$\mathbb{P}(214720 - x_0 \leq S \leq 214720 + x_0) \geq 0,95$$

Exercice 2 On veut tester l'hypothèse mendelienne : yeux bleus (gènes) récessifs et yeux marrons dominants.

1. Si l'hypothèse est vraie, qu'elle la probabilité pour qu'une personne choisie au hasard ait les yeux bleus.
2. Si l'hypothèse mendelienne est valide, combien de personnes doit on observer pour être certain avec une probabilité de 99,8% que la proportion de personnes aux yeux marrons sera comprise entre 0,7 et 0,8.

Exercice 3 Soit p la proportion de la population favorable à l'adoption d'un texte de loi (donc $1 - p$ la proportion de personnes hostile à ce texte).
On souhaite grâce à un sondage avoir une estimation de p, pour cela on intéroge un échantillon de n personnes et on prend comme approximation de p le nombre $\tilde{p}$ de personnes favorables à ce texte dans l'échantillon.
Déterminer, grâce au théorème de la limite centrée, la taille de l'échantillon pour que $\tilde{p}$ approche p à 10^{-2} prés avec une probabilité de 0,95.

Exercice 4 A précision donnée quelle taille d'échantillon ?
On reprends les notations de l'exercice précédent. On souhaite déterminer le nombre de personnes à interroger pour ce sondage afin d'approcher p à ϵ prés avec une probabilité α. En remarquant que $p(1-p) \leq 1/4$ calculer

1. la taille n de l'échantillon pour approcher p à $0,01$ avec une probabilité de 95%.
2. la taille n de l'échantillon pour approcher p à $0,01$ avec une probabilité de 99%.

Solution Remarque préliminaire :

$$(2p-1)^2 = 4p^2 - 4p + 1 \quad \Rightarrow \quad 4p^2 - 4p + 1 \geq 0 \quad \Rightarrow \quad p(1-p) \leq 1/4$$

Soit X_i la v.a. de Bernoulli représentant le vote de la personne i ; X_i vaut 1 si cette personne est favorable au texte et 0 sinon (le paramètre de cette v.a. de Bernoulli est p). On notera n la taille de l'échantillon et S la v.a. représentant le nombre de personnes favorables au texte : $S = \sum_{i=1}^{n} X_i$.

La proportion de personnes favorables à ce texte dans l'échantillon est donnée par

$$\tilde{p} = \frac{S}{n}$$

$\tilde{p}$ est appelé un estimateur de p.
On cherche n pour que

$$\mathbb{P}(|p - \tilde{p}| \leq 10^{-2}) = 0.95 \quad \Rightarrow \quad \mathbb{P}(p - 10^{-2} \leq \tilde{p} \leq p + 10^{-2}) = 0.95$$

or d'aprés le théorème central-limite pour n grand (ou la proposition 1) $S \sim \mathcal{N}(np; \sqrt{np(1-p)})$ d'où

$$\begin{aligned} \mathbb{P}(p - 10^{-2} \leq \frac{S}{n} \leq p + 10^{-2}) &= \mathbb{P}\left(-\frac{10^{-2}\sqrt{n}}{\sqrt{p(1-p)}} \leq \frac{S - np}{\sqrt{np(1-p)}} \leq \frac{10^{-2}\sqrt{n}}{\sqrt{p(1-p)}}\right) \\ &= \mathbb{P}(-\alpha \leq Z \leq \alpha) \end{aligned}$$

où l'on a posé $\alpha = \dfrac{10^{-2}\sqrt{n}}{\sqrt{p(1-p)}}$ et $Z \sim \mathcal{N}(0,1)$.
On détermine d'abord α grâce à la table :

$$\mathbb{P}(-\alpha \leq Z \leq \alpha) = 2F_Z(\alpha) - 1 = 0.95 \quad \Rightarrow \quad \alpha = 1.96$$

et finalement on a

$$\begin{aligned} 10^{-2}\frac{\sqrt{n}}{\sqrt{p(1-p)}} = 1.96 \quad &\Rightarrow \quad \sqrt{n} = 1.96 \times 10^2 \sqrt{p(1-p)} \\ &\quad \text{et grâce à la remarque préliminaire} \\ &\Rightarrow \quad n \geq 196^2 \times \tfrac{1}{4} \\ &\Rightarrow \quad n \geq 9604 \end{aligned}$$

Il faut donc interroger au moins 9604 personnes pour que $\tilde{p}$ approche p à 10^{-2}.

2. On reprend exactement le même raisonnement avec $\epsilon = 0,01$ et une probabilité de 99%

Exercice 5 On reprends les notations de l'exercice 3. On souhaite, à taille d'échantillon fixée, trouver la précision obtenue. En remarquant que $p(1-p) \leq 1/4$ calculer

1. la précision ϵ sachant que l'échantillon est de taille $n = 1000$ et pour un seuil de confiance de $0,95$ (autrement dit qu'elle est la précision obtenue dans 95% des cas si l'échantillon est de taille $n = 1000$).
2. la précision ϵ sachant que l'échantillon est de taille $n = 1000$ et pour un seuil de confiance de $0,99$

Solution 1. On reprend les notations de la démonstration de l'exercice précédent et on fait le même type de raisonnement, cette fois-ci n est donné et on cherche la précision ϵ pour un seuil de confiance donné.
On cherche ϵ pour que

$$\mathbb{P}(|p - \tilde{p}| \leq \epsilon) = 0.95 \quad \Rightarrow \quad \mathbb{P}(p - \epsilon \leq \tilde{p} \leq p + \epsilon) = 0.95$$

or d'aprés le théorème central-limite (ou la proposition 1) $S \sim \mathcal{N}(np; \sqrt{np(1-p)}$ d'où

$$\begin{aligned}\mathbb{P}(p - \epsilon \leq \frac{S}{1000} \leq p + \epsilon) &= \mathbb{P}\left(-\frac{\epsilon\sqrt{1000}}{\sqrt{p(1-p)}} \leq \frac{S - 1000p}{\sqrt{1000p(1-p)}} \leq \frac{\epsilon\sqrt{1000}}{\sqrt{p(1-p)}}\right) \\ &= \mathbb{P}(-\alpha \leq Z \leq \alpha)\end{aligned}$$

où l'on a posé $\alpha = \dfrac{\epsilon\sqrt{1000}}{\sqrt{p(1-p)}}$ et $Z \sim \mathcal{N}(0,1)$.
On détermine d'abord α grâce à la table :

$$\mathbb{P}(-\alpha \leq Z \leq \alpha) = 2F_Z(\alpha) - 1 = 0.95 \quad \Rightarrow \quad \alpha = 1.96$$

et finalement on a

$$\begin{aligned}\epsilon\frac{\sqrt{1000}}{\sqrt{p(1-p)}} = 1.96 \quad &\Rightarrow \quad \epsilon = \frac{1.96 \times \sqrt{p(1-p)}}{\sqrt{1000}} \\ &\qquad \text{et grâce à la remarque préliminaire de l'exo précédent} \\ &\Rightarrow \quad \epsilon \leq 1.96 \times 0.5 \times \tfrac{1}{\sqrt{1000}} \\ &\Rightarrow \quad \epsilon \leq 0.03\end{aligned}$$

2. même raisonnement en changeant les données.

Exercice 6 Le dé est-il biaisé ?
Weldon a lancé un dé 315 672 fois et tiré 106 602 fois l'une des faces 5 ou 6.

1. Calculer les fréquences théorique et observée d'apparition des faces 5 ou 6.
2. Calculer la fluctuation (différence entre les fréquences observée et théorique)
3. Quelle loi représente cette expérience aléatoire.
4. En utilisant le théorème central-limite, montrer que l'hypothèse d'un dé équilibré est à rejeter