I. Contexte :

Dans une population on veut **estimer** la proportion p d'un caractère.

Exemple : Avant la commercialisation d'un médicament, on veut estimer la proportion de patients susceptibles de présenter des effets secondaires.

II. Procédé :

1. Dans un échantillon de la population, on calcule la fréquence f du caractère.

Exemple : Lors d'essais cliniques avant la commercialisation du médicament, on a observé que sur un échantillon de 3 000 personnes, 59 ont présenté des effets secondaires.

La fréquence observée est donc : $f = \dfrac{59}{3\ 000} \approx 0,02$

2. A partir de la fréquence f obtenue, on définit un **intervalle de confiance** I qui dépend de f et de la taille n de l'échantillon.

L'intervalle de confiance donne une **estimation de p par encadrement**

Propriété :

Dans un échantillon de taille n, on note f_n la fréquence des succès.
Pour n assez grand, la probabilité que p appartienne à l'intervalle $\left[f_n - \dfrac{1}{\sqrt{n}}\ ;\ f_n + \dfrac{1}{\sqrt{n}}\right]$ est supérieure à 0,95

Définition :

Soit f une fréquence observée du caractère étudié sur un échantillon de taille n.
$\left[f - \dfrac{1}{\sqrt{n}}\ ;\ f + \dfrac{1}{\sqrt{n}}\right]$ est appelé **intervalle de confiance de la proportion p au seuil de confiance 95 %.**

On utilise cet intervalle dès que : $n \geqslant 30$; $nf \geqslant 5$ et $n(1-f) \geqslant 5$

Dans notre exemple : A quelle proportion de patients présentant des effets secondaires peut-on s'attendre après commercialisation du médicament avec un seuil de confiance de 95 % ? $f = \ldots\ldots\ldots$

$n = \ldots\ldots\ldots$; $nf = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$ et $n(1-f) = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$

L'intervalle de confiance au seuil de 95% est .....

.....

.....

.....

Exercice 1

Une association de consommateurs décide d'estimer la proportion de personnes satisfaites par l'utilisation d'une crème de soin. Elle réalise un sondage parmi les personnes utilisant ce produit. Sur 140 personnes interrogées, 99 se déclarent satisfaites. Estimer, par intervalle de confiance au seuil de 95 %, la proportion de personnes satisfaites parmi les utilisateurs de cette crème.

Exercice 2 *Les résultats seront arrondies au millième.*

La Hadopi (Haute Autorité pour la diffusion des œuvres et la Protection des droits sur Internet) souhaite connaître la proportion en France de jeunes âgés de 16 à 24 ans pratiquant au moins une fois par semaine le téléchargement illégal sur internet. Pour cela, elle envisage de réaliser un sondage.

Mais la Hadopi craint que les jeunes interrogés ne répondent pas tous de façon sincère. Aussi, elle propose le protocole $(\mathcal{P})$ suivant :

On choisit aléatoirement un échantillon de jeunes âgés de 16 à 24 ans.

Pour chaque jeune de cet échantillon :

- le jeune lance un dé équilibré à 6 faces ; l'enquêteur ne connaît pas le résultat du lancer ;
- l'enquêteur pose la question : « Effectuez-vous un téléchargement illégal au moins une fois par semaine ? » ;

- si le résultat du lancer est pair alors le jeune doit répondre à la question par « Oui » ou « Non » de façon sincère ;
- si le résultat du lancer est « 1 » alors le jeune doit répondre « Oui » ;
- si le résultat du lancer est « 3 ou 5 » alors le jeune doit répondre « Non ».

Grâce à ce protocole, l'enquêteur ne sait jamais si la réponse donnée porte sur la question posée ou résulte du lancer de dé, ce qui encourage les réponses sincères.

On note p la proportion inconnue de jeunes âgés de 16 à 24 ans qui pratiquent au moins une fois par semaine le téléchargement illégal sur internet.

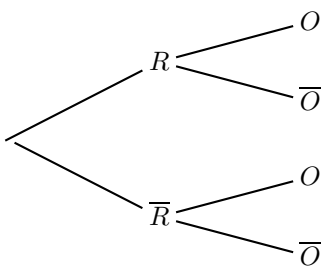
1. *Calculs de probabilités*

On choisit aléatoirement un jeune faisant parti du protocole $(\mathcal{P})$.

On note : R l'évènement « le résultat du lancer est pair »,

O l'évènement « le jeune a répondu Oui ».

Compléter l'arbre pondéré :



En déduire que la probabilité q de l'évènement « le jeune a répondu Oui » est :

$$q = \frac{1}{2}p + \frac{1}{6}$$

2. *Intervalle de confiance*

a. À la demande de l'Hadopi, un institut de sondage réalise une enquête selon le protocole $(\mathcal{P})$. Sur un échantillon de taille 1 500, il dénombre 625 réponses « Oui ».

Donner un intervalle de confiance, au niveau de confiance de 95 %, de la proportion q de jeunes qui répondent « Oui » à un tel sondage, parmi la population des jeunes français âgés de 16 à 24 ans.

b. Que peut-on en conclure sur la proportion p de jeunes qui pratiquent au moins une fois par semaine le téléchargement illégal sur internet ?