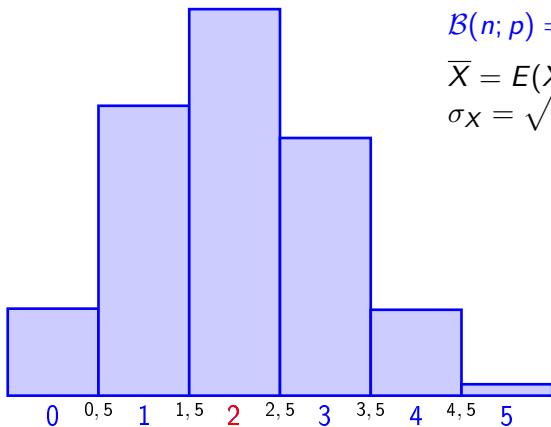


Approximation de la loi Binomiale

I. Approximation $\mathcal{B} \simeq \mathcal{N}$

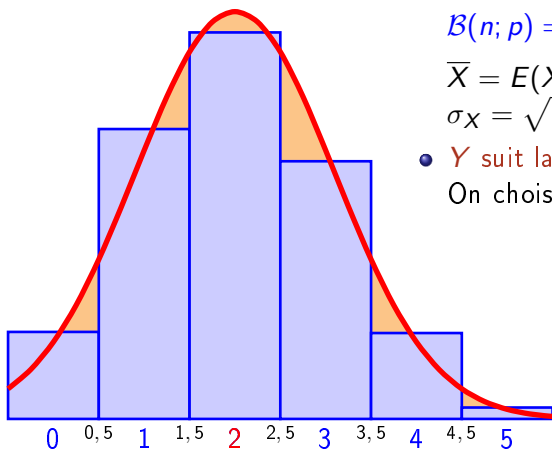


- X suit la loi binomiale

$$\mathcal{B}(n; p) = \mathcal{B}(5; 40\%)$$

$$\bar{X} = E(X) = n \cdot p = 2$$

$$\sigma_X = \sqrt{n \cdot p \cdot (1 - p)} \approx 1,1$$



- X suit la loi binomiale

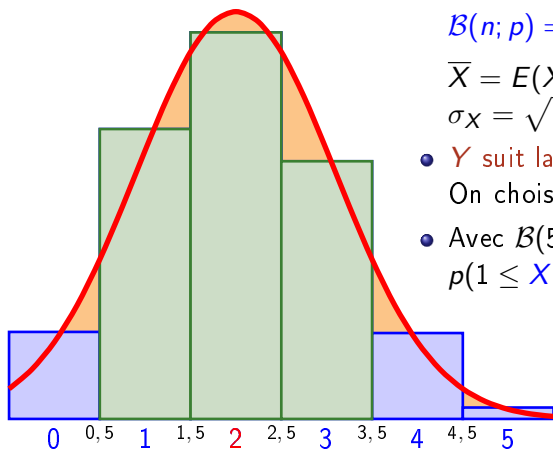
$$\mathcal{B}(n; p) = \mathcal{B}(5; 40\%)$$

$$\bar{X} = E(X) = n \cdot p = 2$$

$$\sigma_X = \sqrt{n \cdot p \cdot (1 - p)} \approx 1,1$$

- Y suit la loi normale $\mathcal{N}(\mu; \sigma^2)$

On choisit $\bar{Y} = \mu = \bar{X}$; $\sigma = \sigma_Y = \sigma_X$



- X suit la loi binomiale

$$\mathcal{B}(n; p) = \mathcal{B}(5; 40\%)$$

$$\bar{X} = E(X) = n \cdot p = 2$$

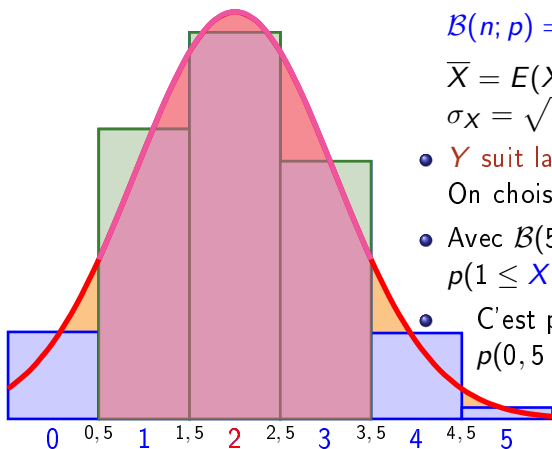
$$\sigma_X = \sqrt{n \cdot p \cdot (1 - p)} \approx 1,1$$

- Y suit la loi normale $\mathcal{N}(\mu; \sigma^2)$

On choisit $\bar{Y} = \mu = \bar{X}$; $\sigma = \sigma_Y = \sigma_X$

- Avec $\mathcal{B}(5; 40\%)$, on peut calculer :

$$p(1 \leq X \leq 3) = 0,8352$$



- X suit la loi binomiale

$$\mathcal{B}(n; p) = \mathcal{B}(5; 40\%)$$

$$\bar{X} = E(X) = n \cdot p = 2$$

$$\sigma_X = \sqrt{n \cdot p \cdot (1 - p)} \approx 1,1$$

- Y suit la loi normale $\mathcal{N}(\mu; \sigma^2)$

On choisit $\bar{Y} = \mu = \bar{X}$; $\sigma = \sigma_Y = \sigma_X$

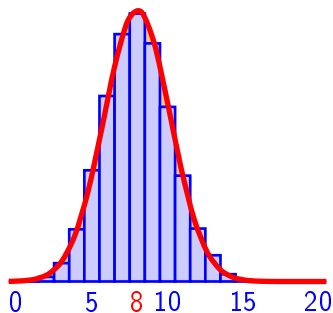
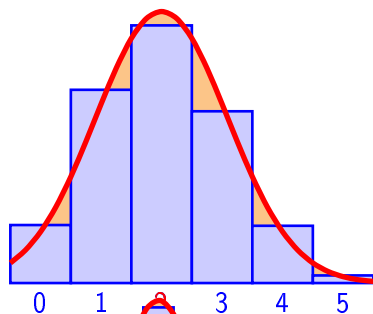
- Avec $\mathcal{B}(5; 40\%)$, on peut calculer :

$$p(1 \leq X \leq 3) = 0,8352$$

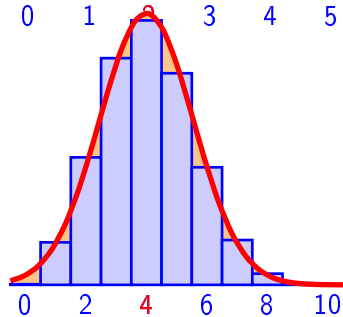
- C'est proche de

$$p(0,5 \leq Y \leq 3,5) \approx 0,8273$$

L'approximation s'améliore quand $n \nearrow$



n successivement : 5, 10 et 20.



- Avec X , on veut calculer des probabilités. Quand on veut remplacer X par Y , l'important est que Y produise les mêmes valeurs de probabilités.
- C'est en regardant comment évolue \mathcal{B} quand $n \nearrow$ que l'on a découvert la loi \mathcal{N} .

Autre exemple avec la calculette

X suit $\mathcal{B}(500; 40\%)$

- $\bar{X} = E(X) = 500 \times 0,4 = 200$ et

$$\sigma(X) = \sqrt{100 \times 0,4 \times 0,6} = \sqrt{120} \approx 10,95$$

Disons alors que Y suit $\mathcal{N}(\mu; \sigma^2)$ avec $\mu = 200$ et $\sigma = 10,95$.

X suit $\mathcal{B}(500; 40\%)$

- $\bar{X} = E(X) = 500 \times 0,4 = 200$ et

$$\sigma(X) = \sqrt{100 \times 0,4 \times 0,6} = \sqrt{120} \approx 10,95$$

Disons alors que Y suit $\mathcal{N}(\mu; \sigma^2)$ avec $\mu = 200$ et $\sigma = 10,95$.

- $p(180 \leq X \leq 190) \approx 0,163$

- $p(180 \leq Y \leq 190) \approx 0,147$

Autre exemple avec la calculette

X suit $\mathcal{B}(500; 40\%)$

- $\bar{X} = E(X) = 500 \times 0,4 = 200$ et

$$\sigma(X) = \sqrt{100 \times 0,4 \times 0,6} = \sqrt{120} \approx 10,95$$

Disons alors que Y suit $\mathcal{N}(\mu; \sigma^2)$ avec $\mu = 200$ et $\sigma = 10,95$.

- $p(180 \leq X \leq 190) \approx 0,163$

- $p(180 \leq Y \leq 190) \approx 0,147$

- On peut mieux faire en calculant

$$p(179,5 \leq Y \leq 190,5) \approx 0,162$$

mais cette correction n'apporte rien dans les circonstances où on utilise la ressemblance entre \mathcal{B} et \mathcal{N} .

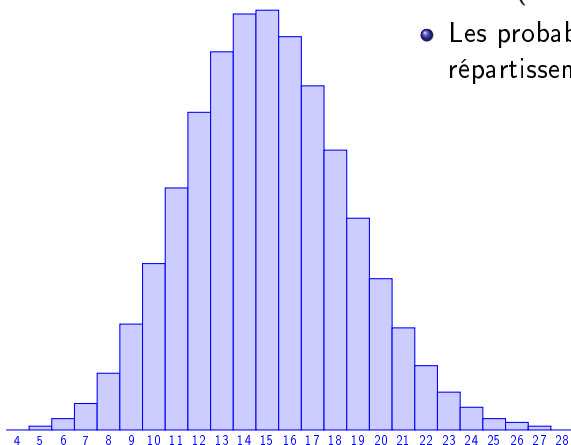
Conclusion : X est très bien approximé par Y .

II. Intervalle de fluctuation asymptotique à 95%

II. 1) Loi binomiale seule

Prenons $\mathcal{B}(100; 0, 15)$

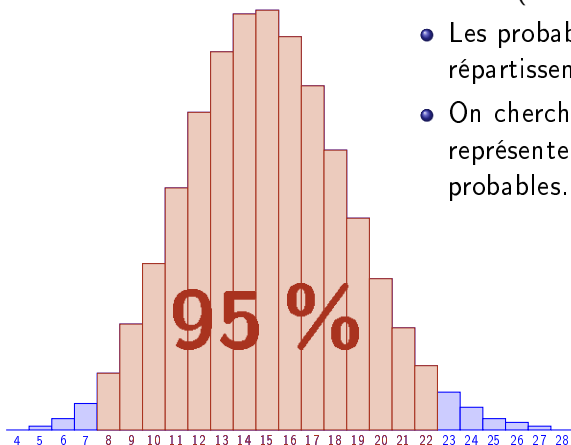
- Les probabilités $P(X = k)$ se répartissent selon la figure.



II. 1) Loi binomiale seule

Prenons $\mathcal{B}(100; 0, 15)$

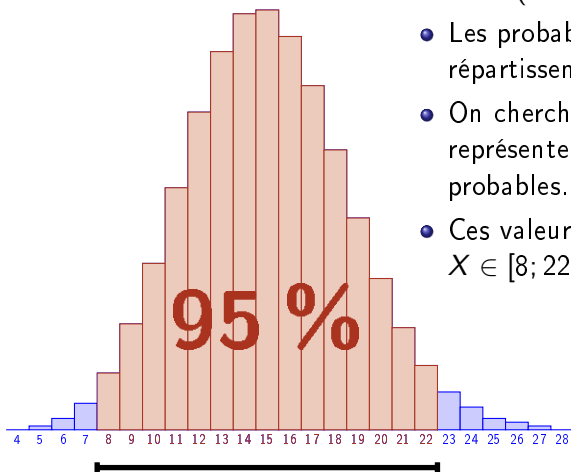
- Les probabilités $P(X = k)$ se répartissent selon la figure.
- On cherche les valeurs de X qui représentent 95% des cas les plus probables.



II. 1) Loi binomiale seule

Prenons $\mathcal{B}(100; 0, 15)$

- Les probabilités $P(X = k)$ se répartissent selon la figure.
- On cherche les valeurs de X qui représentent 95% des cas les plus probables.
- Ces valeurs forment un intervalle : $X \in [8; 22]$.



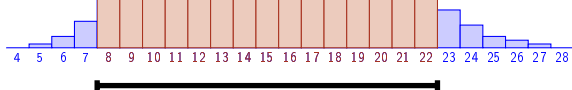
II. 1) Loi binomiale seule

Prenons $\mathcal{B}(100; 0, 15)$

- Les probabilités $P(X = k)$ se répartissent selon la figure.
- On cherche les valeurs de X qui représentent 95% des cas les plus probables.
- Ces valeurs forment un intervalle : $X \in [8; 22]$.

95 %

- En fréquence : $F = \frac{X}{n} \in \left[\frac{8}{100}; \frac{22}{100} \right]$



Tout le problème est de trouver les valeurs 8 et 22...

Normalement : On calcule les valeurs de $p(X \leq k)$ (*tableau*)

k	7	8	...	21	22
$p(X \leq k)$ (%)	1,22	2,75	...	96,07	97,78

Tout le problème est de trouver les valeurs 8 et 22...

Normalement : On calcule les valeurs de $p(X \leq k)$ (*tableau*)

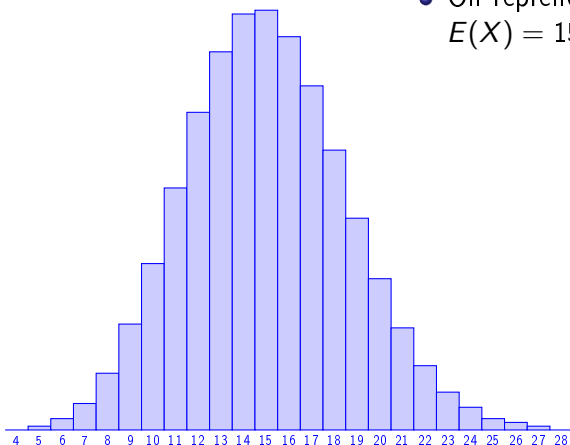
k	7	8	...	21	22
$p(X \leq k)$ (%)	1,22	2,75	...	96,07	97,78

On prend les valeurs pour lesquelles on dépasse **2,5 %** et **97,5 %**.

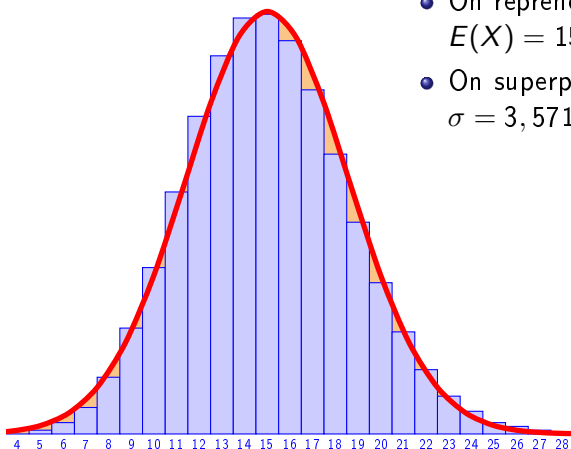
C'est inutilement compliqué. On va préférer obtenir une approximation avec la loi normale, **c'est beaucoup plus simple et rapide.**

II. 2) Intervalle de fluctuation asymptotique

- On reprend $\mathcal{B}(100; 0,15)$
 $E(X) = 15$ et $\sigma(X) \approx 3,571$

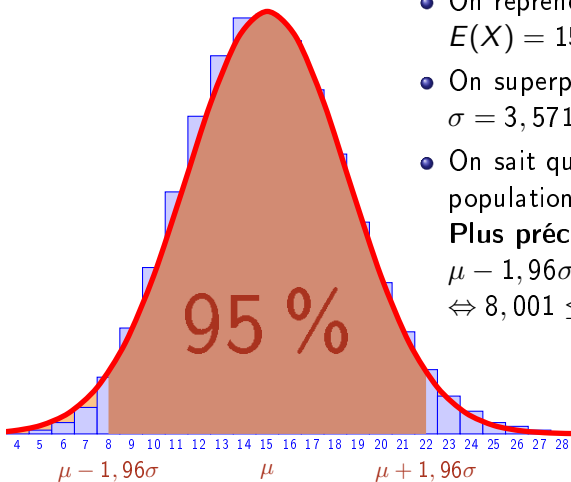


II. 2) Intervalle de fluctuation asymptotique



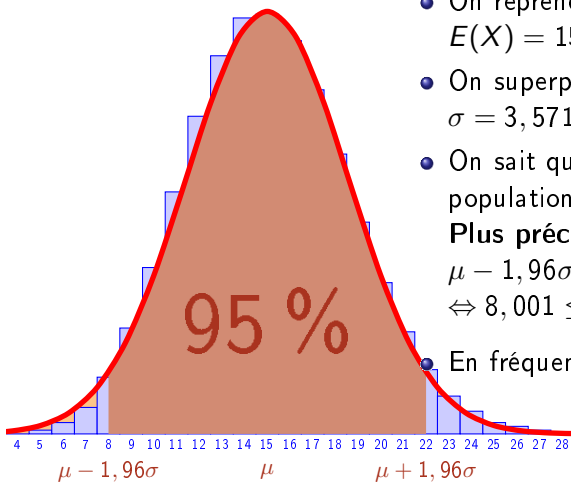
- On reprend $\mathcal{B}(100; 0,15)$
 $E(X) = 15$ et $\sigma(X) \approx 3,571$
- On superpose \mathcal{N} avec $\mu = 15$ et
 $\sigma = 3,571$

II. 2) Intervalle de fluctuation asymptotique



- On reprend $\mathcal{B}(100; 0,15)$
 $E(X) = 15$ et $\sigma(X) \approx 3,571$
- On superpose \mathcal{N} avec $\mu = 15$ et $\sigma = 3,571$
- On sait que sous \mathcal{N} , 95 % de la population est contenue dans $\mu \pm 2\sigma$.
Plus précisément, dans $\mu \pm 1,96\sigma$
 $\mu - 1,96\sigma \leq X \leq \mu + 1,96\sigma$
 $\Leftrightarrow 8,001 \leq X \leq 21,999$

II. 2) Intervalle de fluctuation asymptotique



- On reprend $\mathcal{B}(100; 0,15)$
 $E(X) = 15$ et $\sigma(X) \approx 3,571$
- On superpose \mathcal{N} avec $\mu = 15$ et $\sigma = 3,571$
- On sait que sous \mathcal{N} , 95 % de la population est contenue dans $\mu \pm 2\sigma$.
Plus précisément, dans $\mu \pm 1,96\sigma$
 $\mu - 1,96\sigma \leq X \leq \mu + 1,96\sigma$
 $\Leftrightarrow 8,001 \leq X \leq 21,999$
- En fréquence : $8\% \leq F = \frac{X}{n} \leq 22\%$

Résumons nous :

- On part d'une situation relevant de $\mathcal{B}(n; p)$

Résumons nous :

- On part d'une situation relevant de $\mathcal{B}(n; p)$
- On approxime par $\mathcal{N}(\mu; \sigma^2)$ avec $\mu = n \cdot p$ et $\sigma = \sqrt{n \cdot p \cdot (1 - p)}$

Résumons nous :

- On part d'une situation relevant de $\mathcal{B}(n; p)$
- On approxime par $\mathcal{N}(\mu; \sigma^2)$ avec $\mu = n \cdot p$ et $\sigma = \sqrt{n \cdot p \cdot (1 - p)}$
- On cherche l'intervalle de fluctuation à 95 % pour la **fréquence**

Résumons nous :

- On part d'une situation relevant de $\mathcal{B}(n; p)$
- On approxime par $\mathcal{N}(\mu; \sigma^2)$ avec $\mu = n \cdot p$ et $\sigma = \sqrt{n \cdot p \cdot (1 - p)}$
- On cherche l'intervalle de fluctuation à 95 % pour la **fréquence**
- On l'obtient en calculant $\left[\frac{\mu - 1,96\sigma}{n} ; \frac{\mu + 1,96\sigma}{n} \right]$
C'est l'**intervalle de fluctuation asymptotique**.

Résumons nous :

- On part d'une situation relevant de $\mathcal{B}(n; p)$
- On approxime par $\mathcal{N}(\mu; \sigma^2)$ avec $\mu = n \cdot p$ et $\sigma = \sqrt{n \cdot p \cdot (1 - p)}$
- On cherche l'intervalle de fluctuation à 95 % pour la **fréquence**
- On l'obtient en calculant $\left[\frac{\mu - 1,96\sigma}{n} ; \frac{\mu + 1,96\sigma}{n} \right]$
C'est l'**intervalle de fluctuation asymptotique**.

L'intervalle de fluctuation asymptotique à 95% d'une **fréquence** obtenu sur un échantillon de taille n , dans laquelle on a une proportion p est :

$$\mathcal{I}_F = \left[p - 1,96 \sqrt{\frac{p \cdot (1 - p)}{n}} ; p + 1,96 \sqrt{\frac{p \cdot (1 - p)}{n}} \right]$$

Il faut $n \geq 30$, $n \cdot p \geq 5$ et $n \cdot (1 - p) \geq 5$

Exercice 1

6 % de composants défectueux dans la production globale. L'entreprise vend des boîtes contenant 150 composants.

X est la variable aléatoire qui à chaque boîte associe le nombre de composants défectueux que contient la boîte.

1 Quelle est la loi de probabilité suivie par X ? Donner ses paramètres.

2 On donne ce tableau de probabilités :

c	2	3	4	13	14	15	16
$p(X \leq c)$	0,005	0,019	0,050	0,932	0,963	0,981	0,991

a Donner l'intervalle de fluctuation à 95 % de X .

b Donner cette intervalle en fréquence.

3 Donner l'intervalle de fluctuation asymptotique I_F .

Et si ce n'est pas 95 % ?

1,96 est le coefficient qui fonctionne pour un intervalle de fluctuation à 95 %. Mais si ce n'est pas 95 %, combien faut-il mettre ?

On avait vu la courbe de la loi normale avec le tableau de référence qui permet de calculer le coefficient désiré dans tous les cas. Mais en général ces valeurs devraient suffire :

confiance	90 %	95 %	99 %
risque	10 %	5 %	1 %
coeff	1,645	1,96	2,58

Mais le cas 95 % est de loin le plus important et vous devez vous rappeler de ce cas.

Exemple

Un industriel produit des clous. Il a annoncé : 0,1% des clous produits n'ont pas la solidité demandée.

Pour vérifier cette affirmation, on prélève 10 000 clous dans le stock et on va relever la fréquence f de clous trop fragiles.

Exemple

Un industriel produit des clous. Il a annoncé : 0,1% des clous produits n'ont pas la solidité demandée.

Pour vérifier cette affirmation, on prélève 10 000 clous dans le stock et on va relever la fréquence f de clous trop fragiles.

En fonction de f , on va utiliser la **règle de décision** :

- Si $f \in \mathcal{I}_F$, on accepte l'affirmation.
- Si $f \notin \mathcal{I}_F$ on rejette l'affirmation.

Ces règles sont énoncées avant de faire l'expérience.

On fait l'expérience et on obtient 20 clous trop fragiles dans l'échantillon.
Que conclure ?

On fait l'expérience et on obtient 20 clous trop fragiles dans l'échantillon.

Que conclure ?

D'abord calcul de \mathcal{I}_F :

$$n = 10\,000; p = 0,1\% \Rightarrow \mathcal{I}_F \approx [0,037\%; 0,163\%]$$

Exemple de réalisation

On fait l'expérience et on obtient 20 clous trop fragiles dans l'échantillon.

Que conclure ?

D'abord calcul de \mathcal{I}_F :

$$n = 10\,000; p = 0,1\% \Rightarrow \mathcal{I}_F \approx [0,037\%; 0,163\%]$$

On calcule f et on conclut :

$$f = \frac{20}{10\,000} = 0,2\% \notin \mathcal{I}_F \Rightarrow \textbf{Affirmation rejetée}$$

Remarque : Dans le cas précédent,

- l'**hypothèse** est « La proportion de clous trop fragiles dans la population est $p = 0,1\%$ »
- On dit « au risque de 5%, si $f \in \mathcal{I}_F$ on accepte l'hypothèse... ». Le risque est la **probabilité d'erreur** qu'on s'autorise : Il est possible que $p = 0,1\%$ mais que $f \notin \mathcal{I}_F$.

Exercice 2

Un signal est utilisé pour transmettre des signaux numériques (bits). Dans certaines conditions, des bits peuvent être mal transmis.

On transmet 800 bits durant chaque période. Chaque bit a une probabilité égale à p d'être mal transmis. On note X la variable aléatoire qui associe à chaque période le nombre de bits mal transmis durant cette période.

Le technicien affirme que $p = 1,5\%$.

- 1 En supposant l'affirmation vraie :
 - a quelle est la loi suivie par X ?
 - b par quelle loi peut-on l'approximer ?
 - c calculez \mathcal{I}_F , l'intervalle de fluctuation asymptotique à 95 %.
- 2 On souhaite tester l'affirmation. Énoncez une règle de décision.
- 3 On a réalisé une expérience et on a obtenu 18 erreurs.
 - a Calculez la fréquence f d'erreurs dans l'expérience.
 - b Quelle décision entraîne le résultat de l'expérience ?

II. 4) Intervalle de confiance et estimation

- Avec \mathcal{I}_F on connaît p (ou on pense le connaître). On mesure f pour vérifier.
- Avec \mathcal{I}_C , on ne connaît pas p . On mesure f pour trouver p .

II. 4) Intervalle de confiance et estimation

- Avec \mathcal{I}_F on connaît p (ou on pense le connaître). On mesure f pour vérifier.
- Avec \mathcal{I}_C , on ne connaît pas p . On mesure f pour trouver p .

Si f est la fréquence obtenue sur un échantillon de taille n , l'intervalle de confiance, au **niveau de confiance de 95%** pour la proportion p dans la population est :

$$\mathcal{I}_C = \left[f - 1,96 \sqrt{\frac{f \cdot (1 - f)}{n}}; f + 1,96 \sqrt{\frac{f \cdot (1 - f)}{n}} \right]$$

Faux positif, faux négatif

- On considère une propriété A . Ex : *individu malade*

Faux positif, faux négatif

- On considère une propriété A . Ex : *individu malade*
- Pour savoir si A vrai on fait un test consistant à prélever un échantillon.

Faux positif, faux négatif

- On considère une propriété A . Ex : *individu malade*
- Pour savoir si A vrai on fait un test consistant à prélever un échantillon.
- On calcule une fréquence f pour cet échantillon.

Faux positif, faux négatif

- On considère une propriété A . Ex : *individu malade*
- Pour savoir si A vrai on fait un test consistant à prélever un échantillon.
- On calcule une fréquence f pour cet échantillon.
- On a déterminé un intervalle de fluctuation \mathcal{I}_F . Si $f \in \mathcal{I}_F$, alors on dit que A est vrai. A faux sinon.

Faux positif, faux négatif

- On considère une propriété A . Ex : *individu malade*
- Pour savoir si A vrai on fait un test consistant à prélever un échantillon.
- On calcule une fréquence f pour cet échantillon.
- On a déterminé un intervalle de fluctuation \mathcal{I}_F . Si $f \in \mathcal{I}_F$, alors on dit que A est vrai. A faux sinon.

	$f \in \mathcal{I}_F$	$f \notin \mathcal{I}_F$
A vraie	Ok	Faux négatif
A fausse	Faux positif	Ok

Faux positif, faux négatif

- On considère une propriété A . Ex : *individu malade*
- Pour savoir si A vrai on fait un test consistant à prélever un échantillon.
- On calcule une fréquence f pour cet échantillon.
- On a déterminé un intervalle de fluctuation \mathcal{I}_F . Si $f \in \mathcal{I}_F$, alors on dit que A est vrai. A faux sinon.

	$f \in \mathcal{I}_F$	$f \notin \mathcal{I}_F$
A vraie	Ok	Faux négatif
A fausse	Faux positif	Ok

Avec un \mathcal{I}_F au risque de 5 %, c'est le risque de faux négatif qui est de 5 %.

Le risque de faux positif est difficile à déterminer en général. Mais on peut dire qu'on ne peut pas minimiser simultanément les deux risques, si bien qu'il faut trouver un compromis.

Le risque à 5 % est un bon compromis.

Exercice 3

Lors d'un sondage qualité, on a interrogé 500 personnes.

La question était « Recommanderiez-vous ce produit ? »

150 personnes ont coché la case « Oui ».

Donnez un intervalle de confiance, au risque de 5 %, de la proportion, dans la population totale, des personnes qui recommanderaient le produit.

III. Approximation $\mathcal{B} \simeq \mathcal{P}$

On a dit que $\mathcal{B} \simeq \mathcal{N}$ quand $n \nearrow$. Mais cette approximation ne fonctionne bien que si $n \cdot p \geq 5$. Si p est trop faible, il faut faire une autre approximation. On devra utiliser la loi de **Poisson**, appelée aussi loi des événements rares (p faible).

Cette loi **n'est plus au programme**.