Haute école de gestion de Genève
Geneva School of Business Administration

Distribution de la moyenne
Distribution d'une proportion

Dr Sacha Varone

Objectif

- Connaître la distribution d'une moyenne
- Connaitre la distribution d'une proportion

Rappels

Rappel

Une *estimation ponctuelle*, ou point d'estimation, est une valeur calculée à partir d'un échantillon pour estimer un paramètre d'une population.

Une *distribution d'échantillonnage* d'un estimateur $\hat{\Theta}$ est la distribution des valeurs possible d'une statistique pour un échantillon de taille fixée, sélectionné à partir d'une population.

- comportement moyen :
$$\mathrm{E}(\hat{\Theta})$$
- dispersion :
$$\mathrm{Var}(\hat{\Theta})$$

Rappel

biais : $\mathrm{E}(\hat{\Theta} - \theta) = \mathrm{E}(\hat{\Theta}) - \theta$
Estimateur *non-biaisé* si $\mathrm{E}(\hat{\Theta} - \theta) = 0$

Estimateur *convergent* si, lorsque la taille n de l'échantillon devient grande

1. le biais disparaît : $\lim_{n \to \infty} \mathrm{Biais}(\hat{\Theta}) = 0$
2. la variance devient nulle : $\lim_{n \to \infty} \mathrm{Var}(\hat{\Theta}) = 0$

Un estimateur sans biais et convergent est dit *absolument correct*.

h e g

Distribution d'une moyenne

Moyenne dans une population normale

Si une population est normalement distribuée, de moyenne μ et d'écart type σ, alors la distribution d'échantillonnage de la moyenne $\bar{x}$ est

$$\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$$

Remarque
La distribution d'échantillonnage de la moyenne est composée de toutes les moyennes possibles sur tous les échantillons de même taille.

Erreur standard de la moyenne

L'écart type de la distribution d'échantillonnage de la moyenne, aussi appelée *erreur standard de la moyenne*, est le terme

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Remarque :
$\sigma_{\bar{x}} \leq \sigma$

Variable centrée réduite

La variable centrée réduite associée à la moyenne d'échantillonnage est la variable

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$\bar{x}$ = moyenne de l'échantillon
μ = moyenne de la population
σ = écart type de la population
n = taille de l'échantillon

Exemple

Étiquette : poids en poisson d'un plat surgelé de 1 kg :
$\mathcal{N}(500, 100^2)$ en grammes.
Échantillon : 25 plats achetés pour vérifier la quantité de poisson.
Résultats : moyenne de 490 g par plat.
Question : Y a-t-il tromperie du consommateur ?

1. La moyenne pour cet échantillon :

Exemple

Étiquette : poids en poisson d'un plat surgelé de 1 kg : $\mathcal{N}(500, 100^2)$ en grammes.
Échantillon : 25 plats achetés pour vérifier la quantité de poisson.
Résultats : moyenne de 490 g par plat.
Question : Y a-t-il tromperie du consommateur ?

1. La moyenne pour cet échantillon : $\bar{x} = 490$
2. La distribution d'échantillonnage de la moyenne :

Exemple

Étiquette : poids en poisson d'un plat surgelé de 1 kg : $\mathcal{N}(500, 100^2)$ en grammes.
Échantillon : 25 plats achetés pour vérifier la quantité de poisson.
Résultats : moyenne de 490 g par plat.
Question : Y a-t-il tromperie du consommateur ?

1. La moyenne pour cet échantillon : $\bar{x} = 490$
2. La distribution d'échantillonnage de la moyenne : $\mathcal{N}(500, \frac{100^2}{25} = 400)$
3. L'événement d'intérêt :

Exemple

Étiquette : poids en poisson d'un plat surgelé de 1 kg : $\mathcal{N}(500, 100^2)$ en grammes.
Échantillon : 25 plats achetés pour vérifier la quantité de poisson.
Résultats : moyenne de 490 g par plat.
Question : Y a-t-il tromperie du consommateur ?

1. La moyenne pour cet échantillon : $\bar{x} = 490$
2. La distribution d'échantillonnage de la moyenne : $\mathcal{N}(500, \frac{100^2}{25} = 400)$
3. L'événement d'intérêt : $P(\bar{x} \leq 490) =?$
4. Conversion en une valeur centrée réduite z

h e g

Exemple

Étiquette : poids en poisson d'un plat surgelé de 1 kg : $\mathcal{N}(500, 100^2)$ en grammes.
Échantillon : 25 plats achetés pour vérifier la quantité de poisson.
Résultats : moyenne de 490 g par plat.
Question : Y a-t-il tromperie du consommateur ?

1. La moyenne pour cet échantillon : $\bar{x} = 490$
2. La distribution d'échantillonnage de la moyenne : $\mathcal{N}(500, \frac{100^2}{25} = 400)$
3. L'événement d'intérêt : $P(\bar{x} \leq 490) =?$
4. Conversion en une valeur centrée réduite z
   $z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{490-500}{\frac{100}{\sqrt{25}}} = -0.5$
5. Probabilité désirée

Exemple

Étiquette : poids en poisson d'un plat surgelé de 1 kg : $\mathcal{N}(500, 100^2)$ en grammes.
Échantillon : 25 plats achetés pour vérifier la quantité de poisson.
Résultats : moyenne de 490 g par plat.
Question : Y a-t-il tromperie du consommateur ?

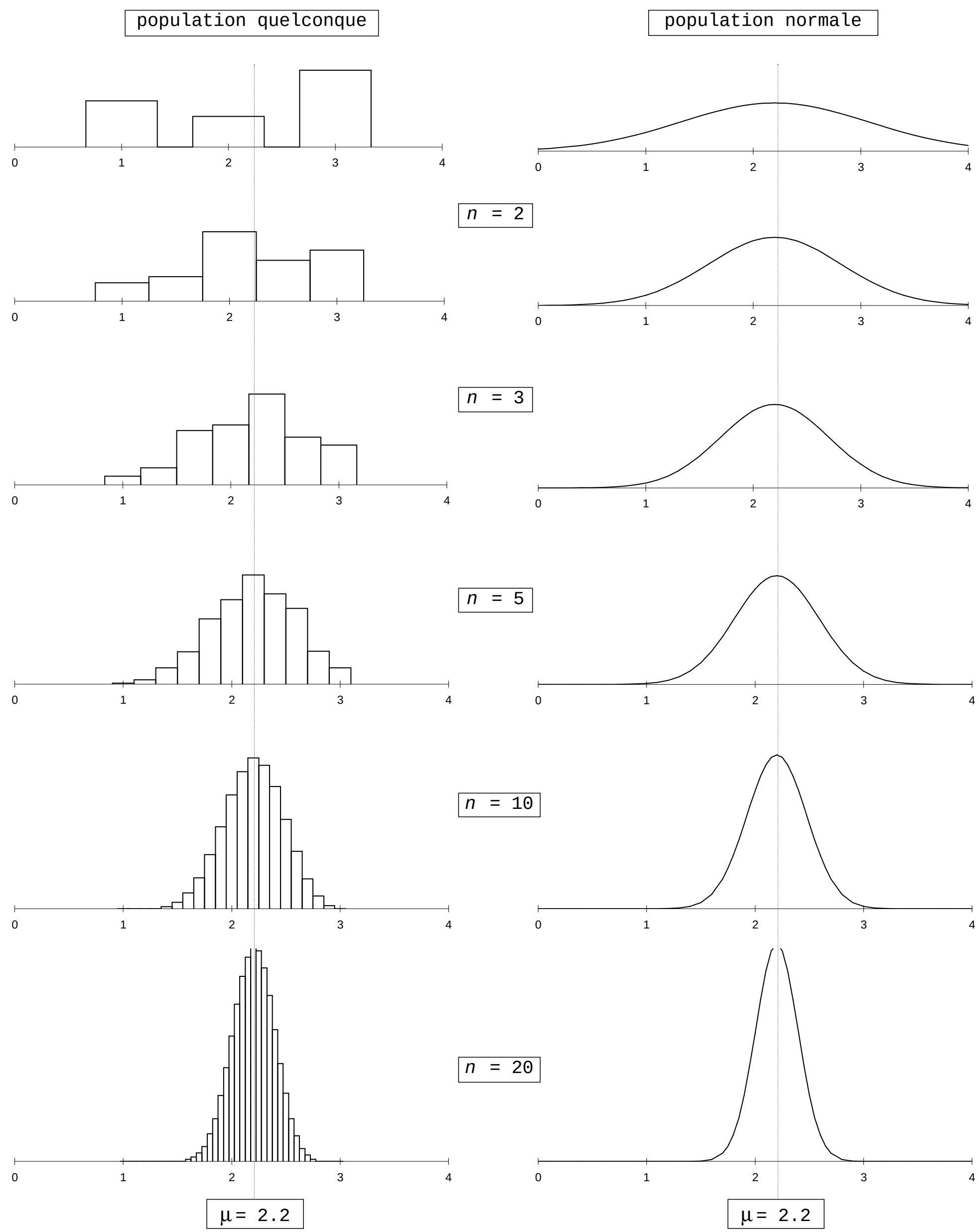
1. La moyenne pour cet échantillon : $\bar{x} = 490$
2. La distribution d'échantillonnage de la moyenne : $\mathcal{N}(500, \frac{100^2}{25} = 400)$
3. L'événement d'intérêt : $P(\bar{x} \leq 490) =?$
4. Conversion en une valeur centrée réduite z
   $z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{490-500}{\frac{100}{\sqrt{25}}} = -0.5$
5. Probabilité désirée
   $P(z \leq -0.5) = 0.3085$

Facteur de correction

Si la taille de l'échantillon est *plus du 5%* de la taille de la population, et que l'échantillon tiré est fait *sans remise*
$\Rightarrow$ facteur de correction sur l'écart type

$$\sqrt{\frac{N-n}{N-1}}$$

où
N = taille de la population
n = taille de l'échantillon


population quelconque
population normale
n = 2
n = 3
n = 5
n = 10
n = 20
μ = 2.2
μ = 2.2

h e g

Rappels
Distribution d'une moyenne
Pop. normale
Erreur standard
z
Correction
TCL
Distribution d'une proportion

Théorème central limite

Soit une suite $(X_1, X_2, \ldots, X_n)$ de n variables aléatoires identiquement et indépendamment distribuées (μ, σ^2). Lorsque $n \to \infty$, la distribution de

$$\overline{X} = \frac{1}{n} \sum_i X_i$$

tend vers la loi $N\left(\mu, \frac{\sigma^2}{n}\right)$

Remarque
Plus la taille de l'échantillon augmente, meilleure est l'approximation par la loi normale.

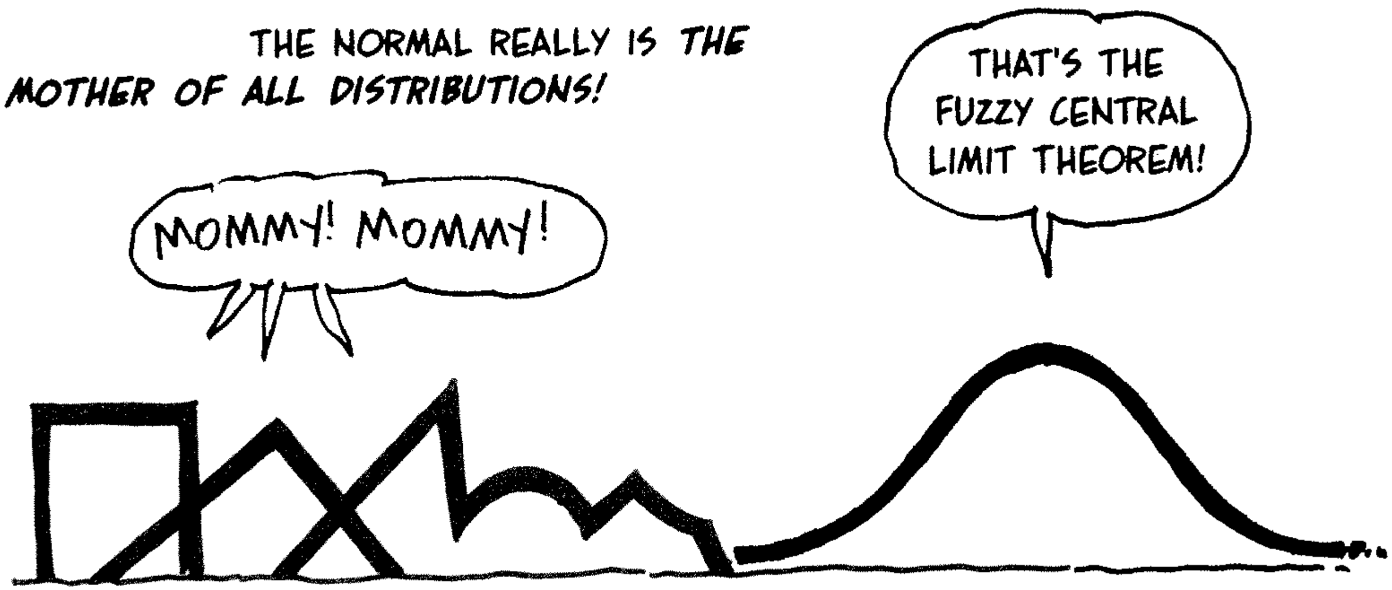
Importance du TCL

Hypothèse de nombreuses méthodes statistiques : loi normale.
En pratique : pas toujours vérifié.

Mais, grâce au **théorème central limite**, même si la population ne satisfait pas à la normalité, la moyenne d'un échantillon de grande taille issu de celle-ci est distribuée de façon normale

Donc : ok pour employer la plupart des outils statistiques.

Illustration



source : "The Cartoon Guide to Statistics", L. Gonick & W. Smith

Jeu 1

Chacun indique un budget pour ses soirées.
Budget moyen ?
Les budgets sont triés par ordre croissant. Quel échantillon semble le plus représentatif ?

1. Les 6 premiers budgets
2. Les 6 derniers budgets
3. les 6 valeurs centrales
4. 6 nombres pris au hasard

Distribution d'une proportion

Qu'est ce qu'une proportion ?

Une proportion est une moyenne particulière

$$Y_i = \begin{cases} 1 & \text{la personne a l'intention d'acheter le produit} \\ 0 & \text{sinon} \end{cases}$$

$$Y_i = \begin{cases} 1 & \text{la personne a l'intention de voter pour Mme L.U.} \\ 0 & \text{sinon} \end{cases}$$

$$Y_i = \begin{cases} 1 & \text{le client est satisfait du service} \\ 0 & \text{sinon} \end{cases}$$

Jeu 2

Chacun joue 10 fois à pile ou face et inscrit la proportion de piles. Graphique de la distribution ?

Propriété

Dans un sondage aléatoire simple, la proportion dans l'échantillon $\hat{p}$ est un estimateur sans biais de la proportion p dans la population.

Lorsque la taille n de l'échantillon est suffisamment grande, i.e.

$$n\pi \geq 5 \qquad \text{et } n(1-\pi) \geq 5$$

L'espérance de l'estimateur $\hat{p}$ est $\mathrm{E}(\hat{p}) = \pi$
L'écart type de la distribution d'échantillonnage est

$$\sqrt{\frac{\pi(1-\pi)}{n}}$$

Théorème

Lorsque la taille n de l'échantillon est suffisamment grande, i.e. $n\pi \geq 5$ et $n(1-\pi) \geq 5$, la distribution d'échantillonnage peut être approchée par une distribution normale centrée en π, avec comme écart type

$$\sigma_{\bar{p}} = \sqrt{\frac{\pi(1-\pi)}{n}}$$

où

π	$=$	proportion dans la population
n	$=$	taille de l'échantillon

h e g

Remarque

π est une paramètre inconnu.
$\hat{p} = \bar{p}$ est l'estimateur de π

Exemple

Le responsable d'une agence immobilière souhaite passer une annonce vantant la rapidité de traitement des affaires. Il pense que le 80% des propriétés à vendre trouvent preneur en au plus 4 mois. Il a sélectionné aléatoirement 100 affaires, et parmi celles-là, 73 se sont terminées en au plus 4 mois. Les étapes suivantes déterminent la probabilité de ce résultat :

1. Déterminer la proportion de la population.

Exemple

Le responsable d'une agence immobilière souhaite passer une annonce vantant la rapidité de traitement des affaires. Il pense que le 80% des propriétés à vendre trouvent preneur en au plus 4 mois. Il a sélectionné aléatoirement 100 affaires, et parmi celles-là, 73 se sont terminées en au plus 4 mois. Les étapes suivantes déterminent la probabilité de ce résultat :

1. Déterminer la proportion de la population. $\pi = 0.8$
2. Calculer la proportion de l'échantillon.

Exemple

Le responsable d'une agence immobilière souhaite passer une annonce vantant la rapidité de traitement des affaires. Il pense que le 80% des propriétés à vendre trouvent preneur en au plus 4 mois. Il a sélectionné aléatoirement 100 affaires, et parmi celles-là, 73 se sont terminées en au plus 4 mois. Les étapes suivantes déterminent la probabilité de ce résultat :

1. Déterminer la proportion de la population. $\pi = 0.8$
2. Calculer la proportion de l'échantillon. $\hat{p} = \frac{73}{100}$
3. Déterminer la moyenne et l'écart type de la distribution d'échantillonnage.

Exemple

Le responsable d'une agence immobilière souhaite passer une annonce vantant la rapidité de traitement des affaires. Il pense que le 80% des propriétés à vendre trouvent preneur en au plus 4 mois. Il a sélectionné aléatoirement 100 affaires, et parmi celles-là, 73 se sont terminées en au plus 4 mois. Les étapes suivantes déterminent la probabilité de ce résultat :

1. Déterminer la proportion de la population. $\pi = 0.8$
2. Calculer la proportion de l'échantillon. $\hat{p} = \frac{73}{100}$
3. Déterminer la moyenne et l'écart type de la distribution d'échantillonnage.

$$\mu_{\hat{p}} = 0.8 \quad \sigma_{\hat{p}} = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.8(1-0.8)}{100}} = 0.04$$

Exemple

4. Définir l'événement d'intérêt.

Exemple

4. Définir l'événement d'intérêt. $P(\hat{p} \leq 0.73) = ?$
5. Vérifier les hypothèses.

Exemple

4. Définir l'événement d'intérêt. $P(\hat{p} \leq 0.73) = ?$
5. Vérifier les hypothèses. $n\pi = 80 > 5$ et $n(1 - \pi) = 20 > 5$
6. Déterminer la probabilité.

h e g

Exemple

4. Définir l'événement d'intérêt. $P(\hat{p} \leq 0.73) =?$
5. Vérifier les hypothèses. $n\pi = 80 > 5$ et $n(1-\pi) = 20 > 5$
6. Déterminer la probabilité.

$$z = \frac{\hat{p} - \pi}{\sigma_{\hat{p}}} = \frac{0.73 - 0.8}{\sqrt{\frac{0.8(1-0.8)}{100}}} = -1.75$$

$$P(\hat{p} \leq 0.73) = P(z \leq -1.75) = 0.0401$$

Exemple

Lors d'une élection, $\theta = \pi = 20\%$ des électeurs choisissent le candidat Schroumpf. Un institut de sondage interroge au préalable certains électeurs choisis au hasard. Soit X la variable aléatoire valant 1 si l'électeur vote pour Schroumpf, et 0 sinon.
Quelle loi suit la variable aléatoire X ?

Exemple

Lors d'une élection, $\theta = \pi = 20\%$ des électeurs choisissent le candidat Schroumpf. Un institut de sondage interroge au préalable certains électeurs choisis au hasard. Soit X la variable aléatoire valant 1 si l'électeur vote pour Schroumpf, et 0 sinon.
Quelle loi suit la variable aléatoire X ? $\mathcal{B}(1, 0.2)$

On interroge $n = 100$ électeurs par un tirage avec remise (ou sans remise si le nombre total d'électeurs est suffisamment grand, car la loi hypergéométrique converge vers la loi binomiale dans ce cas). Le nombre de votes favorables à Schroumpf est donc
$Y = X_1 + \ldots + X_{100}$.
Quelle loi suit la variable aléatoire Y ?

Exemple

Lors d'une élection, $\theta = \pi = 20\%$ des électeurs choisissent le candidat Schroumpf. Un institut de sondage interroge au préalable certains électeurs choisis au hasard. Soit X la variable aléatoire valant 1 si l'électeur vote pour Schroumpf, et 0 sinon.
Quelle loi suit la variable aléatoire X ? $\mathcal{B}(1, 0.2)$

On interroge $n = 100$ électeurs par un tirage avec remise (ou sans remise si le nombre total d'électeurs est suffisamment grand, car la loi hypergéométrique converge vers la loi binomiale dans ce cas). Le nombre de votes favorables à Schroumpf est donc
$Y = X_1 + \ldots + X_{100}$.
Quelle loi suit la variable aléatoire Y ? $\mathcal{B}(100, 0.2)$
Espérance $\mathrm{E}(Y) = 100 * 0.2 = 20$ et variance
$\mathrm{Var}(Y) = 100 * 0.2 * 0.8 = 16$

La proportion de votants pour Schroumpf est heg

Exemple

Lors d'une élection, $\theta = \pi = 20\%$ des électeurs choisissent le candidat Schroumpf. Un institut de sondage interroge au préalable certains électeurs choisis au hasard. Soit X la variable aléatoire valant 1 si l'électeur vote pour Schroumpf, et 0 sinon.
Quelle loi suit la variable aléatoire X ? $\mathcal{B}(1, 0.2)$

On interroge $n = 100$ électeurs par un tirage avec remise (ou sans remise si le nombre total d'électeurs est suffisamment grand, car la loi hypergéométrique converge vers la loi binomiale dans ce cas). Le nombre de votes favorables à Schroumpf est donc
$Y = X_1 + \ldots + X_{100}$.
Quelle loi suit la variable aléatoire Y ? $\mathcal{B}(100, 0.2)$
Espérance $\mathrm{E}(Y) = 100 * 0.2 = 20$ et variance
$\mathrm{Var}(Y) = 100 * 0.2 * 0.8 = 16$

La proportion de votants pour Schroumpf est $\hat{p} = \frac{X_1 + \ldots + X_{100}}{100}$ dont l'espérance est

Exemple

Lors d'une élection, $\theta = \pi = 20\%$ des électeurs choisissent le candidat Schroumpf. Un institut de sondage interroge au préalable certains électeurs choisis au hasard. Soit X la variable aléatoire valant 1 si l'électeur vote pour Schroumpf, et 0 sinon.
Quelle loi suit la variable aléatoire X ? $\mathcal{B}(1, 0.2)$

On interroge $n = 100$ électeurs par un tirage avec remise (ou sans remise si le nombre total d'électeurs est suffisamment grand, car la loi hypergéométrique converge vers la loi binomiale dans ce cas). Le nombre de votes favorables à Schroumpf est donc
$Y = X_1 + \ldots + X_{100}$.
Quelle loi suit la variable aléatoire Y ? $\mathcal{B}(100, 0.2)$
Espérance $\mathrm{E}(Y) = 100 * 0.2 = 20$ et variance
$\mathrm{Var}(Y) = 100 * 0.2 * 0.8 = 16$

La proportion de votants pour Schroumpf est $\hat{p} = \frac{X_1 + \ldots + X_{100}}{100}$ dont l'espérance est $\mathrm{E}(\hat{p}) = 0.2$ et la variance

Exemple

Lors d'une élection, $\theta = \pi = 20\%$ des électeurs choisissent le candidat Schroumpf. Un institut de sondage interroge au préalable certains électeurs choisis au hasard. Soit X la variable aléatoire valant 1 si l'électeur vote pour Schroumpf, et 0 sinon.
Quelle loi suit la variable aléatoire X ? $\mathcal{B}(1, 0.2)$

On interroge $n = 100$ électeurs par un tirage avec remise (ou sans remise si le nombre total d'électeurs est suffisamment grand, car la loi hypergéométrique converge vers la loi binomiale dans ce cas). Le nombre de votes favorables à Schroumpf est donc
$Y = X_1 + \ldots + X_{100}$.
Quelle loi suit la variable aléatoire Y ? $\mathcal{B}(100, 0.2)$
Espérance $\mathrm{E}(Y) = 100 * 0.2 = 20$ et variance
$\mathrm{Var}(Y) = 100 * 0.2 * 0.8 = 16$

La proportion de votants pour Schroumpf est $\hat{p} = \frac{X_1 + \ldots + X_{100}}{100}$ dont l'espérance est $\mathrm{E}(\hat{p}) = 0.2$ et la variance $\mathrm{Var}(\hat{p}) = 0.0016$

Exemple

Lors d'une élection, $\theta = \pi = 20\%$ des électeurs choisissent le candidat Schroumpf. Un institut de sondage interroge au préalable certains électeurs choisis au hasard. Soit X la variable aléatoire valant 1 si l'électeur vote pour Schroumpf, et 0 sinon.
Quelle loi suit la variable aléatoire X ? $\mathcal{B}(1, 0.2)$

On interroge $n = 100$ électeurs par un tirage avec remise (ou sans remise si le nombre total d'électeurs est suffisamment grand, car la loi hypergéométrique converge vers la loi binomiale dans ce cas). Le nombre de votes favorables à Schroumpf est donc
$Y = X_1 + \ldots + X_{100}$.
Quelle loi suit la variable aléatoire Y ? $\mathcal{B}(100, 0.2)$
Espérance $\mathrm{E}(Y) = 100 * 0.2 = 20$ et variance
$\mathrm{Var}(Y) = 100 * 0.2 * 0.8 = 16$

La proportion de votants pour Schroumpf est $\hat{p} = \frac{X_1+\ldots+X_{100}}{100}$ dont l'espérance est $\mathrm{E}(\hat{p}) = 0.2$ et la variance $\mathrm{Var}(\hat{p}) = 0.0016$

Par le théorème central limite, la loi de $\hat{p}$ est donc proche d'une loi normale $\mathcal{N}(0.2, 0.0016)$