

■ Exemple 1.

Pour $N, p = 900, 1/365$ (On a bien N grand et Np de l'ordre de l'unité).

k	B(N, p)	P(Np)
0	0.085	0.085
1	0.209	0.209
2	0.258	0.258
3	0.213	0.212
4	0.131	0.131
5	0.064	0.065
6	0.026	0.027
7	0.009	0.01
8	0.003	0.003
9	0.001	0.001

B) Théorème central-limite

■ Théorème 6 [Central-limite forme 1 : variance connue]

1. Si (X_n) un n -échantillon de variables aléatoires d'espérance μ et de variance $\sigma^2 > 0$, alors la suite VAR de terme général M_n^* , appelée erreur normalisée, où

$$M_n^* = \frac{M_n - \mu}{\frac{\sigma}{\sqrt{n}}}$$

converge en loi vers une variable $\mathcal{N}(0, 1)$. Précisément :

$$\forall (a, b) \in \mathbb{R}^2 \text{ t.q. } a < b \quad \lim_{n \rightarrow +\infty} P(M_n^* \in [a, b]) = \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-\frac{t^2}{2}\right) dt.$$

Même conclusion pour la suite (S_n^*) .

2. Seuil d'approximation (Thm4. 2)) : si $n \geq 30$, il est communément admis qu'une valeur approchée acceptable de $P(a \leq M_n^* \leq b)$ est la limite ci-dessus, c-à-d. $\Phi(b) - \Phi(a)$.

■ Théorème 7 [Central-limite forme 2 : variance inconnue]

Si (X_n) un n -échantillon de variables aléatoires d'espérance μ et de variance inconnue, alors la loi de l'erreur normalisée empirique $M_n^* = \frac{M_n - \mu}{\sqrt{\frac{S_n^2}{n}}}$ converge vers une loi $\mathcal{N}(0, 1)$. En particulier, si $n \geq 30$, il

est communément admis que la limite est une valeur approchée de $P(a \leq M_n^* \leq b)$.

■ Exemple 2.

(illustre l'intérêt de de genre de théorèmes). On lance 50 dés identiques numérotés de 1 à 50. Pour tout entier k , on note X_k la variable aléatoire égale au score du dé k .

- Donner l'expression de la variable aléatoire M donnant la moyenne M des scores observés sur les 50 lancers ?
- Appliquer le théorème central limite pour estimer $P(M \in [3, 4; 3, 6])$.
- La loi de M pourrait-elle se calculer exactement ?
- Estimer la probabilité

Les résultats qui suivent sont des cas particuliers du TCL forme 1

C) Approximation gaussienne de la loi de Poisson

■ Théorème 8 [Approximation gaussienne de la loi de Poisson]

1. Soit $\alpha > 0$ un entier et $(X_n)_n$ une suite de VAR telles que $X_n \rightsquigarrow \mathcal{P}(n\alpha)$. Alors la suite de VAR de t.g. $M_n^* = \frac{X_n - \overbrace{n\alpha}^{E(X_n)}}{\underbrace{\sqrt{n\alpha}}_{\sigma_{X_n}}}$ converge en loi vers une variable de loi $\mathcal{N}(0, 1)$.
↑ le param. augmente avec n (il tend vers +∞)
 Même conclusion pour S_n^* : *la somme centrée + réduite converge vers la N(0,1) en loi*
2. (Approximation Thm4. 2)) : il est communément admis que si $\lambda \geq 18$, la loi de Poisson $\mathcal{P}(\lambda)$ est approximativement une loi $\mathcal{N}(\lambda, \lambda)$.

■ Remarque 4.

1. On a donc une loi normale de paramètres ceux de la loi de Poisson qu'elle approche.
2. La somme de n variables indépendantes de loi $\mathcal{P}(\alpha)$ suit une loi de Poisson de paramètre $n\alpha$, la recette découle du théorème avec $\lambda = n\alpha$ dès lors que $n\alpha \geq 18$ et la proposition 1.

D) Approximation gaussienne de la loi binomiale

■ Théorème 9 [de Moivre-Laplace]

Soit $p \in]0, 1[, q = 1 - p,$

1. Soit (X_n) une suite de VAR de loi $\mathcal{B}(n, p)$. Alors la suite de VAR M_n^* converge en loi vers une VAR de loi $\mathcal{N}(0, 1)$.
2. (Approximation Thm4. 2)) : Il est communément admis que si $n \geq 30$ et $np, nq \geq 5$,

- a) loi de $M_n^* = \frac{X_n - \overbrace{np}^{E(X_n)}}{\underbrace{\sqrt{npq}}_{\sigma_{X_n}}}$ est approximativement une loi $\mathcal{N}(0, 1)$. De même pour S_n^* .
- b) La loi de X_n est approximativement une loi $\mathcal{N}(np, npq)$.

↑ $E(X_n)$ ↓ $V(X_n)$

T_{17.1}

Approximation d'une loi discrète par une loi continue

Comme les lois de Poisson et binomiale sont discrètes à valeurs entières, pour avoir une approximation de $P(X = k)$ par une loi à densité, on utilise ces théorèmes en effectuant une *correction de continuité*, qui consiste à dire que :

1. si Z suit la loi normale qui approche celle de X_n :

$$P(X = k) = P\left(k - \frac{1}{2} \leq X \leq k + \frac{1}{2}\right) \\ \simeq P\left(Z \in \left[k - \frac{1}{2}, k + \frac{1}{2}\right]\right)$$

2. Ensuite, on calcule cette dernière probabilité par centrage-réduction de la loi de Z , ce qui permet de se reporter aux valeurs des quantiles de la loi de $Z^* \rightsquigarrow \mathcal{N}(0, 1)$.

■ Exemple 3.

Estimer $P(X = 3)$ pour $X \rightsquigarrow \mathcal{P}(20)$. Comme $20 \geq 18$, on peut considérer l'approximation de X par une variable Z de loi $\mathcal{N}(m, m)$ où $m = 20$ valide (mais on est quasiment au seuil de validité).

1. On écrit d'abord que

$$P(X = 3) = P(X \in [2, 5; 3, 5]) \stackrel{\text{Thm. 9}}{\simeq} P(Z \in [2, 5; 3, 5])$$

2. Ensuite on centre-réduit : la variable $Z^* = \frac{Z - m}{\sqrt{m}}$ suit une loi $\mathcal{N}(0, 1)$ et de plus :

$$Z \in [2, 5; 3, 5] = Z^* \in \left[\underbrace{\frac{2,5 - m}{\sqrt{m}}}_a; \underbrace{\frac{3,5 - m}{\sqrt{m}}}_b \right].$$

D'après le TCL : $P(Z^* \in [a, b]) = \Phi(b) - \Phi(a)$, et la calculatrice donne $0,048 \simeq \Phi(b) - \Phi(a) \simeq P(X = 3)$, la valeur réelle de $P(X = 3)$ étant environ 0,052. Approximation acceptable.

■ Exemple 4.

Combien de fois N faut-il lancer une pièce de monnaie pour être sûr à 95% que le nombre de pile observés soit dans l'intervalle $I = \left[\frac{9N}{20}, \frac{11N}{20} \right]$? *Noter que $\left[\frac{9N}{20}, \frac{11N}{20} \right] = \left[\frac{N}{2} - \alpha, \frac{N}{2} + \alpha \right]$ avec $\alpha = \frac{N}{20} > 0$.*

La loi de la VAR X_N qui compte le nombre de piles est une loi $\mathcal{B}\left(N, \frac{1}{2}\right)$. Comme $E(X_N) = N/2$, on cherche N tel que $P\left(X_N \in \left[E(X_N) - \frac{N}{20}, E(X_N) + \frac{N}{20}\right]\right) \geq 95\%$. En supposant $N \geq 30$, et comme $p = 1/2$, les conditions du théorème de Moivre-Laplace sont vérifiées : X_N suit approximativement une loi normale de

centre $E(X_N)$ et de variance $V(X_N) = N/4$. On centre-réduit : $X_N \in I = \frac{X_N - E(X_N)}{\sqrt{N}} \in \left[-\frac{\sqrt{N}}{10}, \frac{\sqrt{N}}{10} \right]$. Il suffit

donc de choisir N de sorte que $\sqrt{N}/10 \geq 2,58$ (quantile d'ordre 0,975 de loi normale centrée réduite), c'est à dire : $N \geq 666$. *385* *1,96*

■ Exercice 1.

(méthode inférentielle). C'est le second tour de l'élection présidentielle. Les candidats A et B s'opposent. On interroge 1000 personnes sur leur intention de vote : 52% des sondés annoncent vouloir voter pour A . On souhaite obtenir une majoration de la probabilité de se tromper en annonçant que A sera élu. On numérote pour cela les sondés de $i = 1$ à $i = 1000$ et on note X_i la variable aléatoire égale 1 si i compte voter pour A et 0 sinon. On suppose que les X_i sont mutuellement indépendantes et suivent toutes une loi de Bernoulli de paramètre p .

1. Interpréter le paramètre p , qui est donc inconnu est qui cherche à être estimé.
2. Que mesure la variable $M = \frac{1}{1000}(X_1 + \dots + X_{1000})$?
3. Appliquer l'inégalité de Tchebychev à la variable M avec un paramètre $\varepsilon > 0$, et interpréter la probabilité de l'évènement : $|M - p| \geq \varepsilon$.
4. Donner une majoration de $p(1 - p)$ indépendante de p .
5. Pourquoi s'intéresser à $\varepsilon = 2\%$?
6. Conclure.

4 Application à la démarche des tests statistiques

La démarche des tests statistiques est à rapprocher des raisonnements par absurde.

Exemple 7 Je réévalue l'événement

$$A = X_N \in \left[E(X_N) - \frac{N}{20}, E(X_N) + \frac{N}{20} \right]$$

en centrant-réduisant X_N pour utiliser la théorie de Moivre-Laplace :

$$A = E(X_N) - \frac{N}{20} \leq X_N \leq E(X_N) + \frac{N}{20}$$

Centrage

$$= -\frac{N}{20} \leq X_N - E(X_N) \leq \frac{N}{20}$$

réduction

$$A = -\frac{\sqrt{N}}{40} \leq \frac{X_N - E(X_N)}{\frac{\sqrt{N}}{2}} \leq \frac{\sqrt{N}}{40}$$

$$A = X_N^* \in \left[-\frac{\sqrt{N}}{10}, \frac{\sqrt{N}}{10} \right]$$

donc $P(A) = P\left(X_N^* \in \left[-\frac{\sqrt{N}}{10}, \frac{\sqrt{N}}{10}\right]\right)$

$$\approx P\left(Z \in \left[-\frac{\sqrt{N}}{10}, \frac{\sqrt{N}}{10}\right]\right)$$

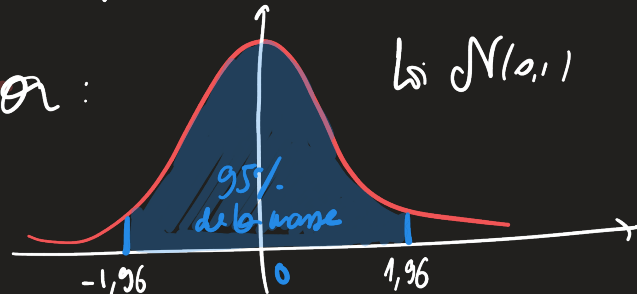
Moivre
Laplace

où $Z \sim \mathcal{N}(0,1)$

$$P(A) \approx \frac{1}{\sqrt{2\pi}} \int_{-\frac{\sqrt{N}}{10}}^{\frac{\sqrt{N}}{10}} e^{-t^2/2} dt$$

On veut $P(A) \geq 95\%$

ou : $\text{soit } \mathcal{N}(0,1)$



donc il suffit de prendre N de sorte

que $\frac{\sqrt{N}}{10} \geq 1.96$

$$\sqrt{N} \geq 19.6$$

$$N \geq 19.6^2$$

$$N \geq 385$$

rem.

$$19.6^2 = (20 - 0.4)^2$$

$$= 400 \left(1 - \frac{0.4}{20}\right)^2 \quad \varepsilon = 0.02$$

$$\approx 400(1 - 2\varepsilon) \quad \text{par équivalent usuel}$$

$$\approx 400 - 800\varepsilon$$

$$(1 - \varepsilon)^2 \approx 1 - 2\varepsilon$$

$$\approx 400 - 16$$

$$\approx 384$$

Test de conformité à la moyenne

Ce test, comme tous les tests, relève d'une application du TCL forme 2

T_{17.2}

Démarche pour tester l'hypothèse (H_0) (dite de conformité)

l'hypothèse que l'on formule est :

$$(H_0) \quad \mu = \mu_0$$

1. On fixe un risque α (souvent $\alpha = 5\%$).
2. Sous (H_0) , la loi théorique est normale d'après le TCL si l'échantillon statistique est assez gros. La moyenne empirique suit alors approximativement une loi normale et de mêmes moments que la loi théorique : donc sous (H_0) une loi normale de moyenne μ_0 et de variance soit connue (TCL forme 1, cas typique : usinage de pièces par machines), ou inconnue (TCL forme 2, cas majoritaire).
 - a) On calcule les quantités empiriques $\bar{x}$ et s_x qui sont des estimations de la moyenne et de l'écart-type théoriques sur l'échantillon de taille n prélevé.
 - b) On calcule $u = \frac{\bar{x} - \mu_0}{\frac{s_x}{\sqrt{n}}}$ qui est la valeur observée du la variable M_n^* sur l'échantillon statistique.
3. Si $u \notin \left[-u_{1-\frac{\alpha}{2}}, u_{1-\frac{\alpha}{2}}\right]$: il y avait une probabilité $\leq \alpha$ (donc trop faible) que cela se produise sous (H_0) : on peut rejeter (H_0) au risque α (c'est-à-dire qu'on a une probabilité α de se tromper en annonçant que (H_0) est fausse).
4. Si $u \in \left[-u_{1-\frac{\alpha}{2}}, u_{1-\frac{\alpha}{2}}\right]$: on n'a pas d'élément permettant de douter de H_0 . Faute d'indices supplémentaires, on conserve (H_0) .

Quantiles de la loi $\mathcal{N}(0,1)$:

- Quantile d'ordre 97.5% : 1.96 (risque à $\alpha = 5\%$)
- Quantile d'ordre 99.5% : 2.58 (risque à $\alpha = 1\%$)
- Quantile d'ordre 99.9% : 3.29 (risque à $\alpha = 0.2\%$)

Test Statistique: conformité à la moyenne.

J'ai lancé une pièce de monnaie 100 fois. J'ai obtenu 41 piles.

Est-ce que ma pièce est truquée?

Raisonnement "par abus" T_{12} .

Je suppose au contraire que ma pièce n'est pas truquée:

La loi du nombre de piles observé est une loi $\mathcal{B}(n, p)$ où $n = 100$
 $p = \frac{1}{2}$.

$$n \geq 30 \checkmark$$

$$np \geq 5 \checkmark$$

$$nq \geq 5 \checkmark$$

donc la loi $\mathcal{B}(n, p)$ est approximativement normale.

Donc si je note X la VAR égale au nombre de piles observés $X^* = \frac{X - np}{\sqrt{npq}}$ $np = 50$
 $npq = 25$

suit approximativement $\mathcal{N}(0, 1)$.

J'effectue une correction de continuité: T_{12}

$$X = 41$$

NECESSAIRE CHAQUE FOIS QU'ON APPROXIME UNE LOI DISCRÈTE par une loi continue.

$$[X = 41] = X \in [40,5 ; 41,5]$$

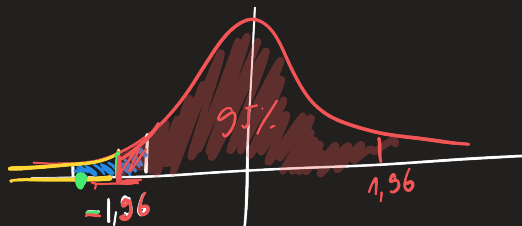
$$= 40,5 \leq X \leq 41,5$$

$$\stackrel{\text{centrage}}{=} -9,5 \leq X - 50 \leq -8,5$$

$$\stackrel{\text{réduction}}{\sqrt{100}} [X = 41] = -\frac{9,5}{5} \leq \frac{X - 50}{5} \leq \frac{-8,5}{5}$$

$$P(X = 41) = P(X^* \in [-1,9; -1,7])$$

$$\approx \frac{1}{\sqrt{2\pi}} \int_{-1,9}^{-1,7} e^{-t^2/2} dt$$



Il faudrait faire un calcul, mais on a, à l'œil, moins de 4% de chances d'obtenir une valeur de X^* dans la zone .

- je prends un risque en disant que la pièce est truquée.

- Si le nombre de piles observé se trouve à gauche de $-1,96$, il y a au moins de 2,5% qu'elle arrive.

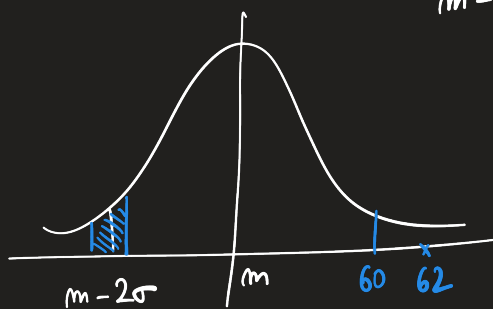
Donc, notre observation aurait très peu de probabilité sous l'hypothèse: "pièce non truquée".

→ Je peux affirmer avec un α d'erreur $\leq 5\%$ que ma pièce est truquée.

$$m = 50$$

$$\sigma = 5$$

41 piles: quasiment $m - 2\sigma$



Principe des tests statistiques

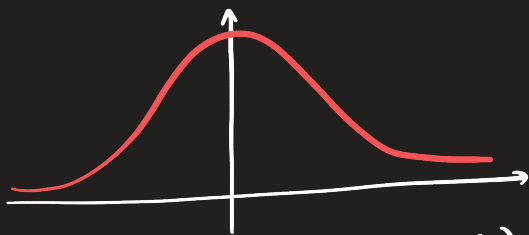
① On suppose le contraire de ce qu'on

pense: la pièce est équilibrée

Cette hypothèse s'appelle *hypothèse nulle*.
Traditionnellement, notée (H_0) .

② Sous cette hypothèse, la loi de la VAR
mesurée est connue.

La loi du nb de pbs est $CL(m, \sigma^2)$
 $m = 50 \quad \sigma^2 = 25 : \sigma = 5$



③ Vous fixez (= votre choix personnel) un
niveau de risque = risque que vous

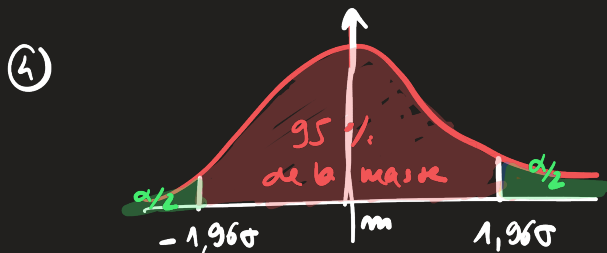
assumez en énonçant votre verdict
sachant qu'il se peut-être faux = α

Traditionnellement $\alpha = 5\%$.

pour des enjeux plus faibles: $\alpha = 1\%$.

voire $\alpha = 1/1000$

J'assume $\alpha = 5\%$ (= niveau de risque)



Si (H_0) est vraie, la mesure faite
de ma VAR se situe dans $[m - 1,96\sigma, m + 1,96\sigma]$
avec 95%.

Avec mon choix de α , je considère que
toute mesure donnant une valeur dans
la zone verte est trop peu probable sous
l'hypothèse (H_0) .

Il est trop peu probable, en supposant
la pièce équilibrée, que le nombre de
pbs observé soit en dehors de

$$\begin{aligned} I &= [50 - 1,96 \times 5, 50 + 1,96 \times 5] \\ &= [50 - 9,8, 50 + 9,8] \\ &= [40,2; 59,8] \end{aligned}$$

⑤ Règle de décision ← standard de la méthode.

• si la valeur observée est
hors de I : il y avait moins de α
chance que cela se produise sous (H_0)

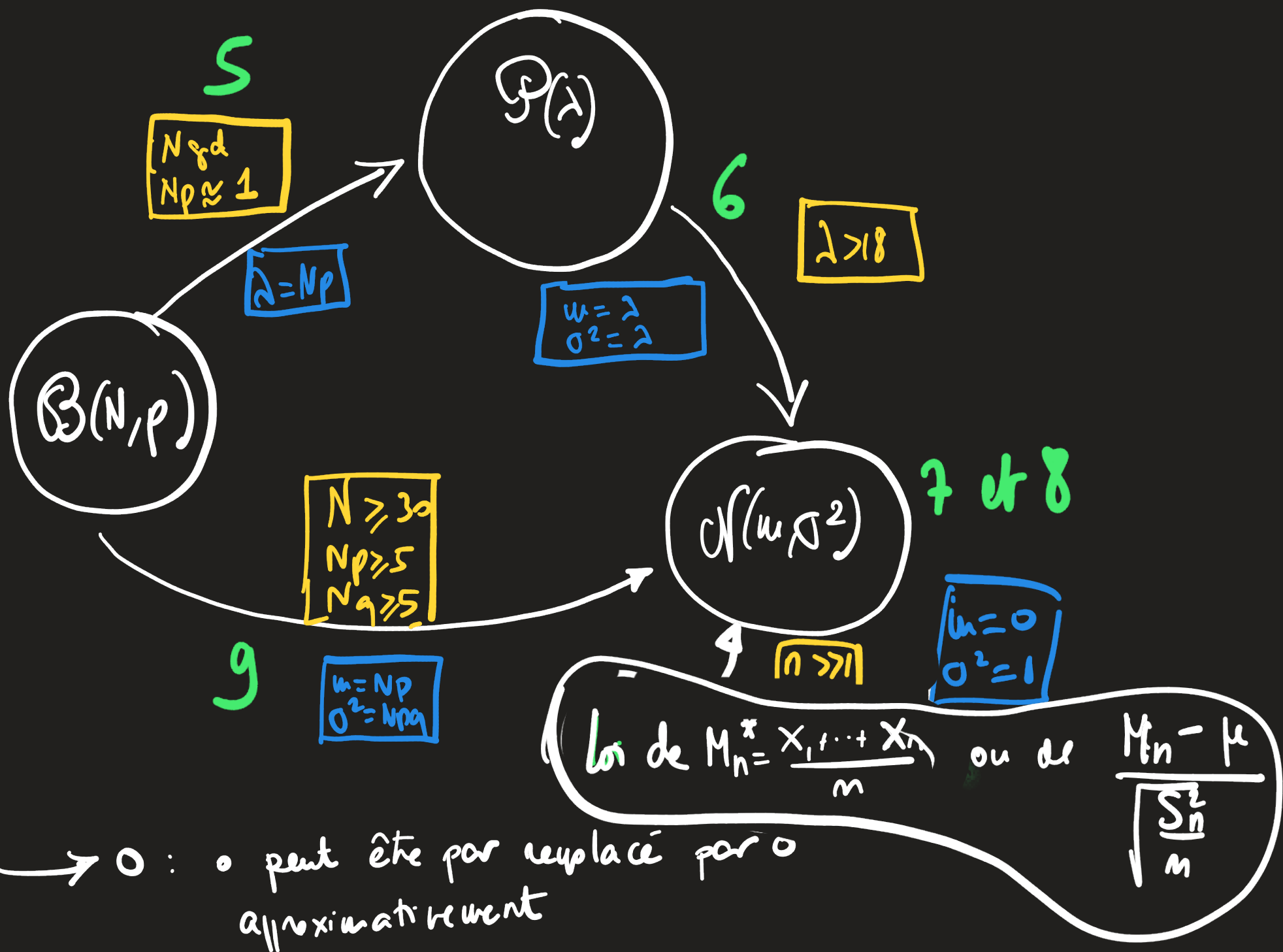
→ on annonce que (H_0) est fautive
et notre verdict est incorrect avec
le risque α .

• Sinon: la valeur observée est probable.

→ On n'a aucune raison d'écarter (H_0)
en l'absence d'éléments supplémentaires.

$41 \in [40,2; 59,8]$: on n'a pas
de raison de penser que la pièce
est truquée.

Schéma de synthèse des théorèmes d'approximation



• $\rightarrow \circ$: $\circ$ peut être remplacé par $\circ$ approximativement

 Conditions de validité de l'approximation

 définition des param. réf. du thm

(X_i) : n échantillon d'une loi q.c.g.

$$E(X_i) = \mu \quad \forall i$$

$$V(X_i) = \sigma^2 \quad \forall i$$

Exemple: Exo 139 TD.

- Un étudiant légendaire ne faisait en moyenne qu'une faute d'orthographe tous les 500 mots.
- Probabilité qu'il ne fasse pas plus de 5 fautes dans une composition de 200 mots.

- Posons $n = 200$ (= nb de mots écrits).
- Notons X_i la V.A.R. égale à 1 si le i^e mot écrit contient une faute, et 0 sinon.
- La variable aléatoire égale au nombre de fautes commises par l'étudiant est:

$$Y = X_1 + \dots + X_n.$$

On suppose que (mais c'est un peu abusé) que les X_i sont mutuellement indépendantes.

$$X_i \rightarrow \mathcal{B}(p) \text{ où } p = \frac{1}{500}.$$

D'après le théorème de stabilité des lois binomiales $Y \sim \mathcal{B}(n, p)$.

On veut calculer $P(Y \leq 5)$

on utilise l'approximation Poissonienne:

$$n = 200 \text{ grand ? } \checkmark$$

$$np = \frac{200}{500} = \frac{2}{5} \text{ de l'ordre de } 1 \checkmark$$

$$\text{Soit } Z \sim \mathcal{P}(np) = \mathcal{P}\left(\frac{2}{5}\right)$$

$$P(Y \leq 5) \simeq P(Z \leq 5)$$

$$\text{Calculons } P(Z \leq 5) = e^{-\frac{2}{5}} \left(1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \frac{x^5}{120} \right)$$

où $x = \frac{2}{5}$.

$$\text{rem: } P(Y \leq 5) = \binom{200}{0} q^{200} + \binom{200}{1} p q^{199} + \binom{200}{2} p^2 q^{198} + \binom{200}{3} p^3 q^{197} + \dots + \binom{200}{5} p^5 q^{195}$$

$p = \frac{1}{500}$
 $q = \frac{499}{500}$

beaucoup mais facile à calculer

② De nos jours: on s'en fiche: on fait ça à l'ordinateur.