Un bref résumé synthétique

Loi binomiale
(n tirages oui ou non)

X = 1 si succès probabilité **p**
X = 0 si échec probabilité **q = 1 – p**

Pour X

E(X) = p
V(X) = p – p² = p(1–p) **= pq**

Pour Z = $\sum X_i$

P(Z=k)=C_n^k**p^k q^(n-k)**
E(Z) = np
V(Z) = npq
σ(Z) = $\sqrt{npq}$

Pour $\frac{Z}{n}$ **fréquence des succès sur n**

E($\frac{Z}{n}$**) =p** **V(**$\frac{Z}{n}$**) =** $\frac{pq}{n}$

Loi hypergéométrique
(n tirages sans remise)

N boules dont F noires, N-F blanches
X = 1 si noire, n tirages sans remise

Exemple 5 noires , 10 blanches, n=5
P(01010) = $\frac{10}{15}\frac{5}{14}\frac{9}{13}\frac{4}{12}\frac{8}{11}$ = P(11000)

P(X=x) = $\frac{C_F^x C_{N-F}^{n-x}}{C_N^n}$

E(X) = p

V(X) = npq $\frac{N-n}{N-1}$

Loi exponentielle
(Durée de vie sans vieillissement)

La probabilité de durer t au-delà de h = probabilité de durer t .
Densité f(t) = $\lambda e^{-\lambda t}$
Répartition P(t <x) = 1 – $e^{-\lambda x}$
Espérance E(t) = $\frac{1}{\lambda}$
Variance V(t) = $\frac{1}{\lambda^2}$
Écart type σ(t)= $\frac{1}{\lambda}$
Médiane $\frac{\ln 2}{\lambda}$

Loi de Poisson
Évènement aléatoire E dans le temps ou l'espace

Si ΔZ = portion de temps ou d'espace
P(E)=pΔZ (P(E) proportionnel à ΔZ)
X =nombre d'événements observés sur Z
Si on pose m = pZ (P(E) sur Z)
P(X=x sur Z) = $\frac{e^{-m}m^x}{x!}$
E(X) = m
V(X) = m
De la loi binomiale à Poisson.
Soit B(n,p) telle que np = 3 ou 4.
Une succession d'épreuves rapides telle que sur Δt E se produise en moyenne 3 ou 4 fois.
Quand n →∞ et p→ 0 (np → m)
P(X=x) → $\frac{e^{-m}m^x}{x!}$
On se sert de
$$\boxed{z \in \mathbb{C} \quad e^z = \lim_{n\to\infty}\left(1+\frac{z}{n}\right)^n}$$

Loi normale
La moyenne (E(x)), la médiane et le mode (P(E) max) coïncident. De plus P(m-2σ < X < m+2σ) > 95%

Densité de probabilité
$$f(x) = \boxed{\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}}$$
Après changement de variable
$$\boxed{T = \frac{x-m}{\sigma}} \quad \text{fx dx devient} \quad \frac{1}{\sqrt{2\pi}}e^{-\frac{T^2}{2}}dT$$
Et $\int_{-\infty}^{+\infty} f(T)\,dT$ = 1 d'après Gauss.
Cas général loi N(m,σ) centrée en m
Médiane = m
Mode = m
Écart type = σ
P(T < x) donnée par des tables .
Ne pas oublier que X =Tσ +m
De la loi binomiale à la normale.
Si n est grand et que p proche de ½.
B(n,p) d'écart type σ moyenne np→ N (np,σ)

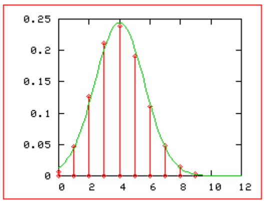


Diagramme en bâtons de B(n,p) et graphe de N(np,σ)

Loi du KHI2 de Pearson

Loi du χ² de Pearson
On calcule **X=**$\sum_i^n \frac{(Oi-Ei)^2}{Ei}$
Oi =valeur observée Ei = estimée
Le nombre de degrés de liberté est égal au nombre n de données indépendantes diminué du nombre de relations liant ces données. (Par exemple n–1 si on veut évaluer une moyenne ou si la somme des variables = 100).
Dans un tableau destiné à étudier le croisement de L modalités de A et de C modalités de B, le nombre de degrés de libertés est (L–1)(C–1).

Dans la table du khi 2 à k degrés de libertés on a en fonction de X la probabilité que les séries Oi et Ei « divergent ». Plus cette probabilité est faible, plus les séries sont probablement de même loi .
Ce test permet d'estimer l'adéquation de la population à une loi ou à une autre, l'homogénéité ou l'indépendance de 2 populations.

Bienaymé Tchebytchev

P(Z ≥ t) ≤ $\frac{E(Z)}{t}$

P(|x–m| ≥ t) ≤ $\frac{\sigma^2}{t^2}$ **(t arbitraire)**
x fréquence mesurée
t Rayon de l'intervalle de confiance
σ²/t² Probabilité pour x dans l'intervalle
Si X obéit à la loi binomiale σ² =$\frac{pq}{n}$.
Ce qui permet de répondre à la question : "Comment faut-il choisir n pour approcher p la moyenne théorique d'au plus 5% = t au risque de 1% = σ²/t²"

Convergence , loi faible

X1,X2,…Xn
Converge en probabilité vers X si
∀ ε > 0 $\lim_{n\to\infty}$ **P (| Xn – X | < ε) = 1**
Converge en moyenne quadratique
E(Xn) → E(X) et $\lim_{n\to\infty}$ **V(Xn – X) = 0**
Converge en loi (f(x) fct de répartition)
f(x)continue en x₀ $\lim_{n\to\infty}$ **P(Xn<t) = f(x)**

Loi faible des grands nombres
La fréquence f(A,n) du caractère A, mesurée dans un échantillon d'effectif n de la population, tend vers P(A) la fréquence effective (réelle) de A dans la population quand n →∞ .

Limite centale

X1,X2,…Xn indépendantes de même loi m = 0 , σ = 1
Yi =$\frac{Xi}{\sqrt{n}}$ suit une loi N(0,1) qd n→∞
X1,X2,…Xn indépendantes de même loi moyenne m écart type σ
Les variables $Y_i = \frac{X_i - m}{\sigma}$ ont pour moyenne **0** et pour variance 1 (donc on se ramène au cas précédent)
$Y = \sum \frac{Y_i}{\sqrt{n}}$ peut aussi s'écrire $\frac{\bar{X}-m}{\frac{\sigma}{\sqrt{n}}}$ et on retrouve la formulation de la loi de la limite centrale : La loi de
X= $\frac{\bar{X}-m}{\frac{\sigma}{\sqrt{n}}}$ **converge vers N(0 , 1) quand n → ∞**

Probabilités et statistiques

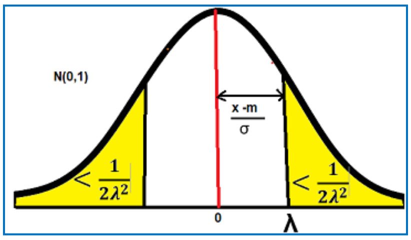
Ce chapitre des probabilités a pour but
◙ de préciser l'ajustement d'une loi de probabilité par une autre et l'approximation d'une distribution donnée par une loi théorique.
◙ de donner les limites d'erreurs possibles dans l'estimation d'un paramètre d'une population ce qui sera utile lors de l'estimation ou de l'échantillonnage.
Il s'agit donc clairement de faire le lien entre probabilités et statistiques.

Inégalité de Bienaymé – Tchebychev (B – T).

◙ Soit une variable aléatoire **X** dont on ne connaît que l'espérance **m** et la variance **V = σ^2**

◙ Soit t un nombre positif arbitraire

$\mathbf{P(Z \geq t) \leq \frac{E(Z)}{t}}$ (Z variable aléatoire réelle ≥0) **Markov**



◙ Définissons de façon arbitraire un écart à la moyenne positif **t** (ce qui revient à choisir **t (ou λ)** arbitrairement)

$$\mathbf{P(|X - m| \geq t) \leq \frac{\sigma^2}{t^2}}$$

Bienaymé – Tchebychev

ou en posant $\lambda = \frac{t}{\sigma}$ $\mathbf{P(\frac{|X-m|}{\sigma} \geq \lambda) \leq \frac{1}{\lambda^2}}$

$Y = \frac{x - m}{\sigma}$ a pour espérance 1 et pour moyenne 0 donc

Espérance 1$= \int y^2 f(y)dy \geq \int_{|y|\geq\lambda} y^2 f(y)dy \geq \lambda^2 \int_{|y|\geq\lambda} f(y)dy$ $\mathbf{\geq \lambda^2 P(|y| \geq \lambda)}$ d'où le résultat.

Souvent
X est la **<u>fréquence mesurée</u>** pour un caractère
m la **<u>fréquence effective</u>**
t l'écart maximum toléré entre fréquence mesurée et fréquence effective (intervalle de confiance) en fonction de **n** nombre d'observations (qui entre dans la composition de σ)
σ^2 / t^2 est le risque maximum (l'erreur) ou la probabilité pour que notre mesure soit bien dans l'intervalle de confiance (ce qui nous permet de déterminer **t)** .

Exemple :

Dans une population le caractère A a la probabilité **p** et le caractère $\bar{A}$ la probabilité **q = 1 – p** .
Quel effectif de la population (n) faut – il examiner pour que la fréquence observée pour **A** dans cet effectif ne s'écarte pas de plus de **0,05** de la fréquence effective avec un **risque** inférieur à **1%** ?

Solution :
◙ on reconnaît une loi binomiale donc la fréquence effective de **A** dans la population doit être m = **p**.
◙ la variable observée **X** est la fréquence de **A** dans la population.
◙ **P (|X – m| ≤ t)** s'écrit **P(|X – p| ≤ 0.05)**
◙ « **avec un risque inférieur à 1%** » s'écrit **P(|X – p| ≤ 0.05) ≥ 0.99**
la probabilité pour que la fréquence mesurée ne diffère pas de plus de 5% de la fréquence effective est de 99%.
On pourrait écrire : **P(|X – p| ≥ 0.05) ≤ 0.01**
◙ Donc **t = 0.05**.
◙ La loi binomiale donne **<u>pour les fréquences</u>**
Si Y est le nombre d'évènements d'une sorte (caractère A observé)

$E(\frac{Y}{n}) = p$
$V(\frac{Y}{n}) = \frac{pq}{n}$
Et **σ^2 = pq / n**
$\mathbf{P(|X - m| \geq t) \leq \frac{\sigma^2}{t^2}}$
◙ t = 5% → $\frac{\sigma^2}{t^2}$= 1% → $\frac{pq}{nt^2}$ = 1% →n = $\frac{0,25x0,75}{0,01x0,0025}$ → n ≥7500.

Convergence

Soit X une variable aléatoire.
Soit X1 , X2 , ..., Xn une suite de variables aléatoires de même champ que X.
Cette suite **converge en probabilité** vers X si

$$\forall \varepsilon > 0 \quad \lim_{n\to\infty} P(|Xn - X| < \varepsilon) = 1$$

Cette suite **converge en moyenne quadratique** vers X si

$$E(Xn) \to E(X)$$

$$\lim_{n\to\infty} V(Xn - X) = 0$$

De plus on a « convergence en moyenne quadratique » implique « convergence en probabilité ».
Cette suite **converge en loi** vers **X** de fonction de répartition **F(x)** si
Tous les **Xi** étant de même loi et de fonction de répartition **Fi(x)**

F(x) étant continue en x_0 $\lim_{n\to\infty}$ **P(Xn<t) = F(x)**

Si (Xn) converge en loi vers X, Alors Xn converge en probabilité vers X (réciproque fausse)
X peut être une constante.
Si (Xn) converge en loi vers X, Alors f(Xn) converge en probabilité vers f(X) (f continue)

Loi faible des grands nombres

Soit X1 , X2 , ... Xn une suite de variables aléatoires indépendantes
- **de même loi (par exemple loi binomiale)**
- **de même espérance E(X)**
- **de même variance σ^2**

Alors la moyenne des Xi converge en probabilité vers E(X)

.

variable	valeurs				moyenne
X1	x11	x12		x1p	E(X)
X2	x21	x22		x2p	E(X)
Xn	xn1	xn2		xnp	E(X)
moyenne des Xi	M1	M2		Mp	E(X)

Si **n** $\to \infty$ on a **lim (M_i) = E(X)**

$$\lim_{n\to+\infty} P\left(\left|\frac{X_1+X_2+...+X_n}{n} - E(X)\right| \geq \epsilon\right) = 0$$

Pour le démontrer on remarque que la variable M a pour espérance **E(X)** et pour variance **V(X) / n**
Donc d'après la loi de B – T

$$P\left(\left|\frac{X_1+X_2+...+X_n}{n} - E(X)\right| \geq \epsilon\right) \leq \frac{V(X)}{n\epsilon^2}$$

Il suffit maintenant de faire tendre **n** vers ∞ .

Corollaire :

Supposons que dans une population la probabilité (la fréquence) du caractère **A** soit **p** (inconnu)
je réalise **n** tests en prélevant chaque fois au hasard un élément de la population.
à chaque test correspond une variable de Bernouilli **Xi** ($1 < i \leq n$) dont la valeur est **1** si on trouve le caractère **A** chez l'élément de la population et **0** si on ne le trouve pas.
J'ai donc **n** variables indépendantes **Xi,** de même loi, dont l'espérance mathématique est **p** et la variance **pq**.
Je me trouve dans les conditions de la loi faible des grands nombres.
Sur **n** tests, la moyenne des **Xi** est égale à **la fréquence $f_{A,n}$ de A dans la population testée**.
La loi des grands nombres me dit que lorsque **n** $\to \infty$

$$\forall \varepsilon > 0 \quad \lim_{n\to\infty} P(|f_{A,n} - p| < \varepsilon) = 1$$

Ce que l'on traduit par :

La fréquence $f_{A,n}$ du caractère **A**, mesurée dans un échantillon d'effectif **n** de la population, tend vers **P(A)** la fréquence effective (réelle) de **A** dans la population quand **n** $\to \infty$.

Théorème de la limite centrale

(Liapounov – Lindeberg – Lévy – Gnedenko – Kolmogorov)

Soit X1, X2, ...Xn une suite de variables aléatoires indépendantes
- **dont la loi de distribution est la même**
- **dont l'espérance mathématique est E(Xi) = m**
- **dont la variance est V(Xi) = σ^2**

Si $\overline{X}$ est la moyenne arithmétique de ces n variables alors la variable

$$X = \frac{\overline{X} - m}{\frac{\sigma}{\sqrt{n}}}$$

converge en loi vers une variable aléatoire de loi N(0,1) lorsque n → ∞ .

Soient **X1, X2, ... Xn** vérifiant les conditions de l'énoncé de la loi

Alors la variable **Sn = X1+X2+......Xn** admet pour espérance **nm** et pour écart type $\sigma\sqrt{n}$
En effet : quand X et Y sont indépendantes on a
- E(X + Y) = E(X) + E(Y) d'où E(Sn) = nm
- V(X+Y) = V(X) + V(Y) d'où V(Sn) = n σ^2

La loi de **Sn** tend vers la loi normale **N(nm , $\sigma\sqrt{n}$)** quand n → ∞

Donc la loi de **Zn** = $\frac{Sn-nm}{\sigma\sqrt{n}}$ tend vers une loi **N(0,1)** (On centre et on réduit)

Or **Zn** peut s'écrire $\frac{\frac{Sn-nm}{n}}{\frac{\sigma\sqrt{n}}{n}} = \frac{\overline{X}-m}{\frac{\sigma}{\sqrt{n}}}$ dont la loi tend vers **N(0,1)** quand **n** → ∞

Limite centrale

⊕ Si X1, X2, ..., Xn sont des variables indépendantes ayant chacune 0 pour moyenne et 1 pour variance

- Posons **Yi** = $\frac{X_i}{\sqrt{n}}$ **Y suit une loi N(0,1)**

⊕ Si X1 , X2, ... Xn sont des variables quelconques de même loi ayant pour moyenne m et pour variance σ

Les variables $Y_i = \frac{X_i - m}{\sigma}$ ont pour moyenne **0** et pour variance 1 (donc on se ramène au cas précédent)

Il suffit de remarquer que $Y = \sum \frac{Y_i}{\sqrt{n}}$ peut aussi s'écrire $\frac{\overline{X}-m}{\frac{\sigma}{\sqrt{n}}}$ et on retrouve la formulation de la loi de la limite centrale :

La loi de $\frac{\overline{X}-m}{\frac{\sigma}{\sqrt{n}}}$ converge vers N(0 , 1) quand n → ∞

De la loi binomiale à la loi normale

Une variable distribuée selon la loi binomiale **B(n,p)** converge en loi vers une variable distribuée selon la loi normale **N(np , $\sqrt{npq}$)** lorsque **n** → ∞ .
La convergence est d'autant plus rapide que **p** est voisin de **0,5.**

Pour la loi binomiale on a **E(X) = np** et **σ** = $\sqrt{npq}$ **.** Il en va de même pour la loi normale.

L'approximation est judicieuse dés que npq dépasse 10.

pq étant maximum et égal à **0,25** pour **p = 0,5** . on devrait avoir en tout état de cause **n > 40**.

Exemple Un centre de transfusion sanguine doit relancer une catégorie de donneuses et donneurs pour un don exceptionnel. L'établissement a en effet besoin de 340 dons supplémentaires.
On considère que chaque donneuse ou donneur contacté€ a une probabilité 0.7 de faire ce don.
Sachant que la relance est très onéreuse, le centre de transfusion contacte n personnes, avec n ≥ 340.
Soit Yn la variable aléatoire correspondant au nombre de personnes contactées se déplaçant pour faire ce don exceptionnel.
En considérant que les personnes acceptent de faire un don indépendamment les unes des autres, Yn suit la loi binomiale de paramètres n et 0.7 .
On considère d'abord que n = 500, n est donc assez grand pour faire l'approximation de la loi de Yn par une loi normale.
Calculons la probabilité qu'il manque au moins 20 dons :
Yn obéit à la loi binomiale B(500,0,7) donc σ(Y500) =$\sqrt{npq}$ = $\sqrt{500x0{,}7x0{,}3}$ =$\sqrt{105}$ et m= E(Y500) = np=0,7x500 =350
Donc 500 étant assez grand on devrait avoir B(50,7) → N(350, $\sqrt{105}$)
Donc on va chercher dans les tables de Z = $\frac{Y500-m}{\sigma}$ = $\frac{Y500-350}{\sqrt{105}}$
Mais avant il faut transformer la probabilité cherchée P(Y500≤340–20) en fonction de Z.
Or Y500≤340–20 équivaut à Y500 –350 ≤ –30 et à P($\frac{Y500-350}{\sqrt{105}}$)≤ $\frac{-30}{\sqrt{105}}$ et on doit chercher 1-P(Z ≤2,93) dans la table.

$$\mathbb{P}(Y_{500} \leq 340-20) = \mathbb{P}(Y_{500}-350 \leq -30) \quad = \mathbb{P}(\frac{Y_{500}-350}{\sqrt{105}} \leq -\frac{30}{\sqrt{105}}) = \mathbb{P}(Z \leq -\frac{30}{\sqrt{105}})$$

où Z suit la loi $\mathcal{N}(0,1)$. Donc on a

$$\mathbb{P}(Y_{500} \leq 340-20) = \mathbb{P}(Z \geq \frac{30}{\sqrt{105}}) = 1-\mathbb{P}(Z \leq \frac{30}{\sqrt{105}}) = 1-\mathbb{P}(Z \leq 2.93) = 1-9983 = 0.0017.$$

On s'intéresse maintenant au nombre minimal n de personnes à contacter pour limiter à 5% la probabilité qu'il manque des donneuses ou donneurs. On doit avoir

$$\begin{aligned}\mathbb{P}(Y_n < 340) \leq 0.05 &\Leftrightarrow \mathbb{P}((Y_n-0.7n)/\sqrt{0.21n} < (340-0.7n)/\sqrt{0.21n}) \leq 0.05\\ &\Leftrightarrow \mathbb{P}(Z < (340-0.7n)/\sqrt{0.21n}) \leq 0.05\\ &\Leftrightarrow 1-\mathbb{P}(Z < -(340-0.7n)/\sqrt{0.21n}) \leq 0.05\\ &\Leftrightarrow \mathbb{P}(Z < -(340-0.7n)/\sqrt{0.21n}) \geq 0.95.\end{aligned}$$

On doit donc avoir

$$\begin{aligned}-(340-0.7n)/\sqrt{0.21n} \geq 1.65 &\Leftrightarrow 0.7n-340 \geq 1.65\sqrt{0.21n}\\ &\Leftrightarrow 115600+0.49n^2-476n \geq 0.57n\\ &\Leftrightarrow 0.49n^2-476.57n+115600 \geq 0.\end{aligned}$$

(avec la contrainte $0.7n-340>0$ ie $n \geq 486$). Ce polynôme a pour discriminant

$$\Delta = (476.57)^2 - 4 \times 0.49 \times 115600 = 542.96,$$

et pour racines

$$\frac{476.57 \pm 23.3}{2 \times 0.49} = 462.52 \text{ et } 510.07$$

Donc finalement on doit avoir n supérieur à la plus grande racine, soit $n \geq 511$.
Pour déterminer n On part de P(Yn<340) ≤0,05 et on cherche l'inégalité pour Zn. Or Yn =Znσ+m = Zn$\sqrt{npq}$ +np
Donc notre inégalité est équivalente à P(Zn < $\frac{340-0{,}7n}{\sqrt{0{,}21n}}$) < 0.5 On passe à P(Z>x) la table donne – $\frac{340-0.7n}{\sqrt{0.21n}}$ ≥ 1.65
Ce qui permet de trouver 2 racines 462 et 511 et on choisit que n doit être plus grand que 511 pour que le risque d'avoir au moins 340 donneurs soit inférieur à 5%.

Connexion des différentes lois

Loi hypergéométrique **n** Tirages sans remise **N** effectif total initial **F** effectif favorable initial **X** nombre de tirages favorables sur n. **P(X = x)** = $\frac{C_F^x C_{N-F}^{n-x}}{C_N^n}$ **E(X) = np** **V(X) = npq** $\frac{N-n}{N-1}$	⇨ Pour **n <** $\frac{N}{10}$ et **N** grand **H(N,F,n)** → **B(n,**$\frac{F}{n}$)	**Loi binomiale B(n,p)** **n** Tirages avec remise Nombre de « succès » **P(X = x) =** C_n^x **$p^x q^{n-x}$** **E(X) = np** **V(X) = npq** Fréquence des « succès ». **E(f) = p** **V(f)** = $\frac{pq}{n}$
	⬃ Pour **p < 0,1** et **n > 50** **B(n,p)** → **P(m = np)** ⬃	⇩ Pour **npq > 10** **B(n,p)** → **N(np ,** $\sqrt{npq}$**)**
Loi de Poisson Aléatoire dans le temps ou l'espace (mesuré par Z) . **m=pZ** probabilité qu'un évènement se produise sur Z. **X**= nombre d'évènements produits sur Z . **P(X = x)** = $\frac{e^{-m} m^x}{x\,!}$ **E(X) = m** **V(X) = m**		**Loi normale N(m,σ)** Toutes les lois à distribution symétrique par rapport au mode lorsque **n** augmente. **f(x)** = $\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$ **E(X) = m** médiane et mode = **m** **V(X) = σ^2**