

Université Claude Bernard · Lyon
FACULTE DE PHARMACIE

**COURS
DE
STATISTIQUES**

C.Z. Paultre

1991-1992

INSTITUT DES SCIENCES PHARMACEUTIQUES ET BIOLOGIQUES

Universite Claude Bernard • Lyon
FACULTE DE PHARMACIE

**COURS
DE
STATISTIQUES**

C.Z. Paultre

1990 - 1991

INSTITUT DES SCIENCES PHARMACEUTIQUES ET BIOLOGIQUES

PREMIERE PARTIE

CALCUL DES PROBABILITES

CHAPITRE I

INTRODUCTION AUX METHODES STATISTIQUES

I - PHENOMENE ALEATOIRE

La notion de reproductibilité d'une observation s'appréhende assez bien dans les sciences expérimentales.

Exemple 1 :

On mesure le PH d'une solution d'acétate de Na 0,1 N à 27°C. Si on relève les valeurs mesurées à des intervalles de temps on obtient 8,2 ; 8,3 ; 8,4 ; 8,2 ; 8,2 ; 8,2 ; 8,3 ; 8,2.

La non reproductibilité d'un phénomène observable est un phénomène tout à fait général. Les causes de cette non reproductibilité sont complexes et leur analyse approfondie sort du cadre de cet exposé. Schématiquement on peut considérer les facteurs liés au système d'observation, les facteurs liés au système observé, parmi ces facteurs il y a ceux que l'on contrôle plus ou moins bien (réglages du PH mètre, température de la solution etc...) et ceux plus ou moins connus que l'on ne contrôle pas du tout (vieillesse des électrodes, hétérogénéité de mélange etc...)

Cette insuffisance de contrôle qui traduit toujours un manque d'information est à la base de la non reproductibilité observée.

Remarque : Deux valeurs identiques d'une mesure données par un même appareil ne signifient pas qu'elles sont exactement égales. En effet elle sont entachées d'une incertitude implicite qui n'apparaît pas toujours clairement et donne l'impression d'une fausse reproductibilité.

Exemple 1 : 8,2 et 8,2 signifient 8,2 ? et 8,2 ? la fluctuation s'inscrit dans le point d'interrogation.

La reproductibilité peut aussi s'appréhender dans des situations plus complexes où la mesure n'intervient pas.

Exemple 2 :

Effet d'un médicament X sur une maladie Y. L'effet peut être apprécié par des critères complexes que l'on exprime par des expressions telles-que : amélioration, pas d'amélioration, aggravation.

Ici encore on observe une fluctuation du résultat selon les malades et cette non reproductibilité du résultat comme précédemment traduit un manque de connaissance sur les facteurs susceptibles d'influencer la guérison des malades.

A partir de ces deux exemples et de leur commentaire on peut essayer de dégager quatre points fondamentaux.

1) **Définition d'un phénomène aléatoire :**

Phénomène dont la réalisation ne peut être prévue exactement.

2) La plus part des phénomènes observables dans les sciences expérimentales sont aléatoires.

3) La reproductibilité est plus ou moins bonne selon les phénomènes observés.

4) Le manque de reproductibilité traduit toujours en dernière analyse un manque d'information sur le phénomène observé.

Ces deux derniers points permettent de comprendre les deux attitudes complémentaires que l'on a dans les sciences expérimentales.

1) Réduire la fluctuation en augmentant la qualité et quantité de l'information pour mieux contrôler l'observation. *réduire cours cochés P.*

2) Mesurer la fluctuation des phénomènes pour mieux les prévoir. Le deuxième point de vue fait l'objet de la méthode statistique.

II - La méthode de statistique

Elle peut être décomposée assez artificiellement en trois domaines :

- la statistique descriptive qui s'attache à représenter les résultats d'observations en mettant l'accent sur leur fluctuation.

- la statistique déductive qui étudie des modèles mathématiques pour représenter la fluctuation on l'appelle aussi calcul des probabilités. (*→ modèles math.*)

- La statistique inductive qui applique les modèles aux résultats observés pour représenter la fluctuation des observations, faire des prévisions et permettre de prendre des décisions. (*pratique*)

III - DEFINITION STATISTIQUE D'UN EVENEMENT

III.1 - Epreuve

Une épreuve est une expérience dont le résultat n'est pas connu à l'avance.

Exemple 1 : Dans le tirage d'une ampoule injectable dans un lot de fabrication on peut définir une épreuve comme prendre et mesurer le volume d'une ampoule.

*R → volume
P → casser et voir*

Exemple 2 : Dans l'effet produit par un médicament X sur une maladie Y, on peut définir une épreuve comme prendre un individu malade, le traiter et étudier l'effet du traitement.

Une épreuve conduit à établir un résultat sur un phénomène aléatoire.

III.2 - Evènement élémentaire

Un évènement élémentaire est le résultat exact d'une épreuve.

Exemple 1 = nombre d'ampoules défectueuse dans un lot.

Exemple 2 = volume observé d'une ampoule prélevée dans un lot n'est pas un évènement élémentaire car le nombre exact est inconnu à cause des incertitudes expérimentales.

Exemple 3 = L'amélioration d'un signe clinique en est un, le résultat peut être quantitatif ou qualitatif et il s'exprime dans ce dernier cas par une expression du langage parlé (syntagme).

III.3 - Ensemble fondamental

C'est l'ensemble de tous les évènements élémentaires possibles pour une épreuve donnée. On peut rencontrer divers types d'ensembles, dénombrables ou non, finis ou infinis, avec des évènements qualitatifs ou quantitatifs.

Exemples :

$E = \{1, 2, 3, 4, 5, 6, \}$ fini, dénombrable, quantitatif.

$E = \{\text{amélioration, pas amélioration, stationnaire}\}$ fini, dénombrable, quantitatif.

$E = \{0 < V < 1,2\}$ fini, indénombrable, quantitatif. *= intervalle continu.*

(pas accès à ét volume)

etc...

Pour une grandeur continue, on ne peut exprimer un évènement élémentaire. Mais il est très facile d'exprimer un ensemble fondamental.

III.4 - Evènements

On appelle évènement A un sous ensemble de l'ensemble fondamental E

$$A \subset E$$

Exemple 1 : Notes de mathématiques supposées entières $E = \{0,1,2,\dots,39,40\}$

L'évènement A être passable en maths peut être défini par $A = \{20,21,22\}$. On dit que l'évènement A s'est réalisé lorsqu'un évènement élémentaire de A est réalisé.

Exemple 2 : Volume d'une ampoule $E = \{1,0 < V < 1,5\}$. On observé $V = 1,25 \text{ cm}^3$ ce résultat est bien un évènement car il est sous entendu ici ($\{1,24 < V_{\text{exact}} < 1,26\}$) et cet intervalle est bien un sous ensemble de E.

Sur les grandeurs continues, on définit simplement des évènements par des intervalles.

Le langage ensembliste est largement utilisé en statistique pour les évènements où il s'avère efficace et expressif.

Ainsi : Φ ensemble vide est l'évènement impossible.

E l'ensemble fondamental est un évènement certain.

On peut aussi combiner des évènements pour en former de nouveaux.

$C = A \cup B$	(ou)	union
$D = A \cap B$	(et)	intersection
$C_A = \bar{A}$	(non)	complément

si $A \cap B = \emptyset$, A et B sont dits incompatibles ou disjoints.

Exemple : Soit la glycémie sur un animal $E = \{0_{g/l} < G < 10_{g/l}\}$ On peut définir selon les besoins, différents évènements et leur donner une définition verbale ainsi :

$$G_1 = \{0_{g/l} < G < 0,8_{g/l}\} \rightarrow \text{hypoglycémie}$$

$$G_2 = \{0,8_{g/l} < G < 1,2_{g/l}\} \rightarrow \text{glycémie normale}$$

$$G_3 = \{1,2_{g/l} < G < 10_{g/l}\} \rightarrow \text{hyperglycémie}$$

Il est alors facile de vérifier que E est un évènement certain, $G_4 = \{11_{g/l} < G < 12_{g/l}\}$ est un évènement impossible, que $G_1 \cap G_2 = \emptyset$ est un évènement impossible et que $G_1 \cup G_2 = C_{G_3} = \bar{G}_3$ ce qui équivaut bien à ne pas être hyperglycémique.

IV - INDIVIDU POPULATION ECHANTILLON

IV - Individu

Un individu est un système identifiable sur le quel on peut observer un (ou plusieurs) caractères (qualitatifs ou quantitatifs) \Rightarrow connaissance et identification
individualisation du résultat. $n = 4,7 \rightarrow$ individu

Exemple : Une souris dans une boîte est un individu, s'il existe un procédé qui permet de la reconnaître au milieu de ses semblables. Sur un tel individu on pourrait observer son poids (caractère quantitatif) ou son anxiété (caractère qualitatif).

Mais en statistique un individu est aussi la valeur observée d'un caractère quantitatif, ou la modalité d'un caractère qualitatif.

Exemple : La 10ème valeur d'intensité mesurée sur une même résistance, dans les mêmes conditions physiques, est un individu (caractère quantitatif). Ici encore la valeur est identifiable (individualisable) indépendamment de sa valeur, le critère choisi dans cet exemple est le rang de la valeur. La définition statistique de l'individu n'est donc pas exactement celle du langage courant.

IV.2 - Population

Une population est un ensemble d'individus. Cette définition très simple appelle quelques remarques et commentaires. *nb variable $z \rightarrow \infty$*

Une population peut être de petite taille et ce n'est pas le sens courant du terme population. Ainsi trois souris dans une boîte forment une population au sens statistique. Une population est parfois difficile à définir car les critères d'appartenance ne sont pas toujours nets. *individualisable / individualise*

Exemple : population des Français atteints d'hypertension artérielle.

Il est dans certains cas difficiles de savoir si un individu est atteint d'hypertension.

On est parfois conduit à définir la population par les critères d'appartenance des individus sans trop s'intéresser aux individus eux-mêmes. On définit alors une population de taille grande et indéterminée que l'on pourra dans certains cas considérer comme infinie.

On appellera population parente infinie d'une population donnée, une population fictive qui a les mêmes caractéristiques que la population réelle, mais dont la taille est infinie.

Exemple 1 : Population des souris de la race Swiss Star.

C'est une population parente infinie de la population réelle de l'ensemble des individus obtenus dans cette race. *(on prend des sauteuses jaunes)*

Exemple 2 : Population formée de l'ensemble des mesures possibles de courant sur une résistance donnée dans des conditions physiques bien déterminées .

Dans ce cas on a une population fictive de taille ∞ et indénombrable.

Si on étudie tous les individus d'une population, on dit que l'étude est exhaustive. Cette situation est assez rare dans les faits et concerne les populations de petites tailles.

Sur les populations de grande taille on n'observera généralement qu'un nombre restreint d'individus.

IV.3 - Echantillon

On appelle échantillon tout sous ensemble d'individus appartenant à une population.

Pour que les observations sur un échantillon soient le reflet de la population d'où elles proviennent, il faut que l'échantillon soit représentatif de la population.

Cette représentativité, très importante en statistique, nécessite des règles très strictes d'échantillonnage qui seront étudiées dans un chapitre ultérieur.

CHAPITRE II

VARIABLES ALEATOIRES - LOIS DE PROBABILITES

I - DEFINITION DE LA PROBABILITE D'UN EVENEMENT

I.1 - Notion de probabilité expérimentale

Si pour chaque événement d'un ensemble fondamental on peut, par répétition de l'épreuve trouver une fréquence limite d'apparition lorsque le nombre de répétition tend vers l' ∞ on dit que l'ensemble fondamental est probabilisé. On peut dans ce cas donc, à tout événement A, associer un nombre égal à sa fréquence limite qu'on appelle probabilité de l'événement A désigné par P_A .

Cette notion de probabilité suggère que la répétition d'une épreuve permette de mesurer la probabilité d'un événement de manière approchée en mesurant sa fréquence d'apparition $f_A = \frac{n_A}{n}$. L'approximation sera d'autant meilleure que le nombre de répétition n sera plus grand et la convergence plus rapide comme le montre le schéma ci-dessous :

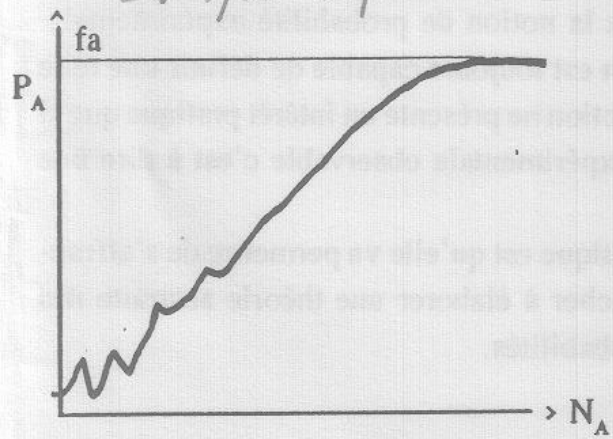
soit :

$$P_A = \lim_{n \rightarrow \infty} \frac{n_A}{n}$$

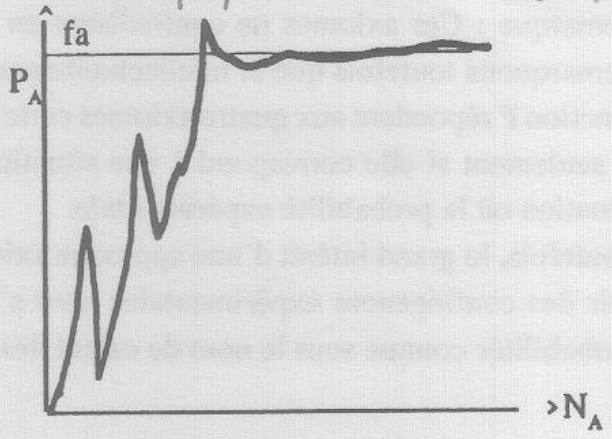
n = nombre total d'épreuves
 n_A = nombre de fois où l'évènement A s'est

eliminer riture de convergence.

pas definition Math. mais def. phenomenologique.



Convergence lente



Convergence rapide

*la mesure experimentale d'une proba est tjrs approchie
 $\lim_{n \rightarrow \infty} f_A = l$ pas tjrs experimentalement.*

Il arrive souvent qu'on ne puisse pas observer de fréquence limite d'apparition d'un événement ce qui apporte une restriction cruciale pour la méthode statistique. Ceci provient en général d'une modification non aléatoire des facteurs qui concourent à la réalisation de l'événement. Ces modifications sont souvent liées à l'évolution des conditions expérimentales au cours des répétitions d'une épreuve et justifient une certaine prudence dans l'approche expérimentale d'une probabilité.

I.2 - Probabilité axiomatique

On peut aussi donner de la probabilité d'un événement une définition purement axiomatique :

Soient E un ensemble fondamental, ϵ la famille des événements et P une fonction réelle définie sur ϵ , on dit que P est une fonction de probabilité et P_A la probabilité de l'événement A si l'on a les axiomes suivants :

- 1) Pour chaque événement A $0 \leq P_A \leq 1$ *P_A qualifie l'événement*
- 2) La probabilité de l'ensemble fondamental $P_E = 1$ *= 1 seul / événement*
- 3) Si 2 événements A et B sont disjoints on dit aussi incompatibles
 $(A \cap B) = \emptyset$ alors $P(A \cup B) = P_{(A)} + P_{(B)}$
- 4) Si la suite des événements A_1, A_2, A_3, \dots est une suite d'événements incompatibles alors
 $P(A_1 \cup A_2 \cup A_3 \dots) = P_{A_1} + P_{A_2} + P_{A_3} + \dots$

Ces quatre axiomes permettent de démontrer les propriétés élémentaires suivantes :

- 1) $P(\emptyset) = 0$
- 2) $P(C_A) = 1 - P_A$ *$C_A \rightarrow \bar{A}$ $P_A \rightarrow A$*
- 3) Si $A \subset B \rightarrow P_A < P_B$
- 4) $P(A \cup B) = P_A + P_B - P(A \cap B)$

Remarque : Ces axiomes ne contredisent en rien la notion de probabilité expérimentale. Remarquons toutefois que si intellectuellement on est toujours capable de définir une telle fonction P répondant aux quatre axiomes cette fonction ne présente un intérêt pratique que si et seulement si elle correspond à une situation expérimentale observable c'est à dire une situation où la probabilité expérimentale.

Toutefois, le grand intérêt d'une approche axiomatique est qu'elle va permettre de s'affranchir des contingences expérimentales pour s'attacher à élaborer une théorie abstraite des probabilités connue sous le nom de calcul des probabilités.

I.3 - Calcul des probabilités : Ensembles équiprobables

Où est ensemble événements équiprobable ?

Dans certains cas la situation expérimentale suggère qu'on puisse attribuer à priori une probabilité à certains événements. L'axiomatique des probabilités permet alors de définir des ensembles fondamentaux probabilisés et de calculer la probabilité de nombreux événements.

L'ensemble le plus fréquent est celui pour lequel les conditions expérimentales permettent de définir un ensemble fondamental E où aucun événement élémentaire ne peut être privilégié. Dans ces conditions on attribue à priori à chaque événement élémentaire une probabilité égale. On appelle de tels ensembles fondamentaux ensembles équiprobables. Dès lors la probabilité de tout événement $A \subset E$, par application des axiomes précédents, peut être directement calculée par dénombrement selon la formule :

$$P_A = \frac{\text{nombre d'événements élémentaires } \in A}{\text{nombre d'événements élémentaires } \in E \text{ fondamental.}}$$

Ce qui s'énonce aussi plus classiquement :

$$P_A = \frac{\text{nombre de cas favorables à la réalisation de A}}{\text{nombre total de cas}} \quad \neq \text{ fréquence } \text{ dénombrement théorique}$$

Remarque 1 : Historiquement cette formule a permis de définir la probabilité d'un événement. On peut la considérer aujourd'hui comme une propriété particulière des ensembles équiprobables.

Remarque 2 : Les exemples d'ensembles équiprobables abondent dans la théorie des jeux.

Remarque 3 : Les ensembles équiprobables jouent un rôle très important en physique statistique et permettent d'énoncer le deuxième principe de la thermodynamique.

Remarque 4 : Le calcul des probabilités par dénombrement sur des ensembles équiprobables fait appel aux formules de l'analyse combinatoire que nous rappelons brièvement.

Le nombre :

a) d'arrangements de n objets distincts pris r à r vaut :

(ordre présentation objet)

b) de permutations de n objets distincts vaut :

$$A_r^n = \frac{n!}{(n-r)!}$$

$$P = A_n^n = n!$$

1^{er} n façon
2^e n-1 "
3^e n-2 "
⋮

c) de permutations de n objets dont n_1 sont semblables, n_2 sont semblables... n_r sont semblables vaut :

$$P = \frac{n!}{n_1! n_2! \dots n_r!}$$

d) de combinaisons de n objets distincts pris r à r l'ordre n'intervenant pas vaut :

$$C^n = \frac{n!}{r! (n-r)!} = \frac{A_n^r}{r!} \quad \text{(ordre intervient pas)} \\ \text{divisé par nb permutation}$$

(coefficients du binôme)

e) de combinaisons d'objets pris r à r parmi n groupes d'objets semblables l'ordre intervenant vaut : n^r

Exemple 1 : Le jeu de dé. Si toutes les faces sont physiquement semblables aucune des faces ne doit être privilégiée et l'ensemble fondamental défini par le résultat d'un lancé,

$E = \{1, 2, 3, 4, 5, 6\}$, doit être considéré comme équiprobable

Soit

$P(1) = P(2) = P(3) = P(4) = P(5) = P(6)$ 6 événements

et comme

$P(1) + P(2) + P(3) + P(4) + P(5) + P(6) = 1$ élémentaire

on en déduit bien que $P(1) = P(2) = P(3) = \dots = 1/6$

On peut sur cet ensemble calculer la probabilité d'autres événements par exemple obtenir un multiple de 3.

$$A = \{3, 6\} \quad E = \{1, 2, 3, 4, 5, 6\} \quad P_A = \frac{1 \cdot 2}{6 \cdot 3} = \frac{1}{3}$$

tirage de l'urne (pharmaceutique) et sans remise.

Exemple 2 : Dans un lot de 12 unités on sait que 4 unités sont défectueuses.

Calculer la probabilité d'avoir 2 unités défectueuses en tirant au hasard 2 unités.

Ici l'expression choisir au hasard signifie que l'ensemble des événements élémentaires est équiprobable.

× L'ensemble fondamental E comprend 66 événements élémentaires équiprobables (nombre total de cas) puisqu'il y a C_2^{12} façons de choisir 12 objets distincts pris 2 à 2 l'ordre n'intervenant pas.

$$C_2^{12} = \frac{12!}{2! \cdot 10!} = \frac{11 \times 12}{2} = 66$$

*ensemble équiprobable : 2 unités
↳ 12 individus*

*sans remise → C
avec remise → A
ordre → A
sans ordre → C*

× Le nombre d'événements élémentaires ε à l'événement avoir 2 unités défectueuses (nombre de cas favorables) est ici $C_2^4 = 6$ puisqu'il y a C_2^4 façons de choisir 4 objets distincts (les unités défectueuses) pris 2 à 2 l'ordre n'intervenant pas.

La probabilité cherchée vaut donc : $\frac{C_2^4}{C_2^{12}} = \frac{6}{66} = \frac{1}{11}$

Remarque : ce type d'exemple est appelé parfois tirage dans une urne sans remise.

Exemple 3 : Même problème mais cette fois lorsqu'on tire au hasard les deux unités avant le deuxième tirage on remet dans le lot l'unité tirée la première fois (tirage avec remise). Le nombre d'évènements élémentaires de E va changer puisqu'il a cette fois $\frac{12^2}{2!}$ façons de prendre 12 objets distincts deux à deux l'ordre n'intervenant pas et qu'il y en a $\frac{4^2}{2!}$ de prendre 4 objets distincts, 2 à 2, l'ordre n'intervenant pas.

La probabilité cherchée vaut donc : $\left(\frac{4}{12}\right)^2 = \frac{1}{9}$

Exemple 4 : Plus subtil emprunté au jeu de dés.

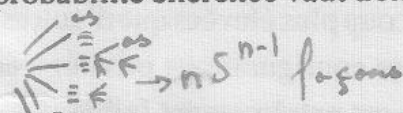
Trouver la probabilité pour qu'un as apparaisse au bout de n lancers.

Pour cela remarquons qu'il y a 6^n façons de prendre 6 objets distincts n à n l'ordre intervenant ce qui définit le nombre d'évènements élémentaires de l'ensemble fondamental équiprobable, alors qu'il n'y a qu'un seul évènement élémentaire pour l'évènement considéré.

La probabilité cherchée vaut donc :

$P_{(A_n)} = \left(\frac{1}{6}\right)^n$

*dim. ombres des év. él. de A
n 5^{n-1}*



Exemple 5 : Important mais franchement compliqué emprunté à la thermodynamique. Soit un ensemble de N molécules discernables. A un instant donné chacune présente un niveau d'énergie appartenant à l'ensemble quantifié ($\epsilon_0, \epsilon_1, \epsilon_2, \dots$). Le système peut donc être, caractérisé par une répartition des molécules sur les niveaux d'énergie = N_0 molécules présentent l'énergie ϵ_0, N_1 l'énergie ϵ_1 etc.... Les molécules étant discernables, à une telle répartition de l'énergie correspond en fait un grand nombre de configurations distinctes chacune d'elles est appelée état accessible e du système. Une répartition de l'énergie va donc définir un évènement R dont les évènements élémentaires e sont les états accessibles, leur nombres sera désigné par Ω .

Si le système est isolé, l'énergie totale est constante mais la répartition sur les niveaux peut être différente, on définit ainsi de nouveaux évènements R_1, R_2, R_3, R_4 ici encore chaque répartition permettra de définir de nouveaux états accessibles (nouveaux évènements élémentaires) dont le nombre sera ($\Omega_1, \Omega_2, \Omega_3, \dots, \Omega_n$)

L'ensemble de tous les évènements R définit l'ensemble fondamental qui est ainsi constitué de l'ensemble de tous les états accessibles appartenant à chaque évènement R

$$R_i = \begin{cases} n_0 \rightarrow \epsilon_0 \\ n_1 \rightarrow \epsilon_1 \\ \vdots \\ n_n \rightarrow \epsilon_n \end{cases}$$

$$E_{\text{therm. T}} = \sum_i n_i \epsilon_i$$

$$N = \sum_i n_i \quad \text{nb molécules}$$

$$E = R_1 \cup R_2, \text{ etc... } \cup R_n$$

$$R_1 = \underbrace{\{e_{11}, e_{12}, \dots\}}_{\Omega_1}, \quad R_2 = \underbrace{\{e_{21}, e_{22}, \dots\}}_{\Omega_2}, \quad \text{etc...}$$

$$R_n = \underbrace{\{e_{n1}, e_{n2}, \dots\}}_{\Omega_n}$$

Du fait des collisions thermiques la répartition des molécules sur les différents niveaux d'énergie va continuellement changer. On considère alors que l'ensemble fondamental E est équiprobable et que chaque événement élémentaire (chaque état accessible) a même probabilité. Dès lors la probabilité d'observer une répartition R_i est donnée par :

$$P(R_i) = \frac{\Omega_i}{\sum_{j=1}^n \Omega_j} = \frac{\Omega_i}{\Omega_{\text{total}}} \quad \Omega_i \ll \Omega_{\text{total}}$$

Le principe d'équiprobabilité des états accessibles constitue le 2ème principe de la thermodynamique et permet d'expliquer très simplement l'évolution thermodynamique des systèmes isolés.

En effet si certaines configurations R ont des probabilités beaucoup plus élevées que d'autres, comme le montre un calcul que nous n'effectuerons pas ici, ces configurations privilégiées seront plus souvent observées que d'autres dont la probabilité est extrêmement faible.

Autrement dit, si on abandonne un système préparé dans une configuration R où le nombre d'états Ω est faible (probabilité faible) celui-ci évoluera inexorablement vers des configurations R où ce nombre est élevé (probabilité forte).

En posant que pour une configuration R_i l'entropie du système vaut $S_i = k \ln \Omega_i$, on voit qu'un système isolé évolue vers des configurations où l'entropie est la plus forte.

I.4 - Probabilités conditionnelles

h = cte de Wolman
S : entropie du système des états
 Ω : nb état accessible $S \propto \Omega$

Sur un ensemble fondamental sur lequel on a défini 2 événements A et B, si on ne s'intéresse à l'événement A que lorsque B s'est réalisé, on définit le nouvel événement A si B noté (A|B). Cet événement a lui aussi une probabilité que l'on définit par :

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Remarque 1 : Cette définition est bien sur compatible avec une définition fréquentielle des probabilités .

$$\text{Ainsi } F_{(A|B)} = \frac{N(A \cap B)}{N(B)} = \frac{\frac{N(A \cap B)}{N}}{\frac{N(B)}{N}}$$

Remarque 2 : cette définition permet de calculer la probabilité de $A \cap B$ soit :

$$P(A \cap B) = P(A|B) P(B).$$

Ayant ainsi défini l'évènement $(A|B)$ on peut définir aussi l'évènement distinct $(B|A)$ avec;

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

On peut de la sorte exprimer $P(A|B)$ en fonction de $P(B|A)$

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

En remarquant que l'évènement B est obtenu soit avec $(B|A)$ soit avec $(B|\bar{A})$ on a $P(B) = P(B|A) + P(B|\bar{A})$

Il vient $P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) + P(B|\bar{A})}$ expression classique du théorème de Bayes

Intérêt : alors si A est un évènement hypothétique, B un évènement observé pour le quel on connaît $P(B)$ alors

$P(A|B)$ donne la valeur de l'hypothèse A à partir de la connaissance de B

Ainsi dans $P(A|B) = \frac{P(A) \cdot P(B|A)}{P(B)}$

$P(A|B)$ est appelé probabilité à postériori de l'hypothèse A

$P(A)$ est appelé probabilité à priori de l'hypothèse A

$P(B|A)$ est appelé fonction de vraisemblance de l'hypothèse A

Exemple : Si avoir la grippe est l'évènement hypothétique A et si l'évènement observé B est avoir la fièvre, alors la probabilité d'avoir la grippe quand on a la fièvre $P(A|B)$ donne la valeur diagnostique de la fièvre observée pour la grippe et est intéressante à connaître. Elle dépend de la probabilité d'avoir la fièvre avec et sans grippe $P(B)$ et d'avoir la grippe avec et sans fièvre $P(A)$ (probabilité à priori) Elle dépend aussi de la probabilité d'avoir la fièvre quand on a la grippe.

si $P(A|B) = P(A)$
 \Leftrightarrow indépendance statistique
 des évènements A et B

si $P(A|B) = 1 \Leftrightarrow$ A totalement lié
 à l'autre

si $P(A|B) = 0 \Leftrightarrow$ A et B totalement exclus

I.5 - Indépendance statistique

La probabilité conditionnelle permet aussi de définir d'une manière très simple l'indépendance statistique de 2 événements A et B.

Si $P(A|B) = P(A)$ il est facile de comprendre que la réalisation de B est sans effet sur celle de A et que A et B sont indépendants et inversement.

Dans ces conditions si A et B sont indépendants, on a les relations fondamentales :

$$P(B|A) = P(B)$$

$$P(A|B) = P(A)$$

$$P(B \cap A) = P(A) \cdot P(B)$$

V - VARIABLE ALEATOIRE

V.1 - Définition

On appelle variable aléatoire X une fonction de l'ensemble fondamental E dans \mathcal{R} , telle que l'inverse de chaque intervalle \mathcal{R} soit un événement de E.

a) Cas des ensembles fondamentaux dénombrables (finis ou infinis)

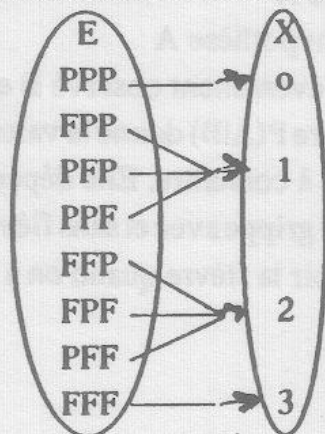
Pour de tels ensembles on dit que la variable aléatoire est discrète ou discontinue.

Le problème qui se pose ici est de savoir quelle valeur numérique il faut attribuer à un résultat qualitatif. On peut ainsi sur un même ensemble fondamental définir plusieurs variables aléatoires.

Exemple : l'ensemble fondamental est défini par les trois résultats successifs obtenus en lançant trois fois une pièce de monnaie :

$$E = \{PPP, FPP, FPF, PPF, FFP, FPF, PFF, FFF\}$$

Définissons la variable aléatoire X comme le nombre de faces tirées, on a alors l'application:



nb associés à des résultats

On voit ici qu'un intervalle sur X
a comme inverse un événement de E

ainsi $0 < x < 1$ correspond à l'évènement tirer au moins 2 piles

Remarque : On aurait pu sur le même ensemble définir d'autres variables aléatoires. Par exemple on peut attribuer à une nouvelle variable aléatoire Y la valeur 1 lorsqu'on a tiré FFF et la valeur 0 pour tout autre résultat, etc...

Lorsqu'on a pour un évènement A , attribué la valeur 1 et pour \bar{A} la valeur 0 on dit qu'on a défini une variable de Bernouilli. Ces variables sont parfois très commodes, on le verra.

b) Cas des ensembles fondamentaux non dénombrables

Pour de tels ensembles on dit que la variable aléatoire est continue.

Ici le résultat d'une épreuve est le résultat de la mesure d'une grandeur continue, l'application de E dans R ne pose aucun problème particulier.

Exemple : l'ensemble fondamental est défini par le poids des comprimés fabriqués par une machine.

$$E = \{0,8g < p < 1,2g\}$$

Il est facile de définir une variable aléatoire X par une application identique de E dans R mais on aurait aussi pu définir une autre variable, par exemple $Y = \log(p + 2)$

Ici encore il est facile de voir qu'un intervalle sur X a comme inverse un évènement de E . Ainsi $0,9 < X < 1,1$ correspond à l'évènement «obtenir un comprimé dont le poids est 1g à 0,1g près»

V.2 - Lois de probabilités

Les variables aléatoires précédemment définies sont des outils commodes pour définir des évènements sur E . Elles vont permettre aussi une représentation simple des probabilités associées à ces évènements sous la forme de lois de probabilités.

a) Cas des variables aléatoires discrètes *discrètes*

La variable aléatoire $x = \{x_1, x_2, \dots, x_n\}$ définie sur un ensemble E , permet de définir la probabilité $P(X = x_i)$ des x_i que l'on écrira $f(x_i)$, puisqu'à chaque valeur de x_i correspond un évènement ayant une probabilité $P(X = x_i) = f(x_i)$ est appelé la distribution ou la loi de probabilité X .

La loi de probabilité f satisfait les conditions :

$$f(x_i) > 0 \quad \text{et} \quad \sum_{i=1}^{i=n} f(x_i) = 1$$

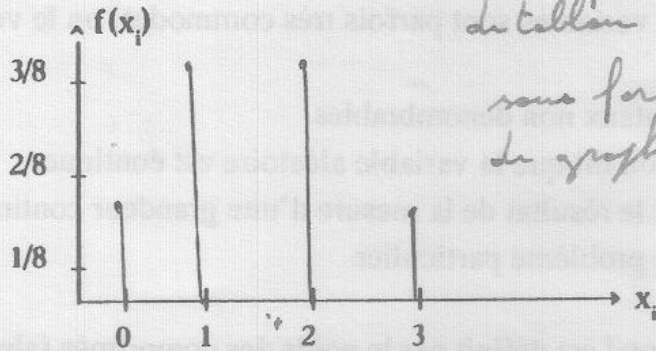
condition de normalisation.
↳ ensemble fondamental $P=1$

La loi de probabilité de x est donnée soit sous forme d'un tableau d'un graphique ou d'une formule :

jeu pile ou face

Exemple :

x_i	0	1	2	3
$f(x_i)$	1/8	3/8	3/8	1/8



sans forme de tableau
sans forme de graph.

$$f(x_i) = C_{x_i}^3 \left(\frac{1}{2}\right)^{x_i} \cdot \left(\frac{1}{2}\right)^{3-x_i} \quad 0 < x_i < 3, \quad x_i \text{ entier}$$

b) cas des variables aléatoires continues

La difficulté pour définir une loi de probabilité provient du fait que $P(X = x_0) = 0$ puisque la probabilité de trouver une valeur exacte est nulle comme nous l'avons déjà souligné alors que $P(x_a < X < x_b)$ est bien définie et généralement non nulle. On suppose alors qu'il existe une fonction $f(x)$ continue par morceau telle que

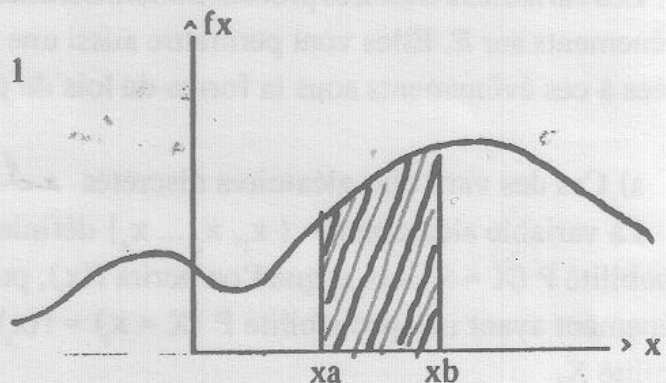
$$P(x_a < X < x_b) = \int_{x_a}^{x_b} f(x) dx \quad f(x) : \text{densité de probabilité}$$

Cette fonction $f(x)$ s'appelle loi de probabilité continue de X ou distribution de X . Pour satisfaire aux axiomes des probabilités il faut avoir :

$$f(x) > 0 \quad \text{et} \quad \int_{-\infty}^{+\infty} f(x) = 1$$

l'aire hachurée est égale à :

$$P(x_a < X < x_b)$$



Cette définition respecte l'axiome d'additivité des événements incompatibles. On appellera probabilité élémentaire dPx :

$$P(x < X < x + dx) = dPx = f(x)dx$$

La fonction $f(x)$ est donnée soit sous forme d'une table, d'une fonction analytique, d'un graphique.

Exemple :

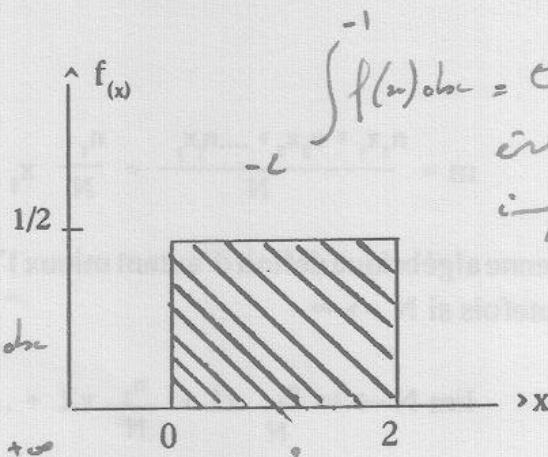
Loi de probabilité uniforme

$$f(x) = 0; x < 0$$

$$f(x) = 1/2; 0 < x < 2$$

$$f(x) = 0; x > 2$$

ex. $P(1,5 < X < 1,6) = \int_{1,5}^{1,6} f(x) dx$
 $\hookrightarrow 0,05$



$\int_{-\infty}^{+\infty} f(x) dx = 1$
 -L
 évidemment impossible.

V.3 - Fonction de répartition

$$\int_{-\infty}^{+\infty} f(x) dx = 0 + \int_0^2 f(x) dx + 0 = 1$$

Pour une variable aléatoire X on définit la fonction de répartition F(x) comme étant égale à P(X < x) variable aléatoire: X valeur précise: x
 Ainsi pour une variable aléatoire discrète on aura

$$F(x_n) = \sum_{i=1}^{i=n} f(x_i)$$

Pour une variable aléatoire continue

$$F(x) = \int_{-\infty}^x f(t) dt$$

Dans ce cas on voit que F(x) est l'intégrale impropre de la loi de probabilité f(x).

C'est une fonction monotone croissante (puisque f(x) > 0) qui sera souvent utilisée en statistique.

V.4 - Espérance mathématique : moyenne
 l'expérience de gain du jeu

Appelée parfois moyenne, l'espérance mathématique d'une distribution est définie dans le cas d'une distribution discrète par :

$$E(X) = \mu = \sum_i x_i f(x_i)$$

et dans le cas d'une distribution continue par

$$E(X) = \mu = \int_{-\infty}^{+\infty} x f(x) dx$$

Cette grandeur caractéristique de la distribution, appelée aussi paramètre de position, définit l'ordre de grandeur du résultat de X auquel on peut s'attendre pour une expérience.

Par exemple dans le cas d'une variable discrète où l'on a répété N fois l'expérience avec n₁ fois le résultat x₁, n₂ fois le résultat x₂ etc..., la moyenne algébrique m des résultats vaut :

$$m = \frac{n_1 x_1 + n_2 x_2 + \dots + n_j x_j}{N} = \frac{n_1}{N} x_1 + \frac{n_2}{N} x_2 + \dots + \frac{n_j}{N} x_j$$

Cette moyenne algébrique définit d'autant mieux l'ordre de grandeur du résultat que N est plus grand. Toutefois si $N \rightarrow \infty$

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{n_1}{N} x_1 + \frac{n_2}{N} x_2 + \dots + \frac{n_j}{N} x_j \\ = f(x_1)x_1 + f(x_2)x_2 + \dots + f(x_j)x_j \end{aligned}$$

$$E_{(x)} = \mu = \lim_{N \rightarrow \infty} m$$

L'espérance mathématique représente donc la valeur limite de la moyenne algébrique des résultats observés pour un nombre ∞ d'épreuves.

Dans le cas d'une distribution continue on obtient la même interprétation.

Exemple 1 : jeu de dé

$$X = \{1, 2, 3, 4, 5, 6\}$$

$$P(X = x_i) = f(x_i) = 1/6$$

pour toute valeur de x

$$\text{soit : } E(X) = \mu = \frac{1}{6} (1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6} = 3,5$$

*≈ moy algébrique
de tous les résultats
obtenus*

Exemple 2 : Dans la loi de probabilité uniforme définie plus haut

$$f(x) = 1/2 \quad 0 < x < 2$$

$$f(x) = 0 \text{ partout ailleurs}$$

$$E(X) = \mu = \int_0^2 \frac{1}{2} x dx \quad \left[\frac{1}{4} x^2 \right]_0^2 = 1$$

V.5 - Variance

Avec l'espérance mathématique, ayant défini l'ordre de grandeur du résultat de x, on peut aussi définir comment les résultats sont dispersés autour de l'espérance mathématique. Cette caractéristique de dispersion est définie par la variance.

Dans le cas d'une distribution continue elle est donnée par :

$$\text{variance} = V(X) = \sigma^2 = \sum_i f(x_i) (x_i - \mu)^2$$

et dans le cas d'une distribution continue par :

$$\text{variance } V(X) = \sigma^2 = \int_{-\infty}^{\infty} f(X) (X - \mu)^2 dx$$

La racine carrée de la variance définit l'écart type (σ). Ici encore la variance représente la valeur limite de la moyenne algébrique observée du carré des écarts à l'espérance mathématique pour un nombre d'épreuves qui tend vers l' ∞

Exemple 1 = jeu de dé

Variance $V(X) = \sigma^2 = \frac{1}{6} (1-3,5)^2 + (2-3,5)^2 + 3(3-3,5)^2 + (4-3,5)^2 + (5-3,5)^2 + (6-3,5)^2 = 2,9$
écart type = $\sqrt{2,9}$

Exemple 2 : loi de probabilité uniforme

$f(x) = 1/2 \quad 0 < x < 2$

$f(x) = 0$ partout ailleurs

$E(X) = \mu = 1$

Variance $V(X) = \sigma^2 = \int_0^2 \frac{1}{2} (x-1)^2 dx = \left[\frac{(x-1)^3}{6} \right]_0^2 = \frac{1}{3}$

Propriété importante pour le calcul de la variance

Cas d'une variable discrète :

$$V(X) = \sigma^2 = \sum_i f(x_i) (x_i - \mu)^2 = \sum_i f(x_i) x_i^2 - 2\mu \sum_i f(x_i) x_i + \mu^2 \sum_i f(x_i)$$

$$= E(X^2) - 2\mu^2 + \mu^2$$

$\rightarrow \sigma^2 = E(X^2) - [E(X)]^2$

VI - OPERATION ET PROPRIETES DES VARIABLES ALEATOIRES

VI.1 - Changement de variable linéaire

Soit la variable aléatoire X d'espérance $E(X) = \mu$ et de variance $V(X) = \sigma^2$ et soit a et b deux constantes arbitraires.

La variable $X' = aX + b$ est encore une variable aléatoire dont on se propose de trouver l'espérance $E(X')$ et la variance $V(X')$. En remarquant que le changement de variable ne change pas la probabilité associée à l'événement on peut donc écrire :

- dans le cas d'une variable discrète

$$E(X') = \sum_i f(x_i) x'_i = \sum_i f(x_i) (ax_i + b) = a \sum_i f(x_i) x_i + b \sum_i f(x_i)$$

soit : $E(X') = a E(X) + b = a\mu + b$

- et dans le cas d'une variable continue

$$E(X') = \int_{-\infty}^{+\infty} f(x) x' dx = \int_{-\infty}^{+\infty} f(x) (ax + b) dx = a \int_{-\infty}^{+\infty} f(x) x dx + b \int_{-\infty}^{+\infty} f(x) dx$$

soit :

$$E(X') = a E(X) + b = a\mu + b$$

Pour la variance, dans le cas d'une variable discrète, il vient :

$$V(X') = \sum_i f(x_i) (x'_i - E(x'))^2 = \sum_i f(x_i) (ax_i - a\mu)^2 = a^2 \sigma^2$$

soit :

$$V(X') = a^2 V(X) = a^2 \sigma^2$$

On trouvera un résultat identique pour une variable continue. Ces changements de variables linéaires permettent de simplifier le calcul de μ et σ^2

Pour cela, il suffit de chercher un changement de variable pour x , pour lequel les calculs de $E(X')$ et $V(X')$ sont plus simples, les formules précédentes permettent alors de trouver $E(X)$ et $V(X)$.

VI.2 - Addition de deux variables aléatoires

Soient deux variables aléatoires X et Y définies sur le même ensemble fondamental, il est possible de définir une nouvelle variable aléatoire $T = X + Y$. En effet à tout couple de valeur (x, y) correspond un évènement de l'ensemble fondamental pour lequel correspond une probabilité définie par $P(x \cap y)$.

Exemple pour les variables aléatoires discrètes :

On tire 3 fois à pile ou face ce qui définit l'ensemble fondamental :

$$E = \text{FFF, FPP, PFP, PPF, FFP, PPF, PFF, FFF}$$

Soit la variable aléatoire X égale au nombre de faces consécutives qu'on a tiré :

	PPP	FPP, PFP, PPF, PFP	FFP, PFF	FFF
$X =$	0	1	2	3

Soit la variable aléatoire Y égale au nombre total de faces qu'on a tiré :

	PPP	FPP, PFP, FFP	FFP, PFP, PFF	FFF
$Y =$	0	1	2	3

On voit bien ici que pour tout couple de valeur (x, y) correspond un évènement de E , pour lequel est définie une probabilité que l'on désignera par $P(x, y)$. Ici par exemple $P(1, 1)$ vaut $1/8$ et $P(T = 2) = P((x=2 \cap y = 0) \cup (x' = 0 \cap y = 2) \cup (x = 1 \cap y = 1))$ les évènements x et y n'étant pas ici indépendants $P(T = 2) = 0 + 0 + 1 \times 3/8 = 3/8$

Il est donc possible de calculer $E(T)$ en écrivant :

$$E_T = \sum_i \sum_j P(x_i, y_j) (x_i + y_j)$$

$$= \sum_i \sum_j P(x_i, y_j) x_i + \sum_i \sum_j P(x_i, y_j) y_j$$

Or $P(x_i) = \sum_j P(x_i, y_j)$ et $P(y_j) = \sum_i P(x_i, y_j)$
 En effet $P(x_i) = P(x_i \cap y_1) \cup (x_i \cap y_2) \cup \dots \cup (x_i \cap y_j) = \sum_j P(x_i, y_j)$
 Par exemple $P(x=1) = P(1 \cap 0) \cup (1 \cap 1) \cup (1 \cap 2) \cup (1 \cap 3)$

$$= 0 + 1 \cdot \frac{3}{8} + \frac{1}{3} \cdot \frac{3}{8} + 0 = \frac{12}{24} = \frac{1}{2}$$

On obtient ainsi : $E_T = E(X) + E(Y)$
 $E(T)$

Remarque : 1) Ce théorème est général et ne suppose nullement que les variables soient indépendantes.

2) Ce même théorème se démontre facilement pour les variables continues

3) Ce même théorème peut s'étendre à la somme de plusieurs variables aléatoires

échantillons :

Cas particulièrement intéressant, s'appliquant à la répétition d'une épreuve : dans ce cas on considère que le premier résultat définit une première variable aléatoire X_1 , le second, la variable X_2 jusqu'à la n ième qui définit X_n . Un ensemble de n résultats constitue un échantillon de taille n . Chacune des variables ainsi définie a même distribution et même espérance mathématique μ .

En répétant l'opération on peut ainsi définir une nouvelle variable aléatoire

$$M = \sum_i \frac{X_i}{n} \quad \text{moyenne algébrique de l'échantillon de taille } n$$

Le théorème précédent nous montre que :

$$E(M) = E\left(\sum_i \frac{X_i}{n}\right) = \frac{1}{n} E \sum_i (x_i) = \frac{n\mu}{n} = \mu = E(X)$$

Ainsi l'espérance mathématique de la moyenne algébrique μ d'un échantillon d'une variable aléatoire X est égale à l'espérance mathématique μ de cette variable X .

La moyenne algébrique d'un échantillon de taille n joue, nous le verrons, un rôle très important en statistiques.

De même on peut aussi pour la variable $T = X + Y$ calculer la variance :

$$\begin{aligned}
 \text{Var}(T) &= \text{Var}(X + Y) = E(X + Y)^2 - [E(X + Y)]^2 \\
 &= E(X^2 + Y^2 + 2XY) - [E(X) + E(Y)]^2 \\
 &= E(X^2) + E(Y^2) + 2E(XY) - E(X)^2 - E(Y)^2 - 2E(X)E(Y) \\
 &= \text{Var}(X) + \text{Var}(Y) + 2E(XY) - 2E(X)E(Y) \\
 &= \text{Var}(X) + \text{Var}(Y) + 2\text{cov}(XY)
 \end{aligned}$$

La quantité $E(XY) - E(X)E(Y)$ s'appelle covariance de X, Y et se note $\text{cov}(XY)$ et pose le problème du calcul de $E(XY)$.

Dans le cas de variables discrètes on a :

$$E(XY) = \sum_i \sum_j x_i y_j P(x_i, y_j)$$

Ici on ne peut comme précédemment séparer les variables X et Y. Mais en posant $E(X) = \mu_x$ et $E(Y) = \mu_y$ on peut démontrer que :

$$\text{cov}(XY) = E[(X - \mu_x)(Y - \mu_y)]$$

En effet on a :

$$\begin{aligned}
 E(X - \mu_x)(Y - \mu_y) &= E(XY - X\mu_y - Y\mu_x + \mu_x\mu_y) \\
 &= E(XY) - \mu_y E(X) - \mu_x E(Y) + \mu_x\mu_y \\
 &= E(XY) - \mu_y\mu_x - \mu_x\mu_y + \mu_x\mu_y \\
 &= E(XY) - E(X)E(Y) \text{ cqfd}
 \end{aligned}$$

La covariance de deux variables se définit donc par :

$$\text{cov}(XY) = E(XY) - E(X)E(Y) = E(X - \mu_x)(Y - \mu_y)$$

Cette définition s'applique aussi aux variables continues.

Corrélation de X et Y.

On définit la corrélation de X et Y par :

$$\rho(x, y) = \frac{\text{cov}(XY)}{\sigma_x \sigma_y}$$

On démontre que ρ est un coefficient compris entre -1 et +1.

proche de 1 : corrélation forte, liaison statistique

Si X et Y correspondent à des événements stochastiquement indépendants, on démonte également : $\text{cov}(XY) = 0$ et $\rho = 0$

En effet dans ce cas et pour des variables discrètes :

$$E(XY) = \sum_i \sum_j X_i Y_j P(x_i, y_j)$$

or ici $P(x_i, y_j) = P(x_i) \cdot P(y_j)$
 et dès lors $E(XY) = \sum x_i P(x_i) \cdot \sum y_j P(y_j) = \mu_x \mu_y$
 et $\text{cov}(XY) = 0$

Cette démonstration peut se retrouver pour les variables continues.

Variance de la moyenne algébrique d'un échantillon :

Puisque $\text{Cov}(XY) = 0$ si les variables X et Y sont indépendantes on peut écrire :

$$\boxed{\text{var}(X + Y) = \text{Var } X + \text{Var } Y}$$

Dans un échantillon les variables X_1, X_2, \dots, X_n sont physiquement et donc stochastiquement indépendantes et leurs variances sont égales à σ^2

Soit

$$\begin{aligned} \text{var}\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) &= \text{var}(M) = \text{var}\frac{(X_1)}{n} + \text{var}\frac{(X_2)}{n} + \dots + \text{var}\frac{(X_n)}{n} \\ &= \frac{1}{n^2} \left(\text{var}(X_1) + \text{var}(X_2) + \dots + \text{var}(X_n) \right) \\ &= \frac{n}{n^2} \text{var}(X) = \frac{\sigma^2}{n} \end{aligned}$$

$$\boxed{\text{var}(M) = \frac{\sigma^2}{n}} = \frac{\text{var}(X)}{n}$$

n taille de l'échantillon

Inégalité de Tchebycheff. Loi des grands nombres.

Pour un échantillon de taille n d'une variable aléatoire X continue ou discrète d'espérance mathématique μ on démontre que la moyenne algébrique m tend vers μ lorsque n tend vers $l'∞$

Démonstration :

Soit une variable aléatoire X discrète dont la distribution est désignée par $f(x)$ d'espérance μ et la variance σ^2 et désignons par ϵ un nombre positif aussi petit que l'on veut.

On a $\sigma^2 = \sum_i (x_i - \mu)^2 f(x_i)$
 en éliminant dans cette somme les valeurs de x_i telles que $|x_i - \mu| < \epsilon$:

$$\sigma^2 > \sum_i^{\circ} (x_i - \mu)^2 f(x_i)$$

$$\sigma^2 > \sum_i^{\circ} \epsilon^2 f(x_i) \quad \text{or ici} \quad \sum_i^{\circ} f(x_i) = P(|X - \mu| > \epsilon)$$

$$\sigma^2 >> \epsilon^2 P(|X - \mu| > \epsilon) \quad \rightarrow \quad P(|X - \mu| > \epsilon) < \frac{\sigma^2}{\epsilon^2}$$

En reprenant ce même calcul pour la moyenne algébrique μ d'un échantillon de taille n défini sur cette même variable X il vient en remplacement σ^2 par $\frac{\sigma^2}{n}$ et X par M

$$P(|M - \mu| > \epsilon) < \frac{\sigma^2}{n\epsilon^2}$$

Si $n \rightarrow \infty$: $P(|M - \mu| > \epsilon) \rightarrow 0$ et $P(|M - \mu| < \epsilon) \rightarrow 1$ et $m \rightarrow \mu$ cqfd

Cette même démonstration peut se faire sur une variable continue.

Cette loi des grands nombre doit être évidemment rapprochée de la notion de probabilité expérimentale d'un évènement définie rappelons le par :

$$P_A = \lim_{n \rightarrow \infty} \frac{nA}{n}$$

En effet nous venons de démontrer que :

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{1}{n} x_i = E(M) = E(X) = \sum_i f(x_i) x_i \text{ en désignant par } x_i \text{ l'évènement } A$$

ceci démontre que $P_A = \lim_{n \rightarrow \infty} \frac{nA}{n}$

On a donc démontré en utilisant uniquement les axiomes de probabilités que la probabilité d'un évènement peut être considérée comme sa fréquence limite d'apparition. Ceci signifie donc que si expérimentalement on peut montrer qu'un évènement possède une fréquence limite d'apparition, c'est que cet évènement peut être probabilisé, qu'on peut lui appliquer l'axiomatique des probabilités et l'ensemble du calcul des probabilités. La loi des grands nombres permet donc le passage de l'expérience à la théorie des probabilités d'où son très grand intérêt.

CHAPITRE III

LOIS DE PROBABILITES USUELLES

modèles mathématiques

I - LOI BINOMIALE OU LOI DE BERNOUILLI

I.1 Définition

Soit une épreuve définissant deux évènements élémentaires A et \bar{A} dont les probabilités respectives sont p et $q = 1 - p$. La loi de probabilité d'une variable aléatoire K_n définie comme le nombre d'évènements A obtenus par n épreuves répétées et indépendantes s'appelle loi binomiale et vaut :

$$P(K_n = k) = C_k^n p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k} \quad P(A) = p \quad P(\bar{A}) = 1 - p$$

Démonstration :

n fixe
 $n = 1 \quad E = \{A, \bar{A}\}$

$K_1 = \{0, 1\}$

$P(AA) = p \cdot p$

$n = 2 \quad E = \{AA, A\bar{A}, \bar{A}A, \bar{A}\bar{A}\}$

$K_2 = \{0, 1, 2\}$

$n = 3 \quad E = \{AAA, A\bar{A}\bar{A}, \bar{A}A\bar{A}, \bar{A}\bar{A}A, A\bar{A}A, \bar{A}AA, \bar{A}\bar{A}\bar{A}\}$

$K_3 = \{0, 1, 2, 3\}$

La probabilité d'un évènement élémentaire telle que $K_n = k$ peut s'écrire :

$$P(\underbrace{A \cap A \cap \dots \cap A}_k \cap \underbrace{\bar{A} \cap \bar{A} \cap \dots \cap \bar{A}}_{n-k})$$

Les évènements A et \bar{A} étant indépendants cette probabilité vaut donc $p^k q^{n-k}$

Le nombre d'évènements élémentaires ainsi défini vaut C_k^n puisqu'il y a C_k^n façons de choisir n objets (le rang du résultat) pris k à k l'ordre n'intervenant pas.

Par exemple pour $n = 3, k = 2$, le rang des résultats A déterminent trois objets distinct 1, 2, 3, les combinaisons 2 à 2 de ces rangs, l'ordre n'intervenant pas, définissent les évènements élémentaires correspondants :

$$\begin{matrix} 1 & 2 \\ \bar{A} & A \end{matrix} \quad , \quad \begin{matrix} 1 & 3 \\ \bar{A} & A \end{matrix} \quad , \quad \begin{matrix} 2 & 3 \\ \bar{A} & A \end{matrix} \quad , \quad \text{evén. indépendant équiprobables}$$

Ces évènements élémentaires tels que $K_n = k$ étant incompatibles, la probabilité de l'évènement

$K_n = k$ est donc égale à la somme des probabilités des événements qui le composent.

$$\text{soit } P(K_n = k) = \frac{p^k q^{n-k} + p^k q^{n-k} + \dots + p^k q^{n-k}}{C_k^n}$$

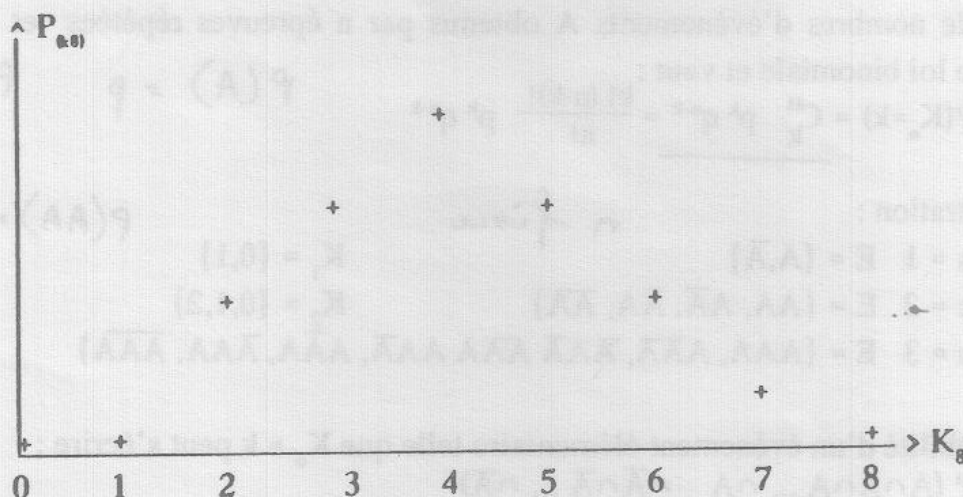
$$= C_k^n p^k q^{n-k} \quad \text{cqfd}$$

*ensemble fondamental
non équiprobable.*

Remarque : cette loi de probabilité discrète est appelée binomiale car C_k^n définit les coefficients du binôme.

Exemple : $n = 8, p = q = 0,5$

K_8	0	1	2	3	4	5	6	7	8
$P(K_8 = k)$	1/256	8/256	28/256	56/256	70/256	56/256	28/256	8/256	1/256



1.2 - Paramètres caractéristiques

a) Espérance mathématique

On démontre que $E(K_n) = np$

En effet si aux deux événements élémentaires A et \bar{A} , on associe la variable de Bernoulli $X = \{1, 0\}$ on a $E(X) = p$ et $\text{var}(X) = pq$; $(p - p^2) = p(1 - p) = pq$. Pour n tirages successifs indépendants on définit n variables X_1, X_2, \dots, X_n d'espérance p et de variance pq dans ces conditions

$$K_n = X_1 + X_2 + \dots + X_n \text{ et donc d'après le chapitre précédent}$$

$$E(K_n) = E(X_1) + E(X_2) + \dots + E(X_n) = nE(X) = np \quad \text{cqfd}$$

$$K_n \neq \text{moyenne}$$

$$K_n = \sum_{i=1}^n X_i$$

b) Variance

On démontre que $\text{var}(X) = npq$

En effet $\text{var}(K_n) = \text{var}(X_1 + X_2 + \dots + X_n)$ les variables X_1, X_2, \dots, X_n étant indépendantes $\text{var}(K_n) = n \text{Var}(X) = npq$ cqfd

I.3 - Usages

La loi binomiale intervient dans tous les problèmes équivalents à un tirage d'urne à deux caractères avec remise. Cette loi est souvent donnée dans des tables, sinon elle peut facilement être calculée à l'aide de machines. Si n est grand $n!$ devient impossible à calculer directement (nombre trop grand).

On peut donc dans ce cas utiliser l'approximation de Stirling :

$$\ln n! \approx n(\ln n) - n + \frac{1}{2} \ln(2\pi n)$$

II - LOI DE POISSON

II.1 Définition

La loi binomiale dans le cas où la probabilité est faible et où le nombre d'épreuves n est important, peut être approchée par la loi suivante appelée loi de Poisson :

lien général

$$P(K_n = k) = \frac{n!}{k!} p^k q^{n-k} \quad p < 0,1$$

$$P(K_n = k) = \frac{(np)^k e^{-np}}{k!} \quad n > 50$$

La démonstration est assez longue :

En posant $np = \lambda$ la loi binomiale peut s'écrire :

\triangle approximation

$$P(K_n = k) = C_k^n p^k q^{n-k} = C_k^n \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$E(K_n) = np$$

et $P(K_n = k-1) = C_{k-1}^n p^{k-1} q^{n-k+1}$ or $C_k^n = C_{k-1}^n \frac{(n-k+1)}{k}$

$$V(K_n) = npq$$

soit $P(K_n = k) = \frac{(n-k+1)}{kq} P(K_n = k-1)$

avec $p \approx 0$ et $q = 1$

$$= \frac{\lambda}{n} P(K_n = k-1)$$

et $P(K_n = k-1) = \frac{\lambda}{k-1} P(K_n = k-2)$ etc...

soit $P(K_n = k) = \frac{\lambda^k}{k.(k-1).(k-2)... \times 1} P(K_n = 0)$

or $P(K_n = 0) = (1 - p)^n = (1 - \frac{\lambda}{n})^n$

$$\ln (1 - \frac{\lambda}{n})^n = n \ln (1 - \frac{\lambda}{n})$$

$\frac{\lambda}{n}$ étant petit en développant le logarithme en série de Taylor il vient :

$$\ln (1 - \frac{\lambda}{n}) = - \frac{\lambda}{n} - \frac{\lambda^2}{2n^2} - \frac{\lambda^3}{3n^3} \dots \text{etc} = \frac{-\lambda}{n} \text{ si } n \text{ grand}$$

soit $P(K_n = 0) = e^{-\lambda}$

soit $P(K_n = k) = \frac{\lambda^k}{k!} e^{-\lambda}$ cqfd

II.2 - Paramètres caractéristiques

Espérance mathématique :

On démontre que : $E(K_n) = np = \lambda$ la démonstration est longue

Variance :

On démontre que : $\text{Var}(K_n) = np = \lambda$ ici également la démonstration est longue et comme les deux résultats ne choquent pas l'intuition nous ne donnons pas ces démonstrations.

II.3 - Usages

Ce sont les mêmes que pour la loi binomiale. Notons toutefois que des événements rares et de grands échantillons se rencontrent assez fréquemment dans les problèmes de contrôles de fabrication.

III - LOI MULTINOMIALE

C'est une extension de la loi binomiale à plus de deux événements

$E = \{A_1, A_2, A_3, \dots, A_j\}$ événements élémentaires
 $P_1, P_2, P_3, \dots, P_j$ probabilités
 $n_1, n_2, n_3, \dots, n_j$ nombres de fois où l'événement est observé

$$P(n_1 \cap n_2 \cap n_3 \cap \dots \cap n_j) = \frac{n!}{\prod_{i=1}^j n_i!} \times \prod_{i=1}^j p_i^{n_i} = \frac{n!}{\prod_{i=1}^j n_i!} \cdot \prod_{i=1}^j p_i^{n_i}$$

Ici $\prod_{i=1}^j n_i! = n_1! n_2! n_3! \dots n_j!$ et $\prod_{i=1}^j p_i^{n_i} = p_1^{n_1} \times p_2^{n_2} \times \dots \times p_j^{n_j}$

avec $\sum_{i=1}^j n_i = n$ nombre total d'épreuves.

Remarque : pour $j = 2$ on retrouve bien la loi binomiale. La loi multinomiale se rencontre en thermodynamique statistique.

IV - LA LOI NORMALE OU LOI DE GAUSS

IV.1 - Définition

C'est une loi de probabilité continue donnée par :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

X est la variable aléatoire telle que $P(x < X < x + dx) = f(x)dx$

μ et σ sont deux nombres réels, avec $\sigma > 0$

L'établissement de cette loi sort du cadre de notre exposé.

IV.2 - Paramètres caractéristiques

On démontre que $E(X) = \mu$

et que $V(X) = \sigma^2$

En effet : $E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$

en faisant le changement de variable $u = \frac{x - \mu}{\sigma}$ soit $x = u\sigma + \mu$ et $dx = \sigma du$

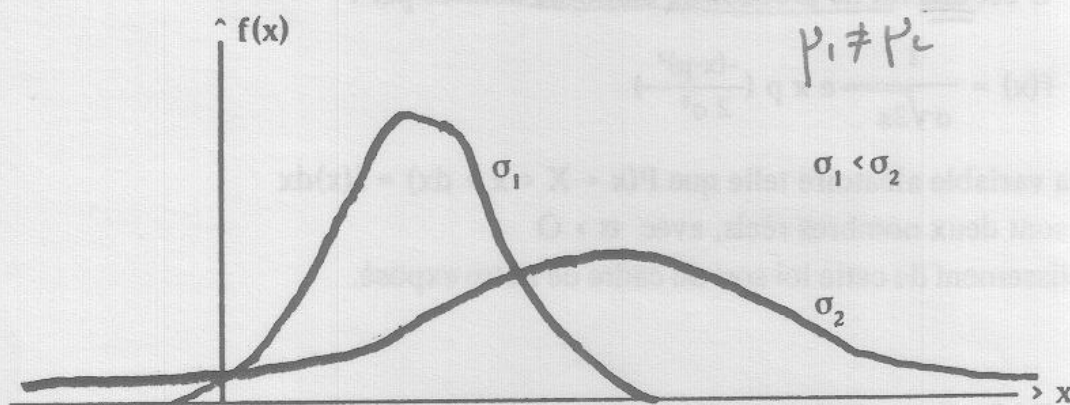
$$\begin{aligned} E(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (u\sigma + \mu) e^{-\frac{u^2}{2}} du \\ &= \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}} du + \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u e^{-\frac{u^2}{2}} du = \mu \\ &= \sqrt{2\pi} \text{ (résultat classique)} \quad = 0 \text{ (par parties)} \end{aligned}$$

et $\text{var} = E(X^2) - E(X)^2$

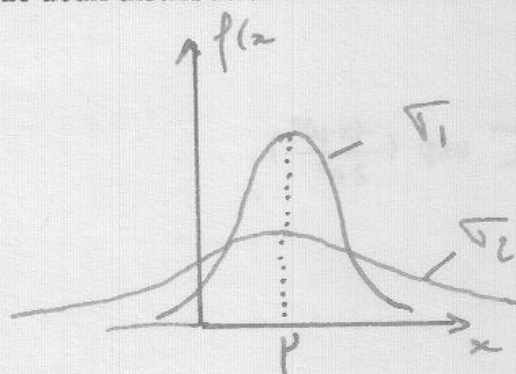
$$\begin{aligned} \text{avec } E(x^2) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (u\sigma + \mu)^2 e^{-\frac{u^2}{2}} du \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u^2 e^{-\frac{u^2}{2}} du + \frac{2\mu\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u e^{-\frac{u^2}{2}} du + \frac{\mu^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}} du = \mu^2 + \sigma^2 \\ &= \sqrt{2\pi} \text{ (par parties)} \quad = 0 \quad = \sqrt{2\pi} \end{aligned}$$

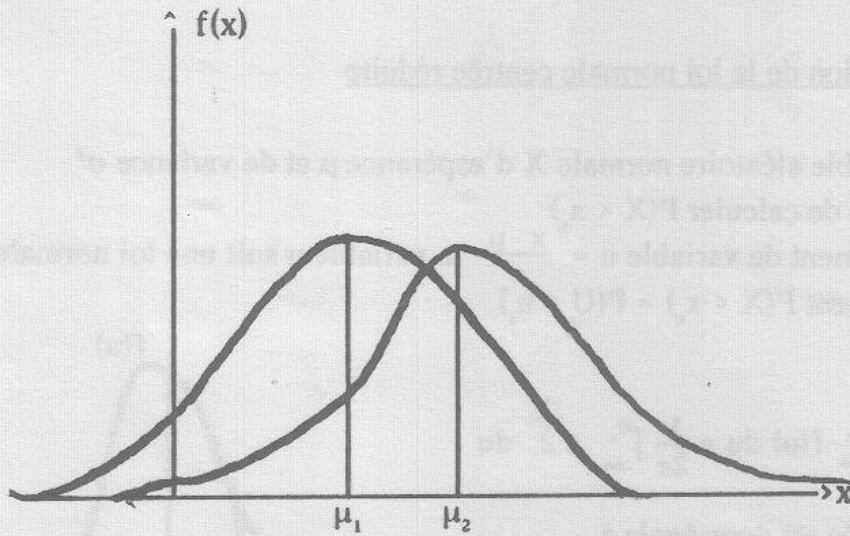
Soit $E(x^2) - (E(x))^2 = \sigma^2$ cqfd

IV.3 Représentation graphique



Comparaison de deux distributions normales de même espérance μ et de variance inégales.





Comparaison de deux distributions normales de même variance σ et d'espérances inégales.

IV.4 - Loi normale centrée réduite

Le changement de variable $u = \frac{x - \mu}{\sigma}$ transforme la loi normale en

$$f(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$$

$$x = u\sigma + \mu$$

$$u = \frac{1}{\sigma} x - \frac{\mu}{\sigma}$$

*changement
variable
linéaire*

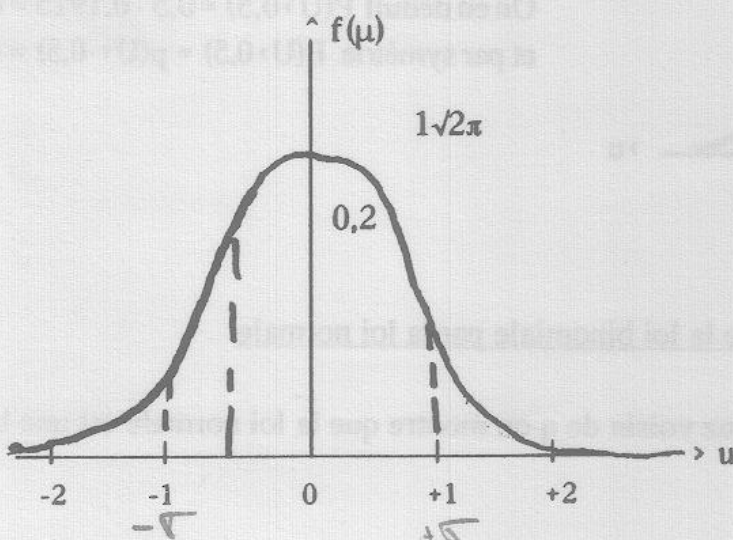
Cette distribution est appelée loi normale centrée réduite elle devient indépendante des valeurs prises par μ et σ^2 d'où son intérêt.

La variable aléatoire U est appelée variable centrée réduite

On voit alors que :

$$E(U) = 0 \quad (= \frac{1}{\sigma} E(X) - \frac{\mu}{\sigma})$$

et que
$$Var(U) = 1 \quad (= \frac{1}{\sigma^2} Var(X) = \frac{\sigma^2}{\sigma^2})$$



IV.5 - Utilisation de la loi normale centrée réduite

Soit une variable aléatoire normale X d'espérance μ et de variance σ^2

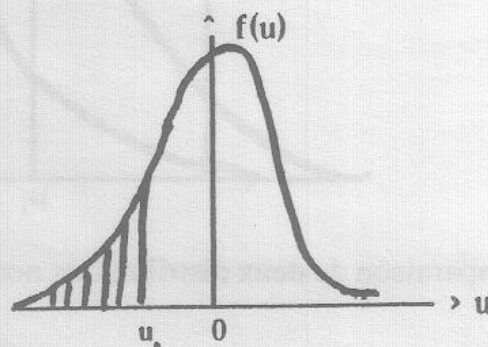
On se propose de calculer $P(X < x_0)$

Par le changement de variable $u = \frac{x - \mu}{\sigma}$ la variable u suit une loi normale centrée réduite et par conséquent $P(X < x_0) = P(U < u_0)$

avec $u_0 = \frac{x_0 - \mu}{\sigma}$

$$P(U < u_0) = \int_{-\infty}^{u_0} f(u) du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u_0} e^{-\frac{u^2}{2}} du$$

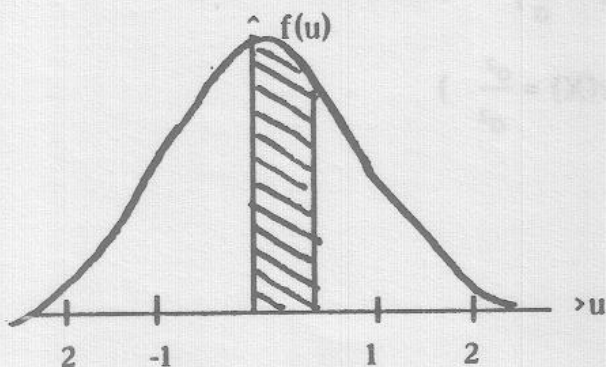
La probabilité cherchée est donc égale à l'aire hachurée sur le dessin



Malheureusement la fonction de répartition $F(u) = \int_{-\infty}^u f(t) dt$ est impossible à intégrer directement. Il faut donc avoir recours à des tables.

Pour utiliser ces tables il faut faire attention à la façon dont sont prises les bornes d'intégration dans la table. En se souvenant que la surface totale sous la courbe vaut 1 et que la courbe est symétrique par rapport à l'axe des abscisses il est facile de trouver la probabilité cherchée.

Exemples : $P(U < -0,5)$



On trouve dans la table

$$P(0 < U < 0,5) = 0,1915 = \int_0^{0,5} f(u)$$

qui correspond à l'aire hachurée sur le dessin.

On en déduit $P(U > 0,5) = 0,5 - 0,1915 = 0,3085$

et par symétrie $P(U > 0,5) = p(U < -0,5) = 0,3085$

IV.6 - Approximation de la loi binomiale par la loi normale

Si n grand et si p est assez voisin de q on montre que la loi normale est une bonne

approximation de la loi binomiale.

$$P(k_1 < K_n < k_2) \cong P(x_1 < X < x_2)$$

$$\mu = np$$

$$\sigma^2 = npq$$

où $x = k_n$, $k_1 = x_1$, $k_2 = x_2$, $f(x)$ loi normale d'espérance np de variance npq .

Exemple : soit une variable aléatoire K_n suivant une loi binomiale avec $p=q=0,5$ et un échantillon de taille $n = 8$

La loi binomiale nous permet de trouver $P(3 < K_8 < 5) = \frac{56}{256} + \frac{70}{256} + \frac{56}{256} = 0,711$

L'approximation par la loi normale nous permet d'écrire :

approximer binom

$$P(2,5 < X < 5,5) = P\left(\frac{2,5-4}{2} < u < \frac{5,5-4}{2}\right) = P(-1,06 < U < +1,06)$$

$$= 0,3355 + 0,3355 = 0,710$$

Remarque : on a transformé l'intervalle $3 < K_8 < 5$ en $2,5 < X < 5,5$ pour tenir compte de la continuité de X . En effet $K_8 = 3$ signifie $2,5 < X < 3,5$ et $K_8 = 5$ signifie $4,5 < X < 5,5$ dès lors $3 < K_8 < 5$ $2,5 < X < 5,5$

IV.7 - Théorème central limite

ADDITION

Soit une variable aléatoire $Y = X_1 + X_2 + \dots + X_n$ où les variables aléatoires X sont indépendantes de même espérance μ de même variance σ^2 et suivent une même loi de distribution qui peut être quelconque. On démontre que lorsque n augmente indéfiniment, Y suit une loi normale d'espérance $E(Y) = n\mu$ et de variance $V(Y) = n\sigma^2$

De même la variable aléatoire $M = \frac{Y}{n}$ suit une loi normale avec $E(M) = \mu$ et $Var(M) = \frac{\sigma^2}{n}$ et la variable $U = \frac{M - \mu}{\sigma/\sqrt{n}}$ suit une loi normale centrée réduite. $U = \frac{n - \mu}{\sigma/\sqrt{n}}$

Le théorème exprime donc également que la loi de distribution de la moyenne algébrique d'un échantillon de taille n suit une loi normale si n est suffisamment grand quelle que soit la loi de distribution de la variable X . On considère en général que n est grand si $n > 30$.

Ce théorème est encore applicable dans le cas où X_1, X_2, \dots, X_n suivent des lois de probabilités différentes.

loi normale ← Y ← inconnus, somme, indépendants

L'existence du théorème central limite est la justification de la place prépondérante qu'occupe la loi normale dans le domaine de la statistique. En particulier chaque fois qu'un évènement est la résultante d'un grand nombre de facteurs aléatoires indépendants agissant par effets sommables non contrôlés, la variable aléatoire associée à l'épreuve suivra une loi normale.

Ce cas est fréquent dans les erreurs de mesure.

En statistique on aura fréquemment recours à des mesures sur des moyennes algébriques qui permettent si la taille de l'échantillon est suffisante d'utiliser la loi normale. Mais il existe d'autres lois de probabilités que nous allons rapidement envisager.

IV.8 - Autres lois de probabilités

a) Loi log-normale

NEPERIEN!

Une variable aléatoire continue X suit une loi log-normale si la variable

$Y = \log(X - a)$ suit une loi normale, a étant une constante.

$$Y = Y_1 + Y_2 + \dots + Y_n$$

$$= \ln(X_1 \cdot X_2 \cdot \dots \cdot X_n)$$

Si $E(Y) = \mu$ $Var(Y) = \sigma^2$ on montre alors que :

$$E(X) = \exp\left(\mu + \frac{\sigma^2}{2}\right) + a, \text{ et } V(X) = (\mu - a)^2(e^{\sigma^2} - 1)$$

$$X = X_1 \cdot X_2 \cdot \dots \cdot X_n$$

MULTIPLICATION

Cette loi se rencontre en pharmacologie et toxicologie où la dose suit une loi log-normale.

b) Loi de χ^2 ou de Pearson

$$\chi^2 = U_1^2 + U_2^2 + \dots + U_n^2 \quad k_i = \chi$$

La variable aléatoire $\chi^2 = U_1^2 + U_2^2 + \dots + U_n^2$ où U sont des variables normales, centrées, réduites a pour loi de distribution :

$$f(\chi^2) = g(y) (\chi^2)^{(y-2)/2} \exp\left(-\frac{\chi^2}{2}\right)$$

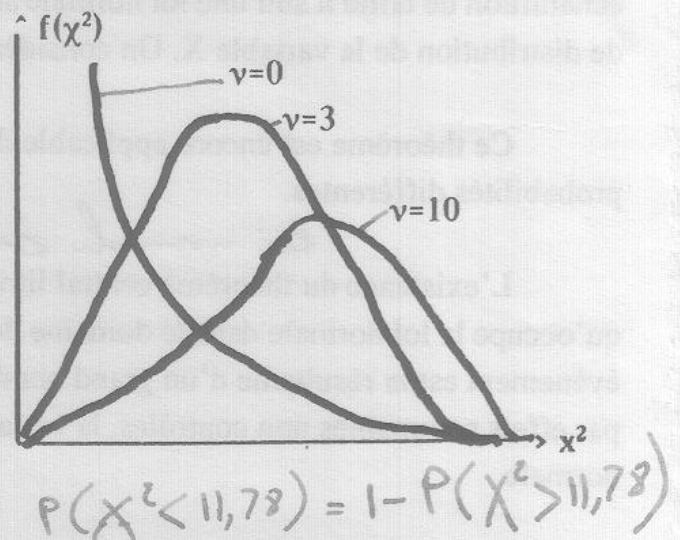
v est appelé nombre de degrés de liberté et représente le nombre de variables indépendantes U dont on a sommé les carrés.

$g(v)$ est un coefficient qui dépend de v

$$E(\chi^2) = v \quad Var(\chi^2) = 2v$$

Les valeurs de la fonction de répartition $P(\chi^2 > a)$ sont données dans des tables en fonction de v

Exemple : pour $v = 5; P(\chi^2 > 9,29) = 0,10$



E: *table*

$v=10$

la moyenne arithmétique de l'échantillon.

$$S^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n-1} \quad \text{III.11}$$

On appelle variance d'échantillonnage la quantité $\hat{S}^2 = \sum_{i=1}^{i=n} \frac{(x_i - \bar{x})^2}{n-1}$ où x_1, x_2, \dots, x_n

sont les valeurs prises par les variables aléatoires X_1, X_2, \dots, X_n associées aux n épreuves indépendantes qui constituent l'échantillon.

Si les variables X sont normales, avec même espérance μ et même variance σ^2 alors on montre que la variable aléatoire : $\hat{S}^2 \times \frac{(n-1)}{\sigma^2}$ suit une loi du χ^2 à $(n-1)$ degrés de liberté.

Exemple soit un échantillon de 10 valeurs x_1, x_2, \dots, x_{10} dont chacune provient d'une variable aléatoire normale d'espérance μ et de variance σ^2

On cherche la probabilité pour que la variance de l'échantillon \hat{S}^2 soit supérieure à deux fois la variance de X soit $P(\hat{S}^2 > 2\sigma^2)$

Dans ces conditions on sait que $\chi^2 = \hat{S}^2 \frac{(n-1)}{\sigma^2}$

$$\text{soit } P(\hat{S}^2 > 2\sigma^2) = P\left(\chi^2 \frac{\sigma^2}{(n-1)} > 2\sigma^2\right) = P(\chi^2 > 2(n-1))$$

$$= P(\chi^2 > 18) = 1 - 0,965 = 0,045$$

$\chi^2 > \chi^2_{(n-1)}$

c) Loi de Student

Si U est une variable normale centrée réduite et X^2 une variable aléatoire indépendante de U distribuée selon la loi du χ^2 à ν degrés de libertés la variable

$$T = \frac{U \sqrt{\nu}}{\sqrt{\chi^2}}$$

suit une loi de distribution de Student

$$\text{donnée par } f(t) = g(\nu) \left(1 + \frac{t^2}{\nu}\right)^{-\frac{(\nu+1)}{2}}, \quad -\infty < t < +\infty$$

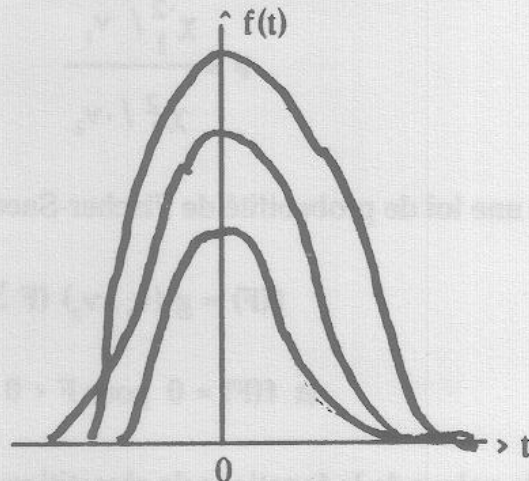
ici encore $g(\nu)$ est un coefficient qui dépend de ν

si $\nu \rightarrow \infty$ $f(t) \rightarrow$ loi normale

$$E(T) = 0 \text{ pour } \nu \neq 1$$

$$\text{Var}(T) = \frac{\nu}{\nu-2} \text{ pour } \nu > 2$$

Les valeurs de la fonction de répartition $P(T > a)$ sont données dans des tables en fonctions de ν



Exemple $\nu = 19, P(T > 0,39) = 0,65$

X, Y e var indépendantes, loi normale
 avec $E(X) = \mu_x$
 $E(Y) = \mu_y$

Au paragraphe IV.6 il a été montré que la moyenne algébrique M d'un échantillon de n variables aléatoires indépendantes suit une loi normale si n est suffisamment grand et que $U = \frac{M - \mu}{\sigma/\sqrt{n}}$ suit une loi normale centrée réduite.

On montre que $T = \frac{M - \mu}{s/\sqrt{n}}$ est une variable aléatoire qui suit une loi de Student à $(n - 1)$ degrés de liberté où S est l'écart type de l'échantillon.

On montre aussi que :

$$T = \frac{(m_x - m_y) - (\mu_x - \mu_y)}{\left(\frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2} \right)^{1/2} \left(\frac{1}{n_x} + \frac{1}{n_y} \right)^{1/2}}$$

suit également une loi de Student à $(n_x + n_y - 2)$ degrés de liberté. m_x et m_y sont les moyennes algébriques de deux échantillons issus de deux variables X et Y indépendantes, normales, avec

$$E(x) = \mu_x \quad E(Y) = \mu_y \quad \text{et} \quad \text{var}(X) = \text{var}(Y) = \sigma^2$$

n_x et n_y sont les nombres d'observations pour X et pour Y

S_x^2 et S_y^2 sont les variances des échantillons issus de X et de Y

Cette loi sera utilisée pour comparer deux moyennes algébriques sur deux échantillons.

d) Loi de Fischer-Snedecor

Si χ_1^2 et χ_2^2 sont deux variables aléatoires indépendantes avec respectivement ν_1 et ν_2 degrés de liberté la variable aléatoire

$$F = \frac{\chi_1^2 / \nu_1}{\chi_2^2 / \nu_2}$$

suit une loi de probabilité de Fischer-Snedecor, donnée par :

$$f(F) = g(\nu_1, \nu_2) (F \frac{\nu_1 - \nu_2}{2}) / (\nu_2 + 1 F)^{\frac{\nu_1 - \nu_2}{2}} \text{ pour } F > 0$$

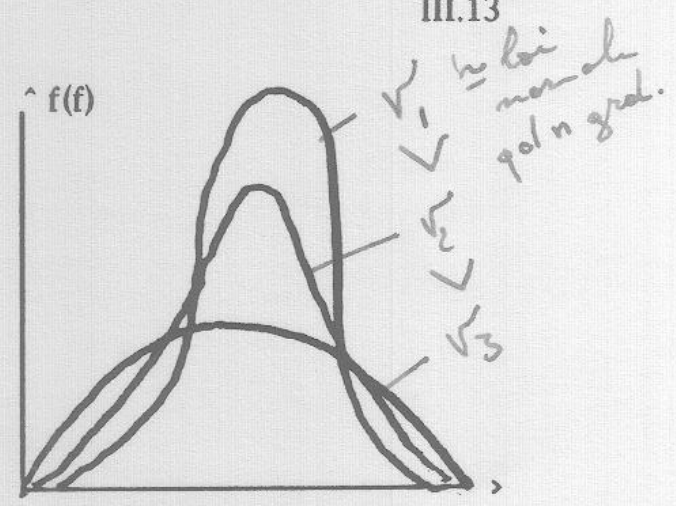
et $f(F) = 0$ pour $F < 0$

Handwritten note: $g(\nu_1, \nu_2) = \frac{\Gamma(\frac{\nu_1 + \nu_2}{2})}{\Gamma(\frac{\nu_1}{2}) \Gamma(\frac{\nu_2}{2})} \frac{1}{2} F^{\frac{\nu_1 - \nu_2}{2}}$

Ici la valeur de la fonctions de répartition $P(F < a)$ sont données dans des tables en fonction des degrés de liberté ν_1 et ν_2

On montre que la variable aléatoire $F =$
 suit une loi de Fischer a ν_x et ν_y degrés de liberté
 Ici \hat{s}_x^2 et \hat{s}_y^2 sont les variances des 2 échantillons
 de taille n_x et n_y obtenus sur deux variables
 normales indépendantes X et Y ; $\nu_x = n_x - 1$
 $\nu_y = n_y - 1$

III.13



DEUXIEME PARTIE

STATISTIQUE INDUCTIVE

PRINCIPE DE BASE

L'étude d'un phénomène aléatoire nécessite que l'expérimentateur et le statisticien collectent sur le phénomène toute l'information disponible pour la traduire en un langage qui se prête à l'utilisation du modèle statistique.

Ces opérations de recueil de l'information portent sur les points suivants.

- 1) La population
- 2) Le plan d'échantillonnage
- 3) L'échantillon
- 4) Les paramètres statistiques
- 5) La distribution d'échantillonnage.

Ces opérations qui font l'objet du premier chapitre vont permettre de porter un jugement statistique sur les résultats de l'échantillon.

On abordera :

- l'estimation d'un paramètre statistique (chapitre II)
- les tests d'hypothèses (chapitre III)
- la corrélation et la régression (chapitre IV)

Dans tous les chapitres de cette deuxième partie nous serons toujours contraints de répondre aux trois questions fondamentales :

- 1) que connaît-on avec certitude ?
- 2) quelles hypothèses fait-on ?
- 3) les conclusions que l'on donne sont-elles compatibles avec ce que l'on connaît ou ce que l'on suppose ?

CHAPITRE I

RECUEIL DE L'INFORMATION STATISTIQUE

I - LA PROBLEMATIQUE

L'expérimentateur doit définir aussi clairement que possible l'objet de son étude et ce qu'il veut savoir sur le phénomène qu'il observe.

Cette évidente nécessité n'est pas toujours dans la pratique suffisamment respectée. En particulier pour des situations complexes l'expérimentateur se contente de quelques intuitions et pense que la méthode statistique va poser et résoudre les problèmes. Cette situation est à l'origine de nombreux malentendus et déboires entre statisticien et expérimentateur.

Exemple I - Contrôle de fabrication

Un responsable de fabrication veut savoir si un lot de comprimés physiquement défini est acceptable du point de vue de sa présentation. Plus précisément si la proportion de comprimés ébréchés est à un niveau acceptable.

Exemple II - Expérimentation pharmacologique

Un pharmacologue veut étudier sur le rat l'effet hypotenseur d'une drogue bien définie.

Ainsi exprimé l'objet de ces deux études est à peu près correctement énoncé, toutefois il faudra encore préciser de nombreux points et en particulier ce qu'on entend par niveau acceptable (exemple 1) par effet hypotenseur (exemple 2).

II - LA POPULATION

II.1 - L'épreuve, la variable aléatoire

L'épreuve définit un résultat élémentaire permettant de définir la variable aléatoire à l'épreuve.

Exemple 1 : L'épreuve consiste à déterminer si un comprimé est défectueux ou non selon des critères donnés sans ambiguïté.

La variable aléatoire associée à l'épreuve est discrète et peut être une variable de Bernoulli.

$$X = \{0,1\} \quad \begin{array}{l} X = 1 \text{ si le comprimé est défectueux} \\ X = 0 \text{ si le comprimé n'est pas défectueux} \end{array}$$

On évite de définir d'emblée la situation expérimentale par des épreuves de répétition.

Exemple 2 : L'épreuve consiste à mesurer la pression artérielle chez un rat auquel on a administré la drogue dans des conditions expérimentales bien précisées. La variable aléatoire associée à l'épreuve est continue, elle peut être égale à la valeur de pression artérielle.

Pour des grandeurs continues, compte tenu de l'imprécision de la mesure, la valeur observée à l'issue d'une épreuve n'est jamais une valeur exacte. Elle constitue un évènement non élémentaire.

II.2 - Individus, règles d'appartenance

Chaque résultat observé doit permettre, grâce à des règles d'appartenance, de définir les individus de la population.

Exemple : Les comprimés soumis à l'étude (individus de la population) sont définis par :

1) les caractéristiques de la machine et ses réglages ainsi que tous les facteurs techniques que l'on contrôle (fabrication du granulé etc...)

2) les conditions d'existence physique la taille du lot, le jour et l'heure de fabrication. Dans ces conditions on a une population fictive de taille infinie si l'on ne précise pas les conditions d'existence physique. Ces populations s'appellent populations parentes infinies.

Exemple : 1

Si l'on ne précise pas le jour, l'heure de fabrication, on définit l'ensemble de tous les comprimés que l'on pourrait obtenir avec les caractéristiques et réglages que l'on a précisé. Cette population n'a pas d'existence réelle, sa taille peut être infinie.

Dans le cas où le résultat est une grandeur continue, la définition des critères d'appartenance des individus à la population est assez semblable.

Exemple : 2

Les pressions artérielles de rat soumises à l'étude sont définies par :

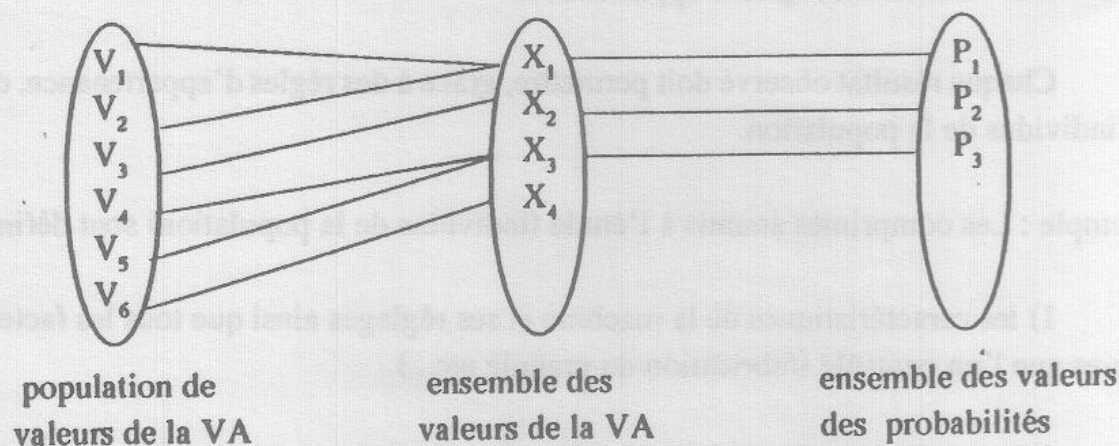
- 1) souche, sexe, age, poids, etc... précisés,
- 2) administration de la drogue bien définie,
- 3) conditions d'existence physique précisées (rats présents dans l'animalerie tel jour, telle heure...)

Les critères 1, 2, 3 définissent une population finie d'individus dénombrables ayant une existence physique.

La suppression du critère 3 définit ici encore une population parente infinie.

Les populations d'individus définissent des populations de valeurs prises par la variable aléatoire. Dans la suite de ce cours on utilisera le mot population avec pour sens population de valeurs.

Attention : il ne faut pas confondre population de valeurs et ensemble de définition de la variable aléatoire. Le schéma suivant fait comprendre la nature des applications que l'on réalise.



Plusieurs individus peuvent prendre une même valeur de la VA qui leur est associée. On ne peut donc caractériser complètement un individu par la valeur de la VA qui lui est associée.

Par contre, la loi de probabilité de la VA caractérise la composition de la population de valeurs de la VA. L'espérance mathématique et la variance de la VA n'apportent qu'une information partielle sur cette composition.

Exemple 1 : (comprimés)

La probabilité d'obtenir 1 comprimé défectueux dans le lot de taille finie vaut :

$$P(X = 1) = \lim_{N \rightarrow \infty} \frac{N_1}{N}$$

N_1 = nombre de fois où l'on obtient 1 comprimé défectueux
 N = nombre de fois où l'on a répété l'épreuve

La composition du lot est caractérisée par :

$$\frac{n_1}{n}$$

n_1 = nombre de comprimés défectueux dans le lot
 n = nombre total de comprimés dans le lot

Si l'on a $P(X = 1) = \frac{n_1}{n}$ la loi de distribution caractérise bien la composition de la population.

La loi de probabilité de la V.A. ne caractérise la composition de la population que sous certaines conditions.

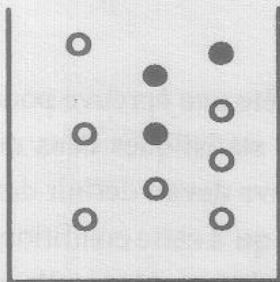
En particulier, il faut que l'épreuve définisse des événements probabilisables on dit alors que les résultats (individus) sont obtenus au hasard.

Pour ce faire on devra respecter certaines règles appelées règles d'échantillonnages qui seront envisagées ultérieurement.

II.3 Modèle de l'urne

C'est une représentation des individus d'une population finie.

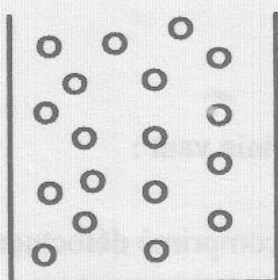
Exemple des comprimés :



Les comprimés défectueux sont représentés par n_1 «boules» marquées 1 et les comprimés non défectueux par n_2 boules marquées 2.

L'épreuve est représentée par le tirage d'une boule au hasard ce qui signifie dans ce cas :

$$P(X = 1) = \frac{n_1}{n}; P(X = 2) = \frac{n_2}{n}$$

Exemple 2 : Le médicament

On marque les boules de la valeur prise par la pression artérielle avec une précision donnée par un animal.

La précision des mesures discrétise sur intervalle la variable continue et nous ramène au cas précédent.

$$P(7,2 < X < 7,4) = \frac{n_{(7,3)}}{n} ; P(7,3 < X < 7,5) = \frac{n_{(7,4)}}{n} ; \text{etc...}$$

pour toute la famille complète d'évènements si les individus sont prélevés au hasard.

Pour les ensembles parents infinis le modèle de l'urne devient physiquement problématique.

La composition de l'urne est définie par une proportion limite.

Exemple des comprimés :

Dans la population parente ∞ on peut définir la composition de l'urne par les proportions limites

$$\lim_{n' \rightarrow \infty} \frac{n'_1}{n'} = \frac{n_1}{n} ; \quad \lim_{n' \rightarrow \infty} \frac{n'_0}{n'} = \frac{n_0}{n}$$

Ces deux valeurs sont égales respectivement à la composition d'une urne de taille finie n avec n_1 boules marquées 1, n_0 boules marquées 0.

Une population parente ∞ est une population qui a même composition qu'une population finie qui permet de la définir. Elle pourra être caractérisée par la même loi de probabilité.

III - PLAN D'ECHANTILLONNAGE

On appelle plan d'échantillonnage, la façon dont on répète une épreuve pour constituer un échantillon de valeurs permettant d'utiliser les modèles statistiques dans de bonnes conditions. Nous avons déjà signalé que la répétition de l'épreuve devait définir des évènements probabilisables au sens phénoménologique (fréquentiel) et qu'à cette condition, la loi de probabilité de la VA caractérisait la composition de la population. Nous allons préciser ces notions et voir leurs conséquences pratiques au niveau de l'expérimentation.

III.1 - Echantillonnage aléatoire

Un échantillon est dit aléatoire si pour une épreuve chaque individu a même probabilité d'apparaître. Le modèle de l'urne permet de bien comprendre cette notion.

Exemple des comprimés

Soit une population de taille finie n et soit n' le nombre de répétitions de l'épreuve constituant la taille de l'échantillon. L'échantillonnage est dit aléatoire s, i à chaque tirage, chacune des boules présentes dans l'urne a même probabilité d'être tirée, indépendamment de la valeur inscrite sur la boule.

Dans ces conditions, à chaque tirage, la probabilité d'observer une valeur de X est égale à la proportion de cette valeur dans l'urne au moment du tirage.

III.2 - Echantillonnage indépendant

On peut considérer qu'un échantillon aléatoire de taille n' est une succession de n' valeurs distinctes correspondant à n' variables aléatoires.

Echantillon = X_1, X_2, \dots, X_n

X_1 Variable aléatoire correspondant au résultat de la 1ère épreuve

X_2 Variable aléatoire correspondant au résultat de la 2ème épreuve

X_n Variable aléatoire correspondant au résultat de la 3ème épreuve

Si ces n' VA ont même loi de probabilité on dit qu'on a réalisé un échantillonnage aléatoire indépendant

$$f(x) = f(x_1) = f(x_2) = f(x_n)$$

Exemple : les comprimés

Si à chaque tirage de l'urne chaque comprimé dans l'urne, a même probabilité d'être tiré (échantillonnage aléatoire) et si à chaque tirage, la composition de l'urne ne change pas on réalise un échantillonnage aléatoire indépendant.

Ceci est obtenu en remettant dans l'urne, après chaque épreuve, la boule tirée (tirage avec remise).

Si la taille de l'échantillon est très petite devant la taille de la population dans l'urne, un tirage sans remise ne modifie pratiquement pas la composition de l'urne et peut dans ce cas être assimilé à un tirage avec remise.

III.4 - Réalisation pratique d'un plan d'échantillonnage aléatoire simple aléatoire (plan EASI)

Pour réaliser un tel plan, le choix des individus doit être déterminé par une épreuve spéciale construite sur un ensemble équiprobable. Chacun des n individus de la population doit être numéroté de 1 à n puis on réalise une épreuve dont l'ensemble fondamental équiprobable comprend les n valeurs.

Le résultat des n épreuves de choix définit les individus qui devront constituer l'échantillon de taille n .

Exemple du médicament

Soit à prélever un échantillon aléatoire de 10 rats dans une population de 250 rats. On repère chacun des 250 rats par un numéro et on organise une épreuve où chacun des 250 numéros a même chance d'être tiré. Pour cela on peut utiliser une loterie à l'aide d'une urne contenant 250 morceaux de papiers numérotés de 1 à 250 ou encore une table de nombres au hasard. On répète dix fois l'épreuve et l'on définit ainsi les 10 animaux qui serviront à l'expérimentation formant l'échantillon de taille 10. Ici bien que le tirage soit sans remise, la taille de l'échantillon est petite devant la population, l'on a réalisé un EASI.

Cette procédure compliquée, selon un plan expérimental strict ne doit pas être considérée comme un luxe de précautions inutiles ou une loterie dérisoire. On peut en effet penser qu'un tirage au hasard des individus est possible directement par le manipulateur. Ceci est illusoire car le manipulateur est très souvent inconsciemment influencé dans son prélèvement par des facteurs non aléatoires.

Exemple du médicament

Lorsqu'on choisit directement des animaux dans une boîte on a tendance à choisir malgré soi les animaux les moins agressifs ou les moins craintifs. Ces animaux généralement plus calmes sont généralement une pression artérielle plus basse que les autres. L'échantillon ainsi choisi devient représentatif d'une autre population avec un nouveau critère d'appartenance : «rats calmes».

Un bon échantillon est donc représentatif d'une population on dit aussi qu'il est sans biais.

Il existe des techniques plus sophistiquées d'échantillonnage lorsque les caractéris

tiques des populations sont complexes. Elles sont utilisées principalement par les instituts de sondage d'opinion et sortent du cadre de cet exposé.

Dans tout ce qui suit nous supposons que les échantillons ont été obtenus par :

Echantillonnage	Aléatoire	Simple	Indépendant
(E A S I)			

Il ne faudra jamais perdre de vue que l'on ne peut jamais être sûr d'avoir un EASI.

III - ECHANTILLON - PARAMETRES CARACTERISTIQUES

Les informations recueillies sur l'échantillon sont des informations certaines qui ne représentent qu'une partie des informations sur la population. Elles se présentent sous forme d'un ensemble fini de valeurs prises par une (ou plusieurs) variables aléatoires.

A partir de ces valeurs, avant d'utiliser le modèle statistique on peut calculer de nombreux paramètres caractéristiques de l'échantillon. Les plus utilisés en pratique sont la moyenne et la variance d'échantillonnage. Si l'échantillon est de taille importante on peut présenter les résultats sous forme de rangement ordonné (rangement par classe, histogrammes).

III.1 - Paramètres caractéristiques

a) Moyenne observée :

C'est la moyenne algébrique des valeurs prises par la variable aléatoire dans l'échantillon :

$$m = \sum_{i=1}^{i=n} \frac{x_i}{n} \quad n = \text{taille de l'échantillon}$$

Si les valeurs de l'échantillon sont regroupées par classe (cf infra) en désignant par i une classe, n_i le nombre d'individus dans la classe, i et x_i la valeur du centre de la classe alors:

$$m \# \frac{\sum_{i=1}^{i=j} n_i x_i}{n} \quad j = \text{nombre de classes}$$

$$n = \sum_{i=1}^{i=j} n_i$$

\hat{S}^2
b) Variance d'échantillonnage observée (ou variance estimée)

Elle est définie par la quantité

$$\hat{S}^2 = \frac{1}{n-1} \left(\sum_{i=1}^{i=n} (x_i - m)^2 \right) = \frac{1}{n-1} \left[\sum_{i=1}^{i=n} x_i^2 - \frac{\left(\sum_{i=1}^{i=n} x_i \right)^2}{n} \right]$$

Si l'échantillon est réparti en classes comme précédemment :

$$\hat{S}^2 = \frac{1}{n-1} \left(\sum_{i=1}^{i=j} n_i (x_i - m)^2 \right) = \frac{1}{n-1} \left[\sum_{i=1}^{i=j} n_i x_i^2 - \frac{\left(\sum_{i=1}^{i=j} n_i x_i \right)^2}{n} \right]$$

Remarque 1 : Dans les calculs de variance éviter d'arrondir les résultats intermédiaires.

Remarque 2 : La variance empirique $S^2 = \frac{1}{n} \sum (x_i - m)^2$ n'est pratiquement jamais utilisé en statistique nous justifierons cette remarque ultérieurement. Il convient donc d'éviter d'utiliser cette formule.

III.2 Histogrammes

- 1) Tableau de répartition, rangement par classe (caractère quantitatif continu = x)
 - a) Déterminer l'étendue $E = x_{\max} - x_{\min}$
 - b) Déterminer le nombre de classes k
 - c) Calculer l'intervalle de classe $x = E/k$, l'arrondir sans trop modifier le nombre de classe
 - d) Dénombrer pour chaque classe le nombre d'individus dont la valeur appartient à la classe. Par convention la borne supérieure de la classe est exclue.
 - e) Dresser le tableau de répartition

rang de la classe	1	2	3	j
centre de la classe	x'_1	x'_2	x'_3	x'_j
limite de classe	x_1, x_2	x_2, x_3	x_3, x_4	x_{j-1}, x_j
effectifs	n_1	n_2	n_3	n_j
fréquence %	f_1	f_2	f_3	f_j
fréquences cumulées %	F_1	F_2	F_3	100%

$$N = \sum_{i=1}^{i=j} n_i \quad \text{effectif total}$$

$$f_i = \frac{n_i}{N} \times 100 \quad \text{effectif relatif ou fréquence}$$

$$F_i = \sum_{l=1}^{l=i} f_l \quad \text{fréquence cumulée}$$

f) Le nombre de classe est réputé correct s'il est supérieur à 6 et si l'effectif le plus élevé dans les classes est au moins égal à 3 fois le nombre total de classe.

nb classe ≥ 6 effectif le plus élevé (mode) $> 3 \times$ nb de classe

2) Histogrammes de répartition

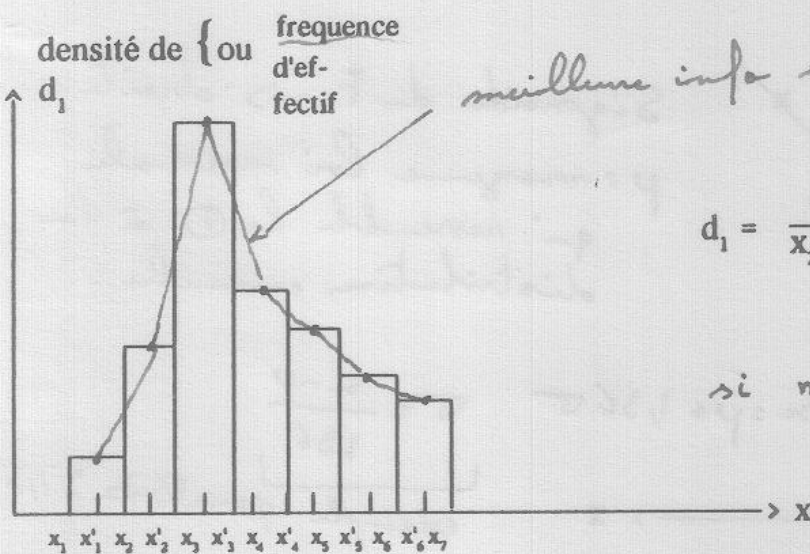
Diagrammes en bâtons de la répartition des valeurs

La répartition peut s'exprimer en utilisant

- les effectifs = n_i Histogramme de répartition d'effectifs
- les fréquences = f_i Histogramme de répartition des fréquences
- les fréquences cumulées = F_i = Histogramme cumulatif des fréquences

Attention : Lorsqu'on utilise les effectifs ou les fréquences on portera en ordonnée pour chaque classe, la densité d'effectif ou la densité de fréquence, obtenue en divisant l'effectif ou la fréquence, par l'intervalle de classe correspondant. De la sorte la surface de chaque rectangle sera proportionnelle à l'effectif ou à la fréquence de la classe correspondante

Lorsque tous les intervalles de classe sont égaux on peut porter directement les effectifs ou les fréquences

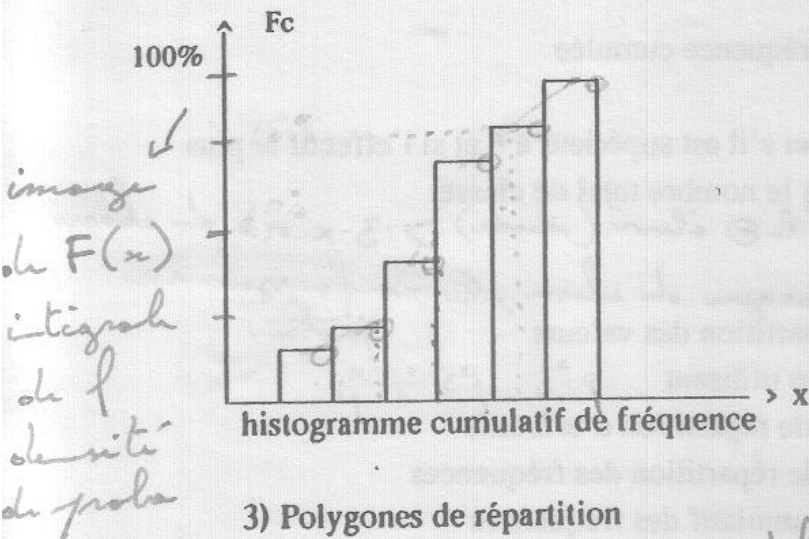


$$d_1 = \frac{n_1}{x_2 - x_1} \quad \text{ou} \quad d_1 = \frac{n_1}{n(x_2 - x_1)}$$

si n grand \rightarrow usage de la loi de probabilité discrète

histogramme de densité de distribution d'effectifs
ou
de fréquence

Pour les histogrammes cumulatifs on portera directement en ordonnée les fréquences cumulées. (FC)



$EIQ = \text{écart interquartile}$
 $= x_{75\%} - x_{25\%}$
 rang normalisé
 quartile inférieur (1^{er})
 valeur individuelle dont
 rang = 25%
 = 25% des valeurs sont inférieures à x

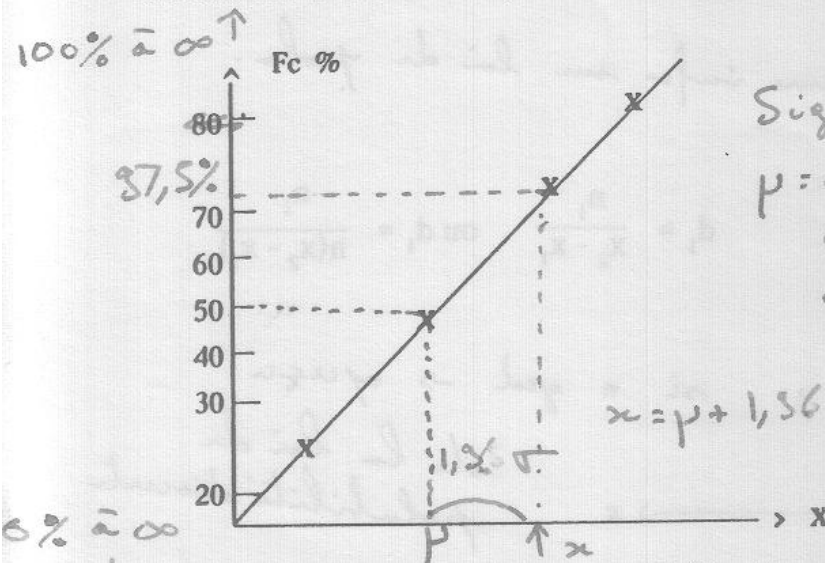
3) Polygones de répartition

2^e quartile = 50% (médiane) 3^e q = 75%

C'est une présentation particulière des histogrammes obtenus en traçant des lignes polygonales comme on les a figurées en pointillé sur les histogrammes

4) Représentation de la distribution sur une échelle probit

Une échelle probit est une échelle d'anamorphose pour les fréquences cumulées qui donne une représentation linéaire d'un histogramme cumulatif dans le cas d'une distribution suivant une loi de Gauss (ou loi normale).



Sigmoïde de $F \rightarrow$ droite
 μ : moyenne loi normale
 qui ressemble à la distribution cumulée.

$x = \mu + 1,56 \sigma$

$\sigma = \frac{x - \mu}{1,56}$

exemple pour $F(x) = 57,5\%$

On utilise ces échelles pour vérifier la normalité d'un échantillon.

On construit le polygone de fréquences cumulées en portant directement sur l'échelle probit les fréquences cumulées.

Si la ligne polygonale est voisine d'une droite c'est que la distribution observée est voisine d'une distribution normale.

Dans certains cas (toxicologie, pharmacologie) l'échelle des x pourra être logarithmique on parlera alors d'une représentation log-probit.

IV - FLUCTUATION D'ECHANTILLONNAGE

Afin de pouvoir sur l'épreuve construire des estimations ou des tests il est nécessaire d'établir des hypothèses à priori plausibles. Ces hypothèses concernent la loi de probabilité de la population $f(x)$. Par exemple :

- On peut en utilisant le théorème central limite (causes variées, nombreuses, indépendantes) faire l'hypothèse d'une population normale dont les paramètres sont inconnus.
- On peut en utilisant des arguments physiques faire l'hypothèse d'une distribution uniforme.

Ces hypothèses peuvent également s'appuyer sur des informations recueillies antérieurement et permettre de déterminer ainsi a priori la moyenne ou la variance de la population.

Ces hypothèses concernent aussi le caractère aléatoire et indépendant des variables X_1, X_2, \dots, X_n de l'échantillon

Ces hypothèses étant admises on va utiliser les modèles statistiques pour calculer les lois de probabilités de variables aléatoires construites sur les variables de l'échantillon.

On étudiera principalement les fluctuations d'échantillonnage des variables aléatoires suivantes :

Moyenne algébrique d'un échantillon $M = \frac{1}{n} (x_1 + x_2 + \dots + x_n)$

Variance d'échantillonnage $\hat{S}^2 = \frac{1}{n-1} \sum_i (x_i - M)^2$

Exemple des comprimés : on a tiré un échantillon de 20 comprimés, 2 sont défectueux. Pour la population la loi de distribution est inconnue.

$$P(x=1) = p \quad P(x=0) = 1 - p$$

Si l'on a tiré un échantillon par EASI de 20 individus (hypothèse plausible) Alors la fluctuation d'échantillonnage de $M = \frac{1}{n} (x_1 + x_2 + \dots + x_{20})$ suit une loi binomiale (cf Ch. III).

d'espérance $E(M) = p$; $\text{Var}(M) = \frac{pq}{n}$ et l'on a sa disposition une seule valeur de M obtenue sur un seul échantillon. Avec cette information et les hypothèses de fluctuation d'échantillonnage de M peut-on en déduire la valeur de p ? La réponse à cette question fait l'objet du chapitre suivant.

Exemple du médicament :

- On tire un échantillon de 20 valeurs de pression artérielle
- Compte-tenu des précautions prises on fait l'hypothèse a priori d'un échantillon (EASI)
- Des études antérieures sur de très grands échantillons de rats ont montré que la variable aléatoire (pression artérielle) est distribuée à peu près normalement.
- Il est très plausible que la population parente ∞ de rats traités par la drogue suive aussi une loi normale.
- Par contre on ignore tout à fait l'espérance $E(X)$ et $\text{Var}(X)$. Ces hypothèses a priori étant faites; le calcul des probabilités nous montre que M suit une loi normale: et que $\frac{M - \mu}{\hat{S}/\sqrt{n}} = T$ suit une loi de t de Student.

Ce qui donne la loi de fluctuation d'échantillonnage de M et \hat{S}

A partir de la valeur de m et de s obtenue sur un échantillon peut-on connaissant la loi de fluctuation de M et \hat{S} en déduire la valeur de μ et s de la population et la comparer aux valeurs μ_0 et s_0 connus d'une population d'animaux non traités ?

Cette double déduction comme dans l'exemple précédent fera l'objet des chapitres suivants.

CHAPITRE II

L'ESTIMATION

I - GENERALITES - POSITION DU PROBLEME

Soit une population de valeurs, caractérisée par une variable aléatoire X et sa loi de distribution $f(x)$.

Soit un échantillon de cette variable obtenu par EASI, peut-on si $f(x)$ est inconnu estimer les paramètres caractéristiques de cette loi de distribution?

Cette situation appelle plusieurs commentaires.

a) Elle est très fréquemment rencontrée en pratique

Exemple des comprimés : sur un lot de fabrication de plusieurs milliers de comprimés, on a prélevé un échantillon par un plan EASI de 100 comprimés, chaque comprimé de l'échantillon ayant été pesé peut-on estimer l'espérance mathématique du poids des comprimés dans le lot de fabrication?

Répondre à cette question est d'un intérêt évident pour le responsable du contrôle. On peut multiplier ce genre d'exemple.

b) Choix des paramètres à estimer.

Dans la pratique on cherche à estimer essentiellement la variance et l'espérance mathématique de la loi de distribution caractéristique de la population.

c) Choix de l'estimateur

Ayant choisi le paramètre de la population à estimer, il convient alors de rechercher la grandeur mesurée sur l'échantillon, susceptible de donner de façon correcte et efficace une estimation du paramètre cherché.

Pour estimer l'espérance mathématique d'une population faut il mieux prendre, la moyenne algébrique des valeurs de l'échantillon, la moyenne algébrique de la plus grande et de la plus petite valeur obtenue sur l'échantillon, la somme des valeurs divisée par $n-1$ ou par $n+1$ etc...?

La grandeur choisie est appelée estimation, cette opération s'appelle estimation ponctuelle, elle nécessite de définir clairement ce qu'on entend par estimation correcte et efficace.

d) Qualité de l'estimation

On peut aussi dans certains cas, déterminer la précision de l'estimation en délimitant un intervalle dans le quel peut se trouver la valeur exacte du paramètre estimé et en précisant pour cet intervalle, le risque pour que la valeur vraie se trouve à l'extérieur. Cette opération plus complète que la précédente s'appelle estimation par intervalle de confiance.

II - L'ESTIMATION PONCTUELLE

II.1 - Critères de l'estimation

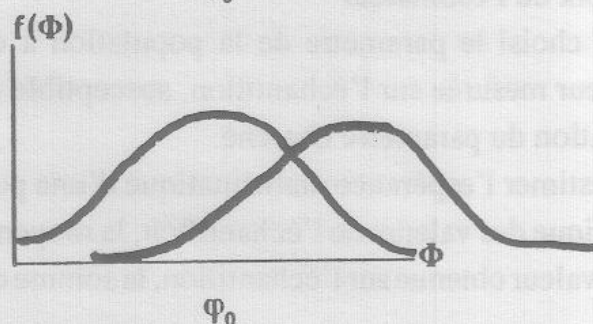
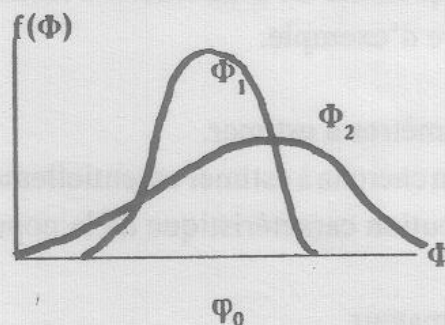
meilleure info possible.

A partir d'un échantillon l'estimation fournit une valeur unique φ du paramètre inconnu à estimer. φ_0 qui en général ne coïncide pas avec elle. Toutefois cette valeur doit être la plus proche possible compte tenu de l'information dont on dispose sur l'échantillon. Pour définir l'estimation optimale on va considérer que la quantité choisie sur l'échantillon est elle même une variable aléatoire Φ fonction à priori des variables aléatoires $(x_1, x_2 \dots x_n)$ de l'échantillon.

Pour être un «bon» estimateur cette variable aléatoire Φ doit avoir des caractéristiques suivantes :

Biais : défini par l'espérance $E(\Phi) - \varphi_0$, qui doit être minimum.
Si $E(\Phi) = \varphi_0$ on dit que l'estimateur est sans biais

Dispersion ou efficacité = définie par $E(\Phi - \varphi_0)^2$ qui doit être minimum



Consistance = l'estimateur doit être consistant c'est à dire que son efficacité et son biais, doivent tendre vers zéro quand la taille de l'échantillon tend vers l'infini.

II.2 - Erreur quadratique moyenne d'un estimateur

On peut apprécier de manière combinée l'efficacité et le biais d'un estimateur en définissant l'erreur quadratique moyenne

$$\text{Erreur} = \underbrace{(E(\Phi) - \varphi_0)^2}_{\text{biais}} + \underbrace{E(\Phi - E(\Phi))^2}_{\text{variance de l'estimateur}}$$

plus l'erreur quadratique est faible et plus l'estimateur est efficace

II.3 - Estimation de l'espérance mathématique

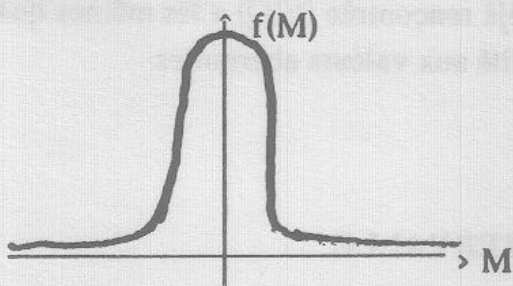
Parmi tous les estimateurs possibles de l'espérance mathématique $E(X)$ d'une variable X sur un échantillon (EASI) l'estimateur optimal est :

$\hat{M} = M = \frac{1}{n} \sum_i X_i$ où $(X_1, X_2 \dots X_n)$ représentent les n variables aléatoires indépendantes échantillonnées sur X

Ce résultat intuitivement raisonnable se démontre.

Remarque 1 : on a démontré que $\text{Var}(M) = \frac{\sigma^2}{n}$ où $\sigma^2 = \text{Var}(X)$ quelle que soit la loi de distribution $f(x)$

Remarque 2 : quand la distribution de la moyenne présente des extrémités très allongées les échantillons peuvent alors présenter facilement des valeurs aberrantes qui viennent influencer de manière gênante l'estimation ponctuelle de la moyenne.



On peut alors utiliser une moyenne corrigée des valeurs aberrantes appelée moyenne pondérée facile à programmer sur ordinateur.

$$M \text{ corrigée} = \frac{\sum xW(x)}{\sum W(x)}$$

avec (w) coefficient de pondération qui varie avec x selon les formules

$$W(x) = 1 - Z^2 \quad \text{si } Z < 1$$

$$W(x) = 0 \quad \text{si } Z > 1$$

avec

$$Z = X - \frac{M \text{ corrigé précédent}}{3EIQ}$$

Avec EIQ = étendue interquantile entre le 3^e et le 1^{er} quantile. Cette valeur représente dans l'histogramme cumulatif des fréquences pour l'échantillon la différence entre la valeur de la VA correspondant à la fréquence cumulée 75% et celle correspondant à 25%.

La formule de calcul est itérative c'est à dire que le calcul de M corrigée fait intervenir la valeur de M corrigée du calcul précédent.

Pour le premier calcul de M corrigée on prend M observé comme valeur du calcul précédent. La valeur de M corrigée est rapidement convergente ; lorsqu'elle atteint une valeur suffisamment précise on arrête le calcul.

II.4 - Estimation de la variance

Parmi tous les estimateurs possibles de la variance $\text{Var}(X)$ d'une loi de distribution $f(x)$, obtenues sur un échantillon (EASI) l'estimation optimale est la variance d'échantillonnage.

$$\hat{S}^2 = \frac{1}{n-1} \sum_{i=1}^{i=n} (X_i - M)^2$$

Ce résultat se démontre

Remarque 1 : $\hat{S}^2 = \frac{1}{n} \sum_{i=1}^{i=n} (X_i - M)$ est un estimateur biaisé et n'est pas l'estimateur optimal, ce résultat se démontre aussi.

Remarque 2 : Il existe pour estimer la dispersion d'une distribution d'autres estimateurs utilisables. Par exemple l'étendue interquantile déjà rencontrée (EIQ) a les mêmes qualités que la moyenne b pondérée quant à son insensibilité aux valeurs aberrantes.

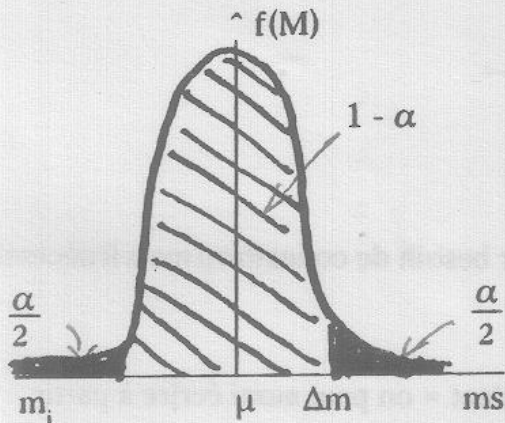
III - PRINCIPE DE L'ESTIMATION PAR INTERVALLE

III.1 - Exemple de la moyenne d'un échantillon de grande taille

Comme on l'a déjà montré la moyenne M d'un tel échantillon est une V.A. qui suit une loi normale de moyenne M et d'écart type σ^2 Dans un premier temps pour comprendre le

principe de l'intervalle de confiance supposons que M et $\frac{\sigma}{\sqrt{n}}$ sont connus.

Il est alors facile pour une probabilité $1-\alpha$ donnée de trouver l'intervalle symétrique (m_i, m_s) tel que =



$$P(m_i < M < m_s) = 1 - \alpha$$

et représenté par la surface hachurée

Cela signifie qu'une proportion $1-\alpha$ des valeurs de μ observées se situe dans cet intervalle $\mu \pm \Delta m$. D'autre part si à chaque valeur m_{obs} on construit l'intervalle $m_{\text{obs}} \pm \Delta m$ l'intervalle contient la valeur de μ si m observé appartient à l'intervalle $\mu \pm \Delta m$

On une valeur observée de m a une probabilité $1-\alpha$ appartenir à l'intervalle $\mu \pm \Delta m$

Ainsi l'intervalle $m_{\text{obs}} \pm \Delta m$ construit sur une telle valeur de m_{obs} a la probabilité $1-\alpha$ de contenir la valeur et la probabilité α de ne pas le contenir.

Ce que l'on peut aussi exprimer par

$$\mu = m_{\text{obs}} \pm \Delta m \text{ avec un risque de } \alpha \text{ se tromper}$$

Ici le risque X de se tromper signifie simplement que l'on est dans le cas où l'intervalle ne contient pas μ .

L'intervalle $m_{\text{obs}} \pm \Delta m$ s'appelle intervalle de confiance de la moyenne.

III.2 - Réalisation pratique d'un intervalle de confiance de la moyenne pour un échantillon de grande taille.

Cette réalisation pose un double problème le calcul de Δm et le choix de α

Le calcul de Δm est ici possible car l'on connaît la loi de probabilité sur M qui est une loi normale. Il suffit d'écrire comme on l'a déjà montré.

$$\begin{aligned} P(m_i < M < m_s) &= 1 - \alpha \\ = P(-u_{\frac{\alpha}{2}} < U < u_{\frac{\alpha}{2}}) \end{aligned}$$

ou U est la variable centrée réduite et où les valeurs $\pm u_{\frac{\alpha}{2}}$ sont données par la table de la loi normale pour la probabilité $1-\alpha$

$$\text{et avec } U_{\frac{\alpha}{2}} = \frac{m_s - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\Delta m}{\frac{\sigma}{\sqrt{n}}}$$

$$\text{et } \Delta m = u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

On voit donc que le calcul de Δm s'effectue sans avoir besoin de connaître μ mais il nécessite de connaître σ

Remarque : Autre façon de présenter le calcul précédent = on peut aussi écrire à partir de
 $P(u_{\frac{\alpha}{2}} < U < +u_{\frac{\alpha}{2}}) = 1 - \alpha$
 on peut écrire

$$P\left(u_{\frac{\alpha}{2}} < \frac{M - \mu}{\frac{\sigma}{\sqrt{n}}} < +u_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$P\left(m - u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < m + u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Attention ceci est une présentation commode, pas une démonstration.

III.3 - Choix de la probabilité α

Comme on vient de le voir, la probabilité s'interprète comme la probabilité de se tromper lorsqu'on donne l'intervalle de confiance c'est à dire la probabilité pour que le paramètre n'appartienne pas à l'intervalle.

Cette probabilité peut être choisie a priori et l'on peut être tenté de penser qu'il convient de la prendre aussi petite que possible.

Malheureusement on s'aperçoit que plus α est petit et plus l'intervalle est grand il suffit de faire le calcul comme nous le verrons.

Ainsi plus le risque de se tromper est faible et plus l'estimation est imprécise.

En pratique on considère :

- $\alpha = 10\%$ moyenne prudence
- $\alpha = 5\%$ bonne prudence
- $\alpha = 1\%$ très bonne prudence
- $\alpha = 1\%$ extrême prudence

Le choix du risque dépend des conséquences auxquelles peuvent conduire une erreur d'estimation.

En pratique, le risque $\alpha = 5\%$ est le plus couramment employé.

III.4 - Exemple numérique

On a trouvé sur l'échantillon de l'exemple précédent

$$n = 30 \quad m = 1,8 \quad \sigma = 0,6$$

on demande de calculer l'intervalle pour $\alpha = 10\%, 5\%, 1\%, 1\%$.

On calcule $\frac{\sigma}{\sqrt{n}} = 0,11$. Puis dans une table de la loi normale centrée réduite on cherche les valeurs de $U \left(\frac{\alpha}{2} \right)$ correspondant aux valeurs $\left(\frac{\alpha}{2} \right)$

$$\mu = 1,8 \pm 1,64 \times 0,11 = 1,80 \pm 0,18$$

$$\mu = 1,8 \pm 1,36 \times 0,11 = 1,80 \pm 0,22$$

$$\mu = 1,8 \pm 2,57 \times 0,11 = 1,80 \pm 0,28$$

$$\mu = 1,8 \pm 3,29 \times 0,11 = 1,80 \pm 0,36$$

IV - INTERVALLE DE CONFIANCE DE LA MOYENNE

IV.1 Cas où $f(M)$ est une loi normale

μ suit une loi normale $N \left(\mu, \frac{\sigma}{\sqrt{n}} \right)$ dans les cas suivants :

$f(X)$	Var $X = \sigma^2$	taille de l'échantillon	$f(M)$	Var M	nature de l'échantillon
loi normale	comme par ailleurs	petite ou grande	loi normale	$\frac{\sigma^2}{n}$ $\frac{\sigma^2}{n}$	EASI
loi quelconque	connue par $\hat{\sigma}^2$	grande	loi normale	$\frac{\sigma^2}{n}$	EASI

$$U = \frac{M - \mu}{\frac{\sigma}{\sqrt{n}}}, \text{ suit une loi normale centrée réduite.}$$

$$\frac{\sigma^2}{n}$$

faire avec u_2

$$\rightarrow u_i = u_2 = \frac{m_i - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$\rightarrow \Delta m = \mu - m_i$$

$$= -\mu + \frac{u_i \sigma}{\sqrt{n}} + \mu$$

$$= \frac{u_i \sigma}{\sqrt{n}}$$

Construction de l'intervalle (déjà vu) :

$$P(-u_{\frac{\alpha}{2}} < U < u_{\frac{\alpha}{2}}) = P\left(-u_{\frac{\alpha}{2}} < \frac{S-\sigma}{\sqrt{2n}} < u_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$\mu = m \pm u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

IV.2 Cas où f(M) suit une loi de Student

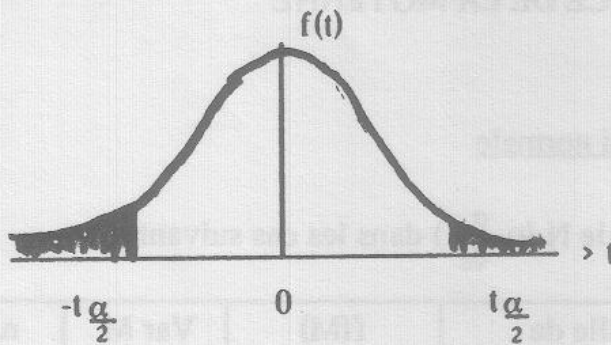
Conditions f(X)	Var (X)	nature de l'échantillon	taille de l'échantillon	fluctuation d'échantillonnage
loi normale	inconnue	EASI	quelconque grande ou petite	$\frac{(M-\mu)\sqrt{n}}{\hat{s}} = T$ suit une loi de T

Condition de l'intervalle

T suit une loi de Student avec $T = \frac{(M - \mu)}{\hat{s}/\sqrt{n}}$

Pour une risque donné

On cherche dans la table de t à $\gamma = n - 1$ ddl les valeurs symétriques $-t_{\frac{\alpha}{2}}$ et $t_{\frac{\alpha}{2}}$ correspondantes à l'évènement sur t telle que =



$$P(-t_{\frac{\alpha}{2}} < T < t_{\frac{\alpha}{2}}) = 1 - \alpha$$

$$P(-t_{\frac{\alpha}{2}} < \frac{m - \mu}{\hat{s}/\sqrt{n}} < t_{\frac{\alpha}{2}})$$

On en déduit l'intervalle de confiance pour μ

$$\mu = m \pm t_{\left(\frac{\alpha}{2}\right)} \frac{\hat{s}}{\sqrt{n}}$$

ex: $n=10$ $m=1,24$ $\alpha=0,04$ $\alpha=5\%$
 table pour $\nu=n-1=9$, $\alpha=5\%$ donc $t_{\alpha}=2,26$
 $\rightarrow \mu = 1,24 \pm 0,14$

V - INTERVALLE DE CONFIANCE DE LA VARIANCE

V.1 - Cas où l'on peut utiliser une loi de χ^2 * voir feuille

Conditions

f(X)	Var (X)	nature de l'échantillon	taille de l'échantillon	fluctuation d'échantillonnage
loi normale	inconnue	EASI	quelconque	$\frac{(n-1) \hat{s}^2}{\sigma^2} = \chi^2$ soit une loi du χ^2

avec DDL $\nu = n-1$

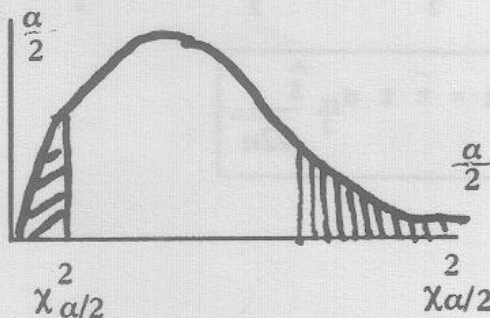
Construction de l'intervalle

Le procédé est identique aux exemples précédents mais attention ici l'intervalle n'est pas symétrique car f(χ^2) ne l'est pas

Pour 1 risque α donné on calcule

$$P(\chi_i^2 < \chi^2 < \chi_s^2) = 1-\alpha$$

$$P(\sqrt{\chi_i^2} < \sqrt{\chi^2} < \sqrt{\chi_s^2}) = 1-\alpha$$



Cette intervalle est équivalent à =

$$\chi_i^2 < \frac{(n-1)\hat{s}^2}{\sigma^2} < \chi_s^2$$

ou encore

$$1/\chi_i^2 < \frac{\sigma^2}{(n-1)\hat{s}^2} < 1/\chi_s^2$$

$\frac{(n-1)\hat{s}^2}{\chi_s^2}$	<	σ^2	<	$\frac{(n-1)\hat{s}^2}{\chi_i^2}$
-----------------------------------	---	------------	---	-----------------------------------

V.2 - Cas où l'on peut utiliser une loi normale

Condition

f(x)	E(x=)	Var (x)	taille de l'échantillon	fluctuation d'échantillonnage
quelconque	inconnue	connue estimée par $\hat{\sigma}^2$	très grand $n > 100$	$\frac{U = \hat{S} - \sigma}{\sigma_s}$ Suit une loi normale centrée réduite

$$\text{avec } \sigma_s = \frac{\sigma}{\sqrt{2n}} = \frac{\hat{\sigma}}{\sqrt{2n}}$$

Construction de l'intervalle

$$P\left(-u_{\frac{\alpha}{2}} < U < +u_{\frac{\alpha}{2}}\right) = P\left(-u_{\frac{\alpha}{2}} < \frac{S - \sigma}{\frac{\hat{\sigma}}{\sqrt{2n}}} < u_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$\alpha = \hat{s} \pm u_{\frac{\alpha}{2}} \frac{\hat{s}}{\sqrt{2n}}$$

VI - INTERVALLE DE CONFIANCE D'UNE PROPORTION OBSERVEE D'UN CARACTERE

C'est la cas où l'on a un caractère qualitatif à i modalités (A et \bar{A}) et où la variable aléatoire est une variable de Bernouilli

$$(X = 0 ; X = 1)$$

$$M = \frac{1}{n} \sum_i X_i$$

M représente ici la fréquence observée P du caractère sur un échantillon et $nM = K$ représente le nombre d'individu ayant le caractère A dans l'échantillon.

Conditions

$f(n\mu) = f(K)$	taille échantillon	$f(X)$	$Var(M)$	$E(M)$
loi binomiale	quelq.	$p_0 (q_0)$	$\frac{p_0 q_0}{n}$	P_0
loi de Poisson	grand	$p_0 \ll q_0$	$\frac{p_0 q_0}{n}$	P_0
loi normale	grand	$p_0 \approx q_0$	$\frac{p_0 q_0}{n}$	

Construction de l'intervalle à l'aide de l'une des 3 tables, on construit l'intervalle sur M pour un risque α donné.

$$P(X_{n\ inf} < nM < X_{n\ sup}) = 1 - \alpha$$

L'intervalle sur M définit l'intervalle sur P_0

$$P_{\text{observé}} + n k_{\text{inf}} < P < P_{\text{observé}} + n k_{\text{sup}}$$

VII - INTERVALLES DE CONFIANCE UNI LATERAUX

Pour un risque α donné connaissant la loi de distribution de l'estimateur $f(\Phi)$

au lieu de rechercher pour un risque donné α

$P(\varphi_1 < \Phi < \varphi_2)$ et construire l'intervalle de confiance

on peut chercher

$P(-\infty < \Phi < \varphi_1)$ et construire l'intervalle de confiance

ou $P(\bar{\Phi} > \varphi_1) \geq 1 - \alpha$

On dit alors qu'on a un intervalle unilatéral. La technique de construction est exacte-

ment la même

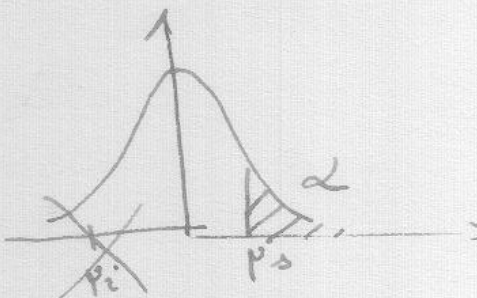
Exemple :

$n = 30$

$m = 1,8$

$\sigma = 0,6$

$\frac{\sigma}{\sqrt{n}} = 0,11$



On demande de trouver la valeur m_1 telle que $\mu < m_1$ au risque $\alpha = 5\%$

ds. tests d'hypothèse.

Il suffit d'écrire

$$P(-\infty < U < u_\alpha) = 1 - \alpha = P\left(-\infty < \frac{m - \mu}{\frac{\sigma}{\sqrt{n}}} < u_\alpha\right)$$

$$\rightarrow -\infty < \mu < m + u_\alpha$$

La table pour $\alpha = 5\%$ donne $u_\alpha = 1,64$
et non 1,36 comme dans le cas d'un intervalle unilatéral.

m a donc

$$\mu < 1,80 + 1,64 \times 0,11$$

$$\mu < 1,80 + 0,18$$



CHAPITRE III

TESTS D'HYPOTHESES

I - PROBLEMATIQUE SUR UN EXEMPLE

Un contrôleur veut vérifier sur un lot de 10^6 comprimés que leur poids moyen est égal à 2g.

Pour ce faire il prélève, par exemple, un échantillon de 30 comprimés et observe $m = 1,8g$ $\hat{s} = 0,6$ comment peut-il conclure ?

Si l'on désigne par μ la moyenne de la population réelle mais inconnue, et par $\mu_0 = 2g$ la moyenne de référence l'hypothèse que l'on veut tester est $\mu = \mu_0$ on l'appelle nous le verrons l'hypothèse H_0

Si la réponse est non et que l'on conclut $\mu \neq \mu_0$ on dit que l'on vérifie une hypothèse alternative H_1

*— peut s'interpréter à σ / σ_0
populations X X
 $\sigma_x = \sigma_y ? \mu_x = \mu_y ?$
 \hookrightarrow Hypothèse H_0 double.*

II - INTERVALLE DE CONFIANCE

Comme on l'a vu au chapitre précédent, l'intervalle de confiance, ici de μ , construit sur la valeur observée m , avec un risque α s'interprète comme l'intervalle contenant μ dans une proportion $(1-\alpha)$ d'observations.

Si ce même intervalle ne contient pas μ_0 on peut considérer que l'hypothèse $\mu = \mu_0$ n'est pas acceptable au seuil de confiance α , inversement si l'intervalle contient μ_0 l'hypothèse sera déclarée acceptable au seuil de confiance α .

De ce point de vue on peut encore interpréter ce même intervalle de confiance comme un ensemble d'hypothèses acceptables.

avec l'exemple précédent :

$$\begin{array}{l} m = 1,8 \quad s = 0,6 \quad n = 30 \quad \mu_0 = 2g \\ \text{On voit que pour } \alpha = 5\% \quad \mu = 1,80 \pm 0,18 \end{array}$$

μ_0 est l'intérieur de l'intervalle et l'hypothèse est inacceptable au seuil de confiance $\alpha = 5\%$ et que $1,58 < \mu_0 < 2,02$ est un ensemble d'hypothèse acceptables au seuil $\alpha = 5\%$ Mais on voit aussi que pour $\alpha = 10\%$ $\mu = 1,80 \pm 0,18$ μ_0 est alors à l'intérieur de l'intervalle et l'hypothèse est inacceptable au seuil de confiance $\alpha = 10\%$.
On voit sur cet exemple que plus le seuil de confiance est faible, plus l'intervalle est large et plus l'ensemble d'hypothèses acceptables est grand.

III - PROBABILITE CRITIQUE BILATERALE

On peut retourner le problème et se poser la question si l'hypothèse H_0 est vrai $\mu = \mu_0 = 2g$, dans notre précédent exemple que vaut :

$$P(\mu_0 - \Delta m_{\text{observé}} < M < \mu_0 + \Delta m_{\text{observé}})?$$

$$\text{avec } \Delta m_{\text{observé}} = |\mu_0 - m_{\text{observé}}|$$

$$\text{soit } P(1,8 < M < 2,2) = P(-1,82 < U < 1,82) = 6,9\%$$

Cette probabilité est appelée probabilité critique et définit un seuil de décision pour accepter H_0 ou H_1

En effet pour tout risque $\alpha < 6,9\%$ l'hypothèse H_0 $\mu_0 = 2g = \mu$ est acceptable et pour tout risque $\alpha > 6,9\%$ l'hypothèse H_1 , $\mu_0 = 2g \neq \mu$, est acceptable

On voit ainsi que la méthode de l'intervalle de confiance permet, pour une valeur de α fixée, de déterminer un ensemble d'hypothèses acceptables, tandis que la méthode de la probabilité critique permet, pour une hypothèse fixée, de déterminer un ensemble de probabilités α rendant l'hypothèse acceptable.

Ces deux points de vue sont en fait le reflet du même phénomène la fluctuation du résultat d'échantillonnage de m

Nous allons envisager une 3ème présentation appelée test d'hypothèse classique.

IV - TEST D'HYPOTHESE CLASSIQUE.

Il s'opère en 3 étapes :

- définition de H_0
- calcul de l'intervalle d'acceptation
- décision statistique

1) - Définition de H_0

On fixe l'hypothèse H_0 appelée hypothèse nulle car si $\mu = \mu_0 \rightarrow \mu - \mu_0 = 0$

l'hypothèse alternative H_1 pourra s'écrire soit $\mu = \mu_0$ soit $> \mu_0$ soit $\mu < \mu_0$

On se fixe aussi α seuil d'erreur du test et n taille de l'échantillon.

exemple $\mu = \mu_0 = 2g$ $\alpha = 5\%$ $H_1, \mu \neq 2g$

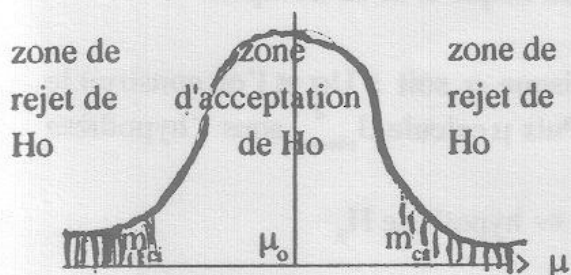
2) - Calcul de l'intervalle d'acceptation et de refus de H_0

Ayant déterminé la loi de probabilité de M , loi normale dans notre exemple, et le changement de variable qui permet d'utiliser les tables, appelé fonction discriminante,

$$\text{ici } U = \frac{m - \mu}{\sigma/\sqrt{n}}$$

Pour utiliser la table de la loi normale centrée réduite on considère provisoirement que l'hypothèse H_0 est vraie et que $\mu = \mu_0$ et l'on détermine $\pm m_c$ critique sous cette hypothèse tel que:

$$P(-m_c < M < +m_c) = 1 - \alpha$$



$$\text{soit } \pm U_{\alpha/2} = \frac{m_c - \mu_0}{\sigma/\sqrt{n}}$$

et avec notre exemple numérique

$$\alpha = 5\%, m_c = 2, \sigma/\sqrt{n} = 0,11$$

$$\pm 1,96 = \frac{m_c - 2}{0,11} \quad \begin{matrix} m_c = 2,22 \\ -m_c = 1,78 \end{matrix}$$

3) décision statistique

On considère alors le résultat observé sur l'échantillon, $m_{\text{observée}}$: 2 cas peuvent se produire.

$m_{\text{observée}}$ n'appartient pas à l'intervalle $[-mc, +mc]$ on considère que l'observation est un conflit avec H_0 et celle-ci doit être rejetée.

$m_{\text{observée}}$ appartient à l'intervalle $[mc, +mc]$ appelé intervalle d'acceptation on ne peut rejeter H_0 , H_0 est acceptable.

Dans notre exemple si $m_{\text{observé}} = 1,8$ l'hypothèse H_0 ne peut être rejetée.

V - LES ERREURS DE DECISION DANS LE TEST D'HYPOTHESE CLASSIQUE

Dans le cas où $m_{\text{observé}}$ n'appartient pas à l'intervalle d'acceptation on peut considérer deux alternatives.

- la première H_0 est vraie et l'on a pas eu de chance dans notre observation bien que le cas soit rare sa probabilité valant justement α

- la seconde H_0 est faux

Si dans la décision statistique on choisit la seconde alternative se tromper signifie qu'on aurait du choisir la première.

Autrement dit en rejetant l'hypothèse H_0 la probabilité de se tromper vaut α , probabilité d'avoir raison dans la première alternative.

Dans ce cas α est appelé risqué la première espèce.

On dit classiquement que l'on rejette l'hypothèse H_0 au risque α de se tromper.

Variante du test strictement équivalente :

On définit les bornes de la fonction discriminante au risque α soit $\pm U_{\alpha/2}$ et l'on construit le domaine d'acceptation et de refus U et non sur M - Puis μ calcule $U_{\text{observé}}^2$ sous l'hypothèse

$$H_0 ; \text{ici } U_{\text{observé}} = \frac{M_{\text{obs}} - \mu_0}{\sigma/\sqrt{n}} \quad \leftarrow \text{hypothèse } H_0$$

La décision statistique s'opère selon la valeur de $u_{\text{observé}}$ par rapport à l'intervalle $\pm U_{\alpha/2}$; $\mu_{\text{observé}}$ est à l'extérieur de l'intervalle on refuse H_0 , il est à l'intérieur on ne peut refuser...

Dans le cas où $m_{\text{observé}}$ appartient à l'intervalle d'acceptation du test M peut encore considérer deux alternatives

- la première H_0 est vraie

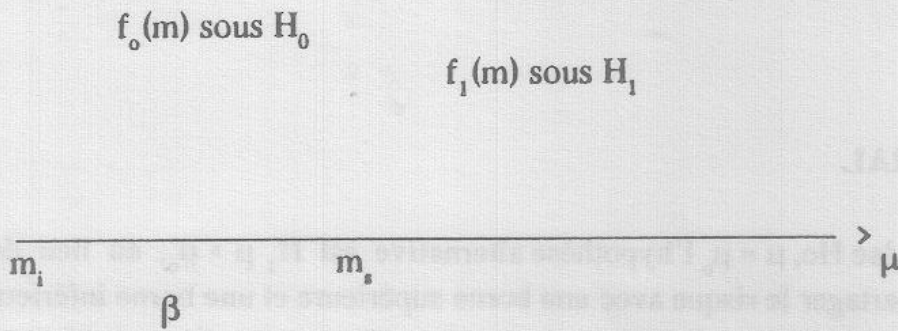
- la seconde H_0 est faux et il existe une autre hypothèse H_1 qui une probabilité

$$B = P(m_1 < M < m_2).$$

Cette hypothèse H_1 signifie que $f(m)$ a une autre loi de probabilité.

Si dans la décision statistique on choisit la première alternative se tromper signifie qu'on aurait du choisir la seconde alternative. Autrement dit en acceptant l'hypothèse H_0 la probabilité de se tromper vaut β , probabilité d'avoir raison dans l'hypothèse H_1 . β est appelé risque de 2eme espèce.

La figure suivante exprime la notion du risque β



Exemple : Dans l'exemple numérique précédent supposons que l'hypothèse alternative à

$$\mu_0 = 2g \text{ soit } \mu_1 = 2,2g$$

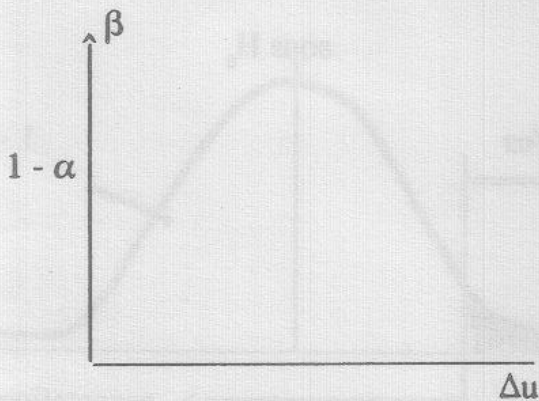
$$\beta = P(m_i < M < m_s) = (1,78 < M < 2,22)$$

$$\text{avec } u_i = \frac{(1,78 - 2,2)}{0,11} = -3,8 \quad U_s = \frac{2,22 - 2,2}{0,11} = 0,18$$

$$\beta = F(0,18) - F(-3,8) = F(0,18) - 1 + F(3,8) = 0,57 - 1 + 0,999 = 0,57$$

Dans le cas où l'hypothèse alternative H_1 ne porte que sur la moyenne alternative μ_1 les autres caractéristiques de la loi de distribution restant inchangées. On peut exprimer β en fonction de la valeur

$$\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} = \Delta u$$



$$\beta = F(u_s) - F(u_i)$$

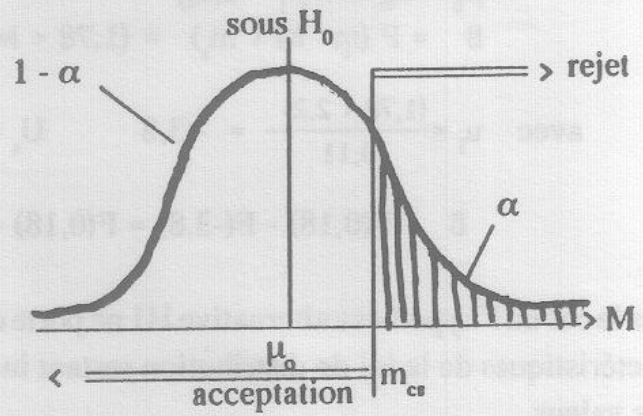
$$u_i = \frac{m_i - \mu_1}{\sigma/\sqrt{n}} = \frac{\mu_0 - u_{\frac{x}{2}} \frac{\sigma}{\sqrt{n}} - \mu_1}{\sigma/\sqrt{n}} = \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - u_{\frac{x}{2}}$$

$$Us = \Delta u + u_{\frac{x}{2}}$$

VI - TEST UNILATERAL

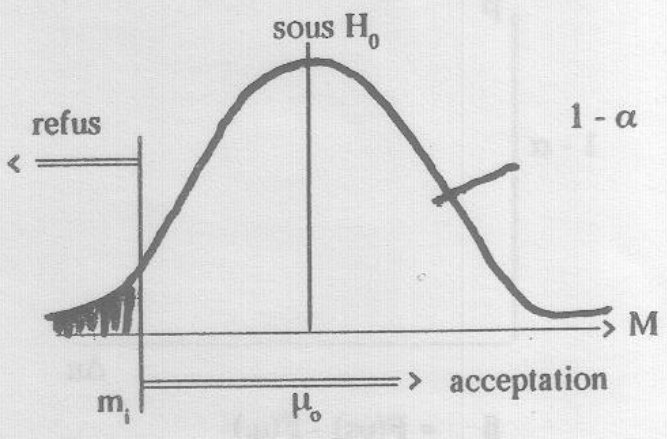
Si pour l'hypothèse $H_0, \mu = \mu_0$ l'hypothèse alternative est $H_1, \mu > \mu_0$, au lieu de $\mu = \mu_0$ on ne pourra plus partager le risque avec une borne supérieure et une borne inférieure. En pratiquant le test dans sa forme classique on calculera l'intervalle d'acceptation sans partager le risque symétriquement soit sous H_0 .

$P(-\infty < M < m_c) = 1 - \alpha$



Une hypothèse H_1 inverse $\mu_0 > \mu$ aurait donné un intervalle inversé

$P(m_i < M < +\infty) = 1 - \alpha$



VII - GENERALISATION

VII - hypothèse simple

Les test d'hypothèses que nous avons présentés pour une moyenne d'un grand échantillon suivant une loi normale peut également s'appliquer à d'autres moyennes suivant d'autres lois.

- moyenne simple suivant une loi de Student
- moyenne appariée suivant une loi de Student ou une loi normale
- pourcentage suivant une loi binomiale, ou une loi de Poisson ou une loi normale

Ils peuvent aussi s'appliquer non plus à la moyenne mais à la variance

- variance suivant une loi du χ^2
- variance suivant une loi normale

Dans tous les cas on teste un hypothèse simple H_0 : le paramètre φ de la population est-il égal à une valeur de référence φ_0 ?

VII - 2 Hypothèse double

Etant donnés deux échantillons de grande taille prélevés indépendamment dans des condition différentes mais connues ont peut considérer qu'ils proviennent de deux population X et Y de moyenne μ_x et μ_y et l'on peut se pose la question $\mu_x = \mu_y$? ou encore $\mu_x - \mu_y = 0$? qui définit une hypothèse H_0 appelée hypothèse double.

Par exemple on a prélevé un échantillon d'un premier lot produit par une machine avec un premier réglage connu et un second lot produit par cette même machine avec un second réglage connu.

On a bien deux populations distinctes et l'on peut se demander si leur moyenne est égale, hypothèse H_0

Ayant fait un test d'hypothèse avec m_x observé et m_y observée

Si l'hypothèse H_0 est refusée on peut conclure que le réglage explique la difference de moyenne .

Si l'hypothèse H_0 est acceptée on ne peut pas conclure d'un effet du réglage sur les moyennes. Cette méthode est très intéressante pour mettre en évidence des effets lorsqu'on en contrôle les causes.

Des hypothèses doubles peuvent être testées sur des moyennes mais aussi sur des variances, comme pour les hypothèses simples il faudra préciser les lois de probabilités des paramètres d'échantillonnage, moyennes et variances des échantillons.

VIII - COMPARAISON PORTANT SUR DES MOYENNES

VIII - 1 Comparaison d'une moyenne à une valeur de référence

a) Utilisation de la loi normale centré réduite

Condition de validité :

Var X = σ^2 connue antérieurement, ou estimée avec une précision suffisante pour le test, par s^2 , si $n > 30$. *n > 30*
 Taille de l'échantillon.

quelconque et notamment petite si f(x) suit une loi normale

- $n > 5$ si f(x) suit une loi symétrique
- $n > 10$ si f(x) suit une loi faiblement asymétrique
- $n > 20$ si f(x) suit une loi nettement asymétrique
- $n > 30$ si f(x) suit une loi inconnue et notamment très asymétrique

fonction discriminante

$$U = \frac{M - \mu}{\sigma/\sqrt{n}}$$
 suit une loi normale centrée réduite

b) Utilisation de la loi de Student

Condition de validité

La loi de probabilité doit être normale ou très symétrique la variance est inconnue
 L'échantillon peut avoir une taille quelconque en particulier la loi de Student est intéressante pour les petits échantillons

la fonction discriminante vaut :

$$T = \frac{M - \mu}{\sigma/\sqrt{n}} \quad T = \frac{M - \mu}{\frac{s}{\sqrt{n}}}$$
 σ^2 inconnu.
petits échantillons
f(x) : loi normale ou symétrique
 avec $vddl = n - 1$
vddl = n - 1

VIII- 2 Comparaison de 2 moyennes appariées

Exemple :

Sur un même individu on mesure la glycémie avant et après traitement par un médicament supposé hypoglycémiant. Chaque individu i donne 2 valeurs de glycémie x_i et y_i

essai à double aveugle

correspondant aux deux situations avant et après traitement, la différence $y_i - x_i$ pour chaque paire de valeurs est une variable aléatoire D.

Pour savoir si le traitement produit ou non un effet on peut faire l'hypothèse que si le traitement est sans effet la moyenne de la différence est nulle.

Ainsi pour résoudre le problème on considère la variable aléatoire D a une moyenne $\mu = \mu_0 = 0$ ceci constitue l'hypothèse $H_0 \rightarrow$ pas d'effets $\mu_x = \mu_y$

Pour tester l'hypothèse sur l'échantillon de taille $2n$, on calcule les différences appariées D la moyenne algébrique m_d observée connaissant la loi de distribution de cette moyenne Md

On peut alors réaliser un test classique.

Conditions d'applications.

Elles sont assez peu rigoureuses car si X et Y sont assez symétriques si n est grand μ_d suit une loi normale.

Dans ces conditions le plus simple est d'utiliser la fonction discriminante

$$T = \frac{M_d - \mu_0}{\hat{\sigma}_d / \sqrt{n}}$$

qui suit une loi du Student

$\mu_0 = E(D)$

$\hat{\sigma}_d$ variance d'échantil. de D

D variable aléatoire (rigueur)

μ_0 " " "

avec $n =$ nombre de paires et $nddl = n - 1$

Remarque on pourrait être tenter de tester une hypothèse double $\mu_x - \mu_y = 0$ qui ne tienne pas compte de l'appariement des valeurs de x. Cette méthode aurait au moins deux inconvénients Les deux échantillons n'étant pas indépendants il est difficile de trouver une fonction discriminante rigoureuse, de plus on aurait une fluctuation d'échantillonnage plus forte qu'avec l'utilisation des différences. Ceci, éliminant une partie des facteurs individuels permet, plus facilement de faire apparaître des différences significatives.

VIII-3 Comparaison de 2 moyennes non appariées

a) Utilisation de la loi normale

Les conditions de validités sont les même que celles pour une moyenne mais elle doivent s'appliquer aux deux populations.

Var X et Var Y connues antérieurement ou estimée avec une précision suffisante pour le test par s_x et s_y si $n_x > 30$ et $n_y > 30$ les échantillons doivent être grands ou alors X et Y doivent suivre une loi normale ou très symétrique

fonction discriminante :

$$U = \frac{(M_x - M_y) - (\mu_x - \mu_y)}{\sigma_c}$$

$$U = \frac{(\pi_x - \pi_y) - (\mu_x - \mu_y)}{\sigma_c}$$

variance de cette différence?

$\pi_x - \pi_y$: nouvelle variable aléatoire.

avec $\sigma_c = \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$ *écart type commun*

Sous l'hypothèse H_0 on fera $\mu_x - \mu_y = 0$

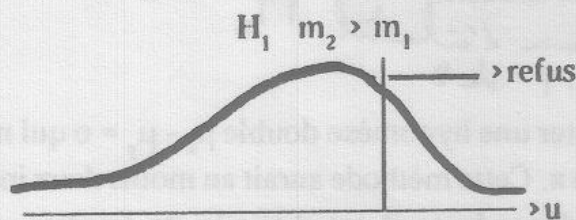
La réalisation du test ne pose aucune difficulté particulière, peut s'effectuer selon les différentes façons que nous avons déjà présentées.

Exemple	$m_1 = 1,9$	$m_2 = 2,1$
	$N_1 = 30$	$N_2 = 40$
	$S_1 = 0,4$	$S_2 = 0,5$

La moyenne m_2 est-elle plus grande que m_1 au seuil $\alpha = 5\%$?

$$H_0 \Rightarrow m_2 - m_1 = 0$$

$$u_\alpha = 1,64$$



$$u \text{ observée sous } H_0 = \frac{2,1 - 1,9 - 0}{0,108} = 1,85$$

On rejette H_0 au seuil $\alpha = 5\%$

On remarquera qu'avec la même valeur de α on ne peut refuser l'hypothèse H_0 pour $H_1 m_2 \neq m_1$ (test unilatéral)

b) Utilisation de la loi de Student

Les conditions d'utilisation sont les mêmes que celles que nous avons vues pour une seule moyenne :

Distribution de X et de Y normales ou symétriques - Variances de X et de Y inconnues, mais condition supplémentaire les deux variances doivent être égales.

La loi de Student implique donc que l'on fasse au préalable un test d'égalité sur les variances que nous envisagerons ultérieurement.

Le test de Student est surtout utilisé pour les petits échantillons.

La fonction discriminante est un peu compliquée

$$T = \frac{(M_x - M_y) - (\mu_x - \mu_y)}{\hat{S}_{com}}$$

avec $S_{\infty} = \sqrt{\frac{\hat{S}_x^2}{n_x} + \frac{\hat{S}_y^2}{n_y}}$

et $\hat{S}_y^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n_x} + \sum y_i^2 - \frac{(\sum y_i)^2}{n_y}}{n_x + n_y - 2}$

$$\sigma_y^2 = \sigma_x^2$$

X et Y indépendants

Le nombre de ddl de T est $\nu = n_x + n_y - 2$

La réalisation du test est tout à fait semblable à celle des autres tests

IX - COMPARAISON PORTANT SUR DES VARIANCES

IX-1 Comparaison d'une variance à une variance de référence

- Condition de validité = la population doit être normale
- Fonction discriminante =

$$\chi^2 = \frac{(n-1)\hat{s}^2}{\sigma^2}$$

Suit une loi du χ^2 à $\nu = n - 1$ ddl

- L'hypothèse H_0 est donnée par $\sigma^2 = \sigma_0^2$
- La réalisation du test ne pose aucune difficulté particulière.

IX-2 Comparaison de variables test de Snedecor

*de 2 variables
X et Y indépendants*

- Conditions de validité = Les deux populations doivent être normales
- Fonction discriminante =

$$F = \frac{S_x^2 / \sigma_x^2}{S_y^2 / \sigma_y^2} \rightarrow \text{inconnus}$$

2 valeurs F :

$$F = \frac{x}{y} \quad \text{ou} \quad \frac{y}{x}$$

F suit une loi de Fischer Snédecor à $\nu_x = n_x - 1$ ddl pour le numérateur et $\nu_y = n_y - 1$ pour le dénominateur L'hypothèse H_0 $\sigma_x^2 = \sigma_y^2$

L'hypothèse H_1 pose un problème car dans la fonction discriminante on pourrait tout aussi bien avoir :

$$F = \frac{\hat{S}_y^2 / \sigma_y^2}{\hat{S}_x^2 / \sigma_x^2}$$

Donc si l'on veut tester $H_1 = \sigma_x^2 = \sigma_y^2$ il faut tester les 2 valeurs F et $1/F$

Pour éviter cela on préfère pratiquer une hypothèse H_1 unilatérale équivalente à la précédente. Pour cela on prend la plus grande des 2 valeurs observés de s^2 appelons la s_1^2 et appelons σ_1^2 la variance correspondante de la population on a donc :

$$s_1^2 > s_2^2 \text{ observé et } \sigma_1^2 > \sigma_2^2 \text{ hypothèse } H_1$$

L'hypothèse H_1 sera une hypothèse unilatérale si l'on choisit systématiquement le numérateur supérieur au dénominateur.

Soit $\sigma_1^2 > \sigma_2^2$, dans le cas H_1 est équivalent à $\sigma_x^2 \neq \sigma_y^2$ si l'on teste F et $1/F$

Réalisation pratique

Pour α on détermine F_s

Pour calculer F observé sous H_0 on choisit la valeur la plus grande au numérateur et la plus petite au dénominateur

$$F_o = \frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2}$$

$$\text{avec } H_0 = \sigma_1^2 = \sigma_2^2$$

$$\hat{s}_1^2 > \hat{s}_2^2$$

$$F_o = \hat{s}_1^2 > \hat{s}_2^2$$

La décision se prend en comparant les valeurs de F_o et F_s

X - COMPARAISON PORTANT SUR DES PROPORTIONS :

Nous n'envisageons ici que la comparaison d'une proportion à une proportion de référence.

Nous avons déjà signalé que la proportion d'un caractère A dans une population était égale à la probabilité $P(A)$ qu'en désignant par X la variable de Benouilli de dénombrement

de A telle que

$$P(X=1) = P(A) = p$$

$$P(X=0) = P(\bar{A}) = 1 - p = q$$

La proposition observée P_{obs} sur un échantillon de taille N est une V.A. telle que $P_{obs} = Mx$ où Mx est la moyenne algébrique de X sur l'échantillon de taille n

Dans ces conditions

$n_{obs} = nMx = Kn$ et la fonction discriminante

est

$$P_{obs} = K_n n$$

qui suit une loi binomiale dans le cas général

avec $E(P_{obs}) = nP$; var $P npq$

une loi de Poisson si n grand p petit

avec $E(P_{obs}) = nP$; var $P npq$

une loi Normale si n grand et $p \neq q$

avec $E(P_{obs}) = nP$; var $P npq$

Réalisation pratique du test (version bilatérale)

Pour une valeur α donnée ou calcule sous l'hypothèse $H_0 \Rightarrow p = p_0$ et $H_1 \Rightarrow p \neq p_0$, les valeurs K_{inf} R_{sup} avec l'une des 3 lois précédentes selon le cas dans le quel on se trouve.

On compare alors la valeur $Np_{obs} = K_{obs}$ avec celles de l'intervalle et l'on prend la décision statistique.

A.A.E.P.L.
Association Amicale des Etudiants en Pharmacie de Lyon
8 avenue Rockefeller - 69373 LYON CEDEX 08
Tél. : 78-74-40-37