Faculté de psychologie

Statistique II
Inférence pour une et deux variables
Probabilités

Support à l'exposé oral

Titulaire
Bernadette Govaerts
ISBA, LSBA et SMCS
UCLouvain

Eléments de Probabilités

P1 : Introduction
P2 : Calcul de probabilités sur des événements
P3 : Variables aléatoires : généralités et lois classiques
P4 : Théorème central limite et combinaisons de variables aléatoires

Eléments de Probabilités

P1 : Introduction
P2 : Calcul de probabilités sur des événements
P3 : Variables aléatoires : généralités et lois classiques
P4 : Théorème central limite et combinaisons de variables aléatoires

Eléments de Probabilités
P1 : Introdution

Concepts nouveaux
- Modèle probabiliste
- Échantillon – population
- Evénement - Variable aléatoire

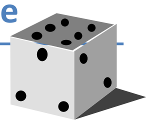
P1 : «Théorie» des probabilités : Quoi et pour quoi ?

Définition
- La «théorie des probabilités» ou «Les probabilités» désignent une branche des mathématiques qui fournit un formalisme et des outils de calcul pour analyser ou traiter des phénomènes aléatoires.
- Exemples de phénomènes aléatoires : le résultat du jet d'un dé ou le comportement d'un consommateur.

Motivation dans un cours de stat pour les psychologues
- Les probabilités sont indispensables pour faire de l'inférence statistique
- Elles fournissent les outils pour décrire des « modèles probabilistes » pour les variables observées sur les population d'intérêt et les manipuler.

Approche adoptée dans ce chapitre
- Présentation des notions indispensables pour le psychologue et pour faire de l'inférence statistique.
- Utilisation d'exemples simplistes pour garder du temps pour la suite. Plus d'exemples aux TPs.

P1 : Petite observation d'un phénomène aléatoire



Expérience : on jette un dé 12 fois

Résultats

Tableau de fréquences

Résultats descriptifs

$\bar{X} =$

$q_{0.5} =$ médiane = Etendue =

P1 : Quelques concepts de base

Phénomène ou expérience aléatoire

Phénomène ou expérience qu'on peut organiser dont le résultat ne peut pas être prévu à l'avance. Ex: jet de 1 dé

Evènement

Sous-ensemble de résultats possibles d'une expérience aléatoire répondant à une certaine caractéristique.

- Ex: A = Obtenir un nombre pair = {2, 4, 6}
- Notation habituelle : lettre du début de l'alphabet

Variable aléatoire

Représentation numérique des résultats d'une expérience aléatoire. On distingue les v.a. discrètes et continues.

- Ex: X = résultat du jet d'un dé, Y = QI ou poids d'un individu
- Notation habituelle : lettre majuscule de la fin de l'alphabet

Probabilité

Nombre entre 0 et 1, qui donne la fréquence théorique ou attendue d'un événement. On la détermine soit sur base de fréquences observées ou sur base subjective.

P1 : Modèle probabiliste pour le jet d'un dé

X: variable aléatoire résultat du jet d'un dé

Quel tableau de fréquence idéal et moyenne s'attend-on à avoir pour cette variable ?

Tableau de fréquence « attendu »
Ou Distribution de probabilité

Moyenne arithmétique attendue
Ou Moyenne μ ou Espérance E(X)

$$\mu =$$

$$= E(X) = \sum_{i=1}^{q} x_i P(X = x_i)$$

Exemple de calcul de probabilité

P(A=obtenir un nombre pair) =

P1 : Modèle probabiliste et Inférence

Modèle probabiliste

- Une distribution de probabilité est un modèle probabiliste pour décrire le comportement d'une simple variable aléatoire. Des modèles plus sophistiqués existent pour décrire «mathématiquement» le comportement de phénomènes aléatoires plus complexes.
- Les données mais aussi des connaissances subjectives sont utilisées pour construire un modèle répondant le mieux possible à la réalité.

Inférence

L'inférence a pour but de généraliser à une population des résultats observés sur une échantillon extrait de cette population.

Inférence et modèle probabiliste

- Un modèle probabiliste est utilisé pour décrire le phénomène d'intérêt pour la population. Il comporte souvent des paramètres inconnus.
- L'inférence donne des outils pour estimer des paramètres du modèle à partir des données de l'échantillon et/ou évaluer la pertinence d'hypothèses prises dans le modèle. Affaire à suivre !!!!

P1 : Exemple d'inférence pour variable continue

X : QI, Taille, poids… des étudiants de Bac 2 en psychologie de l'UCL

Population

Tous les étudiants de Bac 2

Modèle probabiliste

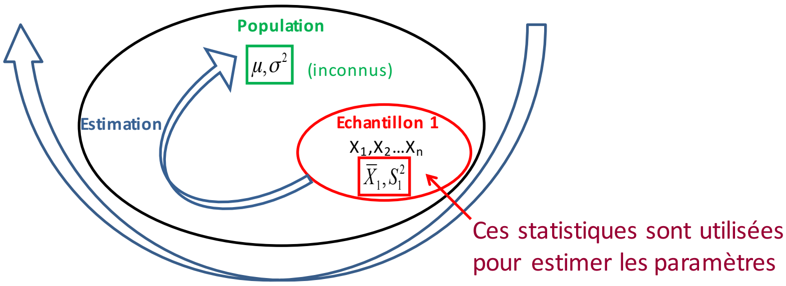
X est Normale N(μ, σ^2)

Les paramètres μ, σ^2 sont inconnus

Echantillon

Une partie des étudiants de Bac 2

On peut calculer la moyenne arithmétique et variance d'échantillon des observations.



Eléments de Probabilités

P2: Calcul de probabilités sur des événements

Concepts/formules

- Probabilité
- Evénements complémentaires
- Intersection et union d'événements
- Calcul de probabilités par dénombrement, combinaison
- Probabilité conditionnelle, indépendance
- Calcul de probabilités composées et probabilités totales
- Théorème de Bayes

P2 : Définition de base

Ω : Ensemble des résultats possibles d'une expérience aléatoire

Ex : Jet d'un dé : Ω = { }

Evènement

Sous-ensemble de Ω répondant à une certaine caractéristique.

Ex: A = Obtenir un nombre pair = {2, 4, 6} ou B = Obtenir un chiffre ≥ 5

Evénement complémentaire A^C

Ensemble des résultats possibles Ω sans les résultats appartenant à A

A^C =

Intersection ou union de 2 événements

A∩B = Intersection de A et B = ensemble des résultats répondant aux caractéristiques de A et B. A∩B =

A∪B = Union de A et B = ensemble des résultats répondant aux caractéristiques de A ou de B (le ou n'est pas exclusif !!!). A∪B =

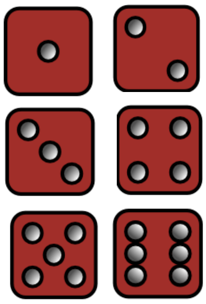
Probabilité P(A)

Nombre entre 0 et 1, noté P qui donne la fréquence théorique ou attendue d'un événement A.

P2: Exemple : jet d'un dé

A = Obtenir un nombre pair = {2, 4, 6}

B = Obtenir un chiffre ≥ 4



P(A) =

P(B) =

P(A^C) =

P(B^C) =

P(A∩B) =

P(A∪B) =

P(Ω) =

P2 : Formules de base de calcul de probabilité

P(Ω) = 1

Evénement complémentaire : P(A) = 1-P(A^c)

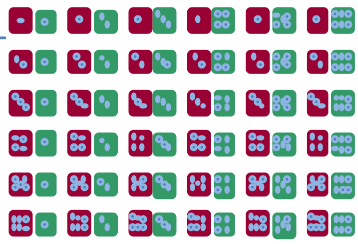
Union : P(A∪B) = P(A) + P(B) – P(A∩B)

Si les événements élémentaires de Ω sont équiprobables

$$P(A) = \frac{NCF}{NCP} = \frac{\text{Nombre de cas favorables}}{\text{Nombre de cas possibles}}$$
$$= \frac{\text{Nombre d'éléments de A}}{\text{Nombre d'éléments de }\Omega}$$

On parle de calcul de probabilité par « dénombrement »

P2: Exemple : jet de deux dés



On jette un dè rouge et un dè vert

A1 = Obtenir un nombre pair avec le dé rouge

A2 = Obtenir un nombre pair avec le dé vert

B1 = Obtenir un résultat ≥ 4 avec le dé rouge

C1 = Obtenir un nombre impair avec le dé rouge

P(A1∩B1) =

P(A1∩C1) =

P(A1∩A2) =

P(A1 | B1) =

P(A2 | A1) =

P(A1∪A2) =

P2 : Probabilité conditionnelle. Evénements indépendants et incompatibles

Soient A et B sont deux événements tels que P(B) et P(A) ≠0,

La probabilité conditionnelle de A si B notée P(A | B) est définie par

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

Les deux événements A et B sont indépendants si et seulement si

$$P(A \text{ et } B) = P(A \cap B) = P(A).P(B)$$

$$\text{ce qui et équivalent à} \quad P(A|B) = P(A)$$

$$\text{ou} \quad P(B|A) = P(B)$$

Les deux événements A et B sont incompatibles si et seulement si

$$P(A \text{ et } B) = P(A \cap B) = 0$$

P2 : Exemple de la classe

On tire au hasard 2 élèves SANS REMISE dans une classe de 3 filles et 2 garçons.

F1 = le 1er élève tiré est une fille

F2 = le 2d élève tiré est une fille

P(F1) =

P(F2 | F1) =

$P(F1 \cap F2) =$

P(F2) =

P(une fille et un garçon dans les 2 élèves tirés)

P2 : Probabilité composée et probabilité totale

Formule de probabilité composée

Quand il est plus facile de calculer P(A | B) ou P(B | A) que la probabilité $P(A \cap B)$

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

Probabilité totale

Quand on doit calculer P(A) et qu'on connaît P(A | B) et $P(A | B^C)$

$$P(A) = P(A|B)P(B) + P(A|B^C)P(B^C)$$

P2: Formule de combinaison:

La formule de combinaison est indispensable en calcul de probabilité par combinatoire avec la formule NCF/NCP. Elle permet de calculer le nombre de façons de tirer m objets dans n sans remise et sans tenir compte de l'ordre. Elle est donnée par :

$$C_n^m = \frac{n!}{m!(n-m)!} \quad \text{avec} \quad n! = n.(n-1).(n-2)...1$$

Exemples

$5! =$ $\qquad$ $1! =$ $\qquad$ $0! =$

$C_5^2 =$ $\qquad$ $C_5^5 =$ $\qquad$ $C_5^3 =$

On tire 3 élèves dans une classe qui contient 3 filles et 2 garçons. Calculer:
P(tirer 2 filles et 1 garçon)=

P(tirer au moins 1 un garçon)=

P2 : La formule de Bayes

La formule de Bayes est une formule fondamentale en théorie des probabilités qui permet de réviser la valeur de la probabilité d'une événement (B par ex) quand on a plus d'informations lié à cet événement (A).

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)} = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid B^C)P(B^C)}$$

- Ce résultat est très utilisé dans le domaine du diagnostic médical (ou autre domaine décisionnel). B = Un patient a une maladie, A = un test de dépistage a un résultat positif. Voir exemples aux TPs
- Le théorème de Bayes est la base d'une branche des statistiques: la statistique bayésienne.

P2 : Application de la formule de Bayes (Howell p124)

- Problème de Tversky et Kahneman (1980) dans leur travaux sur la prise de décision humaine
- Problème : Décider quelle société de taxi est impliquée dans un accident de la route (de nuit). Information disponible :
  - Il y a 2 sociétés de taxi en ville : 85% de taxis verts et 15% de taxis bleus
  - Un témoin déclare avoir vu un taxi bleu
  - On sait qu'une personne a une probabilité de 80% de bien identifier une couleur la nuit.
- Question précise : quelle est la probabilité que le taxi responsable soit bleu tenant compte de l'information du témoin ?
- Etapes de résolution
  - Estimez une valeur intuitive pour cette probabilité.
  - Définissez la question en terme d'événements et donnez les probabilités connues associées.
  - Calculez la probabilité demandée et comparez là à votre valeur intuitive.

P2 : Matériel et activités pour travailler le chapitre P2

Indispensable

- Transparents du cours
- Formulaire page 13
- Syllabus de TPs partie « TP1 » et TP associé
- QCM d'exercice et QCM méthodologique 1

En plus si vous en voulez plus ou ne vous en sortez pas

- Exercices supplémentaires en fin de syllabus de TP
- Syllabus de probabilités pp 3 à 16 sur I-campus : dans « Documents et lien - Autres documents utiles »
- Livre de Howell. Chapitre 5 sections 5.1, 5.2, 5.7

Eléments de Probabilités

P3 : Variables aléatoires :
Généralités et lois classiques

Concepts/formules

- Variable aléatoire
- Distribution et densité de probabilité, moyenne et variance
- Lois uniforme discrète, Bernouilli, binomiale,
- Lois uniforme continue, normale réduite et normale générale
- Formule de combinaison

P3: Variables aléatoires discrètes et continues

Variable aléatoire (ou v.a.)

Représentation numérique des résultats d'une expérience aléatoire

Variable aléatoire discrète

•Prend un nombre fini ou dénombrable de valeurs

•Définie par la liste des valeurs possibles et les probabilités associées

Exemples : X = résultat du jet d'un dé,
Y = nombre de questions réussies dans un QCM

Variable aléatoire continue

•Prend un nombre infini non dénombrable de valeurs

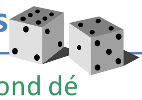
•Définie par un domaine et une fonction de densité qui permet le calcul de probabilités

Exemples : X = poids d'un enfant à la naissance
Y = taille d'un étudiant

Caractéristiques typiques d'une v.a.

On la résume souvent par 2 paramètres : sa moyenne et sa variance

P3: Exemple de v.a. discrète: lancé de deux dés



X1 = Résultat du premier dé X2 = Résultat du second dé

Y = Somme des résultats des deux dés

Distribution de probabilité de Y

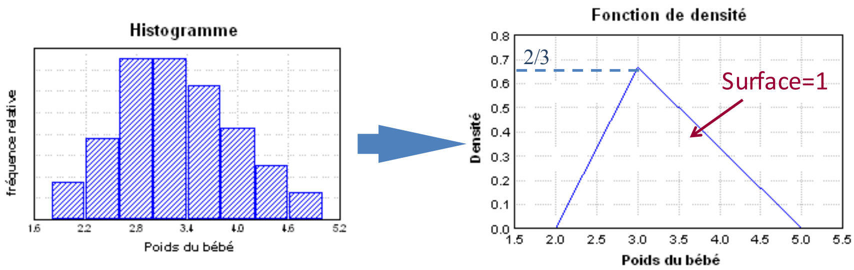
Moyenne de Y

$$\mu = \sum_{i=1}^{k} x_i P(X = x_i)$$

P3: Exemple de v.a. continue : poids d'enfants à la naissance

On observe le poids de 100 enfants à la naissance.

Représentation sous la forme d'un histogramme :



Densité f(X) = "modèle" pour l'histogramme idéal

Domaine = Valeurs de X telles que f(X) > 0

Probabilité = surface sous la fonction de densité

P(X 3) = P(X=3) = P(X>5) =

P(2 X 5) = P(X) = P(3 X 4) =

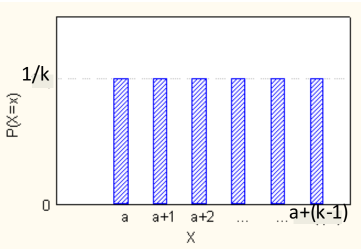
P3: Variables aléatoires : lois classiques

- Il existe une infinité de variables aléatoires possibles et donc une infinité de distributions de probabilité pour les caractériser.
- Mais il existe un petit nombre de distributions qui sont très utilisées dans un grand nombre d'applications. Ces distributions de probabilité très courantes sont définies systématiquement pour faciliter leur utilisation dans les applications.

- Lois vues dans ce cours
  - **Uniforme discrète** — Ex 1 : Résultat du jet dun dé
  - **Indicatrice / Bernoulli** — Ex 2 : Une question à choix multiples
  - **Binomiale** — Ex 3 : Questionnaire à choix multiple
  - **Uniforme continue** — Ex 4 : Attente d' un bus
  - **Normale réduite**
  - **Normale**

P3: Variable aléatoire uniforme discrète UD(a,k)

Une variable aléatoire X de distribution de probabilité uniforme discrète prend k valeurs possibles x = a, a+1, a+2... a+(k-1) avec pour probabilités associées : P(X = x) = 1/k



$$E(X) = a + \frac{k\text{-}1}{2}$$ et

$$V(X) = \frac{k^2 - 1}{12}$$

Exemple : Résultat du jet d'un dé

a = k =

E(X) = V(X) =

P3: v.a. Indicatrice: Ex : une question à choix multiple

- Expérience aléatoire : vous répondez au hasard à une question à choix multiples qui a 4 réponses possibles. La question est sur un point
- Variable aléatoire X : Vos points à la question
- Distribution de probabilité de X

- Moyenne :

$$\mu = \sum_{i=1}^{k} x_i P(X = x_i)$$

P3: Variable aléatoire indicatrice / Bernoulli Be(π)

Une variable aléatoire de distribution indicatrice ou Bernoulli X ~ Be(π) prend pour valeurs 0 ou 1 avec des probabilités associées (1-π) et π :

x	0	1
P(X=x)	(1–π)	π

$$E(X) = \pi \qquad \text{et} \qquad V(X) = \pi(1-\pi)$$

Généralisation de l'exemple de la question à choix multiples:

La question à k réponses possibles

π =

E(X) = V(X) =

P3: V.A. Binomiale : Ex 3 : Questionnaire à choix multiples

Contexte : Questionnaire à choix multiples de 5 questions et 4 réponses possibles par question. 1 point par réponse correcte

Question 1 :

A B C D

Question 2 :

A B C D

Question 3 :

A B C D

Question 4 :

A B C D

Question 5

A B C D

Vos points : /5

Mais comment calculer la probabilité de réussir en tapant au hasard ?

P3: V.A. Binomiale : Expérience et schéma de Bernoulli

Expérience aléatoire de Bernoulli : une expérience aléatoire qui peut avoir deux résultats possibles : un événement donné "échoue" ou "réussi". Le résultat sera noté 0 si l'expérience échoue et 1 si elle réussit.

Schéma de Bernoulli : expérience aléatoire qui consiste à exécuter un certain nombre de fois (n) une expérience de Bernoulli et qui répond aux conditions suivantes :

- la probabilité d'avoir un succès reste la même tout au long de l'expérience.
- les expériences ou essais sont indépendants (le résultat d'un essai n'influence pas le résultat du suivant)

Variable aléatoire binomiale : nombre d'expériences réussies dans une schéma de Bernouilli

Exemple

- Expérience de Bernoulli : répondre au hasard à UNE question d'un questionnaire à choix multiples.
- Schéma de Bernoulli : répondre au hasard à n questions

P3: V.A. Binomiale : QCM : distribution de poba

Contexte : Questionnaire à choix multiples de 5 questions et 4 réponses possibles par question. 1 point par réponse correcte

Expérience aléatoire : On répond au hasard au questionnaire

Variables aléatoires :

- Qi : Résultat à la question i
- X : résultat total au questionnaire : X = Q1+Q2+Q3+Q4+Q5

Distribution de probabilité de Qi

E(Qi) = 1/4

q	0	1
P(Q=q)	3/4	1/4

Distribution de probabilité de X

Valeurs possibles ? Probabilités associées ? E(X) ?

P3: V.A. Binomiale : QCM – distribution de proba

Distribution de probabilité de X

Moyenne de X = E(X) =

Généralisation des résultats pour n=5 questions à k=4 réponses par question

Valeurs possibles de X =

Probabilité de succès à une expérience : π =

P(X=i) = E(X) =

P3: Variable aléatoire Binomiale Bi(n,π)

Une variable aléatoire binomiale X ~ Bi(n, π) est définie comme le nombre de réussites dans n expériences de Bernoulli exécutées suivant un schéma de Bernoulli à n essais ou par la somme de n variables aléatoires de Bernoulli indépendantes.

L'ensemble des valeurs possibles de X est 0, 1, 2, … n.

Si π est la probabilité de réussite d'une expérience :

$$P(X=i)=C_n^i\pi^i(1-\pi)^{n-i}=\frac{n!}{i!.(n-i)!}\pi^i(1-\pi)^{n-i}$$

$E(X)=n\pi$ et $V(X)=n\pi(1-\pi)$

Exemple :

Questionnaire à choix multiple 5 questions, 4 réponses par question

n = π = P(X=i) =

E(X) = V(X) =

P3: V.A. Binomiale : utilisation de tables

Les tables de la binomiale facilitent le calcul de probabilités.

Les tables donnent la fonction de distribution ou de répartition $P(X \le x_i)$.

Exemple : Bi(5,1/4) : calcul de la fonction de distribution $P(X=4)=C_5^4\left(\frac{1}{4}\right)^4\left(\frac{3}{4}\right)^1$

x_i	0	1	2	3	4	5
$P(X=x_i)$	0.2373	0.3955	0.2637	0.0879	0.0146	0.001
$P(X\le x_i)$						

Table de la Bi(5,π)

p	x	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9
n=5	0	0.5905	0.3277	0.2373	0.1681	0.0778	0.0312	0.0102	0.0024	0.0010	0.0003	0.0000
	1	0.9185	0.7373	0.6328	0.5282	0.3370	0.1875	0.0870	0.0308	0.0156	0.0067	0.0005
	2	0.9914	0.9421	0.8965	0.8369	0.6826	0.5000	0.3174	0.1631	0.1035	0.0579	0.0086
	3	0.9995	0.9933	0.9844	0.9692	0.9130	0.8125	0.6630	0.4718	0.3672	0.2627	0.0815
	4	1.0000	0.9997	0.9990	0.9976	0.9898	0.9688	0.9222	0.8319	0.7627	0.6723	0.4095
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

P3: V.A. Binomiale : exercice d'utilisation de tables

Questionnaire à choix multiples : 20 questions et 5 réponses possibles par question

1.Paramètres de la v.a. binomiale associée ?

1.P(N'obtenir aucune bonne réponse)

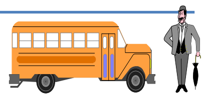
2.P(Obtenir au moins 10 réponses correctes)

3.P(Obtenir exactement 1 bonne réponse)

4.P(Obtenir de 5 à 9 bonnes réponses)

p		0.1	0.2
n=20	x		
	0	0.1216	0.0115
	1	0.3917	0.0692
	2	0.6769	0.2061
	3	0.8670	0.4114
	4	0.9568	0.6296
	5	0.9887	0.8042
	6	0.9976	0.9133
	7	0.9996	0.9679
	8	0.9999	0.9900
	9	1.0000	0.9974
	10	1.0000	0.9994
	11		0.9999
	12		1.0000
	13		1.0000
	14		
	15		
	16		
	17		
	18		
	19		
	20		

P3 : V.A. Uniforme continue : Ex : Attente d'un bus…



Une personne se rend à un arrêt de bus pour prendre un bus qui passe exactement toutes les 20 minutes.

X : Temps d'attente du bus si la personne ne connaît pas les horaires

Type de la v.a. X =

Domaine de la v.a. X =

$P(X \leq 20)$ =

$P(X \leq 5)$ =

$P(X=5)$ =

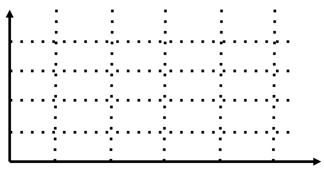
$P(X>5)$ =

Si $P(X \leq x)$ = 3/4, que vaut x ?

E(X) =

$P(X \leq x)$ =

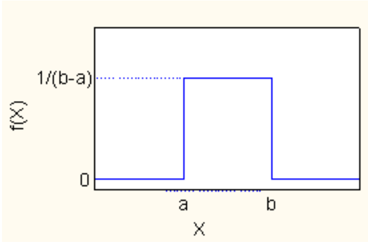
Fonction de densité de X



P3 : Variable aléatoire uniforme continue U(a,b)

Une variable aléatoire uniforme continue X ~ U(a,b) est une v.a. continue de domaine [a,b] et de fonction de densité :

$$f(X) = \begin{cases} \dfrac{1}{b-a} & \text{si } x \in [a,b] \\ 0 & \text{sinon} \end{cases}$$



$$E(X) = \frac{a+b}{2} \quad \text{et} \quad V(X) = \frac{(b-a)^2}{12}$$

Exemple : Attente du bus

a = b =

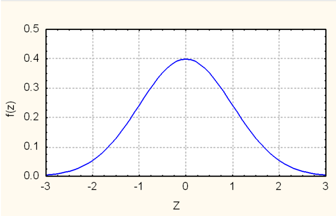
E(X) = V(X) =

P3 : Variable aléatoire normale N(μ,σ²)

- C'est la v.a. la plus utilisée en statistique car un très grand nombre de variables continues répondent à la loi Normale en conséquence du théorème central limite (voir P4).
- Sa densité a la forme de la courbe en cloche de Gauss. Un courbe symétrique de forme constante mais dont le centre et l'aplatissement dépendent des paramètres μ, la moyenne et σ, l'écart-type.
- La Normal réduite, noté Z est la normale de moyenne 0 et variance 1. Elle est utilisée pour faire des calculs sur la normale générale
- La loi Normale a été vue dans le chapitre 10 du cours de bac 1. A revoir absolument !!
- Savoir faire des calcul de probabilité sur une v.a. normale est indispensable pour
  - Situer un individu par rapport à une population d'individus selon une caractéristique donnée (QI, score à un test…)
  - Faire de l'inférence statistique

P3 : Variable aléatoire normale réduite Z ~ N(0,1)

Une variable aléatoire Normale Réduite Z ~ N(0,1) est une v.a. de domaine $\Re$, l'ensemble de réels, et de fonction de densité :



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

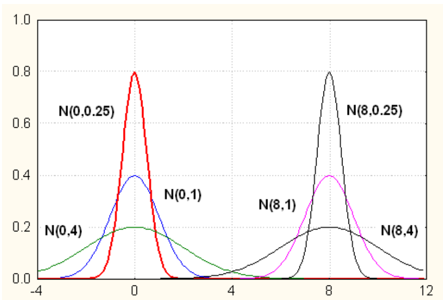
Cette fonction de densité est appelée courbe de Gauss.
La moyenne de Z est
E(Z) = μ = 0 et
sa variance V(Z) = σ^2 = 1

P3 : Variable aléatoire normale générale N(μ,σ^2)

Une variable aléatoire Normale X de paramètres μ et σ^2 est une v.a. de domaine $\Re$, l'ensemble de réels, et de fonction de densité :

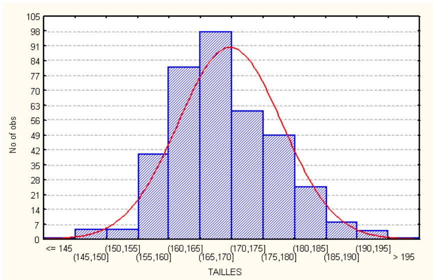
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

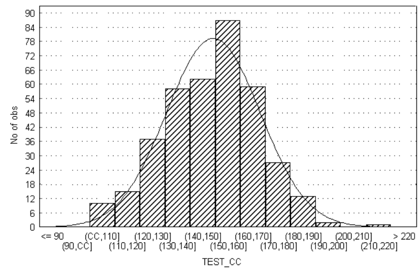
La moyenne de X est E(X) = μ et sa variance V(X) = σ^2



P3 : Exemples de v.a. réelles proches de la Normale

Enquête menée chez 400 étudiants de PSP 11 :
histogrammes pour les variables taille (en cm) et
résultat total au test de “Cognitive closure” marqué sur 250 points





P3: Calculs de probabilités sur variables normales

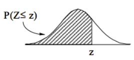
Ce qu’il faut savoir faire :

- Calculer une probabilité sur une N(0,1)
- Calculer un percentile ou quantile d’une N(0,1)
- Calculer une probabilité sur une Normale générale
- Connaître les percentiles courants de la normale réduite

Les tables disponibles

Table de la variable aléatoire Normale réduite

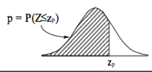
Fournit la probabilité $P(Z \leq z)$
pour $Z \sim N(0,1)$



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852

Table des quantiles de la v. a. Normale réduite

Fournit les quantiles x_p tels que $P(Z \leq z_p) = p$
pour $Z \sim N(0,1)$



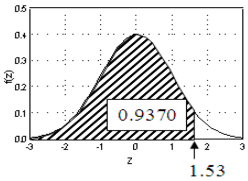
p	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00		-2.3263	-2.0537	-1.8808	-1.7507	-1.6449	-1.5548	-1.4758	-1.4051	-1.3408
0.10	-1.2816	-1.2265	-1.1750	-1.1264	-1.0803	-1.0364	-0.9945	-0.9542	-0.9154	-0.8779
0.20	-0.8416	-0.8064	-0.7722	-0.7388	-0.7063	-0.6745	-0.6433	-0.6128	-0.5828	-0.5534
0.30	-0.5244	-0.4959	-0.4677	-0.4399	-0.4125	-0.3853	-0.3585	-0.3319	-0.3055	-0.2793
0.40	-0.2533	-0.2275	-0.2019	-0.1764	-0.1510	-0.1257	-0.1004	-0.0753	-0.0502	-0.0251
0.50	0.0000	0.0251	0.0502	0.0752	0.1004	0.1257	0.1510	0.1764	0.2019	0.2275
0.60	0.2533	0.2793	0.3055	0.3319	0.3585	0.3853	0.4125	0.4399	0.4677	0.4959
0.70	0.5244	0.5534	0.5829	0.6128	0.6433	0.6745	0.7063	0.7388	0.7722	0.8064
0.80	0.8416	0.8779	0.9154	0.9542	0.9945	1.0364	1.0803	1.1264	1.1750	1.2265
0.90	1.2816	1.3408	1.4051	1.4758	1.5548	1.6449	1.7507	1.8808	2.0537	2.3263

P3 : Calcul de probabilités pour v.a. N(0,1)

Principe de calcul

- Se ramener un calcul du type $P(Z \leq z)$ avec $z>0$ en utilisant les propriétés de symétrie de la normale et le fait que la surface totale en dessous de la densité = 1.
- Rechercher le z puis le p associé dans la table de la normale réduite

Exemples



$P(Z \leq 1.53) =$

$P(Z \geq 1.53) =$

$P(Z \geq -1.53) =$

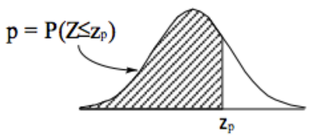
$P(Z \leq -1.53) =$

$P(-1.53 \leq Z \leq 1.53)$

P3 : Calcul de percentiles ou quantiles d'un v.a. N(0,1)

But

On donne une probabilité p et on recherche la valeur de Z notée z_p telle que $P(Z \leq z_p)=p$



Principe de calcul

- Se ramener à une expression du type $p=P(Z \leq z_p)$
- Rechercher le p et le z associé dans la table des quantiles de la normale réduite

Exemples

$P(Z \leq z_p)=0.5$ $z_p=$

$P(Z \leq z_p)=0.9$ $z_p=$

$P(Z \leq z_p)=0.1$ $z_p=$

$P(Z > z)=0.2$ $z =$

$P(-z<Z<z) = 0.8$ $z =$

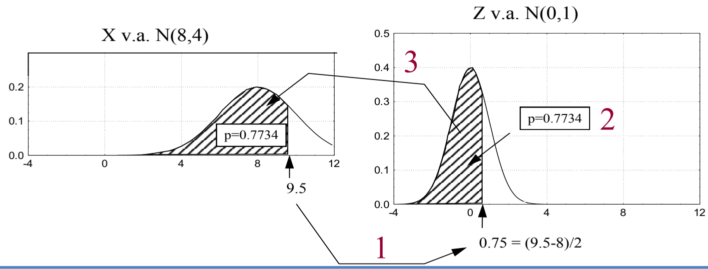
P3: Calcul de probabilité sur une variable normale générale

Le calcul d'une probabilité relative à une v.a. N(μ,σ^2) s'effectue par "réduction" à une Z ~ N(0,1) :

$$P(X \le x) = P(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}) = P(Z \le \frac{x-\mu}{\sigma})$$

Où z=(x-μ)/σ est un chiffre à rechercher dans la table de la N(0,1).

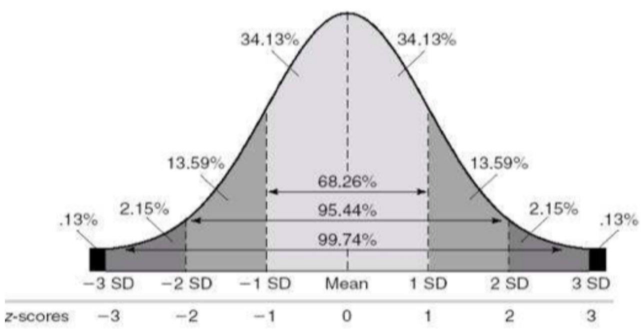
Exemple : Soit X une v.a. N(8,4), P(X≤9.5) =



P3 : Percentiles importants de la Normale

Si une v.a. X suit une N(μ,σ^2) alors

- 68.3% des individus se trouvent dans un intervalle de +/- 1σ autour de l'individu «moyen» de moyenne μ, 95.4% dans une intervalle de +/-2σ autour de μ et 99.7% dans une intervalle de +/-3σ autour de μ
- Le score z=(x-μ)/σ d'un individu permet de le situer par rapport au reste de la population.



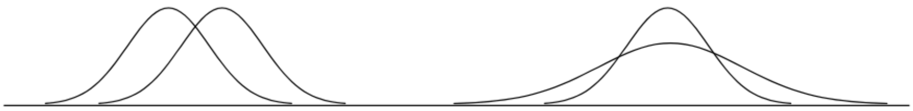
P3: Exercices de calcul sur la normale générale

Soit $X \sim N(75,100)$
- P(X<60) =
- P(60<X≤100) =
- Trouver x tel que P(X ≤ x) = 0.25

Sans faire de calcul

Soit $X \sim N(5,4)$ et $Y \sim N(6,4)$
- Quelle est la plus grande probabilité : P(X≥6) ou P(Y≥6)

Soit $X \sim N(5,4)$ et $Y \sim N(5,9)$
- Quelle est la plus grande probabilité : P(X≥6) ou P(Y≥6)



P2: Lois classiques de base : résumé

Nom	Valeurs possibles ou Domaine	P(X=x) ou densité	Moyenne E(X)	Variance V(X)
Uniforme discrète X ~ UD(a,k)	x = a, a+1, ...,a+(k-1)	1/k	a+(k-1)/2	$(q^2-1)/12$
Bernoulli X ~ Be(π)	x = 0, 1	$\pi^x(1-\pi)^{1-x}$	π	π (1-π)
Binomiale X ~ Bi(n, π)	x = 0, 1, 2, ...,n	$C_n^x \pi^x (1-\pi)^{(n-x)}$	nπ	nπ(1-π)
Uniforme Continue X ~ U(A,b)	Domaine = [a,b]	*f(x) = 1/(b-a)*	(a+b)/2	$(b-a)^2/12$
Normale Réduite Z ~ N(0,1)	Domaine = R =]-∞,+∞[	$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$	0	1
Normale X ~ N(μ,σ²)	Domaine = R = =]-∞,+∞[	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ²

P3 : Matériel et activités pour travailler le chapitre P3

Indispensable

- Transparents du cours
- Formulaire page 15
- Syllabus de TPs partie « TP2 » et séance de TP associée
- QCM de TP et méthodologique sur e-Test

En plus si vous en voulez plus ou ne vous en sortez pas

- Exercices supplémentaires en fin de syllabus de TP
- Syllabus de probabilités pp 19 à 55 sur I-campus : dans « Documents et lien - Autres documents utiles »
- Livre de Howell. Chapitre 3 et Chapitre 5, sections 5.3 à 5.6 et 5.8

Eléments de Probabilités

P4 : Combinaisons de variables aléatoires et Théorème central limite

Concepts/formules

- Moments de combinaisons linéaires de variables aléatoires
- Théorème central limite
- Approximation d'une binomiale par une normale
- Transformations linéaires de v.a. normales
- Loi chi-carré, loi de Student, loi de Fischer

P4 : Motivations

- Très souvent en statistique, on est amené à combiner les valeurs de plusieurs variables observées : faire la somme, le rapport… Si on connaît les propriétés des variables de départ (leur moyenne, variance, distribution …), il est important de connaître les propriétés de la combinaison calculée.
- Les combinaisons très souvent utilisées sont la somme ou la moyenne de variables aléatoires qui ont la même distribution.

Cette section présente les résultats liés à ces questions qui sont indispensables à comprendre quand on fait de l'inférence statistique en psychologie

P4 : Exemple : lancé de 4 dés

Expérience : on lance 4 dés à 6 faces

Variables aléatoires : X_i = résultat du $i^{ème}$ dé

S_4 = somme des résultats $X_1 + X_2 + X_3 + X_4$

M_4 = moyenne arithmétique $(X_1 + X_2 + X_3 + X_4)/4$

Propriétés de ces variables ?

Variable	Valeurs possibles	Distribution	Espérance	Variance
X_i				35/12 = 2.92
S_4		?	?	?
M_4		?	?	?

Cette section présente les étapes pour trouver les distributions de S_4 et M_4

P4 : Simulation du lancé de 4 dés

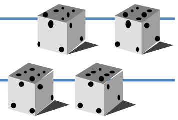
Simulation

- On lance 5000 fois 4 dés et on calcule,
- Pour chaque lancé, la moyenne arithmétique et la somme des résultats
- Puis on réalise un histogramme des 5000 sommes et moyennes pour visualiser la distribution de probabilité de ces 2 variables.

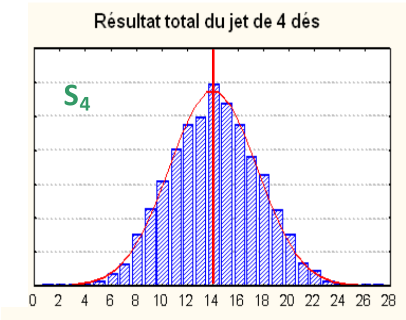
Résultats numériques

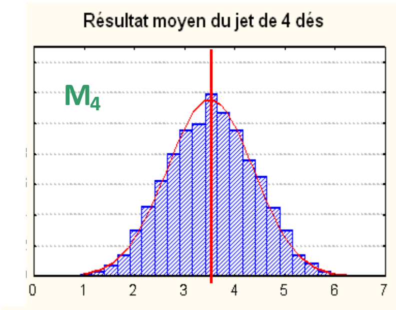
	X_1	X_2	X_3	X_4	S_4	M_4
Lancé 1	4	5	2	1	12	3.0
Lancé 2	1	6	6	3	16	4.0
....						
Lancé 5000	2	2	3	5	15	3.75

P4 : Simulation du lancé de 4 dés : résultats



Histogramme des 5000 sommes et 5000 moyennes





Distribution de S_4 ? $E(S_4)$ = $V(S_4)$=

Distribution de M_4 ? $E(M_4)$ = $V(M_4)$=

P4: Combinaison linéaire de variables aléatoires

Espérance et variance de combinaisons de variables aléatoires quelconques

- $E(a+bX) = a + b\ E(X)$
- $E(X_1+X_2 +\ldots+ X_n) = E(X_1)+E(X_2)+\ldots+E(X_n)$
- $V(a+bX) = b^2\ V(X)$
- $V(X_1+X_2 +\ldots+ X_n) = V(X_1)+V(X_2)+\ldots+V(X_n)$ si les v.a. sont indépendantes

Exemple : jet de 4 dés

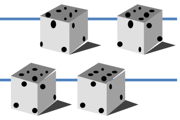
$E(X_i)$ = $E(S_4)$ =

$E(M_4)$ =

$V(X_i) = 2.92$ $V(S_4)$ =

$V(M_4)$ =

P4 : Théorème central limite : Exemple

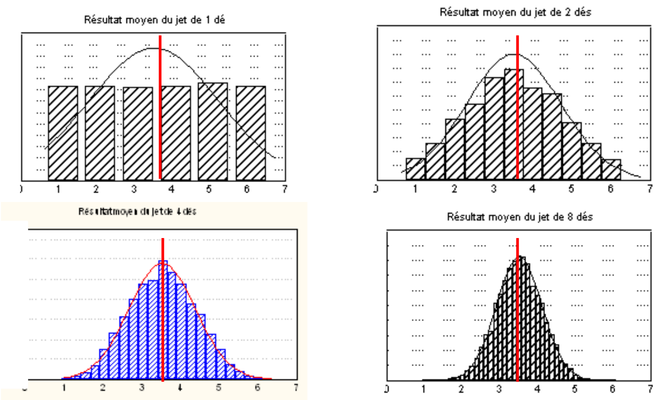


Comment évolue la **DISTRIBUTION** de M_4 quand le nombre de dés augmente ?

Distributions de la moyenne des résultats de n dés pour 1, 2, 4 et 8 dés

1 dé **M_1** **2 dés** **M_2**

4 dés **M_4** **8 dés** **M_8**



Conclusions : $E(M_i)$ $V(M_i)$

Distribution :

P4 : Théorème central limite : Enoncé

Théorème central limite :

Soient $X_1, X_2, \ldots X_n$ n variables aléatoires indépendantes et identiquement distribuées de même moyenne μ et de même variance σ^2.

alors, quand n est suffisamment grand,

•La variable aléatoire $S_n = X_1 + X_2 + \ldots + X_n$ est approximativement une v.a. Normale de moyenne $n\mu$ et de variance $n\sigma^2$

•La variable aléatoire $\bar{X} = (X_1 + X_2 + \ldots + X_N)/n$ est approximativement une v.a. Normale de moyenne μ et de variance σ^2/n

$$S_n = \sum_{i=1}^{n} X_i \overset{A}{\approx} N(n\mu, n\sigma^2) \qquad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \overset{A}{\approx} N(\mu, \frac{\sigma^2}{n})$$

Exemple

- X_i résultat du jet d'un dé i = 1,2,…n $\mu =$ $\sigma^2 =$
- $S_n \approx$
- $\bar{X} \approx$

•Si N=4, calculer la probabilité $P(S_4 \leq 18) =$

P4: TCL: Approximation d'une binomiale par une Normale

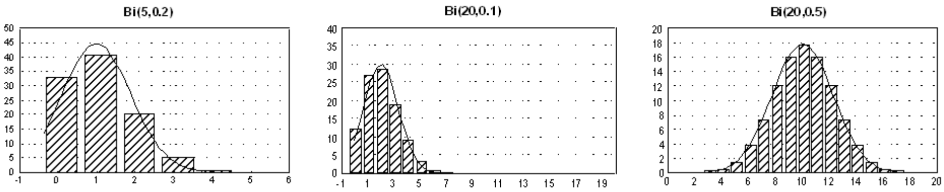
- Une variable aléatoire binomiale Bi(n,π) est la somme de n v.a. indépendantes de Bernoulli de paramètre π (de moyenne π et de variance π (1-π)). Ex: somme des résultats de n questions QCM

- Le théorème central limite peut être appliqué pour affirmer qu'une v.a. Binomiale ressemble à une Normale quand n est grand:

  Si n est suffisamment grand, une v.a. X de distribution binomiale X~ Bi(n,π) suit approximativement une distribution Normale :

  X_{Norm} de moyenne nπ et de variance nπ(1-π) : X_{Norm} ~ N(nπ,nπ(1-π))

  L'approximation est bonne dès que nπ ≥ 5 et n(1-π) ≥ 5.



nπ =　　n(1-π)=　　π=　　n(1-π)=　　π=　　n(1-π)=

P4: TCL: Approximation d'une binomiale par une Normale : calcul

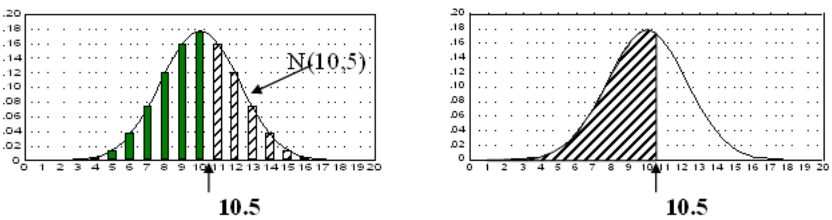
Le TCL peut s'utiliser pour faire des calculs sur des v.a. Binomiales quand les tables ne sont pas disponibles mais il faut faire une correction de continuité:

$$P(X_{Binomial} \leq x) \cong P(X_{Normal} \leq x + 0.5)$$

Soit X une Bi(20,.5). Quelle est la probabilité P(X 10) ?

- Résultats des tables : P(X≤10) =
- X est approximativement une Normal X_{Norm}~N( )
- L'approximation Normale donne donc :

P(X_{bi} ≤ 10) =



P4: TCL: : Approximation d'une Bi par une N : exercices

Soit X une v.a. Bi(100,0.2), Calculez :

a. $P(X \leq 25)$

b. $P(X<25)$

c. $P(X \geq 25)$

d. $P(X=25)$

P4 : Combinaison linéaire de variables aléatoires Normales

Combinaison linéaire de variables aléatoires Normales

Toute somme, moyenne ou combinaison linéaire de variables aléatoires normales indépendantes suit aussi exactement une distribution Normale

Soient X_1 et X_2 deux variables aléatoires indépendantes $N(\mu_1,\sigma_1^2)$ et $N(\mu_2,\sigma_2^2)$,

$$a+bX_1+cX_2 \quad \text{est} \quad N(a+b\mu_1+c\mu_2, b^2\sigma_1^2+c^2\sigma_2^2)$$

Soient $X_1, X_2, \ldots, X_n$ n variables aléatoires $N(\mu,\sigma^2)$ indépendantes

$$\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2) \qquad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \frac{\sigma^2}{n})$$

Exemple

- Un test de sélection à un concours se base sur les résultats à 2 épreuves
- X_1 : le résultat à la première épreuve qui suit une normale N(20,16)
- X_2 : le résultat à la seconde épreuve qui suit une normale N(10,9)
- Un candidat est sélectionné si son résultat total $T = X_1+2*X_2$ est plus grand que 50.
- Quelle est la probabilité qu'un candidat soit sélectionné ?

P4 : Combinaisons non linéaires de variables aléatoires Normales

En inférence statistique on suppose le plus souvent la normalité des données et des formules comme les statistiques de tests font intervenir des transformations parfois compliquées de ces normales.

Exemples : la variance $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

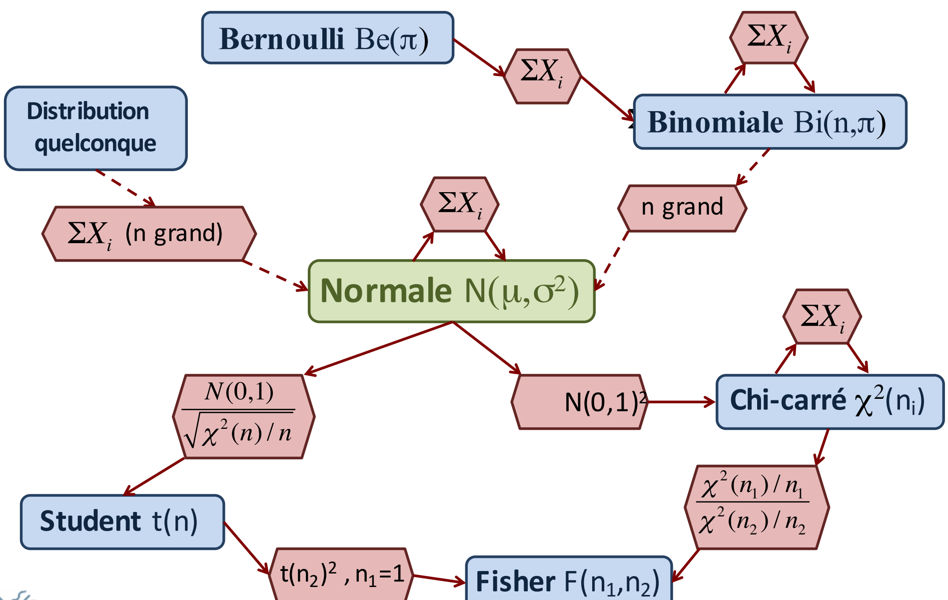
La statistique de test sur une moyenne $t_{obs} = \frac{\bar{X}-\mu_0}{s/\sqrt{n}}$

Des nouvelles lois de probabilités sont utilisées pour manipuler ces statistiques :

- La loi chi-carré quand on traite une somme de N(0,1) au carré
- La loi de student quand on divise une N(0,1) par la racine d'une khi-carré
- La loi de Fischer quand on fait le rapport entre 2 v.a. chi-carré

Des tables statistiques existent pour les quantiles de ces lois

P4: Relations entre les principales lois de probabilités



P4: Distribution chi-carrée $\chi^2(n)$

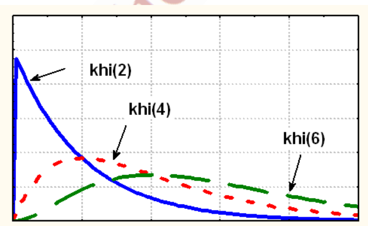
Chi-carré : v.a. positive dépendant d'un paramètre : le degré de lilberté

$X \sim \chi^2(dl)$

dl = degré de liberté

Domaine de X = R^+

Fonction de densité ---->>>



E(X)= dl V(X)= 2xdl

Tables : Les tables ne fournissent les quantiles de la variables chicarré

Exemple : $P(\chi^2(10) < x_p) = 0.95$

$x_p = \chi^2_{10;.95} =$

P4: utilité de la distribution chi-carrée

Variable chi-carré et Normale

Soit $X \sim N(\mu,\sigma^2)$ alors $Z^2 = \dfrac{(X-\mu)^2}{\sigma^2} \sim \chi^2(1)$

Le carré d'une N(0,1) est une chi-carré à 1 degré de liberté

Soient $X_1, X_2, \ldots X_n$ n v.a. $N(\mu,\sigma^2)$ indépendantes

alors $\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \dfrac{(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n)$

Variable chi-carré et variance s^2

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 = \frac{\sigma^2}{n-1}\sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{\sigma^2} \Rightarrow \frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

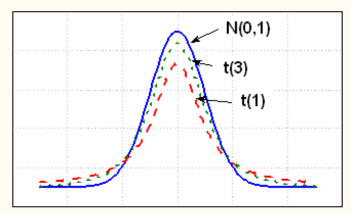
(1 degré de liberté perdu pour l'estimation de μ)

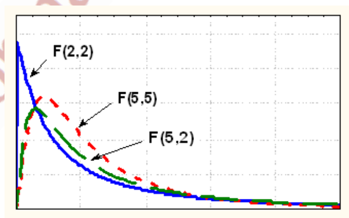
Utilisé dans le test d'hypothèse sur une variance

P4: Distributions *t* et *F*

La loi Student resemble à la Normale mais a des queues plus épaisses. Elle a un paramètre : son degré de liberté. Elle tend vers une normale quand dl augmente.

La loi de Fischer est toujours positive et dépend de 2 paramètres : le degré de liberté du numérateur et du dénominateur.





Tables : Les tables fournissent les quantile

Exemples : **P(t(10) < x_p) = 0.95** x_p =

P(F(5,10) < x_p) = 0.95 x_p =

P4: Relation entre v.a. N, t, F et χ^2

Relation N(0,1), χ^2(n) et t(n)

Soient deux v.a. indép. $Z \sim N(0,1)$ et $W \sim \chi^2(n) \Rightarrow \dfrac{Z}{\sqrt{W/n}} \sim t(n)$

La loi de t s'utilise dans construire un test t sur une moyenne

Soient $X_1, X_2 ..., X_n$ n v.a.i. $N(\mu,\sigma^2) \Rightarrow \dfrac{\overline{X}-\mu}{\sqrt{s^2/n}} \sim t(n-1)$

Relation χ^2(n) et F(n_1,n_2)

Soient deux v.a. indép. $W_1 \sim \chi^2(n_1)$ et $W_2 \sim \chi^2(n_2) \Rightarrow \dfrac{W_1/n_1}{W_2/n_2} \sim F(n_1,n_2)$

Loi de Fisher s'utilise pour construire un test pour comparaison de 2 variances

Soient $X_1,...,X_{n_1}$ n_1 v.a.i. $N(\mu_X,\sigma^2)$ et $Y_1,...,Y_{n_2}$ n_2 v.a.i. $N(\mu_Y,\sigma^2)$

$$\Rightarrow \frac{s_X^2}{s_Y^2} \sim F(n_1-1, n_2-1)$$

P4 : Matériel et activités pour travailler le chapitre P4

Indispensable

- Transparents du cours
- Formulaire partie « formules de probabilité » et tables liées
- Syllabus de TPs partie « TP2 » et TPs associés
- QCM d'exercice et QCM méthodologique 2

En plus si vous en voulez plus ou ne vous en sortez pas

- Exercices supplémentaires en fin de syllabus de TP
- Syllabus de probabilités pp 51 à 63 sur I-campus : dans « Documents et lien - Autres documents utiles »
- Livre de Howell. Chapitre 3 et Chapitre 5, sections 5.3 à 5.6 et 5.8