UMF
UNIVERSITATEA DE
MEDICINĂ ŞI FARMACIE
IULIU HAŢIEGANU
CLUJ-NAPOCA

Variables aléatoires.
Les plus importantes distributions des probabilités

Plan du cours

01 Variable aléatoire: definition

02 Types de variables aléatoires

03 Exemples de distributions théorétiques discrets: Binomiale, Poisson

04 Exemples de distributions théorétiques: Normale, de Student (t)

05 Conclusions…

L'utilité des variables aléatoires

- modélisation de la distribution théorique d'une variable d' intérêt
- extrapoler la distribution des données mesurées sur un échantillon extrait d'un population a la population entière

EXPERIENCE ALÉATOIRE, ÉVÈNEMENT ALÉATOIRE…..

- Une expérience (épreuve) peut être aléatoire ou non.
- Une expérience est aléatoire si ses résultats sont imprédictibles
- Même si on répète l'expérience dans des conditions identiques on peut observer des résultats différents.
- Exemples des **expériences**:
  - ✓ le jeux de dé
  - ✓ mesurer la température d'un individu
  - ✓ mesurer le taux du cholestérol d'un patient,
  - ✓ savoir si un patient pris au hasard a du diabète ou non
  - ✓ savoir s'il souffre d'une maladie parodontale ou non

EXPERIENCE ALÉATOIRE, ÉVÈNEMENT ALÉATOIRE…..

- Un seul **résultat d'une expérience** est un événement élémentaire.

  - exemple:

    ✓ pour le jeux de dé: la face 1, 2, …;

    ✓ pour la température: 36,5 ou 38,2;

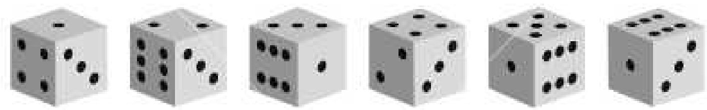
    ✓ pour la présence de la maladie parodontale: oui/non, …

- L'ensemble des toutes événements élémentaires possibles d'une expérience aléatoire représente **l'espace fondamental**.

  - ex. Pour le jeux de dé: l'ensemble {1, 2, 3, 4, 5, 6},

Exemple

Résultats d'une expérience aléatoire



Espace fondamental

On associe des nombres réels

1 2 3 4 5 6

Variable aléatoire

Distribution de probabilité

On associe des probabilités

1/6 1/6 1/6 1/6 1/6 1/6

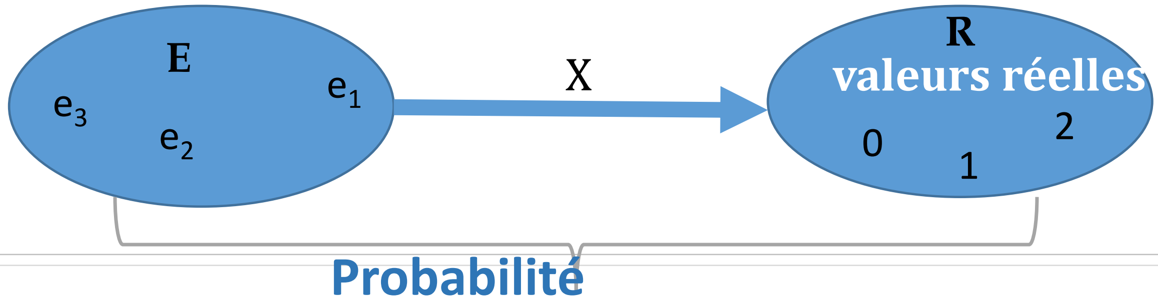
On sélection toutes les valeurs inferieures a x (ex. 3)

$Pr(X \leq 3)$

Fonction de répartition
$\Rightarrow$ $Pr(1)+Pr(2)+Pr(3)$
$=3/6=0,5$

Variable aléatoire

- Une variable est **aléatoire =** variable dont la valeur est un nombre déterminé par l'événement d'une éxperience aleatoire (**les valeurs de la variable sont mutuellement exclusives ayant une probabilité donnée**)
- Notée en lettres majuscules: X, M, N,…tandis que leurs valeurs sont notées en lettres minuscules (x, m, n…)
- **Probabilité que la variable X prenne la valeur x**: Pr(X=x)=p(x)
- **Probabilité que la variable X prenne une valeur inferieure ou égale a x**: Pr(X≤x)=F(x) (fonction de répartition)



Variable aléatoire

- Etant donnée E = espace fondamental qui correspond a un expérience aléatoire.
- **Définition:**
  - Une variable aléatoire (v.a.) dans un espace fondamental E, noté X est <u>un fonction définit sur E qui prend des valeurs dans l'ensemble des nombres réels</u>
  - Une variable aléatoire X est une variable numérique qui prend différents valeurs avec des probabilités spécifiées
- À une variable aléatoire X on peut associe des probabilités avec lesquelles cette variable aléatoire peut prendre certaines valeurs
  - Pr(X = a) → la probabilité que X va prendre la valeur *a*
  - Pr(a ≤ X ≤b) → la probabilité que X prends des valeur dans l'intervalle [a, b]

Distribution de probabilité d’une variable aleatoire

- **Distribution de probabilité**: ensemble (ou énumération) des toutes les valeurs possibles de la variable avec leurs probabilités correspondantes

ou

- **Loi de probabilité:** relation (mathématique) entre les valeurs de la variable et leur probabilités
- Chaque distribution de probabilité est définie par certains **paramètres** (moyenne, variance) qui sont des mesures synthétiques caractérisant la distribution (leur connaissance permet décrire la distribution)
- la distribution de probabilité peut être soit discrète ou continue (selon le type de la variable aléatoire)

Distribution de probabilité d'une variable aléatoire

- Distribution de probabilité d'une variable aleatoare finite X il est aussi noté par le tableau suivant:

$$\boldsymbol{X}: \begin{pmatrix} \boldsymbol{x_1} & \boldsymbol{x_2} & \dots & \boldsymbol{x_n} \\ \boldsymbol{Pr(x_1)} & \boldsymbol{Pr(x_2)} & \dots & \boldsymbol{Pr(x_n)} \end{pmatrix}$$

- Les probabilités qui apparaissent dans la distribution d'une variable aléatoire finie X vérifient la condition suivante:

$$\mathbf{Pr}(\boldsymbol{x_1}) + \boldsymbol{Pr}(\boldsymbol{x_2}) + \cdots . + \boldsymbol{Pr}(\boldsymbol{x_n}) = \mathbf{1}$$

Variables aléatoires DISCRÈTES

DEFINITION:

- La variable ne peut prendre que des valeurs entiers
- il y a un **ensemble discret de valeurs**
  - avec des probabilités spécifiées
- la variable prends un nombre finit ou au moins dénombrable des valeurs

Exemples:

- Pour la variable X: variable caractérisant le résultat du jet d'un dé
- distribution de probabilité de X:

$$\boldsymbol{X}: \begin{pmatrix} \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} & \mathbf{6} \\ \mathbf{1/6} & \mathbf{1/6} & \mathbf{1/6} & \mathbf{1/6} & \mathbf{1/6} & \mathbf{1/6} \end{pmatrix}$$

La somme de probabilités: 1/6+1/6+...+1/6=1

- Pour la variable X: apparition d'un résultat positif d'un traitement parodontal appliqué aux 4 sujets (0 sur 4, 1 sujet sur 4, 2 sujets sur 4 ...)
- distribution de probabilité de X:

$$\boldsymbol{X}: \begin{pmatrix} \mathbf{0} & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} \\ \mathbf{0,008} & \mathbf{0,076} & \mathbf{0,256} & \mathbf{0,411} & \mathbf{0,24} \end{pmatrix}$$

VARIABLES ALÉATOIRES DISCRÈTES:
esperance, variance et écart-type

- **Esperance mathématique** ou la moyenne théorique d'une v.a. X

$$\mathrm{M}(X) = x_1 \cdot Pr(x_1) + x_2 \cdot Pr(x_2) + \cdots . + x_n \cdot Pr(x_n) \sim m = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

Exemple: Soit la variable caractérisant le résultat d'un traitement

$$\boldsymbol{X}: \begin{pmatrix} \mathbf{0} & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} \\ \mathbf{0,008} & \mathbf{0,076} & \mathbf{0,256} & \mathbf{0,411} & \mathbf{0,24} \end{pmatrix}$$

E(X) = 0*0,008 + 1*0,076+2*0,265 + 3*0,411 + 4*0,24 = 1,31

- **Variance** de X:

$$V(X) = \sum_{i=1}^{n} \left[x_i - E(X)\right]^2 * \Pr(x_i) \sim \boxed{\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}}$$

- **Ecart-type** (deviation standard) de X $\sigma(X) = \sqrt{V(X)} = V(X)^{1/2}$

variables aléatoires DISCRÈTES: Loi Bernoulli

- **B (1, p)**
  - L'expérience consiste dans **1** épreuve
  - **La épreuve a seulement les résultats 1 ("succès") et 0 ("échec")**
  - La probabilité de "succès" est **p** et "d' échec" est 1 – p = **q**

    Pr(succès) = p, Pr(échec) = q = 1 -p
  - Exemple:

    – avec/sans maladie parodontale, avec/san obésité, fumeur/non fumeur, avec/sans ulcère gastrique

    – C'est un cas spécial du distribution Binomiale avec n = 1

Variables aléatoires discrètes: Loi Bernoulli

Distribution de X:

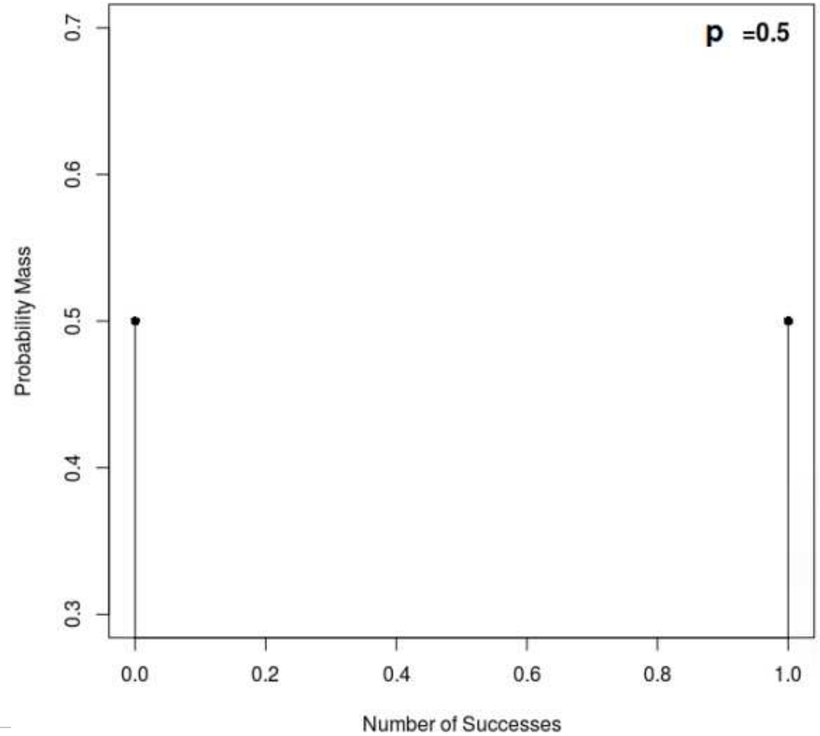
$$\boldsymbol{X}: \begin{pmatrix} \mathbf{0} & \mathbf{1} \\ \boldsymbol{q} & \boldsymbol{p} \end{pmatrix}$$

$$\Pr(X = 1) = 1 - \Pr(X = 0) = p = 1 - q$$

E(X) = p

Var (X) = p × q = p ×(1 – p)

$$\sigma(X) = \sqrt{Var(X)} = \sqrt{p \times q} = \sqrt{p \times (1 - p)}$$



Variables aléatoires DISCRÈTES: Loi Binomiale

- **Bi(n, p)**
  - L'expérience consiste dans **n épreuves indépendantes**
  - Une épreuve a seulement les résultats "succès" et "échec"
  - La probabilité de "succès" est **p** et "d' échec" est 1 – p = q
    Pr(succès) = p, Pr(échec) = q = 1 - p
  - **X = le nombre de "sucées" : {0, 1, 2, … , n}**
  - Exemple:
    - Le nombre de personnes parmi 30 - saines/ ou sans caries
- Application:
  - test binomial pour un proportion
  - Intervalle de confiance pour une proportion

Variables aléatoires discrètes: Loi Binomiale

Application:

- Intervalle de confiance pour un proportion
  - Exemple: on prend un échantillon d’une population des personnes âgées .
  - On compte combien des sujets ont des prothèses dentaires dans un échantillon.
  - On trouve le pourcentage (5%).
  - La question de recherche: Quelle est le vrai pourcentage dans la population??
  - On peut calculer un intervalle dans lequel on va trouver la vrai pourcentage avec un probabilité de 0.95 (par convention).
  - Pour trouver cet intervalle on a besoin de savoir quelle est la lois de probabilité.
  - Dans ce cas on peut observer que la variable X = **Nombre de sujets ayant des prothèse dentaire suit la loi binomiale**
  - Ici le succès est la présence de la prothèse, et l’échec est l’absence.

Variables aléatoires discrètes: Loi Binomiale

- On peut calculer la probabilité qu'un variable aléatoire
- est égale au nombre des ”succès” (k) parmi les « n » épreuves
- **Distribution de X:**

$$\boldsymbol{X}: \begin{pmatrix} \mathbf{0} & \mathbf{1} & \boldsymbol{k} & \dots . & \boldsymbol{n} \\ \boldsymbol{Pr}(\boldsymbol{X} = \mathbf{0}) & \boldsymbol{Pr}(X = 1) & \dots \boldsymbol{Pr}(\boldsymbol{X} = \boldsymbol{k}) & \dots . & \boldsymbol{Pr}(\boldsymbol{X} = \boldsymbol{n}) \end{pmatrix}$$

$$\mathbf{Pr(X = k) = C_n^k p^k q^{n-k}}$$

n – nombre des expériences, k nombre des résultats «succès»

E(X) = n × p

Var (X) = n × p × q = n × p × (1 – p)

$$C_n^k = \frac{n!}{k!\,(n-k)!}$$

$$\sigma(X) = \sqrt{Var(X)} = \sqrt{n \times p \times q} = \sqrt{n \times p \times (1 - p)}$$

ex: 3! = 3 * 2 * 1 = 6

Variables aléatoires discrètes: Loi Binomiale

Exemple:

Quelle est la probabilité d'avoir 2 garçons dans une famille avec 5 enfants.

On connait que:

- a chaque naissance la probabilité d'avoir un garçon est 0.51

- et les sexes d'enfants - des événements independantes

- ***Solution:***
- En utilisant une distribution binomiale avec le paramètres

$$\boldsymbol{X}: \begin{pmatrix} \mathbf{0} & \mathbf{1} & \dots. \quad \mathbf{5} \\ \boldsymbol{Pr}(\boldsymbol{X} = \mathbf{0}) & \boldsymbol{Pr}(X = 1) & \dots. \boldsymbol{Pr}(\boldsymbol{X} = \mathbf{5}) \end{pmatrix}$$

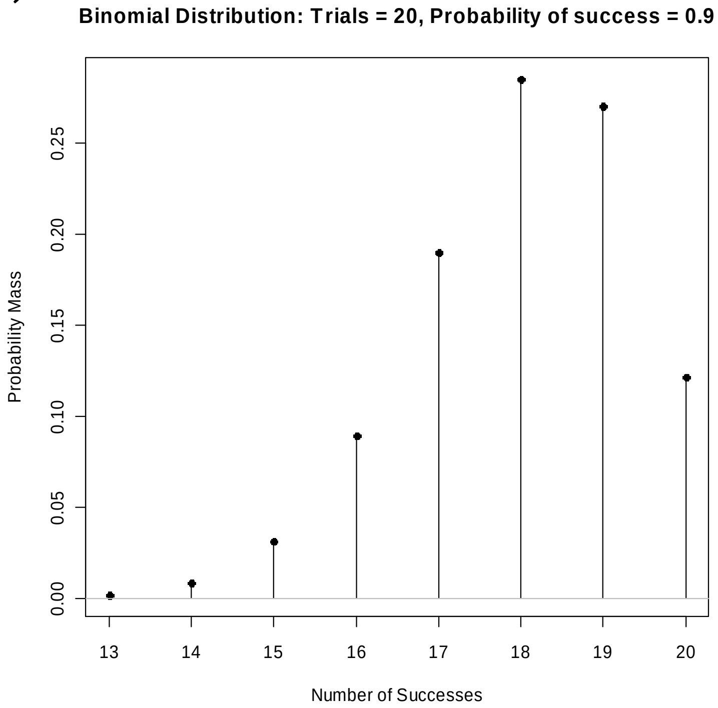
- **n = 5** et p = 0.51 pour **k = 2**, nous avons:

$$\Pr(X = 2) = C_5^2 (0.51)^2 (0.49)^3 = \frac{5!}{2! \times 3!} (0.51)^2 (0.49)^3 = 0.306$$

Variables aléatoires discrètes: Loi Binomiale

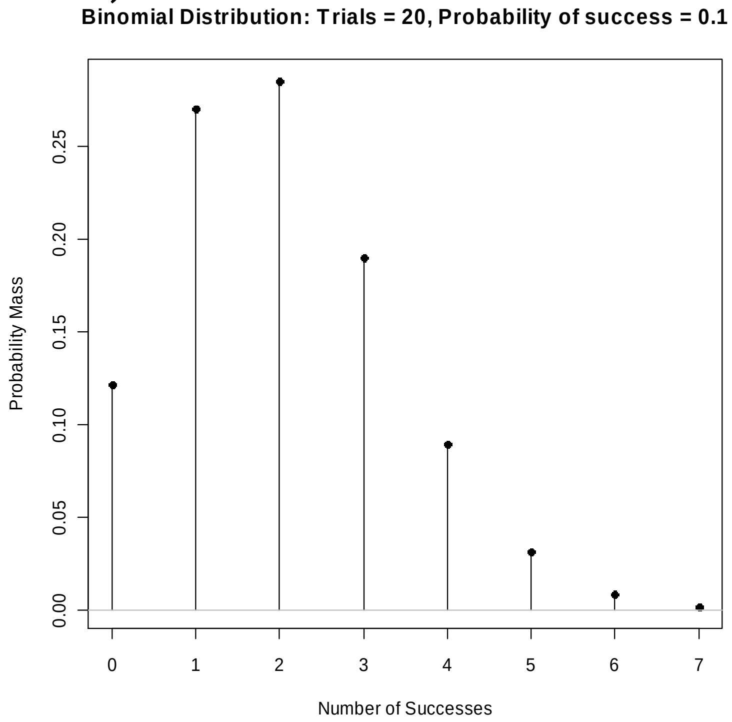
M(X) = 20 * 0,9 = 18; Var (X)= 20 * 0,9 * 0,1 = 1,8

σ = 1,34



M (X)= 20 * 0,1 = 2; Var(X) = 20 * 0,1 * 0,9 = 1,8;

σ = 1,34



Voici deux simulations de 20 expérimentes binomiales avec un probabilité de succès p = 0.90 a la gauche et p = 0.10 a la droite. La variance est identique, mais la moyenne est différente.

Variables aléatoires discrètes: Loi de Poisson

Po(λ) = la loi d'événements rares

- La variable aléatoire de Poisson:
  - variable discrète qui a un infinité dénombrable des valeurs 0, 1, …
- Représente le nombre des réalisations dans un intervalle de temps ou d'espace
  - nombre des entrées dans un cabinet/nombre interventions/nombre extractions,…
  - nombre des bactéries dans un millilitre d'eau, nombre des cellules blanc dans 1 ml de sang…
- On connais le nombre **attendu** λ d'événements dans une intervalle unité
- X = le nombre d'événements ayant lieu (**observé**) dans une intervalle unité
- Application:
  - Prédiction des phénomènes
  - Régression Poisson
    - Prédire le nombre des malades dans une région, dans une unité de temps

Variables aléatoires discrètes: Loi de Poisson

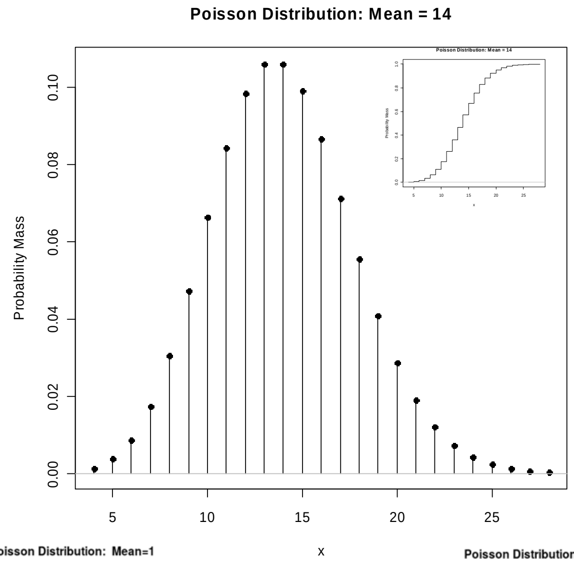
Distribution:

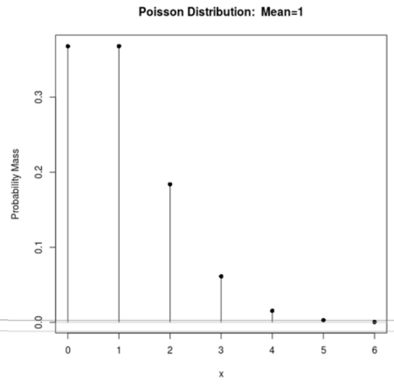
$Pr(X=k) = e^{-\mu}\mu^{k}/k!$

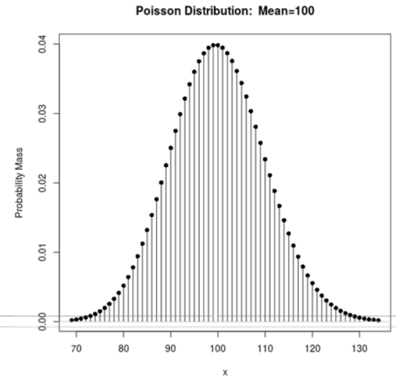
$\mu = \lambda\, t$, $e = 2{,}71828$

$E(X) = \mu$

$Var\ (X) = \mu$







Variables aléatoires discrètes: Loi de Poisson

- Le nombre des naissances dans un maternité d'un hôpitaux dans un année est 1000. La moyenne des naissances par jour est 1000/365, donc 2,74.
- On peut calculer la probabilité d'avoir 0, 1, 2, … naissances par jour, avec la distribution du Poisson. Ex:
- $P(0) = e^{-2,74}\, 2,74^{0} / 0! = 0,065$
- $P(1) = e^{-2,74}\, 2,74^{1} / 1! = 0,177$
- $P(2) = e^{-2,74}\, 2,74^{2} / 2! = 0,242$
- $P(3) = e^{-2,74}\, 2,74^{3} / 3! = 0,221$ (22,1%)

Variables aléatoires CONTINUES

DEFINITION:

- La variable qui prenne n'importe quelle valeure réelle dans un intervalle donné.
  - les valeurs ont des probabilités spécifiées
- peut prendre un ensemble non dénombrable des valeurs

- Exemples:
- X: le poids
- X: la taille
- X: indice de teinte des dents (engl. Tooth shade index)
- X: profondeur de poche (P.P.D Probing Pocket Depth)*

Mesurés sur des sujets tirés au hasard d'une population

* *la distance séparant le sommet de la gencive marginale du fond de la poche*

Les plus importantes distributions des probabilités

- En général, on ne connait pas les loi des distributions des variables dans le domaine médical/médical dentaire/biologique, …
- On cherche d'encadrer ces distributions dans des distributions théorétiques
- Ces distributions théorétiques sont des modèles avec lesquelles on va travailler

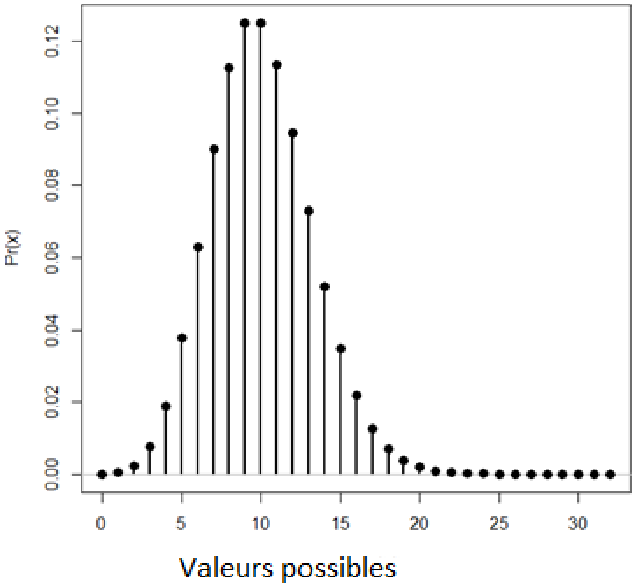
Distributions théorétiques:

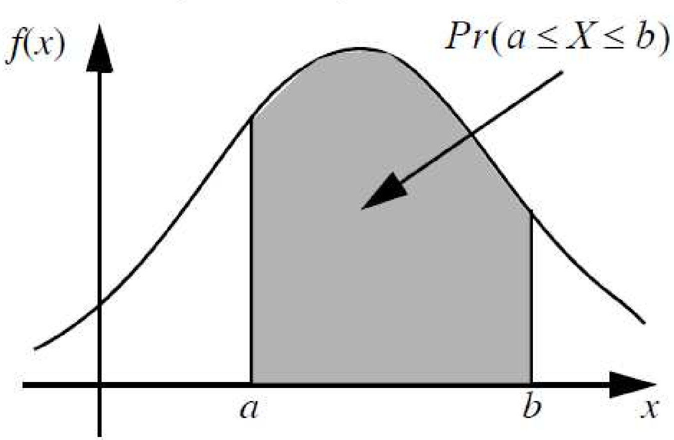
- **discrets:**
  - Binomiale: (le nombre de fois qu'un résultat d'interet se produit dans un nombre donné de tentatives).
  - Poisson: (probabilité d'événements rares: $p<0,05$)
- **continues:**
  - Normale
  - Student
  - Khi carré

Variables aléatoires discrètes vs. Variables continues

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

$$\sum_{i=1}^{\infty}$$



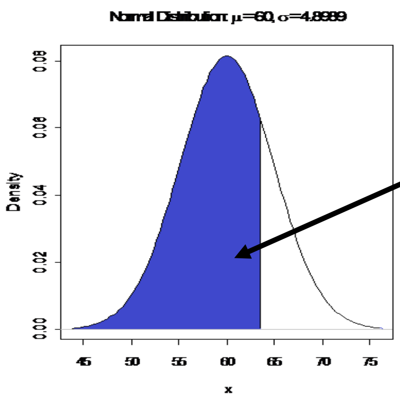


Variables aleatoires continues: DENSITÉ DE PROBABILITÉ

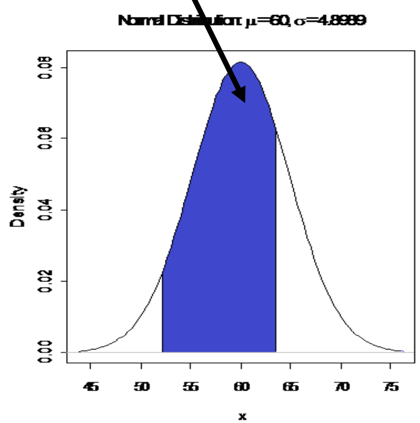
- La **fonction densité de Probabilité** f(x) d'une v.a. X
- est une fonction f positive telle que:
  - Pr(a <X ≤b) = la surface entre le graphique f et les droites x = a et x = b
  - La surface totale en dessous du graphique de f est 1,
- La **fonction de répartition F** associé a la v.a. X:

F(x) = Pr(X ≤ x)

$$\mathbb{P}(a < X \leq b) = \int_a^b f(u)\, du.$$



$$\mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(u)du.$$



Variables aleatoires CONTINUES: Esperance mathématique et Variance

- Esperance mathématique :

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

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n} x_i$$

$$E(X)=\sum_{i=1}^{k} x_i \Pr(X=x_i)$$

- Variance:
  - Var(X)= σ^2 =

$$V(X)=\int_{-\infty}^{+\infty}[x-E(X)]^2 f(x)dx$$

$$\sigma^2=\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$$

- Ecart type

$$V(X)=\sum_{i=1}^{n}[x_i-E(X)]^2\times\Pr(x_i)$$

$$\sigma(X)=\sqrt{Var(X)}=Var(X)^{1/2}$$

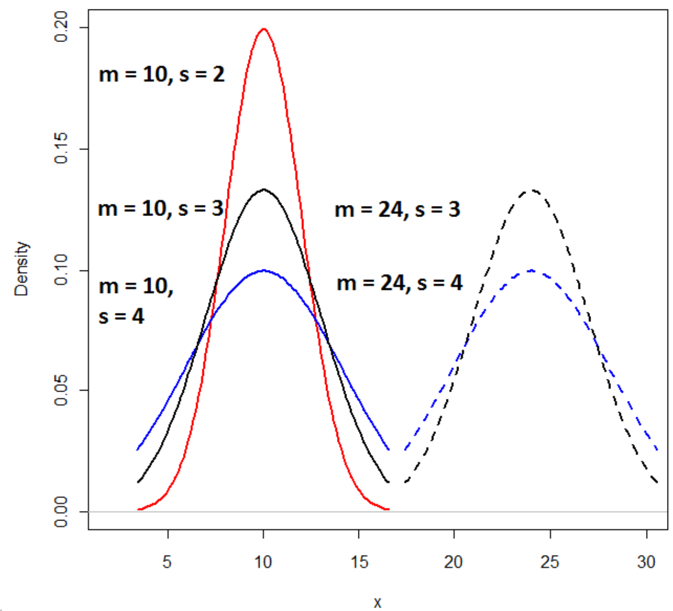
Variables aléatoires continues: Loi Normale (de GAUSS)

Model: **N(m, σ ²)**, dépend de la moyenne et la variation/déviation standard
la concentration des substances dans le sang (glycémie, triglicérides…)
l'erreur de mesure d'un certain objet, quantité (hauteur…), etc.

Fonction densité de probabilité: probabilité d'observer un événement donné

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

$$E(X) = \mu$$
$$V(X) = \sigma^2$$



Applications: le test Z pour comparaison des moyennes/fréquences; intervalles de confiance pour les moyennes/fréquences

Variables aléatoires continues: Loi Normale centrée réduite

Loi normale centrée réduite N(0,1)

Pour simplifier l'utilisation de la lois Normale

On fait une transformation: $$Z = \frac{X - \mu}{\sigma}$$

Donc pour la variable aléatoire Z:

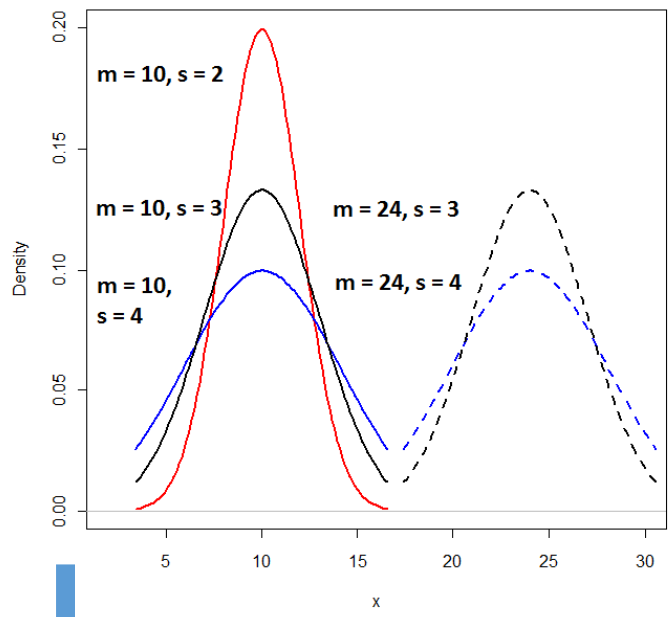
μ = 0

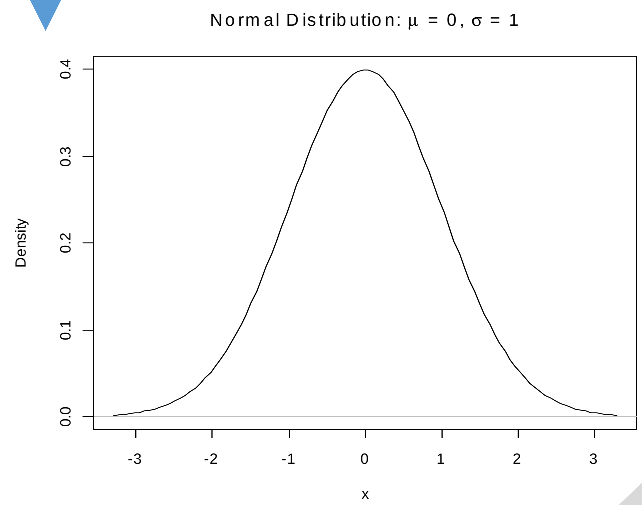
σ = 1

Z: sans unité de mesure

La fonction densité de probabilité:

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





VARIABLES ALÉATOIRES CONTINUES: LOI NORMALE CENTRÉE RÉDUITE

Pour le calcul de probabilités sans utiliser la fonction densité de probabilité, il y a des tables qui permet un utilisation pratique de ce lois.

Propriétés:

1) $Pr(Z \geq u) = 1 - Pr(Z < u)$
2) $Pr(Z \geq -u) = 1 - Pr(Z < -u) = 1 - (1 - Pr(Z \geq u)) = Pr(Z \leq u)$
3) $Pr(u \leq Z \leq v) = Pr(Z \leq v) - Pr(Z \leq u)$
4) $Pr(-u \leq Z \leq u) = Pr(Z \leq u) - Pr(Z \leq -u) = 2Pr(Z \leq u) - 1$

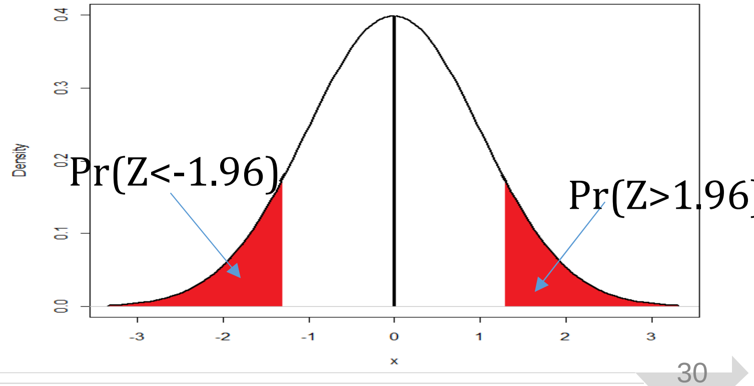
Exemple: $Pr(-1 \leq Z \leq 1) = ???$

??? % des valeurs sont dans l'intervalle [μ-σ, μ +σ]

$Pr(Z < -1{,}96$ ou $Z > 1{,}96)$:

?? % des valeurs sont dans l'intervalle [μ-1,96σ; μ+1,96

- utile pour construire des intervalles de confiance



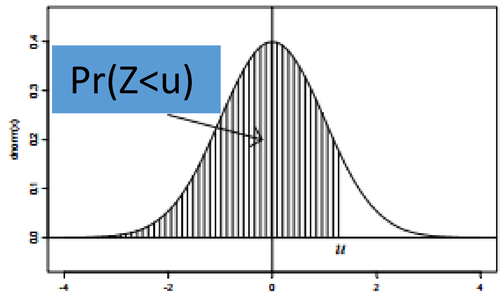
Variables aléatoires continues: Loi Normale centrée réduite

- Ce tableau donne l'aire sous la courbe normale centrée réduite

	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
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	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998

deuxième décimale

Première décimale

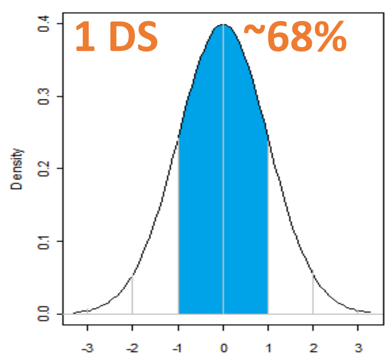


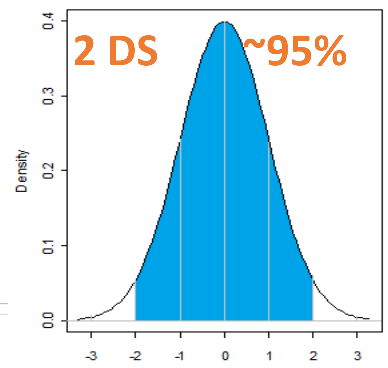
Exemple: $u=1.96 \Rightarrow \Pr(Z<1.96)=0.9750$
$\Pr(-1,96 \leq Z \leq 1,96)=\Pr(Z \leq 1.96)-\Pr(Z \leq -1.96)=$
$2 \cdot \Pr(Z \leq 1,96)-1=2*0,975-1=0,95$

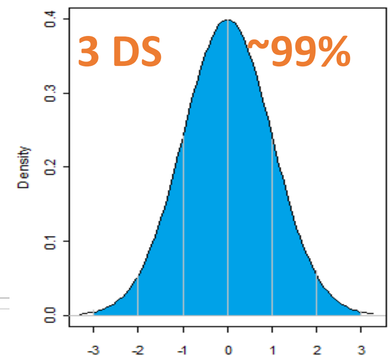
Variables aléatoires continues:
Loi Normale centrée réduite

- **Propriétés de la distribution Normale (Gauss)**
- Est une loi symétrique, unimodale, en forme de cloche
- Dans l'intervalle [moyenne -**1**·DS; moyenne +**1**·DS] on trouve **~68**,3 **%** de la population.
- Dans l'intervalle [moyenne -**2**·DS; moyenne +**2**·DS] on trouve **~95**,45 **%** de la population.
- Dans l'intervalle [moyenne -**1,96**·DS; moyenne +**1,96** ·DS] on trouve **~**95 **%** de la population.
- Dans l'intervalle [moyenne -**3**·DS; moyenne +**3**·DS] **~99**,7 **%** de la population.

DS =déviation standard







Loi Normale (GAUSS)

Applications:

- Intervalle de confiance pour un moyenne
  - Exemple: on prend un échantillon d'une population des personnes âgées .
  - ✓ On mesure le taux du cholestérol. On calcule la moyenne (ex. 184,3).
  - ✓ Notre question est: quelle est le vrai moyenne dans la population.
  - ✓ On peut calculer un intervalle dans lequel on va trouver la vrai moyenne avec un probabilité de 95% (par convention).
  - ✓ Pour trouver cet intervalle on a besoin de savoir quelle est la lois de probabilité dans la population.
  - ✓ On observe qu'il y a une variable continue, et on pense a la lois Normale. On vérifie la normalité des données. Si le données sont proche a la distribution normale on considère qu'on peut utiliser cet lois.
  - ✓ Avec la moyenne et l' écart type du cholestérol (si on suppose que le taux du cholestérol est normale distribuée), en utilisant la lois normale, on peut trouver que la vrai moyenne est compris entre 164,3 et 204,3 avec une probabilité de 95%.

Loi Normale (GAUSS)

Applications:

- Intervalle de confiance pour un fréquence (voir la lois Binomiale) – c'est identique.
- D'habitude si le nombre des sujets (expérimentes) est réduit on utilise la loi Binomiale

MAIS s'il est grand on utilise la lois Normale.

- Le test statistique Z pour comparaison des moyennes

  Exemple:

  - On donne a l'un des deux échantillons un traitement hypocholestérolémiant.
  - On mesure le taux du cholestérol.
  - On calcule les moyennes des deux échantillons (ex. 174,1 pour le group avec le traitement, et 194,4 pour le group sans traitement).
  - Question de recherche: il y a une différence réelle entre les deux groups? Ou mieux il y a une différence réelle entre les deux les populations? Ou en réalité les deux populations (traitée et non traitée) sont identiques?

Loi Normale (GAUSS)

Applications:

- On peut calculer quelle est la probabilité d'avoir une différence identique a celle qu'on a observe 194,4 – 174,1 si dans la réalité dans les populations depuis lesquelles on a extrait nos deux échantillons (traitée et non traitée) la différence est zero (0).
- Pour trouver cet probabilité on a besoin de savoir quelle est la lois de probabilité.
- On observe qu'il y a une variable continue, et on pense a la lois Normale.
- On vérifie la normalité des données.
- Si le données sont proche a la distribution normale on considère qu'on peut utiliser cet lois.
- On peut trouver que cette probabilité est 0.03 donc 3%.
- Par convention cette probabilité est petite (inferieure a 0.05), et on peut conclure que en fait il y a une différence statistiquement signifiante entre les deux populations.

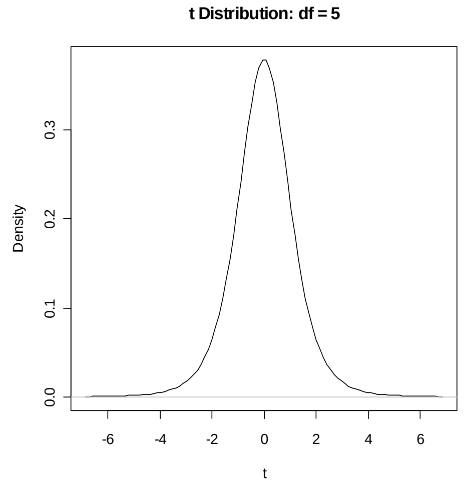
Variables aléatoires continues: La loi de Student t_n

- La variable aléatoire Student t_n

  - variable aléatoire continue – pour estimer la moyenne d'une population

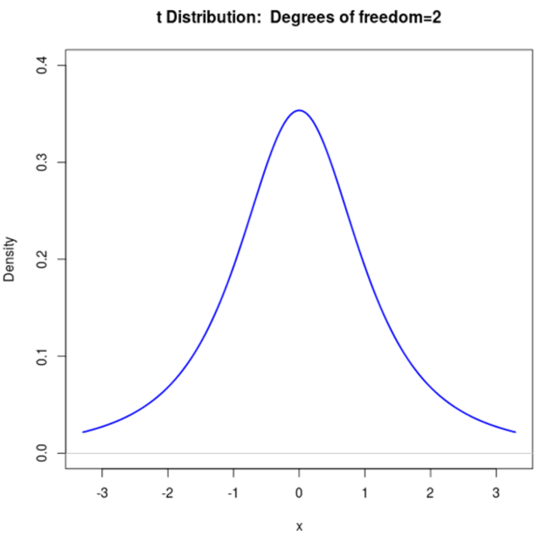
  - qui prend des valeurs dans l'intervalle $(-\infty, +\infty)$

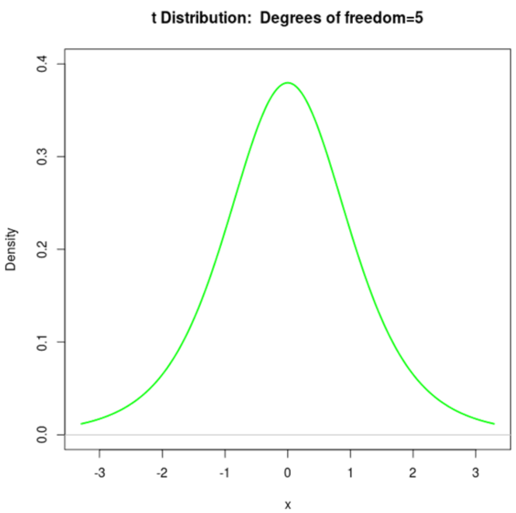
  - dépend d'un paramètre *n* appelée degrés de liberté (n=d.d.l).

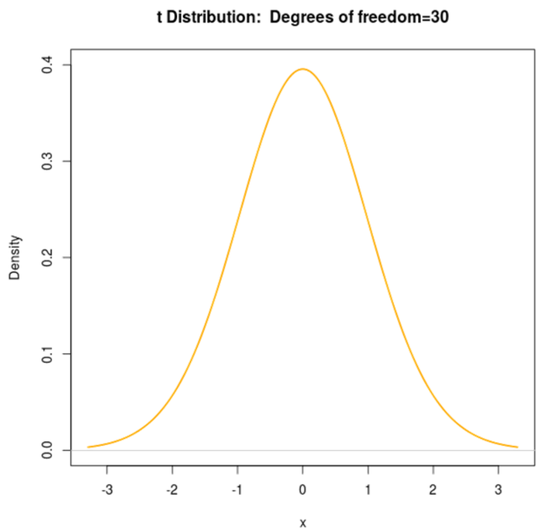


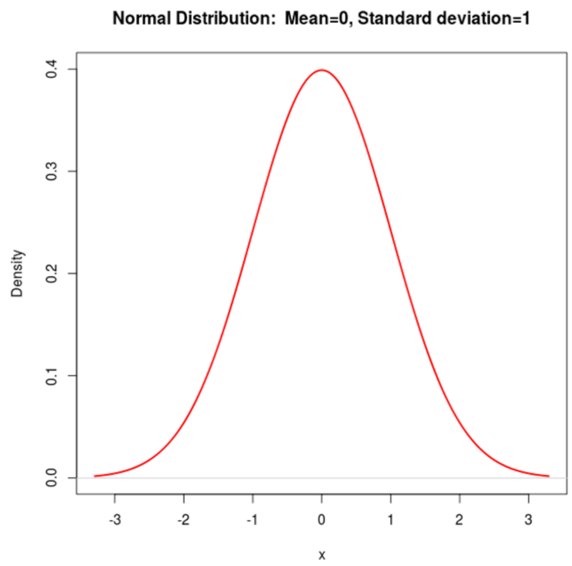
Variables aléatoires continues:La loi de Student t_n

- **$t_n \sim N(0, 1)$, quand $n \to \infty$,**
- **Si n > 30 la lois Normale et la lois Student sont très proches**
- Student (n = 2) Student (n = 5) Student (n = 30) Normale





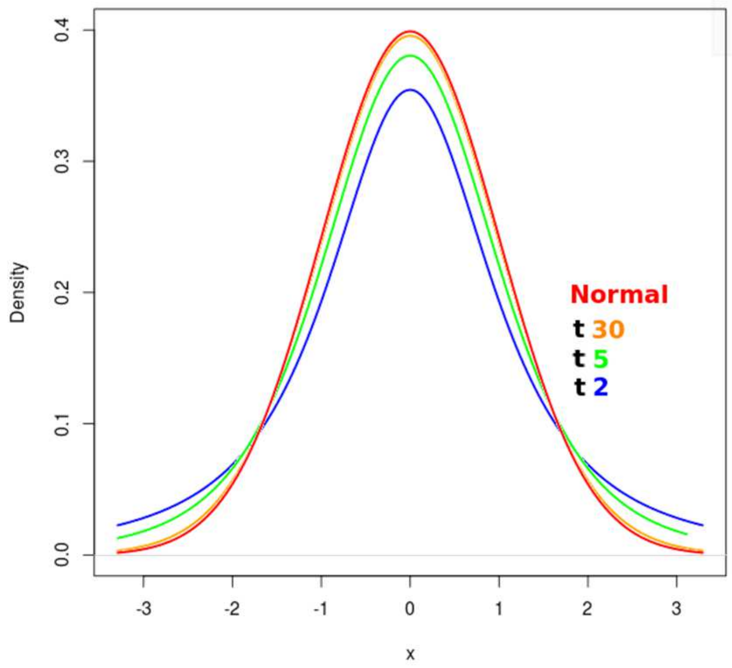




Variables aléatoires continues:La loi de Student t_n

- **Propriétés de la distribution t**
- symétrique
- unimodale
- en forme de cloche
- les queues sont plus longues

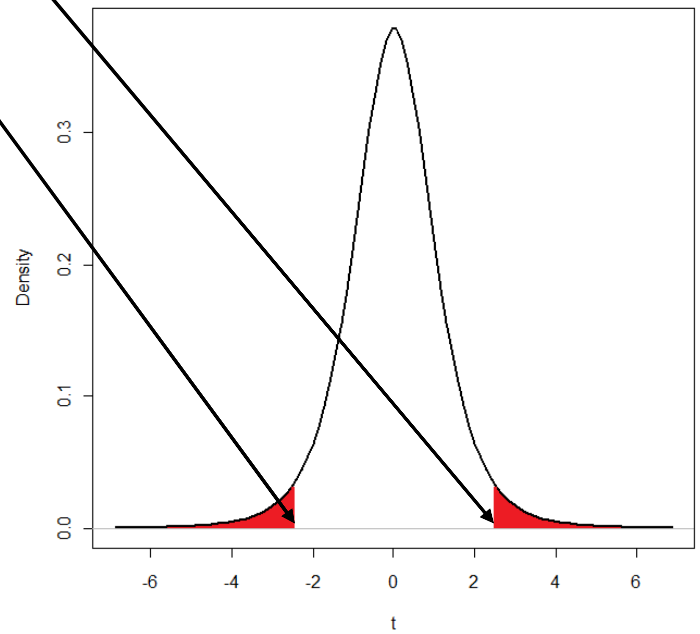
que la distribution normale



Variables aléatoires continues:La loi de Student t_n

- Pr(t>|x|)=P ou   Pr(t<-u ou t>u)=P
- Ou n =d.d.l., P une probabilité, t– valeur du v.a. Student

	P						
ν	0.50	0.40	0.30	0.20	0.10	0.05	0.01
1	1.000	1.376	1.963	3.078	6.314	12.706	63.657
2	0.816	1.061	1.386	1.886	2.920	4.303	9.925
3	0.765	0.978	1.250	1.638	2.353	3.182	5.841
4	0.741	0.941	1.190	1.533	2.132	2.776	4.604
5	0.727	0.920	1.156	1.476	2.015	2.571	4.032
6	0.718	0.906	1.134	1.440	1.943	2.447	3.707
7	0.711	0.896	1.119	1.415	1.895	2.365	3.499
8	0.706	0.889	1.108	1.397	1.860	2.306	3.355
9	0.703	0.883	1.100	1.383	1.833	2.262	3.250
10	0.700	0.879	1.093	1.372	1.812	2.228	3.169
11	0.697	0.876	1.088	1.363	1.796	2.201	3.106
12	0.695	0.873	1.083	1.356	1.782	2.179	3.055
13	0.694	0.870	1.079	1.350	1.771	2.160	3.012
14	0.692	0.868	1.076	1.345	1.761	2.145	2.977
15	0.691	0.866	1.074	1.341	1.753	2.131	2.947
16	0.690	0.865	1.071	1.337	1.746	2.120	2.921



Pr(t<-2.228 *ou* t>2.228)=0.05

La loi de Student t_n

- **Applications:**
- Intervalle de confiance
  - <u>d'une moyenne </u>pour un échantillon de taille n
    - avec n-1 dégrées de liberté (voir cours)
  - <u>d'une différence entre deux moyennes </u>pour deux échantillons indépendants
    - avec $n_1 + n_2 - 2$ dégrées de liberté
- Le test de Student pour la comparaison de moyennes,
  - Test t pour <u>un echantillon </u>(voir les cours suivants)
  - Test t pour <u>échantillons apparies</u>, (voir cours suivants)
    - (on comparer le même échantillon (n sujets) avant et après un intervention
    - avec n - 1 dégrées de liberté
  - Test t pour des <u>échantillons indépendants </u>(avec n_1, et n_2 sujets) (voir les cours suivants)
    - avec $n_1 + n_2 - 2$ dégrées de liberté
- Le test Student pour le <u>coefficient de corrélation </u>(voir les cours suivants)

La loi de Student t_n

Applications:

- Intervalle de confiance pour un moyenne

  Exemple: on prend un échantillon de 25 sujets d'une population des personnes âgées .

  – On mesure le taux des triglycérides. On calcule la moyenne (ex. 162,8).

  – Notre question est: quelle est le vrai moyenne dans la population.

  On peut calculer un intervalle dans lequel on va trouver la vrai moyenne avec un probabilité de 95% (par convention).

  On observe qu'il y a une variable continue, et on pense a la lois Normale. On vérifie la normalité des données. Si le données sont proche a la distribution normale on considère qu'on peut utiliser cet lois. Mais on n'a pas la moyenne et l'écart type dans la population. En plus l' échantillon est petit < 30. On va utiliser la lois t du Student.

  Avec la moyenne et l' écart type du cholestérol (si on suppose que le taux du cholestérol est normale distribuée), en utilisant la lois t, on peut trouver que la vrai moyenne est compris entre 152,8 et 172,8 avec une probabilité de 95%.

La loi de Student t_n

Applications:

- Le test Student pour comparaison des moyennes des échantillons indépendants

  Exemple:

  – On donne a l'un des deux échantillons (de 12 sujets) un traitement et un placebo pour baiser le taux des triglycérides. On mesure le taux du triglycérides.

  – On calcule les moyennes des deux échantillons (ex. 145,1 pour le group avec le traitement, et 187,9 pour le group sans traitement).

  – Notre question est: il y a une différence réelle entre les deux groups? Ou mieux il y a une différence réelle entre les deux les populations? Ou en réalité les deux populations (traitée et non traitée) sont identiques?

> The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data.
>
> — *John Tukey, Exploratory Data Analysis*

Les Graphiques (avec des simulations numériques) du cours ont été réalisés par
Mihaela Iancu & Daniel Leucuța avec:
R software envinronment for statistical computing and graphics