Mathematics for Data Science

Lecture 2: Continuous Random Variables, Sampling and Limit Theorems

Mohamad GHASSANY

EFREI PARIS

Variables Aléatoires Continues

- Précédemment nous avons traité des variables aléatoires discrètes, c'est-à-dire de variables dont l'univers est fini ou infini dénombrable.
- Il existe cependant des variables dont l'univers est infini non dénombrable.
- Exemples:
  - L'heure d'arrivée d'un train à une gare donnée.
  - La durée de vie d'un transistor.

Definition

X est une variable aléatoire continue s'il existe une fonction f non négative définie pour tout $x \in \mathbb{R}$ et vérifiant pour tout ensemble B de nombres réels la propriété

$$P(X \in B) = \int_B f(x)dx$$

La fonction f est appelée densité de probabilité de la variable aléatoire X.

- Tous les problèmes de probabilité relatifs à X peuvent être traités grâce à f.
- Par exemple pour $B = [a, b]$, on obtient:

$$P(a \leqslant X \leqslant b) = \int_a^b f(x)dx$$

Graphiquement, $P(a \leqslant X \leqslant b)$ est l'aire de la surface entre l'axe de x, la courbe correspondante à $f(x)$ et les droites $x = a$ et $x = b$.

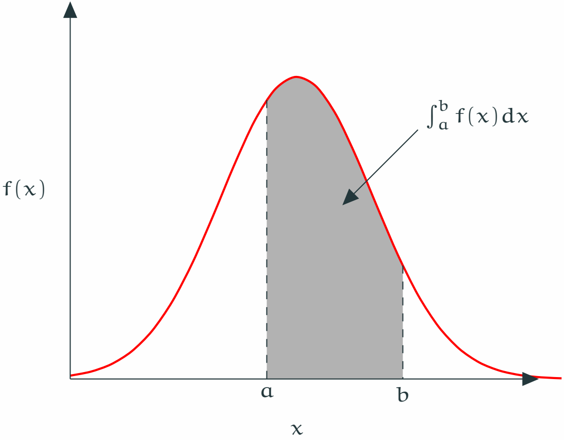


Figure 1: $P(a \leqslant X \leqslant b)$ = *surface grisée*

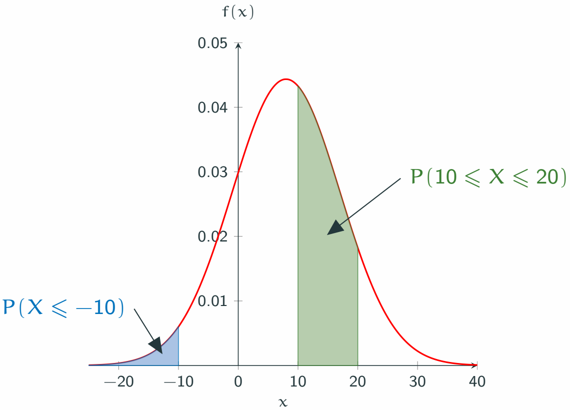
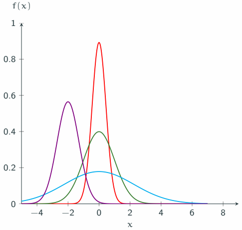
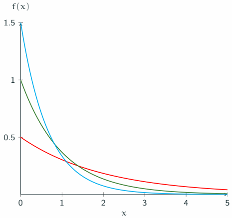
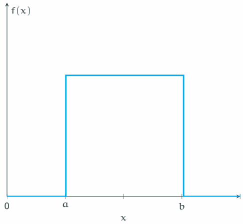


Figure 2: L'aire hachurée correspond à des probabilités. $f(x)$ étant une fonction densité de probabilité.







Properties

Pour toute variable aléatoire continue X de densité f:

- $f(x) \geqslant 0 \quad \forall x \in \mathbb{R}$
- $\int_{-\infty}^{+\infty} f(x)dx = 1$
- Comme $P(a \leqslant X \leqslant b) = \int_a^b f(x)dx$, si l'on pose $a = b$ il résulte $P(X = a) = \int_a^a f(x)dx = 0$
- Ceci siginifie que la probabilité qu'une variable aléatoire continue prenne une valeur isolée fixe est toujours nulle.

Exemple

Soit X la variable aléatoire réelle de densité de probabilité

$$f(x) = \begin{cases} kx & \text{si} \quad 0 \leqslant x \leqslant 5 \\ 0 & \text{sinon} \end{cases}$$

1. Calculer k.
2. Calculer: $P(1 \leqslant X \leqslant 3), P(2 \leqslant X \leqslant 4)$ et $P(X < 3)$.

Exemple

Soit X une variable aléatoire réelle continue ayant pour densité de probabilité

$$f(x) = \begin{cases} \frac{1}{6}x + k & \text{si} \quad 0 \leqslant x \leqslant 3 \\ 0 & \text{sinon} \end{cases}$$

1. Calculer k.
2. Calculer $P(1 \leqslant X \leqslant 2)$

Fonction de répartition d'une v.a.c

Definition

Si comme pour les variables aléatoires discrètes, on définit la fonction de répartition de X par:

$$\begin{aligned} F_X : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto F_X(a) = P(X \leqslant a) \end{aligned}$$

alors la relation entre la fonction de répartition F_X et la fonction densité de probabilité $f(x)$ est la suivante:

$$\forall \quad a \in \mathbb{R} \quad F_X(a) = P(X \leqslant a) = \int_{-\infty}^{a} f(x)dx$$

Properties

Pour une variable aléatoire continue X:

- $F'_X(x) = \frac{d}{dx} F_X(x) = f(x)$.
- Pour tous réels $a \leqslant b$,

$$\begin{aligned} P(a < X < b) &= P(a < X \leqslant b) = P(a \leqslant X < b) \\ &= P(a \leqslant X \leqslant b) = F_X(b) - F_X(a) = \int_a^b f(x)dx \end{aligned}$$

La fonction de répartition correspond aux probabilités cumulées associées à la variable aléatoire continue sur l'intervalle d'étude.

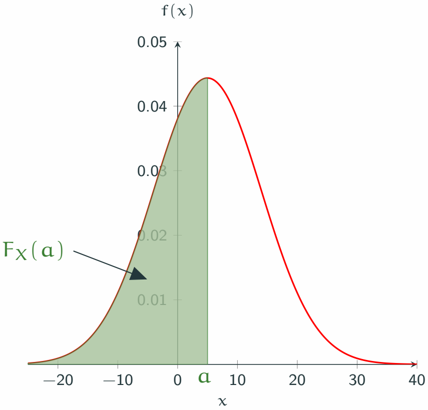


Figure 3: L'aire hachurée en vert sous la courbe de la fonction densité de probabilité correspond à la probabilité $P(X < a) = F_X(a)$ *et vaut* $0,5$ *car ceci correspond exactement à la moitié de l'aire totale sous la courbe.*

Fonction de répartition d'une v.a.c

Properties

Les propriétés associées à la fonction de répartition sont les suivantes:

1. F_X est continue sur $\mathbb{R}$, dérivable en tout point où f est continue.
2. F_X est croissante sur $\mathbb{R}$.
3. F_X est à valeurs dans $[0, 1]$.
4. $\lim\limits_{x \to -\infty} F_X(x) = 0$ et $\lim\limits_{x \to +\infty} F_X(x) = 1$.

Exemple

Soit X et Y deux variables aléatoires réelles de densités de probabilité

$$f_X(x) = \begin{cases} kx & \text{si} \quad 0 \leqslant x \leqslant 5 \\ 0 & \text{sinon} \end{cases}$$

et

$$f_Y(y) = \begin{cases} \frac{1}{6}y + k & \text{si} \quad 0 \leqslant y \leqslant 3 \\ 0 & \text{sinon} \end{cases}$$

Calculer $F_X(a)$ et $F_Y(a)$ pour tout $a \in \mathbb{R}$.

Fonction d'une variable aléatoire continue

Fonction d'une variable aléatoire continue

- Soit X une variable aléatoire continue de densité f_X et de fonction de répartition F_X.
- Soit h une fonction continue définie sur $X(\Omega)$, alors $Y = h(X)$ est une variable aléatoire.
- Pour déterminer la densité de Y, notée f_Y, on commence par calculer la fonction de répartition de Y, notée F_Y, ensuite nous dérivons pour déterminer f_Y.

Calcul de densités

Soit X une variable aléatoire continue de densité f_X et de fonction de répartition F_X. Calculer la densité des variables aléatoires suivantes:

- $Y = aX + b$
- $Z = X^2$
- $T = e^X$

Example

Soit la v.a.c X ayant la fonction de densité

$$f_X(x) = 2x \times \mathbb{1}_{[0,1]}(x)$$

Déterminer la densité de: $Y = 3X + 1$, $Z = X^2$ et $T = e^X$.

Moments des variables aléatoires continues

Definition

Si X est une variable aléatoire absolument continue de densité f, on appelle espérance de X, le réel $E(X)$, défini par:

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

si cette intégrale est convergente.

Les propriétés de l'espérance d'une variable aléatoire continue sont les mêmes que pour une variable aléatoire discrète.

Properties

Soit X une variable aléatoire continue,

- $E(aX+b) = aE(X)+b \qquad a \geqslant 0$ et $b \in \mathbb{R}$.
- Si $X \geqslant 0$ alors $E(X) \geqslant 0$.
- Si X et Y sont deux variables aléatoires définies sur un même univers Ω alors

$$E(X+Y) = E(X)+E(Y)$$

Théorème de transfert

Theorem

Si X est une variable aléatoire de densité $f(x)$, alors pour toute fonction réelle g on aura

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

Example

Soit la v.a.c X ayant la fonction de densité

$$f_X(x) = \begin{cases} 2x & \text{si} \quad 0 \leqslant x \leqslant 1 \\ 0 & \text{sinon} \end{cases}$$

Calculer l'espérance des variables aléatoires $Y = 3X + 1$, $Z = X^2$ et $T = e^X$.

La variance d'une variable aléatoire $V(X)$ est l'espérance mathématique du carré de l'écart à l'espérance mathématique. C'est un paramètre de dispersion qui correspond au moment centré d'ordre 2 de la variable aléatoire X.

Definition

Si X est une variable aléatoire ayant une espérance $E(X)$, on appelle variance de X le réel

$$V(X) = E\left([X - E(X)]^2\right) = E(X^2) - [E(X)]^2$$

Si X est une variable aléatoire continue, on calcule $E(X^2)$ en utilisant le théorème de transfert,

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

Example

Calculer la variance de la variable aléatoire X définie dans l'exemple précédent.

Properties

Si X est une variable aléatoire admettant une variance alors:

- $V(X) \geqslant 0$, si elle existe.
- $\forall\ a \in \mathbb{R}, V(aX) = a^2V(X)$
- $\forall\ (a, b) \in \mathbb{R}, V(aX + b) = a^2V(X)$
- Si X et Y sont deux variables aléatoires indépendantes, $V(X + Y) = V(X) + V(Y)$

Definition

Si X est une variable aléatoire ayant une variance $V(X)$, on appelle écart-type de X, le réel:

$$\sigma_X = \sqrt{V(X)}$$

Lois Usuelles de Variables Aléatoires Continues

Loi Uniforme $\mathcal{U}(a, b)$

Définition

La variable aléatoire X suit une loi uniforme sur le segment $[a, b]$ avec $a < b$ si sa densité de probabilité est donnée par

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{si} \quad x \in [a, b] \\ 0 & \text{si} \quad x \notin [a, b] \end{cases} = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x)$$

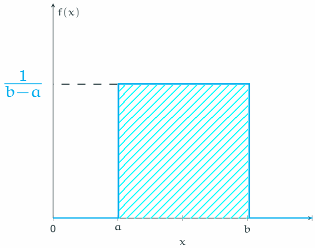


Figure 4: Fonction de densité de $\mathcal{U}([a, b])$

Loi Uniforme $\mathcal{U}(a, b)$

- La *fonction de répartition* associée à la loi uniforme continue est

$$F_X(x) = \begin{cases} 0 & \text{si} \quad x < a \\ \dfrac{x-a}{b-a} & \text{si} \quad a \leqslant x \leqslant b \\ 1 & \text{si} \quad x > b \end{cases}$$

- $E(X) = \dfrac{b+a}{2}$
- $V(X) = \dfrac{(b-a)^2}{12}$

Loi exponentielle $\mathcal{E}(\lambda)$

Loi exponentielle $\mathcal{E}(\lambda)$

Définition

On dit qu'une variable aléatoire X est **exponentielle** (ou suit la loi exponentielle) de paramètre λ si sa densité est donnée par

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si} \quad x \geqslant 0 \\ 0 & \text{si} \quad x < 0 \end{cases} = \lambda e^{-\lambda x} \mathbb{1}_{\mathbb{R}^+}(x)$$

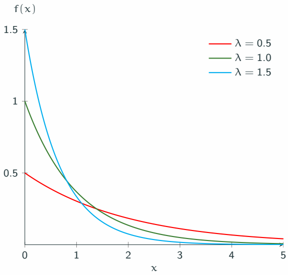
On dit $X \sim \mathcal{E}(\lambda)$

- La fonction de répartition F d'une variable aléatoire exponentielle est donnée par

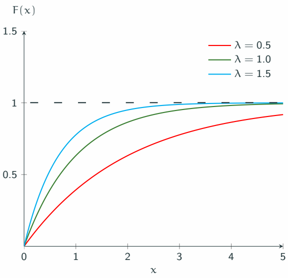
$$\text{Si } x \geqslant 0 \quad F(x) = P(X \leqslant x) = 1 - e^{-\lambda x}$$

- $E(X) = \dfrac{1}{\lambda}$ et $V(X) = \dfrac{1}{\lambda^2}$

Loi exponentielle $\mathcal{E}(\lambda)$



(a) Représentation graphique de la densité d'une loi exponentielle



(b) Représentation graphique de la fonction de répartition d'une loi exponentielle

Cas d'utilisations de la loi exponentielle:

- Représenter le temps d'attente avant l'arrivée d'un événement spécifié.
- Modéliser la durée de vie d'un phénomène **sans mémoire**, ou sans vieillissement, ou sans usure.
- On dit qu'une variable aléatoire non négative X est *sans mémoire* lorsque

$$P(X > t + h | X > t) = P(X > h) \qquad \forall \quad t, h \geqslant 0$$

- Par exemple, la durée de vie de la radioactivité ou d'un composant électronique.

Loi Normale ou de Laplace-Gauss

$\mathcal{N}(\mu, \sigma^2)$

Définition

Une variable aléatoire X est dite **normale** avec paramètres μ et σ^2 si la densité de X est donnée par

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \qquad \forall\, x \in \mathbb{R}$$

Avec $\mu \in \mathbb{R}$ et $\sigma \in \mathbb{R}^+$. On dit que $X \sim \mathcal{N}(\mu, \sigma^2)$.

Moments de la loi Normale $\mathcal{N}(\mu, \sigma^2)$

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

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \qquad \forall\, x \in \mathbb{R}$$

- La fonction f est paire autour d'un axe de symétrie $x = \mu$ car $f(x+\mu) = f(\mu - x)$.
- $f'(x) = 0$ pour $x = \mu$, $f'(x) < 0$ pour $x < \mu$ et $f'(x) > 0$ pour $x > \mu$

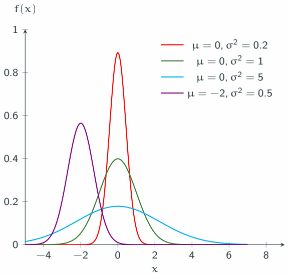


Figure 6: ***Remarque:*** *Le paramètre μ représente l'axe de symétrie et σ le degré d'aplatissement de la courbe de la loi normale dont la forme est celle d'une courbe en cloche.*

Théorème

- $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$
- $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$
- X_1 et X_2 sont indépendantes.

Alors $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Loi Normale centrée réduite $\mathcal{N}(0, 1)$

Loi Normale centrée réduite $\mathcal{N}(0,1)$

Définition

Une variable aléatoire continue X suit une **loi normale centrée réduite** si sa densité de probabilité est donnée par

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \qquad \forall\, x \in \mathbb{R}$$

On dit $X \sim \mathcal{N}(0,1)$.

Moments de la loi Normale $\mathcal{N}(0,1)$

- $E(X) = 0$
- $V(X) = 1$

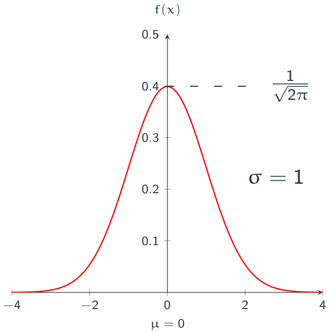


Figure 7: *Densité d'une loi normale centrée réduite* $\mathcal{N}(0,1)$.

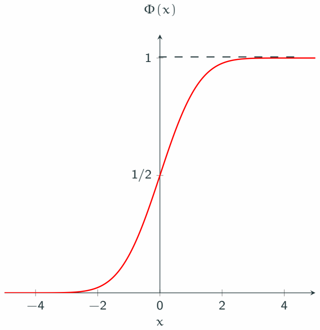


Figure 8: *Fonction de répartition de* $\mathcal{N}(0,1)$

Relation entre loi normale et loi normale centrée réduite

Relation entre loi normale et loi normale centrée réduite

Théorème

Si X suit une loi normale $\mathcal{N}(\mu, \sigma^2)$, alors $Z = \dfrac{X-\mu}{\sigma}$ est une variable centrée réduite qui suit la loi normale centrée réduite $\mathcal{N}(0,1)$.

La fonction de répartition de la loi normale centrée réduite permet d'obtenir les probabilités associées à toutes variables aléatoires normales $\mathcal{N}(\mu, \sigma^2)$ après transformation en variable centrée réduite.

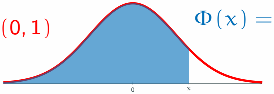
Définition

On appelle fonction Φ, la fonction de répartition de la loi normale centrée réduite $\mathcal{N}(0,1)$, telle que

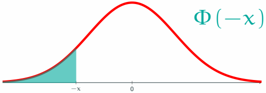
$$\forall\, x \in \mathbb{R} \quad \Phi(x) = P(X \leqslant x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} f(t)\,dt$$

Fonction de répartition de la loi normale centrée réduite $\mathcal{N}(0,1)$

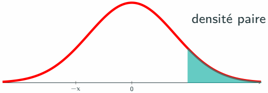
$X \sim \mathcal{N}(0,1)$



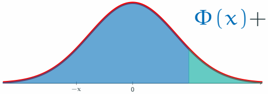
$\Phi(x) = P(X \leqslant x)$



$\Phi(-x) = P(X \leqslant -x)$



densité paire $\Rightarrow$ courbe symétrique % $x = 0$



$\Phi(x) + \Phi(-x) = 1$

Fonction de répartition de la loi normale centrée réduite $\mathcal{N}(0,1)$

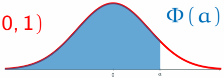
Propriétés de Φ

Les propriétés associées à la fonction de répartition Φ sont:

1. Φ est croissante, continue et dérivable sur $\mathbb{R}$ et vérifie:
   $\lim\limits_{x\to-\infty} \Phi(x) = 0$ et $\lim\limits_{x\to\infty} \Phi(x) = 1$
2. $\forall\, x \in \mathbb{R} \quad \Phi(x) + \Phi(-x) = 1$
3. $\forall\, x \in \mathbb{R} \quad \Phi(x) - \Phi(-x) = 2\Phi(x) - 1$

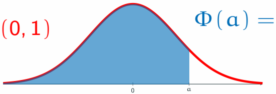
Table de la loi normale centrée réduite $\mathcal{N}(0,1)$

$X \sim \mathcal{N}(0,1)$



$\Phi(a) = P(X \leqslant a)$

a	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
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	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
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
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



a	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
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	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
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
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

Par exemple, pour $x = 1.23$ (intersection de la ligne 1.2 et de la colonne 0.03), on obtient : $\Phi(1.23) \approx$ 0.8907.

Calcul des probabilités d'une loi normale

Exemple 1

Soit X une variable aléatoire normale centrée réduite. Calculer:

1. $P(X > 2)$
2. $P(2 < X < 5)$

Exemple 2

Soit X une variable aléatoire normale de paramètres $\mu = 3$ et $\sigma^2 = 4$. Calculer:

1. $P(X > 0)$
2. $P(2 < X < 5)$
3. $P(|X - 3| > 4)$

Approximation normale d'une répartition binomiale

Théorème de Moivre Laplace

On suppose que pour tout n, X_n suit la loi binomiale $\mathcal{B}(n, p)$ avec $p \in]0, 1[$.

Alors la variable $Z_n = \frac{X_n - np}{\sqrt{np(1-p)}}$ converge en loi vers une loi normale centrée réduite $\mathcal{N}(0, 1)$.

- Ce résultat a été progressivement généralisé par Laplace, Gauss et d'autres pour devenir le théorème actuellement connu comme théorème centrale limite qui est un des deux résultats les plus importants de la théorie de probabilités.
- En pratique de très nombreux phénomènes aléatoires suivent approximativement une distribution normale.

Approximation normale d'une répartition binomiale

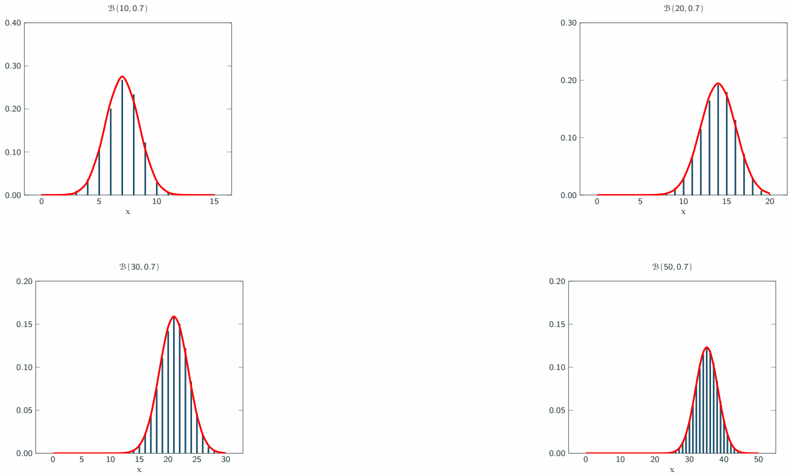


Figure 9: *La loi de probabilité d'une variable aléatoire $\mathcal{B}(n, p)$ devient de plus en plus "normale" à mesure que n augmente.*

Lois déduites de la loi normale

Loi de χ^2 (Khi-deux) de Pearson

Définition

Soit $X_1, X_2, \ldots, X_n$, n variables **normales centrées réduites**, et Y la variable aléatoire définie par

$$Y = X_1^2 + X_2^2 + \ldots + X_i^2 + \ldots + X_n^2 = \sum_{i=1}^{n} X_i^2$$

On dit que Y suit la loi de χ^2 (ou loi de Pearson) à n degrés de liberté, $Y \sim \chi^2(n)$

- La loi de χ^2 trouve de nombreuses applications dans le cadre de la comparaison de proportions, des tests de conformité d'une distribution observée à une distribution théorique et le test d'indépendance de deux caractères qualitatifs. Ce sont les *tests du khi-deux*.
- Remarque: Si $n = 1$, la variable du χ^2 correspond au carré d'une variable normale centrée réduite $\mathcal{N}(0, 1)$.

Définition

Soit U une variable aléatoire suivant une loi **normale centrée réduite** $\mathcal{N}(0,1)$ et V une variable aléatoire suivant une loi de $\chi^2(n)$, U et V étant indépendantes, on dit alors que $T_n = \frac{U}{\sqrt{\frac{V}{n}}}$ suit une **loi de Student** à n degrés de liberté. $T_n \sim St(n)$

- La loi de Student est utilisée lors des tests de comparaison de paramètres comme la moyenne et dans l'estimation de paramètres de la population à partir de données sur un échantillon (Test de Student).

Définition

Soit U et V deux variables aléatoires indépendantes suivant une loi de χ^2 respectivement à n et m degrés de liberté.

On dit que $F = \dfrac{U/n}{V/m}$ suit une loi de Fisher-Snedecor à (n, m) degrés de liberté. $F \sim \mathcal{F}(n, m)$

- La loi de Fisher-Snedecor est utilisée pour comparer deux variances observées et sert surtout dans les très nombreux tests d'analyse de variance et de covariance.

Quelques références – interactives

- A visual introduction to probability and statistics: http://students.brown.edu/seeing-theory/

- Distribution Calculator: https://gallery.shinyapps.io/dist_calc/

Introduction to Statistical Inference

- **Statistics** is the science of collecting, processing and analyzing data derived from the observation of random phenomena.
- Data analysis is used to **describe** the phenomena studied, **make predictions** and **make decisions** about them. In this way, statistics is an essential tool for understanding and managing complex phenomena.
- The data studied can be of any nature, which makes statistics useful in all disciplinary fields.

The fundamental point is that the data present uncertainties and **variations**.

Statistical methods are divided into two classes:

- **Descriptive statistics**, **exploratory statistics** or **data analysis**, aims to summarize the information contained in the data in a synthetic and efficient way. Probabilities play only a minor role here.
- **Inferential statistics** goes beyond the simple description of data. Its purpose is to **make predictions** and **make decisions** based on observations. In general, it is necessary to propose **probabilistic models** of the studied random phenomenon and to know how to manage the risks of errors. Probabilities play a fundamental role here.

The statistical approach i

- **Probability** can be considered as a branch of pure mathematics, based on the theory of measurement, abstract and completely disconnected from reality.
- **Applied probability** proposes **probabilistic models** of the course of concrete random phenomena. One can then, **prior to any experiment**, make predictions about what will happen.

Example: it is usual to model the duration of the good functioning or life of a system, let's say a light bulb, by a random variable X of exponential law of parameter λ. Having adopted this probabilistic model, we can perform all the calculations we want. For example:

- The probability that the bulb has not yet failed at date t is $P(X > t) = e^{-\lambda t}$.
- The average lifetime is $E(X) = 1/\lambda$.
- If n identical light bulbs are turned on at the same time, and they work independently of each other, the number N_t of light bulbs that will fail before a time t is a random variable of binomial distribution $\mathcal{B}(n, P(X \leqslant t)) = \mathcal{B}(n, 1 - e^{-\lambda t})$. Thus we expect that, on average, $E(N_t) = n(1 - e^{-\lambda t})$ bulbs will fail between 0 and t.

The statistical approach ii

In practice, if we want to use the theoretical results stated above, we have to make sure that we have chosen a good model, i.e.that the life span of these bulbs is a random variable with an exponential law, and, on the other hand, we have to be able to calculate the value of the parameter λ in some way. It is statistics that will allow us to solve these problems. To do this, we need to do an **experiment**, **collect data** and **analyze** them.

We therefore set up what we call a **test** or an **experiment**. We run $n = 10$ identical bulbs in parallel and independently of each other, under the same experimental conditions, and we record their lifetimes. Let's say that we obtain the following lifetimes, expressed in hours: $91.6, 35.7, 251.3, 24.3, 5.4, 67.3, 170.9, 9.5, 118.4, 57.1$

Let us note $x_1, \ldots, x_n$ these observations. We will therefore consider that $x_1, \ldots, x_n$ are the ***samples*** of random variables $X_1, \ldots, X_n$.

This means that after the experiment, the lifetime has been observed. We say that x_i is a sample (a realization) of X_i on the test performed.

Since the bulbs are identical, it is natural to suppose that X_i have the same law. This means that the same random phenomenon is observed several times.

We can also assume that the X_i are independent random variables. We can then ask the following questions:

The statistical approach iii

1. With respect to these observations, is it reasonable to assume that the lifetime of a light bulb is a random variable with an exponential distribution? If not, what other law would be more appropriate? This is a **fit test** (Chi-square test) problem.
2. If the exponential distribution model has been chosen, how can we propose a good value (or set of values) for the parameter λ? This is a parametric **estimation** problem.
3. In this case, can we guarantee that λ is less than a fixed value λ_0? This will guarantee that $\mathbb{E}(X) = 1/\lambda \geqslant 1/\lambda_0$, in other words that the bulbs will be sufficiently reliable. This is a **parametric hypothesis testing** problem.
4. If we have 100 light bulbs, how many failures can we expect in less than 50 hours? This is a **prediction** problem.

Sampling

Sampling

Random Sample

The random variables $X_1, X_2, \ldots, X_n$ are a **random sample** of size n if (a) the X_i's are independent random variables, and (b) every X_i has the same probability distribution.

The observed data are also referred to as a random sample, but the use of the same phrase should not cause any confusion.

Statistic

A **statistic** is any function of the observations in a random sample.

For example, if $X, X_2, \ldots, X_n$ is a random sample of size n, the **sample mean** X, the **sample variance** S^2, and the **sample standard deviation** S are statistics. Since a statistic is a random variable, it has a probability distribution.

Sampling Distribution

The probability distribution of a statistic is called a **sampling distribution.**

For example, the probability distribution of X is called the **sampling distribution of the mean.**

The statistic $\overline{X}_n$

Consider determining the sampling distribution of the sample mean $\overline{X}$. Suppose that a random sample of size n is taken from a normal population with mean μ and variance σ^2. Now each observation in this sample, say, $X_1, X_2, \ldots, X_n$, is a normally and independently distributed random variable with mean μ and variance σ^2. Then, because linear functions of independent, normally distributed random variables are also normally distributed (Chapter 5), we conclude that the sample mean

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

has a normal distribution with mean

$$\mu_{\overline{X}} = \frac{\mu + \mu + \cdots + \mu}{n} = \mu$$

and variance

$$\sigma^2_{\overline{X}} = \frac{\sigma^2 + \sigma^2 + \cdots + \sigma^2}{n^2} = \frac{\sigma^2}{n}$$

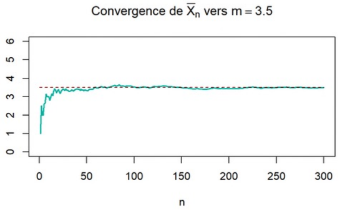
Limit Theorems

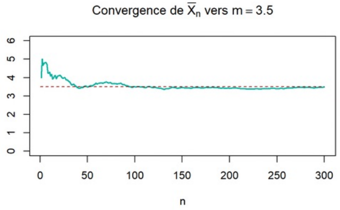
Strong law of large numbers

Let $X_1, X_2, \ldots, X_n$ be a set of independent random variables having a common distribution, and let $E[X_i] = \mu$. then, with probability 1

$$\frac{X_1 + X_1 + \ldots + X_n}{n} \to \mu \quad \text{as } n \to \infty.$$

x_i	1	2	3	4	5	6
$P(X = x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$





Central Limit Theorem

Central Limit Theorem

If $X_1, X_2, \ldots, X_n$ is a random sample of size n taken from a population (either finite or infinite) with mean μ and finite variance σ^2, and if $\overline{X}$ is the sample mean, the limiting form of the distribution of

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \qquad (7\text{-}1)$$

as $n \to \infty$, is the standard normal distribution.

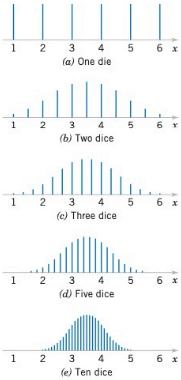


Figure 7-1 Distributions of average scores from throwing dice. [Adapted with permission from Box, Hunter, and Hunter (1978).]

Application 1:

Pour une taille d'échantillon n suffisamment grande, on peut considérer que $\overline{X}_n$ a pour loi:

$$\overline{X}_n \sim \mathcal{N}(m, \frac{\sigma^2}{n})$$

Dans la notation de la loi normale ci dessus $\frac{\sigma^2}{n}$ est la variance. $\frac{\sigma}{\sqrt{n}}$ est l'écart-type.

Application 2: la loi d'un pourcentage, étudiée dans la section suivante.

The statistic P_n (percentage)

Soit X la variable aléatoire représentant le nombre de succès au cours d'une suite de n répétitions indépendantes d'une même épreuve dont la probabilité de succès est p.

La loi de X est la loi binomiale de paramètres n et p notée $\mathcal{B}(n,p)$. X est la somme de n variables indépendantes de Bernoulli de paramètre p.

Notons P_n la *fréquence empirique* du nombre de succès parmi les n épreuves:

$$P_n = \frac{X}{n}$$

> $P_n = \overline{X}_n$ car X est la somme de n variables indépendantes de Bernoulli de paramètre p.

P_n a pour espérance et pour variance:

$$E(P_n) = p \qquad \text{et} \qquad V(P_n) = \frac{p(1-p)}{n}$$

En appliquant le théorème central limite à X somme des variables de Bernoulli:

Pour n suffisamment grand, on peut considérer que P_n suit la loi normale :

$$P_n \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$$

Définition 2.4 (Variance empirique) La statistique S_n^2 ou variance empirique d'échantillon est définie par:

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$$

Propriétés:

- $S_n^2 = \frac{1}{n}\left(\sum_{i=1}^{n} X_i^2\right) - \left(\overline{X}_n\right)^2$.
- $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - m\right)^2 - \left(\overline{X}_n - m\right)^2$.
- S_n^2 converge presque sûrement vers σ^2.

Espérance de S_n^2:

- L'espérance de S_n^2 est:

$$E(S_n^2) = \frac{n-1}{n}\sigma^2$$

démonstration:

$$\begin{aligned}
E(S_n^2) &= \frac{1}{n}\sum_{i=1}^{n} E\left(X_i - m\right)^2 - E\left(\overline{X}_n - m\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} V(X_i) - V(\overline{X}_n) \\
&= \frac{1}{n}\sum_{i=1}^{n}\sigma^2 - \frac{\sigma^2}{n} = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2
\end{aligned}$$

On peut remarquer que si on pose: $S_n^{*2} = \frac{n}{n-1}S_n^2$ alors $E(S_n^{*2}) = \sigma^2$.