

Variables aléatoires à densité

Outils Mathématiques & Statistique Décisionnelle
Licence 3 Management, Economie de la firme, CCF
Faculté d'économie, gestion & AES
Université de Bordeaux - Collège DSPEG

Automne 2024

A. Lourme
alexandre.lourme@u-bordeaux.fr

Outline

- Généralités
- Lois continues usuelles

Définitions

X est une variable aléatoire à densité¹ s'il existe une application $f : \mathbb{R} \rightarrow \mathbb{R}$ positive tq : si $-\infty \leq a < b \leq +\infty$, alors : $\mathbb{P}(a < X \leq b) = \int_a^b f(x)dx$.

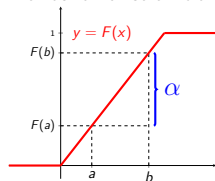
La fonction de répartition de X est alors $F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f(t)dt$.

Propriétés.

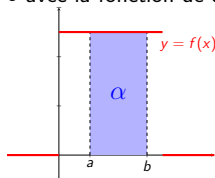
- La fonction de répartition F est une primitive de la densité f
- La densité f est la dérivée de la fonction de répartition F
- En dehors du support $X(\Omega)$ (ensemble des valeurs prises par X) : la fonction de répartition F est constante par morceaux et la fonction de densité f est nulle
- $\mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X < b) = F(b) - F(a)$

Ainsi, quand X est une variable à densité, $\alpha = \mathbb{P}(a < X \leq b)$ se calcule :

- avec la fonction de répartition
- avec la fonction de densité



$$\alpha = F(b) - F(a)$$



$$\alpha = \int_a^b f(x)dx$$

¹ on dit aussi variable absolument continue que l'on abrège par : v.a.c.

Espérance & Variance

X est une variable aléatoire de densité f .

L'*espérance* de X : $E(X) = \int_{\mathbb{R}} x.f(x)dx$ représente la valeur attendue de X .

La *variance* de X est :

$$V(X) \stackrel{(a)}{=} E((X - E(X))^2) \stackrel{(b)}{=} E(X^2) - E(X)^2 \stackrel{(c)}{=} \int_{\mathbb{R}} x^2.f(x)dx - E(X)^2.$$

(a) $V(X)$ est l'écart quadratique moyen entre X et sa valeur attendue

(b) Propriété de Koenig-Huygens

(c) $E(g(X)) = \int_{\mathbb{R}} g(x).f(x)dx$ (Théorème de transfert)

Aspects informatiques.

Déterminer l'espérance, la variance d'une variable à densité passe par la recherche d'une primitive et le calcul d'une intégrale.

◦ Déterminer une primitive sous maxima.

La primitive de $x \mapsto x^2$ sous maxima¹ : `integrate(x^2,x)`

◦ Calculer une intégrale sous R².

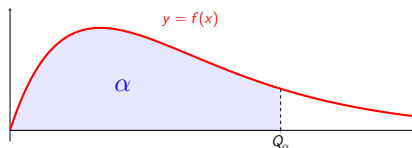
$\int_1^3 x^2 dx$ sous R : `f = function(x){x^2}` ; `integrate(f,lower=1,upper=3)`

¹<https://maxima.sourceforge.io/download.html>

²<https://cran.r-project.org/>

Quantiles

X : une variable aléatoire de densité f , de fonction de répartition F .



Le quantile d'ordre α ($0 < \alpha < 1$) de la loi de X est la valeur Q_α à laquelle X est inférieure ou égale avec probabilité α :

$$\mathbb{P}(X \leq Q_\alpha) = \alpha \quad (1)$$

Comment déterminer Q_α ?

Si on dispose :

- de la fonction de répartition F , on résout : $F(Q_\alpha) = \alpha$
- de la fonction de densité f , on résout : $\int_{-\infty}^{Q_\alpha} f(x)dx = \alpha$
- de R, on utilise une fonction quantile : `qunif`, `qexp`, `qnorm`, `qt`, `qchisq`, etc

La définition d'un quantile selon (1) est un peu restrictive, mais elle suffit à déterminer les quantiles de la plupart des lois usuelles associées à une densité

Exemple

La quantité de pluie (en ul/m^2) sur Dublin en une journée est une variable aléatoire X admettant pour fonction de densité : $f(x) = 2x$ si $0 \leq x \leq 1$; 0 sinon.

◦ Quelle est la fonction de répartition de X ?

$$F(x) = x^2 \text{ si } 0 \leq x < 1 ; 0 \text{ si } x < 0 ; 1 \text{ si } x \geq 1.$$

◦ Avec quelle probabilité tombe-t-il de 0,2 à 0,5 ul/m^2 sur Dublin en un jour ?

$$(a) \mathbb{P}(0,2 \leq X \leq 0,5) = \int_{0,2}^{0,5} 2x dx = [x^2]_{0,2}^{0,5} = 0,21.$$

$$(b) \mathbb{P}(0,2 \leq X \leq 0,5) = F(0,5) - F(0,2) = 0,5^2 - 0,2^2 = 0,21.$$

◦ Avec quelle probabilité tombe-t-il au moins 0,7 ul/m^2 sur Dublin en un jour ?

$$(a) \mathbb{P}(X \geq 0,7) = \mathbb{P}(0,7 \leq X \leq 1) = \int_{0,7}^1 2x dx = [x^2]_{0,7}^1 = 0,51.$$

◦ Quelle quantité de pluie attend-on sur Dublin, chaque jour ?

$$E(X) = \int_0^1 x \times 2x dx = \int_0^1 2x^2 dx = [2x^3/3]_0^1 = 2/3.$$

◦ Quelle est la variance de la quantité quotidienne de pluie sur Dublin ?

$$E(X^2) = \int_0^1 x^2 \times 2x dx = \int_0^1 2x^3 dx = [2x^4/4]_0^1 = 1/2$$

$$V(X) = E(X^2) - E(X)^2 = 1/2 - (2/3)^2 = 1/18.$$

Outline

- Généralités
- Lois continues usuelles

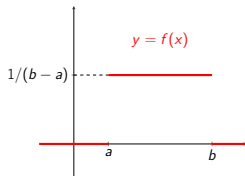
Loi uniforme

Une v.a.c. X est distribuée selon une loi uniforme sur $[a, b]$ ($-\infty < a < b < +\infty$) si elle a pour densité la fonction f définie par : $f(x) = 1/(b - a)$ si $a \leq x \leq b$; 0 sinon.

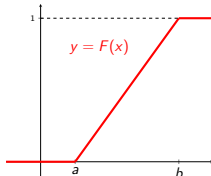
Notation. $X \sim \mathcal{U}_{[a,b]}$

Espérance et variance. $E(X) = (a + b)/2$; $V(X) = (b - a)^2/12$.

Fonction de répartition. $F(x) = (x - a)/(b - a)$ si $a \leq x < b$; 0 si $x < a$; 1 si $x \geq b$.



Fonction de densité de $\mathcal{U}_{[a,b]}$



Fonction de répartition de $\mathcal{U}_{[a,b]}$

Contexte. La loi uniforme modélise les expériences dans lesquelles une quantité continue tombe dans un intervalle avec une probabilité proportionnelle à la longueur de l'intervalle.

Exercice

Le temps d'attente du tram est uniformément réparti entre 0 et 5 mins.

(a) Déterminez les fonctions de probabilité et de répartition du temps d'attente.

X : temps d'attente du tram. Densité de X : $f(x) = 1/5$ si $0 \leq x \leq 5$; 0 sinon.

Fct de répartition : $F(x) = (x - 0)/(5 - 0)$ si $0 \leq x < 5$; 0 si $x < 0$; 1 si $x \geq 5$.

(b) Avec quelle probabilité attend-on le tram au moins trois minutes ?

$$\mathbb{P}(X \geq 3) = 1 - \mathbb{P}(X < 3) = 1 - \mathbb{P}(X \leq 3) = 1 - F(3) = 1 - 3/5 = 0,4.$$

(c) Combien de temps attend-on le tram en moyenne ?

$$E(X) = (0 + 5)/2 = 2,5.$$

(d) Quelle est la variance du temps d'attente ?

$$V(X) = (5 - 0)^2/12 = 25/12.$$

(e) Déterminez et interprétez le quantile d'ordre 0,8 du temps d'attente¹.

$$\mathbb{P}(X \leq Q) = 0,8 ; F(Q) = 0,8 ; (Q - 0)/(5 - 0) = 0,8 ; Q = 5 \times 0,8 = 4.$$

Le quantile d'ordre 0,8 de $\mathcal{U}_{[0;5]}$ vaut 4 : 80% des usagers attendent le tram au plus quatre minutes.

¹ la valeur à laquelle le temps d'attente est inférieur ou égal dans 80% des cas

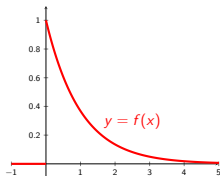
Loi exponentielle

Une v.a.c. X est distribuée selon une loi exponentielle de paramètre $\lambda > 0$ si elle admet pour densité la fonction f définie par : $f(x) = \lambda e^{-\lambda x}$ si $x \geq 0$; 0 sinon.

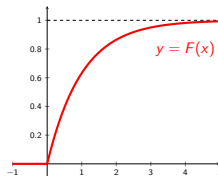
Notation. $X \sim \mathcal{E}(\lambda)$

Espérance et variance. $E(X) = 1/\lambda$; $V(X) = 1/\lambda^2$.

Fonction de répartition. $F(x) = 1 - e^{-\lambda x}$ si $x \geq 0$; 0 sinon.



Fonction de densité de $\mathcal{E}(\lambda)$



Fonction de répartition de $\mathcal{E}(\lambda)$

Contexte. La loi exponentielle modélise les durées de vie sans mémoire : la probabilité de mourir après n années est à la même à tous les âges de la vie.

Exercice

La durée de vie (en années) d'une entreprise est distribuée selon $\mathcal{E}(0,1)$.

(a) Déterminez les fonctions de probabilité et de répartition d'une telle durée de vie.

X : durée de vie d'une entreprise. Densité de X : $f(x) = 0,1e^{-0,1x}$ si $x \geq 0$; 0 sinon.

Fct de répartition : $F(x) = 1 - e^{-0,1x}$ si $x \geq 0$; 0 sinon.

(b) Quelle proportion des entreprises durent-elles au moins deux ans ?

$$\mathbb{P}(X \geq 2) = 1 - \mathbb{P}(X < 2) = 1 - \mathbb{P}(X \leq 2) = 1 - F(2) = 1 - (1 - e^{-0,1 \times 2}) \approx 0,18.$$

NB. Sous R, $\mathbb{P}(\mathcal{E}(0,1) \leq 2)$ se calcule par la commande : `pexp(2,0.1)`

(c) Quelle est la durée de vie attendue d'une entreprise ?

$E(X) = 1/0,1 = 10$. Une entreprise disparaît au bout de dix ans en moyenne.

(e) Déterminez et interprétez le quantile d'ordre 0,8 de la durée de vie¹.

$$\mathbb{P}(X \leq Q) = 0,8 ; F(Q) = 0,8 ; 1 - e^{-0,1Q} = 0,8 ; Q = -\ln(0,2)/0,1 \approx 16,1.$$

NB. Sous R, le quantile d'ordre 0,8 de $\mathcal{E}(0,1)$ se calcule par : `qexp(0.8,0.1)`

Le quantile d'ordre 0,8 de $\mathcal{E}(0,1)$ vaut 16,1 : 80% des entreprises ont une durée de vie inférieure ou égale à 16,1 ans.

¹ la valeur à laquelle la durée de vie est inférieure ou égale dans 80% des cas

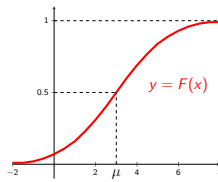
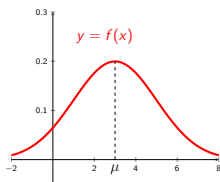
Loi normale

Une v.a.c. X est distribuée selon une loi normale¹ de paramètres : μ et σ^2 ($\sigma^2 > 0$) si elle admet pour densité : $f(x) = [\exp(-0,5(x - \mu)^2/\sigma^2)]/[\sigma\sqrt{2\pi}]$.

Notation. $X \sim \mathcal{N}(\mu, \sigma^2)$

Espérance et variance. $E(X) = \mu$; $V(X) = \sigma^2$.

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



Fonction de densité de $\mathcal{N}(\mu, \sigma^2)$ Fct de répartition de $\mathcal{N}(\mu, \sigma^2)$

Contexte. La loi normale occupe une place centrale en statistique en raison du Théorème Central Limit : *une moyenne empirique est approximativement gaussienne quelle que soit la population dont est issu l'échantillon.*

¹ on dit alors que la variable est normale ou gaussienne

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

La densité de X est : $f(x) = \exp(-x^2/2)/\sqrt{2\pi}$ et sa fonction de répartition : $\phi(x) = \int_{-\infty}^x f(t)dt$.

○ Comment calculer $\mathbb{P}(X \leq 1,78)$?

- avec la fonction de répartition de X : $\mathbb{P}(X \leq 1,78) = \phi(1,78)$; Hélas, on ne connaît pas ϕ explicitement.

- avec les valeurs de ϕ dans une table de la loi normale centrée réduite¹.

- avec R : `pnorm(1.78)`

Ainsi $\mathbb{P}(\mathcal{N}(0, 1) \leq 1,78) \approx 0,96$.

○ Comment calculer $z_{0,95}$: le quantile d'ordre 0,95 de X ?

- avec la réciproque de la fonction de répartition : $z_{0,95} = \phi^{-1}(0,95)$; Hélas, on ne connaît pas plus explicitement ϕ^{-1} que ϕ .

- avec les valeurs de ϕ dans une table de la loi normale centrée réduite¹.

- avec R : `qnorm(0.95)`

Ainsi, $z_{0,95} \approx 1,64$

¹<https://onlinepubs.trb.org/onlinepubs/nchrp/cd-22/manual/v2appendixc.pdf>

Calculs sur $X \sim \mathcal{N}(\mu, \sigma^2)$

X est une variable aléatoire distribuée selon $\mathcal{N}(3, 4)$.

o Comment calculer $\mathbb{P}(X \leq 5)$?

- en se ramenant à une variable normale centrée réduite :

$$\mathbb{P}(X \leq 5) = \mathbb{P}((X - 3)/\sqrt{4} \leq (5 - 3)/\sqrt{4}) = \mathbb{P}(\mathcal{N}(0, 1) \leq 1) = \Phi(1) \approx 0,84.$$

- avec R : `pnorm(5,3,sqrt(4))`

Ainsi $\mathbb{P}(\mathcal{N}(3, 4) \leq 5) \approx 0,84$.

o Comment calculer le quantile d'ordre 0,95 de X ?

- avec la formule : $Q_\alpha(\mathcal{N}(\mu, \sigma^2)) = \mu + z_\alpha \times \sigma$.

$$\text{Ainsi, } Q_{0,95}(\mathcal{N}(3, 4)) = 3 + z_{0,95} \times \sqrt{4} \approx 3 + 1,64 \times 2 = 6,28.$$

- avec R : `qnorm(0.95,3,sqrt(4))`

Probabilités gaussiennes

Rappel. $\phi(x) = \int_{-\infty}^x [\exp(-t^2/2)/\sqrt{2\pi}]dt$ est la fonction de répartition d'une variable normale centrée réduite.

Propriétés de ϕ .

$$(p_1) \quad \phi(-x) = 1 - \phi(x)$$

$$(p_2) \quad \phi(x) - \phi(-x) = 2\phi(x) - 1$$

Justification. La densité de $\mathcal{N}(0, 1)$ est paire.

Règles de calcul usuelles. $X \sim \mathcal{N}(\mu, \sigma^2)$ et $a < b$

$$(r_1) \quad \mathbb{P}(X \leq b) = \phi((b - \mu)/\sigma)$$

$$(r_2) \quad \mathbb{P}(X > a) = 1 - \phi((a - \mu)/\sigma)$$

$$(r_3) \quad \mathbb{P}(a < X \leq b) = \phi((b - \mu)/\sigma) - \phi((a - \mu)/\sigma)$$

Justification. Si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors $(X - \mu)/\sigma \sim \mathcal{N}(0, 1)$.

Applications

◦ Calculer $\mathbb{P}(\mathcal{N}(3, 4) \leq 1)$.

$$\mathbb{P}(\mathcal{N}(3, 4) \leq 1) \stackrel{r_1}{=} \phi((1 - 3)/\sqrt{4}) = \phi(-1) \stackrel{p_1}{=} 1 - \phi(1) \approx 0,16.$$

◦ Calculer $\mathbb{P}(\mathcal{N}(3, 4) > 2)$.

$$\mathbb{P}(\mathcal{N}(3, 4) > 2) \stackrel{r_2}{=} 1 - \phi((2 - 3)/\sqrt{4}) = 1 - \phi(-0,5) \stackrel{p_1}{=} \phi(0,5) \approx 0,69.$$

◦ Calculer $\mathbb{P}(-1 < \mathcal{N}(3, 4) \leq 7)$.

$$\mathbb{P}(-1 < \mathcal{N}(3, 4) \leq 7) \stackrel{r_3}{=} \phi((7 - 3)/\sqrt{4}) - \phi((-1 - 3)/\sqrt{4}) = \phi(2) - \phi(-2).$$

$$\stackrel{p_2}{=} 2 \times \phi(2) - 1 \approx 0,95.$$

Loi de Student

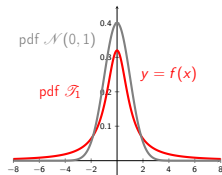
Une v.a.c. X est distribuée selon une loi de Student à ν ($\nu > 0$) degrés de liberté si elle admet pour densité : $f(x) = \gamma \cdot (1 + x^2/\nu)^{-(\nu+1)/2}$ ($\gamma > 0$ tq $\int_{-\infty}^{+\infty} f(x)dx = 1$).

Notation. $X \sim \mathcal{T}_\nu$

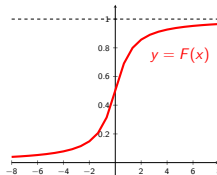
Espérance et variance. $E(X) = 0$ si $\nu > 1$; NA sinon

$V(X) = \nu/(\nu - 2)$ si $\nu > 2$; $+\infty$ si $1 < \nu \leq 2$; NA sinon.

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



Fonction de densité de $\mathcal{T}_\nu$



Fct de répartition de $\mathcal{T}_\nu$

Contexte. La loi de Student est une distribution à queues lourdes permettant de modéliser des phénomènes avec des valeurs extrêmes. La loi de Student intervient aussi dans les tests d'hypothèses et les intervalles de confiance.

Calculs sur $\mathcal{T}_\nu$

Supposons $X \sim \mathcal{T}_3$

◦ Comment calculer $\mathbb{P}(X \leq 1,78)$?

- avec la fonction de répartition de X ; on ne la connaît pas explicitement.
- avec les valeurs de la fct de répartition dans une table de la loi de Student¹.
- avec R : `pt(1.78,3)`

Ainsi $\mathbb{P}(\mathcal{T}_3 \leq 1,78) \approx 0,91$.

◦ Comment calculer $t_{3;0,95}$: le quantile d'ordre 0,95 de $\mathcal{T}_3$?

- avec la réciproque de la fct de répartition ; on ne la connaît pas explicitement.
- avec les valeurs de la fct de répartition dans une table de la loi de Student¹
- avec R : `qt(0.95,3)`

Ainsi, $t_{3;0,95} \approx 2,35$

Remarque. Comme la fonction de densité de $\mathcal{T}_\nu$ est paire, quel que soit x :

$$\circ \mathbb{P}(\mathcal{T}_\nu \leq x) = 1 - \mathbb{P}(\mathcal{T}_\nu \leq -x)$$

$$\circ \mathbb{P}(\mathcal{T}_\nu \leq x) - \mathbb{P}(\mathcal{T}_\nu \leq -x) = 2 \times \mathbb{P}(\mathcal{T}_\nu \leq x) - 1$$

¹<https://onlinepubs.trb.org/onlinepubs/nchrp/cd-22/manual/v2appendixc.pdf>

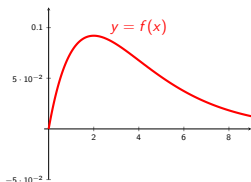
Loi de χ^2

Une v.a.c. X est distribuée selon une loi de χ^2 à ν ($\nu > 0$) degrés de liberté si elle admet pour densité : $f(x) = \gamma \cdot x^{\nu/2-1} e^{-x/2}$ si $x \geq 0$; 0 sinon ($\gamma > 0$ tq $\int_{-\infty}^{+\infty} f(x)dx = 1$).

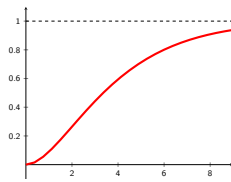
Notation. $X \sim \chi^2_\nu$

Espérance et variance. $E(X) = \nu$; $V(X) = 2\nu$.

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



Fonction de densité de χ^2_ν



Fct de répartition de χ^2_ν

Contexte. La loi de χ^2 donne la distribution de statistiques intervenant dans des tests d'hypothèses et dans les intervalles de confiance portant sur une variance.

Calculs sur χ^2_ν

Supposons $X \sim \chi^2_3$

◦ Comment calculer $\mathbb{P}(X \leq 1,78)$?

- avec la fonction de répartition de X ; on ne la connaît pas explicitement.
- avec les valeurs de la fct de répartition dans une table de la loi de χ^2 .¹
- avec R : `pchisq(1.78,3)`

Ainsi $\mathbb{P}(\chi^2_3 \leq 1,78) \approx 0,38$.

◦ Comment calculer $\chi^2_{3;0,95}$: le quantile d'ordre 0,95 de χ^2_3 ?

- avec la réciproque de la fct de répartition ; on ne la connaît pas explicitement.
- avec les valeurs de la fct de répartition dans une table de la loi de χ^2 .¹
- avec R : `qchisq(0.95,3)`

Ainsi, $\chi^2_{3;0,95} \approx 7,8$

¹<https://onlinepubs.trb.org/onlinepubs/nchrp/cd-22/manual/v2appendixc.pdf>