Projet de statistiques

Maxime Peterlin - Gabriel Vermeulen

ENSEIRB-MATMECA, Bordeaux

28, avril 2014

Table des matières

1 Autour des variables aléatoires gaussiennes **3**

1.1 Variables aléatoires gaussiennes réelles 3

1.1.1 Densité de probabilité d'une variable aléatoire gaussienne de de moyenne μ et de variance σ^2 3

1.1.2 Histogramme de 100 réalisations de la loi $\mathcal{N}_{\mathbb{R}}(5,1)$ 4

1.1.3 Histogramme de 1000 réalisations de la loi $\mathcal{N}_{\mathbb{R}}(5,1)$ 4

1.2 Variables aléatoires gaussiennes vectorielles 4

1.2.1 Fonction caractéristique de la variable $t^T X$ 4

1.2.2 Montrons que les $X_1, ..., X_n$ sont indépentantes 4

1.2.3 Densité de probabilité de la loi $\mathcal{N}_{\mathbb{R}^n}(0, I)$ 4

1.2.4 Densité de probabilité de la loi $\mathcal{N}_{\mathbb{R}^n}(\mu, \Gamma)$ 5

1.2.5 Montrons que les composantes d'un vecteur gaussien sont indépendantes si et seulement si sa matrice de covariance est diagonale 5

1.2.6 Tracé de la densité de la loi $\mathcal{N}_{\mathbb{R}^2}(\mu, \Gamma)$ 5

1.2.7 Histogramme de 1000 réalisations de la loi $\mathcal{N}_{\mathbb{R}^2}(\mu, \Gamma)$ . . . 5

1.2.8 Montrons que $X \sim \mathcal{N}_{\mathbb{R}^d}(0, I) \Rightarrow UX \sim \mathcal{N}_{\mathbb{R}^d}(0, I), \forall U \in \mathbb{O}_d(\mathbb{R})$ 5

1.2.9 Histogramme de 1000 réalisations du vecteur UX 6

1.2.10 Démonstration et vérification expérimentale du caractère gaussien d'un vecteur donné 6

1.3 Variables aléatoires gaussiennes complexes 7

1.3.1 Variance et densité de probabilité de $Z \sim \mathcal{N}_{\mathbb{C}}(\mu, \sigma^2)$ . . . 7

1.3.2 Histogramme de 1000 réalisations de la loi $\mathcal{N}_{\mathbb{C}}(0,1)$ 7

1.4 Théorème de la limite centrale 7

1.4.1 Loi de W_n lorsque $X_k \sim \mathcal{N}_{\mathbb{R}}(\mu, \sigma^2)$ 7

1.4.2 Histogramme de 1000 réalisations de la loi W_n lorsque $X_k \sim \mathcal{N}_{\mathbb{R}}(\mu, \sigma^2)$ 8

1.4.3 Histogramme de 1000 réalisations de la loi W_n lorsque $X_k \sim \mathcal{U}([0,1])$ 8

1.4.4 Histogramme de 1000 réalisations de la loi W_n lorsque $X_k \sim \mathcal{B}(\frac{1}{2})$ 8

1.5 Loi du χ^2 8

1.5.1 Densité de probabilité d'une variable suivant une loi $\chi^2(n)$ 8

1.5.2 Moyenne et variance de la loi $\chi^2(n)$ 9
1.5.3 Tracé de la densité de probabilité de la loi $\chi^2(n)$ 9

Chapitre 1

Autour des variables aléatoires gaussiennes

1.1 Variables aléatoires gaussiennes réelles

1.1.1 Densité de probabilité d'une variable aléatoire gaussienne de de moyenne μ et de variance σ^2

La densité de probabilité d'une loi normale de moyenne 0 et de variance 1 est :

$$f(x) = \frac{1}{\sqrt{2\pi}} \cdot exp(-\frac{x^2}{2})$$

Pour obtenir la densité de probabilité d'une variable aléatoire gaussienne de de moyenne μ et de variance σ^2, on effectue le changement de variable $y = \frac{x-\mu}{\sigma}$:

$$\int f(x)dx = \int \frac{1}{\sqrt{2\pi}} \cdot exp(-\frac{x^2}{2})dx = \int \frac{1}{\sqrt{2\pi}} \cdot exp(-\frac{(y-\mu)^2}{2\cdot\sigma^2})\frac{dy}{\sigma} = \int f(y)dy$$

Ainsi, on a la densité de probabilité $f(y)$ recherchée :

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} \cdot exp(-\frac{(y-\mu)^2}{2\cdot\sigma^2})$$

1.1.2 Histogramme de 100 réalisations de la loi $\mathcal{N}_{\mathbb{R}}(5,1)$

1.1.3 Histogramme de 1000 réalisations de la loi $\mathcal{N}_{\mathbb{R}}(5,1)$

1.2 Variables aléatoires gaussiennes vectorielles

1.2.1 Fonction caractéristique de la variable $t^T X$

On a

$$\Phi_{t^T X}(u) = \mathbb{E}[e^{iut^T X}] = \mathbb{E}[e^{iu \sum\limits_{k=1}^{n} t_k X_k}]$$

1.2.2 Montrons que les $X_1, ..., X_n$ sont indépentantes

En se servant du lemme donné, si

$$\mathbb{E}[e^{i \sum\limits_{k=1}^{n} t_k X_k}] = \prod_{k=1}^{n} \mathbb{E}[e^{it_k X_k}]$$

alors les variables aléatoires $X_1, ..., X_n$ seront indépendantes. Soit $X \sim \mathcal{N}_{\mathbb{R}^{\mathbb{N}}}(\mu, \sigma^2)$ et $Y \sim \mathcal{N}_{\mathbb{R}^{\mathbb{N}}}(O, I)$, on a la relation suivante entre X et Y :

$$X = m + \sigma Y$$

On en déduit que

$$\Phi_X(u) = \mathbb{E}[e^{iuX}] = \mathbb{E}[e^{iu(m+\sigma Y)}] = e^{ium}\mathbb{E}[e^{iu\sigma Y}] = e^{ium - \frac{\sigma^2 u^2}{2}}$$

Comme $X \sim \mathcal{N}_{\mathbb{R}^{\mathbb{N}}}(O, I)$, alors $t^T X \sim \mathcal{N}_{\mathbb{R}^{\mathbb{N}}}(O, \sum\limits_{k=1}^{n} t_k^2)$.

Ainsi $\Phi_{t^T X}(u) = e^{-\frac{u^2}{2} \sum\limits_{k=1}^{n} t_k^2}$ et $\Phi_{t^T X}(1) = e^{-\frac{1}{2} \sum\limits_{k=1}^{n} t_k^2}$.

De plus, $\Phi_{X_k}(t_k) = e^{-\frac{t_k^2}{2}} = \mathbb{E}[e^{it_k X_k}]$, donc $\prod\limits_{k=1}^{n} \Phi_{X_k}(t_k) = \Phi_{t^T X}(1) = \mathbb{E}[e^{i \sum\limits_{k=1}^{n} t_k X_k}]$.

On a bien montré que

$$\mathbb{E}[e^{i \sum\limits_{k=1}^{n} t_k X_k}] = \prod_{k=1}^{n} \mathbb{E}[e^{it_k X_k}]$$

Donc les variables aléatoires $X_1, ..., X_n$ sont indépendantes.

1.2.3 Densité de probabilité de la loi $\mathcal{N}_{\mathbb{R}^n}(0, I)$

On a $\forall k$, $X_k \sim \mathcal{N}_{\mathbb{R}}(0,1)$ indépendantes, ainsi

$$f(X) = f(X_1, ...X_n) = \prod_{k=1}^{n} f_k(X_k) = (\frac{1}{\sqrt{2\pi}})^n e^{-\frac{1}{2} \sum\limits_{k=1}^{n} x_k}$$

1.2.4 Densité de probabilité de la loi $\mathcal{N}_{\mathbb{R}^n}(\mu, \Gamma)$

1.2.5 Montrons que les composantes d'un vecteur gaussien sont indépendantes si et seulement si sa matrice de covariance est diagonale

Si les composantes d'un vecteur gaussien sont indépendantes, alors $\forall i \neq j$ on a $COV(X_i, X_j) = 0$, ainsi, la matrice de covariance a tous ses éléments non diagonaux qui sont nuls. Cette dernière est donc diagonale.
Réciproquement, si la matrice de covariance est diagonale, alors :

$$f_k(X_k) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{1}{2}\sum\limits_{k=1}^{n}\frac{(x_1-\mu_1)^2}{\sigma_k^2}}$$

et

$$f(X) = \frac{1}{\sqrt{2\pi}^n \prod\limits_{k=1}^{n}\sigma_k} e^{-\frac{1}{2}(x-\mu)^T \begin{bmatrix} \sigma_1^2 & ... & 0 \\ \cdot & \cdot & \cdot \\ 0 & ... & \sigma_n^2 \end{bmatrix}^{-1} (x-\mu)} = \frac{1}{\sqrt{2\pi}^n \prod\limits_{k=1}^{n}\sigma_k} e^{-\frac{1}{2}\sum\limits_{k=1}^{n}\frac{(x_k-\mu_k)}{\sigma_k^2}}$$

On a bien $\prod f_k(X_k) = f(X)$, les variables sont donc indépendantes.

1.2.6 Tracé de la densité de la loi $\mathcal{N}_{\mathbb{R}^2}(\mu, \Gamma)$

1.2.7 Histogramme de 1000 réalisations de la loi $\mathcal{N}_{\mathbb{R}^2}(\mu, \Gamma)$

1.2.8 Montrons que $X \sim \mathcal{N}_{\mathbb{R}^d}(0, I) \Rightarrow UX \sim \mathcal{N}_{\mathbb{R}^d}(0, I), \forall U \in \mathbb{O}_d(\mathbb{R})$

On a $UX = [a_1, ..., a_d]$, avec $\forall k, a_k = \sum\limits_{i=1}^{d} u_{ki}X_i$.
Calculons l'espérance de ce vecteur :

$$\mathbb{E}[UX] = \mathbb{E}\left[[\sum_{i=1}^{d} u_{1i}X_i, ..., \sum_{i=1}^{d} u_{di}X_i]\right] = \left[\sum_{i=1}^{d} u_{1i}\mathbb{E}[X_i], ..., \sum_{i=1}^{d} u_{di}\mathbb{E}[X_i]\right] = 0$$

car $\forall i, \mathbb{E}[X_i] = 0$.

Calculons à présent sa variance

$$Var[UX] = \mathbb{E}[(UX)(UX)^T] = \mathbb{E}[UXX^TU^T] = U\mathbb{E}[XX^T]U^T$$

Nous savons également que

$$Var[X] = \mathbb{E}[XX^T] - \mathbb{E}^2[X] = \mathbb{E}[XX^T] = I$$

Ainsi, $Var[X] = UU^T = I$, car $U \in \mathbb{O}_d(\mathbb{R})$.
Nous avons donc montré que $X \sim \mathcal{N}_{\mathbb{R}^d}(0, I) \Rightarrow UX \sim \mathcal{N}_{\mathbb{R}^d}(0, I), \forall U \in \mathbb{O}_d(\mathbb{R})$.

1.2.9 Histogramme de 1000 réalisations du vecteur UX

1.2.10 Démonstration et vérification expérimentale du caractère gaussien d'un vecteur donné

On a $X \sim \mathcal{N}_{\mathbb{R}}(0,1)$.
On pose $\mathbf{X} = [X_1, X_2]^T$, avec $X_1 = X et X_2 = ZX$. Z est une variable aléatoire indépendante de X de loi uniforme sur l'intervalle $\{-1, 1\}$

On veut montrer que X_1 et X_2 suivent des lois gaussiennes.
Il est évident que c'est le cas pour X_1, montrons le pour X_2. La densité de probabilité que suit Z est

$$f_Z(x) = \frac{1}{2}\delta(x+1) + \frac{1}{2}\delta(x-1)$$

La densité de probabilité que suit X est

$$f_X(x) = \frac{1}{\sqrt{2\pi}} exp(-\frac{x^2}{2})$$

La densité de probabilité que suit ZX est la convolution de leur densité respective.

$$f_{ZX}(x) = (f_X * f_Z)(x) = [f * (\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1)](x) = \frac{1}{2}f(x+1) + \frac{1}{2}f(x-1)$$

La fonction caractéristique d'une variable $X \sim \mathcal{N}_{\mathbb{R}}(0,1)$ est

$$\Phi_X(u) = \mathbb{E}[e^{iuX}] = e^{-\frac{u^2}{2}}$$

Ainsi

$$\begin{aligned}\Phi_{ZX}(u) = \mathbb{E}[e^{iuZX}] &= \int_{\mathbb{R}} f_{ZX}(y)e^{ity}dy = \int_{\mathbb{R}} (\frac{1}{2}f_X(x+1) + \frac{1}{2}f_X(x-1))e^{ity}dy \\ &= \int_{\mathbb{R}} \frac{1}{2}f(x+1)e^{ity}dy + \int_{\mathbb{R}} \frac{1}{2}f(x-1)e^{ity}dy \\ &= \frac{1}{2}e^{-\frac{u^2}{2}} + \frac{1}{2}e^{-\frac{u^2}{2}} = e^{-\frac{u^2}{2}}\end{aligned}$$

car f_X suit une loi $\mathcal{N}_{\mathbb{R}}(0,1)$.
Nous avons donc prouvé que $X_2 \sim \mathcal{N}_{\mathbb{R}}(0,1)$.
A présent, nous allons voir si le vecteur $\mathbf{X}$ est gaussien ou non.
Si le vecteur $\mathbf{X}$ était gaussien, il aurait la fonction caractéristique suivante :

$$\Phi_{\mathbf{X}}(u) = e^{-\frac{u^2}{2}}$$

Calculons sa fonction caractéristique.

$$\Phi_{\mathbf{X}}(u) = \mathbb{E}(e^{i<u,X>}) = \int_{\mathbb{R}}\int_{\mathbb{R}} f_{\mathbf{X}}(x,z)e^{iu_1x+iu_2xz}dxdz$$

X et ZX n'étant pas indépendant, on ne peut retrouver la bonne fonction caractéristique.
Le vecteur $\mathbf{X}$ n'est donc pas gaussien.

1.3 Variables aléatoires gaussiennes complexes

1.3.1 Variance et densité de probabilité de $Z \sim \mathcal{N}_{\mathbb{C}}(\mu, \sigma^2)$

Calculons la variance de la variable aléatoire $Z = X + iY$.

$$\begin{aligned} V[Z] &= \mathbb{E}\left[|Z - \mathbb{E}[Z]|^2\right] = \mathbb{E}\left[|Z|^2 + |\mathbb{E}[Z]|^2 - 2\Re(Z\overline{\mathbb{E}[Z]})\right] \\ &= \mathbb{E}[|Z|^2] - |\mathbb{E}[Z]|^2 = \mathbb{E}[|X+iY|^2] - |\mathbb{E}[X+iY]|^2 \\ &= \mathbb{E}[X^2 + Y^2] - (\mathbb{E}^2[X] + \mathbb{E}^2[Y]) = Var[X] + Var[Y] = \sigma^2 \end{aligned}$$

Calculons à présent la densité de probabilité de Z.
On sait que X et Y sont indépendantes, ainsi

$$\begin{aligned} f_{XY}(X,Y) &= f_X(X) f_Y(Y) \\ &= \frac{1}{\sqrt{2\pi}\frac{\sigma}{\sqrt{2}}} exp(-\frac{1}{2}\frac{(x - \Re(\mu))^2}{\frac{\sigma^2}{2}}) \cdot \frac{1}{\sqrt{2\pi}\frac{\sigma}{\sqrt{2}}} exp(-\frac{1}{2}\frac{(y - \Im(\mu))^2}{\frac{\sigma^2}{2}}) \\ &= \frac{1}{\sigma^2 \pi} e^{-\frac{|z-\mu|^2}{\sigma^2}} \end{aligned}$$

1.3.2 Histogramme de 1000 réalisations de la loi $\mathcal{N}_{\mathbb{C}}(0,1)$

1.4 Théorème de la limite centrale

1.4.1 Loi de W_n lorsque $X_k \sim \mathcal{N}_{\mathbb{R}}(\mu, \sigma^2)$

On sait que $\forall k, X_k \sim \mathcal{N}_{\mathbb{R}}(0,1)$, de plus, on a

$$W_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{X_k - \mu}{\sigma}$$

Nous cherchons la loi de W_n en sachant que $\forall k, X_k$ suit une loi normale.
Posons

$$Y_k = \frac{X_k - \mu}{\sqrt{n}\sigma}$$

Commençons par calculer l'espérance de Y_k

$$\mathbb{E}[Y_k] = \mathbb{E}\left[\frac{X_k - \mu}{\sqrt{n}\sigma}\right] = \frac{\mathbb{E}[X_k] - \mu}{\sqrt{n}\sigma} = 0$$

Calculons à présent la variance de Y_k

$$Var[Y_k] = Var\left[\frac{X_k - \mu}{\sqrt{n}\sigma}\right] = \frac{1}{n\sigma^2} Var[X_k - \mu] = \frac{1}{n\sigma^2}\sigma^2 = \frac{1}{n}$$

Nous avons donc $Y_k \sim \mathcal{N}_{\mathbb{R}}(0, \frac{1}{n})$ Ainsi

$$W_n = \sum_{k=1}^{n} Y_k \sim \mathcal{N}_{\mathbb{R}}(0, \sqrt{\sum_{k=1}^{n} \frac{1}{n}} = 1)$$

1.4.2 Histogramme de 1000 réalisations de la loi W_n lorsque $X_k \sim \mathcal{N}_{\mathbb{R}}(\mu, \sigma^2)$

1.4.3 Histogramme de 1000 réalisations de la loi W_n lorsque $X_k \sim \mathcal{U}([0,1])$

Calcul de l'espérance et de la variance de $X_k, \forall k$

$$\mathbb{E}[X_k] = \int_0^1 x dx = \frac{1}{2}$$

$$Var[X_k] = \mathbb{E}[X_k^2] - \mathbb{E}^2[X_k] = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

1.4.4 Histogramme de 1000 réalisations de la loi W_n lorsque $X_k \sim \mathcal{B}(\frac{1}{2})$

Calcul de l'espérance et de la variance de $X_k, \forall k$

$$\mathbb{E}[X_k] = p = \frac{1}{2}$$

$$Var[X_k] = p(1-p) = \frac{1}{2}$$

1.5 Loi du χ^2

1.5.1 Densité de probabilité d'une variable suivant une loi $\chi^2(n)$

Montrons par récurrence que la densité de probabilité de $Y \sim \chi^2(n)$ est

$$f_n(x) = \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \mathbb{1}_{\mathbb{R}^+}(x)$$

Tout d'abord, montrons le pour $n = 1$. On a $Y = X^2$.
Pour toutes fonctions ϕ bornée, on a :

$$\begin{aligned}
\mathbb{E}[\phi(y)] = \mathbb{E}[\phi(x^2)] &= \int_{\mathbb{R}} \phi(x^2) f_X(x) dx \\
&= \int_{-\infty}^{0} \phi(x^2) f_X(x) dx + \int_0^{\infty} \phi(x^2) f_X(x) dx \\
&= \int_0^{\infty} \phi(x^2) f_X(-x) dx + \int_0^{\infty} \phi(x^2) f_X(x) dx
\end{aligned}$$

Posons le changement de variable $x = \sqrt{u}$, $dx = \frac{du}{2\sqrt{u}}$.

$$\begin{aligned}\mathbb{E}[\phi(y)] &= \int_0^\infty \phi(u) f_X(-\sqrt{u}) \frac{du}{2\sqrt{u}} + \int_0^\infty \phi(u) f_X(\sqrt{u}) \frac{du}{2\sqrt{u}} \\ &= \int_0^\infty \phi(u)[f_X(-\sqrt{u}) + f_X(\sqrt{u})] \frac{du}{2\sqrt{u}}\end{aligned}$$

Ainsi, pour $Y = X^2$, la densité de probabilité s'exprime de la façon suivante :

$$f_Y(y) = \frac{1}{2\sqrt{y}}[f_X(-\sqrt{y}) + f_X(\sqrt{y})]\mathbb{1}_{\mathbb{R}^+}(x)$$

On en déduit donc que

$$f_1(x) = \frac{x^{-\frac{1}{2}}}{2\sqrt{2\pi}}[e^{-\frac{x}{2}} + e^{-\frac{x}{2}}]\mathbb{1}_{\mathbb{R}^+}(x) = \frac{x^{-\frac{1}{2}} e^{-\frac{x}{2}}}{\sqrt{2\pi}} \mathbb{1}_{\mathbb{R}^+}(x)$$

Supposons que la propriété soit vraie au rang n et montrons le au rang $n+1$:

$$\begin{aligned}f_{n+1}(x) = f_n(x) * f_1(x) &= \int_{\mathbb{R}} f_n(t) f_1(x-t) dt \\ &= \frac{1}{2^{\frac{n+1}{2}}\Gamma(\frac{n}{2})\Gamma(\frac{1}{2})} e^{-\frac{x}{2}} \mathbb{1}_{\mathbb{R}^+}(x) \int_0^x t^{\frac{n}{2}-1}(x-t)^{-\frac{1}{2}} dt\end{aligned}$$

Posons le changement de variable $t = xu$, $dt = xdu$.

$$\begin{aligned}f_{n+1}(x) &= \frac{1}{2^{\frac{n+1}{2}}\Gamma(\frac{n}{2})\Gamma(\frac{1}{2})} e^{-\frac{x}{2}} \mathbb{1}_{\mathbb{R}^+}(x) \int_0^1 (xu)^{\frac{n}{2}-1}(x-xu)^{-\frac{1}{2}} du \\ &= \frac{1}{2^{\frac{n+1}{2}}\Gamma(\frac{n}{2})\Gamma(\frac{1}{2})} e^{-\frac{x}{2}} \mathbb{1}_{\mathbb{R}^+}(x) \int_0^1 (xu)^{\frac{n}{2}-1}(x-xu)^{-\frac{1}{2}} xdu \\ &= \frac{1}{2^{\frac{n+1}{2}}\Gamma(\frac{n}{2})\Gamma(\frac{1}{2})} e^{-\frac{x}{2}} x^{\frac{n+1}{2}-1} \beta(\frac{n}{2}, \frac{1}{2}) \mathbb{1}_{\mathbb{R}^+}(x) \\ &= \frac{x^{\frac{n+1}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n+1}{2}}\Gamma(\frac{n+1}{2})} \mathbb{1}_{\mathbb{R}^+}(x)\end{aligned}$$

1.5.2 Moyenne et variance de la loi $\chi^2(n)$

1.5.3 Tracé de la densité de probabilité de la loi $\chi^2(n)$