Modèles Linéaires Appliqués

Arthur Charpentier

Automne 2020

Rappels #2 (loi normale et vecteur Gaussien)

Soit $\mathbf{X}$ un vecteur aléatoire de dimension d

- L'espérance de $\mathbf{X}$, notée $\mathbb{E}(\mathbf{X})$ est définie (si elle existe) par le vecteur de dimension d $\mathbb{E}(\mathbf{X}) = (\mathbb{E}(\mathbf{X}_1), \ldots, \mathbb{E}(\mathbf{X}_d))^\top$.
- La matrice de covariance (appelée aussi matrice de variance-covariance de $\mathbf{X}$) est définie (si elle existe) par la matrice de taille (d, d)

$$\text{Var}(\mathbf{X}) = \mathbb{E}\Big((\mathbf{X} - \mathbb{E}(\mathbf{X}))(\mathbf{X} - \mathbb{E}(\mathbf{X}))^\top\Big).$$

Ainsi le terme ij de cette matrice représente la covariance entre X_i et X_j,

$$\text{Cov}(X_i, X_j) = \mathbb{E}\left[(X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))\right].$$

Random Vectors

Soit $\mathbf{X}$ un vecteur aléatoire de dimension d, de moyenne $\boldsymbol{\mu}$ et de matrice de covariance $\boldsymbol{\Sigma}$.
Soient $\mathbf{A}$ et $\mathbf{B}$ deux matrices réeeles de taille (d, p) et (d, q) et enfin soit $\mathbf{a} \in \mathbb{R}^p$ alors

- $\text{Var}(\mathbf{X}) = \mathbb{E}\left((\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^\top\right) = \mathbb{E}(\mathbf{X}\mathbf{X}^\top) - \boldsymbol{\mu}\boldsymbol{\mu}^\top$.
- $\mathbb{E}\Big(\mathbf{A}^\top\mathbf{X} + \mathbf{a}\Big) = \mathbf{A}^\top\boldsymbol{\mu} + \mathbf{a}$.
- $\text{Var}\Big(\mathbf{A}^\top\mathbf{X} + \mathbf{a}\Big) = \mathbf{A}^\top\boldsymbol{\Sigma}\mathbf{A}$.
- $\text{Cov}\Big(\mathbf{A}^\top\mathbf{X}, \mathbf{B}^\top\mathbf{X}\Big) = \mathbf{A}^\top\boldsymbol{\Sigma}\mathbf{B}$.

The Gaussian Distribution

A Gaussian variable, with distribution $\mathcal{N}(\mu,\sigma^2)$, where $\mu \in \mathbb{R}$ and $\sigma > 0$, has density

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right), \text{ for all } x \in \mathbb{R}.$$

Then $\mathbb{E}(X) = \mu$ and $\text{Var}(X) = \sigma^2$.
Observe that if $Z \sim \mathcal{N}(0,1)$, $X = \mu + \sigma Z \sim \mathcal{N}(\mu,\sigma^2)$.
The Gaussian **vector** $\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$: $\mathbf{X} = (X_1,...,X_n)$ is a Gaussian vector with mean $\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}$ and covariance matrix $\text{Var}(\mathbf{X}) = \boldsymbol{\Sigma} = \mathbb{E}\left((\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^{\mathsf{T}}\right)$ non-degenerated ($\boldsymbol{\Sigma}$ is invertible) if its density is

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det\boldsymbol{\Sigma}}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right), \ \mathbf{x} \in \mathbb{R}^n,$$

see multivariate Gaussian distribution

The Gaussian Distribution

If $\mathbf{X}$ is a Gaussian vector, then for any i, X_i has a (univariate) Gaussian distribution, but its converse it not necessarily true.

Let $\mathbf{X} = (X_1, ..., X_n)$ be a random vector with mean $\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}$ and with covariance matrix $\boldsymbol{\Sigma}$, if $\mathbf{A}$ is a $k \times n$ matrix, and $\mathbf{b} \in \mathbb{R}^k$, then $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$ is a Gaussian vector $\mathbb{R}^k$, with distribution $\mathcal{N}\left(\mathbf{A}\mu + \mathbf{b}, \mathbf{A\Sigma A}^\mathsf{T}\right)$.

Observe that if (X_1, X_2) is a Gaussian vector, X_1 and X_2 are independent if and only if

$$\text{Cov}(X_1, X_2) = \mathbb{E}\left((X_1 - \mathbb{E}(X_1))(X_2 - \mathbb{E}(X_2))\right) = 0.$$

The Gaussian Distribution

Let $\mathbf{Z} = (Y, X) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_Y \\ \mu_X \end{pmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{Y,Y} & \Sigma_{Y,X} \\ \Sigma_{X,Y} & \Sigma_{X,X} \end{pmatrix}$$

Then

$$Y|X = x \sim \mathcal{N}(\mu_{Y|x}, \sigma^2_{Y|x}) \text{ where } \begin{cases} \mu_{Y|x} = \mu_Y + \Sigma_{Y,X}\Sigma_{X,X}^{-1}(x - \mu_X) \\ \sigma^2_{Y|x} = \Sigma_{Y,Y} - \Sigma_{Y,X}\Sigma_{X,X}^{-1}\Sigma_{X,Y} \end{cases}$$

Hence, $\mathbb{E}[Y|X = x] = \mu_{Y|x}$ is linear in x, with slope

$$\text{Corr}(X, Y)\sqrt{\Sigma_{Y,Y}\Sigma_{X,X}^{-1}}$$

and $\text{Var}[Y|X = x] = \sigma^2_{Y|x}$ is constant
xw(furthermore $\text{Var}[Y|X = x] \leq \text{Var}[Y]$)

Chi-Square Distribution

The chi-squared distribution $\chi^2(\nu)$, with $\nu \in \mathbb{N}^*$ has density

$$x \mapsto \frac{(1/2)^{\nu/2}}{\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}, \text{ where } x \in [0; +\infty),$$

where Γ denotes the Gamma function ($\Gamma(n+1) = n!$). Observe that $\mathbb{E}(X) = \nu$ et $\text{Var}(X) = 2\nu$. ν are the degrees of freedom, see chi-squared distribution

If $X_1, \cdots, X_\nu \sim \mathcal{N}(0,1)$ are independent variables, then $Y = \sum_{i=1}^{\nu} X_i^2 \sim \chi^2(\nu)$, when $\nu \in \mathbb{N}$.

This is a particular case of the Gamma distribution, $X \sim \mathcal{G}\left(\frac{k}{2}, \frac{1}{2}\right)$, see see gamma distribution

Student's t Distribution

The Student's-t distribution $\mathcal{S}t(\nu)$, has density

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{t^2}{\nu}\right)^{-(\frac{\nu+1}{2})},$$

Observe that

$$\mathbb{E}(X) = 0 \text{ and } \mathrm{Var}(X) = \frac{\nu}{\nu-2} \text{ when } \nu > 2.$$

If $X \sim \mathcal{N}(0,1)$ and $Y \sim \chi^2(\nu)$ are independents, then

$$T = \frac{X}{\sqrt{Y/\nu}} \sim \mathcal{S}t(\nu).$$

see Student's t

Student's t Distribution

Let $X_1, \cdots, X_n$ be $\mathcal{N}(\mu, \sigma^2)$ independent random variables. Let

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n} \text{ and } {S_n}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2.$$

Then $\dfrac{(n-1)S_n^2}{\sigma^2}$ has a $\chi^2(n-1)$ distribution, and furthermore

$$T = \sqrt{n}\frac{\overline{X}_n - \mu}{S_n} \sim \mathcal{S}t(n-1).$$

Fisher's F Distribution

The Fisher distribution $\mathcal{F}(d_1, d_2)$, has density

$$x \mapsto \frac{1}{x\,\mathrm{B}(d_1/2, d_2/2)} \left(\frac{d_1\,x}{d_1\,x + d_2}\right)^{d_1/2} \left(1 - \frac{d_1\,x}{d_1\,x + d_2}\right)^{d_2/2}$$

for $x \geq 0$ and $d_1, d_2 \in \mathbb{N}$, where B denotes the Beta function.
$\mathbb{E}(X) = \dfrac{d_2}{d_2 - 2}$ when $d_2 > 2$ and $\mathrm{Var}(X) = \dfrac{2\,d_2^2\,(d_1 + d_2 - 2)}{d_1(d_2 - 2)^2(d_2 - 4)}$
when $d_2 > 4$.
If $X \sim \mathcal{F}(\nu_1, \nu_2)$, then $\dfrac{1}{X} \sim \mathcal{F}(\nu_2, \nu_1)$.

If $X_1 \sim \chi^2(\nu_1)$ and $X_2 \sim \chi^2(\nu_2)$ are independent

$$Y = \frac{X_1/\nu_1}{X_2/\nu_2} \sim \mathcal{F}(\nu_1, \nu_2)$$

see Fisher's $\mathcal{F}$ on wikipedia

Gaussian Sampling

Suppose X_i's i.i.d. from a $\mathcal{N}(\mu,\sigma^2)$ distribution,

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

then

- $\overline{X}_n$ and S_n^2 are independent random variables
- $\overline{X}_n$ has distribution $\mathcal{N}\left(\mu, \dfrac{\sigma^2}{n}\right)$
- $(n-1)S_n^2/\sigma^2$ has distribution $\chi^2(n-1)$. Assume that X_i's are i.i.d. random variables with distribution $\mathcal{N}(\mu,\sigma^2)$, then
- $\sqrt{n}\dfrac{\overline{X}_n - \mu}{\sigma}$ has a $\mathcal{N}(0,1)$ distribution
- $\sqrt{n}\dfrac{\overline{X}_n - \mu}{S_n}$ has a Student-t distribution with $n-1$ degrees of freedom

Théorème de Cochrane (1)

Note: $\mathbf{A}$ est idempotente si $\mathbf{A}^2 = \mathbf{AA} = \mathbf{A}$.
Si $\mathbf{A}$ est idempotente, $\exists\mathbf{P}$ de taille (d, r) tq $\mathbf{P}^\top\mathbf{P} = \mathbb{I}_r$ et $\mathbf{A} = \mathbf{PP}^\top$
Soit $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbb{I}_d)$ et soit $\mathbf{A}$ une matrice réelle symétrique. Alors,

- si $\mathbf{A}$ est une matrice idempotente, de rang r, $\mathbf{X}^\top\mathbf{AX} \sim \chi^2(r)$ où $r = \text{rg}(\mathbf{A})$.
- si $\mathbf{X}^\top\mathbf{AX} \sim \chi^2(s)$ alors $\mathbf{A}$ est une matrice idempotente et $s = \text{rg}(\mathbf{A})$.

Soit $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbb{I}_d)$ et F un sous-espace vectoriel de $\mathbb{R}^d$ de dimension $p < d$. Soient $F^\perp$ son orthogonal, $\mathbf{P}_F$ et $\mathbf{P}_{F^\perp}$ les matrices de projection sur F et $F^\perp$, alors

- $\mathbf{P}_F\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_F)$ et $\mathbf{P}_{F^\perp}\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_{F^\perp})$ et $(\mathbf{P}_F\mathbf{X})$ et $(\mathbf{P}_{F^\perp}\mathbf{X})$ sont indépendants.
- $\|\mathbf{P}_F\mathbf{X}\|^2 \sim \chi^2_p$, $\|\mathbf{P}_{F^\perp}\mathbf{X}\|^2 \sim \chi^2_{d-p}$ et ces deux variables sont indépendantes.

Théorème de Cochrane (2)

Soient $Y_1, \dots, Y_n$ n v.a. i.i.d. de loi $N(0,1)$ et soient

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \text{ et } S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

$\overline{Y}$ et S^2 sont indépendantes et $(n-1)S^2 \sim \chi^2_{n-1}$.

- $F = \text{Vec}(\mathbf{1}_n)$, $\mathbf{P}_F = \mathbf{1}_n(\mathbf{1}_n^\top\mathbf{1}_n)^{-1}\mathbf{1}_n^\top = \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top$ et donc

$$\mathbf{P}_{F^\perp} = \mathbb{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top$$

- $\mathbf{P}_F$ est de rang 1 donc $\mathbf{P}_{F^\perp}$ est de rang $n-1$.
- Enfin $\mathbf{P}_F\mathbf{Y} = (\overline{Y}, \dots, \overline{Y})^\top$ et $\mathbf{P}_{F^\perp} = (Y_1 - \overline{Y}, \dots, Y_n - \overline{Y})^\top$.
- $\|\mathbf{P}_{F^\perp}\mathbf{Y}\|^2 = (n-1)S^2$.