

Éléments du Calcul des Probabilités

Chapitre 5: Vecteurs et processus aléatoires

Louis Wehenkel

Département d'Electricité, Electronique et Informatique - Université de Liège
B24/II.93 - L.Wehenkel@ulg.ac.be



MATH0062-1 : 2BacIng, 2BacInf - 7/5/2013

Find slides: <http://montefiore.ulg.ac.be/~lwh/Probas/>

Vecteurs aléatoires gaussiens

- 5.1 Notion générale de vecteur aléatoire
- 5.2 Vecteurs aléatoires gaussiens
- 5.3 Application: réseau de capteurs
- 5.4 Marginalisation/conditionnement de v.a. gaussiens

Fonctions aléatoires et processus stochastiques

Motivation

- ▶ Dans de nombreuses applications on est amené à étudier les relations entre un certain nombre (parfois plusieurs centaines, ou plusieurs milliers) de variables aléatoires à valeurs réelles.
 - ▶ Par exemple la surveillance des installations industrielles (réacteurs chimiques, réseaux électriques) passe par l'installation de capteurs (pour mesurer des débits, pressions, températures, courants, tensions, puissances, ...) à certains endroits accessibles, afin de déterminer des grandeurs internes non observables directement.
- ▶ Dans ce chapitre on va formaliser l'étude systématique de plusieurs v.a. réelles au moyen de la notion de **vecteur aléatoire**.
 - ▶ Il s'agit surtout de généraliser les notions vues au chapitre 4, et de développer le langage mathématique adéquat pour formuler de façon compacte et facilement exploitable les principales propriétés.
 - ▶ Les vecteurs aléatoires **gaussiens** sont étudiés plus en détails, car ils jouissent de propriétés particulièrement intéressantes en pratique.

5.1 Notion générale de vecteur aléatoire

Définition. Une v.a. vectorielle ou vecteur (colonne) aléatoire \mathcal{X} est une application mesurable de (Ω, \mathcal{E}, P) dans \mathbb{R}^p muni de sa σ -algèbre $\mathcal{B}_{\mathbb{R}^p}$.

NB: $\mathcal{B}_{\mathbb{R}^p}$ est la σ -algèbre produit de p σ -algèbres boréliennes $\mathcal{B}_{\mathbb{R}}$. Il s'agit de l'ensemble des parties de \mathbb{R}^p qui peuvent être obtenues en combinant un nombre au plus dénombrable de sous-ensembles du type $[a_1, b_1] \times \cdots \times [a_p, b_p]$ par intersection, union et complémentation.

On peut se convaincre que le vecteur \mathcal{X} est $(\mathcal{E}, \mathcal{B}_{\mathbb{R}^p})$ -mesurable si et seulement si chacune de ses composantes \mathcal{X}_i est $(\mathcal{E}, \mathcal{B}_{\mathbb{R}})$ -mesurable. L'étude de vecteurs aléatoires de dimension p est donc essentiellement équivalente à l'étude conjointe de p v.a. à valeurs réelles.

Lois, fonctions de répartition, densités

La **loi** de \mathcal{X} est définie pour tout $E \in \mathcal{B}_{\mathbb{R}^p}$ par

$$P_{\mathcal{X}}(E) \triangleq P(\{\omega \in \Omega : \mathcal{X}(\omega) \in E\}). \quad (1)$$

La **fonction de répartition** d'un vecteur aléatoire est une fonction de \mathbb{R}^p dans l'intervalle $[0, 1]$ définie par

$$F_{\mathcal{X}}(x_1, x_2, \dots, x_p) \triangleq P(\mathcal{X}_1 < x_1, \dots, \mathcal{X}_p < x_p), \quad (2)$$

où \mathcal{X}_i désigne la i -ème composante de \mathcal{X} .

Comme dans le cas scalaire, la fonction de répartition définit la loi de probabilité de \mathcal{X} , c'est-à-dire qu'elle permet de calculer la probabilité de tout événement du type $\{\omega \in \Omega : \mathcal{X}(\omega) \in E\}$ pour tout $E \in \mathcal{B}_{\mathbb{R}^p}$.

Si la **densité** existe (on dit alors que le vecteur aléatoire \mathcal{X} est continu, ou encore que les \mathcal{X}_i sont **conjointement** continues), elle est définie par

$$f_{\mathcal{X}}(x_1, x_2, \dots, x_p) \triangleq \frac{\partial^p F_{\mathcal{X}}}{\partial x_1 \dots \partial x_p}. \quad (3)$$

Vecteur moyen et matrice de covariance

On note par $\boldsymbol{\mu}$ (ou $\boldsymbol{\mu}_{\mathcal{X}}$, si nécessaire) le **vecteur moyen**

$$\boldsymbol{\mu} \triangleq E\{\boldsymbol{\mathcal{X}}\} = \begin{bmatrix} E\{\mathcal{X}_1\} \\ \vdots \\ E\{\mathcal{X}_p\} \end{bmatrix}. \quad (4)$$

On note par $\boldsymbol{\Sigma}$ (ou $\boldsymbol{\Sigma}_{\mathcal{X}}$, si nécessaire) la **matrice de covariance**

$$\boldsymbol{\Sigma} \triangleq E\{(\boldsymbol{\mathcal{X}} - \boldsymbol{\mu})(\boldsymbol{\mathcal{X}} - \boldsymbol{\mu})^T\}, \quad (5)$$

une matrice $p \times p$ dont l'élément i, j vaut $E\{(\mathcal{X}_i - \mu_i)(\mathcal{X}_j - \mu_j)\}$.
L'élément i, j de $\boldsymbol{\Sigma}$ vaut donc

$$\boldsymbol{\Sigma}_{i,j} = \text{cov}\{\mathcal{X}_i, \mathcal{X}_j\}. \quad (6)$$

En particulier, on a $\boldsymbol{\Sigma}_{i,i} = V\{\mathcal{X}_i\}$, puisque $\text{cov}\{\mathcal{X}_i, \mathcal{X}_i\} = V\{\mathcal{X}_i\}$.

Dans ce qui suit, nous supposons que toutes les composantes de $\boldsymbol{\mu}$ et de $\boldsymbol{\Sigma}$ sont finies, ce qui revient à supposer que $\mathcal{X}_i \in L_{\Omega}^2, \forall i = 1, \dots, p$.

Propriétés de base

La matrice Σ est par définition **symétrique** et il est important de noter qu'elle est aussi **semi-définie positive (s.d.p.)**. On a en effet $\forall y \in \mathbb{R}^p$ que

$$y^T \Sigma y \stackrel{\Delta}{=} y^T E\{(\mathcal{X} - \mu)(\mathcal{X} - \mu)^T\} y \quad (7)$$

$$= E\{y^T (\mathcal{X} - \mu)(\mathcal{X} - \mu)^T y\} \quad (8)$$

$$= E\{\|y^T (\mathcal{X} - \mu)\|^2\} \quad (9)$$

$$\geq 0. \quad (10)$$

Puisque l'espérance d'une variable positive (équation (9)) est positive. Notons que le passage de (7) à (8) est autorisé, car l'espérance d'une combinaison linéaire de variables est égale à la combinaison linéaire des espérances de ces variables, dès lors que ces dernières sont finies, ce qui est le cas étant données nos hypothèses.

Remarquons aussi que

$$\Sigma = E\{\mathcal{X}\mathcal{X}^T\} - \mu\mu^T. \quad (11)$$

En effet, $\Sigma_{i,j} = \text{cov}\{\mathcal{X}_i, \mathcal{X}_j\} = E\{(\mathcal{X}_i - \mu_i)(\mathcal{X}_j - \mu_j)\}$ et $E\{(\mathcal{X}_i - \mu_i)(\mathcal{X}_j - \mu_j)\} = E\{\mathcal{X}_i \mathcal{X}_j\} - \mu_i \mu_j$, puisque $E\{(\mathcal{X}_i - \mu_i)\mu_j\} = 0$, que $E\{\mu_i(\mathcal{X}_j - \mu_j)\} = 0$, et que $E\{\mu_i \mu_j\} = \mu_i \mu_j$.

Transformations linéaires

- ▶ Soit A une matrice $r \times p$ et \mathcal{X} un v.a. de \mathbb{R}^p .
 - ▶ Alors $\mathcal{Y} = A\mathcal{X}$ est un vecteur aléatoire de \mathbb{R}^r .
 - ▶ On a $\mu_{\mathcal{Y}} = A\mu_{\mathcal{X}}$ et $\Sigma_{\mathcal{Y}} = A\Sigma_{\mathcal{X}}A^T$.
- ▶ Soit b un vecteur de \mathbb{R}^p et \mathcal{X} un v.a. de \mathbb{R}^p .
 - ▶ Alors $\mathcal{Z} = b + \mathcal{X}$ est un vecteur aléatoire de \mathbb{R}^p .
 - ▶ On a $\mu_{\mathcal{Z}} = b + \mu_{\mathcal{X}}$ et $\Sigma_{\mathcal{Z}} = \Sigma_{\mathcal{X}}$.

Théorème de Cramer-Wold. La loi de \mathcal{X} est entièrement déterminée par celles de toutes les combinaisons linéaires de ses composantes $a^T \mathcal{X}$, $\forall a \in \mathbb{R}^p$.

5.2 Vecteurs aléatoires gaussiens

Définition. $\mathcal{X} \in \mathbb{R}^p$ est (par définition) un vecteur aléatoire gaussien à p dimensions (on note $\mathcal{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, où $\boldsymbol{\Sigma}$ est la matrice de covariance de \mathcal{X} , et $\boldsymbol{\mu}$ sa moyenne), si $\forall a \in \mathbb{R}^p$, la combinaison linéaire $a^T \mathcal{X}$ de ses composantes suit une loi de Laplace-Gauss $\mathcal{N}(a^T \boldsymbol{\mu}, a^T \boldsymbol{\Sigma} a)$.

(Dans cette définition, nous autorisons le cas limite où $a^T \boldsymbol{\Sigma} a = 0$).

La propriété d'être gaussien est donc **invariante vis-à-vis de toute transformation linéaire (rotation, dilatation, translation, ...)** de l'espace \mathbb{R}^p .

Cette propriété implique en particulier que chacune des v.a. \mathcal{X}_i suit une loi gaussienne $\mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_{i,i})$ (*mais la réciproque est fautive*).

Le théorème de Cramer-Wold implique donc que la loi d'un vecteur aléatoire gaussien est entièrement déterminée par son vecteur moyen $\boldsymbol{\mu}$ et sa matrice de covariance $\boldsymbol{\Sigma}$.

Densité conjointe et indépendance mutuelle

- ▶ Lorsque Σ est régulière, et seulement dans ce cas, la densité existe et vaut

$$f_{\mathcal{X}}(x) = \frac{1}{(2\pi)^{p/2} \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right), \quad (12)$$

où $|\Sigma|$ désigne le déterminant de la matrice Σ .

- ▶ Par conséquent, les composantes de \mathcal{X} sont **mutuellement indépendantes si, et seulement si, Σ est une matrice diagonale**, c'est-à-dire si les composantes sont décorrélées deux à deux.

En effet, on peut se convaincre que dans ce cas la densité conjointe se factorise en le produit des densités marginales :

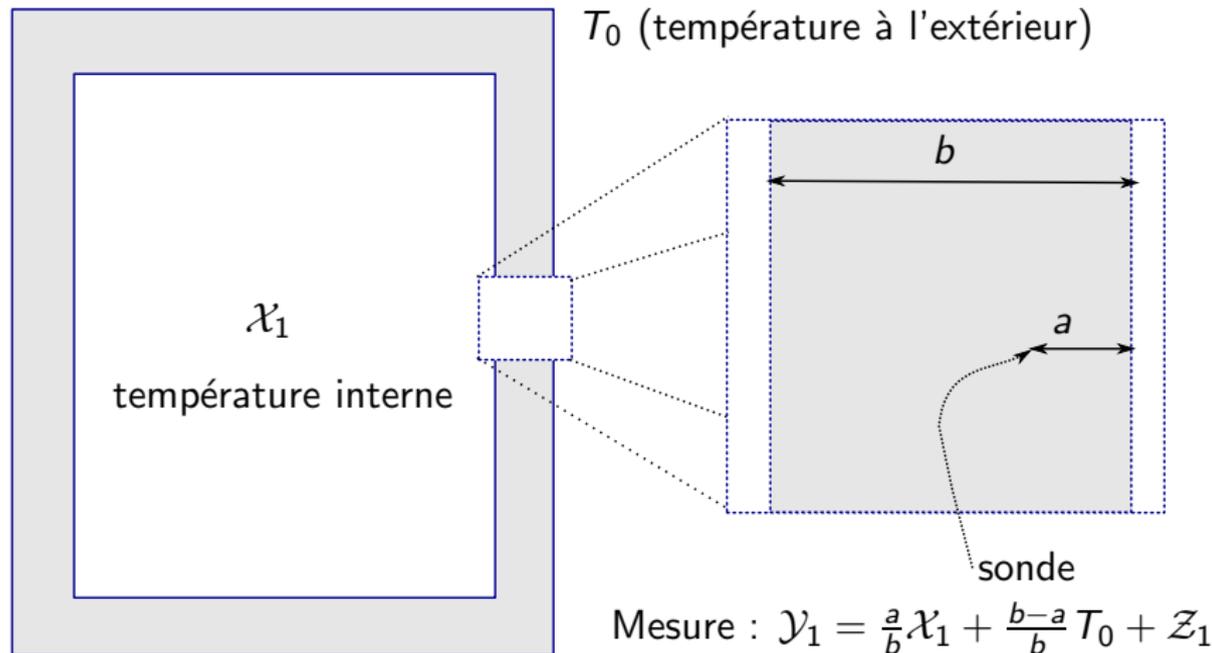
$$f_{\mathcal{X}}(x) = \prod_{i=1}^p f_{\mathcal{X}_i}(x_i), \text{ avec } f_{\mathcal{X}_i}(x_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right).$$

Pour rappel: Les termes non-diagonaux $\Sigma_{i,j}$ correspondent aux covariances $\text{cov}\{\mathcal{X}_i; \mathcal{X}_j\}$, $\forall i \neq j$, et les termes diagonaux correspondent aux variances σ_i^2 des variables \mathcal{X}_i .

5.3 Application: réseau de capteurs

- ▶ On souhaite estimer la température à l'intérieur d'un réacteur, à l'aide d'une sonde de température, qui peut être placée à une certaine profondeur a dans l'enceinte (voir figure).
- ▶ On sait que la température interne est une variable aléatoire \mathcal{X}_1 qui suit une loi gaussienne $\mathcal{N}(\mu_{\mathcal{X}_1}, \sigma_{\mathcal{X}_1}^2)$, avec $\sigma_{\mathcal{X}_1}^2 > 0$.
- ▶ Le modèle physique du système permet d'affirmer que la température réelle au niveau de la sonde vaut $\frac{a}{b}\mathcal{X}_1 + \frac{b-a}{b}T_0$ ou T_0 représente la température à l'extérieur du réacteur, que nous supposons connue.
- ▶ La valeur mesurée est entachée d'une erreur de mesure \mathcal{Z}_1 également gaussienne de moyenne $\mu_{\mathcal{Z}_1} = 0$ et de variance $\sigma_{\mathcal{Z}_1}^2 > 0$. La variable \mathcal{Z}_1 est indépendante de la variable \mathcal{X}_1 .
- ▶ On demande de déterminer une estimation de \mathcal{X}_1 en fonction de \mathcal{Y}_1 et qui minimise l'erreur quadratique.
- ▶ On demande de déterminer a de manière à minimiser l'erreur quadratique d'estimation, sachant que a peut varier entre 0 et a_{\max} , avec $a_{\max} < b$.

Illustration graphique: réacteur et système de mesure



Première analyse du problème

- ▶ $\mathcal{Y}_1 = \frac{a}{b}\mathcal{X}_1 + \frac{b-a}{b}T_0 + \mathcal{Z}_1$.
 - ▶ \mathcal{Y}_1 est une combinaison linéaire des deux variables aléatoires gaussiennes indépendantes \mathcal{X}_1 et \mathcal{Z}_1 ; elle suit donc une loi gaussienne $\mathcal{N}(\mu_{\mathcal{Y}_1}, \sigma_{\mathcal{Y}_1}^2)$.
 - ▶ On calcule $\mu_{\mathcal{Y}_1} = \frac{a}{b}\mu_{\mathcal{X}_1} + \frac{b-a}{b}T_0$, et $\sigma_{\mathcal{Y}_1}^2 = \frac{a^2}{b^2}\sigma_{\mathcal{X}_1}^2 + \sigma_{\mathcal{Z}_1}^2$.
 - ▶ On calcule $\text{cov}\{\mathcal{X}_1, \mathcal{Y}_1\} = \frac{a}{b}\sigma_{\mathcal{X}_1}^2$.
- ▶ Toute combinaison linéaire de \mathcal{X}_1 et \mathcal{Y}_1 peut s'écrire comme une combinaison linéaire de \mathcal{X}_1 et \mathcal{Z}_1 et est donc gaussienne.
 - ▶ En conclusion \mathcal{X}_1 et \mathcal{Y}_1 sont **conjointement gaussiennes**.
 - ▶ La moyenne et la matrice de covariance du vecteur $[\mathcal{X}_1, \mathcal{Y}_1]^T$ valent

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_{\mathcal{X}_1} \\ \frac{a}{b}\mu_{\mathcal{X}_1} + \frac{b-a}{b}T_0 \end{bmatrix} \quad \text{et} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{\mathcal{X}_1}^2 & \frac{a}{b}\sigma_{\mathcal{X}_1}^2 \\ \frac{a}{b}\sigma_{\mathcal{X}_1}^2 & \frac{a^2}{b^2}\sigma_{\mathcal{X}_1}^2 + \sigma_{\mathcal{Z}_1}^2 \end{bmatrix}.$$

- ▶ On a $\det(\boldsymbol{\Sigma}) = \sigma_{\mathcal{X}_1}^2 \sigma_{\mathcal{Z}_1}^2 > 0$.
- ▶ Les deux variables **possèdent donc une densité conjointe**.

Construction du meilleur estimateur linéaire

- ▶ Par application des formules du chapitre 4, on obtient le meilleur estimateur de \mathcal{X}_1 sous forme d'une fonction affine de \mathcal{Y}_1 par
 - ▶ $\hat{\mathcal{X}}_1 = \mu_{\mathcal{X}_1} + \lambda(\mathcal{Y}_1 - \mu_{\mathcal{Y}_1})$
 - ▶ avec $\lambda = \rho_{\mathcal{X}_1;\mathcal{Y}_1} \frac{\sigma_{\mathcal{X}_1}}{\sigma_{\mathcal{Y}_1}} = \frac{\text{COV}\{\mathcal{X}_1;\mathcal{Y}_1\}}{\sigma_{\mathcal{Y}_1}^2} = \frac{\frac{a}{b}\sigma_{\mathcal{X}_1}^2}{\frac{a^2}{b^2}\sigma_{\mathcal{X}_1}^2 + \sigma_{\mathcal{Z}_1}^2}$.
 - ▶ NB: on a $E\{\hat{\mathcal{X}}_1\} = \mu_{\mathcal{X}_1}$.
- ▶ NB: on peut aussi appliquer cette formule pour calculer l'estimateur de \mathcal{Y}_1 sous la forme d'une fonction affine de \mathcal{X}_1 . On trouve
 - ▶ $\hat{\mathcal{Y}}_1 = \mu_{\mathcal{Y}_1} + \rho_{\mathcal{X}_1;\mathcal{Y}_1} \frac{\sigma_{\mathcal{Y}_1}}{\sigma_{\mathcal{X}_1}} (\mathcal{X}_1 - \mu_{\mathcal{X}_1})$ avec $\rho_{\mathcal{X}_1;\mathcal{Y}_1} \frac{\sigma_{\mathcal{Y}_1}}{\sigma_{\mathcal{X}_1}} = \frac{\text{COV}\{\mathcal{X}_1;\mathcal{Y}_1\}}{\sigma_{\mathcal{X}_1}^2} = \frac{a}{b}$.
 - ▶ Autrement dit, $\hat{\mathcal{Y}}_1 = \frac{a}{b}\mathcal{X}_1 + \frac{b-a}{b}T_0$, c'est-à-dire $\hat{\mathcal{Y}}_1 = E\{\mathcal{Y}_1|\mathcal{X}_1\}$!
- ▶ Nous allons voir que **de même** $\hat{\mathcal{X}}_1 = E\{\mathcal{X}_1|\mathcal{Y}_1\}$, cette propriété étant une propriété générale des variables conjointement gaussiennes.

Placement optimal de la sonde

- ▶ Calculons l'erreur quadratique moyenne $E\{(\mathcal{X}_1 - \hat{\mathcal{X}}_1(\mathcal{Y}_1))^2\}$. On a

$$\begin{aligned} E\{(\mathcal{X}_1 - \hat{\mathcal{X}}_1)^2\} &= E\{((\mathcal{X}_1 - \mu_{\mathcal{X}_1}) - \lambda(\mathcal{Y}_1 - \mu_{\mathcal{Y}_1}))^2\} \\ &= E\{(\mathcal{X}_1 - \mu_{\mathcal{X}_1})^2\} + \lambda^2 E\{(\mathcal{Y}_1 - \mu_{\mathcal{Y}_1})^2\} - 2\lambda \text{cov}\{\mathcal{X}_1; \mathcal{Y}_1\} \\ &= \sigma_{\mathcal{X}_1}^2 \left(1 + \lambda^2 \frac{\sigma_{\mathcal{Y}_1}^2}{\sigma_{\mathcal{X}_1}^2} - 2\lambda \frac{\text{cov}\{\mathcal{X}_1; \mathcal{Y}_1\}}{\sigma_{\mathcal{X}_1}^2} \right) \\ &= \sigma_{\mathcal{X}_1}^2 \left(1 - \rho_{\mathcal{X}_1; \mathcal{Y}_1}^2 \right), \end{aligned}$$

avec

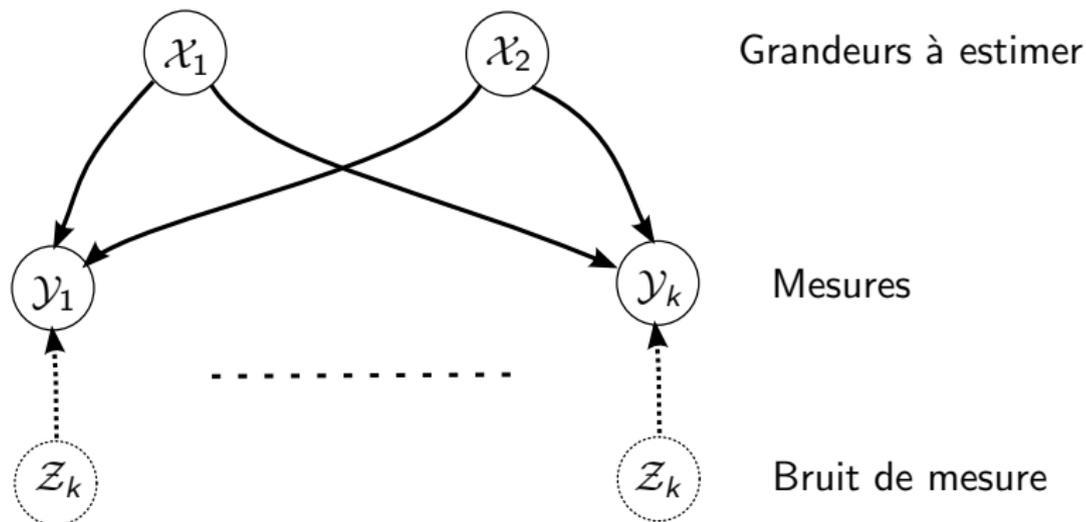
$$\rho_{\mathcal{X}_1; \mathcal{Y}_1} = \frac{\frac{a}{b} \sigma_{\mathcal{X}_1}}{\sqrt{\frac{a^2}{b^2} \sigma_{\mathcal{X}_1}^2 + \sigma_{\mathcal{Z}_1}^2}}.$$

- ▶ Pour minimiser l'erreur d'estimation, il faut maximiser $\rho_{\mathcal{X}_1; \mathcal{Y}_1}$, ce qui revient ici à choisir la valeur maximale de a , i.e. $a = a_{\max}$.
- ▶ La sonde doit donc être placée à la profondeur maximale autorisée.

Discussion/élaboration

- ▶ Quid si la température extérieure T_0 n'est pas connue ?
 - ▶ Introduire dans le modèle une nouvelle v.a. \mathcal{X}_2 pour modéliser la température extérieure : $\mathcal{Y}_1 = \frac{a}{b}\mathcal{X}_1 + \frac{b-a}{b}\mathcal{X}_2 + \mathcal{Z}_1$ et refaire les calculs...
- ▶ Comment étendre le système capteurs ?
 - ▶ Comment placer un second capteur pour améliorer la précision de façon optimale ?
 - ▶ Comment placer d'emblée k capteurs de façon optimale ?
 - ▶ Est-ce que cela revient au même de placer d'abord $k - 1$ capteurs de façon optimale, puis d'y ajouter un nouveau capteur "conditionnellement" optimal, étant donné le placement des $k - 1$ précédents ?
- ▶ Quid, si on veut utiliser le réseau de capteurs pour estimer plusieurs grandeurs simultanément, p.ex. dans notre exemple à la fois la température dans la cuve et la température externe ?

Modèle graphique générique du problème



$$f_{x_1, x_2, y_1, \dots, y_k}(x_1, x_2, y_1, \dots, y_k) = f_{x_1}(x_1)f_{x_2}(x_2)f_{y_1|x_1, x_2}(y_1|x_1, x_2) \cdots f_{y_k|x_1, x_2}(y_k|x_1, x_2)$$

5.4 Marginalisation/conditionnement de v.a. gaussiens (a)

- Cas où $p = 2$: on a $f_{\mathcal{X},\mathcal{Y}}(x, y) = A \exp \left\{ -\frac{1}{2} [\tilde{x}, \tilde{y}] \begin{bmatrix} a & c \\ c & b \end{bmatrix} \begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} \right\}$ avec

$$\tilde{x} = x - \mu_{\mathcal{X}} \quad (13)$$

$$\tilde{y} = y - \mu_{\mathcal{Y}} \quad (14)$$

$$\begin{bmatrix} a & c \\ c & b \end{bmatrix} = \mathbf{\Sigma}^{-1} = \begin{bmatrix} \sigma_{\mathcal{X}}^2 & \rho \sigma_{\mathcal{X}} \sigma_{\mathcal{Y}} \\ \rho \sigma_{\mathcal{X}} \sigma_{\mathcal{Y}} & \sigma_{\mathcal{Y}}^2 \end{bmatrix}^{-1}. \quad (15)$$

NB: la constante A est telle que $\iint_{\mathbb{R}^2} f_{\mathcal{X},\mathcal{Y}}(x, y) dx dy = 1$. On trouve que $A = \frac{1}{2\pi\sqrt{|\mathbf{\Sigma}|}}$ avec $|\mathbf{\Sigma}| = \det(\mathbf{\Sigma}) = \sigma_{\mathcal{X}}^2 \sigma_{\mathcal{Y}}^2 (1 - \rho^2)$.

- Donc $f_{\mathcal{X},\mathcal{Y}}(x, y) = \frac{1}{2\pi\sigma_{\mathcal{X}}\sigma_{\mathcal{Y}}\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2} (a\tilde{x}^2 + 2c\tilde{x}\tilde{y} + b\tilde{y}^2) \right\}$

$$\text{avec } a = \frac{1}{\sigma_{\mathcal{X}}^2(1-\rho^2)}, b = \frac{1}{\sigma_{\mathcal{Y}}^2(1-\rho^2)}, c = \frac{-\rho}{\sigma_{\mathcal{X}}\sigma_{\mathcal{Y}}(1-\rho^2)}. \quad (16)$$

5.4 Marginalisation/conditionnement de v.a. gaussiens (b)

On peut réécrire l'argument de l'exponentielle sous la forme

$$-\frac{1}{2}(a\tilde{x}^2 + 2c\tilde{x}\tilde{y} + b\tilde{y}^2) = \frac{-\tilde{x}^2}{2\sigma_x^2(1-\rho^2)} + \frac{\rho\tilde{x}\tilde{y}}{\sigma_x\sigma_y(1-\rho^2)} + \frac{-\tilde{y}^2}{2\sigma_y^2(1-\rho^2)} = \frac{-\tilde{y}^2}{2\sigma_y^2} - \frac{(\tilde{x} - \rho\frac{\sigma_x}{\sigma_y}\tilde{y})^2}{2\sigma_x^2(1-\rho^2)}.$$

On en déduit que la densité conjointe $f_{X,Y}(x,y)$ peut s'écrire sous la forme

$$f_{X,Y}(x,y) = A_1 \exp\left\{-\frac{(y - \mu_y)^2}{2\sigma_y^2}\right\} A_2 \exp\left\{-\frac{\left(x - \mu_x - \rho\frac{\sigma_x}{\sigma_y}(y - \mu_y)\right)^2}{2\sigma_x^2(1-\rho^2)}\right\}$$

avec $A_1 = \left(\int_{\mathbb{R}} \exp\left\{-\frac{(y - \mu_y)^2}{2\sigma_y^2}\right\} dy\right)^{-1} = \frac{1}{\sqrt{2\pi}\sigma_y}$ et

$$A_2 = \left(\int_{\mathbb{R}} \exp\left\{-\frac{\left(x - \mu_x - \rho\frac{\sigma_x}{\sigma_y}(y - \mu_y)\right)^2}{2\sigma_x^2(1-\rho^2)}\right\} dx\right)^{-1} = \frac{1}{\sqrt{2\pi}\sigma_x\sqrt{1-\rho^2}}, \text{ et}$$

donc $A_1 A_2 = A$.

5.4 Marginalisation/conditionnement de v.a. gaussiens (c)

- Puisque

$$f_{\mathcal{X},\mathcal{Y}}(x, y) = A_1 \exp \left\{ -\frac{(y - \mu_{\mathcal{Y}})^2}{2\sigma_{\mathcal{Y}}^2} \right\} A_2 \exp \left\{ -\frac{\left(x - \mu_{\mathcal{X}} - \rho \frac{\sigma_{\mathcal{X}}}{\sigma_{\mathcal{Y}}} (y - \mu_{\mathcal{Y}}) \right)^2}{2\sigma_{\mathcal{X}}^2 (1 - \rho^2)} \right\},$$

que $A_1 \exp \left\{ -\frac{(y - \mu_{\mathcal{Y}})^2}{2\sigma_{\mathcal{Y}}^2} \right\} = f_{\mathcal{Y}}(y)$, et que $f_{\mathcal{X}|\mathcal{Y}}(x|y) \triangleq \frac{f_{\mathcal{X},\mathcal{Y}}(x,y)}{f_{\mathcal{Y}}(y)}$, on conclut que

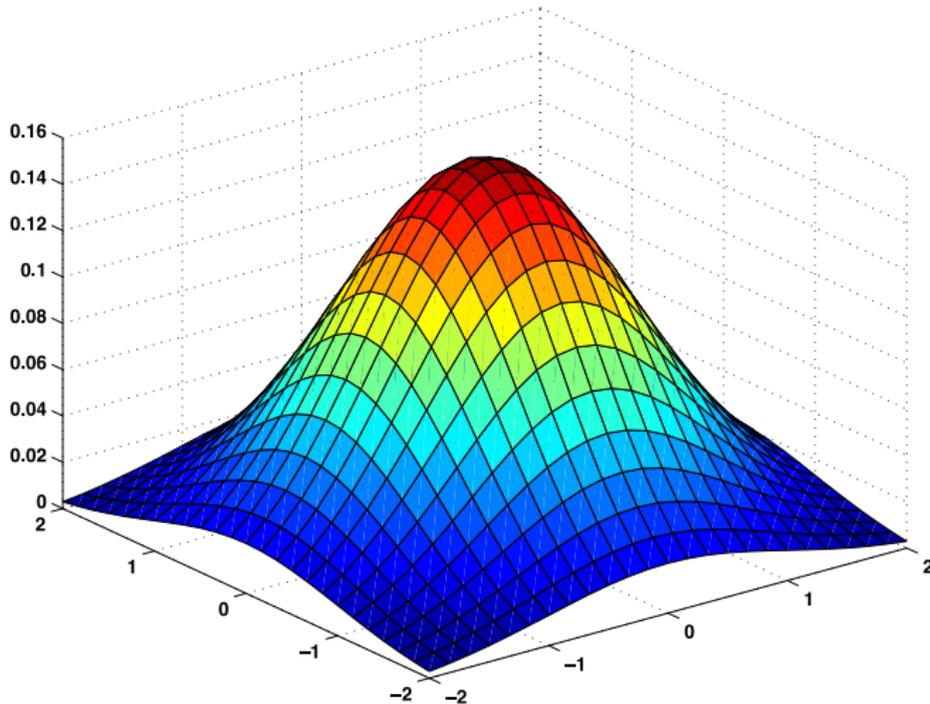
$$f_{\mathcal{X}|\mathcal{Y}}(x|y) = A_2 \exp \left\{ -\frac{\left(x - \mu_{\mathcal{X}} - \rho \frac{\sigma_{\mathcal{X}}}{\sigma_{\mathcal{Y}}} (y - \mu_{\mathcal{Y}}) \right)^2}{2\sigma_{\mathcal{X}}^2 (1 - \rho^2)} \right\}. \quad (17)$$

- En d'autres termes :

- La densité conditionnelle est gaussienne
- $E\{\mathcal{X}|\mathcal{Y}\} = \mu_{\mathcal{X}} + \rho \frac{\sigma_{\mathcal{X}}}{\sigma_{\mathcal{Y}}} (y - \mu_{\mathcal{Y}}) = \mu_{\mathcal{X}} + \text{cov}\{\mathcal{X}; \mathcal{Y}\} (\mathbf{V}\{\mathcal{Y}\})^{-1} (y - \mu_{\mathcal{Y}})$
- $\mathbf{V}\{\mathcal{X}|\mathcal{Y}\} = \sigma_{\mathcal{X}}^2 (1 - \rho^2) = \mathbf{V}\{\mathcal{X}\} - \text{cov}\{\mathcal{X}; \mathcal{Y}\} (\mathbf{V}\{\mathcal{Y}\})^{-1} \text{cov}\{\mathcal{X}; \mathcal{Y}\}$.

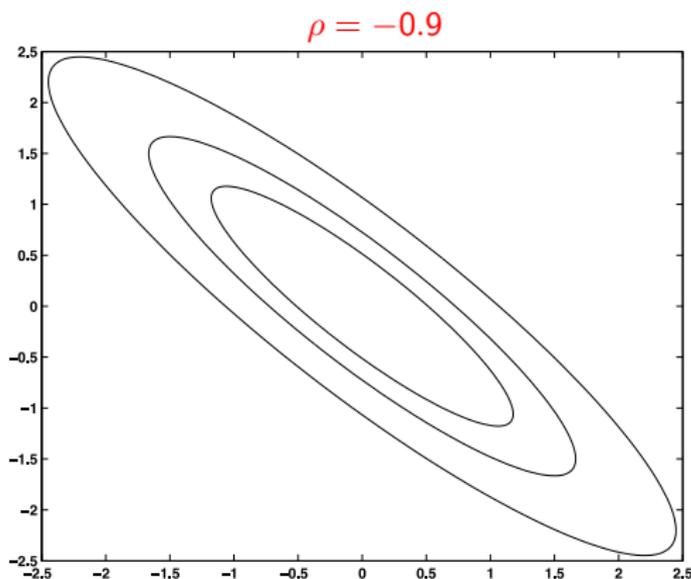
- On vérifiera les théorèmes de l'espérance et de la variance totales, en notant que puisque $\mathbf{V}\{\mathcal{X}|\mathcal{Y}\}$ est constante on a $\mathbf{V}\{\mathcal{X}\} = E\{\mathbf{V}\{\mathcal{X}|\mathcal{Y}\}\}$.

Densité conjointe $f_{X,Y}(x,y)$ avec $\mu = \mathbf{0}$ et $\Sigma = I$



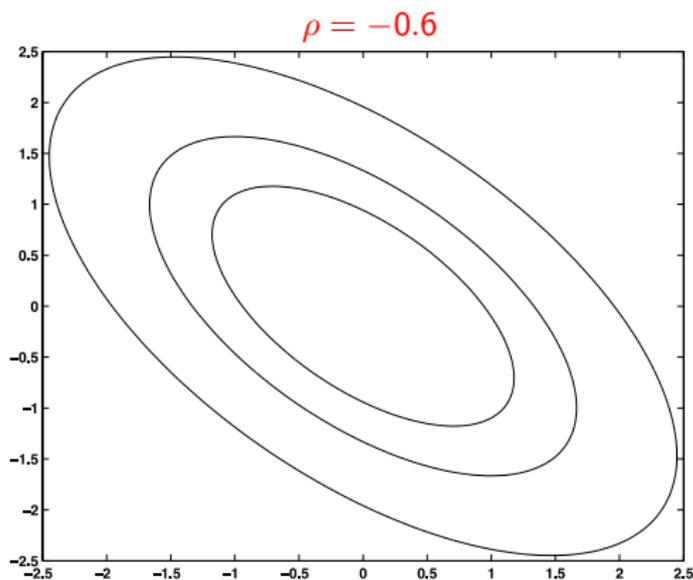
Lieux de points tels que $A \exp \left\{ -\frac{1}{2} [\tilde{x}, \tilde{y}] \Sigma^{-1} \begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} \right\} = cste$

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}. \text{ Supposons que } \sigma_x = \sigma_y = 1, \text{ i.e. } \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$



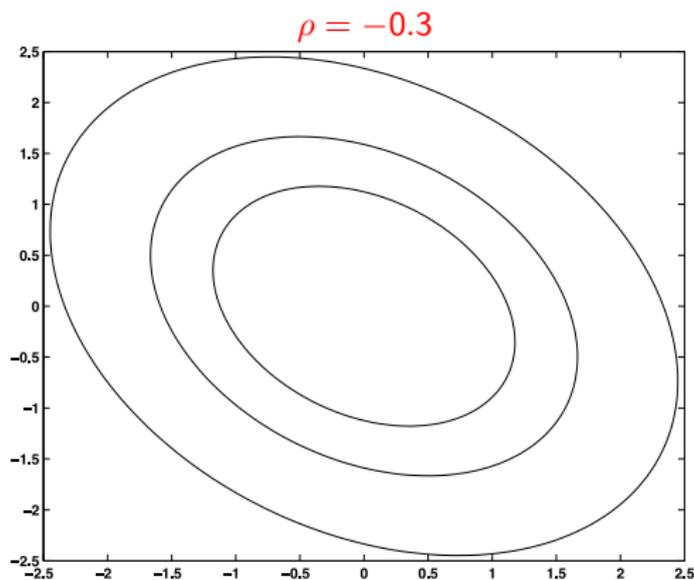
Lieux de points tels que $A \exp \left\{ -\frac{1}{2} [\tilde{x}, \tilde{y}] \Sigma^{-1} \begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} \right\} = cste$

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho \sigma_x \sigma_y \\ \rho \sigma_x \sigma_y & \sigma_y^2 \end{bmatrix}. \text{ Supposons que } \sigma_x = \sigma_y = 1, \text{ i.e. } \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$



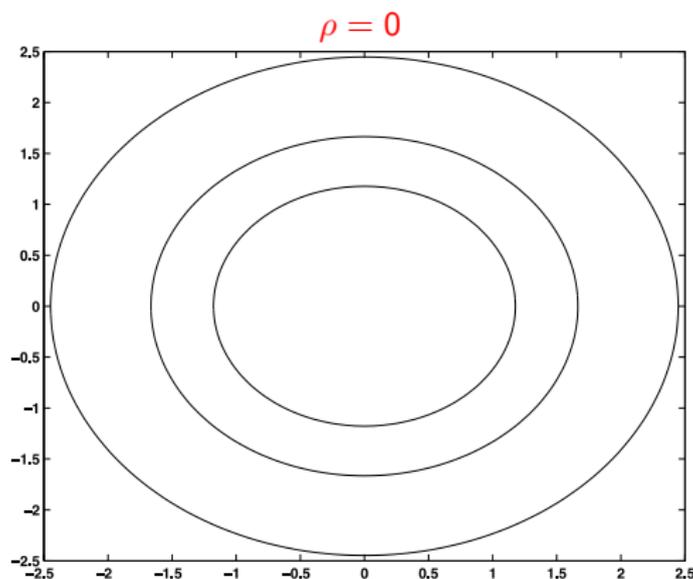
Lieux de points tels que $A \exp \left\{ -\frac{1}{2} [\tilde{x}, \tilde{y}] \Sigma^{-1} \begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} \right\} = cste$

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}. \text{ Supposons que } \sigma_x = \sigma_y = 1, \text{ i.e. } \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$



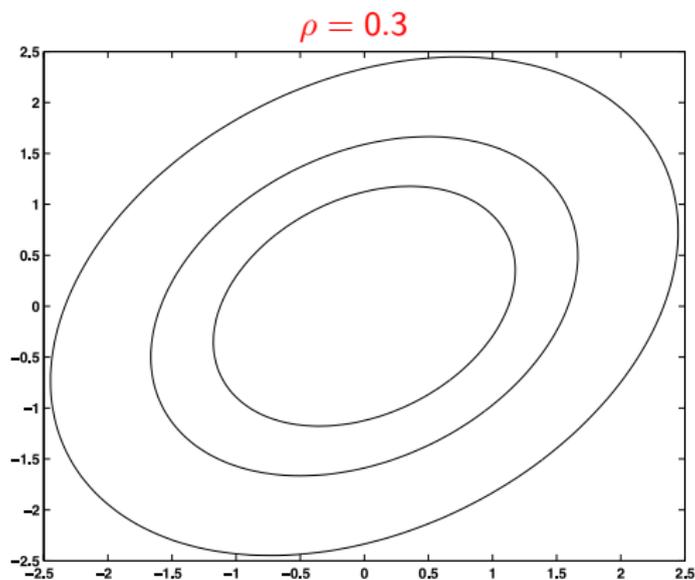
Lieux de points tels que $A \exp \left\{ -\frac{1}{2} [\tilde{x}, \tilde{y}] \Sigma^{-1} \begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} \right\} = cste$

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}. \text{ Supposons que } \sigma_x = \sigma_y = 1, \text{ i.e. } \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$



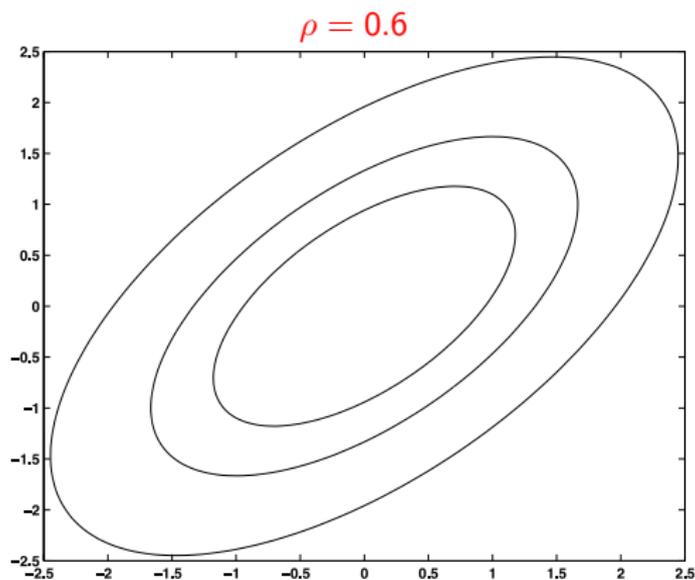
Lieux de points tels que $A \exp \left\{ -\frac{1}{2} [\tilde{x}, \tilde{y}] \Sigma^{-1} \begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} \right\} = cste$

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}. \text{ Supposons que } \sigma_x = \sigma_y = 1, \text{ i.e. } \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$



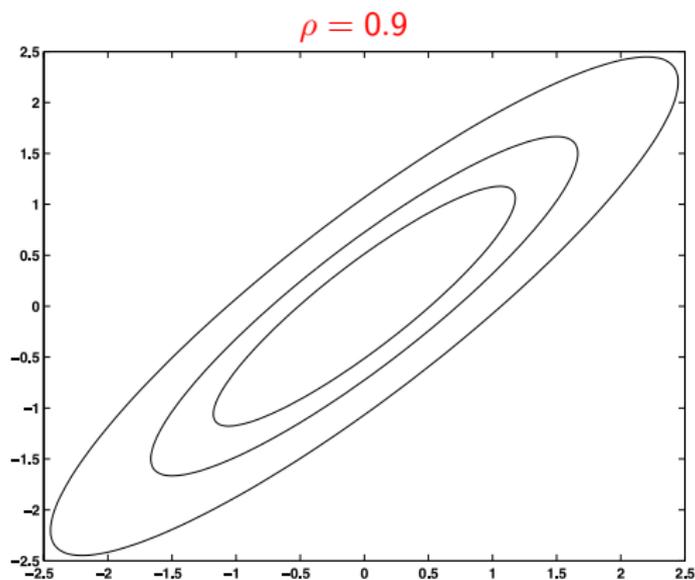
Lieux de points tels que $A \exp \left\{ -\frac{1}{2} [\tilde{x}, \tilde{y}] \Sigma^{-1} \begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} \right\} = cste$

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}. \text{ Supposons que } \sigma_x = \sigma_y = 1, \text{ i.e. } \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$

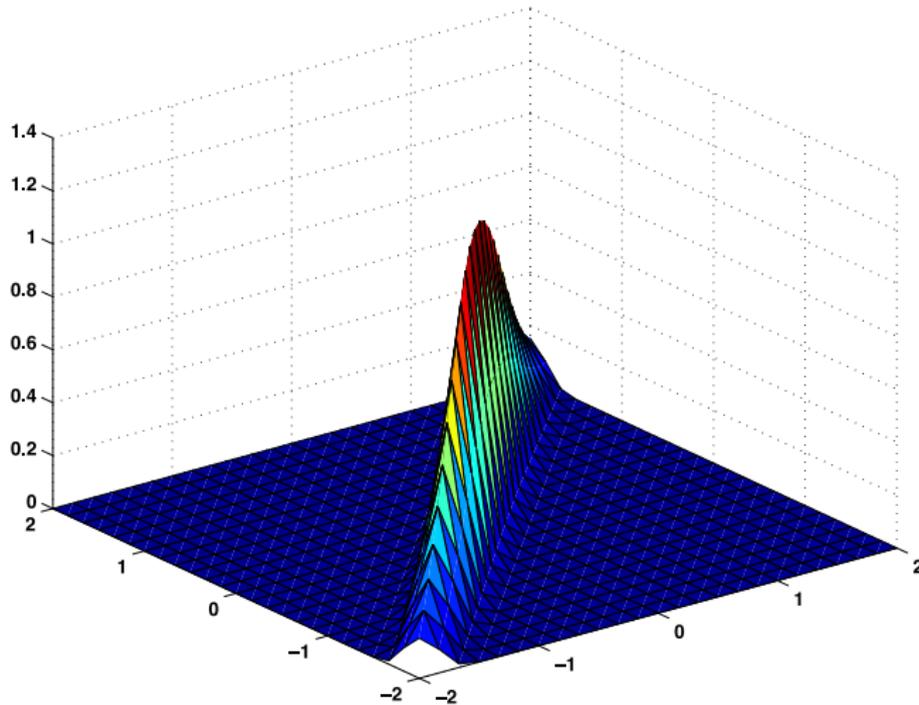


Lieux de points tels que $A \exp \left\{ -\frac{1}{2} [\tilde{x}, \tilde{y}] \Sigma^{-1} \begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} \right\} = cste$

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}. \text{ Supposons que } \sigma_x = \sigma_y = 1, \text{ i.e. } \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$

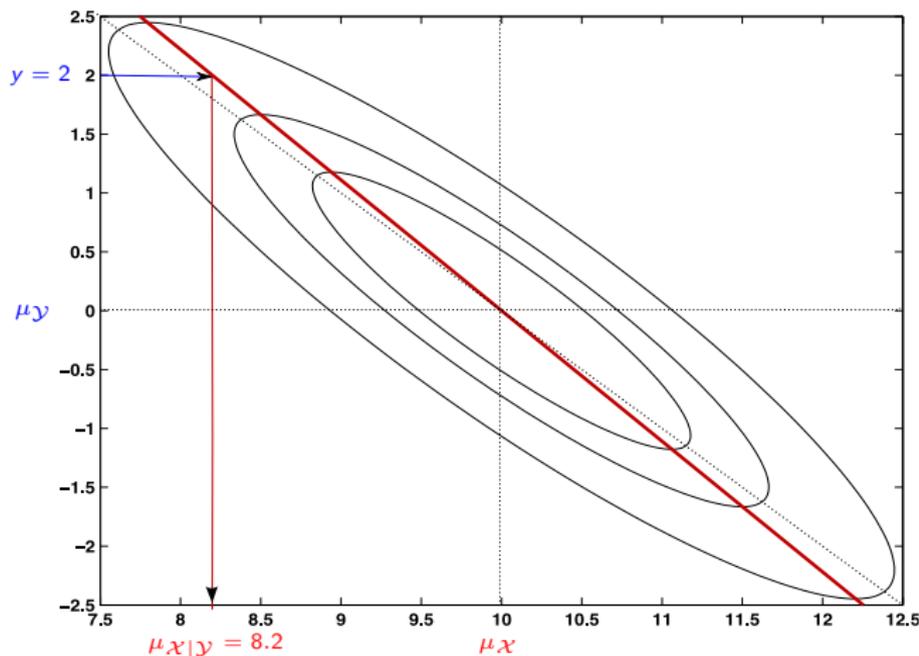


Densité conjointe $f_{X,Y}(x,y)$ pour $\rho = 0.99$



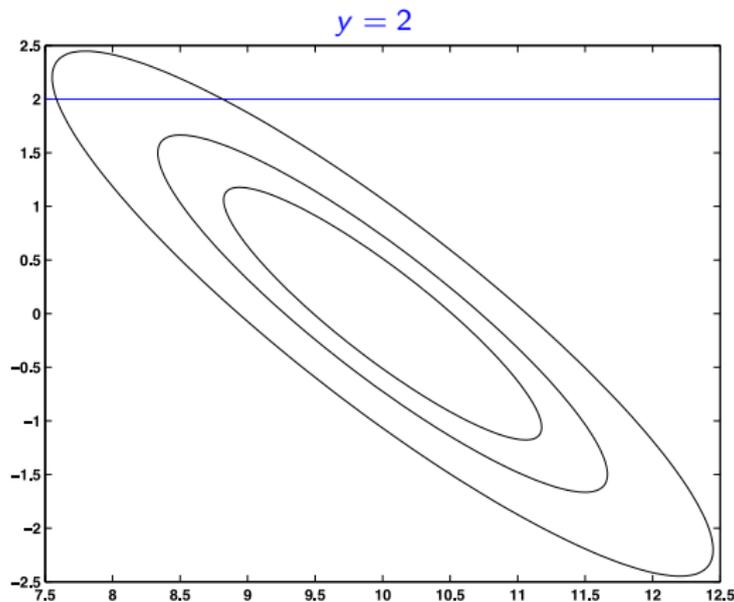
Esp. cond. $\mu_{x|y}(x|y)$: $\rho = -0.9$, $\mu_x = 10$, $\mu_y = 0$

► $\sigma_{x|y}^2 = \sigma_x^2(1 - \rho^2) = 0.19$; $\mu_{x|y} = \mu_x + \rho \frac{\sigma_x}{\sigma_y}(y - \mu_y) = 10 - 0.9y$



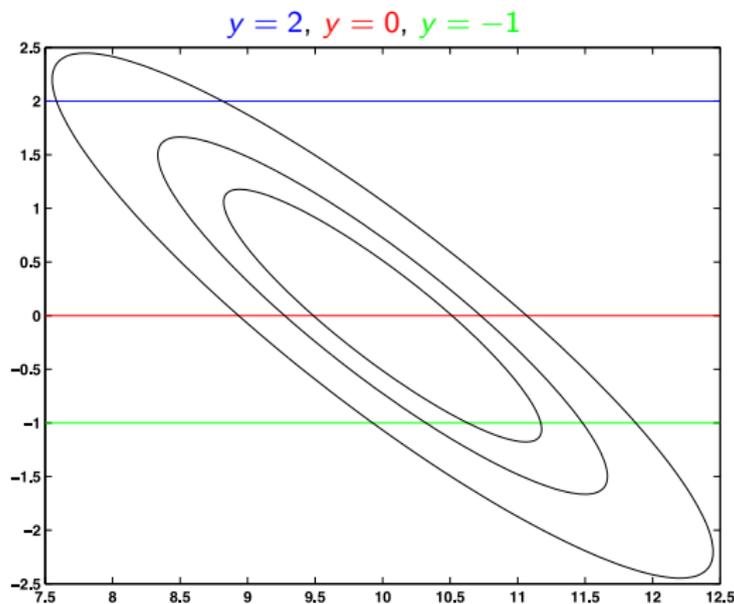
Densités cond. $f_{\mathcal{X}|\mathcal{Y}}(x|y)$: $\rho = -0.9$, $\mu_{\mathcal{X}} = 10$, $\mu_{\mathcal{Y}} = 0$

- ▶ En couleurs : $f_{\mathcal{X}|\mathcal{Y}}(x|2)$, $f_{\mathcal{X}|\mathcal{Y}}(x|0)$; $f_{\mathcal{X}|\mathcal{Y}}(x|-1)$.
- ▶ En noir : $f_{\mathcal{X}}(x)$.



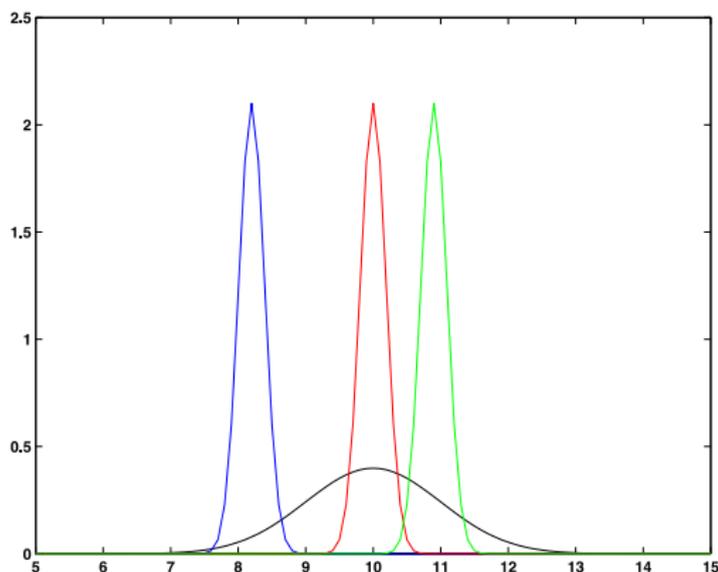
Densités cond. $f_{\mathcal{X}|\mathcal{Y}}(x|y)$: $\rho = -0.9$, $\mu_{\mathcal{X}} = 10$, $\mu_{\mathcal{Y}} = 0$

- ▶ En couleurs : $f_{\mathcal{X}|\mathcal{Y}}(x|2)$, $f_{\mathcal{X}|\mathcal{Y}}(x|0)$; $f_{\mathcal{X}|\mathcal{Y}}(x|-1)$.
- ▶ En noir : $f_{\mathcal{X}}(x)$.



Densités cond. $f_{X|Y}(x|y)$: $\rho = -0.9, \mu_X = 10, \mu_Y = 0$

- ▶ En couleurs : $f_{X|Y}(x|2), f_{X|Y}(x|0); f_{X|Y}(x|-1)$.
- ▶ En noir : $f_X(x)$.



5.4 Marginalisation/conditionnement de v.a. gaussiens (d)

Généralisation au cas d'un v.a. de \mathbb{R}^p , $p > 2$

(nous supposons ici que la densité conjointe existe, c'est-à-dire que Σ est non singulière)

- ▶ Si on partitionne \mathcal{X} en deux sous-vecteurs \mathcal{X}_1 et \mathcal{X}_2 , respectivement à k et à $p - k$ composantes, et respectivement de moyennes μ_1 et μ_2 :

$$\mathcal{X} = \begin{bmatrix} \mathcal{X}_1 \\ \mathcal{X}_2 \end{bmatrix}, \quad (18)$$

la moyenne se partitionne selon

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad (19)$$

et la matrice de variance-covariance se partitionne selon

$$\Sigma = \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{bmatrix}, \quad (20)$$

où $\Sigma_{2,1} = \Sigma_{1,2}^T$.

NB: Si Σ est s.d.p. et non singulière, alors $\Sigma_{1,1}$ et $\Sigma_{2,2}$ sont aussi s.d.p. et non singulières.

5.4 Marginalisation/conditionnement de v.a. gaussiens (e)

Les densités marginales **sont gaussiennes** et obtenues comme suit (cf notes):

- ▶ la densité marginale de \mathcal{X}_1 est donnée par

$$f_{\mathcal{X}_1}(x_1) = \frac{1}{(2\pi)^{k/2} \sqrt{|\Sigma_{1,1}|}} \exp\left(-\frac{1}{2}(x_1 - \mu_1)^T \Sigma_{1,1}^{-1}(x_1 - \mu_1)\right). \quad (21)$$

- ▶ la densité marginale de \mathcal{X}_2 est donnée par

$$f_{\mathcal{X}_2}(x_2) = \frac{1}{(2\pi)^{(p-k)/2} \sqrt{|\Sigma_{2,2}|}} \exp\left(-\frac{1}{2}(x_2 - \mu_2)^T \Sigma_{2,2}^{-1}(x_2 - \mu_2)\right). \quad (22)$$

Les densités conditionnelles **sont aussi gaussiennes**. Par exemple, celle de \mathcal{X}_1 lorsque \mathcal{X}_2 est connu est à k dimensions, et est obtenue comme suit:

- ▶ l'espérance conditionnelle: $E\{\mathcal{X}_1|\mathcal{X}_2\} = \mu_1 + \Sigma_{1,2}\Sigma_{2,2}^{-1}(\mathcal{X}_2 - \mu_2)$,
- ▶ la matrice de covariance conditionnelle: $\Sigma_{1,1|2} = \Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1}$,
- ▶ et la densité conditionnelle s'écrit donc sous la forme

$$f_{\mathcal{X}_1|\mathcal{X}_2}(x_1|x_2) = \frac{1}{(2\pi)^{k/2} \sqrt{|\Sigma_{1,1|2}|}} \exp\left(-\frac{1}{2}(x_1 - \mu_{1|2}(x_2))^T \Sigma_{1,1|2}^{-1}(x_1 - \mu_{1|2}(x_2))\right),$$

avec $\mu_{1|2}(x_2) = \mu_1 + \Sigma_{1,2}\Sigma_{2,2}^{-1}(x_2 - \mu_2)$.

5.4 Marginalisation/conditionnement de v.a. gaussiens (f)

Synthèse

- ▶ Etant donnée la densité conjointe d'un vecteur aléatoire gaussien continu, les formules qui précèdent permettent d'obtenir la densité marginale de n'importe quel sous-ensemble de variables
 - ▶ Le vecteur moyen est le sous-vecteur μ correspondant aux indices des variables retenues
 - ▶ La matrice de covariance est obtenue en conservant les lignes et les colonnes de Σ correspondant aux indices des variables retenues.
- ▶ Etant donnée la densité conjointe d'un vecteur aléatoire gaussien continu, les formules qui précèdent permettent d'obtenir la densité conditionnelle de n'importe quel sous-ensemble de variables par rapport à n'importe quel autre sous-ensemble de variables en formant les vecteurs μ_1, μ_2 , et les matrices $\Sigma_{1,1}, \Sigma_{2,2}, \Sigma_{1,2}$ correspondant aux indices des ces deux sous-ensembles de variables.

NB: Les formules peuvent être étendues au cas où les variables ne sont pas conjointement continues, mais le traitement complet de cela sort du cadre de ce cours.

Théorèmes de l'espérance et de la covariance totales

► Théorème de l'espérance totale sous forme vectorielle

- On montre en toute généralité que si μ_1 est fini, alors

$$E\{\mu_{1|2}\} = \mu_1. \quad (23)$$

- Dans le cas gaussien, on peut le vérifier en partant du fait que $\mu_{1|2} = \mu_1 + \Sigma_{1,2}\Sigma_{2,2}^{-1}(\mathcal{X}_2 - \mu_2)$.

► Théorème de la covariance totale

- On montre aussi en toute généralité que si $\Sigma_{1,1}$ est finie, alors

$$E\{\Sigma_{1,1|2}\} + \Sigma\mu_{1|2} = \Sigma_{1,1}. \quad (24)$$

- Dans le cas gaussien (non singulier) on peut le vérifier en tenant compte

1. du fait que $\Sigma_{1,1|2} = \Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1}$ est constante et donc égale à $E\{\Sigma_{1,1|2}\}$, et
2. du fait que la matrice de covariance du vecteur $\mu_{1|2} = \mu_1 + \Sigma_{1,2}\Sigma_{2,2}^{-1}(\mathcal{X}_2 - \mu_2)$ vaut $(\Sigma_{1,2}\Sigma_{2,2}^{-1})\Sigma_{2,2}(\Sigma_{1,2}\Sigma_{2,2}^{-1})^T = \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{1,2}^T = \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1}$.

Vecteurs aléatoires gaussiens

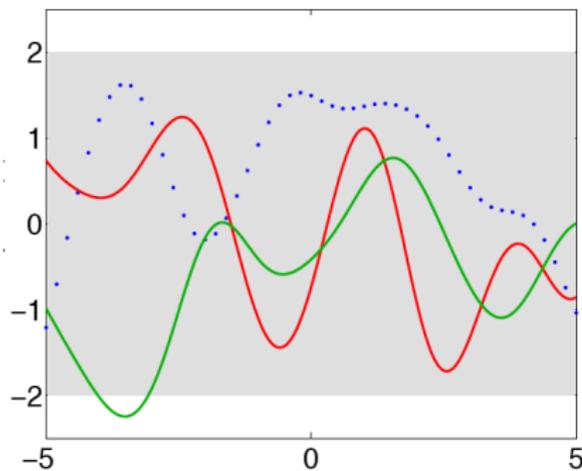
- 5.1 Notion générale de vecteur aléatoire
- 5.2 Vecteurs aléatoires gaussiens
- 5.3 Application: réseau de capteurs
- 5.4 Marginalisation/conditionnement de v.a. gaussiens

Fonctions aléatoires et processus stochastiques

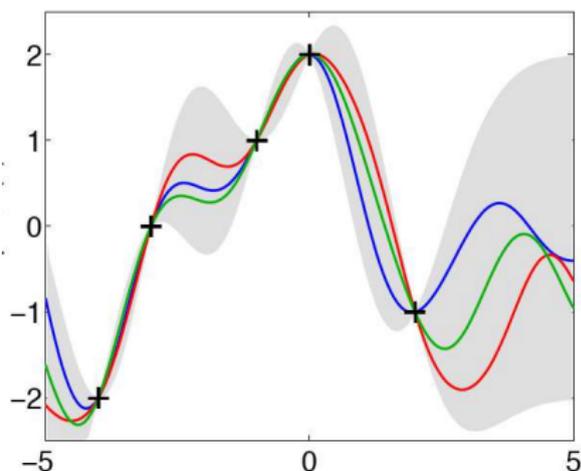
Preview cours BAC3: notion de fonction aléatoire

- ▶ On définit la notion de fonction aléatoire comme suit :
 - ▶ on se donne un ensemble Λ , et un espace de probabilité (Ω, \mathcal{E}, P)
 - ▶ on associe à chaque élément λ de Λ une v.a. \mathcal{X}_λ définie sur (Ω, \mathcal{E}, P) et à valeurs dans un même ensemble $\Omega_{\mathcal{X}}$.
 - ▶ Pour une valeur fixée de ω , cette notion définit donc une fonction de Λ dans $\Omega_{\mathcal{X}}$, d'où le nom de *fonction aléatoire*.
 - ▶ NB: quand $\Lambda = \{1, 2, \dots, p\}$ et $\Omega_{\mathcal{X}} = \mathbb{R}$, cette notion se réduit à la notion de vecteur aléatoire.
- ▶ Dans les applications on considère
 - ▶ le cas où Λ désigne l'axe du temps, et on utilise alors le nom de *processus stochastique* (temporel),
 - ▶ le cas où Λ désigne (une partie de) l'espace physique \mathbb{R}^3 et on parle alors de *champ aléatoire spatial*.
 - ▶ Lorsque Λ est le produit cartésien d'un axe du temps et d'un espace physique on parle de *processus spatio-temporel*.

Suggestion de la notion de "processus stochastique"



Quelques réalisations possibles a priori
 $x(t, \omega_1), x(t, \omega_2), x(t, \omega_3) \dots$



Trajectoires possibles étant données
les observations aux points +

Suggestion de la notion de “champ aléatoire”



Synthèse sur les vecteurs et processus aléatoires

- ▶ Les notions introduites dans ce chapitre sont essentiellement des extensions directes des notions du chapitre 4.
- ▶ La notation vectorielle/matricielle est pratique pour écrire de façon compacte les versions générales des formules les plus importantes.
- ▶ Les vecteurs aléatoires gaussiens sont intimement liés à la structure des espaces euclidiens (linéarité, et orthogonalité).
- ▶ Les modèles gaussiens sont à la base de très nombreuses applications du calcul de probabilités dans le domaine de l'ingénieur, en particulier pour aborder des problèmes complexes:
 - ▶ Informatique, géologie, électricité, mécanique, chimie, constructions
 - ▶ Traitement de données biomédicaux, audio, vidéo etc.
 - ▶ Diagnostic, systèmes de capteurs, contrôle optimal
 - ▶ Prédiction de séries temporelles