

Gestion de l'incertain

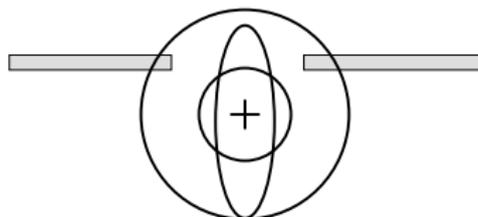
Marie-Odile Berger

<http://members.loria.fr/moberger>

Pourquoi est ce important?

- les données de base sont acquises grace à des capteurs ou elles sont extraites (algorithmiquement) et donc sujettes à des erreurs d'extraction (segmentation de signal de paroles en phonèmes, extraction de caractéristiques à partir d'images, capteurs avec du bruit . . .)
- les données présentes peuvent donc être entachées d'une certaine erreur, voir aberrantes.
- Il est nécessaire de représenter cette incertitude, de la propager et de la prendre en compte dans le processus de classification, d'estimation ou de décision

Exemple: Le robot passe-t-il par la porte étant donnée l'incertitude sur sa position?



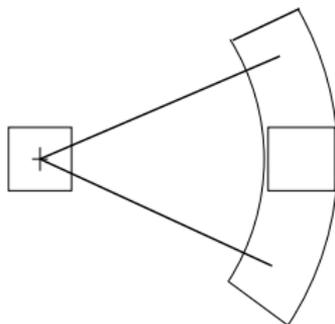
- Représenter l'erreur entachant une donnée ou une estimation
- Rappels de probabilité
- propager ces erreurs dans un processus complexe
- combiner (fusionner) ces mesures
- mesurer l'adéquation modèle/mesure tenant compte de l'incertitude sur la mesure et sur le modèle.

Part I

Représentation de l'incertain

Représentation par région

- La représentation doit permettre de composer et de fusionner facilement les erreurs.
- **Représentation par des régions:** Définition d'une zone où la vraie grandeur doit se trouver. En général: polygone, disque, ellipse
- problème: par composition, la zone peut devenir très complexe et/ou ne plus appartenir à la famille de représentation choisie.



comment fusionner deux observations?

Représentation par une distribution de probabilité

une mesure = réalisation d'une variable aléatoire.

(en répétant plusieurs fois une mesure, on obtient des résultats plus ou moins différents, dont la répartition est aléatoire).

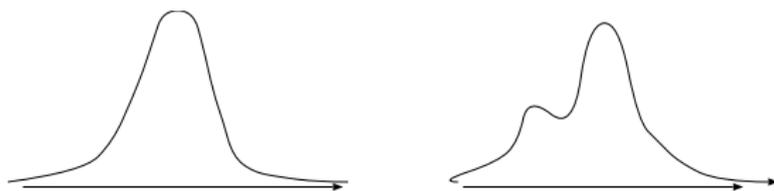


Figure: exemple de distributions d'erreur

Intérêt: facilité de manipulation (composition, comparaison, transformation de deux distributions de probabilité).

Remarque: représentation par zone d'incertitude \longleftrightarrow utilisation d'une distribution uniforme.

Trois utilisations possibles de ces représentations:

- indépendamment d'un modèle: on considère un certain nombre de statistiques descriptives de la distribution (moyenne, médiane, variance. . .) **sans avoir** un modèle de la distribution.
- le modèle de la distribution est **supposé connu** (souvent loi normale)
- on se place dans une approche **non paramétrique** permettant de modéliser des distributions variées, en particulier non uni-modales.

Difficulté: la modélisation utilisée est elle adaptée aux données?

Part II

Rappels de probabilité

- Le but n'est pas de faire un cours de probabilité complet mais d'insister sur certains aspects utiles dans ce cours.
- Voici plusieurs liens de cours de probabilité et statistiques pour l'ingénieur que vous pouvez consulter
 - Un cours de Anne Perrut qui fournit les notions de base des probabilités
 - Cours de probabilité et statistique pour l'ingénieur par B.Jourdain à l'ENPC
 - Méthodes statistiques pour l'ingénieur par O. Gaudouin à l'ENSIMAG

En pratique, on dispose d'un vecteur de caractéristique $x \in R^n, n > 1$). Il faut étendre les notions de moyenne et variances aux dimensions $n > 1$.

Inégalité de Bienaymé-Tchebyshev:

$$P(|X - E(X)| > k\sigma) \leq \frac{1}{k^2}$$

On en déduit que $P(|X - m| > 3\sigma) \leq \frac{1}{9}$. C'est à dire que l'intervalle $[m - 3\sigma, m + 3\sigma]$ contient au moins $8/9$, c'est à dire l'essentiel de la distribution.

Cette majoration est souvent excessive. Pour une gaussienne centrée réduite, $[-1.65, 1.65]$ contient 90% de la distribution.

soit X le vecteur aléatoire $X = [X_1, \dots, X_n]^t$.

Espérance

$$E(X) = [E(X_1), \dots, E(X_n)]^t$$

Matrice de covariance (variance-covariance)

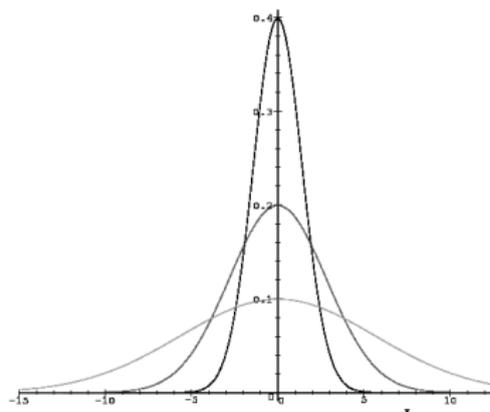
$$\text{var}(X) = E((X - E(X))(X - E(X))^t) = \begin{bmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) & \text{cov}(X_1, X_3) \\ \text{cov}(X_1, X_2) & \text{var}(X_2) & \text{cov}(X_2, X_3) \\ \text{cov}(X_1, X_3) & \text{cov}(X_2, X_3) & \text{var}(X_3) \end{bmatrix}$$

Cette matrice résume la structure des dépendances linéaires entre les n variables prises deux à deux.

La loi normale

en dimension 1: une v.a. gaussienne de moyenne m et de variance σ^2 (notée $\mathcal{N}(m, \sigma^2)$) a la densité de probabilité:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$$



Calcul de $P(X \in [1, 2])$?

$$P(X \in [1, 2]) = \int_1^2 f(x)dx = \int_{-\infty}^2 f(x)dx - \int_{-\infty}^1 f(x)dx]$$

en matlab $\int_{-\infty}^t f(x)dx = \text{cdf}(t, m, \sigma)$ (cumulative distribution function)

$$P(X \in [1, 2]) = \text{cdf}(2, m, \sigma) - \text{cdf}(1, m, \sigma)$$

Définition: un **vecteur aléatoire gaussien** de moyenne m et de matrice de covariance Λ a pour densité

$$f(x) = \frac{1}{(2\pi)^{n/2} \sqrt{\det \Lambda}} \exp\left\{-\frac{1}{2}(x - m)^t \Lambda^{-1} (x - m)\right\}$$

bruit blanc gaussien: v.a suivant une loi de Gauss de moyenne nulle.

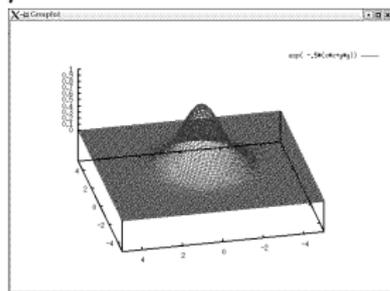
Gaussienne en dimension 2

Exemple de la densité en dimension 2 en fonction des corrélations entre variables.

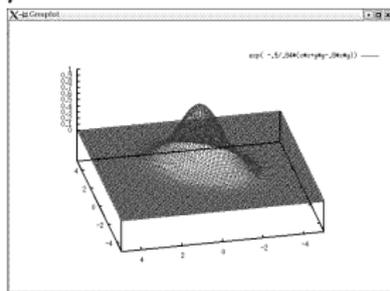
$$\Lambda = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

pour $\sigma_1 = \sigma_2 = 1$,

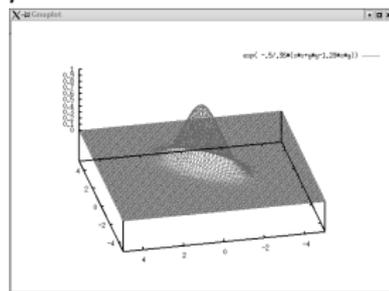
$\rho = 0$



$\rho = .4$



$\rho = .8$



- Mesure de la distance qui prend en compte la covariance des données (à la différence de la distance euclidienne où toutes les composantes sont traitées de la même façon)
- soit une v.a de variance Λ diagonale $[1, 2, 10]$. La distance $d_{\Lambda}(x, y) = \sum \frac{1}{\sigma_i^2} (x_i - y_i)^2$ va faire en sorte que les écarts sur la 3^{ième} coordonnée, qui est très imprécise ($\sigma_3^2 = 10$), seront moins pris en compte dans le calcul de la distance
- Plus généralement, $d_{\Lambda}(x, y) = \sqrt{(x - y)^t \Lambda^{-1} (x - y)}$

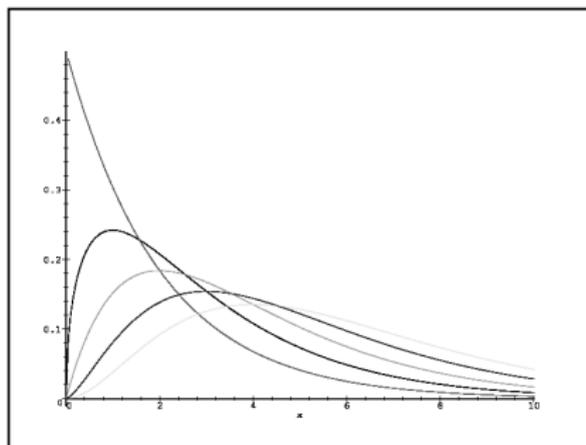
La loi du χ^2

Definition: Soient X_1, \dots, X_n n variables aléatoires gaussiennes $\mathcal{N}(0, 1)$ indépendantes, la v.a.

$$\chi^2 = X_1^2 + \dots + X_n^2$$

suit une loi du χ^2 à n degrés de liberté.

Propriétés: La somme de deux variables de χ^2 indépendantes à p et q degrés de liberté est encore une variable de χ^2 à $p + q$ degrés de liberté



Moyenne et variance

$$E(\chi^2) = n$$

$$\text{var}(\chi^2) = 2n$$

prop 1: si $X_i : \mathcal{N}(\mu, \sigma)$, $\sum (X_i - \mu)^2 / \sigma^2$ suit une loi du χ^2 .

prop 2 Si X est un vecteur aléatoire gaussien de moyenne nulle et de matrice de covariance Λ , alors $X^t \Lambda^{-1} X$ suit une loi du χ^2 à n degrés de liberté ($n = \text{dimension de } X$).

Cette propriété est très utilisée pour tester la compatibilité d'une mesure avec un modèle

Part III

Composer l'incertain

Soit X une v.a. Soit f une fonction déterministe $f : R^n \rightarrow R$.

Comment calculer (ou approcher) $E(f(X))$ et $var(f(X))$?

Si la densité $p(x)$ de X est connue, c'est direct:

$$E(f(X)) = \bar{f} = \int f(x)p(x)dx$$

$$var(f(X)) = \int (f(x) - \bar{f})^2 p(x)dx$$

Ces valeurs peuvent en général (attention si support infini) être calculées par intégration numérique. Mais cela pose un problème si on veut faire des développements formels.

Ex: si X dépend d'un paramètre a , trouver a tel que $cov(X(a))$ soit minimale.

Il y a donc un intérêt à disposer de formules, même approchées, pour la propagation de l'erreur.

le Cas linéaire: $Y = f(X) = AX$

D'après la linéarité de l'espérance

$$E(Y) = E(AX) = AE(X)$$

$$\text{cov}(Y) = \text{cov}(AX) = E((A(X - E(X)))(X - E(X))^t A^t)) = A \text{cov}(X) A^t$$

$$\boxed{\text{cov}(AX) = A \text{cov}(X) A^t}$$

Propagation de l'incertitude: cas non linéaire

en l'absence d'autres informations, on approche f par l'application linéaire tangente.

Soit $X_0 = E(X)$

à l'ordre 1

$$Y = f(X) \approx f(X_0) + J_f(X_0)(X - X_0)$$

avec $J_f(X_0) = [\frac{\partial f}{\partial X_1}(X_0), \dots, \frac{\partial f}{\partial X_n}(X_0)]$

alors

$$E(Y) \approx f(X_0)$$

$$\begin{aligned} \text{var}(Y) &= E((f(X) - E[f(X)])(f(X) - E[f(X)])^t) \\ &\approx E((f(X) - f(X_0))(f(X) - f(X_0))^t) \\ &\approx E((J_f(X_0)(X - X_0))(J_f(X_0)(X - X_0))^t) \\ &= J_f(X_0)\text{cov}(X)J_f^t(X_0) \end{aligned}$$

$$\boxed{\text{cov}(f(X)) \approx J_f(X_0)\text{cov}(X)J_f^t(X_0)}$$

Exemple: Soit (x, y) un vecteur gaussien de moyenne $(0, 0)$ et de matrice de covariance $\sigma^2 \text{diag}(1, 4)$. Soit la variable aléatoire $x' = f(x, y) = x^2 + 3x - 2y + 5$. Comparons les valeurs exactes et les valeurs approchées de la covariance.

$$p(x, y) = \frac{1}{4\pi\sigma^2} e^{-(x^2+y^2/4)/2\sigma^2}$$

en intégrant (utiliser Maple):

$$E(x') = 5 + \sigma^2$$

$$\text{var}(x') = 25\sigma^2 + 2\sigma^4$$

En utilisant les formules d'approximation, on a $J = [2x + 3, -2]$. Donc au point $(0, 0)$, $J = [3, -2]$. D'où

$$E(x') \approx f(0, 0) + J(0, 0) = 5$$

$$\text{var}(x') = \sigma^2 \begin{bmatrix} 3 & -2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix} = 25\sigma^2$$

Conclusion: si σ est petit, σ^4 est négligeable devant σ^2 et l'approximation fournie est correcte.

Simulation de la covariance par méthode de Monte Carlo

A utiliser lorsque l'hypothèse de linéarité est visiblement mal adaptée. Il s'agit d'une technique de simulation probabiliste.

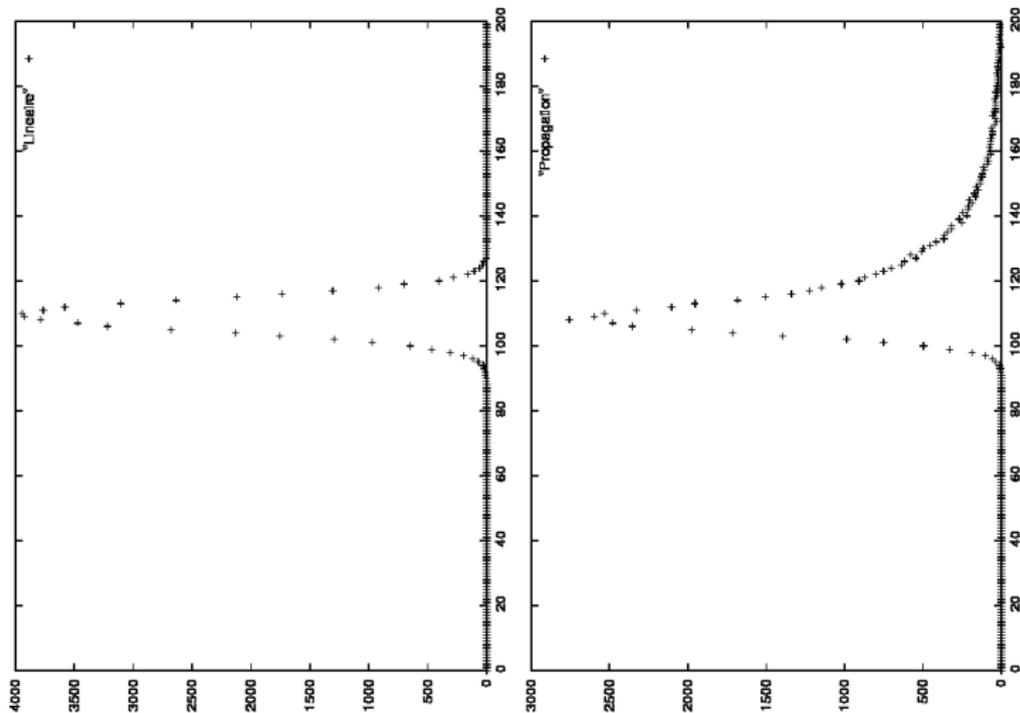
méthode: en supposant que le bruit affectant les données a une distribution connue, ajouter aux données un bruit [respectant cette loi](#).

Faire N fois (N grand) cette simulation et calculer la moyenne et la variance avec la formulation discrète $\bar{x} = \frac{1}{N} \sum x_i$

remarques:

- il faut juste savoir générer selon une loi (voir générateur de loi uniforme, gaussienne...)
- Cela a pu paraître coûteux à une certaine époque, mais beaucoup moins maintenant!
- cela permet de prendre en compte des [fonctions quelconques](#)

Exemple dans le cas de $f(x, y) = x^2 + 3x - 2y + 5$



AVR/TD/proba/genereDistrib.m