

Gestion de l'incertain

Marie-Odile Berger

Oui ! car

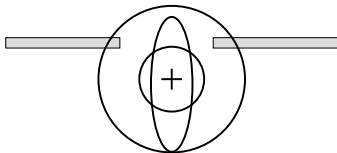
- les données de base sont acquises grace à des capteurs ou elles sont extraites (algorithmiquement) et donc sujettes à des erreurs d'extraction (segmentation de signal de paroles en phonèmes, extraction de caractéristiques à partir d'images, capteurs avec du bruit . . .)
- les données présentes peuvent 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

Modéliser l'incertain: les objectifs

Modéliser l'incertain entachant les mesures et savoir

- 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.

Exemple: un robot se déplace dans un environnement connu muni d'un capteur de position. Passe-t-il par une porte ?



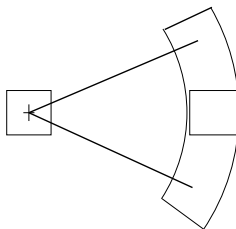
Questions: comment estimer la position du robot à chaque instant? avec quelle précision ?

Les différents type d'erreur

- erreurs systématiques
 - une erreur systématique est soit constante (**biais**), soit à variation lente en fonction du temps (**dérive**) → décalage entre la valeur vraie et la valeur mesurée.
 - exemples: erreur sur la valeur d'une grandeur de référence, erreur due à la surchauffe d'un appareil (caméra)
- erreurs accidentelles (aléatoires)
 - exemples: *bruit* induit par une carte de numérisation, par un microphone...
- erreur aberrante:
 - écart énorme entre vraie valeur et valeur mesurée
 - exemple: erreur de mise en correspondance.

Quelle représentation de l'incertain?

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?

Quelle représentation de l'incertain?

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).

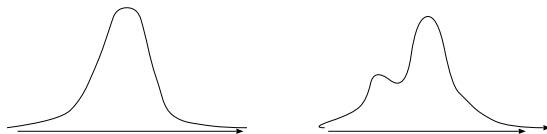


Figure : exemple de distributions d'erreur

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?

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)

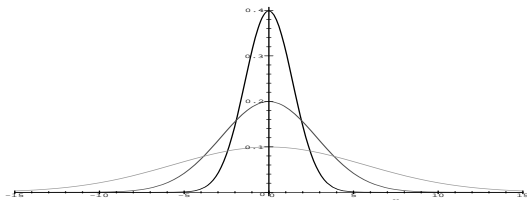
$$\text{var}(X) = E((X - E(X))(X - E(X))^t) = \begin{bmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) & \text{cov}(X_1, X_3) \\ \text{cov}(X_1, X_2) & \text{var}(X_2) & \text{cov}(X_2, X_3) \\ \text{cov}(X_1, X_3) & \text{cov}(X_2, X_3) & \text{var}(X_3) \end{bmatrix}$$

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

La loi multinormale

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:

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



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.

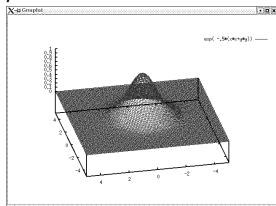
Gaussienne en dimension 2

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

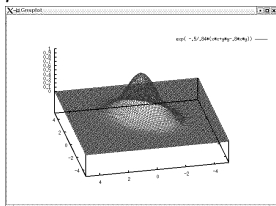
$$\Lambda = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

pour $\sigma_1 = \sigma_2 = 1$,

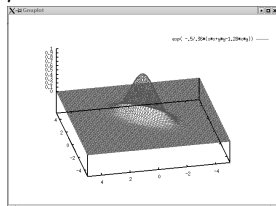
$\rho = 0$



$\rho = .4$



$\rho = .8$



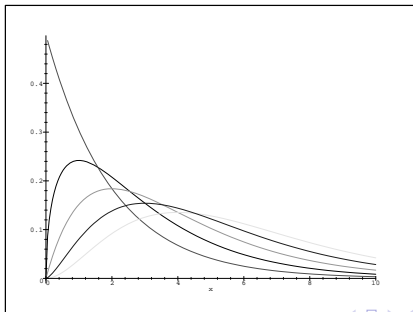
La loi du χ^2

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 ($n =$ dimension de X).

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

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:

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

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

Ces valeurs peuvent toujours 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.

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

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.

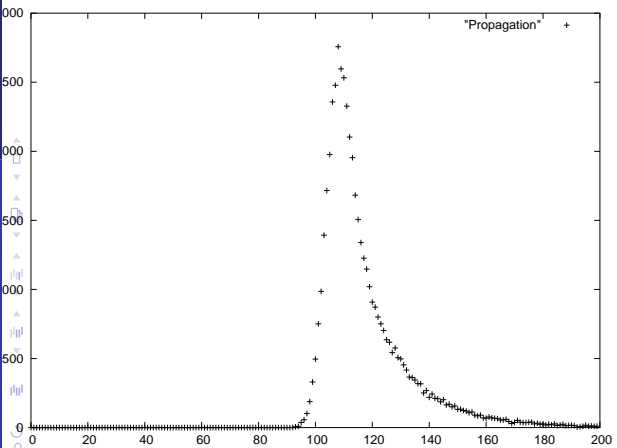
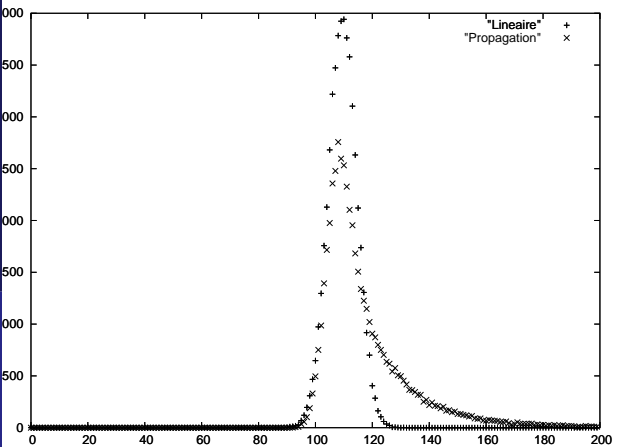
A utiliser lorsque l'hypothèse de linéarité est visiblement mal adaptée. Il s'agit d'une technique de simulation exhaustive.

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.

remarques:

- très couteux car on réalise N fois le processus d'estimation
- On ajoute ainsi du bruit à des données déjà bruitées, mais en pratique, cela a peu d'influence sur le résultat [3].

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



Mesure de la compatibilité d'une mesure avec un modèle : tests statistiques

Exemple introductif Les faiseurs de pluie (Tiré de [1]). Des relevés de pluie effectués en Beauce ont permis d'établir que le niveau naturel des pluies dans la Beauce en mm/an suit une loi de Laplace Gauss LG(600,100). Des entrepreneurs surnommés *faiseurs de pluie* prétendaient pouvoir augmenter le niveau moyen de pluie de 50mm en inséminant les nuages avec de l'iodure d'argent. Ce procédé fut mis à l'essai entre 1951 et 1959 et on releva les hauteurs suivantes:

Année	1951	1952	1953	1954	1955	19956	1957	1858	1959
mm	510	614	780	512	501	534	603	788	650

Que peut on conclure? Il faut confronter les hypothèses

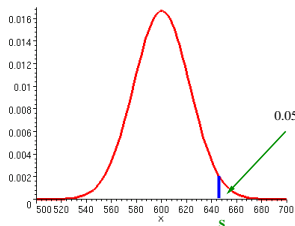
$$H_0 : m = 600mm$$

$$H_1 : m = 650mm$$

Démarche des tests statistiques

Démarche: on choisira H_1 si les faits contredisent nettement H_0 . (H_0 est l'hypothèse nulle et H_1 l'hypothèse alternative). H_0 sera abandonnée si les faits expérimentaux traduisent une éventualité improbable compte tenu de H_0 .

Mise en oeuvre du test: la variable de décision est naturellement la moyenne des observations \bar{X} . Les précipitations théoriques suivant une loi $LG(600,100)$, \bar{X} devrait suivre une loi $LG(600, 100/3)$.



Principe: Certaines valeurs sont rares (improbables) et on prend donc la règle de décision suivante:

- Si \bar{X} est trop grand, ie supérieur à un seuil s qui n'a que 5 chances sur 100 d'être dépassé, alors on optera pour l'hypothèse H_1 .
- Si $\bar{X} < s$, alors on conservera l'hypothèse H_0 faute d'évidence permettant de conclure à H_1 .

Les faiseurs de pluie...

Sur l'échantillon: on a $\bar{X} = 610.2mm$ Et la consultation d'une table gaussienne montre que $s = 600 + \sigma * 1.64 = 654.7$. Ceci conduit donc à rejeter l'hypothèse H_1 et à conserver H_0 .

Remarques:

- L'ensemble $\bar{X} > 655$ est appelé **région critique**.
- Les deux hypothèses H_0 et H_1 ne jouent pas des rôles symétriques. En particulier, le test ne dépend que de l'hypothèse H_0 .
- Calcul du risque de se tromper : croire les faiseurs de pluies alors qu'ils n'étaient pour rien dans le résultat ($r_1 = 0.05$) **ou** ne pas croire les faiseurs de pluie alors que leur méthode est bonne. Si les faiseurs de pluie ont raison $\bar{X} : LG(650, 100/3)$. On commet une erreur lorsque \bar{X} est inférieure à 655 (cad quand H_0 est acceptée) avec une probabilité

$$\beta = P(\bar{X} < 655) = P(U < \frac{655 - 650}{100/3}) = P(U < .15) = 0.56$$

β est appelé risque de deuxième espèce et est très élevé ici... (U est la variable centrée réduite).

On formule une hypothèse H_0 portant sur la valeur d'un paramètre d'une population.

But: porter un jugement sur cette hypothèse sur la base d'un échantillon.

Démarche: décider si l'écart entre l'estimation du paramètre sur l'échantillon et le paramètre théorique est uniquement dû au hasard de l'échantillonnage ou si l'hypothèse doit être rejetée.

Exemples:

- Un échantillon suit-il une loi donnée?
- On approxime un ensemble de points par une droite. Peut-on vérifier que le modèle linéaire est bien adapté?
- Étant données plusieurs sources émettrices, peut-on utiliser si un test pour savoir si une donnée est émise par la source 1 ou la source 2 ?

Principe des tests statistiques

Soient H_1 et H_2 deux hypothèses dont une seule est vraie. Le processus du test aboutit à choisir H_1 ou H_2 avec 4 cas possibles.

verite \rightarrow	H_0	H_1
H_0	$1 - \alpha$	β
H_1	α	$1 - \beta$

α et β sont les risques de première et de deuxième espèce:

- α est la probabilité de choisir H_1 alors que H_0 est vraie
- β est la probabilité de retenir H_0 alors que H_1 est vraie.

En pratique, on se donne α ($=0.005$ par exemple). H_0 joue un rôle particulier et correspond souvent à une hypothèse de prudence.

La région critique W est l'ensemble des valeurs de la variable de décision qui conduisent à écarter H_0 au profit de H_1 : $P(W|H_0) = \alpha$

Étapes d'un test

- Choix de H_0 et H_1
- Détermination de la variable de décision
- Calcul de la région critique en fonction de α
- Calcul de la valeur expérimentale de la variable de décision
- Rejet ou acceptation de H_0

exemple 1: Visualisation de l'incertitude

But: définir un ellipsoïde contenant *l'essentiel* de la distribution à partir de la matrice de covariance.

Soit X une grandeur à estimer. Soit X_0 une estimation de X de matrice de covariance Λ . X_0 peut être assimilé à la réalisation d'une variable aléatoire d'espérance X (ie $X_0 = X + \text{bruit blanc}$).

problème: représenter le domaine dans lequel X peut se trouver avec *une probabilité donnée*.

Sous l'hypothèse d'un bruit blanc, $X - X_0 : \mathcal{N}(0, \Lambda)$.

Donc $(X - X_0)^t \Lambda^{-1} (X - X_0)$ suit une loi du χ^2 à $n (= \dim(X))$ degrés de liberté.

→ il existe un seuil α tel que

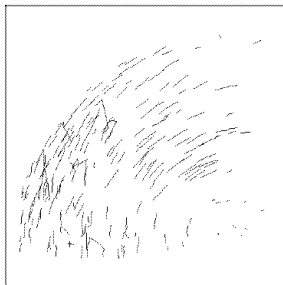
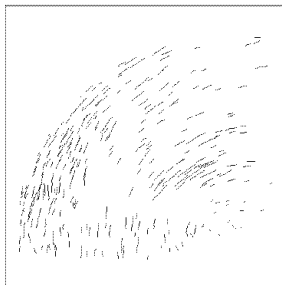
$$\mathcal{P}((X - X_0)^t \Lambda^{-1} (X - X_0) < \alpha) = 90\%$$

→ définit un ellipsoïde de centre X_0 . ex: si $n = 2$: pour 90%, $\alpha = 4.60$,
pour 99%, $\alpha = 9.21$

l'intérêt du test est de fournir un seuil homogène

Compatibilité d'une mesure avec un modèle: distance de Mahalanobis

Exemple: dans une application de suivi dans une sequence d'images, on détecte des particules sans connaitre leur association temporelle. Pour déterminer le candidat potentiel dans l'image suivante, il faut savoir pour chaque point candidat s'il est compatible avec la trajectoire préalablement déterminée.



On a un vecteur d'état $\hat{\omega} = (\theta, t_x, t_y)$ de covariance Λ_ω (le mouvement 2D du marqueur), un vecteur de mesure \hat{x} de covariance Λ_x (détection du marqueur dans la scène). équation de mesure $f(x, a) = 0$.

$$f(x, \omega) = \begin{cases} m'_x - \cos\theta m_x + \sin\theta m_y + t_x = 0 \\ m'_y - \sin\theta m_x - \cos\theta m_y + t_y = 0 \end{cases}$$

objectif: déterminer si \hat{x} est une mesure **plausible** ou une mesure **aberrante**.

Si les bruits entachant les mesures sont blancs alors $x - \hat{x}$ et $\omega - \hat{\omega}$ sont des v.a. gaussiennes de moyenne nulle et de variance Λ_x et Λ_ω .

En utilisant un développement à l'ordre 1:

$$f(x, \omega) = 0 = f(\hat{x}, \hat{\omega}) + \frac{\partial f}{\partial x}(\hat{x}, \hat{\omega})(x - \hat{x}) + \frac{\partial f}{\partial \omega}(\hat{x}, \hat{\omega})(\omega - \hat{\omega}) + \epsilon$$

Les bruits étant gaussiens et supposés indépendants, $f(\hat{x}, \hat{\omega})$ est gaussienne et:

$$E[f(\hat{x}, \hat{\omega})] = 0$$

$$\begin{aligned}
 Q = \text{var}(f(\hat{x}, \hat{\omega})) &= E[f(\hat{x}, \hat{\omega})f(\hat{x}, \hat{\omega})^t] \\
 &\approx \frac{\partial f}{\partial x}(\hat{x}, \hat{\omega})\Lambda_x \frac{\partial f}{\partial x}(\hat{x}, \hat{\omega})^t + \frac{\partial f}{\partial \omega}(\hat{x}, \hat{\omega})\Lambda_\omega \frac{\partial f}{\partial \omega}(\hat{x}, \hat{\omega})^t
 \end{aligned}$$

Alors


$$\delta(\hat{x}, \hat{\omega}) = f(\hat{x}, \hat{\omega})^t Q^{-1} f(\hat{x}, \hat{\omega})$$

suit une distribution du χ^2 à $q(= \dim(f))$ degrés de liberté. Il existe donc α tel que


$$\mathcal{P}(\delta(\hat{x}, \hat{\omega}) < \alpha) = 90\%$$


test: les mesures vérifiant $\delta(\hat{x}, \hat{\omega}) < \alpha$ sont estimées cohérentes avec le vecteur d'état.


Voir les références [4, 5, 2]

 G. Saporta.
Probabilités analyse des données et statistique.
Editions Technip, 1990.

 I. Cox.
A Review of Statistical Data Association for Motion Correspondance.
International Journal of Computer Vision, 10(1):53–66, 1993.

 Csurka, C. Zeller, Z.Y. Zhang, and O.D. Faugeras.
Characterizing the Uncertainty of the Fundamental matrix.
Computer Vision and Image Understanding, 68(1):18–36, May 1997.

 I. K. Sehti and R. Jain.
Finding Trajectories of Feature Points in a Monocular Image
Sequence.
IEEE Transactions on PAMI, 9(1):56–73, 1987.

 Y. Yao and R. Chelappa.
Dynamic Feature Point Tracking in an Image Sequence. ICPR 94.