Initiation à l'Estimation Statistique et Applications Astrométriques

Frédéric Arenou
UMR 8111 du CNRS et Gépi (Observatoire de Paris)

Support de cours au
TD bruit et signaux
Revision: 1.1.1 , Date: 2004/10/15 13:53:31
Mise à jour sur `http://wwwhip.obspm.fr/~arenou`

1 L'estimation statistique

1.1 À quoi servent les statistiques ?

Le problème de l'estimation statistique est aussi ancien que les différentes sciences observationnelles, dès lors qu'il a fallu synthétiser des mesures répétées d'une même grandeur inconnue. Les différentes applications de l'estimation sont les suivantes :

- *inférer les propriétés d'une population à partir d'un échantillon représentatif:* on suppose que l'échantillon que l'on a est extrait d'une population parente dont on connaît la forme de la distribution, et dont on cherche la meilleure valeur des paramètres qui la caractérisent (estimation *ponctuelle*) ou un intervalle de confiance qui contienne ces paramètres avec une certaine probabilité (estimation *d'intervalle*). Extraire un signal du bruit de mesure en est une application analogue.

- *choisir entre différentes hypothèses:* par exemple savoir si la loi des erreurs sur des données est ou non gaussienne, ou bien être certain qu'une observation que l'on vient d'obtenir est (significativement) différente de celle qu'un modèle prédirait.

1.2 Quelques remarques

- *inférence:* le but de l'estimation statistique est d'analyser les évènements passés, et éventuellement de prédire les évènements futurs. Le mot « prédire » n'est pas innocent, car il sous-entend le risque de se tromper : même si les méthodes utilisées proviennent des mathématiques, où c'est la *déduction* qui est utilisée, l'estimation statistique utilise l'*inférence*, tout comme la physique. C'est donc une interprétation du monde (une modélisation réductrice), le but est que c'en soit la plus probable...

- *grandeurs physiques:* toute variable observée a une erreur de mesure aléatoire (dont on doit donc indiquer la dispersion). Ex : si l'on mesure des parallaxes trigonométriques, on doit logiquement s'attendre à obtenir des parallaxes négatives, même si l'on sait que la vraie parallaxe est positive. Toute analyse de données doit donc prendre les erreurs de mesure en compte.

- *échantillon:* les résultats d'une analyse statistique dépendent clairement de la taille de l'échantillon et de sa représentativité ; les sondages d'opinion en sont un exemple.

- *interprétation:* lors de l'interprétation des résultats d'une analyse, il ne faut pas confondre corrélation et causalité.

- *biais:* en plus des erreurs aléatoires sur les données observationnelles, on peut parfois s'attendre à des erreurs systématiques.

Parmi les différentes causes de ces biais : problèmes instrumentaux (par exemple déformation de plaques photographiques), échantillon non représentatif, et en particulier à cause d'une censure (ex : on observe des étoiles jusqu'à une certaine magnitude apparente limite, donc on privilégie les plus intrinsèquement brillantes, donc la magnitude absolue moyenne observée est biaisée), présence de points aberrants (si l'estimateur que l'on utilise y est sensible)

---

2 Probabilités

2.1 Conventions

Une variable aléatoire (v.a.) est une fonction à valeur réelle (ou un vecteur dont les composantes sont à valeurs réelles). Pour une v.a. X et sa réalisation x, on notera souvent $f(x)$ au lieu de $f_X(x)$ sa densité de probabilité.

On s'intéressera essentiellement à des fonctions continues. Dans le cas discret où $\Omega = (x_1, \ldots, x_n)$ est l'ensemble des valeurs possibles de la v.a. X, et en notant $p_i = P(X = x_i)$ la probabilité de réalisation, il suffit de substituer les sommations aux intégrales et p_i à $f(x)$.

On utilisera les lettres grecques pour les paramètres inconnus, les autres lettres pour les estimations empiriques : $m \rightarrow \mu$, $s \rightarrow \sigma$; $\widehat{\theta}$ désigne un estimateur de la valeur inconnue θ, le signe $\rightsquigarrow$ indique qu'une v.a. suit une certaine loi de probabilité. On note souvent $\overline{x}$ ou $< x >$ la valeur moyenne. Pour qu'il n'y ait pas de confusion, on note dans ce qui suit ϖ la parallaxe d'une étoile et π le nombre utile aux sages.

2.2 Fonction de répartition (distribution)

$$F_X(x) = P(X \leq x), \quad x \in [-\infty, +\infty]$$

- *Propriétés:*

$$F(x) \in [0,1], \quad F(-\infty) = 0, \quad F(+\infty) = 1$$

$$F(x) \leq F(x'), \quad \forall x \leq x'$$

2.3 Densité de probabilité (p.d.f.)

$$f(x) = \frac{dF}{dx}$$

Pour qu'une fonction f soit une densité, il faut donc au moins que $f(x) \geq 0$ et d'intégrale 1.

- *densité marginale en X d'une loi $f(x, y)$:*

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y)dy$$

$$P(X \leq a) = \int_{-\infty}^{a} f_X(x)dx$$

- *indépendance:* X et Y sont indépendantes $\Longleftrightarrow$

$$f(x, y) = f_X(x)f_Y(y), \quad \forall(x, y)$$

- *densité conditionnelle:*

$$f(x \mid y) = \frac{f(x, y)}{f_Y(y)}$$

$$P(a \leq X \leq b \mid Y = y) = \int_a^b f(x \mid y)dx$$

---

3 Moments

3.1 Espérance mathématique

$$E[X] = \int_{-\infty}^{+\infty} xf(x)dx = \mu$$

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)dx$$

Quand g n'est pas une fonction linéaire, il faudra donc s'attendre en général à ce que $E[g(X)] \neq g(E[X])$, ou, formulé autrement, si $E[X] = \mu$, ceci signifie que $g(X)$ peut être un estimateur biaisé de $g(\mu)$.

3.1.1 Application

Soit ϖ_0 une parallaxe mesurée avec la précision σ, estimateur supposé non biaisé de la parallaxe ϖ d'une étoile. Si l'on cherche la distance de l'étoile ($r = \frac{1}{\varpi}$), il paraît naturel d'utiliser $\frac{1}{\varpi_0}$. Vérifions si cet estimateur de la vraie distance r est ou non biaisé, dans le cas où les erreurs sont gaussiennes :

$$\begin{aligned} E[\frac{1}{\varpi_0}] - \frac{1}{\varpi} &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \frac{1}{\varpi_0} e^{-\frac{(\varpi_0-\varpi)^2}{2\sigma^2}} d\varpi_0 - \frac{1}{\varpi} \\ &= \frac{-1}{\varpi\sqrt{2\pi}} \int_{-\infty}^{+\infty} (1 - \frac{1}{1 + u\frac{\sigma}{\varpi}}) e^{-\frac{u^2}{2}} du \quad (1) \\ &\neq 0 \text{ en général, dès que } \frac{\sigma}{\varpi} \neq 0 \end{aligned}$$

La distance calculée avec la parallaxe observée est donc biaisée, avec un biais $\approx \frac{\sigma^2}{\varpi^3} + 3\frac{\sigma^4}{\varpi^5} + \ldots$ aux premiers ordres en $\frac{\sigma}{\varpi}$. Ce biais est aggravé quand on ne conserve que les parallaxes positives.

La démonstration est analogue pour le calcul de la magnitude absolue en utilisant la loi de Pogson.

3.2 Variance

$$\begin{aligned} \sigma^2(X) &= E\left[(X - E[X])^2\right] &= E[X^2] - E^2[X] \\ &= \int_{-\infty}^{+\infty} (x - \mu)^2 f(x)dx &= \int_{-\infty}^{+\infty} x^2 f(x)dx - \mu^2 \end{aligned}$$

3.3 Écart-type $\sigma(X)$

C'est la racine carrée de la variance. Pour la désigner, on rencontrera souvent les termes d'erreur, de précision, de dispersion. L'erreur interne (ou formelle) est celle qui est obtenue par la méthode d'estimation utilisée, par opposition à l'erreur externe.

3.4 Covariance

$$\begin{aligned} \mathrm{Cov}(X, Y) &= E\left[(X - E[X])(Y - E[Y])\right] \\ &= \int_{-\infty}^{+\infty} (x - \mu_X)(y - \mu_Y) f(x, y)dxdy \\ &= \int_{-\infty}^{+\infty} xyf(x, y)dxdy - \mu_X\mu_Y \end{aligned}$$

On a $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$ et, si a et b sont des réels,

$$\mathrm{Cov}(aX + bY, Z) = a\mathrm{Cov}(X, Z) + b\mathrm{Cov}(Y, Z)$$

Dans le cas multidimensionnel d'un vecteur $\mathbf{X} = (X_i)$, on introduit la matrice de variance-covariance

$$\mathbf{V} = E\left[(\mathbf{X} - E[\mathbf{X}])(\mathbf{X} - E[\mathbf{X}])^T\right] = (\mathrm{Cov}(X_i, X_j))$$

dont la diagonale est formée des variances, et qui est définie non-négative.

3.5 Corrélation

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

Dans le cas multidimensionnel, soit $\mathbf{X} = (X_i)$, d'écart-type (σ_i), et soient $\Delta_i = \frac{X_i - E[X_i]}{\sigma_i}$ les données normalisées. La matrice $\mathbf{R} = E\left[\mathbf{\Delta}\mathbf{\Delta}^T\right]$ est la matrice de corrélation, de terme général $(\rho(X_i, X_j))$, et également définie non-négative.

On a toujours $-1 \leq \rho(X, Y) \leq 1$, et d'autre part $\rho(X, Y) = 0$ si les deux variables ne sont pas corrélées. Noter que l'indépendance implique la non-corrélation, mais que l'inverse n'est pas forcément vrai (sauf dans le cas Gaussien). Si X et Y sont complètement corrélées, $|\rho(X, Y)| = 1$, c'est qu'il existe des réels a, b, c tels que $aX + bY = c$.

Si X, Y sont des v.a. et a, b des réels, on a

$$\mathrm{Var}(aX + bY) = a^2\sigma^2(X) + 2ab\rho(X, Y)\sigma(X)\sigma(Y) + b^2\sigma^2(Y)$$

3.5.1 Application

Les données d'Hipparcos furent réduites par deux Consortiums, avec des résultats donc corrélés. Comment déterminer les variations du coefficient de corrélation des parallaxes ?

On ne peut pas utiliser l'estimateur empirique

$$R = \frac{\sum_{i=1}^n (\varpi_{\mathrm{F}i} - \overline{\varpi_{\mathrm{F}}})(\varpi_{\mathrm{N}i} - \overline{\varpi_{\mathrm{N}}})}{\sqrt{\sum_{i=1}^n (\varpi_{\mathrm{F}i} - \overline{\varpi_{\mathrm{F}}})^2 \sum_{i=1}^n (\varpi_{\mathrm{N}i} - \overline{\varpi_{\mathrm{N}}})^2}}$$

parce que les données proviennent de populations différentes : d'abord toutes les étoiles n'ont pas la même parallaxe, ensuite les précisions sont variables d'une étoile l'autre (se dégradant essentiellement avec la magnitude).

Par contre, comme $\varpi_{\mathrm{F}i}$ et $\varpi_{\mathrm{N}i}$ sont de moyenne ϖ_i (la vraie parallaxe de l'étoile i) et de précision respective $\sigma_{\mathrm{F}i}$ et $\sigma_{\mathrm{N}i}$, on remarque que toutes les différences normalisées

$$\Delta_i = \frac{\varpi_{\mathrm{F}i} - \varpi_{\mathrm{N}i}}{\sqrt{\sigma^2_{\mathrm{F}i} - 2\rho\sigma_{\mathrm{F}i}\sigma_{\mathrm{N}i} + \sigma^2_{\mathrm{N}i}}}$$

suivent la même loi de moyenne nulle et d'écart-type 1.

Pour chaque étoile, en notant

$$\delta_i = \frac{\varpi_{\mathrm{F}i} - \varpi_{\mathrm{N}i}}{\sqrt{\sigma^2_{\mathrm{F}i} + \sigma^2_{\mathrm{N}i}}}$$

on a

$$\Delta_i = \delta_i \frac{1}{\sqrt{1 - \rho\left(\frac{2\sigma_{\mathrm{F}i}\sigma_{\mathrm{N}i}}{\sigma^2_{\mathrm{F}i}+\sigma^2_{\mathrm{N}i}}\right)}}$$

Comme $\mathrm{Var}(\Delta_i) = 1$, et en supposant l'indépendance des termes du produit, on peut estimer la corrélation sur un échantillon par

$$\widehat{\rho} \approx (\frac{1}{n}\sum_{i=1}^{n} \frac{\sigma^2_{\mathrm{F}i} + \sigma^2_{\mathrm{N}i}}{2\sigma_{\mathrm{F}i}\sigma_{\mathrm{N}i}})(1 - \frac{1}{n}\sum_{i=1}^{n}\delta_i^2)$$

Si l'approximation ci-dessus est contestable, on peut toujours la vérifier en faisant des bins en σ_{F} et σ_{N}, dans lesquels on peut calculer

$$\widehat{\rho_j} \approx \frac{\sigma^2_{\mathrm{F}} + \sigma^2_{\mathrm{N}} - \frac{1}{n}\sum_{i=1}^{n}(\varpi_{\mathrm{F}i} - \varpi_{\mathrm{N}i})^2}{2\sigma_{\mathrm{F}}\sigma_{\mathrm{N}}}$$

et vérifier s'il reste constant pour tous les bins j.

3.6 Moment d'ordre r

$$E\left[X^r\right] = \int_{-\infty}^{+\infty} x^r f(x)dx$$

et le moment centré est :

$$E\left[(X-\mu)^r\right] = \int_{-\infty}^{+\infty} (x-\mu)^r f(x)dx$$

3.7 Quantiles

Q_α est un quantile $(1-\alpha)$ si $P(X \leq Q_\alpha) = 1-\alpha$. En particulier, la médiane est $Q_{0.5}$

3.8 Mode

C'est un maximum de la pdf (il peut y en avoir plusieurs). Une distribution multimodale est souvent le signe d'un mélange de populations. Pour une distribution unimodale, on a dans l'ordre (croissant ou décroissant suivant l'asymétrie) : mode, médiane, moyenne. Les trois sont confondus si la distribution est symétrique.

4 Estimation bayésienne

4.1 Théorème de Bayes

Pour 2 évènements A et B, la probabilité conjointe est

$$P(A \cap B) = P(A \mid B)P(B) = P(B \mid A)P(A)$$

$$\text{donc } f(\theta \mid x) = \frac{f(x \mid \theta)f(\theta)}{\int_{-\infty}^{+\infty} f(x \mid \theta)f(\theta)d\theta}$$

$f(\theta)$ est la loi *a priori*, $f(\theta \mid x)$ est la loi *a posteriori* et $f(x \mid \theta)$ est nommée la vraisemblance.

Il ne s'agit pas d'une méthode statistique parmi d'autres, mais bien d'un concept différent : ici on probabilise l'inconnu (θ, considéré constant d'ordinaire), d'une part, et on choisit une loi *a priori*, avec l'aspect subjectif que cela comporte (mais toute connaissance n'est-elle pas subjective ?). D'où le conflit entre bayésiens et fréquentistes.

4.2 loi *a priori*

si la loi $f(\theta)$ est inconnue, on la choisit en général :

- uniforme ($f(\theta)$ constant) pour un paramètre de position (comme la moyenne)
- inverse ($f(\theta) \propto 1/\theta$) pour un paramètre d'échelle (comme l'écart-type)

4.3 Espérance *a posteriori*

$$E[\theta \mid X] = \frac{\int_{-\infty}^{+\infty} \theta f(x \mid \theta)f(\theta)d\theta}{\int_{-\infty}^{+\infty} f(x \mid \theta)f(\theta)d\theta}$$

C'est l'estimateur qui minimise l'espérance *a posteriori* de la perte quadratique

$$\min_{\widehat{\theta}} \int_{-\infty}^{+\infty} (\widehat{\theta} - \theta)^2 f(\theta \mid x)d\theta$$

4.3.1 Application

Si la loi conditionnelle est gaussienne $N(\varpi_0\,;\varpi,\sigma)$, on a alors

$$E[\varpi \mid \varpi_0] = \varpi_0 + \sigma^2 \frac{f'(\varpi_0)}{f(\varpi_0)}$$

Utile quand on a fait une censure sur les parallaxes observées ϖ_0 (e.g. en sélectionnant un échantillon d'étoiles plus près que p parsecs) cet estimateur de la vraie parallaxe a l'avantage de ne pas nécessiter de se donner de loi *a priori*, puisque seule la distribution observée intervient.

4.3.2 Application

On a vu que l'estimation de la distance d'une étoile en prenant l'inverse de la parallaxe observée était biaisée (eq. 1). Mais d'autres estimateurs sont possibles, comme l'espérance *a posteriori* de la distance :

$$E[\frac{1}{\varpi} \mid \varpi_0] = \frac{\int_0^{+\infty} \frac{1}{\varpi} f(\varpi_0 \mid \varpi)f(\varpi)d\varpi}{\int_0^{+\infty} f(\varpi_0 \mid \varpi)f(\varpi)d\varpi}$$

À partir de là, soit l'on suppose ne rien connaître de la loi des vraies parallaxes, et l'on prend $f(\varpi)$ uniforme (entre $\varpi_{\min}$ et

$\varpi_{\max}$), et, dans le cas gaussien, il reste à calculer numériquement

$$E[\frac{1}{\varpi} \mid \varpi_0] = \frac{1}{\sigma\sqrt{2\pi}} \int_{\varpi_{\min}}^{\varpi_{\max}} \frac{1}{\varpi} e^{-\frac{(\varpi_0-\varpi)^2}{2\sigma^2}} d\varpi$$

Mais il serait faux d'affirmer que l'on n'a aucune connaissance *a priori* de la distribution en parallaxe d'un échantillon d'étoiles : pour une distribution sphérique, de densité spatiale uniforme, le nombre d'étoile croît comme r^3 avec la distance r, soit $f(r) \propto r^2$, et $f(\varpi) = f(r) \mid \frac{d\varpi}{dr} \mid$, donc $f(\varpi) \propto \frac{1}{\varpi^4}$. Ou encore plus réaliste avec une distribution exponentielle dans le plan galactique, et une autre en Z, et en tenant compte des informations sur la luminosité de l'étoile, sa vitesse spatiale, etc. Il est clair que plus on aura introduit d'information *a priori*, plus précis seront les paramètres obtenus, mais à condition que cette modélisation soit correcte...

---

5 En pratique

On suppose que l'on a un n-échantillon $x_1, .., x_n$, réalisation des v.a. $X_1, .., X_n$.

5.1 Statistique

C'est une fonction $g(x_1, .., x_n)$, qui est une réalisation de la v.a. $g(X_1, .., X_n)$, comme par exemple la moyenne de l'échantillon

$$m = \frac{1}{n} \sum_{i=1}^{n} x_i$$

estimation de l'espérance de la population parente. Un estimateur de $E[X] = \mu$ est

$$M = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Ne pas oublier qu'un estimateur est lui-même une v.a., donc possédant une distribution.

5.2 Distribution empirique

$$F_n(x) = \frac{(\text{nombre de } x_i \leq x)}{n}$$

On représente classiquement une distribution empirique à l'aide d'un histogramme. Mais quel pas faut-il choisir, et en commençant à partir de quelle valeur ? Il vaut mieux estimer la densité à l'aide d'autres méthodes (par ex. noyau de convolution)

5.3 Moments empiriques

d'ordre 1 : $m = \frac{1}{n} \sum_{i=1}^{n} x_i$
Si μ n'est pas connu, il est estimé par m, et alors un estimateur non biaisé de la variance est

$$s_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - m)^2$$

et une approximation non biaisée de l'écart-type est

$$s_n \approx \sqrt{\frac{1}{n-1.45} \sum_{i=1}^{n} (x_i - m)^2}$$

5.4 Statistique d'ordre

On note $x_{(1)} \leq .. \leq x_{(n)}$ l'échantillon trié par ordre croissant : $x_{(1)}$ et $x_{(n)}$ sont alors les extrêmes de l'échantillon, et $x_{(n)} - x_{(1)}$ en est l'étendue.

5.5 Quantile q_α

q_α est un quantile $(1-\alpha)$ de l'échantillon si

$$\frac{(\text{nombre de } x_i < q_\alpha)}{n} \leq 1 - \alpha \leq \frac{(\text{nombre de } x_i \leq q_\alpha)}{n}$$

En particulier la médiane est alors

$$q_{0.5} = \begin{cases} x_{(\frac{n+1}{2})} & \text{si } n \text{ est impair,} \\ \frac{1}{2}(x_{(\frac{n}{2})} + x_{(\frac{n+2}{2})}) & \text{sinon.} \end{cases}$$

---

6 Loi binomiale

6.1 Définition

Probabilité de x succès sur n essais ayant deux résultats possibles, de probabilités respectives p et $1-p$

6.2 Densité

$$b(x\,;n,p) = C_n^x p^x (1-p)^{n-x}$$

6.3 Moments

$$E[X] = np \qquad \sigma(X) = \sqrt{np(1-p)}$$

6.4 Propriétés

La somme de variables binomiales indépendantes de probabilité p est binomiale : si $X_i \rightsquigarrow b(x\,;n_i,p)$, alors

$$Y = \sum_{i=1}^{k} X_i \rightsquigarrow b(y\,;\sum_{i=1}^{k} n_i, p)$$

6.5 Convergence

- Vers loi de Poisson : si $n \to +\infty$ et $np \to \lambda \neq 0$, alors

$$b(x\,;n\,;p) \to p(x\,;\lambda) \qquad (n \gtrsim 20)$$

- Vers loi normale : si $n \to +\infty$ et $p \in]0,1[$, alors

$$b(x\,;n\,;p) \to N(x\,;np,\sqrt{np(1-p)}) \qquad (n \gtrsim 36)$$

---

7 Loi de Poisson

7.1 Définition

Probabilité d'apparition d'un évènement rare (en moyenne $\lambda \neq 0$) sur un grand nombre d'observations

7.2 Densité

$$p(x\,;\lambda) = \frac{\lambda^x}{x!}e^{-\lambda}$$

7.3 Moments

$$E[X] = \lambda \qquad \sigma(X) = \sqrt{\lambda}$$

7.4 Propriétés

La somme de variables de Poisson indépendantes est de Poisson : si $X_i \rightsquigarrow p(x\,;\lambda_i)$, alors

$$Y = \sum_{i=1}^{k} X_i \rightsquigarrow p(y\,;\sum_{i=1}^{k}\lambda_i)$$

7.5 Convergence

Vers loi normale : si $\lambda \to +\infty$, alors

$$p(x\,;\lambda) \to N(x\,;\lambda,\sqrt{\lambda}) \qquad (\lambda \gtrsim 20)$$

7.6 Applications

- Files d'attente
- Nombre de photons reçus sur un récepteur

---

8 Loi Uniforme

- Définition
Equiprobabilité de se trouver dans un intervalle $[a, b]$
- Densité

$$u(x\,;a,b) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a,b] \\ 0 & \text{sinon.} \end{cases}$$

- Moments

$$E[X] = \frac{a+b}{2} \qquad \sigma(X) = \frac{b-a}{2\sqrt{3}}$$

- Propriétés
la loi la plus simple en l'absence d'autres informations...
- Applications
erreur d'arrondi dans les calculs

---

9 Loi Exponentielle

- Définition
Probabilité d'attendre un temps $> x$ quand $\frac{1}{\alpha}$ est le temps moyen
- Densité

$$e(x\,;\alpha) = \alpha e^{-\alpha x} \text{ pour } x \geq 0 \text{ et } \alpha > 0$$

- Moments

$$E[X] = \frac{1}{\alpha} \qquad \sigma(X) = \frac{1}{\alpha}$$

- Propriétés

- Loi sans mémoire : $P(X > x + x' \mid X > x') = P(X > x)$
- Si le nombre d'apparitions d'un phénomène pendant le temps t suit une loi $p(x\,;\alpha t)$ alors la distribution du temps entre deux apparitions suit une loi $e(t\,;\alpha)$
- Si $X \rightsquigarrow u(x\,;a,b)$ alors

$$Y = \frac{-\log[(b-X)/(b-a)]}{b-a} \rightsquigarrow e(y\,;b-a)$$

- Y=Min$(e(x\,;\alpha_1),\ldots,e(x\,;\alpha_k)) \rightsquigarrow e(y\,;\sum_{i=1}^{k}\alpha_i)$

- Applications

- Durée de vie d'une pièce
- Intervalle de temps entre deux pannes
- Durée de service dans une file d'attente
- Distribution stellaire: doublement exponentielle (loi de Laplace) dans le plan galactique et en Z (la hauteur au-dessus du plan)

---

10 Loi Normale ou Gaussienne

- Définition
Influence d'un grand nombre de facteurs aléatoires, indépendants, petits et additifs
- Densité

$$N(x\,;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \text{ pour } \sigma \geq 0$$

$$N(x\,;0,1) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \text{ est la loi centrée réduite}$$

Dans le cas multidimensionnel (vecteur moyenne μ de dimension p, matrice $p \times p$ de variance-covariance $\mathbf{V}$),

$$N(\mathbf{X}\,;\mu,\mathbf{V}) = \frac{1}{(2\pi)^{p/2}|\mathbf{V}|^{1/2}}e^{-\frac{(\mathbf{X}-\mu)^T\mathbf{V}^{-1}(\mathbf{X}-\mu)}{2}}$$

- Moments

$$E[X] = \mu \qquad \sigma(X) = \sigma$$

- Propriétés
La somme de variables normales indépendantes est normale : si $X_i \rightsquigarrow N(x\,;\mu_i,\sigma_i)$ et si les a_i sont des constantes, alors

$$\sum_{i=1}^{k} a_i X_i \rightsquigarrow N(x\,;\sum_{i=1}^{k} a_i\mu_i, \sqrt{\sum_{i=1}^{k} a_i^2\sigma_i^2})$$

- Convergence
Théorème Central Limite (TCL) : soient $X_1 \dots X_n$ des variables indépendantes et identiquement distribuées (suivant n'importe quelle loi), de moyenne μ et de variance σ^2. Quand $n \to +\infty$, alors

$$Y = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \rightsquigarrow N(y\,;0,1) \qquad \begin{cases} n \gtrsim 30 & \text{si loi symétrique,} \\ n \gtrsim 60 & \text{sinon} \end{cases}$$

Ce théorème est fondamental, en particulier car : 1) la moyenne est un estimateur non biaisé de la valeur centrale, 2) sa précision améliore d'un facteur $\sqrt{n}$ celle des données individuelles, 3) la distribution des moyennes tend vers une gaussienne.
- Applications
La loi la plus utilisée (à cause du TCL), parfois abusivement.

---

11 Loi de Cauchy

- Définition
Rapport de deux variables iid $N(x\,;0,1)$
- Densité

$$C(x) = \frac{1}{\pi(x^2+1)}$$

ou pour généraliser

$$C(x\,;a,b) = \frac{b}{\pi((x-a)^2+b^2)} \text{ où } b>0$$

- Moments
Aucun ! Même si a et b ressemblent respectivement à des facteurs de position et d'échelle, la moyenne et l'écart-type de cette loi ne sont pas définis.
- Propriétés
La somme de variables de Cauchy indépendantes est de Cauchy : si $X_i \rightsquigarrow C(x\,;a_i,b_i)$, alors

$$Y = \sum_{i=1}^{k} X_i \rightsquigarrow C(y\,;\sum_{i=1}^{k} a_i, \sum_{i=1}^{k} b_i)$$

11.0.1 Application

Cette loi, que l'on préfèrerait éviter, compte-tenu de son absence de moments, se rencontre par exemple dans le cas suivant : pour des étoiles lointaines ($\varpi \approx 0$) avec un petit mouvement propre en ascension droite ou en déclinaison ($\mu \approx 0$), et dont la précision de mesure de ces deux quantités est du même ordre de grandeur ($\sigma_{\mu_0} \approx \sigma_{\varpi_0}$), la vitesse tangentielle est proportionnelle au rapport de $N(\mu_0\,;0,\sigma_{\mu_0})$ sur $N(\varpi_0\,;0,\sigma_{\varpi_0})$, donc elle suit une distribution qui devrait ressembler à celle de Cauchy ; on verra plus loin que la médiane serait un estimateur plus approprié que la moyenne arithmétique pour calculer la moyenne des vitesses tangentielles d'un groupe d'étoiles.

---

12 Loi du Khi-deux (χ^2)

- Définition
Somme des carrés de ν variables iid $N(x\,;0,1)$
- Densité

$$\chi^2(x\,;\nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{(\nu-2)/2} e^{-x/2} \text{ pour } x \geq 0$$

où $\Gamma(k) = \int_0^{+\infty} y^{k-1}e^{-y}dy$ (si k entier, $\Gamma(k) = (k-1)\,!$)
- Moments

$$E[X] = \nu \qquad \sigma(X) = \sqrt{2\nu}$$

- Propriétés
La somme de variables χ^2 indépendantes est χ^2 : si $X_i \rightsquigarrow \chi^2(x\,;\nu_i)$, alors

$$Y = \sum_{i=1}^{k} X_i \rightsquigarrow \chi^2(y\,;\sum_{i=1}^{k} \nu_i)$$

- Convergence
Si $X \rightsquigarrow \chi^2(x\,;\nu)$, quand $\nu \to +\infty$:

- $X \to N(x\,;\nu,\sqrt{2\nu})$
- $Y = \sqrt{2X} \to N(y\,;\sqrt{2\nu-1},1) \qquad (n \gtrsim 30)$

---

13 Simulations

Dans de nombreux cas, on peut être amené à effectuer des simulations des lois que suivent les données sur lesquelles on travaille. C'est le cas quand ces lois sont trop compliquées pour obtenir des résultats analytiques (dans le cas contraire, il n'est jamais inutile de vérifier les résultats obtenus analytiquement...)

13.1 Générateur d'une loi uniforme

On utilise en général une méthode congruentielle (mais il faut une longue période). C'est une fonction qui s'appelle `rand` dans la plupart des langages informatiques.

$$a_{n+1} = ia_n + j \qquad (\text{mod } m)$$

$X = \frac{a_n}{m}$ suit alors une loi (pseudo-aléatoire) uniforme entre 0 et 1, et $Y = a + (b-a)X \rightsquigarrow u(y\,;a,b)$.

13.2 Autres distributions :

Une fois que l'on sait générer une valeur uniforme, on peut générer une valeur suivant une autre loi, par l'une des méthodes ci-dessous :

- *cas où l'on peut utiliser des propriétés de la loi:* Par exemple pour les lois classiques ci-dessous :

  – exponentielle : $Y = \frac{-\log[(b-X)/(b-a)]}{b-a} \rightsquigarrow e(y\,;b-a)$ si $X \rightsquigarrow u(x\,;a,b)$

  – Poisson : loi $p(x\,;\alpha t)$ si $\Delta t \rightsquigarrow e(t\,;\alpha)$

  – Normale : en utilisant le TCL, si u_i suit une loi uniforme entre 0 et 1, $\sqrt{12/n}\sum_{i=1}^{n}(u_i - 0.5) \rightsquigarrow N(0,1)$. Mais il y a d'autres algorithmes plus efficaces pour cette loi.

  – Cauchy : rapport de deux $N(0,1)$

- *cas où $F^{-1}(y)$ est facile à calculer:* On utilise la propriété que $F(x)$ suit une loi uniforme. On tire $Y \rightsquigarrow u(y\,;0,1)$, puis on calcule $x = F^{-1}(y)$ pour obtenir une variable X qui suivra la loi désirée (ex : Cauchy)
- *sinon, méthode du rejet:* tirer x uniforme dans l'intervalle $[x_{\min}, x_{\max}]$, puis tirer y uniforme dans $[0, y_{\max}]$, où $y_{\max} > \max f(x)$, et garder la réalisation x si $y \leq f(x)$. Cette méthode est évidemment pénalisante en temps-calcul.

---

14 Estimation ponctuelle

14.1 Détermination d'un paramètre (ou vecteur)

L'estimation ponctuelle consiste à associer une valeur unique obtenue de l'échantillon à un paramètre de la population. Parfois, l'échantillon ne représente pas bien la population parente (données censurées ou erronées), et l'estimateur doit être choisi en conséquence.

14.2 Qualité des estimateurs

Quand on veut connaître la valeur centrale d'un échantillon, le premier réflexe est d'en calculer la moyenne arithmétique. En fait, il existe bien d'autres estimateurs. Nous verrons que la méthode d'estimation est fonction de la distribution des erreurs. On peut donc se demander ce que sont en général les qualités que l'on peut demander à un estimateur.

14.2.1 Application

Si l'on veut calculer la distance moyenne d'un amas en utilisant les parallaxes observées des étoiles ϖ_{0i}, deux estimateurs sembleraient à première vue équivalents : la moyenne des distances individuelles $m_1 =< \frac{1}{\varpi_{0i}} >$ ou bien l'inverse de la moyenne des parallaxes $m_2 = \frac{1}{<\varpi_{0i}>}$. Lequel des deux choisir ?

On a calculé au §3.1.1 le biais B_i sur la distance individuelle de l'étoile i. Sur la moyenne m_1 de ces distances, le biais est donc $< B_i >$, globalement équivalent à chaque biais individuel. L'estimateur m_2 est également biaisé, et son biais s'obtient en substituant $\frac{\sigma}{\sqrt{n}}$ à σ dans l'equation 1 (car c'est la précision sur la moyenne des parallaxes). Quand la taille de l'échantillon augmente, le biais de m_2 tend ainsi vers 0, rendant cet estimateur nettement préférable à m_1. De plus, on peut montrer que la variance de m_2 est également plus petite que celle de m_1.

- *biais contre precision:* intuitivement, on préférerait un estimateur non-biaisé. Mais parfois, il vaut mieux disposer d'un estimateur biaisé mais de petite variance. Plusieurs méthodes existent pour corriger du biais (...au risque d'augmenter la variance !).

---

15 Qualité des estimateurs

Soit $\widehat{\theta}_n$ un estimateur de θ, calculé à partir d'un n-échantillon.

15.1 Convergence

$\widehat{\theta}_n$ est un estimateur convergent si

$$\forall \epsilon > 0\,; \lim_{n\to+\infty} P(|\widehat{\theta}_n - \theta| \geq \epsilon) = 0$$

15.2 Absence de biais

Le biais d'un estimateur $\widehat{\theta}_n$ est

$$B_n(\theta) = E[\widehat{\theta}_n] - \theta$$

Si $\lim_{n\to+\infty} B_n(\theta) = 0$, l'estimateur est dit asymptotiquement correct.

15.3 Optimalité

$\widehat{\theta}_n$ est un estimateur optimal s'il est à la fois convergent, non-biaisé, et de variance inférieure à celle de tout autre estimateur.

15.4 Robustesse (ou fiabilité)

$\widehat{\theta}_n$ est robuste s'il a une faible sensitivité en cas d'écart aux hypothèses initiales. Par exemple, s'il y a des points aberrants, ou une contamination par une autre loi.

---

16 Efficacité d'un estimateur

16.1 Information de Fisher

Si les v.a. X_i sont indépendantes, la vraisemblance du n-échantillon est le produit des vraisemblances individuelles

$$\mathcal{L}(x_1, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta)$$

$$I_n(\theta) = E\left[\left(\frac{\partial \log \mathcal{L}}{\partial \theta}\right)^2\right] = -E\left[\frac{\partial^2 \log \mathcal{L}}{\partial \theta^2}\right]$$

est nommée l'information de Fisher contenue dans le n-échantillon

16.2 Inégalité de Fréchet-Darmois-Rao-Cramer

Si $\widehat{\theta}_n$ est un estimateur non biaisé de θ, alors sa variance est supérieure à la borne de Fréchet :

$$\mathrm{Var}_\theta(\widehat{\theta}_n) \geq \frac{1}{I_n(\theta)}$$

avec $I_n(\theta) = nI_1(\theta)$ si l'échantillon est IID. Ce théorème est vrai sous certaines conditions, en particulier que l'on puisse échanger intégration (due à l'espérance) et dérivation (par rapport à θ).

16.3 Estimateur efficace (MVB)

C'est un estimateur non biaisé dont la variance atteint la borne de Fréchet (donc le plus précis des estimateurs non-biaisés). Par exemple, dans le cas d'un n-échantillon suivant une loi normale $N(x\,;\mu,\sigma)$, la moyenne arithmétique m est MVB. En effet, dans ce cas, $\frac{\partial^2 \log \mathcal{L}}{\partial \mu^2} = \frac{n}{\sigma^2}$, donc $\mathrm{Var}(m) = \frac{\sigma^2}{n} = \frac{1}{I_n(\theta)}$

---

17 Quelques estimateurs

17.1 Du centre de la distribution

Estimateur	Définition	variance asymptotique		
		normale $N(0,1)$	uniforme $u(0,1)$	Cauchy $C(0,1)$
Moyenne arithmétique	$\frac{1}{n}\sum_{i=1}^{n} x_i$	$\frac{1}{n}$	$\frac{1}{12n}$	∞
Médiane	$q_{(0.5)}$	$\frac{\pi}{2n}$	$\frac{1}{4n}$	$\frac{\pi^2}{4n}$
Milieu	$\frac{x_{(1)}+x_{(n)}}{2}$	$\frac{\pi^2}{24\log n}$	$\frac{1}{2n^2}$	∞
Moyenne tronquée (on a choisi $\frac{r}{n} = 0.2$)	$\frac{1}{n-2r}\sum_{i=r+1}^{n-r} x_i$	$\frac{1.14}{n}$	$\frac{0.15}{n}$	$\frac{2.87}{n}$
Moyenne winsorisée (on a choisi $\frac{r}{n} = 0.2$)	$\frac{1}{n}\left(rx_{(r+1)} + \sum_{i=r+1}^{n-r} x_{(i)} + rx_{(n-r)}\right)$	$\frac{1.1}{n}$	$\frac{0.12}{n}$	$\frac{4.36}{n}$

D'autres estimateurs comme le mode, la moyenne pondérée, la moyenne géométrique n'ont pas été mentionnés. C'est la moyenne arithmétique qui l'emporte dans le cas d'une loi normale, le milieu dans le cas d'une loi uniforme, la médiane ou la moyenne tronquée pour une loi de Cauchy.

17.2 De la dispersion

Estimateur	Définition	variance asymptotique si loi $N(\mu,\sigma)$
Écart-type	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2}$	$\frac{\sigma^2}{2n}$
Écart absolu	$\sqrt{\frac{\pi}{2}} \times \frac{1}{n}\sum_{i=1}^{n}\lvert x_i-\mu\rvert$	$(\pi-2)\frac{\sigma^2}{2n}$
Intervalle semi-interquartile	$0.741(x_{(0.75)} - x_{(0.25)})$	$2.72\frac{\sigma^2}{2n}$

Les termes multiplicatifs dans la définition de ces derniers estimateurs sont introduits afin que leur espérance soit égale à l'écart-type dans le cas gaussien.

17.3 Choix d'estimateur

Il existe donc beaucoup d'estimateurs différents, avec des performances dépendant de la loi en présence. Mais le problème n'est pas seulement la question de l'efficacité, mais aussi de la robustesse, parce que dans la pratique, les lois que l'on rencontre peuvent être contaminées, voire différentes de celles que l'on suppose. Avec de moins bonnes performances, la médiane s'avère par exemple beaucoup plus robuste que la moyenne arithmétique.

---

18 Méthodes d'estimation

Nous avons vu les différentes qualités d'un estimateur, mais, face à un échantillon, le premier problème est déjà d'en trouver un. Les trois principales méthodes sont les suivantes :

18.1 Moments

Si un paramètre θ peut s'exprimer en fonction des k premiers moments $h(\mu_1,\ldots,\mu_k)$,

- calculer les moments empiriques $\widehat{\mu}_j = \frac{1}{n}\sum_{i=1}^{n} x_i^j$
- estimer $\widehat{\theta} = h(\widehat{\mu}_1,\ldots,\widehat{\mu}_k)$
- d'où résolution de k équations à k inconnues.

Cette méthode sous-entend que la loi en présence a des moments qui existent jusqu'à l'ordre k. D'autre part, il est préférable numériquement de travailler avec les moments centrés.

Asymptotiquement, les estimateurs trouvés sont non-biaisés, gaussiens et de variance décroissant en $\frac{1}{n}$, mais ils ne sont pas les plus efficaces. Enfin, compte-tenu de la sensibilité des moments empiriques aux observations extrêmes (et *a fortiori* aux points aberrants), cette méthode est peu robuste.

18.2 Maximum de vraisemblance (ML)

Maximiser $\mathcal{L}(x_1,\ldots x_n;\theta) = f(x_1;\theta)\times\ldots f(x_n;\theta)$ par rapport aux paramètres. On recherche les solutions de

$$\frac{\partial \mathcal{L}}{\partial \theta} = 0 \text{ avec } \frac{\partial^2 \mathcal{L}}{\partial \theta^2} < 0$$

Pour la simplicité de calcul, on utilise en général le log (Néperien) de la vraisemblance.

L'estimateur est asymptotiquement gaussien, non biaisé et MVB car on a

$$\lim_{n\to+\infty} \sqrt{nI_1(\theta)}(\theta_n - \theta) \rightsquigarrow N(0,1)$$

La matrice de variance asymptotique de l'estimateur est l'inverse du Hessien

$$\mathbf{\Sigma}_{\widehat{\mathbf{\Theta}}} = \left(\frac{\partial^2 \log \mathcal{L}}{\partial\theta_i \partial\theta_j}(\theta)\right)^{-1}$$

Quand on parle de propriétés asymptotiques, il ne faut néanmoins pas oublier que bien souvent l'estimateur peut être biaisé pour un petit échantillon.

18.2.1 Application

Quel est le meilleur estimateur de la parallaxe moyenne ϖ d'un amas ? On suppose avoir un échantillon de parallaxes ϖ_{0i} indépendantes, dont les erreurs sont gaussiennes de précision individuelle σ_i, et que l'amas est suffisamment lointain pour que sa profondeur soit négligeable.

La vraisemblance individuelle est donc

$$f(\varpi_{0i} \mid \varpi) = \frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{(\varpi_{0i}-\varpi)^2}{2\sigma_i^2}}$$

et on calcule $\frac{\partial \log \mathcal{L}}{\partial \varpi} = 0$, d'où l'on trouve, en posant $p_i = \frac{1}{\sigma_i^2}$, l'estimateur de ϖ :

$$\widehat{\varpi} = \frac{\sum_{i=1}^n p_i \varpi_{0i}}{\sum_{i=1}^n p_i}$$

C'est donc la moyenne pondérée par l'inverse des variances individuelles. Sa précision se calcule par l'information de Fisher, ou bien par

$$\mathrm{Var}(\widehat{\varpi}) = \frac{\sum_{i=1}^n p_i^2 \mathrm{Var}(\varpi_i)}{(\sum_{i=1}^n p_i)^2}$$

D'où $\sigma_{\widehat{\varpi}} = \frac{1}{\sqrt{\sum_{i=1}^n p_i}}$. Si tous les σ_i sont égaux à σ, on retrouve bien que $\sigma_{\widehat{\varpi}} = \frac{\sigma}{\sqrt{n}}$

18.3 Moindres carrés (LS)

Comme son nom l'indique, il s'agit de minimiser l'écart quadratique entre un modèle et les observations censées le représenter :

$$\min_{\mathbf{\Theta}} (\mathbf{Y} - h(\mathbf{X}\,;\mathbf{\Theta}))^T \mathbf{V}^{-1} (\mathbf{Y} - h(\mathbf{X}\,;\mathbf{\Theta}))$$

où $\mathbf{V}$ est la matrice de variance-covariance des observations Y_i, observations modélisées en fonction de coefficients X_i à l'aide des paramètres Θ_j que l'on cherche à déterminer. Dans le cas particulier où les variances sont toutes égales et les covariances nulles, il s'agit donc de minimiser

$$\sum_{i=1}^n (y_i - f(x_i\,;\Theta))^2$$

Dans le cas gaussien, on retrouve l'estimateur du ML.

Un résultat important concerne le cas linéaire (par rapport à Θ, et non pas par rapport à x !), où le modèle s'écrit

$$\mathbf{Y} = \mathbf{X}\mathbf{\Theta} + \epsilon \tag{2}$$

où ϵ est un vecteur d'espérance nulle et de matrice de variance-covariance $\mathbf{V}$, la solution est

$$\widehat{\mathbf{\Theta}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}$$

et la matrice de variance-covariance de cet estimateur est

$$\mathbf{\Sigma}_{\widehat{\mathbf{\Theta}}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}$$

Dans ce cas, l'estimateur par moindre carré est non-biaisé, et de variance minimum parmi tous les estimateurs linéaires en $\mathbf{Y}$.

18.3.1 Application

Dans le cas d'Hipparcos, les parallaxes des étoiles dans une petite zone du ciel (cas d'un amas), sont corrélées, suite au mode d'observation du satellite. L'estimation de la parallaxe moyenne faite au 18.2.1 était sous l'hypothèse d'indépendance des ϖ_i et ne peut donc convenir, car cet estimateur serait alors sous-optimal, et de variance sous-estimée. Il faut revenir aux observations de base (abscisses a sur des grands cercles), et calculer par LS la parallaxe moyenne. Au premier ordre, on a dans l'équation 2, $\mathbf{Y} = \delta\mathbf{a}$, qui sont les résidus (différence entre les abscisses observées et celles prédites pour chaque étoile i avec des paramètres de référence $(\alpha_{0i}, \delta_{0i}, \varpi_0, \mu_{\alpha_{0*i}}, \mu_{\delta_{0i}})$), les $\mathbf{X}$ sont les dérivées partielles des abscisses

$$(\frac{\partial \mathbf{a}}{\partial \alpha_i}, \frac{\partial \mathbf{a}}{\partial \delta_i}, \frac{\partial \mathbf{a}}{\partial \varpi_i}, \frac{\partial \mathbf{a}}{\partial \mu_{\alpha*i}}, \frac{\partial \mathbf{a}}{\partial \mu_{\delta i}})$$

par rapport aux paramètres astrométriques, les paramètres recherchés $\mathbf{\Theta}$ étant les corrections $(\ldots \delta\alpha_i, \delta\delta_i, \delta\mu_{\alpha*i}, \delta\mu_{\delta i} \ldots)$ à ajouter aux paramètres de référence, ainsi que la parallaxe moyenne ϖ.

---

19 Propagation des erreurs

19.1 Changement de variable

Si l'on connaît la densité de X, celle de $Y = h(X)$ est $f_Y(y) = f_X(x)|\frac{dx}{dy}|$. Pour le montrer, on utilise le fait que $P(Y \leq y) = P(X \leq x)$ si h est croissante, d'où $f_Y(y)dy = f_X(x)dx$

19.2 Quelle est l'erreur sur $Y = h(X)$ sachant celle sur X ?

Si $\mathbf{\Sigma_X} = (\sigma_{ij})$ est la matrice de variance-covariance des $\mathbf{X}$, alors la matrice de variance-covariance des $\mathbf{Y}$ est

$$\mathbf{\Sigma_Y} = \mathbf{J}\mathbf{\Sigma_X}\mathbf{J}^T$$

où $\mathbf{J} = (\frac{\partial h(X_i)}{\partial X_j})$ est le jacobien.

19.3 cas unidimensionnel

si $\mathrm{Var}(X) = \sigma^2$, la précision de $g(X)$ est alors $|g'(\mu)|\sigma$

19.3.1 Application

Soit $V_{\mathrm{T}\delta} = 4.74\frac{\mu_{\delta 0}}{\varpi_0}$ la vitesse tangentielle observée. La précision sur cette vitesse est donc

$$\sigma_{V_{\mathrm{T}\delta}} \approx |V_{\mathrm{T}\delta}| \sqrt{\frac{\sigma^2_{\mu\delta 0}}{\mu^2_{\delta 0}} + \frac{\sigma^2_{\varpi_0}}{\varpi_0^2} - 2\rho_{\mu\varpi}\frac{\sigma_{\mu\delta 0}}{\mu_{\delta 0}}\frac{\sigma_{pi_0}}{\varpi_0}}$$

19.4 précautions :

- ce n'est valable qu'au premier ordre (ex : avec la distance $r = \frac{1}{\varpi}$, alors $\frac{\sigma_r}{r} \approx \frac{\sigma_\varpi}{\varpi}$ est une mauvaise approximation dès que l'erreur relative sur ϖ est plus grande que 20% environ)
- μ étant souvent inconnu, on utilise $g'(x)$, au lieu de $g'(\mu)$, rendant ce terme aléatoire, donc dégradant la précision, et pouvant introduire des biais.

---

20 Estimation d'intervalles

Jusqu'à présent, nous avons vu l'estimation ponctuelle, où l'on essayait de trouver la valeur d'un paramètre. Mais parfois, ce qui est important, ce n'est pas tant la valeur elle-même que l'intervalle dans lequel elle se situe.

20.1 Intervalle de confiance

L'intervalle de confiance $[m_{\inf}, m_{\sup}]$ contient le paramètre recherché μ avec la probabilité γ si

$$P(m_{\inf} \leq \mu \leq m_{\sup}) = \gamma \tag{3}$$

Dans le cas Gaussien bidimensionnel, on parle d'ellipse de confiance.

Il faut bien voir que, dans le cadre fréquentiste, μ est une constante, donc on ne peut pas dire « μ a la probabilité 0.95 de se trouver dans cet intervalle » : ce sont les bornes de l'intervalle qui sont des variables aléatoires. Quant au cas bayésien, il sera abordé dans l'application 20.2.2.

20.2 Types d'intervalle

Il y a bien sûr une infinité de solutions à l'équation 3. Les plus courantes sont les suivantes :

- l'intervalle minimal : tend vers le mode de $f(x)$ quand $\gamma \to 0$
- l'intervalle central symétrique : tend vers la moyenne si $\gamma \to 0$
- l'intervalle bilatéral symétrique : tend vers la médiane quand $\gamma \to 0$

Quand on ne précise pas, c'est de l'intervalle bilatéral symétrique qu'il s'agit, et on prend souvent $\gamma = 0.95$. Dans le cas gaussien, si l'on reproduit plusieurs fois l'expérience, dans 95% des cas on aura $\mid m - \mu \mid \leq 1.96\frac{\sigma}{\sqrt{n}}$; l'intervalle de confiance de la moyenne est

$$\mu \in [m - 1.96\frac{\sigma}{\sqrt{n}}, m + 1.96\frac{\sigma}{\sqrt{n}}]$$

20.2.1 Application

On veut vérifier s'il n'y a pas d'erreur systématique dans les parallaxes d'Hipparcos, en particulier si leur point-zéro global z (décalage systématique) est bien nul (ou en tous cas négligable comparé à l'erreur aléatoire sur les parallaxes dont la dispersion est de l'ordre de 1 mas). On peut considérer que les Nuages de Magellan sont à une telle distance (0.02 mas) que la parallaxe d'Hipparcos devrait être approximativement nulle. On prend donc l'échantillon des 46 étoiles des Nuages de Magellan qui ont été observées, et la parallaxe moyenne calculée avec la moyenne pondérée est -0.1 ± 0.23 mas. L'intervalle de confiance à 95% contenant z est donc $[-0.1-0.02-1.96\times 0.23, -0.1-0.02+1.96\times 0.23] = [-0.57, 0.33]$ mas.

20.2.2 Application

L'étoile HIP 3366 a une parallaxe mesurée $\varpi_0 = -4.09 \pm 1.41$ mas. Dans quel intervalle de confiance à 95% se trouve la vraie parallaxe ϖ ?

En supposant l'erreur gaussienne, on trouve $[-6.85, -1.31]$ mas. C'est plutôt fâcheux, parce que l'on sait que la vraie parallaxe est positive, et que l'intervalle trouvé a une probabilité quasiment nulle (à vrai dire, l'exemple est choisi à dessein pour montrer que dans un échantillon de 118 000 étoiles, on en trouvera probablement quelques-unes dans les queues de distribution, à près de $3\sigma\ldots$).

Dans le cadre bayésien, la situation est différente : ϖ est une v.a. : dans le cas non-informatif, on met le minimum de connaissance que l'on a sur la vraie parallaxe:

$$f(\varpi) \propto \begin{cases} 1 & \text{si } \varpi > 0 \\ 0 & \text{sinon.} \end{cases}$$

On indique généralement le signe $\propto$ de proportionnalité, parce que ce doit être une densité (d'intégrale 1), mais le facteur est peu important parce que ce terme disparaît dans la densité *a posteriori*. Celle-ci vaut

$$f(\varpi \mid \varpi_0) = \begin{cases} \frac{e^{-\frac{(\varpi_0-\varpi)^2}{2\sigma^2}}}{\int_0^{+\infty} e^{-\frac{(\varpi_0-\varpi)^2}{2\sigma^2}} d\varpi} & \text{si } \varpi > 0 \\ 0 & \text{sinon.} \end{cases}$$

L'intervalle de confiance bayésien $[\varpi_-, \varpi_+]$ est tel que

$$\int_{-\infty}^{\varpi_-} f(\varpi \mid \varpi_0) d\varpi = \int_{\varpi_+}^{+\infty} f(\varpi \mid \varpi_0) d\varpi = \frac{1-0.95}{2}$$

ce qui revient à résoudre

$$\int_{-\frac{\varpi_0}{\sigma}}^{\frac{\varpi_- - \varpi_0}{\sigma}} e^{-\frac{t^2}{2}} dt = \int_{\frac{\varpi_+ - \varpi_0}{\sigma}}^{+\infty} e^{-\frac{t^2}{2}} dt = .025 \int_{-\frac{\varpi_0}{\sigma}}^{+\infty} e^{-\frac{t^2}{2}} dt$$

donc l'intervalle bayésien $[0.01, 1.42]$ mas à 95%, qui est quand même plus satisfaisant que l'intervalle trouvé dans le cadre fréquentiste.

Mais c'est évidemment dépendant de ce que l'on a mis comme *a priori* sur $f(\varpi)$, et l'on pourrait par exemple faire valoir que la distribution des parallaxes devrait être choisie croissante (plutôt que constante) à partir de 0 si l'échantillon est limité en magnitude, etc.

---

21 Tests d'hypothèses

21.1 Test

C'est une procédure de décision à partir d'un échantillon, conduisant à choisir entre deux hypothèses, par ex :

$$\left| \begin{array}{ll} H_0 : \theta = 3 & \text{(hypothèse nulle) contre} \\ H_1 : \theta \neq 3 & \text{(hypothèse alternative)} \end{array} \right.$$

21.2 Erreurs de :

- première espèce : rejet de H_0 alors qu'elle est vraie. C'est le seuil α du test. On prend souvent $\alpha = 0.05$.
- seconde espèce : acceptation de H_0 alors qu'elle est fausse (de probabilité β ; $1-\beta$ est alors appelé puissance du test)

21.3 Types de tests

- tests paramétriques :
  on connaît la loi en présence et on teste un ou plusieurs de ses paramètres. Ex : pour une loi normale, on teste si $\sigma = 1$.
- tests non paramétriques :
  on ne fait pas de supposition sur la loi en présence. Ex : tester si deux échantillons sont indépendants.
- tests d'adéquation :
  on teste le type de la loi en présence. Ex : mon échantillon suit-il une loi Gaussienne ?
  On utilise classiquement le test du χ^2, qui regroupe les données en classes, ou, mieux, le test de Kolmogorov, calculant la statistique

$$D_n = \max |S_n(x) - F(x)|$$

21.3.1 Application

Dans l'exemple 20.2.1, le point-zéro global z n'est pas significativement différent de 0. Cela ne veut pas dire qu'il est réellement non-nul, mais que les données ne permettent pas de rejeter l'hypothèse nulle $H_0 : z = 0$. Par contre, si l'hypothèse nulle est $H_0 :\mid z \mid > 1$, on peut la rejeter avec nettement moins de 5% chances de se tromper.

21.3.2 Application

Pour la réduction astrométrique des données d'Hipparcos, des étoiles (probalement des binaires astrométriques à longue période) pouvaient avoir un mouvement non-linéaire et un terme d'accélération devait alors être pris en compte. Le problème était de savoir pour quelles étoiles cette accélération était significativement non nulle.

L'accélération est calculée suivant l'ascension droite ($g_{\alpha *}$) et la déclinaison (g_δ). En l'absence d'accélération réelle (hypothèse nulle $H_0 : \mathbf{G} = 0$), à cause des erreurs de mesure, et des corrélations entre paramètres astrométriques, l'accélération observée $\mathbf{G} = (g_{\alpha *}, g_\delta)$ suit une loi gaussienne bidimensionnelle, de moyenne (0,0) et de matrice de variance-covariance

$$\mathbf{V} = \begin{pmatrix} \sigma^2_{g_{\alpha *}} & \rho\sigma_{g_{\alpha *}}\sigma_{g_\delta} \\ \rho\sigma_{g_{\alpha *}}\sigma_{g_\delta} & \sigma^2_{g_\delta} \end{pmatrix}$$

Donc la statistique $F^2 = \mathbf{G}^T\mathbf{V}^{-1}\mathbf{G}$ suit une loi du χ^2 à 2 degrés de liberté.

Le seuil qui a été choisi pour le test est $\alpha = 0.0027$, qui correspondrait à un test à 3σ pour une gaussienne. Dans une table du $\chi^2(2)$, ceci correspond à la valeur 11.83 ($= 3.44^2$). Pour chaque étoile, on a ainsi calculé l'accélération, puis la statistique F^2, et on a considéré que l'étoile avait une accélération significative quand $F > 3.44$; dans ce cas l'hypothèse alternative était donc adoptée avec moins de 0.27% de risque d'erreur.

---

22 Bibliographie sommaire

22.1 Cours de stat aux DEA d'astronomie

- *Introduction aux statistiques et à la théorie des estimateurs*, D. Pelat, Comptes-Rendus de l'École de Goutelas 1988
- *Remarques sur l'estimation statistique*, F. Mignard, École d'Aussois de Structure Interne 1996

22.2 Aspects numériques

- *Numerical Recipes*, Press *et al.*, ed. Cambridge University Press (en C, ISBN 0-521-35465-X)

22.3 Côtés mathématiques

- *Méthodes statistiques*, Tassi, ed. Economica, ISBN 271-7816232
- *Modern Mathematical Statistics*, Dudewicz & Mishra, ed. Wiley & sons, ISBN 0-471-60716-9
- *L'analyse statistique bayésienne*, C. Robert, ed. Economica, ISBN 2-7178-2199-6 (pour les bayésiens purs et durs)

22.4 Applications en astronomie

- *Errors, Bias and Uncertainties in Astronomy*, Jaschek & Murtagh, Cambridge University Press, ISBN 0-521-39300-0
- *Statistical Challenges in Modern Astronomy*, Feigelson & Babu, ed. Springer Verlag, Vol I : ISBN 0-387-97911-5, Vol II : ISBN 0-387-98203-0
- *On-line statistical software for astronomy & related fields,* `http://www.astro.psu.edu/statcodes/`
- *Statistical Consulting Center for Astronomy,* `http://www.stat.psu.edu/scca/homepage.html`

22.5 Pour les physiciens

- *Statistics in theory and in practice*, Lupton, ed. Princeton University Press,
- *Statistics for physicists*, B.R. Martin, ed. Academic Press, ISBN 0-12-474750-7

22.6 Et pour les autres

... qui veulent des formules rapides :

- *Guide de Statistique appliquée*, Manoukian, ed. Hermann, ISBN 2705660224
