Statistique Bayésienne

Estimation Ponctuelle

Anna Simoni[2]

[2]CREST - Ensae and CNRS

Outline

L'estimateur MAP

- Un estimateur de référence de θ fondé sur $\pi(\theta|x)$ est l'*estimateur du maximum a posteriori* (MAP), défini comme le mode à posteriori :

$$\delta^\pi(x) = \arg\max_\theta \pi(\theta|x).$$

- L'estimateur MAP maximise aussi $\ell(\theta|x)\pi(\theta)$ et, par conséquent, ne requiert pas le calcul de la loi marginale.

- Le MAP peut s'exprimer comme un estimateur du maximum de vraisemblance pénalisée au sens classique.

Précision des estimateurs de Bayes

Puisque la loi à posteriori $\pi(\theta|x)$ est complètement disponible, il est possible d'associer à un estimateur $\delta^\pi(x)$ de $h(\theta)$ une évaluation de la précision de l'estimation via, e.g., l'erreur quadratique a posteriori :

$$\mathbf{E}\left[\left(\delta^\pi(x) - h(\theta)\right)^2 \Big| \, x \right]$$

égale à $Var(h(\theta)|x)$ quand $\delta^\pi(x) = \mathbf{E}^\pi[h(\theta)|x]$.

Dans un cadre multidimensionnel, la matrice de covariance caractérise la performance des estimateurs.

Rappelons que, pour une fonction de coût $L(\theta, \delta)$ et une loi a priori π, la règle de Bayes $\delta^\pi(x)$ est solution de

$$\min_\delta \mathbf{E}^\pi [L(\theta, \delta)|x].$$

Considerons le coût quadratique et l'estimateur moyenne a posteriori :

Lemma

Soit $f(x|\theta) = h(x)e^{\theta x - \psi(\theta)}$, une distribution d'une famille exponentielle. Pour toute loi a priori π, la moyenne a posteriori de θ est donnée par

$$\delta^\pi(x) = \nabla \log m_\pi(x) - \nabla \log h(x),$$

où ∇ est l'opérateur gradient et m_π est la loi marginale associée à π.

Le lemme est satisfait pour tout π. Son intérêt pratique est limité, car le calcul de m_π est généralement assez délicat.

Rappelons que, pour une fonction de coût $L(\theta, \delta)$ et une loi a priori π, la règle de Bayes $\delta^\pi(x)$ est solution de

$$\min_\delta \mathbf{E}^\pi[L(\theta, \delta)|x].$$

Considerons le coût quadratique et l'estimateur moyenne a posteriori :

Lemma

Soit $f(x|\theta) = h(x)e^{\theta x - \psi(\theta)}$, une distribution d'une famille exponentielle. Pour toute loi a priori π, la moyenne a posteriori de θ est donnée par

$$\delta^\pi(x) = \nabla \log m_\pi(x) - \nabla \log h(x),$$

où ∇ est l'opérateur gradient et m_π est la loi marginale associée à π.

Le lemme est satisfait pour tout π. Son intérêt pratique est limité, car le calcul de m_π est généralement assez délicat.

Rappelons que, pour une fonction de coût $L(\theta, \delta)$ et une loi a priori π, la règle de Bayes $\delta^\pi(x)$ est solution de

$$\min_\delta \mathbf{E}^\pi[L(\theta, \delta)|x].$$

Considerons le coût quadratique et l'estimateur moyenne a posteriori :

Lemma

Soit $f(x|\theta) = h(x)e^{\theta x - \psi(\theta)}$, une distribution d'une famille exponentielle. Pour toute loi a priori π, la moyenne a posteriori de θ est donnée par

$$\delta^\pi(x) = \nabla \log m_\pi(x) - \nabla \log h(x),$$

où ∇ est l'opérateur gradient et m_π est la loi marginale associée à π.

Le lemme est satisfait pour tout π. Son intérêt pratique est limité, car le calcul de m_π est généralement assez délicat.

Les lois à priori conjuguées. I

- Dans le cas particulier des lois a priori conjuguées, les espérances a posteriori des paramètres naturels admettent des expressions explicites.

- Quand plusieurs observations de $f(x|\theta)$ sont disponibles, on retrouve les mêmes lois à priori conjuguées et seuls les paramètres dans l'estimateur sont modifiés.

Loi de x	Loi conjuguée	Moyenne à posteriori
Normale $\mathcal{N}(\theta, \sigma^2)$	Normale $\mathcal{N}(\mu, \tau^2)$	$\frac{\mu\sigma^2+\tau^2 x}{\sigma^2+\tau^2}$
Poisson $\mathcal{P}(\theta)$	Gamma $\mathcal{G}(\alpha, \beta)$	$\frac{\alpha+x}{\beta+1}$
Gamma $\mathcal{G}(\nu, \theta)$	Gamma $\mathcal{G}(\alpha, \beta)$	$\frac{\alpha+\nu}{\beta+x}$
Binomiale $\mathcal{B}(n, \theta)$	Beta $\mathcal{B}e(\alpha, \beta)$	$\frac{\alpha+x}{\alpha+\beta+n}$

TABLE – n =nombre d'observations.

Les lois à priori conjuguées. II

Loi de x	Loi conjuguée	Moyenne à posteriori
Binomiale Négative $\mathscr{N}eg(n,\theta)$	Beta $\mathscr{B}e(\alpha,\beta)$	$\frac{\alpha+n}{\alpha+\beta+x+n}$
Multinomiale $\mathscr{M}_k(n;\theta_1,\ldots,\theta_k)$	Dirichlet $\mathscr{D}(\alpha_1,\ldots,\alpha_k)$	$\frac{\alpha_i+x_i}{\sum_j \alpha_j+n}$
Normale $\mathscr{N}(\mu,1/\theta)$	Gamma $\mathscr{G}(\alpha/2,\beta/2)$	$\frac{\alpha+1}{\beta+(\mu-x)^2}$

TABLE – n =nombre d'observations.

Outline

Le modèle normal. I

- Supposons que le modèle soit normale : $x|\theta \sim \mathcal{N}_p(\theta, \Sigma)$ de matrice de covariance Σ connue.

- La loi conjuguée est aussi normale : $\theta \sim \mathcal{N}_p(\mu, A)$ et la loi à posteriori $\pi(\theta|x)$ est
$$\mathcal{N}_p(x - \Sigma(\Sigma + A) - 1(x - \mu), (A^{-1} + \Sigma^{-1})^{-1}).$$

- Sous un coût quadratique, l'estimateur de Bayes est la moyenne à posteriori
$$\delta^\pi(x) = (A^{-1} + \Sigma^{-1})^{-1}(\Sigma^{-1}x + A^{-1}\mu).$$

- Pour des observations répétées du modèle normal, $x_1, \ldots, x_n$, la statistique exhaustive est $\bar{x} \sim \mathcal{N}_p(\theta, \frac{1}{n}\Sigma)$.

Le modèle normal. II

Estimation de la variance.

Dans la plupart des cas, la variance du modèle est partiellement ou totalement inconnue. Prior pour les paramètres (θ, Σ).

- Si la variance est connue à une constante multiplicative près, σ^2, il est possible de revenir à un cadre unidimensionnel (i.e. $x_1, \ldots, x_n$ sont i.i.d. $\mathcal{N}(\theta, \sigma^2)$) pour des raisons d'exhaustivité.

- Soient $\bar{x} := \frac{1}{n} \sum_{i=1}^{n} x_i$ et $s^2 := \sum_{i=1}^{n} (x_i - \bar{x})^2$, alors l'estimateur de Bayes ne depend que de $\bar{x}$ et σ^2 (statistiques exhaustives).

- Vraisemblance :

$$\ell(\theta, \sigma | \bar{x}, s^2) \propto \sigma^{-n} \exp\left[-\frac{1}{2\sigma^2} (s^2 + n(\bar{x} - \theta)^2) \right].$$

1) Prior de Jeffreys : $\pi(\theta, \sigma) = 1/\sigma^2$.

- Posteriori :

$$\begin{aligned} \theta | \sigma, \bar{x}, s^2 &\sim \mathcal{N}(\bar{x}, \sigma^2/n) \\ \sigma^2 | \bar{x}, s^2 &\sim \mathcal{IG}\left(\frac{n-1}{2}, \frac{s^2}{2} \right). \end{aligned}$$

Le modèle normal. III

- La loi à posteriori marginale de σ^2 est du même type que lorsque θ est connu. En revanche, la loi marginale à posteriori de θ diffère :
$$\pi(\theta|\bar{x}, s^2) \propto \{s^2 + n(\bar{x} - \theta)^2\}^{-n/2}$$
i.e. $\theta|\bar{x}, s^2 \sim t_1(n-1, \bar{x}, s^2/n(n-1))$.

2) Prior conjuguée : $\pi(\theta, \sigma^2) = \pi_1(\theta|\sigma^2)\pi_2(\sigma^2)$ où π_1 est une $\mathscr{N}(\mu, \sigma^2/n_0)$ et π_2 est une $\mathscr{IG}(\nu/2, s_0^2/2)$. Remarque : θ et σ^2 ne sont pas indépendants a priori.

- Les lois a posteriori conjuguées ont naturellement la même forme que dans le cas precedent.

- La loi à posteriori satisfait :
$$\pi(\theta, \sigma^2|x) \propto \sigma^{-n-\nu-3} \exp\left\{-\frac{1}{2}[s_1^2 + n_1(\theta - \theta_1)^2/\sigma^2]\right\}$$
où $n_1 = n + n_0$, $\theta_1 = \frac{1}{n_1}(n_0\theta_0 + n\bar{x})$, $s_1^2 = s^2 + s_0^2 + (n_0^{-1} + n^{-1})^{-1}(\theta_0 - \bar{x})^2$.

- Ces lois à posteriori sont conjuguées car : $\theta|\sigma, \bar{x}, s^2 \sim \mathscr{N}(\theta_1, \sigma^2/n)$ et $\sigma^2|\bar{x}, s^2 \sim \mathscr{IG}(\nu_1/2, s_1^2/2)$.

- La loi à posteriori marginale de θ est une loi de Student.

Le modèle normal. III

- La loi à posteriori marginale de σ^2 est du même type que lorsque θ est connu. En revanche, la loi marginale à posteriori de θ diffère :

$$\pi(\theta|\bar{x}, s^2) \propto \{s^2 + n(\bar{x} - \theta)^2\}^{-n/2}$$

  i.e. $\theta|\bar{x}, s^2 \sim t_1(n-1, \bar{x}, s^2/n(n-1))$.

2) Prior conjuguée : $\pi(\theta, \sigma^2) = \pi_1(\theta|\sigma^2)\pi_2(\sigma^2)$ où π_1 est une $\mathcal{N}(\mu, \sigma^2/n_0)$ et π_2 est une $\mathcal{IG}(\nu/2, s_0^2/2)$. Remarque : θ et σ^2 ne sont pas indépendants a priori.

- Les lois a posteriori conjuguées ont naturellement la même forme que dans le cas precedent.

- La loi à posteriori satisfait :

$$\pi(\theta, \sigma^2|x) \propto \sigma^{-n-\nu-3} \exp\left\{-\frac{1}{2}[s_1^2 + n_1(\theta - \theta_1)^2/\sigma^2]\right\}$$

  où $n_1 = n + n_0$, $\theta_1 = \frac{1}{n_1}(n_0\theta_0 + n\bar{x})$, $s_1^2 = s^2 + s_0^2 + (n_0^{-1} + n^{-1})^{-1}(\theta_0 - \bar{x})^2$.

- Ces lois à posteriori sont conjuguées car : $\theta|\sigma, \bar{x}, s^2 \sim \mathcal{N}(\theta_1, \sigma^2/n)$ et $\sigma^2|\bar{x}, s^2 \sim \mathcal{IG}(\nu_1/2, s_1^2/2)$.

- La loi à posteriori marginale de θ est une loi de Student.

Le modèle normal. IV

- n_0/n caractérise la précision de la détermination de la loi à priori, relativement à la précision des observations.

- Si n_0/n tend vers 0, nous obtenons le cas limite $\theta|\bar{x}, \sigma^2 \sim \mathcal{N}(\bar{x}, \sigma^2/n)$, correspondant à la loi a posteriori associée à la loi a priori de Jeffreys.

- Donc, les lois non informatives se présentent souvent comme des limites de lois conjuguées.

- L'inférence statistique fondée sur la loi conjuguée ci-dessus nécessite une détermination précise des hyperparamètres $(\theta_0, s_0^2, n_0, \nu)$.

Estimation de la variance : (θ, Σ) totalement inconnu

- Soit $x_1, \ldots, x_n$ un échantillon aléatoire simple de $\mathcal{N}_p(\theta, \Sigma)$.
- Statistique exhaustive : $\bar{x} := \frac{1}{n}\sum_{i=1}^n x_i$ et $S = \sum_{i=1}(x_i - \bar{x})(x_i - \bar{x})'$.
- Vraisemblance :

$$\ell(\theta, \Sigma|\bar{x}, S) \propto |\Sigma|^{-n/2} \exp\left\{-\frac{1}{2}\left(n(\bar{x}-\theta)'\Sigma^{-1}(\bar{x}-\theta) + tr(\Sigma^{-1}S)\right)\right\}.$$

- Prior conjuguée :

$$\begin{aligned} \theta|\Sigma &\sim \mathcal{N}_p(\mu, \Sigma/n_0) \\ \Sigma^{-1} &\sim \mathscr{W}_p(\alpha, W) \end{aligned}$$

où $\mathscr{W}_p$ denote la loi de Wishart.

Le modèle normal. VI

- Posteriori :

$$
\begin{aligned}
\theta|\Sigma, \bar{x}, S &\sim \mathcal{N}_p\left(\frac{n_0\mu + n\bar{x}}{n_0 + n}, \Sigma/(n_0+n)\right)\\
\Sigma^{-1}|\bar{x}, S &\sim \mathscr{W}_p(\alpha + n, W_1(\bar{x}, S)),
\end{aligned}
$$

avec $W_1(\bar{x}, S)^1 = W^{-1} + S + \frac{nn_0}{n+n_0}(\bar{x}-\mu)(\bar{x}-\mu)'$.

- Loi à priori de Jeffreys : limite de la loi Wishart $\mathscr{W}_p(\alpha, W)$ pour Σ^{-1} lorsque $W^{-1} \to 0$. Elle est donnée par :

$$
\pi^J(\theta, \Sigma^{-1}) = \frac{1}{|\Sigma|^{-(p+1)/2}}.
$$

Let $D = (x_1, \ldots, x_n)$

I. $x_i \sim \mathcal{N}(\mu, \sigma^2)$, $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$

II. $x_i \sim \mathcal{N}(\mu, \sigma^2)$, $\mu \sim \mathcal{N}(\mu_0, \sigma^2/\kappa_0)$. Remarque que κ_0 joue le rôle de n (equivalent sample size).

III. $x_i \sim \mathcal{N}(\mu, \sigma^2)$, $\mu \sim \mathcal{N}(\mu_0, \infty)$ (non-informative).

IV. $\lambda := \sigma^{-2}$ unknown. $x_i \sim \mathcal{N}(\mu, \sigma^2)$, $\mu|\sigma^2 \sim \mathcal{N}(\mu_0, \sigma^2/\kappa_0)$, $\lambda \sim Ga(\alpha_0, \beta_0)$.

V. $\lambda := \sigma^{-2}$ unknown. $x_i \sim \mathcal{N}(\mu, \sigma^2)$, $\mu|\sigma^2 \sim \mathcal{N}(\mu_0, \sigma^2/0)$, $\lambda \sim Ga(\alpha_0, 0)$ (non-informative, i.e. $\mu, \lambda \propto \lambda^{-1}$).

VI. $\lambda := \sigma^{-2}$ unknown, μ known. $x_i \sim \mathcal{N}(\mu, \sigma^2)$, $\lambda \sim Ga(\alpha_0, \beta_0)$.

VII. Different parametrization : σ^2 unknown. $x_i \sim \mathcal{N}(\mu, \sigma^2)$, $\mu|\sigma^2 \sim \mathcal{N}(\mu_0, \sigma^2 v_0)$, $\sigma^{-2} \sim IG(\alpha_0, \beta_0)$.

VIII. Σ^{-1} unknown. $x_i \sim \mathcal{N}_p(\mu, \Sigma)$, $\mu|\Sigma \sim \mathcal{N}_p(\mu_0, \Sigma)$, $\Sigma^{-1} \sim \mathcal{W}_p(\alpha, T)$.

IX. Σ^{-1} unknown. $x_i \sim \mathcal{N}_p(\mu, \Sigma)$, $\mu, \Sigma^{-1} \propto |\Sigma|^{(p+1)/2}$ (non-informative).

X. Different parametrization : Σ unknown. $x_i \sim \mathcal{N}_p(\mu, \Sigma)$, $\mu|\Sigma \sim \mathcal{N}_p(\mu_0, \Sigma/\kappa_0)$, $\Sigma \sim \mathcal{IW}_p(\alpha, \Lambda_0^{-1})$.

Le modèle normal : Normal-Gamma Prior

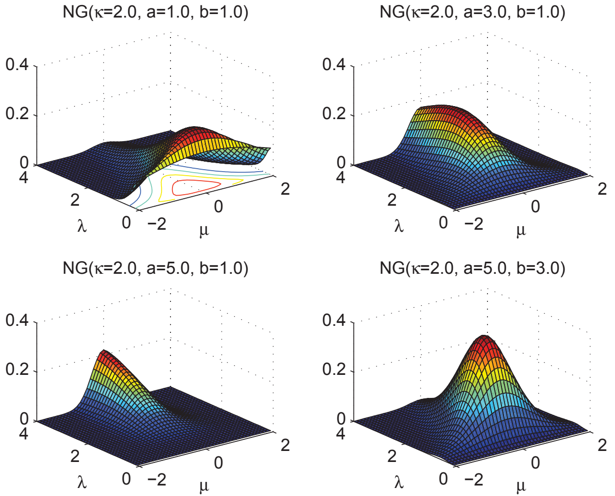

Figure 3: Some Normal-Gamma distributions. Produced by `NGplot2`.

Modèle linéaire et G-priors :

Considerons le modèle de régression linéaire :

$$y = X\beta + \epsilon, \qquad \epsilon \sim \mathcal{N}_n(0, \Sigma), \quad \beta \in \mathbb{R}^p. \tag{1}$$

on travaille conditionnellement à X (X is $n \times p$ a matrix with stochastic entries).

1) Σ connue :

- Statistique exhaustive : $\widehat{\beta} := (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$. $\widehat{\beta}$ est l'estimateur de MV et de MCO. En plus, $\widehat{\beta}|X \sim \mathcal{N}_p(\beta_*, (X'\Sigma^{-1}X)^{-1})$.
- Prior conjuguée : (Lindley & Smith (1972)) $\beta \sim \mathcal{N}_p(A\theta, C)$ où $\theta \in \mathbb{R}^q$, $q \leq p$, et $C \in \mathbb{R}^{p\times p}$. Remarque : A, C, θ peuvent dépendre de X.
- Si la variance Σ est totalement inconnue, il n'est pas possible de construire des lois a priori conjuguées.

2) Σ inconnue :

- Des observations indépendantes sont disponibles.

- Vraisemblance :

$$\ell(\beta, \Sigma|y) \propto |\Sigma|^{-n/2} \exp\left\{-\frac{1}{2} tr\left[\Sigma^{-1}\sum_{i=1}^{n}(y_i - X_i\beta)(y_i - X_i\beta)'\right]\right\}.$$

2.1) **Prior de Jeffrey :** $\pi(\beta, \Sigma) = |\Sigma|^{-(n+1)/2}$

2.2) **Σ est connue à un facteur multiplicatif σ^2 près**. Il est alors possible de réécrire le modèle comme $\epsilon \sim \mathcal{N}_n(0, \sigma^2 I_n)$.

- L'estimateur de MCO satisfait : $\widehat{\beta}|X \sim \mathcal{N}_p(\beta_*, \sigma^2(X'X)^{-1})$.

- Prior conjuguée (G-prior, Zellner 1971, 1986) :

$$\begin{aligned} \beta|\sigma^2 &\sim \mathcal{N}_p\left(\mu, \frac{\sigma^2}{n_0}(X'X)^{-1}\right) \\ \sigma^2 &\sim \mathcal{IG}(\nu/2, s_0^2/2). \end{aligned}$$

Le modèle de régression linéaire. III

- Posteriori :

$$\beta|\widehat{\beta}, s^2, \sigma^2 \quad \sim \quad \mathcal{N}_p\left(\frac{n_0\mu + \widehat{\beta}}{n_0 + 1}, \frac{\sigma^2}{n_0 + 1}(X'X)^{-1}\right)$$

$$\sigma^2|\widehat{\beta}, s^2 \quad \sim \quad \mathcal{IG}\left(\frac{n - p + \nu}{2}, \frac{s^2 + s_0^2 + \frac{n_0}{n_0+1}(\mu - \widehat{\beta})'X'X(\mu - \widehat{\beta})}{2}\right).$$

- Loi a priori alternative : $\beta|\sigma \sim \mathcal{N}_p(\beta_0, \sigma^2 A)$.

- Le modèle de régression est entièrement conditionnel aux variables explicatives X. La G-priori peut se voir comme une loi a posteriori par rapport à X.

- Un G-prior suggère une distribution pour la moyenne de y, $\theta = \mathbf{E}[y|X]$, plutôt que pour β.

- La G-prior est adéquat pour la prise en compte des problèmes de multicolinéarité, car il permet d'assigner une grande variance a priori aux composantes affectées par la multicolinéarité.

- Des points de vue pratique et subjectif, la détermination a priori d'une matrice A plutôt que d'un scalaire n_0 nécessite une plus grande quantité d'information a priori.

D_n =données.
Prior :

$$\beta \sim \pi(\beta)$$

Posterior :

$$\pi(\beta|D_n) \propto \pi(\beta) \prod_{i=1}^{k} \Phi(\beta' x_i)^{y_i} (1 - \Phi(\beta' x_i))^{n_i - y_i}.$$

Difficile à traiter. Albert & Chib (1993) ont proposé une méthode computationnelle basé sur l'algorithme de Gibbs :

Example :

- non-informative (uniform) prior on β.
- the binary response y_i is an indicator of survival ($y_i = 1$ if the person survived the ordeal).
- Suppose that there exists a continuous measurement z_i of health such that if $z_i > 0$, then the person survives ; otherwise the person does not survive.

- Moreover, the health measurement is related to the covariates x_i by the normal regression model

$$z_i = x_{i1}\beta_1 + \ldots + x_{ik}\beta_k + \epsilon_i, \qquad \epsilon_i \sim i.i.d.\mathcal{N}(0,1).$$

- So, this is a missing data problem (with latent variables z_i).

- An automatic Gibbs sampling algorithm is constructed by adding the (unknown) latent data $z = (z_1, \ldots, z_n)$ to β and sampling from the joint posterior distribution of (z, β) :

  - $\beta | z, D_n \sim \mathcal{N}_k\left((X'X)^{-1}X'Z, (X'X)^{-1}\right)$;

  - if we are given a value of β, then $(z_1, \ldots, z_n)$ are independent, with (truncated Normal distributions)

$$\begin{aligned} z_i | \beta, D_n &\sim \mathcal{N}(\beta' x_i, 1) I(z_i > 0), \qquad \text{if } y_i = 1 \\ z_i | \beta, D_n &\sim \mathcal{N}(\beta' x_i, 1) I(z_i < 0), \qquad \text{if } y_i = 0. \end{aligned}$$

- R code : `bayes.probit`

- Variation : informative prior. $\beta \sim \mathcal{N}_k(\beta_0, V_0)$. Now, the β-step of the Gibbs sampler is :

$$\beta | z, D_n \sim \mathcal{N}_k\left((X'X + V_0)^{-1}(X'Z + V_0^{-1}\beta_0), (X'X + V_0)^{-1}\right)$$

Outline

Les modèles dynamiques (ou de séries temporelles) sont des modèle paramétrique où la distribution des variables observées $x_1, \ldots, x_T$ varie dans le temps :

$$f(x_1, \ldots, x_T|\theta) = \prod_{t=1}^{T} f_t(x_t|x_{1:(t-1)}, \theta) \tag{2}$$

où $x_{1:(t-1)} := (x_1, \ldots, x_{t-1})$.

Le modèle AR. I

Le modèle $AR(p)$ exprime la distribution de x_t conditionnellement au passé $x_{1:(t-1)}$ comme une régression linéaire normale sur les p variables les plus récentes :

$$x_t \sim \mathcal{N}\left(\mu - \sum_{i=1}^{p} \rho_i (x_{t-1} - \mu), \sigma^2\right). \tag{3}$$

- Ce modèle est markovien, car la distribution de x_t ne dépend que d'un nombre fixe de valeurs passées, $x_{(t-p):(t-1)}$.
- Vraisemblance conditionnelle :

$$L(\mu, \rho_1, \ldots, \rho_p, \sigma | x_{1:T}, x_{0:(-p+1)}) \\ = \sigma^{-T} \prod_{t=1}^{T} \exp\left\{ -\left(x_t - \mu + \sum_{i=1}^{p} \rho_i (x_{t-i} - \mu) \right)^2 / 2\sigma^2 \right\}.$$

- Paramètres : $\theta = (\mu, \rho_1, \ldots, \rho_p, \sigma^2)$.

- Prior : $(\mu, \rho_1, \ldots, \rho_p) \sim \mathcal{N}_{p+1}$, $\sigma^2 \sim \mathcal{IG}$.
- Imposer la contrainte de *stationnarité* : la récurrence de Durbin-Levinson propose une reparamétrisation des paramètres ρ_i en les autocorrélations partielles ψ_i qui satisfont, sous la contrainte de stationnarité, $\psi_i \in (-1, 1)$, $i = 1, \ldots, p$ et permettent une loi à priori uniforme.

Lemma

Sous la stationnarité du modèle (3)*, les coefficients ρ_i se déduisent des coefficients ψ_i par l'algorithme suivant :*

Récurrence de Durbin-Levinson :

1. *Définir $\varphi^{ii} = \psi_i$ et $\varphi^{ij} = \varphi^{(i-1)j} - \psi_i \varphi^{(i-1)(i-j)}$, pour $i > 1$ et $j = 1, \ldots, i-1$.*
2. *Prendre $\rho_i = \varphi^{pi}$ pour $i = 1, \ldots, p$.*

Le modèle $MA(q)$, (moving average ou moyenne mobile), est un cas spécial de la décomposition de Wold pour un processus stationnaire x_t :

$$x_t = \mu + \epsilon_t - \sum_{j=1}^{q} \vartheta_j \epsilon_{t-j}, \qquad \epsilon_t \sim \mathcal{N}(0, \sigma^2). \tag{4}$$

- En contraste avec le modèle $AR(1)$, où les covariances entre les termes de la série décroissent exponentiellement vers 0 mais sont toujours non nuls, le processus $MA(q)$ est tel que les autocovariances $\gamma_s = cov(x_t, x_{t+s})$ sont égales à 0 pour $|s| > q$.
- Le processus $MA(q)$ est stationnaire, quel que soit le vecteur $(\vartheta_1, \ldots, \vartheta_q)$.
- Cependant, des considérations d'inversibilité et d'identifiabilité impliquents que le polynôme

  $$Q(x) = 1 - \sum_{j=1}^{q} \vartheta_j x^j$$

  doit avoir toutes ses racines en dehors du cercle unité.

- Vraisemblance conditionnelle :

$$L(\mu, \vartheta_1, \ldots, \vartheta_p, \sigma | x_{1:T}, \epsilon_0, \epsilon_{-q+1})$$
$$= \sigma^{-T} \prod_{t=1}^{T} \exp\left\{ - \left(x_t - \mu + \sum_{j=1}^{q} \vartheta_j \widehat{\epsilon}_{t-j} \right)^2 / 2\sigma^2 \right\}$$

où $\widehat{\epsilon}_t = x_t - \mu + \sum_{j=1}^{q} \vartheta_j \widehat{\epsilon}_{t-1}$ et $\widehat{\epsilon}_0 = \epsilon_0, \ldots, \widehat{\epsilon}_{1-q} = \epsilon_{1-q}$.

Le modèle ARMA.

Une extension des modèles précédents est le modèle $ARMA(p,q)$:

$$x_t = \mu - \sum_{j=1}^{p} \rho_j (x_{t-j} - \mu) + \epsilon_t - \sum_{j=1}^{q} \vartheta_j \epsilon_{t-j}, \qquad \epsilon_t \sim \mathcal{N}(0, \sigma^2). \tag{5}$$

- Le but de tels modèles est de permettre une plus forte parcimonie, *i.e.* p et q plus petits, que dans un modèle uniquement AR ou uniquement MA.

Outline

Inférence prédictive.

Définition

La densité prédictive de $Z \sim f_z(z|\theta)$ (ou $Z \sim f_z(z|x,\theta)$) étant donnée $x \sim f_x(x|\theta)$, quand l'a priori pour θ est π est définie par :

$$f_z(z|x) = \int_{\Theta} f_z(z|\theta)\pi(\theta|x)d\theta.$$