Asymptotisch gedrag van de a-posteriori verdeling: Bernstein-von Mises

1 Bernstein-von Mises

Laat $X_1, X_2, \ldots, X_n$ onafhankelijke stochastische variabelen zijn met dichtheid p_θ, voor $\theta \in \mathbb{R}$. We nemen weer de bekende "regulariteitsvoorwaarden" aan, zodat zoals gebruikelijk kunnen definiëren $\ell_\theta(x) = \log p_\theta(x)$, $\dot{\ell}_\theta(x) = \frac{\partial}{\partial\theta}\ell_\theta(x)$ en $\ddot{\ell}_\theta(x) = \frac{\partial}{\partial\theta}\dot{\ell}_\theta(x)$. We brengen in herinnering dat de Fisher informatie in één waarneming gegeven wordt door

$$i_\theta = \mathbb{V}\mathrm{ar}_\theta \dot{\ell}_\theta(X_1).$$

We nemen aan dat deze eindig is. De maximum likelihood schatter voor θ noteren we met $\hat{\theta}_n$.

Voor een gegeven a-priori dichtheid π wordt de a-posteriori dichtheid gegeven door

$$p_{\bar{\Theta}\,|\,X_1,\ldots,X_n}(\theta) = \frac{\prod_{i=1}^n p_\theta(X_i)\pi(\theta)}{\int_{\mathbb{R}} \prod_{i=1}^n p_\theta(X_i)\pi(\theta)\,d\theta}.$$

De volgende stelling beschrijft de asymptotische vorm van de a-posteriori verdeling. Voor grote n is dit bij benadering een normale verdeling met verwachting de maximum likelihood schatter $\hat{\theta}_n$ en optimale variantie $1/(ni_{\theta_0})$.

Stelling 1.1 (Bernstein-von Mises)**.** *Stel dat de a-priori dichtheid π positief en continu is. Onder regulariteitsvoorwaarden geldt onder $\mathbb{P}_{\theta_0}$ dat*

$$\sup_B \Big| \int_B p_{\bar{\Theta}\,|\,X_1,\ldots,X_n}(\theta)\,d\theta - N\Big(\hat{\theta}_n, \frac{1}{ni_{\theta_0}}\Big)(B)\Big| \to 0$$

in kans als $n \to \infty$.

Schets van bewijs. We schetsen het bewijs van convergentie voor een vaste verzameling B. We schrijven eerst

$$\int_{\theta:\sqrt{n}(\theta-\theta_0)\in B} p_{\bar{\Theta}\,|\,X_1,\ldots,X_n}(\theta)\,d\theta = \frac{\int_{\theta:\sqrt{n}(\theta-\theta_0)\in B} \prod_{i=1}^n \frac{p_\theta}{p_{\theta_0}}(X_i)\pi(\theta)\,d\theta}{\int_{\mathbb{R}} \prod_{i=1}^n \frac{p_\theta}{p_{\theta_0}}(X_i)\pi(\theta)\,d\theta}. \tag{1}$$

Een Taylor expansie geeft, voor iedere $x \in \mathbb{R}$,

$$\ell_\theta(x) - \ell_{\theta_0}(x) \approx (\theta - \theta_0)\dot{\ell}_{\theta_0}(x) + \frac{1}{2}(\theta-\theta_0)^2\ddot{\ell}_{\theta_0}(x).$$

Vanwege de wet van grote aantallen hebben we onder $\mathbb{P}_{\theta_0}$ de bijna zekere convergentie

$$-\frac{1}{n}\sum_{i=1}^n \ddot{\ell}_{\theta_0}(X_i) \to -\mathbb{E}_{\theta_0}\ddot{\ell}_{\theta_0}(X_1) = \mathbb{V}\mathrm{ar}_{\theta_0}\dot{\ell}_{\theta_0}(X_1) = i_{\theta_0}.$$

Voor de likelihood hebben we daarom de benadering

$$\prod_{i=1}^n \frac{p_\theta}{p_{\theta_0}}(X_i) \approx e^{-\frac{1}{2}i_{\theta_0}(n(\theta-\theta_0)^2-2\sqrt{n}(\theta-\theta_0)\Delta_n)} = e^{-\frac{1}{2}i_{\theta_0}(\sqrt{n}(\theta-\theta_0)-\Delta_n)^2}e^{\frac{1}{2}i_{\theta_0}\Delta_n^2},$$

met

$$\Delta_n = \frac{1}{i_{\theta_0}\sqrt{n}}\sum_{i=1}^n \dot{\ell}_{\theta_0}(X_i).$$

Als we deze benadering substituëren in (1) en een substitutie $h = \sqrt{n}(\theta-\theta_0)$ in de integralen uitvoeren, vinden we dat

$$\int_{\theta:\sqrt{n}(\theta-\theta_0)\in B} p_{\bar{\Theta}\,|\,X_1,\ldots,X_n}(\theta)\,d\theta \approx \frac{\int_B e^{-\frac{1}{2}i_{\theta_0}(h-\Delta_n)^2}\pi(\theta_0+h/\sqrt{n})\,dh}{\int e^{-\frac{1}{2}i_{\theta_0}(h-\Delta_n)^2}\pi(\theta_0+h/\sqrt{n})\,dh}.$$

Vanwege continuïteit is $\pi(\theta_0+h/\sqrt{n}) \approx \pi(\theta_0)$ voor grote n. We concluderen dat voor iedere verzameling B

$$\int_{\theta:\sqrt{n}(\theta-\theta_0)\in B} p_{\bar{\Theta}\,|\,X_1,\ldots,X_n}(\theta)\,d\theta \approx N(\Delta_n, i_{\theta_0}^{-1})(B)$$

voor grote n, maar dan ook

$$\int_B p_{\bar{\Theta}\,|\,X_1,\ldots,X_n}(\theta)\,d\theta \approx N\Big(\theta_0+\frac{\Delta_n}{\sqrt{n}}, \frac{1}{ni_{\theta_0}}\Big)(B)$$

voor alle B (ga na!). De benadering van de likelihood impliceert ook dat de maximum likelihood schatter $\hat{\theta}_n$ bij benadering de kwadratische functie $\theta \mapsto (\sqrt{n}(\theta-\theta_0)-\Delta_n)^2$ minimaliseert, oftwel

$$\hat{\theta}_n \approx \theta_0 + \frac{\Delta_n}{\sqrt{n}}.$$

We concluderen dat de a-postiori verdeling bij benadering gelijk is aan de normale verdeling met verwachting $\hat{\theta}_n$ en variantie $(ni_{\theta_0})^{-1}$. □

Een precies bewijs van de Bernstein-von Mises stelling kan gevonden worden in Van der Vaart (1998).

2 Gevolgen voor credible sets

Het (bewijs van de) BvM stelling impliceert dat onder $\mathbb{P}_{\theta_0}$ de a-posteriori verdeling en de $N(\theta_0+\Delta_n/\sqrt{n}, (ni_{\theta_0})^{-1})$-verdeling asymptotisch dezelfde kwantielen hebben. Voor $\alpha \in (0,1)$ is het α-kwantiel van deze normale verdeling gelijk aan

$$\theta_0 + \frac{\Delta_n}{\sqrt{n}} + \frac{1}{\sqrt{ni_{\theta_0}}}\xi_\alpha,$$

met ξ_α het α-kwantiel van de standaard normale verdeling. Een $1-\alpha$ credible interval is dus asymptotisch van de vorm

$$I = \Big[\theta_0 + \frac{\Delta_n}{\sqrt{n}} + \frac{1}{\sqrt{ni_{\theta_0}}}\xi_a, \theta_0 + \frac{\Delta_n}{\sqrt{n}} + \frac{1}{\sqrt{ni_{\theta_0}}}\xi_b\Big],$$

met $b - a = 1 - \alpha$ (meestal $a = \alpha/2$ en $b = 1 - \alpha/2$). We hebben

$$\mathbb{P}_{\theta_0}(\theta_0 \in I) = \mathbb{P}_{\theta_0}(\xi_a \leq -\sqrt{i_{\theta_0}}\Delta_n \leq \xi_b).$$

Vanwege de Centrale Limiet Stelling geldt dat $-\sqrt{i_{\theta_0}}\Delta_n \xrightarrow{\mathrm{d}} N(0,1)$ onder $\mathbb{P}_{\theta_0}$ (ga na!). Er volgt dat $P_{\theta_0}(\theta_0 \in I) \to 1 - \alpha$. Een Bayesiaans $(1-\alpha)$-credible interval is dus asymptotisch ook een betrouwbaarheidsinterval met betrouwbaarheid $1 - \alpha$.

Bibliografie

Van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge University Press, Cambridge.