ON BERNSTEIN - VON MISES THEOREM AND SURVIVAL ANALYSIS

Jana Timková

Robust 2010, Králiky

31.1. - 5.2.2010

MLE

- $X_i, i = 1, .., n$ be iid $\sim f(x, \theta), \theta \in \Theta \subset \mathbb{R}^p$, Θ open
- let $I(\theta) = -\mathrm{E}\left[\partial^2 \log f(x, \theta)/\partial\theta^2\right]$ be continuous, with $0 \leq I(\theta) \leq \infty$.
- let θ_0 be the true value of θ and $\hat{\theta}_n$ be **a maximum-likelihood estimator of** θ based on $X_1, .., X_n$

MLE asymptotics: Under certain regularity conditions

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathscr{D}} \mathscr{N}(0, I(\theta_0)^{-1}).$$

- let $\pi(\theta)$ be the prior and $\pi_n^\star(t|x_1, .., x_n)$ be **the a posteriori density of the rescaled parameter** $t = \sqrt{n}(\theta - \hat{\theta}_n)$ and $\Pi^\star(dt|x_1, .., x_n)$ be the probability measure with density $\pi^\star(t|x_1, .., x_n)$
- P_θ^n be the joint distribution of $X_1, .., X_n$.

Bayesian asymptotics

Parametric Bernstein - von Mises: Let $\{P_\theta, \theta \in \Theta\}$ be differentiable in quadratic mean at θ_0 with nonsingular Fisher information I_{θ_0}, and suppose that for every sequence of balls $(K_n)_{n\geq 1} \subset \mathbb{R}^p$ with radii $M_n \to \infty$, we have

$$\Pi^\star(K_n|X_1, .., X_n) \xrightarrow{P^n_{\theta_0}} 1.$$

Then the posterior distribution of the scaled parameter $t = \sqrt{n}(\theta - \hat{\theta})$, given $X_1, .., X_n$, converges in total variation to the normal distribution with mean zero and variance $I(\theta_0)^{-1}$, in probability as $n \to \infty$.

$$\sup_{B \in \mathscr{B}^p} |\Pi^\star(B|X_1, .., X_n) - \Phi(B)| \xrightarrow{P^n_{\theta_0}} 0,$$

Example

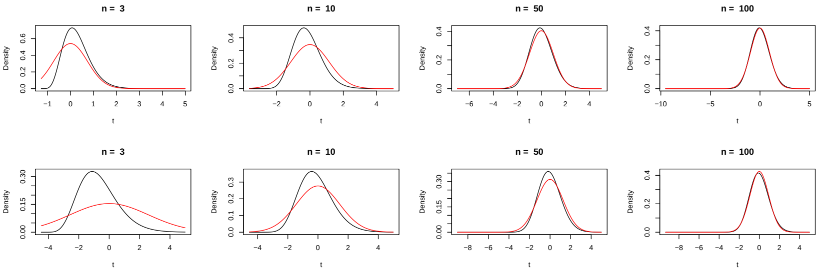


Figure: Generate $x_1, .., x_n$, an n-sample from exponential distribution with the true parameter $\lambda_0 = 1$ (first row) and $\lambda_0 = 3$ (second row). ML estimate of λ is the mean $\overline{x} = \sum_i x_i / n$. Take a family of cunjugate priors $Gamma(a, b)$ for λ. The posterior density of $\sqrt{n}(\lambda - \overline{x})$ and corresponding normal density are in black and red, respectively. The size of sample is $n = 3, 10, 50$ and 100 from left to right.

In semiparametric setting: Cox model

- observed dataset is a set of triplets $(T_i,\ \delta_i,\ \mathbf{Z}_i,\ i = 1, .., n)$
- particular form of the hazard rate which is assumed to satisfy

$$\Lambda_i(t) = \Lambda(t, \mathbf{Z}_i) = \int_0^t \exp\{\boldsymbol{\beta}^\top \mathbf{Z}_i\} d\Lambda(s), \quad i = 1, .., n,$$

- two **unknown parameters**: $\boldsymbol{\beta}$ finite-dimensional regression parameter and $\Lambda(\cdot)$ functional parameter
- traditional approach: **partial likelihood estimators** $\hat{\boldsymbol{\beta}}$, $\hat{\Lambda}(\cdot)$

Frequentist asymptotics for Cox model:
Let some regularity conditions be fulfilled. Then the following is true:

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{\mathscr{D}} \mathscr{N}(0, \Sigma(\boldsymbol{\beta}_0, \tau)^{-1}) \quad \text{and}$$

$$\mathscr{L}(\sqrt{n}(\hat{\Lambda}(\cdot) - \Lambda_0(\cdot)) | \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = x) \xrightarrow{\mathscr{D}} W(V_0(\cdot) - xE_0(\cdot))$$

on the space of cadlag functions. W denotes standard Brownian motion.

Bayesian solution

- $\boldsymbol{\beta}_0$ and $\Lambda_0(\cdot)$ the true values
- A prior process on $\Lambda(\cdot)$: a positive nondecreasing independent increment process with the **Lévy measure** equal to

$$\nu(dt, dx) = \frac{1}{x} g_t(x)\phi(t)\ dx\ dt, \quad t \geq 0, x \in [0,1],$$

  where $\int_0^1 g_t(x)dx = 1,\ \forall t$, and ϕ is bounded and positive on $[0, \tau]$.
- **let** $\pi(\boldsymbol{\beta})$ **be prior distribution for** $\boldsymbol{\beta}$ continuous at $\boldsymbol{\beta}_0$ with $\pi(\boldsymbol{\beta}_0) > 0$

Bernstein - von Mises for Cox model:
Under some conditions the following is true

$$\lim_{n\to\infty} \int_{\mathbb{R}^p} |f_n(x) - \phi(x)| dx = 0$$

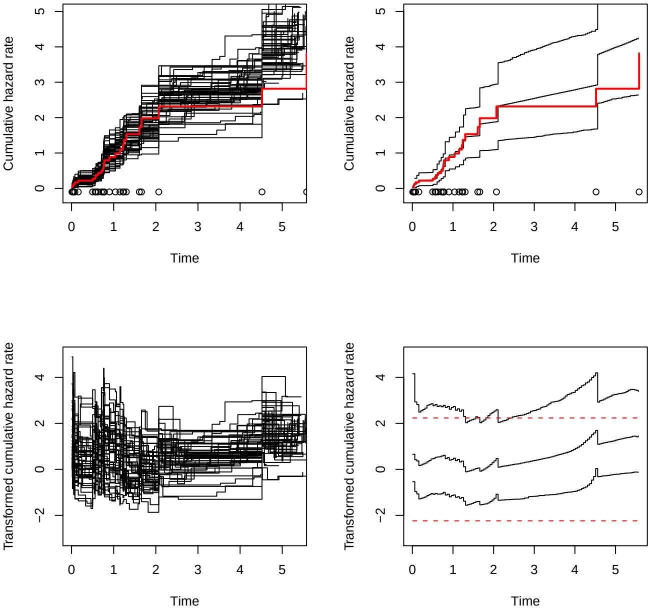
with probability 1, where f_n is the marginal posterior density of $x = \sqrt{n}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$ and ϕ is the normal density with mean 0 and variance $\Sigma(\boldsymbol{\beta}_0, \tau)^{-1}$. Further

$$\begin{aligned} \mathscr{L}(\sqrt{n}(\Lambda(\cdot) \quad & - \quad \hat{\Lambda}(\cdot)|\sqrt{n}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) = x, (T_i, \mathbf{Z}_i, \delta_i)_{i=1}^n) \\ & \xrightarrow{\mathscr{D}} W(V_0(\cdot) - xE_0(\cdot)) \end{aligned}$$

on the space of cadlag functions, with probability 1, as $n \to \infty$.

Illustration

- noncensored simulated data without covariates, $n = 25$
- the only unknown parameter $\Lambda(\cdot) \Longrightarrow$ the asymptotic distribution simplifies into $W(U_0(\cdot))$ where $U_0(t) = \int_0^t d\Lambda_0 / Pr(T \geq t)$.
- Prior: **compound Poisson process** with Lévy measure $\nu(dt, dx) = c\sigma(x)\ dx\ dt$ and Beta distribution with parameters $a = 0.1$ and $b = 0.2$ for $\sigma(\cdot)$.
- Markov Chain Monte Carlo methods


Cumulative hazard rate
Time
Cumulative hazard rate
Time
Transformed cumulative hazard rate
Time
Transformed cumulative hazard rate
Time