Solution sketch

TMA4300 exam

June 2016

Problem 1 (see Givens Hoeting Section 9.2.3)

Ordinary multiple regression model

$$Y_i = \vec{x}_i^T \vec{\beta} + \varepsilon_i \qquad \text{for } i=1,\dots,n$$

iid mean zero RV.

Bootstrap the residuals:

1.) Fit the regression model to the observed data and obtain the fitted responses.

$$\hat{Y}_i = \vec{x}_i^T \vec{\beta}$$

and the residuals $\hat{\varepsilon}_i = Y_i - \hat{Y}_i$

2.) Sample a bootstrap set of length n with replacement i.e. $\hat{\varepsilon}_1^*, \dots, \hat{\varepsilon}_n^*$, from the set of fitted residuals completely at random.

3.) Generate a bootstrap set of pseudo observations

$$Y_i^* = \hat{Y}_i + \hat{\varepsilon}_i^* \qquad i=1,\dots,n$$

fitted NOT original responses!

4.) Regress Y^* on x to obtain a bootstrap estimate $\hat{\beta}^*$.

Repeat this process to get an empirical distribution for $\hat{\beta}^*$

Problem 1 (cont.)

Contrast to paired bootstrap:

Bootstrapping residuals relies on the chosen model providing an appropriate fit to the data, and on the assumption that the residuals have constant variance. If you have doubts about this paired bootstrap will be less sensitive to violations of these assumptions.

In paired bootstrap we consider the data pairs

$z_i := (\vec{x}_i, y_i)$ as values observed for i.i.d RV

$Z_i := (\vec{X}_i, Y_i)$ drawn from a joint response-predictor distribution.

To bootstrap sample $z^*_1, \dots, z^*_n$ completely at random and with replacement from $\{z_1, \dots z_n\}$

Apply the regression model to the generated pseudo-data set to obtain $\hat{\beta}^*$ and repeat this many times.

Problem 2

a) Given: Density $f(x)$ with CDF $F(x)$

Inverse transform technique:

- Draw $u \sim \text{Unif}(0,1)$
- Compute the inverse CDF $F^{-1}(y) = x$
- Return $F^{-1}(u)$ as a random draw from $f(x)$

Why does it work?

Proof: Use the change of variables rule.

We know $u \sim \text{Unif}(0,1)$ and are interested in the distribution of $F^{-1}(u) = g(u) = X$

$$\Rightarrow \quad f_X(x) = \underbrace{f_u((\underbrace{F^{-1}(u))^{-1}}_{F(x)})}_{1} \cdot \underbrace{\left| \frac{d}{dx} F(x) \right|}_{f_X(x)}$$

$\square$

b)

$$f(x) = \frac{1}{2} \exp(-|x|), \quad x \in \mathbb{R}$$

$$\Rightarrow \quad f(x) = \frac{1}{2} \exp(-x) \quad x > 0$$

$$f(x) = \frac{1}{2} \exp(x) \quad x \leq 0$$

$$\Rightarrow \quad F(x) = \int_{-\infty}^{x} \frac{1}{2} \exp(z) dz = \frac{1}{2} \exp(z) \Big|_{-\infty}^{x} = \frac{1}{2} \exp(x)$$

for $x \leq 0$

$$F(x) = \int_{-\infty}^{0} \frac{1}{2} \exp(z) dz + \int_{0}^{x} \frac{1}{2} \exp(-z) dz =$$

$$\frac{1}{2} \exp(z) \Big|_{-\infty}^{0} + \left(-\frac{1}{2} \exp(-z)\right) \Big|_{0}^{x} =$$

3

$$= \frac{1}{2} - \frac{1}{2}\exp(-x) + \frac{1}{2} = 1 - \frac{1}{2}\exp(-x) \quad , \quad x > 0$$

$$\Rightarrow \quad F^{-1}(u) = \log(2u) \qquad 0 < u < \frac{1}{2}$$

$$F^{-1}(u) = -\log(2(1-u)) \qquad \frac{1}{2} < u < 1$$

$\Rightarrow$ sample $u \sim \mathrm{Unif}(0,1)$

if $u < \frac{1}{2}$

return $\log(2u)$

else

return $-\log(2(1-u))$

Problem 3

$$y_i \mid \theta_i \sim \text{Poisson}(\theta_i)$$

$$\theta_i \mid \alpha, \beta \sim \text{Gamma}(\alpha, \beta)$$

$$\alpha \sim \text{Exp}(a) \qquad \beta \sim \text{Gamma}(b, c)$$

a) Derive the full conditional distributions

- $$f(\theta_i \mid \cdot) = \frac{(\theta_i)^{y_i}}{y_i!} \exp(-\theta_i) \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \theta_i^{\alpha-1} \exp(-\beta\theta_i)$$

$$\propto \theta_i^{y_i+\alpha-1} \exp(-\theta_i(\beta+1))$$

$$\Rightarrow \theta_i \mid \cdot \sim \text{Gamma}(y_i+\alpha, \beta+1) \quad \text{for } i=1,\dots,n$$

- $$f(\beta \mid \cdot) = \prod_{i=1}^{n} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta_i^{\alpha-1} \exp(-\beta\theta_i) \cdot \frac{c^b}{\Gamma(b)} \beta^{b-1} \exp(-c\beta)$$

$$\propto \beta^{n\alpha+b-1} \exp\left(-\beta\left(\sum_{i=1}^{n}\theta_i + c\right)\right)$$

$$\Rightarrow \beta \mid \cdot \sim \text{Gamma}\left(n\alpha+b, \sum_{i=1}^{n}\theta_i + c\right)$$

- $$f(\alpha \mid \cdot) = \prod_{i=1}^{n} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta_i^{\alpha-1} \exp(-\beta\theta_i) \cdot a\exp(-a\alpha)$$

$$= \frac{\beta^{n\alpha}}{(\Gamma(\alpha))^n} \prod_{i=1}^{n} \theta_i^{\alpha-1} \cdot \exp(-a\alpha)$$

This is no standard density.

3b) - Set inital values for all $\theta_i^{(0)}$ $i=1,\dots,n$, $\alpha^{(0)}$ and $\beta^{(0)}$

- for (j in 1:numIterations) {

  for (i in 1:n) {

  draw $\theta_i^{(j)} \sim Ga(y_i + \alpha^{(j-1)}, \beta^{(j-1)} + 1)$

  }

  draw $\beta^{(j)} \sim Ga(n\alpha^{(j-1)} + b, \sum_{i=1}^{n} \theta_i^{(j)} + c)$

  # the full conditional for α is non-standard!

  # Here, we need a Metropolis-Hastings step.

  propose a value α^* from a proposal distribution Q, e.g. using a random-walk proposal

  Compute the acceptance ratio p [more details needed here!].

  Sample $u \sim Unif(0,1)$

  if ($u < p$)

  $\alpha^j = \alpha^*$

  else $\alpha^j = \alpha^{j-1}$

  # Important $\alpha^j > 0$ must be guaranteed!

  # i.e. either reject negative proposals or

  # update α on log-scale

}

numIterations is chosen to reach convergence

3b) It might be further necessary to remove a burn-in period.

Important: In the algorithm it should be clear that for all updates always the most recent parameter values are used.

Problem 4

a) One of the crucial properties to apply INLA is that the second stage of the Bayesian hierarchical model is Gaussian

Thus, let us check this:

- 1. stage $\quad y_i \mid p_i \sim \text{Binomial}(n_i, p_i) \qquad i = 1, \ldots, I$

2. stage $\quad \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_i + u_i$

$\Rightarrow$ latent vector $x = (\beta_0, \beta_1, u_1, \ldots, u_I)$

with $\beta_0 \sim N(0, 0.001^{-1})$

$\beta_1 \sim N(0, 0.001^{-1})$

$u_i \mid \tau \sim N(0, \tau^{-1})$

3. stage $\quad \tau \sim \text{Gamma}(1, 0.05)$

Question is $x \mid \tau \sim$ Gaussian?

$$(\beta_0, \beta_1, u_1, \ldots, u_n) \mid \tau \sim N\left(\vec{0}, \begin{pmatrix} 0.001^{-1} & & 0 \\ & 0.001^{-1} & \\ 0 & & \tau^{-1} I \end{pmatrix}\right)$$

$\Rightarrow$ We only have one hyper parameter and the latent field is Gaussian

$\Rightarrow$ we can use INLA

- We proceed similar to the first case

However, here we see that one of the latent variables namely v_{ij}, $i=1,\dots,n$, $j=1\dots,m$ is not Gaussian but Bernoulli distributed. The consequence is that the latent field

$$x = (\beta_0, u_1 \dots, u_n, v_{11} \dots\dots, v_{nm})$$

will not be Gaussian $\Rightarrow$ We cannot use INLA.

Problem 5

a) Using Bayes rule we derive the posterior distribution as

$$f(\theta|y) = \frac{f(y|\theta) \cdot f(\theta)}{f(y)}$$

where $f(y) = \int f(y|\theta) f(\theta) \, d\theta$ is the marginal likelihood.

The posterior density is the target density, so $h(\theta) = f(\theta|y)$ and the prior density is the proposal, so $g(\theta) = f(\theta)$

b) We have

$$p = \frac{h(\theta^*)}{a \cdot g(\theta^*)} = \frac{\frac{f(y|\theta^*) \cdot f(\theta^*)}{f(y)}}{a \cdot f(\theta^*)} = \underline{\underline{\frac{f(y|\theta^*)}{a \cdot f(y)}}}$$

$$\Rightarrow \quad \text{for } a = \frac{f(y|\hat{\theta}_{ML})}{f(y)} \quad \text{we get}$$

$$p = \frac{f(y|\theta^*)}{f(y|\hat{\theta}_{ML})}.$$

To determine the best a we consider

$$a \geq \frac{h(\theta)}{g(\theta)} = \frac{f(y|\theta)}{f(y)}$$

The term $f(y)$ is the same for all θ, thus we have to consider the term $f(y|\theta)$

$$\boxed{h(\theta) \leq a \cdot g(\theta) \text{ with } a = \frac{f(y \mid \hat{\theta}_{ML})}{f(y)}}$$

5b) Since $f(y \mid \theta) \leq f(y \mid \hat{\theta}_{ML})$ for all θ the inequality follows directly. Further the expression for $a = \dfrac{f(y \mid \hat{\theta}_{ML})}{f(\theta)}$ is the smallest constant so that $h(\theta) \leq a \cdot g(\theta)$. We cannot further improve the choice of a.

Problem 6

a) $$f_Y(y) = f\Big(y \mid \underbrace{U \le \frac{h(Z)}{a \cdot g(Z)}}_{\text{acceptance}}\Big) =$$

$$\frac{P\left(U \le \frac{h(Z)}{a g(Z)} \mid Z = y\right) g(y)}{\underbrace{P\left(U \le \frac{h(Z)}{a \cdot g(Z)}\right)}_{d}}$$

$$= \frac{P\left(U \le \frac{h(y)}{a \cdot g(y)} \mid Z = y\right) g(y)}{d}$$

$$\overset{Z \,\&\, U \text{ indep.}}{=} \frac{P\left(U \le \frac{h(y)}{a \cdot g(y)}\right) \cdot g(y)}{d}$$

$$= \begin{cases} \dfrac{g(y)}{d} & \text{if } h(y) \ge a \cdot g(y), \text{ i.e } y \notin C \\[2ex] \dfrac{h(y)}{a \cdot d} & \text{if } h(y) \le a \cdot g(y), \text{ i.e } y \in C \end{cases}$$

6b)

$$d = P\left(U \leq \frac{h(Z_1)}{a \cdot g(Z_1)}\right) =$$

$$= \int_{-\infty}^{\infty} P\left(U \leq \frac{1}{a} \frac{h(Z_1)}{g(Z_1)} \,\Big|\, Z_1 = x\right) \cdot g(x)\, dx$$

$$= \int_{-\infty}^{\infty} P\left(U \leq \frac{1}{a} \frac{h(x)}{g(x)} \,\Big|\, Z_1 = x\right) g(x)\, dx$$

$$= \int_{-\infty}^{\infty} P\left(U \leq \frac{1}{a} \frac{h(x)}{g(x)}\right) g(x)\, dx$$

$$= \int_{x \in C_1} \frac{1}{a} \frac{h(x)}{\cancel{g(x)}} \cdot \cancel{g(x)}\, dx + \int_{x \notin C_1} 1 \cdot g(x)\, dx$$

$$= \frac{1}{a} \int_{x \in C_1} h(x)\, dx + \int_{x \notin C} g(x)\, dx$$

If $a \cdot g(x)$ would be dominating over the whole range than $x \in C_1$ for all $x \Rightarrow$ we loose the second term and the first integral is equal to $1 \Rightarrow d = \frac{1}{a}$.