Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Løsningsforslag - Eksamen mai 2008

Problem 1

a) The cumulative distribution function becomes

$$F(x) = \int_0^x f(u)\mathrm{d}u = \int_0^x 2\alpha u e^{-\alpha u^2}\mathrm{d}u = \left[-e^{-\alpha u^2}\right]_0^x = 1 - e^{-\alpha x^2}$$

Finds the corresponding inverse function by solving $u = F(x)$ with respect to x.

$$1 - e^{-\alpha x^2} = u$$
$$-\alpha x^2 = \ln(1-u)$$
$$x = \sqrt{-\frac{1}{\alpha}\ln(1-u)}$$

Thus, we can generate a sample from $f(x)$ by

1. Generate $u \sim \mathrm{Unif}[0,1]$.
2. Compute
$$x = F^{-1}(u) = \sqrt{-\frac{1}{\alpha}\ln(1-u)}.$$

b) The proposal distribution is given to be

$$g(x) = \lambda e^{-\lambda x}, x > 0.$$

The acceptance probability must then be

$$\gamma(x) = c\frac{f(x)}{g(x)} = \frac{2c\alpha}{\lambda} x e^{\lambda x - \alpha x^2},$$

where c is a constant that must be decided. Differentiate with respect to x to find the maximal value of $\gamma(x)$.

$$\gamma'(x) = \frac{2c\alpha}{\lambda}\left[e^{\lambda x - \alpha x^2} + x e^{\lambda x - \alpha x^2}(\lambda - 2\alpha x))\right] = \frac{2c\alpha}{\lambda} e^{\lambda x - \alpha x^2}(1 + \lambda x - 2\alpha x^2).$$

Equating this to zero we get

$$2\alpha x^2 - \lambda x - 1 = 0$$
$$x = \frac{\lambda \pm \sqrt{\lambda^2 + 8\alpha}}{4\alpha}.$$

Only positive values of x is of interest for us so we get

$$x = \frac{\lambda + \sqrt{\lambda^2 + 8\alpha}}{4\alpha}.$$

This must give a maximum for $\gamma(x)$ as $\gamma(x) > 0$ for all $x > 0$ and $\gamma(0) = 0 = \lim_{x\to\infty} \alpha(x)$. Thus, defining

$$x_0 = \frac{\lambda + \sqrt{\lambda^2 + 8\alpha}}{4\alpha},$$

the maximal value for $\gamma(x)$ becomes

$$\gamma(x_0) = \frac{2c\alpha}{\lambda} x_0 e^{\lambda x_0 - \alpha x_0^2}.$$

Equating this maximal value to 1 and solving for c we get

$$c = \frac{\lambda}{2\alpha} e^{-\lambda x_0 + \alpha x_0^2},$$

which gives acceptance probability

$$\gamma(x) = x \exp\left\{\lambda(x - x_0) - \alpha(x^2 - x_0^2)\right\}.$$

Thus, we can generate a sample from $f(x)$ by

1. Generate $x \sim \text{Exponential}(\lambda)$.
2. Compute
$$x_0 = \frac{\lambda + \sqrt{\lambda^2 + 8\alpha}}{4\alpha}$$
   and
$$\gamma = x \exp\left\{\lambda(x - x_0) - \alpha(x^2 - x_0^2)\right\}.$$
3. Generate $u \sim \text{Unif}[0, 1]$.
4. If $u \leq \gamma$ return x as the realisation, otherwise goto 1.

The expected number of tries per acceptance is equal to c, so to find the optimal value for λ we need to maximise c with respect to λ. Thus, we need to solve the equation

$$\frac{\partial c}{\partial \lambda} = 0$$

with respect to λ.

c) One possibility is to adopt the random number generator in item **a**) and obtain antithetic variates by defining

$$X_i = \sqrt{-\frac{1}{\alpha}\ln(1-U_i)} \quad \text{and} \quad Y_i = \sqrt{-\frac{1}{\alpha}\ln(U_i)}$$

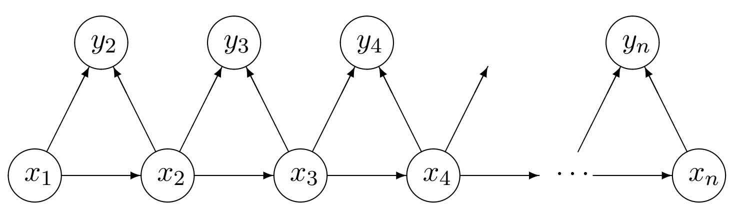
for $i = 1, \ldots, n$. When U_i is uniformly distributed on $[0,1]$, $1-U_i$ will also be uniformly distributed on $[0,1]$, so X_i and Y_i are both distributed according to $f(x)$. Clearly X_i and Y_i are negatively correlated. Thus, we can define

$$\widetilde{\theta} = \frac{1}{2n}\sum_{i=1}^{n}(X_i + Y_i).$$

[An alternative solution to the problem is to define a control variate. Again adopting the random number generator in **a**) one can use either $u-1$ or $-\ln(1-u)-1$ as a control variate. Both of these control variates has mean zero.]

Problem 2

a) The graphical model becomes



and the posterior distribution becomes

$$\pi(x|y) = \frac{\pi(x)\pi(y|x)}{\pi(y)} \propto \pi(x)\pi(y|x)$$

$$\propto p(x_1)\prod_{i=2}^{n} p(x_i|x_{i-1}) \prod_{i=2}^{n}\left[\frac{1}{\sqrt{\sigma^2(x_i)}}\exp\left\{-\frac{1}{2}\frac{(y_i - (v(x_i) - v(x_{i-1})))^2}{\sigma^2(x_i)}\right\}\right]$$

Use the Gibbs sampler algorithm, i.e. use as proposal distribution the full conditional for a randomly chosen node, x_i. For $i > 1$, the full conditional distribution for x_i becomes

$$\pi(x_i|y, x_{-i}) \propto \pi(x|y) \propto p(x_i|x_{i-1})p(x_{i+1}|x_i)\frac{1}{\sqrt{\sigma^2(x_i)}}$$
$$\cdot \exp\left\{-\frac{1}{2}\left[\frac{(y_i - (v(x_i) - v(x_{i-1})))^2}{\sigma^2(x_i)} + \frac{(y_{i+1} - (v(x_{i+1}) - v(x_i)))^2}{\sigma^2(x_{i+1})}\right]\right\}.$$

For $i = 1$ the expression becomes

$$\pi(x_1|x_{-1}, y) \propto \pi(x|y) \propto p(x_1)p(x_2|x_1)\exp\left\{\frac{1}{2}\frac{(y_2 - (v(x_2) - v(x_1)))^2}{\sigma^2(x_2)}\right\}.$$

As we use a Gibbs sampler the acceptance probability becomes equal to one. Pseudo code for the algorithm is as follows.

1. Draw initial values for $x^0 = (x_1^0, \ldots, x_n^0)^T$
2. Iterate for $t = 1, 2, \ldots$
   (a) Draw what node to update, $i \sim \text{Unif}(1, \ldots, n)$.
   (b) Draw new value $x_i^t \sim \pi(x_i^t|x_{-i}^{t-1}, y)$ and set $x_j^t = x_j^{t-1}$ for $j \neq i$.

b) First one needs to find the length of the burn-in phase of the chain. This can be done by output analysis. Assume the Markov chain has (essentially) converged after $T < M$ iterations. The posterior probability for rock type k in node i is then estimated by

$$\widehat{\alpha}_{ik} = \frac{1}{M - T + 1}\sum_{t=T}^{M} I(x_i^t = k),$$

where $I(\cdot)$ is the indicator function. The posterior expected fraction of nodes with rock type k is

$$\beta_k = \text{E}\left[\frac{1}{n}\sum_{i=1}^{n} I(x_i = k)\middle| y\right]$$

and this is estimated by

$$\widehat{\beta}_k = \frac{1}{(M - T + 1)n}\sum_{i=1}^{n}\sum_{t=T}^{M} I(x_i^t = k).$$

Finally, the posterior probability for the fraction of nodes with rock type k is larger than a given threshold r is

$$\gamma_k(r) = \text{P}\left(\frac{1}{n}\sum_{i=1}^{n} I(x_i = k) > r\middle| y\right)$$

and the corresponding estimate becomes

$$\widehat{\gamma}_k(r) = \frac{1}{M-T+1}\sum_{t=T}^{M} I\left(\frac{1}{n}\sum_{i=1}^{n} I(x_i^t = k) > r\right).$$

Problem 3

a) As the data that are used to estimate the prediction error also have been used to estimate the regression model the estimated prediction errors will tend to be too small. Thus, the apparent prediction error is biased and gives a too optimistic estimate for θ.

Cross validation can be used to avoid using the same data to estimate the prediction error as was used to estimate the regression model. Leave-one-out cross validation is described by the following. Let $\widehat{\beta}_{-i}$ denote the estimator for β based on all the data except data i, i.e.

$$\widehat{\beta}_{-i} = \operatorname*{argmin}_b \left\{ \sum_{j\neq i} (y_j - m(b, x_j))^2 \right\}.$$

The leave-one-out cross validation estimator for θ is then

$$\widehat{\theta}_{cv} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - m(\widehat{\beta}_{-i}, x_i)\right)^2.$$

b) Let F denote the (joint) distribution of x and y. The prediction error θ is then a function of F, so we write $\theta = \theta(F)$. To stress the dependence of the apparent error rate on the data we write

$$\widehat{\theta}_a = \widehat{\theta}_a((x_1, y_1), \ldots, (x_n, y_n)).$$

The bias of $\widehat{\theta}_a$ is then defined as

$$\text{bias}_F(\widehat{\theta}_a, \theta) = \text{E}_F(\widehat{\theta}_a((x_1, y_1), \ldots, (x_n, y_n)) - \theta(F).$$

The ideal bootstrap estimator for the bias is obtained by adopting the plug-in principle,

$$\widehat{\text{bias}} = \text{E}_{\widehat{F}}(\widehat{\theta}_a((x_1^\star, y_1^\star), \ldots, (x_n^\star, y_n^\star)) - \theta(\widehat{F}),$$

where $\widehat{F}$ is a discrete distribution giving probability $1/n$ to each of $(x_1, y_1), \ldots, (x_n, y_n)$.

For each of $i = 1, \ldots, n$, the number of possible states for $(x_i^\star, y_i^\star)$ is n. Thereby the number of possible states for $((x_1^\star, y_1^\star), \ldots, (x_n^\star, y_n^\star))$ is n^n and the above expectation is given by a sum of n^n terms. Thus, the is not pratical to evaluate the expectation except when n is small.

Pseudo code for estimating the ideal bootstrap estimator:

1. Generate B sets of bootstrap samples $((x_1^{b\star}, y_1^{b\star}), \dots, (x_n^{b\star}, y_n^{b\star})), b = 1, \dots, B$, where $(x_i^{b\star}, y_i^{b\star})$ for $b = 1, \dots, B$ and $i = 1, \dots, n$ are sampled independently (with replacement) from $(x_1, y_1), \dots, (x_n, y_n)$.

2. Evaluate the bootstrap sample estimates,
$$\widehat{\beta}^{b\star} = \underset{b}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} (y_i^{b\star} - m(b, x_i^{b\star}))^2 \right\}$$
and
$$\theta_a^{b\star} = \frac{1}{n} \sum_{i=1}^{n} \left(y_i^{b\star} - m(\widehat{\beta}^{b\star}, x_i^{b\star}) \right)^2$$
for $b = 1, \dots, B$.

3. Approximate the bootstrap expectation $\mathrm{E}_{\widehat{F}}(\widehat{\theta}_a((x_1^\star, y_1^\star), \dots, (x_n^\star, y_n^\star))$ by
$$\widehat{\theta}_a(\cdot) = \frac{1}{B} = \sum_{b=1}^{B} \widehat{\theta}_a^{b\star}$$
and approximate $\theta(\widehat{F})$ by
$$\widehat{\theta(\widehat{F})} = \frac{1}{B} \sum_{b=1}^{B} \left[\frac{1}{K} \sum_{k=1}^{K} (y_0^{k\star} - m(\widehat{\beta}^{b\star}, x_0^{k\star}))^2 \right],$$
where $(x_0^{k\star}, y_0^{k\star}), k = 1, \dots, K$ are sampled independently (with replacement) from $(x_1, y_1), \dots, (x_n, y_n)$.

4. Approximate the ideal bootstrap estimator with
$$\mathrm{bias}_B = \widehat{\theta}_a(\cdot) - \widehat{\theta(\widehat{F})}.$$

c) The bias corrected estimate for θ is
$$\widehat{\theta}_{bc} = \widehat{\theta}_a - \left[\widehat{\theta}_a(\cdot) - \widehat{\theta(\widehat{F})} \right].$$

As discussed in **a**), we expect $\widehat{\theta}_a$ to have a strong bias. We can expect the bias corrected estimator to be approximately unbiased. However, the uncertainty in the estimated bias gives that $\mathrm{Var}[\widehat{\theta}_{bc}]$ is somewhat larger than $\mathrm{Var}[\widehat{\theta}_a]$. In total the bias corrected estimator is preferable (because the bias of $\widehat{\theta}_a$ is large).