Jo Eidsvik, May 2012

Exam: TMA4300 COMPUTATIONAL STATISTICS

1. a) $U = F(X) = 1 - e^{-\left(\frac{x}{\lambda}\right)^k}$

$-\ln(1-U) = \left(\frac{x}{\lambda}\right)^k$, $X = \lambda \sqrt[1/k]{-\ln(1-U)}$

$k=2$: $Y = \lambda\sqrt{-\ln(1-u)}$ Exponential

$k=1$: $X = -\lambda \ln(1-u)$ Rayleigh

$Y = \sqrt{\lambda X} = \sqrt{\lambda \cdot \lambda \cdot -\ln(1-u)} = \lambda\sqrt{-\ln(1-u)}$

b) $T^b = X_1^b + \dots X_{10}^b$, where $X_j^b \sim \text{Rayleigh}(\lambda)$

$b = 1, \dots, B$.

Sort $T^1, \dots T^B$ from smallest to largest.

$T^{(1)} < T^{(2)} < \dots < T^{(B)}$

$$\hat{\text{median}} = \begin{cases} T^{(B/2)} + T^{(B/2+1)} & B \text{ even} \\ T^{\left(\frac{B+1}{2}\right)} & B \text{ odd} \end{cases}$$

c) $T^b = X_1^b + \dots + X_{10}^b$ like in b).

$$\hat{p} = \frac{\sum_{b=1}^{B} I(T^b > 5)}{B}$$

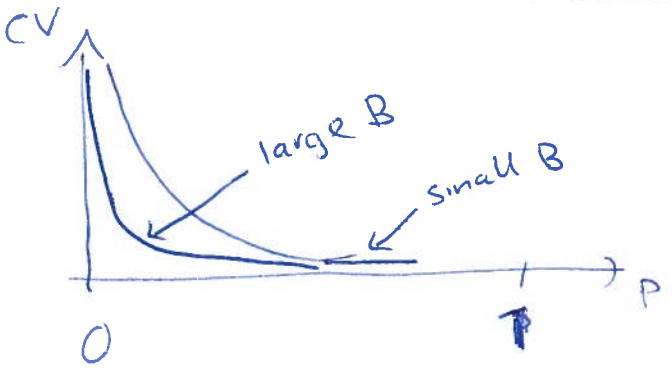
B. $\hat{p}$ is binomial

$\sum_{b=1}^{B} I(T^b > 5) \sim Bin(B, p)$, This means:

$E(\hat{p}) = p$

$Var(\hat{p}) = \frac{p(1-p)}{B}$.

$$CV = \frac{\sqrt{\frac{p(1-p)}{B}}}{p} = \sqrt{\frac{1-p}{pB}}$$



For moderate CV for small p's, the number of samples B needs to be very large.

d) $\pi(x) \sim$ Rayleigh $\pi(x) = \frac{dF}{dx}\Big|_{k=2} = 2 \cdot \frac{x}{\lambda} e^{-\left(\frac{x}{\lambda}\right)^2}$.

$q(x) \sim$ Exponential $q(x) = \frac{dF}{dx}\Big|_{k=1} = \frac{1}{\lambda} e^{-\frac{x}{\lambda}}$

Draw $x_1^b \dots x_{10}^b$ from $q(x)$. $T^b = x_1^b + \dots x_{10}^b$.

Evaluate $w(x_1^b \dots x_{10}^b) = \prod_{i=1}^{10} \frac{\pi(x_i^b)}{q(x_i^b)}$

$$\hat{p}_{IS} = \sum_{b=1}^{B} I\left(T^b > \tau\right) \cdot w\left(x_1^b, \ldots x_{10}^b\right)$$

which is unbiased $E\left(\hat{p}_{IS}\right) = p$, since

$$p = E\left(I\left(T(x_1, \ldots x_{10}) > \tau\right)\right)$$

$$= \int_{x_1} \cdots \int_{x_{10}} I\left(x_1 + \ldots x_{10} > \tau\right) \cdot \pi(x_1) \cdots \pi(x_{10})\, dx_1 \cdots dx_{10}$$

$$= \int_{x_1} \cdots \int_{x_{10}} I(x_1 + \ldots x_{10} > \tau) \cdot w(x_1 \ldots x_{10}) \cdot q(x_1) \cdots q(x_{10})\, dx_1 \cdots dx_{10}$$

2. a) $f(\beta | y, q, \sigma) \propto f(y | \beta, q, \sigma) \cdot f(\beta)$

$$\propto \exp\left(-\frac{q}{2}(y - X\beta)^t R^{-1}(y - X\beta) - \beta^t S^{-1} \beta\right)$$

$$\propto \exp\left(-\frac{1}{2}\beta^t \left(q X^t R^{-1} X + S^{-1}\right)\beta + \beta^t \left(q X^t R^{-1} y\right)\right)$$

This is a Gaussian with covariance $\Sigma_{\beta | y, q, \sigma} = \left(q X^t R^{-1} X + S^{-1}\right)^{-1}$ and mean

$$\mu_{\beta | y, q, \sigma} = \Sigma_{\beta | y, q, \sigma} \left[q X^t R^{-1} y\right].$$

Since Gaussian is $y \sim N(\mu, \Sigma)$

$$\exp\left(-\frac{1}{2}(y - \mu)' \Sigma^{-1} (y - \mu)\right)$$

$$\propto \exp\left(-\frac{1}{2} y' \Sigma^{-1} y + y' \Sigma^{-1} \mu\right)$$

iii)

a) $f(\tau \mid \beta, y, e) \propto f(y \mid \beta, \tau, e) \cdot f(\tau)$

$$\propto \frac{\tau^{n/2}}{|R|^{1/2}} \exp\left(-\frac{\tau}{2}(y-X\beta)^t R^{-1}(y-X\beta)\right) \cdot \tau^{a-1} e^{-b\tau}$$

$$\propto \tau^{\frac{n}{2}+a-1} \exp\left[-\tau\left(b + \frac{1}{2}(y-X\beta)^t R^{-1}(y-X\beta)\right)\right]$$

This is Gamma$\left(\frac{n}{2}+a,\ \frac{1}{2}(y-X\beta)^t R^{-1}(y-X\beta)+b\right)$

b) A Gibbs sampler iterates between β and τ updates:

Initiate $\{$ β^0, τ^0

while $b = 1, \ldots, B$

$\{$

$\beta^b \sim f(\beta \mid y, \tau^{b-1}, e)$

$\tau^b \sim f(\tau \mid y, \beta^b, e)$

$\}$

This defines a Markov chain of $(\beta^0, \tau^0), (\beta^1, \tau^1), \ldots (\beta^B, \tau^B)$ values that converges to $f(\beta, \tau \mid y, e)$.

c)

Proposal $g(\varphi|\varphi(b)) = \text{Beta}(c,d)$.

Akseptrate:

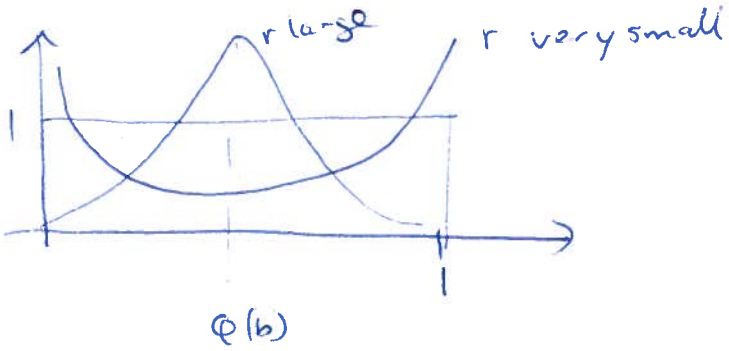
$$\alpha = \min\left(1, \frac{f(y|\varphi,\beta,q)\, f(\varphi) \cdot g(\varphi(b)|\varphi)}{f(y|\varphi(b),\beta,q)\, f(\varphi(b))\, g(\varphi|\varphi(b))}\right)$$

β, q fixed, so $f(\beta)f(q)$ cancels.

$f(\varphi) = f(\varphi(b)) = 1$ also cancels.

$$\alpha = \min\left(1, \frac{|R(\varphi)|^{-1/2} e^{-\frac{1}{2}(y-X\beta)^t R(\varphi)^{-1}(y-X\beta)} \cdot \varphi(b)^{c(\varphi)-1} (1-\varphi(b))^{d(\varphi)-1}}{|R(\varphi(b))|^{-1/2} e^{-\frac{1}{2}(y-X\beta)^t R(\varphi(b))^{-1}(y-X\beta)} \cdot \varphi^{c(\varphi(b))-1} (1-\varphi)^{d(\varphi(b))-1}}\right)$$

The proposal is beta-distributed, centered at the current value. If r is very small, the variance in the proposal increases.

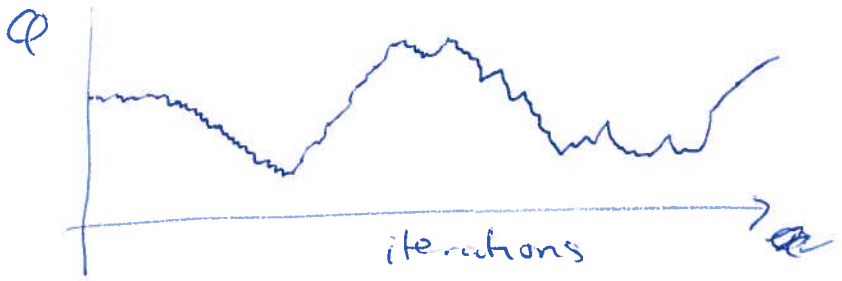


r large gives smaller changes $|\varphi - \varphi(b)|$, and the proposal is almost always accepted.

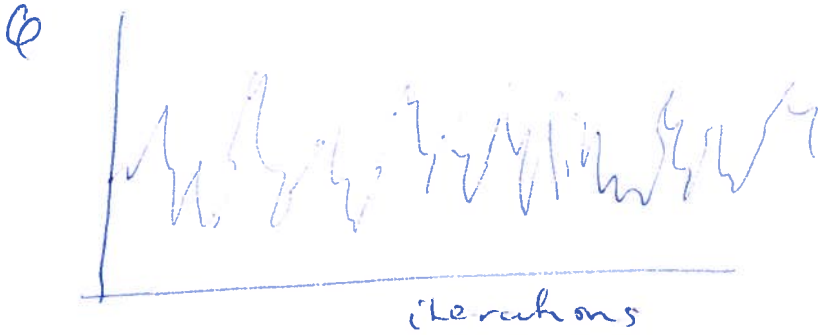
r small gives larger changes $|\varphi - \varphi(b)|$, and the proposal is often rejected.

V)

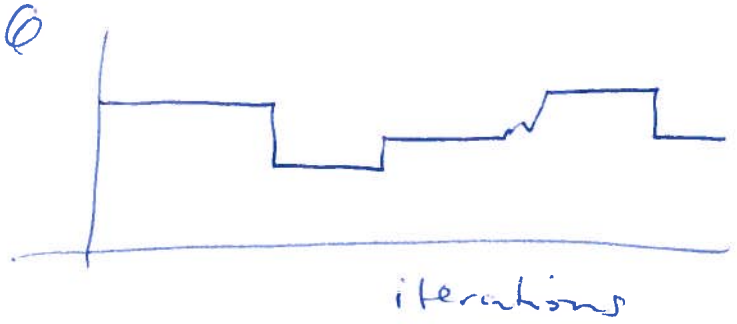
For intermediate r values, the mean acceptance rate is ~ 0.25 - 0.5, which tends to give reasonable mixing.



r large



r intermediate



r small

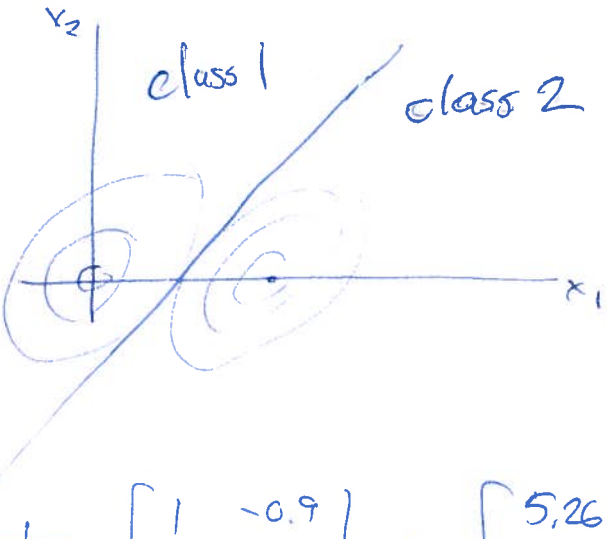
Illustration (traceplot) of $\theta^1, \ldots \theta^B$ from a MH sampler with varying r-values.

3

a) LDA picks k - class with largest density. When $\pi_l = \frac{1}{K}$:

$$k = \underset{l}{\operatorname{argmax}}\ f(x|l), \quad f(x|l) \sim N(x; \mu_l, \Sigma_l)$$

For $\Sigma_1 = \Sigma_2 = \dots = \Sigma_K$, this gives linear boundaries for classification



$$\Sigma^{-1} = \frac{1}{1-0.9^2}\begin{bmatrix} 1 & -0.9 \\ -0.9 & 1 \end{bmatrix} = \begin{bmatrix} 5.26 & -4.74 \\ -4.74 & 5.26 \end{bmatrix}$$

Boundary (since $\pi_1 = \pi_2 = 0.5$):

$$\underbrace{x^t \Sigma^{-1} \mu_1}_{0} - \frac{1}{2}\underbrace{\mu_1^t \Sigma^{-1} \mu_1}_{0} = x^t \Sigma^{-1} \mu_2 - \frac{1}{2}\mu_2^t \Sigma^{-1} \mu_2$$

$$= (x_1\ x_2)\begin{bmatrix} 5.26 & -4.74 \\ -4.74 & 5.26 \end{bmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} - \frac{1}{2}(1\ 0)\begin{bmatrix} 5.26 & -4.74 \\ -4.74 & 5.26 \end{bmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

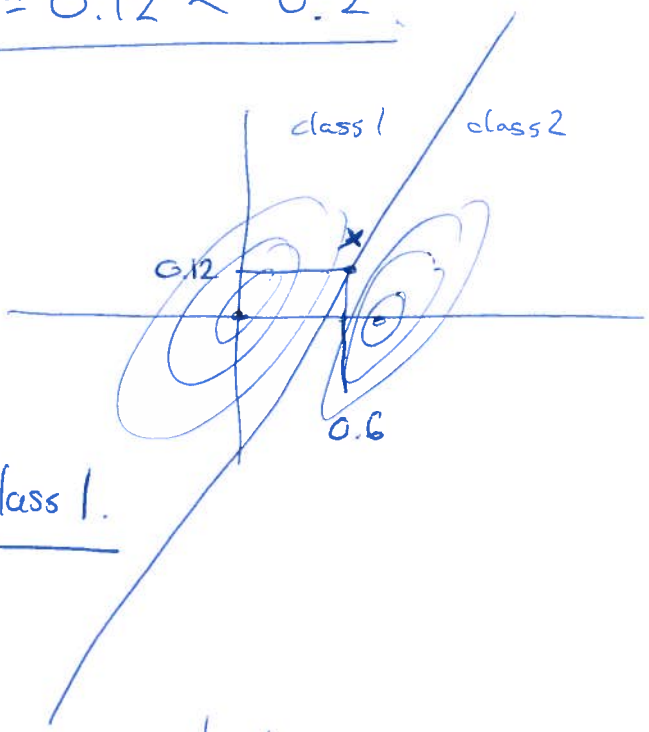
$$= 5.26 \cdot x_1 - 4.74\, x_2 - \frac{5.26}{2}$$

$$x_2 = \frac{5.26}{4.74} x_1 - \frac{5.26}{4.74} \cdot \frac{1}{2} = 1.11\, x_1 - 0.55$$

vii)

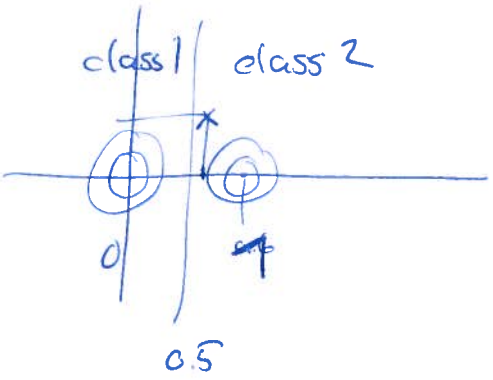
$x = (0.6, 0.2)$

$1.11 \cdot 0.6 - 0.55 = 0.12 < 0.2$



Response is classified in class 1.

If $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$,



the separation is made only on the first component, where $\mu_1 \neq \mu_2$.

Then x is classified in class 2, since $0.6 > 0.5$.

Viii)

3b) $\hat{\mu}_1, \hat{\mu}_2, \hat{\Sigma}, \hat{\pi}_1$ are MLE

A parametric bootstrap uses the model (based on MLEs) to generate "new" data. From these "new" data, we compute the MLE's. As the number of sample increases, these are representative of the uncertainty of the $\hat{\mu}_1, \hat{\mu}_2, \hat{\Sigma}, \hat{\pi}_1$.

Idea

for $b = 1, \ldots, B$

{ Simulate $x^{*b} \sim \sum_{k=1}^{K} \hat{\pi}_k N(x; \hat{\mu}_k, \hat{\Sigma})$.

Compute $\hat{\mu}_1^{*b}, \hat{\mu}_2^{*b}, \hat{\pi}_1^{*b}, \hat{\Sigma}^{*b}$ from x^{*b}

The MLE's from data x^{*b}.

}

$\hat{\mu}_1^{*1}, \ldots \hat{\mu}_1^{*B}$ represent uncertainty in $\hat{\mu}_1$

$\hat{\mu}_2^{*1}, \ldots \hat{\mu}_2^{*B}$ ——— " ——— $\hat{\mu}_2$

$\hat{\pi}_1^{*1}, \ldots \hat{\pi}_1^{*B}$ ——— " ——— $\hat{\pi}_1$

$\hat{\Sigma}^{*1}, \ldots \hat{\Sigma}^{*B}$ ——— " ——— $\hat{\Sigma}$.