Norwegian University of Science and Technology
Department of Mathematicak Sciences

Solution - Exam June 2010

Problem 1

a) We have

$$\mathrm{P}(X_i = x) = \exp\{x \ln p + (1-x)\ln(1-p)\} = \exp\{x(\ln p - \ln(1-p)) + \ln(1-p)\}$$
$$= \exp\left\{x \ln \frac{p}{1-p} + \ln(1-p)\right\}$$

Thus, we may choose for example

$$\underline{\underline{a(x) = 1 \;\; , \;\; \phi(p) = \ln \frac{p}{1-p} \;\; , \;\; t(x) = x \;\; \text{and} \;\; b(p) = \ln(1-p).}}$$

The conjugate prior distribution becomes

$$\pi(p) \propto \exp\{\phi(p)a + b(p)b\} = \exp\left\{a \ln \frac{p}{1-p} - b\ln(1-p)\right\}$$
$$= \left(\frac{p}{1-p}\right)^a (1-p)^b = p^a(1-p)^{b-a},$$

where a and b are two parameters. After the reparameterisation $\alpha = a+1$ and $\beta = b-a+1$ we get

$$\underline{\underline{\pi(p) \propto p^{\alpha-1}(1-p)^{\beta-1}}}$$

as required. This is called a beta distribution.

b) We get

$$\pi(p|x_1,\ldots,x_n) \propto \pi(p)\mathrm{P}(X_1 = x_1,\ldots,X_n = x_n|p) = \pi(p)\prod_{i=1}^{n} \mathrm{P}(X_i = x_i|p)$$

$$\propto p^{\alpha-1}(1-p)^{\beta-1}\prod_{i=1}^{n}\left[p^{x_i}(1-p)^{1-x_i}\right] = p^{\alpha+\sum_{i=1}^{n} x_i - 1}(1-p)^{\beta+\sum_{i=1}^{n}(1-x_i)-1}$$

$$= p^{\alpha+\sum_{i=1}^n x_i-1}(1-p)^{\beta+n-\sum_{i=1}^n x_i-1}.$$

Thus,

$$\underline{\underline{\widetilde{\alpha} = \alpha + \sum_{i=1}^n x_i \quad \text{and} \quad \widetilde{\beta} = \beta + n - \sum_{i=1}^n x_i.}}$$

For a rejection sampling algorithm with proposal distribution $g(p)$ and target distribution $\pi(p|x_1,\ldots,x_n)$, the acceptance probability for a proposed value p is

$$r = c\frac{\pi(p|x_1,\ldots,x_n)}{g(p)},$$

where c is a constant ensuring that $r \leq 1$ for all values p. With $g(p) = 1$ and the $\pi(p|x_1,\ldots,x_n)$ given above we get

$$r = cp^{\widetilde{\alpha}-1}(1-p)^{\widetilde{\beta}-1}.$$

To find a legal value for c we must maximise r with respect to p, set the maximal value equal to 1 and solve the resulting equation with respect to c. We start by finding the derivative

$$\frac{\partial \ln r}{\partial p} = \frac{\partial}{\partial p}[(\widetilde{\alpha}-1)\ln p + (\widetilde{\beta}-1)\ln(1-p)]$$

$$= \frac{\widetilde{\alpha}-1}{p} - \frac{\widetilde{\beta}-1}{1-p} = \frac{\widetilde{\alpha}-1-(\widetilde{\alpha}+\widetilde{\beta}-2)p}{p(1-p)}.$$

Equating the derivative to zero we get

$$p = \frac{\widetilde{\alpha}-1}{\widetilde{\alpha}+\widetilde{\beta}-2}.$$

It is given that $\alpha, \beta \geq 1$, so we also have $\widetilde{\alpha}, \widetilde{\beta} \geq 1$. Thereby

$$\frac{\widetilde{\alpha}-1}{\widetilde{\alpha}+\widetilde{\beta}-2} \in [0,1],$$

so we ensure the acceptance probability is less than or equal to one by setting

$$\max_{p\in[0,1]}\left[cp^{\widetilde{\alpha}-1}(1-p)^{\widetilde{\beta}-1}\right] = c\left(\frac{\widetilde{\alpha}-1}{\widetilde{\alpha}+\widetilde{\beta}-2}\right)^{\widetilde{\alpha}-1}\left(\frac{\widetilde{\beta}-1}{\widetilde{\alpha}+\widetilde{\beta}-2}\right)^{\widetilde{\beta}-1} = 1$$

$$\Rightarrow c = \left(\frac{\widetilde{\alpha}+\widetilde{\beta}-2}{\widetilde{\alpha}-1}\right)^{\widetilde{\alpha}-1}\left(\frac{\widetilde{\alpha}+\widetilde{\beta}-2}{\widetilde{\beta}-1}\right)^{\widetilde{\beta}-1}.$$

Thereby the acceptance probability becomes

$$r = \left(p \cdot \frac{\widetilde{\alpha} + \widetilde{\beta} - 2}{\widetilde{\alpha} - 1}\right)^{\widetilde{\alpha}-1} \left((1-p) \cdot \frac{\widetilde{\alpha} + \widetilde{\beta} - 2}{\widetilde{\beta} - 1}\right)^{\widetilde{\beta}-1}.$$

Pseudo-code for the rejection sampling algorithm:

1. Generate a potential value for p, $p \sim \text{Unif}(0,1)$.
2. Compute the acceptance probability
$$r = \left(p \cdot \frac{\widetilde{\alpha} + \widetilde{\beta} - 2}{\widetilde{\alpha} - 1}\right)^{\widetilde{\alpha}-1} \left((1-p) \cdot \frac{\widetilde{\alpha} + \widetilde{\beta} - 2}{\widetilde{\beta} - 1}\right)^{\widetilde{\beta}-1}.$$
3. Generate $u \sim \text{Unif}(0,1)$.
4. If $u \leq r$ return p, otherwise goto 1.

Problem 2

a) The full conditional distribution for p becomes

$$\begin{aligned}\pi(p|\text{everything else}) &\propto \pi(p, c_1, \ldots, c_k, \mu_0, \theta_0, \mu_1, \theta_1, x_1, \ldots, x_n, y_1, \ldots, y_m, z_1, \ldots, z_k) \\ &= \pi(p)\left[\prod_{i=1}^{k} \pi(c_i|p)\right] \pi(\mu_0)\pi(\theta_0)\pi(\mu_1)\pi(\theta_1)\left[\prod_{i=1}^{n} \pi(x_i|\mu_0,\theta_0)\right]\left[\prod_{i=1}^{m} \pi(y_i|\mu_1,\theta_1)\right] \\ &\quad \cdot \left[\prod_{i=1}^{k} \pi(z_i|c_i,\mu_0,\theta_0,\mu_1,\theta_1)\right] \\ &\propto \pi(p)\prod_{i=1}^{k} \pi(c_i|p),\end{aligned}$$

where the proportionalities are as a function of p. Here $\pi(p)$ is a uniform distribution on $[0,1]$ and $\pi(c_i|p) = p^{c_i}(1-p)^{1-c_i}$. Thus, we have the same situation as considered in Problem 1 with $\alpha = \beta = 1$ and where our c_i corresponds to x_i in Problem 1. Thus, the full conditional distribution becomes a beta distribution with parameters $\widetilde{\alpha} = 1 + \sum_{i=1}^{k} c_i$ and $\widetilde{\beta} = 1 + k - \sum_{i=1}^{k} c_i$,

$$\pi(p|\text{everything else}) \propto p^{\widetilde{\alpha}-1}(1-p)^{\widetilde{\beta}-1}$$

b) As a function of c_i we get

$$\pi(c_i|\text{everything else}) \propto \pi(p, c_1, \ldots, c_k, \mu_0, \theta_0, \mu_1, \theta_1, x_1, \ldots, x_n, y_1, \ldots, y_m, z_1, \ldots, z_k)$$

$$= \pi(p) \left[\prod_{j=1}^{k} \pi(c_j|p)\right] \pi(\mu_0)\pi(\theta_0)\pi(\mu_1)\pi(\theta_1) \left[\prod_{j=1}^{n} \pi(x_j|\mu_0, \theta_0)\right] \left[\prod_{i=j}^{m} \pi(y_j|\mu_1, \theta_1)\right]$$

$$\cdot \left[\prod_{i=1}^{k} \pi(z_i|c_i, \mu_0, \theta_0, \mu_1, \theta_1)\right] \propto \pi(c_i|p)\pi(z_i|c_i, \mu_0, \theta_0, \mu_1, \theta_1).$$

Thus, for some normalising constant v we have

$$\pi(c_i = 0|\text{everything else}) = v \cdot \pi(c_i = 0|p)\pi(z_i|c_i = 0, \mu_0, \theta_0, \mu_1, \theta_1)$$

$$= v(1-p)\frac{1}{\sqrt{2\pi\theta_0}} \exp\left\{-\frac{(z_i - \mu_0)^2}{2\theta_0}\right\}$$

and

$$\pi(c_i = 1|\text{everything else}) = v \cdot \pi(c_i = 1|p)\pi(z_i|c_i = 1, \mu_0, \theta_0, \mu_1, \theta_1)$$

$$= vp\frac{1}{\sqrt{2\pi\theta_1}} \exp\left\{-\frac{(z_i - \mu_1)^2}{2\theta_1}\right\}.$$

Thereby, since we must have $\pi(c_i = 0|\text{everything else}) + \pi(c_i = 1|\text{everything else}) = 1$, we get

$$\pi(c_i|\text{everything else}) = \frac{\left(\frac{1-p}{\sqrt{\theta_0}} \exp\left\{-\frac{(z_i-\mu_0)^2}{2\theta_0}\right\}\right)^{1-c_i} \left(\frac{p}{\sqrt{\theta_1}} \exp\left\{-\frac{(z_i-\mu_1)^2}{2\theta_1}\right\}\right)^{c_i}}{\left(\frac{1-p}{\sqrt{\theta_0}} \exp\left\{-\frac{(z_i-\mu_0)^2}{2\theta_0}\right\}\right) + \left(\frac{p}{\sqrt{\theta_1}} \exp\left\{-\frac{(z_i-\mu_1)^2}{2\theta_1}\right\}\right)}.$$

Pseudo-code for simulating from the full conditional:

1. Compute the probability for $c_i = 1$:

$$\pi(c_1 = 1|\text{everything else}) = \frac{\frac{p}{\sqrt{\theta_1}} \exp\left\{-\frac{(z_i-\mu_1)^2}{2\theta_1}\right\}}{\left(\frac{1-p}{\sqrt{\theta_0}} \exp\left\{-\frac{(z_i-\mu_0)^2}{2\theta_0}\right\}\right) + \left(\frac{p}{\sqrt{\theta_1}} \exp\left\{-\frac{(z_i-\mu_1)^2}{2\theta_1}\right\}\right)}.$$

2. Generate $u \sim \text{Unif}(0, 1)$.

3. If $u \leq \pi(c_i = 1|\text{everything else})$ return $c_i = 1$, otherwise return $c_i = 0$.

c) First one needs to find the length of the burn-in phase of the Markov chain. This can be done by output analysis. Assuming the Markov chain has (essentially) converged after $T < S$ iterations, the natural estimator for the first quantity is

$$\widehat{\mathrm{P}}(c_i = 1|x_1, \ldots, x_n, y_1, \ldots, y_m, z_1, \ldots, z_k) = \frac{1}{S-T+1} \sum_{s=T}^{S} c_i^s.$$

No additional assumptions are necessary.

The natural classification rule for observation z_i is

$$\widehat{c_i} = \begin{cases} 0 & \text{if } \widehat{\mathrm{P}}(c_i = 1|x_1, \ldots, x_n, y_i, \ldots, y_m, z_1, \ldots, z_k) \leq 0.5 \\ 1 & \text{otherwise.} \end{cases}$$

No additional assumptions are necessary here.

To estimate the error rate one needs to make an assumption about the distribution of the test data, (c, z). A natural assumption is to assume we want to estimate the error rate for test data that are coming from the same distribution as the observed $z_1, \ldots, z_k$. If so, the natural estimator of the error rate is

$$\widehat{\mathrm{err}} = \frac{1}{k} \sum_{i=1}^{k} \widehat{\mathrm{P}}(c_i \neq \widehat{c_i}|x_1, \ldots, x_n, y_1, \ldots, y_m, z_1, \ldots, z_k).$$

One should note that one then use the same data both for training and testing, so this estimator would tend to be optimistic.

d) By stating that the c_i's have a common probability p of being equal to 1, as we are doing in the first model but not in the alternative model, we increase the probability for the c_i's to be equal. Thus, if for example 75% of the z_i's are classified as class 0 by the first model we would expect that slightly less than 75% are classified as class 0 in the alternative model. Conversely, if 25% of the z_i's are classified to class 0 in the first model we would expect that slightly more than 25% of the cases are classified as class 0 in the alternative model.

Problem 3

a) The bias of $\widehat{\theta}$ is defined as

$$\mathrm{bias}_F = \mathrm{bias}_F(\widehat{\theta}, \theta) = \mathrm{E}_F[\widehat{\theta}] - \theta = \mathrm{E}_F[s(x)] - t(F) = \mathrm{E}_F\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right] - \mathrm{Var}_F(x).$$

The ideal bootstrap estimator for the bias of $\widehat{\theta}$ is then given as

$$\begin{aligned}\text{bias}_{\widehat{F}} &= \text{E}_{\widehat{F}}\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i^\star - \bar{x}^\star)^2\right] - \text{Var}_{\widehat{F}}(x^\star),\\ &= \text{E}_{\widehat{F}}\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i^\star - \bar{x}^\star)^2\right] - \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2,\end{aligned}$$

where $\widehat{F}$ is the empirical distribution of $x_1, \ldots, x_n$.

The ideal bootstrap estimator can be estimated by Monte Carlo simulation by the following algorithm:

1. For $b = 1, \ldots, B$ and $i = 1, \ldots, n$ draw $x_i^{\star(b)}$ from the values $x_1, \ldots, x_n$ independently at random. For $b = 1, \ldots, B$, form the bootstrap samples $x^{\star(b)} = (x_1^{\star(b)}, \ldots, x_n^{\star(b)})$.
2. For $b = 1, \ldots, B$ compute
$$\widehat{\theta}^\star(b) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i^{\star(b)} - \bar{x}^{\star(b)})^2 \quad \text{where} \quad \bar{x}^{\star(b)} = \frac{1}{n}\sum_{i=1}^{n} x_i^{\star(b)}.$$
3. Estimate the ideal bootstrap estimator for the bias of $\widehat{\theta}$ by
$$\widehat{\text{bias}}_B = \frac{1}{B}\sum_{b=1}^{B}\widehat{\theta}^\star(b) - \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

b)
From the solution of **a**) we have that

$$\text{bias}_{\widehat{F}} = \text{E}_{\widehat{F}}\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i^\star - \bar{x}^\star)^2\right] - \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2. \tag{1}$$

Thus, we need to evaluate analytically the first of the two terms. We have

$$\text{E}_{\widehat{F}}\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i^\star - \bar{x}^\star)^2\right] = \frac{1}{n-1}\sum_{i=1}^{n}\text{E}_{\widehat{F}}\left[(x_i^\star - \bar{x}_i^\star)^2\right] = \frac{n}{n-1}\text{E}_{\widehat{F}}\left[(x_i^\star - \bar{x}_i^\star)^2\right]. \tag{2}$$

Expanding the square, and inserting for $\bar{x}^\star = (1/n)\sum_{j=1}^{n} x_j^\star$, we get

$$\text{E}_{\widehat{F}}\left[(x_i - \bar{x}_i^\star)^2\right] = \text{E}_{\widehat{F}}[(x_i^\star)^2] - 2\text{E}_{\widehat{F}}[x_i^\star \bar{x}_i^\star] + \text{E}_{\widehat{F}}[(\bar{x}^\star)^2]$$

$$
\begin{aligned}
&= \mathrm{E}_{\widehat{F}}[(x_i^\star)^2] - \frac{2}{n}\sum_{j=1}^{n}\mathrm{E}_{\widehat{F}}[x_i^\star x_j^\star] + \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\mathrm{E}_{\widehat{F}}[x_j^\star x_k^\star] \\
&= \mathrm{E}_{\widehat{F}}[(x_i^\star)^2] - \frac{2}{n}\left[\mathrm{E}_{\widehat{F}}[(x_i^\star)^2] + \sum_{j\neq i}\mathrm{E}_{\widehat{F}}[x_i^\star x_j^\star]\right] + \frac{1}{n^2}\left[\sum_{j=1}^{n}\mathrm{E}_{\widehat{F}}[(x_j^\star)^2] + \sum_{j=1}^{n}\sum_{k\neq j}\mathrm{E}_{\widehat{F}}[x_j^\star x_k^\star]\right] \\
&= \left(1-\frac{2}{n}\right)\mathrm{E}_{\widehat{F}}[(x_i^\star)^2] + \frac{1}{n^2}\sum_{j=1}^{n}\mathrm{E}_{\widehat{F}}[(x_j^\star)^2] - \frac{2}{n}\sum_{j\neq i}\mathrm{E}_{\widehat{F}}[x_i^\star]\mathrm{E}_{\widehat{F}}[x_j^\star] + \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k\neq j}\mathrm{E}_{\widehat{F}}[x_j^\star]\mathrm{E}_{\widehat{F}}[x_k^\star].
\end{aligned}
$$

Inserting this into (2) and using that $x_i^\star$ has the same distribution for all $i = 1, \ldots, n$ we get

$$
\begin{aligned}
&\mathrm{E}_{\widehat{F}}\left[\frac{1}{n-1}\sum_{i=1}^{n}\sum_{i=1}^{n}(x_i^\star - \bar{x}^\star)^2\right] \\
&= \frac{n}{n-1}\left[\left(1-\frac{2}{n}+\frac{n}{n^2}\right)\mathrm{E}_{\widehat{F}}[(x_i^\star)^2] + \left(\frac{n(n-1)}{n^2} - \frac{2(n-1)}{n}\right)\left(\mathrm{E}_{\widehat{F}}[x_i^\star]\right)^2\right] \\
&= \mathrm{E}_{\widehat{F}}[(x_i^\star)^2] - \left(\mathrm{E}_{\widehat{F}}[x_i^\star]\right)^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2.
\end{aligned}
$$

Inserting this into (1) we get

$$
\underline{\underline{\mathrm{bias}_{\widehat{F}} = 0.}}
$$