Petter Mostad
Applied Mathematics and Statistics
Chalmers

Suggested solutions for
MSA100 / MVE186 Computer Intensive Statistical Methods
Re-exam 2 January 2017

1. A random number is a mathematical concept whereas a pseudo-random number is a number generated by a computer in such a way that it appears to have the properties of a random number in the contexts we investigate. In other words, although a sequence of pseudo-random numbers is completely deterministic, looking at the numbers or functions of them, one should not be able to detect any correlations or patterns.

2. (a) The cumulative dennsity function of a Logistic$(-1, 1)$ distribution is

$$F(x) = \frac{1}{1 + \exp\left(-(x+1)\right)}$$

Inverting this function gives

$$F^{-1}(u) = -1 - \log\left(u^{-1} - 1\right)$$

and thus with $U \sim$ Uniform$(0, 1)$, $F^{-1}(U)$ is a sample from Logistic$(-1, 1)$.

(b) First compute, for $i = 0, 1, \ldots, 12$,

$$c_i = \sum_{j=0}^{i} \binom{12}{j} 0.42^j 0.58^{12-j}.$$

Then simulate $U \sim$ Uniform$(0, 1)$ and set

$$k = \min_k c_k \geq U.$$

(c) One possibility is to use rejection sampling: The density of a Beta$(4.5, 9.2)$ distribution is given by $(0 \leq x \leq 1)$

$$f(x) = \frac{\Gamma(4.5 + 9.2)}{\Gamma(4.5)\Gamma(9.2)} x^{4.5-1}(1-x)^{9.2-1}$$

and attains it maximum at the distribution mode

$$m = \frac{4.5 - 1}{4.5 + 9.2 - 2}$$

so this maximum can be computed as $M = f(m)$. To use rejection sampling, sample $U \sim$ Uniform$(0, 1)$, and accept it with the probability $f(u)/M$.

3. (a) A Gibbs sampling would in essence iterate between simulating from the conditional distribution of x given y, and the conditional distribution of y given x. First of all, these conditional distributions are very difficult to handle and to simulate from in this case. Secondly, even if one manages to simulate from them, the resulting Markov Chain would be badly behaved, as x and y are highly correlated, as indicated in the figure.

   (b) A first step would be to do a variable transformation, substituting $u = x + y$ and $v = x - y$. As this is a linear transformation it would change the density only with a constant, which can be ignored as we know the density only up to a constant.
   The new density function

$$g(u, v) = \frac{1}{(|u|/10 + 1)^4} \cdot (|v| + 0.1)^2 \cdot e^{-7|u|}$$

   separates into a product of functions of u and v, so these can be simulated independently.
   Because of symmetry, it is clear that we may simulate from the density proportional to

$$h(u) = \frac{1}{(u/10 + 1)^4}$$

   for $u > 0$ and the density proportional to

$$k(v) = (v + 0.1)^2 \exp(-7v)$$

   for $v > 0$, annd then choose the signs randomly. Again we may make linear transformations $w = u/10 + 1$ and $s = u + 0.1$ to obtain the functions

$$A(w) = w^{-4}$$

   for $w > 1$ and

$$B(s) = s^2 \exp(-7s)$$

   for $s > 0.1$. For the first function, the cumulative distribution can be computed explicitly and used for simulation. For the second function, it is proportional to a Gamma density, so one may simulate from it by simulating from a Gamma density with parameters 3 and 7, and throwing away results below 0.1.

4. Let X denote a random variable with density $\pi(x)$ on $[a, b]$. Provided that $\mathbb{E}[g(X)]$ exists, we may write I as $I = \mathbb{E}[g(X)]$ and the Central Limit Theorem tells us that

$$\hat{I} = \frac{1}{n}\sum_{i=1}^{n} g(x_i) \sim \text{Normal}(I, \sigma^2/n)$$

   provided we assume $\sigma^2 = \text{Var}[g(X)]$ exists. Making these assumptions, we get that $\hat{I}$ as defined above is an estimate for I, and that an approximate 95% confidence interval is given by

$$\hat{I} \pm 1.96 \cdot \hat{\sigma}/\sqrt{n}$$

   where $\hat{\sigma}^2$ is the sample variance computed from $g(x_1), g(x_2), \ldots, g(x_n)$.

5. Importance sampling is a method to improve the accuracy in the evaluation of integrals of the form

$$\mathbb{E}_\pi\left(h(X)\right) = \int h(x)\pi(x)\,dx$$

where $\pi(x)$ is some density function. Specifically, one assumes there is an instrumental density $g(x)$ which is strictly positive whenever $h(x)\pi(x)$ is positive, and write

$$\mathbb{E}_\pi\left(h(X)\right) = \int h(x)\pi(x)\,dx = \int \frac{h(x)\pi(x)}{g(x)} g(x)\,dx \approx \frac{1}{n}\sum_{i=1}^{n}\frac{h(x_i)\pi(x_i)}{g(x_i)}$$

where $x_1, \ldots, x_n$ is a sample from the distribution whose density is $g(x)$. Using the Central Limit Theorem, we see that the approximation is better than the approximation $1/n\sum_{i=1}^{n} h(y_i)$ (where $y_1, \ldots, y_n$ is a sample from the distribution with density $\pi(x)$) if

$$\mathrm{Var}\left[\frac{h(X)\pi(X)}{g(X)}\right] < \mathrm{Var}\left[h(Y)\right].$$

Thus, generally, $\frac{h(x)\pi(x)}{g(x)}$ should be as stable as possible. Note that it needs to be practical to simulate from the density $g(x)$; when it is easier to simulate from $g(x)$ than from $\pi(x)$ it is another advantage of importance sampling.

6. (a) We get for the posterior

$$\pi(p \mid x) \propto_p \pi(x \mid p)\pi(p) \propto_p p^x(1-p)^{n-x}$$

which is proportional to a Beta$(x+1, n-x+1)$ distribution. The posterior expectation is, from the properties of the Beta distribution,

$$\mathbb{E}\left[p \mid x\right] = \frac{x+1}{n-x+1+x+1} = \frac{x+1}{n+2}$$

while the maximum likelihood estimate for p is $\frac{x}{n}$.

(b) We have for the predictive distribution

$$\begin{aligned}
\pi(y \mid x) &= \frac{\pi(y \mid p)\pi(p \mid x)}{\pi(p \mid x, y)} \\
&= \frac{\mathrm{Binomial}(y; m, p)\cdot \mathrm{Beta}(p; x+1, n-x+1)}{\mathrm{Beta}(p; x+y+1, n+m-x-y+1)} \\
&= \frac{\binom{m}{y}p^y(1-p)^{m-y}\frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}p^x(1-p)^{n-x}}{\frac{\Gamma(n+m+2)}{\Gamma(x+y+1)\Gamma(n+m-x-y+1)}p^{x+y}(1-p)^{n+m-x-y}} \\
&= \binom{m}{y}\frac{\Gamma(n+2)\Gamma(x+y+1)\Gamma(n+m-x-y+1)}{\Gamma(x+1)\Gamma(n-x+1)\Gamma(n+m+2)}
\end{aligned}$$

(c) We get $\mathbb{E}[y] = \mathbb{E}[\mathbb{E}[y \mid x]] = \mathbb{E}[mp \mid x] = m\frac{x+1}{n+2}$, and also

$$\begin{aligned}
\mathrm{Var}[y] &= \mathrm{Var}[\mathbb{E}[y \mid x]] + \mathbb{E}[\mathrm{Var}[y \mid x]] \\
&= \mathrm{Var}[mp \mid x] + \mathbb{E}[mp(1-p) \mid x] \\
&= m^2 \mathrm{Var}[p \mid x] + m\,\mathbb{E}[p \mid x] - m\,\mathbb{E}\left[p^2 \mid x\right] \\
&= m^2 \mathrm{Var}[p \mid x] + m\,\mathbb{E}[p \mid x] - m\,\mathrm{Var}[p \mid x] + m\,\mathbb{E}[p \mid x]^2 \\
&= (m^2 - m)\frac{(x+1)(n+1-x)}{(n+2)^2(n+3)} + m\frac{x+1}{n+2} - m\left(\frac{x+1}{n+2}\right)^2
\end{aligned}$$

7. (a) We get

$$\begin{aligned}
L(\theta) &= \log\left(\pi\left(X_0, X_1, X_2, Z \mid \mu_0, \mu_1, p\right)\right) \\
&= \log\left[\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(X_0-\mu_0)^2\right)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(X_1-\mu_1)^2\right)\right. \\
&\qquad \left.\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(X_2-\mu_Z)^2\right)p^Z(1-p)^{1-Z}\right] \\
&= C - \frac{1}{2}(X_0-\mu_0)^2 - \frac{1}{2}(X_1-\mu_1)^2 - \frac{1}{2}(X_2-\mu_Z)^2 + Z\log(p) + (1-Z)\log(1-p)
\end{aligned}$$

(b) We get

$$\begin{aligned}
z' &= \Pr\left[Z = 1 \mid X_2, \theta'\right] \\
&= \frac{\Pr\left[X_2 \mid Z=1,\theta'\right]\Pr\left[Z=1 \mid \theta'\right]}{\Pr\left[X_2 \mid Z=1,\theta'\right]\Pr\left[Z=1 \mid \theta'\right] + \Pr\left[X_2 \mid Z=0,\theta'\right]\Pr\left[Z=0 \mid \theta'\right]} \\
&= \frac{\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(X_2-\mu_1'\right)^2\right)p'}{\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(X_2-\mu_1'\right)^2\right)p' + \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(X_2-\mu_0'\right)^2\right)(1-p')} \\
&= \frac{1}{1+\exp\left(-\frac{1}{2}\left(X_2-\mu_0'\right)^2 + \frac{1}{2}\left(X_2-\mu_1'\right)^2\right)\left(\frac{1}{p'}-1\right)}
\end{aligned}$$

(c) As Z is a binary 0/1 variable, the posterior probability that $Z = 1$ is equal to its expectation given fixed θ' and Z_2. Thus

$$\begin{aligned}
Q(\theta,\theta') &= \mathbb{E}\left[L(\theta) \mid \theta'\right] \\
&= -\frac{1}{2}(X_0-\mu_0)^2 - \frac{1}{2}(X_1-\mu_1)^2 - \mathbb{E}\left[\frac{1}{2}(X_2-\mu_Z)^2 \mid \theta'\right] \\
&\quad + \mathbb{E}[Z]\log(p) + (1-\mathbb{E}[Z])\log(1-p) \\
&= -\frac{1}{2}(X_0-\mu_0)^2 - \frac{1}{2}(X_1-\mu_1)^2 - \frac{z'}{2}(X_2-\mu_1)^2 - \frac{1-z'}{2}(X_2-\mu_0)^2 \\
&\quad + z'\log(p) + (1-z')\log(1-p)
\end{aligned}$$

(d) Setting $\frac{\partial Q}{\partial p} = 0$ gives us directly that the maximizing value for p is $\hat{p} = z'$. Setting $\frac{\partial Q}{\partial \mu_1} = 0$ gives us $(X_1 - \mu_1) + z'(X_2 - \mu_1) = 0$ and thus $\hat{\mu}_1 = \frac{X_1 + z'X_2}{1+z'}$. Correspondingly we get $\hat{\mu}_0 = \frac{X_0 + (1-z')X_2}{1+1-z'}$.

8. (a) Assume you have a random sample $x_1, \ldots, x_n$ from a probability distribution with density π, and a statistic θ, computing from a sample $y_1, \ldots, y_m$ some value $\theta(y_1, \ldots, y_m)$. The idea of Bootstrapping is to approximate the properties of the statistic by studying the properties of θ applied to samples from the discrete distribution wich assignes the probability $1/n$ to each of the values x_i.

   (b) For $j = 1, \ldots, N$, for some large number N, resample with replacement sequences

$$(x_{1j}, y_{1j}), (x_{2j}, y_{2j}), \ldots, (x_{nj}, y_{nj})$$

from the original pairs $(x_1, y_1), \ldots, (x_n, y_n)$. For each sequence, compute $\hat{\rho}_j$ in the same way as $\hat{\rho}$ is computed from $(x_1, y_1), \ldots, (x_n, y_n)$. Now, compute a Bootstrap estimate for the standard error of $\hat{\rho}$ as

$$\sqrt{\frac{1}{N-1} \sum_{j=1}^{N} \left(\hat{\rho}_j - \overline{\hat{\rho}}\right)^2}$$

where $\overline{\hat{\rho}}$ is the averate of the $\hat{\rho}_j$'s.