UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Examination in:	STK4050/9050 — Statistical simulations and computation
Day of examination:	Trial exam fall 2009
Examination hours:	?? – ??
This examination set consists of 8 pages.	
Appendices:	None.
Permitted aids:	??

Make sure that your copy of the examination set is complete before you start solving the problems.

Problem 1.

Consider first the standard Weibull distribution with density function

$f_0(x_0) = \alpha x_0^{\alpha-1} e^{-x_0^\alpha}$

and cummulative distribution function

$F_0(x_0) = 1 - e^{-x^\alpha}$.

(a) Explain how the inversion method can be used to generate samples from f_0.

Consider now the general Weibull distribution with density function

$$p(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-(x/\beta)^\alpha}$$

(b) Show that if x_0 has a standard Weibull density then $x = \beta x_0$ has a general Weibull density. Discuss how this result can be used to generate random variables from the general Weibull distribution.

(Continued on page 2.)

Assume now we want to generate two *dependent* random variables that have marginal distributions that are of the Weibull form. Direct specification of dependence for the Weibull distribution can be difficult, but can be greatly simplified through transformation (this is called a *copula* approach in the literature).

(c) Let $\Phi()$ be the cummulative distribution function for the standard Normal distribution. Show that if $y \sim N(0,1)$, then

$$x = F_0^{-1}(\Phi(y))$$

has a standard Weibull distribution.

(d) Assume now that you are able to simulate $\boldsymbol{y} = (y_1, y_2)$ from a bivariate Normal distribution $N(\mathbf{0}, \boldsymbol{\Sigma})$ where $\Sigma_{1,1} = \Sigma_{2,2} = 1$ and $\Sigma_{1,2} = \Sigma_{2,1} = \rho$. Explain how you can use this to simulate two dependent Weibull distributed variables.

Problem 2.

Consider the following algorithm, which we will call Barker's algorithm (after Barker (1965) who suggested it):
Given the current state $\boldsymbol{x}^{(t)}$:

- Draw $\boldsymbol{y}$ from the proposal distribution $K(\boldsymbol{x}^{(t)}, \boldsymbol{y})$ (or transition kernel).
- Draw $U \sim \text{Uniform}[0,1]$ and update

$$\boldsymbol{x}^{(t+1)} = \begin{cases} \boldsymbol{y}, & \text{if } U \leq r_B(\boldsymbol{x}^{(t)}, \boldsymbol{y}) \\ \boldsymbol{x}^{(t)} & \text{otherwise} \end{cases}$$

where

$$r_B(\boldsymbol{x}, \boldsymbol{y}) = \frac{\pi(\boldsymbol{y})K(\boldsymbol{y}, \boldsymbol{x})}{\pi(\boldsymbol{x})K(\boldsymbol{x}, \boldsymbol{y}) + \pi(\boldsymbol{y})K(\boldsymbol{y}, \boldsymbol{x})}.$$

We will assume that $K(\boldsymbol{x}, \boldsymbol{y}) > 0$ for all $\boldsymbol{x}, \boldsymbol{y}$.

(a) Show that $\{\boldsymbol{x}_t\}$ is a Markov chain with invariant distribution $\pi(\boldsymbol{x})$.

(b) Explain how we can used the simulations $\{\boldsymbol{x}^{(t)}\}$ to estimate $E_\pi h = \int_{\boldsymbol{x}} h(\boldsymbol{x})\pi(\boldsymbol{x})d\boldsymbol{x}$.

What kind of properties of the Markov chain will influence on the precision of such an estimate?

(Continued on page 3.)

Assume A_1 and A_2 are two transition-kernels for Markov chains with the same stationary distribution π. Let v_1 be the variance of the estimate on $E_\pi h$ based on simulations using A_1 and v_2 the variance of the estimate of $E_\pi h$ using A_2.
Assume $A_1(\boldsymbol{x}, \boldsymbol{y}) \geq A_2(\boldsymbol{x}, \boldsymbol{y})$ for all $\boldsymbol{y} \neq \boldsymbol{x}$. One can then show that $v_1 \leq v_2$ (this you do not have to prove).

(c) Let

$$r_M(\boldsymbol{x}, \boldsymbol{y}) = \min\left\{1, \frac{\pi(\boldsymbol{y})K(\boldsymbol{y}, \boldsymbol{x})}{\pi(\boldsymbol{x})K(\boldsymbol{x}, \boldsymbol{y})}\right\}.$$

Show that $r_M(\boldsymbol{x}, \boldsymbol{y}) \geq r_B(\boldsymbol{x}, \boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y}$.

Use this to argue that the Metropolis-Hastings algorithm is more efficient than Barker's algorithm.

Based on the differences between these two algorithms, do you think this is a reasonable result?

Problem 3.

The *Gibbs sampler* applies to vectors of random variables. We shall in this exercise consider random pairs (X, Y). The algorithm is as follows:

Algorithm

Select X (initialization)
Repeat
Sample Y from its conditional distribution given X.
Sample X from its conditional distribution given Y.

It can under general conditions be proved that a simulation of (X, Y) appears in the limit as the loop is continued on and on. We shall below actually *prove* this result in the simple example considered.
Let (X, Y) be bivariate normal, with means $E(X) = E(Y) = 0$, variances $\text{var}(X) = \text{var}(Y) = 1$ and correlation $\text{corr}(X, Y) = \rho$. The conditional distribution of Y given $X = x$ is then normal with mean ρx and variance $1 - \rho^2$, and the conditional distribution of X given $Y = y$ is defined by symmetry. Let $\{Z_n\}$ and $\{V_n\}$ be sequences of independent normal variables $(0, 1)$. Also assume independence between sequences.

(a) Show that the Gibbs sampler sets up the double recursion

$$Y_n = \rho X_n + \sqrt{1-\rho^2}Z_n, \qquad X_n = \rho Y_{n-1} + \sqrt{1-\rho^2}V_n.$$

(Continued on page 4.)

It will be proved that as $n \to \infty$, (X_n, Y_n) converges to a sample of (X, Y) for any starting point $X_0 = \mu_0$. We shall also study the rate of convergence. Consider $\{X_n\}$ first.

(b) Show that $X_n = \rho^2 X_{n-1} + \varepsilon_n$, where $\varepsilon_n = \sqrt{1-\rho^2}(\rho Z_{n-1} + V_n)$.

Note that ε_n is normal with mean 0 and variance $\sigma_\varepsilon^2 = 1 - \rho^4$. Stochastic processes of the form $X_n = aX_{n-1} + \varepsilon_n$ is known as *autoregressive* of order one (or AR(1) for short). They are known to converge in distribution to a limit if $|a| < 1$. Take this result for granted. Explain why it applies here.

(c) Why is $E(X_n) = \rho^2 E(X_{n-1})$? Use this to establish that $E(X_n) = \rho^{2n}\mu_0$.

(d) Show that $\text{var}(X_n) = \rho^4 \text{var}(X_{n-1}) + \sigma_\varepsilon^2$. Since $\text{var}X_0 = 0$, this yields

$$\text{var}(X_n) = \frac{\sigma_\varepsilon^2}{1-\rho^4}(1-\rho^{4n}).$$

Prove it.

(e) What is the limit for $E(X_n)$ and $\text{var}(X_n)$ when $n \to \infty$? Insert for σ_ε^2.

(f) Explain by reason of symmetry that the same results applies to $\{Y_n\}$.

(g) Show that $E(X_nY_n) = \rho E(X_n^2)$ and use this to show that $E(X_nY_n)$ converges to the right value. (Note that in this case $E(X_nY_n) = \text{corr}(X_n, Y_n)$.)

(h) Summarize your findings. What is the limit distribution of (X_n, Y_n)? Discuss the convergence speed. What is its dependence on ρ?

Problem 4.

Consider the following state space model:

$x_t = \phi x_{t-1} + \varepsilon_t$	state equation
$y_t \sim \text{Poisson}(\exp\{1 + x_t\})$	observation equation

where x_0 and $\varepsilon_1, \varepsilon_2, ...$ are independent and standard normal distributed. We want to estimate ϕ based on observations $y_1, ..., y_T$. We will do this in a Bayesian way and assume we have a prior distribution $N(0, \sigma_\phi^2)$ on ϕ.

(Continued on page 5.)

A possible way to estimate ϕ in such situations is to extend the state model to the following model:

$\phi_t = \phi_{t-1}$	state equation 1
$x_t = \phi_{t-1} x_{t-1} + \varepsilon_t$	state equation 2
$y_t \sim \text{Poisson}(\exp\{1 + x_t\})$	observation equation

where $\phi_0 \sim N(0, \sigma_\phi^2)$. Non-linear filters try to compute the posterior distribution for (ϕ_t, x_t) based on $y_1, ..., y_t$. Since $\phi = \phi_T$, the posterior distribution for (ϕ_T, x_T) based on $y_1, ..., y_T$ gives us the posterior distribution for ϕ given $y_1, ..., y_T$.
Simulation methods for non-linear filters can therefore be used on the *bivariate* state vector (ϕ_t, x_t).

(a) Explain the general principles behind sequential importance sampling (SIS).

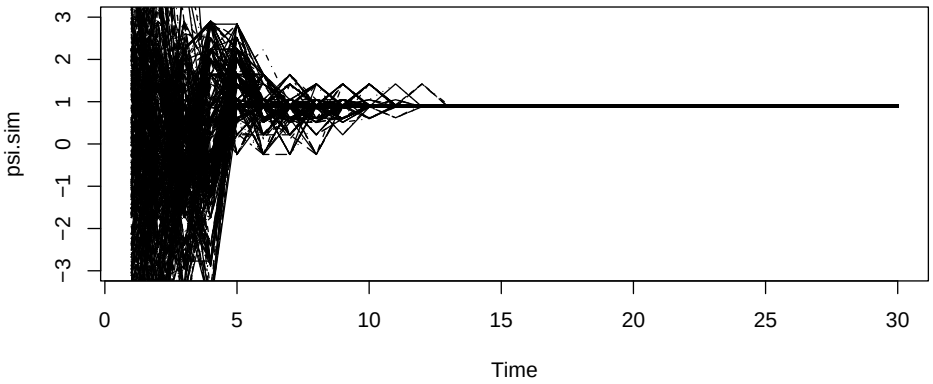
Discuss why resampling in general is important in connection to SIS algorithms.

(b) Simulations from the posterior distribution for (ϕ_t, x_t) based on $y_1, ..., y_t$ was performed through the following SIS algorithm:

- Draw $\tilde{x}_t^j$ from $N(\phi_{t-1}^j x_{t-1}^j, 1)$ for $j = 1, ..., M$
- Put $\widetilde{\phi}_t^j = \phi_{t-1}^j$ for $j = 1, ..., M$.
- Calculate the weights $w_t^j = p(y_t | x_t = \tilde{x}_t^j)$ for $j = 1, ..., M$ and the normalized weights $q_t^j = w_t^j / \sum_{j'} w_t^{j'}$.
- Draw $(x_t^1, \phi_t^1), ..., (x_t^M, \phi_t^M)$ from $\{(\tilde{x}_t^1, \tilde{\phi}_t^1), ..., (\tilde{x}_t^M, \widetilde{\phi}_t^M)\}$ with replacement and with probabilities $q_t^1, ..., q_t^M$.

The figure below shows simulations of ϕ_t for $t = 1, ..., T$ based on a SIS algorithm *with* resampling. Each curve corresponds to a sequence of simulated ϕ's, $\phi_1^j, ..., \phi_T^j$. The different simulations $\phi_t^j, j = 1, ..., M$ for a fixed t are (approximately) from the posterior distribution for ϕ_t given $y_1, ..., y_t$. Here $T = 30$ and $M = 50$.

(Continued on page 6.)



Why do the number of different values of the simulated ϕ's decrease with t? What kind of problems do this make in the estimation of ϕ?

(c) A more efficient algorithm can be obtained by integrating out the unknown ϕ when simulating the x-process.

One can show (you do not have to do this) that

$$p(\phi|x_1,...,x_t,y_1,...,y_t) = N(\hat{\phi}_t, \sigma_t^2)$$

where

$$\hat{\phi}_t = \frac{\sigma_\phi^2 \sum_{i=2}^t x_i x_{i-1}}{1+\sigma_\phi^2 \sum_{i=2}^t x_{i-1}^2}, \quad \sigma_t^2 = \frac{\sigma_\phi^2}{1+\sigma_\phi^2 \sum_{i=2}^t x_{i-1}^2}$$

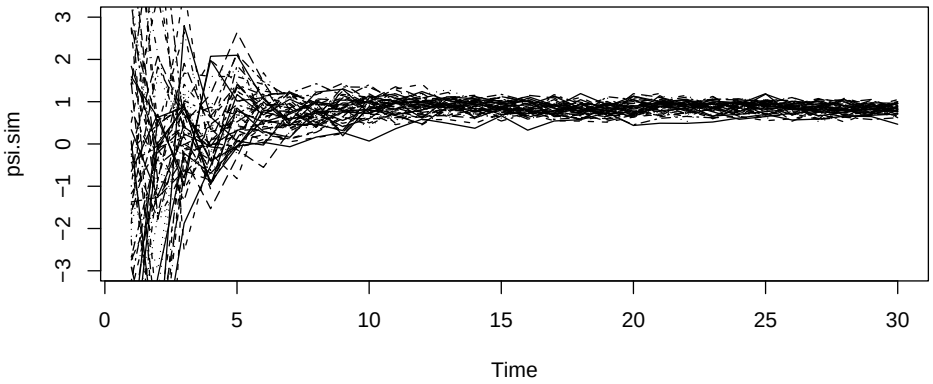
Use this to explain how you can simulate from the distribution $p(x_{t+1}|x_1,...,x_t,y_1,...,y_t)$.

(d) Consider now the SIS algorithm (with resampling) which at time t goes through the following steps:

- Draw $\tilde{x}_t^j$ from $p(x_t|x_1,...,x_{t-1},y_1,...,y_{t-1})$ for $j=1,...,M$.
- Calculate the weights $w_t^j = p(y_t|x_t=\tilde{x}_t^j)$ for $j=1,...,M$ and the normalized weights $q_t^j = w_t^j / \sum_{j'} w_t^{j'}$.
- Draw $x_t^1,...,x_t^M$ from $\{\tilde{x}_t^1,...,\tilde{x}_t^M\}$ with replacement and the probabilities $q_t^1,...,q_t^M$.
- Draw $\phi_t^j \sim p(\phi|x_1^j,...,x_t^j,y_1,.,y_t)$

The figure below shows simulations of ϕ_t based on this algorithm.

(Continued on page 7.)



Which advantages does this algorithm have compared to the one given in (*b*)?

In order to estimate the posterior expectation of ϕ, is it necessary to simulate the ϕ's at all? If not, explain how inferense on ϕ then can be performed. What is this technique called?

(Continued on page 8.)

END