Economet-II SS12

k. utikal

Final Exam (26.07.12)

Solutions:

1. using a uniform random number generator $U_1, \ldots, U_n$ iid $\sim \mathcal{U}(0,1)$ explain (algorithm) how to:

   a) generate random numbers from

   $$X_1, \ldots, X_n \text{ iid} \sim EXP(\lambda)$$

   with $\lambda = 2$

   answer: *probability transform ("inverse sampling")*

   $$x_i = -\frac{1}{2}\log(u_i)$$

   b) compute by MC simulation

   $$\int_0^\infty \sqrt{x}e^{-2x}dx$$

   answer: $\int_0^\infty \sqrt{x}e^{-2x}dx = \int_0^\infty \left(\frac{1}{2}\sqrt{x}\right)\ \left(2e^{-2x}\right)dx.$

   Algorithm 1 *generate* $X_1, \ldots, X_n$ *iid*$\sim EXP(\lambda)$
   compute $\frac{1}{n}\sum_1^n \frac{1}{2}\sqrt{x_i}$

   then $\frac{1}{n}\sum_1^n \frac{1}{2}\sqrt{x_i} \to \int_0^\infty \sqrt{x}e^{-2x}dx$ for n *large*
   more than one person tried to argue with MCMC using Metropolis-Hastings. this was misguided.

   c) give a plausible reason why the method described in **b)** should work.

   answer: *LLNs*

   d) compute by MC simulation

   $$\int\int_{x+y<1} e^{-\frac{1}{2}\left(x^2+y^2\right)}dx\ dy$$

answer: *generate* $(x_1, y_1), \ldots, (x_n, y_n)$ iid $\sim f(x,y)$ *by what ever method from independent normal distribution:*

$$f(x,y) = \frac{1}{2\pi} e^{-\frac{1}{2}\left(x^2+y^2\right)}$$

then

$$\frac{2\pi}{n} \sum_i^n I(x+y<1) \to 2\pi\ \mathbb{E}(X+Y<1)$$
$$= 2\pi\ \mathbb{P}(X+Y<1) = \int\int_{x+y<1} e^{-\frac{1}{2}\left(x^2+y^2\right)} dx\ dy$$

half of all found this very hard.

2. consider the model

$$x_i = \theta_i + \epsilon_i$$
$$\theta_i \text{ iid } \sim N\left(\mu, \sigma^2\right),\ \epsilon_i \text{ iid } \sim N\left(0, \sigma_x^2\right)$$
$$\theta_i \text{ indep } \epsilon_j,\ i,j = 1, \ldots n$$

answer: *this is just a special case of*

$$x_{ij} = \theta_i + \epsilon_{ik}$$
$$\theta_i \text{ iid } \sim N\left(\mu, \sigma^2\right),\ \epsilon_{ik} \text{ iid } \sim N\left(0, \sigma_x^2\right)$$
$$\theta_i \text{ indep } \epsilon_{ik},\ i,k = 1, \ldots n$$

a) sketch a DAG (directed acyclic graph) for Bayesian inference on $\mu,\ \sigma^2$ from observations $x_1, \ldots, x_n$. with $\theta_1, \ldots, \theta_n$, σ_x^2 unknown. (you may assume independent priors, do not specify them, do not explain the distribution on the knots of the graph).

b) sketch the collapsed graph for Bayesian inference on $\mu,\ \sigma^2$ if $\theta_1, \ldots, \theta_n$ are observed.

c) describe the Gibbs sampler for drawing random numbers from the posterior distribution of (μ, σ^2) for problem **b)** (you do not need to compute explicitly the distribution on the knots of the graph).

d) continuation of **c)**: expert opinion: "μ around 50, certainly positive, but not as large as 100". What prior on $\mu,\ \sigma^2$ do you propose (see **e)**)?

answer:

$$\mu \sim \mathcal{N}(50, 625)$$

will cover 95% of the interval, do not worry about the remaining 5% will give an easy answer. second best is

$$\mu \sim \mathcal{U}(0, 100)$$

because it ignores the vital information that the center is around 50. for the variance σ^2 of $\theta_1, \ldots, \theta_n$ *use a vage prior:*

$$f\left(\sigma^2\right) \sim \frac{1}{\sigma^2}$$

or $f\left(\sigma^2\right) \sim Gamma\left(\alpha, \beta\right)$ with $\alpha = \beta = 10E-6$ *since we don't have any expert opinion here.*

e) compute the full conditionals required for Gibbs sampling explicitly, using the prior in **d)**

answer: *standard formulas of the mean/variance of the posterior distribution of the mean/variance as given in the formula sheet.*

3. time changing voting patterns: define p_t : population % of "yes" voters for a certain candidate at time t.

   model: $p_t = p^{(1)}$ for $t = t_1, \ldots, t_\kappa$, $p_t = p^{(2)}$ for $t = t_{\kappa+1}, \ldots, t_n$

   data: draw independently a sample voters each time and ask their preferences

time	t_1	$\cdots$	t_n
"yes" counts	x_1	$\cdots$	x_n
sample size	m	$\cdots$	m

   problem: Bayesian inference on $\left(p^{(1)}, p^{(2)}, \kappa\right)$

   a) find the likelihood function

   b) define priors on the unknown parameters of interest

   c) find posterior distribution

   compute posterior by Gibbs sampling:

   d) identify full conditionals

   e) write down the Gibbs sampler (algorithm) explicitly

   answer: *erverybody got this right!*

4. Given a sample $(x_1, y_1), \ldots, (x_{15}, y_{15})$ iid $\sim F$, where F is a distribution with positive support: $F(x, y) > 0 \iff x > 0$ and $y > 0$. Define the parameter of interest $\eta = \mathbb{E}\frac{Y}{X}$.

   Using the bootstrap method:

   a) write an algorithm to construct a 90% percentile bootstrapped conf int for η

   answer: *everybod knows how to do this one. though some people didn't and gave the basic CI instead. we know, e.g. from our simulations, that both intervals are different.*

   b) give plausible reasons why this method described in **a)** should work

   answer: *it all derives from convergence of the ecdf of the sample to the population ecdf. nobody said that. other reason were accepted.*

   c) write an algorithm to construct a 90% bootstrap-t-interval

   answer: *this was pretty standard. some people complicated things by bootstrapping the bootstrap to obtain the variance. this is only second best. easier and more accurate is to use the undergraduate formula for variance estimation.*

   d) describe a situation when interval **c)** could be problematic while **a)** remains valid.

   answer: *nobody gave a satisfactory answer to this. the purpose was not to come up with examples from other bootstrap problems. i just wanted to read that the bootstrap-t can overshoot the limit, like in the case of the correlation estmate. what if η is close to zero. what if your lower limit reaches into the negative? that would be pretty embarrassing.*