Bayesian Inference and Computational Methods   2009

Homework 3

1. Use **R** to generate a (pseudo-)random sample of size 10000 from the $U(0,1)$ distribution (i.e. a realisation of 10000 independent $U(0,1)$ random variables):

   ```
   z = runif(10000)
   ```

   Use graphical techniques to investigate the uniformity and independence of the sample. For example, to check uniformity:

   ```
   plot(sort(z))
   ```

   while independence could be (partially) investigated by plotting each realisation against the previous one.

   Repeat with a further sample of the same size.

2. Let Y_1 and Y_2 be *independent identically distributed* random variables. Use simulation to investigate the distributions of both $\max(Y_1, Y_2)$ and $\min(Y_1, Y_2)$ in each of the following cases:

   (a) each $Y_i \sim U(0,1)$,

   (b) each $Y_i \sim \text{Exp}(1)$,

   (c) each $Y_i \sim N(0,1)$

   Where possible compare the empirical distributions with the theoretical ones, using appropriate graphical techniques and comparing also sample means. *Hint:* appropriate **R** code for generating samples of size 10000 and plotting histograms and empirical distribution functions in case (a) is

   ```
   y1 = runif(1000)
   y2 = runif(1000)

   z1 = pmax(y1, y2)
   hist(z1)
   plot(sort(z1), ppoints(z1))

   z2 = pmin(y1, y2)
   hist(z2)
   plot(sort(z2), ppoints(z2))
   ```

   However, in case (b) an appropriate probability plot will be more satisfactory for investigating the distribution the minimum.

3. Use the *inverse transform* method to generate a random sample of size 10000 from the *logistic* distribution with distribution function $F(y) = [1 + \exp(-y)]^{-1}$. Display the distribution graphically, and estimate both its mean and variance, comparing the former with the true value.

4. Use the (discrete) *inverse transform* method to generate a random sample of size 10000 assigning probability 1/6 to each of the numbers 1 to 6. Display the distribution of the sample graphically and use a chi-squared test to investigate its acceptability. *Hint:* appropriate code (which you should make sure you understand) to generate the sample is:

   ```
   u = runif(10000)
   sample = 1 + (u>1/6) + (u>2/6) + (u>3/6) + (u>4/6) + (u>5/6)
   ```

Obtain two further samples each of the same size from this distribution, without overwriting the first, and hence obtain a random sample of the total shown by three dice. Compare with the theoretical distribution.

5. Use both the *inverse transform* method and *rejection sampling* to generate a large random sample (say of size 10000) from each of the *continuous* distributions below—where in each case the probability density function f is given. In each case display the distribution of the sample graphically and satisfy yourself that it is indeed compatible with the given density function f. Also, in the case of the rejection method, determine the theoretical overall acceptance probability and verify the compatibility of the final sample size with this.

   (a) $f(y) = e^y/(e-1), \quad 0 \leq y \leq 1;$
   (b) $f(y) = 3(1-y)^2, \quad 0 \leq y \leq 1.$

   Hints: In case (a) appropriate code to generate a sample by *rejection sampling* (using the $\mathrm{U}(0,1)$ distribution as the *envelope*) is:

```
u1 = runif(10000)
acc = runif(10000) < exp(u1)/exp(1)      #acceptance decisions
sample = u1[acc]                         #sample with reqd dist
```

   You may (or may not!) also find it helpful to recall that in the case of *rejection sampling*, the function f may be redefined at the outset so as to omit any normalising constant.

6. Suppose that a random sample is required from the $\Gamma(2,1)$ distribution, i.e. from the distribution with density

$$f(y) = ye^{-y}, \quad y \geq 0.$$

   Explain why it is difficult to use the *inverse transform* method for this task. Use *rejection sampling* to generate such a sample by taking as *envelope* the $\mathrm{Exp}(1/2)$ distribution. Display the distribution of the $\Gamma(2,1)$ sample graphically.

   Generate a further sample by using instead the result that a gamma random variable with integer shape parameter can be regarded as a sum of independent exponential random variables.

7. Use the *composition method* to generate a random sample of size 10000 from the distribution of a random variable which with probability 3/4 has an $\mathrm{Exp}(1)$ distribution and with probability 1/4 has an $\mathrm{Exp}(1/10)$ (mean 10) distribution. Verify that the sample generated is as you would expect.

8. Use simulation (using, e.g., the **R** function `rpois`) to verify experimentally that the sum of *independent* Pois(4) and Pois(6) random variables is a Pois(10) random variable.

9. Use the algorithm of the lectures to generate a random sample of size 10000 from the Geo(1/3) distribution. Display the distribution of the sample graphically, and verify that its mean does not differ significantly from its theoretical value of 3. Calculate also the standard deviation of the sample and compare with the true value.

   Generate a similar example directly using the `rgeom` function.

10. Investigate the central limit theorem by simulating and adding independent random variables uniformly distributed on $(0,1)$.

11. Use **R** to simulate a single realisation of an inhomogeneous Poisson process of rate $1+\sin(2\pi t/100)$ over the interval $[0,100]$. Display a plot of $N(t)$ against t, where $N(t)$ is the number of events which have occurred by time t.