STAT 2011 Probability and Estimation Theory – Semester 1, 2024

Computer Practical Sheet Week 8

Computer Problems for Week 8

For this week's lab report: Please submit your code, output, and any comment required, for Q3 *ONLY*: the pdf or html file must include code, output and any comment, otherwise you will be penalised. An assignment item has been set up on Canvas; file upload format is limited to pdf and html.

1. Reconsider the Week 8 tutorial problem 9: Let X and Y have the joint pdf $f_{X,Y}(x,y) = 2e^{-(x+y)}$, $0 < x < y$, $0 < y$. Confirm $P(Y < 3X) = 1/2$ using *Monte Carlo Integration* with the number of uniformly distributed points J ranging over 100, 1000 and 10000 and restricting Y to be less than 10. The main challenges will be in finding out how to generate uniform (pseudo)random numbers over the (restricted) region $R_{y<10} = \{(x,y) : 0 < y < 3x < 10\}$, how to calculate $V(R)$ and ensuring that $f_{X,Y}(x,y) = 0$ whenever $0 < x < y$ and $0 < y$ does not hold.

   (Hint: Generate uniform (pseudo) random numbers (x,y) from $[0,c_1] \times [0,c_2]$ and then only use those that satisfy $y < 3x$ and $x < y$. Check with `plot(x,y)` if the sampled points look uniform over the desired region.)

2. We will visualise the joint cdf and pdf for two continuous random variables X and Y. The joint cdf is defined as $F_{X,Y}(x,y) = \frac{1}{3}x^2(2y+y^2)$ for $0 \le x,y \le 1$.

   (a) Generate a grid of x,y values on the interval $[0,1]$:

   ```
   x = seq(0, 1, length= 30)
   y = x
   ```

   (b) Define a function `f` equal to the joint cdf, and plot the surface.

   ```
   f = function(x, y) { z = x^2*(2*y+y^2)/3 }
   z = outer(x, y, f)
   persp(x,y,z,theta=45,col = "lightblue")
   ```

   (c) Repeat the above for the joint pdf, $f_{X,Y}(x,y) = \frac{4}{3}x(1+y)$, $0 \le x,y \le 1$.

3. Consider two RVs, $X \sim Bin(m, p)$, $Y \sim Bin(n, p)$, with $m = 10$, $n = 20$, $p = 1/3$. We investigate the distribution of the sum $W = X + Y$, in particular how close the empirical probability distribution (obtained via simulation) comes to the theoretical probability distribution $Bin(n + m, p)$.

   (a) Set `B=100`. Generate a random sample of size `B` from $Bin(m, p)$ and store the result in `X`. Similarly, generate a random sample of size `B` from $Bin(n, p)$ and store the result in `Y`. Compute W and tabulate the values using `table(W)`.

   (b) Convert the column names of the table to numeric, store in `W.obs`, and compute the empirical frequencies of different values of W, store in `P.obs`:

   ```
   Tab=table(W)
   W.obs = as.numeric(names(Tab))
   P.obs = as.numeric(Tab/B)
   ```

   (c) Plot the empirical frequencies and the theoretical probabilities from $Bin(n + m, p)$ (obtained using `dbinom(0:(m+n),m+n,p=1/3)`), and comment on how close they are.

   (d) Repeat for $B = 10,000$. Comment.