Statistics - Homework 4 (Due Nov. 29, 2017)

1. The exponential density $g(x) = \lambda e^{-\lambda x}$ with $\lambda = 1$ is everywhere higher than the standard normal density $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ for $0 \le x < +\infty$.

   (1) Prove and show that the rejection method with $cg(x) = \sqrt{\frac{2e}{\pi}} e^{-x}$ can be used to generate random samples of the standard normal distribution.

   (2) Develop an R code to generate 10,000 random numbers from the standard normal density $(z_1,\ z_2, \cdots,\ z_n;\ n = 10{,}000)$ using the above rejection method. Calculate the mean and standard deviation of $(z_1,\ z_2, \cdots,\ z_n;\ n = 10{,}000)$. Also, show the histogram of $(z_1,\ z_2, \cdots,\ z_n;\ n = 10{,}000)$.

   > Hint:
   >
   > (1) Generate $X \sim Exp(1)$ and $Y \sim U\left(0, \sqrt{\frac{2e}{\pi}} e^{-x}\right)$.
   >
   > (2) Reject *X*, if $Y > \phi_+(X)$, where $\phi_+(x) = \sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}x^2}$
   >
   > (3) If $Y \le \phi_+(X)$, generate $S \sim U[-1,\ 1]$. Then, $Z = SX$.

   (3) Let *N* be the total number of random numbers of the exponential density which were generated to achieve $(z_1,\ z_2, \cdots,\ z_n;\ n = 10{,}000)$. What is the value of n/N in your simulation? Can you give a theoretical explanation of this ratio?

2. *Monte Carlo integration*. Integration of a function *f*(*x*) over a range, say [a, b], can be approximated by the following method.

   $\int_a^b f(x) g_X(x) dx = E[f(X)]$, where $g_X(x)$ is the probability density function of X.

   Let $g_X(x) = \frac{1}{(b-a)} I_{[a,b]}(x)$, it yields $\frac{1}{b-a} \int_a^b f(x) dx = E[f(X)]$ and thus $\int_a^b f(x) dx = (b-a) E[f(X)]$.

   (1) Write an R code to estimate $\Phi(0.42)$ and $\Phi(-0.38)$ by using the above method. [$\Phi(x)$ is the cumulative distribution function of the standard normal density.]

   (2) Compare the your estimates in (1) against the values of $\Phi(0.42)$ and $\Phi(-0.38)$ based on *pnorm* in R.

3. The bivariate normal density (joint density of two normally distributed random variables) can be expressed as

$$f_{XY}(x, y) = \frac{1}{2\pi\sqrt{(1-\rho^2)}\sigma_X \sigma_Y} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\cdot\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}.$$

   Thus, $f_{XY}(x, y) = C$ (a constant) is equivalent to

$$\frac{1}{(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\cdot\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right] = C^* \text{ (a constant).}$$

   The above equation represents a rotated ellipse centered at $(\mu_X,\ \mu_Y)$. Let $\mu_X = 60,\ \mu_Y = 100,\ \sigma_X = 25,$ $\sigma_Y = 40,$ and $\rho = 0.75$.

   (1) Generate 1000 random number pairs of the above bivariate normal distribution by using the conditional density method.

   (2) Given the following values of *C**, plot the corresponding rotated ellipses and the data pairs generated in (1).

0.2107	0.4463	0.7134	1.0217	1.3863	1.8326	2.4079	3.2189	4.6052

(3) Calculate the number of data points falling inside each individual ellipse.

4. Two insect species (predator and prey) coexist in a study area. The numbers of insects of the two species found in the study area can be described by a bivariate normal distribution. Let *X* and *Y* represent the numbers of the predator species and the prey species, respectively. A long term ecological survey found that

   $P(Y > 86 \mid X = 80) = 0.04779,\ \ P(Y < 50 \mid X = 80) = 0.09121$

   $P(Y > 55 \mid X = 150) = 0.20233,\ \ P(Y < 30 \mid X = 150) = 0.10565$

   $\rho_{XY} = -0.6$ (correlation coefficient of *X* and *Y*).

   (1) Find the standard deviations of *X* and *Y*, respectively.
   (2) What should the maximum number of the predator species be in order to achieve a higher than 75% chance for the number of prey species to be more than 70?

5. A city experiences 3.26 typhoons on average annually. The event-total rainfalls drawn by typhoons can be characterized by a gamma distribution with expected value of 210 mm and standard deviation of 190 mm.

   (1) Let *N* be the annual count of typhoons occurred at the city and *X* be the event-total rainfalls drawn by a typhoon. Prove that $E\left[\sum_{i=1}^{N} X_i\right] = E(N)E(X)$. [Hint: Use the property $E(Y) = E_X(E(Y \mid X))$]
   (2) What is the expected value of the annual-total typhoon rainfalls at the city?
   (3) Assuming the annual count of typhoons can be characterized by a Poisson distribution. Develop an R code to generate a random sample of size 1000 of the annual-total typhoon rainfalls for the city.
   (4) Plot the histogram of the random sample generated in (3).
   (5) Calculate the sample mean of the annual-total typhoon rainfalls and compare it against the result in (2).