GradSTAT 2019-20: Computational Statistics
Instructor: P.T. Besbeas
E-mail: besbeasaueb.gr

Assessment: Paper 1

Due: 5:00pm Tuesday 4 February 2020

1. Develop two functions, one that calculates s^2 using the formation definition, and one that calculates s^2 based on the hand calculation equation. Generate data from a normal distribution with a large mean, eg $\mu = 1000$. Use your functions to calculate s^2 for your generated data. Try this for various variances so that your data become closer and closer to the mean (less variance). What do you find, comment.

2. The following data are an i.i.d. sample from a Cauchy(θ, 1) distribution: 1.77, -0.23, 2.76, 3.80, 3.47, 56.75, -1.34, 4.24, -2.44, 3.29, 3.71, -2.40, 4.53, -0.07, -1.05, -13.87, -2.53, -1.75, 0.27, 43.21.

   Plot the log likelihood function. Find the MLE for θ using the Newton-Raphson method. Try all of the following starting points: -11, -1, 0, 1.5, 4, 4.7, 7, 8, and 38. Discuss your results. Is the mean of the data a good starting point?

GradSTAT 2019-20: Computational Statistics
Instructor: P.T. Besbeas
E-mail: besbeasaueb.gr

Final Assessment: Paper 2

Due: 5:00pm Tuesday 4 February 2020

A random variable Y has a generalised gamma distribution if its density function is given by

$$f(y;\alpha,\kappa,\lambda) = \frac{\alpha\lambda^{\kappa}y^{\alpha\kappa-1}}{\Gamma(\kappa)}\exp\left(-\lambda y^{\alpha}\right), \quad y>0, \quad \alpha,\kappa,\lambda>0. \tag{1}$$

This arises on supposing that for some α, Y^{α} has a gamma distribution, and reduces to the gamma density when $\alpha=1$, to the Weibull density when $\kappa=1$, and to the exponential density when $\alpha=\kappa=1$.

The data below correspond to times (hours) spent by women giving birth in the delivery suite at the John Radcliffe Hospital, Oxford, during the first seven of 92 successive days of observation. The researchers analysing the data propose to model the times using a generalised gamma distribution.

	Day						
Woman	1	2	3	4	5	6	7
1	2.10	4.00	2.60	1.50	2.50	4.00	2.00
2	3.40	4.10	3.60	4.70	2.50	4.00	2.70
3	4.25	5.00	3.60	4.70	3.40	5.25	2.75
4	5.60	5.50	6.40	7.20	4.20	6.10	3.40
5	6.40	5.70	6.80	7.25	5.90	6.50	4.20
6	7.30	6.50	7.50	8.10	6.25	6.90	4.30
7	8.50	7.25	7.50	8.50	7.30	7.00	4.90
8	8.75	7.30	8.25	9.20	7.50	8.45	6.25
9	8.90	7.50	8.50	9.50	7.80	9.25	7.00
10	9.50	8.20	10.40	10.70	8.30	10.10	9.00
11	9.75	8.50	10.75	11.50	8.30	10.20	9.25
12	10.00	9.75	14.25	10.25	12.75	10.70	
13	10.40	11.00	14.50	12.90	14.60		
14	10.40	11.20	14.30				
15	16.00	15.00					
16	19.00	16.50					

1. Calculate maximum-likelihood estimates of λ,α,κ, and give approximate standard errors for your estimates. Hence or otherwise provide a 95% confidence interval (CI) for α. [You may find it easier to perform the optimisation in terms of the unconstrained parameters $(\theta_1=\log\lambda, \theta_2=\log\alpha, \theta_3=\log\kappa)$.]

2. Using the parametric bootstrap, provide a 95% percentile bootstrap CI for α (or θ_2) based on 800 simulations, for comparison with the asymptotic CI above. Hence explain whether the researchers proposal to use a generalised gamma distribution appears justified.

3. The researchers become interested in the possibility of modelling the data with an exponential distribution. Calculate the likelihood ratio statistic for comparison of the exponential and generalised gamma densities, and assess the significance of your test statistic using both asymptotic theory and Monte Carlo simulation.

4. For planning purposes, the hospital administrator is interested in estimating the probability that a delivery time will not exceed 10 hours. Estimate this probability using the model supported by your findings above and provide a standard error for your estimate based on simulation.

GradSTAT 2019-20: Computational Statistics
Instructor: P.T. Besbeas
E-mail: besbeasaueb.gr

Assessment: Paper 3

Due: 5:00pm Tuesday 4 February 2020

The random variables X, Y have joint probability density function (pdf)

$$f(x, y) = 24y(1 - x - y), \quad x > 0,\ y > 0,\ x + y < 1.$$

1. The marginal pdf of X can be shown to be

$$f(x) = 4(1 - x)^3, \quad 0 < x < 1.$$

   Implement a function to simulate from $f(x)$ in R, using `runif` as a source of uniform random variables.

2. The conditional pdf of Y given $X = x$ is given by

$$f(y \mid x) = \frac{6y(1 - x - y)}{(1 - x)^3}, \quad 0 < y < 1 - x.$$

   Implement a function to simulate from $f(y \mid x)$, for a given $X = x$, using `runif` as a source of uniform random variables.

3. Hence, or otherwise, explain how you could simulate from $f(x, y)$. Implement your method in R, and use your program to simulate a sample (X_i, Y_i), $i = 1, \ldots 1000$. Compute the sample correlation between X and Y, and compare it with the theoretical value $\rho = -\sqrt{6}/6$. [You are not asked to prove this result.]