Prof. Donna Ankerst, Katharina Selig

July 14th & 15th 2019

Problem 1

The natural logarithm and exponential functions are inverses of each other, so that mathematically

$$\log(\exp(x)) = \exp(\log(x)) = x.$$

`In R:`

Show by example that this property does not hold exactly in computer arithmetic. Does the identity hold with near equality (see `all.equal`)?

Problem 2

The density of the Cauchy(η, θ) distribution is given by

$$\frac{1}{\theta\pi\left(1 + [(x-\eta)/\theta]^2\right)},$$

with $\theta > 0$ and $\eta, x \in \mathbb{R}$. η is denoted the location parameter and θ the scale parameter.

`In R:`

Write a function to compute the cdf of the Cauchy$(0, \theta)$ distribution for $\theta = 1, 10$, and 100. Use numerical integration and not MC integration. Try various options of the integrate function and compare whole plots of cdfs against `pcauchy`. Explain briefly what the different options in integrate are as well as what the documentation states about the calculation of `pcauchy`.

Problem 3

Let $x_1, x_2, \ldots, x_n$ be a random sample from a Gamma(α, β) distribution with density function

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0,\ \alpha, \beta > 0.$$

The goal of this exercise is to find the maximum likelihood estimator (MLE) of $\theta = (\alpha, \beta)$ by reducing the two-dimensional optimization problem to a one-dimensional root-finding problem.

(a) Write down the likelihood and the log-likelihood function.

(b) Determine the partial derivatives of the log-likelihood function with respect to α and β and set them to 0.

(c) Solve the equation for β and substitute the result into the equation for α.

(d) Write `R` Code to calculate the MLE estimates for a given $\bar{x}$ (`xbar`), $\overline{\log(x)}$ (`log.xbar`) using the `R` function `uniroot`. The function $\psi(t) = \frac{d}{dt} \log \Gamma(t) = \Gamma'(t)/\Gamma(t)$ is implemented in the `digamma`-function in `R`.

`In R:`

(e) Repeat 2000 times: Generate 200 random samples from a Gamma$(5, 2)$-distribution. Estimate the parameters α and β as in (d). Give the average estimates of the parameters α and β along with the standard deviations.

Problem 4

The Poisson distribution is often used to model counts $X = 0, 1, 2, \ldots$. The formula for a K-component mixture of Poisson distributions with rate parameters $\lambda_1, \ldots, \lambda_K$ and mixing probabilities $p_1, \ldots, p_K$, with $\sum_{j=1}^{K} p_j = 1$ is given by

$$f^*(X) = \sum_{j=1}^{K} p_j \frac{\lambda_j^X e^{-\lambda_j}}{X!}.$$

(a) Given data $X_1, \ldots, X_n$ from this distribution, write down the procedure of an EM algorithm to estimate the p's and λ's.

(b) Write out the steps for an EM algorithm to estimate the mixing probability p and the rate parameters λ_1 and λ_2 for a 2-component mixture of Poisson distributions.

`In R:`

(c) Simulate 600 observations from a 2-component mixture of a Poisson(2) and Poisson(6) distribution with mixing probabilities $p_1 = p = 0.65$ and $p_2 = 1 - p = 0.35$, respectively. Create a barplot of the sample and overlay the true probabilities.

(d) Implement the EM algorithm of (b) to estimate the mixing probability p and the rate parameters λ_1 and λ_2. Use the starting values $p = 0.4$, $\lambda_1 = 1$, $\lambda_2 = 7$ and 200 iterations.

(e) Overlay the estimated and the true probabilities on the barplot from (c).