Zuse Institute Berlin (ZIB)
Freie Universität Berlin
Dr. Vikram Sunkara, Dr. Max von Kleist

12. Homework
Numerics for Bioinformaticians
WS 2016/17

Deadline: February 01, 10:00 (**before** the lecture)

The homework should be worked out in groups of two or three students. Each solution sheet must contain the names and student numbers ('Matrikulationnummer') of all group members and the exercise group (Wednesday/Friday). Solutions to exercise 1.b & 3 must be handed in in paper form, either hand written or printed out if generated electronically. Please staple all sheets. Programming tasks must be submitted to max.kleist2@fu-berlin.de or BioInfNumerik@hotmail.com by email. Before sending it, please 'zip' it.

Exercise 1 (unnormalized posterior I, 2 points)
Consider the structural model $\mathcal{M}$:

$$x(t_i) = \frac{\lambda}{\delta}\left(1 - e^{-\delta t_i}\right).$$

The likelihood of observing y_i at t_i for a parameter choice $\lambda, \delta > 0$, is given by,

$$\mathcal{L}_y = \prod_i \frac{x(t_i)^{y_i} e^{-x(t_i)}}{y_i!}.$$

Given the measurements are $y = \{108, 108, 101, 108, 109, 91, 108, 97, 92, 98\}$, all taken at the same time $t = 100$.
Consider the following *gaussian* priors (assuming the two parameters are uncorrelated)

$$p(\theta_j) = \frac{1}{\sqrt{(2\sigma_j^2 \cdot \pi)}} e^{-\frac{(\theta_j - \mu_j)^2}{2 \cdot \sigma_j^2}}$$

with $\mu_1 = 10$ and $\sigma_1^2 = 100$ for $\theta_1 = \lambda$ and $\mu_2 = 1$ and $\sigma_2^2 = 10$ for $\theta_2 = \delta$.

a) Compute the unnormalized posterior $p(\lambda) \cdot p(\delta) \cdot \mathcal{L}_y(\theta) = p(\theta) \cdot \mathcal{L}_y(\theta)$ and make a contour plot of it, for the given data, with axis $\lambda = [0.1, 30]$ and $\delta = [0.01, 0.3]$.

Exercise 2 (unnormalized posterior II, 3 Points)
You are given the following model:

$$x_{i|k} = x_0 + k \cdot t_i$$

with additive error

$$y_i = x_i + \eta_i \ ; \ \eta_i \sim \mathcal{N}(0, \sigma^2)$$

with $\sigma^2 = 5$ and $x_0 = 5$. The data is $t = (0, 1, 2)$ and $y = (5, 7.1, 8.8)$ and the prior is a log-normal distribution with mean $m = 5$ and variance $v = 5$ (the mean and standart deviation of the associated normal distribution are $\mu = 1.43$ and $\sigma = 0.43$).
1) Sample 1000 parameters k from the lognormal distribution above and compute the likelihood of the data.
2) Compute the unnormalized posterior

$$p(k) \cdot \mathcal{L}_y(k)$$

where

$$p(k) = \frac{1}{k \cdot \sigma \sqrt{(2 \cdot \pi)}} e^{-\frac{(\log(k)-\mu)^2}{2 \cdot \sigma^2}}$$

3) Plot the lognormal distribution along with the sample frequencies, plot the Likelihood for each sampled k and plot the posterior.

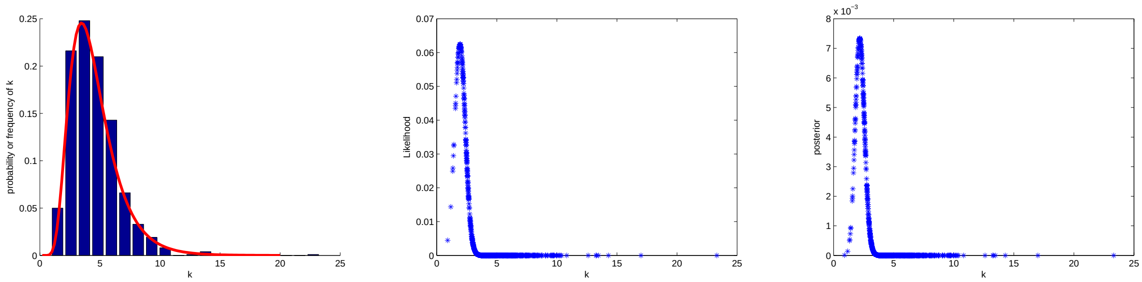

Figure 1: Exemplary graphics from exercise 2.

Exercise 3 (Algorithm, 5 Points)
1) Implement the Gauss-Newton method to estimate the parameters k_1 and k_2 of the model

$$x_i = k_1 \cdot e^{k_2 \cdot t_i}$$

with $t = (1, 2.5, 5, 7.5, 10)^T$ and $y = (2.98, 4.41, 8.44, 16.17, 30.97)^T$, assuming an additive *gaussian* measurement error $\eta_i \sim \mathcal{N}(0, \sigma_i^2)$ with $\sigma^2 = (2.27, 2.26, 1.14, 1.70, 0.23)$. Stop the algorithm, if a maximum of 30 updates have been exceeded, or the parameters change by less than $\epsilon = 10^{-8}$, i.e. $||\theta_j^{s+1} - \theta_j^s||_1 = \sum_{j=1}^m |\theta_j^{s+1} - \theta_j^s| < \epsilon$. Let your initial parameter guess be $\theta_0 = (9.38, 0.96)^T$.
2) Plot the data and the prediction with the optimal parameter set θ^* determined in the Gauss-Newton method the into the same plot.

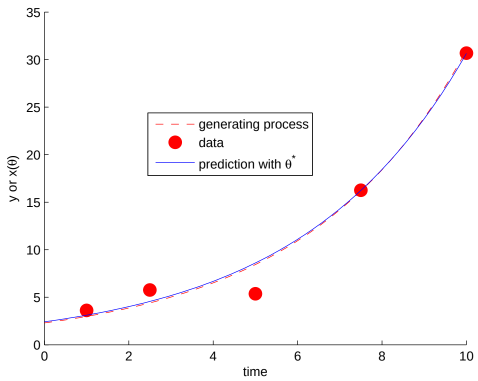

Figure 2: Exemplary graphics from exercise 3.