

STAT 513/413, ASSIGNMENT 5

1. Widely used in reliability and survival analysis is the Weibull distribution with the density

$$f(y; \lambda, \kappa) = \begin{cases} \lambda^{-\kappa} \kappa x^{\kappa-1} e^{-(x/\lambda)^\kappa} & \text{for } y \geq 0, \\ 0 & \text{for } y < 0, \end{cases}$$

Suppose that the data y_1, y_2, \dots, y_n are modeled as the outcomes of independent random variables Y_1, Y_2, \dots, Y_n , each with the Weibull distribution as above.

(a) (Theoretical.) Derive the equations that the maximum likelihood estimates of the parameters λ and κ have to satisfy. Solve as much as you can of those in closed form.

(b) (Computational.) For the rest, design and implement a numerical procedure to finish the problem; in order to learn something by doing that, do it from the first principles: hence, please, no packages or downloaded code, no functions `optim()`, `optimize()`, `nls()`, and `uniroot()`. Test it on (at least) two examples: in each of these choose n (preferably not very small one), choose λ and κ , generate the data (you may use the R function `rweibull()` for this), compute the estimates, assess the convergence, and compare to the values you started with. Report the results.

2A. (Only for STAT 413) Construct test example for the *simultaneous* (that is, both in μ and σ) location-scale model with Cauchy standard distribution: select μ and σ , and then generate 30 random numbers using `rcauchy()` (and store them).

(a) Write R functions implementing the negative loglikelihood and its gradient, for your particular generated sample (you can transfer those then into the final function conveniently given the R scoping rules). These functions are both functions of two variables (μ and σ); the first returns one number, the second a vector of two. Do `persp()` and `contour()` plots of your negative loglikelihood functions for judiciously selected ranges for μ and σ (that is, showing the interesting part of the function in the illuminative way).

Using R function `optim()` now,

(b) Find the solutions for μ and σ using the Nelder-Mead method. Plot the result into the `contourplot`.

(c) Find the solutions for μ and σ using the BFGS method; you *must* supply the gradient here (if you forget, the method will run anyway, but will be less precise; and your solution will be deemed not satisfactory). Plot the result into the `contourplot`; comment on all results and possibly other aspects.

2B. (For STAT 513; STAT 413 may choose to do this one instead of 2A, the previous one; STAT 513 have no choice.) Suppose that the data y_1, y_2, \dots, y_n can be modeled as the outcomes of independent random variables Y_1, Y_2, \dots, Y_n , which have all the same distribution – a mixture of two normal distributions with density

$$f(y_i; p, \mu_1, \mu_2, \sigma_1, \sigma_2) = \frac{p}{\sigma_1 \sqrt{2\pi}} e^{-\frac{(y_i - \mu_1)^2}{2\sigma_1^2}} + \frac{1-p}{\sigma_2 \sqrt{2\pi}} e^{-\frac{(y_i - \mu_2)^2}{2\sigma_2^2}}.$$

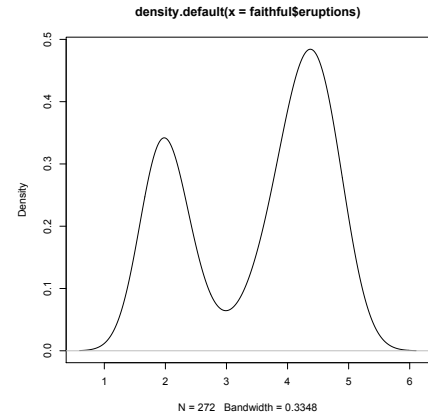
The distribution, for the data that show length of eruptions of the geyser Old Faithful in the Yellowstone park (R dataset `faithful`), can be interpreted as switching

between two possible regimes: with probabilities p and $1 - p$, the length of the eruption follows respectively the normal distribution with μ_1 and σ_1 , and that with μ_2 and σ_2 . The plot of the kernel estimate of the probability density, invoked by the R command

```
> plot(density(faithful$eruptions))
```

can be seen on the right.

We want to estimate all five unknown parameters involved: $\mu_1, \mu_2, \sigma_1, \sigma_2$, and p . To this end, we consider a hypothetical situation that apart from y_1, y_2, \dots, y_n , we would also possess certain additional data z_1, z_2, \dots, z_n which would specify whether y_1 follow from the first ($z_i = 1$) or the second regimen ($1 - z_i = 1$). These additional data would be supposed to be the outcomes of 0–1 random variables Z_1, Z_2, \dots, Z_n .



(a) (Theoretical.) Assuming the hypothetical situation, write, or given y_i and z_i , the negative log-likelihood of the parameters $\mu_1, \mu_2, \sigma_1, \sigma_2$ and p , and derive the maximum likelihood estimates of these parameters.

(b) (Theoretical.) Repeat (a) and z_i by corresponding random variables now. For fixed $\mu_1, \mu_2, \sigma_1, \sigma_2$ and p , what is the probability of $Z_i = 1$ given $Y_i = y_i$? Note that is is also the conditional expected value of Z_i under the same condition, as Z_i is an indicator variable. Use this to derive the conditional expected value of the negative loglikelihood you derived in (a).

(c) (Computational.) Piecing together (a) and (b), design and implement an EM algorithm to find the parameters of the the mixture model, on the basis of the data y_1, y_2, \dots, y_n only (and *not* z_1, z_2, \dots, z_n now, of course – as those pertain to the situation that was purely hypothetical). Apply this algorithm to the eruptions variable of the R dataset faithful and plot the resulting estimated density into the picture shown above.

As indicated in Lecture 22, page 410 of the Rizzo textbook contains a formula which looks almost like a solution of (b). While this formula is correct for what is done with it in the textbook, beware: it is NOT the solution of our problem – although its form is very similar.

4. The dataset cars records braking distances, dist, depending on the initial velocities, speed. (If you look more closely at the units, you will find that the dataset is pretty ancient.)

(a) Design and implement function for computing the parameters of *quadratic* relationship ($y = a + bx + cx^2$) between a quantile (for given p) of braking distance and initial speed – via the IRLS method.

(b) Draw the picture of the datapoints and plot the fitted curves into it.

Do not use any package for this problem. (You may, however, use it for checking the correctness of the results, but this is not to be reported in the solution.)

Due Monday, April 12, 2021