
MLE from a Competing Risks Model of an Exponential Failure and a Lognormal Survival

Nikola Chochkov, MSc Statistics, Humboldt University Berlin

10 November 2010

1 Introduction.

1.1 Assumptions.

Two usage-measured lifetime random variables are discussed in a fixed-time life test. The items under study are considered to accumulate usage independently and with a different rate. The survival population is considered *Lognormally distributed* (with parameters μ and σ), while the failure population - *Exponential* (with parameter λ).

An estimation of the parameter vector $(\lambda, \mu, \sigma)'$ is sought using all usage data collected (from both survival and failure cases)

- $\eta \sim \text{Lognormal}$
- $\psi \sim \text{Exponential}$
- $f_\eta(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$ is the *probability density function* of the Lognormal distribution;
- $\bar{F}_\eta(x) = 1 - \frac{1}{\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$ is its *survival function*
- $f_\psi(x) = \lambda e^{-\lambda x}$ is the *pdf* of the Exponential distribution;
- $\bar{F}_\psi(x) = e^{-\lambda x}$ is its *survival function*;
- \mathbf{x} is the data vector
- $N_f = \{i : x_i \text{ is failed}\}$ and $n_f = \#N_f$ (i.e. number of failures)
- $N_s = \{i : x_i \text{ is survived}\}$ and $n_s = \#N_s$ (i.e. number of survivals)
- n is the total number of units under test ($n = n_f + n_s$)
- $\lambda, \mu, \sigma, x > 0$

1.2 Maximum Likelihood Method.

The Likelihood function is given by (since η and ψ are independent):

$$L(\lambda, \mu, \sigma | \mathbf{x}) = \prod_{i \in N_f} (f_\psi(x_i) \bar{F}_\eta(x_i)) \prod_{i \in N_s} (f_\eta(x_i) \bar{F}_\psi(x_i)) \quad (1)$$

MLE method applied in this case prescribes that $\log L^*$ be derived with respect to λ , μ and σ up to its second derivatives. The zeros $(\hat{\lambda}, \hat{\mu}, \hat{\sigma})$ of the first derivatives vector, aka *Gradient* will be the sought estimates if they turn the matrix of second derivatives, aka *Hessian* to a negative definite. The Hessian being negative definite will imply that $\log L^*$ is convex in some neighbourhood around the found extremum point, i.e. that point maximizes the Likelihood.

I derived the Gradient and Hessian and numerically optimized the function $\log L^*$ for λ , μ and σ using:

- *Newton-Raphson-like* iterative algorithm for numerical optimization. I used the routine *nlminb()* in R;
- *Gauss-Hermite* method for numerical integration;

This was repeated a number of times for different data set, under the above assumptions and the average of the estimates was taken.

1.3 Correction of Bias.

Such an estimation is the ultimate goal of the current research. However, up to now the results, to which I arrived are biased with the cause or trend of this bias being yet unconfirmed. Attempting to justify or remove the bias, and following Professors' comments, I adjusted the estimates by ignoring data sets, which produced absurd results in:

- *Theoretical Variance/Covariance Matrix* - this is the Inverse of the negative Hessian of the Log-Likelihood, taken at the true parameter values. I found and ignored some *data sets* that produce negative variance or larger than one correlation between parameters;
- *Eigen values of Hessian matrix* - I found and ignored some *data sets* that produce non-negative definite Hessian;
- *Convergence of Optimization Iterations* - I ignored *data sets*, from which the *Newton-Raphson-like* iterative method I use - *nlminb()* - did not converge. In case convergence is not achieved for some data set, we could attempt to change the initial point for iterations, so ignoring these cases would not lead to loss of generality;

2 Maximum Likelihood Estimators' properties.

2.1 Mean of Score.

Testing the estimates and the data I collected about the $\text{Log}L^*$ during the simulations, I tested the average of the *Gradient* elements taken at true parameter values (aka as Score). A well known property of the Maximum Likelihood Estimator is that the asymptotic average of the Score tends to zero. The simple proof comes below (Θ is the vector of parameters):

$$E\left(\frac{\partial \log L^*(\Theta|x)}{\partial \Theta}\right) = \int \frac{\partial L^*(\Theta|x)/\partial \Theta}{L^*(\Theta|x)} L^*(\Theta|x) \partial x = \int \frac{\partial L^*(\Theta|x)}{\partial \Theta} \partial x = \frac{\partial}{\partial \Theta} \int L^*(\Theta|x) \partial x \quad (2)$$

For the last equality to hold, however, regularity conditions need to be met. Specifically, the function $L^*(\Theta|x)$ and its first partial derivatives have to be continuous in all arguments.

Table 1. shows the simulation details and histograms of the Score functions taken at the true parameter values.

3 Full Likelihood.

According to the histograms on Table 1, the condition (2) holds only for parameter λ , whereas the scores with respect to μ and σ do not tend to average at zero.

To confirm this, I constructed a different likelihood based also on the generated data for censored mileages. The so constructed likelihood is a *Full Likelihood* in the sense that every observation contributes to it and there's *no unknown mileage*.

$$L(\lambda, \mu, \sigma, \eta, m, \mathbf{x}, \mathbf{y}) = \prod_{t=1}^T \left[\prod_{i=1}^{r_t^{(a)}} \left(f_a(x_i^{(a)}) \bar{F}_b(x_i^{(a)}) \bar{G}(x_i^{(a)}) \right) \prod_{i=1}^{r_t^{(b)}} \left(f_b(x_i^{(b)}) \bar{F}_a(x_i^{(b)}) \bar{G}(x_i^{(b)}) \right) \prod_{i=i}^{c_t} (g(y_i) \bar{F}_a(y_i) \bar{F}_b(y_i)) \right] \beta$$

I apply the same approach as above to estimate all *five* parameters simultaneously. I derived the Gradient and Hessian functions in this case and performed the same simulation - now with *no practical use*, but only useful to compare my estimations.

4 Conclusion.

As expected, the results in the Full Likelihood case seem very good and seem to possess the properties of MLE; Table 2 shows them. However, the same trend is noticed with respect to the condition (2) - it holds for the *Exponential* and *Weibull* parameters, but does not seem to hold for the *Log-normal* parameters.

Furthermore, the estimates for μ and σ appear not to conform with another MLE property - the one for normality around the true value. Table 2 shows the results for one parameter set from Likelihood (3) as well as the histograms of the Score functions. The same trends I obtained in other parameter sets as well.

These results make me doubt that the functions L^* and L or their first-order derivatives might be discontinuous with respect to parameters μ and σ .

5 Next steps.

I hope to check whether the departure from zero of the mean of the Score w.r.t. μ and σ is significant. If so then, is it due to discontinuousness of the likelihood or their derivatives. At the same time I continue to try to find errors in my simulation code or calculations, in which I have so far been unsuccessful.