Maximum Likelihood and Bayesian Estimation

Artin Armagan and Sayan Mukherjee

Lab assignment Three

March 12, 2010
The lab will be due on the 20th of March. You will be expected to write a short: 1-2 page lab report and also turn in code. The 1-2 pages do not include plots or graphs that illustrate ideas.
Data comes from your Facebook experiment. Your observations will be the elapsed time between subsequent FB log-ons between the times 8 AM - 12 AM. In your report, indicate how many samples you have. We will assume these observations are coming from an exponential distribution with parameter $\lambda$ and try to estimate the value from our observations.
We will estimate the parameter $p$ using two approaches

1. Maximum likelihood estimation (MLE).
2. Bayesian estimation.
Remember how we derived the MLE for $\lambda$.
The MLE is computed by maximizing the log-likelihood with respect to $\lambda$,

$$n \log \lambda - \lambda \sum_{i=1}^{n} x_i,$$

where $x_i$ the $i$th observation in your data set and $n$ is the number of samples collected.
In the frequentist framework, if we do not know the distribution of
the estimator (e.g. we do not have enough samples to claim that \( \hat{\lambda} \)
is approximately normally distributed) what should we do? One approach is
the idea of the bootstrap. This involves sampling from the data in hand
with replacement several times and to obtain several estimates \( \{ \hat{\lambda}_1, ..., \hat{\lambda}_T \} \) and then using this empirical
distribution to provide confidence intervals.
The bootstrap algorithm

Given data $D := \{(x_1), \ldots, (x_n)\}$

1. For $i = 1$ to $T$
   1. Sample $n$ points from $D$ with replacement and call this $D_i$.
   2. Given $D_i$, compute $\hat{\lambda}_i$ (this is a bootstrap estimate).

2. Construct a histogram from $\{\hat{\lambda}_i\}_{i=1}^T$.

3. Compute the $(1 - \alpha) \times 100\%$ confidence interval by finding the values $\ell$ and $u$ surrounding the $(1 - \alpha) \times 100\%$ of all the empirical observations made, $\{\hat{\lambda}_i\}_{i=1}^T$. 
Bayesian inference can be made by combining the information coming from the observations (the likelihood) and a prior distribution specified on the parameter of interest which conveys a personal opinion (if we have any) about the parameter.

\[ \text{posterior} \propto \text{likelihood} \times \text{prior} \]

After normalization, we have a posterior distribution on the parameter which we can use for inference.

In this simple problem, if we specify a uniform distribution on \( \lambda \) as our choice of prior distribution which conveys no prior information, the posterior distribution of the parameter is proportional to

\[ \text{posterior} \propto \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i} . \]

This posterior distribution, after normalization, is a Gamma distribution with parameters \( \alpha = n + 1 \) and \( \beta = \sum_{i=1}^{n} x_i \). (Recall that the pdf for Gamma distribution was \( p(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\lambda \beta} \)).
Bayes estimate is the mean of $\lambda$ with respect to its posterior distribution.

$$E(\lambda) = \int \lambda \, p(\lambda|x) \, d\lambda$$

The mean of a gamma distribution is given by $\alpha/\beta$ which yields $E(\lambda) = (n + 1)/\sum_{i=1}^{n} x_i$.

Another estimator than can be obtained from the Bayesian approach is the maximum a posteriori, MAP, estimator. MAP estimator is the maximizer of the posterior distribution, i.e. the $\lambda$ value that maximizes the posterior distribution (instead of the likelihood function this time). If one takes the derivative of the above-mentioned posterior distribution with respect to $\lambda$ and sets it equal to zero, this will yield the estimator $\tilde{\lambda} = n/\sum_{i=1}^{n} x_i$. 
Compute the maximum likelihood, MAP and Bayes estimates. What do you observe? What are the similarities and differences.

Obtain 95% and 99% confidence intervals for the MLE of $\lambda$ using bootstrap samples as explained earlier.

Obtain 95% and 99% credible intervals over $\lambda$ using the posterior distribution of $\lambda$, $p(\lambda|x)$.

What do you conclude from the point and interval estimates obtained on your FB log-on tendencies?