A Brief Introduction to Parameter Estimation

Gianfranco Balbo, András Horváth

Scuola di Dottorato in Scienza ed Alta Tecnologia
Dipartimento di Informatica
Università di Torino
Italia

April-May, 2009

Outline

Estimation Based on Statistical Indices

- given
  - a parametrisable stochastic model (a distribution or a stochastic process) and
  - a set of observations, O, where O can be samples of a distribution or trajectories of a stochastic process
- find such set of parameters that given statistical indices of the observations are identical (or similar) to those of the model
- possible indices are
  - moments
  - tail decay
  - correlation

Exponential Distribution

- exponential distribution has a single parameter, λ
- given a set of samples from an exponential distribution, $O = \{x_1, \ldots, x_n\}$
- to catch the mean set λ according to

$$\frac{1}{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- to catch the probability that a sample is greater than x set λ according to

$$\exp(-x\lambda) = \frac{\text{number of samples larger than } x}{n}$$

Erlang distributions

- Erlang distribution is the convolution of k exponential distributions with parameter λ
- i.e., it has two parameters
- the mean and the variance are according to

$$E[X] = \frac{k}{\lambda}$$

$$E[(X - E[X])^2] = \frac{k}{\lambda^2}$$

- two equations with two unknowns determine the two parameters

A Simple Counting Process

- a two state continuous time Markov chain with infinitesimal generator

$$Q = \begin{vmatrix} -\lambda_1 & \lambda_1 \\ \lambda_2 & -\lambda_2 \end{vmatrix}$$

- state 1 generates events according to a Poisson process with intensity γ

A Simple Counting Process

- some statistical properties:
  - moments of inter-event times:

$$E[X] = \frac{\lambda_1 + \lambda_2}{\gamma\lambda_2}$$

$$E[X^2] = \frac{2(\gamma\lambda_1 + (\lambda_1 + \lambda_2)^2)}{\gamma^2\lambda_2^2}$$

  - joint moment of two consecutive inter-event times:

$$E[X_1 X_2] = \frac{(\lambda_1 + \lambda_2)^2}{\gamma^2\lambda_2^2}$$

- three equations with three unknowns determine the three parameters

Estimation Based on Statistical Indices

- choice of the applied indices is crucial
- relation of the parameters of the model and the statistical indices are usually not simple
- realizable region of statistical indices depend on the model and might be unknown

Maximum Likelihood Estimation

- given
  - a parametrisable stochastic model (a distribution or a stochastic process) and
  - a set of observations, O, where O can be samples of a distribution or trajectories of a stochastic process
- find such set of parameters, λ, that the likelihood that model produces O is maximal
- the likelihood can be
  - the joint density of the observations
  - the probability of the observations, i.e., $P(O|\lambda)$
  - mixture of density and probability

A Simple Example: ML for Exponential Distribution

- **situation 1:** we are given a set of samples from an exponential distribution, $O = \{x_1, \ldots, x_n\}$
- the likelihood of O is the joint density of n samples at $x_1, \ldots, x_n$
- assuming that the parameter of the exponential distribution is λ, the likelihood is

$$L(\lambda) = \prod_{i=1}^{n} \lambda \exp(-\lambda x_i) = \lambda^n \prod_{i=1}^{n} \exp(-\lambda x_i)$$

- our aim is to find such λ that $L(\lambda)$ is maximal

A Simple Example: ML for Exponential Distribution

- taking the logarithm of $L(\lambda)$ we have

$$\ln(L(\lambda)) = n \ln(\lambda) - \lambda \sum_{i=1}^{n} x_i$$

- for the derivative of the logarithm we have

$$\frac{d \ln(L(\lambda))}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i \begin{cases} > 0 & \text{if } \lambda < n / \sum_{i=1}^{n} x_i \\ = 0 & \text{if } \lambda = n / \sum_{i=1}^{n} x_i \\ < 0 & \text{if } \lambda > n / \sum_{i=1}^{n} x_i \end{cases}$$

- consequently, $L(\lambda)$ is maximal if

$$\lambda = \frac{n}{\sum_{i=1}^{n} x_i}$$

A Simple Example: ML for Exponential Distribution

- **situation 2**: instead of having n samples themselves, we know only how many of them are smaller than a given limit, T
- let r denote the number of observations that are smaller than T
- the likelihood is the probability of the event that out of n samples r are smaller then T

$$L(\lambda) = \binom{n}{r} (1 - \exp(-\lambda T))^r (\exp(-\lambda T))^{n-r}$$

- it is easy to verify that a quantity like $(1-x)^r x^{n-r}$ with $0 \le x \le 1$ is maximal with $x = (n-r)/n$
- hence we must have $\exp(-\lambda T) = (n-r)/n$, from which $\lambda = -\ln((n-r)/n)/T$

A Simple Example: ML for Exponential Distribution

- **situation 3 (censored samples)**: we know the value of those samples that are smaller than T, for those samples that are larger we know only that they are larger
- the total number of samples is n
- r denotes the number of samples that are smaller than T and we organise the samples such that $x_1 \le x_2 \le \cdots \le x_r < T$
- $n - r$ samples are larger than T but their values are unknown
- the likelihood is the joint density of the r known values joint with the probability that $n - r$ values are larger than T, i.e.,

$$L(\lambda) = \left(\prod_{i=1}^{r} \lambda \exp(-\lambda x_i) \right) (\exp(-\lambda T))^{n-r}$$

A Simple Example: ML for Exponential Distribution

- by considering the derivative of the logarithm of $L(\lambda)$, it can be verified that $L(\lambda)$ is maximal at

$$\lambda = \frac{r}{\sum_{i=1}^{r} x_i + (n-r)T}$$

A Simple Example: ML for Exponential Distribution

- for situation 1 and 2 the resulting ML estimate has a simple probabilistic interpretation:
  - situation 1: the parameter of the exponential distribution is one divided by the mean
  - situation 2: the parameter is such that the probability of having a value larger than T is $(n-r)/n$
- the result in situation 3 does not seem to have such simple interpretation
- for more complex models with many parameters to find, the ML estimate can rarely be written as simple closed form expressions
- when a closed form solution is not available, numerical techniques can be applied to maximise the likelihood or the log-likelihood function

Expectation Maximisation Method

- the EM method is a technique to obtain the ML estimate in case of incomplete data problems
- it starts from an initial guess and follows and iterative scheme to improve the parameters step by step
- notation:
  - X is the random variable (rv) describing the complete data
  - Y is the rv describing the known data
  - Z is the rv describing the missing data (i.e., Y and Z together form X)
- the initial set of parameters are denoted by λ_0
- λ_{i+1} is obtained from λ_i by applying the Expectation (E) step and the Maximisation (M) step

Expectation Maximisation Method

- the E step reconstructs the missing data in expectation applying the actual set of parameters λ_i
- formally, denoting the known data by y, the E step computes the missing data, z, according to the conditional expectation
$$z_i = E[Z|Y = y, \lambda_i]$$
- the M step reestimates the parameters by maximising the likelihood function, $L(\lambda)$, using the complete data formed by y and z, i.e.,
$$\lambda_{i+1} = \text{argmax}_\lambda L(\lambda, y, z_i)$$
- convergence toward (a possibly local) maximum of the likelihood function is guaranteed for a wide range of situations

A Simple Example: EM for Exponential Distribution

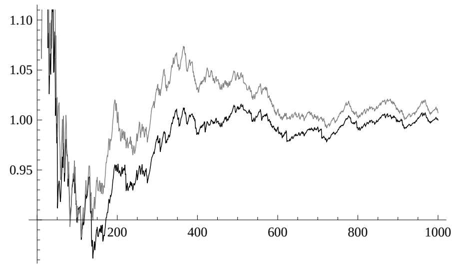
- we apply the EM method to have an ML estimate for situation 3 seen before
- the known data, y, consists in $x_j, 1 \leq j \leq r$
- we know that $x_j > T$ for $j > r$
- E-step: based on the memoryless property of the exponential distribution, we reconstruct $x_j, j > r$ as $x_j = T + 1/\lambda_i$
- M-step: $\lambda_{i+1} = n / \sum_{j=1}^{n} x_j$

A Variant: the Monte Carlo EM Method

- if the computation of the E step is difficult, the missing data can be generated by simulation
- convergence is slower and/or not guaranteed in this case
- for the last example the E step becomes: we reconstruct $x_i, i > r$ as $x_i = T + R(\lambda_i)$ where $R(\lambda)$ denotes a random number generated according to an exponential distribution with parameter λ

Numerical Examples

- estimation of parameter of exponential samples
- $\lambda = 1$, x-axis: number of samples used, y-axis: estimate
- black line: closed form ML estimate for situation 1,
- gray line: closed form ML estimate for situation 3 censored at $T = 1.5$ (about 22% of the samples are not known)



Numerical Examples

- black line: ML estimate by the EM method for situation 3 (identical to the gray line of the previous figure)
- gray line: ML estimate by the Monte Carlo EM method for situation 3 which oscillates because of the random reconstruction of the missing data