A Bayesian Look at Classical Estimation: The Exponential Distribution

Abdulaziz Elfessi and David M. Reineke
University of Wisconsin - La Crosse

Journal of Statistics Education Volume 9, Number 1 (2001)

Copyright © 2001 by Abdulaziz Elfessi and David M. Reineke, all rights reserved.
This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

Key Words: Bayes estimator; Classical estimator; Credibility interval; Improper prior distribution.

Abstract

Many undergraduate students are introduced to frequentist or classical methods of parameter estimation such as maximum likelihood estimation, uniformly minimum variance unbiased estimation, and minimum mean square error estimation in a reliability, probability, or mathematical statistics course. Rossman, Short, and Parks (1998) present some thought provoking insights on the relationship between Bayesian and classical estimation using the continuous uniform distribution. Our aim is to explore these relationships using the exponential distribution. We show how the classical estimators can be obtained from various choices made within a Bayesian framework.

1. Introduction

The one-parameter exponential distribution is often used to illustrate concepts such as parameter estimation in undergraduate courses in mathematical statistics, probability, and reliability. The exponential density is easy to manipulate analytically and provides a good starting point for discussions of more general distributions. Furthermore, its analytical tractability allows exploration of the relationships between classical and Bayesian estimation.

Consider a random sample of independent observations \( X_1, \ldots, X_n \) from an exponential distribution with probability density function

\[
f(x; \theta) = \theta \exp(-\theta x), \quad x > 0, \quad \theta > 0.
\]

Among the classical estimators of \( \theta \), it is easy to show that the maximum likelihood estimator (MLE)
and the uniformly minimum variance unbiased estimator (UMVUE) are $\frac{n}{\sum_{i=1}^{n} X_i}$ and 

$\frac{(n-1)/\sum_{i=1}^{n} X_i}$, respectively. In the class of estimators of the form $\frac{c}{\sum_{i=1}^{n} X_i}$, the one that minimizes the mean squared error $E\left[\left(\frac{c}{\sum_{i=1}^{n} X_i} - \theta\right)^2\right]$ is $(n-2)/\sum_{i=1}^{n} X_i$.

In Section 2 we consider the problem of estimating the parameter $\theta$ using the Bayesian approach. Bayesian estimators derived from an improper prior distribution can be used to derive the classical estimators given above. The technique of deriving the classical estimators from the Bayesian estimator is not new. Rossman, Short, and Parks (1998) present a very helpful paper for teaching connections between Bayesian and classical estimators using the continuous uniform distribution.

The roots of Bayesian analysis lie in Bayes's Theorem:

$$P\left(B_j | A\right) = \frac{P\left(B_j\right)P\left(A | B_j\right)}{\sum_{j=1}^{m} P\left(B_j\right)P\left(A | B_j\right)}.$$

where $A$ is an event and the $B_j$'s, $j = 1, ..., m$, are mutually exclusive and collectively exhaustive events in a sample space with $P\left(B_j\right) > 0$ for all $j$. The same results can be translated to random variables, both discrete and continuous. Let $U$ and $Y$ be continuous random variables and let $f_U(\cdot)$ be the prior density of $U$ and $g(\cdot | u)$ be the conditional density of $Y$ given $U$. Bayes' Theorem for continuous random variables then can be represented by

$$g(\cdot | y) = \frac{f_U(\cdot)g(\cdot | \cdot)}{\int_{-\infty}^{\infty} f_U(\cdot)g(\cdot | \cdot)d\mu},$$

where $g(\cdot | y)$ is called the posterior density function of $U$. For more information see Rohatgi (1984) or Berger (1988). In the next section we illustrate the continuous case using the exponential distribution with a single parameter $\theta$. Additionally, interval estimators for $\theta$ are compared and an example is given.

### 2. Derivation of Point and Interval Estimators

Bayesian statistics have traditionally been dominated by the notion of conjugate priors. A class $C$ of prior distributions is a conjugate family for $F$, where $F$ denotes a class of density functions, if the posterior distribution is also in the class $C$ for all density functions in $F$ and all prior density functions in $C$. When dealing with conjugate priors, the posterior distribution can be easily calculated. In this section we derive the posterior distribution by using an improper prior distribution for the parameter $\theta$. 

Page 2 of 7
Consider the improper prior distribution (i.e. \( \int_0^\infty \pi(\theta) \, d\theta = \infty \)) for \( \theta \) of the form

\[
\pi(\theta) = \theta^{\alpha-1} \exp(-\beta \theta), \quad \theta > 0, \quad -\infty < \alpha < \infty, \quad \beta \geq 0.
\]

Notice that this prior distribution is the kernel of a gamma distribution when \( \alpha \geq 0 \). However, such a restriction on \( \alpha \) is not necessary and decreases the flexibility of the resulting parameter estimator.

Applying Bayes' Theorem

\[
\pi(\theta \mid X_1, \ldots, X_n) \propto \frac{\pi(\theta) f(X_1, \ldots, X_n \mid \theta)}{f(X_1, \ldots, X_n)}
\]

where \( f(X_1, \ldots, X_n) \) is the marginal distribution of \( X \), it follows that the posterior distribution of \( \theta \) is

\[
\pi(\theta \mid X_1, \ldots, X_n) \propto \theta^{n+\alpha-1} \exp\left(-\theta \left(\beta + \sum_{i=1}^{n} x_i\right)\right)
\]

The posterior distribution \( \pi(\theta \mid X_1, \ldots, X_n) \) is proper when \( \alpha + n > 0 \) and has a constant of proportionality given by

\[
\left(\beta + \sum_{i=1}^{n} x_i\right)^{\alpha+n} / \Gamma(\alpha + n).
\]

The estimator \( \hat{\theta} \) is derived by choosing that value of \( \theta \) which minimizes \( E\left[(\hat{\theta} - \theta)^2\right] \) (assuming the squared error loss). The Bayes estimator of \( \theta \) is given by

\[
\hat{\theta}_B = E\left[\theta \mid X_1, \ldots, X_n\right]
\]

\[
= \frac{\int_0^\infty \theta \, \theta^{n+\alpha-1} \exp\left(-\theta \left(\beta + \sum_{i=1}^{n} x_i\right)\right) \, d\theta}{\int_0^\infty \theta^{n+\alpha-1} \exp\left(-\theta \left(\beta + \sum_{i=1}^{n} x_i\right)\right) \, d\theta}
\]

\[
= \frac{\alpha + n}{\beta + \sum_{i=1}^{n} x_i}
\]

The classical estimators derived in \textit{Section 1} can be obtained from the Bayes estimator \( \hat{\theta}_B = (n + \alpha) / \left(\beta + \sum_{i=1}^{n} x_i\right) \) by choosing different values of \( \alpha \) and \( \beta \). If \( \alpha = 0 \) and \( \beta = 0 \), then the estimator corresponds to the MLE and the prior distribution is the Jeffreys' prior, \( \pi(\theta) \propto 1/\theta \), a
standard noninformative prior as well as an improper prior. For more information on the Jeffreys' prior see Berger (1988). Choosing $\alpha = -1$ and $\beta = 0$ yields the UMVUE. The mean square estimator corresponds to setting $\alpha = -2$ and $\beta = 0$. For $\alpha = 1$ and $\beta = 0$ the prior density function is $\pi(\theta) = 1$ (the flat improper prior). The resulting estimator in the case of the flat prior is
\[(n+1)/\sum_{i=1}^{n} X_i.\]

A 100\% confidence interval for a parameter $\theta$ is obtained by finding $L$ and $U$ such that
\[P(L < \theta < U) = C.\] When $X_1, \ldots, X_n$ are independent and identically distributed exponential random variables, Kapur and Lamberson (1977) show that $2\theta \sum_{i=1}^{n} X_i$ has a chi-square distribution with $2n$ degrees of freedom. Using this transformation, the interval estimate is developed by solving for $\theta$ in
\[P\left(\chi^2_{2n}(1 - \frac{C}{\theta}) < 2\theta \sum_{i=1}^{n} X_i < \chi^2_{2n}(\frac{1-C}{\theta})\right) = C\]
and the resulting 100\% confidence interval for $\theta$ is
\[\left[\frac{\chi^2_{2n}(1 - \frac{C}{\theta})}{2 \sum_{i=1}^{n} X_i}, \frac{\chi^2_{2n}(\frac{1-C}{\theta})}{2 \sum_{i=1}^{n} X_i}\right]\]

The Bayesian analog to the confidence interval is called a credibility interval. In general, a 100\% credibility interval for a parameter $\theta$ given a random sample $X_1, \ldots, X_n$ is an interval $(l(X_1, \ldots, X_n), u(X_1, \ldots, X_n))$ such that
\[P(l(X_1, \ldots, X_n) < \theta < u(X_1, \ldots, X_n)) = \int_{l(X_1, \ldots, X_n)}^{u(X_1, \ldots, X_n)} f(\theta | X_1, \ldots, X_n) d\theta = C.\]

Kapur and Lamberson (1977) show that $2\theta \left(\beta + \sum_{i=1}^{n} X_i\right)$ has a chi-square distribution with $2(\alpha + n)$ degrees of freedom. By using the posterior distribution in (1) a 100\% Bayesian credibility interval is easily developed beginning with
\[P\left(\chi^2_{2(\alpha+n)}(1 - \frac{C}{\theta}) < 2\theta \left(\beta + \sum_{i=1}^{n} X_i\right) < \chi^2_{2(\alpha+n)}(\frac{1-C}{\theta})\right) = C\]
which gives the interval

\[
\left( \frac{\chi^2_{\alpha+n}(1-\frac{1-C}{2})}{2\left(\beta + \sum_{i=1}^{n} X_i\right)}, \frac{\chi^2_{\alpha+n}(1-\frac{1-C}{2})}{2\left(\beta + \sum_{i=1}^{n} X_i\right)} \right)
\]

So, when \( \alpha = 0 \) and \( \beta = 0 \) the Bayesian and classical interval estimates are the same.

3. Example

Consider the following random sample of cycles to failure (in ten thousands) for 20 heater switches subject to an overload voltage:

\[
0.0100, 0.0340, 0.1940, 0.5670, 0.6010, 0.7120, 1.2910, 1.3670 \\
1.9490, 2.3700, 2.4110, 2.8750, 3.1620, 3.2800, 3.4910, 3.6860 \\
3.8540, 4.2110, 4.3970, 6.4730
\]

These data are from Kapur and Lamberson (1977, p. 240). Table 1 summarizes the Bayesian point and interval estimates of \( \theta \). It also identifies the values of \( \alpha \) and the Bayesian interpretation of the prior distribution as well as the corresponding classical counterpart to each point estimate of \( \theta \). Notice that negative values of \( \alpha \) produce lower values for the posterior mean for \( \beta = 0 \). Negative values of \( \alpha \) and positive values of \( \beta \) put more prior weight on the small values of \( \theta \), resulting in lower estimates of the posterior mean.

Table 1. Bayes Estimates for Various Values of \( \alpha \) and \( \beta \)

| \( \alpha \) | \( \beta \) | Posterior mean | Classical counterpart | 95% Credibility interval |
|-------------|--------------|----------------|-----------------------|-------------------------|
| -2          | 0            | 0.3835         | minimum MSE estimate  | (0.2273, 0.5799)        |
| -1          | 0            | 0.4048         | UMVUE                 | (0.2437, 0.6061)        |
| -1          | 1            | 0.3964         |                        | (0.2386, 0.5935)        |
| -1          | 2            | 0.3883         |                        | (0.2338, 0.5813)        |
| -1          | 3            | 0.3805         |                        | (0.2291, 0.5697)        |
4. Conclusion

We have shown the relationship of Bayesian estimators of the scale parameter of the one-parameter exponential distribution to three classical estimators, namely the MLE, UMVUE, and minimum MSE estimator. We considered both point and interval estimators. Our Bayesian estimators were derived from an improper prior distribution that is rather general. In practice, $\alpha$ and $\beta$ are parameters whose values depend on the experimenter's a priori knowledge of the unknown parameter $\theta$ and its distribution. An example was used to demonstrate the methods presented and to illustrate how Bayesian methods can yield classical estimators.

Acknowledgements

The authors are very grateful to the editor and to three anonymous referees, all of whom contributed toward the improvement of the manuscript.

References

Berger, J. O. (1988), *Statistical Decision Theory and Bayesian Analysis* (2nd ed.), New York: Springer-Verlag.

Kapur, K. C. and Lamberson, L. R. (1977), *Reliability in Engineering Design*, New York: John Wiley & Sons, Inc.

Rohatgi, V. K. (1984), *Statistical Inference*, New York: John Wiley & Sons, Inc.

Rossman, A. J., Short, T. H. and Parks, M. T. (1998), "Bayes Estimators for the Continuous Uniform Distribution," *Journal of Statistics Education*, [Online], 6(3). ([http://ww2.amstat.org/publications/jse/v6n3/rossman.html](http://ww2.amstat.org/publications/jse/v6n3/rossman.html))
