Using the Uniform Distribution in Teaching the Foundations of Statistics

Charles Rohde
Department of Biostatistics, Johns Hopkins University
615 North Wolfe Street
Baltimore, Maryland 21205 USA
crohde@jhsph.edu

1 Abstract

The uniform distribution has long been used to provide examples where standard frequentist inference procedures are not applicable. In this expository paper we elaborate on these examples which can be used in teaching a variety of topics in the foundations of statistical inference. The paper by Wittinghill and Hogg (2001) discussed this subject from a frequentist viewpoint. Generalizations discussed here include two examples due to Basu (Ghosh (1988)) which greatly expand the instructional value.

2 General Uniform (Rectangular) Model

Let $X_1, X_2, \ldots, X_n$ be iid with pdf

$$f(x;\theta_1,\theta_2) = \begin{cases} \frac{1}{\Delta} & \theta_1 \le x \le \theta_2 \\ 0 & \text{elsewhere} \end{cases} \tag{1}$$

where $\Delta = \theta_2 - \theta_1$. It is easy to show that the minimum and maximum of $X_1, X_2, \ldots, X_n$ are minimal sufficient statistics for θ_1 and θ_2. The joint density of Y_1 and Y_n is given by

$$f(y_1, y_n; \theta_1, \theta_2) = \frac{n(n-1)(y_n - y_1)^{n-2}}{(\theta_2 - \theta_1)^n} \quad \theta_1 \le y_1 \le y_n \le \theta_2 \tag{2}$$

We note that the minimal sufficient statistic is not complete and that the maximum likelihood estimators of θ_1 and θ_2 are given by $\widehat{\theta}_1 = y_1$ and $\widehat{\theta}_2 = y_n$.

The joint likelihood function for θ_1 and θ_2 is thus

$$f(y_1, y_n, \theta_1, \theta_2) = \frac{1}{(\theta_2 - \theta_1)^n} \quad \theta_1 \le y_1 \le y_n \le \theta_2 \tag{3}$$

and the profile likelihoods for θ_1 and θ_2 are

$$\mathbb{L}_p(\theta_2) = \left[\frac{y_n - y_1}{\theta_2 - y_1}\right]^n \quad \theta_2 \ge y_n \text{ and} \mathbb{L}_p(\theta_1) = \left[\frac{y_n - y_1}{y_n - \theta_1}\right]^n \quad \theta_1 \le y_1 \tag{4}$$

3 Special Case 1: Uniform $\theta - \rho, \theta + \rho$

3.1 ρ Known

This example is discussed in Cox (2006) and various other places. Let $\theta_1 = \theta - \rho$ and $\theta_2 = \theta + \rho$ then we have that the joint density is

$$f(y_1, y_n; \theta) = \frac{n(n-1)(y_n - y_1)^{n-2}}{(2\rho)^n} \quad \theta - \rho \le y_1 \le y_n \le \theta + \rho \tag{5}$$

It follows that the (relative) likelihood function is

$$\mathbb{L}(\theta) = 1 \quad y_n - \rho \leq \theta \leq y_1 + \rho \tag{6}$$

i.e. all values of θ in the interval $y_n - \rho, y_1 + \rho$ are equally supported by the data.

For the special case where $\rho = 1/2$ it is easy to show that $[Y_1, Y_n]$ is a $100\left(1 - \frac{1}{2^{n-1}}\right)$ confidence interval for θ. If we take $n = 5$, $y_1 = .01$ and $y_n = .99$ then the $100(1 - \frac{1}{16})\% = 93.75\%$ confidence interval for θ is .01 to .99. But with these values of y_1 and y_n we are certain that $.49 \leq \theta \leq .51$ and yet our 93.75% confidence interval is $.01 \leq \theta \leq .99$. From **?**.

The distribution of the range $R = Y_n - Y_1$ is given by

$$f_R(r) = \frac{n(n-1)r^{n-2}(2\rho - r)}{(2\rho)^n} \quad 0 \leq r \leq 2\rho \tag{7}$$

and it follows that R is ancillary for θ. As Cox (2006) points out it is imperative to condition on this ancillary statistic.

3.2 ρ Unknown

In this case the maximum likelihood estimate of ρ is $\widehat{\rho} = (y_n - y_1)/2$ with density function

$$f_{\widehat{\rho}}(t) = \frac{n(n-1)t^{n-2}(\rho - t)}{\rho^{n-1}} \quad 0 \leq t \leq \rho \tag{8}$$

This does not depend on θ so it can be used as a marginal likelihood for ρ. This marginal likelihood is not the same as the profile likelihood for ρ which is proportional to $1/(\rho^n$ for $\rho \geq y_n$. We thus have a simple example of a marginal likelihood differing from a profile likelihood. Which to use is an open question.

4 Special Case 3: Basu Example 1

Suppose now that $\theta_1 = \theta$ and $\theta_2 = 2\theta$. Then the joint density is given by

$$[f(y_1, y_2; \theta) = \frac{n(n-1)(y_n - y_1)^{n-2}}{\theta^n} \tag{9}$$

for $\theta \leq y_1 \leq y_n \leq 2\theta$. The maximum likelihood estimate of θ is $\widehat{\theta} = y_n/2$ and hence the likelihood function is

$$\mathbb{L}(\theta; y_1, y_n) = \left[\frac{\widehat{\theta}}{\theta}\right]^n \quad ; \quad \frac{y_n}{2} \leq \theta \leq y_1 \tag{10}$$

The minimal sufficient statistic for θ is again the minimum and and the maximum of the order statistics and the minimal sufficient statistic is not complete. An ancillary statistic is Y_n/Y_1

Using results in the appendix the mean square error of the maximum likelihood estimator is given by

$$\text{MSE}(\widehat{\theta}) = \left[\frac{1}{2(n+1)(n+2)}\right]\theta^2 \tag{11}$$

so that $\widehat{\theta}_n$ is consistent. Also note that

$$\mathbb{P}\left\{W_n = n(\theta - \widehat{\theta}_n) \leq w\right\} = 1 - \left(1 - \frac{2w}{n\theta}\right)^n \tag{12}$$

so that W_n converges in distribution to an exponential rather than a normal distribution.

Another estimator for θ is Y_1 which is also consistent since its mean square error is

$$\text{MSE}(Y_1) = \frac{2\theta^2}{(n+1)(n+2)} \tag{13}$$

The ratio of the mean square error of the estimator Y_1 to that of the MLE is 4. Following Basu (Ghosh (1988)) the ratio of the mean square error of the estimator $\widetilde{\theta} = (4\widehat{\theta}_n + Y_1)/5$ to that of the maximum likelihood estimator is

$$\frac{\text{MSE}(\widetilde{\theta}_n)}{\text{MSE}(\widehat{\theta}_n)} = \frac{12}{25}\left(\frac{1+\frac{1}{2n}}{1+\frac{1}{n}}\right) \tag{14}$$

which tends to $12/25$ as $n \to \infty$ i.e. the MLE has asymptotic relative efficiency of $12/25$, slightly less than 50%. Thus we have a simple example of a non-efficient maximum likelihood estimator.

The distribution of the range $R = Y_n - Y_1$ is given by

$$f(r;\theta) = \frac{n(n-1)r^{n-2}(\theta - r)}{\theta^n} \quad \text{for } 0 < r < \theta \tag{15}$$

$U = R/\theta$ has a Beta$(n-2,1)$ distribution and hence is a pivot. Hence we can find u_1 and u_2 such that $\mathbb{P}(u_1 \leq U \leq u_2) = 1 - \alpha$ and a $100(1-\alpha)\%$ confidence interval for θ is given by

$$\left(\frac{R}{u_2}, \frac{R}{u_1}\right) \tag{16}$$

which can be compared to the likelihood interval.

5 Special Case 4: Basu Example 2

Suppose that $\theta_1 = \theta$ and $\theta_2 = \theta^2$ The joint density is now given by

$$f(y_1, y_2; \theta) = \frac{n(n-1)(y_n - y_1)^{n-2}}{[\theta(\theta-1)]^n} \tag{17}$$

for $\theta \leq y_1 \leq y_n \leq \theta^2$. It follows that the maximum likelihood estimator is $\widehat{\theta} = \sqrt{y_n}$ and that the (relative) likelihood function is

$$\mathbb{L}(\theta; y_1, y_n) = \left[\frac{\widehat{\theta}(\widehat{\theta}-1)}{\theta(\theta-1)}\right]^n \quad \sqrt{y_n} \leq \theta \leq y_1 \tag{18}$$

There is no ancillary statistic in this case.

6 Special Case 5:Uniform $(0, \theta)$

In this case the minimal sufficient statistic is Y_n, the maximum of the X_i's. Its distribution is complete. The maximum likelihood estimator of θ is Y_n which, from the appendix has expected value and variance

$$\mathbb{E}(Y_n) = \frac{n\theta}{n+1} \quad \mathbb{V}(Y_n) = \frac{n\theta^2}{(n+1)^2(n+2)} \tag{19}$$

Note that $\widehat{\theta}$ is consistent but with a rate of order n rather than the usual $\sqrt{n}$ associated with regular maximum likelihood estimators. It is easily shown that Y_n/θ^n is a pivot which can be used to construct a $100(1-\alpha)\%$ confidence interval for θ of the form

$$\{\theta\ : y_n \leq \theta \leq y_n/\alpha^{1/n}\} \tag{20}$$

7 Appendix

7.1 Moments

Letting $\Delta = \theta_2 - \theta_1$ the expected values, variances and covariances of Y_1 and Y_n are given by

$$\begin{array}{lcl} \mathbb{E}(Y_1) = \frac{\Delta}{n+1} + \theta_1 & ; & \mathbb{V}(Y_1) = \frac{n\Delta^2}{(n+1)^2(n+2)} \\ \mathbb{E}(Y_n) = \frac{n\Delta}{n+1} + \theta_1 & ; & \mathbb{V}(Y_n) = \frac{n\Delta^2}{(n+1)^2(n+2)} \\ \mathbb{C}(Y_1, Y_n) = \frac{\Delta^2}{(n+1)^2(n+2)} & & \end{array} \tag{21}$$

7.2 Asymptotics

The asymptotic distributions of the smallest and largest order statistics and the range obey the following:

$$\begin{array}{rcl} W_1 = n(Y_1 - \theta_1) & \xrightarrow{d} & \text{exponential}(\theta_1) \\ W_2 = n(\theta_2 - Y_n) & \xrightarrow{d} & \text{exponential}(\theta_2) \\ W_3 = n(\theta_2 - \theta_1 - R_n) & \xrightarrow{d} & \text{Gamma}(2, 2(\theta_2 - \theta_1)) \end{array} \tag{22}$$

Moreover W_1 and W_2 are asymptotically independent with the above distributions.

8 RESUME

In this expository paper we indicate how the uniform distribution can provide instructive examples in the foundations of statistics including conditioning, profile likelihood, likelihood, marginal likelihood, reference priors and confidence intervals.

9 REFERENCES

1 D. R. Cox. Principles of Statistical Inference. Cambridge University Press, 2006.

2 Jayanta K. Ghosh, editor. Statistical Information and Likelihood : A Collection of Critical Essays by Dr. D. Basu, volume 45 of Lecture Notes in Statistics. Springer-Verlag, 1988.

3 Constantinos Goutis and George Casella. Frequentist post-data inference. International Statistical Review, 63(3):325.344, 1995.

4 E.L. Lehmann. Elements of Large Sample Theory. Springer, first edition, 1999.

5 Dexter C. Whittinghill and Robert V. Hogg. A little uniform density with big instructional potential. Journal of Statistics Education, 9(2), 2001.