A Little Uniform Density With Big Instructional Potential

Dexter C. Whittinghill
Rowan University

Robert V. Hogg
University of Iowa

Journal of Statistics Education Volume 9, Number 2 (2001)

Copyright © 2001 by Dexter C. Whittinghill and Robert V. Hogg, 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: Confidence intervals; Efficiency; Estimation; Maximum likelihood; Sufficiency; Tests of hypotheses.

Abstract

We explore the varied uses of the uniform distribution on $[\theta - 1/2, \theta + 1/2]$ as an example in the undergraduate probability and statistics sequence or the mathematical statistics course. Like its cousin, the uniform distribution on $[0, \theta]$, this density provides tractable examples from the topic of order statistics to hypothesis tests. Unlike its cousin, which appears in many probability and statistics books, this uniform is less well known or used. We discuss maximum likelihood estimators, likelihood ratio tests, confidence intervals, joint distributions of order statistics, use of Mathematica®, sufficiency, and other advanced topics. Finally, we suggest a few exercises deriving likelihood ratio tests when the range is unknown as well, or for the uniform on $[\theta - \rho, \theta + \rho]$.

1. Introduction

Anyone who has taught undergraduate probability and statistics knows that it is sometimes difficult to find good examples of families of density functions that can be used in the classroom or for challenging homework for the better students. The 'common' families of one-parameter probability functions (such as the geometric or exponential) are almost surely used as an example or in one of the textbook problems. When trying to 'cook up' an example, often the instructor finds that the author of the book has already used the recipe! For illustration, we often think of a uniform density on $[0, \theta]$, but it is in nearly every probability and statistics text. When instructors get more creative ('the author will never have thought of this!'), they find that their wonderful mass or density function has a maximum likelihood estimator with a nearly intractable distribution. Exasperation sets in soon thereafter.
Here is a simple distribution that appears in two textbooks by the second author. It is also in Miller and Miller (1999) and Mood, Graybill, and Boes (1974), but not in many other texts. It is the uniform distribution on \([\theta - 1/2, \theta + 1/2]\), with density

\[
f(x; \theta) = I_{[\theta-\frac{1}{2}, \theta+\frac{1}{2}]}(x), \quad \theta \in \mathcal{R}.
\] (1)

\(I_{[a,b]}(x)\) is an indicator function that is 1 when the argument is in the interval, and 0 when it is not. Other texts sometimes use the uniform on other fixed ranges, such as \([\theta - 1, \theta + 1]\), but this is essentially the same function. Often more interesting problems can be created when the uniform distribution is on \([\theta - \rho, \theta + \rho]\), with both \(\theta\) and \(\rho\) unknown, and we will discuss these briefly.

The density (1) is a location-parameter alternative to the uniform on \([0, \theta]\), and it provides a rich assortment of material for discussions or examples. There are many unbiased and consistent estimators of \(\theta\) to compare, including the familiar \(\bar{x}\) as the method of moments estimator. The rather simple likelihood function yields an interesting uncountable set of maximum likelihood estimators. The density also provides fairly simple distributions for the order statistics. Using a modified likelihood ratio test, sensible and intuitive tests of hypotheses and confidence intervals can easily be derived. Finally, for more advanced students, the sufficiency results are not straightforward (we have joint sufficiency), and completeness and minimal sufficiency can also be discussed.

\section{The Uniform Distribution on \([\theta - 1/2, \theta + 1/2]\)}

When teaching probability and statistics, often the greatest need for help is in finding examples when the course reaches 'parametric point estimation.' That is, we want to estimate the value of the unknown parameter or parameters in a family of distributions, and to make statistical inferences we need to know the distributions of the estimators. Before presenting the flexibility of the uniform distribution on \([\theta - 1/2, \theta + 1/2]\), we begin with a discussion of how such a model might arise.

Consider the situation where you are waiting to catch a bus. You want to estimate \(\theta\), the average number of minutes that the bus takes to travel from the previous stop to your stop (you can see that other stop up the road). Each day you can take an observation on how long the bus takes and use the observations to estimate \(\theta\). Let us assume that under 'usual conditions' the variation induced by the traffic causes the bus to take between \(\theta\)-minus and \(\theta\)-plus a fixed unit of time, say one-half a minute for convenience, to arrive at your stop. Moreover, it is reasonable to assume that the distribution is uniform. If we let the random variable \(X\) be the number of minutes for the trip, then \(X\) has a probability density function (density) given by (1).

We note two things. First, the interval can be open or closed, and we will use a closed interval to avoid discussion of supremum versus maximum. Second, it is an easy homework problem for students to show that the mean of the distribution is \(\bar{x}\), the variance is 1/12, and the cumulative distribution function is

\[
F(x; \theta) = \begin{cases} 
0, & x < \theta - \frac{1}{2}; \\
\frac{x - \theta}{1/2}, & \theta - \frac{1}{2} \leq x \leq \theta + \frac{1}{2}; \\
1, & \theta + \frac{1}{2} < x.
\end{cases}
\]

The joint density for a random sample of \(n\) observations is the deceivingly simple
where \( y_1 \leq y_2 \leq \ldots \leq y_n \) are the order statistics. If an instructor does not stress the derivation of distributions of order statistics (we recommend covering it), the distributions of the minimum \( Y_1 \) and maximum \( Y_n \), along with their joint distributions, can be given to the students. The students can also simply work with data. (See Hogg and Tanis 1997, problems 6.1.15 and 10.1.11.)

In general, we present our students with different ways of finding estimators in probability and statistics. Because \( \theta \) is the mean of the distribution for \( X \), our students usually suggest \( \bar{X} \) as an estimator (which we will call \( W_1 \)). This intuitive result is the method of moments estimator of \( \theta \) and leads nicely into a discussion of that topic. \( \bar{X} \) is also the least-squares estimator of \( \theta \). (See Hogg and Tanis 1997, problem 6.1.15.)

If we are lucky, our students will suggest other intuitive estimators, such as the median (which we will call \( W_3 \)) or the midrange, which is the average of the minimum and maximum order statistics, say \( W_2 = (Y_1 + Y_n) / 2 \). While discussing desirable properties of estimators, such as unbiasedness with smallest variance, \( W_2 \) and \( W_3 \) can be compared with \( W_1 = \bar{X} \). All three of them are unbiased for \( \theta \). Comparison of the variances leads to the ideas of minimum variance and efficiency. At this point, the instructor can go in many different directions, but, because of the desirable properties of the maximum likelihood estimators, these are often discussed next. (See Miller and Miller 1999, problems 10.3 and 10.30.)

3. Maximum Likelihood and Efficiency

The likelihood function for a random variable is the joint density of the random sample, but we consider it as a function of \( \theta \) for a given sample, not as a function of the sample for a given \( \theta \). We seek maximum likelihood estimators (MLEs) because they can possess useful properties. For instance, if \( \hat{\theta} \) is an MLE for \( \theta \), and \( h \) is a function with a single-valued inverse, then \( h(\hat{\theta}) \) is the MLE of \( h(\theta) \). (See Casella and Berger 1990, Theorem 7.2.1, for a more advanced invariance property, where \( h \) is any function, not necessarily with a single-valued inverse.) Also, under certain regularity conditions, the sequence of MLEs (one for each \( n \)) is best asymptotically normal, or BAN. Textbooks are full of examples for which these results apply, but they sometimes lack examples that fail to satisfy regularity conditions. One of the ‘charms’ of our example is that it does not meet the regularity conditions needed for many results.

The likelihood function of a random sample from a distribution with density (1) is

\[
L(\theta; x_1, \ldots, x_n) = \prod_{i=1}^{n} I_{[\theta-1/2, \theta+1/2]}(x_i) = I_{[\theta-1/2, \theta+1/2]}(y_1)I_{[\theta-1/2, \theta+1/2]}(y_n).
\]

Unlike many other situations, here the MLE of \( \theta \) is not unique. A careful look at the likelihood function shows that any statistic \( \hat{\theta} \) satisfying
is an MLE of $\theta$. This includes the midrange, $W_2$. By way of examples, the student should demonstrate that $W_1 = \bar{X}$ and $W_3$ (the median) are not maximum likelihood estimators of $\theta$ even though they are unbiased.

Of course every estimator of the form $\omega_1(Y_n - \frac{1}{2}) + \omega_2(Y_1 + \frac{1}{2})$, with $\omega_1 + \omega_2 = 1$ and $\omega_i \geq 0$, $i = 1, 2$, is an MLE, but it is not unbiased unless $\omega_1 = \omega_2$. (See Hogg and Craig 1995, problem 6.3; Hogg and Tanis 1997, problems 6.1.15 and 10.1.11; Miller and Miller 1999, problems 10.75 and 76; and Mood, Graybill, and Boes 1974, example VII.7.)

Which of our three unbiased estimators, $W_1$, $W_2$, or $W_3$, is best? Are there better estimators? If class discussion has not already led to the notion that we are searching for unbiased estimators with small variance, then the idea of minimum variance unbiased estimators or efficiency can be raised here. When comparing $W_1$, $W_2$, and $W_3$, we must calculate the mean and variance of each of the three estimators. The case of $W_1 = \bar{X}$ is straightforward, as it is a consequence of standard formulas like $\text{Var}(\bar{X}) = \sigma^2/n$.

Each of the midrange ($W_2$) and the median ($W_3$) requires knowledge of distributions of order statistics. Depending on the level of the students or the book, the calculations for the cases involving order statistics can be assigned or not (with the results being given to the students) in various combinations. For the better students these are challenging homework problems, and we note some of the difficulties.

### 3.1 The Sample Midrange and the Joint Density for Two Order Statistics

Because $W_2 = (Y_1 + Y_n)/2$, we need the joint density for $Y_1$ and $Y_n$, as well as their marginal densities. In this section, we make the change of variables $Z_i = Y_i - (\theta - \frac{1}{2})$, $i = 1, 2, ..., n$, to simplify the calculations. The $Z_1 \leq Z_2 \leq \ldots \leq Z_n$ are order statistics and have the distributions of the respective order statistics for a random sample from the uniform distribution on $[0,1]$. Of course the means change, but those of $Y_i$ are easily found from those of $Z_i$, and the variances and covariances are the same. The density for the $k^{th}$ order statistic is given on page 198 of Hogg and Craig (1995) as

$$g_k(z_k) = \frac{n!}{(k-1)!(n-k)!}[F(z_k)]^{k-1}[1-F(z_k)]^{n-k}f(z_k), \quad (3)$$

where $F$ and $f$ are the respective distribution and density function of the uniform distribution on $[0,1]$. Using $k = 1$ and $k = n$, the densities for $Z_1$ and $Z_n$ are, respectively,

$$g_1(z_1) = n(1-z_1)^{n-1}, \quad 0 \leq z_1 \leq 1, \quad \text{and}$$

$$g_n(z_n) = nz_n^{n-1}, \quad 0 \leq z_n \leq 1.$$ 

It is easy to show that $E[Z_n] = 1 - \frac{1}{n+1}$ and $E[Z_1] = \frac{1}{n+1}$, so $E[Y_n] = \theta + \frac{1}{2} - \frac{1}{n+1}$ and
Thus

\[ E[Y_1] = \theta - \frac{1}{2} + \frac{1}{n+1}. \]

Thus

\[ E[W_2] = E \left[ \frac{(Y_1 + Y_n)}{2} \right] = \frac{1}{2} \left[ \theta - \frac{1}{2} + \frac{1}{n+1} + \theta + \frac{1}{2} - \frac{1}{n+1} \right] = \theta, \]

making \( W_2 \) unbiased. Also it is an easy exercise to find the variances, namely

\[ \text{Var}[Y_1] = \text{Var}[Y_n] = \text{Var}[Z_1] = \text{Var}[Z_n] = \frac{n}{(n+1)^2(n+2)}. \]

To find the \( \text{Var}[W_2] \), we need \( \text{Cov}[Y_1, Y_n] = \text{Cov}[Z_1, Z_n] \), and to find these we need the joint density of \( Z_1 \) and \( Z_n \). From page 199 of \textit{Hogg and Craig (1995)}, the joint density for the order statistics \( Z_i \), and \( Z_j \) is

\[ g_{ij}(z_i, z_j) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \times [F(z_i)]^{i-1} [1 - F(z_j)]^{n-j} f(z_i) f(z_j). \]

Hence the joint density for \( Z_1 \), and \( Z_n \) is

\[ g_{1,n}(z_1, z_n) = \frac{n!}{(n-2)!} (z_n - z_1)^{n-2}, \quad 0 \leq z_1 \leq z_n \leq 1. \]

\( \text{Cov}[Z_1, Z_n] = E[Z_1 Z_n] - E[Z_1] E[Z_n], \) and here

\[ E[Z_1 Z_n] = \int_0^1 \int_0^{2z_1} z_1 z_n (n-1)(z_n - z_1)^{n-2} dz_1 dz_n = \frac{1}{(n+1)^2}. \]

after some straightforward integration. Thus

\[ \text{Cov}[Y_1, Y_n] = \text{Cov}[Z_1, Z_n] = \frac{1}{n+2} - \left( \frac{n}{n+1} \right) \left( \frac{1}{n+1} \right) = \frac{1}{(n+1)^2(n+2)}. \]

Finally, the variance of \( (Y_1 + Y_n)/2 \) is

\[ \text{Var}[W_2] = \frac{1}{4} \left[ \text{Var}[Y_1] + \text{Var}[Y_n] + 2\text{Cov}[Y_1, Y_n] \right] = \frac{1}{4} \left[ \frac{n}{(n+1)^2(n+2)} + \frac{1}{(n+1)^2(n+2)} \right] = \frac{1}{2(n+1)(n+2)}. \]

\textbf{Remarks.} The transformation used employed the function \( Z_i - Y_i - (\theta - \frac{1}{2}) \), which is a pivotal quantity, because it made the distribution of \( Z_i \) independent of \( \theta \), whatever \( \theta \) was. Pivotal quantities, generally
discussed in the more rigorous undergraduate texts, are also very helpful in constructing confidence intervals. (See Casella and Berger 1990 or Mood, Graybill, and Boes 1974 for further discussion.) Another way of finding $\text{Var}(W_2)$ is to first find the distribution of $W_2$ using the 'distribution function technique' (see page 222 of Hogg and Tanis 1997).

### 3.2 The Sample Median

Again in this discussion, we let $Z_i = Y_i - (\theta - \frac{1}{2})$. For $n = 2m + 1$, the sample median is $W_3 = Y_{m+1}$. Using (3), the density for $Z_{m+1}$ is

$$g_{m+1}(z) = \frac{n!}{m!m!} z^m (1-z)^m, \quad 0 \leq z \leq 1,$$

which is that of a beta distribution with $\alpha = \beta = m + 1$. Hence $E(Z_{m+1}) = \frac{1}{2}$; so

$$E(W_3) = E(Y_{m+1}) - \frac{1}{2} + \theta - \frac{1}{2} - \theta.$$ Moreover,

$$\text{Var}(W_3) = \text{Var}(Y_{m+1}) = \text{Var}(Z_{m+1}) = \frac{(m+1)^2}{(2m+3)(2m+2)^2} = \frac{1}{4(n+2)}.$$

If the students have not studied the beta distribution, a computer algebra system (CAS) such as Mathematica® or Maple® can be used to establish the results.

For $n = 2m$, the median is $W_3 = (Y_m + Y_{m+1})/2$, but as before we will work with $Z_m$ and $Z_{m+1}$. From (4) the joint density of $Z_m$ and $Z_{m+1}$ is

$$g_{m,m+1}(u,v) = \frac{n!}{(m-1)!(m-1)!} u^{m-1} (1-v)^{m-1}, \quad 0 \leq u \leq v \leq 1.$$

Moreover, the marginal distributions of $Z_m$ and $Z_{m+1}$ are, respectively,

$$g_m(u) = \frac{n!}{(m-1)!m!} u^{m-1} (1-u)^m, \quad 0 \leq u \leq 1,$$

and

$$g_{m+1}(v) = \frac{n!}{m!(m-1)!} v^m (1-v)^{m-1}, \quad 0 \leq v \leq 1.$$

These are beta distributions with $\alpha = m$, $\beta = m + 1$ and $\alpha = m + 1$, $\beta = m$, respectively. Hence

$$E[Z_m] = \frac{m}{2m+1} \quad \text{and} \quad E[Z_{m+1}] = \frac{m+1}{2m+1}.$$
and
\[
E\left( \frac{Z_m + Z_{m+1}}{2} \right) = \frac{1}{2} \quad \text{and} \quad E(W_2) = \frac{1}{2} + (\theta - \frac{1}{2}) = \theta.
\]

Also,
\[
\text{Var}[Y_m] = \text{Var}[Y_{m+1}] = \text{Var}[Z_m] = \text{Var}[Z_{m+1}] = \frac{n}{4(n+1)^2}.
\]

As before, to find \( \text{Var}[W_3] \) we must evaluate
\[
E[Z_m Z_{m+1}] = \int_0^1 \int_0^u uv \frac{n!}{(m-1)!(m-1)!} u^{m-1}(1-v)^{m-1} du \, dv = \frac{n(n+4)}{4(n+2)(n+1)}.
\]

The integration is straightforward, or a CAS can be used. Thus
\[
\text{Cov}[Y_m, Y_{m+1}] = \text{Cov}[Z_m, Z_{m+1}] = \frac{n(n+4)}{4(n+1)(n+2)} - \left( \frac{n}{2n+2} \right) \left( \frac{n+2}{2n+2} \right) = \frac{n^2}{4(n+1)^2(n+2)}.
\]

Finally, algebra yields
\[
\text{Var}[W_3] = \text{Var}\left[ \frac{Y_m + Y_{m+1}}{2} \right] = \frac{1}{4} \left[ \text{Var}(Y_m) + \text{Var}(Y_{m+1}) + 2\text{Cov}(Y_m, Y_{m+1}) \right]
\]
\[
= \frac{n}{n+1} \cdot \frac{1}{4(n+2)}.
\]

In most cases you would not make the students do all of the work indicated above, and again Mathematica® or Maple® could be used. However, after they have suggested the various estimators, and found the mean and variance of at least two (including \( \bar{X} \)), an instructor could present all of the variances and have the students compare them by calculating the relative efficiency of one estimator to another. Having the students derive or check some of the efficiency results requires them to work with inequalities, something they may not have done for quite some time in their undergraduate careers.

### 3.3 Comparison of the Three Estimators of \( \theta \)

Table 1 gives the variances and shows the relative efficiencies of the estimators \( W_1 = \bar{X} \), the midrange \( W_2 \), and the median \( W_3 \). It contains some interesting results. Of course, for \( n = 2 \) the mean, median, and midrange are equivalent. For \( n \geq 3 \), the midrange is more efficient than the mean, which is more efficient than the median. That is, the average of the two extreme values is better than the average of all of the values, which is better than the median for estimating the center of this fixed-width density. At first this might seem amazing, but the midrange is based on two order statistics, each of which has a smaller variance than any \( Y_i \). It should also be noted that for the sample median, the case of \( n = 2m \) is more efficient than for \( n = 2m + 1 \), which has one more observation. In this case, two order statistics are better than one, even if the two are associated with a smaller sample size!
Table 1. Variances and Relative Efficiencies of the Mean, Median, and Midrange

| Estimator: | $W_1$, mean | $W_2$, midrange | $W_3$, median |
|-----------|-------------|-----------------|---------------|
| Variance: | $\frac{1}{12n}$ | $\frac{1}{2(n+1)(n+2)}$ | $\frac{n}{2(n+1)(n+2)}$ if even; $\frac{n}{2(4(n+2))}$ if odd |
| Efficiency of $W_1$ to above: | | $\frac{6n}{(n+1)(n+2)}$ if even; $\frac{3n}{n+2}$ if odd |
| Efficiency of $W_2$ to above: | | | $\frac{n}{2}$ if even; $\frac{n+1}{2}$ if odd |

NOTE: See Hogg and Tanis (1997), problem 10.1.11, for a comparison of the mean, median, and midrange for $n = 3$.

**Remark.** If the median is to be used for an odd sample size, don't throw away an observation! Statisticians do not recommend throwing away or ignoring any data. Instead, for a sample of size $2m + 1$, combine the strength of the median for an even sample size and the information from the odd observation. Taking a linear combination of the median for the first $2m$ observations and the last observation gives yet a better statistic:

$$\frac{2m}{2m + 1} W_3(2m) + \frac{1}{2m + 1} X_{2m+1}.$$

Finally, when discussing the Cramér-Rao Lower Bound, this example can be used as one where the conditions of the theorem are not met: the domain depends on the parameter being estimated. The students can also show that all three estimators are consistent; each is unbiased and its variance converges to zero as $n$ increases. In that regard, it is interesting to note that the variances of the mean and median are of order $1/n$, while that of the midrange is $1/n^2$. That is, for large $n$, the variance of the midrange is much smaller.

### 4. Sufficiency and Other Advanced and Varied Topics

The topic of sufficiency is very important for finding good estimators, but it can be very difficult for students to grasp. It is important because sufficient statistics are associated with good estimators. Unique MLEs are functions of sufficient statistics, if they exist (see Rice 1995, p. 284, Corollary A). Unbiased estimators that are functions of sufficient statistics have smaller variances, and unbiased estimators derived from sufficient statistics with 'complete' families are uniformly minimum variance unbiased estimators. The concept that a sufficient statistic contains all of the information from the sample necessary for estimating the parameter sounds simple and intuitive. However, the definition of a sufficient statistic is very technical and not always written intuitively. Many authors use "$W(X_1, ..., X_n)$ is sufficient for $\theta$ if and only if the conditional distribution of $X_1, ..., X_n$, given $W = w$, does not depend on $\theta$ for any value $w$ of $W$." (See Miller and Miller 1999, Definition 10.3, and Mood, Graybill, and Boes 1974, Definition 15, p. 301.) The intuition in the
definition is that given you have calculated the sufficient statistic, it has already sucked out all information about \( \theta \) from the sample. Yet because sufficiency is so important, authors go to great lengths, committing much text space with description and examples, in order to explain the concept. To help apply the concept of sufficiency to finding good estimators, authors present one or two factorization theorems. Even then, for a student new to the material, it can be difficult seeing that a factorization theorem follows intuitively from the concept. In Hogg and Craig (1995), over two pages of discussion precedes the definition, and then the definition is presented in a functional form like a factorization theorem. Hogg and Tanis (1997) wait until the end of the book to discuss the topic, and use the factorization theorem as the definition.

Let us assume that sufficiency has been defined in the course, and that a one-statistic-one-parameter factorization theorem like Theorem 1 (p. 318) of Hogg and Craig (1995) has been presented. The random variable of (1) provides a good second or third example for discussing sufficiency. The likelihood function in (2) is:

\[
L(\theta; x_1, \ldots, x_n) = \prod_{i=1}^{n} I[\theta-1/2, \theta+1/2](x_i) = I[\theta-1/2, y_n](y_1)I[y_1, \theta+1/2](y_n). \tag{5}
\]

From this expression, some students will say that you can't rewrite (5) so that there is a \( k_1(u_1(x_1, x_2, \ldots, x_n); \theta)k_2(x_1, x_2, \ldots, x_n) \) factorization, and hence there is no single sufficient statistic \( u_1 \).

Some may in fact come up with the idea that \( Y_1 \) and \( Y_n \) are jointly sufficient statistics. Example 1, page 348 of Hogg and Craig (1995), shows that \( Y_1 \) and \( Y_n \) are jointly sufficient for \( \theta \). They also point out that these are minimally sufficient and state that there can't be one sufficient statistic. So the density of (1) shows the student that all examples don't work out nicely. The fact that \( Y_1 \) and \( Y_n \) are joint sufficient statistics for \( \theta \) explains why the variance of the midrange is so much less than for \( W_1 \) and \( W_3 \).

**Remark.** First, Mood, Graybill, and Boes (1974) give an alternate definition of sufficiency (Definition 16, p. 306). Paraphrased it reads, "\( W(X_1, \ldots, X_n) \) is sufficient for \( \theta \) if and only if the conditional distribution of \( T \) given \( W = w \) does not depend on \( \theta \) for any statistic \( T(X_1, \ldots, X_n) \) [and any value \( w \) of \( W \)]. They point out that it is 'particularly useful in showing that a statistic is not sufficient.' Second, in principle, if you cannot see a factorization, this does not mean that a factorization is impossible. Thus, sometimes you need to derive the distribution of \( T \), given \( W \), to demonstrate that \( W \) is not sufficient. As an exercise, the first author showed that \( W_2 \) was not sufficient by considering the joint density for \( Y_1, Y_2, \) and \( Y_n \). He then used \( Y_2 \) as the \( T \) statistic, \textit{Mathematica®} for pictures, and calculus to find the distribution of \( T = Y_2 \) given \( W_2 \). Of course it depended on \( \theta \).

The density for (1) is also useful for illustrating other topics. Mood, Graybill, and Boes (1974) show that \( \theta \) is a location parameter (example 37, p. 333), and then that \( W_2 \) is the uniformly smallest mean-squared error estimator in the class of location invariant estimators (example 39, p. 335). Hogg and Craig (1995) essentially show that \( Y_1 \) and \( Y_n \) are not complete (example 2, p. 349), and discuss the idea of ancillary statistics (example 5, p. 357). We have noted that the MLE is not unique, as any weighted mean of \( Y_1 + \frac{1}{2} \) and \( Y_n - \frac{1}{2} \) will serve as an MLE. However, none of these are sufficient alone; so there is no minimal sufficient statistic for \( \theta \), because to enjoy that property the MLE must be sufficient.

### 5. Confidence Intervals and Tests of Hypotheses

After discussing the properties of good estimators, it is easy to forget that we wanted to estimate the time the
bus takes to travel between stops. Let us agree that \( W = W_2 \) is our best estimator of \( \theta \). It is unbiased, a maximum likelihood estimator, a function of sufficient statistics (if not sufficient itself), and has as small a variance as we have seen. Therefore we proceed being confident that we are doing a good job of estimation, if not 'the best' job. This suggests that \( W_2 \pm c \), where \( c \) is a constant such that

\[
P(W_2 - c \leq \theta \leq W_2 + c) = 1 - \alpha,
\]

might be the best \( 1 - \alpha \) confidence interval for \( \theta \). However, this is not the case, although a good student might find it an interesting exercise to find that constant \( c \) by determining the distribution of \( W_2 \).

We illustrate a better way by first finding a good test of \( H_0 : \theta = \theta_0 \) against \( H_1 : \theta \neq \theta_0 \), and then we use the relationship between tests and confidence intervals to find a confidence interval for \( \theta \). The likelihood ratio is given by

\[
\lambda = \frac{\prod_{i=1}^{n} I[\hat{\theta}_0 - 1/2, \hat{\theta}_0 + 1/2](x_i)}{\prod_{i=1}^{n} I[\hat{\theta} - 1/2, \hat{\theta} + 1/2](x_i)},
\]

where \( \hat{\theta} \) is a maximum likelihood estimator such as \( W_2 \). Note that \( \lambda = 1 \) provided \( \theta_0 - 1/2 \leq y_1 \leq y_n \leq \theta_0 + 1/2 \), but \( \lambda = 0 \) if \( y_1 < \theta_0 - 1/2 \) or \( \theta_0 + 1/2 < y_n \). Thus, strictly following the likelihood ratio criterion, we would reject \( H_0 \) if \( \theta_0 < Y_n - 1/2 \) or \( Y_1 + 1/2 < \theta_0 \). That is, we would accept \( H_0 \) if \( Y_n - 1/2 < \theta_0 < Y_1 + 1/2 \). Clearly, if \( H_0 : \theta = \theta_0 \) is true, we would never reject \( H_0 \) with this test, and the significance level is \( \alpha = 0 \). However, we must be concerned about the power of the test when \( \theta \neq \theta_0 \). This suggests that we could improve the power if we selected a constant \( c \) slightly less than \( 1/2 \) and accepted \( H_0 \) if \( Y_n - c < \theta_0 < Y_1 + c \); otherwise reject \( H_0 \).

Let us find \( c \) so that the significance level is \( \alpha > 0 \). Thus we want, when \( \theta = \theta_0 \), that

\[
1 - \alpha = P(\theta_0 - c < Y_1 < Y_n < \theta_0 + c)
\]

\[
= P(\theta_0 - c < X_1, X_2, \ldots, X_n < \theta_0 + c)
\]

\[
= \left( \int_{\theta_0 - c}^{\theta_0 + c} 1 \, dx \right)^n = (2c)^n.
\]

Accordingly, \( c = \frac{1}{2} \sqrt{1 - \alpha} \). Going 'backwards' from this test, the corresponding \( 1 - \alpha \) confidence interval for \( \theta \) is \( (Y_n - c, Y_1 + c) \), one which is based upon the joint sufficient statistics for \( \theta \). (In problem 4 of their Chapter VIII, Mood, Graybill, and Boes (1974) ask the student to verify that \( (Y_1, Y_n) \) is a confidence interval, and to find the confidence coefficient. In problem 21 of Chapter VIII, they give the student the more open-ended 'find a good confidence interval' for \( \theta \).

The power of this test is easy for a student to compute by evaluating
\[ K(\theta) = 1 - P(\theta - c < X_1, X_2, \ldots, X_n < \theta + c) = 1 - [P(\theta - c < X_i < \theta + c)]^n. \]

Clearly, when \( \theta_0 - \frac{1}{2} + c < \theta < \theta_0 + \frac{1}{2} - c \), \( K(\theta) = \alpha \). Of course, \( K(\theta) \) is symmetric about \( \theta = \theta_0 \) and hence we need only consider when \( \theta_0 + \frac{1}{2} - c < \theta \). If \( \theta \geq \theta_0 + \frac{1}{2} + c \), then all \( X_i \geq \theta_0 + c \) and \( K(\theta) = 1 \). So when \( \theta_0 + \frac{1}{2} - c < \theta < \theta_0 + \frac{1}{2} + c \),

\[ K(\theta) = 1 - \left( \int_{\theta - 1/2}^{\theta_0 + c} 1 \, dx \right)^n = 1 - (\theta_0 + c + \frac{1}{2} - \theta)^n. \]

6. Possible Likelihood Ratio Exercises When the Range Is Unknown

We present several exercises when considering the uniform distribution on \([\theta - \rho, \theta + \rho]\), where \( \rho \) is unknown. Thus, \( f(x; \theta, \rho) = I_{[\theta-\rho,\theta+\rho]}(x)/(2\rho) \). These are analogous to problems associated with the normal distribution when \( \sigma^2 \) is unknown, and bright students can often think of the two-sample and \( k \)-sample problems on their own (possibly with a little direction from the instructor). In each of the following situations, the student is asked to show that the likelihood ratio \( \lambda \) equals the form given. Students may find these 'easy' or 'hard' depending on their insight.

6.1 One Sample: Test \( H_0 : \theta = \theta_0 \) Against \( H_1 : \theta \neq \theta_0, \rho \) Unknown

Then

\[ \lambda = \left[ \frac{y_n - y_1}{2 \cdot \max(\theta_0 - y_1, y_n - \theta_0)} \right]^n, \]

where \( y_1 \leq y_2 \leq \ldots \leq y_n \) are the order statistics of a random sample of size \( n \).

6.2 One Sample: Test \( H_0 : \rho = \rho_0 \) Against \( H_1 : \rho \neq \rho_0, \theta \) Unknown

Using the same notation as in Exercise 6.1,

\[ \lambda = \left( \frac{y_n - y_1}{2\rho_0} \right)^n. \]

6.3 Two Samples: Test \( H_0 : \rho_1 = \rho_2 \) Against \( H_1 : \rho_1 \neq \rho_2, \theta_1 \) and \( \theta_2 \) Unknown

For convenience, let \( n = n_1 = n_2 \) and \( x_1 \leq x_2 \leq \ldots \leq x_n \) and \( y_1 \leq y_2 \leq \ldots \leq y_n \) be the order statistics of the independent random samples. Then
\[ \lambda = \left( \frac{(x_n - x_1)^n (y_n - y_1)^n}{\max(x_n - x_1, y_n - y_1)^{2n}} \right) = \left( \frac{\min(x_n - x_1, y_n - y_1)}{\max(x_n - x_1, y_n - y_1)} \right)^n. \]

6.4 Two Samples: Test \( H_0 : \theta_1 = \theta_2 \) Against \( H_1 : \theta_1 \neq \theta_2 \), Given That \( \rho_1 = \rho_2 \), but Unknown

Using the same assumptions and notation of Exercise 6.3,
\[ \lambda = \left( \frac{\max(x_n - x_1, y_n - y_1)}{\max(x_n, y_n) - \min(x_1, y_1)} \right)^{2n}. \]

6.5 \( k \) Samples: Test \( H_0 : \rho_1 = \rho_2 = \ldots = \rho_k \) Against All Alternatives, When \( \theta_1, \theta_2, \ldots, \theta_k \) Are Unknown

For convenience, let all \( k \) independent random samples have the same sample size \( n \), and \( x_{1i} \leq x_{2i} \leq \ldots \leq x_{ni} \) be the order statistics of the \( i^{th} \) sample. Then
\[ \lambda = \left( \frac{\Pi_{i=1}^k (x_{ni} - x_{1i})^n}{\max(x_{ni} - x_{1i}; i = 1, 2, \ldots, k)^{kn}} \right). \]

6.6 \( k \) Samples: Test \( H_0 : \theta_1 - \theta_2 - \ldots - \theta_k \) Against All Alternatives, Given That \( \rho_1 = \rho_2 = \ldots = \rho_k \), but Unknown

Using the same assumptions and notation of Exercise 6.5,
\[ \lambda = \left( \frac{\max(x_{ni} - x_{1i}; i = 1, 2, \ldots, k)}{\max(x_{ni}; i - 1, 2, \ldots, k) - \min(x_{1i}; i - 1, 2, \ldots, k)} \right)^{kn}. \]

To determine the significance level of each of the above tests that rejects \( H_0 \) if \( \lambda \leq \lambda_0 \), we need to determine the distribution of \( \lambda \) or some appropriate function of \( \lambda \). These are much more difficult exercises.

**Acknowledgments and Requests**

First we must thank the referees who made great suggestions, and one of whom caught a major error in one of our statements. We certainly appreciate their time and effort. The estimator in the Remark in Section 3.3 is theirs. We also thank the other authors whom we have referenced and who have dabbled with the interesting, one-parameter example. Their work has contributed to this paper.

Although the first author was dismayed when he found out that 'his' wonderful example was already in some of the literature (some published over 40 years ago; see Hogg and Craig (1956) and the 1959 edition of their book), he realized that he was not the only person to think this example was a good one. He then contacted the second author about presenting many of the applications of these uniform distributions in one document. When the first author later found the example in Mood, Graybill, and Boes (1974), a book he had used at least eleven years prior to his 'creating' the example, he almost thought that he had subconsciously plagiarized! However,
with the continued rise in importance of statistics education and the fact that no single textbook has used this uniform distribution to any extent, the authors felt the project was still worthwhile.

The authors have one request: readers who find uniform distributions of the type considered here, referenced in any book or article, should contact the first author with the references. This is especially true of those who have written the book or article themselves! We have looked at many texts, but certainly not all of them.

References

Casella, G., and Berger, R. L. (1990), *Statistical Inference*, Belmont, CA: Duxbury.

Hogg, R. V., and Craig, A. T. (1956), "Sufficient Statistics In Elementary Distribution Theory," *Sankhya*, 17, 209.

----- (1995), *Introduction to Mathematical Statistics* (5th ed.), Englewood Cliffs, NJ: Prentice Hall.

Hogg, R. V., and Tanis, E. A. (1997), *Probability and Statistical Inference* (5th ed.), Englewood Cliffs, NJ: Prentice Hall.

Miller, I., and Miller, M. (1999), *John E. Freund's Mathematical Statistics* (6th ed.), Upper Saddle River, NJ: Prentice Hall.

Mood, A. M., Graybill, F. A., and Boes, D. C. (1974), *Introduction to the Theory of Statistics* (3rd ed.), NY: McGraw-Hill.

Rice, J. A. (1995), *Mathematical Statistics and Data Analysis* (2nd ed.), Belmont, CA: Duxbury.

Dexter C. Whittinghill
Department of Mathematics
Rowan University
201 Mullica Hill Rd.
Glassboro, NJ 08028
whittinghill@rowan.edu

Robert V. Hogg
Department of Statistics and Actuarial Science
University of Iowa
Iowa City, IA 52242