Note

Application of the Simple Saddlepoint Approximation to Estimate Probability Distributions in Wildlife Research

STEPHEN J. DEMASO,1 Caesar Kleberg Wildlife Research Institute, Texas A&M University-Kingsville, MSC 218, Kingsville, TX 78363-8202, USA

FIDEL HERNÁNDEZ,2 Caesar Kleberg Wildlife Research Institute, Texas A&M University-Kingsville, MSC 218, Kingsville, TX 78363-8202, USA

LEONARD A. BRENNAN Caesar Kleberg Wildlife Research Institute, Texas A&M University-Kingsville, MSC 218, Kingsville, TX 78363-8202, USA

RALPH L. BINGHAM,3 Department of Mathematics, Texas A&M University-Kingsville, MSC 172, Kingsville, TX 78363-8202, USA

ABSTRACT The simple saddlepoint approximation (SSA) uses the mean, variance, and skewness (a measure of the asymmetry of the distribution) of a data set to algebraically approximate the probability density function of a selected variable. We compared habitat-suitability bounds estimated with SSAs and continuous selection functions. Habitat-suitability bounds for bobwhite nesting based on the SSA method were biologically comparable to the results of the method based on continuous selection functions. The SSA approach allows habitat-suitability bounds to be estimated using algebra and can be calculated in computer spreadsheets. © 2011 The Wildlife Society.

KEY WORDS Colinus virginianus, habitat suitability, northern bobwhite, probability distribution, simple saddlepoint approximation, skewness, variance.

Estimating probability distributions of demographic variables is an important aspect of wildlife research. For example, many wildlife studies involving simulation modeling often require estimates of probability distributions to invoke stochasticity. Accurately characterizing the underlying probability distribution for a particular variable, however, can be difficult because many studies have small sample sizes. Consequently, a variety of methods have been used to estimate probability distributions including bootstrapping (Fieberg and Ellner 2001), using a uniform distribution and truncating values at the minimum and maximum (Sandercock et al. 2008), or matching moments to develop an approximate convolutions of commonly used distributions (Huzurbazar 1999). The SSA also has been used to approximate the equilibrium distribution (i.e., the carrying capacity) of a

Therefore, skewness is an additional measure that provides valuable information for describing distributions (J. H. Matis, Department of Statistics, Texas A&M University, personal communication).

The SSA uses the mean, variance, and skewness to find a general (i.e., family free) saddlepoint approximation (SA) of a probability distribution (i.e., an approximation of the probability density function [pdf]). Renshaw (1998) was the first to incorporate the third cumulant \( k_3 \) into the SA. This formula is general, can be evaluated algebraically, and provides a considerable improvement over the normal approximation (i.e., \( k_3 = 0 \)) because it incorporates skewness (Renshaw 2000). Being a completely general technique, it does not require selection of an assumed underlying distribution fitted according to some statistical goodness-of-fit criteria (Renshaw 2000). We refer the reader to Reid (1998), Jensen (1995), Kolassa (1997), and Goutis and Casell (1999) for thorough reviews of the theory and derivation of SAs used to obtain accurate expressions for probability distributions.

Saddlepoint approximations have been a useful theoretical tool for describing probability distributions in statistics since the mid 1950s (Daniels 1954, Daniels 1987, Goutis and Casell 1999, Huzurbazar 1999). Recently, the availability of personal computers has made SAs accessible to the broader research community. Saddlepoint approximations have been used to obtain accurate expressions for probability distributions (Goutis and Casell 1999), compute \( P \)-values for some test statistics, approximate finite mixture distributions, and approximate convolutions of commonly used distributions (Huzurbazar 1999). The SSA has also been used to approximate the equilibrium distribution (i.e., the carrying capacity) of a

Received: 30 March 2010; Accepted: 11 August 2010

1Present Address: Gulf Coast Joint Venture, U.S. Fish and Wildlife Service, 700 Cajundome Blvd., Lafayette, LA 70506, USA.
2E-mail: fidel.hernandez@tamuk.edu
3Deceased.
stochastic-logistic model of population growth (Matis et al. 2003).
To our knowledge, SSA has been rarely if ever used in wildlife science. However, we believe the simple saddlepoint approximation has the potential for applications in wildlife research. Our goal was to show how the SSA can be used as an analytical tool in wildlife research. Specifically, our objective was to present an intuitive description of the SSA and illustrate its application in wildlife research using long-term precipitation data and nesting-ecology data from a northern bobwhite (Colinus virginianus) radio telemetry project in south Texas. We used SSAs to estimate 1) probability distributions of July precipitation and bobwhite clutch size and 2) habitat-suitability bounds for bobwhite nesting cover on semiarid rangelands.

STUDY AREA

Data were collected on a private hunting lease on the Encino Division of the King Ranch, Brooks County, Texas situated within the South Texas Plains ecoregion (Gould 1975). This ecoregion was characterized by level to rolling land dissected by streams flowing into the Rio Grande or the Gulf of Mexico (Scifres 1980). Climatic conditions were classified as semi-arid, sub-humid, and with a high rate of evaporation (Williamson 1983). Long-term monthly precipitation ranged from 1.4 cm to 13.0 cm with a mean annual rainfall of 65.4 cm, with June and September receiving the most rainfall (Williamson 1983). January was the coldest month (\(\bar{x} = 13.1^\circ C\); min. = 6.3\(^\circ C\), max. = 19.9\(^\circ C\)), and July was the hottest month (\(\bar{x} = 29.8^\circ C\); min. = 23.1\(^\circ C\), max. = 36.4\(^\circ C\); Williamson 1983). For more specific details regarding the study area, see Arredondo et al. (2007). The telemetry project monitored radiomarked bobwhites year round in each of three spatially independent study sites. Study sites were 1,200–2,000 ha and separated by \(>5\) km. Study areas were \(>200\) km from the study site, and \(8\) km from research sites.

METHODS

We chose the variables used in our examples because they represent key components of northern bobwhite demography and habitat. We used July precipitation data from 1895 to 2006 for the southern climatic division (code 09) of Texas (National Climate Data Center, NCDC, 2007). We obtained data on bobwhite clutch size (number of eggs in a nest after incubation began, regardless of nest fate) and habitat structure around nests and random locations near these nests, from a radio telemetry project in Brooks County, Texas conducted from 2001 through 2005.

We estimated population mean (\(\mu\)), variance (\(\sigma^2\)), and skewness (\(k_3\)) using the sample mean (\(\bar{x}\)), sample variance (\(s^2\)), and sample skewness (\(k_3\)), respectively, using PROC UNIVARIATE (Institute SAS Inc. 2006). We calculated sample skewness as follows:

\[
\hat{k}_3 = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} (x_i - \bar{x})^3 / s^3
\]

where \(n\) is the number of non-missing values for the variable, \(x_i\) is the \(i\)th value of the variable, \(\bar{x}\) is the sample mean, and \(s\) is the sample standard deviation.

We calculated SSAs for each variable by substituting the values of interest (i.e., of \(x\)), along with \(\bar{x}, s^2\), and \(k_3\) (Table 1) for \(\mu, \sigma^2,\) and \(k_3\), respectively, for each variable into the SSA formula:

\[
f(x) = (4\pi^2\psi)^{-1/4} \exp\left\{-\left(\sigma^4 - 3\sigma^2\psi + 2\psi^{3/2}\right) / 6k_3^2\right\}
\]

where,

\[
\psi = \sigma^4 + 2k_3(x-\mu)
\]

Renshaw 1998, Matis et al. 2003). One limitation of the SSA is that it exists only for \(\psi \geq 0\), which in turn restricts the range of \(x\). From the equation for

<table>
<thead>
<tr>
<th>Variable</th>
<th>(n)</th>
<th>(\bar{x})</th>
<th>(s^2)</th>
<th>(k_3)</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jul precipitation (1895–2006)</td>
<td>112</td>
<td>4.41</td>
<td>18.60</td>
<td>2.14</td>
<td>0.00</td>
<td>25.30</td>
</tr>
<tr>
<td>56</td>
<td>4.96</td>
<td>18.69</td>
<td>1.70</td>
<td>0.08</td>
<td>20.12</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>5.11</td>
<td>26.10</td>
<td>2.83</td>
<td>0.58</td>
<td>15.60</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>3.61</td>
<td>14.02</td>
<td>0.54</td>
<td>0.08</td>
<td>10.00</td>
<td></td>
</tr>
<tr>
<td>Clutch size</td>
<td>232</td>
<td>11.99</td>
<td>7.79</td>
<td>0.04</td>
<td>1.00</td>
<td>22.00</td>
</tr>
<tr>
<td>116</td>
<td>11.59</td>
<td>6.68</td>
<td>0.07</td>
<td>1.00</td>
<td>19.00</td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>12.10</td>
<td>6.55</td>
<td>0.51</td>
<td>1.00</td>
<td>19.00</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>11.72</td>
<td>6.29</td>
<td>0.89</td>
<td>7.00</td>
<td>19.00</td>
<td></td>
</tr>
<tr>
<td>Bunchgrass diameter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nest</td>
<td>105</td>
<td>29.59</td>
<td>78.61</td>
<td>2.22</td>
<td>15.00</td>
<td>75.50</td>
</tr>
<tr>
<td>Random</td>
<td>187</td>
<td>22.70</td>
<td>53.83</td>
<td>0.78</td>
<td>8.50</td>
<td>48.00</td>
</tr>
<tr>
<td>Bunchgrass height</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nest</td>
<td>105</td>
<td>23.69</td>
<td>44.74</td>
<td>0.89</td>
<td>9.00</td>
<td>48.00</td>
</tr>
<tr>
<td>Random</td>
<td>187</td>
<td>23.29</td>
<td>54.62</td>
<td>0.54</td>
<td>6.00</td>
<td>46.00</td>
</tr>
<tr>
<td>Herbaceous canopy coverage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nest</td>
<td>105</td>
<td>52.15</td>
<td>370.12</td>
<td>-0.05</td>
<td>13.80</td>
<td>90.00</td>
</tr>
<tr>
<td>Random</td>
<td>203</td>
<td>42.10</td>
<td>483.21</td>
<td>0.75</td>
<td>5.00</td>
<td>100.00</td>
</tr>
<tr>
<td>Radius of complete visual obstruction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nest</td>
<td>105</td>
<td>3.33</td>
<td>2.46</td>
<td>1.49</td>
<td>0.60</td>
<td>9.50</td>
</tr>
<tr>
<td>Random</td>
<td>204</td>
<td>5.47</td>
<td>13.90</td>
<td>1.64</td>
<td>0.60</td>
<td>22.80</td>
</tr>
</tbody>
</table>

Table 1. Sample size (\(n\)), mean (\(\bar{x}\)), sample variance (\(s^2\)), sample skewness (\(k_3\)), and range of values for July precipitation (cm), northern bobwhite clutch size, nest width (cm), nest height (cm), percent herbaceous canopy coverage (%), radius of complete visual obstruction (m), Brooks County, south Texas, 2001–2005. Nest width, height, percent herbaceous canopy coverage, and radius of complete visual obstruction data is from Arredondo et al. (2007).

DeMaso et al. • Simple Saddlepoint Approximation
$\psi$, the range for which the SSA exists is

\[ x > \mu - \sigma^2 / 2k_3 \quad \text{for} \quad k_3 > 0, \]
\[ x < \mu - \sigma^2 / 2k_3 \quad \text{for} \quad k_3 < 0 \quad (\text{Matis et al. 2003}). \]

To examine the effect of sample size on SSAs we estimated SSAs for each variable (Jul precipitation and clutch size) with different sample sizes, which we randomly drew, without replacement, from the total number of observations for each variable.

We used the SSA to obtain the respective pdfs $f(x)$ for nest sites and $g(x)$ for random points. We then calculated a continuous selection function as $u(x) = f(x)/g(x)$ for $g(x) > 0$ (Kopp et al. 1998). We interpret the continuous selection function the same as the discrete selection ratio, defined as proportional use divided by proportional availability. A selection function value of $u(x) > 1$ indicates selection, or use greater than availability. A value of $u(x) < 1$ indicates avoidance, or use less than availability; a value of $u(x) = 1$ indicates no difference in proportional use and proportional availability (Kopp et al. 1998). We interpret random use with respect to a habitat variable as $u(x) = 1$ for all domain $x$.

**RESULTS**

From 1895 to 2006, July precipitation ranged from 0 cm to 25.3 cm, with $\mu = 4.4$ cm (Table 1). Based on the frequency

![Figure 1. Histogram of (A) July precipitation (cm), southern Climatic Division (09), Texas, 1895–2006 and (B) northern bobwhite clutch size, Brooks County, Texas, 2001–2005.](image)
distribution for July precipitation, the SSA appeared to be a reasonable approximation of the underlying probability distribution (Fig. 1). During 1 March 2001 through 28 February 2005, we monitored 337 nesting attempts of radio-marked bobwhites on our study area. We estimated clutch size from 232 bobwhite nests, ranging from 1 to 22, with a mean of 11.99 eggs per clutch (Table 1). Based on the frequency distribution for bobwhite clutch size, the SSA appeared to be a reasonable approximation of the underlying probability distribution (Fig. 2). The probability distribution of clutch size (Fig. 2) was nearly symmetrical. Histograms in addition to descriptive statistics show the skewness of the variables,
which provide additional information about the variables and their probability distributions (Table 1, Fig. 3).

Depending on sample size, mean July precipitation varied from 3.61 cm to 5.11 cm. Mean clutch size varied from 11.6 to 12.1, depending on sample size. Sample skewness ($k_3$) was highest for smaller sample sizes. The probability distribution of July precipitation (Fig. 2) was skewed right. Sample size appeared to have a minimum effect on the SSAs (Fig. 2). The SSA selection function indicated that bobwhites selected for bunchgrass height 18.0–33.0 cm, bunchgrass diameter ≥28.0 cm, herbaceous canopy cover ≥42%, and a radius of complete visual obstruction of 2.00–5.00 m (Fig. 3, Table 2).

**DISCUSSION**

The SSA has the potential to provide additional perspective on the data structure of specific variables, especially when sample sizes are small and distributions are not normal or close to normal. Our sample sizes for July precipitation and bobwhite clutch size were large ($\geq$112). Accuracy of approximations from saddlepoint methodology can hold for small sample sizes and can be accurate in the tails of a distribution (Huzurbazar 1999). The SSA can be more accurate than the normal approximation, even when skewness is small (Matis et al. 2003). In cases of larger relative skewness, as in Renshaw’s (1998) Poisson example, the SSA has even greater utility as it incorporates skewness, which is appealing.
considering that wildlife data may be skewed and often is based on small sample sizes.

Our estimates of habitat-suitability bounds for bobwhite nests using the SSA approach were biologically comparable to those reported by Arredondo et al. (2007). The shapes of three of four selection function graphs (Fig. 3A,C,D) were similar to those presented by Arredondo et al. (2007, figs. 1, 4, 5). Only the shape of the selection function for bunchgrass diameter differed markedly between the SSA and continuous selection function approach of estimating habitat-suitability bounds (Guthery 1997, Kopp et al. 1998). The SSA selection function (Fig. 3B) for bunchgrass diameter increased exponentially, whereas the selection function reported by Arredondo et al. (2007, fig. 2) peaked between 30 cm and 40 cm and then gradually declined. However, both methods resulted in similar lower bounds for selection. One possible explanation for this difference in shapes of selection functions is that different methods were used to estimate $f(x)$ and $g(x)$.

Even though we could not make a direct statistical comparison to the data presented by Arredondo et al. (2007), the shapes of the selection functions we presented appear similar to those based on empirical data presented by Arredondo et al. (2007) and exhibit approximately the same lower and upper bounds. Because knowledge cannot be proven correct, only proven false, we conclude that the simple saddlepoint selection functions we presented are no better or worse than those presented by Arredondo et al. (2007) and that the simple saddlepoint approach is less complex.

Advantages of the SSA include: 1) it is simple and easy to calculate, 2) it works well with small sample sizes, 3) it uses skewness, which is a useful measure of asymmetry and most ecological data have skewed distributions (Young and Young 1998), therefore, skewness is useful for approximating distributions, 4) it provides another way to describe univariate data, and 5) it can be used to estimate selection functions. Disadvantages include: 1) SSAs can only be calculated if $\psi \geq 0$, which restricts the range of $x$ and 2) sensitivity of SSAs to sampling error is unknown. The SSA should work well for any discrete or continuous variable that has a mean, variance, and skewness. However, common sense needs to be used; for example, clutch size can only be a positive integer within the biological reasonable range. It would not work well for categorical values. Additionally, researchers should keep in mind that the SSA should not be extrapolated outside the range of their data.

Use of SSAs in wildlife research could be augmented by computer code that would allow values to be randomly drawn from saddlepoint distributions. Consequently, the SSA would improve stochastic habitat and population models for many species. Future research should address the issue of sampling errors, which could be examined through simulation, assuming a known distribution, taking samples of a given size, finding the corresponding sample cumulants, and then finding the saddlepoint approximations. Also, developing a method for calculating the error bars associated with SSAs would be useful in addressing the sampling error issue.

**Management Implications**

In wildlife science, it is axiomatic that management decisions be based on reliable knowledge. Small sample sizes, highly skewed and variable data, and various other limitations combine to make sound inferences, and thus reliable knowledge, a challenge. The SSA is an analytical tool that wildlife scientists can use to gain insight into the probability distributions of important variables, especially those based on small sample sizes. The SSA is easy to use, assumes no underlying distribution, and can be used to estimate a pdf, as shown in our examples. The SSA can be calculated and graphed using most computer spreadsheets. The SSA can also be used in stochastic simulation modeling, where values are randomly drawn from the probability distribution of a given variable.

**Acknowledgments**

On 28 May 2010, Dr. Ralph L. Bingham passed away while fishing in Baffin Bay along the Texas Gulf Coast. Ralph was a teacher, mentor, colleague, and a friend to many of us who knew him. We will miss him. We thank the Caesar Kleberg Wildlife Research Institute and Texas A&M University-Kingsville for providing financial and logistical support. Cooperative funding was provided by the Texas State Council, South Texas, Houston, East Texas, and Alamo Chapters of Quail Unlimited; George and Mary Josephine Hamman Foundation; Robert J. Kleberg, Jr. and Helen C. Kleberg Foundation; Amy Shelton McNutt Charitable Trust; Amy Shelton McNutt Memorial Fund; Bob and Vivian Smith Foundation; and The William A. and Madeline Welder Smith Foundation. We thank P. Doherty, F. S. Guthery, A. R. Litt, M. J. Peterson, J. P. Sands, and J. H. Matis for providing helpful comments on earlier drafts of this manuscript. This manuscript is Caesar Kleberg Wildlife Research Institute publication number 10–121.

**LITERATURE CITED**


Gould, F. W. 1975. Texas plants: a checklist and ecological summary. Texas A&M University, Agricultural Experiment Station, College Station, USA (MP-555/Revision).


Associate Editor: Christopher K. Williams.