Ann. Inst. Statist. Math.
Vol. 42, No. 1, 115–131 (1990)

SADDLEPOINT APPROXIMATIONS IN RESAMPLING ANALYSIS

SUOJIN WANG*

Department of Statistical Science, Southern Methodist University, Dallas, TX 75275, U.S.A.

(Received May 26, 1988; revised May 29, 1989)

Abstract. The technical validity of the saddlepoint technique for approximating the resampling distribution of the sample mean of i.i.d. and non-i.i.d. random variables is examined. The relative errors are shown to occur at the same rate as in parametric analysis. Discreteness in resampling problems is accounted for. The uniformity of the errors is also explored.

Key words and phrases: Asymptotic expansion, bootstrap, discreteness, randomization.

1. Introduction

Assume that $X_1,\dots,X_n$ are independently and identically distributed (i.i.d.) random variables with distribution F and that the moment generating function (MGF)

$$M(t)=\int_{-\infty}^{\infty} e^{tx}dF(x) \tag{1.1}$$

converges for t in a non-trivial real interval I containing the origin. In statistical analysis, we often need to know the distribution and density, when it exists, of a linear statistic such as the sample mean $\bar{X}=n^{-1}\sum_1^n X_i$. The exact formulas are often intractable. In such cases close approximations are desirable. Saddlepoint approximations provide an excellent method to achieve this goal.

Daniels (1954) derived the saddlepoint formula for the density h_n of $\bar{X}$ as

$$h_s(x)=\{n/[2\pi K''(t_x)]\}^{1/2}\exp\{n[K(t_x)-t_x x]\}\,, \tag{1.2}$$

*This work was completed while the author was at the University of Texas at Austin.

where $K(t) = \log\{M(t)\}$ is the cumulant generating function (CGF) and t_x is the unique solution to $K'(t) = x$, provided that the solution exists.

The saddlepoint technique can also be applied to obtain an approximation of the cumulative distribution H_n of $\bar{X}$, or equivalently the tail probability $G_n = 1 - H_n$:

$$(1.3)\qquad G_s(x) = \begin{cases} 1 - \Phi(w) + \phi(w)(z^{-1} - w^{-1}), & \text{if} \quad x \neq E(X)\,, \\ \dfrac{1}{2} - \dfrac{1}{6}(2\pi n)^{-1/2} K^{(3)}(0)/\{K''(0)\}^{3/2}, & \text{if} \quad x = E(X)\,, \end{cases}$$

where $w = \{2n[t_x x - K(t_x)]\}^{1/2}\,\mathrm{sgn}\,(t_x)$, $z = t_x\{nK''(t_x)\}^{1/2}$, and Φ and ϕ are the standard normal distribution and density, respectively. The relative errors in (1.2) and (1.3) are of order n^{-1}. Moreover, they are uniform over all x in a fairly large family of distributions (see Jensen (1988) for a rigorous discussion). Formula (1.3) is due to Lugannani and Rice (1980). The successive saddlepoint approximations extending (1.2) and (1.3) to reduce the relative errors were also obtained by Daniels (1954) and Lugannani and Rice (1980), respectively (see Daniels (1987)).

For discrete random variables, results similar to (1.2) and (1.3) can be derived with slight modifications. In particular, $G_s(x)$ is as in (1.3) with z replaced by $z = \{1 - \exp(-t_x)\}\{nK''(t_x)\}^{1/2}$.

Saddlepoint approximations have shown great accuracy, more so than normal approximations and Edgeworth approximations. In most practical applications, the first-order approximations are sufficiently accurate. Therefore they are used in a large number of statistical problems (see, for example, Barndorff-Nielsen and Cox (1979) and Daniels (1983)). Reid (1988) gives an excellent review of this topic.

One obvious limitation in using the above formulas is that the MGF is supposed to be known, except for possible unknown parameters. This limitation has restricted most applications in the parametric framework. The first application of the saddlepoint technique in nonparametric analysis appeared in Robinson (1982), who used the method to approximate permutation distributions. Davison and Hinkley (1988) apply saddlepoint approximations in a number of bootstrap and randomization problems. They give no proofs, leaving open the question of technical validity of the approximations when F is replaced by the discrete, but consistent, empirical distribution function (EDF) $\tilde{F}$.

The object of this paper is to study the technical validity of saddlepoint approximations in resampling analysis from the viewpoint of asymptotic theory. The simple statistic sample mean is considered throughout the paper. Note that unlike in parametric analysis, here the sampled distribution $\tilde{F}$, as well as the resampled statistic, depends on n. Such dependence raises the question of the applicability of saddlepoint approximations.

Section 2 shows the details of the proofs for the distribution function of $\bar{X}$ for discrete and continuous variables in bootstrap analysis. The corresponding approximations in randomization analysis are proved to be valid in Section 3. The relative errors in the saddlepoint approximations are not uniform, in contrast to the standard saddlepoint approximation theory, as shown in Section 4. We briefly discuss generalizations in Section 5.

2. Bootstrap distribution function of $\bar{X} - E(X)$

Let $T = t(X_1, \ldots, X_n)$ be an appropriate estimate of some parameter $\theta = t(F)$; e.g., if θ is the population mean μ, then a natural choice of T is $T = \bar{X} = n^{-1} \sum_1^n X_i$. Because the distribution F of X is unknown, we will use the EDF $\tilde{F}$ as an approximation to F, where

$$\tilde{F}(x) = \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i \leq x\}} . \tag{2.1}$$

The following is the basic non-parametric bootstrap idea. Assume that we can write $T = t(\tilde{F})$ as usual. Then the tail probability

$$G(x) = \Pr\left(t(\tilde{F}) - t(F) \geq x \mid F\right) \tag{2.2}$$

is approximated by

$$\tilde{G}(x) = \Pr\left(t(\tilde{F}^*) - t(\tilde{F}) \geq x \mid \tilde{F}\right) , \tag{2.3}$$

where $\tilde{F}^*$ is the EDF of $X_1^*, \ldots, X_n^*$, which is a sample from $\tilde{F}$ under a uniform resampling plan (see Efron (1982) for details). The distribution function defined by $H(x) = \Pr\left(t(\tilde{F}) - t(F) < x \mid F\right)$ is hence approximated by the bootstrap distribution function $\tilde{H}(x) = 1 - \tilde{G}(x)$. The accuracy of $\tilde{G}(x)$ to $G(x)$ has been explored by many authors (see, for example, Singh (1981)). We will not go over this topic here.

$\tilde{G}(x)$ in (2.3) can be approximated by simulation up to any accuracy. One substantial problem is that the bootstrap technique usually requires a large number of simulations. Avoiding this difficulty, the saddlepoint method provides an alternative way to approximate $\tilde{G}(x)$ in certain circumstances.

In this paper, we consider $t(F) = E(X) = \mu$ and $t(\tilde{F}) = n^{-1} \sum_1^n X_i$. Let $Y_i^* = X_i^* - \bar{X}$ and

$$\tilde{M}(t) = \exp\{\tilde{K}(t)\} = n^{-1} \sum_1^n \exp\{(X_i - \bar{X})t\} \tag{2.4}$$

be the MGF of Y_i^* with CGF $\tilde{K}$. The following lemmas, together with Lemma 3.1 in Section 3, are very important in our derivations. Lemma 2.1 generalizes the truncated form of Watson's lemma.

LEMMA 2.1. *Assume that $\{A_n\}$ and $\{B_n\}$ are two sequences of random variables such that $A_n \to A$, $B_n \to B$ in probability, where $A > 0$, $B > 0$ are constants, and that $\{h_n(x)\}$ is a series of stochastic functions satisfying*

(a) *each $h_n(x)$ is differentiable infinitely many times in a closed neighborhood U of $x = 0$;*

(b) $\max_{x \in U} |h_n^{(r)}(x)| = O_p(1)$ *for each* $r = 1, 2, \ldots$;

(c) $\max_{x \in [-A_n, B_n]} |h_n(x)| = O_p(1)$.

Then, for $k = 1, 2, \ldots,$

$$\text{(2.5)} \qquad \{n/(2\pi)\}^{1/2} \int_{-A_n}^{B_n} e^{-nx^2/2} h_n(x)\,dx$$
$$= h_n(0) + \frac{1}{2n} h_n''(0) + \cdots + \frac{1}{(2n)^k k!} h_n^{(2k)}(0) + O_p(n^{-(k+1)}) .$$

PROOF. See Appendix A.

In particular, A_n, B_n and $h_n(x)$ can be fixed numbers, i.e., non-stochastic for fixed n and x. Notice that the maximum in condition (c) is defined on $[-A_n, B_n]$ instead of U. Adding this condition is to include the case where U does not contain $[-A_n, B_n]$, so that one only needs to check the existence of U. No relationship between U and $[-A_n, B_n]$ is needed. A counterexample can be constructed to show that the lemma does not hold if the maximum in condition (c) is defined on U. The main difference between this lemma and Watson's lemma is that $h_n(x)$ here depends on n while the counterpart in Watson's lemma is independent of n.

LEMMA 2.2. *Let $C = [-b_1, -c_1] \cup [c_2, b_2]$ with $b_j > c_j > 0$, $j = 1, 2$. If $d \in I$ defined in* (1.1), *then*

$$\text{(2.6)} \qquad \max_{y \in C} |\tilde{M}(d + iy)/\tilde{M}(d)|^n = O_p(\rho^n)$$

with $0 \leq \rho < 1$, where C is restricted within the principal period of $\tilde{M}$ if X is discrete.

PROOF. See Appendix B.

The proof given in Appendix B is similar to the classical proof of the consistency of the maximum likelihood estimator (but in a different space).

2.1 *Discrete variables*

Suppose that the X's are discrete and take integer values only. Without loss of generality, we may exclude the case that $\Pr(X = r) = 0$ except at multiples of integers bigger than 1. Then

$$\tilde{g}(r/n) = \Pr(\bar{Y}^* = r/n|\tilde{F}) = \frac{1}{2\pi i}\int_{-i\pi}^{i\pi} e^{n[\tilde{K}(t)-tr/n]}dt \tag{2.7}$$

for any integer r. Thus the tail probability $\Pr(\bar{Y}^* \geq c|\tilde{F})$ is

$$\tilde{G}(c) = \sum_{r\geq nc} \tilde{g}(r/n) = \frac{1}{2\pi i}\int_{d-i\pi}^{d+i\pi} e^{n[\tilde{K}(t)-tc/n]}/(1-e^{-t})dt \tag{2.8}$$

for any $d > 0$, where c is such that nc is an integer. Here we apply the method of steepest descents (see Jeffreys and Jeffreys (1962), Chapter 17) to approximate the above integral.

Assume that t_c^0 and t_c are the solutions to $K'(t) = c$ and $\tilde{K}'(t) = c$, respectively. (It is easily seen that t_c exists and is unique for $c \in (X_{(1)}, X_{(n)})$, where $X_{(1)}, \ldots, X_{(n)}$ are the ordered statistics.) From the theory of classic saddlepoint approximations (Daniels (1954)), there is a contour L_0 perpendicular to the real axis at $t = t_c^0$, such that $K(t) - tc/n$ is real and descents most quickly on L_0. Moreover, there exist $\alpha > 0$, $\beta > 0$ such that L_0 and the straight line from $t_c^0 - i\pi$ to $t_c^0 + i\pi$ intersect at $t_c^0 - i\alpha$ and $t_c^0 + i\beta$. Similarly, there are counterparts L_n, $\alpha_n > 0$ and $\beta_n > 0$ for $\tilde{K}(t) - tc/n$ corresponding to L_0, α and β. Because there may be more than two intersection points, we assume $t_c - i\alpha_n$ and $t_c + i\beta_n$ are the points with minimum $|\alpha_n - \alpha|$ and $|\beta_n - \beta|$. It can easily be shown that $\alpha_n \to \alpha$, $\beta_n \to \beta$ in probability. First we assume $c > 0$ and then $t_c > 0$. Let $d = t_c$ in (2.8). We say that an error E_n is exponentially small if $E_n = O_p(\rho^n)$ for $0 \leq \rho < 1$. Ignoring an exponentially small relative error, the contour of the integral in (2.8) can be replaced by L_n without affecting the integral. Let w be a new variable, so that

$$\frac{1}{2}(w - w_c)^2 = \tilde{K}(t) - tc - [\tilde{K}(t_c) - t_c c]$$

and $\operatorname{sgn}(\operatorname{Im}(w - w_c)) = \operatorname{sgn}(\operatorname{Im}(t - t_c))$, where $w_c = \{2[t_c c - \tilde{K}(t_c)]\}^{1/2}\operatorname{sgn}(t_c)$. The transformation $w = w(t)$ maps L_n from $t_c - i\alpha_n$ to $t_c + i\beta_n$ onto the straight line $C(w_c) = \{w|w = w_c + iy, y \in [-a_n, b_n]\}$. Thus

$$\begin{aligned}\tilde{G}(c) &= \frac{1}{2\pi i}\int_{w_c-ia_n}^{w_c+ib_n} e^{n[w^2/2-w_c w]}(1-e^{-t})^{-1}\frac{dt}{dw}dw \\ &= \frac{1}{2\pi i}\int_{w_c-ia_n}^{w_c+ib_n} e^{n[w^2/2-w_c w]}\frac{dw}{w}\end{aligned} \tag{2.9}$$

$$+\frac{1}{2\pi i}\int_{w_c-ia_n}^{w_c+ib_n} e^{n[w^2/2-w_cw]}\{(1-e^{-t})^{-1}dt/dw-1/w\}dw$$
$$=I_1+I_2\,.$$

Clearly, $a_n \to a>0$, $b_n \to b>0$ in probability. We may assume $a\le b$. Then

$$\text{(2.10)}\qquad I_1=\frac{1}{2\pi i}\int_{w_c-i\infty}^{w_c+i\infty} e^{n[w^2/2-w_cw]}\frac{dw}{w}\{1+O_p(e^{-na^2/4})\}$$
$$=\{1-\Phi(\sqrt{n}\,w_c)\}\{1+O_p(e^{-na^2/4})\}\,.$$

On contour $C(w_c)$ write $\Psi_n(w)=(1-e^{-t})^{-1}dt/dw-1/w=\Psi_{n1}(y)+i\Psi_{n2}(y)$, where Ψ_{n1} and Ψ_{n2} are real functions. In a closed neighborhood $V(w_c)$ of $w=w_c$ on the complex plane that may include $w=0$, $\Psi_n(w)$ is analytic. Let U be the projection of $C(w_c)\cap V(w_c)$ on the y-axis so that $U\subset[-a_n,b_n]$. Then $\Psi_{n1}(y)$ satisfies conditions (a) and (b) of Lemma 2.1 by the strong law of large numbers, and

$$d^k\Psi_{n1}(y)/dy^k|_{y=0}\,i^k=d^k\Psi_n(w)/dw^k|_{w=w_c}=\Psi_n^{(k)}(w_c)$$

for $k=0,1,2,\ldots$, since the imaginary part of $\Psi_n(w)$ is zero on the real axis. It is not necessary to check if U is $[-a_n,b_n]$ or not. Furthermore, it is easily seen that Ψ_{n1} satisfies condition (c) also. By Lemma 2.1, for $r=1,2,\ldots$,

$$\text{(2.11)}\quad I_2=\frac{1}{2\pi}\int_{-a_n}^{b_n} e^{-ny^2/2}\Psi_{n1}(y)\,dy\,\phi(\sqrt{n}\,w_c)$$
$$=\phi(\sqrt{n}\,w_c)\{\tilde{d}_0n^{-1/2}+\tilde{d}_1n^{-3/2}+\cdots+\tilde{d}_rn^{-(2n+1)/2}+O_p(n^{-(r+1)})\}\,,$$

where

$$\tilde{d}_0=z_c^{-1}-w_c^{-1},\qquad \tilde{d}_j=\Psi_n^{(2j)}(w_c)/\{(-2)^jj!\},\qquad j=1,\ldots,r\,,$$

and

$$z_c=(1-e^{-t_c})\{\tilde{K}''(t_c)\}^{1/2}\,.$$

Therefore,

$$\text{(2.12)}\qquad \tilde{G}(c)=\{1-\Phi(\sqrt{n}\,w_c)$$
$$+\phi(\sqrt{n}\,w_c)(\tilde{d}_0n^{-1/2}+\tilde{d}_1n^{-3/2}+\cdots+\tilde{d}_rn^{-(2r+1)/2})\}$$
$$\cdot\{1+O_p(n^{-(r+1)})\}\,.$$

Formula (2.12) can be similarly shown to be valid when $c < 0$. It follows that the saddlepoint approximation

$$(2.13)\quad \tilde{G}_s(c) = \begin{cases} 1 - \Phi(\tilde{a}_c) + \phi(\tilde{a}_c)(\tilde{b}_c^{-1} - \tilde{a}_c^{-1}), & \text{if } c \neq 0 \\ \frac{1}{2} - \frac{1}{6}(2\pi n)^{-1/2}\tilde{K}^{(3)}(0)/\{\tilde{K}''(0)\}^{3/2} + \frac{1}{2}\{2\pi n\tilde{K}''(0)\}^{-1/2}, & \text{if } c = 0 \end{cases}$$

with $\tilde{a}_c = \sqrt{n}\, w_c$, $\tilde{b}_c = t_c\{nK''(t_c)\}^{1/2}$, satisfies

$$(2.14)\qquad \tilde{G}(c) = \tilde{G}_s(c)\{1 + O_p(n^{-1})\}\,.$$

These results are parallel to those in parametric analysis.

2.2 *Continuous variables*

Now assume that the unknown distribution F is continuous. First round data $X_1, \ldots, X_n$ to m decimal places and denote the results by $X_1^{(m)}, \ldots, X_n^{(m)}$: these are discrete with steps of 10^{-m}. Let

$$\tilde{K}(m, t) = \log\left(n^{-1} \sum_1^n \exp\{(X_j^{(m)} - \bar{X}^{(m)})t\} \right).$$

Because $10^m X_j^{(m)}$ is an integer for each j, a simple transformation gives as an approximation to

$$\tilde{G}(m, c) = \Pr(\bar{Y}^{(m)*} \geq c \mid \tilde{F}^{(m)})\,,$$

the following:

$$(2.15)\quad \tilde{G}_s(m, c) = \begin{cases} 1 - \Phi(\tilde{a}_c) \\ \quad + \phi(\tilde{a}_c)\{t_c[10^m(1 - \exp\{-t_c/10^m\})]^{-1}\tilde{b}_c^{-1} - \tilde{a}_c^{-1}\}\,, & \text{if } c \neq 0\,, \\ \frac{1}{2} - \frac{1}{6}(2\pi n)^{-1/2}\tilde{K}^{(3)}(m, 0)/\{\tilde{K}''(m, 0)\}^{3/2} \\ \quad + \frac{1}{2}10^{-m}\{2\pi n\tilde{K}''(m, 0)\}^{-1/2}, & \text{if } c = 0\,, \end{cases}$$

where superscript (m) indicates the corresponding rounding of $\tilde{F}$, etc., $c10^m n$ is an integer, and $\tilde{a}_c$ and $\tilde{b}_c$ are the same as in (2.13). Again the relative error rate is n^{-1} in probability. The higher order approximations analogous to (2.12) can be obtained similarly. Note that the error term in general depends on m, but for each fixed m the rate is valid asymptotically, as n increases.

In practice, the X_j's are recorded with a fixed number $m = m_0$ of decimal places, so that the bootstrap estimate $\tilde{G}$ in (2.3) is $\tilde{G}(c) = \tilde{G}(m_0, c)$; then the corresponding saddlepoint approximation to $\tilde{G}(c)$ is $\tilde{G}_s(c) = \tilde{G}_s(m_0, c)$. Thus the asymptotical formula (2.14) holds. But if m is allowed to increase with n with a certain slow rate, the saddlepoint approximation is still valid. It is theoretically unclear whether the approximation is valid when m is infinite. Numerical examples suggest that the increase in m does not affect the approximation significantly.

If one uses $\tilde{G}$ in approximating the true continuous distribution $H(x)$ or $G(x) = 1 - H(x)$ of $\bar{X}$, modification of $\tilde{G}_s(c)$ with continuity correction is desirable; i.e., for real x,

$$\tilde{G}_1(x) = \tilde{G}_s(x + (2n10^m)^{-1}) . \tag{2.16}$$

If we were to assume $\tilde{F}$ continuous, we would have the existing saddlepoint formula (1.3) with $K(t) = \tilde{K}(t)$. Denote the result by $\tilde{G}_2$. Numerical examples show that all three of $\tilde{G}_s$, $\tilde{G}_1$ and $\tilde{G}_2$ are remarkably close to each other, although $\tilde{G}_s$ is asymptotically more accurate for the bootstrap tail probability.

Example 1. Consider the following artificial data with sample size $n = 10$ and $m = 1$ decimal place:

9.6 10.4 13.0 15.0 16.6 17.2 17.3 21.8 24.0 33.8

discussed by Davison and Hinkley (1988). We want to calculate the approximate bootstrap percentage points of $\bar{X} - \mu$. The results are compared in Table 1. The "exact" results in column 2 are simulation approximations from 50,000 simulated samples, taken from Davison and Hinkley (1988). They calculated the approximation $\tilde{H}_2 = 1 - \tilde{G}_2$ for the data. As Table 1 shows, the differences among $\tilde{H}_s = 1 - \tilde{G}_s$, $\tilde{H}_1 = 1 - \tilde{G}_1$ and $\tilde{H}_2$ are negligible. They are all very close to the "exact" results, even in the extreme tails, while normal approximations are not. For comparisons with Edgeworth approximations, see Davison and Hinkley (1988).

The calculations of the saddlepoint approximations in Table 1 are straightforward, except for solving the solution t_c for each fixed c. The method of bisection search was used to find the solution. Let $D(t) = \tilde{K}'(t) - c$ and choose a proper interval (b_L, b_U) such that $D(b_L) < 0$ and

Table 1. Approximations to bootstrap percentage points of $\bar{X} - \mu$.

Probability	"Exact"	Via saddlepoint approximation $\tilde{H}_s$	Via saddlepoint approximation $\tilde{H}_1$	Via saddlepoint approximation $\tilde{H}_2$	Normal approximation
0.0001	−6.34	−6.3180	−6.3130	−6.3130	−8.46
0.0005	−5.79	−5.7892	−5.7842	−5.7842	−7.48
0.001	−5.55	−5.5273	−5.5223	−5.5223	−7.03
0.005	−4.81	−4.8101	−4.8051	−4.8051	−5.86
0.01	−4.42	−4.4381	−4.4331	−4.4331	−5.29
0.05	−3.34	−3.3346	−3.3296	−3.3296	−3.74
0.10	−2.69	−2.6913	−2.6863	−2.6864	−2.91
0.20	−1.86	−1.8606	−1.8556	−1.8557	−1.91
0.80	1.80	1.7906	1.7956	1.7955	1.91
0.90	2.87	2.8466	2.8516	2.8515	2.91
0.95	3.73	3.7430	3.7480	3.7479	3.74
0.99	5.47	5.4715	5.4765	5.4765	5.29
0.995	6.12	6.1162	6.1212	6.1212	5.86
0.999	7.52	7.4584	7.4634	7.4634	7.03
0.9995	8.19	7.9839	7.9889	7.9889	7.48
0.9999	9.33	9.1095	9.1145	9.1145	8.46

$D(b_U) > 0$. Let $b_M = (b_L + b_U)/2$. If $D(b_M) < 0$, then redefine $b_L = b_M$. Otherwise redefine $b_U = b_M$. Repeat the above procedure until $b_U - b_L$ is smaller than a pre-set error bound. We then take b_M in the final recursive procedure as the solution t_c. The convergence of the recursive procedure is usually very quick. Table 2 in the next section was calculated similarly. There is no further numerical approximation involved in the calculation, since when the error bound was reduced from 10^{-6} to 10^{-10}, no further changes in Table 1 were made.

Using the special properties of the saddlepoint approximation and its smoothness, we can show theoretically that when α is close to zero, the difference between the α-th quantiles a_α of $\tilde{H}$ and $a_{s\alpha}$ of $\tilde{H}_s$, $a_{s\alpha} - a_\alpha = \alpha O_p(n^{-3/2})$, where $O_p(n^{-3/2})$ is fairly stable as $\alpha \to 0$ so that $a_{s\alpha}$ is generally more accurate than quantiles from other standard second-order accurate approximations. Similarly, when $\alpha \to 1$ we have $a_{s\alpha} - a_\alpha = (1 - \alpha)O_p(n^{-3/2})$ by considering the tail probabilities.

It is worth mentioning that the appropriate bootstrap density of $\bar{X} - E(X)$ can be constructed by using (1.2) with K replaced by $\tilde{K}$. The good properties of the classical results are retained in the approximation; this is also true in the randomization case. The details are omitted here.

3. Saddlepoint approximations in randomization analysis

Saddlepoint approximations can also be applied to randomization analysis in some situations. In this section we consider a special application in a matched pair design initially proposed by Davison and Hinkley (1988). Let $X_1, \ldots, X_n$ be the differences of two treatments randomly allocated to each pair of experimental units. If no effect exists, $\pm X_j$ appear equally likely for each unit.

To carry out the randomization test of a sample $\bar{X}$, one would need to calculate the distribution of $\bar{X}^* = n^{-1} \sum_1^n X_j^*$ given $\tilde{F}$, or equivalently the tail probability

$$\tilde{G}(c) = 1 - \Pr(\bar{X}^* < c \mid \tilde{F}) = \Pr(\bar{X}^* \geq c \mid \tilde{F}) , \tag{3.1}$$

where X_j^* is randomly selected to be $\pm X_j$ with probability of $1/2$ each.

Notice that there are two special features in this resampling plan. First, the X_j^*'s are independently but not identically distributed; second, each X_j^* takes only two possible values. Therefore any two values of $n\bar{X}^*$ always differ by a multiple of 2×10^{-m_0}. The MGF of $n\bar{X}^*$ is

$$\tilde{M}_{n\bar{X}^*}(t) = \prod_{j=1}^n \{(e^{tX_j} + e^{-tX_j})/2\} = \prod_{j=1}^n \cosh(tX_j) . \tag{3.2}$$

Let

$$\tilde{K}(t) = \frac{1}{n} \log \tilde{M}_{n\bar{X}^*}(t) = \frac{1}{n} \sum_{j=1}^n \log\{\cosh(tX_j)\} . \tag{3.3}$$

The saddlepoint technique used in Section 2 is still applicable here. But the detailed calculations are quite different. Because $\bar{X}^*$ is symmetric, it is sufficient to consider positive values of $\bar{X}^*$. For each value c of $\bar{X}^*$, $c = \left(\sum_{j=1}^n X_j + 2r_c/10^{m_0} \right) \Big/ n$, for some integer r_c, and

$$\begin{aligned} \tilde{M}_{n\bar{X}^*}(t) e^{-ntc} &= \sum_{j=1}^n \{\cosh(tX_j) e^{-tX_j}\} \exp\{-2r_c 10^{-m_0} t\} \\ &= \frac{1}{2n} \sum_{j=1}^n (1 + e^{-2tX_j}) \exp\{-2r_c 10^{-m_0} t\} . \end{aligned} \tag{3.4}$$

Thus, $(d - \alpha i, d + \alpha i)$ with $\alpha = 10^{m_0}\pi/2$ is a principal period of the formula in (3.4). For any $d > 0$,

$$\begin{aligned} \tilde{G}(c) &= \sum_{r \in S} \Pr(\bar{X}^* = r/n \mid \tilde{F}) = \sum_{r \in S} \frac{1}{2\alpha i} \int_{d-\alpha i}^{d+\alpha i} e^{n[\tilde{K}(t) - tr/n]} dt \\ &= \frac{1}{2\alpha i} \int_{d-\alpha i}^{d+\alpha i} \exp\{n[\tilde{K}(r) - tc]\}\{1 - \exp(-2t/10^{m_0})\}^{-1} dt , \end{aligned} \tag{3.5}$$

where $S = \{r \mid r \geq nc \text{ and } r10^{m_0}/2 \text{ is an integer}\}$.

LEMMA 3.1. *For fixed d and $a > 0$, there exists $0 \leq \rho < 1$, so that*

$$\max_{a \leq |y| \leq \alpha} |\tilde{M}_{n\bar{X}^*}(d + iy)/\tilde{M}_{n\bar{X}^*}(d)| = O_p(\rho^n)\,. \tag{3.6}$$

PROOF. See Appendix C.

Lemma 3.1 indicates that the integral of (3.5) in the two tails is exponentially small in probability, as in the i.i.d. case. Hence we can apply the saddlepoint method to (3.5), obtaining

$$\begin{aligned}\tilde{G}_s(c) &= 1 - \Phi(\tilde{a}_c) \\ &\quad + \phi(\tilde{a}_c)\{2t_c[10^{m_0}(1 - \exp[-2t_c/10^{m_0}])\tilde{b}_c]^{-1} - \tilde{a}_c^{-1}\}\,,\end{aligned} \tag{3.7}$$

with

$$\tilde{G}(c) = \tilde{G}_s(c)\{1 + O_p(n^{-1})\}\,,$$

where t_c satisfies $\tilde{K}'(t_c) = c$, $\tilde{a}_c = \{2n[t_c c - \tilde{K}(t_c)]\}^{1/2}\,\mathrm{sgn}\,(t_c)$ and $\tilde{b}_c = t_c\{\tilde{K}''(t_c)\}^{1/2}$. It can be easily seen that (3.7) is also valid for any negative value c of $\bar{X}^*$. To approximate the true tail probability of $\bar{X}$ which is continuous, it would be useful to simply modify (3.7) as

$$\tilde{G}_1(x) = \tilde{G}_s(x + (10^{m_0}n)^{-1}) \tag{3.8}$$

for real $x \in \left(-n^{-1}\sum_1^n |X_j|,\, n^{-1}\sum_1^n |X_j|\right)$, where $(10^{m_0}n)^{-1}$ is the continuity correction.

Example 2. In this example we illustrate the performance of (3.7) and (3.8) using the following $n = 12$ matched pair differences given by Miller ((1986), Exercise 1.11) and used in Davison and Hinkley (1988):

4.5 − 34.2 7.4 12.6 − 2.5 1.7 − 34.0 7.3 15.4 − 3.8 2.9 − 4.2 .

In the matched pair design, we wish to calculate the distribution of $\bar{X}^*$. Table 2 lists the comparisons of the approximations. Davison and Hinkley calculated $\tilde{G}_2$, the saddlepoint approximation defined in (1.3) with $K(t) = \tilde{K}(t)$. As in the bootstrap example, these three versions of saddlepoint approximations are all very close to each other, and close to the exact values which were calculated by exhausting all possible values of $\bar{X}^*$. Table 2 does not show significant effects of discreteness.

Table 2. Approximations to $\Pr(\bar{X}^* \geq c|\tilde{F})$ for matched pair differences data.

c	Exact	Via saddlepoint approx.			Normal approx.
		$\tilde{G}_s$	$\tilde{G}_1$	$\tilde{G}_2$	
0	0.50	0.50	0.50	0.50	0.50
0.90	0.417	0.428	0.427	0.427	0.421
2.24	0.310	0.323	0.323	0.322	0.309
4.04	0.202	0.201	0.200	0.200	0.184
6.28	0.0972	0.0883	0.0880	0.0879	0.0808
7.63	0.0442	0.0432	0.0430	0.0430	0.0446
8.53	0.0205	0.0213	0.0212	0.0215	0.0287
8.97	0.0127	0.0132	0.0131	0.0131	0.0228
10.32	0.00098	0.00104	0.00101	0.00101	0.01072
10.77	0.00024	0.00019	0.00018	0.00018	0.00820

4. Uniformity of relative errors

In the resampling problems discussed in the previous sections, the relative errors in the saddlepoint formulas appear not to be uniform. We will now show this fact.

We only need to prove that $\tilde{d}_1$ in (2.12) is not bounded in the tails for fixed n. Formula (3.7) has the same property. Without loss of generality we assume $X_{(n-1)} < X_{(n)}$. Let $p(t) = \exp\{tX_{(n)}\}/\sum_1^n \exp\{tX_j\}$ and $q(t) = 1 - p(t)$. Then $p(t) \to 1$, $q(t) \to 0$, as $t \to \infty$. Hence

$$\tilde{K}'(t) = \tilde{M}'(t)/\tilde{M}(t) \sim X_{(n)} - \bar{X} + q(t)(X_{(n)} - X_{(n-1)}) , \tag{4.1}$$

$$\begin{aligned}\tilde{K}''(t) &= \tilde{M}''(t)/\tilde{M}(t) - (\tilde{M}'(t)/\tilde{M}(t))^2 \\ &\sim X_{(n)}^2 + q(t)X_{(n-1)}^2 - (p(t)X_{(n)} + q(t)X_{(n-1)})^2 \\ &\sim q(t)(X_{(n)} - X_{(n-1)})^2 ,\end{aligned} \tag{4.2}$$

as $t \to \infty$. Here we ignore $q^2(t)$ and higher order terms. Similarly, $\tilde{K}^{(3)}(t) \sim q(t)(X_{(n)} - X_{(n-1)})^3$ and $\tilde{K}^{(4)}(t) \sim q(t)(X_{(n)} - X_{(n-1)})^4$. From (4.1), $q(t_c) \sim (X_{(n)} - \bar{X} - c)/(X_{(n)} - X_{(n-1)})$, as $c \to X_{(n)} - \bar{X}$. By expressing $\Psi_n''(w)$ in terms of dt/dw, d^2t/dw^2 and d^3t/dw^3, we can find

$$|\tilde{d}_1| \sim 5/\{2(X_{(n)} - X_{(n-1)})[q(t_c)]^{3/2}\} \to \infty , \tag{4.3}$$

as $t_c \to \infty$ and thus as $c \to X_{(n)} - \bar{X}$. We can similarly prove that $|d_1| \to \infty$ as $c \to X_{(1)} - \bar{X}$.

However, the numerical examples show that near the two extremes when the relative errors become apparent, both exact and saddlepoint approximations are practically 0 and 1. Thus while the lack of uniformity in these problems is interesting from the theoretical point of view, it does not cause serious problems in practice.

5. Generalizations

If we consider more general statistics such as $T = \sum_1^n a_j^{(n)} X_j$, it is still possible to use saddlepoint formulas with some restrictions on $a_j^{(n)}$. To be more specific, if $na_j^{(n)} \to c_j > 0$ for each fixed j, and $\sum_1^n a_j^{(n)}$ is bounded for all n, then the saddlepoint formulas are applicable both in parametric and nonparametric analysis. Such statistics include the estimates of coefficients in linear models. In bootstrap and randomization contexts, one should use the saddlepoint formulas derived in the paper with $\tilde{K}(t)$ replaced by $\tilde{K}(t) = n^{-1} \sum_1^n \tilde{K}(na_j^{(n)} t)$. The proofs are parallel to those for the case of sample mean.

Another natural extension is to a smooth function of the sample mean. It is straightforward when the sample mean is univariate. There are no general results of the saddlepoint approximations in the multivariate case in the literature yet, and this is a current research area. However, some special situations have been considered by Wang (1988). A genuine saddlepoint approximation for the distribution function of a nonlinear statistic $T = \bar{X} - \bar{Y}^2$ is derived with excellent accuracy. But the subsequent attempt to approximate the distribution of a studentized t-statistic (this type of statistic is very important in the bootstrap context; see Hall (1988)) is unsuccessful numerically due to other expansions involved. See Chapter 6 of Wang (1988) for details.

Acknowledgements

The author is deeply grateful to Professor David Hinkley for his many helpful ideas and suggestions. Thanks are also due to the editors and referees for helpful comments and efforts on the original draft which lead to much improvement of the paper.

Appendix A

PROOF OF LEMMA 2.1. For $x \in U$ and fixed k,

$$h_n(x) = h_n(0) + h_n'(0)x + \cdots + h_n^{(2k+1)}(0)x^{2k+1}/(2k+1)! + R_{n,2k+2}(x)\,,$$

where $R_{n,2k+2}(x) = h^{(2k+2)}(\xi_n)x^{2k+2}/(2k+2)!$, and ξ_n is between 0 and x. Let

$$g_{n,k}(x) = \{h_n(x) - [h_n(0) + h_n'(0)x + \cdots + h_n^{(2k+1)}(0)x^{2k+1}/(2k+1)!]\}/x^{2k+2} .$$

Then by (b) and (c),

$$C_{n,k} = \max_{x \in [-A_n, B_n]} |g_{n,k}(x)| = O_p(1) .$$

Write $e_n = \exp\{-(n/4)\min(A^2, B^2)\}$. It is easily seen that

$$\{n/(2\pi)\}^{1/2} \int_{-A_n}^{B_n} e^{-nx^2/2} x^{2k} dx = \frac{(2k)!}{2^k k! n^k} + O_p(e_n) ,$$

and

$$\{n/(2\pi)\}^{1/2} \int_{-A_n}^{B_n} e^{-nx^2/2} x^{2k+1} dx = O_p(e_n) .$$

But,

$$\begin{aligned} &\Big| \{n/(2\pi)\}^{1/2} \int_{-A_n}^{B_n} e^{-nx^2/2} \{h_n(x) \\ &\qquad - [h_n(0) + h_n'(0)x + \cdots + h_n^{(2k+1)}(0)x^{2k+1}/(2k+1)!]\} dx \Big| \\ &\qquad \le \{n/(2\pi)\}^{1/2} \int_{-A_n}^{B_n} e^{-nx^2/2} C_{n,k} x^{2(k+1)} dx = O_p(n^{-(k+1)}) . \end{aligned}$$

Thus, for each k,

$$\{n/(2\pi)\}^{1/2} \int_{-A_n}^{B_n} e^{-nx^2/2} h_n(x) dx = h_n(0) + \frac{h_n''(0)}{2n} + \cdots + \frac{h_n^{(2k)}(0)}{(2n)^k k!} + O_p(n^{-(k+1)}) .$$

Appendix B

PROOF OF LEMMA 2.2. Let

$$\rho_1(d) = \max_{y \in C} |M(d+iy)|/M(d) .$$

Then $\rho_1 < 1$. For $d \in I$, there exists $r \in (1, 2]$, such that

$$L(d) = \int_{-\infty}^{\infty} |x|^r e^{rxd} dF(x)$$

converges. It is easily seen that

$$\begin{aligned}
\Pr(|\tilde{M}(d+iy) - M(d+iy)| > \varepsilon) &\\
&\le \varepsilon^{-r} E|\tilde{M}(d+iy) - M(d+iy)|^r \\
&\le D\varepsilon^{-r} n^{1-r} E|\exp\{(d+iy)X_1\} - M(d+iy)|^r \\
&\le D(2/\varepsilon)^r n^{1-r}(E[\exp\{rX_1 d\}] + M^r(d)),
\end{aligned}$$

for any $\varepsilon > 0$, any y and some $D > 0$. The second inequality above comes from a result in Petrov ((1975), Chapter 3, Supplement item 15). The above result and the properties of M assure us that there exist $0 < \rho_2(d) < 1$ and $D_1 > 0$, such that

$$\Pr(|\tilde{M}(d+iy)/\tilde{M}(d)| > \rho_2(d)) \le D_1/n^{r-1},$$

for $y \in C$. Moreover, similarly to the first inequalities, for $a > E\{|X_1| \cdot \exp(dX_1)\}$, there is $D_2 > 0$, so that

$$\Pr\left(\frac{1}{n}\sum_1^n |X_j| e^{dX_j} > a\right) \le D_2/n^{r-1}.$$

Let $A_n(d,y) = \{\omega\colon |\tilde{M}(d+iy)|/\tilde{M}(d) > \rho_2(d)\}$, $B_n(d) = \left\{\omega\colon n^{-1}\sum_1^n |X_j| \exp(dX_j) > a\right\}$. For each $x \in C$, there exists open set C_x relative to C and independent of n, such that for $y \in C_x$ and $w \notin A_n(d,x) \cup B_n(d)$,

$$|\tilde{M}(d+iy)| \le |\tilde{M}(d+ix)| + \frac{2}{n}\sum_1^n |X_j| \exp(dX_j)|y-x| \le \rho(d)\tilde{M}(d),$$

where $\rho(d) = \{1 + \rho_2(d)\}/2 < 1$. Thus,

$$\left\{\omega\colon \sup_{y \in C_x} |\tilde{M}(c+iy)| > \rho(d)\tilde{M}(d)\right\} \subset A_n(d,x) \cup B_n(d).$$

There exist finite subsets of $\{C_x\}$, $C_{x_1},\dots, C_{x_p}$, say, covering C. Then

$$\begin{aligned}
Q_n &= \left\{\omega\colon \sup_{y \in C} |\tilde{M}(c+iy)| > \rho(d)\tilde{M}(d)\right\} \\
&\subset \bigcup_{k=1}^p \left\{\omega\colon \sup_{y \in C_{x_k}} |\tilde{M}(d+iy)| > \rho(d)\tilde{M}(d)\right\} \subset \left(\bigcup_{k=1}^p A_n(d,x_k)\right) \cup B_n(d).
\end{aligned}$$

Therefore,

$$\Pr(Q_n) \le \sum_1^p \Pr\{A_n(d, x_k)\} + \Pr\{B_n(d)\} \le (pD_1 + D_2)/n^{r-1}.$$

Finally,

$$\Pr\left(\rho^{-n} \max_{y \in C} |\tilde{M}(d+iy)/\tilde{M}(d)|^n \le 1\right) \ge 1 - \Pr(Q_n) \to 1,$$

as $n \to \infty$, and this implies Lemma 2.2.

Appendix C

PROOF OF LEMMA 3.1.

$$\begin{aligned}
&\log\{|\tilde{M}_{n\bar{X}^*}(d+iy)/\tilde{M}_{n\bar{X}^*}(d)|^{1/n}\} \\
&= \frac{1}{n}\sum_1^n \log|(e^{dX_j} + e^{-dX_j})\cos(yX_j) + i(e^{dX_j} - e^{-dX_j})\sin(yX_j)| \\
&\quad - \frac{1}{n}\sum_1^n \log(e^{dX_j} + e^{-dX_j}) \\
&= \frac{1}{2n}\sum_1^n \log\{(e^{dX_j} - e^{-dX_j})^2 - 4\sin^2(dX_j)\} - \frac{1}{n}\sum_1^n \log(e^{dX_j} + e^{-dX_j}) \\
&= \frac{1}{2n}\sum_1^n \log\{(1 - 4\sin^2(yX_j))(e^{dX_j} + e^{-dX_j})^{-2}\} \\
&\le -\frac{2}{n}\sum_1^n \sin^2(yX_j)(e^{dX_j} + e^{-dX_j})^{-2}.
\end{aligned}$$

Thus,

$$\begin{aligned}
&\log\left\{\max_{a \le |y| \le \alpha} |\tilde{M}_{n\bar{X}^*}(d+iy)/\tilde{M}_{n\bar{X}^*}(d)|^{1/n}\right\} \\
&\qquad \le -\frac{2}{n}\min_{a \le |y| \le \alpha}\sum_1^n \sin^2(yX_j)(e^{dX_j} + e^{-dX_j})^{-2}.
\end{aligned}$$

Because

$$\frac{1}{n}\sum_1^n \sin^2(yX_j)(e^{dX_j} + e^{-dX_j})^{-2} \to E\{\sin^2(yX_1)(e^{dX_1} + e^{-dX_1})^{-2}\}$$

in probability, and

$$\min_{a \leq |y| \leq \alpha} E\{\sin^2(yX_1)(e^{dX_1} + e^{-dX_1})^{-2}\} = c > 0 ,$$

using the same technique as given in Appendix B, we can easily prove that

$$\max_{a \leq |y| \leq \alpha} |\tilde{M}_{n\bar{X}^*}(d + iy) / \tilde{M}_{n\bar{X}^*}(d)|^{1/n} \leq \rho$$

in probability, where $\rho = e^{-c/4} < 1$. Formula (3.6) then follows.

References

Barndorff-Nielsen, O. and Cox, D. R. (1979). Edgeworth and saddlepoint approximations with statistical applications (with discussion), *J. Roy. Statist. Soc. Ser. B*, **41**, 279–312.

Daniels, H. E. (1954). Saddlepoint approximations in statistics, *Ann. Math. Statist.*, **25**, 631–650.

Daniels, H. E. (1983). Saddlepoint approximations for estimating equations, *Biometrika*, **70**, 89–96.

Daniels, H. E. (1987). Tail probability approximations, *Internat. Statist. Rev.*, **55**, 37–48.

Davison, A. C. and Hinkley, D. V. (1988). Saddlepoint approximations in resampling methods, *Biometrika*, **75**, 417–431.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*, CBMS-NSF Monograph 38, SIAM, Philadelphia.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals (with discussion), *Ann. Statist.*, **16**, 927–953.

Jeffreys, H. and Jeffreys, B. S. (1962). *Methods of Mathematical Physics*, 3rd ed., Cambridge Univ. Press, London-New York.

Jensen, J. L. (1988). Uniform saddlepoint approximations, *Adv. in Appl. Probab.*, **20**, 622–634.

Lugannani, R. and Rice, S. O. (1980). Saddlepoint approximations for the distribution of the sum of independent random variables, *Adv. in Appl. Probab.*, **12**, 475–490.

Miller, R. G., Jr. (1986). *Beyond ANOVA, Basics of Applied Statistics*, Wiley, New York.

Petrov, V. V. (1975). *Sums of Independent Random Variables*, (translated by A. A. Brown), Springer, New York.

Reid, N. (1988). Saddlepoint methods and statistical inference (with discussion), *Statist. Sci.*, **3**, 213–238.

Robinson, J. (1982). Saddlepoint approximations for permutation tests and confidence intervals, *J. Roy. Statist. Soc. Ser. B*, **44**, 91–101.

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap, *Ann. Statist.*, **9**, 1187–1195.

Wang, S. (1988). Saddlepoint approximations in statistics, including resampling, Ph.D. dissertation, The University of Texas at Austin.