Journal of the Korean
Data & Information Science Society
1998, Vol. 9, No. 2, pp. 255 ~ 262

Saddlepoint approximations for the ratio [1] of two independent sequences of random variables

Daehyeon Cho [2]

Abstract

In this paper, we study the saddlepoint approximations for the ratio of independent random variables. In Section 2, we derive the saddlepoint approximation to the probability density function. In Section 3, we represent a numerical example which shows that the errors are small even for small sample size.

Key Words and Phrases: Saddlepoint approximation, Probability density function,Fourier inversion.

1. Introduction

In probability theory, one of the most important problems is to find the probability density function of some statistics. And it is often required to approximate the distribution of some statistics whose distribution can not be exactly obtained. When the first few moments are known, a common procedure is to fit the law of the Edgeworth type having the same moments as far as they are given (Edgeworth(1905),Wallace(1958)). This method is often satisfactory in practice, but can assume negative values in some tail regions of distribution.

Daniels (1954)introduced a new type of idea into statistics by applying saddlepoint techniques to derive a very accurate approximation to the probability density function of $\bar{X}$. He showed that the error incurred by using the saddlepoint approximation method is $O(n^{-1})$ as against the more usual $O(n^{-\frac{1}{2}})$ associated with the normal approximation. Moreover, he showed that the relative error of the approximation is uniformly $O(n^{-1})$ over the whole range of the random variable in an

[1]This paper was supported by Grant from InJe University,1997
[2]Associate Professor, Department of Applied Statistics, Inje University, Kimhae, 621-749, Korea.

important class of cases. For reviews of saddlepoint approximations, see Reid (1988) and Field and Ronchetti (1990).

In this paper, we study the saddlepoint approximations for the ratio of independent sequences of random variables. Cho(1996) derived an saddlepoint approximation formula for the ratio $R_n = \frac{U_n}{S_n}$ where $\{U_n, n \geq 1\}$, $\{S_n > 0, n \geq 1\}$ are two independent sequences of random variables. This methods can be applied to find the approximate density function of statistics such as F distribution with degree of freedom (n, n). In this paper we study for the ratio $R_{n1,n2} = \frac{U_{n1}}{S_{n2}}$ where $\{U_{n1}, n_1 \geq 1\}$, $\{S_{n2} > 0, n_2 \geq 1\}$ are two independent sequences of random variables using the similar method as in Cho(1996). Section 2, we derive the saddlepoint approximation to the density. And in Section 3, we represent a numerical example which shows the errors are small even for the small size. Let $\{U_{n1}, n_1 \geq 1\}, \{S_{n2} > 0, n_2 \geq 1\}$ be independent sequences of random variables with absolutely continuous distribution functions F_{n1}, F_{n2} respectively. Denote $\phi_{n1}(t) = E\{\exp(tU_{n1})\}$ and $\phi_{n2}(t) = E\{\exp(tS_{n2})\}$ be the moment generating functions of U_{n1} and S_{n2}, respectively. And let $\psi_{n1}(t) = (1/n_1) log\phi_{n2}(t)$, and $\psi_{n2}(t) = (1/n_2) log\phi_{n2}(t)$ be their cumulant generating function. Assume that $\phi_{n1}(t)$ and $\psi_{n1}(t)$ exist for real t in some interval (t_1, t_2) containing 0 and that $\phi_{n2}(t)$ and $\psi_{n2}(t)$ exist for real t in some interval (t_3, t_4) containing 0.

2. Saddlepoint approximation

The integrand in Fourier inversion of the density of $R_{n1,n2}$ is of the form exp $[n_2 \{\Psi_{n1,n2}(z)\}]$. This is the starting point to derive the saddlepoint approximation formula for $R_{n1,n2}$.

Let $H_{n1,n2}$ be the distribution function of $R_{n1,n2}$. Then

$$H_{n1,n2}(r) = P_r(R_{n1,n2} \leq r) = \int_0^{+\infty} F_{n1}(ry)\ dF_{n2}(y). \tag{1}$$

And the p.d.f. $h_{n1,n2}$ of $R_{n1,n2}$ is given by

$$h_{n1,n2}(r) = \int_0^{+\infty} y f_{n1}(ry)\ dF_{n2}(y), \tag{2}$$

where f_{n1} is the p.d.f. of U_{n1}.

The characteristic function of $R_{n1,n2}$ is given by

$$\begin{aligned} \hat{h}_{n1,n2}(t) &= \int_{-\infty}^{+\infty} e^{itr} \int_0^{+\infty} y f_{n1}(ry)\ dF_{n2}(y)\ dr \\ &= \int_{-\infty}^{+\infty} \phi_{n1}(\frac{it}{y})\ dF_{n2}(y). \end{aligned} \tag{3}$$

Using the Fourier inversion formula, the p.d.f. $h_{n1,n2}$ is given by

$$\begin{aligned} h_{n1,n2} \quad (r) &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itr}\hat{h}_{n1,n2}(t)dt \\ &= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\{\int_{0}^{+\infty}\phi_{n1}(\frac{it}{y})dF_{n2}(y)\}dt \\ &= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\phi_{n1}(is)\{\int_{0}^{+\infty}e^{-isyr}\cdot ydF_{n2}(y)\}ds \\ &= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\phi_{n1}(it)\phi'_{n2}(-irt)dt \\ &= \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\phi_{n1}(z)\phi'_{n2}(-rz)dz \\ &= \frac{n_2}{2\pi i}\int_{c-i\infty}^{c+i\infty} exp[n_2\{\frac{n_1}{n_2}\psi_{n1}(z)+\psi_{n2}(-rz)\}]\times\psi'_{n2}(-rz)dz \qquad (4) \end{aligned}$$

where c is any real number in the interval where the moment generating function exists.

When n_1, n_2 is large, an approximation is found by passing the path of integration through a saddlepoint τ of the exponential part of integrand given by $n_1\psi'_{n1}(\tau) - rn_2\ \psi'_{n2}(-r\tau) = 0$. We choose c to be τ.

On the contour near τ, we have

$$\begin{aligned} n_2\{\frac{n_1}{n_2} \quad & \psi_{n1}(z)+\psi_{n2}(-rz)\} = n_2\ \Psi_{n1,n2}(z)(say) \\ &= n_2\{\Psi_{n1,n2}(\tau)+\frac{\Psi''(\tau)}{2}(z-\tau)^2+\frac{\Psi^{(3)}(\tau)}{6}(z-\tau)^3+\cdots\} \qquad (5) \end{aligned}$$

and

$$\begin{aligned} \psi'_{n2}(-rz) \quad &= \psi'_{n2}(-r\tau) - r\psi^{(2)}_{n2}(-r\tau)(z-\tau) \\ &+\frac{r^2}{2}\psi^{(3)}_{n2}(-r\tau)(z-\tau)^2-\frac{r^3}{6}\psi^{(4)}_{n2}(-r\tau)(z-\tau)^3+\cdots. \qquad (6) \end{aligned}$$

Let $\sqrt{n_2\Psi''_{n1,n2}(\tau)}(z-\tau) = iy$ and expanding the integrand in (4) near τ, we have

$$\begin{aligned} h_{n1,n2}(r) \quad &= \left(\frac{n_2}{\hat{\Psi}''_{n1,n2}}\right)^{\frac{1}{2}}\frac{\exp[n_1\{\psi_{n1}(\tau)+n_2\psi_{n2}(-r\tau)\}]}{2\pi}\psi'_{n2}(-r\tau) \\ &\times\int_{-\infty}^{+\infty}\exp\left(\frac{y^2}{2}\right)(1-\frac{\lambda_3}{6\sqrt{n_2}}iy^3+\frac{\lambda_4}{24n_2}y^4-\frac{\lambda_3^2}{72n_2}y^6+\cdots) \\ &\times\{1-\frac{r\psi^{(2)}_{n2}(-r\tau)}{\psi'_{n2}(-r\tau)}\frac{iy}{\sqrt{n_2}(\hat{\Psi}''_{n1,n2})^{\frac{1}{2}}}-\frac{r^2\psi^{(3)}_{n2}(-r\tau)}{2n_2\psi'_{n2}(-r\tau)\hat{\Psi}''_{n1,n2}}y^2 \end{aligned}$$

$$+\frac{r^3\psi_{n2}^{(4)}(-r\tau)}{6n_2\sqrt{n_2}\psi'_{n2}(-r\tau)(\hat{\Psi}''_{n1,n2})^{\frac{3}{2}}}y^3+\cdots)dy\}$$

$$=\tilde{h}_{n1,n2}(r)[1+\frac{3}{n_2}\left\{\frac{\lambda_4}{24}-\frac{r\psi_{n2}^{(2)}(-r\tau)\lambda_3}{6\psi'_{n2}(-r\tau)(\hat{\Psi}''_{n1,n2})^{\frac{1}{2}}}\right\}-\frac{15\lambda_3^2}{72n_2}$$

$$-\frac{r^2\psi_{n2}^{(3)}(-r\tau)}{2n_2\psi'_{n2}(-r\tau)\hat{\Psi}''_{n1,n2}}+\cdots], \tag{7}$$

where

$$\tilde{h}_{n1,n2}(r)=\frac{\sqrt{n_2}\psi'_{n2}(-r\tau)\exp[n_1\psi_{n1}(\tau)+n_2\psi_{n2}(-r\tau)]}{\sqrt{2\pi\{\frac{n_1}{n_2}\psi''_{n1}(\tau)+r^2\psi''_{n2}(-r\tau)\}}} \tag{8}$$

and

$$\hat{\Psi}^{(j)}_{n1,n2}=\Psi^{(j)}_{n1,n2}(\tau),\ \lambda_j=\frac{\hat{\Psi}^{(j)}_{n1,n2}}{(\hat{\Psi}''_{n1,n2})^{j/2}}.$$

We call $\tilde{h}_{n1,n2}(r)$ the saddlepoint approximation to $h_{n1,n2}(r)$.

Remark. In the above if $n_1=n_2=n$,then (8) becomes as follows.

$$\tilde{h}_{n,n}(r)=\frac{\sqrt{n}\psi'_{n2}(-r\tau)\exp[n(\psi_{n1}(\tau)+\psi_{n2}(-r\tau))]}{\sqrt{2\pi\{\psi''_{n1}(\tau)+r^2\psi''_{n2}(-r\tau)\}}}, \tag{9}$$

where

$$\psi_{n1}(t)=\frac{1}{n}log\phi_{n1}(t),\quad \phi_{n1}(t)=E(exp(tU_n)),$$

$$\psi_{n2}(t)=\frac{1}{n}log\phi_{n2}(t),\quad \phi_{n2}(t)=E(exp(tS_n)).$$

3. Numerical example

In this section, we present an example to show that the errors of our saddlepoint approximation formular are small.

Example

Assume that $\{U_{n1}\}$ and $\{S_{n2}\}$ have $\chi^2(n_1)$, $\chi^2(n_2)$distribution respectively and are independent. Then $R_{n_1,n_2}=U_{n1}/S_{n2}$ follows F-distribution with (n_1,n_2) degrees of freedom(Denote,$F(n_1,n_2)$). So the exact probability density function of $R_{n1,n2}$ is given by

$$h_{n1,n2}(r)=\frac{\Gamma(\frac{n_1+n_2}{2})r^{\frac{n_1}{2}-1}}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})(1+r)^{\frac{n_1+n_2}{2}}} \tag{10}$$

By definition, we have

$$\begin{aligned}\phi_{ni}(z) &= (1-2z)^{-\frac{n_i}{2}}, \psi_{ni}(z) = -\frac{1}{2}\, log(1-2z),\\ \psi'_{ni}(z) &= (1-2z)^{-1} \text{ and } \psi_{ni}^{(2)} = 2(1-2z)^{-2}\ (i=1,2).\end{aligned}$$

So the saddlepoint equation becomes $n_1\psi'_{n1}(\tau) - n_2 r\psi'_{n2}(-r\tau) = 0$ and the saddlepoint $\tau = \frac{n_2 r - n_1}{2r(n_1+n_2)}$. Therefore the saddlepoint approximation to the p.d.f. of R_{n_1,n_2} is given by

$$\begin{aligned}\tilde{h}_{n1,n2}(r) &= \frac{\sqrt{n}_2\psi'_{n2}(-r\tau)\exp[n_1\psi_{n1}(\tau) + n_2\psi_{n2}(-r\tau)]}{\sqrt{2\pi\{\frac{n_1}{n_2}\psi''_{n1}(\tau) + r^2\psi''_{n2}(-r\tau)\}}}\\ &= \frac{(\frac{n_1+n_2}{n_1})^{\frac{n_1-1}{2}}(\frac{n_1+n_2}{n_2})^{\frac{n_2}{2}}\, r^{\frac{n_1}{2}-1}}{2\sqrt{\pi}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}(1+r)^{\frac{n_1+n_2}{2}}}. \end{aligned} \tag{11}$$

Note that the ratio $B_{n1,n2}(r) = \frac{\tilde{h}_{n1,n2}(r)}{h_{n1,n2}(r)} = \frac{\Gamma(\frac{n_1+n_2}{2})(\frac{n_1+n_2}{n_1})^{\frac{n_1}{2}}(\frac{n_1+n_2}{n_2})^{\frac{n_2}{2}}}{2\sqrt{\pi}\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ does not depend on r. Tables 3.1 $\sim$ Table 3.3 represent the numerical results of (10) ,(11) when $\{(n_1 = 2, n_2 = 18), (n_1 = 18, n_2 = 2)\}, \{(n_1 = 4, n_2 = 16), (n_1 = 16, n_2 = 4)\}, \{(n_1 = 8, n_2 = 16), (n_1 = 16, n_2 = 16)\}$ with increasing r by 0.2.

The exact values of F-distribution's probability density function are computed from *IMSL*. From the tables we can see that the errors of saddlepoint approximation are small even for small sample size.

Reference

1. Cho, D. H. (1996). Tail probability approximations for the ratio of the independent random variables. *Journal of Statistical Theory and Methods*, 7, 189-201.

2. Daniels, H. E. (1954). Saddlepoint approximations in statistics, *Annals of Mathmatical Statistics*, 25, 631-650.

3. Edgeworth, F. Y. (1905). The law of error, *Cambridge Philos. Trans.*, 20, 36-66 and 113-141.

4. Field, C. and Ronchetti, E. (1990). *Small sample asymptotics,* IMS Lecture notes-monograph series 13.

5. Reid, N. (1988). Saddlepoint methods and statistical Inference, *Statist. Sci.*, 3, 213-238.

6. Wallace, D. L. (1958). Asymptotic approximations to distributions, *Annals of Mathmatical Statistics*, 29, 635-654.

Table 3.1 Values of the exact probability density function $h_{n1,n2}(r)$ from (10) and saddlepoint approximation $\tilde{h}_{n1,n2}(r)$ from (11) for $F(2,18)$ and $F(18,2)$ with increasing r by 0.2

$F(2,18)$			$F(18,2)$		
r	$h_{n1,n2}$	$\tilde{h}_{n1,n2}$	r	$h_{n1,n2}$	$\tilde{h}_{n1,n2}$
.1	3.46989	3.76636	.1	3.46989E-8	3.76636E-8
.3	.6528437	.708623	.3	4.28331E-5	4.64928E-5
.5	.1560738	.1694088	.5	6.09663E-4	6.61754E-4
.7	4.46430E-2	4.84574E-2	.7	2.57358E-3	2.79347E-3
.9	1.46793E-2	1.59336E-2	.9	6.31897E-3	6.85887E-3
1.1	5.39573E-3	5.85673E-3	1.1	1.15662E-2	1.25544E-2
1.3	2.17252E-3	2.35815E-3	1.3	1.77219E-2	1.92361E-2
1.5	9.43718E-4	1.02435E-3	1.5	2.41865E-2	2.62530E-2
1.7	4.37124E-4	4.74473E-4	1.7	3.04927E-2	3.30981E-2
1.9	2.13925E-4	2.32204E-4	1.9	3.63322E-2	3.94364E-2
2.1	1.09806E-4	1.19188E-4	2.1	4.15317E-2	4.50802E-2
2.3	5.87629E-5	6.37837E-5	2.3	4.60178E-2	4.99496E-2
2.5	3.26259E-5	3.54134E-5	2.5	4.97831E-2	5.40367E-2
2.7	1.87165E-5	2.03157E-5	2.7	5.28610E-2	5.73775E-2
2.9	1.10560E-5	1.20006E-5	2.9	5.53070E-2	6.00326E-2
3.1	6.70508E-6	7.27797E-6	3.1	5.71870E-2	6.20731E-2
3.3	4.16445E-6	4.52027E-6	3.3	5.85692E-2	6.35735E-2
3.5	2.64312E-6	2.86895E-6	3.5	5.95198E-2	6.46053E-2
3.7	1.71106E-6	1.85725E-6	3.7	6.01005E-2	6.52355E-2
3.9	1.12793E-6	1.22430E-6	3.9	6.03669E-2	6.55246E-2
4.1	7.56033E-7	8.20630E-7	4.1	6.03687E-2	6.55266E-2
4.3	5.14617E-7	5.58586E-7	4.3	6.01494E-2	6.52887E-2
4.5	3.55317E-7	3.85676E-7	4.5	5.97470E-2	6.48518E-2
4.7	2.48596E-7	2.69836E-7	4.7	5.91939E-2	6.42515E-2
4.9	1.76085E-7	1.91130E-7	4.9	5.85182E-2	6.35181E-2
5.1	1.26167E-7	1.36946E-7	5.1	5.77438E-2	6.26775E-2
5.3	9.13770E-8	9.91843E-8	5.3	5.68911E-2	6.17519E-2
5.5	6.68512E-8	7.25630E-8	5.5	5.59771E-2	6.07599E-2
5.7	4.93736E-8	5.35921E-8	5.7	5.50166E-2	5.97172E-2
5.9	3.67919E-8	3.99354E-8	5.9	5.40216E-2	5.86373E-2

Table 3.2 Values of the exact probability density function $h_{n1,n2}(r)$ from (10) and saddlepoint approximation $\tilde{h}_{n1,n2}(r)$ from (11) for $F(4,16)$ and $F(16,4)$ with increasing by 0.2

$F(4,16)$			$F(16,4)$		
r	$h_{n1,n2}$	$\tilde{h}_{n1,n2}$	r	$h_{n1,n2}$	$\tilde{h}_{n1,n2}$
.1	2.77591	2.8991	.1	2.77591E-6	2.89910E-6
.3	1.56683	1.63636	.3	1.14222E-3	1.19290E-3
.5	.6242951	.6520000	.5	9.75461E-3	1.01875E-2
.7	.2500007	.2610953	.7	2.94123E-2	3.07176E-2
.9	.1056913	.1103816	.9	5.61687E-2	5.86613E-2
1.1	4.74824E-2	4.95895E-2	1.1	8.41180E-2	8.78509E-2
1.3	2.25942E-2	2.35969E-2	1.3	.1090580	.1138978
1.5	1.13246E-2	1.18272E-2	1.5	.1289945	.1347190
1.7	5.94489E-3	6.20871E-3	1.7	.1434951	.1498632
1.9	3.25167E-3	3.39597E-3	1.9	.1529775	.1597663
2.1	1.84474E-3	1.92660E-3	2.1	.1582160	.1652374
2.3	1.08124E-3	1.12922E-3	2.3	.1600619	.1671652
2.5	6.52517E-4	6.81475E-4	2.5	.1593060	.1663756
2.7	4.04277E-4	4.22218E-4	2.7	.1566252	.1635758
2.9	2.56498E-4	2.67881E-4	2.9	.1525711	.1593419
3.1	1.66286E-4	1.73666E-4	3.1	.1475794	.1541286
3.3	1.09942E-4	1.14821E-4	3.3	.1419859	.1482871
3.5	7.40075E-5	7.72917E-5	3.5	.1360454	.1420827
3.7	5.06472E-5	5.28948E-5	3.7	.129947	.1357137
3.9	3.51914E-5	3.67531E-5	3.9	.1238295	.1293248
4.1	2.47979E-5	2.58984E-5	4.1	.1177925	.1230200
4.3	1.77028E-5	1.84884E-5	4.3	.1119059	.1168721
4.5	1.27914E-5	1.33591E-5	4.5	.1062168	.1109305
4.7	9.34721E-6	9.76202E-6	4.7	.1007555	.1052269
4.9	6.90253E-6	7.20885E-6	4.9	9.55399E-2	9.97798E-2
5.1	5.14760E-6	5.37603E-6	5.1	9.05786E-2	9.45983E-2
5.3	3.87439E-6	4.04632E-6	5.3	8.58733E-2	8.96842E-2
5.5	2.94145E-6	3.07199E-6	5.5	8.14213E-2	8.50346E-2
5.7	2.25144E-6	2.35135E-6	5.7	7.72162E-2	8.06429E-2
5.9	1.73658E-6	1.81364E-6	5.9	7.32497E-2	7.65003E-2

Table 3.3 Values of the exact probability density function $h_{n1,n2}(r)$ from (10) and saddlepoint approximation $\tilde{h}_{n1,n2}(r)$ from (11) for $F(8,16)$ and $F(16,16)$ with increasing by 0.2

$F(8,16)$			$F(16,16)$		
r	$h_{n1,n2}$	$\tilde{h}_{n1,n2}$	r	$h_{n1,n2}$	$\tilde{h}_{n1,n2}$
.1	.4205925	.4309207	.1	1.12036E-03	1.13799E-03
.3	1.52974	1.56730	.3	.1691976	.1718602
.5	1.27171	1.30294	.5	.6123059	.6219424
.7	.7771072	.7961898	.7	.8712482	.8849595
.9	.4347687	.4454450	.9	.8536475	.8670818
1.1	.2388476	.2447126	1.1	.7012600	.7122960
1.3	.1323335	.1355831	1.3	.5267402	.5350301
1.5	7.47425E-2	7.65778E-2	1.5	.3777785	.3837239
1.7	4.32071E-2	4.42681E-2	1.7	.2648260	.2689939
1.9	2.55893E-2	2.62177E-2	1.9	.1838846	.1867785
2.1	1.55200E-2	1.59011E-2	2.1	.1274636	.1294697
2.3	9.62920E-3	9.86566E-3	2.3	8.86157E-2	9.00103E-2
2.5	6.10348E-3	6.25335E-3	2.5	6.19626E-2	6.29378E-2
2.7	3.94680E-3	4.04371E-3	2.7	4.36473E-2	4.43342E-2
2.9	2.60012E-3	2.66396E-3	2.9	3.10021E-2	3.14899E-2
3.1	1.74282E-3	1.78562E-3	3.1	2.22141E-2	2.25637E-2
3.3	1.18712E-3	1.21627E-3	3.3	1.60598E-2	1.63126E-2
3.5	8.20782E-4	8.40938E-4	3.5	1.17143E-2	1.18986E-2
3.7	5.75447E-4	5.89577E-4	3.7	8.61956E-3	8.75521E-3
3.9	4.08710E-4	4.18746E-4	3.9	6.39666E-3	6.49733E-3
4.1	2.93822E-4	3.01036E-4	4.1	4.78633E-3	4.86166E-3
4.3	2.13633E-4	2.18879E-4	4.3	3.60997E-3	3.66678E-3
4.5	1.56985E-4	1.60840E-4	4.5	2.74362E-3	2.78679E-3
4.7	1.16512E-4	1.19373E-4	4.7	2.10052E-3	2.13357E-3
4.9	8.72847E-5	8.94280E-5	4.9	1.61949E-3	1.64498E-3
5.1	6.59670E-5	6.75869E-5	5.1	1.25705E-3	1.27683E-3
5.3	5.02707E-5	5.15052E-5	5.3	9.82021E-4	9.97475E-4
5.5	3.86102E-5	3.95583E-5	5.5	7.71905E-4	7.84052E-4
5.7	2.98745E-5	3.06081E-5	5.7	6.10332E-4	6.19936E-4
5.9	2.32777E-5	2.38494E-5	5.9	4.85308E-4	4.92945E-4