2001

Interval estimators of parameters for normal one sample and balanced one-way random effects models when data are rounded

Chiang-Sheng Lee
Iowa State University

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction..

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Interval estimators of parameters for normal one sample and balanced one-way random effects models when data are rounded

by

Chiang-Sheng Lee

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Industrial Engineering

Major Professor: Stephen B. Vardeman

Iowa State University

Ames, Iowa

2001

UMI Number: 3016720

UMI Microform 3016720

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Interval estimators of parameters for normal one sample and balanced one-way random effects models when data are rounded

Chiang-Sheng Lee

Major Professor: Stephen B. Vardeman
Iowa State University

In standard statistical analysis, data are typically assumed to be essentially exact. But in fact, all real data are reported to some smallest unit of measure related to the precision of the device used to produce them. We might call such data "rounded" because they are really obtained by "rounding to something." We first discuss the interval estimation of the parameters μ and σ when a single rounded sample comes from the $N(\mu, \sigma^2)$ distribution with both parameters unknown. Then we discuss the interval estimation of variance components σ and σ_τ if rounded data are from a balanced one-way normal random effects model. For each problem rounded-data likelihood-based methods are compared to naive calculations made as if observations were exact. We find that with some modifications the likelihood-based methods provide an effective way to analyze such data.

Graduate College
Iowa State University

This is to certify that the Doctoral dissertation of

Chiang-Sheng Lee

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Major Professor

Signature was redacted for privacy.

For the Major Program

Signature was redacted for privacy.

For the Graduate College

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

Introduction

It is a practical problem that data on hand are sometimes obtained using crude gaging. For example, a scale might read only to the nearest pound while ounces are still important. We might call such data "rounded " since they are really obtained by rounding to the nearest unit. Furthermore, we can call them "rounded Normal data" if underlying exact values are from a Normal distribution.

In standard statistical analyses, data are assumed to be essentially exact. It is of interest to know what happens to the statistical properties of these methods when rounded data are used. Do the traditional methods still work? And if they do not, what are reasonable ways to improve on these methods?

Dissertation Organization

This dissertation contains three papers that focus on finding better methods of interval estimation of distribution parameters when rounded data are collected. When rounded sample is from the $N(\mu, \sigma^2)$ distribution with both parameters unknown, we discuss interval estimators of the parameters μ and σ separately in Chapters 2 and 3. In Chapter 4, similar analyses are made of interval estimators of the two variance components σ and σ_τ, for rounded data from the balanced one-way random effects model.

In each chapter, we start by defining different types of likelihood functions (like the

log-likelihood function and an appropriate profile likelihood function), then discuss some properties of these functions for special data configurations. The properties considered include approximate maximizers, supremum values, and the qualitative nature of the functions.

In each problem two methods are used to construct confidence intervals. One is a traditional method derived as if the data were exact and the other is a likelihood-based method. After the simulation and computation, adjusted versions of the simplest likelihood-based methods are suggested and some related results are provided.

2 INTERVAL ESTIMATORS OF THE PARAMETER μ FOR ROUNDED NORMAL DATA

A paper to be published in the Journal of Quality Technology

Chiang-Sheng Lee and Stephen B. Vardeman
Iowa State University, Ames, IA 50011-1210

Abstract

Standard statistical methods are based on an implicit assumption that numerical data are exact. But in truth, all real data are rounded to some smallest unit of measure related to the precision of the device used to produce them. When the degree of rounding is severe, ignoring the rounding produces statistical methods with operating characteristics far from nominal. We discuss the interval estimation of the parameter μ when rounded data come from the $N(\mu,\sigma^2)$ distribution.

Key Words: crude gaging, interval-censoring, likelihood, profile likelihood, coverage probability, average length

---

Mr. Lee is a Ph. D. Candidate in Industrial Engineering in the Industrial and Manufacturing Systems Engineering Department. email: chiang@iastate.edu

Dr. Vardeman is a Professor in the Statistics and Industrial and Manufacturing Systems Engineering Departments. He is a Senior Member of ASQ.

Usually, we suppose that numerical data are exact. But in truth, all real data are rounded to some smallest unit of measure related to the precision of the device used to produce them. Because of this, it can reasonably be assumed that a sample in hand was collected by "rounding to something." We will discuss the interval estimation of the parameter μ when such rounded data come from the $N(\mu,\sigma^2)$ distribution.

2.1 Introduction

It is an important practical problem that the collection of measurement data is sometimes done using relatively crude gaging. For example, a scale might read only to the nearest pound while ounces are still of some importance. Traditional methods of estimation of distribution parameters and the construction of confidence intervals are really based on an assumption that observed data are essentially "exact." It is of interest to know what happens to the statistical properties of these methods when, in fact, the available data are produced by relatively crude gaging. Do nominal (or exact data) statistical properties carry over to the case of crudely gaged data? And if they do not, what are reasonable replacements for these traditional methods?

The main purpose of this paper is to investigate the properties of interval estimators of the parameter μ based on rounded Normal data. Two methods will be compared. One is the traditional t interval (appropriate for exact Normal data) and the other is obtained from inversion of (rounded data) likelihood ratio tests for μ. Our end goal is to find which method provides better confidence intervals for μ. We first discuss the *likelihood function* for rounded Normal data. Then we discuss the maximum likelihood estimates of μ and σ for two special cases. The construction of the rounded data confidence intervals for μ (and approximate formulas for them) will be provided. Then simulation results are given, and based on these, a correction to the second method for small sample sizes is suggested. We also compute and compare the average interval lengths for these two

methods for various sample sizes.

2.2 The Model for Rounded Normal Data

Without loss of generality, it is convenient to assume that all observations available for data analysis take on integer values. (Measurements can, for example, be expressed in an integer number of smallest possible increments above a nominal value.) One possible model for such data is that they arise from rounding a random sample from a Normal process with mean μ and standard deviation σ. With this model, the probability that n observations $X_1, X_2, ..., X_n$ take the integer values $x_1, x_2, ..., x_n$ is

$$\begin{aligned} f(\mathbf{x};\mu,\sigma) &= Pr(X_1 = x_1,\ X_2 = x_2, \ldots,\ X_n = x_n) \\ &= \prod_{i=1}^{n} \left\{ \Phi\left(\frac{x_i + 0.5 - \mu}{\sigma}\right) - \Phi\left(\frac{x_i - 0.5 - \mu}{\sigma}\right) \right\} \qquad (2.1) \\ &= \prod_{i} \left\{ \Phi\left(\frac{i + 0.5 - \mu}{\sigma}\right) - \Phi\left(\frac{i - 0.5 - \mu}{\sigma}\right) \right\}^{n_i}, \qquad (2.2) \end{aligned}$$

where $\Phi(x)$ is the standard normal cumulative probability function, the product in (2.2) is over integer values i and n_i is the number of observed values which equal the integer i. The expression (2.2) can be termed the *likelihood function.* It will be convenient to work with the natural logarithm of expression (2.2) and thus define the *log likelihood function* by

$$L(\mu,\sigma) = \sum_i n_i * \ln\left\{ \Phi\left(\frac{i + 0.5 - \mu}{\sigma}\right) - \Phi\left(\frac{i - 0.5 - \mu}{\sigma}\right) \right\}. \qquad (2.3)$$

Finally, define

$$L^*(\mu) = \sup_{\sigma>0} L(\mu,\sigma)\ . \qquad (2.4)$$

Then $L^*(\mu) \leq 0$ is often called the *profile loglikelihood function* for μ, and for fixed μ can be explained as the "maximum" or supremum value of $L(\mu,\sigma)$ over $\sigma > 0$.

2.3 Approximate Maximizers of $L(\mu, \sigma)$

Apparently, there are no closed forms for maximizers μ and σ of the expression (2.3). One method to get approximate maximizers $(\hat{\mu}, \hat{\sigma})$ is to apply the mean value theorem, which says that for $F(x)$ a differentiable function on (a, b)

$$F(b) - F(a) = (b-a) * F'(a + \epsilon * (b-a)) \text{ for some } \epsilon \in (0,\ 1),$$

where $F'(x)$ is the first derivative of $F(x)$. Putting $a = \dfrac{(i - 0.5 - \mu)}{\sigma}, b = \dfrac{(i + 0.5 - \mu)}{\sigma}$, $F(x) = \Phi(x)$, and $F'(x) = \phi(x)$ into above expression, we get

$$\left\{\Phi\left(\frac{i+0.5-\mu}{\sigma}\right) - \Phi\left(\frac{i-0.5-\mu}{\sigma}\right)\right\} = \frac{1}{\sigma} * \phi\left(\frac{i-0.5-\mu+\epsilon}{\sigma}\right). \tag{2.5}$$

for some $\epsilon \in (0, 1)$ and $\phi(x)$ the standard Normal probability density function.
As a convenient and simple approximation, we might let $\epsilon = 0.5$ in above equation and get

$$\left\{\Phi\left(\frac{i+0.5-\mu}{\sigma}\right) - \Phi\left(\frac{i-0.5-\mu}{\sigma}\right)\right\} \doteq \frac{1}{\sigma} * \phi\left(\frac{i-\mu}{\sigma}\right). \tag{2.6}$$

It is clear that under some circumstances this approximation is a poor one. For example, putting $\sigma = 0.001$, $\mu = 0.3$ and $i = 0$ into (2.6), we get essentially 1 on the left side and essentially 0 on the right side. As a second example, suppose that again $i = 0$ and $\sigma = 0.001$ but μ changes from 0.3 to 0. Then the right side of (2.6) is 398.9, which is much bigger than 1. The point is that approximation (2.6) can be poor if σ is small. On the other hand approximation (2.6) works very well when $\sigma \geq 2$ and $|\dfrac{(i-\mu)}{\sigma}| \leq 3$.

Substituting expression (2.6) into equation (2.3) produces

$$L(\mu, \sigma) \doteq \sum_i n_i * \ln\left\{\frac{1}{\sigma} * \phi\left(\frac{i-\mu}{\sigma}\right)\right\}. \tag{2.7}$$

(The rounded data log likelihood is approximately what one would get treating the rounded values as if they were exact Normal observations.) If we take partial derivatives with respect to μ and σ in expression (2.7) and set them to 0, we get approximate

maximum likelihood estimates of μ and σ, $\bar{x}$ and $\sqrt{\sum_i \frac{n_i * (i-\bar{x})^2}{n}}$ respectively, the maximum likelihood estimates for a Normal model supposing the integer observations to be exact, not produced by rounding.

2.4 Special Cases in the Maximization of $L(\mu, \sigma)$

There are two special forms of data $(x_1, x_2, ..., x_n)$ that cause problems in the numerical maximization of $L(\mu, \sigma)$. We will call these Case 1 and Case 2.

2.4.1 Case 1

Case 1 is the situation where a sample contains only one distinct value, say i_0 for convenience. When such happens, expression (2.3) strictly speaking has no maximum value for $\sigma > 0$. But the supremum value of $L(\mu, \sigma)$ is nearly achieved for any μ in a particular interval (when σ is small enough). More precisely, if $\mu \in (i_0 - 0.5, i_0 + 0.5)$, then

$$\lim_{\sigma \to 0} \left\{ \Phi\left(\frac{i_0 + 0.5 - \mu}{\sigma}\right) - \Phi\left(\frac{i_0 - 0.5 - \mu}{\sigma}\right) \right\}^n = 1,$$

or

$$\lim_{\sigma \to 0} L(\mu, \sigma) = 0.$$

Figure 2.1 shows a typical graph of $L^*(\mu)$ versus μ when a sample contains only one distinct value. From Figure 2.1, it is easy to see that the supremum value of $L(\mu, \sigma)$ is 0 (i.e. $\sup_{\mu \in R} \sup_{\sigma > 0} L(\mu, \sigma) = 0$), and this value is approached only when μ is between $i_0 - 0.5$ and $i_0 + 0.5$. We can also see that there are two discontinuities in the graph, which occur at the points $\mu = i_0 - 0.5$ and $\mu = i_0 + 0.5$. The reason for these discontinuities is that the supremum value of $L(i_0 \pm 0.5, \sigma)$ is $-n*\ln(2)$, which is much smaller than 0.

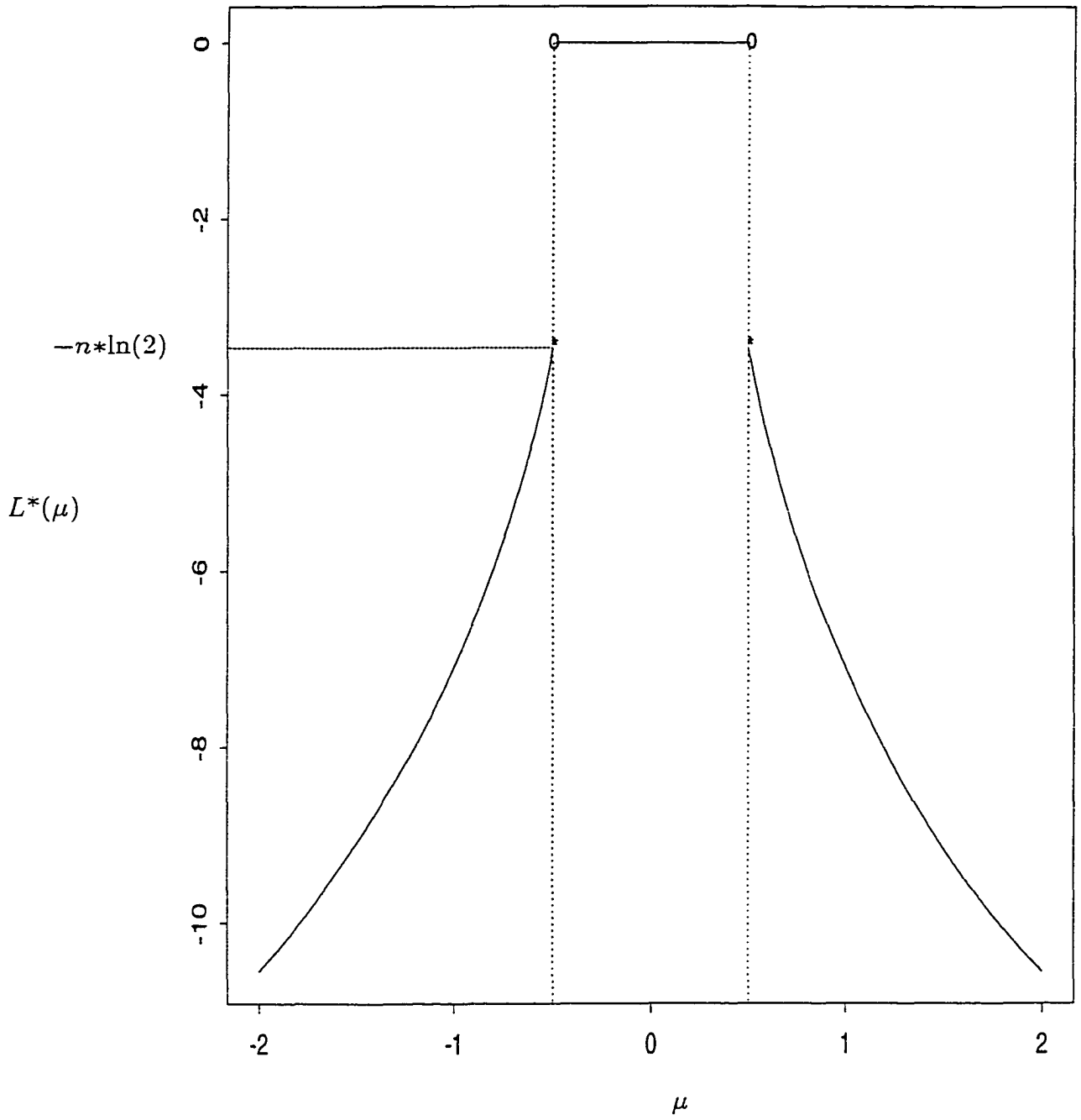

Figure 2.1 Representative graph of $L^*(\mu)$ when a sample contains only one distinct value. This particular graph is for a case where a sample of size $n = 5$ contains only the value 0.

Vardeman and Jensen(1989) concluded that if there is only one value i_0 observed in a given sample, we might still use $\hat{\mu} = \bar{x} = i_0$ and $\hat{\sigma} = 0$ as "maximum likelihood estimates" for the parameters μ and σ, but at the same time should recognize that many different (μ,σ) pairs with small σ and $\mu \in (i_0 - 0.5, i_0 + 0.5)$ are essentially indistinguishable.

2.4.2 Case 2

Case 2 is the situation where a sample contains only two different values with the sample range 1, say the integers i_0 and i_0+1. Again, a maximum value of $L(\mu, \sigma)$ is not achieved. But in this case, the supremum value of $L(\mu, \sigma)$ is

$$n_{i_0} * \ln(\frac{n_{i_0}}{n}) + n_{i_0+1} * \ln(\frac{n_{i_0+1}}{n}) \ , \tag{2.8}$$

which is approached if μ and σ are chosen so that

$$P_0 \doteq \frac{n_{i_0}}{n} \qquad \text{and} \qquad P_1 \doteq \frac{n_{i_0+1}}{n} \ , \tag{2.9}$$

where P_0 and P_1 are the probabilities assigned to integers i_0 and (i_0+1) by the rounded Normal distribution. (That is, $P_k = \Phi\left(\frac{i_0 + k + 0.5 - \mu}{\sigma}\right) - \Phi\left(\frac{i_0 + k - 0.5 - \mu}{\sigma}\right)$ for $k = 0, 1$.)

The two conditions in (2.9) are equivalent to the three constraints $\Phi(\frac{i_0 - 0.5 - \mu}{\sigma}) \doteq 0$, $\Phi(\frac{i_0 + 1.5 - \mu}{\sigma}) \doteq 1$, and $\Phi(\frac{i_0 + 0.5 - \mu}{\sigma}) = \frac{n_{i_0}}{n}$ simutaneously. Using these constraints and the facts that $\Phi(-3) \doteq 0$ and $\Phi(3) \doteq 1$, the following two results hold.

Result(i) If $\frac{n_{i_0}}{n} > \frac{1}{2}$, then $\sigma = \frac{i_0 + 0.5 - \mu}{\Phi^{-1}(\frac{n_{i_0}}{n})}$ makes the function $L(\mu, \sigma)$ approximate its supremum value when μ belongs to the interval

$$(\ i_0 + 0.5 - \frac{\Phi^{-1}(\frac{n_{i_0}}{n})}{\Phi^{-1}(\frac{n_{i_0}}{n}) + 3} \ , \ i_0 + 0.5 \ ) \ .$$

Result(ii) If $\frac{n_{i_0}}{n} < \frac{1}{2}$, then $\sigma = \frac{i_0 + 0.5 - \mu}{\Phi^{-1}(\frac{n_{i_0}}{n})}$ makes the function $L(\mu, \sigma)$ approximate its supremum value when μ belongs to the interval

$$(\ i_0 + 0.5 \ , \ i_0 + 1.5 - \frac{3}{3 - \Phi^{-1}(\frac{n_{i_0}}{n})} \ ) \ .$$

Figures 2.2 and 2.3 show the two types of graphs possible for the function $L^*(\mu)$ in Case 2 when $n_{i_0} \neq n_{i_0+1}$. When $n_{i_0} > n_{i_0+1}$ the graph is similar to Figure 2.2 and when $n_{i_0} < n_{i_0+1}$ the graph is similar to Figure 2.3. There is a point of discontinuity at $\mu = i_0 + 0.5$ in both pictures. This is because $L^*(i_0 + 0.5) = -n*\ln(2)$, and this equals the value in display (2.8) only when $\frac{n_{i_0}}{n} = \frac{1}{2}$. In other words, if the sample contains only two values with a range of 1, the function $L^*(\mu)$ will be continuous only when the two different values have the same observed frequency (i.e. when $n_{i_0} = n_{i_0+1}$).

2.5 Construction of the Intervals for μ

Two methods of making confidence intervals for μ will be discussed in this section.

First, if we ignore rounding and treat the rounded data as "exact" Normal data, then the usual $(1-\alpha)$ level confidence interval for μ is

$$[\ \bar{x} - \frac{s}{\sqrt{n}} * t_{(n-1,\ 1-\frac{\alpha}{2})},\ \bar{x} + \frac{s}{\sqrt{n}} * t_{(n-1,1-\frac{\alpha}{2})}\ ]\ , \tag{2.10}$$

where $s = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{(n-1)}}$, and $t_{(n-1,1-\frac{\alpha}{2})}$ is $(1-\frac{\alpha}{2})$ quantile of the t distribution with $(n-1)$ degrees of freedom.

Second, a method of explicity using the rounded data joint distribution in display (2.1) to construct the confidence intervals for μ is to invert likelihood ratio tests of $H_0 : \mu = \mu_0$ and apply the (asymptotic) chi-square null distribution associated with the likelihood ratio test statistic (see, for example, Bickel and Doksum (1977) page 229). That is, if $\mu = \mu_0$, then

$$-2 * \ln(\ \frac{\sup_{\sigma>0} f(\mathbf{x}; \mu_0, \sigma)}{\sup_{\sigma>0} \sup_{\mu \in R} f(\mathbf{x}; \mu, \sigma)}\ ) \dot{\sim} \chi^2_{(1)}$$

(where $f(\mathbf{x}; \mu, \sigma)$ is the likelihood function described in expression (2.1)). Or using the notation in this paper, if $\mu = \mu_0$

$$-2 * (\ L^*(\mu_0) - \sup_{\mu \in R} L^*(\mu)\ ) \dot{\sim} \chi^2_{(1)}\ . \tag{2.11}$$

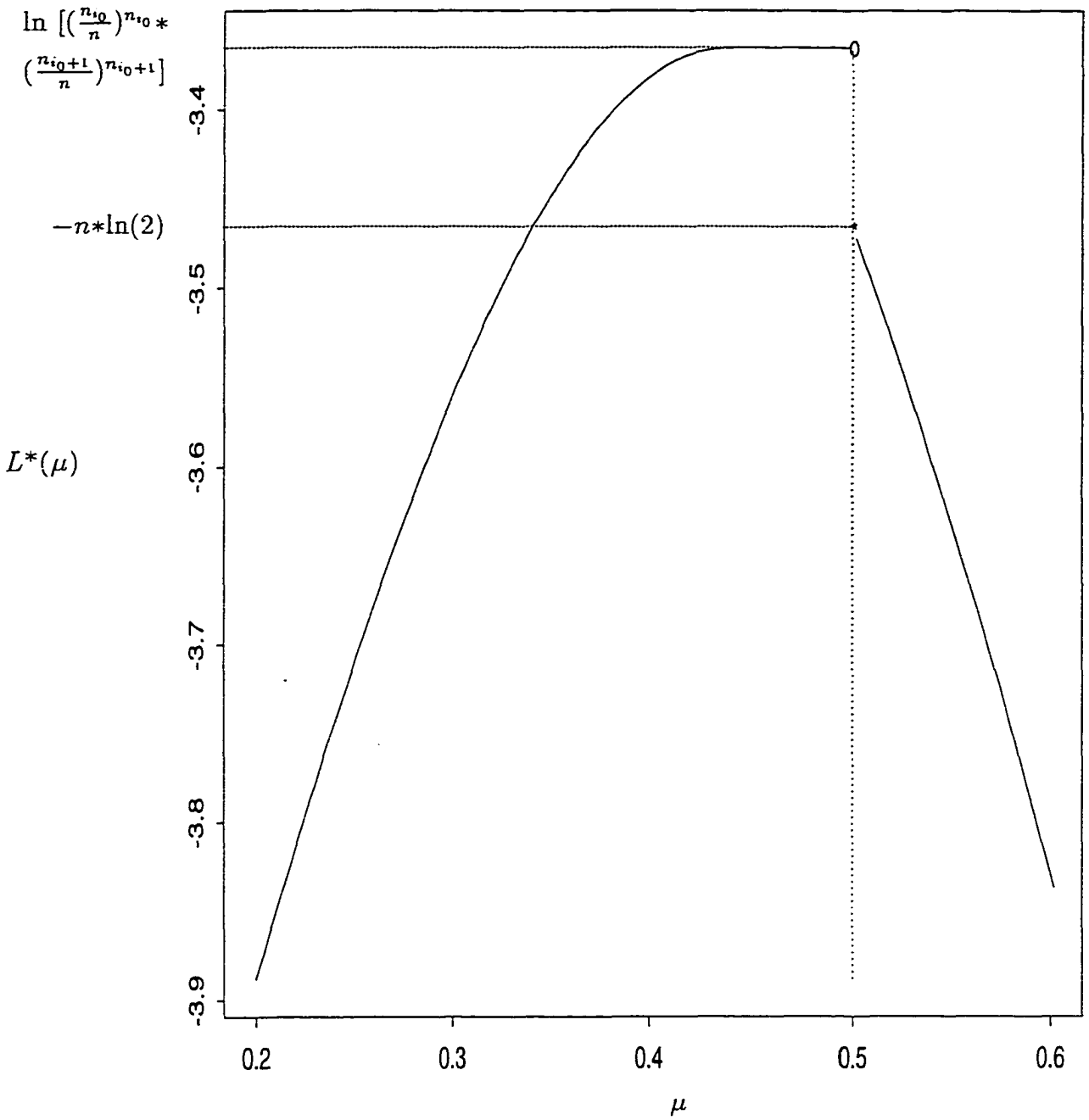

Figure 2.2 Representative graph of $L^*(\mu)$ when a sample contains two distinct values with range of 1. This particular graph is for a case where a sample of size $n = 5$ gives $n_0 = 3$, $n_1 = 2$.

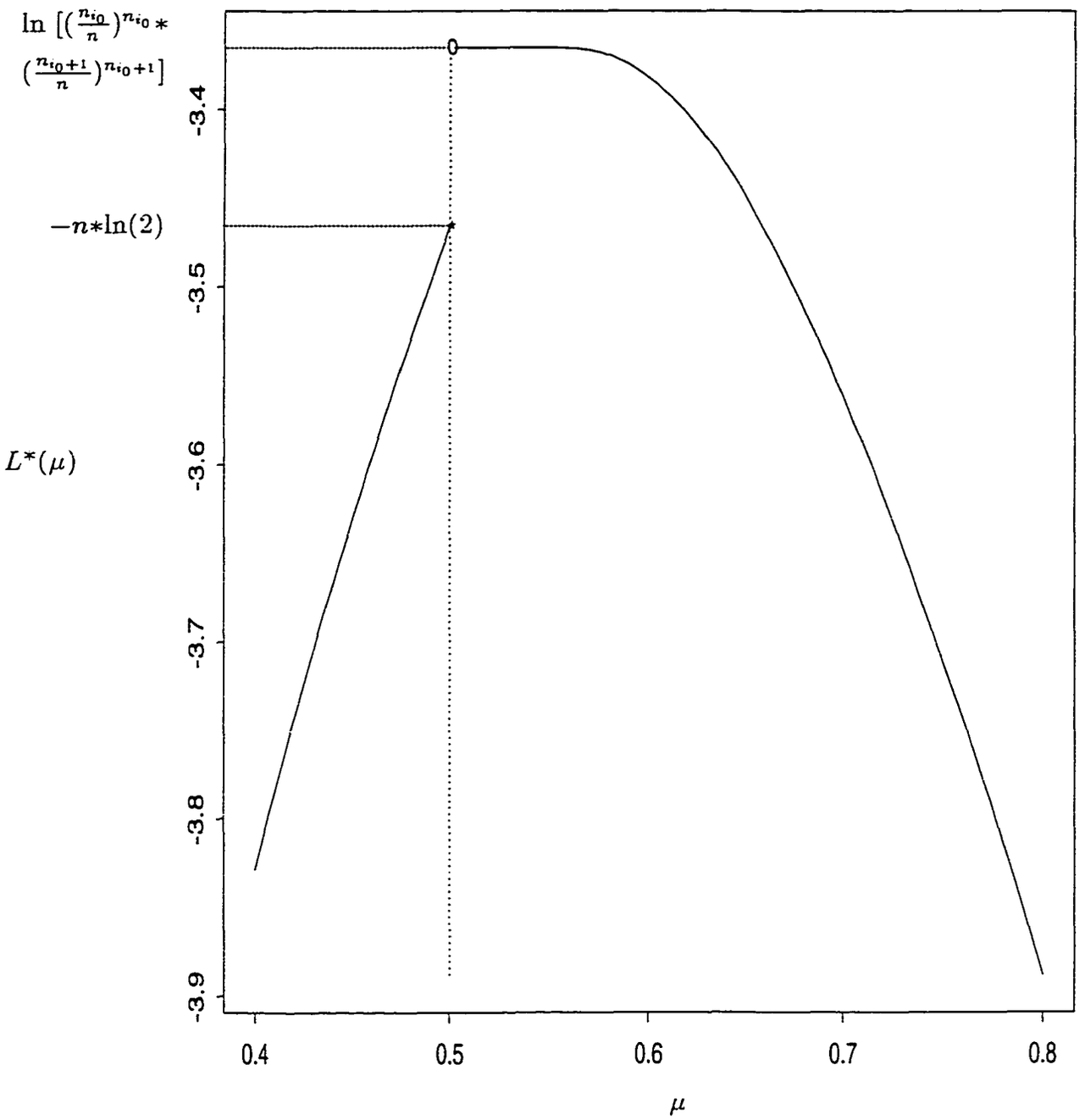

Figure 2.3 Representative graph of $L^*(\mu)$ when a sample contains two distinct values with range of 1. This particular graph is for a case where a sample of size $n = 5$ gives $n_0 = 2$, $n_1 = 3$.

(Note that $\sup_{\mu \in R} L^*(\mu)$ is the supremum log-likelihood value.)

Given a desired significance level α, we can construct an interval of means μ satisfying the inequality

$$\sup_{\mu \in R} L^*(\mu) - L^*(\mu) \leq \frac{1}{2} * \chi^2_{(1,1-\alpha)} , \tag{2.12}$$

and conclude from the approximation (2.11) that the resulting interval has (asymptotic or) approximate coverage probability $(1 - \alpha)$.

For the rest of this section, we consider the nature of the interval specified by (2.12) in Case 1 and Case 2 discussed in the previous section.

2.5.1 Case 1

From Figure 2.1, it is easy to see that if $n * \ln(2) > \frac{1}{2} * \chi^2_{(1,1-\alpha)}$, then any $\mu \in (i_0 - 0.5, i_0 + 0.5)$ will satisfy inequality (2.12). That is, the interval specified by (2.12) is $(i_0 - 0.5, i_0 + 0.5)$. For example, if $i_0 = 0$, $n = 5$ and $\alpha = 0.05$, the above inequality holds and hence the confidence interval for μ is $(-0.5, 0.5)$. In fact, if $n \geq 3$ and $\alpha \geq 0.05$, then the confidence interval (2.12) for Case 1 is always $(i_0 - 0.5, i_0 + 0.5)$. This interval is much wider than the t interval degenerate about i_0 (that is prescribed by equation (2.10) since $\bar{x} = i_0$ and $s = 0$ in Case 1). (Note by the way, that for rounded data with a range of 0, neither of the two methods we're considering produces intervals that change with n or α.)

2.5.2 Case 2

When the range of a sample is 1 (and the graph of $L^*(\mu)$ looks like Figure 2.2 or Figure 2.3) a numerical search is required to find the interval described by display (2.12). However, in part (A) of the Appendix, we provide useful empirical approximations for the end points of the likelihood-based intervals for this case. (Part (B) of the Appendix

provides corresponding approximations for the situation where the range of a sample is 2 or more.)

2.6 Simulations

In this section, we use the two methods discussed in previous section to find intervals for the parameter μ from simulated Normal samples rounded to the nearest integer. First, we randomly select a sample of size n from a Normal distribution with mean μ and standard deviation σ. After getting the exact data, we round these to integers, then apply formulas (2.10) and (2.12) to compute the confidence intervals for the parameter μ. The last step is to check whether the intervals contain the μ or not. If, for a given method, the answer is "yes," then we increment a counter (t or c respectively for methods (2.10) and (2.12)) by 1. If the answer is "no," the counter is not incremented. This is repeated 1,000 times, and so we obtain counts t_{1000} and c_{1000}. The ratios $\frac{t_{1000}}{1000}$ and $\frac{c_{1000}}{1000}$ are then Monte Carlo estimates of the actual coverage probabilities for the nominally (approximately) $(1-\alpha)$ level confidence procedures.

To illustrate, suppose an initial random sample of size $n = 10$ taken from the Normal distribution with $\mu = 1.0$ and $\sigma = 0.25$ produces rounded data with $n_0 = 2$ and $n_1 = 8$. After computing, we can get (0.556,1.044) from formula (2.10) and (0.5,0.996) from formula (2.12) when $\alpha = 0.10$. It is then obvious that only the t-interval contains the true parameter $\mu = 1.0$, so we set the counters $t_1 = 1$ and $c_1 = 0$. Then suppose that a second rounded Normal sample contains $n = 10$ values $i_0 = 1.0$. For this second sample, the t-interval is degenerate at 1. and the interval defined by (2.12) is (0.5, 1.5). In this case we will increment both t and c by 1 and have $t_2 = 2$ and $c_2 = 1$. And so on through 1,000 samples.

Different values of μ, σ, n, and α were used in the simulations to provide a thorough comparison of the two methods. We considered $\mu = [0,\ 1.0](0.1)$, $\sigma \in \{0.01,\ 0.25,\ 0.5,$

1.0, 1.5, 2.0}, $n = [5, 20](5)$, and $\alpha \in \{0.05, 0.10, 0.20\}$, where $[a, b](c)$ means the values from a to b with increment c. Figures 2.4 and 2.5 are graphs of the estimated coverage probabilities for the t-intervals and χ^2-intervals (2.12) for the $n = 5$ and $n = 15$ cases. In those graphs, the solid line indicates the estimated coverage probability for the t-intervals, and the dashed line indicates the estimated coverage probability for the χ^2-intervals. (The actual coverage probabilities are symmetric about $\mu = 0.5$, so for a given σ, n and α we have averaged estimated coverage probabilities for μ and $(1 - \mu)$ before plotting.)

After analyzing these graphs and similar ones for the $n = 10$ and $n = 20$ cases, we can make several conclusions:

(1) When σ is small, say $\sigma = 0.01$, the graphs display basically the same pattern for all combinations of n and α. We can also see that the coverage probability for the likelihood method (2.12) is almost always bigger than that for the t method, except for the special points $\mu = 0$, 0.5, and 1.0. These points deserve explanation.

First, we focus on the coverage probabilities for the likelihood-based intervals (indicated by the dashed lines on the Figures 2.4 and 2.5). If $0.0 \leq \mu < 0.5$ and σ is "small"(here the word "small" means that σ satisfies $\Phi(\frac{0.5 - \mu}{\sigma}) - \Phi(\frac{-0.5 - \mu}{\sigma}) \doteq 1$), then all of the "exact" sample will typically fall below 0.5 and the rounded values will all be $i_0 = 0$. Similar reasoning applies to the interval $0.5 < \mu \leq 1.0$, but this time all rounded data will typically have the value 1. Because the interval for μ in Case 1 is always $(i_0 - 0.5, i_0 + 0.5)$, the true parameter μ is essentially always contained in the interval. That's why the estimated coverage probabilities for the likelihood method (2.12) always have the value 1. But when $\mu = 0.5$ and σ is small, the values in the rounded sample will typically be a (binomial) mixture of 0's and 1's, so the coverage probability will be smaller than 1.

Second, we check the solid lines on the pictures and consider the t interval coverage.

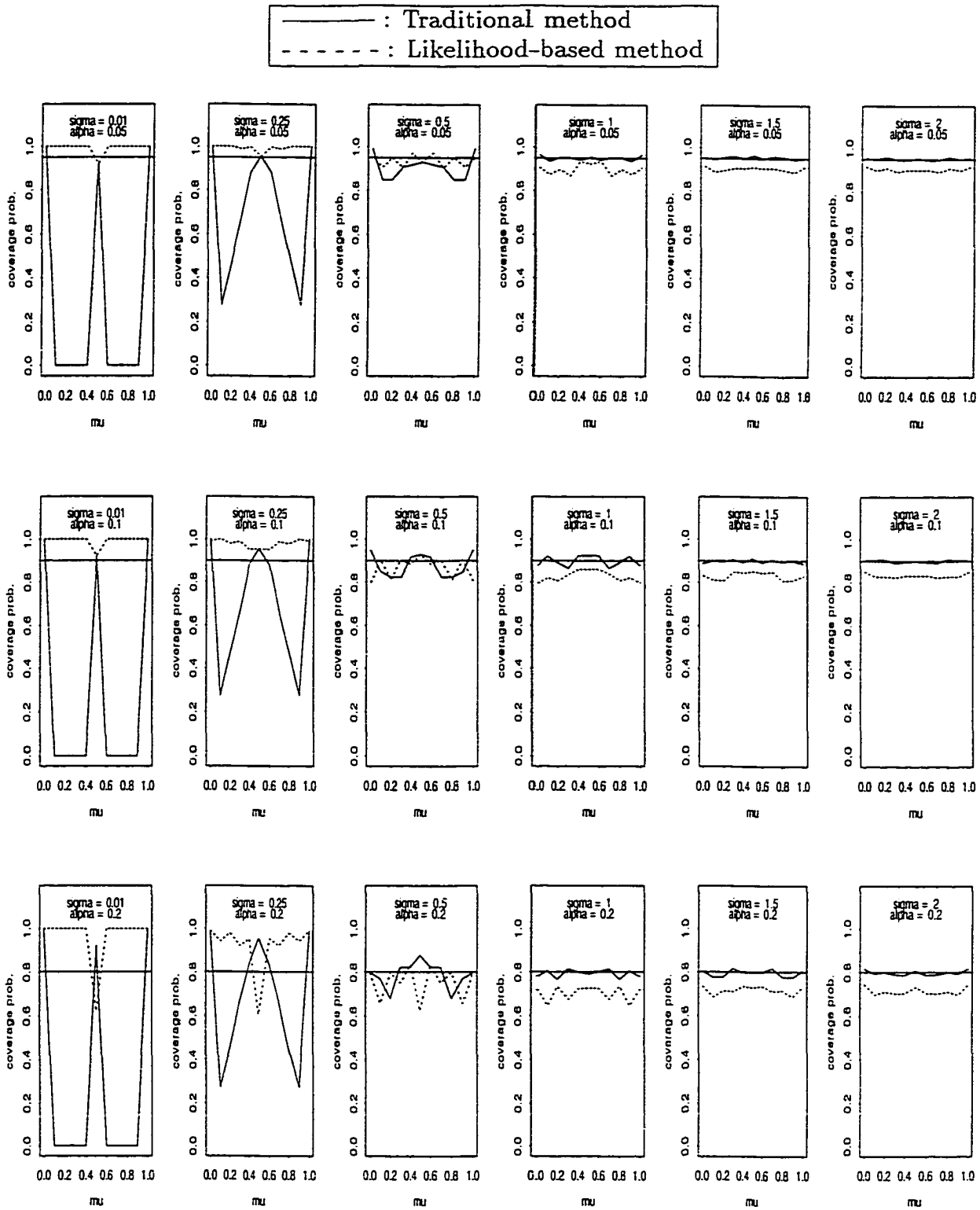

Figure 2.4 Estimated coverage probabilty for sample size $n = 5$.

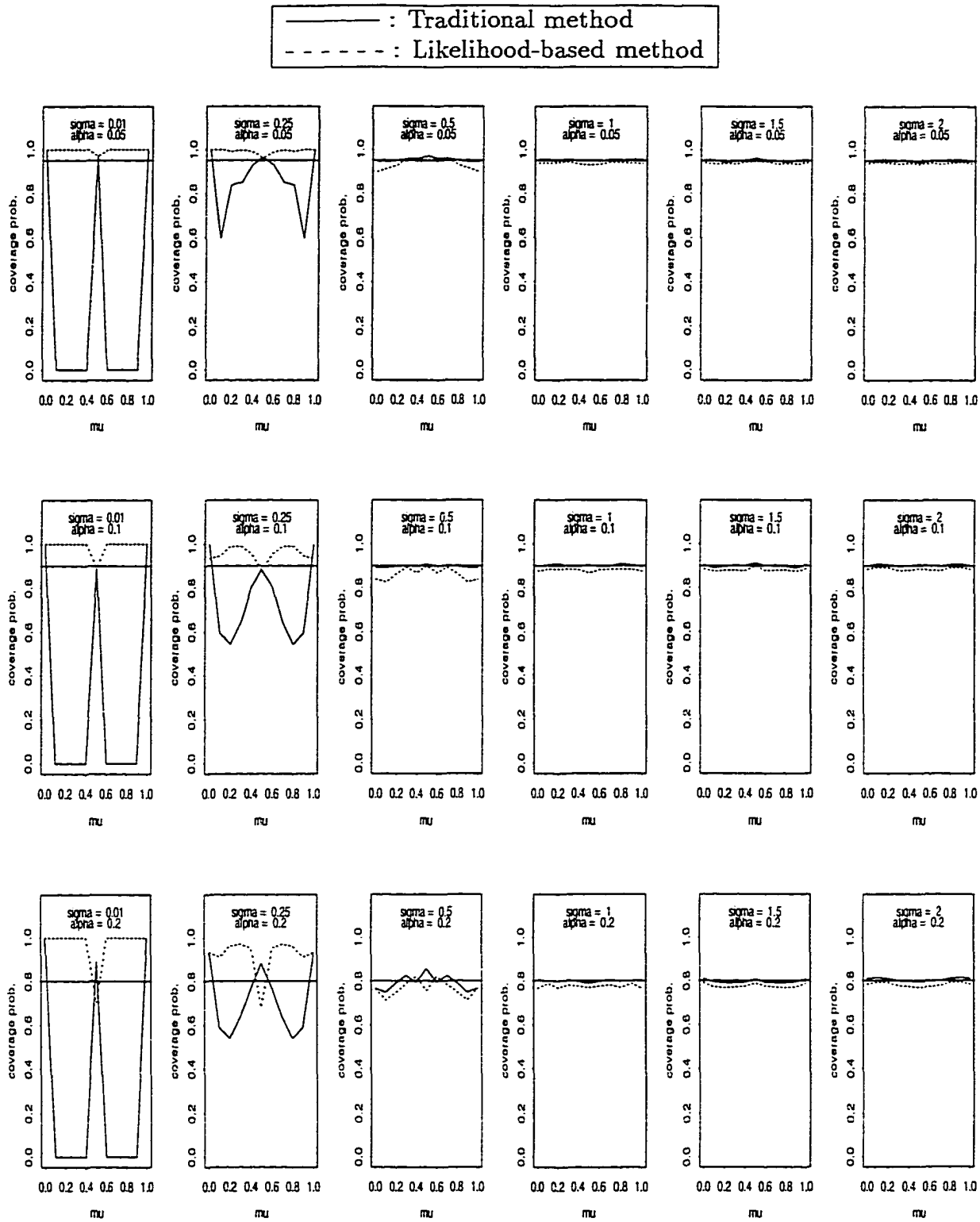

Figure 2.5 Estimated coverage probability for sample size $n = 15$.

The solid lines indicate coverage probabilities larger than 0 at $\mu = 0,\ 0.5$ and 1.0, but 0 probabilities for other μ. Since when σ is small, the rounded samples all tend to contain the single value 0 if $\mu \in [0, 0.5)$ or the value 1 if $\mu \in (0.5, 1.0]$, the t method tends to produce intervals degenerate at $\bar{x} = 0$ or $\bar{x} = 1$. So the method brackets μ with probability near 1 only when $\mu = 0$ or $\mu = 1$. This explains why the coverage probability is always 1 at the two points $\mu = 0$ and $\mu = 1$, but is 0 for $\mu \in (0, 0.5)$ and $\mu \in (0.5, 1)$. As to the situation when $\mu = 0.5$, the same kind of reasoning applies here as was applied to the method (2.12).

(2) When σ grows bigger, say $\sigma \geq 0.5$, then all the graphs indicate that the coverage probability for the t method is closer to the nominal probability $(1 - \alpha)$ than that for the likelihood method. The graphs show that the actual coverage probabilities for the likelihood method approach the nominal level $(1 - \alpha)$ as n increases. But they are still lower than the values one gets from t method.

2.7 Improving the Coverage Probability Calibration of the Likelihood-Based Intervals and Final Comparisons to the t-Method

The simulation results in previous section show that for small n the coverage probability for the method based on inversion of the likelihood ratio tests is much lower than the (desired) nominal value $(1 - \alpha)$ when applied exactly as in (2.12), particularly for large σ. This suggests that for small n, adjustments to the $\chi^2_{(1,1-\alpha)}$ values appearing in (2.12) are needed. In other words, one might try to find an appropriate value $c(n, \alpha)$ to replace $\chi^2_{(1,1-\alpha)}$ in (2.12) so that

$$Pr[\ -2 * (L^*(\mu) - M) \leq c(n, \alpha)\ ] \geq (1 - \alpha)\ ,$$

for most (μ, σ) pairs. (M continues to be the supremum value of the log-likelihood function.) If this can be done, we can then use the likelihood-based intervals defined by

$$\sup_{\mu \in R} L^*(\mu) - L^*(\mu) \leq \frac{1}{2} * c(n, \alpha) \ . \tag{2.13}$$

It might be expected that for large σ, the likelihood ratio test based on "rounded" data is equivalent to the likelihood ratio test of the same hypothesis of $H_0 : \mu = \mu_0$ based on "exact" data. The (standard) development of this exact data test (see Bickel and Doksum (1977, pages 209-212)) shows that the exact data version of $-2 * (L^*(\mu) - M)$ is

$$n * \ln \left[\ 1 + \frac{n}{n-1} (\frac{\bar{x} - \mu}{s})^2 \ \right].$$

Now, with exact normal data $T = \dfrac{\sqrt{n}(\bar{x} - \mu)}{s}$ is well known to have a t_{n-1} distribution. This suggests that a choice of $c(n, \alpha)$ likely to produce correct large σ coverage probabilities is

$$c(n, \alpha) = n * \ln \left(\frac{t^2_{(n-1, 1-\frac{\alpha}{2})}}{n-1} + 1 \right),$$

for $t_{(n-1, 1-\frac{\alpha}{2})}$, the $1 - \frac{\alpha}{2}$ quantile of the t_{n-1} distribution.

Table 2.1 gives $c(n, \alpha)$ values for different combinations of n and α. We have applied these $c(n, \alpha)$ values to make Figures 2.6 to 2.9 giving estimated coverage probabilities for sample sizes $n \in \{2, 5, 10, 15\}$, means $\mu = [0, 1.0](0.1)$, and standard deviations $\sigma \in \{0.01, 0.25, 0.5, 1, 5, 10\}$. In these figures, the solid lines indicate the coverage probability from t-method, and the dashed lines indicate the coverage probability from likelihood-based method when $c(n, \alpha)$ is used. Comparing Figure 2.4 to Figure 2.7 and Figure 2.5 to Figure 2.9, we can see that when $\sigma \geq 0.5$, using the $c(n, \alpha)$ values makes the coverage probabilities much closer to the desired value $(1 - \alpha)$, and does so without changing much when σ is small (e.g. $\sigma = 0.01$).

In addition to estimating coverage probabilities we also ran simulations to compare average interval lengths for the t-method and likelihood-based method. Tables 2.2, 2.3,

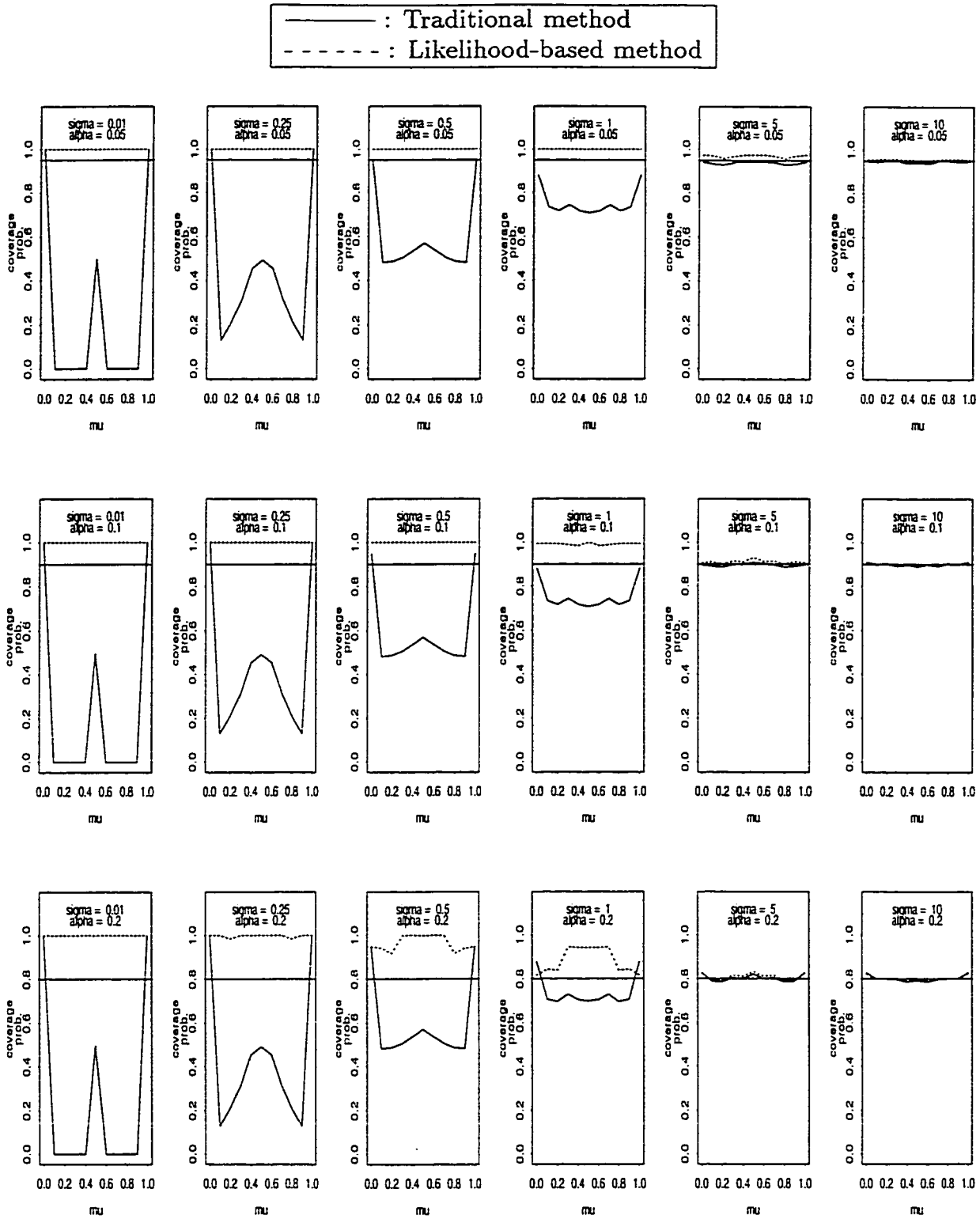

Figure 2.6 Estimated coverage probabilty for sample size $n = 2$.

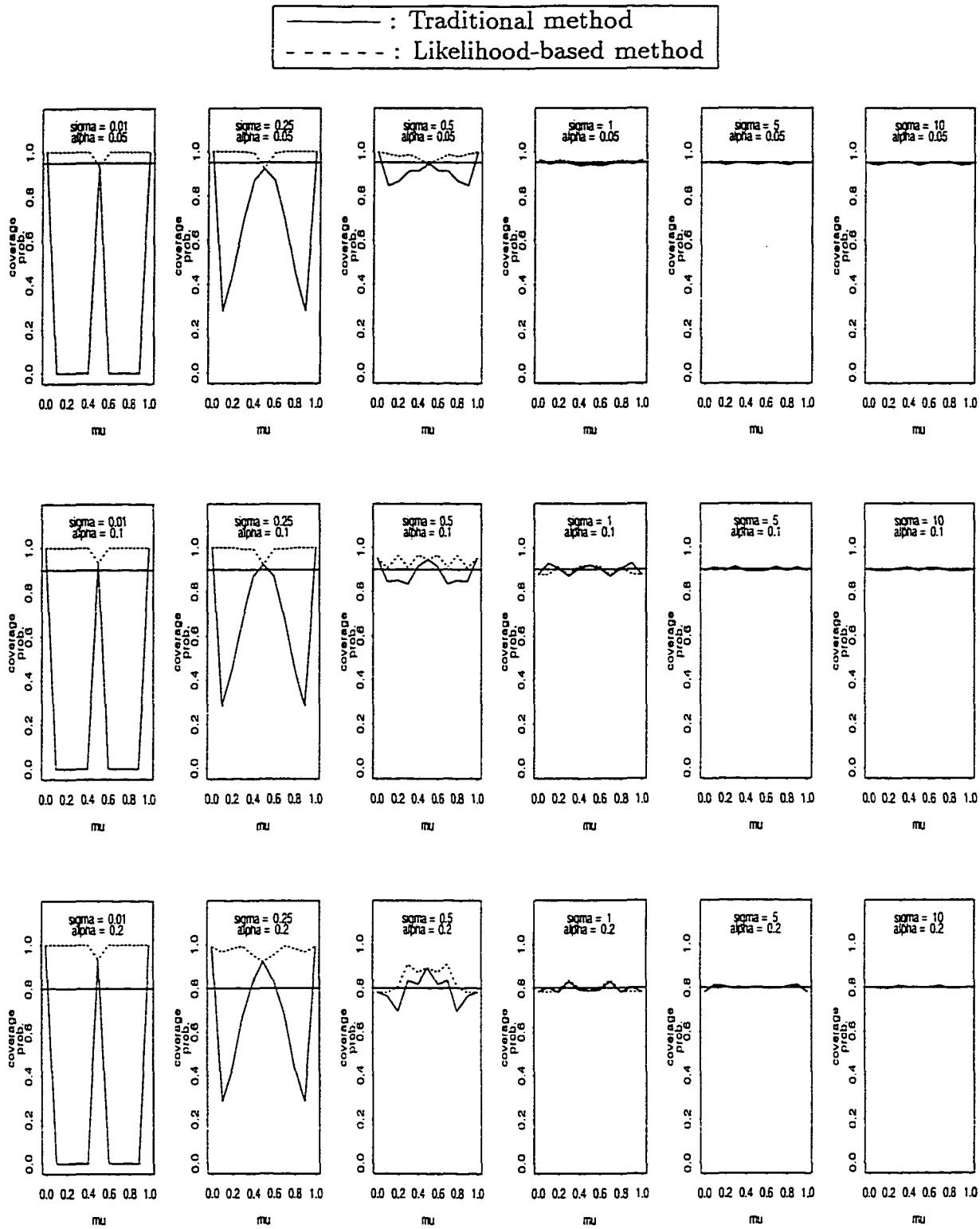

Figure 2.7 Estimated coverage probability for sample size $n = 5$.

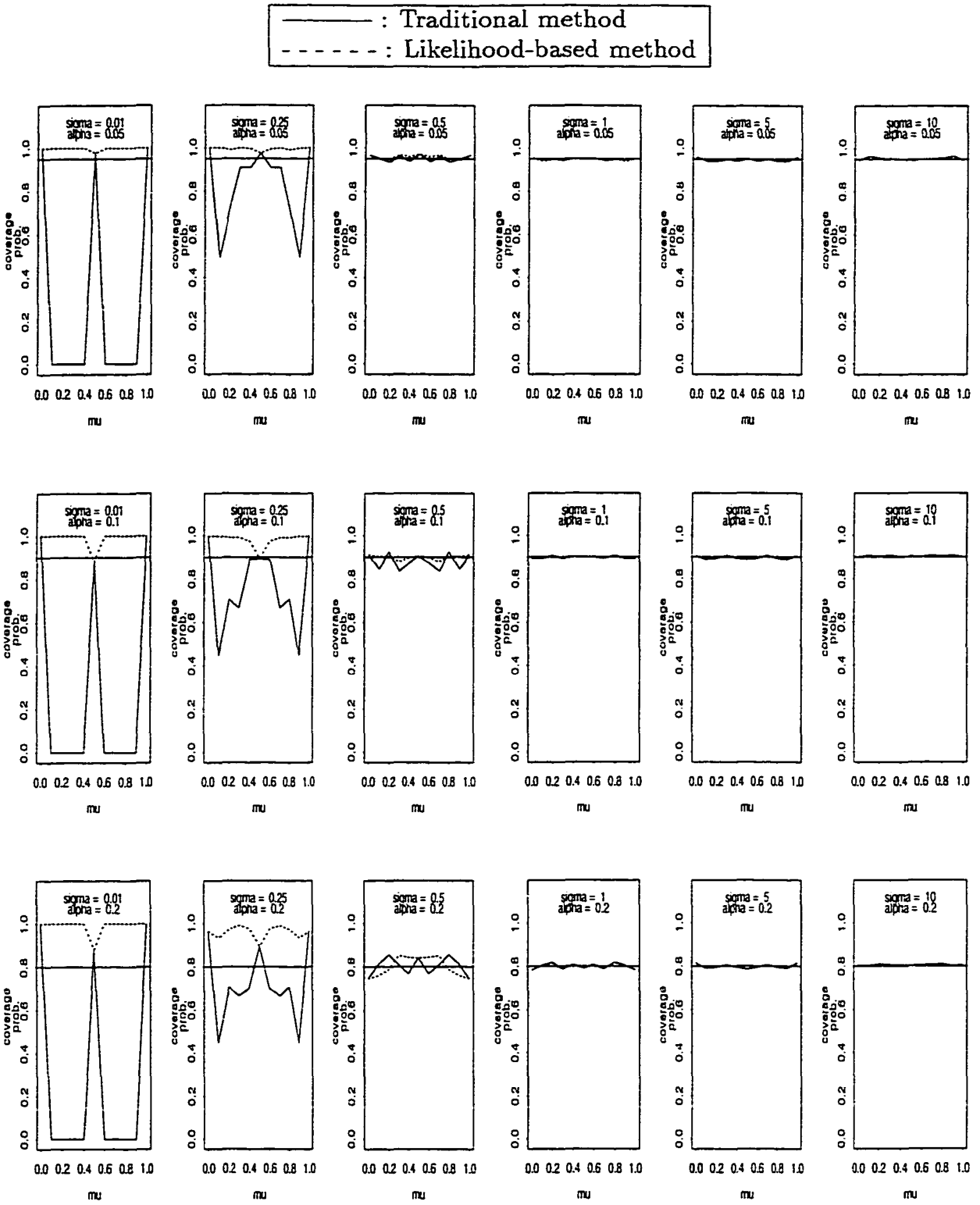

Figure 2.8 Estimated coverage probability for sample size $n = 10$.

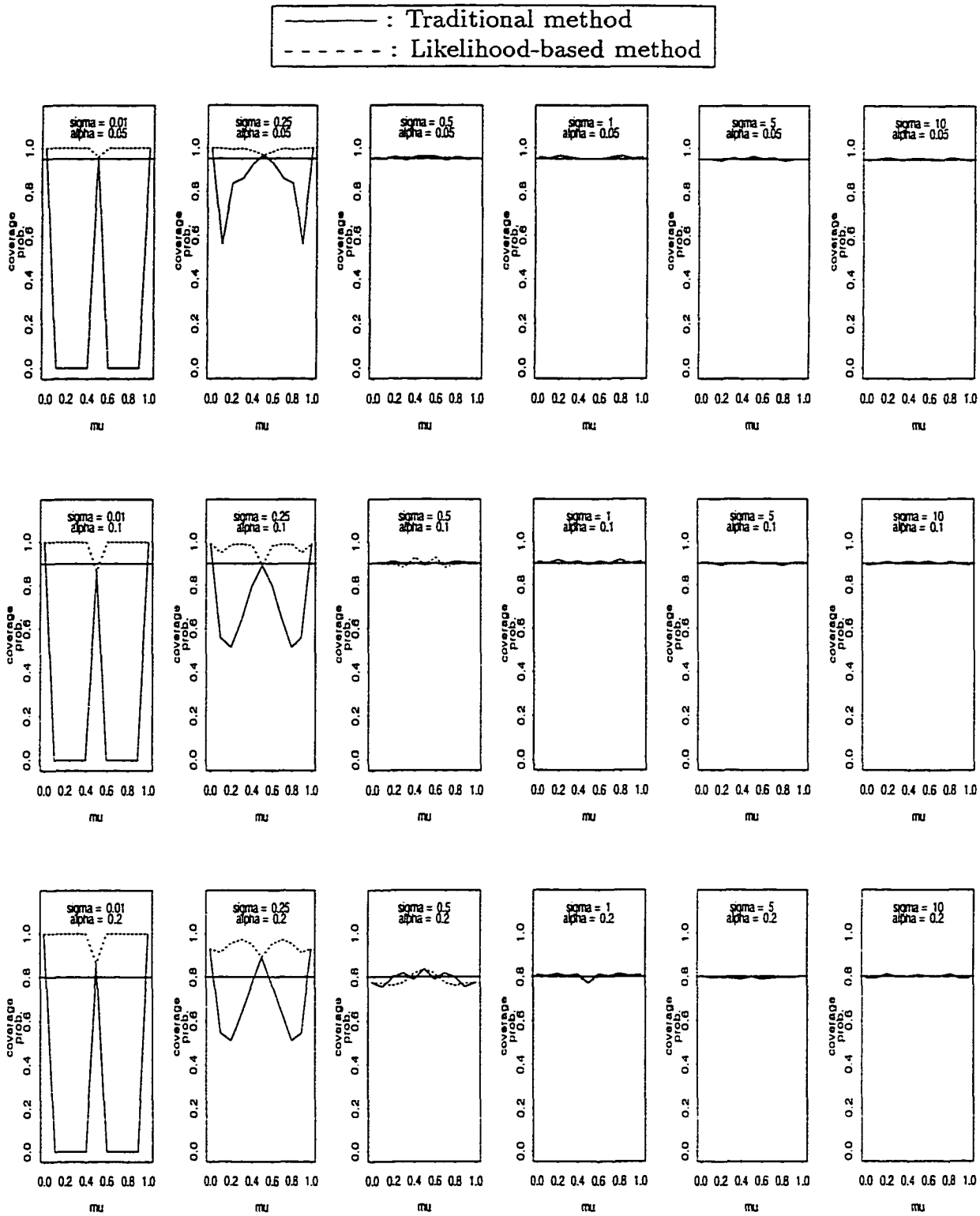

Figure 2.9 Estimated coverage probability for sample size $n = 15$.

Table 2.1 $c(n, \alpha)$ values for different n and α.

n	α		
	0.05	0.10	0.20
2	10.18	7.42	4.70
3	6.98	4.98	3.06
4	5.90	4.18	2.55
5	5.37	3.80	2.31
6	5.05	3.57	2.17
7	4.84	3.42	2.08
8	4.70	3.31	2.01
9	4.59	3.23	1.96
10	4.50	3.17	1.93
15	4.26	3.00	1.82
∞	3.84	2.71	1.64

and 2.4 present average lengths for 1,000 t intervals (from(2.10)) and 1,000 likelihood-based intervals (from(2.13)) for various μ, σ, and n.

The general character of the results in these tables is as follows:

(1) At $\mu = 0.0$ and 0.25, the t method average length is much smaller than the likelihood or "χ^2" method average length for $\sigma = 0.01$ and 0.25. And as σ grows, the mean lengths become quite similar. The difference in lengths when σ is small derives from the fact that many of the samples have range 0, and the poor coverage probabilities for the t method evident in Figures 2.6-2.9 show the impact of the small mean lengths for the t method.

(2) At $\mu = 0.5$ the average lengths for two methods are close to each other.

2.8 Conclusion

In light of the whole discussion in this paper, we reach the following general conclusions about how to handle crudely gaged data in the interval estimation of μ.

Table 2.2 The average simulated lengths for t and the likelihood methods for $\mu = 0.0$.

$\mu = 0.0$											
σ		0.01		0.25		0.50		1.00		5.00	
n	α	t	"χ^2"	t	"χ^2"	t	"χ^2"	t	"χ^2"	t	"χ^2"
2	0.05	0.000	6.167	0.889	6.596	6.760	9.726	14.574	16.050	69.795	70.098
	0.10	0.000	3.094	0.442	3.304	3.359	4.846	7.242	7.982	34.681	34.835
	0.20	0.000	1.571	0.215	1.669	1.637	2.392	3.530	3.907	16.906	16.997
3	0.05	0.000	1.553	0.391	1.727	2.241	2.710	4.508	4.601	21.663	21.651
	0.10	0.000	1.124	0.266	1.232	1.521	1.859	3.060	3.126	14.702	14.697
	0.20	0.000	1.000	0.171	1.032	0.982	1.286	1.976	2.036	9.494	9.483
4	0.05	0.000	1.035	0.258	1.122	1.553	1.753	2.998	3.013	14.639	14.628
	0.10	0.000	1.000	0.191	1.026	1.148	1.345	2.217	2.234	10.826	10.817
	0.20	0.000	1.000	0.133	0.969	0.799	0.998	1.543	1.560	7.534	7.523
5	0.05	0.000	1.000	0.236	1.033	1.276	1.402	2.465	2.459	11.897	11.897
	0.10	0.000	1.000	0.181	0.979	0.980	1.108	1.893	1.889	9.135	9.143
	0.20	0.000	1.000	0.130	0.928	0.704	0.833	1.361	1.356	6.570	6.567

Table 2.3 The average simulated lengths for t and the likelihood methods for $\mu = 0.25$.

$\mu = 0.25$											
σ		0.01		0.25		0.50		1.00		5.00	
n	α	t	" χ^2 "	t	" χ^2 "	t	" χ^2 "	t	" χ^2 "	t	" χ^2 "
2	0.05	0.000	6.167	3.138	7.680	7.446	10.140	14.485	16.027	72.959	73.282
	0.10	0.000	3.094	1.559	3.837	3.700	5.051	7.198	7.971	36.254	36.418
	0.20	0.000	1.571	0.760	1.917	1.804	2.489	3.509	3.902	17.672	17.770
3	0.05	0.000	1.553	1.185	2.056	2.277	2.681	4.507	4.608	21.742	21.727
	0.10	0.000	1.124	0.804	1.434	1.545	1.837	3.059	3.131	14.755	14.748
	0.20	0.000	1.000	0.519	1.082	0.998	1.260	1.975	2.041	9.528	9.516
4	0.05	0.000	1.035	0.822	1.318	1.639	1.757	2.955	2.968	14.805	14.794
	0.10	0.000	1.000	0.608	1.090	1.212	1.330	2.185	2.200	10.948	10.940
	0.20	0.000	1.000	0.423	0.906	0.844	0.964	1.521	1.536	7.619	7.609
5	0.05	0.000	1.000	0.676	1.098	1.314	1.384	2.358	2.362	11.418	11.418
	0.10	0.000	1.000	0.519	0.943	1.009	1.081	1.810	1.817	8.767	8.775
	0.20	0.000	1.000	0.374	0.798	0.726	0.798	1.302	1.307	6.305	6.303

Table 2.4 The average simulated lengths for t and the likelihood methods for $\mu = 0.5$.

$\mu = 0.5$											
σ		0.01		0.25		0.50		1.00		5.00	
n	α	t	" χ^2 "	t	" χ^2 "	t	" χ^2 "	t	" χ^2 "	t	" χ^2 "
2	0.05	6.391	9.249	6.035	9.077	7.459	10.077	14.320	15.658	68.588	68.904
	0.10	3.176	4.608	2.999	4.529	3.706	5.019	7.116	7.786	34.082	34.242
	0.20	1.548	2.275	1.462	2.236	1.807	2.473	3.469	3.810	16.613	16.708
3	0.05	2.169	2.474	2.166	2.473	2.493	2.765	4.590	4.666	22.463	22.446
	0.10	1.472	1.691	1.470	1.691	1.692	1.887	3.115	3.170	15.244	15.237
	0.20	0.950	1.150	0.949	1.150	1.093	1.270	2.011	2.061	9.844	9.831
4	0.05	1.517	1.575	1.464	1.557	1.715	1.779	2.974	2.983	14.725	14.713
	0.10	1.122	1.183	1.083	1.177	1.268	1.334	2.199	2.210	10.889	10.880
	0.20	0.781	0.845	0.753	0.851	0.882	0.951	1.530	1.542	7.577	7.567
5	0.05	1.199	1.221	1.201	1.219	1.382	1.398	2.455	2.453	11.680	11.680
	0.10	0.921	0.947	0.922	0.944	1.061	1.081	1.885	1.885	8.969	8.977
	0.20	0.662	0.689	0.663	0.686	0.763	0.783	1.356	1.354	6.450	6.448

(1) When it is *a priori* clear that σ could be small in comparison to "rounding precision" and one obtains a rounded sample with all values equal to i_0, then there is really no way to estimate μ reliably beyond saying $\mu \in (i_0 - 0.5, i_0 + 0.5)$. (Of course, in such cases, the best option in terms of quality of estimation is to find another gage that is not so crude.)

If obtaining better gaging is not an option and it is *a priori* clear that it is possible that $\sigma < 0.5$, it is best to use the likelihood-based method (2.13), since the simulation results tell us that it gives actual coverage probability closer to $(1 - \alpha)$ than that of the t-method. (At the same time we must remember that the likelihood-based interval covers more often than we expect from its nominal level.)

(2) When one is *a priori* sure that $\sigma \geq 0.5$, both methods (2.10) and (2.13) can be used except for $n = 2$. The simulations show that the likelihood-based method is much better than the t-method when $\sigma = 0.5$ and 1. See Figures 2.6.

Appendix

In some situations it can be helpful to have approximations for the end points of the likelihood-based intervals. We provide such in this Appendix.

(A) Approximations for Case 2.

Continue to let n_{i_0} be the number of values i_0 observed, n_{i_0+1} be the number of values (i_0+1) observed, and take M to be supremum of the log-likelihood given in display (2.8).

To find approximations for the intervals prescribed by display (2.13) in Case 2, we plug $\hat{\sigma}_\mu = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{n}}$ into the approximation (2.7) modified by an empirically derived "correction factor" k to produce the approximation

$$L^*(\mu) \doteq k * \sum_{i=i_0}^{i_0+1} n_i * \ln(\frac{1}{\hat{\sigma}_\mu} * \phi(\frac{i - \mu}{\hat{\sigma}_\mu})) \ .$$

for

$$k = \begin{cases} 1 \ , \text{ if } n_{i_0+1} < n_{i_0} \text{ when computing a lower bound for } \mu \\ 0.975, \text{ if } n_{i_0+1} < n_{i_0} \text{ when computing an upper bound for } \mu \\ 0.975, \text{ if } n_{i_0+1} > n_{i_0} \text{ when computing a lower bound for } \mu \\ \frac{1}{0.975}, \text{ if } n_{i_0+1} > n_{i_0} \text{ when computing an upper bound for } \mu. \end{cases}$$

Substituting this approximation into display (2.13) and solving the quadratic equation in μ that results when there is equality, we get two solutions for μ. For convenience in what follows, let $w = (2 * M - c(n, \alpha))/n$ and $\hat{\sigma}^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}$.

Case 2a : When $n_{i_0+1} < n_{i_0}$ is observed.

If $(n * \ln(2) + M) > \frac{1}{2} * c(n, \alpha)$, then the interval for μ prescribed by display (2.13) is approximately

$$(\ \bar{x} \ - \ \sqrt{(e^{-1-w}/(2 * \pi)) - \hat{\sigma}^2}, \ i_0 \ + \ 0.5 \ ). \tag{2.14}$$

Otherwise, the interval for μ is approximately

$$(\ \bar{x} \ - \ \sqrt{(e^{-1-w}/(2 * \pi)) - \hat{\sigma}^2}, \ \bar{x} \ + \ \sqrt{(e^{-1-(w/0.975)}/(2 * \pi)) - \hat{\sigma}^2} \ ). \tag{2.15}$$

Case 2b : When $n_{i_0+1} > n_{i_0}$ is observed.

If $(n * \ln(2) + M) > \frac{1}{2} * c(n, \alpha)$, then the interval for μ prescribed by display (2.13) is approximately

$$(\ i_0 \ + \ 0.5 \ , \ \bar{x} \ + \ \sqrt{(e^{-1-(0.975*w)}/(2 * \pi)) - \hat{\sigma}^2} \ ). \tag{2.16}$$

Otherwise, the interval for μ is approximately

$$(\bar{x} - \sqrt{(e^{-1-(w/0.975)}/(2 * \pi)) - \hat{\sigma}^2}, \ \bar{x} + \sqrt{(e^{-1-(0.975*w)}/(2 * \pi)) - \hat{\sigma}^2} \ ). \tag{2.17}$$

Consider an example. Suppose a sample consists of data $\mathbf{x} = (0, 0, 1, 1, 1, 1, 1, 1, 1, 1)$ and take $\alpha = 0.10$. Then we have $i_0 = 0, \bar{x} = 0.8, n_{i_0} = 2, n_{i_0+1} = 8, c(10, 0.10) =$

3.17, and $M = -5.00402$. Because $n_{i_0+1} > n_{i_0}$, and $(10 * \ln(2) + M) > \frac{1}{2} * c(10, 0.10)$, we apply expression (2.16) to get the interval (0.5, 1.02718). If we use the same data but take $\alpha = 0.05$, then $(10 * \ln(2) + M) \leq \frac{1}{2} * c(10, 0.05)(= 2.25)$, so we apply expression (2.17) and get the interval (0.48492, 1.08444). The "exact" intervals for μ computed from expression (2.13) are (0.5, 1.02133) and (0.48003, 1.08676) for $\alpha = 0.10$ and $\alpha = 0.05$ respectively, which suggests that formulas (2.16) and (2.17) provide useful approximations.

(B) Approximations when the sample range is 2 or more.

To approximate the likelihood-based intervals for general case, we may apply the Mean Value Theorem and the approximate maximizers μ and σ mentioned before the asymptotic result (2.11). That will give us

$$\begin{aligned} & -2 * (L^*(\mu) - \sup_{\mu \in R} L^*(\mu)) \\ \doteq\ & -2 * \{ \sum_i n_i * \ln(\frac{1}{\hat{\sigma}_\mu} * \phi(\frac{i-\mu}{\hat{\sigma}_\mu})) - \sum_i n_i * \ln(\frac{1}{\hat{\sigma}} * \phi(\frac{i-\bar{x}}{\hat{\sigma}})) \} \\ =\ & -n * \ln(\frac{\hat{\sigma}^2}{\hat{\sigma}_\mu^2}), \end{aligned}$$

where $\hat{\sigma}_\mu^2 = \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{n}$ and $\hat{\sigma}^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}$.

After solving the quadratic equation

$$-n * \ln(\frac{\hat{\sigma}^2}{\hat{\sigma}_\mu^2}) = c(n, \alpha),$$

one has the approximation to interval (2.13)

$$(\bar{x} - \hat{\sigma} * \sqrt{e^{\frac{c(n,\alpha)}{n}} - 1}\ ,\ \bar{x} + \hat{\sigma} * \sqrt{e^{\frac{c(n,\alpha)}{n}} - 1}\ ) .$$

In fact, the reader may verify that substituting the earlier experssion for $c(n, \alpha)$ here, this interval is exactly the usual "t" interval obtained treating the data as exact.

References

[1] Bickel, P.J. and Doksum, K.A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics.* Holden-Day, San Francisco, CA.

[2] Vardeman, S.B. and Jensen, K.L. (1989). *$\bar{x}$ and R control charts for Rounded Data.* Preprint Number 89-33, Statistical Laboratory, Iowa State University, Ames, IA.

3 INTERVAL ESTIMATORS OF THE PARAMETER σ FOR ROUNDED NORMAL DATA

A paper submitted to the Communications in Statistics

Chiang-Sheng Lee and Stephen B. Vardeman
Iowa State University, Ames, IA 50011-1210

3.1 Introduction

Usually in the point estimation of parameters and the making of corresponding confidence intervals, data are assumed to be essentially "exact." But in practice, data are sometimes obtained using crude gaging. For example, a scale might only read to the nearest pound, while the ounces are still important. We will call such crudely gaged data "rounded."

The following questions arise when we analyze crudely gauged or rounded data. Do the traditional estimation methods still work well on these rounded data? What is an alternative method if they do not? In [1], we discussed interval estimation of the parameter μ when rounded data come from the Normal distribution with μ and σ unknown. We found that the answer to the first of these questions depends strongly on the (unknown) value of σ.

Mr. Lee is a Ph. D. Candidate in Industrial Engineering in the Industrial and Manufacturing Systems Engineering Department. email: chiang@iastate.edu

Dr. Vardeman is a Professor in the Statistics and Industrial and Manufacturing Systems Engineering Departments. He is a Senior Member of ASQ.

In this paper, we consider interval estimation of the parameter σ when rounded data come from a Normal distribution with both parameters μ and σ unknown. Two methods will be compared. One is the traditional (exact-data) $\chi^2_{(n-1)}$ method and the other is based on inversion of rounded data likelihood ratio tests concerning σ. Appropriate likelihood and profile loglikelihood functions will be introduced in the next section. Approximate maximizers of the likelihood, two special cases and the nature of the profile loglikelihood will be discussed in Sections 3.3 and 3.4. The initial constructions of the confidence intervals for the parameter σ are shown in Section 3.5. Initial simulations are described in Section 3.6. In Section 3.7 we improve the large σ properties of the second method by replacing the large n critical values for the likelihood ratio tests with more conservative critical values and study the performance of the modified intervals. In Section 3.8 we consider an additional improvement of the second method aimed at correcting remaining small σ deficiencies of the method. Final conclusions are drawn in the last section.

3.2 The Model for Rounded Normal Data

Without loss of generality, we assume observations are integers. We assume they are obtained by rounding the numerical values from a Normal sample. With this model assumption, the probability that n observations $X_1, X_2, ..., X_n$ take the integer values $x_1, x_2, ..., x_n$ is

$$\begin{aligned} f(\mathbf{x};\mu,\sigma) &= Pr(X_1 = x_1,\ X_2 = x_2, ...,\ X_n = x_n) \\ &= \prod_{i=1}^{n} [\Phi(\frac{x_i + 0.5 - \mu}{\sigma}) - \Phi(\frac{x_i - 0.5 - \mu}{\sigma})] \\ &= \prod_{i} [\Phi(\frac{i + 0.5 - \mu}{\sigma}) - \Phi(\frac{i - 0.5 - \mu}{\sigma})]^{n_i}, \end{aligned} \tag{3.1}$$

where $\Phi(x)$ is the standard Normal cumulative probability function, the product in (3.1) is over integer values i, and n_i is the number of observed values which take the value i.

It will be more convenient to work with the natural logarithm of the function in display (3.1), and we thus define the *log-likelihood function* by

$$L(\mu,\sigma) = \sum_i n_i \ln[\Phi(\frac{i+0.5-\mu}{\sigma}) - \Phi(\frac{i-0.5-\mu}{\sigma})].$$

Also, we can let

$$L^*(\sigma) = \sup_{\mu \in R} L(\mu,\sigma)$$

be the profile log-likelihood for σ and denote by M the supremum value of the function $L(\mu,\sigma)$,

$$M = \sup_{\mu \in R} \sup_{\sigma > 0} L(\mu,\sigma).$$

3.3 Approximate Maximizers of $L(\mu,\sigma)$

There are no closed forms for values μ and σ maximizing the function $L(\mu,\sigma)$. An argument presented in [1] says that under some circumstances, μ and σ maximizing $L(\mu,\sigma)$ are approximately $\bar{x}$ and $\sqrt{\sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{n}}$ respectively, the maximum likelihood estimates for a Normal model supposing the (integer) observations are exact, not produced by rounding.

3.4 Special Cases in the Maximization of $L(\mu,\sigma)$ and the Nature of $L^*(\sigma)$

In two special cases, many (μ,σ) pairs nearly maximizing $L(\mu,\sigma)$ will be indistinguishable in practice. We will call these Case 1 and Case 2. (The word "indistinguishable" here means that all of these pairs give $L(\mu,\sigma)$ near M.)

3.4.1 Case 1

Case 1 is the situation where all the observations in a sample have the same value, say i_0. In this case,

$$L(\mu,\sigma) = n\ \ln[\Phi(\frac{i_0+0.5-\mu}{\sigma}) - \Phi(\frac{i_0-0.5-\mu}{\sigma})]. \tag{3.2}$$

The supremum value in display (3.2) is 0 and is approached when the pair (μ,σ) is such that $\Phi(\frac{i_0+0.5-\mu}{\sigma}) \doteq 1$ and $\Phi(\frac{i_0-0.5-\mu}{\sigma}) \doteq 0$. Applying the facts that $\Phi(\geq 3) \doteq 1$ and $\Phi(\leq -3) \doteq 0$, $L(\mu,\sigma)$ in display (3.2) approximates its supremum value 0 if (μ,σ) is in the triangular region of the (μ,σ)-plane where $\mu \in (i_0 - 0.5,\ i_0 + 0.5)$ and $\sigma \in (0,\ \min(\frac{i_0+0.5-\mu}{3},\ \frac{\mu+0.5-i_0}{3})\ ]$.

3.4.2 Case 2

Case 2 is the situation where a sample contains only two different values with sample range 1, say the integers i_0 and i_0+1. As mentioned in [1], the supremum value of the log-likelihood for Case 2 is approached when $\Phi(\frac{i_0-0.5-\mu}{\sigma}) \doteq 0$, $\Phi(\frac{i_0+0.5-\mu}{\sigma}) = \frac{n_{i_0}}{n}$ and $\Phi(\frac{i_0+1.5-\mu}{\sigma}) \doteq 1$, and has the form

$$n_{i_0}\ \ln(\frac{n_{i_0}}{n}) + n_{i_0+1}\ \ln(\frac{n_{i_0+1}}{n}), \tag{3.3}$$

where n_{i_0} and n_{i_0+1} are the numbers of i_0 and i_0+1 observations in the sample.

Applying above three conditions together with the facts that $\Phi(\geq 3) \doteq 1$ and $\Phi(\leq -3) \doteq 0$, the following identifies (effectively indistinguishable) (μ,σ) pairs nearly maximizing $L(\mu,\sigma)$.

Result : In Case 2, the function $L(\mu,\sigma)$ approximates its supremum value (3.3) when $\sigma \in (0,\ \frac{1}{3+|\Phi^{-1}(\frac{n_{i_0}}{n})|})$, and $\mu = i_0 + 0.5 - \Phi^{-1}(\frac{n_{i_0}}{n})\ \sigma$.

3.4.3 Typical Plots of $L^*(\sigma)$

In Figure 3.1, we present representative graphs of $L^*(\sigma)$ for Case 1, Case 2, and the case with sample range ≥ 2 (using the samples (1,1,1,1,1), (0,0,1,1,1), and $(-1,-1.0,1.1)$ respectively). On this plot, $A_1 \doteq \frac{1}{6}$ and $A_2 \doteq \frac{1}{3+|\Phi^{-1}(\frac{n_{10}}{n})|}$ and the intervals $(0, A_1)$ and $(0, A_2)$ are sets of σ's for which $L^*(\sigma)$ is nearly the supremum value. $A_3 \doteq \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}}$ is nearly the maximizer of $L^*(\sigma)$ in this instance of $n = 5$ observations with range 2.

3.5 Confidence Intervals for the Parameter σ

There are two methods that will be used here to set confidence intervals on the parameter σ. We will call these the traditional method and the likelihood-based method.

The traditional method is based on the fact that without rounding.

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{(n-1)}, \tag{3.4}$$

where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$. Applying the property (3.4), one can get the usual ("exact data") intervals for σ^2 to be

$$\left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\chi^2_{(n-1,1-\frac{\alpha}{2})}}, \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\chi^2_{(n-1,\frac{\alpha}{2})}} \right],$$

and (taking square roots) intervals for σ

$$\left[\sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\chi^2_{(n-1,1-\frac{\alpha}{2})}}}, \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\chi^2_{(n-1,\frac{\alpha}{2})}}} \right]. \tag{3.5}$$

($\chi^2_{(n-1,p)}$ is the p quantile of the χ^2_{n-1} distribution.) We will consider naively plugging integer-rounded data into these exact data formulas.

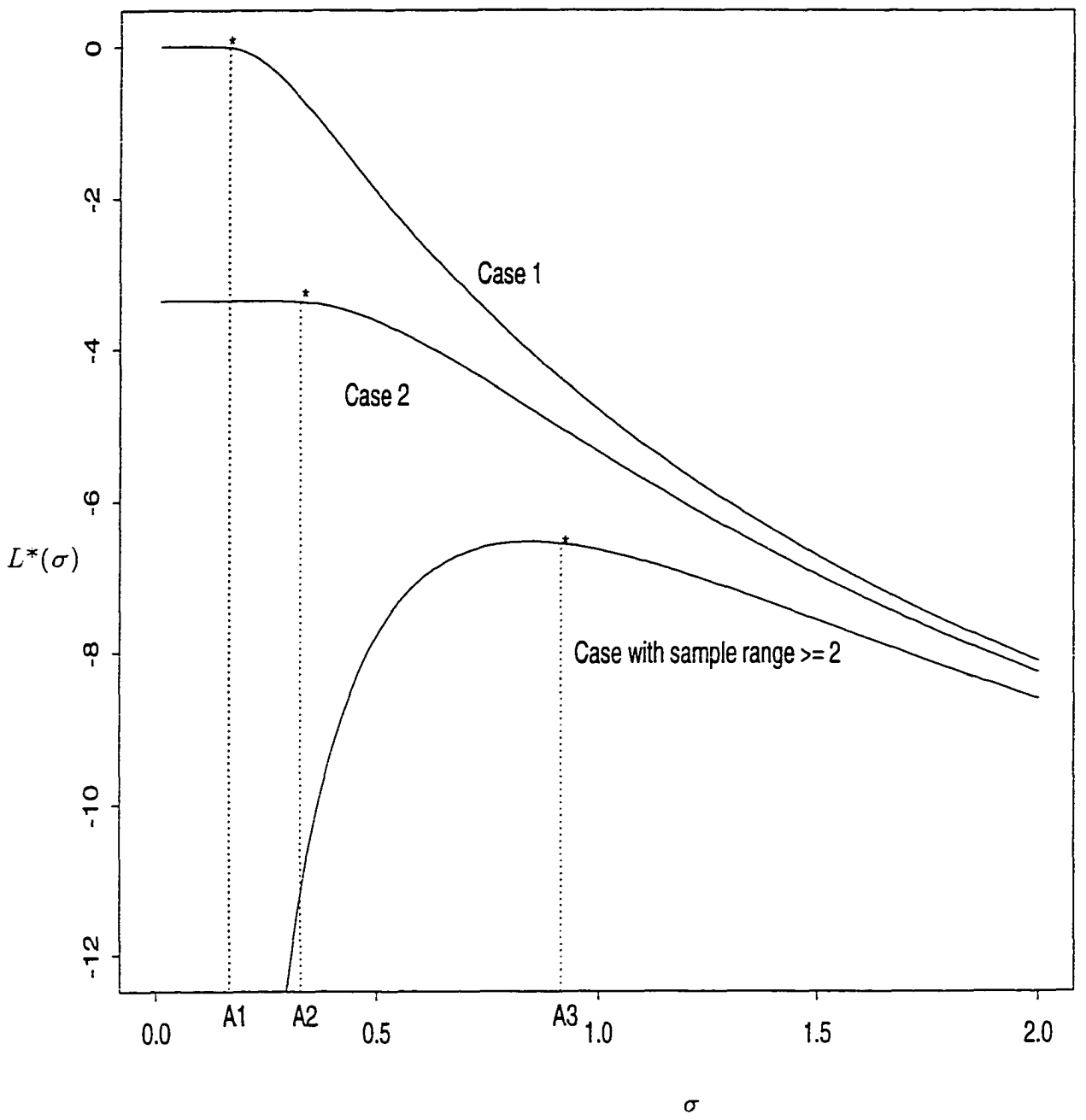

Figure 3.1 Representative graphs of $L^*(\sigma)$ for Case 1, Case 2, and a case with sample range ≥ 2. (These particular graphs are drawn under the samples (1,1,1,1,1), (0,0,1,1,1), and $(-1,-1,0,1,1)$.)

The likelihood-based method is based initially on the asymptotic result that under $H_0 : \sigma = \sigma_0$ and for large n

$$-2 \ln\left(\frac{\sup_{\mu \in R} f(\mathbf{x}; \mu, \sigma_0)}{\sup_{\mu \in R} \sup_{\sigma > 0} f(\mathbf{x}; \mu, \sigma)}\right) \dot{\sim} \chi^2_{(1)} \tag{3.6}$$

Or, using the notations in this paper, for large n

$$-2 \ (L^*(\sigma_0) - M) \dot{\sim} \chi^2_{(1)}.$$

Using approximation (3.6), a likelihood-based confidence interval for σ consists of all those σ for which

$$-2 \ (L^*(\sigma) - M) \leq \chi^2_{(1,1-\alpha)}. \tag{3.7}$$

We now discuss how to find the end-points for the interval defined by (3.7).

3.5.1 The Likelihood-Based Interval in Case 1

As mentioned before, the supremum value of the log-likelihood function $L(\mu, \sigma)$ for Case 1 is 0. The maximum value of $\Phi(\frac{i + 0.5 - \mu}{\sigma}) - \Phi(\frac{i - 0.5 - \mu}{\sigma})$ for fixed σ occurs at $\mu = i$, which gives

$$L^*(\sigma) = n \ \ln(2 \ \Phi(\frac{1}{2 \ \sigma}) - 1).$$

Putting the above expression for the profile log-likelihood into inequality (3.7) we can see that the interval for Case 1 is

$$(0, \ \frac{1}{2 \ \Phi^{-1}(\frac{1}{2} \ (1 + e^{\frac{-\chi^2_{(1,1-\alpha)}}{2n}}))} \ ].$$

3.5.2 The Likelihood-Based Interval in Other Cases

Unlike the situation in Case 1, numerical analysis is needed to find end-points for the interval (3.7) in other circumstances. In Section 3.8, we provide adjusted endpoints for likelihood-based interval for Case 2. (One thing to notice is that the lower end-point of any likelihood-based interval in Case 2 is always 0, just as in Case 1.) In our work in this paper with likelihood-based intervals, we use numerical analysis to find end-points for interval (3.7), but the Appendix also provides approximations to the interval (3.7) for the situation where the sample range is 2 or more.

3.6 Simulations

In this section, Monte Carlo simulations are used to compare the two interval estimation methods introduced in previous sections. First, we randomly select a sample of size n from a Normal distribution with mean μ and standard deviation σ, and then round the observations to integers. Second, we apply formulas (3.5) and (3.7) to make intervals for the parameter σ. The last step is to examine whether the intervals contain the value σ or not. If the answer is "yes," then we increment a counter (t or l respectively for methods (3.5) and (3.7)) by 1. If the answer is "no," the counter is not incremented. This is repeated 1,000 times, and so we obtain counts t_{1000} and l_{1000}. The ratios $\frac{t_{1000}}{1000}$ and $\frac{l_{1000}}{1000}$ are then Monte Carlo estimates of the actual coverage probabilities for the nominally (approximately) $(1-\alpha)$ level confidence procedures.

For example, suppose an initial random sample of size $n = 5$ taken from the Normal distribution with $\mu = 0.1$ and $\sigma = 0.5$ produces rounded data with range 1 and $n_0 = 2$ and $n_1 = 3$. After computing, we can get (0.328, 1.574) from formula (3.5) and (0, 0.982) from formula (3.7) when $\alpha = 0.05$. It is then obvious that both intervals contain the true parameter $\sigma = 0.5$, so we set the counters $t_1 = 1$ and $l_1 = 1$. Then suppose that a second rounded Normal sample contains $n = 5$ values, all $i_0 = 1.0$. For this second

sample, the traditional interval is degenerate at 0, and the interval defined by (3.7) is (0, 0.502). In this case we will only increment l by 1 and have $t_2 = 1$ and $l_2 = 2$. And so on through 1,000 samples.

Different combinations of $\mu, \sigma, n,$ and α were used in the simulations to provide a thorough comparison of the two methods. We considered $\mu \in \{[0,\ 0.5](0.1)\}$, $\sigma \in \{[0.5,\ 1.0](0.1), [2,\ 10](1)\}$, $n \in \{2,\ 5,\ 10,\ 15\}$, and $\alpha \in \{0.05,\ 0.10,\ 0.20\}$, where $[a, b](c)$ means the values from a to b in increments of c. Figures 3.2 to 3.5 provide graphs of the estimated coverage probabilities for the traditional intervals (3.5) and likelihood-based intervals (3.7) for different sample sizes n. In those graphs, the solid lines indicate the estimated coverage probabilities for the traditional method. The dashed lines indicate the estimated coverage probabilities for the likelihood-based method.

After analyzing these figures, we reach the following conclusions.

(1) At sample size $n = 2$, the traditional method is better than the likelihood-based method only when $\sigma \geq 2$. For small σ, Case 1 and Case 2 samples are quite likely and the traditional method fails to produce intervals covering "enough small σ's " for such samples.

(2) At samples sizes $n \geq 5$, the coverage proability for the traditional method is much closer to the nominal value $(1 - \alpha)$ than that for likelihood-based method for $\sigma \geq 1.0$. (1.0 is a compromise value that suits Figures 3.3 through 3.5.)

(3) Generally speaking, the likelihood-based method is better than the traditional method (in terms of converage probability) when σ is small. (For example, we find that when $\sigma \leq 0.5$ all the coverage probabilities for the likelihood-based method are much better than those of the traditional method.)

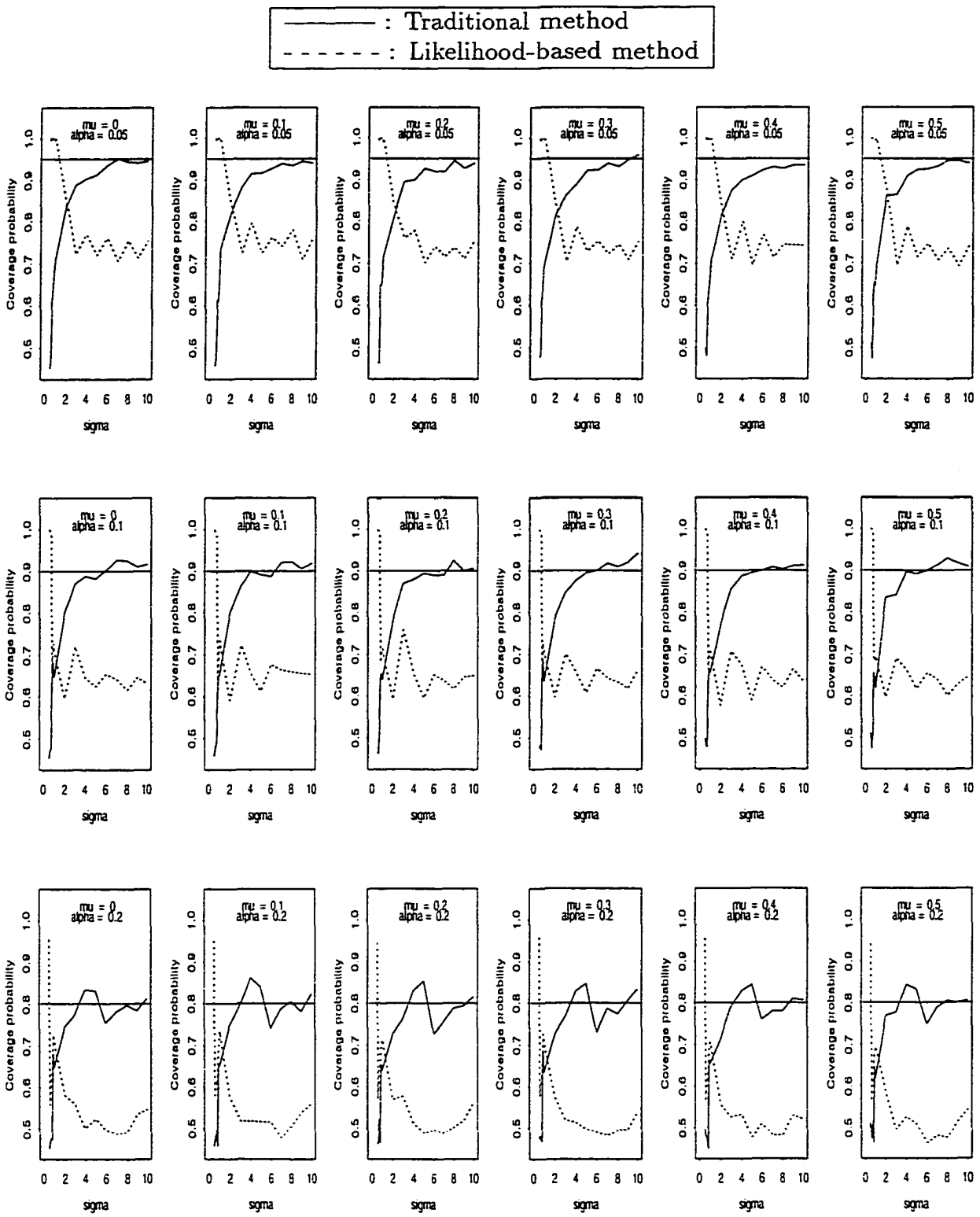

Figure 3.2 The estimated coverage probability for the traditional method and the likelihood-based method at sample size $n = 2$.

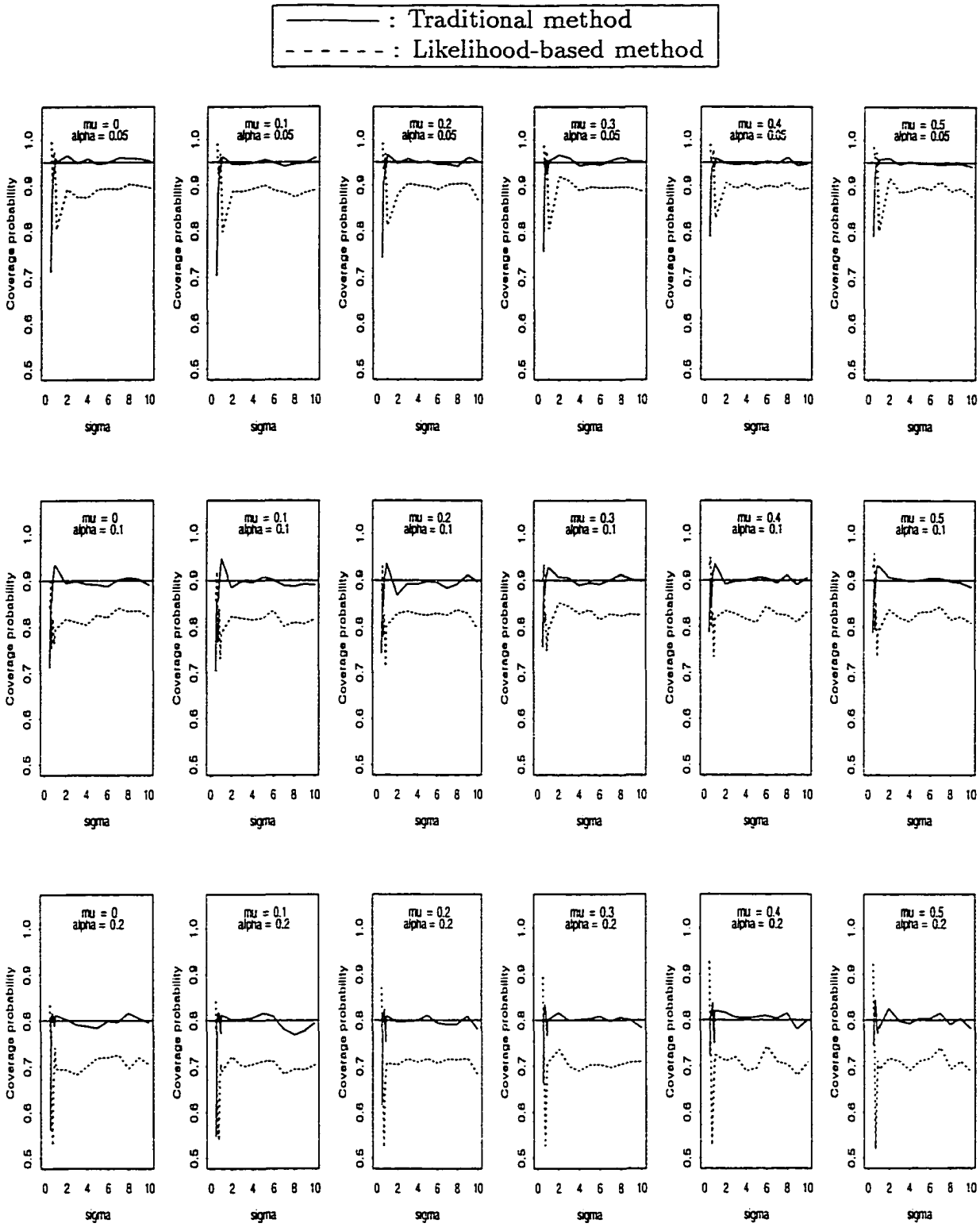

Figure 3.3 The estimated coverage probability for the traditional method and the likelihood-based method at sample size $n = 5$.

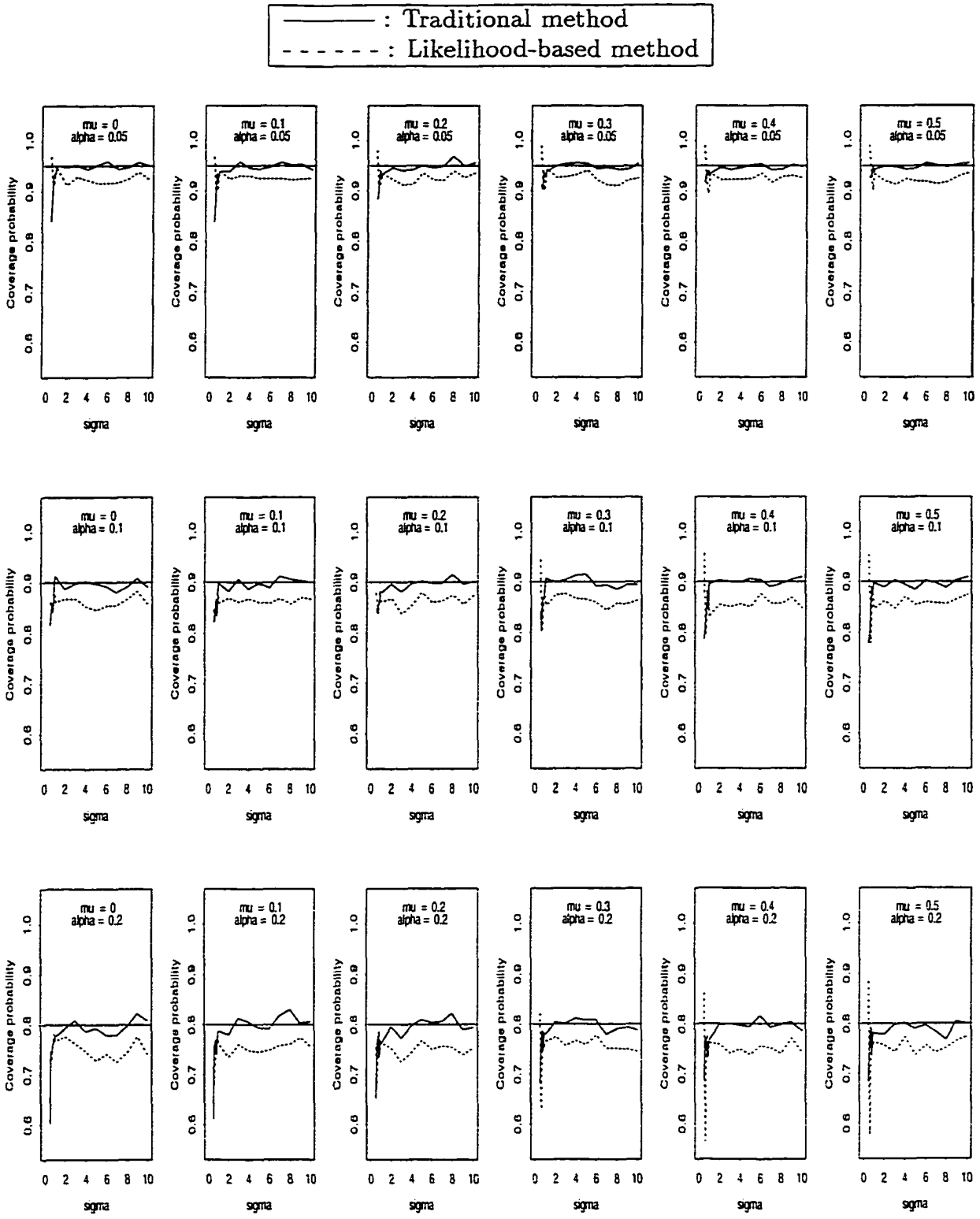

Figure 3.4 The estimated coverage probability for the traditional method and the likelihood-based method at sample size $n = 10$.

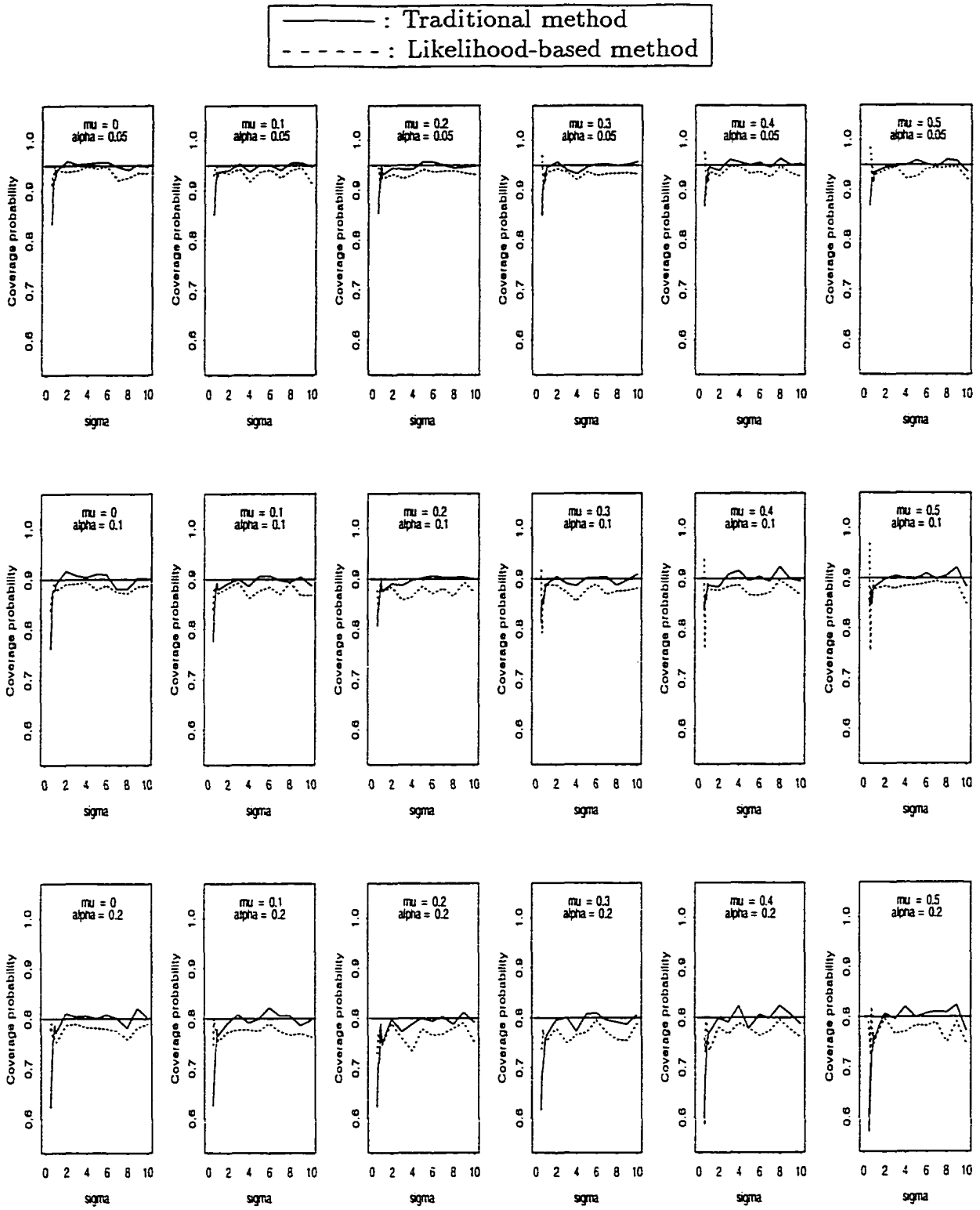

Figure 3.5 The estimated coverage probability for the traditional method and the likelihood-based method at sample size $n = 15$.

3.7 Improving the Coverage Probability Calibration of the Likelihood-Based Intervals

Form the simulations in Section 3.6, we see that although the likelihood-based method is conservative for small σ and guaranteed by standard theory to have coverage probability $(1-\alpha)$ for large n, it can be liberal for small n and (particularly) large σ. A possible means of improving the small n performance of the likelihood-based method is to replace the $\chi^2_{(1,1-\alpha)}$ value in (3.7) with a larger value, $d(n,\alpha)$, chosen to make the coverage probability of the likelihood-based method typically closer to the nominal value $(1-\alpha)$.

It is intuitively plausible that for large σ, the likelihood ratio test for "rounded" data is essentially the same as the one for "exact" data. One can find that for "exact" data (and L^*_{exact} and M_{exact} the exact data analogs of our L^* and M)

$$-2(L^*_{\text{exact}}(\sigma) - M_{\text{exact}}) = Y - n\ln(Y) - n + n\ln(n), \tag{3.8}$$

where $Y = \dfrac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2}$ follows the $\chi^2_{(n-1)}$ distribution. We might therefore consider replacing $\chi^2_{(1,1-\alpha)}$ in display (3.7) with $d(n,\alpha)$ that is the $(1-\alpha)$ quantile of the variable right side of (3.8) when Y is $\chi^2_{(n-1)}$.

To find $d(n,\alpha)$, we need to find $Z(n,\alpha)$ such that for $Y \sim \chi^2_{(n-1)}$

$$1-\alpha = Pr(Y - n\ln(Y) \le Z(n,\alpha))$$

since we may then set $d(n,\alpha) = Z(n,\alpha) - n + n\ln(n)$. To do this, we need to find values y_1 and y_2 such that $y_1 < y_2$, $y_1 - n\ln(y_1) = y_2 - n\ln(y_2)$, and $Pr(y_1 \le Y \le y_2) = 1-\alpha$. $Z(n,\alpha)$ is then $y_1 - n\ln(y_1)$. In Table 3.1 we give such $d(n,\alpha)$ values obtained by numerical methods for different combinations of n and α.

Another set of simulations was conducted, using $d(n,\alpha)$ in place of $\chi^2_{(1,1-\alpha)}$ in (3.7). Figures 3.6 through 3.9 summarize the estimated coverage probabilities for the same

Table 3.1 $d(n, \alpha)$ Values.

n	α		
	0.05	0.10	0.20
2	10.47	7.71	4.97
3	7.26	5.23	3.27
4	6.15	4.39	2.71
5	5.58	3.97	2.43
6	5.24	3.71	2.27
7	5.01	3.54	2.16
8	4.84	3.42	2.09
9	4.72	3.33	2.03
10	4.62	3.26	1.99
15	4.34	3.06	1.86
20	4.21	2.97	1.80
30	4.08	2.88	1.75
∞	3.84	2.71	1.64

combinations of n, μ, and σ used in the previous section. On these graphs, the solid lines indicate estimated coverage probabilities for the traditional method and the dashed lines identify estimated coverage probabilities for the corrected likelihood-based method (with $d(n, \alpha)$ values used in place of $\chi^2_{(1,1-\alpha)}$ in (3.7)). The $d(n, \alpha)$ values make the estimated coverage probabilities closer to the nominal value $(1 - \alpha)$ except in some small σ cases.

3.8 Improvements of Likelihood-Based Method for Small σ Value and Final Comparisons to the Traditional Method

From the figures in previous section, we can see that $d(n, \alpha)$ values improve the estimated coverage probabilities for most σ values, but there are still some small σ values (e.g. $\sigma \leq 1$) where estimated coverage probabilities are much below nominal. For example, in Figure 3.7 with $\mu = 0.5$, the value $\sigma = 1.0$ at $\alpha = 0.10$ and the value $\sigma = 0.8$ at $\alpha = 0.20$ have estimated coverage probabilities well below the value $(1 - \alpha)$. In

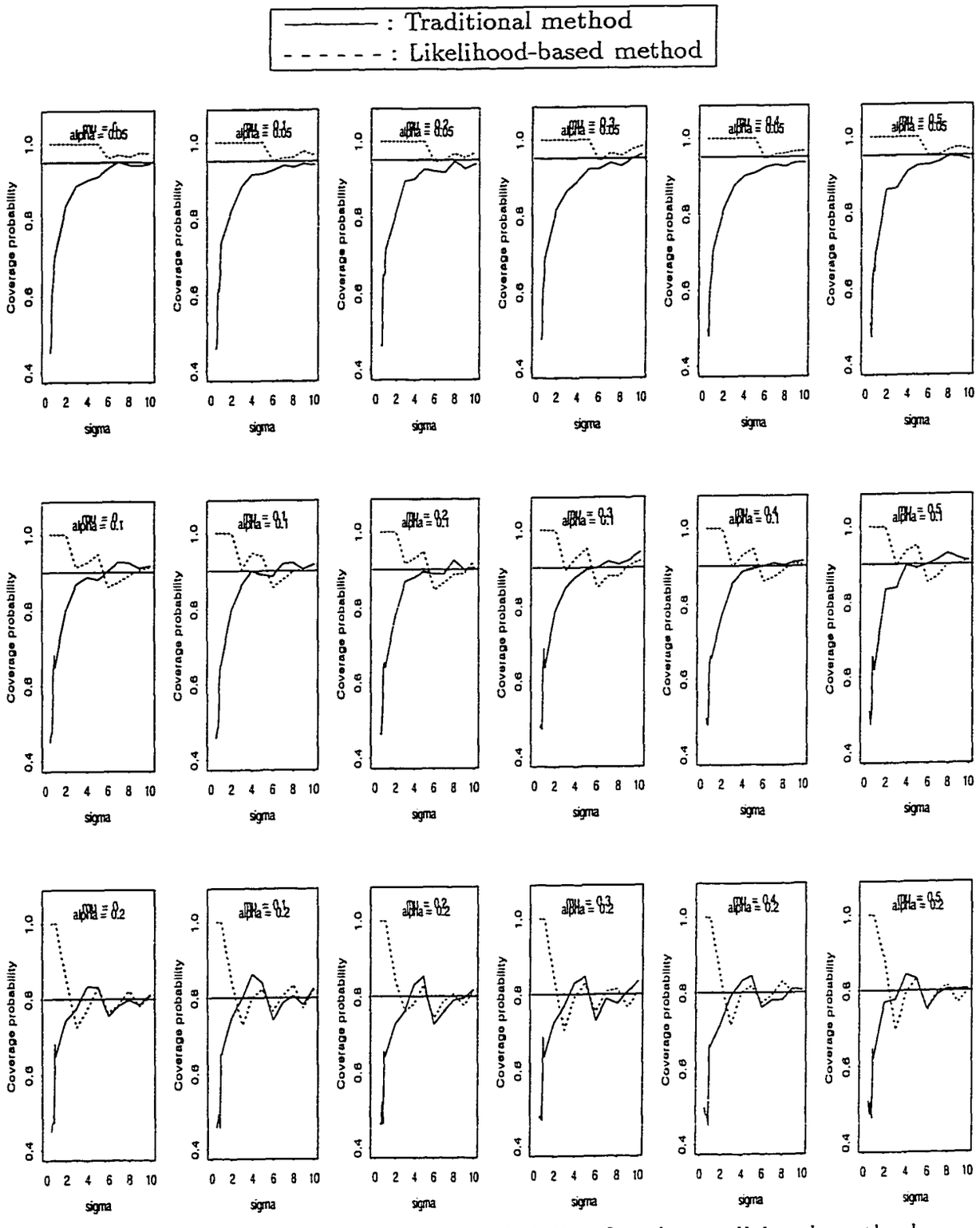

Figure 3.6 The estimated coverage probability for the traditional method and the corrected likelihood-based method (using $d(n, \alpha)$) at sample size $n = 2$.

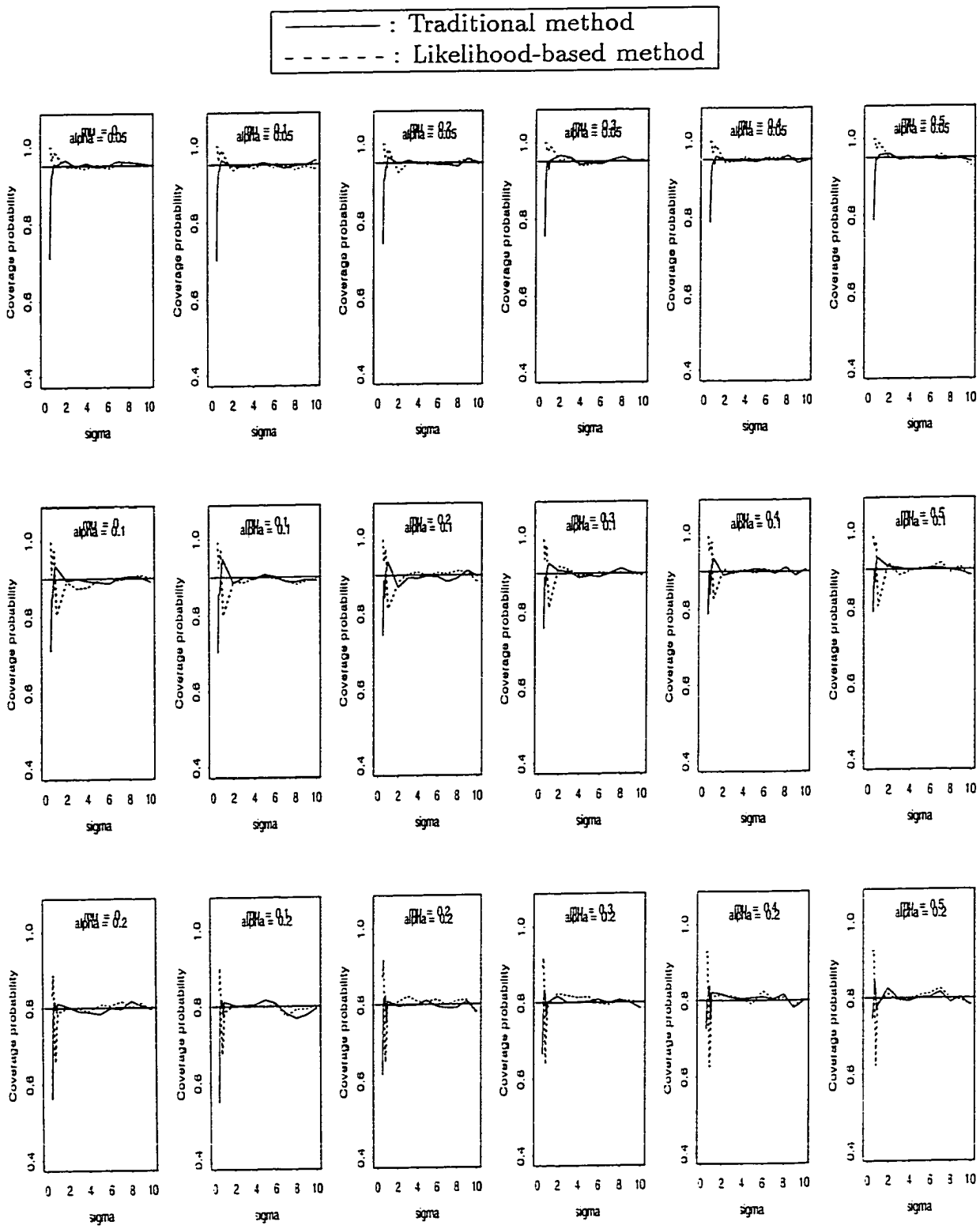

Figure 3.7 The estimated coverage probability for the traditional method and the corrected likelihood-based method (using $d(n, \alpha)$) at sample size $n = 5$.

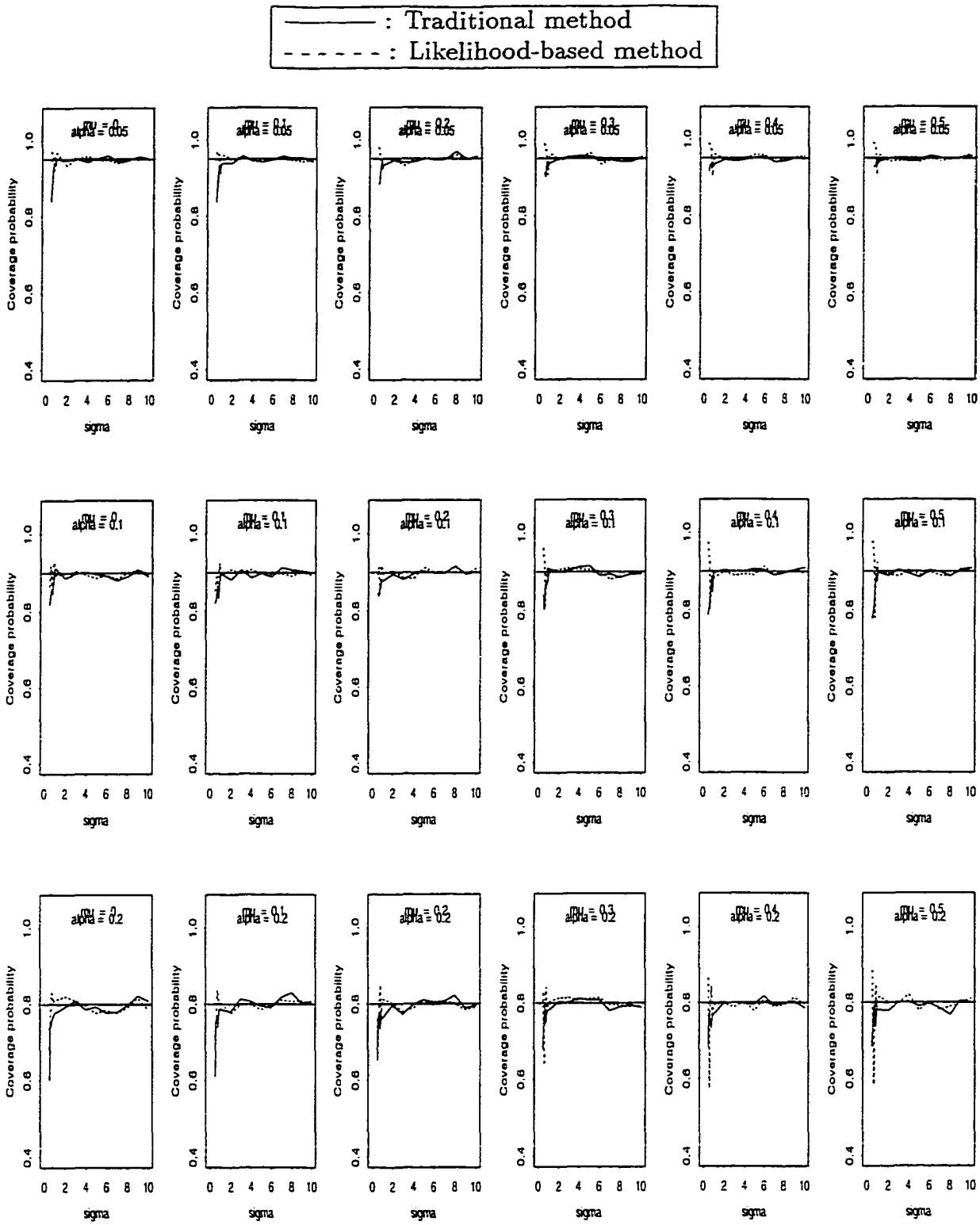

Figure 3.8 The estimated coverage probability for the traditional method and the corrected likelihood-based method (using $d(n, \alpha)$) at sample size $n = 10$.

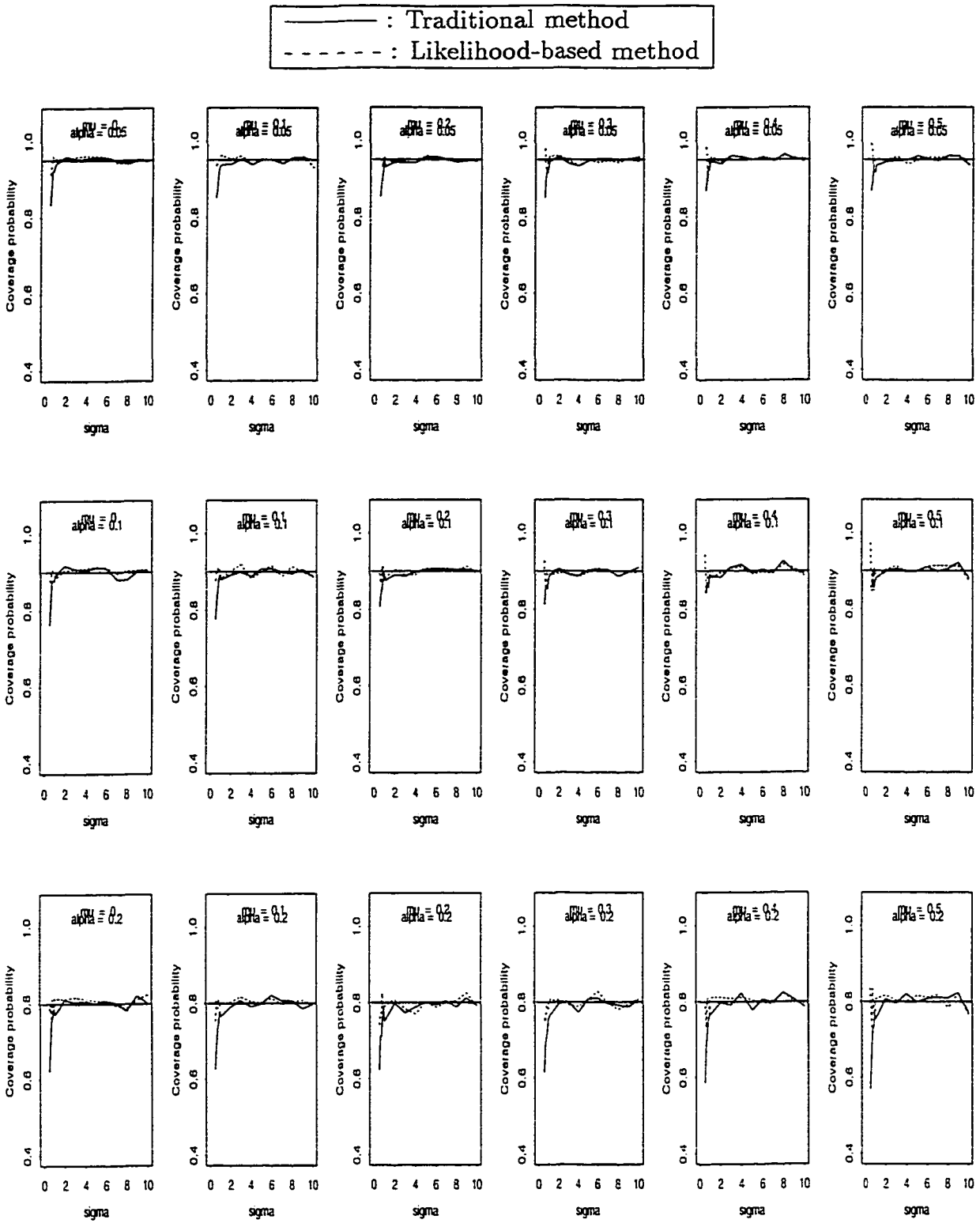

Figure 3.9 The estimated coverage probability for the traditional method and the corrected likelihood-based method (using $d(n, \alpha)$) at sample size $n = 15$.

this section, we discuss how to improve these small σ coverage probabilities.

When σ and n are small, there is a large probability of generating a Case 1 or Case 2 sample. This suggests that if we somehow increase the size of Case 1 and Case 2 intervals, we may well be able to improve the coverage probability and cure the small σ problem seen in the simulation results.

Now it is clear that

$$P_{\mu,\sigma}(\text{the interval fails to cover } \sigma) \geq P^1_{\mu,\sigma} + P^2_{\mu,\sigma}$$

for $P^1_{\mu,\sigma} = P_{\mu,\sigma}(\text{a Case 1 sample is generated and the interval fails to cover } \sigma)$,
and $P^2_{\mu,\sigma} = P_{\mu,\sigma}(\text{a Case 2 sample is generated and the interval fails to cover } \sigma)$.
So to produce $1-\alpha$ coverage probability, it is necessary to have

$$P^1_{\mu,\sigma} + P^2_{\mu,\sigma} \leq \alpha \quad \forall \mu, \sigma. \tag{3.9}$$

Numerically, we also find for the likelihood-based intervals of Section 3.7 that $\sigma_1 < \sigma_{2,1} < \sigma_{2,2} < \cdots < \sigma_{2,\,[\frac{n}{2}]}$, where σ_1 is the upper limit of the interval for a Case 1 sample, and $\sigma_{2,j}$ is the upper limit for a Case 2 sample with $n_{i_0} = j$. The limits $\sigma_1 < \sigma_{2,1} < \sigma_{2,2} < \cdots < \sigma_{2,\,[\frac{n}{2}]}$ derived in Section 3.7 are not large enough to guarantee that inequality (3.9) holds. So, as a step toward correcting the inadequate small σ coverage probabilities of the likelihood method of Section 3.7, we propose to replace $\sigma_1, \sigma_{2,1}, \sigma_{2,2}, \cdots, \sigma_{2,\,[\frac{n}{2}]}$ with the smallest set of numbers satisfying both inequality (3.9) and the order restriction observed by the Case 1 and Case 2 limits of Section 3.7.

Development of replacement Case 1 and Case 2 limits is a "brute force computation" problem. To replace σ_1 we seek the minimum standard deviation σ_1^* so that

$$\max_\mu \; P_{\mu,\sigma_1^*}[\text{ Case 1 sample }] \;\leq\; \alpha\,. \tag{3.10}$$

To replace $\sigma_{2,j}$ we seek the minimum standard deviation $\sigma^*_{2,j}$ so that

$$\max_\mu(\; P_{\mu,\sigma^*_{2,j}}[\text{Case 1 sample}]$$

$$+ \sum_{l=1}^{j} P_{\mu,\sigma^*_{2,j}}[\text{Case 2 sample with } n_{i_0} = l \text{ or } n_{i_0} = n - l\ ]\ )$$
$$\leq \alpha \tag{3.11}$$

Then clearly, $\sigma^*_1 < \sigma^*_{2,1} < \sigma^*_{2,2} < \cdots < \sigma^*_{2,\,[\frac{n}{2}]}$ and we propose to use such limits as replacements for the Case 1 and Case 2 limits of Section 3.7.

Note that for $\sigma \leq \sigma^*_1$ all Case 1 and Case 2 samples have corresponding modified intervals that include σ. So inequality (3.9) holds. For $\sigma^*_1 < \sigma \leq \sigma^*_{2,1}$ all Case 2 samples produce modified intervals including σ, but Case 1 samples do not. But (3.10) guarantees that inequality (3.9) holds. For $\sigma^*_{2,j} < \sigma \leq \sigma^*_{2,j+1}$, Case 1 samples and Case 2 samples with $1 \leq n_{i_0} \leq j$ or $n - j \leq n_{i_0} \leq n - 1$ produce modified intervals that fail to include σ while Case 2 samples with $j < n_{i_0} < n - j$ produced modified intervals that do cover σ. Then inequality (3.11) guarantees that inequality (3.9) holds.

In Tables 3.2 and 3.3, we provide the modified upper limits $\sigma^*_1, \sigma^*_{2,1}, \sigma^*_{2,2}, \cdots, \sigma^*_{2,\,[\frac{n}{2}]}$ for Case 1 and Case 2 samples for different n, n_{i_0}, and α from above method.

In Table 3.3, the value in the parentheses is n_{i_0}, the number of i_0 observations occurring in the sample. If $n_{i_0} > [\frac{n}{2}]$, then the upper limit can be found by looking for $n - n_{i_0}$ in the parentheses. (This is because the data $(n_{i_0},\ n_{i_0+1})$ and $(n - n_{i_0},\ n - n_{i_0+1})$ will produce the same interval.) For example, if a $n = 10$ sample is a Case 2 sample with $i_0 = 1$, $n_1 = 8$, and $\alpha = 0.05$, then the upper limit 0.677 from the table, by finding the value $(n - n_{i_0}) = 2$ in parentheses.

Figures 3.10 to 3.13 compare the estimated confidence levels for the modified method of this section to those of the traditional method. The figures show that the estimated coverage probabilities for small σ are improved over those pictured in Figures 3.6 to 3.9 and that the corrected method of this section can (unlike the traditional method) provide reliable inferences for σ based on rounded data.

We also ran a simulation to compare the average interval lengths for the traditional intervals and the revised likelihood-based intervals of this section for the combinations of

Table 3.2 The modified upper limit for Case 1 samples.

n	α		
	0.05	0.10	0.20
2	5.635	2.807	1.381
3	1.325	0.916	0.638
4	0.822	0.653	0.516
5	0.666	0.558	0.459
6	0.586	0.502	0.422
7	0.533	0.464	0.395
8	0.495	0.435	0.375
9	0.466	0.413	0.360
10	0.443	0.396	0.347
11	0.425	0.381	0.336
12	0.409	0.369	0.327
13	0.396	0.358	0.319
14	0.384	0.349	0.312
15	0.374	0.341	0.306
16	0.366	0.334	0.301
17	0.358	0.328	0.296
18	0.351	0.322	0.291
19	0.344	0.317	0.287
20	0.339	0.312	0.284

Table 3.3 Modified upper limits for Case 2 samples. (The values in the parentheses are values of n_{i_0}.)

n	α		
	0.05	0.10	0.20
2	16.914 (1)	8.439 (1)	4.182 (1)
3	3.535 (1)	2.462 (1)	1.684 (1)
4	1.699 (1) 2.034 (2)	1.303 (1) 1.571 (2)	0.972 (1) 1.189 (2)
5	1.143 (1) 1.516 (2)	0.921 (1) 1.231 (2)	0.728 (1) 0.980 (2)
6	0.897 (1) 1.153 (2) 1.258 (3)	0.752 (1) 0.960 (2) 1.054 (3)	0.620 (1) 0.780 (2) 0.870 (3)
7	0.768 (1) 0.944 (2) 1.106 (3)	0.660 (1) 0.800 (2) 0.949 (3)	0.557 (1) 0.663 (2) 0.802 (3)
8	0.687 (1) 0.819 (2) 0.952 (3) 1.009 (4)	0.599 (1) 0.707 (2) 0.825 (3) 0.880 (4)	0.513 (1) 0.597 (2) 0.698 (3) 0.755 (4)
9	0.629 (1) 0.736 (2) 0.837 (3) 0.941 (4)	0.555 (1) 0.644 (2) 0.726 (3) 0.831 (4)	0.480 (1) 0.551 (2) 0.610 (3) 0.721 (4)
10	0.585 (1) 0.677 (2) 0.747 (3) 0.851 (4) 0.890 (5)	0.520 (1) 0.597 (2) 0.654 (3) 0.753 (4) 0.793 (5)	0.454 (1) 0.516 (2) 0.560 (3) 0.652 (4) 0.694 (5)
11	0.550 (1) 0.630 (2) 0.690 (3) 0.775 (4) 0.851 (5)	0.493 (1) 0.560 (2) 0.609 (3) 0.685 (4) 0.763 (5)	0.434 (1) 0.489 (2) 0.526 (3) 0.587 (4) 0.672 (5)
12	0.522 (1) 0.593 (2) 0.646 (3) 0.708 (4) 0.789 (5) 0.818 (6)	0.470 (1) 0.531 (2) 0.573 (3) 0.626 (4) 0.707 (5) 0.738 (6)	0.417 (1) 0.466 (2) 0.499 (3) 0.542 (4) 0.621 (5) 0.653 (6)
13	0.499 (1) 0.563 (2) 0.610 (3) 0.658 (4) 0.733 (5) 0.791 (6)	0.452 (1) 0.506 (2) 0.544 (3) 0.587 (4) 0.655 (5) 0.716 (6)	0.402 (1) 0.447 (2) 0.477 (3) 0.513 (4) 0.569 (5) 0.638 (6)
14	0.479 (1) 0.537 (2) 0.580 (3) 0.622 (4) 0.681 (5) 0.745 (6) 0.768 (7)	0.436 (1) 0.485 (2) 0.520 (3) 0.558 (4) 0.607 (5) 0.674 (6) 0.698 (7)	0.390 (1) 0.431 (2) 0.458 (3) 0.491 (4) 0.531 (5) 0.597 (6) 0.624 (7)

Table 3.3 (continued)

n	α		
	0.05	0.10	0.20
15	0.463 (1) 0.515 (2) 0.555 (3) 0.593 (4) 0.639 (5) 0.701 (6) 0.748 (7)	0.422 (1) 0.468 (2) 0.499 (3) 0.534 (4) 0.574 (5) 0.632 (6) 0.682 (7)	0.379 (1) 0.417 (2) 0.442 (3) 0.472 (4) 0.506 (5) 0.555 (6) 0.612 (7)
16	0.448 (1) 0.497 (2) 0.533 (3) 0.568 (4) 0.608 (5) 0.659 (6) 0.711 (7) 0.731 (8)	0.410 (1) 0.452 (2) 0.482 (3) 0.514 (4) 0.550 (5) 0.592 (6) 0.647 (7) 0.668 (8)	0.370 (1) 0.406 (2) 0.428 (3) 0.456 (4) 0.487 (5) 0.522 (6) 0.578 (7) 0.601 (8)
17	0.435 (1) 0.480 (2) 0.514 (3) 0.546 (4) 0.583 (5) 0.623 (6) 0.675 (7) 0.715 (8)	0.400 (1) 0.439 (2) 0.466 (3) 0.496 (4) 0.529 (5) 0.564 (6) 0.613 (7) 0.655 (8)	0.362 (1) 0.395 (2) 0.416 (3) 0.442 (4) 0.470 (5) 0.501 (6) 0.543 (7) 0.591 (8)
18	0.424 (1) 0.466 (2) 0.497 (3) 0.528 (4) 0.562 (5) 0.597 (6) 0.640 (7) 0.684 (8) 0.701 (9)	0.390 (1) 0.427 (2) 0.453 (3) 0.480 (4) 0.511 (5) 0.543 (6) 0.579 (7) 0.626 (8) 0.644 (9)	0.354 (1) 0.386 (2) 0.406 (3) 0.429 (4) 0.456 (5) 0.484 (6) 0.514 (7) 0.562 (8) 0.582 (9)
19	0.414 (1) 0.453 (2) 0.482 (3) 0.511 (4) 0.543 (5) 0.576 (6) 0.610 (7) 0.654 (8) 0.688 (9)	0.382 (1) 0.417 (2) 0.440 (3) 0.466 (4) 0.495 (5) 0.525 (6) 0.555 (7) 0.597 (8) 0.633 (9)	0.347 (1) 0.377 (2) 0.396 (3) 0.418 (4) 0.443 (5) 0.469 (6) 0.495 (7) 0.532 (8) 0.574 (9)
20	0.405 (1) 0.442 (2) 0.469 (3) 0.496 (4) 0.526 (5) 0.557 (6) 0.587 (7) 0.625 (8) 0.662 (9) 0.677 (10)	0.374 (1) 0.407 (2) 0.430 (3) 0.454 (4) 0.481 (5) 0.509 (6) 0.536 (7) 0.568 (8) 0.609 (9) 0.624 (10)	0.341 (1) 0.370 (2) 0.388 (3) 0.408 (4) 0.432 (5) 0.456 (6) 0.480 (7) 0.507 (8) 0.549 (9) 0.567 (10)

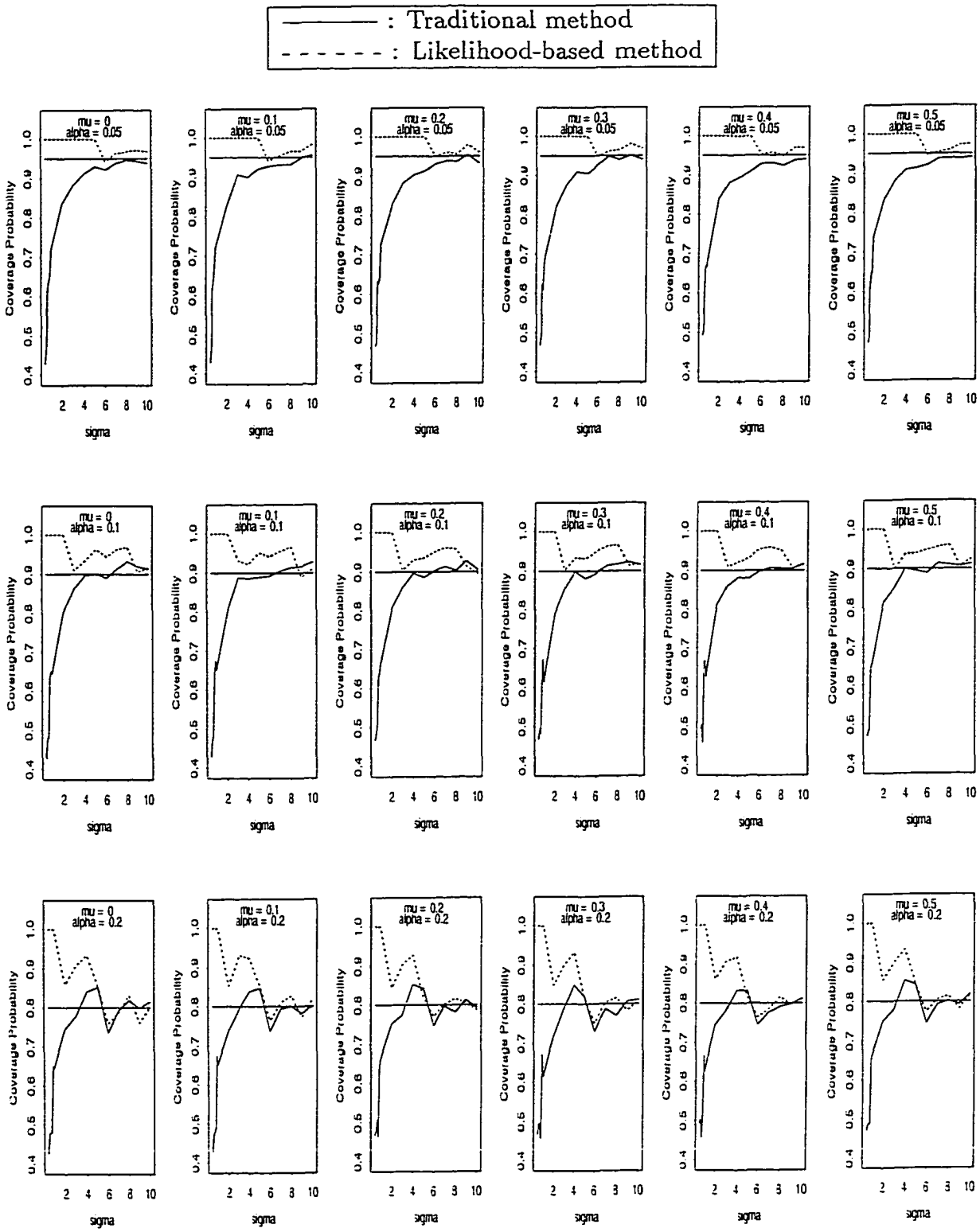

Figure 3.10 The estimated coverage probability for the traditional method and the corrected likelihood-based method (using Table 3.2, Table 3.3, and $d(n, \alpha)$) at sample size $n = 2$.

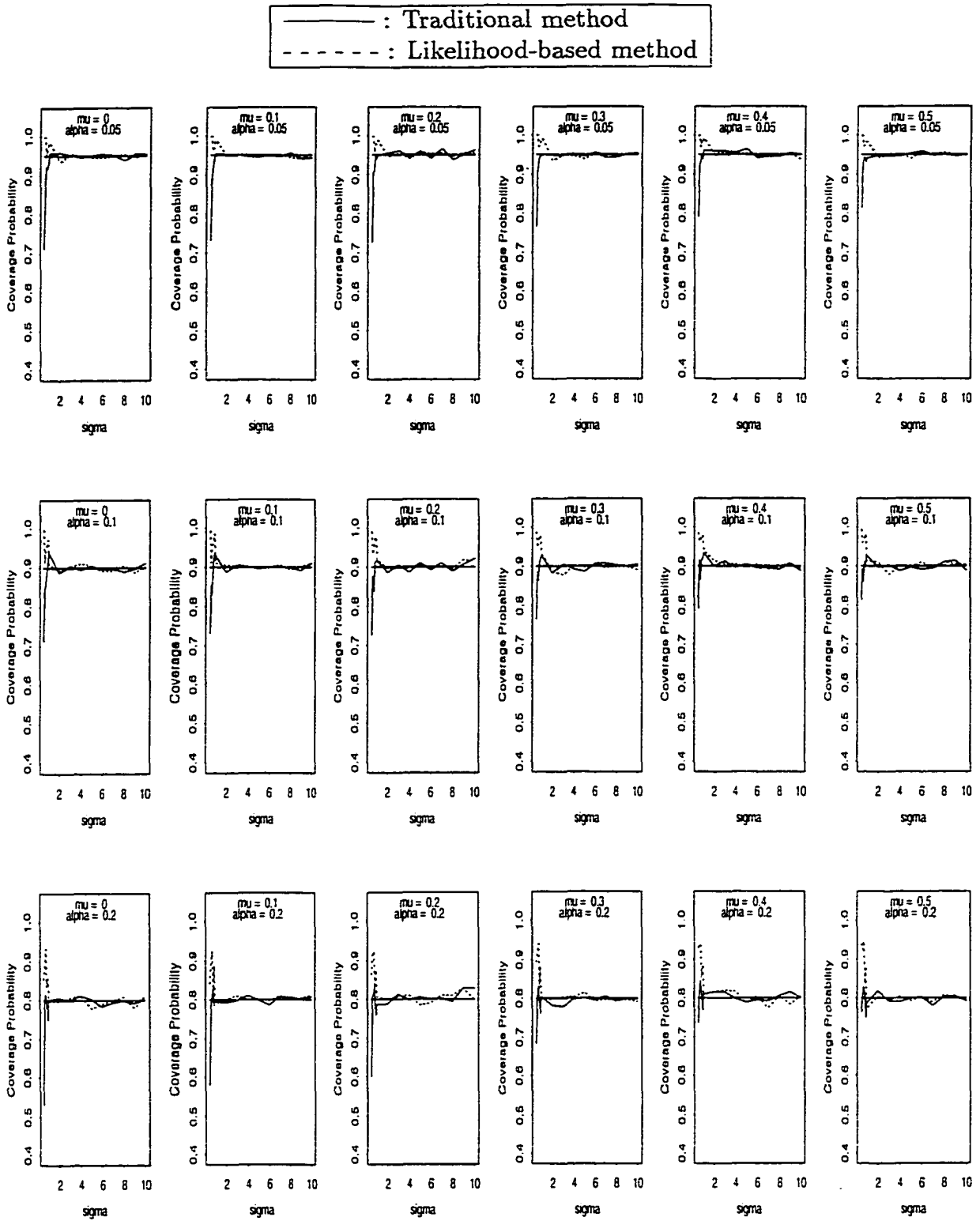

Figure 3.11 The estimated coverage probability for the traditional method and the corrected likelihood-based method (using Table 3.2, Table 3.3, and $d(n, \alpha)$) at sample size $n = 5$.

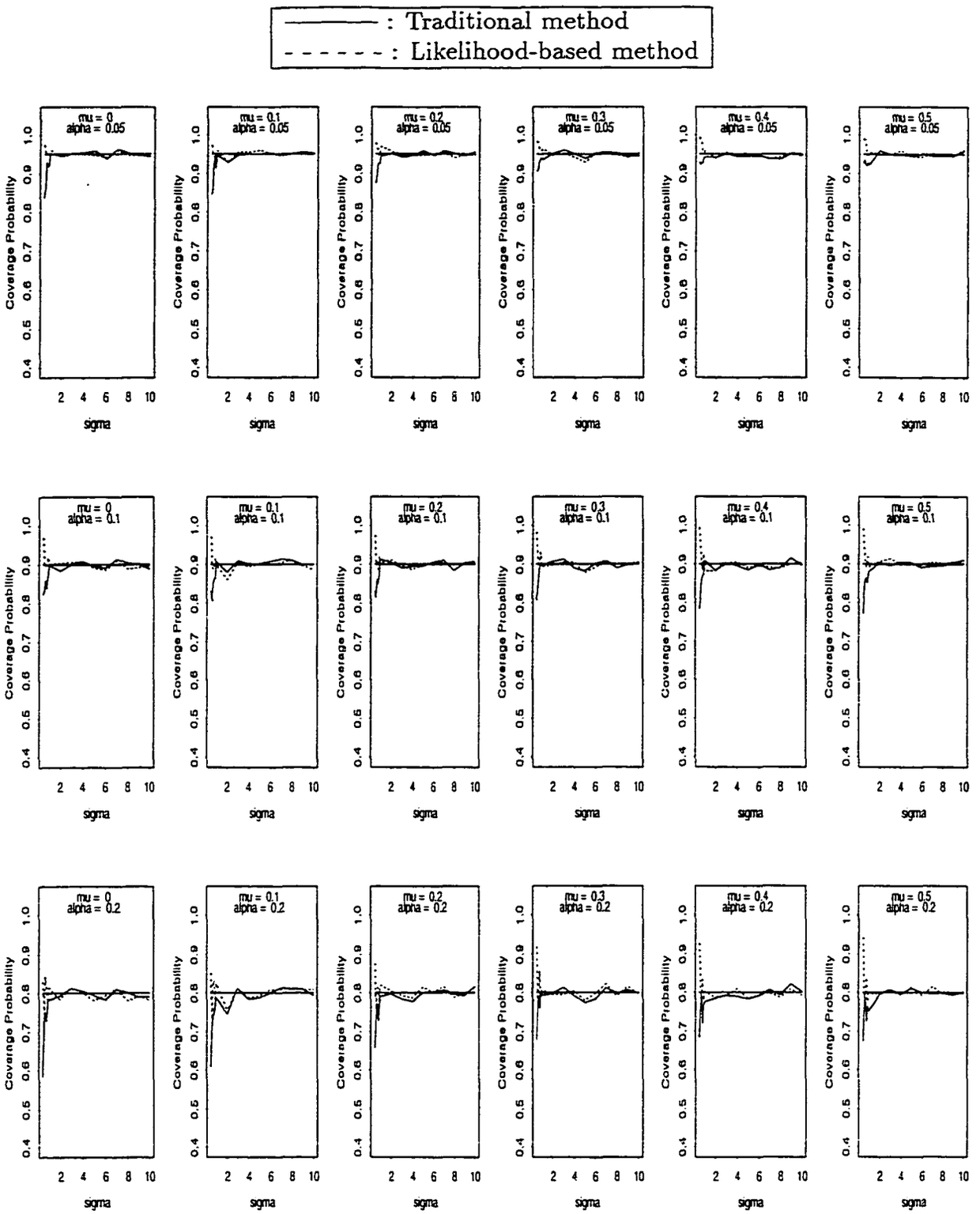

Figure 3.12 The estimated coverage probability for the traditional method and the corrected likelihood-based method (using Table 3.2, Table 3.3, and $d(n, \alpha)$) at sample size $n = 10$.

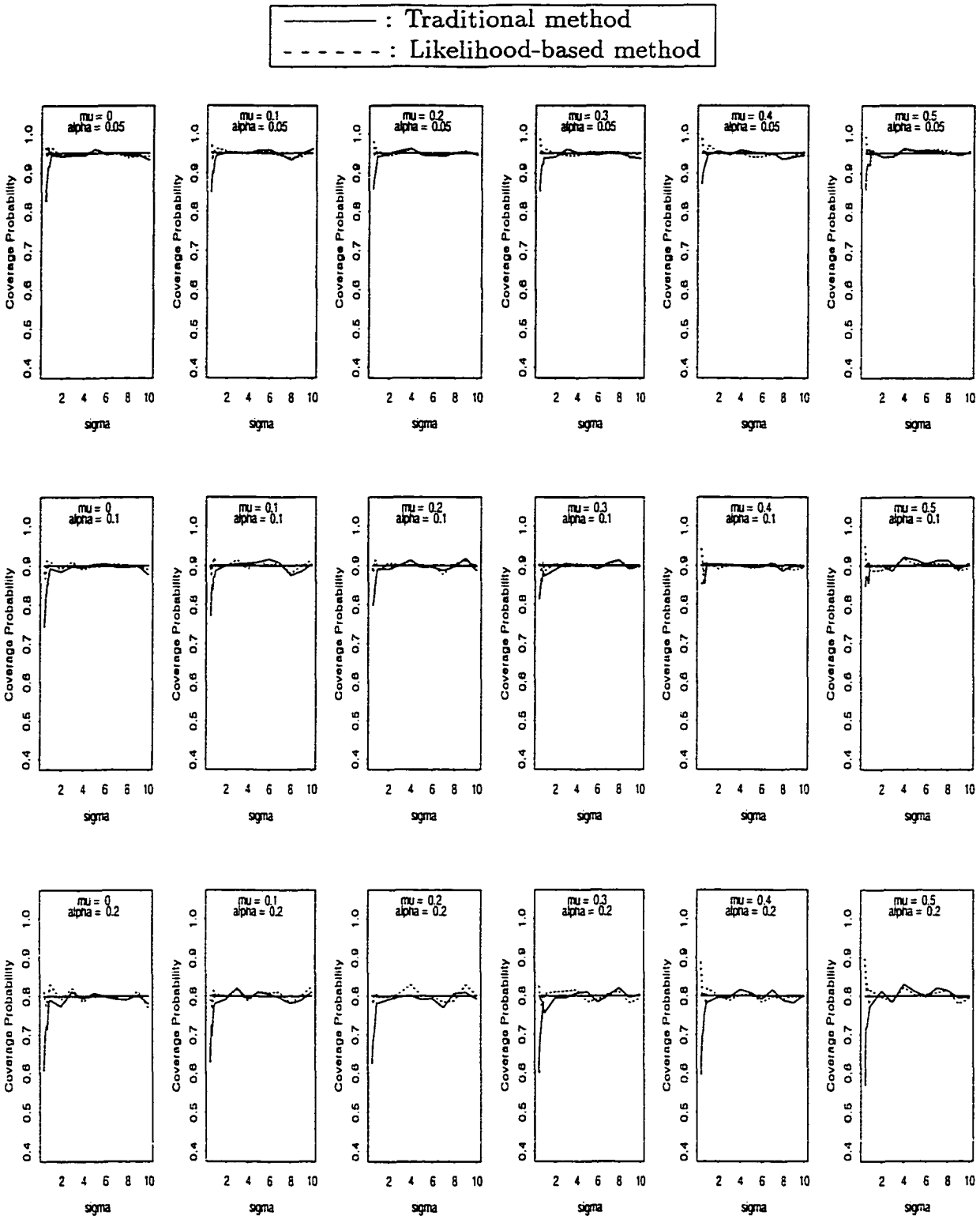

Figure 3.13 The estimated coverage probability for the traditional method and the corrected likelihood-based method (using Table 3.2, Table 3.3, and $d(n, \alpha)$) at sample size $n = 15$.

$\mu \in \{0.0, 0.25, 0.5\}, \sigma \in \{0.3, 0.5, 1.0, 3.0, 5.0\}, n \in \{2, 3, 4, 5\}$, and $\alpha \in \{0.05, 0.10, 0.20\}$. The results are summarized in Tables 3.4, 3.5, and 3.6.

From these simulations, we may observe the following about the average lengths.

(1) At $\sigma = 0.3$, the average length grows as the value μ changes from 0.0 to 0.5. The reason for this is that the probability of obtaining a Case 2 sample increases with μ. We also find that this happens when $\sigma = 0.5$, but the increase in average length is not as pronounced.

For these two small σ cases, the average lengths for the traditional method are less than those from the likelihood-based method, but the simulations shows that the corresponding estimated coverage probabilities for the traditional method are much below nominal.

(2) When $\sigma \geq 1.0$, the likelihood-based method obviously has smaller average lengths than the traditional method. These tables also show that for $\sigma \geq 1.0$, changes in μ do not much affect inferences for σ.

3.9 Conclusion

To sum up the discussion in this paper, we may make the following conclusions.

(1) The simulations show that inferences for the parameter σ will not be much affected by changing the value of location parameter μ.

(2) When we have no prior evidence about the parameter σ, then the likelihood-based method (corrected by the use of the limits in Table 3.2 and Table 3.3 for Case 1 and Case 2 samples and $d(n, \alpha)$ in Table 3.1 when the sample range is at least 2) is suggested for the estimation of σ. When σ is small, say $\sigma \leq 1$, the simulations show that the likelihood-based method has more conservative coverage probabilities than the traditional method. For large σ, the estimated coverage probabilities from the likelihood-

Table 3.4 The estimated average lengths for the traditional method (t) and the final corrected likelihood-based method (l) for $\mu = 0.0$.

$\mu = 0.0$											
σ		0.3		0.5		1.0		3.0		5.0	
n	α	t	l	t	l	t	l	t	l	t	l
2	0.05	4.027	7.659	11.502	11.190	25.875	16.839	75.066	39.136	125.637	63.668
	0.10	1.976	3.817	5.643	5.574	12.695	8.345	36.829	19.249	61.641	31.270
	0.20	0.941	1.883	2.687	2.748	6.044	4.062	17.535	9.197	29.349	14.893
3	0.05	0.974	1.950	2.637	2.946	5.274	4.308	15.478	11.067	25.690	18.218
	0.10	0.648	1.351	1.756	2.035	3.511	2.920	10.305	7.376	17.104	12.124
	0.20	0.409	0.929	1.108	1.377	2.216	1.906	6.503	4.673	10.794	7.662
4	0.05	0.592	1.152	1.503	1.663	2.976	2.568	8.699	6.951	14.560	11.589
	0.10	0.431	0.896	1.093	1.266	2.165	1.897	6.329	5.058	10.592	8.426
	0.20	0.295	0.685	0.747	0.934	1.480	1.331	4.327	3.464	7.243	5.765
5	0.05	0.422	0.873	1.171	1.292	2.231	2.002	6.412	5.422	10.551	8.891
	0.10	0.319	0.715	0.887	1.025	1.690	1.537	4.858	4.112	7.993	6.740
	0.20	0.227	0.574	0.630	0.785	1.199	1.110	3.446	2.915	5.669	4.774

Table 3.5 The estimated average lengths for the traditional method (t) and the final corrected likelihood-based method (l) for $\mu = 0.25$.

$\mu = 0.25$											
σ		0.3		0.5		1.0		3.0		5.0	
n	α	t	l	t	l	t	l	t	l	t	l
2	0.05	7.765	9.565	12.148	11.546	25.163	16.656	72.841	37.944	125.414	63.568
	0.10	3.810	4.769	5.960	5.753	12.346	8.256	35.738	18.665	61.531	31.221
	0.20	1.814	2.357	2.838	2.838	5.878	4.020	17.016	8.922	29.297	14.870
3	0.05	1.641	2.408	2.753	3.051	5.215	4.301	15.253	10.907	25.386	18.004
	0.10	1.093	1.673	1.833	2.111	3.472	2.918	10.156	7.270	16.901	11.981
	0.20	0.689	1.149	1.157	1.433	2.191	1.907	6.409	4.606	10.666	7.572
4	0.05	0.943	1.370	1.592	1.740	2.957	2.560	8.709	6.958	14.501	11.540
	0.10	0.686	1.061	1.158	1.327	2.152	1.893	6.336	5.063	10.550	8.391
	0.20	0.469	0.806	0.792	0.983	1.471	1.329	4.333	3.467	7.214	5.740
5	0.05	0.767	1.093	1.209	1.343	2.225	1.998	6.556	5.543	10.791	9.093
	0.10	0.581	0.889	0.916	1.070	1.685	1.533	4.967	4.203	8.175	6.893
	0.20	0.412	0.709	0.650	0.826	1.196	1.108	3.523	2.979	5.799	4.882

Table 3.6 The estimated average lengths for the traditional method (t) and the final corrected likelihood-based method (l) for $\mu = 0.5$.

$\mu = 0.5$											
σ		0.3		0.5		1.0		3.0		5.0	
n	α	t	l	t	l	t	l	t	l	t	l
2	0.05	11.413	11.415	13.483	12.205	25.519	16.689	74.243	38.680	127.328	64.566
	0.10	5.600	5.693	6.615	6.082	12.520	8.273	36.425	19.027	62.470	31.710
	0.20	2.666	2.816	3.149	3.001	5.961	4.028	17.343	9.095	29.744	15.102
3	0.05	2.453	2.954	2.871	3.139	5.391	4.381	15.753	11.257	25.533	18.108
	0.10	1.633	2.055	1.911	2.174	3.589	2.968	10.488	7.502	17.000	12.051
	0.20	1.030	1.409	1.206	1.476	2.265	1.934	6.619	4.751	10.728	7.616
4	0.05	1.498	1.729	1.705	1.819	2.975	2.562	8.809	7.039	14.749	11.739
	0.10	1.090	1.333	1.240	1.387	2.165	1.892	6.408	5.122	10.730	8.535
	0.20	0.745	1.004	0.848	1.027	1.480	1.327	4.382	3.508	7.337	5.839
5	0.05	1.089	1.341	1.280	1.426	2.226	1.995	6.477	5.477	10.650	8.974
	0.10	0.825	1.087	0.970	1.139	1.686	1.531	4.907	4.154	8.068	6.803
	0.20	0.585	0.865	0.688	0.884	1.196	1.105	3.481	2.944	5.723	4.818

based method are quite close to those for the traditional method (near the nominal value $(1-\alpha)$), while the likelihood intervals have smaller average length.

Appendix

Aids for producing endpoints of the intervals prescribed by (3.7) for cases of range ≥ 2 will be discussed in this appendix.

As mentioned in the Appendix (B) of [1], approximate values of $L^*(\sigma)$ and M for cases with range ≥ 2 are

$$L^*(\sigma) \doteq \sum_i n_i \ln(\frac{1}{\sigma}\phi(\frac{i-\bar{x}}{\sigma}))$$

and

$$M \doteq \sum_i n_i \ln(\frac{1}{\hat{\sigma}}\phi(\frac{i-\bar{x}}{\hat{\sigma}})).$$

Substituting above two approximations into the inequality

$$-2\,(L^*(\sigma) - M) \leq d(n,\alpha)$$

one gets

$$\frac{\hat{\sigma^2}}{\sigma^2} - \ln(\frac{\hat{\sigma^2}}{\sigma^2}) \leq 1 + \frac{1}{n}\,d(n,\alpha)$$

where $\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$. (This is the inequality defining the likelihood-based interval one would obtain ignoring the rounding altogether.) Then let x_1, x_2 be two solutions of the equality $x - \ln(x) = 1 + \frac{1}{n}\,d(n,\alpha)$ with $x_1 < x_2$. An approximation for the corrected likelihood-based interval in the case that the range is 2 or more is $(\frac{\hat{\sigma}}{\sqrt{x_2}}, \frac{\hat{\sigma}}{\sqrt{x_1}})$.

References

[1] Lee, Chiang-Sheng and Vardeman, Stephen B. (1999). *Interval Estimators of the Parameter μ for Rounded Normal Data.* Iowa State University, Ames, IA.

4 ANALYSIS OF ROUNDED DATA FROM THE BALANCED ONE-WAY RANDOM EFFECTS MODEL

A paper to be submitted to the Communications in Statistics

Chiang-Sheng Lee and Stephen B. Vardeman
Iowa State University, Ames, IA 50011-1210

4.1 Introduction

It is a practical problem that data on hand are sometimes obtained using crude gauging. We might call such data "rounded data" since they are really obtained by rounding to the nearest unit. Do traditional statistical methods still provide good estimates of unknown parameters when rounded data are analyzed ? What can we do if they do not?

Building on the discussions of the interval estimation of parameters μ and σ for a single rounded Normal sample in [2] and [3], we extend our discussion of rounded data to the one-way random effects model. Usually, the balanced one-way random effects model is expressed in the form

$$Y_{ij} = \mu_i + \epsilon_{ij}, \quad i = 1, 2, ..., m; \; j = 1, 2, ..., n, \tag{4.1}$$

Mr. Lee is a Ph. D. Candidate in Industrial Engineering in the Industrial and Manufacturing Systems Engineering Department. email: chiang@iastate.edu

Dr. Vardeman is a Professor in the Statistics and Industrial and Manufacturing Systems Engineering Departments. He is a Senior Member of ASQ.

where the μ_i's are a random sample from $N(\mu, \sigma_\tau^2)$, the ϵ_{ij}'s are a random sample from $N(0, \sigma^2)$, and the μ_i's are independent of the ϵ_{ij}'s. The variance components σ_τ and σ are of primary interest in this model and we concentrate our discussion on their interval estimation.

We start with the likelihood function in the next section, then find approximate maximizers for the likelihood function in Section 4.3. Three special cases will be discussed in the fourth section. Inference methods for the parameters σ and σ_τ will be discussed in Sections 4.5 and 4.6, and final conclusions will be drawn in the last section.

4.2 The Rounded Data Likelihood Function

Without loss of generality, we assume that all observed values y_{ij} are integers, and utilize the vector form $\underline{y} = (y_{11}, y_{12}, ..., y_{1,n}, y_{21}, ..., y_{2,n}, ..., y_{m,1}, ..., y_{m,n})'$ for convenience. The rounded data likelihood function $f(\underline{y}; \mu, \sigma_\tau, \sigma)$ can be found by dealing with the $N(\mu_i, \sigma^2)$ conditional distributions of unrounded values given $(\mu_1, \mu_2, ..., \mu_m)$, and has the form

$$\begin{aligned} & f(\underline{y}; \mu, \sigma_\tau, \sigma) \\ &= \prod_{i=1}^{m} \int_{-\infty}^{+\infty} \prod_{j=1}^{n} [\ \Phi(\frac{y_{ij} + 0.5 - \mu_i}{\sigma}) - \Phi(\frac{y_{ij} - 0.5 - \mu_i}{\sigma})\ ] \phi(\mu_i; \mu, \sigma_\tau)\ d\mu_i \\ &= \prod_{i=1}^{m} \mathrm{E}_i[\ \prod_{j=1}^{n} (\ \Phi(\frac{y_{ij} + 0.5 - \mu_i}{\sigma}) - \Phi(\frac{y_{ij} - 0.5 - \mu_i}{\sigma})\ )\ ], \end{aligned} \tag{4.2}$$

where Φ stands for the standard normal cumulative distribution function, $\phi(x; \mu, \sigma_\tau)$ is the $N(\mu, \sigma_\tau^2)$ density, and E_i is expectation with respect to the variable μ_i.

Furthermore, define the log-likelihood function $\mathcal{L}(\mu, \sigma_\tau, \sigma) = \ln(f(\underline{y}; \mu, \sigma_\tau, \sigma))$, two profile log-likelihood functions

$$\mathcal{L}^*(\sigma_\tau) = \sup_{\mu \in R} \sup_{\sigma > 0} \mathcal{L}(\mu, \sigma_\tau, \sigma)$$

and

$$\mathcal{L}^*(\sigma) = \sup_{\mu \in R} \sup_{\sigma_\tau > 0} \mathcal{L}(\mu, \sigma_\tau, \sigma),$$

and the supremum value

$$\mathcal{M} = \sup_{\mu \in R} \sup_{\sigma_\tau > 0} \sup_{\sigma > 0} \mathcal{L}(\mu, \sigma_\tau, \sigma).$$

4.3 Approximate Maximizers of $\mathcal{L}(\mu, \sigma_\tau, \sigma)$

We may sometimes approximate $(\Phi(\frac{y_{ij} + 0.5 - \mu_i}{\sigma}) - \Phi(\frac{y_{ij} - 0.5 - \mu_i}{\sigma}))$, as mentioned in [2], by $\frac{1}{\sigma}\,\phi(\frac{y_{ij} - \mu_i}{\sigma})$, where ϕ is the standard normal density. Substiting this into the log-likelihood function $\mathcal{L}(\mu, \sigma_\tau, \sigma)$, under some circumstances

$$\mathcal{L}(\mu, \sigma_\tau, \sigma) \doteq \ln \Big(\prod_{i=1}^{m} \int_{-\infty}^{+\infty} \prod_{j=1}^{n} [\frac{1}{\sigma}\,\phi(\frac{y_{ij} - \mu_i}{\sigma})]\; \phi(\mu_i; \mu, \sigma_\tau)\; d\mu_i\Big)$$

$$= C - \frac{m(n-1)}{2}\ln(\sigma^2) - \frac{m}{2}\ln(\sigma^2 + n\sigma_\tau^2) - \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2}{2\sigma^2} - \frac{n\sum_{i=1}^{m}(\bar{y}_{i.} - \mu)^2}{2(\sigma^2 + n\sigma_\tau^2)},$$

where $C = -mn\,\ln\sqrt{2\pi}$ and $\bar{y}_{i.} = \frac{\sum_{j=1}^{n} y_{ij}}{n}$. This is the "exact data" log-likelihood, appropriate if there is no rounding.

After setting three partial derivatives with respect to parameters μ, σ_τ^2, and σ^2 equal to 0, one can see that the maximizers $(\hat{\mu}, \hat{\sigma}_\tau^2, \hat{\sigma}^2)$ of this approximate log-likelihood (for parameters $(\mu, \sigma_\tau^2, \sigma^2)$) are

$$(\bar{y}_{..},\ \max\{0,\ \frac{\sum_{i=1}^{m}(\bar{y}_{i.} - \bar{y}_{..})^2}{m} - \frac{\hat{\sigma}^2}{n}\},\ \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2}{m(n-1)}),$$

where $\bar{y}_{..} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} y_{ij}}{m\,n}$. These three approximate maximizers are the maximum likelihood estimates when we treat "rounded data" as "exact."

4.4 Special Cases in the Maximization of $\mathcal{L}(\mu, \sigma_\tau, \sigma)$

In this section, three special data types will be discussed. We will call these Case I, Case II, and Case III data types. For each case, we find an approximate form for the supremum value $\mathcal{M}$, and also indicate $(\hat{\sigma}_\tau^2,\ \hat{\sigma}^2)$ values that result from assuming the data are "exact."

4.4.1 Case I

Case I is the situation where all the observations y_{ij} have the same value, say $y_{ij} = y$ for all i and j. This case occurs often when both σ and σ_τ values are small. For example, $\underline{y} = (1,\ 1;\ 1,\ 1)$ is a Case I sample for $m = n = 2$.

In this case, we find that the supremum value of $\mathcal{M}$ is 0 and conclude that there are many triples $(\mu, \sigma_\tau, \sigma)$ producing nearly this supremum value. A discussion of this point is in Appendix (A). If we treat these Case I data as "exact," then we get $(\hat{\sigma}_\tau^2, \hat{\sigma}^2) = (0, 0)$, which gives us the second characterization this data type.

4.4.2 Case II

Case II is the situation where observations in a given sample all have the same value, $y_{ij} = y_i$ for all j. For example, $\underline{y} = (0,\ 0,\ 0;\ 2,\ 2,\ 2)$ is a Case II sample for $m = 2$ and $n = 3$. This data type occurs often when σ is small.

From the discussion in Appendix (B), the supremum value of $f(\underline{y}; \mu, \sigma_\tau, \sigma)$ approximates the supremum value of $\prod_{i=1}^{m} \left[\ \Phi(\frac{y_i + 0.5 - \mu}{\sigma_\tau}) - \Phi(\frac{y_i - 0.5 - \mu}{\sigma_\tau})\ \right]$, which is the likelihood function with rounded sample $y_1,\ y_2,\ \ldots,\ y_m$ from $N(\mu, \sigma_\tau^2)$. This shows that in Case II an approximate value for $\mathcal{M}$ can be computed from the single sample problem.

Under this case, we get $\hat{\sigma}_\tau > 0$ and $\hat{\sigma} = 0$ when we plug the data into the formulas in Section 4.3, which provides another way to characterize such samples.

4.4.3 Case III

Case III is the situation where if rounding is ignored the sample values cause $\hat{\sigma}_\tau = 0$ and $\hat{\sigma} > 0$. This requires $\frac{\sum_{i=1}^{m}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2}{m} - \frac{\hat{\sigma}^2}{n} \leq 0$. For instance, data $\underline{y} = (-1,\ 0;\ 0,\ 1)$ makes this difference 0 and $\underline{y} = (1,\ 2,\ 3; -1,\ 3,\ 3)$ makes the same difference < 0. One

obvious result is that if any sample gives $\bar{y}_{i\cdot} = y$ for all i, then such a sample always has $\hat{\sigma}_\tau = 0$. Such data occur with high probability when σ is large and σ_τ is small.

In Appendix (C), the discussion shows that in this case the supremum value of $f(\underline{y}; \mu, \sigma_\tau, \sigma)$ is close to the supremum value of $\prod_{i=1}^{m}\prod_{j=1}^{n}[\Phi(\frac{y_{ij}+0.5-\mu}{\sigma}) - \Phi(\frac{y_{ij}-0.5-\mu}{\sigma})]$, the likelihood function for a single rounded sample of size mn from the $N(\mu, \sigma^2)$ distribution.

4.5 Inference for the Parameter σ

In this section, two methods of estimating the parameter σ are compared. The first one is what we call the traditional method, and the other is what we will call the likelihood-based method. They are mainly introduced in Section 4.5.1.

4.5.1 The Construction of Confidence Intervals

The traditional method for estimating the parameter σ in the one-way random effects model is to apply the fact that (with "exact" data)

$$\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\cdot})^2}{\sigma^2} \sim \chi^2_{m(n-1)}.$$

This produces the corresponding $(1-\alpha)$ level confidence interval for σ

$$[\sqrt{\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\cdot})^2}{\chi^2_{(m(n-1),\,1-\frac{\alpha}{2})}}},\ \sqrt{\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i\cdot})^2}{\chi^2_{(m(n-1),\,\frac{\alpha}{2})}}}], \tag{4.3}$$

where $\chi^2_{(r,q)}$ is the q quantile of χ^2 distribution with degrees of freedom r.

The (initial) likelihood-based method is to apply the result that under $H_0 : \sigma = \sigma_0$ and for large n,

$$-2\ln(\frac{\sup_{\mu\in R}\sup_{\sigma_\tau>0} f(\underline{y};\mu,\sigma_\tau,\sigma_0)}{\sup_{\mu\in R}\sup_{\sigma_\tau>0}\sup_{\sigma>0} f(\underline{y};\mu,\sigma_\tau,\sigma)}) \dot{\sim} \chi^2_{(1)},$$

or to use the notation in this paper

$$-2\,(\,\mathcal{L}^*(\sigma_0) - \mathcal{M}\,) \dot{\sim} \chi^2_{(1)}. \tag{4.4}$$

Using this large n result, an approximately $(1-\alpha)$ level confidence interval for the parameter σ is the set of all σ points satisfying

$$-2\ (\ \mathcal{L}^*(\sigma)\ -\ \mathcal{M}\ ) \leq \chi^2_{(1,\ 1-\alpha)}. \tag{4.5}$$

4.5.2 Simulations

In this section, the Monte Carlo method is used to compare the two methods for interval estimation of σ introduced in Section 4.5.1. First we generate m values $\mu_1,\ \mu_2,\ ....,\ \mu_m$ from $N(\mu, \sigma_\tau^2)$ and mn values $\epsilon_{11},\ \epsilon_{12},\ ....,\ \epsilon_{1,n},\ ....,\ \epsilon_{m,n}$ from $N(0, \sigma^2)$, compute $y_{ij} = \mu_i + \epsilon_{i,j}$ and then round to integers. The next step is to find the traditional interval directly from (4.3) and get the value $-2\ (\mathcal{L}^*(\sigma) - \mathcal{M})$. If the interval (4.3) for the ith sample covers the value σ, then we add 1 to a counter t ("t" for the traditional method), otherwise we do not increase it. If the inequality (4.5) is satisfied, then we add 1 to the counter l ("l" for the likelihood-based method), otherwise we do nothing. We repeat these steps 1000 times and finally get the values $\frac{t_{1000}}{1000}$ and $\frac{l_{1000}}{1000}$, which are the estimated coverage probabilities for these two methods.

Take the case $m = n = 2$ for illustration, supposing $\sigma = 0.5$ at $\alpha = 0.05$. Assume the first sample is $\underline{y} = (-1,\ 0;\ 0,\ 1)$. Then the interval (4.3) is $[0.36816,\ 4.44398]$, and $-2\ (\mathcal{L}^*(0.5) - \mathcal{M}) = 0.17061$. It is clear that interval (4.3) covers $\sigma = 0.5$ and that $-2\ (\mathcal{L}^*(0.5) - \mathcal{M}) \leq 3.84146$, so we add 1 to both counters t and l and get $t_1 = l_1 = 1$. Next, suppose the second sample is $\underline{y} = (0,\ 0;\ 1,\ 1)$, then the interval (4.3) degenerates to 0 and $-2\ (\mathcal{L}^*(0.5) - \mathcal{M}) = 3.00615$. Thus we have $t_2 = 1$ and $l_2 = 2$, and so on. We repeat this processes 1000 times and finally get the values $\frac{t_{1000}}{1000}$ and $\frac{l_{1000}}{1000}$.

We ran simulations for parameter/design combinations with $\mu = \{0.0, 0.3, 0.5\}$, $\sigma_\tau = \{0.5, 0.8, 1, 3, 5, 10\}$, $\sigma = \{(0.5, 1.0)[0.1], 3, 5, 7, 10\}$, $(m, n) = \{(2,2), (2,3), (2,5), (3,2), (3,3), (3,5), (4,2), (4,3), (5,2), (5,3)\}$, and $\alpha = \{0.05,\ 0.10,\ 0.20\}$, (where $(a, b)[c]$ means the values from a to b with increment c.) After examining those results, we found

that for fixed (m, n), they were little affected by the choice of μ. Therefore, we here represent those results in Figures 4.1 to 4.4 by presenting only the $\mu = 0$ case for $(m, n) \in \{(2,2), (2,5), (3,3), (5,2)\}$. In these figures, the solid lines identify estimated coverage probabilities for the traditional interval (4.3), and the dashed lines identify the estimated coverage probabilities for the likelihood-based method (4.5).

4.5.3 Improving the Coverage Probability Calibration of Likelihood-Based Method

The figures show that the estimated coverage probabilities for the likelihood-based method can be much lower than the nominal value $(1-\alpha)$ in small samples. Adjustment of the $\chi^2_{(1,1-\alpha)}$ value in (4.5) (that is appropriate asymtotically) is needed to improve these probabilities for small samples.

We reason that with exact data the estimation of σ in the one-way model is in some sense the same problem as estimation of the standard deviation of a single distribution based on a sample of size $m(n-1)+1$. We found in [3] that when estimating σ from a small rounded normal sample, in order to maintain a nominal confidence level it was necessary to replace an asymptotically appropriate $\chi^2_{(1,1-\alpha)}$ value with a larger value we called $d(n, \alpha)$. We have found empirically that applying those values in the present context cures the small-design deficiencies of the likelihood method (4.5). The method is to replace the $\chi^2_{(1,1-\alpha)}$ value in (4.5) by $d(m(n-1)+1,\ \alpha)$ from Table 3.1 of [3]. Figures 4.5 through 4.8 compare the estimated coverage probabilities for those two methods for the same parameter combinations as in Figures 4.1-4.4, but this time the likelihood-based method uses $d(m(n-1)+1,\ \alpha)$ in place of $\chi^2_{(1,1-\alpha)}$ in (4.5). In the new figures, the estimated coverage probabilities are quite close to $(1-\alpha)$.

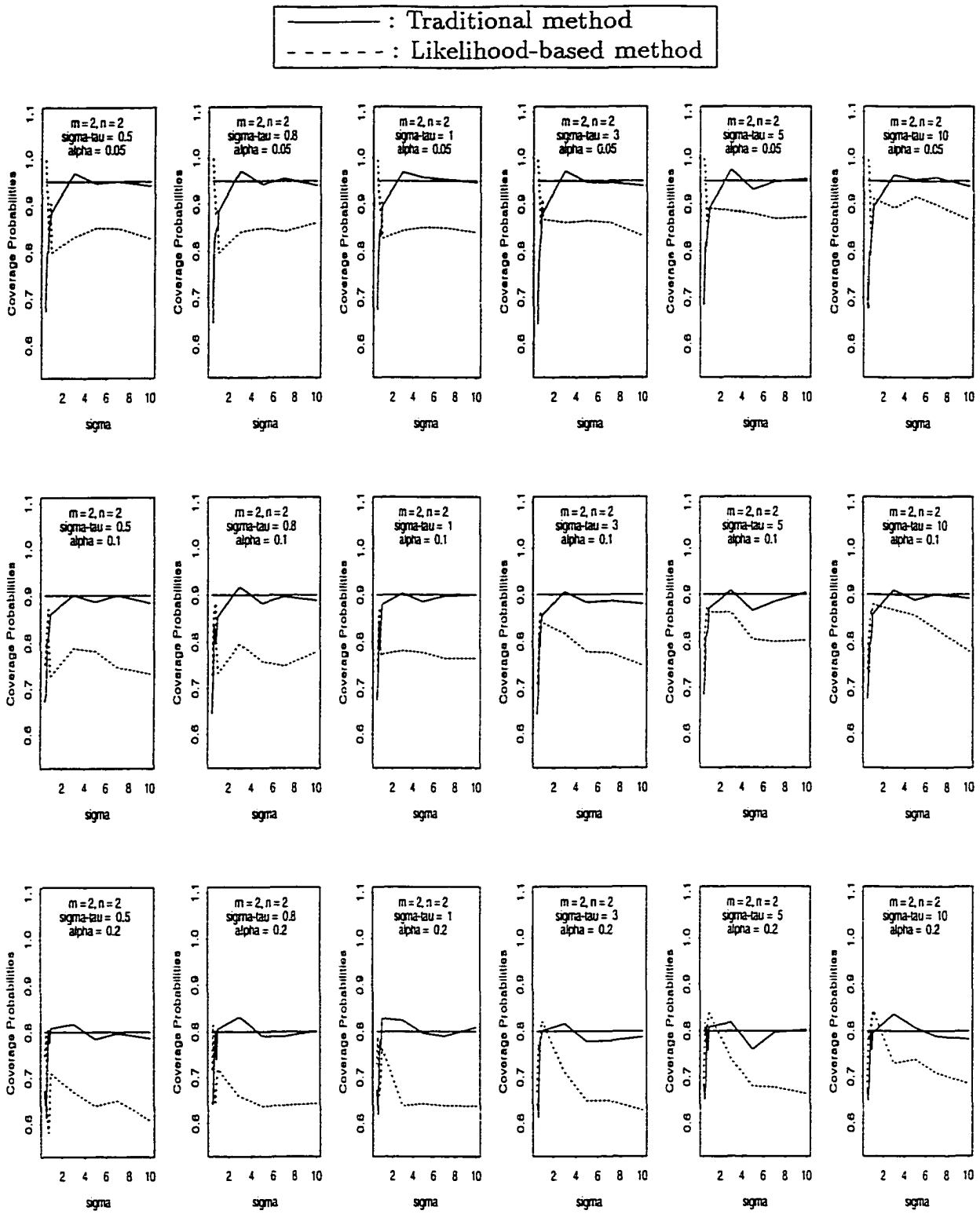

Figure 4.1 Estimated coverage probabilties for σ, $m = 2$ and $n = 2$.

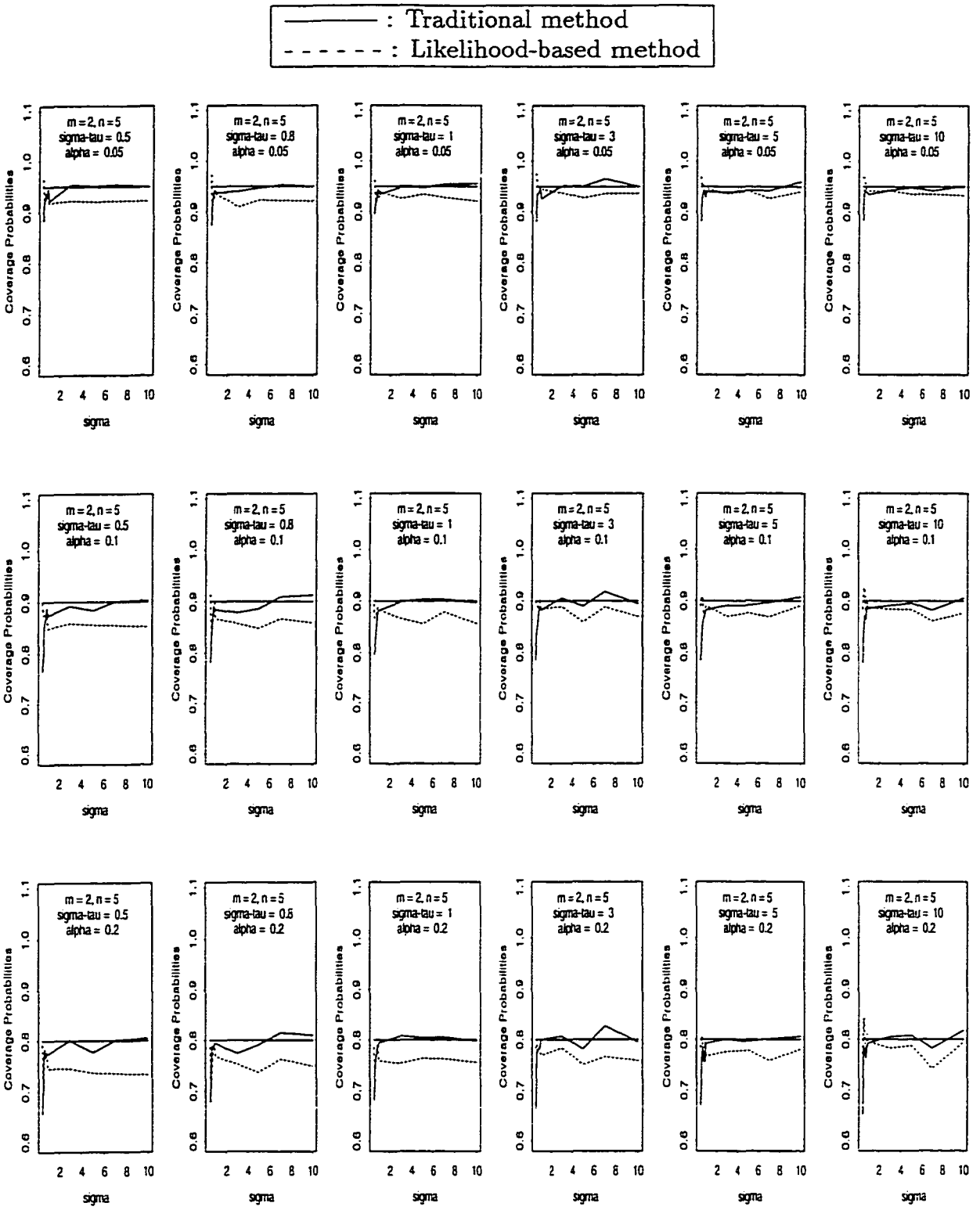

Figure 4.2 Estimated coverage probabilties for σ, $m = 2$ and $n = 5$.

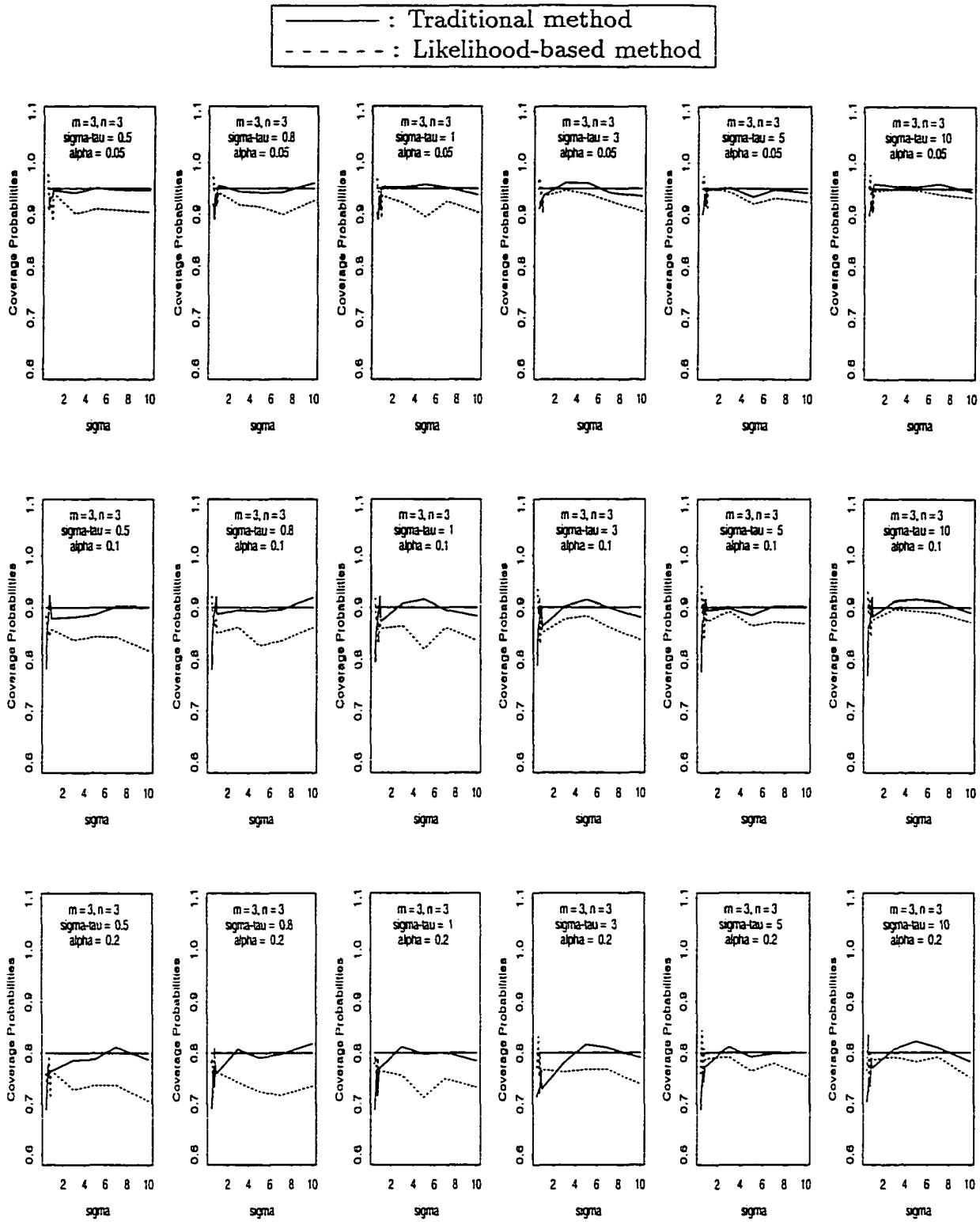

Figure 4.3 Estimated coverage probabilties for σ, $m = 3$ and $n = 3$.

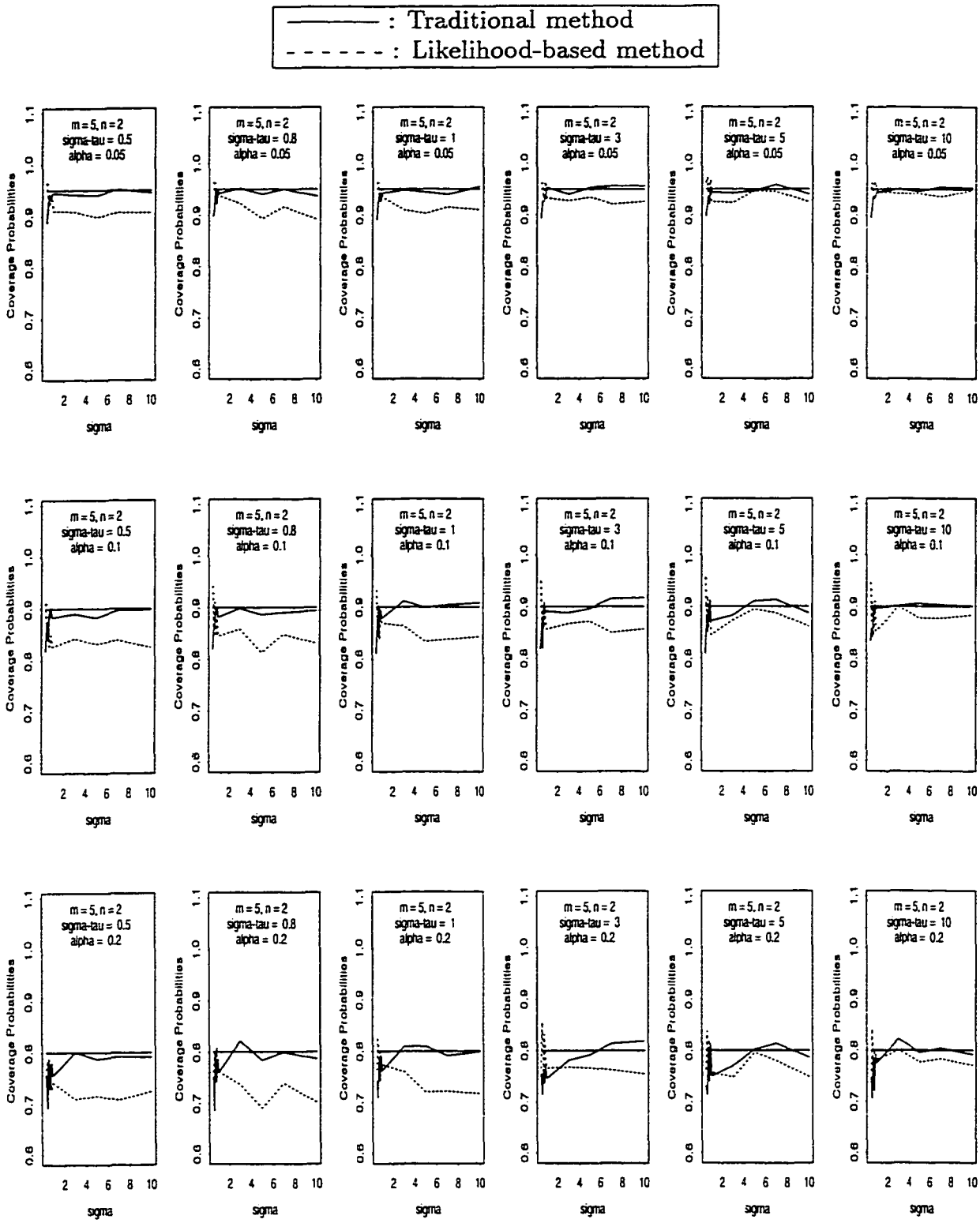

Figure 4.4 Estimated coverage probabilties for σ, $m = 5$ and $n = 2$.

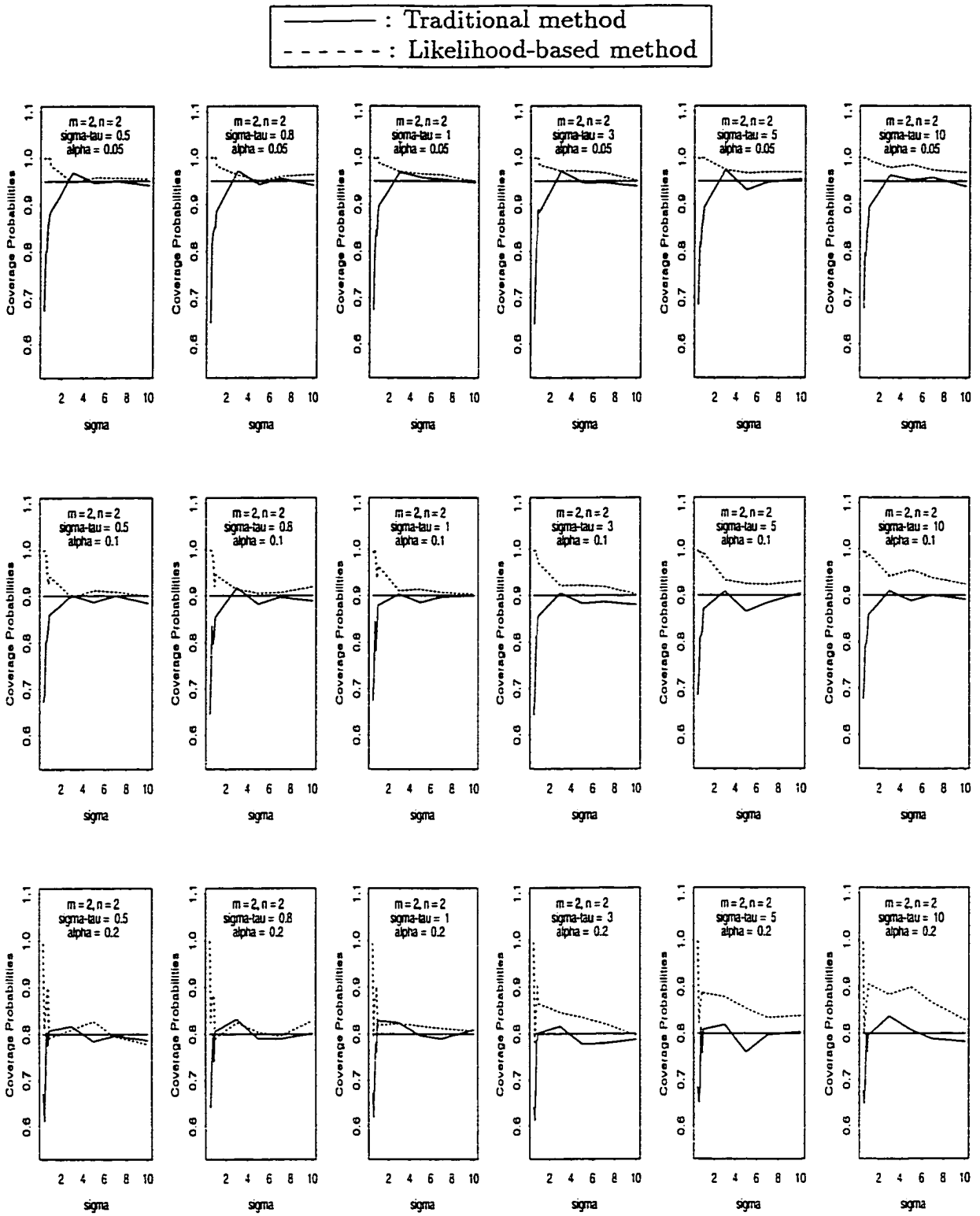

Figure 4.5 Estimated coverage probabilties for σ, $m = 2$ and $n = 2$ (corrected likelihood method).

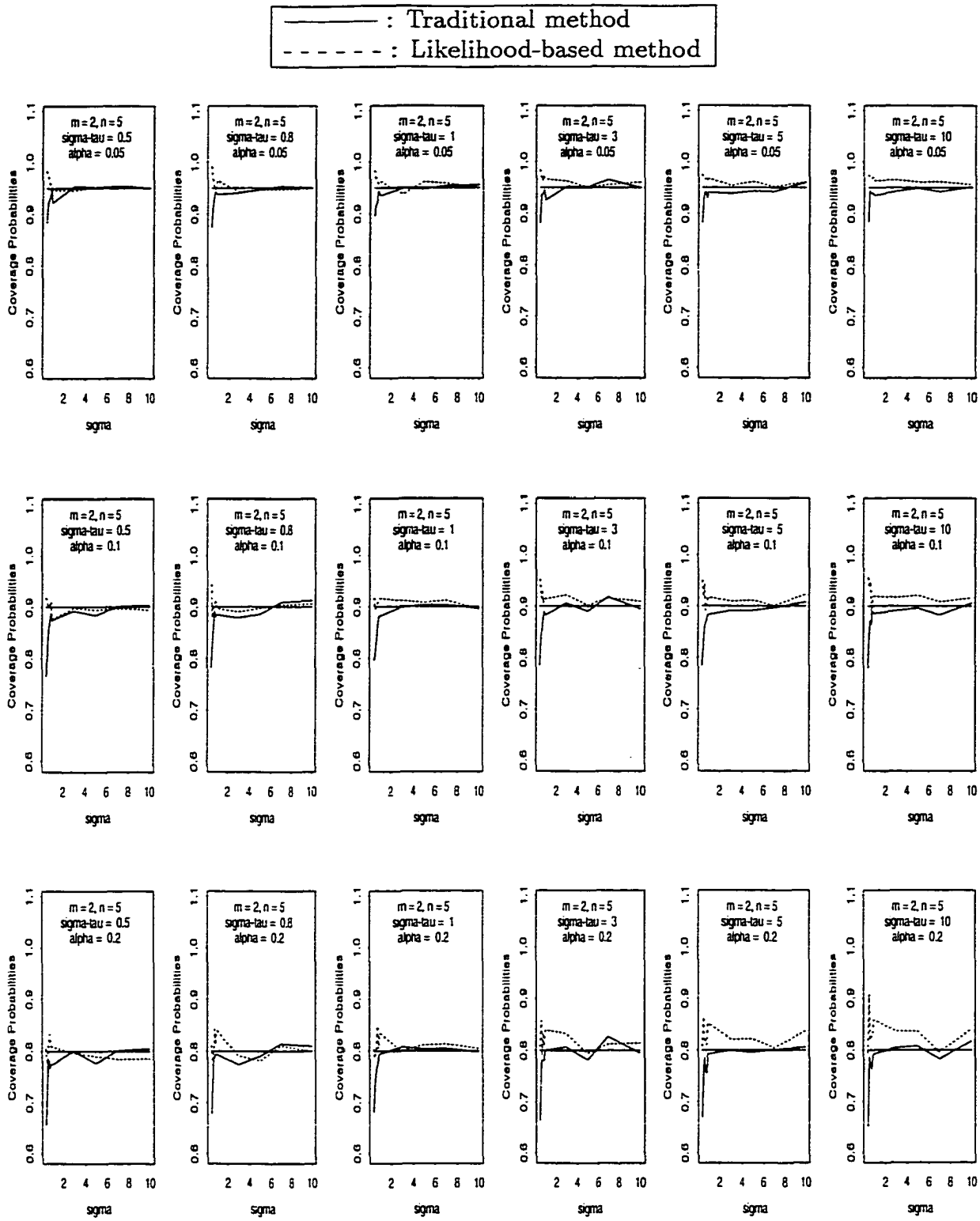

Figure 4.6 Estimated coverage probabilties for σ, $m = 2$ and $n = 5$ (corrected likelihood method).

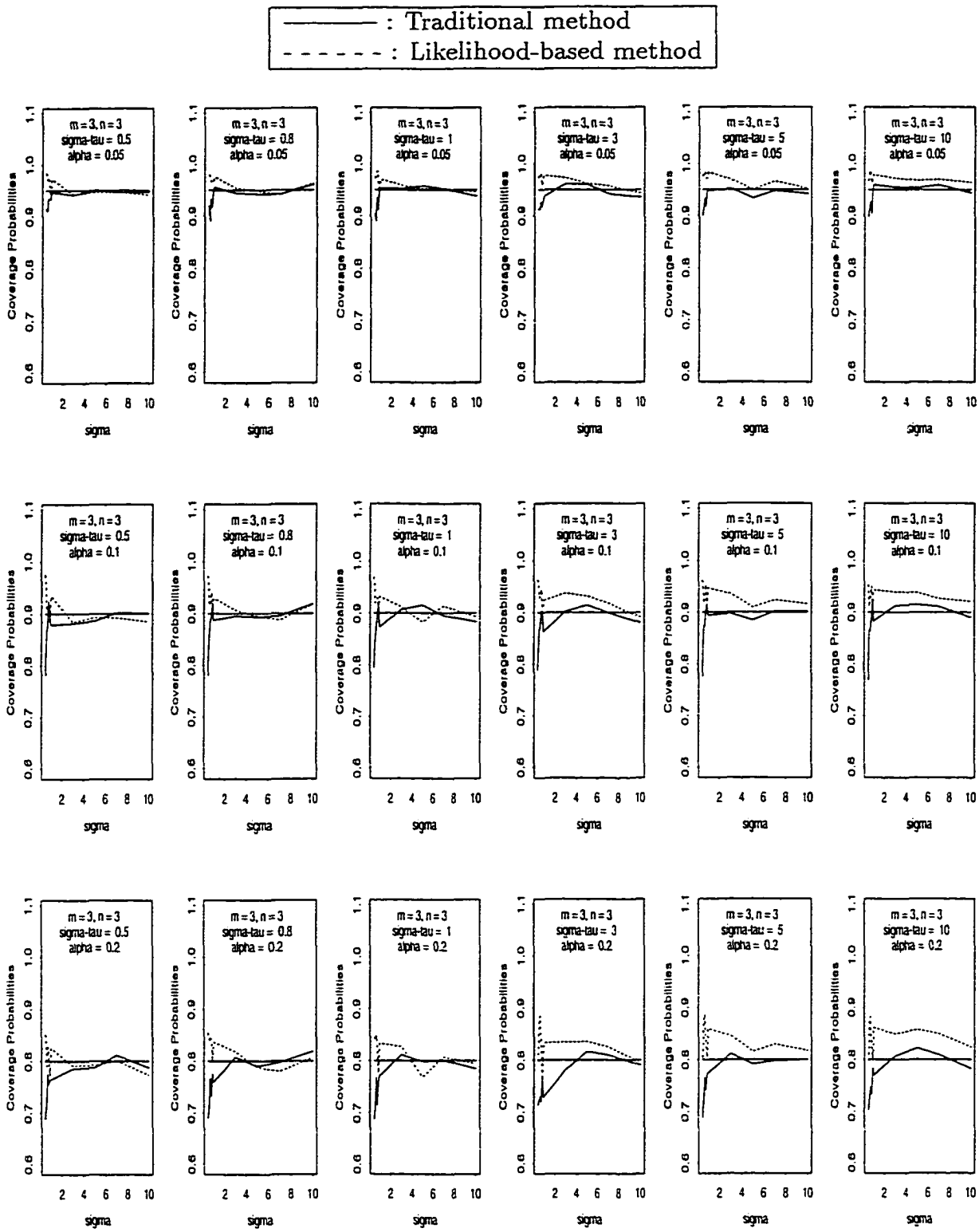

Figure 4.7 Estimated coverage probabilties for σ, $m = 3$ and $n = 3$ (corrected likelihood method).

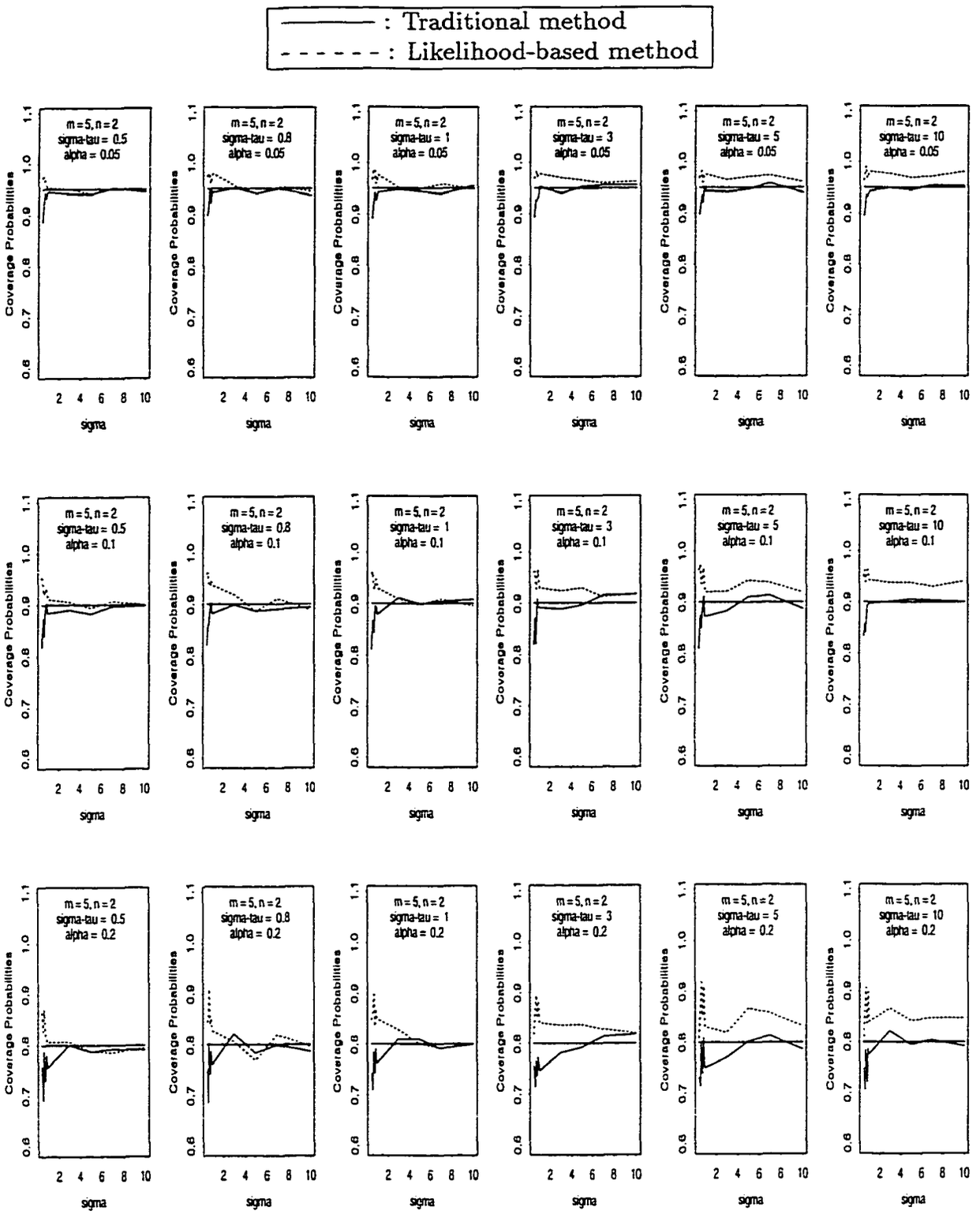

Figure 4.8 Estimated coverage probabilties for σ, $m = 5$ and $n = 2$ (corrected likelihood method).

4.6 Inference for the Parameter σ_τ

We now turn to the problem of estimating the parameter σ_τ.

4.6.1 The Construction of Confidence Intervals

In this discussion, we adopt the method described on page 61 of the text [1] as the "traditional" or exact data method of interval estimation for σ_τ. In the notations of this paper, that confidence interval for σ_τ is

$$[\sqrt{\max\{\ 0,\ \frac{S_1^2 - S_2^2 - \sqrt{V_L}}{n}\}},\ \sqrt{\max\{\ 0,\ \frac{S_1^2 - S_2^2 + \sqrt{V_U}}{n}\}}\ ], \tag{4.6}$$

where

$$\begin{aligned}
S_1^2 &= \frac{n \sum_{i=1}^{m} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2}{m-1}, \\
S_2^2 &= \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i\cdot})^2}{m(n-1)}, \\
V_L &= G_1^2\ S_1^4 + H_2^2\ S_2^4 + G_{12}\ S_1^2\ S_2^2, \\
V_U &= H_1^2\ S_1^4 + G_2^2\ S_2^4 + H_{12}\ S_1^2\ S_2^2, \\
G_l &= 1 - \frac{n_l}{\chi^2_{(n_l,\ 1-\frac{\alpha}{2})}} \qquad (l = 1,\ 2), \\
H_l &= \frac{n_l}{\chi^2_{(n_l,\ \frac{\alpha}{2})}} - 1 \qquad (l = 1,\ 2), \\
G_{12} &= \frac{(F_{(\ n_1,\ n_2;\ \frac{\alpha}{2})} - 1)^2 - G_1^2\ F^2_{(\ n_1,\ n_2;\ \frac{\alpha}{2})} - H_2^2}{F_{(\ n_1,\ n_2;\ \frac{\alpha}{2})}}, \\
H_{12} &= \frac{(1 - F_{(n_1,\ n_2;\ 1-\frac{\alpha}{2})})^2 - H_1^2\ F^2_{(n_1,\ n_2;\ 1-\frac{\alpha}{2})} - G_2^2}{F_{(n_1,\ n_2;\ 1-\frac{\alpha}{2})}}, \\
n_1 &= m - 1, \text{ and} \\
n_2 &= m(n-1).
\end{aligned}$$

Our likelihood-based method is based on $\mathcal{L}^*(\sigma_\tau)$, the profile log-likelihood for σ_τ (rather than for σ). The analogues of expressions (4.4) and (4.5) in this context become

$$-2\ (\ \mathcal{L}^*(\sigma_\tau)\ -\ \mathcal{M}\ ) \dot{\sim} \chi^2_{(1)}, \tag{4.7}$$

and

$$-2\ (\ \mathcal{L}^*(\sigma_\tau)\ -\ \mathcal{M}\ ) \leq \chi^2_{(1,\ 1-\alpha)}. \tag{4.8}$$

4.6.2 Simulations

We conducted simulations of the type described in Section 4.5.2 to compare the traditional and likelihood-based interval estimation methods for σ_τ. The same combinations of m, n, μ, and α used in Section 4.5.2 and new set $\sigma = \{0.5, 0.8, 1, 3, 5, 10\}$ and $\sigma_\tau = \{(0.5, 1.0)[0.1], 3, 5, 7, 10\}$ were employed. As in the estimation of σ, for fixed (m, n), we found little dependence of the results on μ, and therefore only $\mu = 0.5$ is represented in our presentation of the results.

Figures 4.9 through 4.12 are the estimated coverage probabilities for $(m, n) \in \{(2, 2), (3, 3), (4, 2), (5, 3)\}$. As usual, the solid lines identify results for the traditional method (4.6) and dashed lines identify results for the likelihood-based method (4.8).

4.6.3 Improving the Coverage Probability Calibration of Likelihood-Based Method

Figures 4.9 through 4.12 show that the likelihood-based method has unacceptably low estimated coverage probabilities, suggesting the value $\chi^2_{(1,\ 1-\alpha)}$ in inequality (4.8) should be increased to improve these probabilities. We note from Figures 4.9 through 4.12 that the estimated coverage probabilities are particularly deficient when σ_τ is large. Arguing completely heuristically, we reason that for large σ_τ the available information is in some sense "equivalent" to that in a random sample of size m from a $N(0, \sigma_\tau^2)$ distribution. That suggests that once again the values we found to be effective small sample replacements for $\chi^2_{(1,\ 1-\alpha)}$ in the one sample context might be used. After a variety of studies, we have found that it is effective to use $d(m, \alpha)$ from Table 3.1 of [3] in the place of $\chi^2_{(1,\ 1-\alpha)}$ value in inequality (4.8). Figures 4.13 to 4.16 compare the

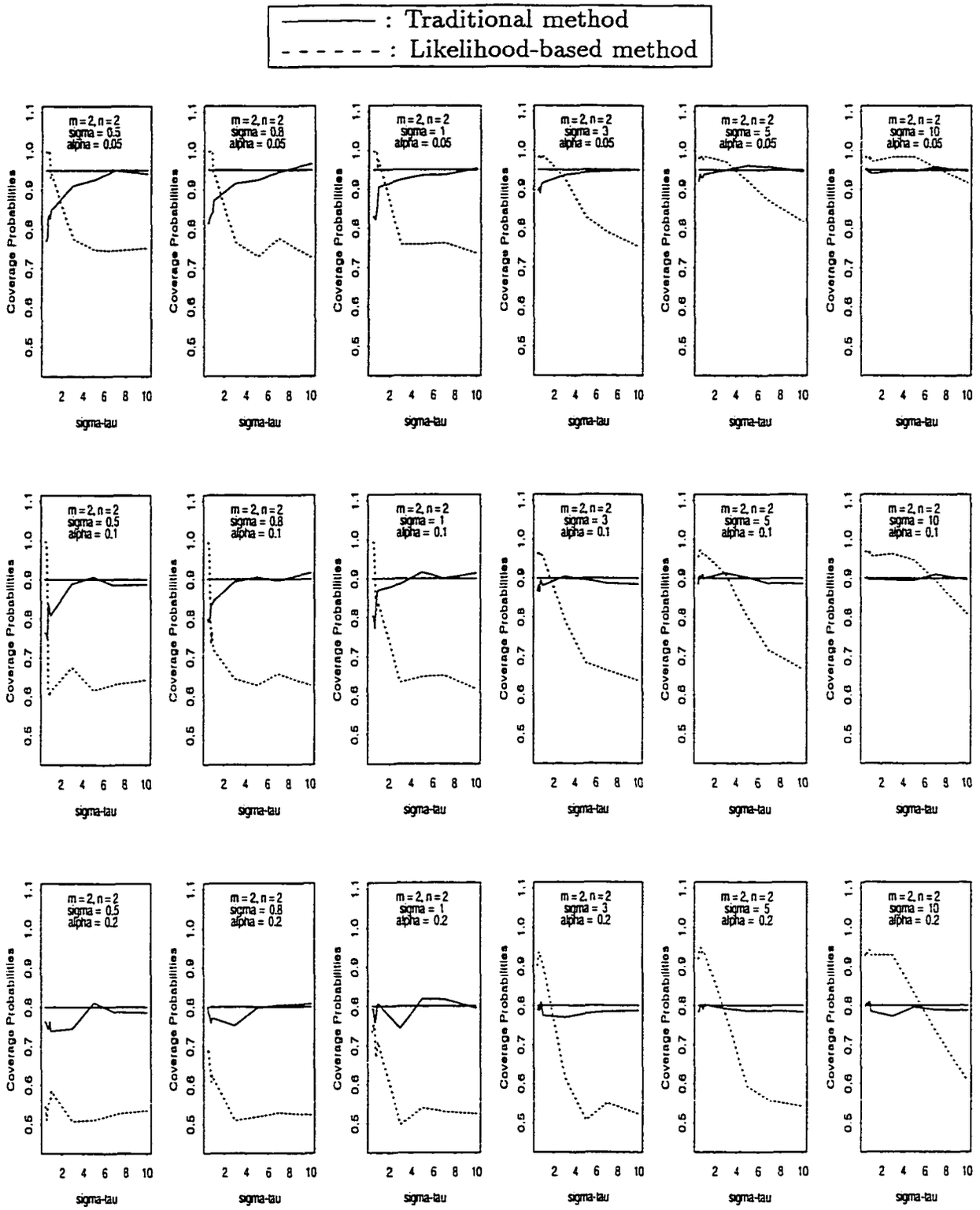

Figure 4.9 Estimated coverage probabilties for σ_τ, $m = 2$ and $n = 2$.

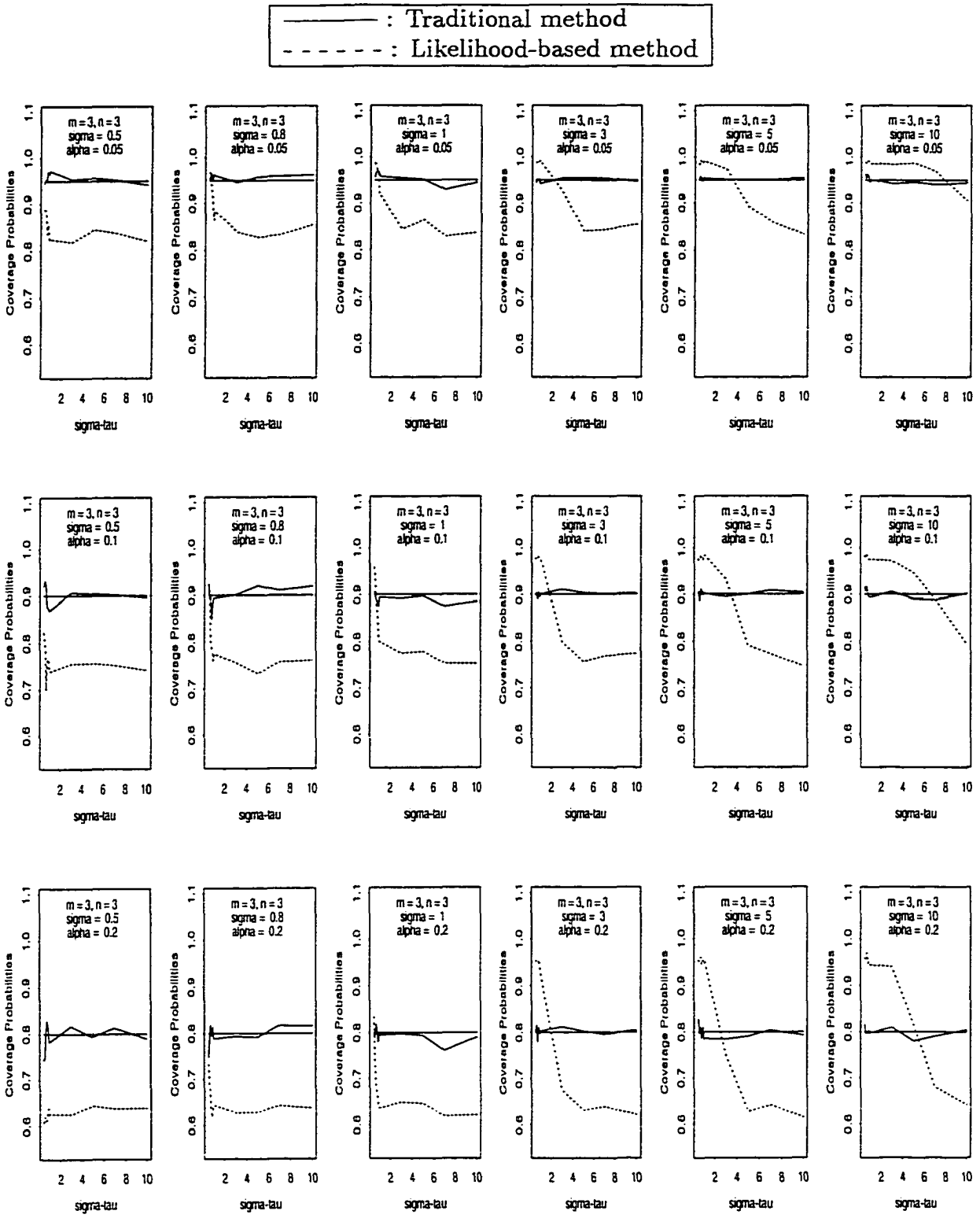

Figure 4.10 Estimated coverage probabilties for σ_τ, $m = 3$ and $n = 3$.

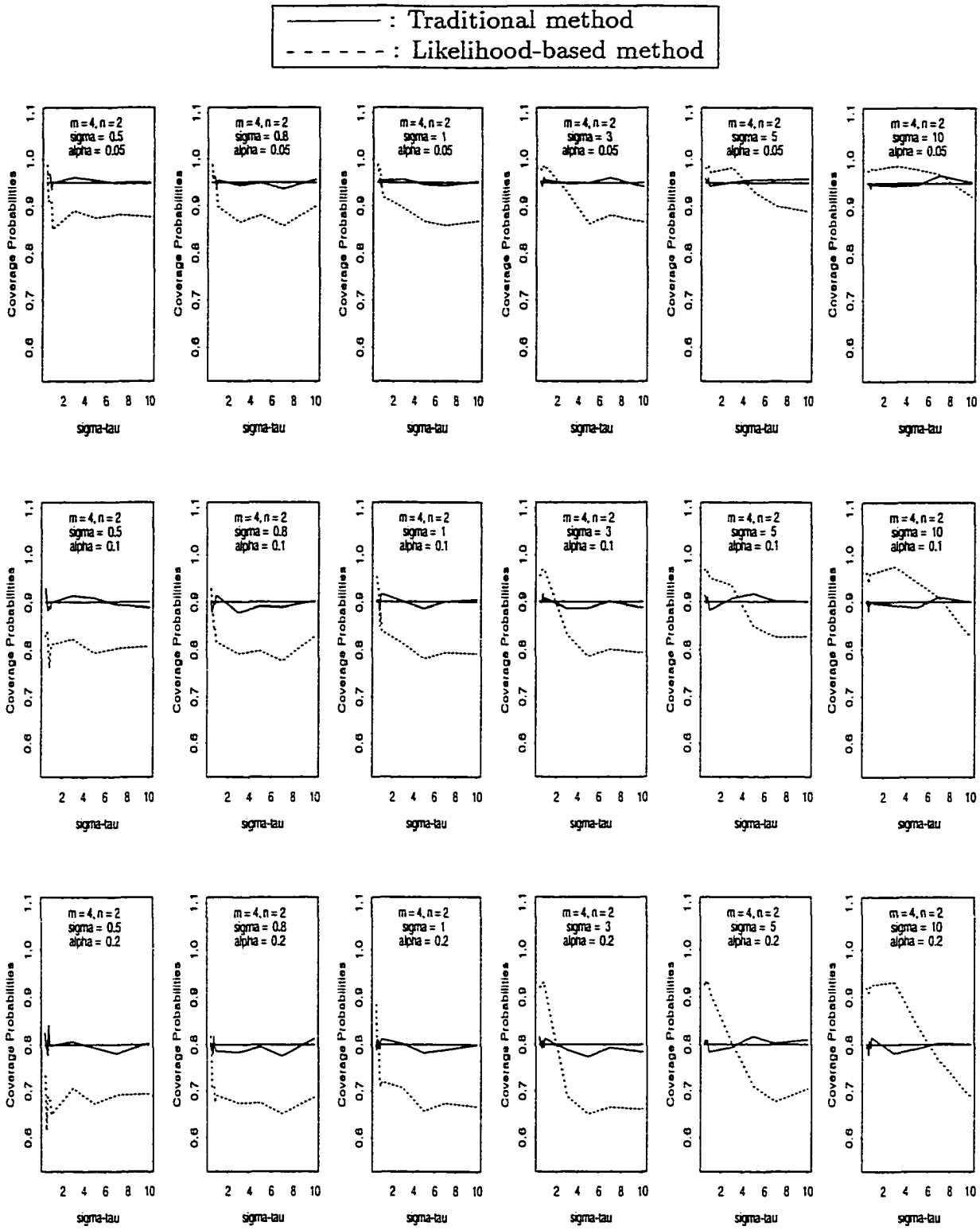

Figure 4.11 Estimated coverage probabilties for σ_τ, $m = 4$ and $n = 2$.

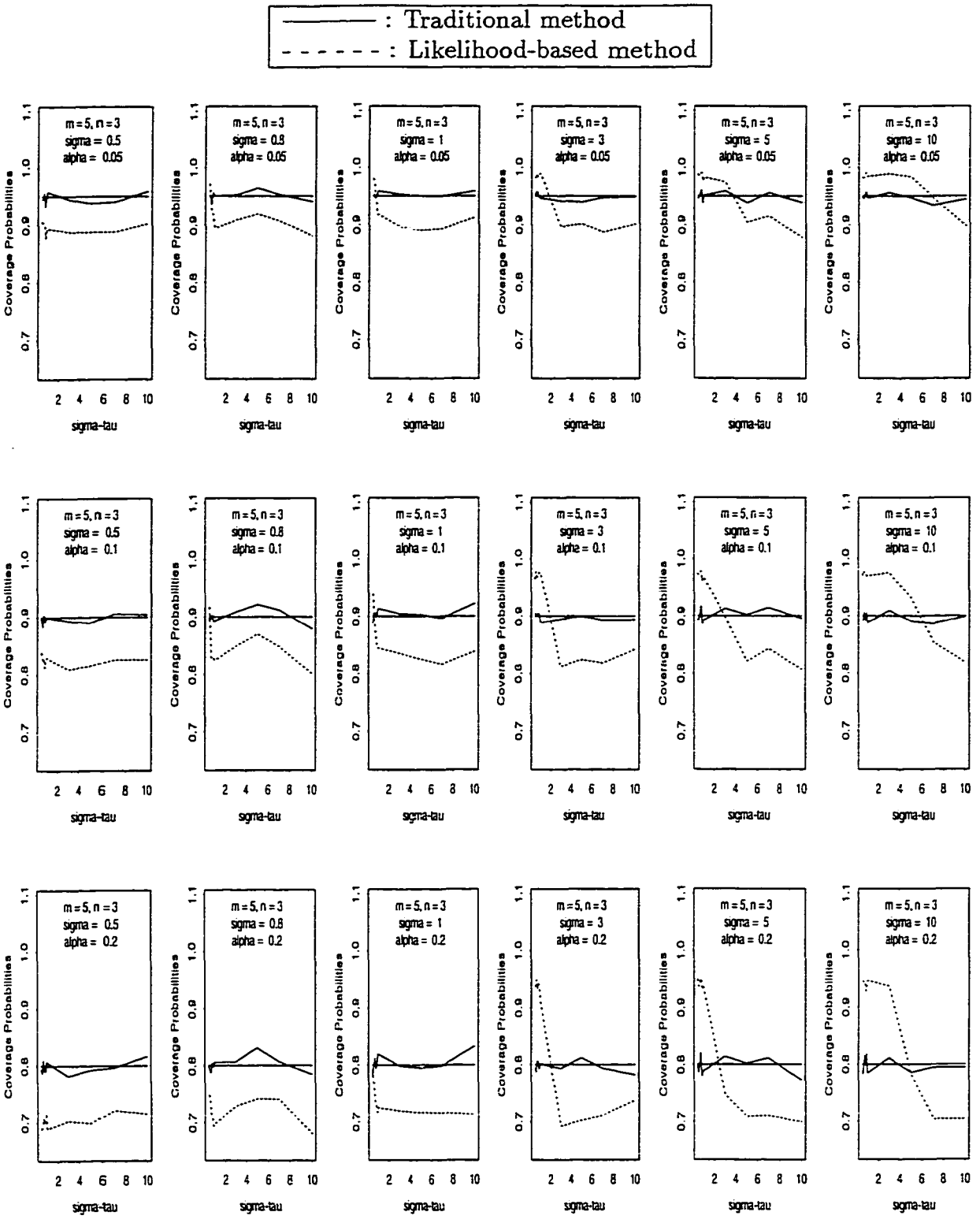

Figure 4.12 Estimated coverage probabilties for σ_τ, $m = 5$ and $n = 3$.

estimated coverage probabilities for the traditional and modified likelihood methods for the same parameter combinations as in Figures 4.9 through 4.12. We can see that coverage probabilities for the likelihood-based method are much improved.

In Tables 4.1 and 4.2, we summarize simulation results of average lengths for the traditional and modified likelihood-based methods when $(m, n) = (2, 2)$ and $(5, 3)$. These results reveal that the traditional method has larger average interval lengths than the modified likelihood-based method, and this situation is pronounced at $(m, n) = (2, 2)$.

4.7 Conclusions

In our simulations we found that it is useful (in computing $\mathcal{M}$ and checking (4.5) and (4.8)) to make use of the cases and analysis of Section 4.5 for special data types. For these data types, the $\mathcal{M}$ can be simply found by reducing the 3-parameter problem to a 2-parameter problem. Further, useful results concerning the shapes of profile likelihoods can be directly applied to these cases from the discussion in [3].

The simulations support the following conclusions:

(1) The analysis on interval estimates of the parameters σ and σ_τ are not much affected by changes in the value of μ.

(2) In the estimation of the parameter σ, no matter what the value of σ_τ is, the traditional method always has coverage probabilities below $(1 - \alpha)$ when σ is small. The modified likelihood-based method (using $d(m(n-1)+1, \alpha)$) doesn't have this problem and performs well even when σ is large.

(3) In the estimation of the parameter σ_τ, it seems that both methods provide good coverage probabilities. But upon closer investigation, the modified likelihood-based method tends to produce somewhat shorter confidence intervals than the traditional method.

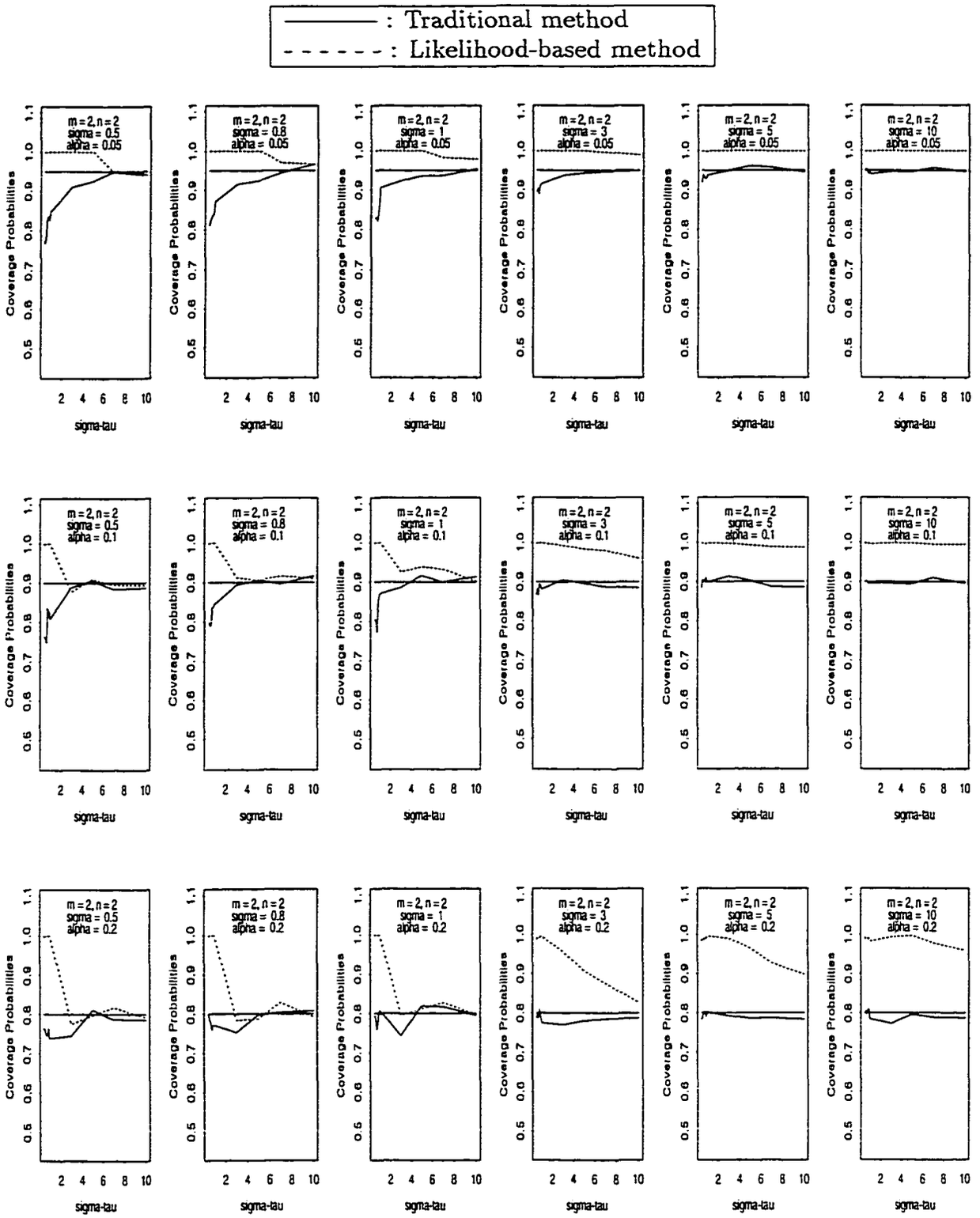

Figure 4.13 Estimated coverage probabilties for σ_τ, $m = 2$ and $n = 2$ (corrected likelihood method).

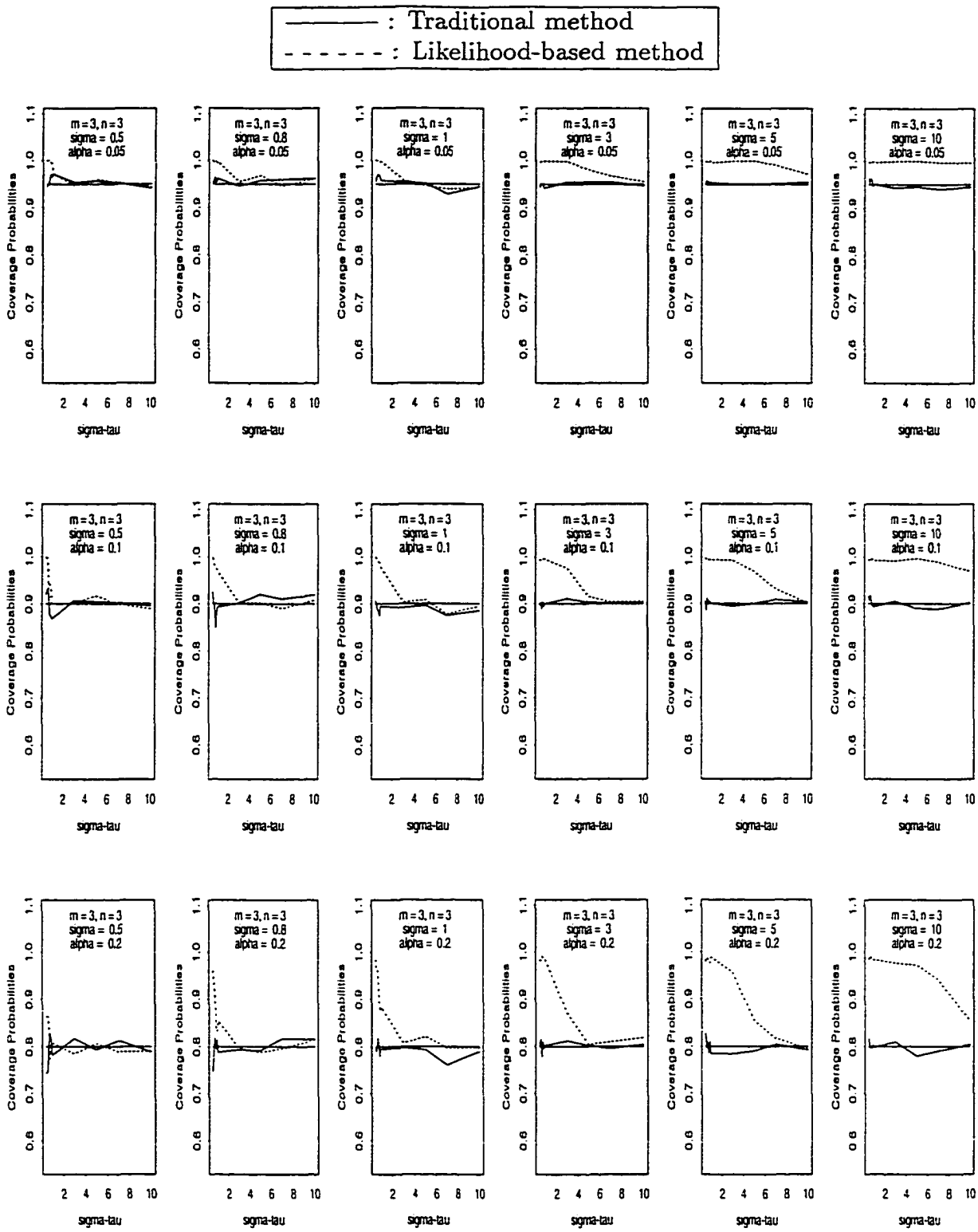

Figure 4.14 Estimated coverage probabilties for σ_τ, $m = 3$ and $n = 3$ (corrected likelihood method).

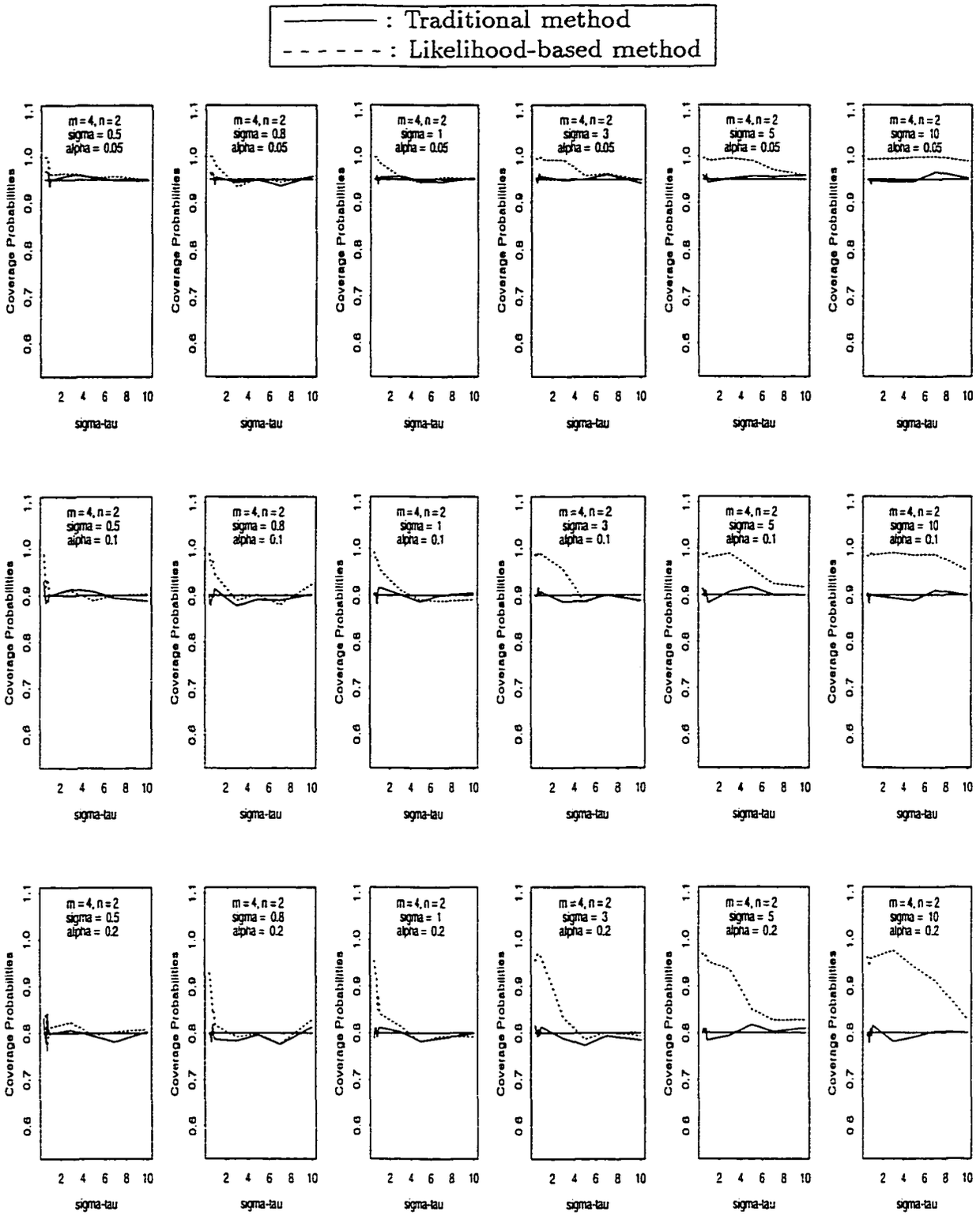

Figure 4.15 Estimated coverage probabilties for σ_τ, $m = 4$ and $n = 2$ (corrected likelihood method).

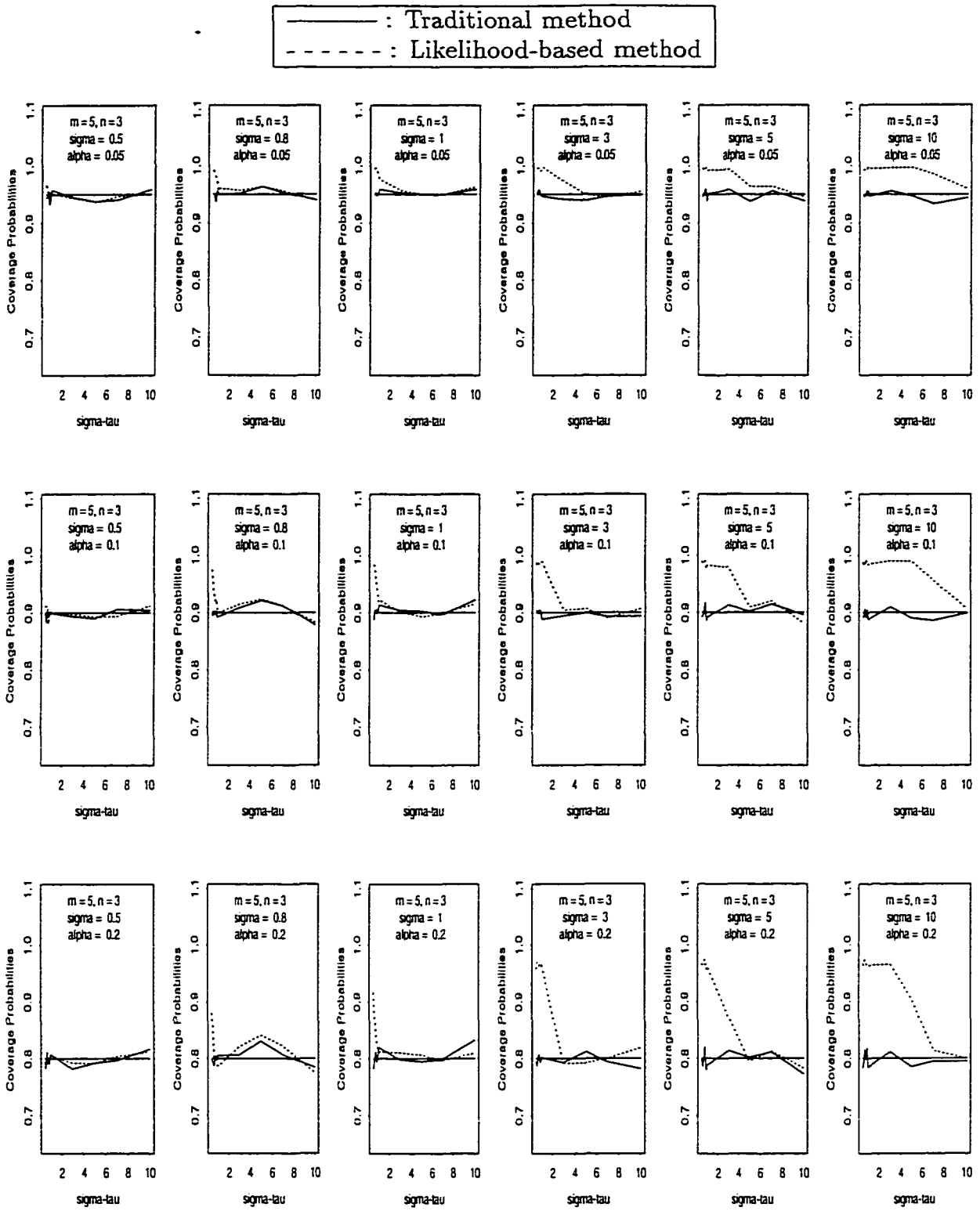

Figure 4.16 Estimated coverage probabilties for σ_τ, $m = 5$ and $n = 3$ (corrected likelihood method).

Table 4.1 Simulated average lengths for traditional method (t) and the modified likelihood-based method (L) for estimating σ_τ, $(m,n) = (2,2)$.

$m = 2,\ n = 2$											
σ		0.5		0.8		1.0		3.0		5.0	
σ_τ	α	t	L	t	L	t	L	t	L	t	L
0.5	0.05	16.022	9.407	19.903	11.886	22.221	13.608	55.149	35.062	92.291	58.877
	0.10	7.963	4.686	9.897	5.920	11.047	6.776	27.340	17.458	45.702	29.314
	0.20	3.889	2.302	4.843	2.902	5.394	3.321	13.207	8.538	22.084	14.320
0.8	0.05	23.470	12.687	25.652	14.404	26.434	15.403	58.434	36.658	94.204	59.701
	0.10	11.639	6.322	12.738	7.180	13.133	7.672	28.993	18.254	46.690	29.718
	0.20	5.647	3.112	6.197	3.531	6.402	3.764	14.034	8.932	22.571	14.527
1.0	0.05	26.471	14.155	28.485	15.632	32.228	17.876	62.958	38.405	93.212	59.890
	0.10	13.100	7.053	14.131	7.793	16.007	8.911	31.247	19.129	46.137	29.816
	0.20	6.327	3.471	6.866	3.838	7.788	4.387	15.110	9.376	22.280	14.586
3.0	0.05	76.961	39.084	78.183	39.776	76.058	39.009	91.948	51.291	114.605	68.163
	0.10	37.794	19.439	38.453	19.825	37.464	19.463	45.573	25.574	56.826	33.959
	0.20	18.022	9.381	18.366	9.644	17.928	9.515	22.031	12.553	27.511	16.648
5.0	0.05	124.222	62.935	123.933	62.981	127.198	64.575	132.516	70.476	151.216	84.532
	0.10	60.944	31.153	60.849	31.294	62.472	32.146	65.552	35.163	74.935	42.137
	0.20	29.021	14.847	29.000	14.989	29.776	15.532	31.503	17.258	36.241	20.683

Table 4.2 Simulated average lengths for traditional method (t) and the modified likelihood-based method (L) for estimating σ_τ, $(m,n) = (5,3)$.

$m = 5,\ n = 3$											
σ		0.5		0.8		1.0		3.0		5.0	
σ_τ	α	t	L	t	L	t	L	t	L	t	L
0.5	0.05	1.475	1.270	1.767	1.530	1.977	1.724	4.518	4.051	7.127	6.524
	0.10	1.143	0.999	1.386	1.212	1.550	1.365	3.535	3.198	5.562	5.152
	0.20	0.831	0.739	1.020	0.912	1.145	1.023	2.601	2.408	4.097	3.872
0.8	0.05	2.014	1.759	2.301	2.004	2.503	2.169	4.810	4.260	7.413	6.640
	0.10	1.535	1.352	1.776	1.568	1.948	1.708	3.774	3.370	5.811	5.249
	0.20	1.095	0.967	1.280	1.149	1.424	1.259	2.792	2.544	4.299	3.960
1.0	0.05	2.384	2.079	2.649	2.314	2.867	2.495	5.027	4.461	7.581	6.809
	0.10	1.812	1.575	2.031	1.776	2.212	1.937	3.943	3.536	5.925	5.379
	0.20	1.287	1.095	1.456	1.261	1.597	1.390	2.915	2.660	4.359	4.048
3.0	0.05	6.537	5.314	6.573	5.520	6.704	5.711	8.418	7.311	10.403	9.006
	0.10	4.954	3.868	4.981	3.952	5.081	4.204	6.498	5.712	8.144	7.151
	0.20	3.515	2.612	3.535	2.616	3.606	2.813	4.695	4.188	5.991	5.376
5.0	0.05	10.646	8.918	10.937	9.213	10.859	9.167	12.091	10.546	14.208	12.354
	0.10	8.066	6.716	8.288	6.949	8.230	6.908	9.182	8.052	10.958	9.660
	0.20	5.722	4.721	5.881	4.894	5.840	4.832	6.520	5.734	7.898	7.076

Appendix

Approximations for $f(\underline{y};\mu,\sigma_\tau,\sigma)$ for the three data types identified in Section 4.4 are discussed in (A), (B), and (C) separately. In these discussions, the approximation

$$1 = \int_{-\infty}^{+\infty} \phi(x;\mu,\sigma)\, dx \quad \doteq \quad \int_{\mu-k\sigma}^{\mu+k\sigma} \phi(x;\mu,\sigma)\, dx, \tag{4.9}$$

often applies, where $\phi(x;\mu,\sigma)$ is the $N(\mu,\sigma^2)$ density and $k \in [3,6]$ is preferred.

(A) Approximation of $f(\underline{y};\mu,\sigma_\tau,\sigma)$ for Case I Samples

For Case I samples, the likelihood function $f(\underline{y};\mu,\sigma_\tau,\sigma)$ is simply

$$\Big(\int_{-\infty}^{+\infty} \Big[\, \Phi\Big(\frac{y+0.5-\mu_i}{\sigma}\Big) - \Phi\Big(\frac{y-0.5-\mu_i}{\sigma}\Big) \,\Big]^n \, \phi(\mu_i;\mu,\sigma_\tau)\, d\mu_i \,\Big)^m,$$

and from (4.9), this is approximately

$$\Big(\int_{\mu-k\sigma_\tau}^{\mu+k\sigma_\tau} \Big[\, \Phi\Big(\frac{y+0.5-\mu_i}{\sigma}\Big) - \Phi\Big(\frac{y-0.5-\mu_i}{\sigma}\Big) \,\Big]^n \, \phi(\mu_i;\mu,\sigma_\tau)\, d\mu_i \,\Big)^m. \tag{4.10}$$

From [3], the value $\Phi\Big(\frac{y+0.5-\mu_i}{\sigma}\Big) - \Phi\Big(\frac{y-0.5-\mu_i}{\sigma}\Big) \approx 1$ if (μ_i,σ) belongs to the triangular region defined by $\mu_i \in (y-0.5,\ y+0.5)$ and $\sigma \in \Big(0, \min\{\frac{y+0.5-\mu_i}{k_1}, \frac{\mu_i+0.5-y}{k_1}\}\Big]$, where $k_1 \in [3,6]$ is selected. So for (μ_i,σ) in the triangular region the value of (4.10) is approximately

$$\Big(\int_{\mu-k\sigma_\tau}^{\mu+k\sigma_\tau} \phi(\mu_i;\mu,\sigma_\tau)\, d\mu_i \,\Big)^m. \tag{4.11}$$

Moreover, the supremum value of (4.11) is 1 and is approximately achieved for (μ,σ_τ) with $(\mu-k\sigma_\tau,\ \mu+k\sigma_\tau) \subset (y-0.5,\ y+0.5)$ or equivalently $\mu \in (y-0.5,\ y+0.5)$ and $\sigma_\tau \in \Big(0,\ \min\{\frac{y+0.5-\mu}{k}, \frac{\mu+0.5-y}{k}\}\Big]$.

To summarize, the supremum value of $f(\underline{y};\mu,\sigma_\tau,\sigma)$ for a Case I sample is 1 and hence $\mathcal{M} = 0$. In addition, points (μ,σ_τ,σ) with $\mu \in (y-0.5,\ y+0.5)$, $\sigma \in$

$(0, \min\{\frac{y+0.5-\mu-k\sigma_\tau}{k_1}, \frac{y+0.5-\mu+k\sigma_\tau}{k_1}, \frac{0.5-y+\mu+k\sigma_\tau}{k_1}, \frac{0.5-y+\mu-k\sigma_\tau}{k_1}\}]$
and $\sigma_\tau \in (0, \min\{\frac{y+0.5-\mu}{k}, \frac{\mu+0.5-y}{k}\}]$ produce $f(\underline{y};\mu,\sigma_\tau,\sigma) \approx \mathcal{M}$.

(B) Approximation of $f(\underline{y};\mu,\sigma_\tau,\sigma)$ for Case II Samples

For Case II samples, the function $f(\underline{y};\mu,\sigma_\tau,\sigma)$ is

$$\prod_{i=1}^{m}\int_{-\infty}^{+\infty} [\ \Phi(\frac{y_i+0.5-\mu_i}{\sigma}) - \Phi(\frac{y_i-0.5-\mu_i}{\sigma})\ ]^n\ \phi(\mu_i;\mu,\sigma_\tau)\ d\mu_i.$$

For any σ, the value $\Phi(\frac{y_i+0.5-\mu_i}{\sigma}) - \Phi(\frac{y_i-0.5-\mu_i}{\sigma})$ approaches 0 when μ_i is outside the interval $[y_i-0.5-k\sigma,\ y_i+0.5+k\sigma]$ with $k \in [3,6]$. So $f(\underline{y};\mu,\sigma_\tau,\sigma)$ can be approximated by

$$\prod_{i=1}^{m}\int_{y_i-0.5-k\sigma}^{y_i+0.5+k\sigma} [\ \Phi(\frac{y_i+0.5-\mu_i}{\sigma}) - \Phi(\frac{y_i-0.5-\mu_i}{\sigma})\ ]^n\ \phi(\mu_i;\mu,\sigma_\tau)\ d\mu_i. \qquad (4.12)$$

Numerically, the σ value which maximizes the function $f(\underline{y};\mu,\sigma_\tau,\sigma)$ in this case is quite small (around 10^{-5} or less). Thus we may approximate the expression (4.12) with

$$\prod_{i=1}^{m}\int_{y_i-0.5}^{y_i+0.5} [\ \Phi(\frac{y_i+0.5-\mu_i}{\sigma}) - \Phi(\frac{y_i-0.5-\mu_i}{\sigma})\ ]^n\ \phi(\mu_i;\mu,\sigma_\tau)\ d\mu_i. \qquad (4.13)$$

Furthermore, for such small σ, the value $\Phi(\frac{y_i+0.5-\mu_i}{\sigma}) - \Phi(\frac{y_i-0.5-\mu_i}{\sigma})$ is nearly 1 for $\mu_i \in [y_i-0.5,\ y_i+0.5]$. The expression (4.13) is then close to

$$\prod_{i=1}^{m}\int_{y_i-0.5}^{y_i+0.5} \phi(\mu_i;\mu,\sigma_\tau)\ d\mu_i = \prod_{i=1}^{m} [\ \Phi(\frac{y_i+0.5-\mu}{\sigma_\tau}) - \Phi(\frac{y_i-0.5-\mu}{\sigma_\tau})\ ]. \qquad (4.14)$$

Therefore, the supremum value of $f(\underline{y};\mu,\sigma_\tau,\sigma)$ for Case II samples can be approximated by computing the supremum value of expression (4.14), which is a likelihood function for a rounded sample of size m from $N(\mu,\sigma_\tau^2)$ based on sample values $y_1,\ y_2,\ \cdots,\ y_m$.

(C) Approximation of $f(\underline{y};\mu,\sigma_\tau,\sigma)$ for Case III Samples

From (4.9), an approximation of the function $f(\underline{y};\mu,\sigma_\tau,\sigma)$ for Case III samples is

$$\prod_{i=1}^{m}\int_{\mu-k\sigma_\tau}^{\mu+k\sigma_\tau}\prod_{j=1}^{n}\left[\,\Phi(\frac{y_{ij}+0.5-\mu_i}{\sigma})-\Phi(\frac{y_{ij}-0.5-\mu_i}{\sigma})\,\right]\phi(\mu_i;\mu,\sigma_\tau)\,d\mu_i. \qquad (4.15)$$

Numerically, we also find that the σ_τ which maximizes $f(\underline{y};\mu,\sigma_\tau,\sigma)$ is a small quantity like 10^{-5} or less. Thus the continuous function $\prod_{j=1}^{n}\left[\,\Phi(\frac{y_{ij}+0.5-\mu_i}{\sigma})-\Phi(\frac{y_{ij}-0.5-\mu_i}{\sigma})\,\right]$ is flat for μ_i in interval $[\mu-k\sigma_\tau,\ \mu+k\sigma_\tau]$. Hence (4.15) is approximately

$$\begin{aligned}&\prod_{i=1}^{m}\prod_{j=1}^{n}\left[\,\Phi(\frac{y_{ij}+0.5-\mu}{\sigma})-\Phi(\frac{y_{ij}-0.5-\mu}{\sigma})\,\right]\int_{\mu-k\sigma_\tau}^{\mu+k\sigma_\tau}\phi(\mu_i;\mu,\sigma_\tau)\,d\mu_i\\ &\doteq\prod_{i=1}^{m}\prod_{j=1}^{n}\left[\,\Phi(\frac{y_{ij}+0.5-\mu}{\sigma})-\Phi(\frac{y_{ij}-0.5-\mu}{\sigma})\,\right]. \qquad (4.16)\end{aligned}$$

To sum up, we can approximate the supremum value of $f(\underline{y};\mu,\sigma_\tau,\sigma)$ in this case by finding the supremum value of (4.16), which is the likelihood function of a rounded sample of size mn from a $N(\mu,\sigma^2)$ distribution based on sample vector $\underline{y}$.

References

[1] Burdick, Richard K. and Graybill, Franklin A. (1992). *Confidence Intervals on Variance Components.* Marcel Dekker Inc., New York, NY.

[2] Lee, Chiang-Sheng and Vardeman, Stephen B. (1999). *Interval Estimators of the Parameter μ for Rounded Normal Data.* Iowa State University, Ames, IA.

[3] Lee, Chiang-Sheng and Vardeman, Stephen B. (2000). *Interval Estimators of the Parameter σ for Rounded Normal Data.* Iowa State University, Ames, IA.

5 CONCLUSION

After the discussions in this dissertation, some general conclusions about the analysis of rounded Normal data can be made.

(1) Changes in the parameter μ will not much affect the properties of confidence intervals for variance parameters σ or σ_τ.

(2) Traditional methods work well only when standard deviations are large.

(3) The (adjusted) likelihood-based methods have better coverage probabilities than the traditional methods when rounding is potentially important.

ACKNOWLEDGEMENTS

I would like to give my thanks to my major professor, Dr. Stephen Vardeman, for his patience in guiding me in every phase of my dissertation.

My wife is the second person I want to thank. Without her encouragement, support, and tolerance, getting my degree would not have gone smoothly. My life would be totally different without her.