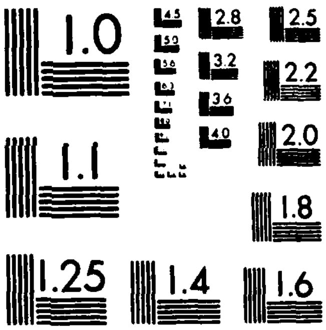
AD-A143 666 ROBUST CONFIDENCE INTERVALS FOR A SCALE PARAMETER: A COMPROMISE BETWEEN T..(U) PRINCETON UNIV NJ DEPT OF STATISTICS S MORGENTHALER SEP 83 TR-255-SER-2 1/1

UNCLASSIFIED ARO-19442.16-MA DAAG29-82-K-0178 F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS 1963-A

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD-A143 666

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARO 19442.16-MA	2. GOVT ACCESSION NO. N/A	3. RECIPIENT'S CATALOG NUMBER N/A
4. TITLE (and Subtitle) Robust confidence intervals for a scale parameter A compromise between the Gaussian and the slash. Technical Report # 255, Series 2.		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Stephan Morgenthaler		8. CONTRACT OR GRANT NUMBER(s) ~~DE-AC02-81ER10841~~ DAAG29-82-K-0178
9. PERFORMING ORGANIZATION NAME AND ADDRESS Statistics Department Princeton University Princeton, N.J. 08544		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE September, 1983
		13. NUMBER OF PAGES 31
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this report we describe a small sample approach parallel to the one for a location parameter (Morgenthaler (1983 1, 1983b). We derive the form of conditional confidence distribution for $\log(\sigma) = \tau$ and then ~~discuss~~ the difficulties in compromising the Gaussian and the slash. ~~We explore~~ both the strong and the bi-optimal procedures. And it becomes clear that there is a difference between inference about location and inference about scale.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

84 07 12 117

DTIC FILE COPY

Robust confidence intervals for a scale parameter:
A compromise between the Gaussian and the slash

by

Stephan Morgenthaler

Technical Report No. 255, Series 2
Department of Statistics
Princeton University
September 1983

Al

*Prepared in connection with research at Princeton University, sponsored by the Army Research Office (Durham). The computing facilities were provided by the Department of Energy, Contract DE-AC02-81ER10841.

September 21, 1983 84 07 12 117

Robust confidence intervals for a scale parameter:
A compromise between the Gaussian and the slash

Stephan Morgenthaler

Technical Report No. 255, Series 2
Department of Statistics
Princeton University
Princeton, New Jersey 08544

ABSTRACT

In this report we describe a small sample approach parallel to the one for a location parameter (Morgenthaler (1983a, 1983b). We derive the form of conditional confidence distribution for $\log(\sigma) = \tau$ and then discuss the difficulties in compromising the Gaussian and the slash. We explore both the strong and the bi-optimal procedures. And it becomes clear that there is a difference between inference about location and inference about scale.

1. Introduction.

Let $y_1 \leq y_2 \leq \dots \leq y_n$ be an ordered sample from the situation $F(\frac{x-\mu}{\sigma})$. In this report we are concerned about inference with regard to the scale parameter σ. From the beginning we will restrict attention to location-invariant and scale-equivariant statistics S, i.e.

$$S(s(t\vec{1} + \vec{x})) = s \quad S(\vec{c}) .$$

remark: On the two-dimensional class of samples $\vec{y}(s,t) = s(t\vec{1} + \vec{c})$ the statistic S is known if only $S(\vec{c})$ is

Prepared in part in connection with research at Princeton University, supported by the Army Research (Durham). The computing facilities were provided by the Department of Energy, Contract DE-AC02-81ER10841.

September 21, 1983

fixed. We call the representing element $\vec{c}$ a <u>configuration</u> (see Morgenthaler (1983a)).

To simplify the situation, we will transform our parameter space by

$$\tau = \log(\sigma) \ .$$

It is well known that this transformation symmetrizes the distributions involved (see e.g. Bartlett and Kendall (1946)). Furthermore it is of mathematical convenience.

For a τ-estimator $T(\ )$ we now require

$$T(s(t\vec{1} + \vec{c})) = T(\vec{c}) + \log(s) \ .$$

This is the starting point for our discussion. In the next chapter we will derive the conditional confidence distribution and examine the resulting strong confidence intervals. The third chapter will be devoted to the study of poly- and bi-optimal confidence interval procedures.

2. Compromising between the Gaussian and the slash: Strong confidence intervals.

2.1. Introduction

Conditioned on any given configuration $\vec{c}$, the distribution of $T(\ )$ is determined by the distribution of $\log(s)$ under the situation F we sample from. The choice $T(\vec{c})$ acts like a location parameter to an otherwise fixed distribution. This implies that the conditional variance is not at all influenced by our choice of $T(\vec{c})$ -- whatever we choose, inside the configuration the variability will be fixed.

For setting confidence limits we are interested in the distribution of log(s) conditioned on the configuration $\vec{c}$ as well.
Let $ds_F(x|\mu,\sigma,\vec{c})$ denote the conditional density of log(s) given the configuration $\vec{c}$ under sampling from $F(\frac{x-\mu}{\sigma})$. Then we have

$$ds_F(x|\mu,\sigma,\vec{c}) = \int_{-\infty}^{\infty} e^x k(e^x,t|\mu,\sigma,\vec{c})\, dt. \tag{2.1}$$

where $k(s,t|\mu,\sigma,\vec{c})$ is the conditional density expressed in terms of the configuration parameters s and t given we are in configuration $\vec{c}$ and the underlying parameter values are μ and σ.

<u>proof</u>:

$$ds_F(x|\mu,\sigma,\vec{c}) = \frac{d}{dx}P[\log(s) \leq x|\mu,\sigma,\vec{c}]$$

$$= \frac{d}{dx}\int_{-\infty}^{\infty}\int_0^{e^x} k(s,t|\mu,\sigma,\vec{c})\, ds\, dt.$$

It then follows that

$$ds_F(x|\mu,\sigma,\vec{c}) = ds_F(x-\tau|0,1,\vec{c}) \tag{2.2}$$

where $\tau = \log(\sigma)$. This is a consequence of a simple change of variables (see Morgenthaler (1983a))

$$ds_F(x|\mu,\sigma,\vec{c}) = \int_{-\infty}^{\infty} e^x k(e^x,t|\mu,\sigma,\vec{c})\, dt$$

$$= \int_{-\infty}^{\infty} e^x k(e^{x-\tau},q|0,1,\vec{c})\frac{dq}{\sigma}$$

$$= ds_F(x-\tau|0,1,\vec{c}).$$

Now we know, what effects changes in the parameter values μ and σ have. The location parameter μ has no effect at all, whereas the scale parameter fixes the location of $ds_F(\ )$, which is otherwise unchanged.

This shows us that the τ-estimation problem is a location-type problem with known scale.

remark: The effects of changing the class-representing configuration $\vec{c}$ are as follows:

$$ds_F(x|0,1,\vec{c} + w\vec{1}) = ds_F(x|0,1,\vec{c})$$

$$ds_F(x|0,1,v\vec{c}) = ds_F(x+\log(v)|0,1,\vec{c})$$

where $w \in \mathbb{R}$ and $v \in \mathbb{R}_+$

2.2. Single situation case: known shape F

If we choose

$$T(\vec{c}) = -ave_F[\log(s)|\mu,\sigma,\vec{c}] = -\tau - ave_F[\log(s)|0,1,\vec{c}]$$

for arbitrary values of μ and σ we will have

$$ave_F[T|\mu^*,\sigma^*,\vec{c}] = ave_F[T(\vec{c})+\log(s)|\mu^*,\sigma^*,\vec{c}] =$$
$$T(\vec{c})+ave_F[\log(s)|0,1,\vec{c}]+\tau^* = \tau^* - \tau$$

where $\tau = \log(\sigma)$ and $\tau^* = \log(\sigma^*)$. Any of these choices of $T(\vec{c})$ leads therefore to estimators whose overall mean is equal to all the conditional means, i.e. it is not functionally dependent on $\vec{c}$. Its variance is therefore the average of the conditional variances which

-- as we have noticed above -- are fixed and can not be influenced by choosing another value for $T(\vec{c})$. This estimate for any choice of σ has therefore the <u>minimal</u> <u>possible</u> <u>variance</u>.

> There is an infinite class of τ - estimators with smallest variance. The difference of two such estimators is constant. On the σ-scale they are multiples of each other, but there the behavior is more complex.

The problem is in one way simpler than the location point-estimation problem, but there is an additional difficulty. We are completely free in choosing the standard form F( ) which is used as a reference to describe the scaling. In this sense the scale parameter σ is a <u>relative</u> <u>parameter</u>, describing the scale relative to a standard form. In the case of the location parameter μ we were able to escape this difficulty by restricting attention to symmetric shapes and choosing the standard form F( ) such that the center of symmetry is at 0.

For the Gaussian situation we could adopt such an escape for the scale parameter too and fix the standard form such that the variance is equal to 1. In this case we have defined a target -- the standard deviation -- for our estimator of σ and it makes sense to ask for the estimator -- now on the τ - scale -- which is unbiased and has smallest variance. In order to be unbiased we need

$$\text{ave}_F[T|0,1,\vec{c}] = \log(1) = 0$$

and hence

$$T(\vec{c}) = -\text{ave}_F[\log(s)|0,1,\vec{c}].$$

Setting confidence limits is straightforward if we have a target in mind. If U( ) is a scale-equivariant upper bound for τ, i.e.

$$U(s(\vec{c}+t\vec{1})) = \log(s) + U(\vec{c}),$$

we are concerned about

$$P[U > \log(\sigma)|\mu,\sigma,\vec{c}] = P[\log(s) + U(\vec{c}) > \tau|\mu,\sigma,\vec{c}] =$$
$$P[\log(s) > \tau-U(\vec{c})|\mu,\sigma,\vec{c}]$$
$$= \int_{\tau-U(\vec{c})}^{\infty} ds_F(x|\mu,\sigma,\vec{c})\,dx = \int_{\tau-U(\vec{c})}^{\infty} ds_F(x-\tau|0,1,\vec{c})\,dx$$

$$= \int_{-U(\vec{c})}^{\infty} ds_F(x|0,1,\vec{c})\,dx = 1 - \int_{-\infty}^{-U(\vec{c})} ds_F(x|0,1,\vec{c})\,dx.$$

There are two natural choices, the <u>balanced</u> and the <u>conditionally shortest</u> choice of upper and lower bound. The <u>length</u> of confidence intervals conditioned on configurations is fixed, since

$$U(\vec{y}) - L(\vec{y}) = \log(s) + U(\vec{c}) - \log(s) - L(\vec{c})$$
$$= U(\vec{c}) - L(\vec{c})$$

if $\vec{y} = s(\vec{c}+t\vec{1})$.

<u>For the balanced confidence interval with conditional confidence level</u> $100(1-\alpha)\%$ we take

$$U(\vec{c}) = -ds_F(\tfrac{\alpha}{2}|0,1,\vec{c})$$
$$L(\vec{c}) = -ds_F(1-\tfrac{\alpha}{2}|0,1,\vec{c}) \qquad (2.3)$$

where $ds_F(\beta|0,1,\vec{c})$ is defined by

$$\int_{-\infty}^{ds_F(\beta|0,1,\vec{c})} ds_F(x|0,1,\vec{c})\ dx = \beta$$

Again there is the problem of specifying a target. We have seen that there is an infinite class of τ - estimatiors with smallest variance. Similarly we can create an infinite class of $100(1-\alpha)\%$ symmetric confidence intervals by moving the one defined in (2.3) by an arbitrary constant. Of course it will then be a $100(1-\alpha)\%$ confidence interval for a different target.

<u>remark</u>: <u>The Gaussian case</u> ($F = \Phi$)

Using (2.1) with the standard Gaussian $\Phi(\ )$ we get

$$ds_\Phi(x|0,1,\vec{c}) \text{ is prop. to } \int_{-\infty}^{\infty} e^x (e^x)^{n-1} \exp(-\frac{e^{2x}}{2}\sum_{i=1}^{n}(t+c_i)^2)\ dt$$

$$\text{i.e. prop. to } e^{nx}\exp(-\frac{e^{2x}}{2}\sum_{i=1}^{n}(c_i-\bar{c})^2)\int_{-\infty}^{\infty}\exp(-\frac{ne^{2x}}{2}(t+\bar{c})^2\ dt.$$

The integral in the last line is proportional to $\frac{1}{e^x}$ and hence

$$ds_\Phi(x|0,1,\vec{c}) \text{ prop. to } e^{(n-1)x}\exp(-\frac{e^{2x}}{2}\sum_{i=1}^{n}(c_i-\bar{c})^2)$$

$$\text{prop. to } (e^{2x})^{\frac{n-1}{2}-1}\exp(-\frac{e^{2x}}{2}\sum_{i=1}^{n}(c_i-\bar{c})^2)e^{2x}.$$

This we recognize as the distribution of a transform of a χ^2_{n-1} random variable.

If X has the density $ds_\Phi(x|0,1,\vec{c})$ then $Y = e^{2X}$ has the density (Jacobian $= \frac{1}{2y}$)

$$t_Y(y|0,1,\vec{c}) \text{ prop. to } y^{\frac{n-1}{2}-1}\exp(-\frac{y}{2}\sum_{i=1}^{n}(c_i-\bar{c})^2)\ y\ \frac{1}{2y}$$

and hence a χ^2_{n-1} scaled by $\frac{1}{\Sigma(c_i-\bar{c})^2}$. This gets us the normalizing constant of $ds_{\Phi}(x|0,1,\vec{c})$ as

$$\frac{1}{\Gamma(\frac{n-1}{2})2^{\frac{n-3}{2}}} \left(\sum_{i=1}^{n}(c_i-\bar{c})^2\right)^{\frac{n-1}{2}}.$$

$ds_{\Phi}(x|0,1,\vec{c})$ has two interesting properties

[1] It contains the configuration only through $S = \Sigma(c_i-\bar{c})^2$.

[2] S only affects the location of $ds_{\Phi}(x|0,1,\vec{c})$.

[2] implies that the single situation confidence intervals in the Gaussian situation will all be of the same length even across configurations. If we sample from the Gaussian, the precision of our knowledge about τ is determined by the sample size and is not dependent on the point pattern of our sample. The interval bounds (2.3) are the usual symmetric χ^2 intervals transformed to the τ - level.

For more general shapes $F(\cdot)$ the above is not true and the length of the single situation confidence intervals will vary from configuration to configuration.

The Gaussian analysis is in fact somewhat naive. The Gaussian confidence intervals are too short for a heavy-tailed situation as for example the slash.

The single situation slash intervals are uniformely, i.e. for all configurations, longer than the Gaussian intervals.

2.3. The two situation case: Gaussian and slash

In order to get a feeling for the problems we face, we intend to study now the slash behavior of the confidence interval for τ based on a "Gaussian analysis".

In order to compute the slash coverage probability we are forced to specify which parameter we want to cover. If we choose our standard in the slash family as

$$f_r(x) = \frac{r}{(2\pi)^{\frac{1}{2}} x^2} \left[1 - \exp\left(- \frac{x^2}{2r^2}\right)\right] \qquad (2.4)$$

values for r around $\frac{1}{2}$ ($f_{\frac{1}{2}}(0) = \frac{1}{(2\pi)^{\frac{1}{2}}}$ as for the standard Gaussian) make sense. From (2.2) we know of course that this implies just a translation of $ds_{slash}(\ )$.

Now we have identified our problem as one of too much freedom. In order to have a compatible meaning of a "scale parameter" in two different location and scale families, i.e. two different shapes, we have to fix the relative scale between the two. More simply put, we have to specify a standard distribution in each family.

remark: There are obviously several ways in which we can do this matching of families (see: Tukey(1980)). If we restrict attention to shapes with finite second moment, one natural choice of the standard form is a member of the family with variance 1. In that case the target of our estimator or confidence interval is the standard deviation.

Another idea is the matching of percentiles -- in the case of the Gaussian and the slash family this leads to smaller values of r if we match further out in the tail (see: Rogers and Tukey(1972)). Finally we need not match at all. We can study estimators like the median absolute deviation MAD and accept whatever "matching" it imposes, i.e. accept whatever it estimates on the population level.

If we try to optimize the slash coverage probability for the Gaussian-balanced τ - intervals, we are lead to values for r around $\frac{1}{4}$, which corresponds to matching the 97.5% - point. The maximal slash coverage of the usual X^2-intervals we can achieve in this way is about 32% for samples of size 20 and 44% for samples of size 10. In all the experimental work we will consider only these two sample sizes and leave sample size 5 aside.

We see from the above numbers how short the Gaussian intervals are from the slash point of view. Furthermore it is clear that $r = \frac{1}{4}$ is a bad choice, since it concentrates on "extreme", slash drawn configurations and tries to make Gaussian estimation compatible to "slash needs". We should rather try to choose r in such a way that the slash estimation is compatible to "Gaussian needs" on "nicely behaved", Gaussian-drawn configurations. In that way we might hope that the <u>slash analysis</u> gives about the right, i.e. compatible answer on Gaussian-drawn samples and can be used to extrapolate in a sensible way to configurations containing outliers, where the Gaussian analysis breaks down quickly.

If we were to allow a <u>conditional choice of r</u> conditioned on each configuration, we would find quite large differences between configurations. If a configuration contains outliers, the value of r such that the families are compatible goes down; in nicely behaved

configurations it is around $\frac{1}{2}$. In point estimation this caus
of problems, since there will be a large part of the varibil
to "conditional bias", which we cannot escape.

2.4. Strong confidence intervals for $\tau = \log(\sigma)$

In this section we want to study the possibilities for intervals which, conditioned on any configuration, reach at 100(1-α)% coverage probability, both for the Gaussian and to slash. For each configuration we get the balanced τ - interv $[L_g, U_g]$ and $[L_s, U_s]$ for the Gaussian and slash situation (se (2.3)). For reasons discussed above, we are free to move all intervals relative to each other by a fixed constant. We wil this by holding the Gaussian intervals fixed and moving the ones. This can be described by choosing a value r in (2.4). Only samples of size 10 and 20 are considered.

It turns out that the slash intervals are longer than t Gaussians in each configuration -- if we were allowed to cho relative scale constant conditioned on the configuration, we always get to a case where the slash interval covers the Ga interval. This is a bit like the confidence intervals for μ samples of size 5, where Student's t interval "dominates" t interval.

A simple strong interval is given by

$$L = \min\{L_g, L_s\}$$
$$U = \max\{U_g, U_s\}$$

But now we have a relative scale constant at our disposal. Table 2.1 contains the fractions of configurations falling into the classes

$$\text{(a)} \qquad L = L_s \quad \text{and} \quad U = U_s$$
$$\text{(b)} \qquad L = L_g \quad \text{and} \quad U = U_s$$
$$\text{(c)} \qquad L = L_s \quad \text{and} \quad U = U_s \ .$$

<u>Table 2.1</u>:
Percentage of cases (b), (c) and (d)

	$\frac{1}{r}$	Gaussian situation (b)	Gaussian situation (c)	Gaussian situation (d)	slash situation (b)	slash situation (c)	slash situation (d)
size=20	2.6	82%	$17\frac{1}{3}$%	$\frac{2}{3}$%	$17\frac{2}{3}$%	0%	$82\frac{2}{3}$%
	2.8	64%	$35\frac{1}{3}$%	$\frac{2}{3}$%	$18\frac{2}{3}$%	2%	$79\frac{1}{3}$%
	3.0	48%	52%	0%	$24\frac{2}{3}$%	$2\frac{2}{3}$%	$72\frac{2}{3}$%
	3.2	$37\frac{1}{3}$%	$62\frac{2}{3}$%	0%	$29\frac{1}{3}$%	$3\frac{1}{3}$%	$67\frac{1}{3}$%
size=10	2.6	$93\frac{1}{3}$%	0%	$6\frac{2}{3}$%	38%	0%	62%
	2.8	$85\frac{1}{3}$%	10%	$4\frac{2}{3}$%	42%	0%	58%
	3.0	78%	$18\frac{2}{3}$%	$3\frac{1}{3}$%	$45\frac{1}{3}$%	$\frac{2}{3}$%	54%
	3.2	$73\frac{1}{3}$%	26%	$\frac{2}{3}$%	$48\frac{2}{3}$%	2%	$49\frac{1}{3}$%

r is as in (2.4)
(b): slash dominates
(c): Gaussian low, slash high
(d): slash low, Gaussian high.

All these percentages are based on 150 sampled configurations.

The two situations behave differently. In slash-drawn configurations the Gaussian interval often supplies the upper bound -- more prominently so for samples of size 20. Of course we expect this behavior which shows how much outliers influence the "Gaussian analysis". In most of the Gaussian-drawn configurations, the slash intervals dominate the Gaussian intervals. We learn that the two situations favor different choices of the relative scale constant $\frac{1}{r}$, low for the Gaussian and high for the slash. Table 2.2 contains expected lengths for the above strong confidence interval procedures.

Table 2.2: estimated expected lengths for strong confidence intervals

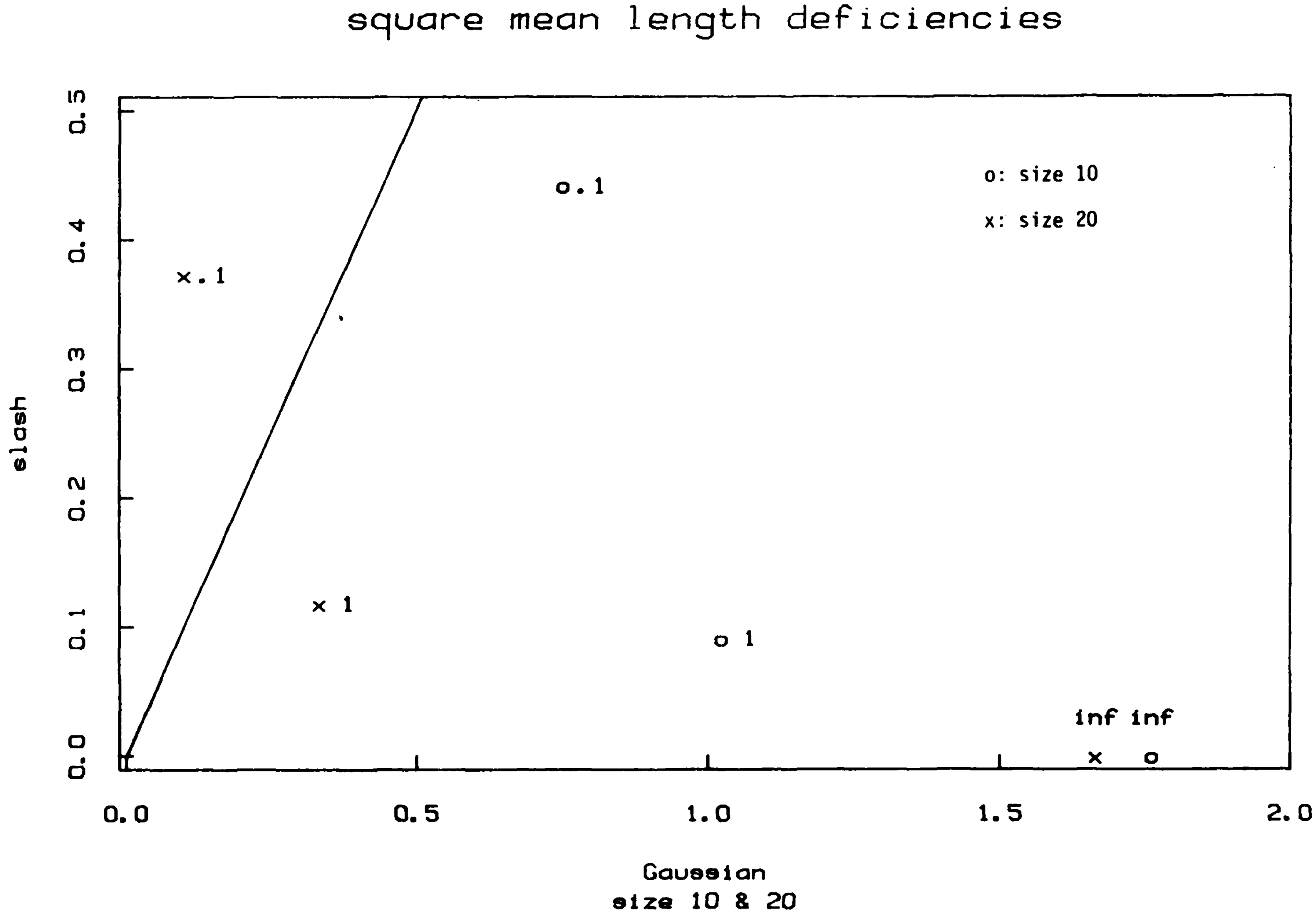
	$\frac{1}{r}$	Gaussian situation	slash situation
	2.6	1.08 (.66)	2.06 (.72)
	2.8	1.10 (.69)	2.00 (.67)
size=20	3.0	1.13 (.74)	1.95 (.63)
	3.2	1.16 (.78)	1.91 (.59)
	single	0.65 (.00)	1.20 (.00)
	2.6	1.63 (.66)	2.41 (.32)
	2.8	1.63 (.66)	2.37 (.30)
size=10	3.0	1.64 (.67)	2.33 (.28)
	3.2	1.65 (.68)	2.29 (.26)
	single	0.98 (.00)	1.82 (.00)

The numbers in parenthesis are $\left(\frac{\text{length}}{\text{single situation shortest}}\right)-1$, i.e. the mean length deficiencies and the row labelled "single" contains the length from the single situation balanced intervals.

Figure 2.1 plots the mean length deficiencies given in Table 2.2. "x" marks the points for $\frac{1}{r}$ = 2.6, 2.8, 3.0 and 3.2 in samples of size 20, "o" in samples of size 10.

$\frac{1}{r}$ = 2.8 seems a good choice for the two sample sizes, in "size=20" it is roughly minimax, in "size=10" it roughly minimizes the Gaussian deficiency (the minimum is rather flat). In Figure 2.1 we see how the strong confidence intervals for τ lose a lot in the

Figure 2.1: Plot of square mean length deficiencies of the bi-optimal intervals for τ



Gaussian situation due to the shortness of the X^2-interval, which is the Gaussian single situation choice. As the sample size decreases, the slash interval more and more dominates the Gaussian one in the slash situation (see Table 2.1). In Figure 2.1 we notice that the strong intervals are really quite good in the slash situation for samples of size 10. The choice $\frac{1}{r} = 2.8$ seems reasonable from what we have just said. In the case of the smaller sample size (10) it minimizes the Gaussian loss, in the case of larger samples (20) it balances the losses in the Gaussian and the slash. In comparison to location-parameter intervals the two situations under consideration exchange places. Now the Gaussian based intervals are optimistically short and the slash ones are long. As the sample size decreases, the slash intervals dominate more prominently.

The relatively big slash loss in samples of size 20 is puzzling. It is due to the fact that the strong intervals described above often are "empty" in the center part for slash-drawn configurations, i.e. the two single situation intervals are separated by a gap = $L_g - U_s$, which has a chance of happening whenever the configuration falls into class (d). For $\frac{1}{r} = 2.8$ such a gap occurs in 42% of the slash-drawn configurations for samples of size 20 and in 19% for samples of size 10. This is a problem which did not occur in the case of confidence intervals for a location parameter. There the strong intervals might have been "overlong" when judged by the slash situation. But here the problem is that neither of the two situations really "needs" these gaps, they are "empty". If we measure the percentage of the total length which is empty, we find that for samples of size 20 as much as 75% of the total conditional length can be made up by empty space and for about 20% of all slash-drawn configurations the percentage of "emptyness" is above $\frac{1}{3}$ of the total length. For samples of size 10 this peculiar problem is not so grave -- about 4% of all slash-drawn configurations are above $\frac{1}{3}$ empty.

The gap problem we have discussed above results from an <u>incompatibility of the meaning of the Gaussian and the slash scale parameters</u> we have chosen. In configurations with outliers, the "Gaussian model" breaks down and it can no longer be connected with the "slash model" in a sensible way. We noticed this in the case of confidence intervals for a location parameter, but it is even more prominent when we discuss the scale parameter.

For the purpose of application, the strong intervals for a scale parameter as given above are not a helpful description of what is going on. We need a definition at the <u>meaning of the scale parameter</u> not guided by one shape (usually the Gaussian) for all configurations, but rather splicing together "meanings" guided by different shapes. In the center section of $d\mu_F(\ )$ (the margingal density across configurations induced by sampling from shape F) the shape F determines the meaning of the scale parameter. Between the shapes there will be a <u>problem or relative scaling</u> similar to the one we have encountered in the case "Gaussian and slash". Solving this solves part of the splicing problem.

From what we have learned about samples of size 20 and 10 we can predict what is happening for size 5. The slash intervals for τ will be much larger than the Gaussian X^2-intervals and it might well be that for $\frac{1}{r}$ around 2.8 the slash intervals in nearly all configurations contain the Gaussian intervals. The strong intervals then would conincide with the slash intervals.

3. Bi-shortest confidence intervals for $\tau = \log(\sigma)$

As we have seen in the previous section the compromise holding the conditional coverage probabilities fixed is not practical. In this section we define intervals for τ which adapt better to the differences in single situation solutions conditioned on the configurations and avoid the empty space we encountered in the previously discussed procedure.

Looking for the bi-shortest interval procedures on the log(σ)-scale leads to a problems similar to the location parameter case as described in Morgenthaler (1983b). The confidence distribution for situation F conditioned on configuration c is

$$Co_F(u) = 1 - \int_{-\infty}^{-u} ds_F(x|0,1,\vec{c})\ dx \tag{3.1}$$

with density

$$co_F(u) = \int_{-\infty}^{-u} ds_F((-u)|0,1,\vec{c})$$

(see (2.3)).

The bi-shortest intervals for τ given the shadow prices p_g and p_s are given by the solution to

$$U_k = h_k^{-1}\left(\frac{p_g w_g^k + p_s w_s^k}{\lambda_g w_g^k + \lambda_s w_s^k}\right)$$

$$L_k = h_k^{-1}\left(\frac{p_g w_g^k + p_s w_s^k}{\lambda_g w_g^k + \lambda_s w_s^k}\right) \tag{3.2}$$

where U_k denotes the largest solution and L_k the smallest and $k=1,\ldots,N$. The Lagrange multipliers λ_g and λ_s are adjusted so that both overall coverage probabilities are at least $100(1-\alpha)\%$. $h_k(\ )$ is the mixture of the conditional confidence densities

$$h_k(\ ) = \frac{\lambda_g w_g^k co_g^k(\ ) + \lambda_s w_s^k co_s^k(\ )}{\lambda_g w_g^k + \lambda_s w_s^k} .$$

The notations and ideas are the same as in Morgenthaler (1983b). Note, however, that $co_g(\ )$ is now the confidence density in the Gaussian situation for the parameter $\tau = \log(\sigma)$.
The solution is simpler than in equations (5.5) where we had to use

$$E_g[s|\vec{c}_k] \text{ and } E_s[s|\vec{c}_k]$$

to adjust for the "scale" differences between configurations. This difficulty disappears in the τ-case since -- as we saw -- we basically deal with a location problem with known scale.

We believe that measuring efficiency by expected length on the logarithmic scale, i.e. after transforming to $\tau = \log(\sigma)$ makes at least some sense. The similar procedures on the original scale, i.e. for σ, are less desirable.

3.1. The slash single-situation confidence interval procedure

We have already pointed out that the Gaussian and slash situation trade places if we move from μ to τ (or σ). And just as Student's t interval was conservative in the slash situation, we have now the slash single-situation interval procedures which are conservative in the Gaussian situation. To keep things simple we will restrict attention to the symmetric slash intervals which have fixed conditional confidence coefficients (note that we have fixed the relative scale between the two families by choosing $\frac{1}{r} = 2.8$ as in the previous sections!). This is not the bi-shortest confidence interval procedure with shadow price ratio $\frac{P_s}{P_g} = \infty$, but is probably not very different from it.

This symmetric interval has a Gaussian coverage probability of 96.2% and 98.6% in samples of size 20 and 10, respectively. The conditional Gaussian confidence levels are most of the time very high and the tail towards low conditional coverages is a lot thicker in samples of size 20. Of course this interval procedure is not balanced if judged from the Gaussian point of view. We can see this in Table 2.1 where the columns headed (c) and (d) show a considerable imbalance. The slash single-situation intervals are frequently too much to the right and hence miss the true τ-value most often by overshooting. The increase in expected length over the symmetric X^2-intervals is considerable. The expected length is increased by about a factor of $1\frac{2}{3}$ for both sample sizes.

Just as Student's t interval should not be applied uncritically, but -- as we have learned -- can be modified successfully, the slash single-situation intervals have undesirable properties. They are -- in Gaussian-drawn configurations -- often too pessimistic and wasteful. Introducing the Gaussian expected length along with the slash expected length hopefully will help us to find procedures which correct this wastefulness. But we must face the need for confidence intervals longer than the common X^2 -based ones. In the next section the slash single-situation interval will be used as a means of comparison to indicate our progress.

3.2. The bi-shortest τ-interval for the shadow price ratio 1

Let us consider the bi-shortest confidence intervals for the shadow prices $p_g = p_s = 1$ (see (3.2)).
Figure 3.1 shows us a plot of the resulting conditional expected lengths vs. the conditional expected lengths of the slash single-situation interval discussed in the previous section. All plots are based on a sample of 150 configurations. The upper half shows the samples of size 20, the lower half the samples of size 10. In both cases we are indeed able to shorten -- and hence "save some information" -- in the Gaussian situation. Note, however, that in samples of size 10 the task seems to be more difficult. Only in configurations where the slash single-situation interval is short are we able to shorten considerably. In samples of size 20, the bi-shortest are quite effectively shortened. In the slash situation we have, of course, to give up something. Most of the bi-shortest intervals are enlarged, thus balancing the configurations where introducing the Gaussian along with the slash leads to "erroneously" short intervals.

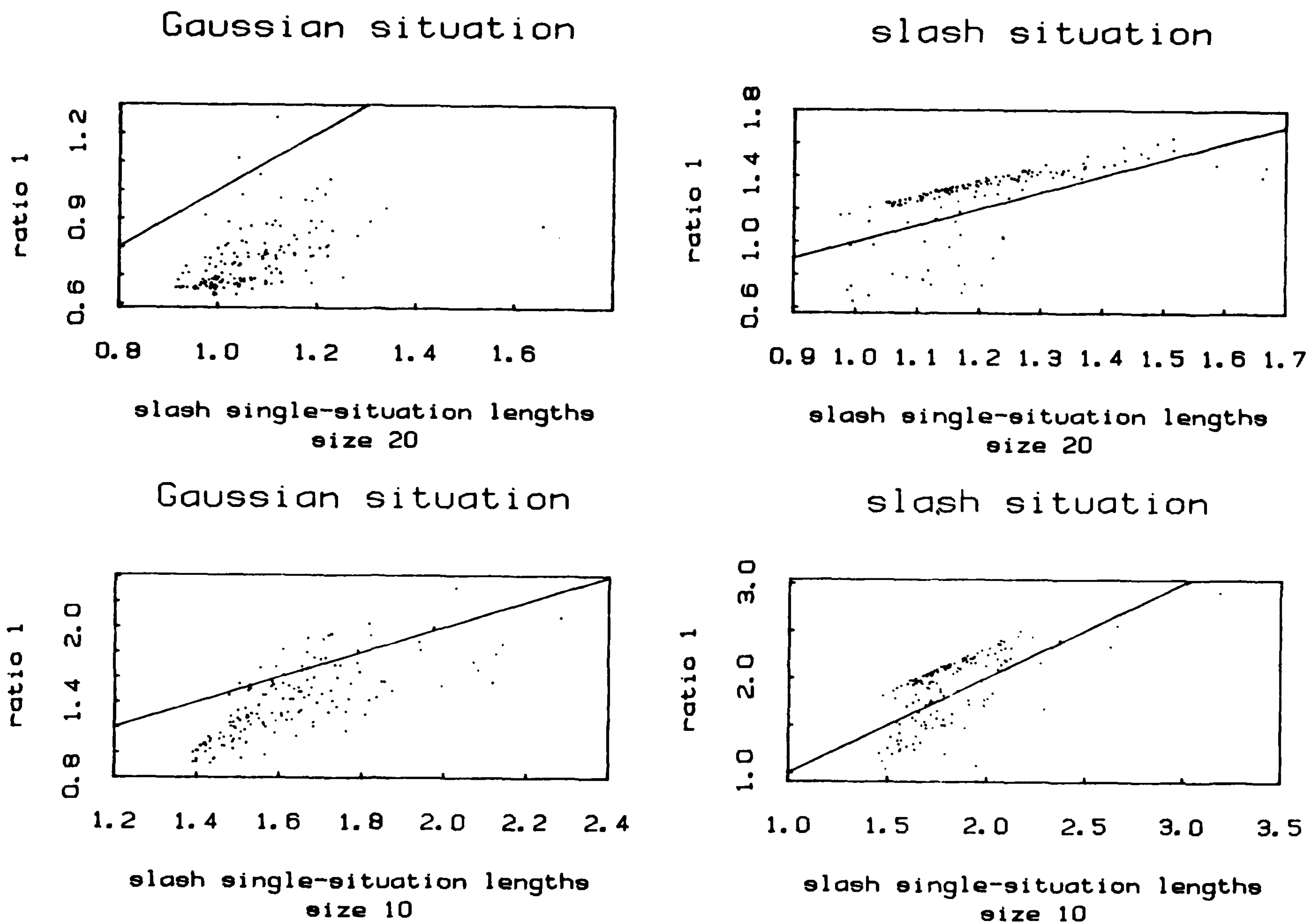
The length of a τ-interval conditioned on the configuration is

fixed, i.e. constant, and is, furthermore, not dependent on the underlying situation. It therefore reflects a property of the configuration, which we can interpret as conditional (degrees of freedom)$^{-\frac{1}{2}}$. The X^2-intervals act as if each configuration had the same number of degrees of freedom. If we introduce the slash situation, we learn that this cannot be tolerated.
In Figure 3.1 we see how the ratio 1 confidence procedure recovers some degrees of freedom in Gaussian-drawn configurations compared to the slash single-situation intervals.

On the average we lose about $\frac{2}{3}$ of the "Gaussian degrees of freedom" for both sample sizes -- a bit less in samples of size 20 -- by going from the X^2-intervals to the slash single-situation intervals. Of course it is true that the slash situation is quite an extreme challenge along with the Gaussian, but degrees of freedom only $\frac{1}{3}$ as large as the usual Gaussian degrees of freedom is not uncommon (see Gosset(1927) especially Table III).

The bi-shortest interval procedure for shadow price ratio 1 recovers most of that loss for Gaussian-drawn configurations in samples of size 20. It leads to a loss of about $\frac{3}{10}$ of the Gaussian degrees of freedom. This recovery, of course, is due partly to the use of a better center for the confidence intervals (see Figure 3.2). In samples of size 10 we still -- even with the bi-shortest intervals -- lose roughly $\frac{1}{2}$ of the Gaussian degrees of freedom in Gaussian-drawn configurations. Again we see that a compromise between the Gaussian and the slash situations is more easily

Figure 3.1: Plot of conditional expected lengths of the slash single-situation intervals vs. the bi-shortest with shadow price ratio 1



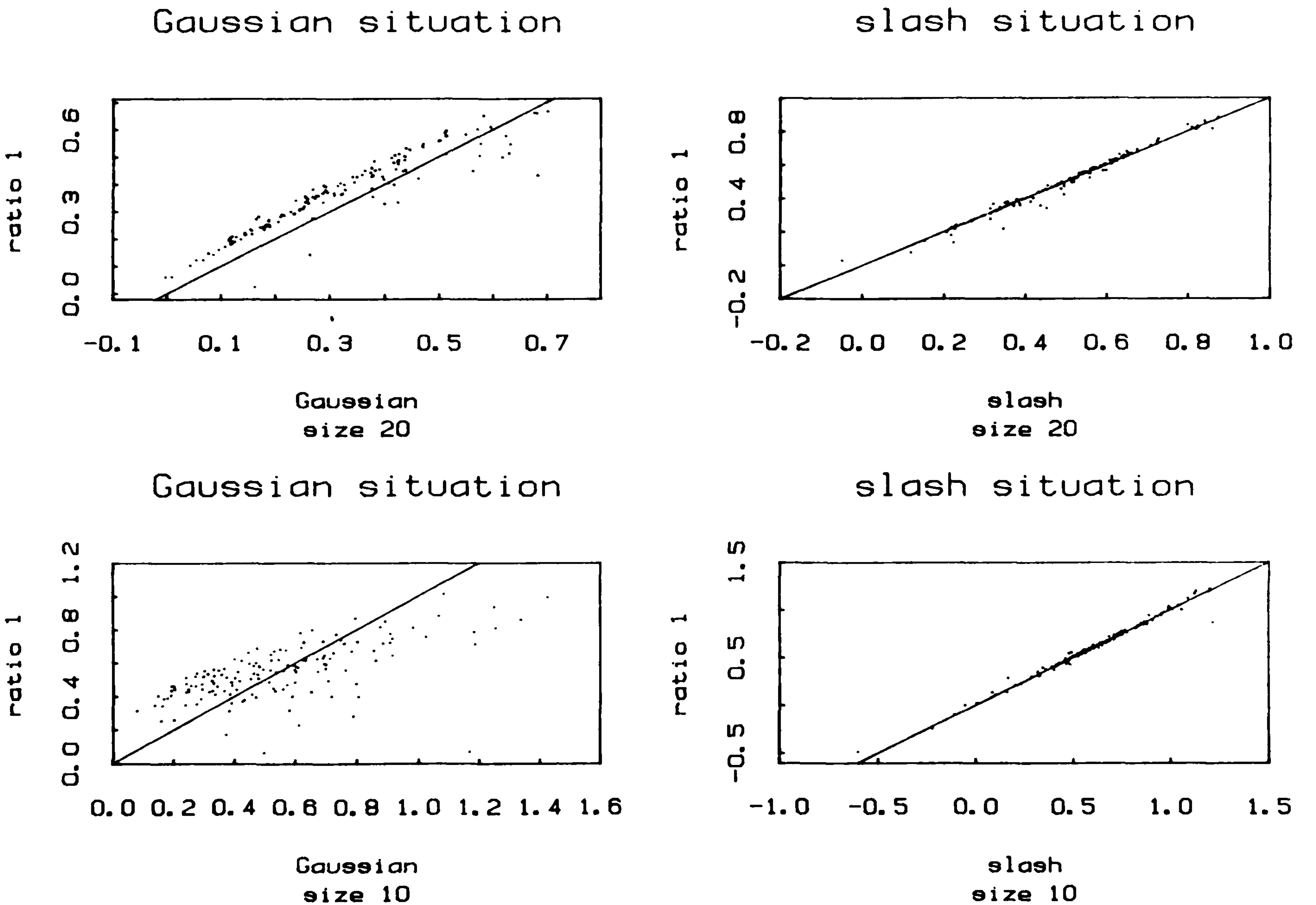
possible in larger samples. In contrast to the location parameter case the sample sizes 10 and 20 are now farther apart in the sense that the "Gaussian loss" is considerable in samples of size 10 no matter what we do.

The above discussion shows us that confidence intervals for τ -- or for σ -- need to be enlarged non-trivially over the "pure Gaussian" ones, if we want to be more realistic about heavy-tailed underlying situations. And while the slash situation is certaily an extreme challenge, the conclusions are by no means unrealistic.

We have already mentioned the importance of the center of our τ-intervals. Figure 3.2 shows us four plots of the bi-shortest interval centers on the configuration-scale. For Gaussian-drawn configurations, the Gaussian single-situation interval centers serve as comparison values. In slash-drawn configurations, the slash symmetric interval centers are used (again, of course, with $\frac{1}{r} = 2.8$ as relative scale between the families). The upper half of the plot shows us what is going on in samples of size 20. The bi-shortest interval with shadow price ratio 1 has a center very nearly the same as the symmetric slash interval in slash-drawn configurations. In the Gaussian situation the bi-shortest interval has a center which is slightly and almost uniformly, i.e. for all of the sampled configurations, moved towards higher values. This again reflects the fact that the choice $\frac{1}{r} = 2.8$ is already too large if judged solely from the Gaussian point of view.

In samples of size 10, the lower half of Figure 3.2 shows us

Figure 3.2 : **Plot of the centers on the configuration-scale of the bi-shortest intervals (ratio 1) vs. the corresponding single-situation intervals**

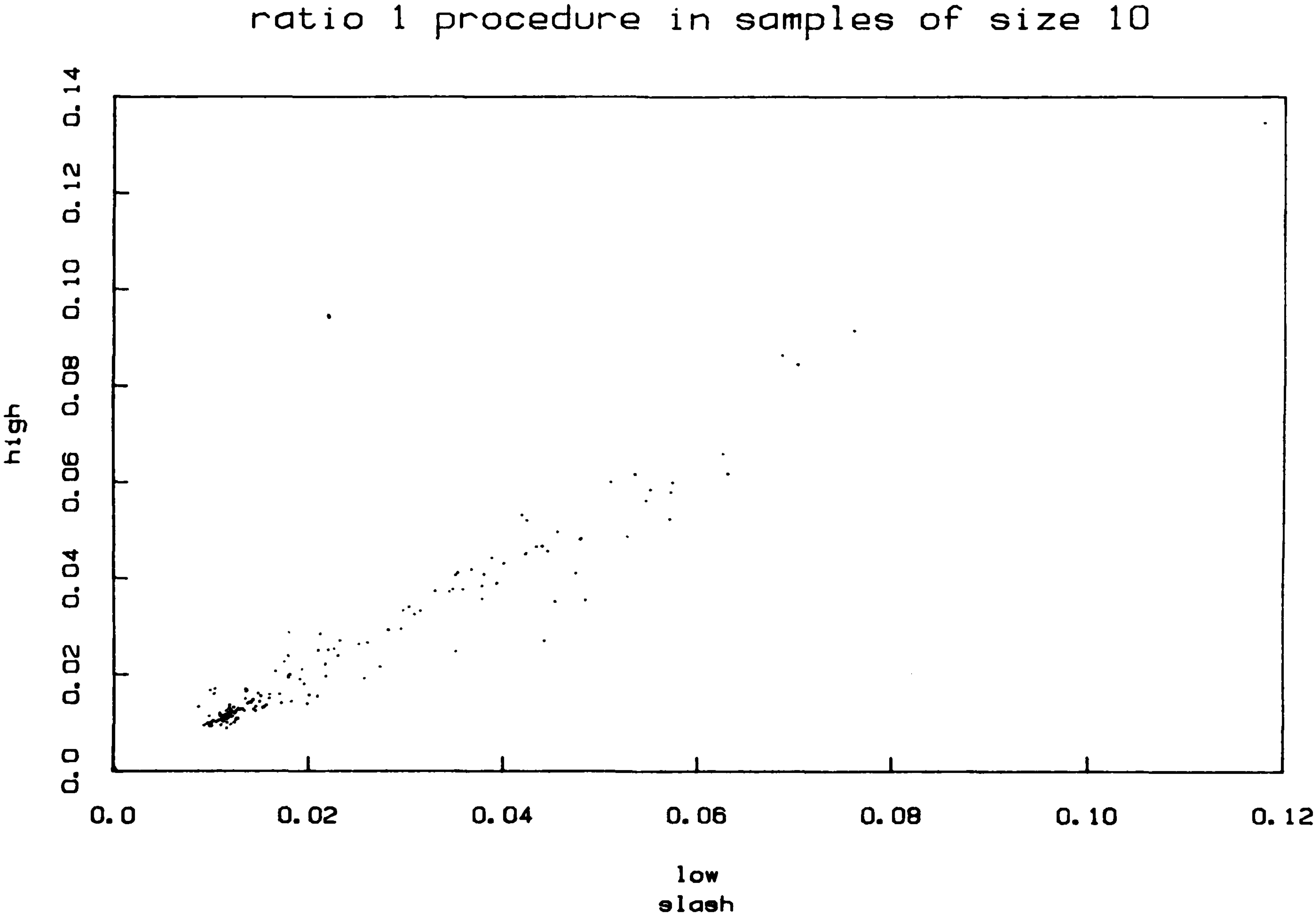


what is going on. Again the slash-drawn configurations b
simple way. Playing the bi-shortest game only affects the
(see Figure 3.1) but not -- or only marginally -- the cen
confidence intervals for τ in slash-drawn configurations.
For Gaussian-drawn configurations the behavior is differe
samples of size 10. Note that we now compare to the Gauss
single-situation interval centers. Clearly for some confi
the bi-shortest center is moved upwards as it was for alm
configurations of size 20. But for a considerable number
Gaussian-drawn configurations the bi-shortest center is l
Such a behavior can be explained by the fact that for siz
distinction between Gaussian-drawn and slash-drawn config
not as clear as it was for size 20.
Note that the interval center on the configuration-scale
the length -- is not the conditional expected center.

Finally, we ought to look at the behavior of the con
confidence levels of our bi-shortest interval procedure.
shows the slash conditional missing-probabilities of the
τ-interval with shadow price ratio 1 in samples of size 1
points in this plot lie around the diagonal. This, of co
follows from what we have seen in Figure 3.2 -- the cente
bi-shortest intervals are near to the centers of the sla
intervals. In samples of size 20, the picture looks some
"neat". There are a few configurations where the bi-shor
interval does not stretch far enough towards high τ-valu

The plot of the missing-probabilities in the Gaussi

Figure 3.3 Plot of the lower vs. the upper conditional missing probabilities of the bi-shortest interval with shadow price ratio 1 in the slash situation



looks according to our expectations. Most configurations are close to the point (0,0), i.e. their conditional coverage is high. From the origin a tail stretches along the x-axis, i.e. if the bi-shortest interval has a low Gaussian conditional coverage probability this is due to the fact that the interval does not stretch far enough towards low τ-values.

3.2.1. Thi bi-optimal curves

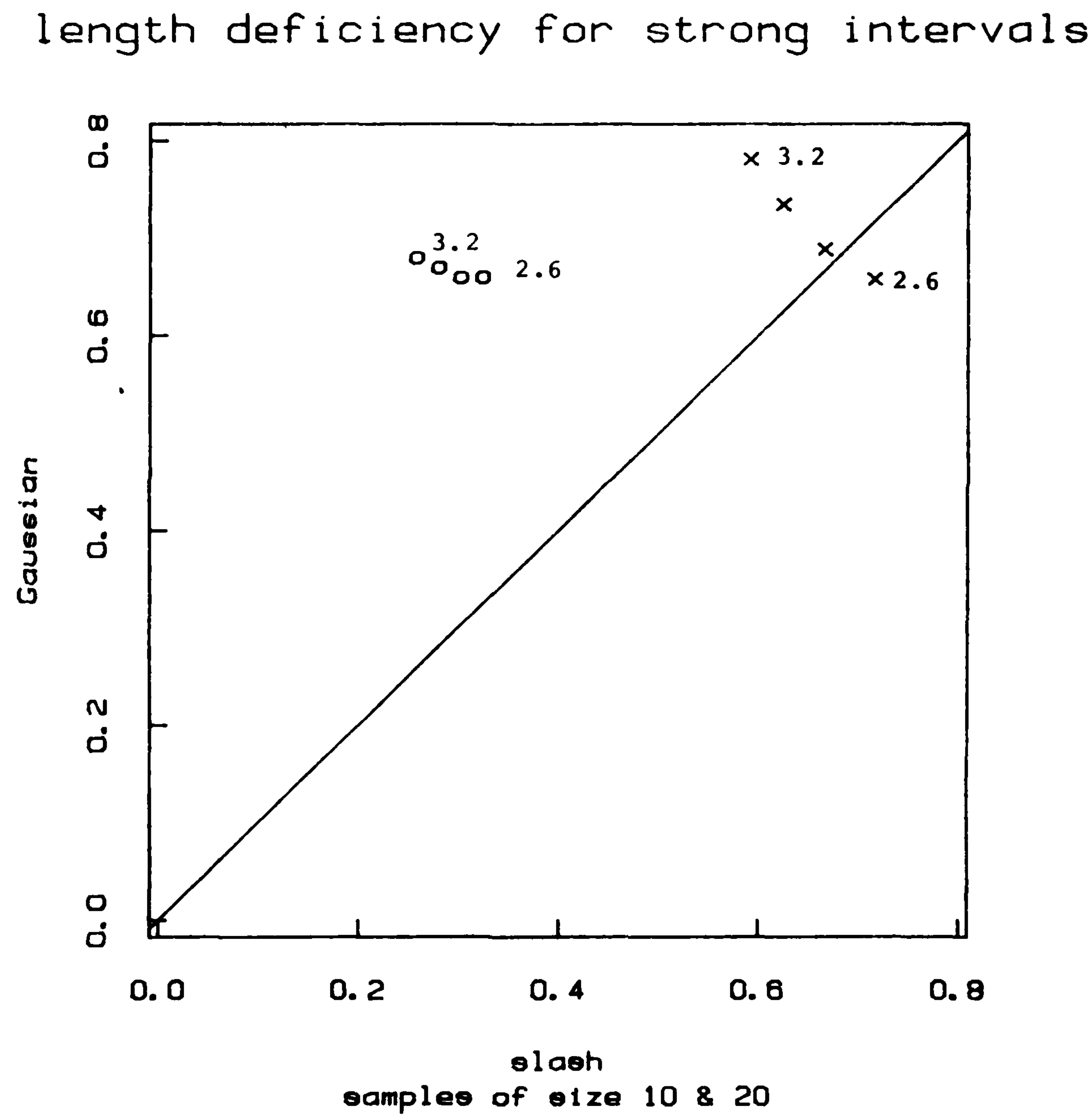
Figure 3.4 shows us the square mean length deficiencies of the bi-shortest confidence intervals for τ in samples of size 10 and 20. These deficiencies are defined by

$$\text{deficiency}_F(I) = \left(\frac{\text{exp. length of I in situation F}}{\text{minimal exp. length in situation F}}\right)^2 - 1,$$

where for the minimal exp. length we use the expected length of the single-situation symmetric intervals for τ. The points for size 20 are marked by "x", the ones for size 10 by "o".
The minimax interval procedure in size 20 has an efficiency of roughly $\frac{1}{1.2}$ or about 83%. In samples of size 10 we still could lower the maximal risk by lowering the shadow price ratio -- the Gaussian risk is dominating the picture for this smaller sample size and the more we can push the Gaussian risk down, the better.

The strong intervals we discussed in section 2 would not fit into the above plot. If we compare Figure 3.4 with corresponding figures in the location parameter case (see Morgenthaler (1983b)), it is obvious that we are no longer able to compromise between our two situations as effectively as in the location parameter case. Figure 3.4 is drawn based on a relative scale of $\frac{1}{r} = 2.8$ between the

Figure 3.4: Plot of deficiencies of strong intervals for τ (o: size 10; x: size 20)



two shapes and we have to keep in mind that for any choice of the relative scale constant there is a different plot.

4. What have we learned about confidence intervals for a scale parameter

It is fair to say that the area of scale estimation has not been explored in a detailed way. This is even more true in the interval-estimation problem. There is hardly any material on robust confidence intervals for a scale parameter in the literature. The present report closes this gap to some extent, but clearly more work is necessary since still more new questions are raised than old questions answered.

Intervals with conditional coverage probability of at least $100(1-\alpha)\%$ in both situations are not realistic because in some configurations the two models are hard to put under one hat and the "strong" intervals are partly "empty".

The bi-shortest interval procedures are a better compromise for the two underlying situations. They define confidence interval procedures which are short in both the Gaussian and the slash situation and reach the 95% confidence level for both situations. They also make us realize that the Gaussian χ^2-interval is too short and has to be enlarged if we want to have a procedure which is safe even in heavy-tailed situations.
For these bi-shortest procedures we need to understand the behavior in other situations better. And it is also not entirely clear which criterion should be adopted for bi-optimization.

It would be of interest to develop other robust confidence intervals for a scale parameter. The only choices available now seem to be based on jackknifing or bootstrapping robust scale statistics like the hinge-spread or the median-absolute-deviation (see Hoaglin et al (1983), chapter 12 for further discussion on robust scale estimators). None of these procedures were tried out in this thesis.

The scale parameter may sometimes be the primary parameter of interest, though most of the time it will be a nuisance parameter. It might well be that the methods developed in this report can help us in setting confidence limits for the location parameter. This idea has not been explored yet.

REFERENCES

Bartlett, M.S. and Kenkall D. G. (1946). "The statistical analysis of variance-heterogeneity and the logarithmic transformation," J. R. Stat. Soc., Suppl. 8, 128-138.

Gosset, W.S. (Student) (1928). "Errors of routine analysis," Biometrika, 19, 151-164.

Hoaglin, D.C., Mosteller, F. and Tukey, J. W. (1983). Understanding Robust and Exploratory Data Analysis, Wiley Publishing Company, New York.

Morgenthaler, S. (1983a). "Strong confidence intervals: A compromise between the Gaussian and the slash," Technical Report No. 253, Series 2, Department of Statistics, Princeton University, New Jersey 08544.

Morgenthaler, S. (1983b). "Bi- and poly- optimal confidence limits for a location parameter," Technical Report No. 254, Series 2, Department of Statistics, Princeton University, Princeton, New Jersey 08544.

Rogers, W.H. and Tukey, J.W. (1972). "Understanding some longtailed symmetrical distributions, Statistica Neerlandica, 26, 211-226.

Tukey, J. W. (1980). Lecture notes to course No. 411, Department of Statistics, Princeton University, Princeton, New Jersey 08544.

September 21, 1983

END

FILMED