MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

②

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARO 19442.14-MA	**2. GOVT ACCESSION NO.** N/A	**3. RECIPIENT'S CATALOG NUMBER** N/A
4. TITLE (and Subtitle) Strong confidence intervals: A compromise between the Gaussian and the slash. Technical Report No. 253, 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~~ DAAG-29-82-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**
CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		**12. REPORT DATE** November, 1983 **13. NUMBER OF PAGES** 53
MONITORING AGENCY NAME & ADDRESS(If different from Controlling Office)		**15. SECURITY CLASS. (of this report)** Unclassified **15a. DECLASSIFICATION/DOWNGRADING SCHEDULE**

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 define strong confidence interval procedures and discuss their properties. Strong confidence means that the reported confidence level is achieved even conditioned on configurations. Furthermore this is true for both the Gaussian and the slash sampling situations. We will show how such a procedure can be obtained and compare its performance to some popular non-parametric confidence intervals.

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE

AD-A142 981

DTIC FILE COPY

DTIC SELECTED JUL 1 3 1984 E

84 07 12 041

Strong confidence intervals: A compromise between
the Gaussian and the slash

by

Stephan Morgenthaler

Technical Report No. 253, Series 2
Department of Statistics
Princeton University
November 1983

November 15, 1983   84 07 12 041

Strong confidence intervals: A compromise between
the Gaussian and the slash

Stephan Morgenthaler
Technical Report No. 253, Series 2
Department of Statistics
Princeton University
Princeton, New Jersey 08544

ABSTRACT

In this report we define strong confidence interval
procedures and discuss their properties. Strong confi-
dence means that the reported confidence level is
achieved even conditioned on configurations. Furthermore
this is true for both the Gaussian and the slash sampling
situations. We will show how such a procedure can be
obtained and compare its performance to some popular
non-parametric confidence intervals.

1. INTRODUCTION

Robust confidence interval estimation is not a heavy researched

area of statistics. Around 1945 the first papers about non-

parametric methods appeared. These procedures provide us with

confidence intervals which reach the reported confidence coefficient

under any -- or any symmetric -- sampling situation. In this sense

they are robust (robustness of validity).

In this report we plan to emphasize (some more) the idea of

having a highly resistant confidence coefficient. In Section 2 we

November 15, 1983

will introduce the formulas and methods for a single, known sampling situation. Section 3, then deals with the case of unknown shape. We are concerned about the dangers of heavy-tailed situations which leads to the choice of the slash-shape as our counterpart to the Gaussian.

2. An equivariant confidence interval procedure for a location parameter.

A lot of work has been done in the area of robust inference about location and scale parameters. Theoretical methods like asymptotic minimax and influence curves handle the case of known scale in a convincing way and the theory is widely accepted. Hand in hand with this development went the recognition of robustness as a research topic in applied statistics. We therefore hardly have to argue in favor of robust procedures.

The need for more work connected with specific sample sizes is less widely acknowledged. The Princeton robustness study (Andrews et al(1972)) is an early example for this kind of research. There several location estimators are tested in a variety of situations and sample sizes.

In the following sections we will discuss an approach which combines the two ideas of robustness and small sample study. The topic is confidence interval estimation, where our knowledge is quite limited.

Let us start our discussion with the classic location and scale problem, where

November 15, 1983

$$\vec{y} = (y_1, y_2, y_3, \ldots, y_n)$$

is a vector of iid random variables drawn from the distribution

$$F(\frac{x - \mu}{\sigma}) \qquad \mu \in \mathbf{R}, \sigma \in \mathbf{R}_+$$

which is fixed and known except for the location μ and the scale σ. Let us assume that the vector $\vec{y}$ is ordered, i.e.

$$y_1 \leq y_2 \leq \cdots \leq y_n .$$

From the beginning we restrict attention to location and scale equivariant procedures. By this we mean the functional property of the procedure T

$$T(s\vec{y} + r\vec{1}) = sT(\vec{y}) + r$$

whenever $s \in \mathbf{R}_+$ and $r \in \mathbf{R}$, i.e. under certain canonical transformations of the sample $\vec{y}$, the value of T transforms canonically too ($\vec{1}$ stands for the vector in $\mathbf{R}^n$ whose components are all 1). Such functions are best studied in reference to location and scale configurations, i.e. the equivalence classes of the equivalence relation defined on $\mathbf{R}^n$

$\vec{x}$ equivalent to $\vec{y}$ iff there is a positive, real s and a real t such that $\vec{y} = s(t\vec{1} + \vec{x})$.

Any location and scale equivariant mapping

$$T : \mathbf{R}^n \longrightarrow \mathbf{R}$$

will be completely specified in any configuration by fixing its value

$T(\vec{c}) \in \mathbb{R}$ for a class-representing element $\vec{c}$, since by the defining relation for equivariance

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

When doing numerical integrations (with respect to conditional distributions given the configuration) it is advisable to choose the class-representing element such that

$$c_a = -1$$

$$c_b = +1$$

and (since we have ordered from the beginning) $c_1 \leq c_2 \leq \ldots \leq c_n$. A good choice seems to be index a around $\frac{n}{4}$ and index b around $\frac{3n}{4}$. This means that we choose as our element representing the configuration, that sample which satisfies the two above restrictions. Of course this element will be uniquely defined.

Another way of thinking about the class-representing elements is to say that they are base points for parametrizing configurations, which are two-dimensional classes of samples.

Any sample $\vec{y} = (y_1, y_2, \ldots, y_n)$ determines uniquely its configuration (we use the word configuration also to denote the class-representing elements) by

$$c_k = \frac{y_k}{s_y} - t_y \quad ; \quad k = 1, \ldots, n$$

where $s_y = \frac{y_b - y_a}{2}$ and $t_y = \frac{y_b + y_a}{y_b - y_a}$. And so we can parametrize $\mathbb{R}^n$ by

$$(t,s,c_1,c_2,\ldots,c_n) \qquad t\in R, s\in R_+$$

Remember that $(c_1,c_2,\ldots,c_n) = \vec{c}$ is $(n-2)$- dimensional since two values are pre-fixed.

It we have a sample $\vec{y}$, we know the configuration $\vec{c}$ and it can be argued that this contains all of the "allowed" information -- we should not be influenced by the actual values of s_y and t_y. The "point pattern" of the sample by itself should fix our ideas about location and scale. J.W. Tukey's picture for this is a code clerk who pre-processes our sample $\vec{y}$ and only gives to us the configuration $\vec{c}$ onto which we have to base our inference ("inference on the $\vec{c}$-scale"). After handing him our estimated values back, the code clerk can use his knowledge of s_y and t_y to transform the answer back to the $\vec{y}$-scale, he simply has to add t_y back in and multiply by s_y.

The configurations partition the sample space R^n into 2-dimensional subsets -- parametrized by $t\in R$ and $s\in R_+$ -- which are the largest sets where a general location and scale equivariant mapping behaves perfectly simply and can be defined by a single real number.

But configurations are also connected with the probability structure on the sample space induced by iid sampling from $F(\frac{x-\mu}{\sigma})$ $\mu\in R$, $\sigma\in R_+$. The idea works because of the equivariance properties of the probability structure, i.e. the canonical changes which take place inside configurations if the parameter values are changed.

Let $k_F(t,s|\mu,\sigma,\vec{c})$ denote the conditional density given the configuration $\vec{c}$. We have to find the Jacobian of the transformation

$$y_i = s(t + c_i) \qquad i = 1,\ldots,n.$$

It we take the coordinates of y in the order $(y_a, y_b, y_1, \ldots, y_n)$ the Jacobian is equal to:

$$\text{abs}(\det(\begin{array}{cccc} s & t-1 & . & \\ s & t+1 & . & \\ s & . & sI_{n-2} \end{array})) = 2s^{n-1}.$$

Hence we have

$k_F(s,t|\mu,\sigma,\vec{c})\, dsdt$ is proportional to

$$s^{n-1} \prod_{i=1}^{n} \{t(\frac{(s(t+c_i)-\mu)}{\sigma})\} \frac{1}{\sigma^n} \, ds\, dt,$$

i.e. proportional to $(\frac{s}{\sigma})^{n-1} \prod_{i=1}^{n} \{t(\frac{s}{\sigma}(\frac{st-\mu}{s} + c_i))\}\frac{ds}{\sigma}dt,$

where t is the density function corresponding to F.

So $k_F(s,t|\mu,\sigma,\vec{c})\, ds\, dt = k_F(p,q|0,1,\vec{c})\, dp\, dq$ it we put

$$p = \frac{s}{\sigma}$$

$$q = \frac{st-\mu}{s} = t - \frac{\mu}{s} \qquad\qquad (2.1)$$

or

$$s = p\sigma$$

$$t = q + \frac{\mu}{p\sigma}$$

We learn from this that the optimal choice of location and scale equivariant procedures is not dependent on the true parameter values μ and σ. By coding the data in terms of a contiguration, we have got

rid of the "true" parameter values, since any choice we make on the $\vec{c}$-level has to be used for all possible "true" parameter values.

Figures 2.1 through 2.4 show a series of contourplots in the (t,s)-plane of the densities $k(s,t|0,1,\vec{c})$ for two configurations in twenty dimensions. Fig. 2.1 and 2.2 show the Gaussian and slash (see 3.1) in a Gaussian- drawn, nicely behaved configuration. Fig. 2.3 and 2.4 do the same with a slash-drawn configuration. The configuration is included in these pictures with "*"'s, the scale on the t-axis is, however, not relevant. We can see how the "outliers" in the second case influence the Gaussian density, which is pushed to small s-values and very stretched in the t-direction. The slash picture does not exhibit such drastic changes. Of special interest is the difference between the two densities in any given, fixed configuration, because a big such difference tells us, that the two sampling situations "interpret" the data quite differently and lead us to quite different conclusions. The configurations used for these plots are :

For the Gaussian-drawn $\vec{c}$ = (-1.4,-1.3,-1.2,-1.1,-1.0,-.91, -.76,-.63,-.30,-.14,-.03,.23,.43,.64,.89,1.0,1.5,1.5,1.6, 3.2)

For the slash-drawn $\vec{c}$ = (-13.8,-6.7,-1.5,-1.1,-1.0,-.90,-.65, -.51,-.39,-.09,-.01,.32,.70,.78,.91,1.0,1.2,1.3,2.8,9.5).

example:

If we want to find the location estimate T with minimal conditional mean square error, we have to choose $T(\vec{c})$ -- i.e. the

Figure 2.1: Contour plot of $k(s,t|0,1,\vec{c})$, the configuration is indicated by "*"'s

Gaussian cond. density in configuration of size 20

Figure 2.2: Contour plot of $k(t,s|0,1,\hat{c})$, the configuration is indicated by "*"'s

slash cond. density in configuration of size 20

Gaussian-drawn configuration

Figure 2.3: Contour plot of $k(t,s|0,1,\vec{c})$, the configuration is indicated by "*"'s
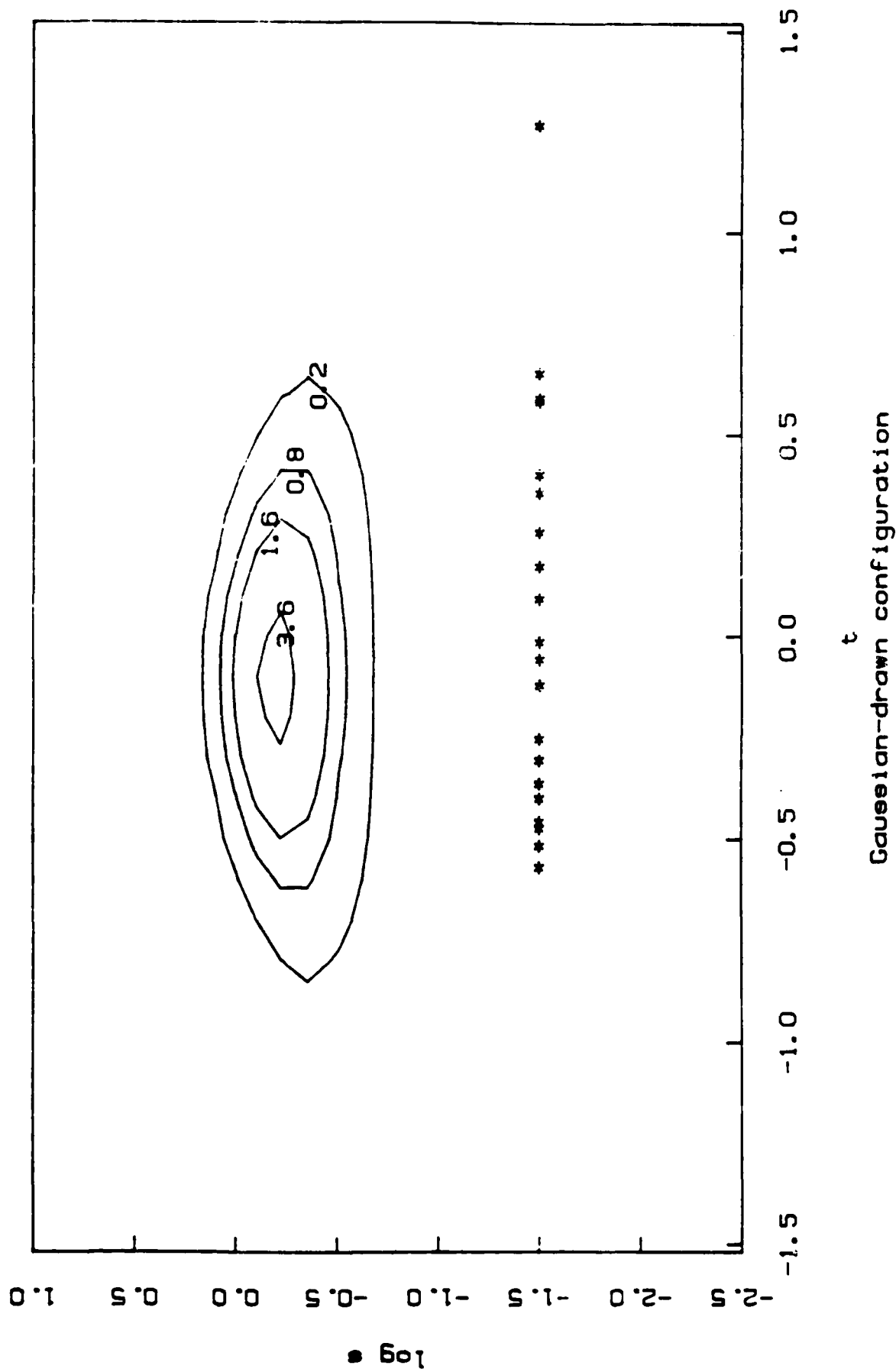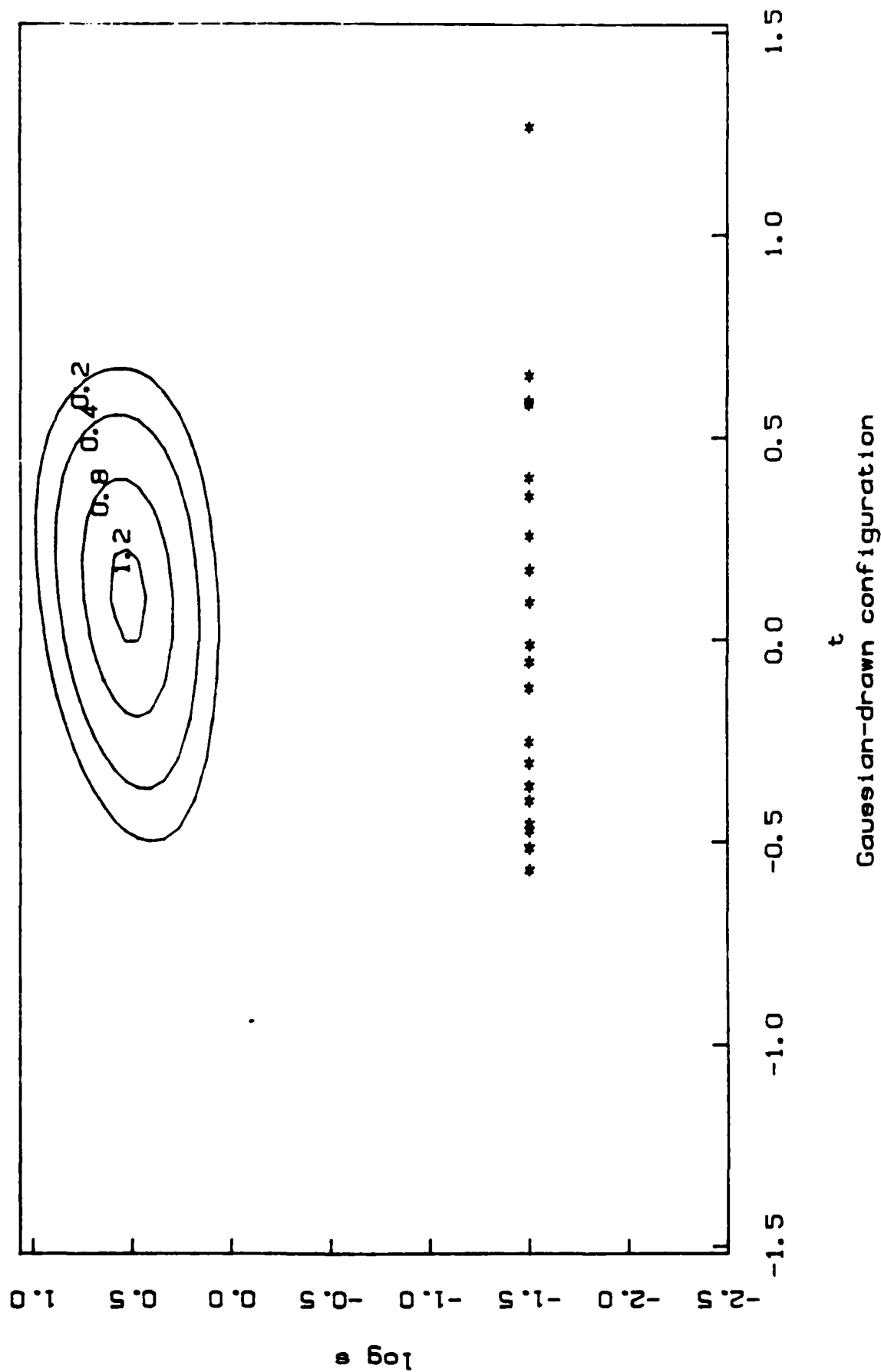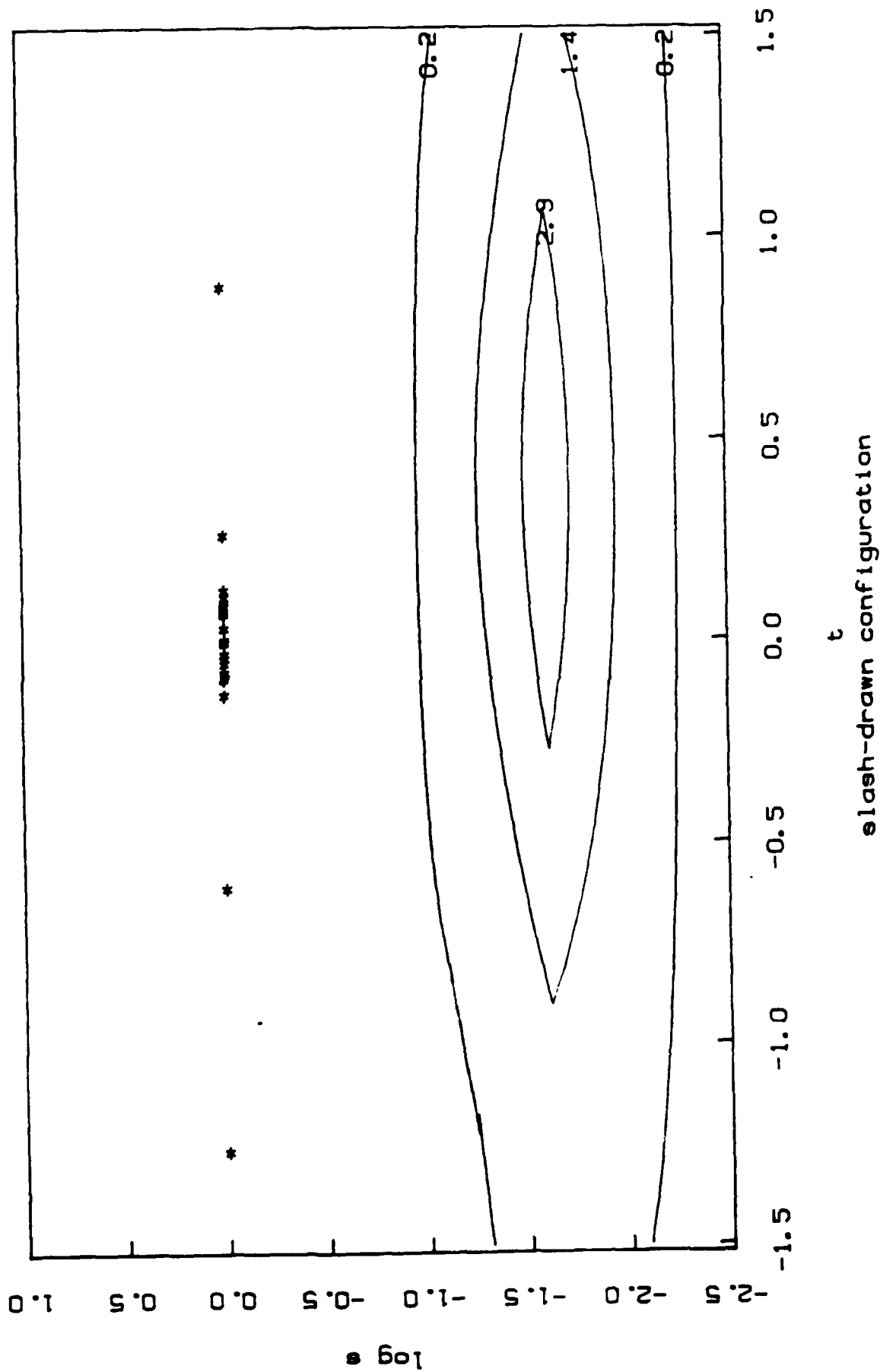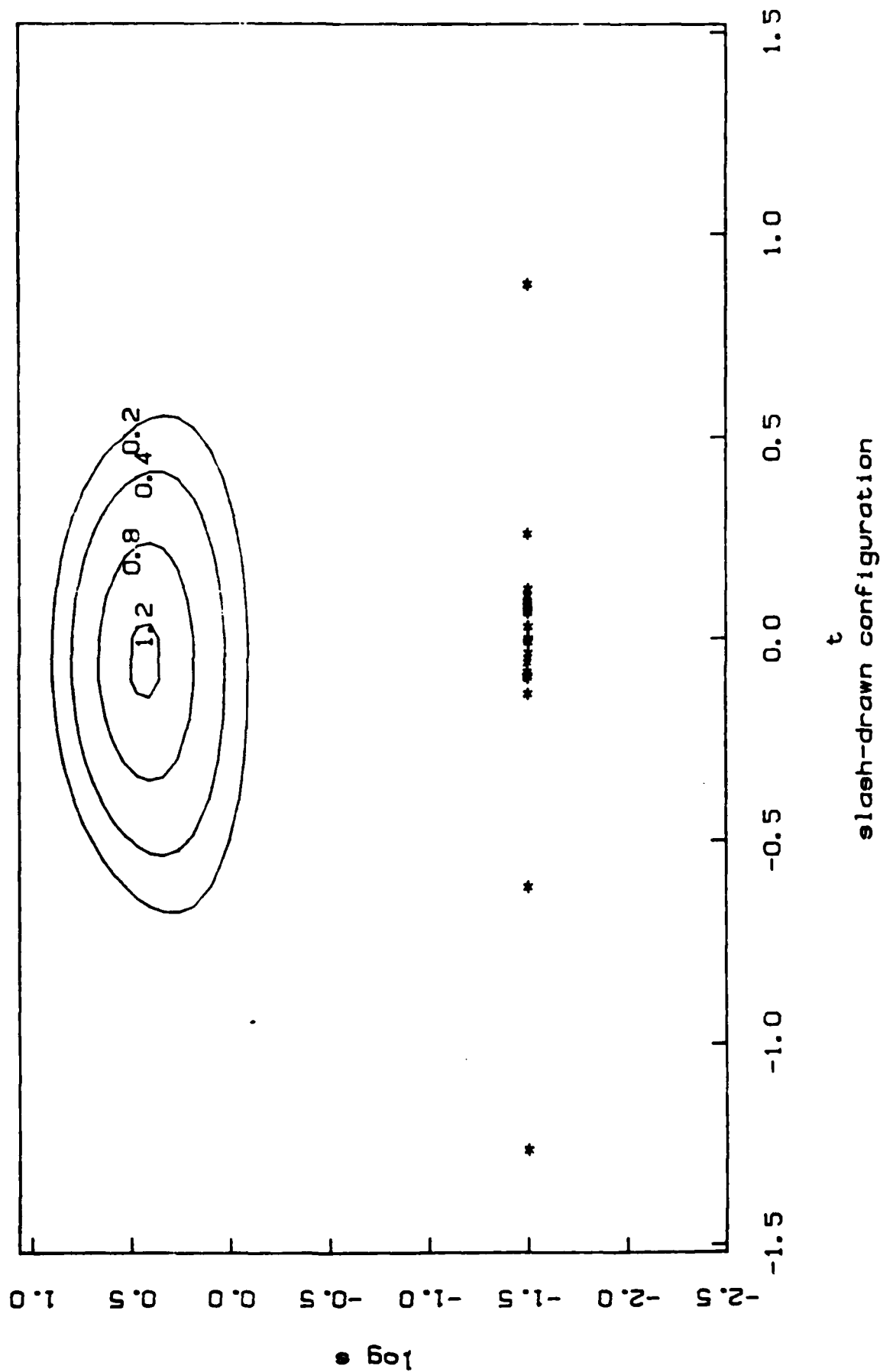
Gaussian cond. density in configuration of size 20

Figure 2.4: Contour plot of $k(t,s|0,1,\vec{c})$, the configuration is indicated by "*"'s

slash cond. density in configuration of size 20

slash-drawn configuration

t

log s

estimated value on the $\vec{c}$-scale -- such that

$$\text{ave}[\,((T(\vec{c}) + t)s - \mu)^2 \,|\, \mu,\sigma,\vec{c}\,] \;\; = \;\;$$

$$\text{ave}[\,(T(\vec{c}) + t)^2 s^2 - 2(T(\vec{c}) + t)s\mu + \mu^2 \,|\, \mu,\sigma,\vec{c}\,]$$

is minimal. Hence

$$\text{ave}[\,(T(\vec{c}) + t)s^2 \,|\, \mu,\sigma,\vec{c}\,] \;=\; \text{ave}[\,s\mu \,|\, \mu,\sigma,\vec{c}\,]$$

or

$$T_{\mu,\sigma}(\vec{c}) \;=\; \frac{\mu\ \text{ave}[\,s\,|\,\mu,\sigma,\vec{c}\,] - \text{ave}[\,ts^2\,|\,\mu,\sigma,\vec{c}\,]}{\text{ave}[\,s^2\,|\,\mu,\sigma,\vec{c}\,]}$$

$$=\; \frac{\displaystyle\int_{0}^{\infty}\!\!\int_{-\infty}^{\infty} s(\mu - ts)\, k_F(s,t\,|\,\mu,\sigma,\vec{c})\ ds\,dt}{\displaystyle\int_{0}^{\infty}\!\!\int_{-\infty}^{\infty} s^2 k_F(s,t\,|\,\mu,\sigma,\vec{c})\ ds\,dt}$$

$$=\; \frac{-\displaystyle\int_{0}^{\infty}\!\!\int_{-\infty}^{\infty} p(qp)\, k_F(p,q\,|\,0,1,\vec{c})\ dp\,dq}{\displaystyle\int_{0}^{\infty}\!\!\int_{-\infty}^{\infty} p^2\ k_F(p,q\,|\,0,1,\vec{c})\ dp\,dq} \qquad \text{(see 1.1)}$$

$$=\; \frac{-\text{ave}[\,qp^2\,|\,0,1,\vec{c}\,]}{\text{ave}[\,p^2\,|\,0,1,\vec{c}\,]} \;=\; T_{0,1}(\vec{c}). \qquad\qquad (2.2)$$

This invariance result is intuitively obvious: if a code clerk gives the configuration $\vec{c}$ to the statistician and asks for a "good" location estimate, he has to come up with the same answer no matter what the true parameter values μ and σ.

For further discussion of this problem see: Pitman(1938),

Fraser(1979), Pregibon and Tukey(1981), Bell and Morgenthaler(1981).

Since the values of the parameters μ and σ do not matter, we will always calculate our integrals with a standard form $\mu = 0$, $\sigma = 1$ and otherwise not mention them.

The observation that the configuration $\vec{c}$ contains all of the allowed information is formalized by the statistical notion of ancillarity. The distribution across the 2-dimensional configurations, i.e. the frequencies of falling into configurations, only depends on the shape F( ) but not on the actual values of μ and σ. This is true since the step from the sample $\vec{y}$ to the configuration $\vec{c}$ involves the subtraction of a "location" estimate and division by a "scale" estimate, so that the distribution of $\vec{c}$ no longer involves the parameters μ and σ (and hence is pure shape). This fact is used as an argument to condition on the ancillary statistic $\vec{c}$ -- the conditional distribution contains the "information" about location and scale (see: Fisher(1934)). These are the ideas we want to use. They greatly simplify all inference problems about μ and σ. Firstly all our mappings will be uniquely defined on each configuration $\vec{c}$ by a single real number t_c. But secondly the choice of t_c is often possible without reference to other configurations and we hence deal with a "simple" 2-dimensional problem no matter how big the sample size.

2.1. The single situation confidence distribution.

Let us fix a configuration $\vec{c}$ and a shape F ( ) in standard form, i.e. $\mu = 0$. Conditioned on $\vec{c}$ we deal with the 2-dimensional

conditional density

$$k_F(s,t|\vec{c}) = \frac{s^{n-1} \prod\limits_{i=1}^{n} t(s(c_i+t))}{\int\limits_{0}^{\infty}\int\limits_{-\infty}^{\infty} s^{n-1} \prod\limits_{i=1}^{n} t(s(c_i+t)) \; dsdt} \tag{2.3}$$

where $\vec{c} = (c_1,\ldots,c_n)$ and $t(x) = \frac{d}{dx}F(x)$.

example: Gaussian case ($F = \phi$)

$$k_\phi(s,t|\vec{c}) = \frac{s^{n-1}\exp(-\frac{s^2}{2}\sum\limits_{i=1}^{n}(c_i+t)^2)}{(\frac{2\pi}{n})^{\frac{1}{2}}(n-3)(n-5)\ldots(2 \text{ or } (\frac{\pi}{2})^{\frac{1}{2}})} \; (\sum\limits_{i=1}^{n}(c_i-\bar{c})^2)^{\frac{n-1}{2}}$$

The normalizing constant in this equation has to be understood as

$$= (n-3)(n-5)\ldots 2 \quad \text{if n odd}$$

$$= (n-3)(n-5)\ldots 1(\frac{\pi}{2})^{\frac{1}{2}} \quad \text{if n even.}$$

Now we want to study the effects of setting an upper bound for the parameter μ. If u is the value of the upper bound statistic U on the $\vec{c}$-scale, i.e. $U(\vec{c}) = u$, the statistic is defined on the whole configuration by

$$U(\vec{y}) = U(s(\vec{c}+t\vec{1})) = s(u+t)$$

(location and scale equivariance).

We are interested in the coverage probability defined by

$$Co_F(u) := P_F[U(y) > 0|\vec{c}]$$

It follows

$$Co_F(u) = P_F[s(u+t) > 0|\vec{c}] = P_F[t > -u|\vec{c}]$$

$$Co_F(u) = \int_0^\infty \int_{-u}^\infty k_F(s,t|\vec{c}) \, dsdt. \qquad (2.4)$$

$Co_F(u)$ is the function which for a choice of an upper bound u on the $\vec{c}$-scale gives as its value the conditional probability of the upper bound statistic actually being an upper bound for the parameter μ on the given configuration.

remark:

Again $Co_F(u)$ is not dependent on the values μ and σ which are used to calculate the integrals. For a shape F( ) not in standard form we would have

$$Co_{\mu,\sigma,F}(u) = P_{\mu,\sigma,F}[U(y) > \mu] = P_{\mu,\sigma,F}[t > \frac{\mu}{s}-u] =$$

$$\int_0^\infty \int_{\frac{\mu}{s}-u}^\infty k_F(s,t|\mu,\sigma,\vec{c}) \, dsdt$$

$$= \int_0^\infty \int_{-u}^\infty k_F(p,q|0,1,\vec{c}) \, dpdq = Co_F(u) \qquad (see\ 2.1)$$

$(t = \frac{\mu}{s}-u \longrightarrow q = \frac{\mu}{s}-u-\frac{\mu}{s} = -u)$.

From (2.4) we see that

$$Co_F(u) \longrightarrow 1 \quad as\ u \longrightarrow \infty$$

$$Co_F(u) \longrightarrow 0 \quad as\ u \longrightarrow -\infty$$

November 15, 1983

which is consistent with our intuition. $Co_F(u)$ is like a distribution function and can be used to define confidence intervals for the location parameter μ.

Whenever we choose $L(\vec{c})$ and $U(\vec{c})$ such that $Co_F(U(\vec{c})) - Co_F(L(\vec{c})) = .95$ we have a 2-dimensional piece of a 95% confidence-interval procedure which has exactly 95% coverage probability conditioned on the configuration. The upper bound of the interval is below μ with probability $1 - Co(U(\vec{c}))$ and the lower bound is above with probability $Co(L(\vec{c}))$ and therefore the interval covers the parameter with conditional probability

$$1 - (1 - Co(U(\vec{c})) + Co(L(\vec{c}))) = Co(U(\vec{c})) - Co(L(\vec{c}))$$

If we choose to do so for each configuration, we will have a 95% confidence-interval procedure, since we get the "overall" coverage probability by averaging the conditional coverage probabilities, which in the above case are constant, across configurations.

The function $Co_F(\ )$ will be called a <u>coverage</u> <u>distribution</u> or a <u>confidence</u> <u>distribution</u>. It has the density

$$\frac{d}{du}Co_F(u) =: co_F(u) = \frac{d}{du} \int_0^\infty \int_{-u}^\infty k_F(s,t|\vec{c}) \, dsdt$$

$$co_F(u) = \int_0^\infty k_F(s,-u|\vec{c}) \, ds. \tag{2.5}$$

remark:

How does the mean of the confidence distribution compare with the minimum-mean-square-error location estimate? The mean m is equal to

$$m = \int\limits_{-\infty}^{\infty} u \int\limits_{0}^{\infty} k_F(s,-u|\vec{c}) \, ds \, du =$$

$$-\int\limits_{-\infty}^{\infty} \int\limits_{0}^{\infty} t \, k_F(s,t|\vec{c}) \, ds dt = -ave(t|\vec{c}),$$

whereas the "optimal" location estimate t_o is

$$t_o = - \frac{ave(ts^2|\vec{c})}{ave(s^2|\vec{c})} \qquad \text{(see 2.2)}.$$

The "optimal" estimate therefore takes the correlation between s and t into account.

The confidence distribution is not dependent upon the values of the parameters μ and σ used for the calculation as we have already seen. Now we want to explore the behavior under changes of the class representing element $\vec{c}$.

Let us first look at the case where $\vec{d} = v\vec{c}$, i.e. our representing element is scaled by $v \in R_+$. From (2.3) we have

$$k_F(s,t|\vec{d}) = \frac{s^{n-1} \prod\limits_{i=1}^{n} f(s(d_i+t))}{\int\limits_{0}^{\infty} \int\limits_{-\infty}^{\infty} s^{n-1} \prod\limits_{1}^{n} f(s(d_i+t)) \, ds dt}$$

$$= \frac{s^{n-1} \prod\limits_{i=1}^{n} f(s(vc_i+t))}{\int\limits_{0}^{\infty} \int\limits_{-\infty}^{\infty} s^{n-1} \prod\limits_{i=1}^{n} f(s(vc_i+t)) \, ds dt}$$

$$= \frac{(vs)^{n-1} \prod\limits_{i=1}^{n} f(vs(c_i+\frac{t}{v}))}{\int\limits_{0}^{\infty} \int\limits_{-\infty}^{\infty} (vs)^{n-1} \prod\limits_{i=1}^{n} f(vs(c_i+\frac{t}{v})) \, ds dt}$$

$$= \frac{(vs)^{n-1} \prod\limits_{i=1}^{n} t(vs(c_i + \frac{t}{v}))}{\int\limits_{0}^{\infty} \int\limits_{-\infty}^{\infty} a^{n-1} \prod\limits_{i=1}^{n} t(a(c_i + b)) \, da \, db} = k_F(vs, \frac{t}{v} | \vec{c})$$

We used the change of variables:

$$a = vs \longrightarrow da = vds \text{ and } b = \frac{t}{v} \longrightarrow db = \frac{dt}{v} \longrightarrow da \, db = ds \, dt$$

The confidence density on the $\vec{d}$-scale is therefore

$$co_F(u|\vec{d}) = \int\limits_{0}^{\infty} k_F(s, -u|\vec{d}) \, ds = \int\limits_{0}^{\infty} k_F(vs, -\frac{u}{v}|\vec{c}) \, ds$$

$$= \int\limits_{0}^{\infty} k_F(a, -\frac{u}{v}|\vec{c}) \frac{da}{v} = \frac{1}{v} co_F(\frac{u}{v}|\vec{c}).$$

We see that the multiplier v behaves like a scale parameter!
Now let us consider the case where $\vec{d} = \vec{c} + w\vec{1}$, i.e. our representing
element is translated in the $(1,1,\ldots,1)$ - direction by $w \in \mathbb{R}$. From
(2.3) we have

$$k_F(s, t|\vec{d}) = \frac{s^{n-1} \prod\limits_{i=1}^{n} t(s(d_i + t))}{\int\limits_{0}^{\infty} \int\limits_{-\infty}^{\infty} s^{n-1} \prod\limits_{i=1}^{n} t(s(d_i + t)) \, ds \, dt}$$

$$= \frac{s^{n-1} \prod\limits_{i=1}^{n} t(s(c_i + w + t))}{\int\limits_{0}^{\infty} \int\limits_{-\infty}^{\infty} s^{n-1} \prod\limits_{i=1}^{n} t(s(c_i + w + t)) \, ds \, dt}$$

$$= \frac{s^{n-1} \prod\limits_{i=1}^{n} t(s(c_i + w + t))}{\int\limits_{0}^{\infty} \int\limits_{-\infty}^{\infty} s^{n-1} \prod\limits_{i=1}^{n} t(s(c_i + b)) \, ds \, dt} = k_F(s, t + w|\vec{c}).$$

We used the change of variable: $b = t+w \longrightarrow db = dt$.

The confidence density on the $\vec{d}$-scale is therefore

$$co_F(u|\vec{d}) = \int_0^\infty k_F(s,-u|\vec{d})ds = \int_0^\infty k_F(s,-(u+w)|\vec{c})ds = co_F(u+w|\vec{c}).$$

We see that a "translator" w behaves like a location parameter. The behavior under changes or transformations of the class-representing elements is again following our intuition:

$$co_F(u|v\vec{c}) = \frac{1}{v} co_F(\frac{u}{v}|\vec{c})$$

(2.6)

$$co_F(u|\vec{c}+w\vec{1}) = co_F(u+w|\vec{c}).$$

example: Gaussian case ($\underline{F} = \phi$)

Using our expression for $k_\phi(s,t|\vec{c})$ in formula (2.3) we get

$$co_\phi(u) = \int_0^\infty \frac{s^{n-1}\exp(-\frac{s^2}{2}\sum_{i=1}^n (c_i-u)^2)}{(\frac{2\pi}{n})^{\frac{1}{2}}(n-3)(n-5)..(2 \text{ or }(\frac{\pi}{2})^{\frac{1}{2}})} (\sum_{i=1}^n (c_i-\bar{c})^2)^{\frac{n-1}{2}} ds$$

$$= \frac{(n-2)(n-4)..(2 \text{ or }(\frac{\pi}{2})^{\frac{1}{2}})}{(\frac{2\pi}{n})^{\frac{1}{2}}(n-3)(n-5)..(2 \text{ or }(\frac{\pi}{2})^{\frac{1}{2}})} \frac{(\sum_{i=1}^n (c_i-\bar{c})^2)^{\frac{n-1}{2}}}{(\sum_{i=1}^n (c_i-u)^2)^{\frac{n}{2}}}$$

$$= \frac{1}{s_{\bar{c}}} t_{n-1}(\frac{u-\bar{c}}{s_{\bar{c}}})$$

(2.7)

where $s_{\bar{c}}^2 = \frac{1}{n(n-1)}\sum_{i=1}^n (c_i-\bar{c})^2$ and $t_{n-1}(\ )$ is the familiar t-density with $(n-1)$ degrees of freedom in its standard form. We see that the confidence density in this case has a fixed shape across

configurations so that we could have chosen our class-representing

elements in such a way that $s_{\bar{c}} = 1$ and $\bar{c} = 0$ in which case the

confidence density would always be $t_{n-1}(\ )$ in standard form.

Figures 2.5 and 2.6 show the functions co( ) (see 2.5) for the

two configurations we used already in Fig. 2.1 through 2.4. The red

function shows the slash coverage density and the black one the

Gaussian coverage density, i.e. a Student's-t density with 19 degrees

of freedom. These densities are calculated by using a linear logistic

fit to Co( ). It is clear how outliers in a configuration greatly

influence the Gaussian coverage density. In the first -- Gaussian

drawn case -- the differences are not big, the slash density is moved

a bit to the left, which is reasonable if we look back at the

configuration.

Once we know the conditional confidence distributions for a

given sampling situation F, we can find a confidence interval

estimator for any prefixed confidence coefficient $100(1-\alpha)\%$ in the

following way.

In each configuration $\vec{c}$ we fix the upper and lower bounds

such that the conditional coverage probability is equal to

$100(1-\alpha)\%$. This implies that the overall confidence level will

indeed be $100(1-\alpha)\%$. We do, however, still have some freedom in the

choice of the upper and lower bounds. The most natural choice treats

the left and right tail of the conditional confidence distribution in

a balanced way, i.e. we leave out $100\frac{\alpha}{2}\%$ from each tail. This now

uniquely -- except possibly in pathological cases -- determines a

confidence interval estimate.

Figure 2.5 : Conditional coverage densities; same configuration as in Fig. 1.1 and 1.2

coverage densities for configuration
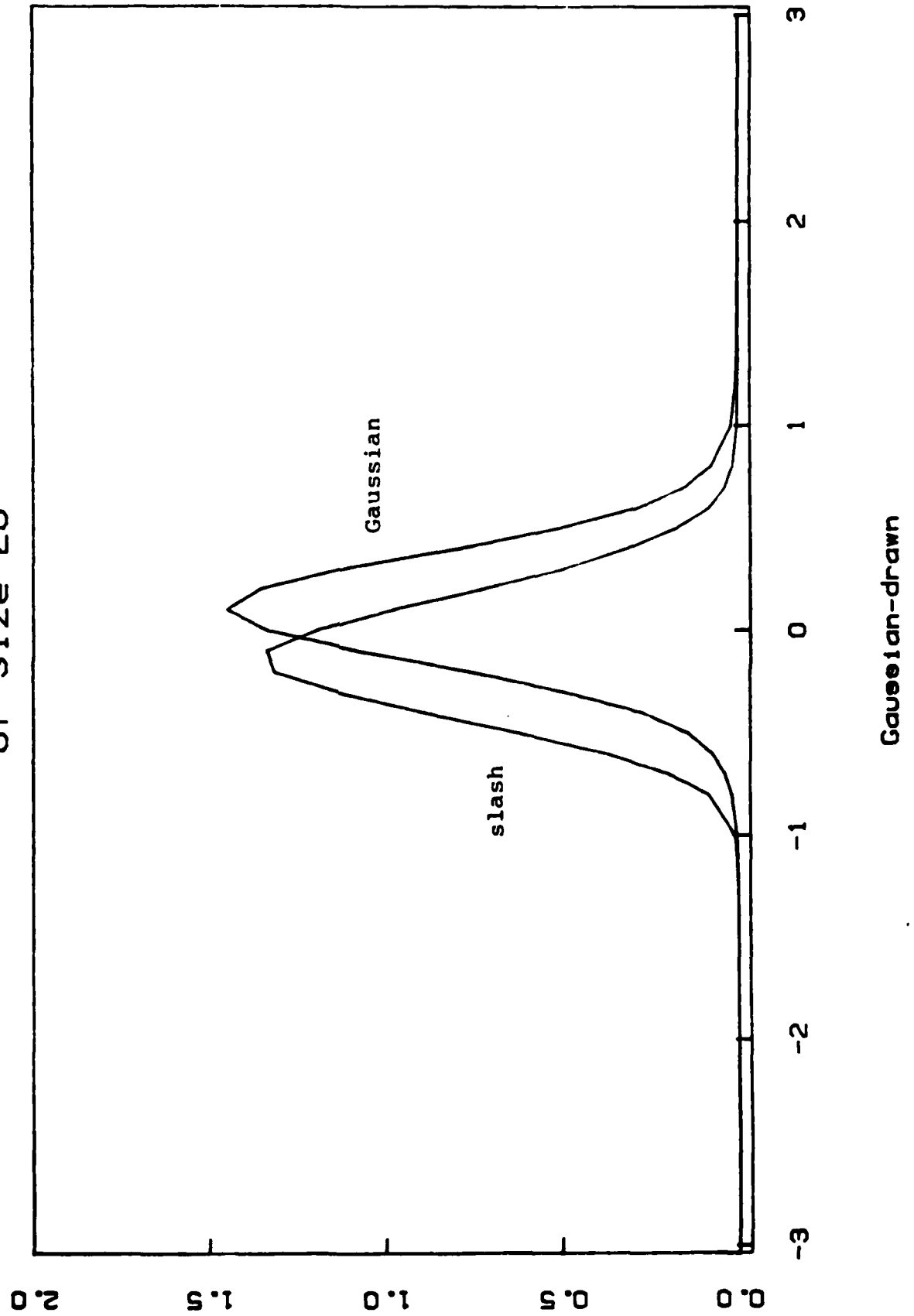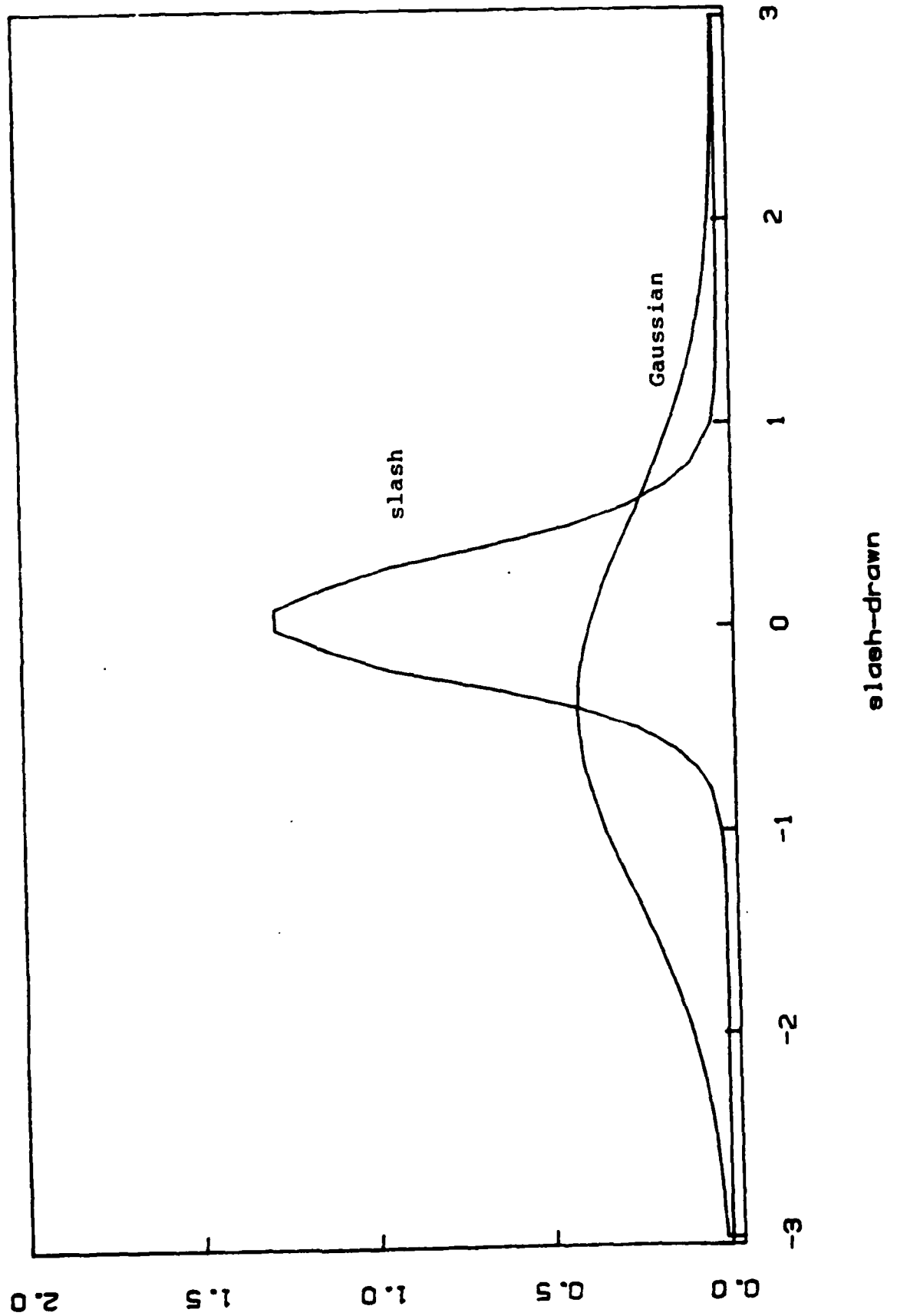of size 20

Gaussian

slash

Gaussian-drawn

Figure 2.6: Conditional coverage densities; same configuration as in Fig. 1.3 and 1.4

coverage densities for configuration
of size 20

Treating the two tails in a balanced way means that we care just as much about missing the true location because the lower bound is too high as missing because the upper bound is too low. Once we have fixed our idea about the allowed conditional missing probability, the balanced division seems the natural thing to do.

<u>example</u>: Gaussian case ($F = \phi$)

The coverage density in this case is always a t_{n-1} -density with location $\bar{c}$ and scale $s_{\bar{c}}$. The above procedure is therefore the usual symmetric t-interval.

3. <u>Strong confidence intervals for a location parameter</u>: A <u>compromise between the Gaussian and the slash</u>.

In this section we want to study the effects of not knowing the shape F of the underlying sampling situation. In order to do this we will look at the simplest possible case where we restrict attention to two possible candidates, the Gaussian and the slash. The latter is the distribution of the ratio of a standard Gaussian and a unit uniform which are independent. The density for the slash is

$$ f(x|\mu,\sigma) = \frac{\sigma}{(2\pi)^{\frac{1}{2}}(x-\mu)^2}[1-\exp(-\frac{(x-\mu)^2}{2\sigma^2})] \ . $$

From the verbal description we recognize the slash as a "continuous" mixture of Gaussians with scales which are like an inverse uniform. The density shows us the tail behavior as $\frac{1}{x^2}$ and is therefore like a t-density with one degree of freedom (see Rogers and Tukey (1972))

November 15, 1983

tor further insights).

Nonparametric interval procedures -- like the sign interval, the
Wilcoxon signed rank interval and so on -- are an attractive choice
of compromise for many statisticians. These intervals are guaranteed
to reach the desired confidence for all symmetric situations and
hence seem to solve the problem of compromising between situations
once and for all. But we would expect that seeking to put such a
vast class of situations under one hat has its disadvantages.
Furthermore it is not at all clear how these procedures behave
conditionally (on configurations). All nonparametric tests -- from
which the corresponding intervals are derived -- need an argument of
equal probability under permutations. They condition on the class of
samples which one gets from the one at hand under permuting around
the hypothesized parameter value. A bit of thought shows us that this
is an operation which does not preserve the configuration. It should
therefore be revealing to learn more about the properties of
nonparametric intervals conditioned on a given configuration.

It has been pointed out in the "robustness literature" that the
stability of the confidence level of nonparametric procedures is only
one aspect which the statistician tries to keep under control -- this
property has been named "robustness of validity". Another aspect of
interest is the efficiency -- which can be expressed in various ways
-- of a statistical procedure.

3.1. Strong confidence intervals

Even if robustness of validity is our goal, we need not stop at

nonparametric intervals.

> We might reasonably ask for a confidence interval which has a
> "robust" confidence coefficient (or coverage probability)
> conditioned on configurations. In our "two-situation world" this
> would mean that the intervals have to be big enough to reach
> $100(1-\alpha)$% coverage probability conditioned on each configuration
> for the slash and for the Gaussian. Of course there will
> usually not be such an interval and we have to settle for at
> least $100(1-\alpha)$% coverage probability conditioned on each
> configuration for the slash and the Gaussian.

We already know that over the sample space the slash coverage of
Student's t interval is conservative. The above approach might then
not be very far from the classical t-interval. We can also say that
the solution to this problem will end up enlarging the t-intervals in
certain configurations.

If we view robustness as a problem in stability of coverage
probabilities, i.e. a problem in how safe is the use of a confidence
interval, we are lead to search for rather long intervals and our
flexibility in choosing upper and lower bounds is quite restricted as
Figure 3.1 tries to show.

Figure 3.1: Solutionspace $[L(\vec{c}), U(\vec{c})]$ conditioned on configuration
$\vec{c} = (c_1, \ldots, c_n)$
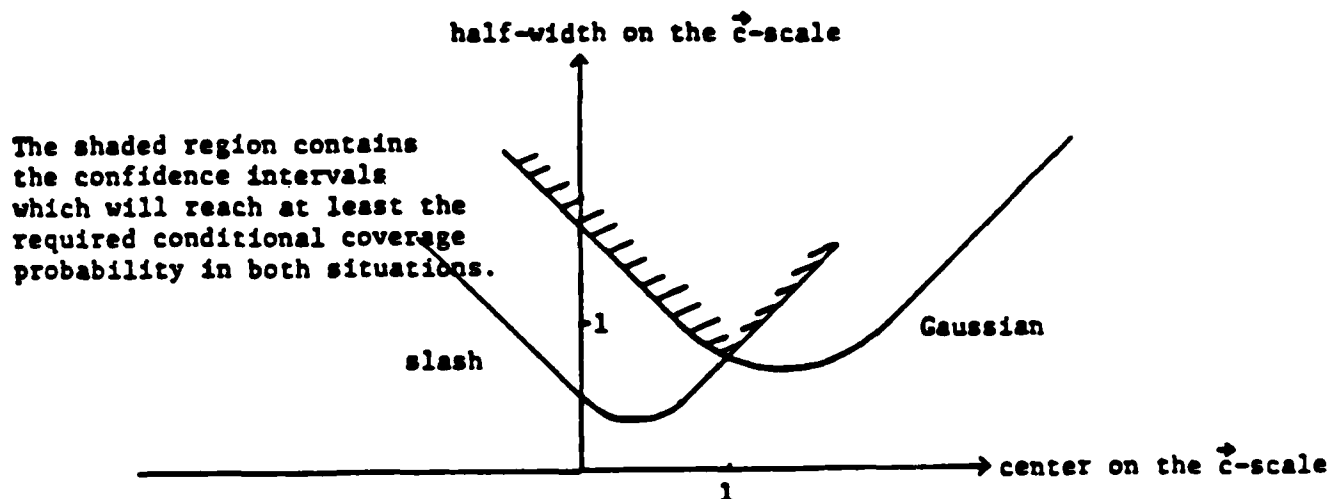
In order to study these strong interval procedures we adopt the following strategy. Our starting point for each configuration consists of the two intervals which have $100(1-\alpha)$% coverage probability and are treating the upper and lower tail of the coverage density symmetrically. Figure 3.2 shows the conditional situation.

Figure 3.2: Conditional coverage densities for the slash and Gaussian situations and the two symmetric intervals
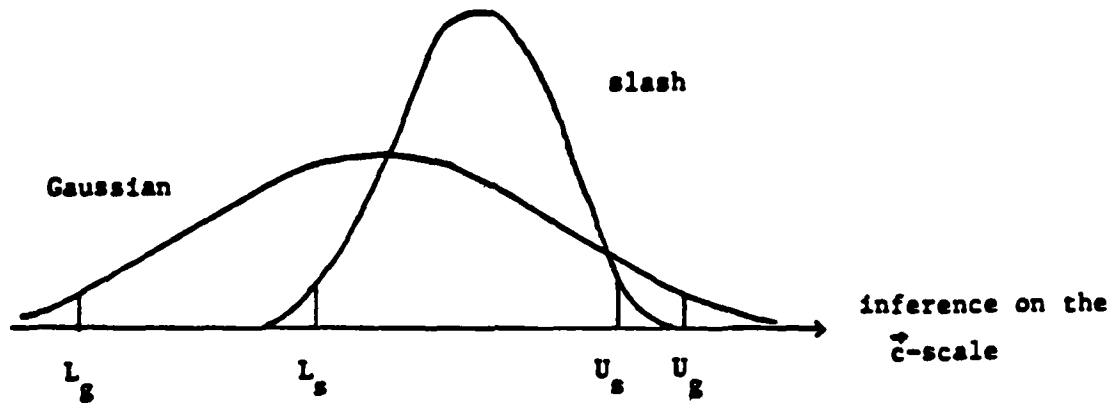
In order to find a "strong" interval, i.e. one which has at least $100(1-\alpha)\%$ conditional coverage probability in both situations we will look at the interval which we get by selecting the maximum of the two upper bounds and the minimum of the two lower bounds.

$$L = \min(L_g, L_s)$$

$$U = \max(U_g, U_s) \qquad (3.1)$$

(see Figure 3.2)

Clearly this will be a strong interval which can be found relatively easily. There are four possible cases into which configurations can fall:

(a) $[L, U] = [L_g, U_g]$

(b) $[L, U] = [L_s, U_s]$

(c) $[L, U] = [L_g, U_s]$

$$(d) \qquad [L,U] = [L_s, U_g] \qquad\qquad (3.2)$$

The cases (a) and (b) are such that one of the two situations dominates the other, and the choice of [L,U] is more or less the best thing one can do. In the mixed cases (c) and (d) however the interval [L,U] could be shortened further and still be kept strong. The possible gains from such a shortening however are small compared to the difficulty in finding them.

We will therefore examine the intervals given by (3.1) for the case of 95% coverage probability. Table 3.1 shows the percentage of observed cases (a), (b) and (c)&(d) in the different sample sizes.

Table 3.1: Percentage of observed cases (a), (b) and (c)&(d) (for definition see (3.2))

sample size	Gaussian situation			slash situation		
	(a)	(b)	(c)&(d)	(a)	(b)	(c)&(d)
20	18%	12%	70%	80%	0%	20%
10	$32\frac{2}{3}$%	$16\frac{2}{3}$%	$50\frac{2}{3}$%	78%	$1\frac{1}{3}$%	$20\frac{2}{3}$%
5	$89\frac{1}{5}$%	$3\frac{1}{5}$%	$7\frac{3}{5}$%	$94\frac{4}{5}$%	1%	$4\frac{1}{5}$%

(In samples of sizes 10 and 20 the numbers are based on 150 slash-drawn and 150 Gaussian-drawn configurations, in samples of size 5 the corresponding numbers are 500 and 500!)

The message from this table is striking. The Gaussian situation is dominating in samples of size 5, where we are most of the time in a case as shown in Figure 3.2 and where Student's t interval is close

to the "best" (in terms of expected length) strong interval procedure. In samples of size 5 then, we do not expect that compromising between Gaussian and heavy-tailed slash is very difficult. The slash intervals are overly optimistic and too short it judged from the Gaussian point of view. The Gaussian intervals, i.e. Student's t intervals, on the other hand seem to do a good job of keeping the conditional coverage probability above 95% even in the slash situation. And this is true whether the samples are drawn from a Gaussian or from a heavy-tailed shape. One might conclude that maybe more serious challenges to the Gaussian model for location confidence intervals have to come from less heavy-tailed shapes than the slash.

Samples of size 10 and 20 behave roughly similarly, but very differently from samples of size 5. In slash drawn samples the Gaussian situation mostly dominates, but not as overwhelmingly as in samples of size 5. In Gaussian drawn configurations the mixed cases are a majority, introducing the slash along with the Gaussian and thus really contributes a new point of view. In a lot of configurations it forces us to acknowledge the fact that Student's t interval doesn't stretch far enough to the right or to the left and has to be enlarged. A more detailed account of the behavior of the "strong" intervals from a configural point of view gives the following results.

For samples of size 20, Figure 3.3 shows the conditional coverage probabilities for the Gaussian and slash situation. Figure 3.4 is the corresponding picture for sample size 10. We see how the

Figure 3.3: Stem-and-leaf plots of the cond. confidence levels of
the strong interval (3.1) for 150 sampled configurations of size 20

N = 150   Median = 0.96295   Hinges = 0.955, 0.9676

Decimal point is 3 place(s) to the left of the colon

```
24    24    952 : 6677777888888888999999999
35    11    953 : 0000111227B
37     2    954 : 44
40     3    955 : 044
48     8    956 : 22334578
49     1    957 : 8
52     3    958 : 777
56     4    959 : 5889
64     8    960 : 23455689
69     5    961 : 11279
75     6    962 : 124699
75    12    963 : 0256667888899
63    11    964 : 00234447899
52     6    965 : 113469
46     6    966 : 005579
40     9    967 : 056669999
31     5    968 : 23578
26     4    969 : 1358
22     4    970 : 0067
18     6    971 : 244799
12     6    972 : 002366
 6     1    973 : 7
 5     2    974 : 29
 3     2    975 : 07
 1     0    976 :
 1     0    977 :
 1     1    978 : 8
```

Gaussian situation

N = 150   Median = 0.9983   Hinges = 0.9809, 1

Decimal point is 2 place(s) to the left of the colon

```
 1     1     95 : 1
 2     1     95 : 7
 3     1     96 : 2
 9     6     96 : 667788
20    11     97 : 11233444444
37    17     97 : 55555555555567789
43     6     98 : 122444
48     5     98 : 55999
57     9     99 : 002224444
      27     99 : 5555566666777888888999999999
66    66    100 : 0000000000000000000000000000000000000000000000000000000000000
```

slash situation

Figure 3.4: Stem-and-leaf plots of the cond. confidence levels of
the strong interval (3.1) for 150 sampled configurations of size 10

N = 150   Median = 0.9559   Hinges = 0.95, 0.962

Decimal point is 3 place(s) to the left of the colon

```
52    52    950 : 00000000000000000000000000000000000000000000000000124
56     4    951 : 6688
62     6    952 : 135678
67     5    953 : 11246
73     6    954 : 344889
75     2    955 : 08
75     9    956 : 012237889
66     7    957 : 0334588
59     5    958 : 02569
54     4    959 : 4557
50     7    960 : 0133469
43     5    961 : 02568
38     9    962 : 013367789
29     9    963 : 013555689
20     3    964 : 579
17     2    965 : 29
15     4    966 : 0267
11     2    967 : 15
 9     1    968 : 2
 8     0    969 :
 8     1    970 : 8
 7     2    971 : 23
 5     0    972 :
 5     4    973 : 0155
 1     1    974 : 9
```

Gaussian situation

N = 150   Median = 0.993   Hinges = 0.9737, 0.9999

Decimal point is 2 place(s) to the left of the colon

```
 4     4     95 : 0023
 8     4     95 : 6899
16     8     96 : 00012234
31    15     96 : 555677888888899
44    13     97 : 0001124444444
54    10     97 : 5556779999
58     4     98 : 0034
70    12     98 : 555566799999
      10     99 : 0012333344
70    24     99 : 55667777778888888889999
46    46    100 : 000000000000000000000000000000000000000000000000
```

slash situation

smaller sample size moves us nearer to Student's t interval, which would have exactly constant 95% conditional coverage probability in the Gaussian situation. In the slash situation this shows in a trend to be increasingly "overlong" in samples of size 20.

The following table gives the expected length of these intervals.

Table 3.2: Expected length of "strong" intervals

	Gaussian situation	slash situation
size=5	2.335 (.0003)	103.0 (75.6)
size=10	1.458 (.0058)	22.12 (7.54)
size=20	1.021 (.0053)	63.93 (52.2)

In comparison to Table 3.1 we can again see what we noticed by looking at conditional coverage probabilities. As the sample size increases one has to enlarge Student's t interval in order to have at least 95% slash conditional coverage.

The next three pictures, Figures 3.5, 3.6 and 3.7, show us something about the conditional unbalance or asymmetry of the strong intervals. Here we plot the conditional probability of missing the true parametervalue at the low (i.e. left) end vs. at the high end.

Two things are quite noticable. In all three cases most of the strong intervals are slightly asymmetric. There is a branch from (0.0, 0.0) to (0.025, 0.025) above and below the diagonal -- which contains the balanced intervals. The correction from

Figure 3.5: Plot of the lower vs. the upper conditional missing probabilities of the strong intervals (4.1) in the slash situation

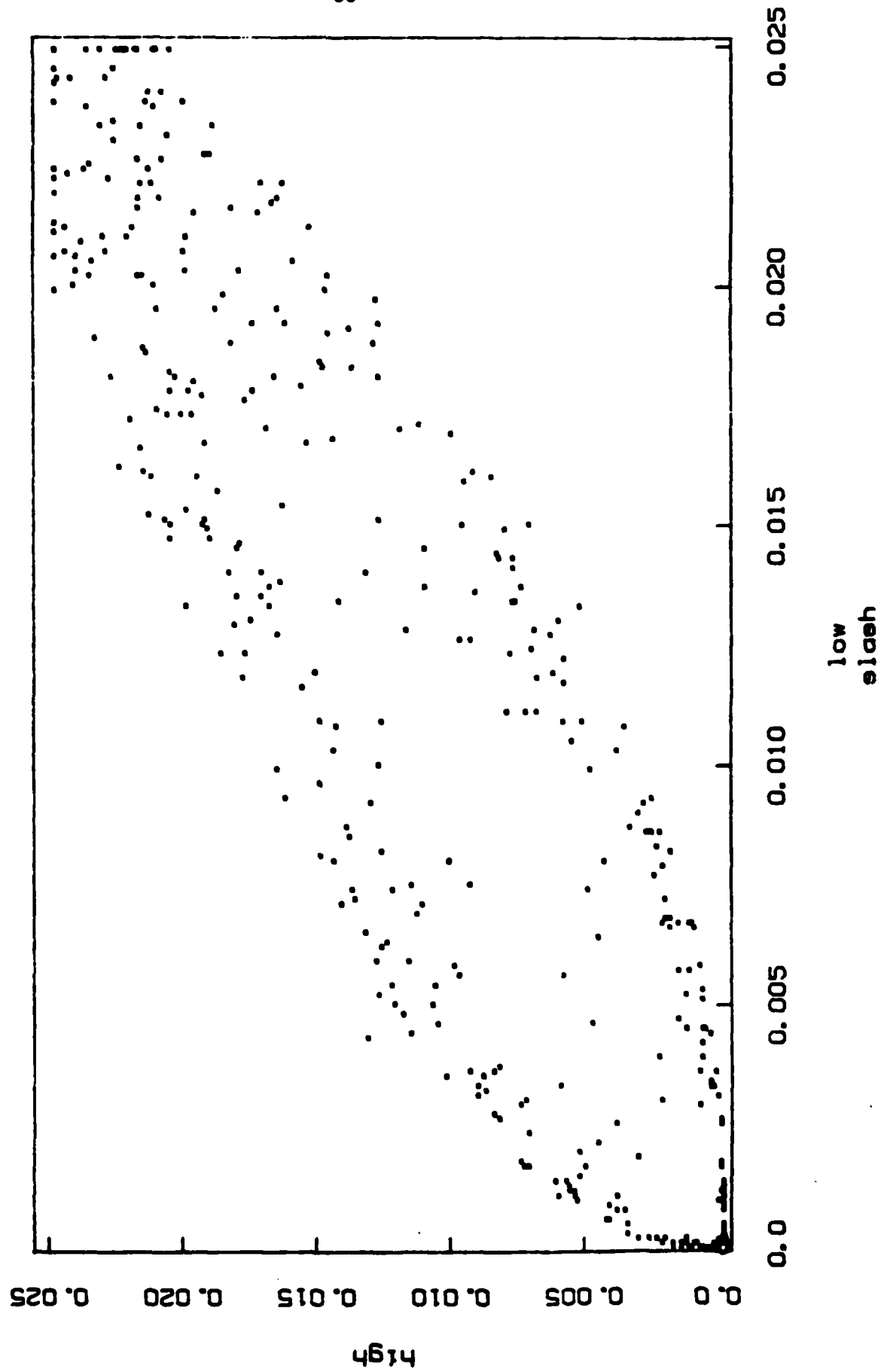strong intervals in samples of size 5

Figure 3.6: Plot of the lower vs. the upper conditional missing probabilities of the strong intervals (4.1) in the slash situation

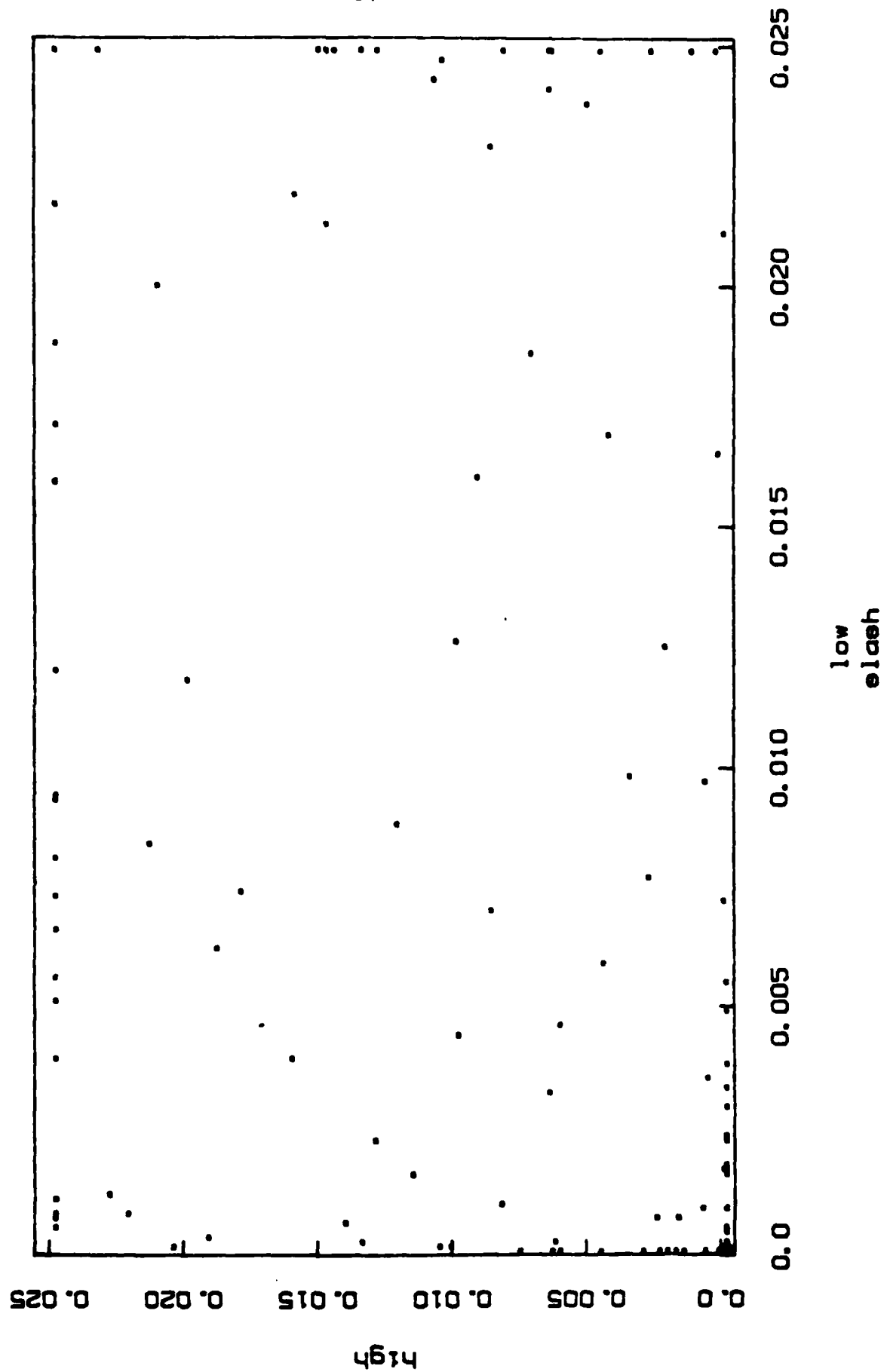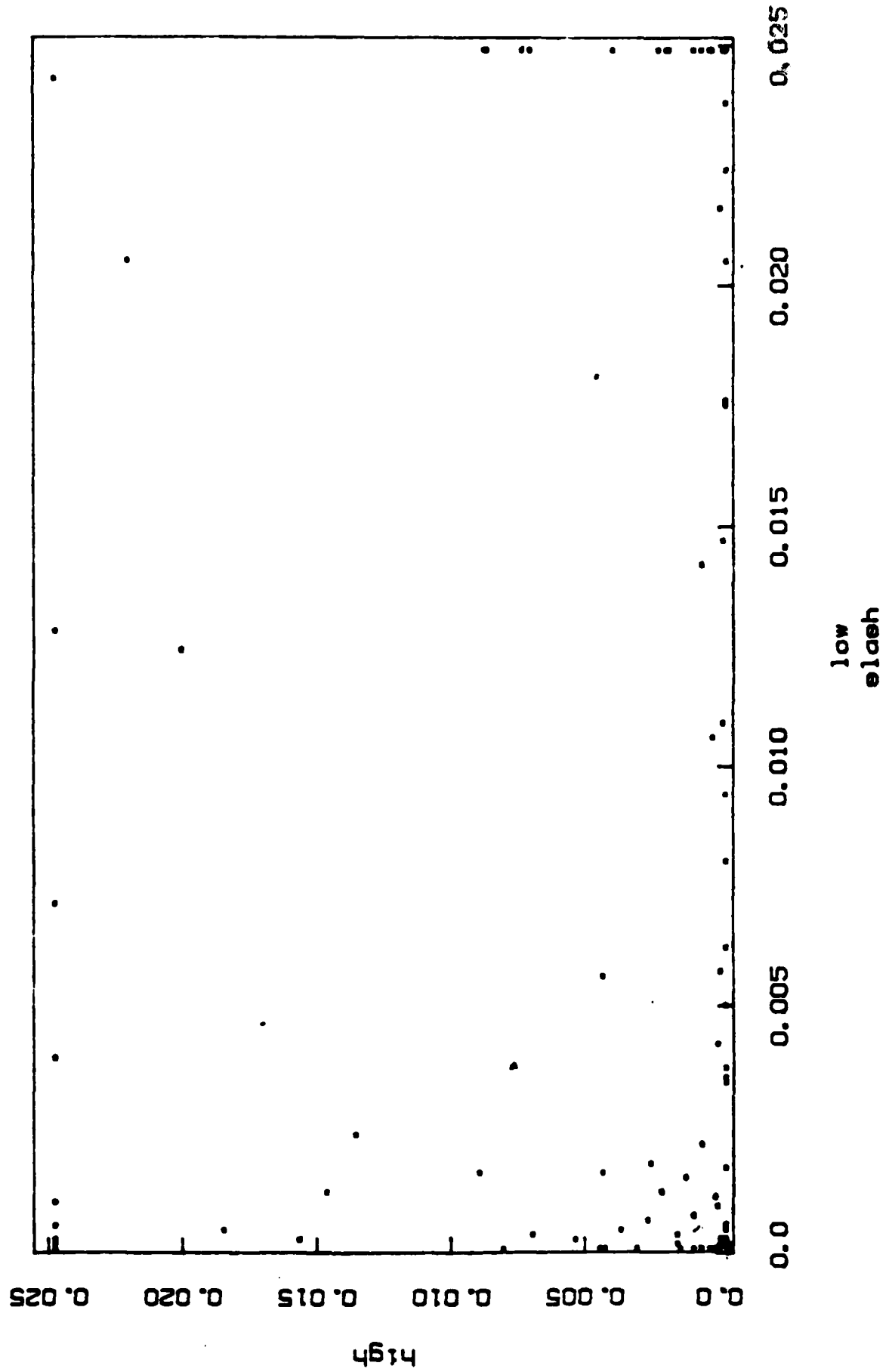strong intervals in samples of size 10

Figure 3.7: Plot of the lower vs. the upper conditional missing probabilities of the strong intervals (4.1) in the slash situation

strong intervals in samples of size 20

Student's t to the strong interval is therefore usually one sided. Furthermore we notice how the increase of sample size influences the picture. More and more points get absorbed into (0.0, 0.0) -- which means they are overlong -- and the two branches we discussed are pushed toward the edges.

The above discussion has an interesting consequence: maybe it makes more sense to compromise with a less extremely heavy-tailed counterpart than the slash in samples of size 5. This finding disagrees with the fact that with bigger sample sizes the distinction between samples drawn from the slash and samples drawn from the Gaussian becomes more "obvious" -- and that compromising these two situations is therefore simpler in larger samples (see: Bell & Morgenthaler(1981)). To try this idea the "slacu"-distribution was used together with the Gaussian. This is the distribution of the ratio of a standard Gaussian and the cube root of a unit uniform -- it's density has tails like $\frac{1}{x^3}$. But still Student's t interval comes out to be very close to the strong interval -- now strong for Gaussian and slacu! Table 3.3 shows the numbers.

Table 3.3: Percentage of observed cases (a), (b) and (c)&(d) (for definition see (3.2) note, however, that the slash is replaced by the slacu)

sample size	Gaussian situation			slash situation		
	(a)	(b)	(c)&(d)	(a)	(b)	(c)&(d)
5	78 $\frac{4}{5}$%	$\frac{3}{5}$%	20 $\frac{3}{5}$%	91 $\frac{2}{5}$%	1 $\frac{1}{5}$%	7 $\frac{2}{5}$%

Surprisingly little changed by replacing the slash with the slacu. It

we want to get robustness in the sense of stable conditional coverage
probability under heavy-tailed challenges, Student's t interval
behaves well in samples of size 5 and leaves little scope for
modifications.

3.2. Conditional behavior of nonparametric confidence intervals

As we pointed out in the introduction to this chapter,
nonparametric confidence intervals, which are assured to have a fixed
confidence level "over the sample space" for all symmetric
situations, need not have a fixed coverage probability conditioned on
the configuration. In this section we will see in a descriptive
fashion how the intervals derived from Wilcoxon's signed rank
statistic and from the sign test statistic behave conditioned on
configurations.

We are interested -- as always -- in 95% confidence intervals
and of course neither of the above procedures will be able to create
a 95% confidence interval for samples of size 5, where the interval
defined by the minimum and maximum of the sample has $1 - \frac{2}{32} = .9375$
coverage probability. In order to compare over the full range of
sample sizes, we will include the "range" - interval for sample size
5. For samples of size 10 and 20 we use logistic interpolation in the
Wilcoxon and binomial tables to get approximately a 95% confidence
interval. On the configuration scale, i.e. expressed as intervals for
$c_1 \leq c_2 \leq \ldots \leq c_n$, we use the following intervals:

	sign interval	Wilcoxon-interval
size=5	$[c_1, c_5]$	$[c_1, c_5]$
size=10	$[.5c_2+.5c_3, .5c_8+.5c_9]$	$[.26w_8+.74w_9, .74w_{47}+.26w_{48}]$
size=20	$[.82c_6+.18c_7, .82c_{15}+.18c_{14}]$	$[.34w_{54}+.66w_{53}, .34w_{157}+.66w_{158}]$

$(w_1 \leq \ldots \leq w_{\frac{n(n+1)}{2}}$ denote the Walsh averages, i.e. $\dfrac{y_i + y_j}{2}$ ordered by value.)

It turns out that the sign interval and the Wilcoxon interval have quite different behaviors as one looks across situations and sample sizes. The following table contains the estimated variation of the conditional coverage probabilities.

Table 3.4: Hinge-spreads (see Tukey(1977)) for conditional coverage probability in %

		Gaussian	slash
sign intervals	size=20	3.50%	3.73%
	size=10	3.66%*	2.58%*
	size=5	2.44%	5.35%
Wilcoxon intervals	size=20	1.25%	3.32%
	size=10	1.30%*	4.00%*
	size=5	2.44%	5.35%

(*: Entries for sample size 5 are for the "range"-interval)

The sign interval procedure is getting worse in the slash situation from samples of size 10 to 20 as far as stability of the conditional coverage level is concerned. The Wilcoxon interval seems to improve. On the whole it is surprising how little the increase in sample size stabilizes the conditional coverage levels.

November 15, 1983

Both interval procedures are bad for very small samples in the
heavy-tailed slash situation. Looking across situations, the two
confidence intervals are complementary as well. The sign interval is
more stable in the slash for the intermediate sample size -- whereas
the Wilcoxon interval is better in the Gaussian. If we adopt a
measure of variation which uses more of the information in the tails
of the distribution -- like the standard deviation -- the effects
come out even more clearly. They can also be seen in Figures 3.8 and
3.9, which show box plots (see: Tukey(1977)) of the conditional
coverage probabilities.

In these Figures we use the logistic transforms of the conditional
coverage probabilities p, defined by

$$\text{logit}(p) \quad = \quad \log(\frac{p}{(1-p)}\frac{.05}{.95}) \tag{3.3}$$

so that, on the transformed scale, a value of zero corresponds to
exactly 95% conditional coverage, whereas positive values indicate
conditional coverage bigger than 95% and negative values indicate
coverage smaller than 95%. We hope that the distributions will be
made more symmetric by this re-expression.

In these pictures we can also see how the numbers of Table 3.4
came about. The Wilcoxon interval in samples of size 10 produces
quite low conditional slash coverage probabilities in some "extreme"
configurations. This is clear from the fact, that information from
the smallest and largest observation are sometimes used, since they
can contribute up to the 10th and 46th Walsh average. The median of
the conditional coverage probabilities can be quite substantially
bigger than 95% (the mean over configurations). This reflects the

Figure 3.8:

Logistic transforms for 150 sampled configurations of the sign interval

cond. coverage for sign in samples of size 10 & 20
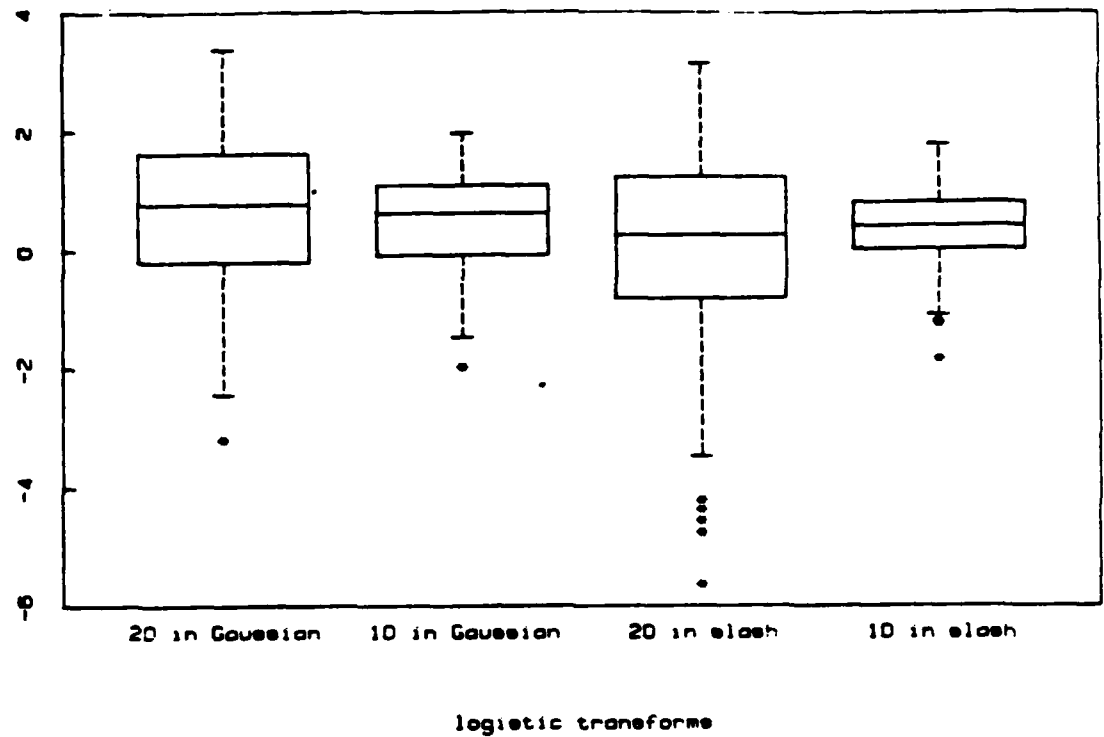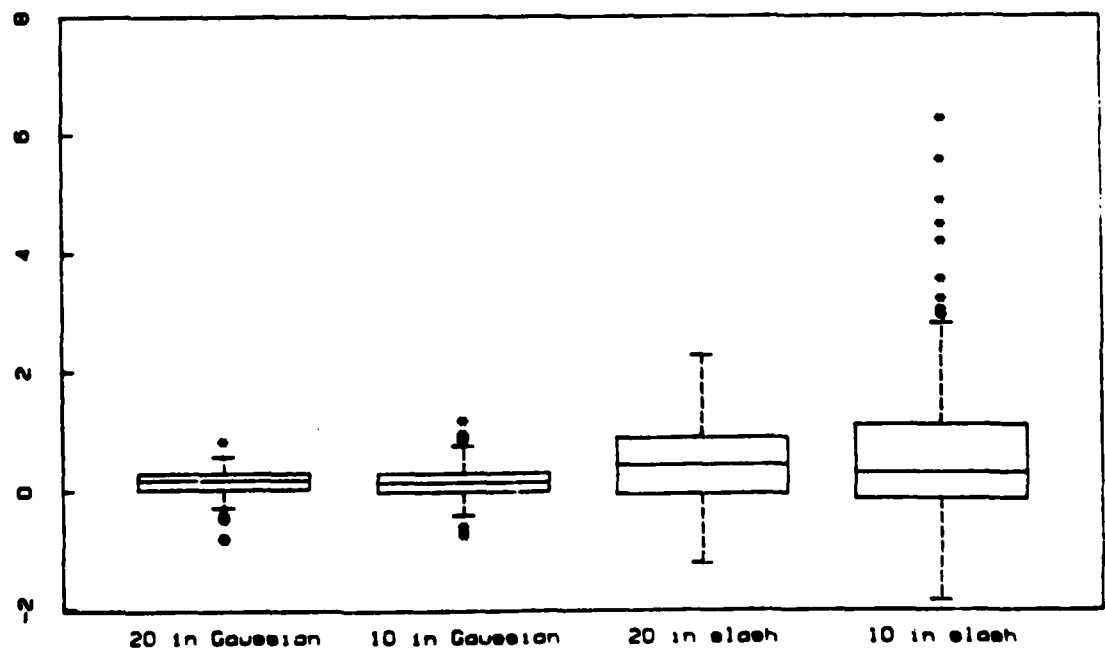

logistic transforms

Figure 3.9:

Logistic transforms for 150 sampled configurations of the Wilcoxon interval

cond. coverage for Wilcoxons in samples of size 10 & 20

fact that the distribution of the conditional coverage probabilities is skewed towards low values.

Asymptotically, these differences between configurations we have observed here, disappear. If we repeatedly take large samples from a fixed shape F( ), the configurations we fall into will usually be relatively similar and the conditional coverage probability for this situation F( ) will therefore be nearly constant. All the comments we have made are hence about phenomena observed frequently only in small samples.

Figures 3.10 through 3.13 show the conditional probability of missing the true location parameter at the lower and upper tail for the two nonparametric intervals and the two situations in samples of size 10. It becomes clear now that the two situations are indeed quite different. Both procedures create rather unbalanced intervals in the slash situation -- the Wilcoxon more so than the sign. Many of the Wilcoxon intervals are too long in one direction and often too short in the other. The lines drawn in the pictures correspond to intervals with exact 95% conditional coverage probability. Most of the intervals are below this line -- and are therefore overlong. The fewer points above the line still bring the overall coverage to the nominal 95%. In Gaussian samples the Wilcoxon clearly does quite a good job and is substantially better than the sign intervals, which show a similar behavior as in the slash situation. These plots look similar in samples of size 20. In that case they are, however, a bit more concentrated for both situations.

We can summarize what we have learned up to now by saying that,

Figure 3.10: Plot of the lower vs. the upper conditional missing probabilities of the sign interval in the slash situation

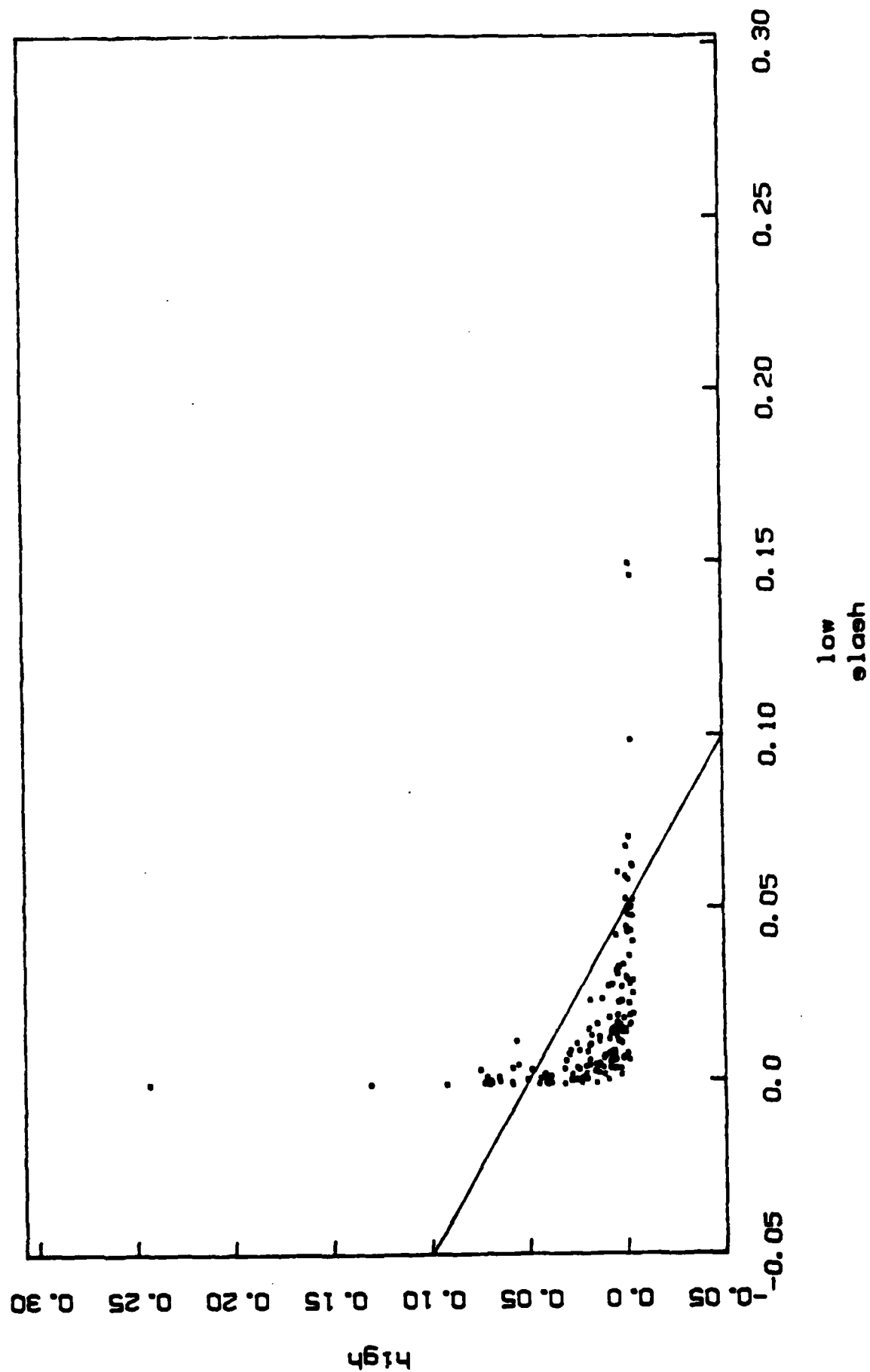sign intervals in samples of size 10

- 42 -

Figure 3.11: Plot of the lower vs. the upper conditional missing probabilities of the
sign interval in the Gaussian situation
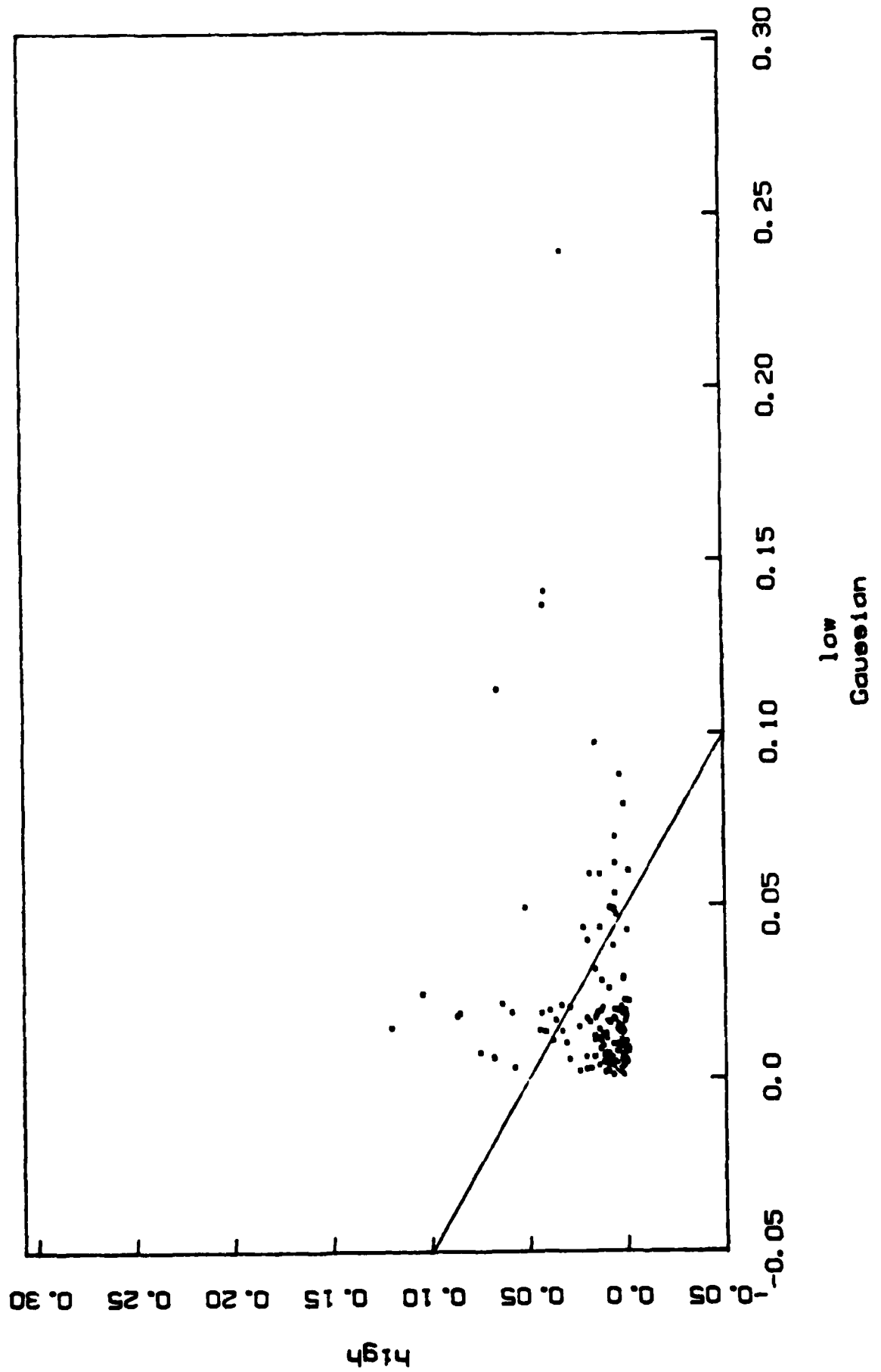
sign intervals in samples of size 10

low
Gaussian

high

Figure 3.12: Plot of the lower vs. the upper conditional missing probabilities of the wilcoxon interval in the slash situation
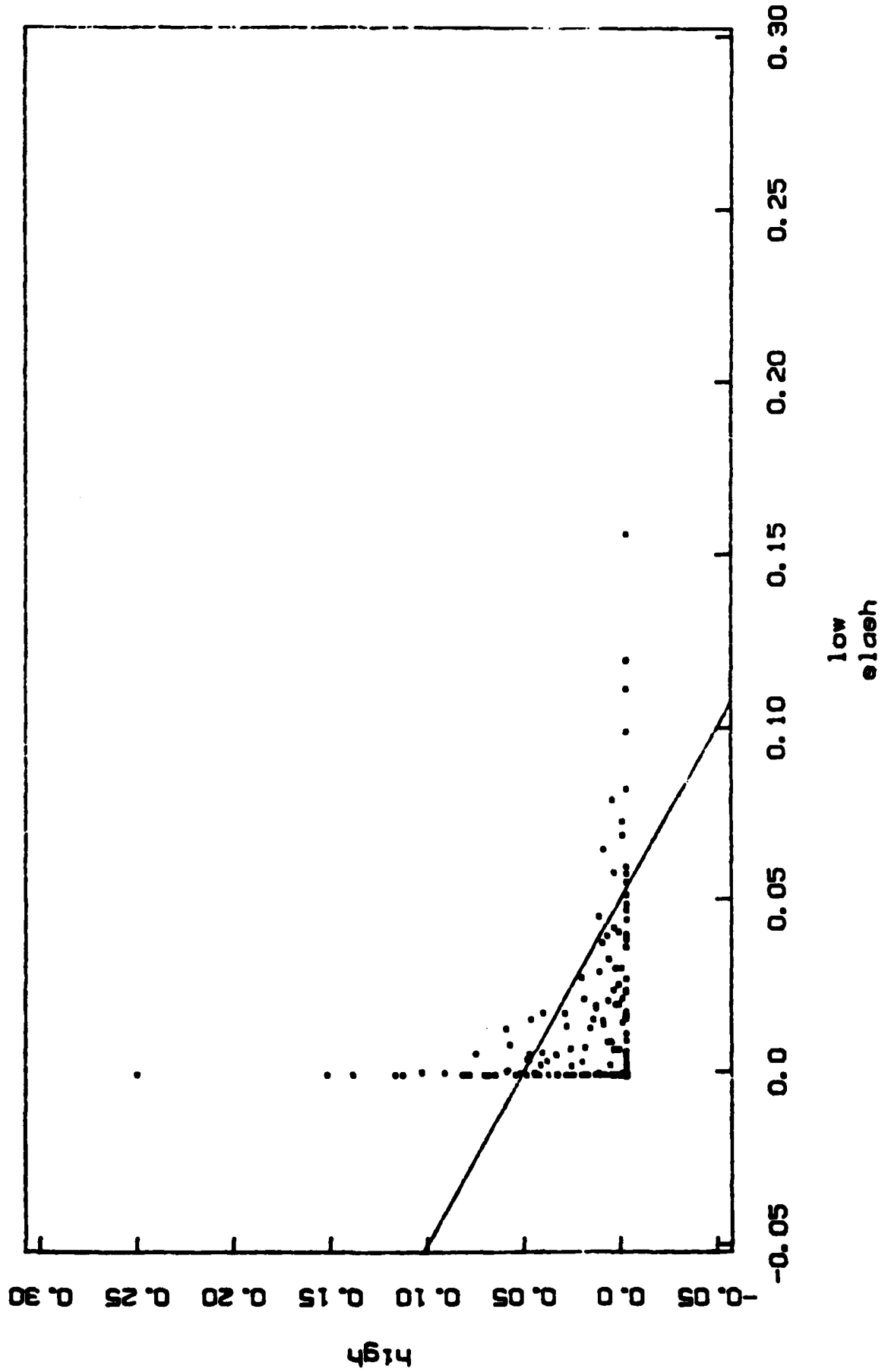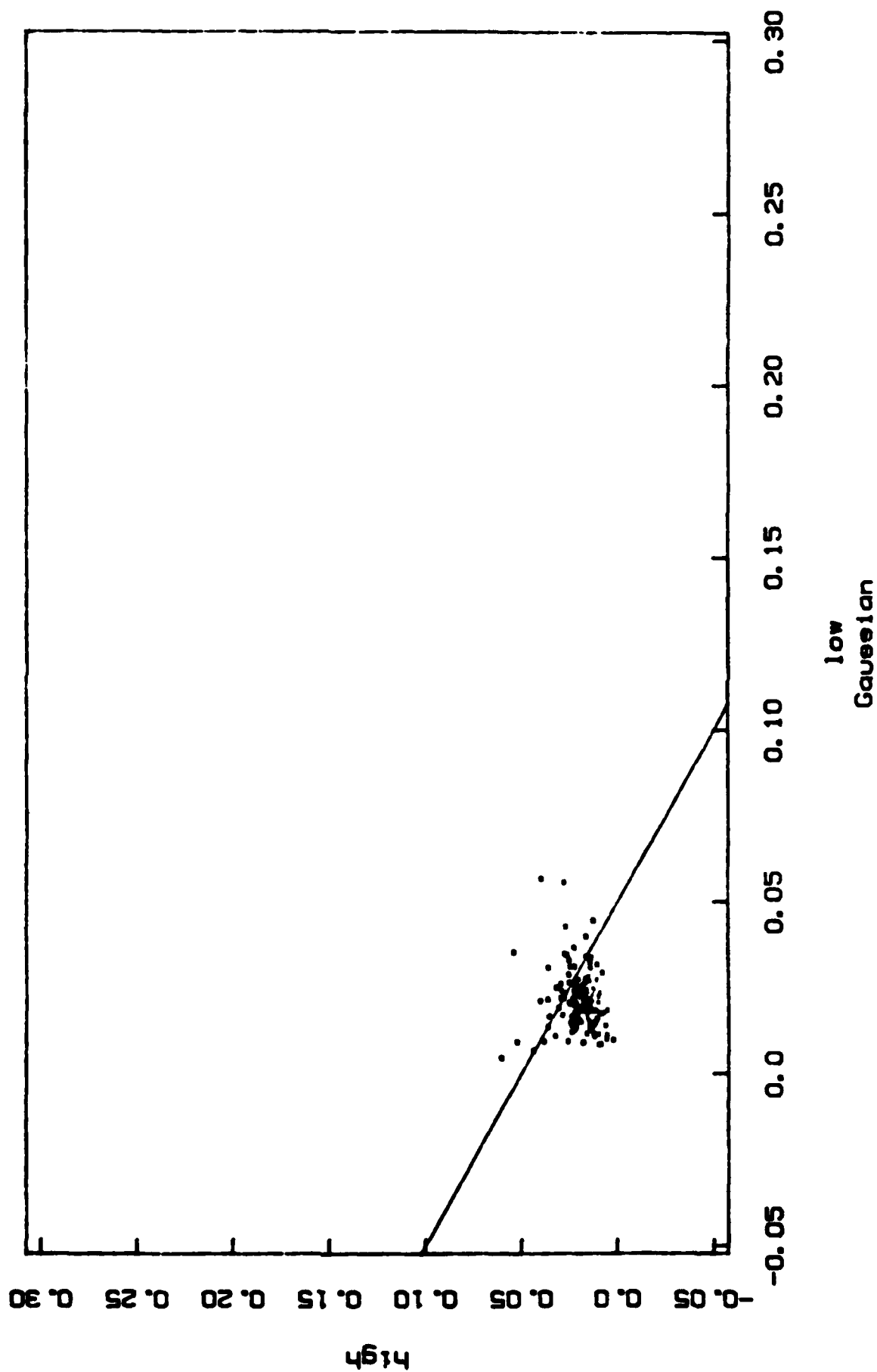
wilcoxon in samples of size 10

Figure 3.13: Plot of the lower vs. the upper conditional missing probabilities of the Wilcoxon interval in the Gaussian situation

wilcoxon in samples of size 10

even though both the sign and the Wilcoxon intervals reach a fixed
confidence level across the whole sample space in all symmetric
situations, there are differences as soon as we look at the
conditional confidence levels. The sign intervals are more stable
than the Wilcoxon intervals if we sample from the heavy-tailed slash.
The opposite is true if we believe strictly in the Gaussian model. As
a compromise between these two extremes, we would probably choose the
Wilcoxon.

The second important thing about confidence intervals besides
coverage probability is the length distribution. Table 3.5 gives the
estimated mean lengths and coefficients of variation.

Table 3.5: Estimated mean length and estimated coefficient of varia-
tion

		Gaussian situation		slash situation	
sign intervals	size=20	1.16	26.5%	2.65	23.3%
	size=10	1.64	29.0%	5.39	75.2%
Wilcoxon intervals	size=20	.955	17.5%	3.34	57.1%
	size=10	1.46	25.0%	24.51	476%

It is quite obvious that the Wilcoxon intervals are extremely long in
extreme configurations of size 10. Otherwise this table confirms our
view that the sign intervals are to be favored in the slash and the
Wilcoxon in the Gaussian. This fact also appears in an asymptotic
theory via Pitman efficiencies. The square root of these efficiencies
applies to asymptotic ratios of mean lengths. The values for the
Gaussian and slash are summarized in Table 3.6.

Table 3.6: Square roots of ratios of Pitman efficiencies

methods	Gaussian	slash
sign vs. Wilcoxon	.8165	1.082
sign vs. Student's t	.7975	∞
Wilcoxon vs. Student's t	.9772	∞

We can check how well these asymptotic numbers work by comparing them to the small sample estimated ratios, as we do in Table 3.7.

Table 3.7: Estimated ratios of mean length

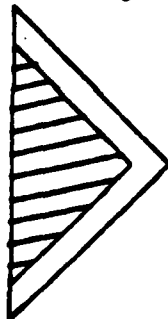|		Gaussian	slash
sign vs. Wilcoxon	size=10	.8901	4.55
	size=20	.8233	1.2604
	asympt	.8165	1.082
sign vs. Student's t	size=10	.8514	13.47*
	size=20	.7966	53.71*
	asympt	.7975	∞
Wilcoxon vs. Student's t	size=10	.9566	15.14*
	size=20	.9675	65.24*
	asympt	.9772	∞

*: theoretical value is ∞

A comparison between the two sets of numbers shows that the asymptotic expression is a good approximation as far as samples of size 20 are concerned, but for sample size 10 the agreement is not good except in "Wilcoxon vs. Student's t".

Both nonparametric intervals we have considered up to now do not leave us entirely happy and a procedure between these two extremes

might well be a better compromise. There are at least two ways in which one can bridge the sign and Wilcoxon. One was suggested by J. W. Tukey and it works by trimming observations from the ordered sample and computing Wilcoxon's signed rank statistic on what is left over, thus omitting certain i and j entirely. The other (Policello and Hettmansperger(1976)) works by "winsorizing" the scores in Wilcoxon's signed-rank statistic. This is equivalent to omitting Walsh averages with $|i-j|$>bound. (If, for example, we put the extreme bound of 1 on the ranks, the sign statistic, i.e. the number of positive observations, comes out.)

Both of these procedures can be explained in terms of the triangle of Walsh averages.
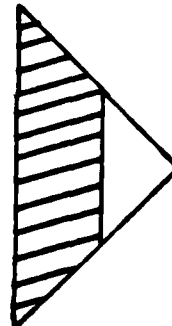
(1) trimmed Wilcoxon   (2) winsorized scores

Both only take the Walsh averages in the corresponding shaded region into account. These methods are expected to do a bit worse in the Gaussian than the Wilcoxon but to improve in the slash.

It turns out that in samples of size 10 the procedure which trims the largest and smallest observation very nearly gives the same confidence intervals as the procedure which puts a bound of 3 on the ranks. Both are worse in the Gaussian situation than the Wilcoxon, which is no surprise -- but they seem to be rather close to the sign

intervals and even a bit worse in terms of stability of the
conditional coverage probabilities. In the slash situation, they are
quite close to the sign intervals. In terms of expected length these
"robust" intervals are an improvement (see Morgenthaler (1983)).
In samples of size 20 the two proposed modifications of the Wilcoxon
test statistic produce nearly identical results. Now these
"robustified" rank intervals really make an improvement. In the
Gaussian situation they are between the Wilcoxon and the sign, in the
slash situation they improve over both of them. Instead of having a
hinge-spread of confidence levels across configurations of roughly 3%
(see Table 3.4), these procedures go down to about 2.5%. Figure 3.14
shows a boxplot similar to the Figures 3.8 and 3.9.
We would therefore recommend the use of these confidence intervals
which bridge the gap between the sign and the Wilcoxon procedures for
larger sample sizes because in a heavy-tailed case they show an
improved behavior.

4. Conclusions.

Strong confidence intervals are appealing from the point of view
of validity of an interval estimator. Their conditional confidence
coefficients given any configuration is kept above the nominal level
for both situations. We only took two situations into consideration,
but this is already a conclusive case. If we safeguard against heavy
tailedness in the way described in this report, we do not have a
difficult job. Student's t interval is already resistant as far as
validity is concerned. But of course many statisticians will
criticize Student's t interval as being too long in configurations

Figure 3.14: Logistic transforms for 150 sampled configurations for three nonparametric intervals in the slash situation

cond. coverage in slash samples of size 20
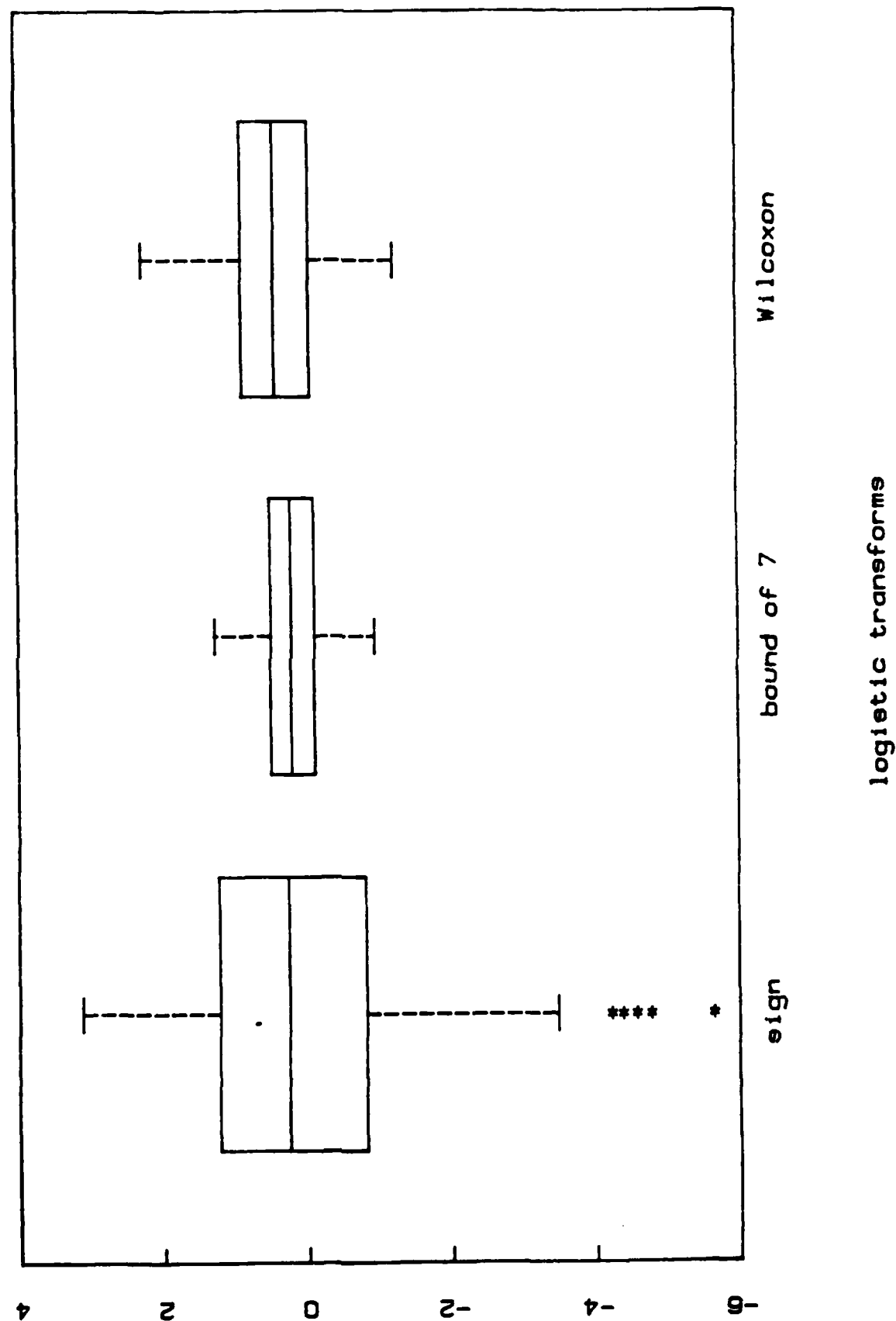
with outliers. The research described shows us the need for compromising between validity and some measure of efficiency. If we only consider validity, we automatically sacrifice efficiency in a rather extreme way. We considered three sample sizes (u = 5, 10 and 20) and it is rather striking how the problem of compromising between shapes changes with changing sample size. For samples of size 5 it turns out that Student's t intervals are nearly "optimal". As the sample size increases this no longer holds. In order to be strong, Student's t interval needs to be enlarged (usually in only one direction).

In the last section of this report we examine non-parametric confidence intervals. It becomes clear that a stable overall confidence coefficient does not ensure a good behavior of the conditional confidence coefficients.

November 15, 1983

REFERENCES

Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J.,
Rogers, W.H. and Tukey, J.W. (1972). Robust
Estimates of Location: Survey and Advances,
Princeton, University Press, Princeton, New Jersey.

Bell Krystinik, K. and Morgenthaler, S. (1981). "Comparison
of the bi-optimal curve with curves for two robust
estimators," Technical Report No. 195, Series 2,
Department of Statistics, Princeton University,
Princeton, New Jersey.

Fisher, R.F. (1934). "Two new properties of maximum likelihood,"
Proc. Royal Society, A, 144, 285-307.

Fraser, D.A.S. (1979). Interence and Linear Models,
McGraw-Hill, New York.

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

Pitman, E. J. G. (1938). "The estimation of the location and scale
parameters of a continuous population of any given torm,"
Biometrika, 30, 391-421.

Policello, G.E., and Hettmansperger, T. P. (1976). "Adaptive robust
        procedures for the one-sample location problem,"
        J. Am. Stat. Assoc., 71, 624-633.

Pregibon, D. and Tukey, J. W. (1981). "Assessing the behavior
        of robust estimates of location in small samples:
        Introduction to configural polysampling," Technical Report
        No. 185, Series 2, Department of Statistics, Princeton
        University, Princeton, New Jersey.

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

Tukey, J. W. (1977). Exploratory Data Analysis, Addison-Wesley,
        Reading, Massachusetts,

END

FILMED

8