Parameter estimation for storm response distributions using fiducial inference

Technische
Universiteit
Eindhoven

Opleiding Wiskunde voor de Industrie Eindhoven

STUDENT REPORT 90-05

PARAMETER ESTIMATION FOR STORM RESPONSE DISTRIBUTIONS USING FIDUCIAL INFERENCE

P.C.J. Jeurissen
J.G. Kester
J.P.M. Laumen

March 1990

ECMI

Den Dolech 2
Postbus 513
5600 MB Eindhoven

PARAMETER ESTIMATION FOR STORM RESPONSE DISTRIBUTIONS USING FIDUCIAL INFERENCE

By
P.C.J. Jeurissen
J.G. Kester
J.P.M. Laumen

Eindhoven,
march 1990

PREFACE

This report describes our work on a project called "the storm", which we carried out within the compass of the modelling colloquium of the postgraduate training "Mathematics for Industry" at the Technical University of Eindhoven. We would like to thank Stef van Eijndhoven and Nico Linssen for their help and advise concerning this project.

This account consists of three parts. In the first part we give an introduction to the problem and the main conclusions. In the second part there is a description of the mathematics which lead to these conclusions. The third part consist of two appendices, which contain the numerical results and some more detailed information about the mathematical method we used: fiducial inference. References to the bibliography are indicated by square brackets.

CONTENTS

0. INTRODUCTION

Offshore structures like drilling platforms have to be strong enough to ride out heavy storms. Both the behaviour of the sea during a storm and hence the response of the platform will have a random character. However, even if the distribution of the sea is known, the response of the platform is too complicated to determine its distribution analytically. Therefore the behaviour of severe storms, and in particular the effect they have on a platform is simulated on a computer using realistic models. We want to use those simulations to estimate the distribution of the maximal response. With this estimate we calculate with what probability the maximal response of the platform will be greater than a certain critical value. Another purpose is to diminish the computation time, hence we want to examine how many observations are needed for a good estimate. More specifically, we want to estimate for a 3-hour storm, using only the observations from an L-hour simulation, with $L \leq 3$.

1. MAIN CONCLUSIONS

For a given simulation of a 1-hour storm we can estimate the α-percentage y_α corresponding to the maximal response of a drilling platform during a 3-hour storm. Further, we can give confidence intervals for the y_α. What does this all mean?

Suppose that we have a critical value for the maximal response of the drilling platform. Now we can find what α corresponds to this value, or in other words, what percentage α of 3-hour storms yield a maximal response of the platform that remains smaller than the critical value. Moreover we can say something about the reliability of this statement by means of the confidence intervals for y_α. (See also the figures from Appendix A.8, A.15 and A.22.)

2. PROBLEM DEFINITION

In this section we will give some definitions and state the problem in a mathematical way. We denote stochastic variables by underlined symbols, while fiducial variables can be recognized by a dashed superscript. Further, probability densities (distributions) will be denoted by small (capital) p's, whereas fiducial densities (distributions) will be represented by small (capital) f's. Finally, all following notations will be equivalent.

$$P(x) = P_{\underline{x}}(x) = P(\underline{x} \le x) = P(\underline{x} \le x\,;\theta) = P_{\underline{x}}(x\,;\theta) = P(x\,;\theta)$$

$$F(\theta) = F_{\bar{\theta}}(\theta) = F_{\bar{\theta}}(\theta\,;x) = F(\theta\,;x)\,.$$

Here θ is a parameter, and we shall use the simplest notations as possible.

The behaviour of the sea and the response of the platform are stochastic processes. Let $\underline{r}(t)$ represent the stochastic response of the drilling platform at time t. Then we define the maximal response $\underline{x}^T$ over a time period with length T as

$$\underline{x}^T = \max_{0 \le t \le T} \underline{r}(t)\,.$$

Since we can assume that $\underline{r}(t)$ has the same distribution at any time t during the storm, we can take the starting time anywhere we want to, as long as the interval $[0,T]$ lies entirely within the storm period.

Define the probability distribution of $\underline{x}^T$ by

$$P^T(x) := \text{Prob}\,(\underline{x}^T < x)\,.$$

Now the problem can be formulated as follows. Estimate the distribution $P^3(x)$ of the 3-hour maximum $\underline{x}^3$, using only the observations of an L-hour simulation. The value of $L \le 3$ should be chosen small but effective.

3. ASSUMPTIONS AND MODELLING

Let the L-hour period be subdivided into n non-overlapping periods of length $l = L/n$. We assume that $\underline{x}^l$ is the maximum of a large number of independent observations. This means that the stochastic process $\underline{r}(t)$ peaks a sufficiently large number of times in a period of length l and that the peak values are independent. This is called the stability assumption.

Further we assume that the individual observations $r(t_1), r(t_2), \ldots, r(t_m)$ (with $t_0 < t_1 < t_2 < \cdots < t_m < t_0 + l$) all arise from a distribution for which all moments exists. This choice does not seem to be too restrictive.

If $\underline{x}^l = \max_{1 \le j \le m} \underline{r}(t_j)$, then the limiting distribution

$$\lim_{m \to \infty} P^l(x) = G^l(x) \equiv \exp(-\exp\left[-\frac{x-\mu}{\sigma}\right]) \quad (-\infty < x < \infty) \tag{3.1}$$

is called the Gumbel distribution with location parameter μ and scale parameter σ, both depending on l (cf. [1] and [3]). In our case we can assume m to be large enough to state that $P^l(x) = G^l(x)$, that is, we have enough observations.

Furthermore we assume that the observations for the different non-overlapping time periods are independent so that

$$P^3(x) = (G^l(x))^{3/l}\ . \tag{3.2}$$

This means that if we estimate the Gumbel parameters μ and σ from (3.1), we have found an estimate for the distribution $P^3(x)$ by (3.2).

Note that we have n observations to estimate μ and σ from , namely the n maxima over the subperiods of length l. This means that we would like to take n as large as possible, in order to enhance the accuracy of the estimates. However, a larger value of n means a smaller value of l, which will in the long run lead to a contradiction with the stability assumption. So we have to be careful in choosing the value of n.

4. SOLUTION METHOD

We are interested in the distribution $P^3(x)$, and more specifically, in its α-percentage y_α, defined by

$$P^3(y_\alpha) = \alpha \,,$$

i.e., the maximal response during a 3-hour storm will be less than or equal to y_α with a probability α. We can solve y_α from this definition using equations (3.1) and (3.2). This yields

$$\begin{cases} y_\alpha = \mu + c_\alpha \cdot \sigma \\ c_\alpha = -\ln(-l/3 * \ln \alpha) \,. \end{cases} \tag{4.1}$$

We will calculate these percentage points with fiducial inference. With this method we can not only obtain estimates $\hat{y}_\alpha$ for y_α, but also a q-confidence interval $[L_\alpha, U_\alpha]$ which contains the real value of y_α with probability q.

The method of fiducial inference is described in more detail in Appendix B. In broad outline we can say that the probability of an observation x given a parameter θ, is transformed into a fiducial density $f(\theta\,; x)$, denoting the fiducial probability of a parameter θ given an observation x. For more observations or parameters analogous descriptions hold.

Let x_i for $i = 1 \cdots n$ be realizations of $\underline{x}^l$ obtained from a simulation of n consecutive time periods of length l.

With the technique described in Appendix B we can calculate the joint fiducial density $f_{\bar{\mu},\bar{\sigma}}(\mu,\sigma)$ for the parameters μ and σ from the Gumbel distribution. This yields

$$\begin{cases} f_{\bar{\mu},\bar{\sigma}}(\mu,\sigma) \approx \dfrac{1}{\sigma^{n+1}} \exp\left(\sum_{i=1}^{n} \phi\left(-\dfrac{\mu - x_i}{\sigma}\right)\right) \\ \phi(t) = -t - e^{-t} \,. \end{cases} \tag{4.2}$$

The symbol ≈ means "is proportional to", i.e. (4.2) determines the fiducial density $f_{\bar{\mu},\bar{\sigma}}(\mu,\sigma)$ up to a constant β. This constant β can be calculated from the requirement that the total fiducial probability, obtained by integrating $f_{\bar{\mu},\bar{\sigma}}(\mu,\sigma)$ over μ and σ, has to be equal to 1.

Since we are interested in the fiducial density of $\bar{y}_\alpha = \bar{\mu} + c_\alpha \cdot \bar{\sigma}$, we transform $f_{\bar{\mu},\bar{\sigma}}(\mu,\sigma)$ to

$$f_{\bar{\mu}+c_\alpha\bar{\sigma},\bar{\sigma}}(\mu + c_\alpha\sigma,\, \sigma) = f_{\bar{y}_\alpha,\bar{\sigma}}(y_\alpha,\, \sigma) \,.$$

If we denote this transformation by

$$\begin{cases} \bar{x}_1 = \bar{\mu} + c_\alpha \cdot \bar{\sigma} \\ \bar{x}_2 = \bar{\sigma} \end{cases}$$

then the Jacobian matrix J is given by

$$J = \begin{bmatrix} \frac{\partial \bar{x}_1}{\partial \bar{\mu}} & \frac{\partial \bar{x}_1}{\partial \bar{\sigma}} \\ \frac{\partial \bar{x}_2}{\partial \bar{\mu}} & \frac{\partial \bar{x}_2}{\partial \bar{\sigma}} \end{bmatrix} = \begin{bmatrix} 1 & c_\alpha \\ 0 & 1 \end{bmatrix}$$

and hence $\det(J) = 1$.

This leads to the following relation.

$$f_{\bar{y}_\alpha, \bar{\sigma}}(y_\alpha, \sigma) = \det(J) * f_{\bar{\mu}, \bar{\sigma}}(y_\alpha, \sigma) = f_{\bar{\mu}, \bar{\sigma}}(y_\alpha, \sigma) .$$

Integration of $f_{\bar{y}_\alpha, \bar{\sigma}}(y_\alpha, \sigma)$ with respect to σ yields the marginal fiducial density $f_{\bar{y}_\alpha}(y_\alpha)$.

$$f_{\bar{y}_\alpha}(y_\alpha) = \int_0^\infty f_{\bar{y}_\alpha, \bar{\sigma}}(y_\alpha, \sigma)\, d\sigma . \tag{4.3}$$

Finally, in order to get an estimate $\hat{y}_\alpha$ and to obtain the lower- and upperbound L_α and U_α of the q-confidence interval for y_α, we have to solve the following integral equations.

$$\begin{cases} \int_{-\infty}^{\hat{y}_\alpha} f_{\bar{y}_\alpha}(s)\, ds = 0.5 \\ \int_{-\infty}^{L_\alpha} f_{\bar{y}_\alpha}(s)\, ds = \frac{1-q}{2} \\ \int_{-\infty}^{U_\alpha} f_{\bar{y}_\alpha}(s)\, ds = 1 - \frac{1-q}{2} . \end{cases} \tag{4.4}$$

Let us summarize the required calculations. Combining (4.2), (4.3) and (4.4) gives us the following set of equations to be solved for β, $\hat{y}_\alpha$, L_α and U_α.

$$\int_0^\infty \int_{-\infty}^\infty \frac{1}{\sigma^{n+1}} * \exp \sum_{i=1}^n \left\{ \frac{\mu - x_i}{\sigma} - \exp\left[\frac{\mu - x_i}{\sigma} \right] \right\} d\mu\, d\sigma = 1/\beta$$

$$\int_0^\infty \int_{-\infty}^{\hat{y}_\alpha} \frac{\beta}{\sigma^{n+1}} * \exp \sum_{i=1}^n \left\{ \frac{\mu - x_i}{\sigma} + c_\alpha - \exp\left[\frac{\mu - x_i}{\sigma} + c_\alpha \right] \right\} d\mu\, d\sigma = 0.5 \tag{4.5}$$

$$\int_0^\infty \int_{-\infty}^{L_*} \frac{\beta}{\sigma^{n+1}} * \exp \sum_{i=1}^{n} \left\{ \frac{\mu - x_i}{\sigma} + c_\alpha - \exp\left[\frac{\mu - x_i}{\sigma} + c_\alpha \right] \right\} d\mu\, d\sigma = \frac{1-q}{2}$$

$$\int_0^\infty \int_{-\infty}^{U_*} \frac{\beta}{\sigma^{n+1}} * \exp \sum_{i=1}^{n} \left\{ \frac{\mu - x_i}{\sigma} + c_\alpha - \exp\left[\frac{\mu - x_i}{\sigma} + c_\alpha \right] \right\} d\mu\, d\sigma = 1 - \frac{1-q}{2} .$$

In the next section the equations (4.5) will be solved numerically. Moreover, the influence of the ancillary statistics will be investigated.

In Appendix B ancillary statistics are defined as stochastic variables whose probability distribution does not depend on the parameters. In our case we have n observations $x_1, \ldots, x_n$ to estimate the parameters μ and σ. However, in general the observations have to be sufficiently reduced. So every ancillary statistic is a function of the minimal sufficient statistic. In this case, the observations themselves are minimal sufficient. Therefore we have $n-2$ ancillary statistics. It is interesting to know in what way the values of these ancillary statistics influence the calculated values $\hat{y}_\alpha$, L_α and U_α.

5. NUMERICAL ANALYSIS

This section is devoted to the numerical computation of the integrals given in (4.5).
We shall first give a general method for computing two-dimensional integrals. Before applying this method we will rewrite the integrals. Thus we get much better results.
Let us, for sake of clarity, write down once more the integrals to be calculated. In our calculations we will take $L = 1$ and $n = 4$, i.e., we will use one hour of simulation and the computations are based on four samples. The choice of the value of n is based on experience in order to make a trade-off between the stability assumptions and the accuracy of the estimate as described in Section 3. The integrals become

$$\int_0^\infty \int_{-\infty}^\infty \sigma^{-5} \exp\{\sum_{i=1}^4 \left[\frac{\mu - x_i}{\sigma} - \exp\left[\frac{\mu - x_i}{\sigma}\right]\right]\}\, d\mu\, d\sigma \tag{5.1}$$

and

$$\int_0^\infty \int_{-\infty}^{ub} \sigma^{-5} \exp\{\sum_{i=1}^4 \left[\frac{\mu - x_i}{\sigma} + c(\alpha) - \exp\left[\frac{\mu - x_i}{\sigma} + c(\alpha)\right]\right]\}\, d\mu\, d\sigma \tag{5.2}$$

where

$$c(\alpha) = -\ln \ln(1/\alpha)^{l/3} \tag{5.3}$$

and ub (upperbound) is equal to $\hat{y}_\alpha$, L_α or U_α.

Furthermore, the integral in (5.2) will be denoted by

$$I(ub, \alpha; x1, x2, x3, x4)\,. \tag{5.4}$$

We will now describe the method by which we compute the integrals in (5.1) and (5.2). Suppose that the function f is "smooth enough", then

$$I = \int_{-h}^{h} \int_{-h}^{h} f(x,y)\, dx\, dy \approx I_h \tag{5.5}$$

where

$$I_h = \frac{h^2}{3}\, [8f(0,0) + f(h,h) + f(h,-h) + f(-h,h) + f(-h,-h)]\,. \tag{5.6}$$

Without proof we state (cf. [5])

$$I - I_h = \frac{-8h^6}{72}\left[\frac{1}{5}\left\{\frac{\partial^4 f(0,0)}{\partial x^4} + \frac{\partial^4 f(0,0)}{\partial y^4}\right\} + 2\,\frac{\partial^4 f(0,0)}{\partial x^2 \partial y^2}\right] + O(h^8) \quad (h \to 0) \tag{5.7}$$

When applied to a rectangle which is subdivided into squares of length $2h$, this method leads to an integration scheme as depicted in Figure 5.1. In this figure we denoted the number of times that the function value in that knot occurs in the integration formula.

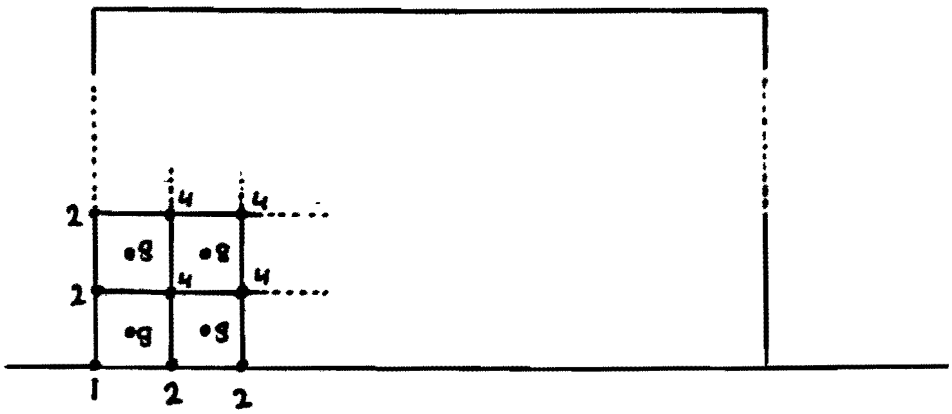


Figure 5.1. Integration over a rectangle

We now want to apply this method to the integrals as given in (5.1) and (5.2). Because the numerical calculations of the integrals in (5.1) and (5.2) are similar we will first look only at the integral in (5.1).
As we shall see later on, the values of the variables $x1, \ldots, x4$ are in the interval [500, 1000]. Also the integrand decreases very slowly in the σ-direction. Because of these facts, application of (5.6) to (5.1) takes a long computing time and does not give very satisfactory results. One way to overcome these problems is the following; suppose $x1 \le x2 \le x3 \le x4$ and $x1 < x4$. The number k is defined by

$$k = x4 - x1 . \tag{5.8}$$

The variables $y1, y2, y3, y4, \mu'$ and γ are now defined by

$$\begin{array}{ll} a & y1 = 0 \\ b & y2 = (x2 - x1)/k \\ c & y3 = (x3 - x1)/k \\ d & y4 = 1 \\ e & \mu' = (\mu - x1)/k \\ f & \gamma = k/\sigma . \end{array} \tag{5.9}$$

Note that $y0 = 0 \le y2 \le y3 \le 1 = y4$, and that $y2$ and $y3$ are ancillary statistics. Using (5.9) the integral in (5.1) transforms into (where the prime of μ' is omitted)

$$\frac{1}{k^3} \int_0^\infty \gamma^3 \, d\gamma \int_{-\infty}^{\infty} \exp[\sum_{i=1}^{4} \gamma(\mu - yi) - \exp\{\gamma(\mu - yi)\}] \, d\mu . \tag{5.10}$$

Because the integral has infinite integration intervals we have to find a rectangle in the (μ,γ)-plane such that the contribution of the integral over the complement of this rectangle is very small (see Figure 5.2).

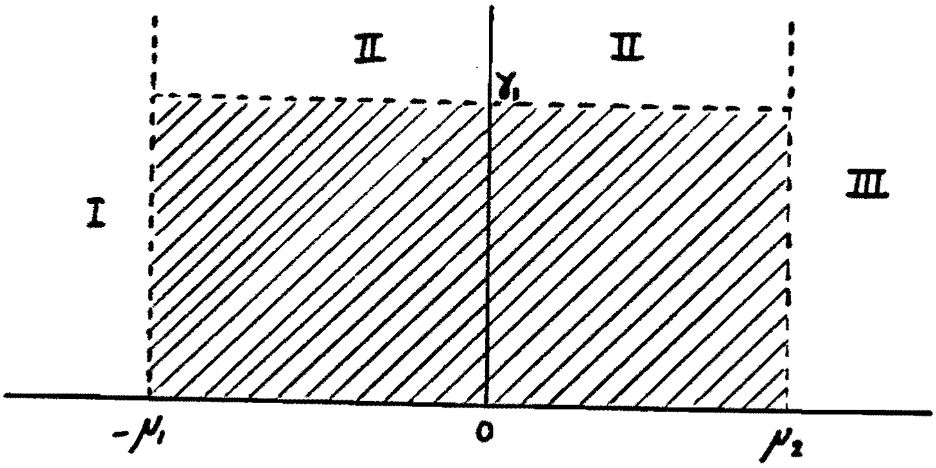


Figure 5.2. Numerical integration area.

In order to get an impression of the values of μ_1, μ_2 and γ_1 we can make the following estimates for the regions I, II and III. For region I we obtain the integral

$$\frac{1}{k^3}\int_0^{\infty}\gamma^3\,d\gamma\int_{-\infty}^{-\mu_1}\exp\left[\sum_{i=1}^{4}\gamma(\mu-yi)-\exp\{\gamma(\mu-yi)\}\right]d\mu\,. \tag{5.11}$$

To estimate this integral we first look at the graph of the function $g(t)$, which is defined by

$$g(t)=t-e^t \quad -\infty<t<\infty\,, \tag{5.12}$$

(see Figure 5.3).

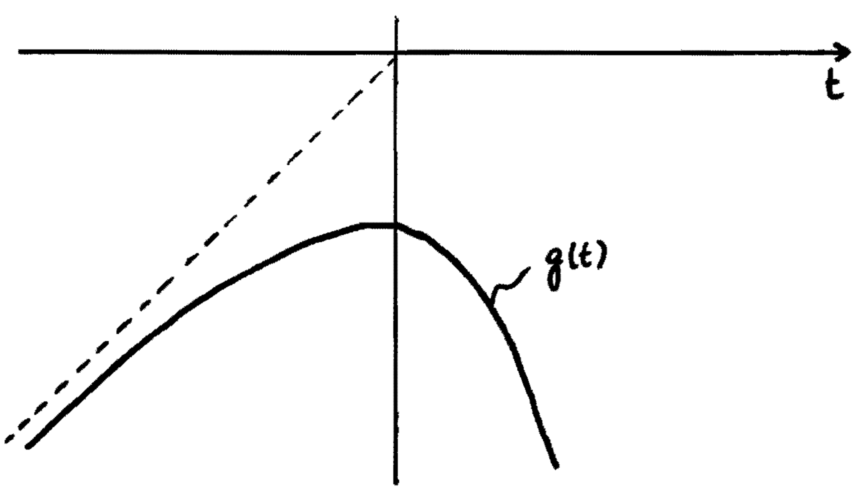


Figure 5.3. Graph of the function $g(t)$.
For $t<0$ almost linear increase and
for $t>0$ almost exponential decrease.

From Figure 5.3 we obtain the inequality

$$\frac{1}{k^3}\int_0^\infty \gamma^3 d\gamma \int_{-\infty}^{-\mu_1} \exp\left[\sum_{i=1}^{4} \gamma(\mu-yi) - \exp\{(\gamma(\mu-yi))\}\right] d\mu \le \tag{5.13}$$

$$\le \frac{1}{k^3 a^3}\int_0^\infty \exp\left[4\{-\omega - \exp(-\omega)\}\right] d\omega .$$

From expression (5.13) it follows that this integral is small if μ_1 takes approximately the value 100. For μ_2 and γ_1 we can make similar estimates and we find $\mu_2 = 100$ and $\gamma_1 = 80$ ($k \approx 50$).
As may be expected, the region over which the integral (5.10) must be integrated numerically to get a good approximation, is still smaller than the rectangle we indicated above. It turns out that the integration region has the form as drawn in Figure 5.4.

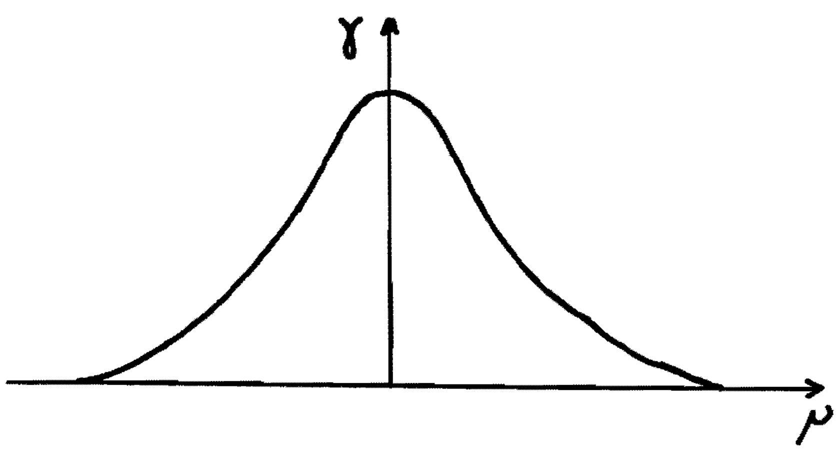


Figure 5.4. Integration area such that numerical integration gives good results.

Without going into details we mention that similar transformations and estimates can be made for the integral (5.2).
In the next section we will look at numerical results obtained by using the above method for computing the integral.

6. NUMERICAL RESULTS

In this section we will give some numerical results on

i) estimates and confidence intervals for $\hat{y}_\alpha$, $0.1 \leq \alpha \leq 0.75$,

ii) the influence of the ancillary statistic.

In order to obtain these numerical results we need the values of the response maxima, i.e., the numbers $x1, x2, x3$ and $x4$. In Table 6.1 5-minutes maxima are given for a period of length 3 hours. These numbers were obtained by simulation.

first hour	second hour	third hour
674.0	634.7	700.7
653.2	649.6	756.9 *
830.9 *	687.4 *	673.7
712.8 *	655.0	716.8 *
608.2	661.0	691.9
669.1	696.9 *	715.4
622.3	703.6 *	702.9
787.6 *	654.4	639.4
641.3	665.1	704.2 *
638.4	622.8	696.3
716.6 *	673.5	691.1
676.1	706.2 *	723.4 *

Figure 6.1. 5-minutes maxima for a time period of 3 hours. The 15-minutes maxima are indicated.

The values of α for which $\hat{y}_\alpha$ and the corresponding confidence intervals will be calculated are $\alpha = 0.1, 0.25, 0.40, 0.50, 060$ and 0.75.

For given ξ, $i = 1, \ldots, 4$, and given α the integral in (5.2) depends only on the parameter β. The bound $y_{\alpha,\beta}$ is then defined as follows

$$I(\beta, \alpha; x1, x2, x3, x4) = \beta\,. \tag{6.1}$$

Using this notation we have the following

$$\begin{cases} \hat{y}_\alpha = y_{\alpha,0.5} \\ 80\% \text{ confidence interval for } \hat{y}_\alpha \quad (y_{\alpha,0.1}\,,\, y_{\alpha,0.9})\,. \end{cases} \tag{6.2}$$

The results for $\hat{y}_\alpha$ the corresponding confidence intervals are given in Appendix A (for the three consecutive hours). The curves in all the figures are obtained by spline interpolation. The figures of the fiducial distributions have on the x-axis the bound $y_{\alpha,\beta}$ (of the integral (5.2)) and on the y-axis the corresponding β times 100. As we can see from (4.1) y_α increases with increasing α. This feature also follows from the pictures.

The figures of the 80% confidence intervals have on the x-axis the bound $y_{\alpha,\beta}$ and on the y-axis the corresponding α. These figures are easily obtained using the figures of the fiducial distributions.

Next we will look at the influence of the ancillary statistics. Because all 4-tuples $(x1, \ldots, x4)$ are scaled to $(0, y2, y3, 1)$ in the computer program $(0 \leq y2 \leq y3 \leq 1)$ we will choose $(x1, \ldots, x4)$ only in such a way that $0 = x1 \leq x2 \leq x3 \leq x4 = 1$. We calculated the 80% confidence intervals in the following cases:

$x1$	$x2$	$x3$	$x4$
0	0	0	1
0	0	½	1
0	0	1	1
0	½	½	1
0	½	1	1
0	1	1	1

The graphs of these calculations are given in Appendix A. From these figures we can clearly see the influence of the ancillary statistics (the ancillary statistics depend on the numbers $x1, \ldots, x4$, see also Appendix B). To what extent the ancillary statistics influence the $\hat{y}_\alpha$ and the confidence intervals can be calculated as follows. As a reference state we choose $(x1, x2, x3, x4) = (0, \frac{1}{2}, \frac{1}{2}, 1)$ and $\alpha = 0.5$. Looking at the bounds of the 80% confidence interval in this case we find $(0.9, 2.9)$. Next we calculate the integral in (5.2) but with the lower and upper bound of the second integral replaced by 0.9 and 2.9 resp. The value of this integral is the confidence level of the interval $(0.9, 2.9)$ corresponding to given $(x1, \ldots, x4)$. The results of these calculations can be seen in Table 6.2.

$x1$	$x2$	$x3$	$x4$	confidence level
0	0	0	1	53 %
0	0.25	0.25	1	67 %
0	0	0.5	1	73 %
0	0.5	0.5	1	81 %
0	1	1	1	72 %

Table 6.2. Confidence levels of the interval (0.9, 2.9) for different ancillary statistics

From Table 6.2 it is clear that the values of the ancillary statistics certainly influence the estimates and corresponding confidence intervals for the number $\hat{y}_\alpha$. This result also implies that the influence of the ancillary statistics on the conditional confidence level is substantial.

7. CONCLUSIONS

We have seen that it is possible to calculate the α-percentage point $\hat{y}_\alpha$ for the 3-hour storm response maximum using just four samples from one hour of simulation. Furthermore we also can compute confidence intervals for y_α. It turns out that for α approaching 100% it becomes more and more difficult to determine $\hat{y}_\alpha$, L_α and U_α numerically.
This can be explained from the fact that we use only four samples to get information on α% of the 3-hour storms. If nevertheless we would like to obtain $\hat{y}_\alpha$, L_α and U_α for large α, we can do the following. First, we could split up the hour of simulation into more than four intervals. However, this may violate the stability assumption. Second, we could use more than one hour of simulation.

Finally we showed that the ancillary statistics do influence the numerical results. Hence the method of fiducial inference proves its usefulness, since most other methods, like, e.g., *ML*-estimations, neglect this information.

APPENDIX A: FIGURES

Results for first hour simulation

Results for second hour simulation

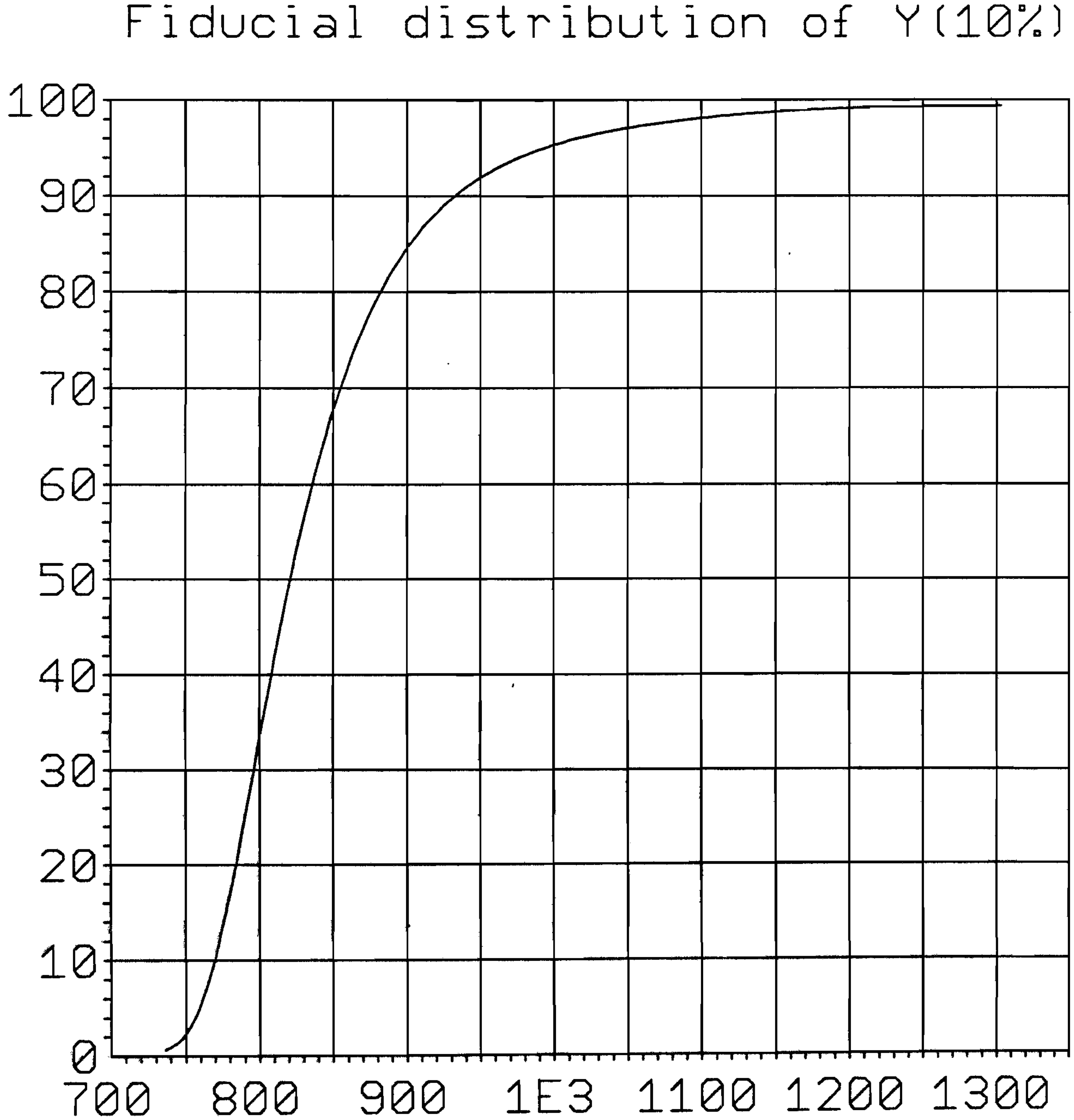
Results for third hour simulation

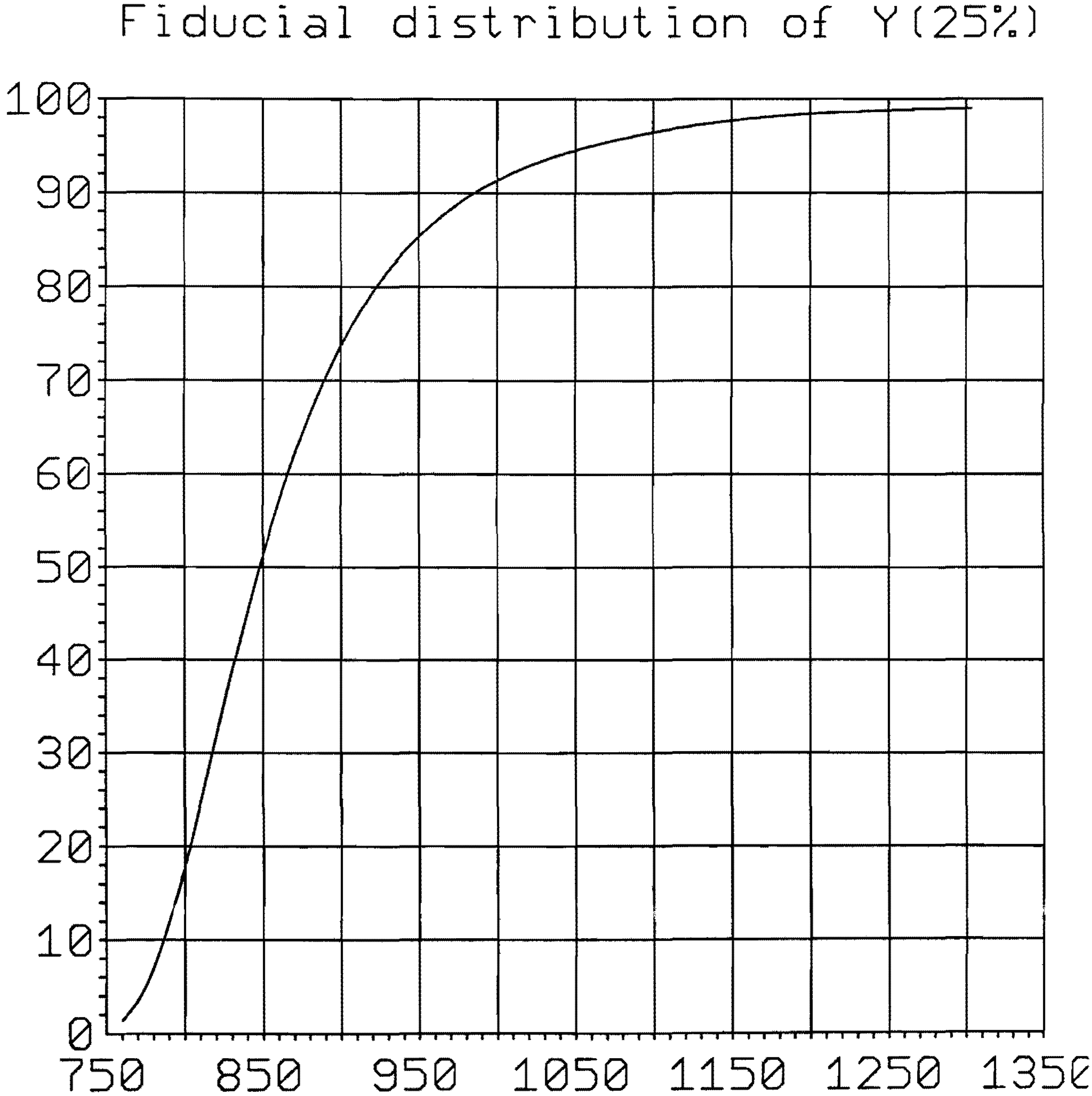
Influence of the ancillary statistic

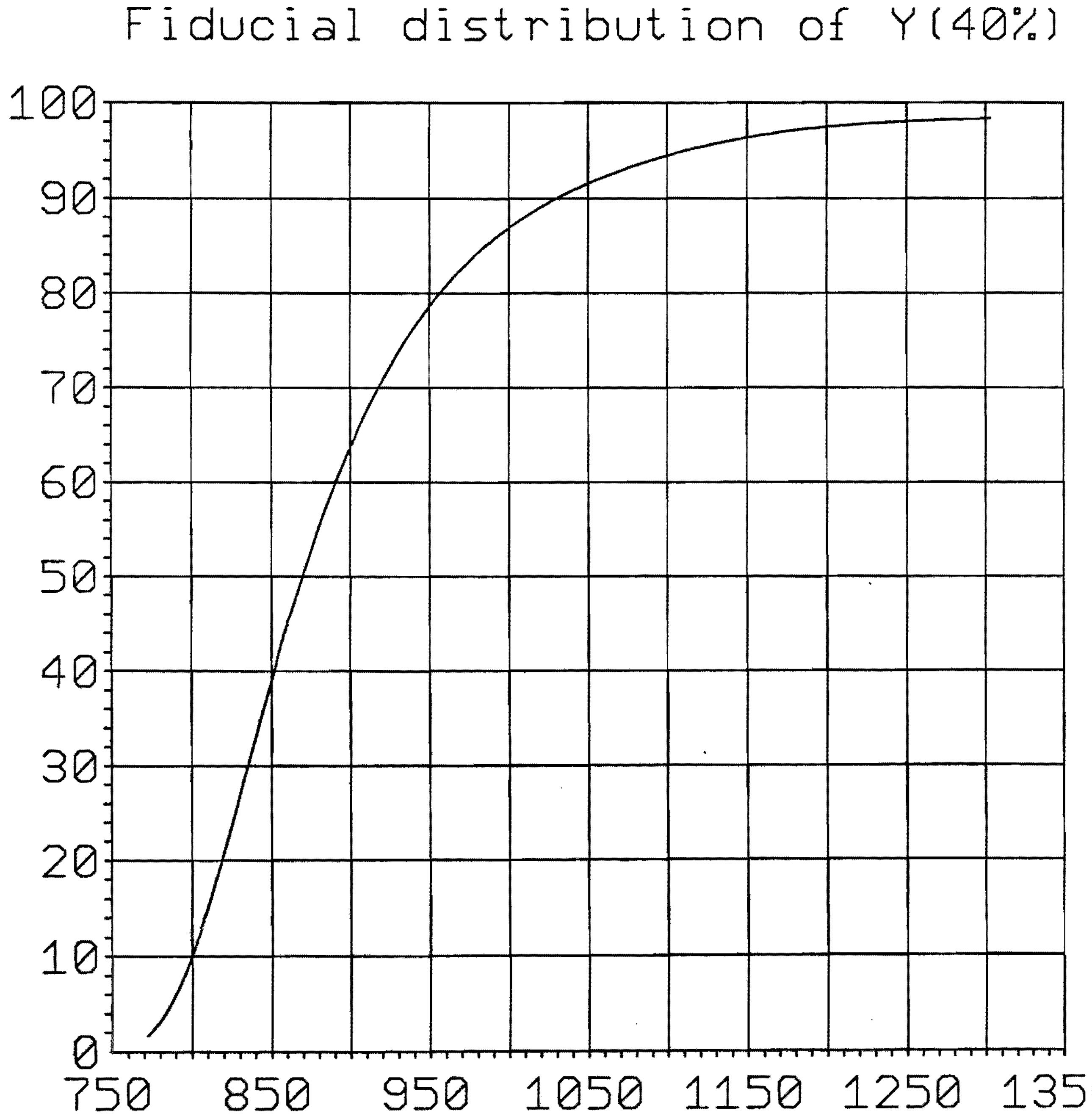
Distribution of tempest maxima

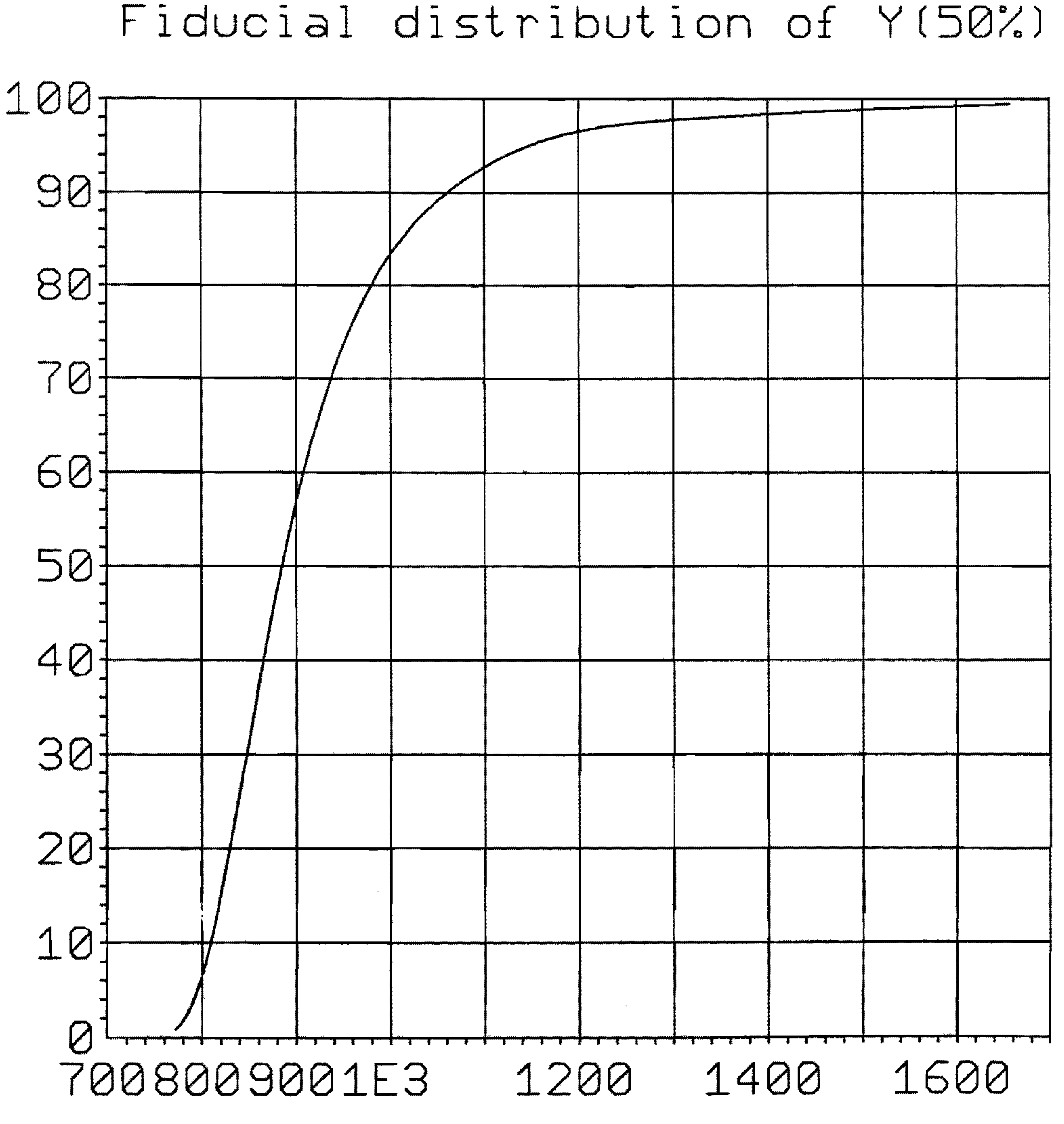
80 %-confidence intervals

$x_1, x_2, x_3, x_4 = 0, 0, 0, 1$	A23
$x_1, x_2, x_3, x_4 = 0, 0, ½, 1$	A24
$x_1, x_2, x_3, x_4 = 0, 0, 1, 1$	A25
$x_1, x_2, x_3, x_4 = 0, ½, ½, 1$	A26
$x_1, x_2, x_3, x_4 = 0, ½, 1, 1$	A27
$x_1, x_2, x_3, x_4 = 0, 1, 1, 1$	A28

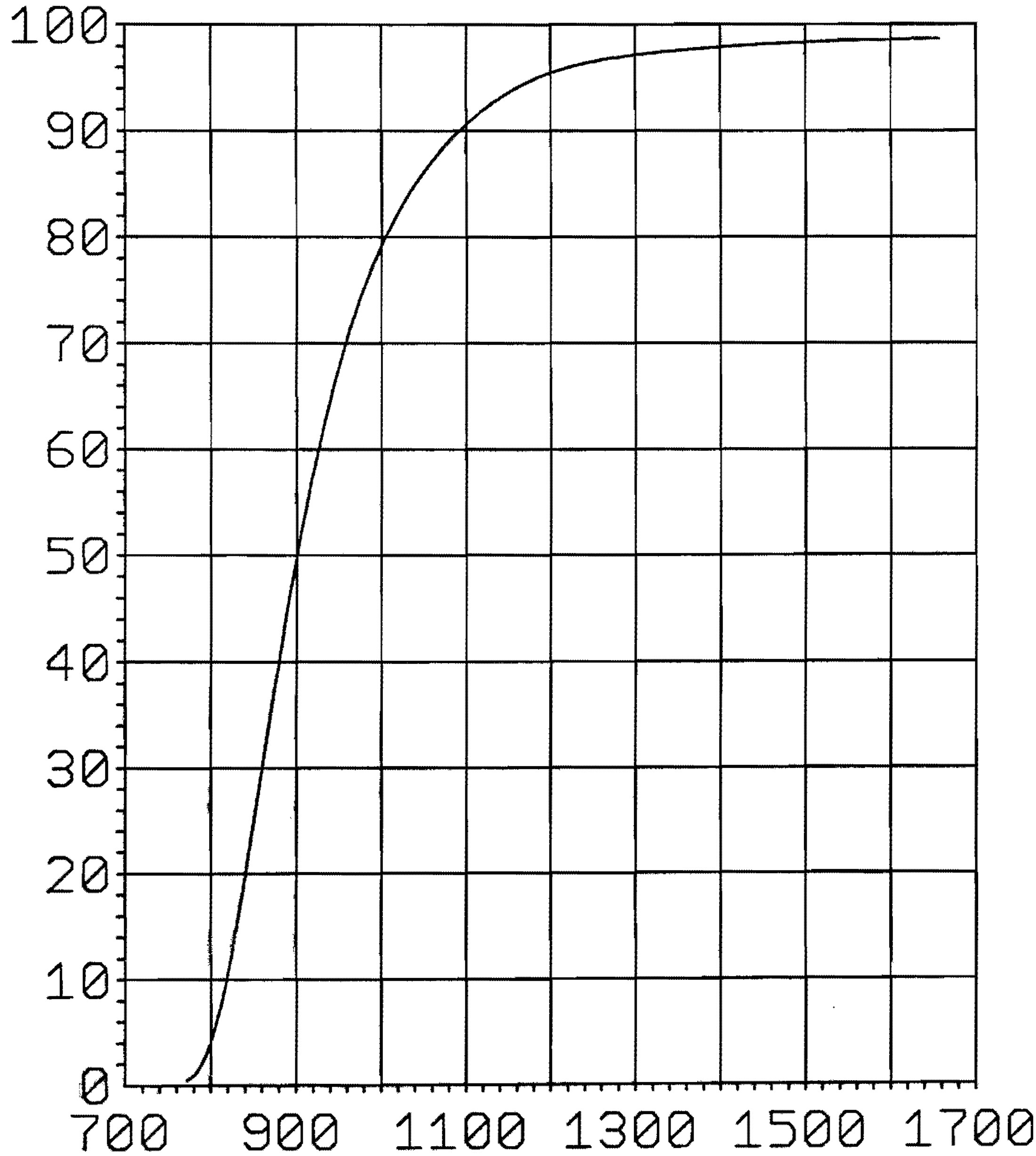

Fiducial distribution of Y(10%)
100
90
80
70
60
50
40
30
20
10
0
700
800
900
1E3
1100
1200
1300

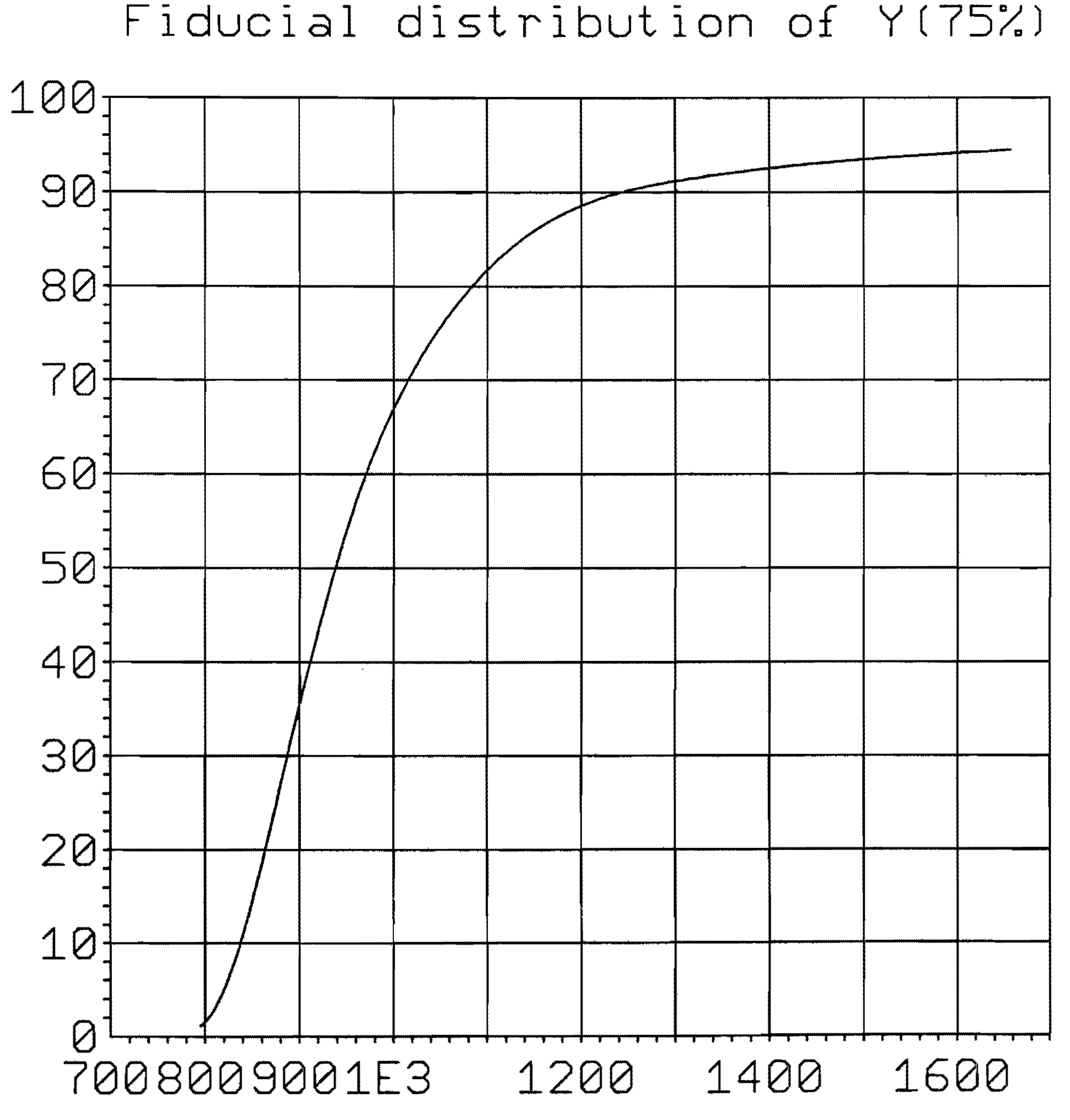

Fiducial distribution of Y(25%)
100
90
80
70
60
50
40
30
20
10
0
750
850
950
1050
1150
1250
1350


Fiducial distribution of Y(40%)
100
90
80
70
60
50
40
30
20
10
0
750
850
950
1050
1150
1250
135

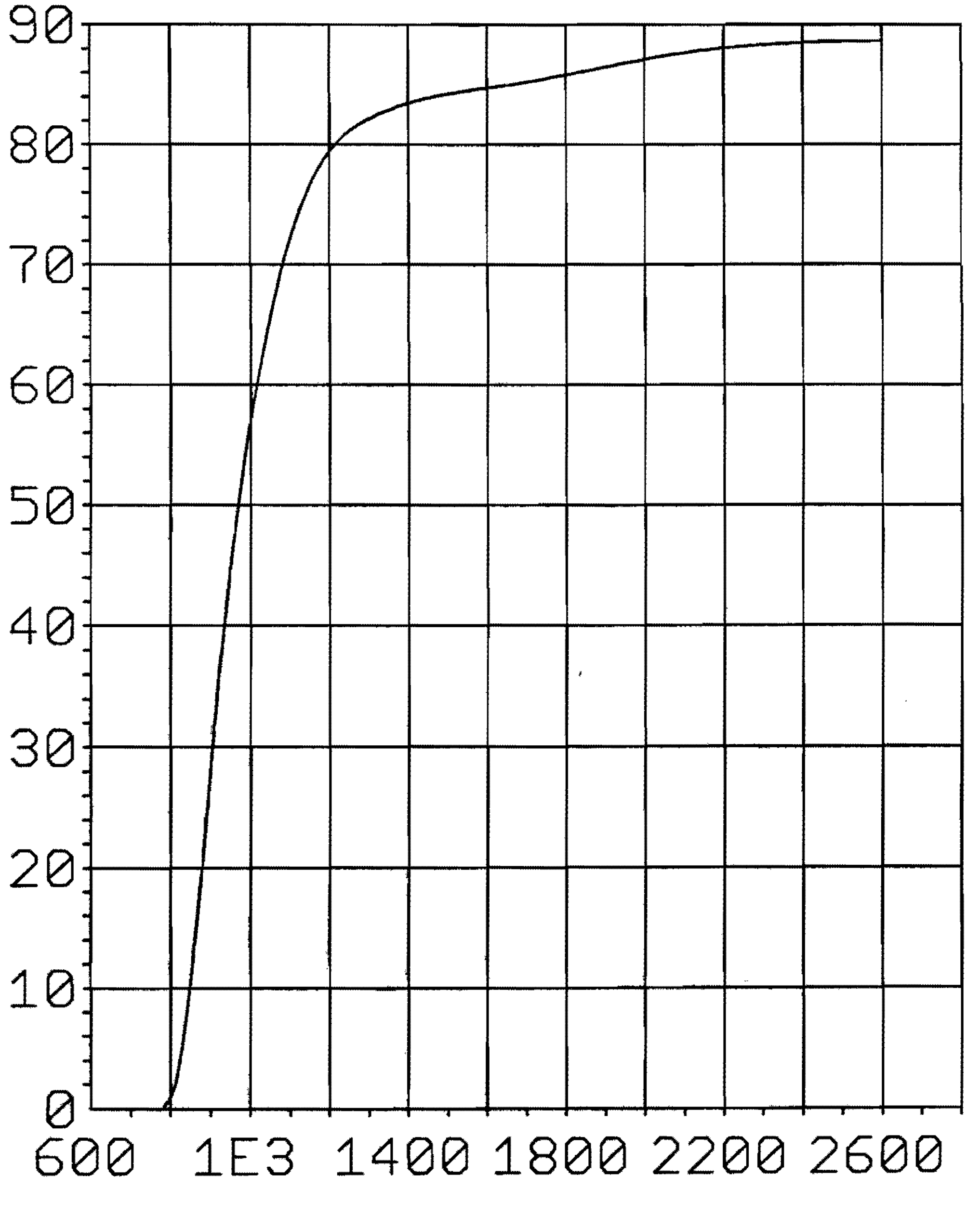

Fiducial distribution of Y(50%)
100
90
80
70
60
50
40
30
20
10
0
700800900 1E3
1200
1400
1600

Fiducial distribution of Y(60%)

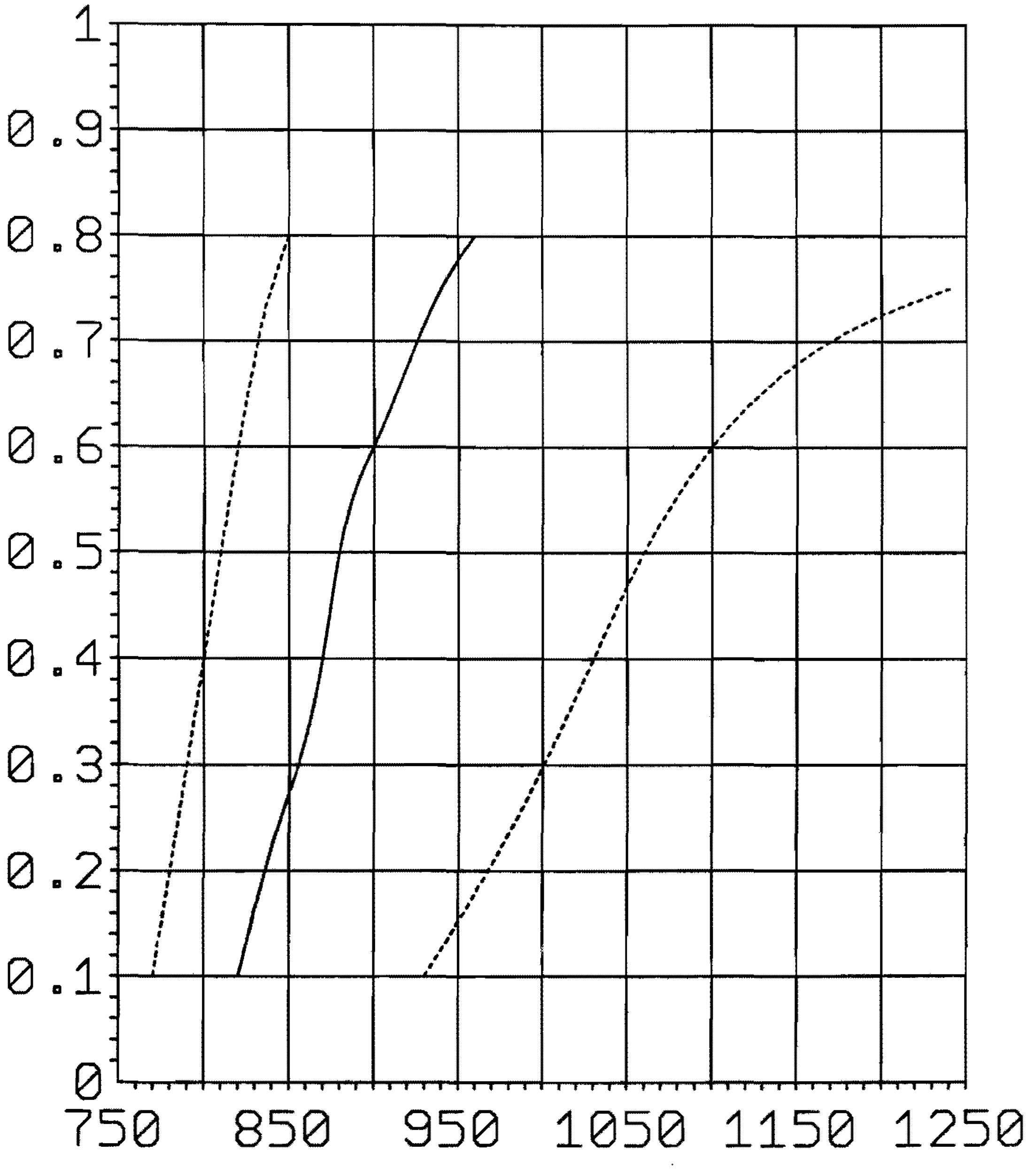



Fiducial distribution of Y(75%)
100
90
80
70
60
50
40
30
20
10
0
700800900 1E3
1200
1400
1600

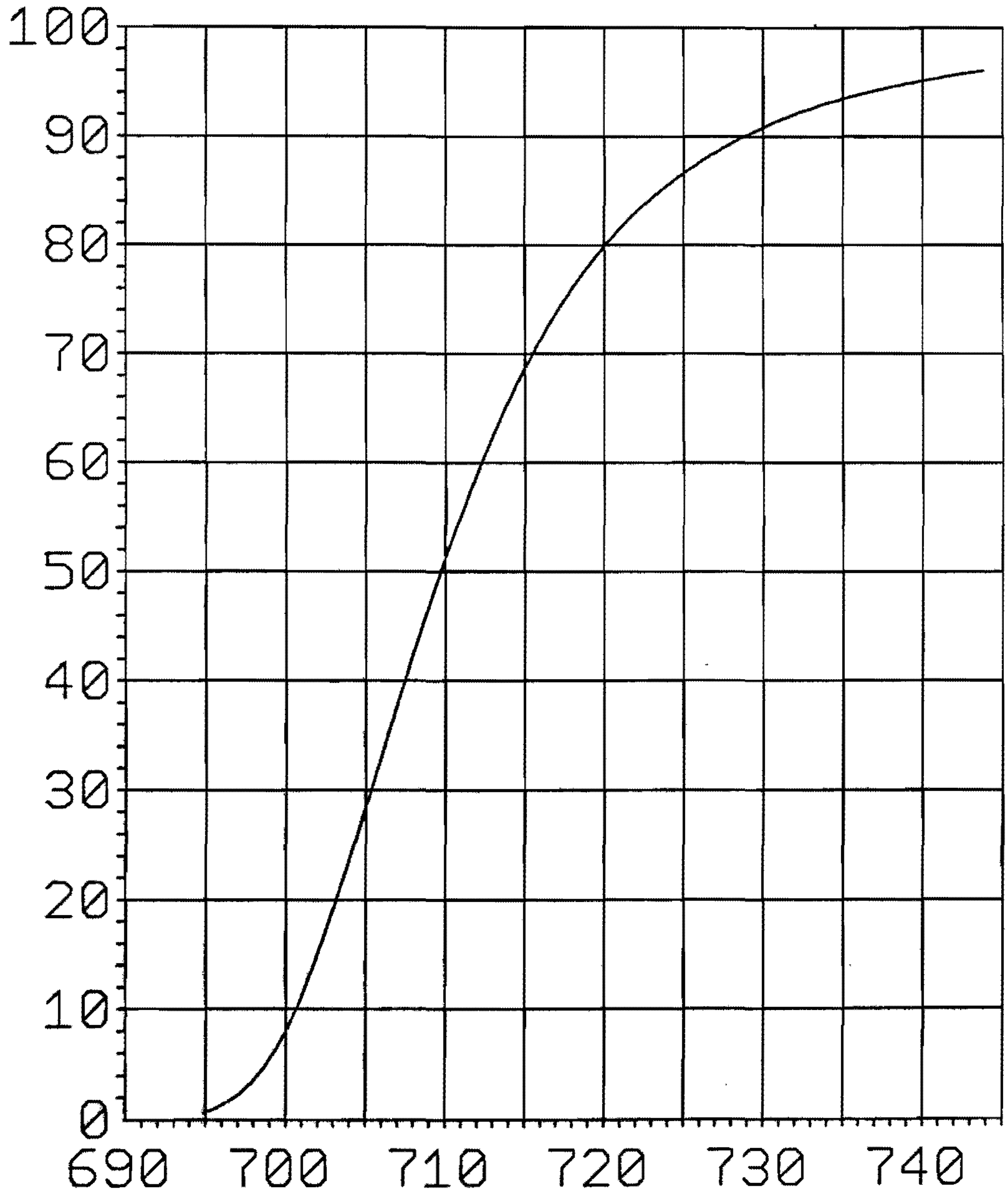
Fiducial distribution of Y(80%)



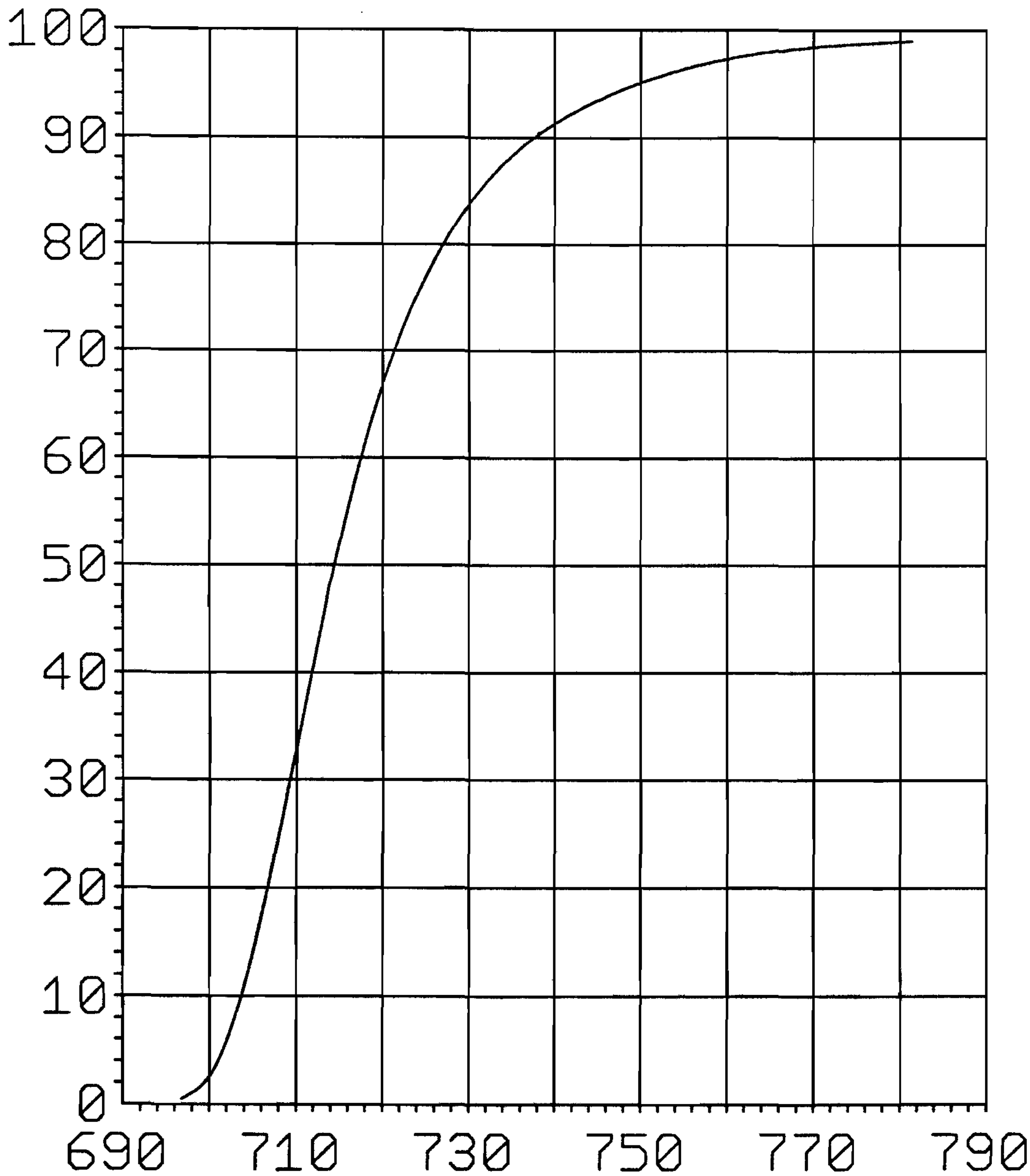
Distribution of tempest maximum
and 80%-confidence intervals: 1



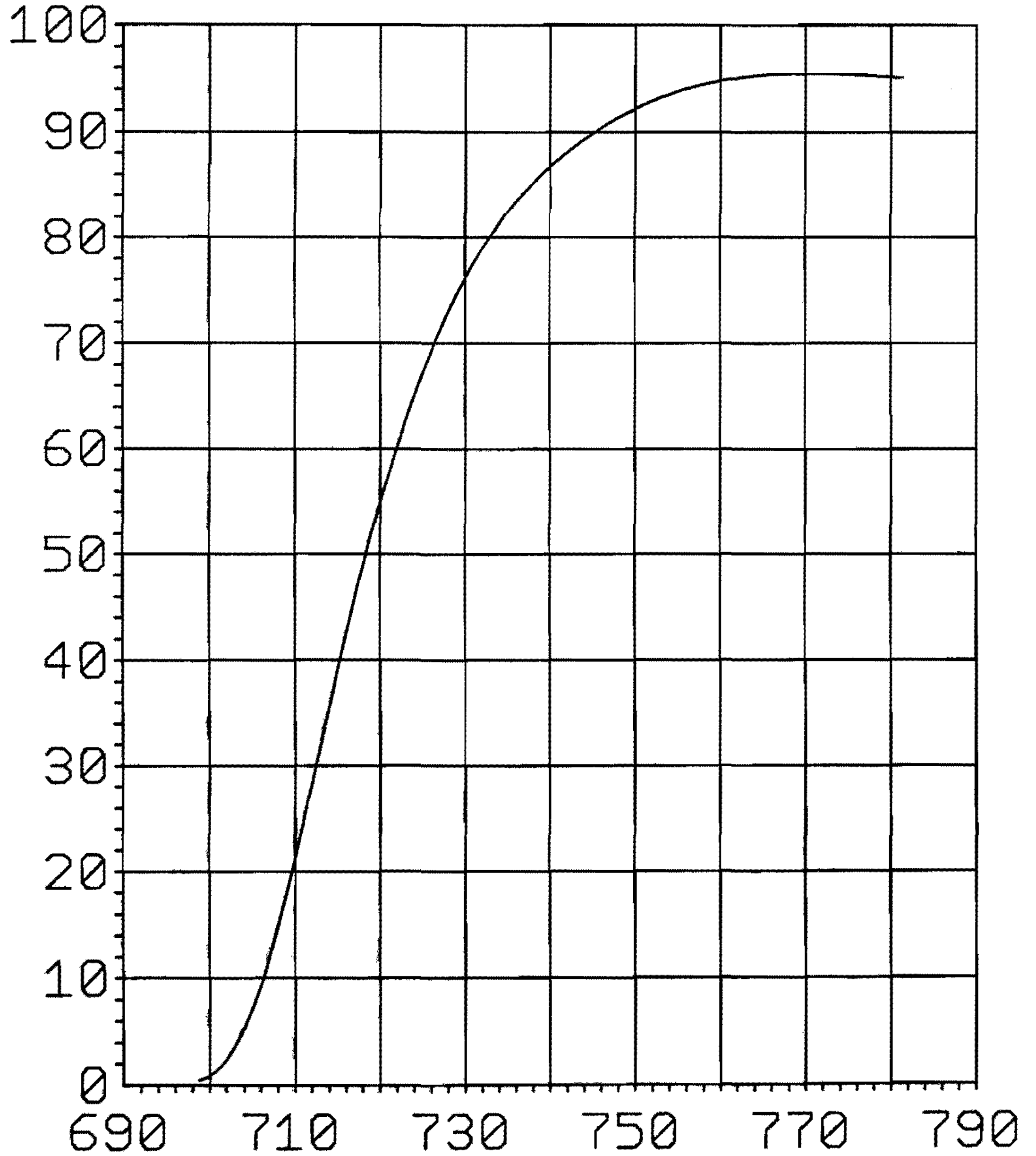
Fiducial distribution of Y(10%)

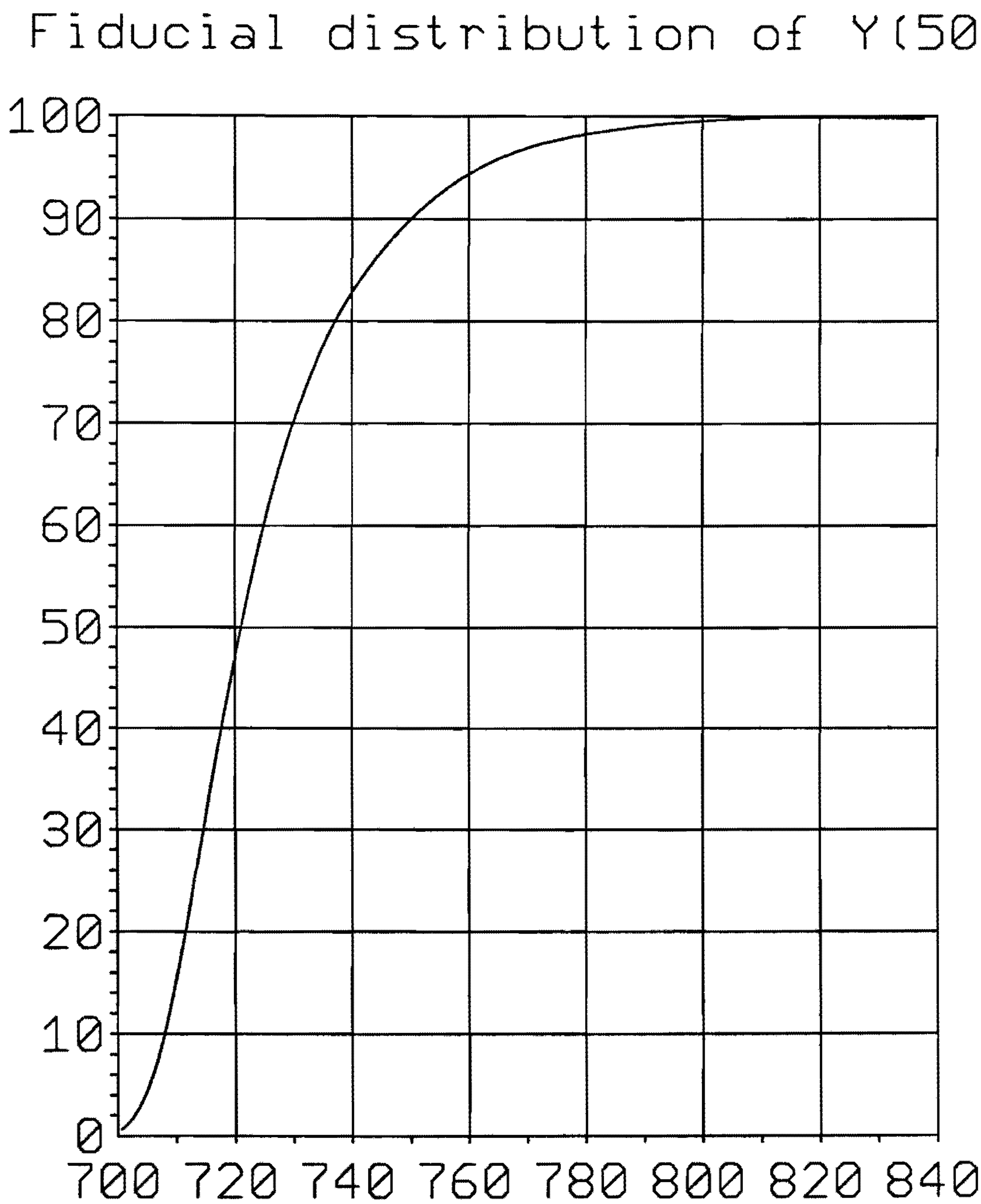


Fiducial distribution of Y(25%)

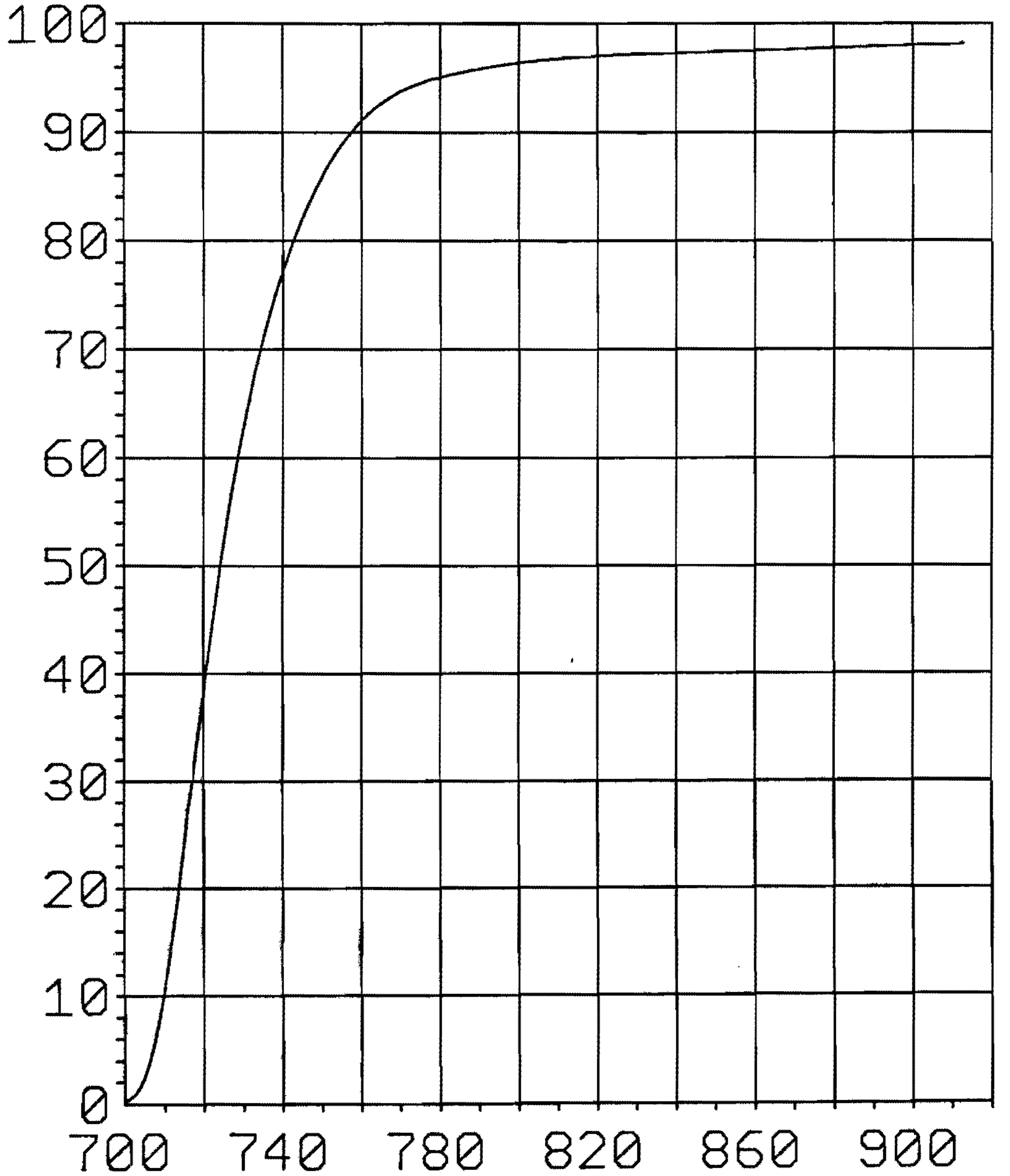


Fiducial distribution of Y(40%)

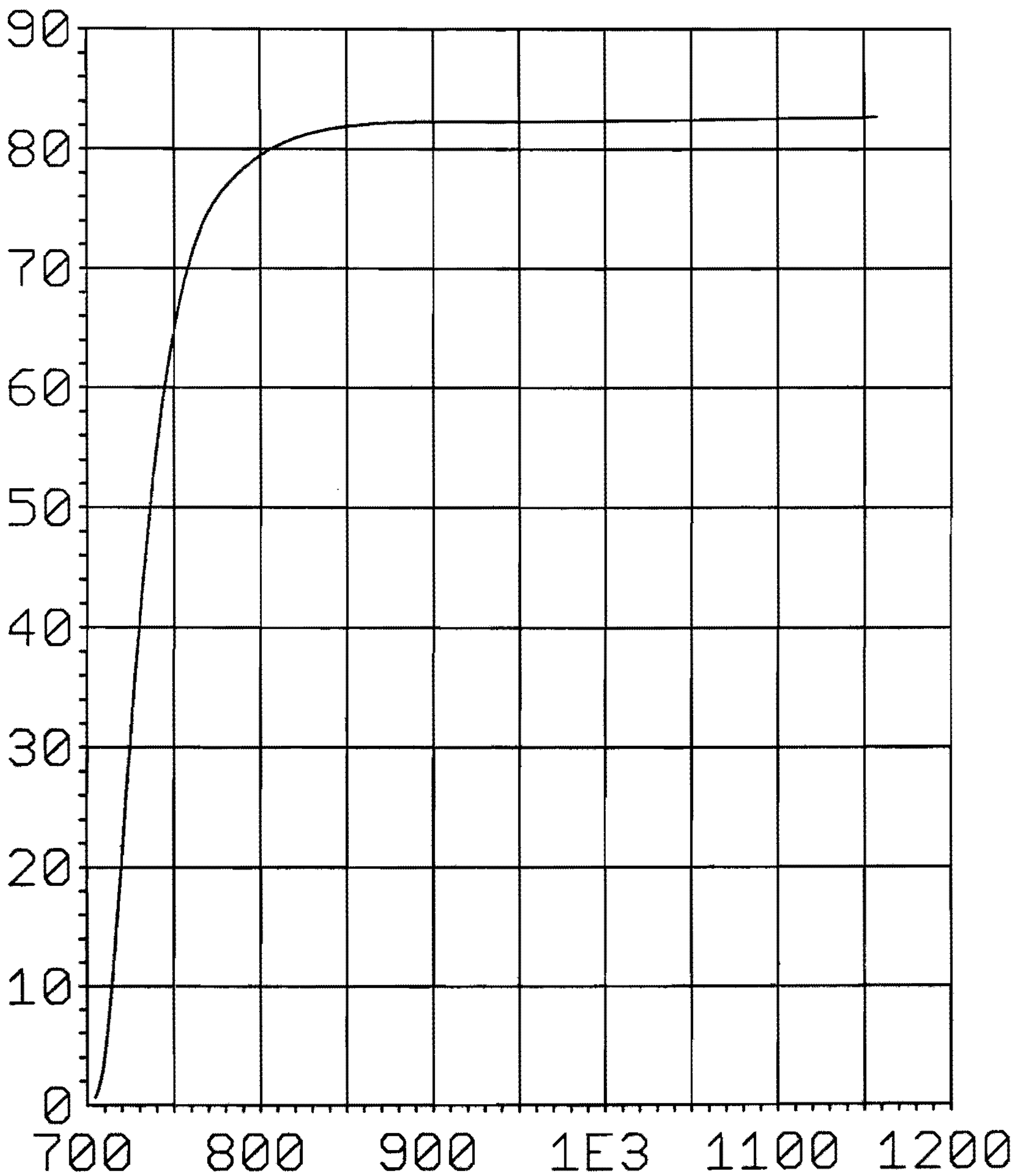


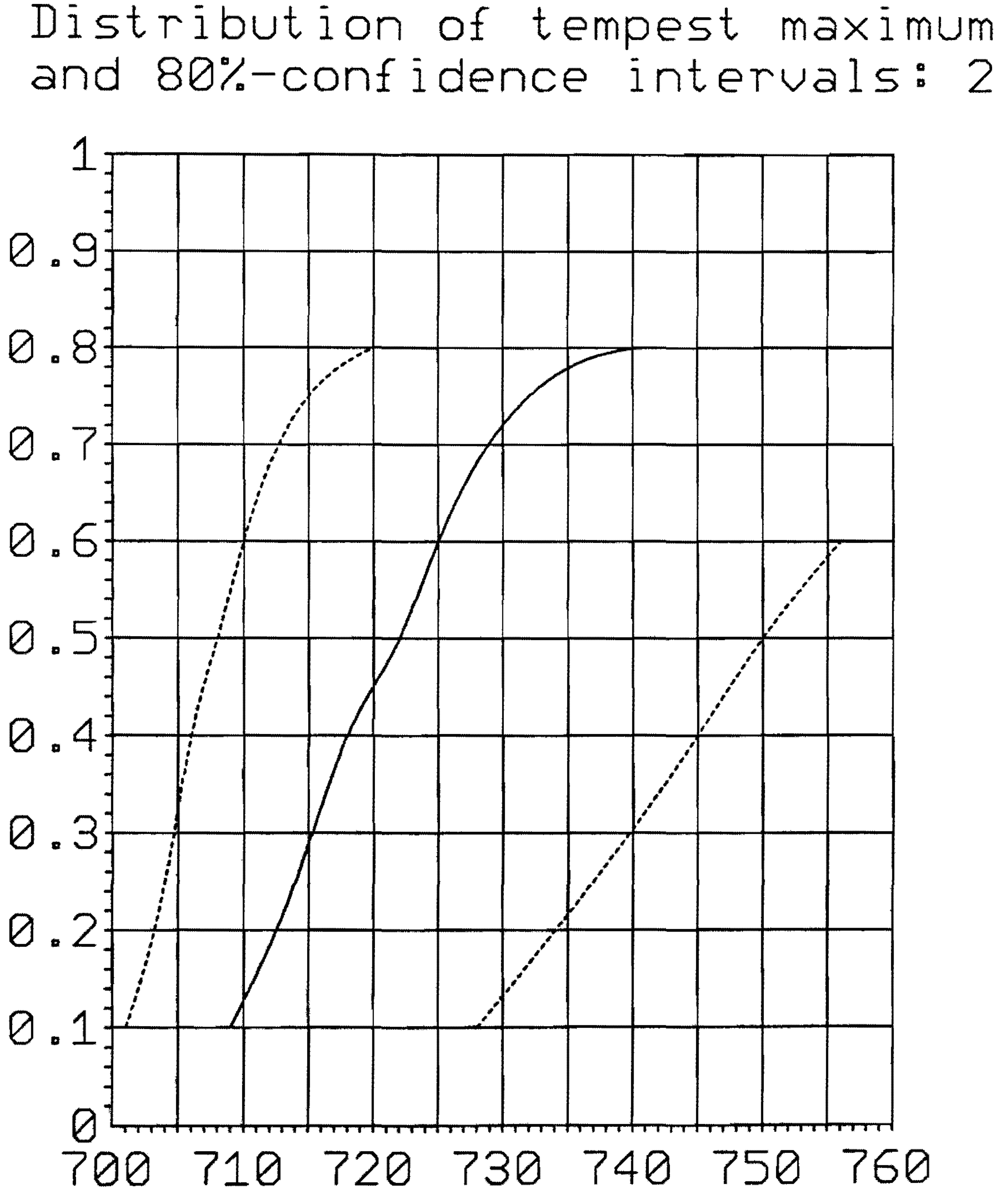

Fiducial distribution of Y(50%)
100
90
80
70
60
50
40
30
20
10
0
700
720
740
760
780
800
820
840

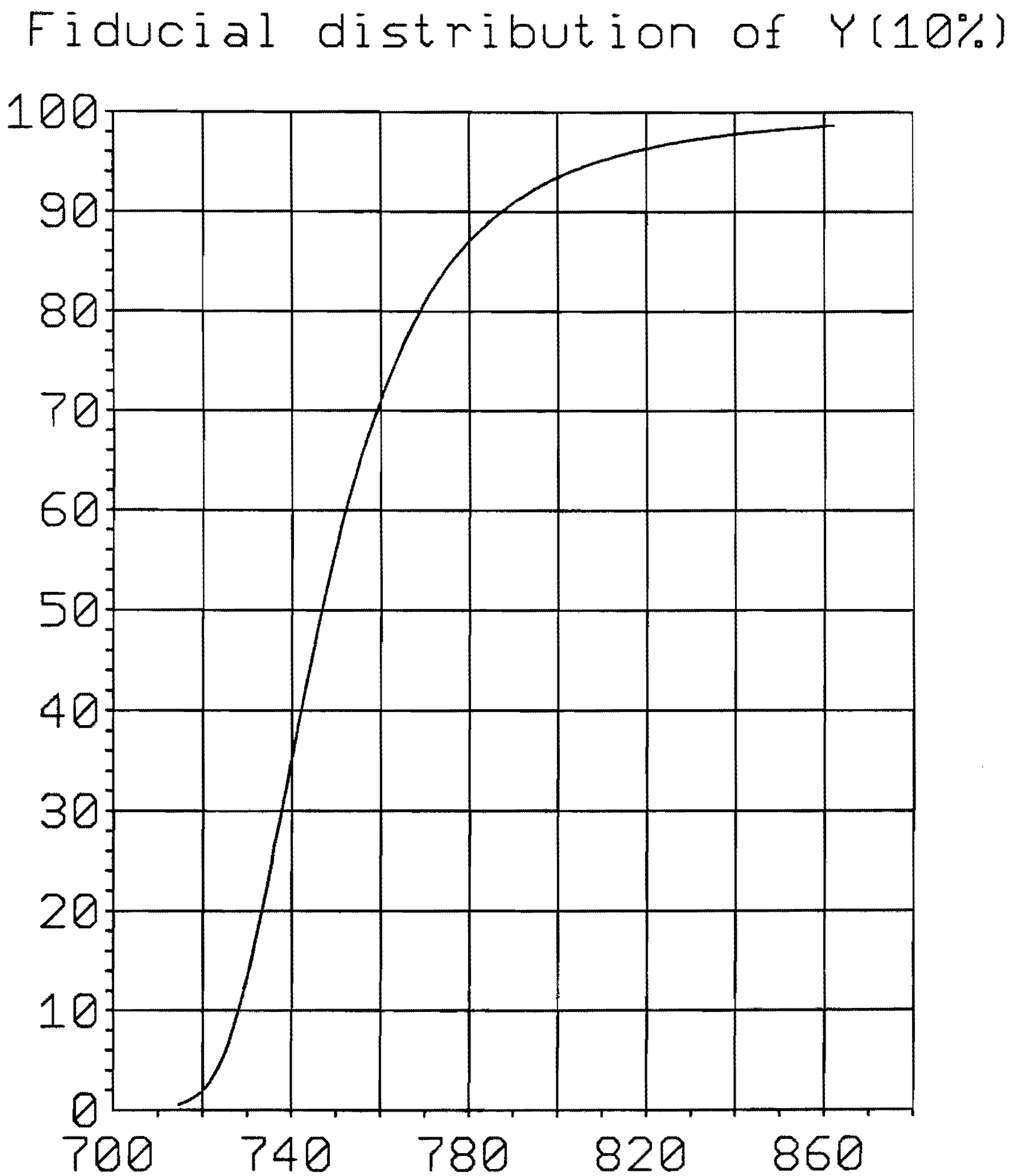
Fiducial distribution of Y(60%)

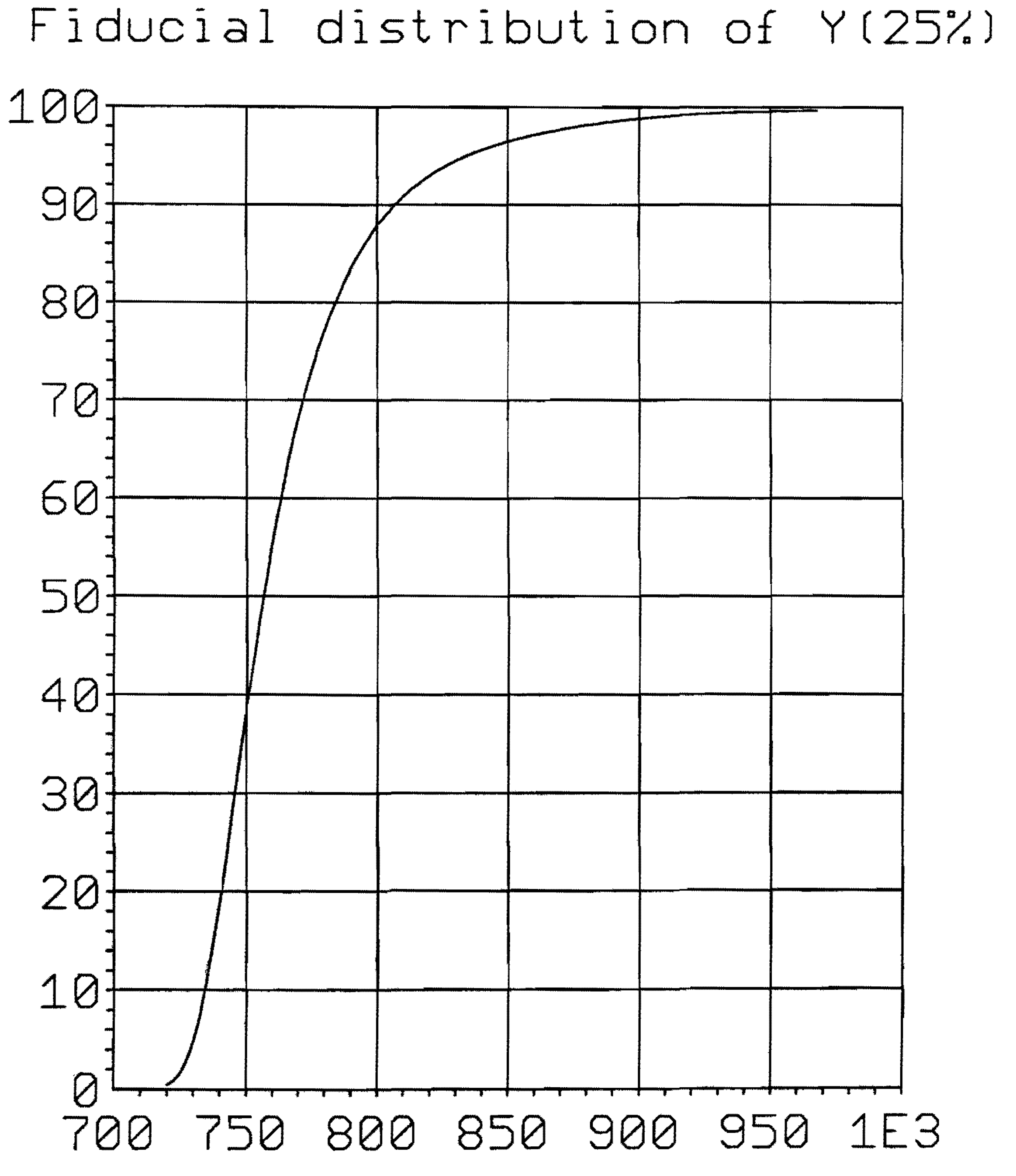


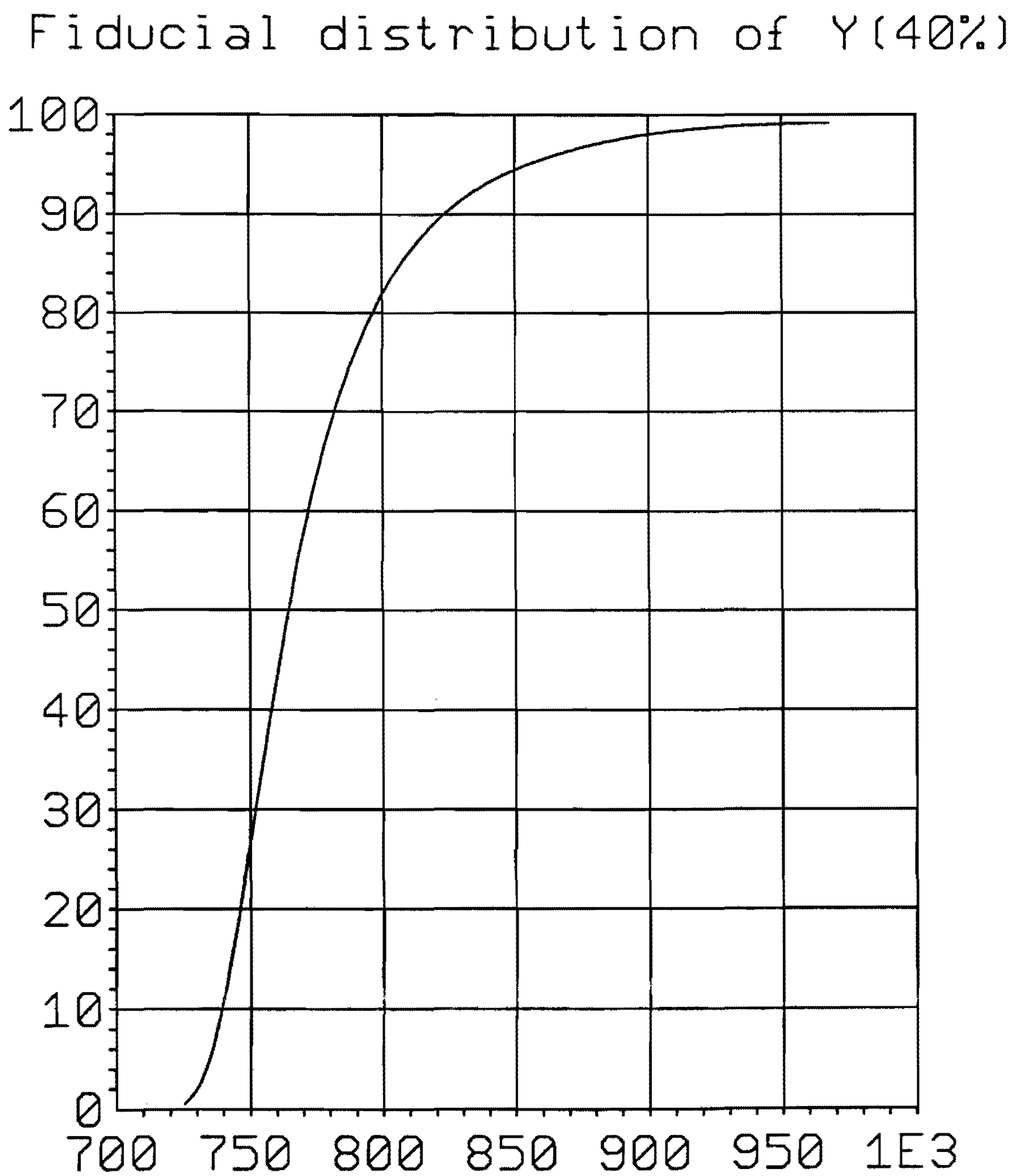
Fiducial distribution of Y(75%)

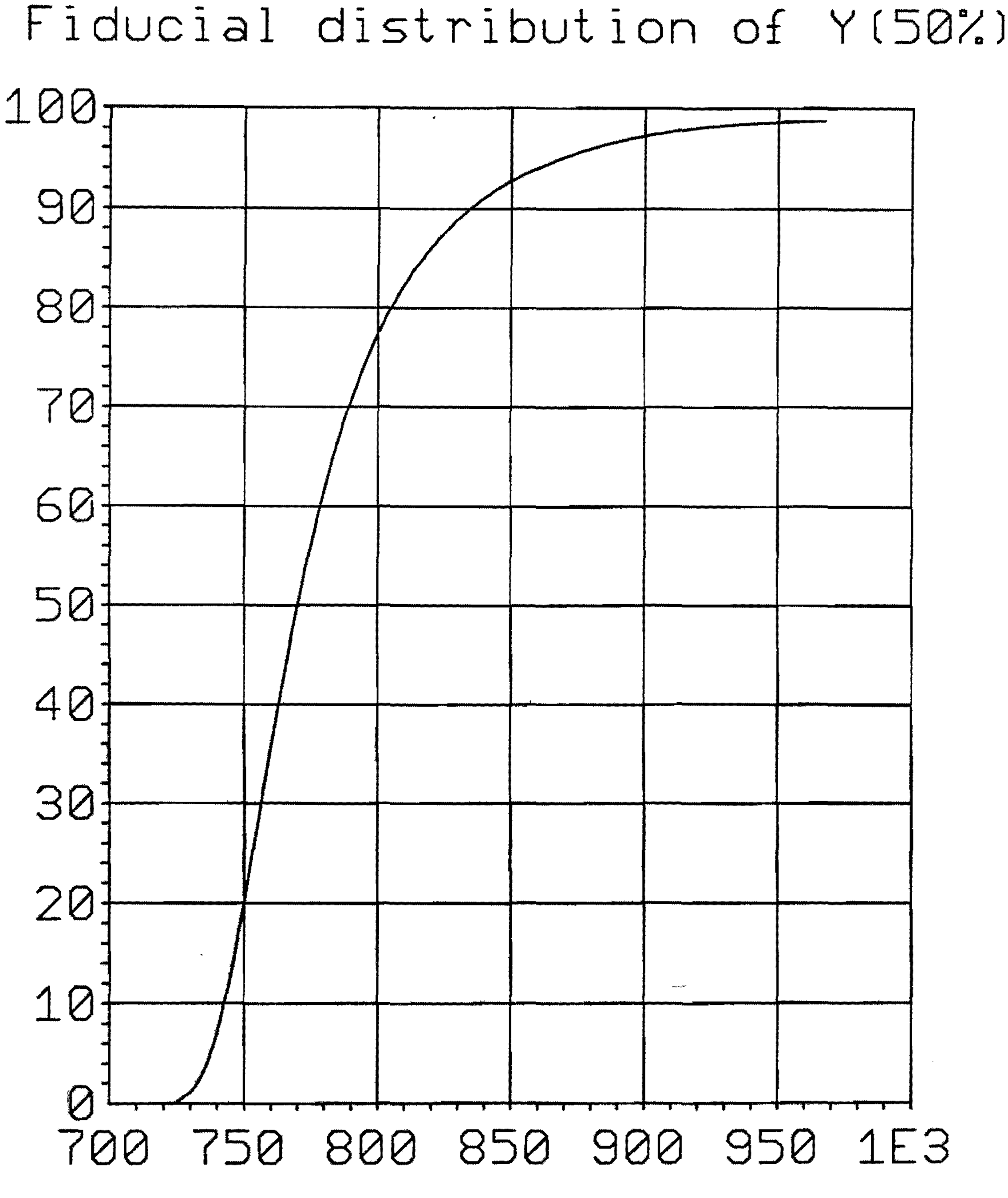



Distribution of tempest maximum
and 80%-confidence intervals: 2
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
700
710
720
730
740
750
760

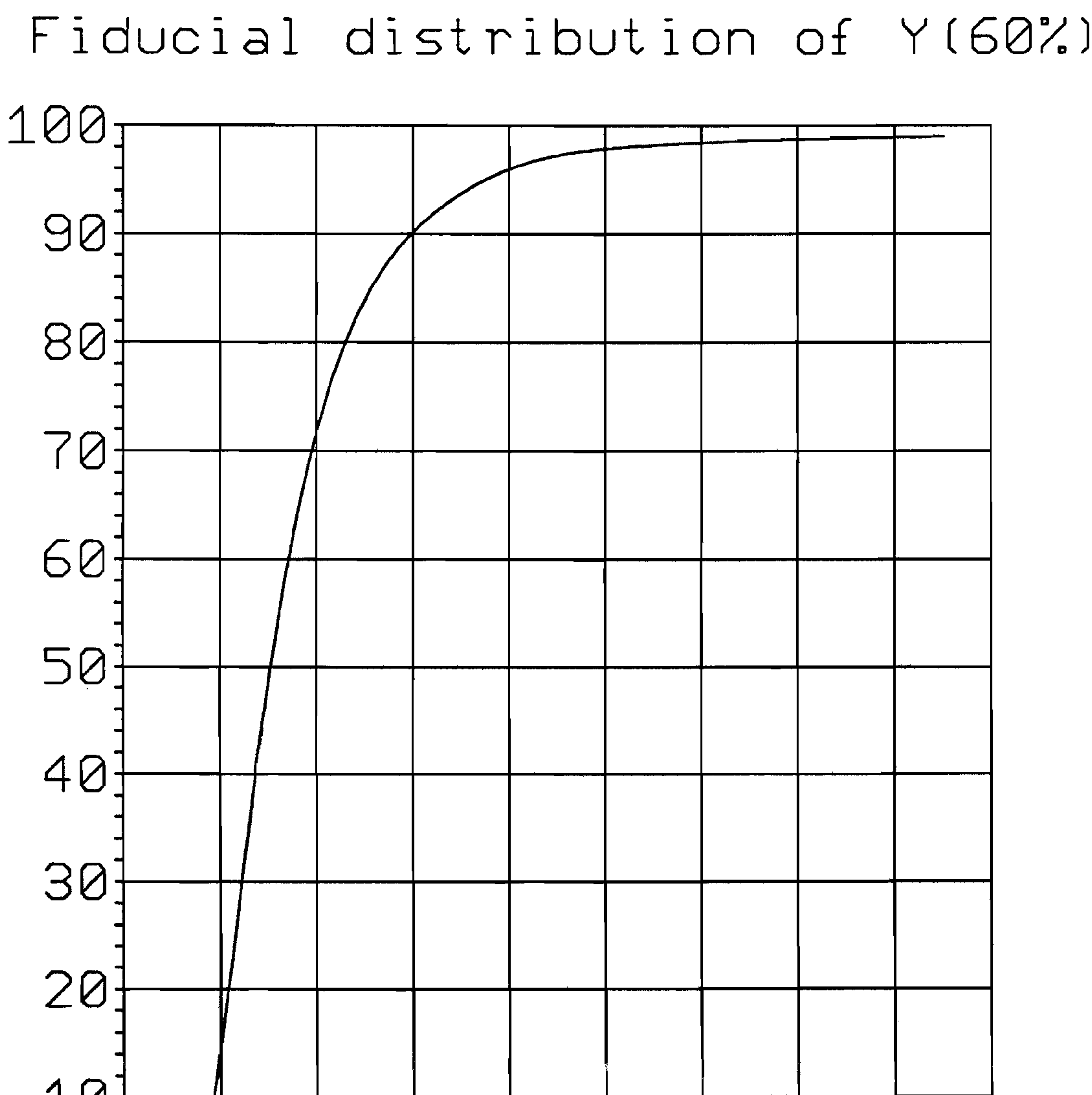

Fiducial distribution of Y(10%)
100
90
80
70
60
50
40
30
20
10
0
700
740
780
820
860


Fiducial distribution of Y(25%)
100
90
80
70
60
50
40
30
20
10
0
700
750
800
850
900
950
1E3

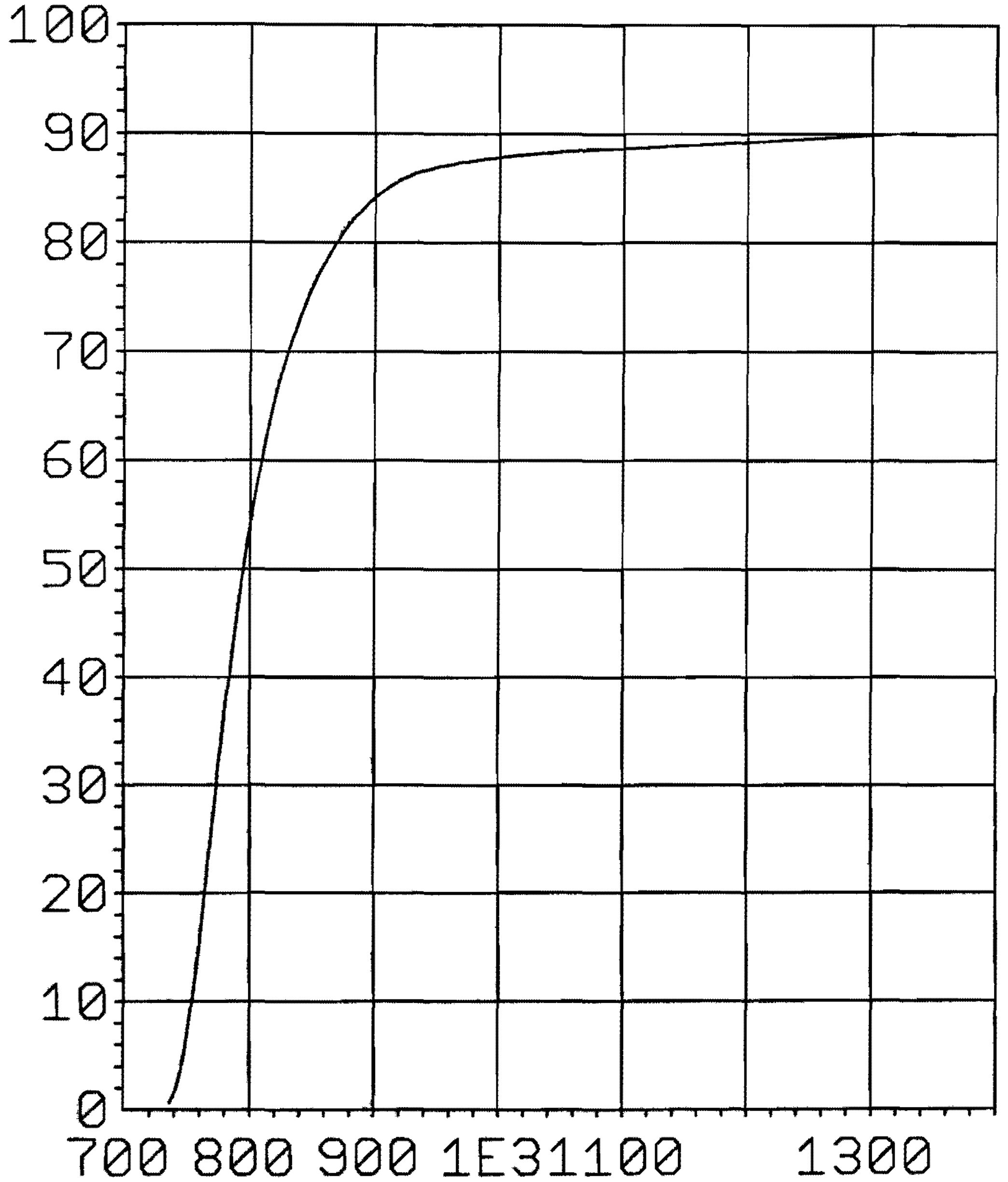

Fiducial distribution of Y(40%)
100
90
80
70
60
50
40
30
20
10
0
700
750
800
850
900
950
1E3



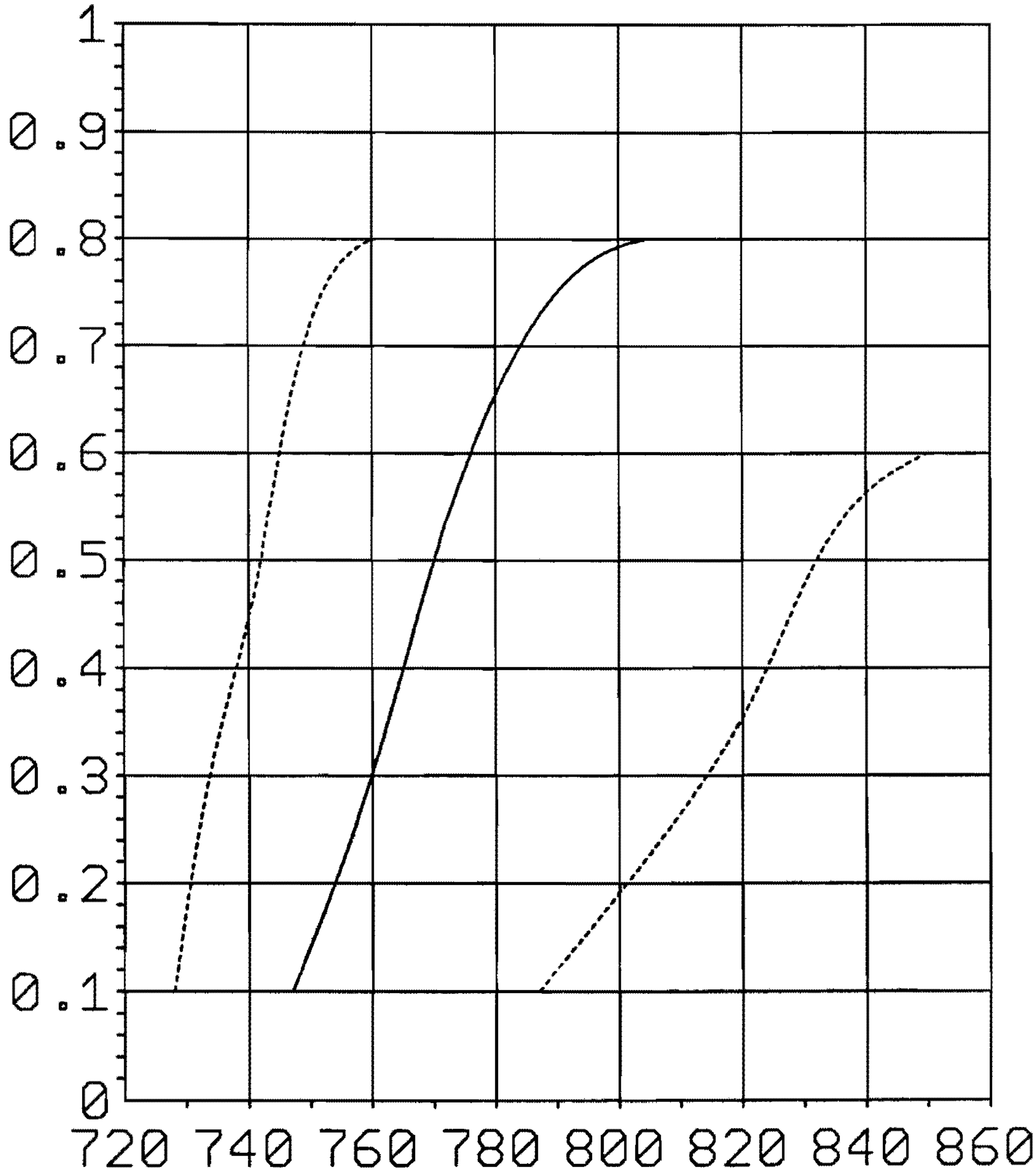
Fiducial distribution of Y(50%)
100
90
80
70
60
50
40
30
20
10
0
700
750
800
850
900
950
1E3


Fiducial distribution of Y(60%)
100
90
80
70
60
50
40
30
20
10
0
700
800
900
1E3
1100

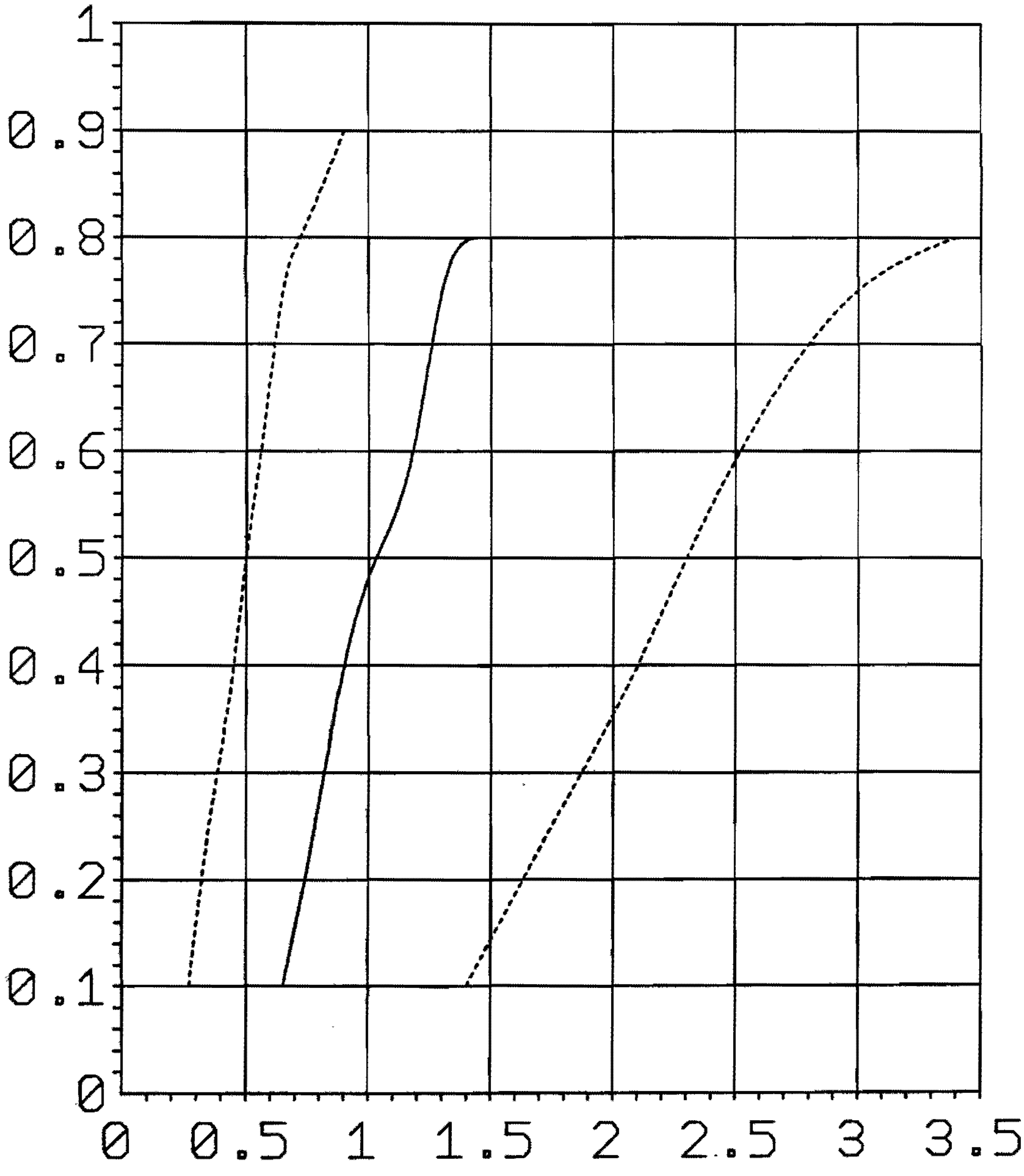
Fiducial distribution of Y(75%)



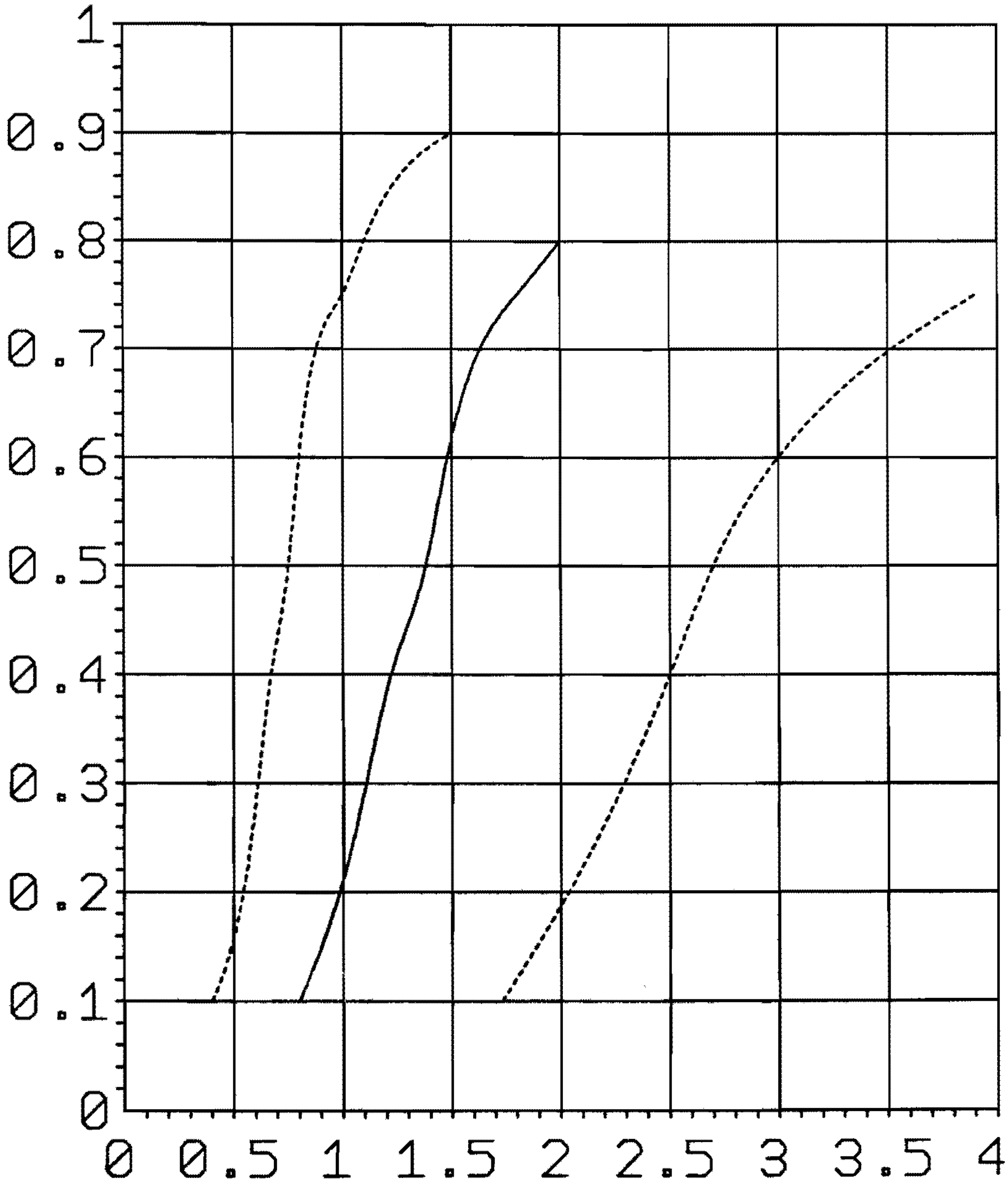
Distribution of tempest maximum and 80%-confidence intervals: 3



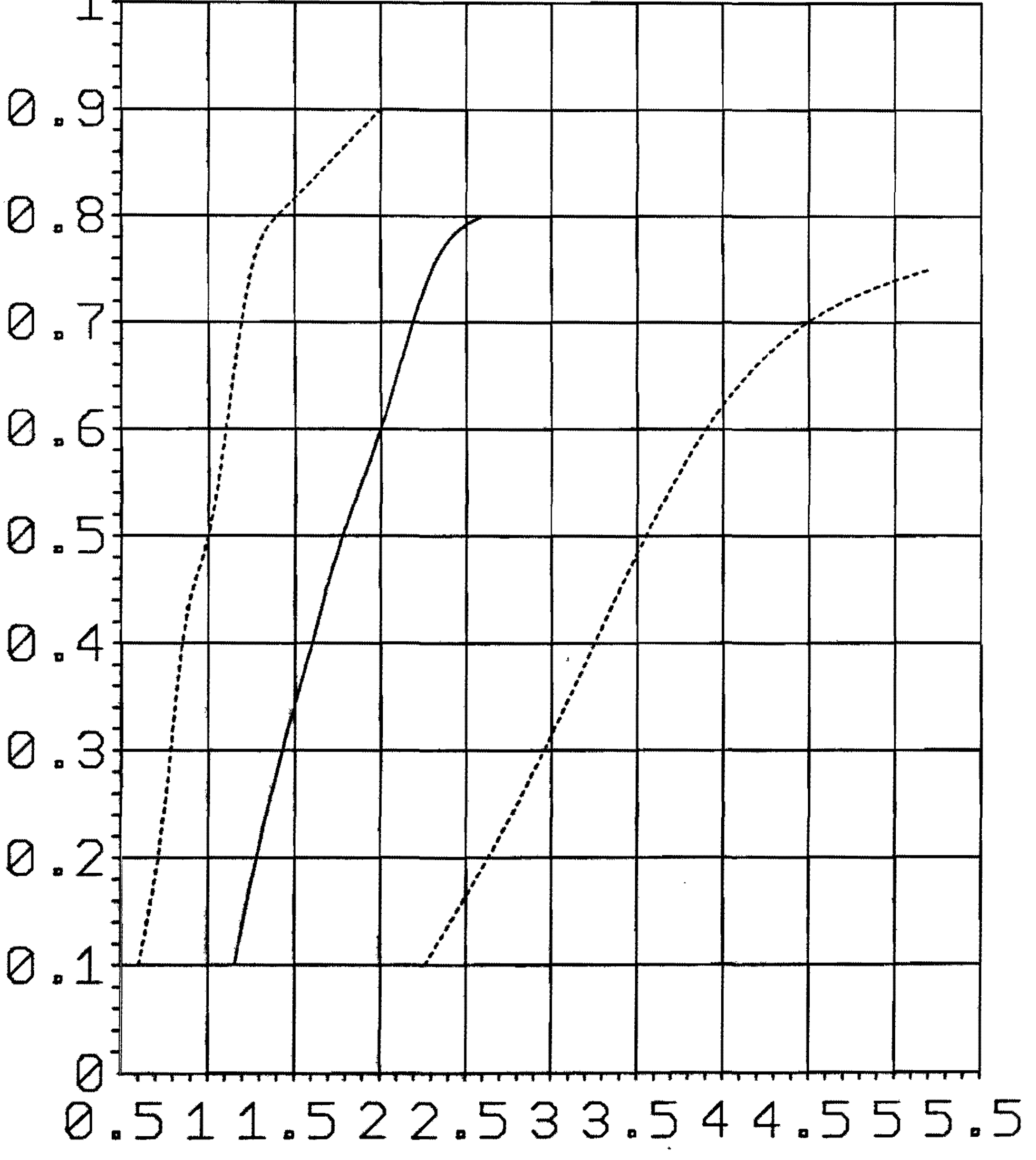
Distribution of tempest maximum
80%-confidence intervals 0001



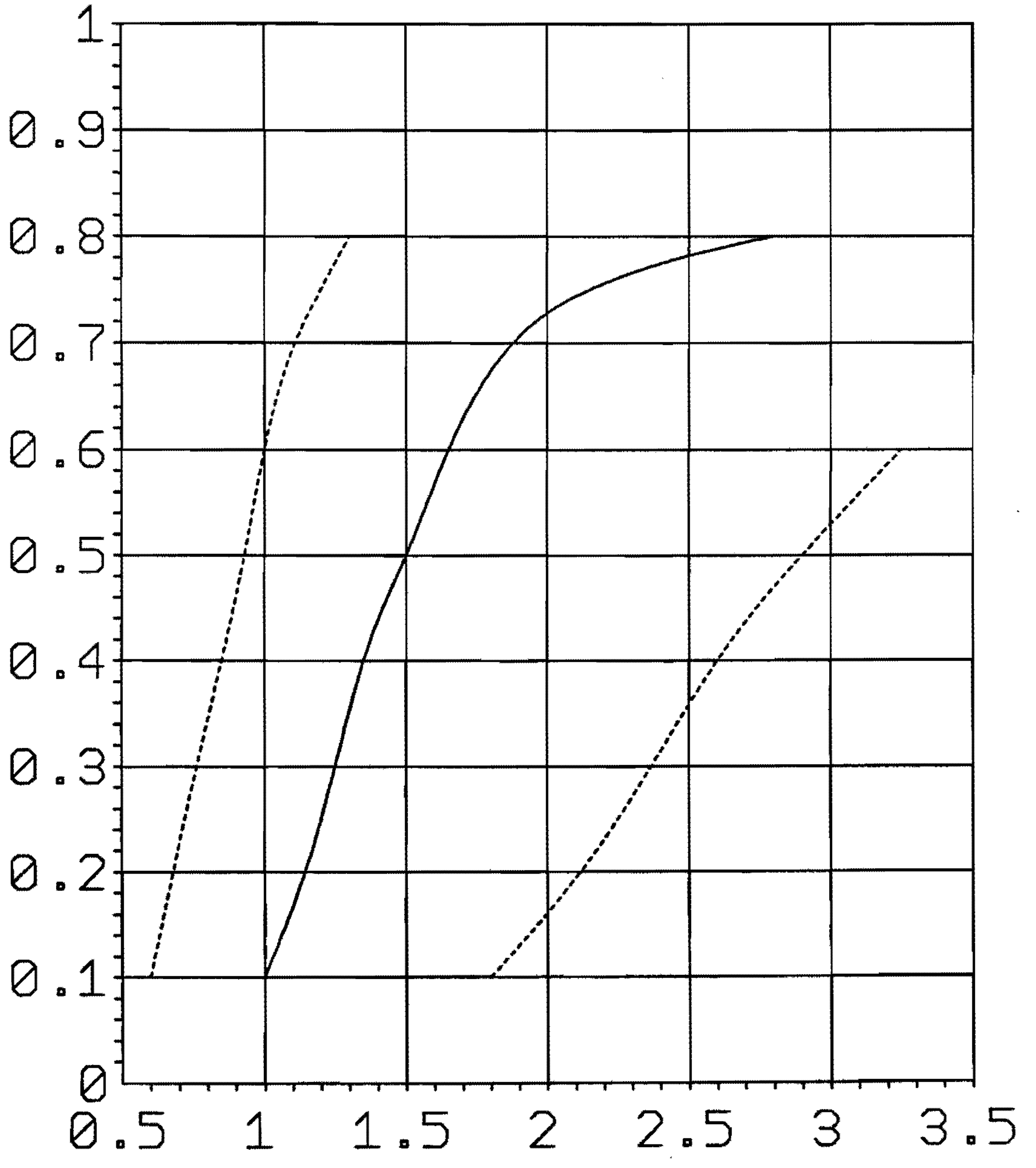
Distribution of tempest maximum
80%-confidence intervals 00%1



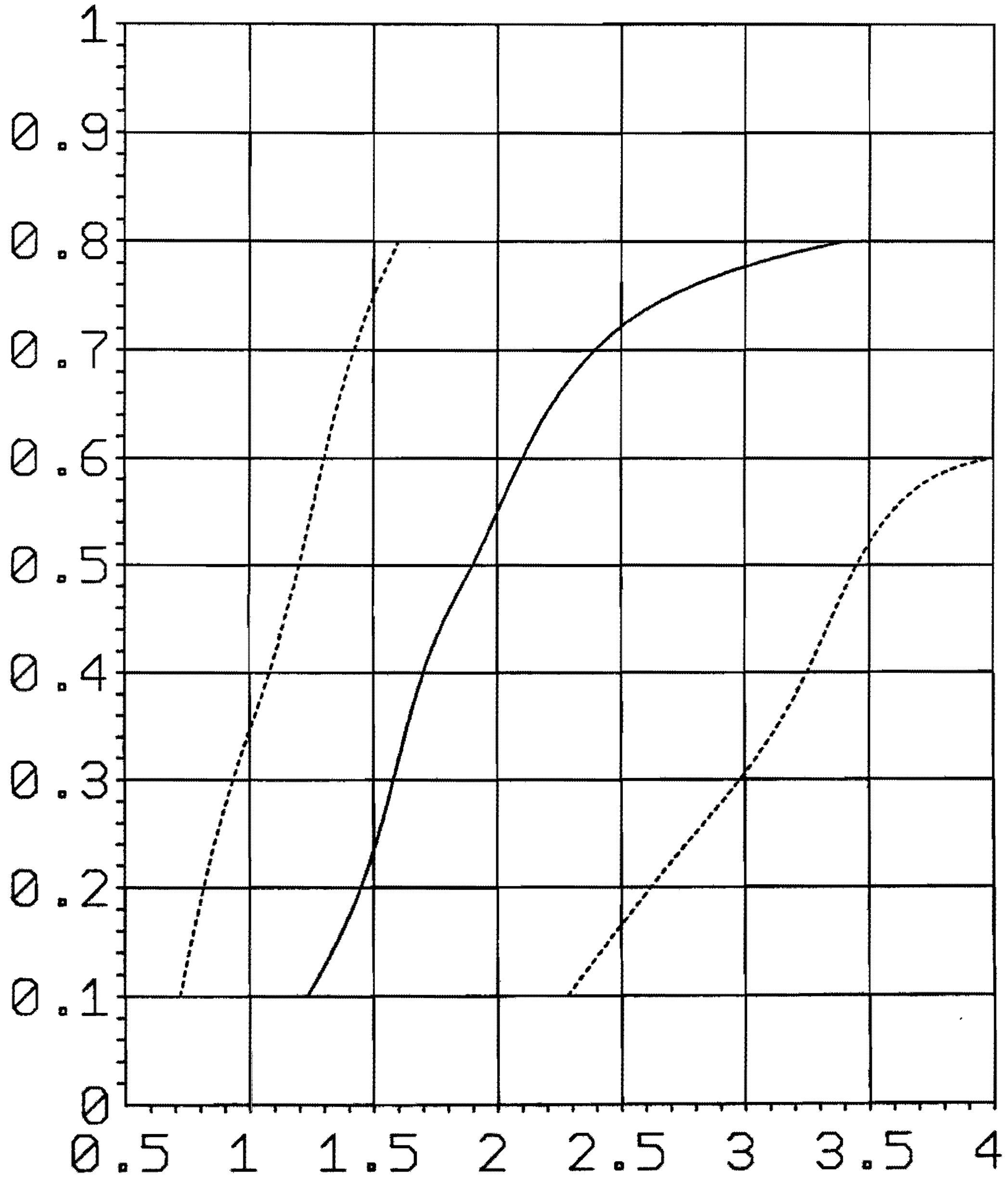
Distribution of tempest maximum
80%-confidence intervals 0011



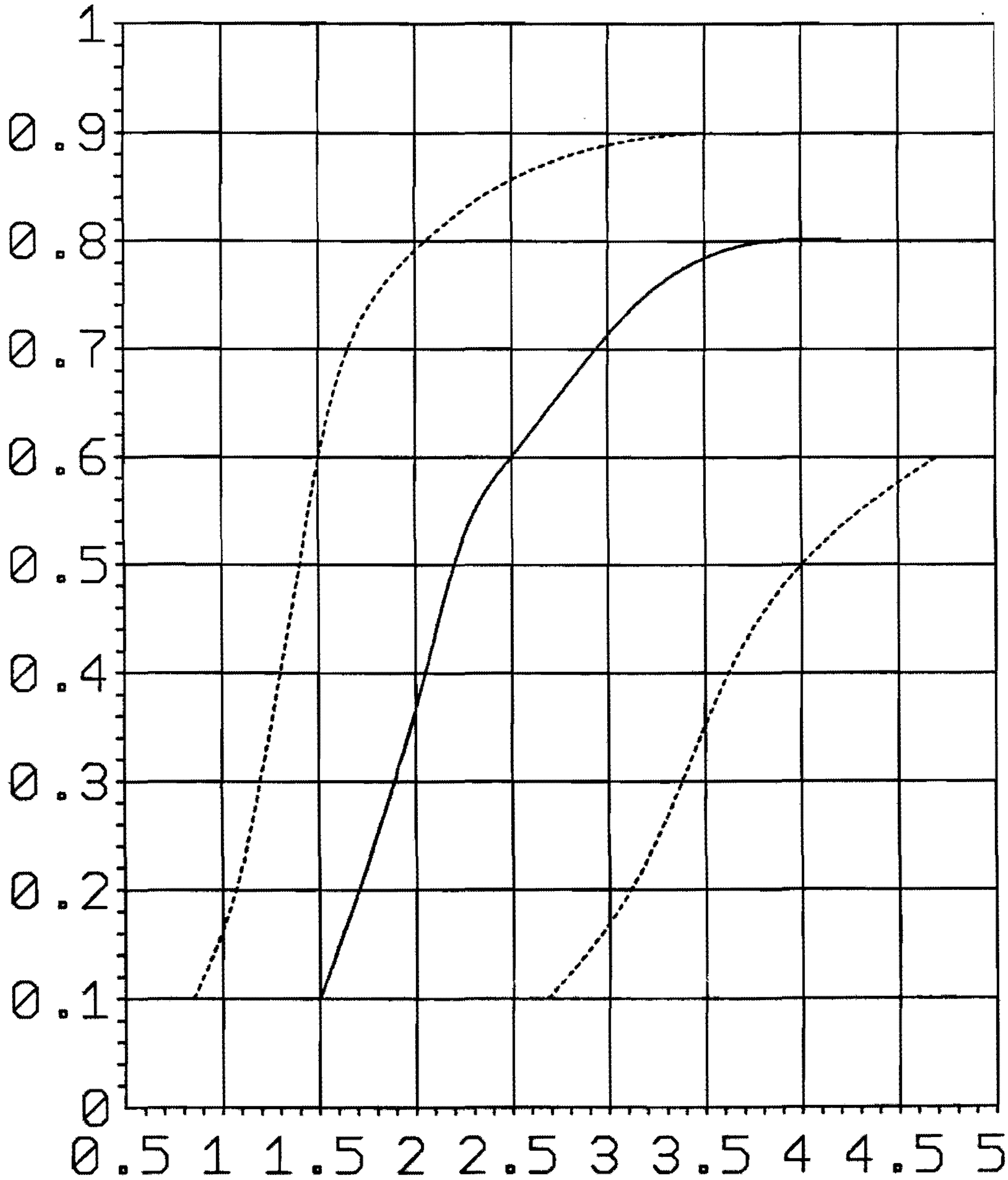
Distribution of tempest maximum
80%-confidence intervals 0½½1



Distribution of tempest maximum
80%-confidence intervals 0½11



Distribution of tempest maximum
80%-confidence intervals 0111



APPENDIX B: FIDUCIAL INFERENCE

In this appendix we will give some background information on fiducial inference. It serves as a useful means for a fuller understanding of the work described in this report. For more details we refer to [2] and [4].

Probability distributions do not only depend on the stochastic variable, but also on one or more parameters. In statistics, we are often interested in the parameters. With the help of observations we can for instance try to calculate estimates or confidence intervals for the parameters. Fiducial inference transforms the probability density into a fiducial density for the parameters, given some observations.

Suppose that the equation $P(x;\theta) = q$ defines a one-to-one correspondence between x and θ for every $0 < q < 1$. Further, we will assume that with fixed θ, $P(x;\theta)$ is strictly increasing in x, and with fixed x, $P(x;\theta)$ is strictly increasing or decreasing in θ.

Now the fiducial density $f(\theta;x)$ is defined as follows (cf. [2]).

$$f(\theta;x) = \left| \frac{\partial P(x;\theta)}{\partial \theta} \right| . \tag{B.1}$$

The fiducial distribution is calculated analogous to the probability distribution as

$$F(\theta;x) = \int_{-\infty}^{\theta} f(\phi;x)\, d\phi . \tag{B.2}$$

Before we will consider fiducial inference for two parameters, we will formulate three important principles of fiducial inference.

The first principle is the *Sufficiency Principle*, that says that in calculating fiducial densities one should use the minimal sufficient observations. We can represent a minimal sufficient sample $(x_1, \ldots, x_n)$ from a probability distribution depending on a parameter θ as

$$(x_1, \ldots, x_n) \equiv (x,a) \text{ with } x \in \mathbb{R} \text{ and } a \in \mathbb{R}^{n-1} . \tag{B.3}$$

The vector a is called an ancillary statistic, that is, a stochastic variable whose probability density does not depend on θ. The representation (B.3) has to be invertible.

The second principle is the *Conditioning Principle*. According to this principle statistical inference has to be based on $p(x \mid a;\theta)$ and moreover, if $a = (a_1, a_2)$ and $p(x \mid a;\theta) = p(x \mid a_1;\theta)$, then a_2 is irrelevant and must therefore not be used in the calculations.

The third principle is the *Non-coherence Principle*, that says that we cannot use the fiducial density of $\bar{\theta}$ to calculate the fiducial density of $g(\bar{\theta})$ if the function $g(\theta)$ is not invertible. In this case the fiducial density of $g(\bar{\theta})$ has to be calculated from the relevant part of the observations. The fiducial distributions for $\bar{\theta}$ and $g(\bar{\theta})$ are then called non-coherent (cf. [4]).

If we want to perform fiducial inference for two parameters $\bar{\alpha}$ and $\bar{\beta}$, we have to go through the following steps. It is assumed that sufficient reduction has taken place.

(1) Transform the sample $(x_1, \ldots, x_n)$ to (y_1, y_2, z) where $y_1, y_2 \in \mathbb{R}$ and where $z \in \mathbb{R}^{n-2}$ is an ancillary statistic. From now on, all densities will be conditional on z.

(2) The (conditional) probability density $p_{y_1}(y_1 ; \beta)$ of y_1 has to depend only on β. By fiducial inversion (cf. (B.1)) we find the fiducial density of $f_{\bar{\beta}}(\beta ; y_1)$ of $\bar{\beta}$.

(3) Now we take the probability density $p(y_2 \mid y_1 ; \alpha, \beta)$ of $y_2 \mid y_1$ and with β fixed, we obtain (again by (B.1)) the fiducial density $f_{\bar{\alpha}|\beta}(\alpha \mid \beta ; y_1, y_2)$ of $\bar{\alpha} \mid \bar{\beta}$.

(4) The joint fiducial density $f(\alpha,\beta) = f(\alpha,\beta ; y_1, y_2)$ is defined by

$$f(\alpha,\beta) = f(\beta) * f(\alpha \mid \beta) .$$

(5) The marginal fiducial density for $\bar{\alpha}$ can be determined by

$$f(\alpha) = \int_\beta f_{\bar{\alpha},\beta}(\alpha, \phi)\, d\phi = \int_\beta f_{\bar{\alpha}|\beta}(\alpha \mid \phi)\, f_{\bar{\beta}}(\phi)\, d\phi .$$

If in step (3) the conditional density of α depends on a non-invertible function $g(\bar{\beta})$ of $\bar{\beta}$, then the marginal density of $\bar{\alpha}$ must be constructed by taking expectation with respect to the fiducial density for $g(\bar{\beta})$:

$$f(\alpha) = \int_\beta f_{\bar{\alpha}|g(\beta)}(\alpha \mid g(\phi)) * f_{g(\bar{\beta})}(g(\phi))\, d\phi$$

(cf. Non-coherence principle).

REFERENCES

[1] Fisher, R.A. and Tippett, L.H.C., (1928), "Limiting forms of the frequency distribution of the largest of smallest member of a sample", Proc. Camb. Phil. Soc. 24, 180-190.

[2] Fisher, R.A. (1973), "Statistical Methods and Scientific Inference", 3rd edition, Hafner Press, New York.

[3] Gumbel, E.J. (1958), "Statistics of extremes", Columbia University Press, New York.

[4] Wilkinson, G.N. (1977), "On resolving the controversy in statistical inference", Journal of the Royal Statistical Society, series B 39, 119-143.

[5] Atkinson, K.E. (1978), "An introduction to numerical analysis", John Wiley & Sons, New York.