SKB TECHNICAL REPORT 88-19

A Bayesian nonparametric estimation of distributions and quantiles

Kurt Pörn

Studsvik AB

November 1988

SVENSK KÄRNBRÄNSLEHANTERING AB
SWEDISH NUCLEAR FUEL AND WASTE MANAGEMENT CO

BOX 5864 S-102 48 STOCKHOLM
TEL 08-665 28 00 TELEX 13108-SKB

A BAYESIAN NONPARAMETRIC ESTIMATION OF DISTRIBUTIONS AND QUANTILES

Kurt Pörn

Studsvik AB

November 1988

Information on KBS technical reports from 1977-1978 (TR 121), 1979 (TR 79-28), 1980 (TR 80-26), 1981 (TR 81-17), 1982 (TR 82-28), 1983 (TR 83-77), 1984 (TR 85-01), 1985 (TR 85-20), 1986 (TR 86-31) and 1987 (TR 87-33) is available through SKB.

ABSTRACT

The report describes a Bayesian, nonparametric method for the estimation of a distribution function and its quantiles. The basic theory behind the method has been presented in /Ferguson, 1973/. The method, presupposing random sampling, is nonparametric so the user has to specify a prior distribution on a space of distributions (and not on a parameter space). In the current application, where the method is used to estimate the uncertainty of a parametric calculational model, the Dirichlet prior distribution is to a large extent determined by the first batch of Monte Carlo-realizations. In this case the result of the estimation technique is very similar to the conventional empirical distribution function.

The resulting posterior distribution is also Dirichlet, and thus facilitates the determination of probability (confidence) intervals at any given point in the space of interest. Another advantage is that also the posterior distribution of a specified quantile can be derived and utilized to determine a probability interval for that quantile.

The method was devised for use in the PROPER code package for uncertainty and sensitivity analysis.

TABLE OF CONTENTS

SUMMARY

The work reported in this paper has been performed on account of Swedish Nuclear Fuel and Waste Management Co. The aim of the work has been to find a method to estimate the uncertainty of the output of a parametric calculational model, when the input values of appropriate parameters are not exactly known. The method described was devised for use in the PROPER code package for uncertainty and sensitivity analysis.

Probability distributions are suitable tools for the description of the uncertainty associated with both the model parameters and the final results. Thereby, the main problem is to estimate a distribution function, based on available observations and desired quantiles of that distribution. Because of a limited number of observations there is also a statistical uncertainty associated with these estimates.

The method applied in this study is a Bayesian nonparametric method, presented in /Ferguson, 1973/. Being nonparametric means that the method does not presuppose any given type of distribution. However, the basic theory implies random sampling, which is not the case when variance reducing techniques are used. The method itself as well as an illustrating example are presented in this report.

1 INTRODUCTION

The estimation of a distribution function or some specific quantile based on available observations is a general statistical problem. In this paper our treatment of the problem is concerned with uncertainty analysis, where the uncertainty attached to a parametric model is studied by the use of Monte Carlo sampling. Thus observations are generated by repeated application of the model, where in each application (realization) each uncertain parameter is assigned a value sampled from its probability distribution. In an earlier report /Pörn, Åkerlund, 1985/ different deterministic or stochastic sampling techniques were surveyed.

If we assume that the unknown distribution is of a certain type (normal, lognormal etc), the corresponding estimation technique is called parametric. In the nonparametric case, no specific distribution type is presupposed. For example the empirical distribution and its quantiles are nonparametric estimators of the unknown distribution and the unknown quantiles respectively.

In this report we will briefly describe and then implement a nonparametric Bayesian approach, suggested in /Ferguson, 1973/. The Bayesian methodology greatly facilitates the determination of uncertainty (probability) intervals around the estimated distribution and estimated quantiles in particular. For the convenience of the reader, Ferguson's method is described in the next section in a brief and simplified manner.

2 THE DIRICHLET DISTRIBUTION

The Bayesian method is generally used for inferences about some unknown parameter. Then one needs a prior distribution on the parameter space. To find a good prior distribution is in many cases a difficult task. The problem is even worse in the nonparametric case, where one has to specify a prior distribution on a space of probability distributions.

Let us, for example, consider the space of all probability distributions $\Omega = \{F(t) \text{ for } t \in R^1\}$ where each distribution corresponds to a probability measure P through

$$F(T) = P((-\infty,t]) \qquad \text{(Eq 1)}$$

Then we have to define a prior distribution on Ω. If $A_1,\ldots, A_k$ is an arbitrary partition of the space R^1, each distribution $F \in \Omega$ corresponds to a probability vector $(P(A_1),\ldots, P(A_k))$ (where $\Sigma\ P(A_k) = 1$).

A distribution is assigned to Ω if we assume that the vector $(P(A_1),\ldots, P(A_k))$ is Dirichlet distributed with a given parameter. Ferguson shows that a Dirichlet distribution has two desirable properties as a prior distribution:

1. It is broad and flexible. Every given distribution in Ω can, with positive probability, be approximated arbitrarily well by a Dirichlet process P.

2. Given a sample of observations from the true probability distribution, the posterior distribution will be analytically manageable.

To define a Dirichlet distribution we start with a gamma distribution $G(\alpha, \beta)$ with shape parameter $\alpha > o$ and scale parameter $\beta > o$. Its probability density function is

$$g(Z;\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} Z^{\alpha-1}e^{-\beta Z};\ Z \geq 0 \qquad \text{(Eq 2)}$$

Let $Z_1, Z_2,\ldots, Z_n$ be independent random variables with $Z_j \sim G(\alpha_j,1)$.

Definition. The Dirichlet distribution, with parameter $\underline{\alpha} = (\alpha_1, \ldots, \alpha_k)$, $D(\alpha_1, \ldots, \alpha_k)$, is defined as the joint distribution of $(Y_1, \ldots, Y_k)$, where

$$Y_j = Z_j \Big/ \sum_{i=1}^{k} Z_i \quad \text{for } j = 1, \ldots, k \qquad \text{(Eq 3)}$$

Because $\Sigma Y_j \equiv 1$, $D(\alpha_1, \ldots, \alpha_k)$ is a (k-1)-dimensional distribution of $(Y_1, \ldots, Y_{k-1})$ and absolutely continuous with the density function

$$d(y_1, \ldots, y_k; \alpha_1, \ldots, \alpha_k) =$$

$$\frac{\Gamma(\alpha_1+\ldots+\alpha_k)}{\Gamma(\alpha_1)\ldots\Gamma(\alpha_k)} y_1^{\alpha_1-1} y_2^{\alpha_2-1} \ldots y_{k-1}^{\alpha_{k-1}-1} \left(1-\sum_{j=1}^{k-1} y_j\right)^{\alpha_{k-1}}$$

(Eq 4)

In the special case k=2 we have

$$D(\alpha_1, \alpha_2) \equiv Be(\alpha_1, \alpha_2) \qquad \text{(Eq 5)}$$

where $Be(\alpha_1, \alpha_2)$ denotes the Beta distribution. Also, the marginal distribution of each Y_j is

$$Be(\alpha_j, \sum_i \alpha_i - \alpha_j).$$

More generally, arbitrary sums of Y_j-variables are still Dirichlet distributed. For example

$$(Y_1+Y_2, Y_3+Y_4+Y_5, \ldots, Y_k) \sim D(\alpha_1+\alpha_2, \alpha_3+\alpha_4+\alpha_5, \ldots, \alpha_k)$$

The first two moments of the Dirichlet distribution $D(\alpha_1, \ldots, \alpha_k)$ are

$$E(Y_i) = \alpha_i/\alpha,$$

$$E(Y_i^2) = \alpha_i(\alpha_i+1)/[\alpha(\alpha+1)], \qquad \text{(Eq 6)}$$

$$E(Y_i Y_j) = \alpha_i \alpha_j/[\alpha(\alpha+1)],$$

where

$$\alpha = \sum_{i=1}^{k} \alpha_i.$$

3 THE DIRICHLET PROCESS

Let X denote the sample space of interest for the distribution we consider. We say that $(B_1,\ldots,B_k)$ is a measurable partition of X if $B_i \cap B_j = \emptyset$ (empty set) for $i \neq j$, and

$$\bigcup_{j=1}^{k} B_j = X.$$

This simply means that the sets B_i do not intersect each other and their union constitutes the whole sample space.

Definition. Let α be a finite measure - positive and finitely additive - on X. The random probability measure P is a Dirichlet process on X with parameter α if for every $k = 1,2,\ldots$ and measurable partition $(B_1,\ldots, B_k)$ of X, the distribution of $(P(B_1),\ldots, P(B_k))$ is Dirichlet $D(\alpha(B_1),\ldots, \alpha(B_k))$.

Thus α is the parameter of the Dirichlet distribution and need not be a probability measure in the sense that $\alpha(X) = 1$. The probabilities $P(B_1),\ldots, P(B_k)$, which are random, correspond to a distribution through (Eq 1). The randomness of this distribution follows the Dirichlet, $D(\alpha(B_1),\ldots, \alpha(B_k))$.

Let us denote n observations by $X_1,\ldots,X_n$. These are said to be a sample of size n from P if, given $P(C_1),\ldots,P(C_n)$, the events $\{X_1 \in C_1\},\ldots, \{X_n \in C_n\}$ are independent of the rest of the process and are independent among themselves.

Ferguson proves the theorem that makes the Dirichlet process so suitable for Bayesian use. The theorem states that if P is a Dirichlet process with parameter α, and if $X_1,\ldots, X_n$ is a sample from P, then the posterior distribution of P given $X_1,\ldots,X_n$ is also a Dirichlet process with parameter

$$\alpha + \sum_{1}^{n} \delta_{x_i},$$

where δ_x denotes the delta function.

4 ESTIMATION OF A DISTRIBUTION FUNCTION

To repeat our assumptions, we consider a parameter space consisting of all probability measures P on X. We have to choose an action a in some space suffering a loss, L(P, a). Our choice of action can be based on a sample $X_1, \ldots, X_n$ from P, and we seek a Bayes rule with respect to the prior distribution, $P \sim D(\alpha)$ (Dirichlet process with parameter α known).

The loss function is assumed to be of a quadratic type

$$L(P,\hat{F}) = \int [F(t) - \hat{F}(t)]^2 dW(t) \qquad \text{(Eq 7)}$$

where W is a given weight function and $F(t) = P((-\infty,t])$. For the no-sample problem $F(t) \sim Be(\alpha(-\infty,t],\alpha(t,\infty))$ for each t and the Bayes risk is

$$E\{L(P,\hat{F})\} = \int E[F(t) - \hat{F}(t)]^2 dW(t). \qquad \text{(Eq 8)}$$

This Bayes risk is minimized by choosing, for each t,

$$\hat{F}(t) = E\{F(t)\} = F_o(t), \qquad \text{(Eq 9)}$$

where, according to (Eq 6),

$$F_o(t) = a((-\infty,t])/\alpha(X) \qquad \text{(Eq 10)}$$

representing our "best" prior guess at the unknown F(t).

As was stated earlier, the posterior distribution of F(t) given a sample $X_1,\ldots,X_n$ is also a Dirichlet distribution with the parameter

$$\alpha + \sum_i \delta_{x_i}.$$

Therefore, in case of a sample of size n, the Bayes rule is

$$\hat{F}_n(t|X_1,\ldots,X_n) = \frac{\alpha((-\infty,t]) + \sum_{i=1}^{n} \delta_{x_i}((-\infty,t])}{\alpha(X) + n}$$

$$= p_n \cdot F_o(t) + (1 - p_n)F_n(t|X_1,\ldots, X_n), \qquad \text{(Eq 11)}$$

where the mixing factor is

$$p_n = \alpha(X)/(\alpha(X)+n) \qquad \text{(Eq 12)}$$

and the empirical distribution of the sample is

$$F_n(t|X_1,\ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}((-\infty,t]) \qquad \text{(Eq 13)}$$

Thus the Bayes rule is a mixture of the prior guess and the empirical distribution, where the mixing factor is determined by the measure $\alpha(X)$ compared to the sample size n. For a very small $\alpha(X)$ compared to n, the Dirichlet prior is rather "noninformative" and the Bayes estimate converges to the empirical distribution function.

In the application considered here, where the observations are generated by batchwise Monte Carlo simulation, the first batch is used to support our prior guess. If we have no further prior information in addition to the first batch of size n_o, it is quite reasonable to choose $\alpha(X) = n_o$. By that way the information contained in the first batch is given its appropriate weight.

As the application of this estimation technique proceeds one gets successively more knowledge about the type of distribution that reasonably well can approximate the true distribution. Let us at this stage assume that the lognormal type of distribution is a good prior guess. Then, for the choice of the α-measure, a lognormal distribution is by its first and second order moments adjusted to the observation of the first batch. This means that $\alpha(A) = n_o \cdot P_{LN}(A)$, where A is an arbitrary measurable set in X and $P_{LN}(\cdot)$ is the lognormal probability measure. By proceeding in this way one can say that a lognormal distribution, supported by the first batch of observations, represents our very first guess at the shape of the unknown F(t).

5 ESTIMATION OF QUANTILES

The qth quantile t_q of P is defined as

$$P((-\infty,t_q)) \leq q \leq P((-\infty,t_q]) \qquad \text{(Eq 14)}$$

The task is to estimate t_q of an unknown probability measure P or of the corresponding unknown distribution F(t). In /Ferguson, 1973/ it is shown that for a Dirichlet distributed P, $P \varepsilon D(\alpha)$, t_q is unique with probability one, so that t_q is a well-defined random variable.

To estimate t_q means that we try to express our knowledge about t_q in terms of a distribution of t_q. That distribution can be derived from the formula

$$P\{t_q \leq t\} = P\{F(t) > q\} = \int_q^1 \frac{\Gamma(M)}{\Gamma(uM)\Gamma((1-u)M)} z^{uM-1}(1-z)^{(1-u)M-1}dz \qquad \text{(Eq 15)}$$

because, as we have pointed out earlier, F(t) is Beta distributed for given t. In Eq 15, we have

$$M = \alpha(X)$$
$$u = \alpha((-\infty,t])/\alpha(X) = F_o(t) \qquad \text{(Eq 16)}$$

so for given α-measure and q, Eq 15 expresses a function of t. If we are interested in a certain quantile of this distribution, the pth say, we set (Eq 15) equal to p and solve for u. Then the uth quantile of F_o is the searched quantile of the t_q-distribution, in the no-sample situation.

For a sample of size n, the same formulæ apply with α updated to

$$\alpha + \sum_i \delta_{x_i}$$

and the uth quantile determined from the Bayes distribution estimate, given by (Eq 11).

In the application discussed here the estimation of the median ($q = 1/2$) and an upper quantile ($q = 0.95$) is of major interest. The estimation of the distributions of these quantities, according to the procedure above, is therefore a main issue.

6 AN EXAMPLE

To illustrate the methods described above it is quite natural to start with an example where we know the solution. In this case we choose observations obtained by simulating a gamma (10, 0.2)-distribution. The simulation has been performed by using the PROPER package of random variate generators, which is based on the principle of distribution inversion.

According to section 4 the observations are generated batchwise, in this example one hundred observations in each batch (n_o = 100). The first batch is used to provide the initial Dirichlet parameter $\alpha(A) = n_o \cdot P_{LN}(A)$, where $P_{LN}(\cdot)$ denotes a lognormal distribution adjusted to the observations. Thus we have chosen the lognormal distribution to represent our very first guess at the shape of the unknown F(x). From the lognormal distribution we determine a set of quantiles (X) (see Table 1), in which points we subsequently want to look at the estimated distribution with associated uncertainty interval. The median (0.50-quantile) and the upper quantile (0.95-quantile) are also estimated with uncertainty intervals at the level of 90 %.

Table 1 shows the distribution estimates at the given X-points with associated uncertainty bounds on 90% level for various sample sizes. The first part of the Table (after 0 observations) shows the prior distribution - which in reality is based on the first hundred observations while the second part presents the corresponding quantities after 100 additional observations etc. The uncertainty bounds at the estimated median are shown on separate lines in the Table. The last two lines of each subtable reveal the estimates and associated bounds for the median and the 0.95-quantile.

Some of the information in Table 1 is graphically displayed by Figures 1 and 2, where the true distribution is shown by the continuous curve and the dots display the mean value and the associated uncertainty bounds of the distribution at the given X-points. The uncertainty intervals of the estimated median and the estimated 0.95-quantile have also been marked in Figs. 1 and 2.

In Figure 3 we can see the tendency of decreasing uncertainty intervals for the median and the 0.95-quantile for increasing sample sizes, as well as the true values of these quantiles.

7 DISCUSSION AND CONCLUSIONS

It has to be noted that we cannot get uncertainty bounds for the whole distribution by connecting the discrete points, even if these were infinitely close to each other. The only correct interpretion is that for each specific value of X, the probability of having the true distribution within the corresponding uncertainty bounds is 0.9. Bayesian confidence bands for a distribution function assuming a Dirichlet process as prior have been studied by /Breth, 3/.

From the estimate of the distribution function, Eq 11, it is readily seen that it is discrete because the Dirichlet parameter α is changing discontinuously at the observed sample points. As is said in /Dalal, 4/, "it would be appealing to have a prior which increases the probability of a neighbourhood instead". In this reference, some modifications of the Dirichlet prior process are proposed to overcome this defect. In the application which is of interest here, however, the discreteness of the Dirichlet process does not cause any significant disadvantage.

REFERENCES

1 Pörn K, Åkerlund O
Procedures for uncertainty and sensitivity analysis in repository performance assessment.
SKB Technical Report 85-18, October 1985.

2 Ferguson T S
A Bayesian Analysis of Some Nonparametric Problems.
The Annals of Statistics, 1973, Vol 1 No 2, 209-230.

3 Breth M
Bayesian Confidence Bands for a Distribution Function. The Annals of Statistics, Vol. 6, No. 3, p.649-657, 1978.

4 Dalal S R
Nonparametric Bayes Decision Theory.
Bayesian Statistics (Edt. Bernardo, DeGroot, Lindley, Smith)
University Press, Valencia, 1980.

Table 1

Estimates of distribution and quantiles based on data simulated from gamma (10, 0.2).

ESTIMATE OF CDF AND QUANTILE AFTER 0 OBSERVATIONS

X	FL	F	FU	DF
2.141E+01	1.981E-05	5.000E-03	1.917E-02	1.915E-02
2.590E+01	5.813E-03	2.500E-02	5.478E-02	4.897E-02
2.855E+01	2.011E-02	5.000E-02	9.007E-02	6.996E-02
3.195E+01	5.584E-02	1.000E-01	1.533E-01	9.744E-02
3.661E+01	1.381E-01	2.000E-01	2.688E-01	1.307E-01
4.039E+01	2.273E-01	3.000E-01	3.773E-01	1.500E-01
4.392E+01	3.209E-01	4.000E-01	4.814E-01	1.604E-01
4.750E+01	4.181E-01	5.000E-01	5.819E-01	1.638E-01
5.137E+01	5.186E-01	6.000E-01	6.791E-01	1.604E-01
5.587E+01	6.227E-01	7.000E-01	7.727E-01	1.500E-01
6.163E+01	7.312E-01	8.000E-01	8.619E-01	1.307E-01
7.062E+01	8.467E-01	9.000E-01	9.442E-01	9.744E-02
7.902E+01	9.099E-01	9.500E-01	9.799E-01	6.996E-02
8.712E+01	9.452E-01	9.750E-01	9.942E-01	4.897E-02
1.054E+02	9.808E-01	9.950E-01	1.000E+00	1.915E-02
4.750E+01	4.181E-01	5.000E-01	5.819E-01	1.638E-01

	XL	XQ	XU	DX
.50-QUANTIL	4.456E+01	4.750E+01	5.064E+01	6.075E+00
.95-QUANTIL	7.161E+01	7.833E+01	8.854E+01	1.693E+01

ESTIMATE OF CDF AND QUANTILE AFTER 100 OBSERVATIONS

X	FL	F	FU	DF
2.141E+01	8.848E-04	7.500E-03	1.947E-02	1.858E-02
2.590E+01	1.490E-02	3.250E-02	5.539E-02	4.049E-02
2.855E+01	3.130E-02	5.500E-02	8.375E-02	5.244E-02
3.195E+01	6.342E-02	9.500E-02	1.312E-01	6.776E-02
3.661E+01	1.463E-01	1.900E-01	2.372E-01	9.093E-02
4.039E+01	2.292E-01	2.800E-01	3.333E-01	1.042E-01
4.392E+01	3.244E-01	3.800E-01	4.370E-01	1.126E-01
4.750E+01	3.975E-01	4.550E-01	5.131E-01	1.156E-01
5.137E+01	4.920E-01	5.500E-01	6.075E-01	1.155E-01
5.587E+01	6.041E-01	6.600E-01	7.141E-01	1.099E-01
6.163E+01	7.250E-01	7.750E-01	8.218E-01	9.683E-02
7.062E+01	8.516E-01	8.900E-01	9.240E-01	7.236E-02
7.902E+01	9.163E-01	9.450E-01	9.687E-01	5.244E-02
8.712E+01	9.512E-01	9.725E-01	9.884E-01	3.722E-02
1.054E+02	9.904E-01	9.975E-01	1.000E+00	9.584E-03
4.938E+01	4.420E-01	5.000E-01	5.580E-01	1.161E-01

	XL	XQ	XU	DX
.50-QUANTIL	4.656E+01	4.938E+01	5.167E+01	5.112E+00
.95-QUANTIL	7.236E+01	7.919E+01	8.639E+01	1.403E+01

F = estimated distribution at the given X-value

FL = lower uncertainty bound

FU = upper uncertainty bound

DF = FU - FL, uncertainty interval on 90 %-level

XQ = estimated quantile (0.50 and 0.95 respectively)

XL= lower uncertainty bound

XU = upper uncertainty bound

DX = XU - XL, uncertainty interval on 90 %-level

Table 1, cont.

ESTIMATE OF CDF AND QUANTILE AFTER 200 OBSERVATIONS

X	FL	F	FU	DF
2.141E+01	5.887E-04	5.000E-03	1.299E-02	1.241E-02
2.590E+01	2.204E-02	3.833E-02	5.812E-02	3.608E-02
2.855E+01	4.477E-02	6.667E-02	9.185E-02	4.708E-02
3.195E+01	7.896E-02	1.067E-01	1.374E-01	5.840E-02
3.661E+01	1.756E-01	2.133E-01	2.532E-01	7.764E-02
4.039E+01	2.478E-01	2.900E-01	3.338E-01	8.603E-02
4.392E+01	3.572E-01	4.033E-01	4.502E-01	9.303E-02
4.750E+01	4.459E-01	4.933E-01	5.408E-01	9.482E-02
5.137E+01	5.262E-01	5.733E-01	6.200E-01	9.380E-02
5.587E+01	6.316E-01	6.767E-01	7.203E-01	8.870E-02
6.163E+01	7.503E-01	7.900E-01	8.275E-01	7.719E-02
7.062E+01	8.701E-01	9.000E-01	9.269E-01	5.675E-02
7.902E+01	9.278E-01	9.500E-01	9.688E-01	4.106E-02
8.712E+01	9.544E-01	9.717E-01	9.854E-01	3.107E-02
1.054E+02	9.870E-01	9.950E-01	9.994E-01	1.241E-02
4.772E+01	4.513E-01	5.000E-01	5.461E-01	9.483E-02

	XL	XQ	XU	DX
.50-QUANTIL	4.528E+01	4.772E+01	4.994E+01	4.662E+00
.95-QUANTIL	7.229E+01	7.845E+01	8.430E+01	1.201E+01

ESTIMATE OF CDF AND QUANTILE AFTER 300 OBSERVATIONS

X	FL	F	FU	DF
2.141E+01	1.437E-03	6.250E-03	1.380E-02	1.237E-02
2.590E+01	2.431E-02	3.875E-02	5.581E-02	3.150E-02
2.855E+01	4.606E-02	6.500E-02	8.641E-02	4.036E-02
3.195E+01	7.655E-02	1.000E-01	1.257E-01	4.918E-02
3.661E+01	1.727E-01	2.050E-01	2.390E-01	6.630E-02
4.039E+01	2.414E-01	2.775E-01	3.149E-01	7.355E-02
4.392E+01	3.625E-01	4.025E-01	4.431E-01	8.057E-02
4.750E+01	4.564E-01	4.975E-01	5.386E-01	8.215E-02
5.137E+01	5.317E-01	5.725E-01	6.129E-01	8.128E-02
5.587E+01	6.232E-01	6.625E-01	7.009E-01	7.768E-02
6.163E+01	7.294E-01	7.650E-01	7.991E-01	6.964E-02
7.062E+01	8.577E-01	8.850E-01	9.101E-01	5.234E-02
7.902E+01	9.222E-01	9.425E-01	9.603E-01	3.809E-02
8.712E+01	9.594E-01	9.738E-01	9.854E-01	2.600E-02
1.054E+02	9.902E-01	9.963E-01	9.996E-01	9.310E-03
4.757E+01	4.569E-01	5.000E-01	5.391E-01	8.215E-02

	XL	XQ	XU	DX
.50-QUANTIL	4.561E+01	4.757E+01	4.984E+01	4.230E+00
.95-QUANTIL	7.685E+01	7.998E+01	8.404E+01	7.184E+00

ESTIMATE OF CDF AND QUANTILE AFTER 400 OBSERVATIONS

X	FL	F	FU	DF
2.141E+01	1.149E-03	5.000E-03	1.105E-02	9.899E-03
2.590E+01	2.589E-02	3.900E-02	5.420E-02	2.830E-02
2.855E+01	4.533E-02	6.200E-02	8.066E-02	3.534E-02
3.195E+01	7.710E-02	9.800E-02	1.207E-01	4.364E-02
3.661E+01	1.713E-01	2.000E-01	2.301E-01	5.876E-02
4.039E+01	2.437E-01	2.760E-01	3.093E-01	6.569E-02
4.392E+01	3.623E-01	3.980E-01	4.342E-01	7.196E-02
4.750E+01	4.613E-01	4.980E-01	5.348E-01	7.350E-02
5.137E+01	5.415E-01	5.780E-01	6.141E-01	7.259E-02
5.587E+01	6.392E-01	6.740E-01	7.080E-01	6.889E-02
6.163E+01	7.489E-01	7.800E-01	8.098E-01	6.087E-02
7.062E+01	8.683E-01	8.920E-01	9.139E-01	4.556E-02
7.902E+01	9.284E-01	9.460E-01	9.615E-01	3.310E-02
8.712E+01	9.601E-01	9.730E-01	9.837E-01	2.363E-02
1.054E+02	9.922E-01	9.970E-01	9.996E-01	7.451E-03
4.763E+01	4.640E-01	5.000E-01	5.375E-01	7.350E-02

	XL	XQ	XU	DX
.50-QUANTIL	4.583E+01	4.763E+01	4.929E+01	3.459E+00
.95-QUANTIL	7.655E+01	7.927E+01	8.303E+01	6.478E+00

Fig.1 Estimate of cdf and quantiles
prior based on 100 observations

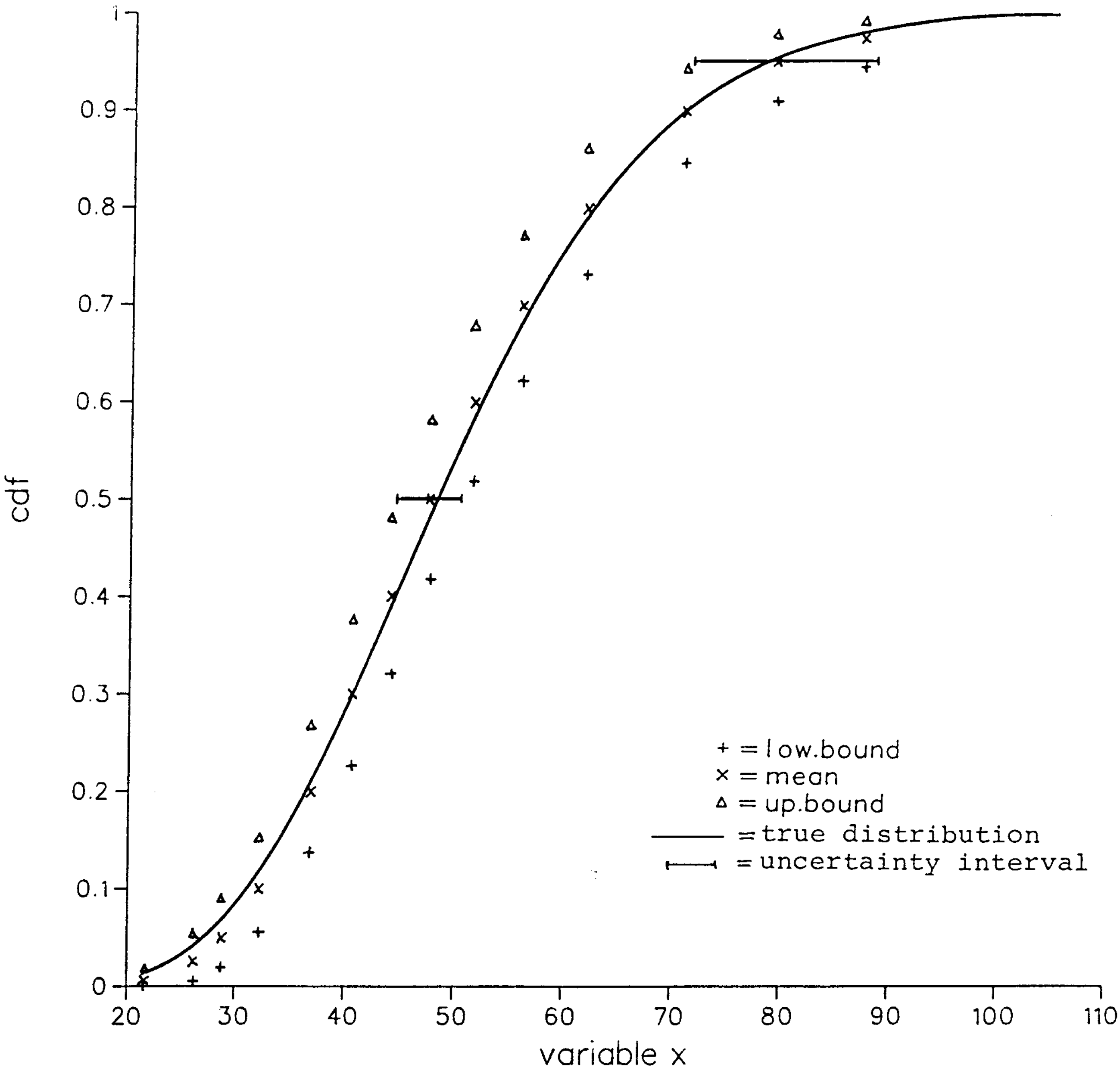


Fig.2 Estimate of cdf and quantiles
400 observations after prior

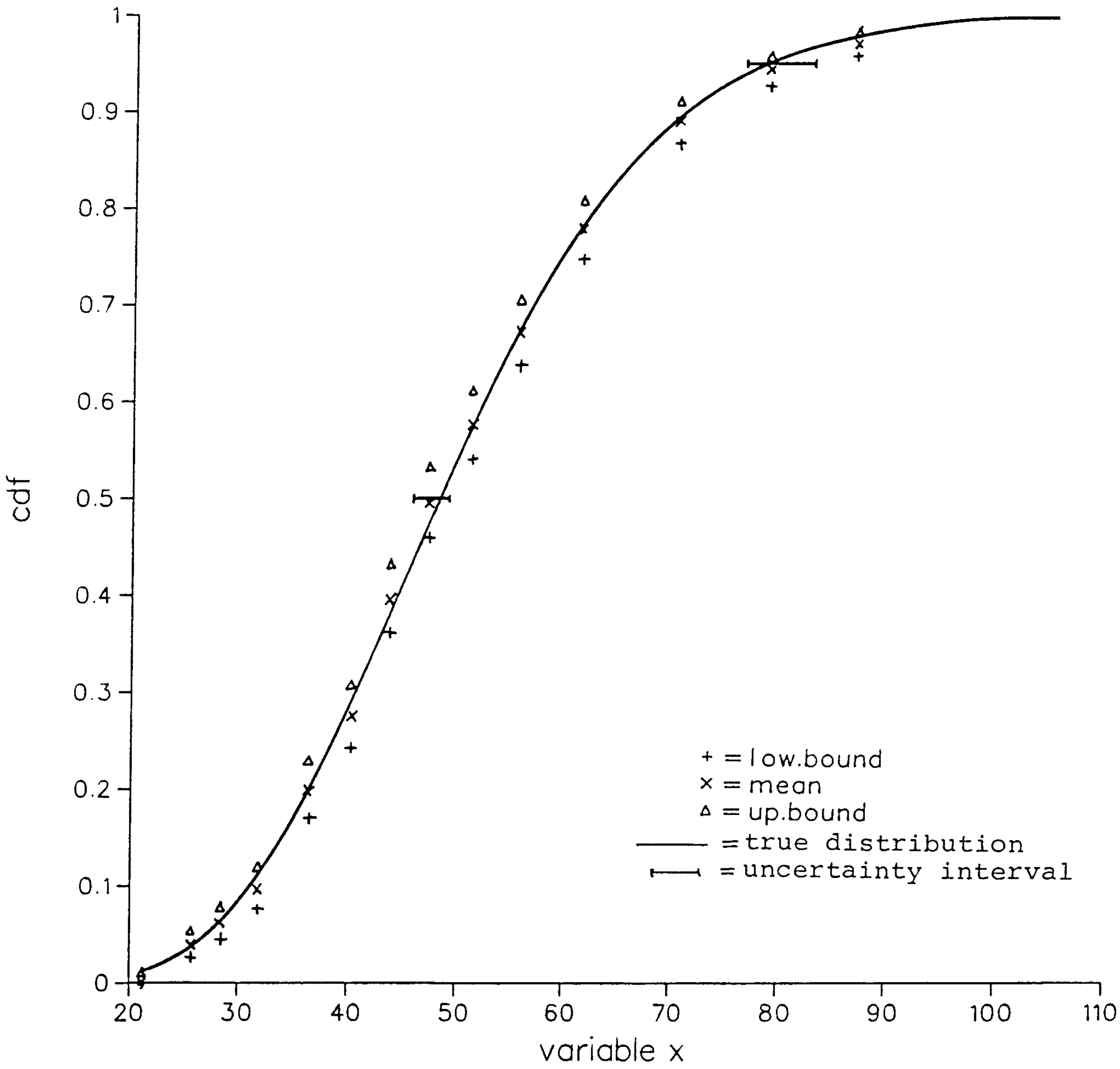
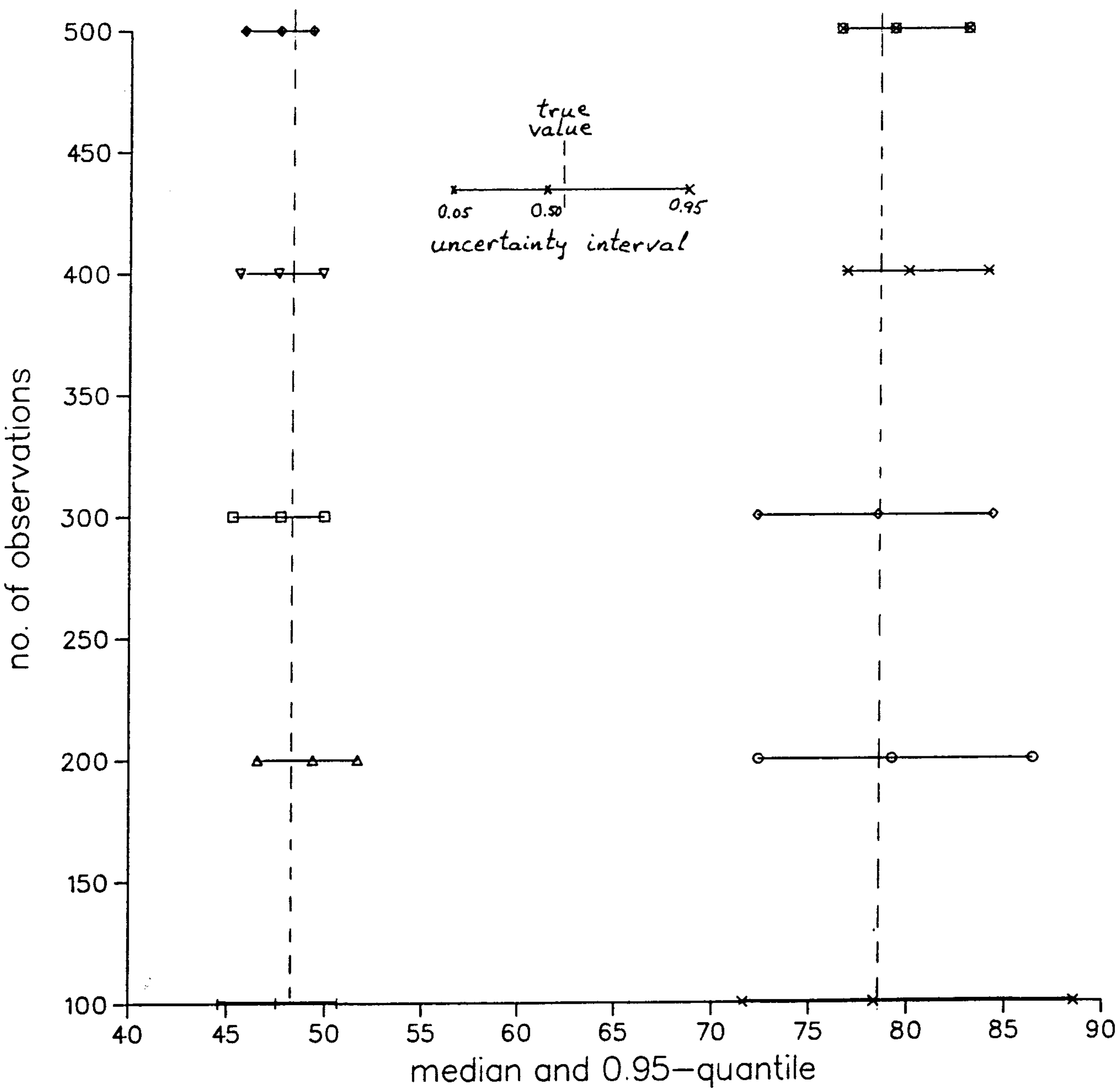


Fig.3 Uncertainty intervals of median and 0.95–quantile vs no. of observations



List of SKB reports

Annual Reports

1977–78
TR 121
KBS Technical Reports 1 – 120.
Summaries. Stockholm, May 1979.

1979
TR 79–28
The KBS Annual Report 1979.
KBS Technical Reports 79-01 – 79-27.
Summaries. Stockholm, March 1980.

1980
TR 80–26
The KBS Annual Report 1980.
KBS Technical Reports 80-01 – 80-25.
Summaries. Stockholm, March 1981.

1981
TR 81–17
The KBS Annual Report 1981.
KBS Technical Reports 81-01 – 81-16.
Summaries. Stockholm, April 1982.

1982
TR 82–28
The KBS Annual Report 1982.
KBS Technical Reports 82-01 – 82-27.
Summaries. Stockholm, July 1983.

1983
TR 83–77
The KBS Annual Report 1983.
KBS Technical Reports 83-01 – 83-76
Summaries. Stockholm, June 1984.

1984
TR 85–01
Annual Research and Development Report 1984
Including Summaries of Technical Reports Issued during 1984. (Technical Reports 84-01–84-19)
Stockholm June 1985.

1985
TR 85-20
Annual Research and Development Report 1985
Including Summaries of Technical Reports Issued during 1985. (Technical Reports 85-01-85-19)
Stockholm May 1986.

1986
TR 86-31
SKB Annual Report 1986
Including Summaries of Technical Reports Issued during 1986
Stockholm, May 1987

1987
TR 87-33
SKB Annual Report 1987
Including Summaries of Technical Reports Issued during 1987
Stockholm, May 1988

Technical Reports

1988

TR 88-01
Preliminary investigations of deep ground water microbiology in Swedish granitic rocks
Karsten Pedersen
University of Göteborg
December 1987

TR 88-02
Migration of the fission products strontium, technetium, iodine, cesium and the actinides neptunium, plutonium, americium in granitic rock
Thomas Ittner[1], Börje Torstenfelt[1], Bert Allard[2]
[1]Chalmers University of Technology
[2]University of Linköping
January 1988

TR 88-03
Flow and solute transport in a single fracture. A two-dimensional statistical model
Luis Moreno[1], Yvonne Tsang[2], Chin Fu Tsang[2], Ivars Neretnieks[1]
[1]Royal Institute of Technology, Stockholm, Sweden
[2]Lawrence Berkeley Laboratory, Berkeley, CA, USA
January 1988

TR 88-04
Ion binding by humic and fulvic acids: A computational procedure based on functional site heterogeneity and the physical chemistry of polyelectrolyte solutions
J A Marinsky, M M Reddy, J Ephraim, A Mathuthu
US Geological Survey, Lakewood, CA, USA
Linköping University, Linköping
State University of New York at Buffalo, Buffalo, NY, USA
April 1987

TR 88-05
Description of geophysical data on the SKB database GEOTAB
Stefan Sehlstedt
Swedish Geological Co, Luleå
February 1988

TR 88-06
Description of geological data in SKBs database GEOTAB
Tomas Stark
Swedish Geological Co, Luleå
April 1988

TR 88-07
Tectonic studies in the Lansjärv region
Herbert Henkel
Swedish Geological Survey, Uppsala
October 1987

TR 88-08
Diffusion in the matrix of granitic rock. Field test in the Stripa mine. Final report.
Lars Birgersson, Ivars Neretnieks
Royal Institute of Technology, Stockholm
April 1988

TR 88-09
The kinetics of pitting corrosion of carbon steel. Progress report to June 1987
G P Marsh, K J Taylor, Z Sooi
Materials Development Division
Harwell Laboratory
February 1988

TR 88-10
GWHRT – A flow model for coupled groundwater and heat flow
Version 1.0
Roger Thunvik[1], Carol Braester[2]
[1] Royal Institute of Technology, Stockholm
[2] Israel Institute of Technology, Haifa
April 1988

TR 88-11
Groundwater numerical modelling of the Fjällveden study site – Evaluation of parameter variations
A hydrocoin study – Level 3, case 5A
Nils-Åke Larsson[1], Anders Markström[2]
[1] Swedish Geological Company, Uppsala
[2] Kemakta Consultants Co, Stockholm
October 1987

TR 88-12
Near-distance seismological monitoring of the Lansjärv neotectonic fault region
Rutger Wahlström, Sven-Olof Linder,
Conny Holmqvist
Seismological Depertment, Uppsala University,
Uppsala
May 1988

TR 88-13
Validation of the rock mechanics HNFEMP code against Colorado school of mines block test data
Ove Stephansson, Tomas Savilahti
University of Luleå, Luleå
May 1988

TR 88-14
Validation of MUDEC against Colorado school of mines block test data
Nick Barton, Panayiotis Chryssanthakis,
Karstein Monsen
Norges Geotekniske Institutt, Oslo, Norge
April 1988

TR 88-15
Hydrothermal effects on montmorillonite. A preliminary study
Roland Pusch
Ola Karnland
June 1988

TR 88-16
Swedish Hard Rock Laboratory
First evaluation of preinvestigations 1986-87 and target area characterization
Gunnar Gustafson
Roy Stanfors
Peter Wikberg
June 1988

TR 88-17
On the corrosion of copper in pure water
T E Eriksen[1], P Ndalamba[1], I Grenthe[2]
[1] The Royal Institute of Technology, Stockholm
Department of nuclear chemistry
[2] The Royal Institute of Technology, Stockholm
Department of inorganic chemistry
March 1988

TR 88-18
Geochemical modelling of the evolution of a granite-concrete-water system around a repository for spent nuclear fuel
Bertrand Fritz, Benoit Madé, Yves Tardy
Université Louis Pasteur de Strasbourg
April 1988