

# A New Method of Measuring the Variability of Estimated Contours by Simulation

Xun-Guo Lin and Alun Pope

Department of Statistics

University of Newcastle

NSW 2308 Australia

Email: {linx, pope}@maths.newcastle.edu.au

**Abstract:** A new method called the coverage plot is proposed to measure and estimate the variability of an estimated contour of a probability density function defined on a region in the plane. Simulation techniques including sectioning and bootstrap methods are compared for a particular problem, which is from the operation of weapons firing ranges. The roles of bias and sample size are also seen in this example.

## 1 INTRODUCTION

Consider a problem in which a probability density function defined over a two-dimensional region is estimated and the outcome of interest is a *contour* of this probability density function. Our goal in this paper is to assess methods of estimating the variability of the estimated contour.

This problem arises frequently for random variables having a spatial distribution, when the question of interest is: for what regions is the probability below some level? For us, this question arose from considering the safety of weapons firing ranges, and the question of interest was: in what regions is the probability that a bullet lands in the region below a specified level? This problem is of practical importance because it is the central issue in assessing the risk to the public and also to the environment resulting from the operation of firing ranges. There has been increasing concern for the safety of the public and military personnel during the operation of weapons firing ranges, which has been expressed through the creation of international working groups such as the International Range Safety Advisory Group and the NATO Range Safety Panel. It also has some features which make it different from usual statistical applications of density estimation. We shall give several examples below, all coming from this source. In this application, we have the advantage that we can compute an approximation to the unknown probability density function by Monte Carlo simulation; thus we can assume that we have available arbitrarily large sample sizes, and in particular, we shall see that it is useful to consider simulating directly the sampling distribution of our estimator. This option will not usually be available in the more traditional statistical setting where one needs to be able to estimate variability of estimates without collecting more data. Nonetheless, it will be seen that our methods will have some relevance for these traditional applications, too. In particular, we shall consider a bootstrap algorithm which will be applicable when the sample has been collected rather than generated by simulation.

Because of the nature of the problem which led to our interest in this question, we need to consider a special type of contour of a probability density function. Let  $f$  be the probability density function of a continuous ran-

dom variable defined on a continuous region in the plane. (In practice we shall deal with discrete approximations below, but here we are discussing the ideal situation.) If  $0 \leq \alpha \leq 1$ , we define the  $\alpha$ -contour of a probability density function  $f$  to be a level set of  $f$  such that the total probability outside the boundary is  $\alpha$ . That is, the  $\alpha$ -contour  $C_\alpha$  satisfies

1. there is  $\ell_\alpha > 0$  such that  $C_\alpha = \{x : f(x) = \ell_\alpha\}$ ;
2.  $\int_{U_\alpha} f = \alpha$ , where  $U_\alpha = \{x : f(x) < \ell_\alpha\}$ .

By the continuity of  $f$ , there is such an  $\ell_\alpha$ , and  $C_\alpha$  will be a curve or set of curves except in pathological cases. We refer to points  $x \in U_\alpha$  as being *outside* the contour  $C_\alpha$  and points  $x$  with  $f(x) > \ell_\alpha$  as being *inside*.

Note that there are two things to estimate here: the probability density  $f$  and the level  $\ell_\alpha$ . A natural way to proceed will be to estimate  $f$  by  $\hat{f}$  say and then obtain  $\hat{\ell}_\alpha$  as the appropriate level of  $\hat{f}$ , so that (2) is satisfied for  $\hat{f}$ . Since this involves estimating the probability outside the required contour, it may be expected that the sampling properties of estimates constructed in this way will differ from those of estimates of the probability density  $f$ . We amplify this in the next section.

It may also be noted that estimation of  $\alpha$ -contours is a natural extension to two-dimensions of estimating quantiles or tails of a univariate density. This extension is not trivial, however, as a quantile is a point but an  $\alpha$ -contour is a curve. Hence the target of our estimation is inherently infinite dimensional, and nonparametric methods will be required (since we do not know the form of  $f$ ).

## 2 SINGLE FIGURE MEASURES OF VARIABILITY

As stated in the introduction, a natural approach to our problem is to estimate  $f$  by  $\hat{f}$  and then construct the  $\alpha$ -contour  $\hat{C}_\alpha$  of  $\hat{f}$ ; we would then use  $\hat{C}_\alpha$  as an estimator of  $C_\alpha$ . We assume throughout that this is the way we construct our estimates.

## 2.1 Approaches based on distances between $f$ and $\hat{f}$

One might hope that knowing how close  $\hat{f}$  is to  $f$ , in some sense, would tell us how close  $\hat{C}_\alpha$  is to  $C_\alpha$  in some — possibly different — sense. Measures of how close  $\hat{f}$  is to  $f$  are typically constructed via integral norms on the space of densities to which  $f$  can belong. Thus typically we would use integrated squared error, or expected integrated squared error  $E\int(f - \hat{f})^2$  as a measure of closeness. Such measures are attractive because estimates of them are well understood and easily obtained as part of standard density estimation procedures. (See the book by Scott, 1992, for example.) Further, estimates of  $\hat{f}$  can be constructed which minimise the estimated measure of closeness.

This seems very satisfactory, but in reality is not, as the goal of getting  $\hat{f}$  close to  $f$  is different from getting  $\hat{C}_\alpha$  close to  $C_\alpha$ . The density estimates  $\hat{f}$  that are good for approximating  $f$  closely will not necessarily be good for approximating  $C_\alpha$ .

The use of a global measure of the variability of  $\hat{f}$  will miss the important point that the variability of  $\hat{C}_\alpha$  will depend on how much smoothing in the low density regions of  $f$  is carried out during the construction of  $\hat{f}$ . Since we need to estimate an integral over the low density region, this is much easier than estimating the values of  $f$  at individual points in the low density region. Note too that the size of  $\alpha$  will be important here: the smaller  $\alpha$  is, the more important such considerations will become.

The main disadvantage of trying to use the variability of estimates of  $f$  as a way of measuring the variability of  $\hat{C}_\alpha$  is that it simply tries to measure something in the wrong place. We are trying to measure the size of an error by looking at the *range* of  $f$  when we should be looking at the *domain*. We therefore abandon approaches based on norms on spaces to which  $f$  might belong and consider an approach based on a natural measure of distance between two sets in the domain of  $f$ .

## 2.2 Approaches based on symmetric difference

Recall that the symmetric difference  $A\Delta B$  between the sets  $A$  and  $B$  is

$$A\Delta B = (A \cup B) \cap (\bar{A} \cup \bar{B})$$

where  $\bar{X}$  denotes the complement of  $X$ . If we want to compare two candidate contours, one method is to look at some measure of the size of the symmetric difference between their interiors. A natural measure of this size is the area of the symmetric difference. Another is the probability content of the symmetric difference. However, there are two difficulties here: although smallness of either of these two measures guarantees closeness of the two contours, there is no information in the measure about where they are close and where they are far apart; secondly, these measures are not tractable analytically. The first of these two difficulties will of course apply to any single figure-of-merit that we try to use. Since in practice we do not expect to be able to estimate equally accurately at all parts of a contour, knowing where we may be close and where not will be important to us.

For these reasons, we are going to introduce a new method of measuring the sampling variability of an estimated contour. But let us describe the techniques of sectioning and bootstrap first.

## 3 SECTIONING AND BOOTSTRAP

We assume for the sake of exposition that we are dealing with a probability density function defined on a compact rectangle in the real plane, and that we have available a method of generating samples from that density. We consider in this part of the paper the method of sectioning (Lewis and Orav, 1989, Chap 9), also known as batching (Kleijnen, 1988, Schmeiser, 1990). We generate  $B$  sections of independent (both within and between sections) observations, all of the same size  $b$ . Thus the sample size  $n = Bb$ . This idea is specifically inspired by the simulation context, in which observations (ie points in the rectangle) are typically generated independently one at a time and stored before being used. The sections are then just the first  $b$  observations, the second  $b$  observations, etc. (The ordering here is the order of generation.) The sectioning method relies on independence of the observations. It can also be employed with a sample obtained by means other than simulation, provided the independence condition is satisfied.

We treat each section as an independent sample of size  $b$  (which is just what the sections are) and compute an  $\alpha$ -contour for each section. This gives us  $B$  independent observations  $\hat{C}_{k\alpha}$ ,  $k = 1, \dots, B$ , from the sampling distribution of  $\hat{C}_\alpha$ , when this estimator is based on a sample of size  $b$ . Thus we have simulated directly the sampling distribution of the estimator  $\hat{C}_\alpha$  with the smaller sample size  $b$ .

Instead of simulating directly the sampling distribution of  $\hat{C}_\alpha$  (with the smaller sample size  $b$ ) by constructing the sectional estimates  $\hat{C}_{k\alpha}$ ,  $k = 1, \dots, B$ , one can instead simulate the bootstrap distribution. (See for example Hall, 1992.) For  $k = 1, \dots, B$ , let  $C_{k\alpha}^*$  be the contour obtained from the  $k$ th resample. (A resample is a sample of size  $n$  drawn independently with replacement and equal probability from the original data.) We then treat the collection of  $C_{k\alpha}^*$  in the same way as the section estimates of the preceding subsection. This technique has the advantage that it uses the correct sample size  $n$  (instead of  $b$  in the case of sectioning); this is also its disadvantage if  $n$  is very large (because of the large CPU time needed). This technique also suffers from bias problems in this context, as we shall see below.

## 4 COVERAGE PLOTS

For each point  $x$  in the plane we can define the coverage probability at  $x$  by

$$CP(x) = P\{x \text{ lies inside } \hat{C}_\alpha\}$$

where  $\hat{C}_\alpha$  is an estimator of the  $\alpha$ -contour. Suppose we have simulated a set  $\{\hat{C}_{k\alpha} : k = 1, \dots, B\}$  of contour estimates (eg by sectioning or bootstrap). Then we can estimate this coverage probability by

$$\widehat{CP}(x) = \text{proportion of } k \text{ such that } x \text{ is inside } \hat{C}_{k\alpha}.$$

This defines a function which we can plot. In practice we shall be working with discrete versions (ie with pixels) so we shall define  $\widehat{CP}$  as a function of pixel location. The result of such a computation can then be presented conveniently as a greyscale plot, as in Figure 2.

In this Figure, the dark grey pixels are those which are inside the contour for all  $k$  ( $B = 200$ ), the white pixels those that are inside for no  $k$ , and the remainder are somewhere in between. (The scale on the right of the Figure shows the precise greyscale mapping.) This picture shows much more vividly the location of high variability regions than do the single figure measures described. In addition, it is easier to compute! However it does have the disadvantage that one cannot easily assign a number to the variability in a region in a way that would lead to a hypothesis test, for example.

The probability plotted in a coverage plot can be thought of as a posterior probability for the event {pixel inside  $C_\alpha$ }.

## 5 EXAMPLES AND DISCUSSION

All of the examples that are discussed below are the results of applying our techniques to the output of a computer program that simulates the firing of many bullets down a firing range and recording their final impacts (after possibly one ricochet). We are interested in describing the density of impact points in terms of the  $\alpha$ -contours of the density of impact points especially for very low  $\alpha$ s. The motivation for using these very low values is that the purpose of the computation is to assess the safety of a firing range (i.e. risk management); hence one is particularly interested in regions of low impact intensity.

Because of the small  $\alpha$ s we use, we have needed large sample sizes and also to employ a weighted sampling scheme to reduce variance. Thus the sample sizes we report are nominal, the effective sample sizes being much larger, because of the weighting.

An important feature of our methods is that we can pool estimates. In fact we estimate the density of impact points by accumulating a histogram of impact points and then applying a linear smoother (Scott's two-dimensional ASH routine, Scott, 1992). This means that the overall density estimate from a number of replications can be obtained as the average of all the estimates for individual replications. However the contours are not linear in this way. Indeed, it was this nonlinearity that led us to the coverage plot idea.

Each pixel in all the figures represents a cell on the firing range measuring 10 metres by 20 metres ( $x$  by  $y$ ). The support of the equivalent convolution kernel (ie filter or mask) is 90 metres by 260 metres ( $x$  by  $y$ ) in all applications of ASH.

A perspective plot of the smoothed probability histogram of 200 million bullets' impacts on the firing range is shown in Figure 1.

We describe first our sectioning examples, then our bootstrap examples.

### 5.1 Coverage plots by sectioning

In Figure 2 we show coverage plot for the  $10^{-3}$ -contours obtained in 200 replications of an experiment in which the sample size was one million. The contour (solid curve) is computed from all the data (ie sample size 200 million). We shall call this *grand* contour, because it is the contour of the grand mean. It will be seen that the grand contour is nicely placed inside the region of high variability (ie the region corresponding to about 60 out of the total 200, value of  $\widehat{CP}$  about 0.3). This is an indication that with a sample size of one million one obtains approximately the same bias as with a sample size of 200 million; hence we may in this case assume the bias has been reduced to negligible levels. We can be confident that the indicated area of high variability represents a properly located confidence region for the  $10^{-3}$ -contour, of a width appropriate to the sample size being used. Since we expect a reduction in variance through increasing the sample size, we may assume that the confidence region is conservative (i.e. too wide). To determine the confidence level for the confidence region is very much more difficult: we would have to interpret  $\widehat{CP}$  values more carefully, take into account nonindependence of the  $\widehat{CP}$  values and allow for the fact that we would require a simultaneous confidence region. This goes beyond the scope of this paper, so we do not pursue it further here.

In Figure 3, the  $10^{-6}$ -contour is plotted for sample sizes of 25 million (dotted) and 200 million (solid). It will be noted that the solid contour lies outside the dotted contour except for the small "island" near (-500, 2900). This Figure, incidentally, demonstrates the need to consider an indicator of variability that shows the spatial distribution of the variability as well as indicating the amount. In parts of the Figure, the two contours are very similar, while in others the main difference is in the greater smoothness of the contour based on the larger sample size.

### 5.2 Coverage plots by bootstrap

In Figure 4 we see a coverage plot as the result of bootstrapping an original sample of size  $n = 300,000$ . This sample size was chosen for computational reasons (see the REMARK in the next subsection). If the sample size had been the one million of the preceding section, the computation would have taken months on a fast workstation! Most of that time would have been spent reading and writing to disk.

When we examine this plot we see that the contour based on the original sample is outside most of the contours produced by the resampling. The reason for this is of course that the histogram produced by the bootstrap resampling has support contained in the support of the original sample histogram. Thus the variation can only be "inwards".

### 5.3 Sectioning versus bootstrap

It is very difficult to compare the bootstrap and sectioning coverage plots, because the two methods are appropriate under different circumstances.

To understand this, we compare the bootstrap and sectioning techniques with the same total sample size

$n$ . Figure 5 shows the result of computing a sectioning-based coverage plot. The sample was divided into  $B = 200$  sections of size  $b = 1,500$ , to give a total sample size of  $n = 300,000$ . The grand  $10^{-3}$ -contour based on all the observations is the solid curve. Bias problems are evident here since the grand contour lies mainly in the region where  $\widehat{CP}$  is zero. The section sample size of  $b = 1,500$  results in section estimates  $C_{k\alpha}$  that are too wiggly to be useful, since it is plain that the grand contour is very wiggly, and hence the estimated contours with the smaller sample size are even worse. (Plotting a few confirms this.)

Compare this with the situation in Figure 4, in which the estimates are based on 500 bootstrap resamples from a sample of size  $n = 300,000$ . There is still some evidence of bias in the estimator based on a sample of size 300,000, but the results are much more satisfactory in this case.

The bias revealed in Figure 5 is of the estimator based on a sample of size only 1,500. We see here the great disadvantage of the sectioning estimator: it uses the wrong sample size. Thus, if the sample size  $n$  and the number of sections  $B$  are chosen so that the section sample size ( $= n/B$ ) is too small, the section estimates may be too biased and too variable to be useful.

On the other hand these examples also show indirectly the disadvantage of the bootstrap estimator. The reason for keeping the sample size down to 300,000 was that this was the largest sample size for which resamples could be computed in a reasonable time. In fact the sectioning method was generally much quicker to compute in this particular problem because of limitations of the hardware configuration employed. Resampling a particular datapoint at random required a random (ie addressed) disk access unless the original sample was small enough to be held in RAM (ie main memory). On the other hand, generating a new data point required solving some differential equations. This could be done in RAM very quickly on a computer on which floating point operations were efficient. The consequence of this was that it was quicker to compute a new point than to find a particular old one from the original sample. (See the REMARK below.) This means that under these circumstances for the estimates using sectioning, the overall sample size  $n$  was limited by the time allowed, while for the bootstrap the sample size was limited by the amount of RAM available. Under these circumstances then the sectioning method outperformed the bootstrap method.

REMARK: Both the sectioning and the bootstrap methods need to have observations generated elsewhere by another code. The CPU time was about one hour to generate a sample of size one million. (For all the computations in this paper, we used an HP 9000/720 workstation with 16 Mb RAM.) Then, we needed about 3 seconds CPU to compute one section estimate.

However, for the bootstrap another story emerged. The maximum sample size was 200,000 if the resampling was done in RAM. The CPU time was 3 seconds per bootstrap (bootstrap sample size 200,000). If a sequential resampling method (generating the random numbers first, combining and sorting them before reading the samples sequentially) was used, the maximum sample size that could be handled in RAM was 300,000. But the time to perform one bootstrap was then 50 seconds CPU. If we removed the restriction imposed by the size of RAM by using a direct access resampling method

(read each sample from disk in direct access mode when a random number is generated), the CPU time for a bootstrap sample size of one million was as much as 14 hours to create one resample!

## 6 CONCLUSION

Which of these methods do we prefer? We regard Figure 2 as our most informative. This plot allows the spatial distribution of variability to be seen clearly, so that regions of high variability can be identified. This would not be possible with use of a single figure-of-merit.

Everywhere that sectioning was used, bootstrap could have been used, on a big enough computer. The limitations caused by our hardware storage strategy make the bootstrap unacceptable to us for the particular problem we have used as illustration, because here sampling is less expensive than resampling. In applications where the data are collected experimentally rather than through simulation, this will rarely be the case, so the bootstrap would probably be favoured for those.

## 7 ACKNOWLEDGMENT

The work was partly supported by a research contract between the Australian Ordnance Council and the University of Newcastle.

## References

- Hall, P. (1992) *The Bootstrap and Edgeworth Expansion*, Springer-Verlag, New York.
- Kleijnen, J.P.C. (1988) *Statistical Tools for Simulation Practitioners*, Marcel Dekker, New York.
- Lewis, P.A.W. and Orav, E.J. (1989) *Simulation Methodology for Statisticians, Operations Analysts and Engineers: Vol. 1*, Wadsworth & Brooks/Cole, Pacific Grove, CA.
- Schmeiser, B. (1990) *Simulation Experiments in Stochastic Models*, D.P. Heyman and M.J. Sobel (eds), North-Holland, Amsterdam.
- Scott, D. W. (1992) *Multivariate Density Estimation*, Wiley, New York.

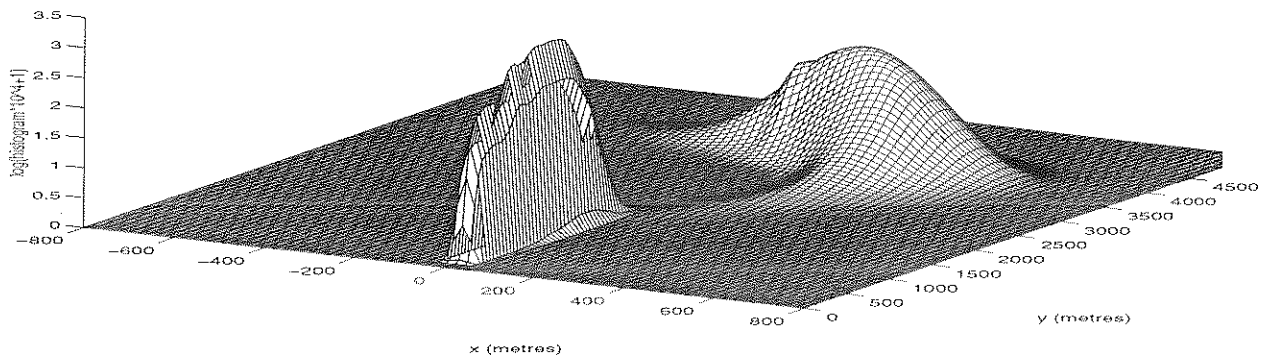


Figure 1: Perspective plot of histogram (logarithmic transformed) of 200 million bullets' impact.

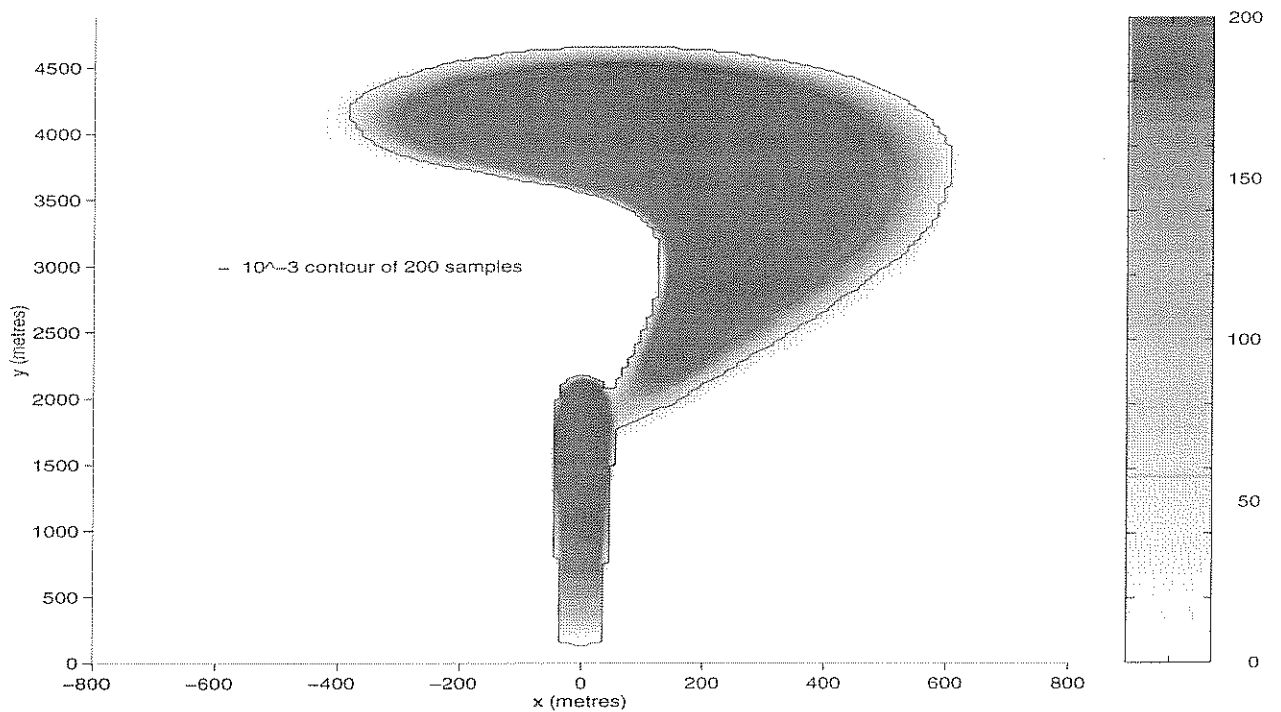


Figure 2: Coverage plot for  $10^{-3}$ -contour from 200 sections ( $B$ ) with sample size ( $b$ ) of one million each. The curve is grand  $10^{-3}$ -contour of overall samples.

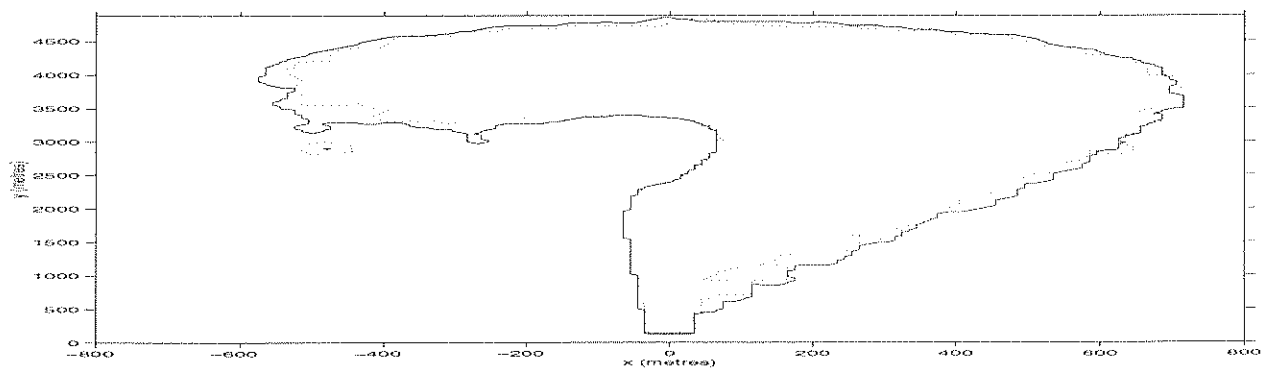


Figure 3:  $10^{-6}$ -contours from sample sizes ( $n$ ) of 25 million (dotted) and 200 million (solid).

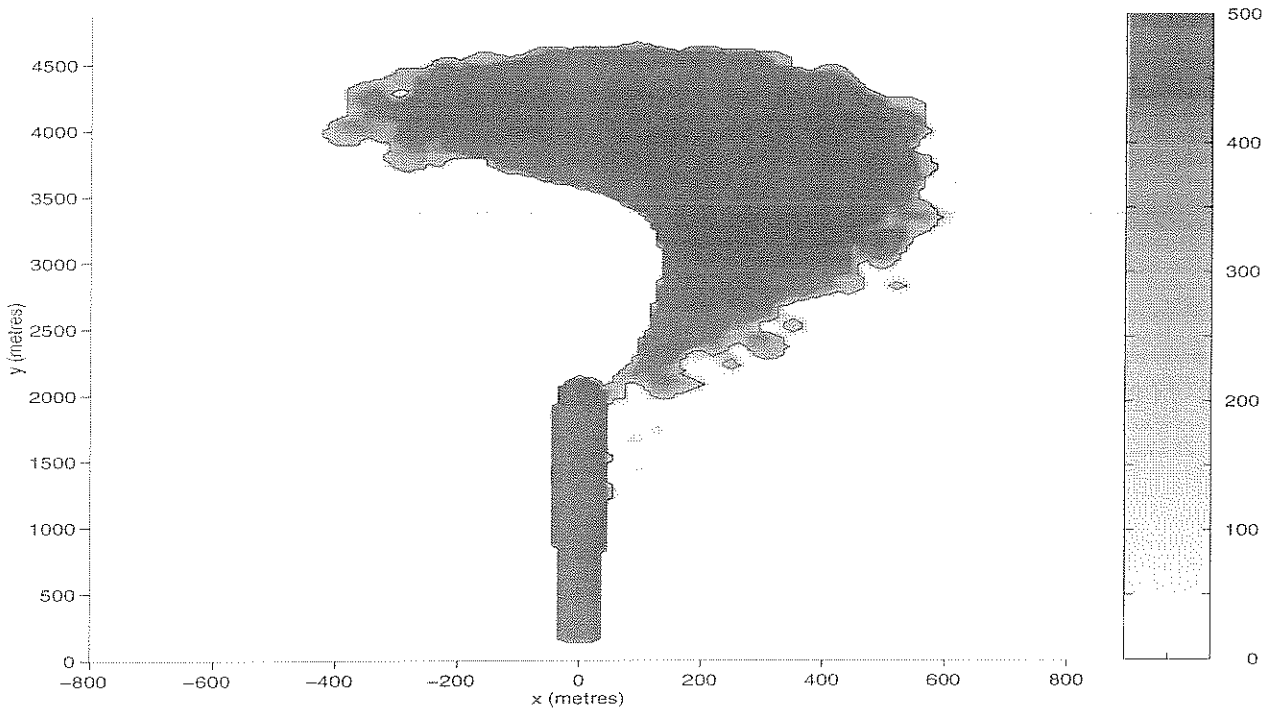


Figure 4: Coverage plot for  $10^{-3}$ -contour from 500 bootstraps with original sample size ( $n$ ) of 300,000. The curve is  $10^{-3}$ -contour of original samples.

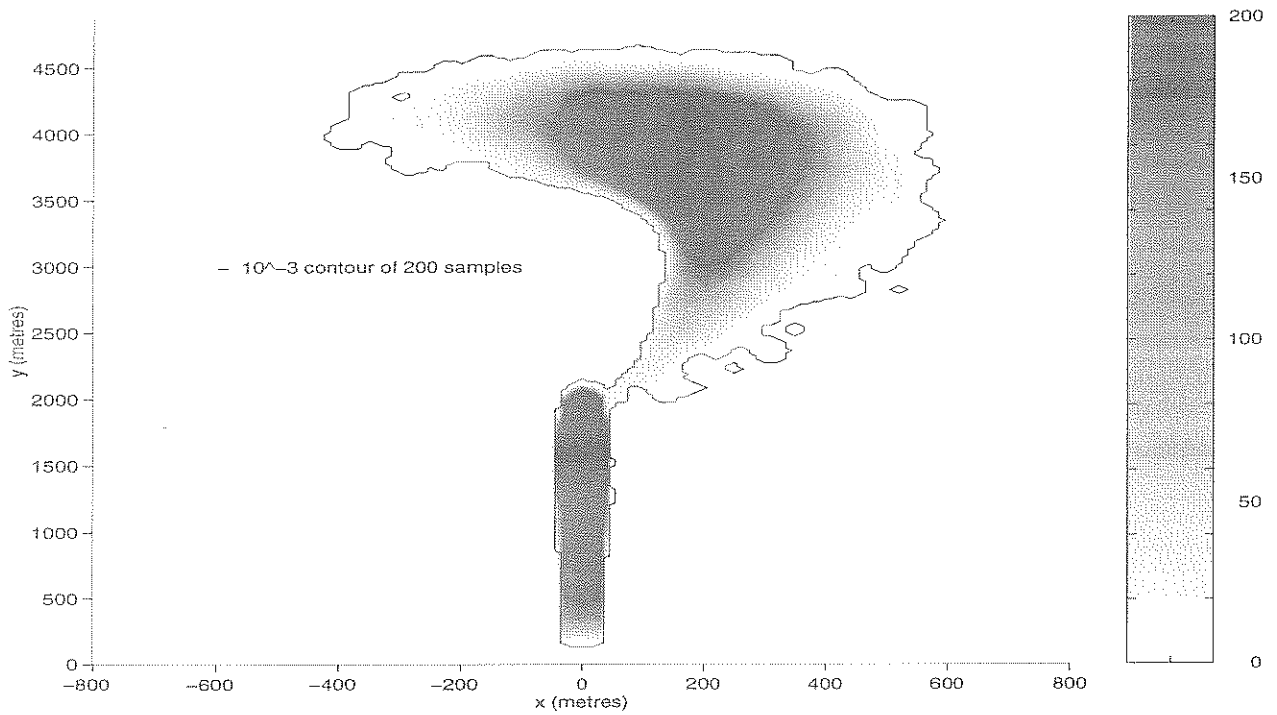


Figure 5: Coverage plot for  $10^{-3}$ -contour from 200 sections ( $B$ ) with sample size ( $b$ ) of 1,500 each. The curve is grand  $10^{-3}$ -contour of overall samples.