METHOD OF MOMENTS ESTIMATION: Simple and Complicated Settings

Michael Sherman and Edward Carlstein
University of North Carolina at Chapel Hill

Abstract:

The method of moments (MOM) is used to estimate parameters in the i.i.d. setup in introductory statistics courses. This method, however, is quickly set aside in favor of the method of maximum likelihood (ML). The main justification for this is the asymptotic efficiency of ML estimates if the assumed model is correct.

We will argue that, in some situations, MOM estimates are actually better than ML estimates for small sample sizes - even if the assumed model is correct. More importantly, the MOM can be extended in a natural way to very general settings where ML would be intractable, and MOM is a robust alternative to ML when we have insufficient knowledge of the underlying population.

Keywords: Maximum Likelihood, Robustness, Asymptotic Efficiency, Subsampling, Variance Estimation

Michael Sherman is a graduate student and Edward Carlstein is Associate Professor in the Department of Statistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599. This research was supported by National Science Foundation Grant #DMS-8902973.

1. INTRODUCTION

Consider the following two problems:

(1) Based on an i.i.d. sample of n observations $\{X_i\}$, we want to estimate $\theta = \mathbf{E}\{e^{X_1}\}$.

(2) Based on observations from a stationary random field on the integer lattice ($\mathbb{Z}^2$), we want to estimate the variance of the α-trimmed mean.

Although both problems are moment estimation problems, there are two fundamental differences between them. In Problem (1), if we choose to specify a likelihood it will be relatively simple due to the i.i.d. assumption, while in Problem (2) the spatial dependence makes any likelihood function more complicated. A second fundamental difference is that in Problem (1) the expectation is of a function ($\exp(\bullet)$) of a single observation (X_1), while in Problem (2) the variance is of a function (trimmed mean) of <u>all</u> the observations in the data set. These differences make estimation more difficult for problems like (2). Problem (2) would be even more difficult if we wanted to estimate the variance of a complicated statistic (e.g., Switzer's adaptive trimmed mean (Efron 1982, p. 28)).

The purpose here is to show that, although these two problems occur in quite different settings, we can use a Method of Moments (MOM) estimator effectively in each. In the i.i.d. setup, Bickel and Doksum (1977, p. 93) say "[MOM estimates] generally lead to procedures that are easy to compute", and "if the sample size is large, [MOM] estimates are likely to be close to the value estimated (consistency)". We show (in Section 3) that these two properties hold, not only for simple problems in the i.i.d. setup , but also for complicated statistics under dependence.

Due to the possibility and the consequences of model misspecification (see Section 2) our MOM approach is nonparametric. We want to estimate θ in Problem (1) without assuming knowledge of F, the distribution of X. For Problem (2), we want to estimate the variance without assuming knowledge of the marginal distribution F, or the dependence mechanism which generated the observations.

In Section 2 we define a MOM estimate of θ and compare it with the Maximum Likelihood estimate derived under the assumption of normality. We compare the two estimates when the data actually are normal and in a case where the data actually are nonnormal. In Section 3 we define an intuitive and consistent MOM estimate for problems like (2) (estimating the moments of a general statistic from dependent data).

2. HOW GOOD IS MAXIMUM LIKELIHOOD UNDER MISSPECIFICATION?

In introductory statistics textbooks, the MOM is quickly set aside in favor of ML (e.g., see Devore (1982) section 6.2). The main reason for this is the asymptotic efficiency of ML estimates if the assumed model is correct. An interesting question is: How "robust" is the ML estimate under misspecification of the likelihood function? For example, if we assume normality when deriving our ML estimate, how will the ML estimate perform when the data actually arise from a different distribution? This question has been studied, from another perspective, by Huber (1981). We address the specific problem of estimating $\theta := \mathbb{E}\{g(X)\}$ where $g(\bullet)$ is some nice function.

Consider the setup discussed in the Introduction in Problem (1): We have n i.i.d. observations $X_1, \ldots, X_n$ and we want to estimate $\theta = \mathbb{E}\{\exp(X_1)\}$, so in this case $g(t) := \exp(t)$. As will be seen, this choice of $g(\bullet)$ allows us to compute all finite sample and asymptotic quantities of interest. Under the assumption that $X_i \sim N(\mu, 1)$, we find $\theta = \exp(\mu + (1/2))$ so that the ML estimate of θ is

$$\hat{\theta}_{MLE} := \exp(\bar{x} + (1/2)), \text{ where } \bar{x} \text{ is the usual sample mean.}$$

A completely nonparametric MOM estimate of θ is:

$$\hat{\theta}_{MOM} := \frac{\sum_{i=1}^{n} \exp(X_i)}{n}.$$

[Note that this is not the typical parametric MOM estimate which would be the same as $\hat{\theta}_{MLE}$ in this case]. In general, the nonparametric MOM estimate for the problem of estimating $\theta := \mathbb{E}\{g(X_1)\}$ is $\sum g(X_i)/n$ (see Hoel, Port and Stone (1971) section 2.6).

We will compare $\hat{\theta}_{MLE}$ and $\hat{\theta}_{MOM}$ in two different situations: (A) the X_i's are distributed $N(\mu, 1)$, and (B) the X_i's have density $f^*(x)$, where $f^*(x) = .697(x-\mu)^2$, $|x-\mu| \leq 1.291$. This latter distribution has the same mean, variance and skewness as the $N(\mu,1)$ distribution, so it is interesting to see how $\hat{\theta}_{MLE}$ will perform under this type of misspecification. As is customary, we will compare the estimates through their MSE's (Mean Square Errors), where $MSE(\hat{\theta}) = Var(\hat{\theta}) + Bias^2(\hat{\theta})$.

Situation A:

Under the assumed normality, for any constant c we have that $\mathbb{E}\{\exp(cX)\} = \exp(c\mu + (c^2/2))$. Using this, one can show that $MSE(\hat{\theta}_{MOM}) = (\exp(2\mu+1))(e-1)/n$ while $MSE(\hat{\theta}_{MLE}) = (\exp(2\mu+1))(\exp(2/n) - 2\exp(1/2n) + 1)$. Listed in Table 1, for different values of n, are the MSE's for each of the two estimates (setting $\theta=1$ for convenience) and the ratio of the two MSE's (which is independent of θ). Notice that even when the X_i's are normally distributed, the nonparametric $\hat{\theta}_{MOM}$ has a smaller MSE than that of the parametric $\hat{\theta}_{MLE}$ for $n \leq 3$. In fact for n=1 we have $Bias(\hat{\theta}_{MOM}) < Bias(\hat{\theta}_{MLE})$ <u>and</u> $Var(\hat{\theta}_{MOM}) < Var(\hat{\theta}_{MLE})$. For larger n this effect wears off and asymptotically we have :

$$\frac{MSE(\hat{\theta}_{MOM})}{MSE(\hat{\theta}_{MLE})} \rightarrow e-1 \approx 1.718.$$

Situation B:

In this case we have the same estimates $\hat{\theta}_{MOM}$ and $\hat{\theta}_{MLE}$, but the MSE's are calculated using the density $f^*(x)$. Let a=.697 and let b=1.291. Then for any constant c we find that $E\{\exp(cX)\}=\exp(c\mu)a[((b^2/c)-(2b/c^2)+(2/c^3))\exp(bc)-((b^2/c)+(2b/c^2)+(2/c^3))\exp(-bc)]$. Using this expression for c=1 and c=2 we find that $\theta=(1.552)\exp(\mu)$ and that $MSE(\hat{\theta}_{MOM})=(\frac{1.537}{n})\exp(2\mu)$. Finally, using this expression for c=1/n and c=2/n we can evaluate $MSE(\hat{\theta}_{MLE})=[\exp(1/2)(E\exp(X/n))^n-\theta]^2+e[(E\exp(2X/n))^n-(E\exp(X/n))^{2n}]$. Table 1 lists the MSE's for the two estimates (setting $\mu=0$ for convenience) and their ratio (which is independent of μ). One can check numerically that $MSE(\hat{\theta}_{MOM}) < MSE(\hat{\theta}_{MLE})$ for each $n \leq 170$. Observe that $\hat{\theta}_{MLE}$ has an asymptotic squared bias of $(e^{.5}-1.552)^2=.009$. Further, by Jensen's inequality, $E\exp(X/n) \geq 1$ so that for any sample size n, $Bias^2(\hat{\theta}_{MLE}) \geq (e^{.5}-1.552)^2$. Thus for $n \geq 171$ we have: $MSE(\hat{\theta}_{MLE}) \geq Bias^2(\hat{\theta}_{MLE}) \geq .009 > 1.537/n = MSE(\hat{\theta}_{MOM})$. Hence $MSE(\hat{\theta}_{MOM}) < MSE(\hat{\theta}_{MLE})$ <u>for all n</u>. Moreover, $\hat{\theta}_{MOM}$ is unbiased for all n and has asymptotic variance equal to 0. This, together with the fact that $\hat{\theta}_{MLE}$ is asymptotically biased, implies:

$$\frac{MSE(\hat{\theta}_{MOM})}{MSE(\hat{\theta}_{MLE})} \rightarrow 0.$$

Thus, asymptotically the MLE is very inefficient.

The example here bears out the efficiency of Maximum Likelihood estimates if the likelihood is correctly specified and if the sample size is sufficiently large (situation A). If the likelihood is misspecified (situation B), however, a simple Method of Moments estimator can be much better.

3. ESTIMATING THE MOMENTS OF A STATISTIC COMPUTED ON A RANDOM FIELD

Now consider the problem of estimating the moments of a general statistic s, when the observations may be dependent. The statistic s may be quite complicated and the presence of any dependence will make the estimation problem even more difficult. Due to insufficient knowledge of F (the marginal distribution of each X_i) and the hazards that come with misspecification (see Section 2), we want to estimate the moments of s without making assumptions on F. Furthermore, if we have insufficient knowledge of the marginal distribution F, it is unrealistic to assume that the joint distribution (i.e., the dependence structure) of the observations is known. For these reasons, we propose a completely nonparametric MOM estimator of θ, analogous to $\hat{\theta}_{MOM}$ in Section 2, which works for a large class of statistics under many different dependence structures. Moreover, our estimator avoids the need for detailed theoretical analysis of the statistic s.

Example:

Suppose that we have equally spaced plants on a large (and possibly irregularly shaped) plot. Let X_j denote the yield of the plant at site j and let D_n be the set of all sites in the plot (see Figure 1). We compute the α-percent trimmed mean (TM_α) yield of the sites as a summary statistic, denoting $s(D_n):=TM_\alpha(X_j : j\epsilon D_n)$. In order to assess the variability of $s(D_n)$, we want an estimate of $\theta_n:=\text{Var}\{s(D_n)\}=\mathbf{E}\{[s(D_n)-\mathbf{E}s(D_n)]^2\}$. The distribution of yields may be complicated (or unknown), and observations close to one another will be dependent due to similar environmental conditions (e.g., moisture in the soil). These factors make the estimation of θ_n a nontrivial problem.

In order to nonparametrically estimate θ_n, using a method analogous to $\hat{\theta}_{\text{MOM}}$ in Section 2, we would like to have k independent plots D_n, compute $s(D_n)$ for each, and then compute the sample variance of these k values. But, because we only have data on a single plot, we need to generate "replicates" of $s(\bullet)$ from D_n. The basic idea is as follows. Compute the statistic s on nonoverlapping subplots $D^i_{l(n)}$ for $i=1,\ldots,k_n$, where $l(n) < n$ determines the common size of each subplot $D^i_{l(n)} \subset D_n$, and k_n denotes the number of subplots (see Figure 1). Then the subplot replicates of s are $s(D^i_{l(n)})=TM_\alpha(X_j : j\epsilon D^i_{l(n)})$, $1 \le i \le k_n$. Let $|D|$ be the cardinality of a set D. Define:

$$\hat{\theta}_n:=\frac{\sum_{i=1}^{k_n} |D^i_{l(n)}|[s(D^i_{l(n)})-\bar{s}_n]^2}{k_n}, \text{ where } \bar{s}_n:=\sum_{i=1}^{k_n}\frac{s(D^i_{l(n)})}{k_n}. \tag{3.1}$$

The nonparametric MOM estimator $\hat{\theta}_n$ is simply the sample variance of the [standardized] replicates $s(D^i_{l(n)})$, $i=1,\ldots,k_n$. This type of MOM estimator has been employed successfully by Carlstein (1986, 1988), Possolo (1986), and Politis and Romano (1992) in simpler scenarios (e.g., for time-series data, rectangular shapes D_n, and linear statistics).

The target parameter here is $\theta:=\lim_{n\to\infty}|D_n|\theta_n$, the asymptotic variance of $s(D_n)$. The following intuitive argument explains why $\hat{\theta}_n$ is a reasonable estimator of θ. Algebraic manipulations show that:

$$\hat{\theta}_n=\sum_{i=1}^{k_n}\frac{t^2(D^i_{l(n)})}{k_n}-\{\sum_{i=1}^{k_n}\frac{t(D^i_{l(n)})}{k_n}\}^2,$$

where $t(D):=|D|^{1/2}(s(D)-\mathbf{E}\{s(D)\})$. Each $D^i_{l(n)}$ is separated from all but a few of the other subplots. This implies that the $t(D^i_{l(n)})$, $i=1,\ldots,k_n$, behave as approximately independent replicates, assuming the dependence is weak at large distances. Therefore, since $\mathbf{E}\{t(D^i_{l(n)})\} \equiv 0$, we expect $\sum_{i=1}^{k_n} t(D^i_{l(n)})/k_n$ to tend towards 0 for large k_n. Similarly, $\sum_{i=1}^{k_n} t^2(D^i_{l(n)})/k_n$ should be close to $\mathbf{E}\{t^2(D^i_{l(n)})\}$ for large k_n. For large $l(n)$, $\mathbf{E}\{t^2(D^i_{l(n)})\}$ is close to θ, because $\lim_{n\to\infty}\mathbf{E}\{t^2(D_n)\}=\theta$. Thus we expect $\hat{\theta}_n\xrightarrow{\mathbf{P}}\theta$, provided $k_n\to\infty$, $l(n)\to\infty$, and provided that dependence decays as distance increases. This is simply a Law of Large Numbers argument (the same argument used to establish consistency of MOM estimates in the i.i.d. case). Essentially the same logic holds here for complicated statistics computed on dependent data.

General Case:

We now address the problem of estimating any moment of a general statistic t. Consider observations from a stationary random field $\{X_i : i \in \mathbb{Z}^2\}$. For each n, let D_n be a finite set of lattice points in $\mathbb{Z}^2$, at which observations are taken. Let f_{D_n} be a function from $\mathbf{R}^{|D_n|}$ to $\mathbf{R}^1$ and let $t(D_n) := f_{D_n}(X_j : j \in D_n)$ be a statistic of interest (f is assumed to be invariant under translations of the set D_n). Assume that $\mathbf{E}\{t^r(D_n)\} \rightarrow \theta \in \mathbf{R}^1$ as $n \rightarrow \infty$, so θ is the asymptotic r^{th} moment of t. Here we give a consistent MOM estimate of the parameter θ. The estimate $\hat{\theta}_n$, analogous to Section 2, is simply the empirical average of subshape replicates $t^r(D^i_{l(n)})$, $i=1,\ldots,k_n$, where $D^i_{l(n)} \subset D_n$; that is,

$$\hat{\theta}_n := \sum_{i=1}^{k_n} \frac{t^r(D^i_{l(n)})}{k_n}. \tag{3.2}$$

We now describe the index sets D_n and $D^i_{l(n)}$ precisely. Let $A \subset (0,1] \times (0,1]$ be the interior of a simple closed curve which will serve as a template for D_n and $D^i_{l(n)}$. To avoid pathological cases we assume that the boundary of A has finite length. Now multiply the set A by n, to obtain the set $A_n \subset (0,n] \times (0,n]$; i.e., A_n is the shape A inflated by a factor of n. The data are observed at the indices in $D_n := \{i \in A_n \cap \mathbb{Z}^2\}$.

Subshape replicates are obtained by partitioning $(0,n]^2$ into $l(n) \times l(n)$ subsquares (of which there are $\lfloor \frac{n}{l(n)} \rfloor^2$). In $(0,l(n)] \times (0,l(n)]$ identify the set $D_{l(n)}$, and do the same in each subsquare by simply translating the origin. Since there is only data in D_n, we only use the k_n subshapes $D^i_{l(n)}$, $i=1,\ldots,k_n$, which are in subsquares completely contained in A_n (see Figure 1). By choosing

$$l(n) = \lfloor \beta n^{\delta} \rfloor \text{ for some } \beta > 0 \text{ and } \delta \in (0,1) \tag{3.3}$$

we get $k_n \rightarrow \infty$ (i.e., an increasing number of replicates in $\hat{\theta}_n$) and $l(n) \rightarrow \infty$ (so that, for each i, $\mathbf{E}\{t^r(D^i_{l(n)})\}$ approaches the target parameter θ).

In order to formally state a consistency result for $\hat{\theta}_n$, the dependence in our random field needs to be quantified. Since our approach is nonparametric, we measure the strength of dependence by a model-free "mixing" coefficient of the type introduced by Rosenblatt (1956). The mixing condition says roughly that observations separated by large distances are approximately independent. Define:

$$\alpha_p(m) := \sup\{\, |\mathbf{P}(AB) - \mathbf{P}(A)\mathbf{P}(B)| : A \in \mathcal{F}(\Lambda_1), B \in \mathcal{F}(\Lambda_2),\ |\Lambda_1| \leq p, |\Lambda_2| \leq p,\ d(\Lambda_1,\Lambda_2) \geq m \}$$

where $\mathcal{F}(\Lambda_i)$ contains the events depending on $\{X_j : j \in \Lambda_i\}$ and $d(\Lambda_1,\Lambda_2)$ is the minimal city-block distance between index sets Λ_1 and Λ_2.

If the observations are independent, then $\alpha_p(m)=0$ for all $m \geq 1$. Here we will need $\alpha_p(m)$ to approach 0 for large m, at some rate depending on the cardinality p. Specifically, we assume:

$$\sup_p \frac{\alpha_p(m)}{p} \leq O(m^{-\epsilon}) \text{ where } \epsilon > \frac{4}{\delta} - 2. \tag{3.4}$$

Condition (3.4) says that, at a fixed distance (m), as the cardinality increases we allow dependence to increase at a rate controlled by p. As the distance increases we need the dependence to decrease at a polynomial rate in m. To interpret the rate (ϵ) recall that δ determines the size of the subshape replicates. For large ϵ, the random field is nearly i.i.d., and in this case δ can be very small; for smaller ϵ, however, we see that δ needs to be larger. This is because small ϵ signifies strong dependence between sites, and the subshape replicates must be relatively large to capture the full extent of this dependence. Mixing conditions of the form (3.4) have been studied and justified by Bradley (1991).

Here is a formal consistency result for the MOM estimator $\hat{\theta}_n$ (a proof is sketched in the Appendix).

<u>Theorem:</u>

Assume that $\mathbf{E}\{t^r(D_n)\} \to \theta \epsilon \mathbf{R}^1$, $l(n)$ is as defined in (3.3), and $\hat{\theta}_n$ is as defined in (3.2).

If $\qquad \sup_n \mathbf{E}\{\,|\,t(D_n)\,|^{r+\gamma}\} < \infty$ for some $\gamma > 0$

and $\qquad$ (3.4) holds,

then

$$\hat{\theta}_n \xrightarrow{\mathbf{P}} \theta \text{ as } n \to \infty.$$

In the agricultural example (above) we can establish consistency of the MOM variance estimator (3.1) by simply applying the Theorem with r=1 and r=2. Note that this MOM variance estimator is consistent for any statistic (not just the α-trimmed mean) satisfying the conditions of the Theorem.

Table 1. MSE's of ML and MOM estimates of $\theta = E\{\exp(X_1)\}$ for Situations A and B

	Situation A			Situation B		
n	MSE($\hat{\theta}_{MOM}$)	MSE($\hat{\theta}_{MLE}$)	Ratio	MSE($\hat{\theta}_{MOM}$)	MSE($\hat{\theta}_{MLE}$)	Ratio
1	1.718	5.092	0.338	1.537	5.191	0.296
2	0.859	1.150	0.747	0.769	2.442	0.315
3	0.573	0.585	0.979	0.512	1.465	0.350
4	0.430	0.382	1.123	0.384	1.019	0.377
10	0.172	0.119	1.446	0.154	0.345	0.445
20	0.086	0.055	1.575	0.077	0.165	0.465
30	0.057	0.035	1.621	0.051	0.111	0.463
40	0.043	0.026	1.645	0.038	0.084	0.456
50	0.034	0.021	1.659	0.031	0.069	0.446
100	0.017	0.010	1.689	0.015	0.039	0.399
∞	0	0	1.718	0	.0094	0

Plot D_n with Subplot Replicates $D^i_{l(n)}$

$n=100,\ l(n)=10,\ k_n=67,\ |D_n|=7440,\ |D_{l(n)}|=68$

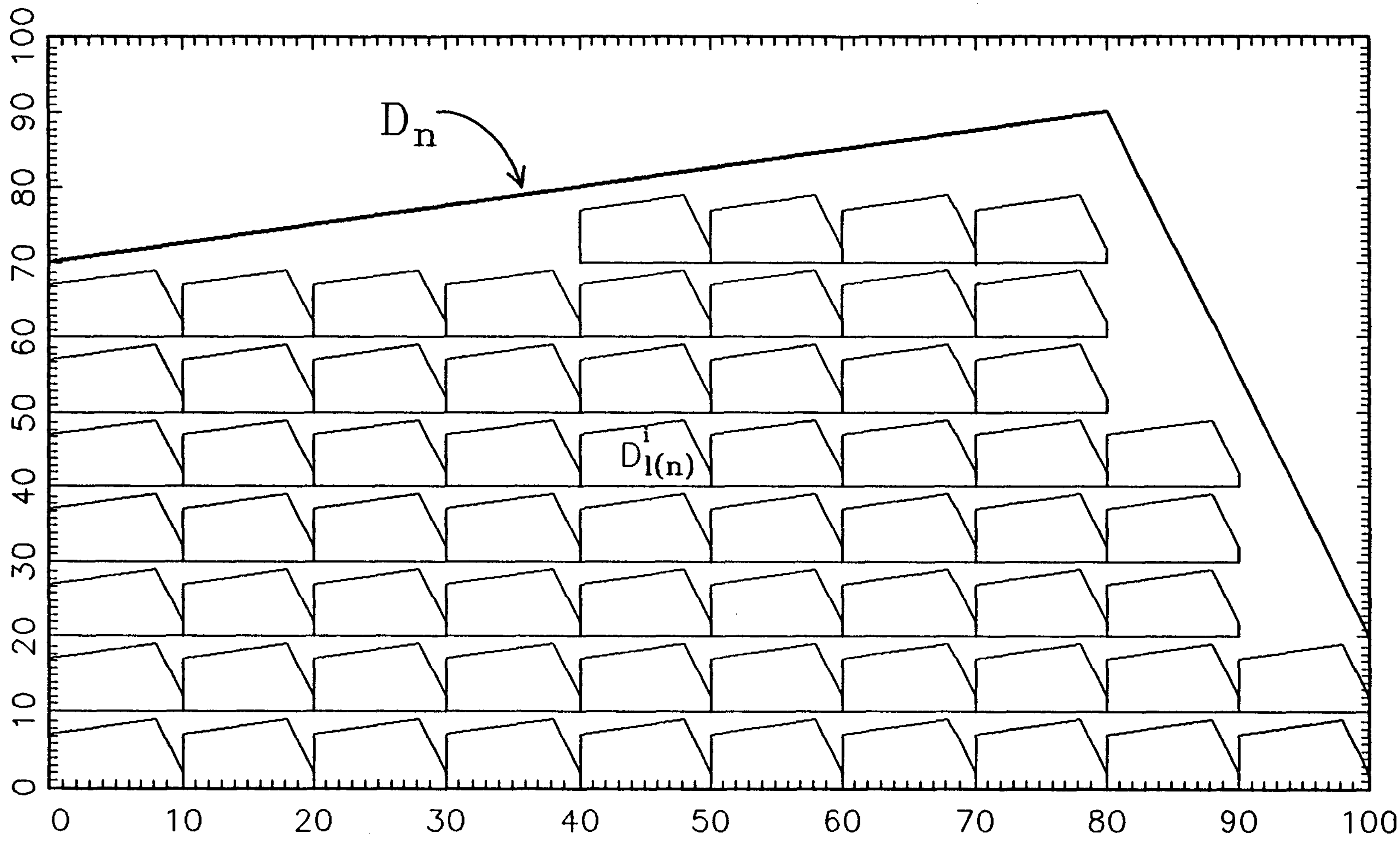

Figure 1

APPENDIX: PROOF OF THE THEOREM

Without loss of generality, take r=1. Let $\lambda(A)$ denote the area of A and let $\| \partial A \|$ denote the length of its boundary. Let μA_n be the number of subsquares that intersect with A_n's boundary. Then $| D_n | \leq l^2(n)(k_n+\mu A_n)$; by Apostol (1957, Theorem 10-42(e)) $\mu A_n \leq 4+4 \| \partial A \| n/l(n)$; and from Lemma A1 of Rudemo, et al. (1990) we have $| D_n | /n^2 \rightarrow \lambda(A) > 0$. These three facts, together with the definition of $l(n)$, imply that $k_n \rightarrow \infty$ (actually $k_n l(n)/n \rightarrow \infty$).

Group the $l(n) \times l(n)$ subsquares in $(0,n]^2$ into disjoint "blocks" of 4, each block being $2l(n) \times 2l(n)$; label the 4 subsquares within a block (1,2,3,4), beginning with "1" in the lower left and proceeding clockwise through the block. Let k_n^j denote the number of usable subshapes with label j (j=1,2,3,4), and denote by $D_{l(n)}^{i,j}$, $i=1,\ldots,k_n^j$, the i^{th} subshape with label j. We have:

$$\hat{\theta}_n = \sum_{j=1}^{4} \sum_{i=1}^{k_n^j} \frac{t(D_{l(n)}^{i,j})}{k_n}.$$

Note that $k_n^j/k_n \rightarrow 1/4$, since $| k_n^u - k_n^v | \leq \mu A_n = o(k_n)$ (this last equality follows from the previous paragraph). Thus it suffices to prove that:

$$T_n := \sum_{i=1}^{k_n^1} \frac{t(D_{l(n)}^{i,1})}{k_n^1} \xrightarrow{\mathbf{P}} \theta;$$

an analogous argument applies to j=2,3,4.

Now let $t_{n,i}^*$, $i=1,\ldots,k_n^1$, have the same marginal distributions as $t(D_{l(n)}^{i,1})$, $i=1,\ldots,k_n^1$, but such that the $t_{n,i}^*$'s are independent for fixed n. Note that $d(D_{l(n)}^{u,1}, D_{l(n)}^{v,1}) \geq l(n)$ for $u \neq v$. Let $\phi_n^*(s)$ and $\phi_n(s)$ be the characteristic functions of

$$T_n^* := \sum_{i=1}^{k_n^1} \frac{t_{n,i}^*}{k_n^1}$$

and T_n, respectively. Then we have:

$$| \phi_n^*(s) - \phi_n(s) | \leq 16 k_n^1 \alpha_{n^2}(l(n)) \leq 16(n/l(n))^2 \alpha_{n^2}(l(n)) \leq C n^{4-\delta(2+\epsilon)} \rightarrow 0.$$

The first inequality follows from a natural extension of Theorem 17.2.1, Ibragimov & Linnik (1971), and from their "telescoping" argument (p. 338). The third inequality follows from (3.4) and from the definition of $l(n)$. It now suffices to show $T_n^* \xrightarrow{\mathbf{P}} \theta$; this follows from the same argument as in Carlstein (1988, p. 297).

REFERENCES:

Apostol, T.M. (1957). Mathematical Analysis: A Modern Approach to Advanced Calculus, Addison-Wesley, Reading.

Bickel, P.J. and Doksum, K.A. (1977). Mathematical Statistics, Holden Day Inc., San Francisco.

Bradley, R.C. (1991). Some Examples of Mixing Random Fields, Technical Report 342, Center for Stochastic Processes, Dept. of Statistics, University of North Carolina, Chapel Hill.

Carlstein, E. (1986). The Use of Subseries Values for Estimating the Variance of a General Statistic from a Stationary Sequence, Annals of Statistics, 14, 1171-1179.

Carlstein, E. (1988). Law of Large Numbers for the Subseries Values of a Statistic from a Stationary Sequence, Statistics, 19, 295-299.

Devore, J.L. (1982). Probability and Statistics For Engineering and the Sciences, Brooks and Cole Publishing Company, Monterey.

Efron, B. (1982). The Jackknife, the Bootstrap and Other Resampling Plans, SIAM, Philadelphia.

Hoel, P.G., Port, S.C. and Stone, C.J. (1971). Introduction to Statistical Theory, Houghton-Mifflin, Boston.

Huber, P.J. (1981). Robust Statistics, John Wiley and Sons, New York.

Ibragimov, I.A. and Linnik, Y.V. (1971). Independent and Stationary Sequences of Random Variables, Wolters-Noordhoff Publishing, Groningen.

Politis, D.N. and Romano, J.P. (1992, to appear). On the Sample Variance of Linear Statistics Derived from Mixing Sequences, Stochastic Processes and Their Applications.

Possolo, A. (1986). Subsampling a Random Field, Technical Report 78, Dept. of Statistics, University of Washington, Seattle.

Rosenblatt, M. (1956). A Central Limit Theorem and a Strong Mixing Condition, Proceedings of the National Academy of Sciences, 42, 43-47.

Rudemo, M., Skovgaard, I. and Stryhn, H. (1990). Maximum Likelihood Estimation of Curves in Images, Technical Report 90-4, Dept. of Mathematics and Physics, Royal Veterinary and Agricultural University, Copenhagen.