Statistics, Probability and Game Theory
IMS Lecture Notes – Monograph Series (1996) Volume 30

EXPONENTIAL PROBABILITY INEQUALITIES WITH SOME APPLICATIONS

George G. Roussas
University of California, Davis

Abstract.

A brief review of the Bennett and Hoeffding inequalities is presented, as they apply to independent random variables, for the purpose of identifying the point where independence is actually utilized. On the basis of an observation, it follows that the inequalities remain in force whenever the expectation of a certain product is bounded by the product of the expectations of the factors involved. This requirement is satisfied, for example, when the underlying random variables are negatively associated. By a counterexample, it is demonstrated that the inequalities need not hold for positively associated random variables. Next, a Hoeffding-type inequality is established for a strong mixing sequence of random variables. The paper is concluded with the utilization of the Hoeffding inequality in order to construct a minimum distance estimate of the probability measure governing a sequence of negatively associated random variables.

1. Introduction and Summary. Let $X_1, X_2, \ldots$ be (real-valued) random variables (r.v.) defined on the underlying probability space $(\Omega, \mathcal{A}, P)$, and set S_n for the sum of the first n r.v.'s, $\sum_{i=1}^{n} X_i$, and $\bar{S}_n$ for S_n/n. The problem of providing exponential bounds for the probabilities $P(|S_n| \geq \varepsilon)$ $(\varepsilon > 0)$ is of paramount importance, both in Probability and Statistics. From a statistical viewpoint, such inequalities can be used, among other things, for the purpose of providing rates of convergence (both in the probability sense and almost surely) for estimates of various quantities. Especially so in a nonparametric setting, where the advantages of a parametric structure are not available to the investigator.

In Section 2, a brief review is presented of the Bennett and the Hoeffding inequalities in the framework of independent r.v.'s, primarily for the purpose of isolating the point, where independence is utilized. This point is inequality (2.12) stated as a corollary to two propositions. In the following section, it is shown that inequality (2.12) is, indeed, satisfied for r.v.'s which are negatively associated. As a consequence of it, such r.v.'s satisfy the Bennett and Hoeffding inequalities. It is then shown, by means of a counterexample, that positively associated r.v.'s need not, in general, satisfy the Hoeffding inequality. This conclusion is an easy consequence of a result for positively associated r.v.'s stated as Proposition 3.1. In the following section, Section 4, a Hoeffding-type inequality is established for a specific mode of mixing sequences of r.v.'s, the so-called α-mixing or strong mixing. This is done in Theorem 4.1 after two auxiliary results are stated as lemmas. The paper is closed with some statistical applications of the Hoeffding inequality for α-mixing and negatively associated r.v.'s.

In this paper, all limits are taken as $n \to \infty$; this fact will not be mentioned explicitly in order to avoid unnecessary repetitions.

2. Bennett and Hoeffding Inequalities. Exponential probability bounds for sums of r.v. are very useful in many probabilistic derivations, and particularly so in many aspects of parametric as well as nonparametric statistical inference. Bennett (1962) provided various forms of such bounds in the framework of independent r.v.'s (see Proposition 2.1) and so did Hoeffding (1963) by way of a different method (see Proposition 2.2). These inequalities are Bernstein-type inequalities, and Bennett's derivations follow the pattern of the original proof of Bernstein's inequality.

Following the steps of the proofs of the above mentioned inequalities, the reader may see that there is essentially only one instance, where independence of the underlying r.v.'s is used. On the basis of this observation, it follows, in effect, that the same bounds hold true in all cases, where such an inequality may be claimed. For the purpose of clarifying this point, we present a brief outline of the basic steps involved in the proofs, and point out explicitly where independence is employed. We start out with one form of Bennett's inequalities.

Proposition 2.1 (Bennett (1962), page 34). *Let $X_1, \ldots, X_n$ be independent r.v.'s almost surely (a.s.) bounded, $|X_i| \leq C_i$ a.s. $i = 1, \ldots, n$, and, without loss of generality, assume them to be centered at their expectations. Set $\sigma_i^2 = \sigma^2(X_i) = \mathcal{E}X_i^2$, $s_n^2 = \sum_{i=1}^n \sigma_i^2 = \sigma^2(S_n)$, where $S_n = \sum_{i=1}^n X_i$. Finally, let $C_0 = \max\{C_i; i = 1, \ldots, n\}$. Then, for every $t > 0$:*

$$P(S_n \geq s_n t) \leq \exp\left[-t^2/(2 + \frac{2}{3} \cdot \frac{C_0 t}{s_n})\right], \tag{2.1}$$

and also:

$$P(|S_n| \geq s_n t) \leq 2\exp\left[-t^2/(2 + \frac{2}{3} \cdot \frac{C_0 t}{s_n})\right]. \tag{2.2}$$

Proof. (brief outline). For any $c > 0$, use Markov inequality to obtain:

$$P(S_n \geq s_n t) \leq \exp(-cs_n t)\mathcal{E}\exp(cS_n) = \exp(-cs_n t)\mathcal{E}\prod_{i=1}^n \exp(cX_i). \tag{2.3}$$

Expanding $\exp(cX_i)$ according to Taylor's formula, taking expectations, and using suitable inequalities involving moments and the function $\exp t$, one has:

$$\mathcal{E}\exp(cX_i) \le \exp\left[\frac{c^2\sigma_i^2}{2}F_i(c)\right], \text{ where } F_i(c) = \sum_{r=2}^{\infty}\left(c^{r-2}\mathcal{E}\left|X_i\right|^r / \frac{1}{2}r!\sigma_i^2\right),$$

for $i = 1\ldots,n$. At this point, suppose (as Bernstein did) that, for $i = 1,\ldots,n$:

$$\mathcal{E}\left|X_i\right|^r \le \frac{1}{2}\sigma_i^2 W_n^{r-2} r!, \quad r \ge 2, \quad W_n = C_0/3.$$

Actually, it is easily seen that these inequalities are satisfied here, and then: $F_i(c) \le (1-cW_n)^{-1}$, provided $cW_n < 1$. In addition to satisfying this last inequality, c is also chosen so that:

$$(1-cW_n)^{-1} \le M_n, \text{ where } \mathrm{M}_n = W_n s_n^{-1}t + 1 = \frac{C_0 t}{3s_n} + 1.$$

Then:

$$\mathcal{E}\exp(cX_i) \le \exp\left(\frac{c^2\sigma_i^2}{2}M_n\right), \quad i = 1,\ldots,n. \tag{2.4}$$

It is at this point where *independence* of the X_i*'s* is used in order to obtain, by means of (2.3) and (2.4):

$$\begin{aligned} P(S_n \ge s_n t) &\le \exp(-cs_n t)\mathcal{E}\prod_{i=1}^{n}\exp(cX_i) \\ &= \exp(-cs_n t)\prod_{i=1}^{n}\mathcal{E}\exp(cX_i) \le \exp\left(\frac{c^2 s_n^2}{2}M_n - cs_n t\right). \end{aligned} \tag{2.5}$$

The proof is then completed by minimizing (with respect to c) the right-hand side in (2.5). ■

Remark 2.1. To be sure, independence of the X_i's also enters the picture, if one wishes to interpret s_n^2 as $\sigma^2\left(\sum_{i=1}^n X_i\right)$ rather than as $\sum_{i=1}^n \sigma^2(X_i)$. However, this does not interfere with the proof. The critical point in establishing the inequalities is relation (2.12) below. In all that follows, s_n^2 will stand for the sum of the variances.

Now consider one of the inequalities obtained by Hoeffding.

Proposition 2.2. *(Theorem 2 in Hoeffding (1963)). Let $X_1,\ldots,X_n$ be independent r.v.'s such that $a_i \le X_i \le b_i$, $i = 1,\ldots,n$. Set $\mu_i = \mathcal{E}X_i$, $i = 1,\ldots,n$ and $\mu = n^{-1}\sum_{i=1}^n \mu_i$. Then, for every $t > 0$:*

$$P\left(\bar{X} - \mu \ge t\right) \le \exp\left[-2n^2t^2 / \sum_{i=1}^{n}\left(b_i - a_i\right)^2\right], \tag{2.6}$$

and also:

$$P\left(|\bar{X}-\mu|\geq t\right)\leq 2\exp\left[-2n^2t^2/\sum_{i=1}^{n}(b_i-a_i)^2\right]. \tag{2.7}$$

Proof. (brief outline). A brief outline of the proof is presented for the purpose of pointing out the only instance where independence is used. The approach is different from that used by Bennett, and is based on convexity properties. To this end, for any arbitrary but fixed $c \in \Re$, the function $g(x) = e^{cx}$ is convex (in $x \in \Re$). Therefore, for each $i = 1, \ldots, n$:

$$\mathcal{E}\exp(cX_i)\leq \frac{b_i-\mu_i}{b_i-a_i}\exp(ca_i)+\frac{\mu_i-a_i}{b_i-a_i}\exp(cb_i), \tag{2.8}$$

and:

$$\begin{aligned}\mathcal{E}\exp\left[c(X_i-\mu_i)\right]&\leq\exp\left[-c(\mu_i-a_i)\right]\exp(ca_i)\\&\quad\cdot\exp\left\{\ell n\left[\frac{b_i-\mu_i}{b_i-a_i}\exp(ca_i)+\frac{\mu_i-a_i}{b_i-a_i}\exp(cb_i)\right]\right\}\\&=\exp\left[-kp+\ell n(1-p+pe^{-k})\right]=\exp\left[L(k)\right],\end{aligned}$$

where: $k = c(b_i - a_i)$, $p = \frac{\mu_i - a_i}{b_i - a_i}$ (so that $1 - p = \frac{b_i-\mu_i}{b_i-a_i}$), and $L(k) = -kp + \ell n(1 - p + pc^k)$. It follows that: $L(0) = L'(0) = 0$, and $L''(k) = u(1-u)$, $0 \leq u \leq \frac{(1-p)e^{-k}}{(1-p)e^k+p} \leq 1$, so that $L''(u) \leq 1/4$. Expand $L(k)$ according to Taylor's formula up to terms involving the second derivative, and use the above results to obtain: $L(k) \leq \frac{k^2}{8} \leq c^2(b_i - a_i)^2/8$. Therefore:

$$\mathcal{E}\exp\left[c(X_i-\mu_i)\right]\leq\exp\left[c^2(b_i-a_i)^2/8\right],\quad i=1,\ldots,n. \tag{2.9}$$

The Markov inequality yields, for $c > 0$:

$$P\left(\bar{X}-\mu\geq t\right)\leq\exp(-cnt)\mathcal{E}\prod_{i=1}^{n}\exp\left[c\left(X_i-\mu_i\right)\right]. \tag{2.10}$$

Once again, it is at this point where *independence* is employed in order to get, by means of (2.9) and (2.10):

$$\begin{aligned}P\left(\bar{X}-\mu\geq t\right)&\leq\exp(-cnt)\mathcal{E}\prod_{i=1}^{n}\exp\left[c\left(X_i-\mu_i\right)\right]\\&=\exp(-cnt)\prod_{i=1}^{n}\mathcal{E}\exp\left[c\left(X_i-\mu_i\right)\right]\\&\leq\exp\left[\frac{c^2}{8}\sum_{i=1}^{n}(b_i-a_i)^2-cnt\right].\end{aligned} \tag{2.11}$$

The desired result then follows by minimizing (with respect to c) the right hand side in (2.11). ■

From (2.5) and (2.11), the following result follows.

Corollary 2.1. *In Propositions 2.1 and 2.2, consider the assumptions made there except for the* **independence** *assumption of the r.v.'s $X_1, \ldots, X_n$, which is replaced by the requirement that:*

$$\mathcal{E}\prod_{i=1}^{n}\exp(\pm cX_i) \leq \prod_{i=1}^{n}\mathcal{E}\exp(\pm cX_i), \quad c > 0. \tag{2.12}$$

Then the conclusions hold true.

An important case, where inequality (2.12) is satisfied, is the case of negatively associated r.v.'s to be considered in the next section.

For the sake of a comparison of Bennett and Hoeffding inequalities, suppose that $|X_i| \leq C$, $i = 1, \ldots, n$, and in (2.2), replace t by nt/s_n. Then inequalities (2.2) and (2.7) become, respectively:

$$P(|\bar{S}_n| \geq t) \leq 2\exp\left(-\frac{3n^2t^2}{6s_n^2 + 2Cnt}\right), \tag{2.13}$$

$$P(|\bar{S}_n| \geq t) \leq 2\exp\left(-\frac{nt^2}{2C^2}\right). \tag{2.14}$$

It follows that, if $t < 3C - \frac{3s_n^2}{nC}$, then Bennett's bound is sharper than Hoeffding's bound. In particular, suppose that X_i, $i = 1, \ldots, n$ have the same second moment σ^2. Then $3C - \frac{3s_n^2}{nC} = 3\left(C - \frac{\sigma^2}{C}\right) > 0$, and Bennett's bound is sharper than Hoeffding's bound, provided $t < 3\left(C - \frac{\sigma^2}{C}\right)$. Of course, the sharpness of the bounds is reversed for $t > 3C - \frac{3s_n^2}{nC}$, or $t > 3\left(C - \frac{\sigma^2}{C}\right)$ for the case of equal variances. In practice, it is more convenient to employ the Hoeffding inequality. An inequality similar to the ones discussed here, albeit in a different context, was obtained by Blackwell and Freedman (1973). Indeed, much of David Blackwell's work is permeated by inequalities; especially that portion of it referring to dynamic programming (see, for example, Blackwell (1962), (1965)).

3. Associated Random Variables. The concept of negative association has been introduced by Joag-Dev and Proschan (1983) and has found significant applications in systems reliability, statistics, and may also be appropriate to model certain biosystems and ecosystems. For the definition of the concept, consider the set $\{1, \ldots, m\}$, and for any subset A of it, let $\mathfrak{R}^A$ denote the cartesian product of $|A|$ copies of $\mathfrak{R}$, where $|A|$ stands for the cardinality of A. Then:

Definition 3.1. The r.v.'s $Y_1, \ldots, Y_m$ are said to be *negatively associated* (NA, for short), if for every nonempty proper subset A of $\{1, \ldots, m\}$, and for every $G : \Re^A \to \Re$ and $H : \Re^{A^c} \to \Re$, which are nondecreasing in each coordinate, the remaining $m-1$ kept fixed and such that $\mathcal{E}G^2(Y_i, i \in A) < \infty$ and $\mathcal{E}H^2(Y_i, i \in A^c) < \infty$, it holds:

$$Cov\,(G\,(Y_i, i \in A)\,, H\,(Y_i, i \in A^c)) \le 0.$$

Remark 3.1. Coordinatewise nondecreasingness of G implies, of course, that, if: $x_{i_1} < x'_{i_1}, \dots, x_{i_k} < x'_{i_k}$ and $x_{j_1}, \dots, x_{j_\ell} (k + \ell = m)$ remain fixed, then $G(x_{i_1}, \dots, x_{i_k}, x_{j_1}, \dots, x_{k_\ell}) \le G(x'_{i_1}, \dots, x'_{i_k}, x_{j_1}, \dots, x_{k_\ell})$. In particular, G is nondecreasing along the main diagonal.

For NA r.v.'s, inequality (2.12) holds; that is:

Proposition 3.1. *If the r.v.'s $Y_1, \dots, Y_m$ are NA, then:*

$$\mathcal{E}\prod_{i=1}^{m} \exp(\pm cY_i) \le \prod_{i=1}^{m} \mathcal{E}\exp(\pm cY_i)\,, \quad c > 0, \tag{3.1}$$

and therefore inequalities (2.1), (2.2) and (2.6), (2.7) hold true.

Proof. That inequality (3.1) holds true with the positive sign is a consequence of property P_6 in Joag-Dev and Proschan (1983), due to the fact that $\exp(cY_i)$, $i = 1, \dots, m$ are NA. However, if Y_i, $i = 1, \dots, m$ are NA, then so are the r.v.'s $-Y_i$, $i = 1, \dots, m$. This is so because, if $G(\cdot)$ and $H(\cdot)$ are nondecreasing (in the sense of Definition 3.1), then so are the functions $-G(-\cdot)$ and $-H(-\cdot)$, and $Cov(-G, -H) = Cov(G, H)$. It follows that the r.v.'s $-cY_i$, $i = 1, \dots, m$ and $\exp(-cY_i)$, $i = 1, \dots, m$ are also NA, and therefore property P_6 applies again and yields the desired result. ∎

If negative association is replaced by positive association (or just association as originally termed by Esary et al (1967)), then the Hoeffding inequality need not be true, in general. This is illustrated by means of a counterexample discussed below (communicated to me by Hong Zhou), after the definition of positive association is given, and an auxiliary result is obtained.

Definition 3.2. *The r.v.'s $Y_1, \dots, Y_m$ are said to be* **positively associated** *(PA, for short), if for every nonempty proper subsets A and B of $\{1, \dots, m\}$, and for every $G : \Re^A \to \Re$ and $H : \Re^B \to \Re$, which are nondecreasing in each coordinate, the remaining $m-1$ kept fixed and such that $\mathcal{E}G^2(Y_i, i \in A) < \infty$ and $\mathcal{E}H^2(Y_i, i \in B) < \infty$, it holds:*

$$Cov\,(G(Y_i, i \in A),\, H(Y_i, i \in B)) \ge 0. \tag{3.2}$$

Infinitely many r.v.'s are said to be PA, if any finite subset is a set of PA r.v.'s.

By property P_3 in Esary et al (1967), the set consisting of a single r.v. is associated. This property generalizes as follows. The essence of this result was also communicated to me by Hong Zhou.

Proposition 3.2. *For any r.v. X and any n, the set consisting of n r.v.'s, all identical to X, $X, \dots, X$ is a set of PA r.v.'s.*

Proof. For any G and H as in Definition 3.2, $m = n$ and $Y_1 = \dots = Y_n = X$, inequality (3.2) must be established. Set $g(X) = G(X, \dots, X)$, $h(X) = H(X, \dots, X)$. By Lemma 3 in Lehmann (1966) (whose proof is attributed to Hoeffding) and $\boldsymbol{X} = (X, \dots, X)$:
(3.3)

$$Cov(G(\boldsymbol{X}), H(\boldsymbol{X})) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{P[G(\boldsymbol{X}) > u, H(\boldsymbol{X}) > v] - P[G(\boldsymbol{X}) > u] P[H(\boldsymbol{X}) > v]\} dudv.$$

However, with $(\Omega, \mathcal{A}, P)$ being the underlying probability space:

$$\begin{aligned}[G(\boldsymbol{X}) > u] &= \{\omega \in \Omega; G(\boldsymbol{X}(\omega)) > u\} = \{\omega \in \Omega; g(X(\omega)) > u\} \\ &= \{\omega \in \Omega; X(\omega) > \inf\{y \in \Re;\ g(y) > u\}\},\end{aligned}$$

and likewise:

$$[H(\boldsymbol{X}) > v] = \{\omega \in \Omega; X(\omega) > \inf\{y \in \Re;\ h(y) > v\}\}.$$

Therefore:

$$\begin{aligned}[G(\boldsymbol{X}) > u, H(\boldsymbol{X}) > v] = &\{\omega \in \Omega; X(\omega) > \inf\{y \in \Re;\ g(y) > u\}\} \\ &\cap \{\omega \in \Omega; X(\omega) > \inf\{y \in \Re;\ h(y) > v\}\},\end{aligned}$$

and hence:

$$[G(\boldsymbol{X}) > u, H(\boldsymbol{X}) > v] = \{\omega \in \Omega;\ X(\omega) > \inf\{y \in \Re;\ g(y) > u\}\},$$

if

$$\inf\{y \in \Re;\ g(y) > u\} > \inf\{y \in \Re;\ h(y) > v\} = [G(\boldsymbol{X}) > u]$$

and:

$$H[G(\boldsymbol{X}) > u, H(\boldsymbol{X}) > v] = \{\omega \in \Omega; X(\omega) > \inf\{y \in \Re;\ h(y) > v\}\},$$

if

$$\inf\{y \in \Re;\ g(y) > u\} > \inf\{y \in \Re;\ h(y) > v\} = [H(\boldsymbol{X}) > v].$$

Then, for all u and v in $\Re$, the integrand in (3.3) is equal either to $P[G(\boldsymbol{X}) > u] \cdot P[H(\boldsymbol{X}) \le v]$ or to $P[H(\boldsymbol{X}) > v] P[G(\boldsymbol{X}) \le u]$, and therefore $Cov(G(\boldsymbol{X}), H(\boldsymbol{X})) \ge 0$, as was to be seen. ∎

Counterexample. (for which the inequalities do not hold). Let X be a r.v. such that: $|X| \le C$, $\mathcal{E}X = 0$ and $\mathcal{E}X^2 = 1$ (for example, $P(X = -1) = P(X = 1) = 1/2$), and consider the Hoeffding inequality (for example, in the form (2.14)). Inequality (2.14) becomes here: $P(|X| \geq t) \leq 2\exp\left(-\frac{nt^2}{2C^2}\right)$. Letting $n \to \infty$, we get $P(|X| \geq t) = 0$, which contradicts the assumption $\mathcal{E}X^2 = 1$. Next, consider the Bennett inequality in the form (2.13), for example, which presently becomes: $P(|X| \geq t) \leq 2\exp\left(-\frac{3nt^2}{6+2Ct}\right)$. The right-hand side, however, tends to 0, as $n \to \infty$, and this leads to a contradiction as before.

4. Mixing Random Variables. The concept of mixing encompasses a large class of stochastic processes and provides an intuitive way of expressing dependence among the r.v.'s involved, which, however, grows weaker as blocks of r.v.'s keep increasing their distancing. This dependence can be formulated in various ways, and is expressed by means of mixing coefficients. Presently, we restrict ourselves to the so-called α-mixing (or strong mixing, introduced by Rosenblatt (1956)), which seems to be the most popular mode of mixing.

Definition 4.1. *Let X_n, $n = 1, 2, \ldots$ be defined on the probability space $(\Omega, \mathcal{A}, P)$, and for $1 \leq i < j \leq \infty$, let $\mathcal{F}_i^j$ be the σ-field induced by the r.v.'s X_n, $n = i, i+1, \ldots, j$. The sequence $\{X_n\}$, $n \geq 1$, is said to be α-mixing (or strong mixing) with mixing coefficients $\alpha(n)$, if:*

$$\sup\left\{|P(A \cap B) - P(A)P(B)|\,;\, A \in \mathcal{F}_1^k,\ B \in \mathcal{F}_{k+n}^\infty,\ k \geq 1\right\} \leq \alpha(n) \downarrow 0;$$

the sup over k may be omitted when the underlying sequence is (strictly) stationary.

Strong mixing sequences of r.v.'s span a wide range. Most Markov processes commonly used are α-mixing, and so is (under suitable conditions) the general linear model employed in time series analysis. For a brief review of various modes of mixing, their relationships and basic results, the interested reader is referred to Roussas and Ioannides (1987). For linear time series models, the references Withers (1981), Pham and Tran (1985), and Athreya and Pantula (1986) are of special interest. The probabilistic literature on mixing is very extensive. However, what is of special interest, from a statistical viewpoint, are exponential probability inequalities. Presently, we focus to such an inequality, whose derivation is given in considerable detail. This is done in Theorem 4.1 below. A version of it may also be found in Roussas and Ioannides (1988).

The proof of the main result hinges on the following two lemmas.

Lemma 4.1. *Let $Y_1, Y_2, \ldots$ be r.v.'s centered at their expectations, bounded by M, say, and forming an α-mixing (but not necessarily stationary) sequence with mixing coefficients $\alpha(n)$ such that $\sum_{n=1}^\infty \alpha(n) = \alpha^* < \infty$. Then:*

$$\mathcal{E}\left(\sum_{i=1}^n Y_i\right)^2 \leq (1 + 8\alpha^*)M^2 n. \qquad (4.1)$$

Proof. The fact that the Y_n's are bounded by M, that Y_i is $\mathcal{F}_1^i$-measurable and Y_j is $\mathcal{F}_j^\infty$-measurable imply that:

$$|Cov(Y_i, Y_j)| \leq 4M^2\alpha(j-i)$$

(see, for example, Theorem 7.1 in Roussas and Ioannides (1987)). Therefore:

$$\mathcal{E}\left(\sum_{i=1}^{n} Y_i\right)^2 = \sum_{i=1}^{n}\mathcal{E}Y_i^2 + 2\sum_{1\leq i<j\leq n}\mathcal{E}(Y_iY_j) \leq nM^2 + 8M^2\sum_{1\leq i<j\leq n}\alpha(j-i).$$

Since $\sum_{1\leq i<j\leq n}\alpha(j-i) = \sum_{j=1}^{n-1}(n-j)\alpha(j) \leq n\sum_{i=1}^{n}\alpha(j) \leq n\alpha^*$, the result follows. ∎

Lemma 4.2. *Let $Y_1, Y_2, \ldots$ be as in Lemma 4.1, and let ξ and η be r.v.'s such that ξ is $\mathcal{F}_1^k$-measurable, η is $\mathcal{F}_{k+n}^\infty$-measurable, $|\eta| \leq M_0$ and $\mathcal{E}|\xi|^p < \infty$ for some $p > 1$. Then:*

$$|\mathcal{E}(\xi\eta) - (\mathcal{E}\xi)(\mathcal{E}\eta)| \leq 6M_0\alpha^{1/q}(n)\|\xi\|_p, \tag{4.2}$$

where $\|\xi\|_p = \mathcal{E}^{1/p}|\xi|^p$ and $\frac{1}{q} = 1 - \frac{1}{p}$.

Proof. See Lemma 2.1 in Davydov (1968). ∎

For the formulation of the theorem in this section, some additional notation is needed. To this effect, let $\nu = \nu(n)$ be a sequence of positive integers tending to ∞ along with n, and let $\mu = \mu(n)$ be defined by $\mu = \left[\frac{n}{2\nu}\right]$, where $[x]$ denotes the integral part of x. Thus, μ is the largest positive integer for which $2\mu\nu \leq n$ and $n/2\mu\nu$ tends to 1 as $n \to \infty$.

Theorem 4.1. *Let $X_1, X_2, \ldots$ be a sequence of r.v.'s centered at their expectations, bounded by M, say, and forming an α-mixing (but not necessarily stationary) sequence with mixing coefficients $\alpha(n)$ such that $\sum_{n=1}^{\infty}\alpha(n) = \alpha^* < \infty$. Set $\bar{S}_n = n^{-1}\sum_{i=1}^{n} Y_i$, and let μ and ν be as above. Then for all $n \geq 1$ (all $\mu \geq 2$ and all ν):*

$$P\left(\left|\bar{S}_n\right| \geq \varepsilon_n\right) \leq K_1\left\{1 + 6e^{1/2}\left[\alpha(\nu)\right]^{1/\mu}\right\}^{\mu}\exp\left(-K_2 n\varepsilon_n^2\right), \tag{4.3}$$

where $0 < \varepsilon_n \leq K_3/\nu$. Here K_1 is a sufficiently large constant (≥ 6), $K_2 = 1/4eM^2(1+8\alpha^*)$, $K_3 = M(1+8\alpha^*)/2$, and $\alpha^* = \sum_{n=1}^{\infty}\alpha(n)$.

Remark 4.1. If $\limsup\left\{1 + 6e^{1/2}\left[\alpha(\nu)\right]^{1/\mu}\right\}^{\mu} < \infty$, inequality (4.2) is a Hoeffding-type inequality. It also provides rates of convergence of $\bar{S}_n$ for a suitable choice of ε_n, subject to the side condition $\varepsilon_n \leq C_3/\nu$.

Proof of Theorem 4.1. For $i = 1, \ldots, \mu$, set: $U_i = Y_{2(i-1)\nu+1} + \cdots + Y_{(2i-1)\nu}$, $V_i = Y_{(2i-1)\nu+1} + \cdots + Y_{2i\nu}$, and $W_\mu = Y_{2\nu\mu+1} + \cdots + Y_n$. Also, set: $U_\mu^* = U_1 + \cdots + U_\mu$, $V_\mu^* = V_1 + \cdots + V_\mu$. Then, for $i = 1, \ldots, \mu$:

$$|U_i| \leq \nu M,\ |V_i| \leq \nu M,\ |W_\mu| \leq \nu M, \text{ and } S_n = U_\mu^* + V_\mu^* + W_\mu.$$

From (4.2), we get:

$$\mathcal{E}(\xi\eta) \leq (\mathcal{E}\xi)(\mathcal{E}\eta) + 6M_0\alpha^{1/q}(n)\|\xi\|_p. \tag{4.4}$$

Set $\bar{U}_\mu^* = U_\mu^*/n$ and let $\lambda > 0$. Then:

$$\mathcal{E}e^{\lambda\bar{U}_\mu^*} = \mathcal{E}\left(e^{\frac{\lambda}{n}U_{\mu-1}^*} \cdot e^{\frac{\lambda}{n}U_\mu}\right) \text{ with } \left|\frac{\lambda}{n}U_\mu\right| \leq \frac{\lambda\nu M}{n}, \text{ so that } e^{\frac{\lambda}{n}U_\mu} \leq e^{\lambda\nu M/n}.$$

Set $\xi = e^{\frac{\lambda}{n}U_{\mu-1}^*}$ and $\eta = e^{\frac{\lambda}{n}U_\mu}$. Then ξ is $\mathcal{F}_1^{(2\mu-3)\nu}$-measurable, η is $\mathcal{F}_{2(\mu-1)\nu+1}^\infty$-measurable, and these two σ-fields are separated by $\nu + 1$ r.v.'s. Then (4.4) yields:

(4.5)

$$\begin{aligned}
\mathcal{E}(e^{\lambda U_{\mu-1}^*/n}e^{\lambda U_\mu/n}) &\leq \mathcal{E}e^{\lambda U_{\mu-1}^*/n} \cdot \mathcal{E}e^{\lambda U_\mu/n} + 6e^{\lambda\nu M/n}\alpha^{1/q}(\nu)\mathcal{E}^{1/p}e^{\lambda pU_{\mu-1}^*/n} \\
&\leq \mathcal{E}^{1/p}e^{\lambda pU_{\mu-1}^*/n} \cdot \mathcal{E}e^{\lambda U_\mu/n} + 6e^{\lambda\nu M/n}\alpha^{1/q}(\nu)\mathcal{E}^{1/p}e^{\lambda pU_{\mu-1}^*/n} \\
&= \mathcal{E}^{1/p}e^{\lambda pU_{\mu-1}^*/n}\left[\mathcal{E}e^{\lambda U_\mu/n} + 6e^{\lambda\nu M/n}\alpha^{1/q}(\nu)\right].
\end{aligned}$$

Now apply the inequality $e^t \leq 1 + t + t^2$ $\left(|t| \leq \frac{1}{2}\right)$ for $t = \frac{\lambda U_\mu}{n}$ to obtain: $e^{\lambda U_\mu/n} \leq 1 + (\lambda U_\mu/n) + (\lambda U_\mu/n)^2$, with $|\lambda U_\mu/n| \leq 1/2$ which is implied by $\lambda \leq n/(2\nu M)$. Since $\mathcal{E}U_\mu = 0$, we have: $\mathcal{E}e^{\lambda U_\mu/n} \leq 1 + \mathcal{E}\left(\frac{\lambda U_\mu}{n}\right)^2$, $\lambda \leq \frac{n}{2\nu M}$. Next, apply the inequality $1 + t \leq e^t$ $(t \geq 0)$ with $t = \mathcal{E}\left(\lambda U_\mu/n\right)^2$ to obtain:

$$1 + \mathcal{E}\left(\frac{\lambda U_\mu}{n}\right)^2 \leq e^{\mathcal{E}(\lambda U_\mu/n)^2}, \quad \lambda \leq n/2\nu M,$$

so that

$$\mathcal{E}e^{\lambda U_\mu/n} \leq e^{\mathcal{E}(\lambda U_\mu/n)^2},\ \lambda \leq n/2\nu M. \tag{4.6}$$

By Lemma 4.1:

$$\mathcal{E}\left(\frac{\lambda U_\mu}{n}\right)^2 = \frac{\lambda^2}{n^2}\mathcal{E}U_\mu^2 \leq \frac{C\lambda^2M^2\nu}{n^2}, \text{ with } C = 1 + 8\alpha^*.$$

Then (4.6) gives $\mathcal{E}e^{\lambda U_\mu/n} \le e^{C\lambda^2M^2\nu/n^2}$, $\lambda \le n/2\nu M$, and therefore inequality (4.5) becomes, for $\lambda \le n/2\nu M$:

$$\begin{aligned}\mathcal{E}e^{\lambda\bar{U}^*_\mu} &\le \mathcal{E}^{1/p}e^{\lambda p\bar{U}^*_{\mu-1}}\left[e^{C\lambda^2M^2\nu/n^2} + 6\alpha^{1/q}(\nu)e^{\lambda\nu M/n}\right]\\ &= \mathcal{E}^{1/p}e^{\lambda p\bar{U}^*_{\mu-1}} \cdot e^{C\lambda^2M^2\nu/n^2}\left[1 + 6\alpha^{1/q}(\nu)e^{\frac{\lambda\nu M}{n} - \frac{C\lambda^2M^2\nu}{n^2}}\right]\\ &\le \mathcal{E}^{1/p}e^{\lambda p\bar{U}^*_{\mu-1}} \cdot e^{C\lambda^2M^2\nu/n^2}\left[1 + 6\alpha^{1/q}(\nu)e^{\lambda\nu M/n}\right]\\ &\le \mathcal{E}^{1/p}e^{\lambda p\bar{U}^*_{\mu-1}} \cdot e^{C\lambda^2M^2\nu/n^2}\left[1 + 6\alpha^{1/q}(\nu)e^{1/2}\right],\end{aligned}$$

since $\lambda \le n/2\nu M$ is equivalent to $\frac{\lambda\nu M}{n} \le \frac{1}{2}$. So:

$$\text{(4.7)}\quad \mathcal{E}e^{\lambda\bar{U}^*_\mu} \le \left[1 + 6e^{1/2}\alpha^{1/q}(\nu)\right]e^{C\lambda^2M^2\nu/n^2} \cdot \mathcal{E}^{1/p}e^{\lambda p\bar{U}^*_{\mu-1}},\ \lambda \le n/2\nu M.$$

Set $\lambda p = \lambda_1$ and apply (4.7) to obtain, for $\lambda_1 \le n/2\nu M$:

$$\mathcal{E}e^{\lambda_1\bar{U}^*_{\mu-1}} \le \left[1 + 6e^{1/2}\alpha^{1/q}(\nu)\right]e^{C\lambda_1^2M^2\nu/n^2} \cdot \mathcal{E}^{1/p}e^{\lambda_1 p\bar{U}^*_{\mu-2}},$$

so that:

$$\mathcal{E}^{1/p}e^{\lambda_1 p\bar{U}^*_{\mu-1}} \le \left[1 + 6e^{1/2}\alpha^{1/q}(\nu)\right]^{1/p} \cdot e^{C\lambda_1^2M^2\nu/n^2p} \cdot \mathcal{E}^{1/p^2}e^{\lambda_1 p\bar{U}^*_{\mu-2}},$$

and hence:

$$\text{(4.8)}\quad \mathcal{E}e^{\lambda\bar{U}^*_\mu} \le \left[1 + 6e^{1/2}\alpha^{1/q}(\nu)\right]^{1+\frac{1}{p}}e^{\frac{C\lambda^2M^2\nu}{n^2} + \frac{C\lambda_1^2M^2\nu}{n^2p}} \cdot \mathcal{E}^{1/p^2}e^{\lambda_1 p\bar{U}^*_{\mu-2}},$$

$$\lambda_1 \le n/2\nu M.$$

From $\lambda_1 = \lambda_p$, we have $\lambda_1^2 = \lambda^2p^2$ and $\frac{C\lambda^2M^2\nu}{n^2} + \frac{C\lambda_1^2M^2\nu}{n^2p} = \frac{C\lambda^2M^2\nu}{n^2}(1+p)$, and $\lambda_1 \le n/2\nu M$ is equivalent to $\lambda \le n/2\nu pM$. Therefore (4.8) becomes:

$$\text{(4.9)}\qquad \mathcal{E}e^{\lambda\bar{U}^*_\mu} \le \left[1 + 6e^{1/2}\alpha^{1/q}(\nu)\right]^{1+\frac{1}{p}}e^{\frac{C\lambda^2M^2\nu}{n^2}(1+p)} \cdot \mathcal{E}^{1/p^2}e^{\lambda_1 p\bar{U}^*_{\mu-2}},$$

$$\lambda \le n/2\nu pM.$$

Next, set $\lambda_2 = \lambda_1 p$ and work as above in order to obtain:

$$\text{(4.10)}\ \mathcal{E}e^{\lambda\bar{U}^*_\mu} \le \left[1 + 6e^{1/2}\alpha^{1/q}(\nu)\right]^{1+\frac{1}{p}+\frac{1}{p^2}} \cdot e^{\frac{C\lambda^2M^2\nu}{n^2}(1+p+p^2)} \cdot \mathcal{E}^{1/p^3}e^{\lambda_2 p\bar{U}^*_{\mu-3}},$$

$$\lambda \le n/2\nu p^2M.$$

Continuing in this manner, we obtain after $\mu - 1$ iterations:
(4.11)

$$\mathcal{E}e^{\lambda \bar{U}_\mu^*} \leq \left[1 + 6e^{1/2}\alpha^{1/q}(\nu)\right]^{1+\frac{1}{p}+\frac{1}{p^2}+\cdots+\frac{1}{p^{\mu-2}}} \cdot e^{\frac{C\lambda^2 M^2 \nu}{n^2}(1+p+p^2+\cdots+p^{\mu-2})}$$

$$\cdot \mathcal{E}^{1/p^{\mu-1}} e^{\lambda p^{\mu-1}\bar{U}_1}, \quad \lambda \leq n/2\nu p^{\mu-2}M.$$

However,

$$\lambda p^{\mu-1}\bar{U}_1 = \frac{\lambda p^{\mu-1}}{n}U_1 \text{ and } \mathcal{E}e^{\lambda p^{\mu-1}\bar{U}_1} = \mathcal{E}e^{\frac{\lambda p^{\mu-1}}{n}U_1} \leq e^{\mathcal{E}(\lambda p^{\mu-1}U_1/n)^2}$$

(by (4.6)), provided $\lambda \leq n/2\nu p^{\mu-1}M$, and furthermore: $\mathcal{E}\left(\frac{\lambda p^{\mu-1}}{n}U_1\right)^2 \leq \frac{C\lambda^2 p^{2(\mu-1)}M^2\nu}{n^2}$. Then:

$$\mathcal{E}e^{\lambda p^{\mu-1}\bar{U}_1} \leq e^{C\lambda^2 p^{2(\mu-1)}M^2\nu/n^2}, \text{ so that } \mathcal{E}^{1/p^{\mu-1}} e^{\lambda p^{\mu-1}\bar{U}_1} \leq e^{C\lambda^2 p^{\mu-1}M^2\nu/n^2},$$

and therefore (4.11) becomes:
(4.12)

$$\mathcal{E}e^{\lambda \bar{U}_\mu^*} \leq \left[1 + 6e^{1/2}\alpha^{1/q}(\nu)\right]^{1+\frac{1}{p}+\frac{1}{p^2}+\cdots+\frac{1}{p^{\mu-2}}} \cdot e^{\frac{C\lambda^2 M^2 \nu}{n^2}(1+p+p^2+\cdots+p^{\mu-1})},$$

$$\lambda \leq n/2\nu p^{\mu-1}M.$$

Now

$$1 + \frac{1}{p} + \frac{1}{p^2} + \cdots + \frac{1}{p^{\mu-2}} < 1 + \frac{1}{p} + \frac{1}{p^2} + \cdots = \frac{p}{p-1} = q,$$

and

$$1 + p + p^2 + \cdots + p^{\mu-1} = \frac{1-p^\mu}{1-p} = \frac{p^\mu - 1}{p-1} < \frac{p^\mu}{p-1}.$$

Therefore (4.12) becomes:

$$\text{(4.13)} \qquad \mathcal{E}e^{\lambda \bar{U}_\mu^*} \leq \left[1 + 6e^{1/2}\alpha^{1/q}(\nu)\right]^q e^{\frac{C\lambda^2 M^2\nu}{n^2}\cdot\frac{p^\mu}{p-1}}, \quad \lambda \leq n/2\nu p^{\mu-1}M$$

At this point, take $p = 1 + \frac{1}{\mu-1} = \frac{\mu}{\mu-1}$, so that $q = \mu$. Also,

$$\frac{p^\mu}{p-1} = (\mu-1)\left(\frac{\mu}{\mu-1}\right)^\mu < \frac{\mu}{(1-\frac{1}{\mu})^\mu} < 2e\mu \quad \text{(for all } \mu \geq 2).$$

Furthermore:

$$\frac{n}{2\nu p^{\mu-1}M} = \frac{n}{2M\nu(1+\frac{1}{\mu-1})^{\mu-1}} > \frac{n}{2M\nu \cdot 2e} = \frac{n}{4Me\nu} \quad \text{(for all } \mu \geq 2).$$

Therefore the inequality $\lambda \leq \frac{n}{2\nu p^{\mu-1}M}$ is implied by $\lambda \leq \frac{n}{4Me\nu}$. Also, $2\nu\mu \leq n$ is equivalent to $\frac{\nu\mu}{n} \leq \frac{1}{2}$. On the basis of these observations, and for all $\mu \geq 2$, (4.13) becomes:

$$\mathcal{E}e^{\lambda\bar{U}_\mu^*} \leq \left[1 + 6e^{1/2}\alpha^{1/\mu}(\nu)\right]^\mu e^{(CeM^2)\lambda^2/n}, \quad \lambda \leq \frac{n}{4Me\nu}. \tag{4.14}$$

At this point, observe that (4.14) is, clearly, true if $\bar{U}_\mu^*$ is replaced by either $\bar{V}_\mu^* = V_\mu^*/n$ or by $\bar{W}_\mu = W_\mu/n$, and proceed to applying Markov inequality to obtain:

$$\begin{aligned} P\left(\bar{U}_\mu^* \geq \varepsilon_n\right) &= P\left(e^{\lambda\bar{U}_\mu^*} \geq e^{\lambda\varepsilon_n}\right) \leq e^{-\lambda\varepsilon_n}\left[1 + 6e^{1/2}\alpha^{1/\mu}(\nu)\right]^\mu e^{\rho\lambda^2/n} \\ &= \left[1 + 6e^{1/2}\alpha^{1/\mu}(\nu)\right]^\mu e^{\rho\frac{\lambda^2}{n} - \lambda\varepsilon_n}, \quad \text{where} \quad \rho = CeM^2. \end{aligned}$$

The function $g(\lambda) = \rho\frac{\lambda^2}{n} - \lambda\varepsilon_n$ is minimized for $\lambda_0 = \frac{n\varepsilon_n}{2\rho}$ and the minimum is $g(\lambda_0) = -\frac{n\varepsilon_n^2}{4\rho}$. Also, it must be checked that the side condition $\lambda \leq \frac{n}{4Me\nu}$ is satisfied by λ_0. This happens, if $\varepsilon_n < CM/2\nu$. Thus:

$$P\left(\bar{U}_\mu^* \geq \varepsilon_n\right) \leq \left[1 + 6e^{1/2}\alpha^{1/\mu}(\nu)\right]^\mu e^{-K_2 n\varepsilon_n^2}, \quad 0 < \varepsilon_n \leq K_3/\nu,$$

where $K_2 = 1/4eM^2(1 + 8\alpha^*)$ and $K_3 = M(1 + 8\alpha^*)/2$. Since the same inequality, clearly, holds when $\bar{U}_\mu^*$ is replaced by -$\bar{U}_\mu^*$, we have:

$$P\left(|\bar{U}_\mu^*| \geq \varepsilon_n\right) \leq 2\left[1 + 6e^{1/2}\alpha^{1/\mu}(\nu)\right]^\mu e^{-K_2 n\varepsilon_n^2}, \quad \varepsilon_n \leq K_3/\nu.$$

Applying this inequality to $|\bar{V}_\mu^*|$ and $|\bar{W}_\mu|$, and utilizing the expression $\bar{S}_n = \bar{U}_\mu^* + \bar{V}_\mu^* + \bar{W}_\mu$, we have:

$$P\left(|\bar{S}_n| \geq \varepsilon_n\right) \leq 6\left[1 + 2e^{1/2}\alpha^{1/\mu}(\nu)\right]^\mu e^{-K_2 n\varepsilon_n^2}, \quad 0 < \varepsilon_n \leq K_3/\nu.$$

This inequality is true for all $n \geq 1$, and all $\mu \geq 2$ and $\nu \geq 1$ for which $2\mu\nu \leq n$. Replacing the number 6 by a potentially larger constant K_1 to take care of the finitely many exceptional $n's$, $\mu's$ and $\nu's$, we have the inequality

$$P\left(|\bar{S}_n| \geq \varepsilon_n\right) \leq K_1\left[1 + 2e^{1/2}\alpha^{1/\mu}(\nu)\right]^\mu e^{-K_2 n\varepsilon_n^2}, \quad 0 < \varepsilon_n \leq K_3/\nu,$$

holding for all n (all $\mu \geq 2$ and all ν).

5. Some Applications. In the framework of nonparametric curve estimation under α-mixing, there is an abundance of results available. Some contributions by this author and his collaborators are results on the asymptotic normality of kernel estimates of a probability density function (p.d.f.) with applications to hazard rate (Roussas (1990a), Roussas and Tran(1992)); asymptotic normality of kernel regression estimates, both in the fixed design and stochastic design cases (Roussas (1990b), Roussas et al (1992)); and uniform strong estimation with rates for d.f.'s, p.d.f.'s and hazard rates (Cai and Roussas (1992)). Similar results have also been obtained for NA r.v.'s in a random field framework (Roussas (1993), (1995)), as well as asymptotic normality again for random fields (Roussas (1994)).

In this section, the Hoeffding inequality for NA r.v.'s is used in order to obtain minimum distance estimates in the fashion of Yatracos (1985). To this effect, let $\mathcal{P}$ be a family of probability measures on $(\Omega, \mathcal{A})$, and for each $P \in \mathcal{P}$, let $X_1, \ldots, X_n$ be N.A. r.v.'s. The objective is to construct a minimum distance estimate $\hat{P}_n$ of P on the basis of $X_1, \ldots, X_n$. In $\mathcal{P}$, consider the total variation distance d defined by:

$$(5.1) \quad d(P, Q) = \|P - Q\| = 2 \sup \{|P(A) - Q(A)|; A \in \mathcal{A}\}, \quad P, Q \in \mathcal{P},$$

and suppose that the space $(\mathcal{P}, d)$ is *totally bounded*; that is, for any $a > 0$, there exists a finite number of balls, $N(a)$ say, centered at some points in $\mathcal{P}$ and having radius a, whose union is $\mathcal{P}$. If $N(a)$ is the most economic number of balls as just described, then the function $\log_2 N(a)$ is called Kolmogorov's *entropy* of the space $(\mathcal{P}, d)$. At this point, it is assumed that $\mathcal{P}$ is dominated by a measure μ, and let $\mathcal{F}_{N(a)}$ be the collection of sets defined by:

$$\left\{\omega \in \Omega;\ \frac{dP_i}{d\mu}(\omega) > \frac{dP_j}{d\mu}(\omega), \quad 1 \leq i < j \leq N(a)\right\},$$

where P_i, $i = 1, \ldots, N(a)$ are the centers of the balls of radius a which cover $\mathcal{P}$. Then it can be shown (see Yatracos (1985)) that, for any P and Q in $\mathcal{P}$:

$$(5.2) \qquad \|P - Q\| \leq 4a + 2 \max \left\{|P(A) - Q(A)|;\ A \in \mathcal{F}_{N(a)}\right\}.$$

At this point, allow the radius a to depend on the number n of the r.v.'s available, and set N_n for $N(a_n)$. Then let μ_n be the empirical measure on $\mathcal{A}$ defined in terms of the r.v.'s $X_1, \ldots, X_n$; that is: $\mu_n(A) = n^{-1} \sum_{i=1}^{n} I_A(X_i)$, and estimate the unknown measure P governing the X_i*'s* by the *minimum distance* measure $\hat{P}_n$, defined to be that measure among the P_i, $i = 1, \ldots, N_n$, which minimizes the quantities:

$$\max \{|\mu_n(A) - P_i(A)|;\ A \in \mathcal{F}_{N_n}, \quad i = 1, \ldots, N_n\}.$$

More formally, $\hat{P}_n$ is defined by:

$$\max\left\{\left|\mu_n(A) - \hat{P}_n(A)\right|; A \in \mathcal{F}_n\right\}$$
$$= \min\left[\max\{|\mu_n(A) - P_i(A)|; A \in \mathcal{F}_{N_n}\}, \quad i = 1, \ldots, N_n\right].$$

Then the following result may be established.

Theorem 5.1. *In the notation introduced above, under the assumptions made, and under the additional condition that a_n is proportional to $(\log_2 N_n/n)^{1/2}$, it holds that, for every $\varepsilon > 0$, there exists $b(\varepsilon) > 0$ such that:*

$$\sup\left\{P\left[\|\hat{P}_n - P\| \geq b(\varepsilon)a_n\right]; P \in \mathcal{P}\right\} < \varepsilon \text{ for all } n.$$

Proof. By setting $Y_i = I_A(X_i)$, the r.v.'s $Y_i - \mathcal{E}Y_i = Y_i - P(A)$, $i = 1, \ldots, n$ are NA, and $\bar{S}_n(A) = n^{-1}\sum_{i=1}^{n}(Y_i - \mathcal{E}Y_i) = \mu_n(A) - P(A)$. From (5.1) and (5.2), it follows by means of the triangular inequality:

$$\|\hat{P}_n - P\| \leq 5a_n + 4\max\{|\mu_n(A) - P(A)|; A \in \mathcal{F}_{N_n}\}. \tag{5.3}$$

Apply inequality (2.14) (with $C = 1$) to $\bar{S}_n(A)$, take into consideration inequality (5.3), and the fact that the cardinality of $\mathcal{F}_{N_n}$ is bounded by N_n^2 in order to obtain, for $\varepsilon_n > 0$ and all n:

$$P\left(\|\hat{P}_n - P\| \geq \varepsilon_n\right) \leq 2N_n^2 \exp\left[-n(\varepsilon_n - 5a_n)^2/32\right], \tag{5.4}$$

and the right-hand side in (5.4) is independent of $P \in \mathcal{P}$. Be selecting ε_n proportional to $(\log_2 N_n/n)^{1/2}$, it is seen that the right-hand side of (5.4) is $< \varepsilon$ for all n. ∎

The theorem just proved can be established when total boundedness of $(\mathcal{P}, d)$ is replaced by the assumption that $\mathcal{P}$ is the countable union of such spaces. Also, one may discuss a regression-type estimation problem. These matters, however, will not be pursued here. In closing, it should be mentioned that the concept of minimum distance method was introduced by Wolfowitz (1957).

Acknowledgments. This work was supported in part by a research grant from the University of California, Davis.

Thanks are due to an anonymous referee for constructive comments.

REFERENCES

ATHREA, K.B. AND PANTULA, S.G. (1986). A note on strong mixing of ARMA processes. *Statist. Probab. Lett.* **4** 187-190.

BENNETT, GEORGE (1962). Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.* **57** 33-45.

BLACKWELL, DAVID (1962). Discrete dynamic programming. *Ann. Math. Statist.* **33** 719-726.

BLACKWELL, DAVID (1965). Discounted dynamic programming. *Ann. Math. Statist.* **36** 226-235.

BLACKWELL, DAVID AND FREEDMAN, DAVID (1973). On the amount of variance needed to escape from a strip. *Ann. Probab.* **1** 772-787.

CAI, Z. AND ROUSSAS, G. G. (1992). Uniform strong estimation under α-mixing, with rates. *Statist. Probab. Lett.* **15** 47-55.

DAVYDOV, YU. A. (1968). Convergence of distributions generated by stochastic processes. *Theory Probab. Appl.* **13** 730-737.

ESARY, J. D., PROSCHAN, F. AND WALKUP, D. W. (1967). Association of random variables, with applications. *Ann. Math. Statist.* **38** 1466-1474.

HOEFFDING, WASSILY (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13-30.

JOAG-DEV, K. AND PROSCHAN, F. (1983). Negative association of random variables, with applications. *Ann. Statist.* **11** 286-295.

LEHMANN, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.* **37** 1137-1153.

PHAM, D. T. AND TRAN, L. T. (1985). Some mixing properties of time series models. *Stochastic Process. Appl.* **19** 297-303.

ROSENBLATT, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U.S.A.* **42** 43-47.

ROUSSAS, G. G. (1990a). Asymptotic normality of the kernel estimate under dependence conditions: Application to hazard rate. *J. Statist. Plann. Inference* **25** 81-104.

ROUSSAS, G. G. (1990b). Nonparametric regression estimation under mixing conditions. *Stochastic Process. Appl.* **36** 107-116.

ROUSSAS, G. G. (1993). Curve estimation in random fields of associated processes. *Nonparametric Statist.* **2** 215-224.

ROUSSAS, G. G. (1994). Asymptotic normality of random fields of positively or negatively associated processes. *J. Multivariate Anal.* **50** 152-173.

ROUSSAS, G. G. (1995). Asymptotic normality of a smooth estimate of a random field distribution function under association. *Statist. Probab. Lett.* **24** 77-90.

ROUSSAS, G. G. AND IOANNIDES, D. (1987). Moment inequalities for mixing sequences of random variables. *Stochastic Anal. Appl.* **5(1)** 61-120.

ROUSSAS, G. G. AND IOANNIDES, D. (1988). Probability bounds for sums in triangular arrays of random variables under mixing conditions. In: K. Matusita, ed., *Statistical Theory and Data Analysis II*, pages 293-308. North-Holland, Amsterdam.

ROUSSAS, G. G., AND TRAN, L. T. (1992). Asymptotic normality of the recursive kernel regression estimate under dependence conditions. *Ann. Statist.* **20** 98-120.

ROUSSAS, G. G., TRAN, L. T. AND IOANNIDES, D. A. (1992). Fixed design regression for time series: Asymptotic normality. *J. Multivariate Anal.* **40** 262-291.

WITHERS, C. S. (1981). Conditions for linear processes to be strong mixing. *Z. Wahrsch. Verw. Gebiete* **57** 479-480.

WOLFOWITZ, J. (1957). The minimum distance method. *Ann. Math. Statist.* **28** 75-88.

YATRACOS, Y. G. (1985). Rates of convergence of minimum distance estimators and Kolmogorov's entropy. *Ann. Statist.* **13** 768-774.

ZHOU, HONG (1995). Personal communication.

DIVISION OF STATISTICS
UNIVERSITY OF CALIFORNIA
DAVIS, CA 95616
ggroussas@ucdavis.edu