The Annals of Mathematical Statistics
1969, Vol. 40, No. 4, 1353–1357

CONSISTENT ESTIMATION OF A LOCATION PARAMETER IN THE PRESENCE OF AN INCIDENTAL SCALE PARAMETER

BY J. PFANZAGL

University of Cologne

Let $P/\mathcal{B}$ be a probability-measure on the Borel algebra of the real line R with median zero. For $\theta \varepsilon R$ and $\sigma \varepsilon (0, \infty)$, let $P_{\theta,\sigma}/\mathcal{B}$ be the probability measure defined by $P_{\theta,\sigma}(B) = P\{(x - \theta)/\sigma : x \varepsilon B\}$, $B \varepsilon \mathcal{B}$. The parameters θ and σ will be called "location parameter" and "scale parameter," respectively. Our problem is that of consistently estimating θ from a realization governed by $\times_{\nu=1}^{\infty} P_{\theta,\sigma^\nu}$, where $(\sigma^\nu)_{\nu \varepsilon N}$ is unknown.

Intuitively speaking, the problem is to estimate an unknown (structural) location parameter θ which is constant for the whole sample sequence $(x_\nu)_{\nu\varepsilon N}$ in the presence of an unknown (incidental) scale parameter σ whose value changes from one sample x_ν to the next.

In statistical theory this problem deserves at least the same interest as the central limit problem. The interest in the asymptotic behavior of $\bar{x}_n = n^{-1}\sum_{\nu=1}^{n} x_\nu$ is rooted in the role which $\bar{x}_n$ plays as an estimate for θ. The generalization which was forced upon this problem in probability theory, namely to study the asymptotic behavior of $b_n^{-1}(\sum_{\nu=1}^{n} x_\nu - a_n)$, was a mistake from the statistical point of view. The relevant generalization is to consider estimates of θ other than $\bar{x}_n$.

Despite its importance, this problem is rarely treated in the literature. The only pertinent result known to the author is due to Wolfowitz (1953, p. 17). It assures the existence of a strongly consistent sequence of estimates for the particular case of P being the normal distribution $N(0, 1)$ under the condition

$$\text{(C)} \qquad \sup_{n\varepsilon N} n^{-1}\sum_{\nu=1}^{n} (\sigma^\nu)^2 < \infty.$$

This result was obtained by the minimum distance method. It does, however, not reveal the full power of this method. It is easy to see that under condition (C), the sequence of means $(\bar{x}_n)_{n\varepsilon N}$ is consistent for θ, and the sequence of estimates defined by $\theta_n(\mathbf{x}) = \bar{x}_{2^{[\log_2 n]}}$, $n \varepsilon N$, is strongly consistent ($[a]$ denotes the largest integer less than or equal to a).

Chebychev's inequality implies

$$\times_{\nu=1}^{\infty} P_{\theta,\sigma^\nu}\{\mathbf{x} \varepsilon R^N : |\bar{x}_{2^k} - \theta| \geqq \epsilon\} \leqq \epsilon^{-2} \cdot 2^{-2k}\sum_{\nu=1}^{2^k} (\sigma^\nu)^2$$
$$\leqq \epsilon^{-2} \cdot 2^{-k} \sup_{n\varepsilon N} n^{-1} \sum_{\nu=1}^{n} (\sigma^\nu)^2.$$

The Lemma of Borel-Cantelli implies that for $\times_{\nu=1}^{\infty} P_{\theta,\sigma^\nu}$ — a.a. $\mathbf{x} \varepsilon R^N$ the relation $|\bar{x}_{2^k} - \theta| \geqq \epsilon$ holds for a finite number of k's only. Hence

$$\times_{\nu=1}^{\infty} P_{\theta,\sigma^\nu} \liminf_{n\varepsilon N} \{\mathbf{x} \varepsilon R^N : |\theta_n(\mathbf{x}) - \theta| < \epsilon\}$$
$$= \times_{\nu=1}^{\infty} P_{\theta,\sigma^\nu} \liminf_{k\varepsilon N} \{\mathbf{x} \varepsilon R^N : |\bar{x}_{2^k} - \theta| < \epsilon\} = 1.$$

Received 5 November 1968.

Though Theorem 5.1 of Wolfowitz (1953, p. 16) would enable us to derive a somewhat stronger result, we shall abstain from doing so because a slight modification of the minimum distance method yields the following much stronger result.

THEOREM. *Assume that P admits a unimodal density with respect to Lebesgue measure. Then there exists a sequence of estimates which is consistent for θ if*

$$\text{(A)} \qquad \lim_{n \varepsilon N} n^{-\frac{1}{2}} \sum_{\nu=1}^{n} (1 + \sigma^{\nu})^{-1} = \infty.$$

If

$$\text{(B)} \qquad \lim_{n \varepsilon N} (n \log \log n)^{-\frac{1}{2}} \sum_{\nu=1}^{n} (1 + \sigma^{\nu})^{-1} = \infty,$$

the sequence is even strongly consistent.

It seems to be at least of theoretical interest that condition (A) [resp. (B)] guarantees the existence of [strongly] consistent estimates even in cases where the variance of $\bar{x}_n$ tends to infinity. Take, e.g., $\sigma^{2k} = k$, $\sigma^{2k-1} = 1$, $k \varepsilon N$

That condition (C) is much stronger than condition (B) can be seen from the following

PROPOSITION 1. *Condition* (C) *implies* $\liminf_{n\varepsilon N} n^{-1} \sum_{\nu=1}^{n}(1 + \sigma^{\nu})^{-1} > 0$.

PROOF. Repeated applications of Schwarz's inequality yield

$$\begin{aligned} 1 &\leqq (n^{-1} \sum_{\nu=1}^{n} (1 + \sigma^{\nu})^{-1}) \cdot (n^{-1} \sum_{\nu=1}^{n} (1 + \sigma^{\nu})) \\ &\leqq (n^{-1} \sum_{\nu=1}^{n} (1 + \sigma^{\nu})^{-1}) \cdot (n^{-1} \sum_{\nu=1}^{n} (1 + \sigma^{\nu})^{2})^{-\frac{1}{2}}. \end{aligned}$$

The assertion now follows immediately from the fact that, because of $(1 + \sigma^{\nu})^2 < 3 + 3(\sigma^{\nu})^2$, condition (C) implies $\sup_{n\varepsilon N} n^{-1} \sum_{\nu=1}^{n} (1 + \sigma^{\nu})^2 < \infty$.

The nature of the conditions (A) and (B) is further explicated by the following Propositions 2 and 3. Roughly speaking, these conditions are only "slightly stronger" than the corresponding conditions with $1 + \sigma^{\nu}$ substituted by σ^{ν}:

PROPOSITION 2. *For arbitrary positive constants* c_0, c_1,

$$\lim_{n\varepsilon N} a_n \sum_{\nu=1}^{n} (c_0 + \sigma^{\nu})^{-1} = \infty$$

if and only if

$$\lim_{n\varepsilon N} a_n \sum_{\nu=1}^{n} (c_1 + \sigma^{\nu})^{-1} = \infty.$$

PROOF. Let $c_0 < c_1$. Then the assertion follows immediately from the relation

$$(c_1 + \sigma^{\nu})^{-1} < (c_0 + \sigma^{\nu})^{-1} < (c_1 + \sigma^{\nu})^{-1} c_1/c_0.$$

PROPOSITION 3. *For arbitrary positive constants* c_0, c_1,

$$\lim_{n\varepsilon N} a_n \sum_{\nu=1}^{n} (c_0 + \sigma^{\nu})^{-1} = \infty$$

if and only if

$$\lim_{n\varepsilon N} a_n \sum_{\nu=1}^{n} (\max (c_1, \sigma^{\nu}))^{-1} = \infty.$$

PROOF. This follows immediately from Proposition 2 and the relation

$$\max(c_1, \sigma^\nu) \leqq c_1 + \sigma^\nu \leqq 2\max(c_1, \sigma^\nu).$$

As the consistency of the sequence of estimates depends on the asymptotic behavior of $(n^{-\frac{1}{2}}\sum_{\nu=1}^n (1+\sigma^\nu)^{-1})_{n\varepsilon N}$, the question naturally arises whether there are realistic conditions under which we know that $\lim_{n\varepsilon N} n^{-\frac{1}{2}}\sum_{\nu=1}^n (1+\sigma^\nu)^{-1} = \infty$ without knowing the sequence $(\sigma^\nu)_{\nu\varepsilon N}$. We shall give two examples.

(1) If there exists a sequence $(a_\nu)_{\nu\varepsilon N}$ such that $\sigma^\nu \leqq a_\nu$ for all $\nu \varepsilon N$ and $\lim_{n\varepsilon N} n^{-\frac{1}{2}}\sum_{\nu=1}^n (1+a_\nu)^{-1} = \infty$, then condition (A) is fulfilled.

(2) If $(\sigma^\nu)_{\nu\varepsilon N}$ is a sequence of independent realizations of (not necessarily identically distributed) random variables such that $\sigma^\nu \leqq c$ with positive probability, say p, then irrespectively of the values of p and c, condition (A) is fulfilled with probability 1. We have $(1+\sigma^\nu)^{-1} \geqq (1+c)^{-1}\chi_{(0,c)}(\sigma^\nu)$. By assumption, $\chi_{(0,c)}(\sigma^\nu)$ is the realization of a random variable assuming the values 1 and 0 with probabilities p and $1-p$, respectively. As

$$\lim_{n\varepsilon N} n^{-\frac{1}{2}}\sum_{\nu=1}^n \chi_{(0,c)}(\sigma^\nu) = \infty$$

with probability 1, this implies the assertion.

PROOF OF THE THEOREM. (i) For notational convenience let $F(t) = P(-\infty, t)$. Let $\{t_1, t_2, \cdots\}$ be an enumeration of the rational numbers. We determine a $\mathcal{B}^n$-measurable map $\mathbf{x} \to (\theta_n(\mathbf{x}), \sigma_n^1(\mathbf{x}), \cdots, \sigma_n^n(\mathbf{x}))$ such that

$$\begin{aligned}(1)\qquad &\sum_{i=1}^\infty 2^{-i}\left|\sum_{\nu=1}^n \chi_{(-\infty,t_i)}(x_\nu) - \sum_{\nu=1}^n F((t_i - \theta_n(\mathbf{x}))/\sigma_n^\nu(\mathbf{x}))\right| \\ &\leqq 2\inf\Big\{\sum_{i=1}^\infty 2^{-i}\left|\sum_{\nu=1}^n \chi_{(-\infty,t_i)}(x_\nu) - \sum_{\nu=1}^n F((t_i - t)/s^\nu)\right| : \\ &\qquad\qquad (t, s^1, \cdots, s^n) \varepsilon R \times (0,\infty)^n\Big\}.\end{aligned}$$

(That this map can be chosen measurable follows as in Pfanzagl (1969), part (1) in the proof of Lemma 2).

We shall show that the sequence of functions $(\theta_n)_{n\varepsilon N}$ thus defined has the properties asserted in the theorem.

(ii) Let $\theta \varepsilon R$ and $\epsilon > 0$ be fixed. We shall show that there exists some $i_0 \varepsilon N$ (depending on θ and ϵ) such that $|\theta_n(\mathbf{x}) - \theta| \geqq \epsilon$ implies

$$\begin{aligned}(2)\qquad & 2^{-i_0}\min\Big\{\sum_{i=1}^n (F(\epsilon/2\sigma^\nu) - \tfrac{1}{2}), \sum_{\nu=1}^n (\tfrac{1}{2} - F(-\epsilon/2\sigma^\nu)\Big\} \\ &\leqq 3\sum_{i=1}^\infty 2^{-i}\left|\sum_{\nu=1}^n \chi_{(-\infty,t_i)}(x_\nu) - \sum_{\nu=1}^n F((t_i - \theta)/\sigma^\nu)\right|.\end{aligned}$$

Let $i_1, i_2 \varepsilon N$ be such that $\theta - \epsilon < t_{i_1} < \theta - \epsilon/2$ and $\theta + \epsilon/2 < t_{i_2} < \theta + \epsilon$ and let $i_0 = \max(i_1, i_2)$.

At first we assume that $\theta + \epsilon \leqq \theta_n(\mathbf{x})$. As P has median zero, we obtain for all $\nu \varepsilon N$

$$F(\epsilon/2\sigma^\nu) \leqq F((t_{i_2} - \theta)/\sigma_\nu)$$

and

$$\tfrac{1}{2} \geqq F((t_{i_2} - \theta_n(\mathbf{x}))/\sigma_n^\nu(\mathbf{x}))$$

Therefore, for all $n \varepsilon N$,

$$\sum_{\nu=1}^{n} (F(\epsilon/2\sigma^{\nu}) - \tfrac{1}{2}) \leqq \sum_{\nu=1}^{n} F((t_{i_2} - \theta)/\sigma^{\nu}) - \sum_{\nu=1}^{n} F((t_{i_2} - \theta_n(\mathbf{x}))/\sigma_n^{\nu}(\mathbf{x})).$$

Using (1), we obtain

$$2^{-i_0}\sum_{\nu=1}^{n} (F(\epsilon/2\sigma^{\nu}) - \tfrac{1}{2}) \leqq \sum_{i=1}^{\infty} 2^{-i}\Big|\sum_{\nu=1}^{n} F((t_i - \theta)/\sigma^{\nu}) - \sum_{\nu=1}^{n} F((t_i - \theta_n(\mathbf{x}))/\sigma_n^{\nu}(\mathbf{x}))\Big| \leqq \sum_{i=1}^{\infty} 2^{-i}\Big|\sum_{\nu=1}^{n} \chi_{(-\infty,t_i)}(x_\nu) - \sum_{\nu=1}^{n} F((t_i - \theta)/\sigma^{\nu})\Big| + \sum_{i=1}^{\infty} 2^{-i}\Big|\sum_{\nu=1}^{n} \chi_{(-\infty,t_i)}(x_\nu) - \sum_{\nu=1}^{n} F((t_i - \theta_n(\mathbf{x}))/\sigma_n^{\nu}(\mathbf{x}))\Big| \leqq 3 \sum_{i=1}^{\infty} 2^{-i}\Big|\sum_{\nu=1}^{n} \chi_{(-\infty,t_i)}(x_\nu) - \sum_{\nu=1}^{n} F((t_i - \theta)/\sigma^{\nu})\Big|.$$

Together with the corresponding argument for $\theta_n(\mathbf{x}) \leqq \theta - \epsilon$, this yields (2).

(iii) Let f be a unimodal density of P with respect to Lebesgue measure and let $c_0 = 2^{-i_0}\cdot\frac{1}{6}\epsilon \min (f(\epsilon/2), f(-\epsilon/2))$. (We remark that c_0 depends not only on ϵ but via i_0 also on θ.) As f is unimodal, $c_0 > 0$ if ϵ is sufficiently small. We have

$$F(\epsilon/2\sigma^{\nu}) - \tfrac{1}{2} \geqq F(\epsilon/2(1 + \sigma^{\nu})) - \tfrac{1}{2} \geqq \tfrac{1}{2}\epsilon(1 + \sigma^{\nu})^{-1}f(\tfrac{1}{2}\epsilon(1 + \sigma^{\nu})^{-1}) \geqq \tfrac{1}{2}\epsilon(1 + \sigma^{\nu})^{-1}f(\epsilon/2) \geqq 3\cdot 2^{i_0}c_0(1 + \sigma^{\nu})^{-1}.$$

Similarly,

$$\tfrac{1}{2} - F(-\epsilon/2\sigma^{\nu}) \geqq 3\cdot 2^{i_0}c_0(1 + \sigma^{\nu})^{-1}.$$

Together with (2) we obtain that $|\theta_n(\mathbf{x}) - \theta| \geqq \epsilon$ implies

$$(3) \quad c_0\sum_{\nu=1}^{n} (1 + \sigma^{\nu})^{-1} \leqq \sum_{i=1}^{\infty} 2^{-i}\Big|\sum_{\nu=1}^{n} \chi_{(-\infty,t_i)}(x_\nu) - \sum_{\nu=1}^{n} F((t_i - \theta)/\sigma^{\nu})\Big|.$$

(iv) We have

$$\int \Big(\sum_{\nu=1}^{n} \chi_{(-\infty,t_i)}(x_\nu) - \sum_{\nu=1}^{n} F((t_i - \theta)/\sigma^{\nu})\Big)\, d\textstyle\bigtimes_{\nu=1}^{\infty} P_{\theta,\sigma^{\nu}} = 0$$

and

$$s_{ni}^2 = \int \Big(\sum_{\nu=1}^{n} \chi_{(-\infty,t_i)}(x_\nu) - \sum_{\nu=1}^{n} F((t_i - \theta)/\sigma^{\nu})\Big)^2 d\textstyle\bigtimes_{\nu=1}^{n} P_{\theta,\sigma^{\nu}} = \sum_{\nu=1}^{n} F((t_i - \theta)/\sigma^{\nu})(1 - F((t_i - \theta)/\sigma^{\nu})) \leqq n/4.$$

Let $a \varepsilon (1, 2)$ be arbitrary. As $(a - 1)\sum_{i=1}^{\infty} a^{-i} = 1$, we obtain from (3) by Chebychev's inequality:

$$\textstyle\bigtimes_{\nu=1}^{\infty} P_{\theta,\sigma^{\nu}}\{\mathbf{x} \varepsilon R^{N}: |\theta_n(\mathbf{x}) - \theta| \geqq \epsilon\}$$
$$\leqq \textstyle\bigtimes_{\nu=1}^{\infty} P_{\theta,\sigma^{\nu}}\{\mathbf{x} \varepsilon R^{N}: c_0\sum_{\nu=1}^{n} (1 + \sigma^{\nu})^{-1} \leqq \sum_{i=1}^{\infty} 2^{-i}|\sum_{\nu=1}^{n} \chi_{(-\infty,t_i)}(x_\nu) - \sum_{\nu=1}^{n} F((t_i - \theta)/\sigma^{\nu})|\}$$
$$\leqq \textstyle\bigtimes_{\nu=1}^{\infty} P_{\theta,\sigma^{\nu}}\bigcup_{i=1}^{\infty} \{\mathbf{x} \varepsilon R^{N}: a^{-i}(a - 1)c_0\sum_{\nu=1}^{n} (1 + \sigma^{\nu})^{-1}$$

$$\leqq 2^{-i}|\sum_{\nu=1}^{n} \chi_{(-\infty,t_i)}(x_\nu) - \sum_{\nu=1}^{n} F((t_i - \theta)/\sigma^\nu)|\}$$
$$\leqq \sum_{i=1}^{\infty} \times_{\nu=1}^{\infty} P_{\theta,\sigma^\nu}\{\mathbf{x} \varepsilon R^N: (a/2)^{-i}(a-1)c_0 \sum_{\nu=1}^{n} (1+\sigma^\nu)^{-1}$$
$$\leqq |\sum_{\nu=1}^{n} \chi_{(-\infty,t_i)}(x_\nu) - \sum_{\nu=1}^{n} F((t_i - \theta)/\sigma^\nu)|\}$$
$$\leqq \sum_{i=1}^{\infty} \tfrac{1}{4} n(a/2)^{2i}(a-1)^{-2} c_0^{-2} (\sum_{\nu=1}^{n} (1+\sigma^\nu)^{-1})^{-2}$$
$$= a^2(a-1)^{-2} 4(4-a^2)^{-1} c_0^{-2} \cdot (n^{-\frac{1}{2}} \sum_{\nu=1}^{n} (1+\sigma^\nu)^{-1})^{-2},$$

which converges to zero by (A). Hence (A) implies consistency of $(\theta_n)_{n \varepsilon N}$.

(v) Now we shall prove strong consistency under (B). By the law of the iterated logarithm, for any $i \varepsilon N$ there exists a $\times_{\nu=1}^{\infty} P_{\theta,\sigma^\nu}$ – null set X_i such that $\mathbf{x} \not\varepsilon X_i$ implies

$$\limsup_{n\varepsilon N} (2s_{ni}^2 \log\log s_{ni}^2)^{-\frac{1}{2}} |\sum_{\nu=1}^{n} \chi_{(-\infty,t_i)}(x_\nu) - \sum_{\nu=1}^{n} F((t_i - \theta)/\sigma^\nu)| = 1.$$

(For the definition of s_{ni}^2 see part (iv).)

As $s_{ni}^2 \leqq n$ for all $n \varepsilon N$, we obtain for $\mathbf{x} \not\varepsilon X_i$:

$$(4) \quad \limsup_{n\varepsilon N} (2n \log\log n)^{-\frac{1}{2}} |\sum_{\nu=1}^{n} \chi_{(-\infty,t_i)}(x_\nu) - \sum_{\nu=1}^{n} F((t_i - \theta)/\sigma^\nu)| \leqq 1.$$

Let $X_0 = \bigcup_{i=1}^{\infty} X_i$. If there exists $\mathbf{x} \not\varepsilon X_0$ such that $|\theta_n(\mathbf{x}) - \theta| \geqq \epsilon$ for infinitely many n, say N_0, we obtain from (3) and (4):

$$2^{-\frac{1}{2}} c_0 \limsup_{n\varepsilon N_0} (n \log\log n)^{-\frac{1}{2}} \sum_{\nu=1}^{n} (1+\sigma^\nu)^{-1} \leqq 1.$$

This, however, contradicts (B). Hence $(\theta_n(\mathbf{x}))_{n\varepsilon N}$ converges to θ for all $\mathbf{x} \not\varepsilon X_0$. As X_0 is a $\times_{\nu=1}^{\infty} P_{\theta,\sigma^\nu}$ – null set, this proves the assertion.

REMARK. It seems natural to replace the distance function

$$\sum_{i=1}^{\infty} 2^{-i} |\sum_{\nu=1}^{n} \chi_{(-\infty,t_i)}(x_\nu) - \sum_{\nu=1}^{n} F((t_i - \theta)/\sigma^\nu)|$$

by the distance function $\sup_{t\varepsilon R} |\sum_{\nu=1}^{n} \chi_{(-\infty,t)}(x_\nu) - \sum_{\nu=1}^{n} F((t - \theta)/\sigma^\nu)|$. This would, however, presuppose to generalize a deep study like that of Chung (1949) to cover the case of not necessarily identical distribution functions.

Acknowledgment. The author wishes to thank Mr. J. Weber for a careful reading of the manuscript and the referee for bringing a slip to his attention.

REFERENCES

[1] CHUNG, K. L. (1949). An estimate concerning the Kolmogorov limit distribution. *Trans. Amer. Math. Soc.* **67** 36–50.

[2] LOÈVE, M. (1963). *Probability Theory.* Van Nostrand, Princeton, N.J.

[3] PFANZAGL, J. (1969). Consistent estimation in the presence of incidental parameters. To appear in *Metrika* **15**.

[4] WOLFOWITZ, J. (1953). Estimation by the minimum distance method. *Ann. Inst. Statist. Math.* **5** 9–23.