XIV. CONSISTENCY OF MAXIMUM-LIKELIHOOD ESTIMATES IN THE EXPLOSIVE CASE

BY HERMAN RUBIN

1. *Introduction*

It was shown by Mann and Wald [1943] that if a temporal stochastic process described by linear difference equations is damped without disturbances, it is stable (the expectations of the squares of all variables are uniformly bounded) with disturbances and the "maximum-likelihood" estimates of the parameters involved are consistent. However, a system which is stable or explosive (the expectations of the squares of some variables being unbounded) without disturbances is explosive with them. Here we prove consistency in a simple example of the explosive case.

Let us consider the equation

$$(1) \qquad x_t = \rho x_{t-1} + u_t, \qquad |\rho| \geq 1 \qquad (t = 1, 2, \dots),$$

where ρ is a real number, the u_t are real stochastic variables independently distributed with mean 0 and variance σ^2, and x_0 is a given real number. (The results derived here hold equally well if ρ, u_t, and x_0 are complex numbers, quaternions, or Cayley numbers.) The maximum-likelihood estimate $\hat{\rho}$ of the parameter ρ is defined as

$$(2) \qquad \hat{\rho} = \frac{\sum x_t x_{t-1}}{\sum x_{t-1}^2},$$

where

$$(3) \qquad \sum = \sum_{t=1}^{T}.$$

[In the case of complex numbers, quaternions, or Cayley numbers, replace (2) by

$$(2') \qquad \hat{\rho} = \frac{\sum x_t \bar{x}_{t-1}}{\sum x_{t-1} \bar{x}_{t-1}} \, ,$$

where $\bar{x}$ denotes the conjugate of x.] We shall show that

$$(4) \qquad \underset{T \to \infty}{\text{plim}} \; \hat{\rho} = \rho \, .$$

2. *Transformation of the Problem*

We can write (2) in the form

$$(5) \qquad \begin{aligned} \hat{\rho} = \frac{\sum (\rho x_{t-1} + u_t) x_{t-1}}{\sum x_{t-1}^2} &= \frac{\sum \rho x_{t-1}^2}{\sum x_{t-1}^2} + \frac{\sum u_t x_{t-1}}{\sum x_{t-1}^2} \\ &= \rho + \frac{\sum u_t x_{t-1}}{\sum x_{t-1}^2} \, . \end{aligned}$$

By the Cauchy-Schwarz inequality,

$$(6) \qquad \left| \sum u_t x_{t-1} \right| \le \left(\sum u_t^2\right)^{1/2} \left(\sum x_{t-1}^2\right)^{1/2} .$$

Hence

$$(7) \qquad \left| \frac{\sum u_t x_{t-1}}{\sum x_{t-1}^2} \right| \le \frac{\left(\sum u_t^2\right)^{1/2}}{\left(\sum x_{t-1}^2\right)^{1/2}} \, .$$

But the u_t^2 are real variables independently distributed with identical distribution functions and means σ^2. Hence [Cramér, Theorem 15]

$$(8) \qquad \text{plim} \, \frac{1}{T} \sum u_t^2 = \sigma^2 \, .$$

Let us consider the quantities x_t. By (1), we have

$$x_t = \rho^t x_0 + \rho^{t-1} u_1 + \cdots + \rho u_{t-1} + u_t . \tag{9}$$

Dividing (9) by ρ^t, we obtain

$$\rho^{-t} x_t = x_0 + \sum_{\tau=1}^{t} \rho^{-\tau} u_\tau . \tag{10}$$

3. *System Stable Without Disturbances*

In this section let us assume $|\rho| = 1$. Then

$$|\rho^{-\tau} u_\tau| = |u_\tau| . \tag{11}$$

We observe that for any $\delta > 0$ there exists a number $A(\delta)$ such that if $B > A(\delta)$ then

$$\int_{|u_\tau| > B} |\rho^{-\tau} u_\tau|^2 \, dF(u_\tau) = \int_{|u_\tau| > B} |u_\tau|^2 \, dF(u_\tau) < \delta , \tag{12}$$

since $\mathcal{E}(|u_\tau|^2) = \mathcal{E}(u_\tau^2) = \sigma^2 < \infty$. Therefore, if $t > A^2(\delta)/\varepsilon^2$, we have

$$\frac{1}{t} \sum_{\tau=1}^{t} \int_{|u_\tau| > \varepsilon\sqrt{t}} |\rho^{-\tau} u_\tau|^2 \, dF(u_\tau) < \frac{1}{t} \sum_{\tau=1}^{t} \delta = \delta . \tag{13}$$

Hence, for every $\varepsilon > 0$,

$$\lim_{t \to \infty} \frac{1}{t} \sum_{\tau=1}^{t} \int_{|u_\tau| > \varepsilon\sqrt{t}} |\rho^{-\tau} u_\tau|^2 \, dF(u_\tau) = 0 . \tag{14}$$

By the central limit theorem [Cramér, Theorem 21a],

$$x = \frac{1}{\sqrt{t}} \sum_{\tau=1}^{t} \rho^{-\tau} u_\tau$$

is asymptotically normally distributed with mean 0 and variance σ^2. Therefore, its cumulative distribution function $G_t(x)$ converges uniformly[1] to the normal distribution $N(x, 0, \sigma^2)$. Let us now prove the following lemma:

LEMMA 3. *There exists a number* λ *such that for every* $\varepsilon > 0$ *there is a* t_0 *such that if* $t > t_0$ *then for every* b, $P_t(|x-b| < \varepsilon) < \lambda\varepsilon$.

Since $G_t(x)$ converges uniformly to $N(x, 0, \sigma^2)$, we see that for $\delta\varepsilon > 0$ there is a t_0 such that if $t > t_0$ then

$$(15) \qquad |G_t(x) - N(x, 0, \sigma^2)| < \delta\varepsilon .$$

Then

$$(16) \quad G_t(b+\varepsilon) - G_t(b-\varepsilon) < N(b+\varepsilon, 0, \sigma^2) - N(b-\varepsilon, 0, \sigma^2) + 2\delta\varepsilon .$$

But $N(b+\varepsilon, 0, \sigma^2) - N(b-\varepsilon, 0, \sigma^2) < 2\varepsilon/\sqrt{2\pi}\sigma$, since $1/\sqrt{2\pi}\sigma$ is the maximum of dN/dx. Therefore

$$(17) \quad P_t(|x-b| < \varepsilon) = G_t(b+\varepsilon) - G_t(b-\varepsilon) < \varepsilon\left(\frac{2}{\sqrt{2\pi}\,\sigma} + 2\delta\right)$$

for $t > t_0$. Take b in the foregoing lemma to be $-x_0/\sqrt{t}$. Then we obtain from (10)

$$(18) \qquad P\left(\left|\frac{\rho^{-t} x_t}{\sqrt{t}}\right| < \varepsilon\right) < \lambda\varepsilon ,$$

or, since $|\rho| = 1$,

$$(19) \qquad P(|x_t| < \varepsilon\sqrt{t}) < \lambda\varepsilon .$$

From (1) we see that

$$(20) \quad \left||x_{t+1}| - |x_t|\right| = \left||x_{t+1}| - |\rho x_t|\right| \le |x_{t+1} - \rho x_t| = |u_{t+1}| .$$

In the Tchebycheff inequality,

$$(21) \qquad P(|u| > \theta\sigma) < \frac{1}{\theta^2} ,$$

[2]Let F_n be a sequence of distribution functions converging to a continuous distribution function F. Then F_n converges uniformly to F.

let us take

$$\theta = \frac{\varepsilon\sqrt{t}}{2\mu\sigma}. \tag{22}$$

Then

$$P(|u_\tau| > \frac{\varepsilon\sqrt{t}}{2\mu}) < \frac{4\mu^2\sigma^2}{\varepsilon^2 t}. \tag{23}$$

Therefore

$$P(\{|u_{t+1}| + |u_{t+2}| + \cdots + |u_{t+\mu}|\} > \frac{\varepsilon\sqrt{t}}{2}) < \frac{4\mu^3\sigma^2}{\varepsilon^2 t}, \tag{24}$$

since if $|u_{t+1}| + \cdots + |u_{t+\mu}| > \varepsilon\sqrt{t}/2$, at least one $|u_{t+i}|$ is greater than $\varepsilon\sqrt{t}/2\mu$, and

$$P(\text{at least one } u_{t+i} > \frac{\varepsilon\sqrt{t}}{2\mu}) \le \sum_{i=1}^{\mu} P(|u_{t+i}| > \frac{\varepsilon\sqrt{t}}{2\mu}). \tag{25}$$

On the other hand it follows that if $k \le \mu$,

$$|x_{t+k}| \ge |x_t| - \sum_{\tau=1}^{k} |u_{t+\tau}| \ge |x_t| - \sum_{\tau=1}^{\mu} |u_{t+\tau}|. \tag{26}$$

Therefore

$$P(\min\{|x_t|, |x_{t+1}|, \ldots, |x_{t+\mu}|\} < \frac{\varepsilon\sqrt{t}}{2}) < \lambda\varepsilon + \frac{4\mu^3\sigma^2}{\varepsilon^2 t}. \tag{27}$$

Hence we see that

$$P(\sum_{\tau=t}^{t+\mu} x_\tau^2 < \frac{\varepsilon(\mu+1)t}{4}) < \lambda\varepsilon + \frac{4\mu^3\sigma^2}{\varepsilon^2 t}. \tag{28}$$

But

$$\sum_{\tau=0}^{t+\mu} x_\tau^2 \ge \sum_{\tau=t}^{t+\mu} x_\tau^2. \tag{29}$$

Take $\mu = [4A\nu/\varepsilon^2]$, $\nu > 1$, where $[x]$ denotes the greatest integer less than or equal to x. Then

$$(30)\qquad P\Big(\sum_{\tau=0}^{t+[4A\nu/\varepsilon^2]} x_\tau < A\nu t\Big) < \lambda\varepsilon + \frac{256\,\sigma^2 A^3\,\nu^3}{\varepsilon^8\, t}\,.$$

And for any $t > ([4A\nu/\varepsilon^2]+1)/(\nu-1)$ we have

$$(31)\qquad t + \left[\frac{4A\nu}{\varepsilon^2}\right] < \nu t - 1\,.$$

Let $\nu t = T$. Since

$$(32)\qquad \sum_{\tau=0}^{t+[4A\nu/\varepsilon^2]} x_\tau^2 < \sum_{\tau=1}^{T} x_{\tau-1}^2\,,$$

it follows that

$$(33)\qquad P\Big(\sum x_{t-1}^2 < A\,T\Big) < \lambda\varepsilon + \frac{256\,\sigma^2 A^3\,\nu^4}{\varepsilon^8\,T}$$

for $T > \max\{t_0,\ ([4A\nu/\varepsilon^2]+1)/(\nu-1)\}$, with t_0 defined as in the preceding lemma. Therefore

$$(34)\qquad \operatorname*{plim}_{T\to\infty} \frac{1}{T}\sum x_{t-1}^2 \geq A$$

for every $A > 0$. Hence

$$(35)\qquad \operatorname*{plim}_{T\to\infty} \frac{\sum u_t^2}{\sum x_{t-1}^2} = \operatorname*{plim}_{T\to\infty} \frac{\frac{1}{T}\sum u_t^2}{\frac{1}{T}\sum x_{t-1}^2} \leq \frac{\sigma^2}{A}\,.$$

Therefore

$$(36)\qquad \operatorname*{plim}_{T\to\infty} \frac{(\sum u_t^2)^{1/2}}{(\sum x_{t-1}^2)^{1/2}} = 0\,,$$

and, from (5),

(37) $$\operatorname*{plim}_{T\to\infty} \hat{\rho} = \rho .$$

4. *System Explosive Without Disturbances*

Let us assume now that $|\rho| > 1$. Denote by $F*G$ the convolution

(38) $$F*G(x) = \int_{-\infty}^{+\infty} F(x-t)\,dG(t)$$

of the distributions F and G. If the distribution function of u_τ is $F(x)$, then the distribution function of $\rho^{-\tau} u_\tau$ is $F(\rho^\tau x)$. Since the mean of $\rho^{-\tau} u_\tau$ is 0 and the variance of $\rho^{-\tau} u_\tau$ is $\rho^{-2\tau}\sigma^2$, it follows [Wintner, Theorem 7.1] that if

(39) $$F(x)*F(\rho x)*\cdots*F(\rho^n x) = G_n(x) ,$$

then $\lim_{n\to\infty} G_n(x)$ exists and is a distribution function $G(x)$.

Let us now prove the following lemma:

LEMMA 4. *$G(x)$ is continuous.*

We shall consider three cases, for although the proof for the third case applies also to the other two, it is the most complicated.

Case 1. $F(x)$ is continuous. Then [Cramér, p. 37, equation (40)] it follows that G is continuous.

For any distribution function $H(x)$ denote by the functional $J\{H(x)\}$ the sum of all jumps of $H(x)$. We observe that $J\{H(\lambda x)\} = J\{H(x)\}$, and that $J\{H(x)\} = 0$ if and only if $H(x)$ is continuous. If $J\{H(x)\} < 1$, we shall say $H(x)$ is partly continuous, and if $J\{H(x)\} = 1$, we shall say that $H(x)$ is discrete.

Case 2. $H(x)$ is partly continuous. Let $J\{H(x)\} = \delta < 1$. From (37) we have

(40) $$G_n(x) = F(x) * G_{n-1}(\rho x) .$$

But

$$J(H * K) = J(H)\, J(K). \tag{41}$$

Therefore

$$J\{G_n(x)\} = \delta^n . \tag{42}$$

We may derive from (37)

$$G(x) = G_n(x) * G(\rho^n x) . \tag{43}$$

Using (39) and (40) we obtain

$$J\{G(x)\} = \delta^n J\{G(x)\} \leq \delta^n . \tag{44}$$

Therefore $J\{G(x)\} = 0$, and $G(x)$ is continuous.

Case 3. $H(x)$ is discrete. Denote by $L\{H(x)\}$ the maximum jump of $H(x)$. We see that $L\{H(\lambda x)\} = L\{H(x)\}$. We have $L\{F(x)\} = \varphi < 1$. For if $L\{F(x)\} = 1$, then

$$\begin{aligned} P(y \leq x) &= 0 \qquad x < x^0 , \\ P(y \leq x) &= 1 \qquad x \geq x^0 , \end{aligned} \tag{45}$$

and the variance of u_τ would be 0, which contradicts its being $\sigma^2 > 0$. Using the fact that [Wintner, Theorem 7.6] $J\{G(x)\} > 0$ if and only if

$$\prod_{k=1}^{\infty} L\{F(\rho^k x)\} > 0 ,$$

we see from

$$\prod_{k=1}^{\infty} L\{F(\rho^k x)\} = \prod_{k=1}^{\infty} \varphi = 0 \tag{46}$$

that $G(x)$ is continuous.

Since the $G_n(x)$ approach $G(x)$, it follows that for any $\varepsilon > 0$, there are a δ and a t_0 such that if $n > t_0$ then

$$P_n(|x + x_0| < \delta) < \varepsilon . \tag{47}$$

We observe that $\rho^{-t}x_t - x_0$ has the distribution G_t. Therefore, for $t > t_0$, we have

$$(48)\qquad P(|\rho^{-t} x_t| < \delta) < \varepsilon .$$

We further observe that for any $\delta > 0$, and for any K, there is a t_1 such that, if $t > t_1$, then

$$(49)\qquad \frac{|\rho|^t \delta}{\sigma\sqrt{t}} > K .$$

From (46), we have for $t > t_0$,

$$(50)\qquad P(|x_t| < |\rho|^t \delta) < \varepsilon .$$

Therefore, for $t > \max(t_0, t_1)$, we see that

$$(51)\qquad P(|x_t| > K\sigma\sqrt{t}) > 1 - \varepsilon .$$

Then

$$(52)\qquad \operatorname*{plim}_{T\to\infty} \frac{\sum u_t^2}{\sum x_{t-1}^2} \leq \operatorname*{plim}_{T\to\infty} \frac{\frac{1}{T}\sum u_t^2}{\frac{1}{T} x_{T-1}^2} \leq \lim_{T\to\infty} \frac{\sigma^2}{K^2\sigma^2 \frac{T-1}{T}} = \frac{1}{K^2} .$$

Therefore

$$(53)\qquad \operatorname*{plim}_{T\to\infty} \frac{(\sum u_t^2)^{1/2}}{(\sum x_{t-1}^2)^{1/2}} = \left(\operatorname*{plim}_{T\to\infty} \frac{\sum u_t^2}{\sum x_{t-1}^2}\right)^{1/2} = 0$$

and

$$(54)\qquad \operatorname*{plim}_{T\to\infty} \hat{\rho} = \rho .$$