EXTRACTA MATHEMATICAE Vol. **11**, Núm. 3, 435 – 442 (1996)

The Integrated Squared Error Estimation of Parameters

JAMAL-DINE CHERGUI

B.P. 706 *Tetouan, Morocco*

(Research paper presented by M. Molina)

AMS *Subject Class.* (1991): 55N22 *Received January* 10, 1996

INTRODUCTION

This paper deals with the problem of estimation in the parametric case for discrete random variables. Their study is facilitated by the powerful method of probability generating function.

In [1] the probability generating function was used instead of the densities or the distribution functions in problemes of inference. Most of the attention has been directed toward the use of the sample characteristic function. We now develop a procedure for estimation of parameter of discrete distributions which probability generating function is known except for the parameter. A similar approach under different conditions was studied in [2] and [3] who have considered the use of the sample characteristic function.

Let $X_1, \ldots, X_n$ be iid, integer-valued rv's with common distribution P_X. Let $P_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$ the empirical measure and G_n its probability generating function. G_n is a sufficient, consistent, unbiased estimator of the probability generating function g of the unknown law P_X. We refer to [1] for more detail.

We now assume that the distribution P_X belong to a family of distributions $\mathcal{P} = \{P_\theta \colon \theta \in \Theta\}$ dominated by some sigma-finite measure μ. Thus, for every given θ, we denote $p_\theta(x) = \frac{dP_\theta}{d\mu}(x)$ a version of the density of P_θ respect μ. We suppose the following regularity conditions:

(i) Θ is an open interval of the real line.

(ii) $\Omega_\theta = \{x \in \mathbb{N} \colon p_\theta(x) > 0\}$ is independent of θ.

(iii) $\forall \theta \in \Theta$, $\sum_{x\geq 0} \frac{\partial}{\partial\theta} p_\theta(x)\mu(x) < \infty$, $\sum_{x\geq 0} \frac{\partial^2}{\partial\theta^2} p_\theta(x)\mu(x) < \infty$, and $E_\theta\left(\left(\frac{\partial}{\partial\theta}\log p_\theta(X)\right)^2\right) < \infty$.

Let $S(X,\theta) = \frac{\partial}{\partial\theta}\log p_\theta(X)$. Under (i)-(ii)-(iii), $E_\theta(S(X,\theta)) = 0$ and the Fisher information

$$I(\theta) = E_\theta(S^2(X,\theta)) = -E_\theta\left(\frac{\partial^2}{\partial\theta^2}\log p_\theta(X)\right).$$

Let $\mathcal{P} = \{g(\cdot,\theta)\colon \theta \in \Theta\}$ be the set of the probability generating functions corresponding to $\mathcal{P}$.

1. The integrated squared error estimation of parameters

DEFINITION 1. A statistic $\tilde{\theta}_n = \tilde{\theta}_n(X_1,\ldots,X_n)$ is said to be an integrated squared error estimator (ISEE) if it minimizes

$$A_n(\theta) = \int_0^1 (G_n(s) - g(s,\theta))^2\,ds.$$

We assume that $A_n(\theta)$ is twice differentiable under the integral sign. The derivative of $A_n(\theta)$ is given by

$$A'_n(\theta) = -\frac{2}{n}\sum_{i=1}^{n} H(X_i,\theta),$$

where

$$H(X,\theta) = \int_0^1 (s^X - g(s,\theta))g'(s,\theta)\,ds,$$

being $g'(s,\theta) = \operatorname{cov}_\theta(s^X, \frac{\partial}{\partial\theta}\log p_\theta(X))$. The estimator will be a root of

$$A'_n(\theta) = 0 \tag{2.1}$$

for which $A''_n(\theta) > 0$.

It is easy to check that

$$E_\theta(H(X,\theta)) = 0, \quad |H(X,\theta)| \le 2I(\theta)^{1/2} \quad \text{and} \quad \operatorname{var}_\theta(H(X,\theta)) \le I(\theta).$$

$$\begin{aligned}\operatorname{Var}_\theta(H(X,\theta)) &= \int_0^1\int_0^1 \operatorname{cov}_\theta(s^X,t^X)g'(s,\theta)\,ds\,dt\\ &= \int_0^1\int_0^1 (g(st,\theta) - g(s,\theta)g(t,\theta))g'(s,\theta)g'(t,\theta)\,ds\,dt,\end{aligned}$$

which ensures that the central limit theorem as well as the law of large numbers apply to the sequence of independent variables $\{H(X_n,\theta)\}$ identically distributed as $H(X,\theta)$.

We now proceed to investigate the existence and the asymptotic behaviour of the ISEE for large values of n.

Let θ_0 denote the unknown true value of the parameter in the distribution which we are sampling.

PROPOSITION 1. *Suppose that*

(i) $E_{\theta_0}(|H(X,\theta)|) < \infty$, $\forall\theta \in \Theta$,

(ii) $H(x,\theta)$ *is, for almost all* x, *a continuous function,*

(iii) $L(\theta_0,\theta) = E_{\theta_0}(H(X,\theta))$ *is monotonically decreasing function over a neighbourhood* U_{θ_0} *of* θ_0.

Then, there exists a sequence $(\tilde{\theta}_n)$ *of roots of* (2.1) *such that* $(\tilde{\theta}_n) \xrightarrow{a.s.} \theta_0$ *when* $n \to \infty$.

Proof. By Kolmogorov theorem it follows that, $\forall\theta \in \Theta$,

$$A'_n(\theta) = -\frac{2}{n}\sum_{i=1}^{n} H(X_i,\theta) \xrightarrow{\text{a.s.}} J(\theta_0,\theta) = -2L(\theta_0,\theta)$$

when $n \to \infty$, which is monotonically increasing. Since $J(\theta_0,\theta) = 0$, we have

$$\begin{cases} A'_n(\theta_0-\varepsilon) \xrightarrow{\text{a.s.}} J(\theta_0,\theta_0-\varepsilon) < 0 \\ A'_n(\theta_0+\varepsilon) \xrightarrow{\text{a.s.}} J(\theta_0,\theta_0+\varepsilon) > 0 \end{cases}$$

where $\varepsilon > 0$ is such that $]\theta_0-\varepsilon,\theta_0+\varepsilon[\subset U_{\theta_0}$. Hence, there exists $n_0(\varepsilon)$ such that for $n > n_0(\varepsilon)$,

$$A'_n(\theta_0-\varepsilon) = -\frac{2}{n}\sum_{i=1}^{n} H(x_i,\theta_0-\varepsilon) < 0 \quad \text{a.s.}$$

and

$$A'_n(\theta_0+\varepsilon) = -\frac{2}{n}\sum_{i=1}^{n} H(x_i,\theta_0+\varepsilon) > 0 \quad \text{a.s.}$$

Since by condition (ii) the function $H(x,\theta)$ is for almost all x a continuous function of θ, then for arbitrary $\varepsilon > 0$, the equation (2.1) have a root $\tilde{\theta}_n$ in $]\theta_0-\varepsilon,\theta_0+\varepsilon[$ as soon as $n > n_0(\varepsilon)$. Then the sequence $\tilde{\theta}_n = \tilde{\theta}_n(X_1,\ldots,X_n)$ converge a.s. to θ_0. ∎

If (i)-(ii)-(iii) of Proposition 1 holds, and we suppose that $H(x,\theta)$ is differentiable for all values of x, the function $\frac{\partial}{\partial\theta}H(x,\theta)$ is continuous in θ, uniformly in x and $E_{\theta_0}\left(\left|\frac{\partial}{\partial\theta}H(X,\theta)\right|\right) < \infty$, $\forall\theta\in\Theta$. Then

PROPOSITION 2. *$A_n(\theta)$ achieves the minimum in the root $\tilde{\theta}_n$ of the equation* (2.1).

Proof. Define

$$B_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}H(X_i,\theta).$$

The regularity assumptions implie that

$$\forall\theta\in\Theta,\ \ B_n(\theta) \xrightarrow{\text{a.s.}} E_{\theta_0}\left(\left|\frac{\partial}{\partial\theta}H(X,\theta)\right|\right) = \beta(\theta,\theta_0), \quad \text{when } n\to\infty.$$

Since the function $\frac{\partial}{\partial\theta}H(x,\theta)$ is continuous in θ uniformly in x, we have

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial}{\partial\theta}H(X_i,\tilde{\theta}_n) - \frac{\partial}{\partial\theta}H(X_i,\tilde{\theta}_0)\right) \longrightarrow 0, \quad \text{when } n\to\infty.$$

Thus,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}H(X_i,\tilde{\theta}_n) \underset{n\to\infty}{\longrightarrow} \beta(\theta_0,\theta_0) = -\int_0^1 g'(s,\theta_0)^2 ds < 0 \quad \text{a.s.}$$

Then $\exists n_0$ such that for $n\geq n_0$, $A_n''(\tilde{\theta}_n) > 0$ and consequently $\tilde{\theta}_n$ is an ISEE strongly consistent (when $n\to\infty$). ∎

Remark. (Heathcote [2]) Suppose that the following identifiability holds: $\theta_1\neq\theta_2 \Longleftrightarrow g(\cdot,\theta_1)\neq g(\cdot,\theta_2)$. For an arbitrary $\delta>0$,

$$E_{\theta_0}(A_n(\theta_0\pm\delta) - A_n(\theta_0)) = \int_0^1 (g(s,\theta_0\pm\delta) - g(s,\theta_0))^2 ds > 0.$$

It follows from the strong law of large numbers that the inequality $A_n(\theta_0\pm\delta) > A_n(\theta_0)$ holds with probability one for n sufficiently large. Furthermore, this is a strict global minimum. Since δ is arbitrary and $A_n(\theta)$ differentiable we conclude that there exists a solution $\tilde{\theta}_n$ of (2.1) which is strongly consistent for θ_0.

PROPOSITION 3. *Suppose the following conditions are satisfied:*

(i) $E_{\theta_0}(|H(X,\theta)|) < \infty$, $E_{\theta_0}\left(|\frac{\partial}{\partial\theta}H(X,\theta)|\right) < \infty$, $\forall\theta \in \Theta$.

(ii) *The function* $\frac{\partial}{\partial\theta}H(X,\theta)$ *is continuous in* θ *uniformly in* x.

Then, $\sqrt{n}(\tilde{\theta}_n - \theta_0) \xrightarrow{L} N(0,\sigma^2(\theta_0))$.

Proof. Let $\tilde{\theta}_n$ a root of the equation (2.1) which is a strongly consistent estimator of θ_0. We have

$$\sum_{i=1}^{n} H(X_i,\theta_0) = \sum_{i=1}^{n} H(X_i,\tilde{\theta}_n) + (\theta_0 - \tilde{\theta}_n)\sum_{i=1}^{n}\frac{\partial}{\partial\theta}H(X_i,\tilde{\theta}_n^*) \tag{2.2}$$

where $\tilde{\theta}_n^*$ is a point on the line segment connecting θ_0 and $\tilde{\theta}_n$.
Now

a) $H(X,\theta_0)$ is a r.v. such that $E_{\theta_0}(H(X,\theta_0)) = 0$ and,

$$\operatorname{var}_{\theta_0}(H(X,\theta_0)) = \int_0^1\int_0^1 (g(st,\theta_0) - g(s,\theta_0)g(t,\theta_0)) \cdot g'(s,\theta_0)g'(t,\theta_0)\,ds\,dt = i(\theta_0) < \infty.$$

Therefore, by the central limit theorem,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} H(X_i,\theta_0) \xrightarrow{L} n(0,i(\theta_0)).$$

b)

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}H(X_i,\tilde{\theta}_n^*) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}H(X_i,\theta_0) + \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial}{\partial\theta}H(X_i,\tilde{\theta}_n^*) - \frac{\partial}{\partial\theta}H(X_i,\theta_0)\right) = B_n(\theta_0) + t_n.$$

We have:

$$B_n(\theta_0) \xrightarrow{a.s} E_{\theta_0}\left(\frac{\partial}{\partial\theta}H(X_i,\theta_0)\right) = \beta(\theta_0,\theta_0), \quad \text{when } n \to \infty,$$

$\tilde{\theta}_n^* = \nu\tilde{\theta}_n + (1-\nu)\theta_0$ for some $0 \le \nu \le 1$ and $\tilde{\theta}_n$ is strongly consistent. Then

$$\tilde{\theta}_n^* \xrightarrow{a.s} \theta_0, \quad \text{when } n \to \infty.$$

Since $\frac{\partial}{\partial\theta}H(x,\theta)$ is continuous in θ uniformly in x, we obtain $t_n \xrightarrow{a.s.} 0$.
Hence,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}H(X_i,\tilde{\theta}_n^*) \xrightarrow{a.s.} \beta(\theta_0,\theta_0)<0$$

and from (2.2) we can write

$$\sqrt{n}(\tilde{\theta}_n-\theta_0)=\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}H(X_i,\theta_0)}{-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}H(X_i,\tilde{\theta}_n^*)},$$

for n sufficiently large.
a) and b) together implie that $\sqrt{n}(\tilde{\theta}_n-\theta_0)\xrightarrow{L} N(0,\sigma^2(\theta_0))$ where $\sigma^2(\theta_0)=i(\theta_0)/C(\theta_0)$, $C(\theta_0)=\beta^2(\theta_0,\theta_0)$. ∎

2. Confidence interval

Now, we suggest a method for obtaining a confidence interval of the parameter θ_0. Suppose the assumptions of Proposition 3 holds. We denote $\vec{X}=(X_1,\dots,X_n)$ and let

$$V_n(\vec{X},\theta_0,\theta)=\frac{H(\vec{X},\theta)}{\sqrt{n\lambda(\theta_0,\theta)}},$$

where $H(\vec{X},\theta)=\sum_{i=1}^{n}H(X_i,\theta)$ and $\lambda(\theta_0,\theta)=E_{\theta_0}(H^2(X,\theta))$.

Proposition 4. *If the function $\frac{\partial}{\partial\theta}V_n(x,\theta,\theta)$ is continuous in θ uniformly in x, and $T_n=\frac{\partial}{\partial\theta}V_n(\vec{X},\tilde{\theta}_n,\tilde{\theta}_n)$, then the interval*

$$\left]\tilde{\theta}_n-\frac{\alpha}{|T_n|},\tilde{\theta}_n+\frac{\alpha}{|T_n|}\right[$$

is, for large n, an asymptotic $100\lambda\%$ confidence interval for θ_0, where α is the quantile of order $(1+\lambda)/2$ of the $N(0,1)$.

Proof. Indeed, an elementary calcul gives:

$$\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}V_n(\vec{X},\theta_0,\theta_0)=\frac{1}{\sqrt{i(\theta_0)}}\frac{1}{n}\frac{\partial}{\partial\theta}H(\vec{X},\theta_0)-\frac{1}{n}H(\vec{X},\theta_0)\frac{1}{i(\theta_0)}\left[\frac{\partial}{\partial\theta}\sqrt{E_{\theta_0}(H^2(\vec{X},\theta))}\right]_{\theta=\theta_0}.$$

According to $\frac{1}{n}H(\vec{X},\theta_0) \xrightarrow{a.s.} 0$ and $\frac{1}{n}\frac{\partial}{\partial\theta}H(\vec{X},\theta_0) \xrightarrow{a.s.} \beta(\theta_0,\theta_0)$, we have

$$\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}V_n(\vec{X},\theta_0,\theta_0) \xrightarrow{a.s.} \frac{\beta(\theta_0,\theta_0)}{\sqrt{i(\theta_0)}}.$$

Furthermore, since $\theta \longrightarrow \frac{\partial}{\partial\theta}V(x,\theta,\theta)$ is continuous in θ uniformly in x, it follows that

$$\frac{1}{\sqrt{n}}T_n = \frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}V_n(\vec{X},\tilde{\theta}_n,\tilde{\theta}_n) \xrightarrow{a.s.} \frac{\beta(\theta_0,\theta_0)}{\sqrt{i(\theta_0)}} = \frac{-1}{\sigma(\theta_0)}.$$

Since,

$$T_n(\theta_0-\tilde{\theta}_n) = \frac{\sqrt{n}(\tilde{\theta}_n-\theta_0)}{\sigma(\theta_0)}\frac{T_n\sigma(\theta_0)}{-\sqrt{n}},$$

in view of Proposition 3 we conclude $T_n(\theta_0-\tilde{\theta}_n) \xrightarrow{L} N(0,1)$ and consequently

$$Pr\left(\theta_0 \in \left]\tilde{\theta}_n - \frac{\alpha}{|T_n|}, \tilde{\theta}_n + \frac{\alpha}{|T_n|}\right[\right) \underset{n\to\infty}{\longrightarrow} \lambda.$$

This complete the proof of the proposition. ∎

EXAMPLE. Let $X \rightsquigarrow B(\theta)$, $\theta \in]0,1[$, the Bernouilli distribution. Then, $g(s,\theta) = \theta(s-1)+1$, $g'(s,\theta) = s-1$.

$$H(X,\theta) = -\frac{1}{4X+2} - \frac{\theta}{3} + \frac{1}{2}, \qquad i(\theta_0) = \frac{\theta_0(1-\theta_0)}{9},$$

$$\beta(\theta_0,\theta_0) = -\frac{1}{3} \qquad \tilde{\theta}_n(X_1,\ldots,X_n) = \frac{3}{2} - \frac{3}{n}\sum_{i=1}^{n}\frac{1}{4X_i+2}.$$

It can be easily shown that, due to the fact that $X_i = 0$ or 1,

$$\tilde{\theta}_n = \frac{1}{n}\sum_{i=1}^{n}, \qquad \mathrm{var}_\theta(\tilde{\theta}_n) = \frac{\left(\frac{\partial}{\partial\theta}E_\theta(\tilde{\theta}_n)\right)^2}{I_n(\theta)}.$$

$\tilde{\theta}_n$ is an efficient, unbiased, strongly consistent estimator of θ and

$$\theta_0 \in \left]\tilde{\theta}_n - \alpha\sqrt{\tilde{\theta}_n(1-\tilde{\theta}_n)}, \tilde{\theta}_n + \alpha\sqrt{\tilde{\theta}_n(1-\tilde{\theta}_n)}\right[$$

is the $100\lambda\%$ confidence asymptotic interval.

ACKNOWLEDGEMENTS

The author would like to express his gratitude to the referee for his precise comments.

References

[1] Chergui, J., Estimation and tests of the discrete probability law based on the empirical generating function (two dimensional case), *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales,* **LXXXVIII** (2-3) (1994), 245–286.

[2] Heathcote, C.R., The integrated squared error estimation of parameters, *Biometrika,* **64** (1977), 255–264.

[3] Paulson, A.S., Holcomb, E.W., Leitch, R.A., The estimation of the paarmeters of the stable laws, *Biometrika,* **62** (1975), 163–170.