The Annals of Mathematical Statistics
1970, Vol. 41, No. 6, 1845–1852

ESTIMATING THE EMPIRIC DISTRIBUTION FUNCTION OF CERTAIN PARAMETER SEQUENCES[1]

BY RICHARD J. FOX[2]

Michigan State University

1. Introduction and notation. A compound decision problem (see Section 6 of Robbins (1951)) consists of a set of n independent statistical decision problems, each having the basic structure of a so-called component problem. The usual objective is to find procedures, which may use all n observations for each decision, whose average risk across problems converges, as $n \to \infty$, to the component problem Bayes risk versus the empiric distribution of the n underlying parameter values. This latter quantity represents the minimum average risk if one employs a procedure where the same decision rule, which does not depend on the set of observations, is applied independently to the individual problems.

One possible solution technique is to use the set of observations to estimate the empiric distribution of the set of parameter values and then use the Bayes procedure versus this estimate in each individual problem. For a very simple example, the estimators developed in Section 2 and Section 3 of this paper, which deal with two specific uniform distributions, could be used in this way to construct procedures which meet the previously specified objective for the compound test of simple hypotheses problem solved by Hannan and Robbins (1955).

Let $\mathbf{x} = (x_1, x_2, \cdots)$ be a sequence of independent random variables with x_i having distribution function F_{θ_i}, henceforth abbreviated to F_i, $\theta_i \in \Omega$ for $i = 1, 2, \cdots$ and Ω a subset of the real line. Suppose that this family, indexed by the parameter $\theta \in \Omega$, is dominated by Lebesgue measure μ and let f_θ be the density of F_θ with respect to μ. Also abbreviate f_{θ_i} by f_i.

Throughout this paper we will occasionally omit the display of the argument of a function of a real variable. We also adopt the convention that distribution functions are right continuous. Let F be a distribution function; we will also use the letter F to denote the corresponding Lebesgue–Stieltjes measure. If A is an event, $[A]$ will be used to denote the indicator function of A.

Let $\mathbf{F} = \times_{i=1}^{\infty} F_i$, i.e. $\mathbf{F}$ is the product measure on the space of $\mathbf{x}$'s corresponding to a particular $\boldsymbol{\theta}$, $\boldsymbol{\theta}$ denoting a parameter sequence: $(\theta_1, \theta_2, \cdots)$. Let G_n be the empiric distribution function of the first n parameters: $\theta_1, \theta_2, \cdots, \theta_n$. We now define the following functions:

$$\bar{F} = \int F_\theta \, dG_n(\theta) = n^{-1} \textstyle\sum_{i=1}^n F_i; \tag{1.1}$$

$$\bar{f} = \int f_\theta \, dG_n(\theta) = n^{-1} \textstyle\sum_{i=1}^n f_i \tag{1.2}$$

Received March 14, 1969; revised April 29, 1970.

[1] Research supported by National Science Foundation, Grant No. GP-13484.

[2] Now at Procter and Gamble.

and note that $\bar{f} = d\bar{F}/d\mu$. Also, for any x, let

$$F^*(x) = n^{-1}\sum_{i=1}^{n}[x_i \leqq x], \tag{1.3}$$

i.e. F^* is the empiric distribution function of $x_1, x_2, \cdots, x_n$. For any distribution function, say F, define, for any $h > 0$ and any x,

$$\Delta F(x) = h^{-1}(F(x+h) - F(x)). \tag{1.4}$$

We allow h to depend on n.

We now make some remarks about the Lévy metric which is discussed on page 215 of Loève (1963). This metric is defined on the space of all distribution functions by the following distance formula. For any two distribution functions F_1 and F_2, letting d denote distance,

$$d(F_1, F_2) = \inf\{\varepsilon > 0 \mid \text{ for all } x, F_1(x-\varepsilon)-\varepsilon \leqq F_2(x) \leqq F_1(x+\varepsilon)+\varepsilon\}.$$

Loève mentions that convergence in Lévy metric of a sequence of distribution functions is equivalent to complete convergence.

In Section 2, we consider the family of uniform on the interval $(0, \theta)$ distributions, $\theta \in (0, \infty)$ and exhibit an estimator of G_n whose Lévy distance from G_n converges to zero a.s. $\mathbf{F}$ for a certain class of $\boldsymbol{\theta}$'s. In Section 3, we deal with the family of uniform on the interval $[\theta, \theta+1)$ distributions, $\theta \in (-\infty, +\infty)$. In this case, we exhibit an estimator whose Lévy distance from G_n converges to zero a.s. $\mathbf{F}$ uniformly in $\boldsymbol{\theta}$.

In Section 4, we again consider the two families mentioned above and assume that the θ's are i.i.d. possessing the distribution function G. Thus, $x_1, x_2 \cdots$ are i.i.d. with a distribution function called the mixed distribution function. We then apply the results of Section 2 and Section 3 to the problem of estimating this prior distribution function G. Estimates of the prior have applications in empirical Bayes decision problems (see Robbins (1964)). In this situation, one is faced with a sequence of independent, identical statistical decision problems in which there is a fixed but unknown prior distribution on the parameter. For the uniform $[\theta, \theta+1)$ family, Theorem 4.3 of Fox (1968) states that in the squared error loss estimation problem, the expected risk of using the Bayes estimator of θ versus a Lévy-convergent estimate of G converges to the Bayes risk versus G.

Robbins (1964) treats the general problem of estimating a prior distribution function, say G. Under certain assumptions, he shows that if the estimate of G is chosen so that the resulting mixed distribution function is within $\varepsilon_n(\varepsilon_n \downarrow 0$ as $n \to \infty)$ of minimizing over the class of possible mixed distribution functions, the sup norm distance from the empiric distribution function of the observations: $x_1, x_2, \cdots x_n$, then the estimator will converge to G in Lévy metric. However, no explicit method is given for obtaining this estimator. The family of Section 2 of this paper is discussed in Robbins' Example 3 and the family of Section 3 is a special case of his Theorem 2.

Deely and Kruse (1968) add to Robbins' assumptions the condition that $F_\theta(x)$ is continuous in x for every θ. They then exhibit a method of finding an estimator satisfying Robbins' condition. Calculating the estimate involves finding an optimal strategy in a certain game.

2. Uniform $(0, \theta)$ case. We now consider the following family of distributions. For $\theta \in \Omega = (0, \infty)$, let $f_\theta(x) = \theta^{-1}[0 < x < \theta]$. It then follows that

$$F_\theta(x) = 0 \qquad x \leqq 0,$$
$$= x\theta^{-1} \qquad 0 < x < \theta,$$
$$= 1 \qquad x \geqq \theta.$$

Thus, for any x, by the definition of $\bar{F}$ in (1.1),

$$\bar{F}(x) = \int([\theta \leqq x] + x\theta^{-1}[\theta > x])\, dG_n(\theta).$$

Hence, recalling the definition of $\bar{f}$ in (1.2), we obtain the following important relationship which suggests an estimator of G_n:

$$\bar{F}(x) = G_n(x) + x\bar{f}(x). \tag{2.1}$$

We estimate G_n at any point $x \geqq 0$ by

$$G_n^*(x) = F^*(x) - x\,\Delta F^*(x) \tag{2.2}$$

where F^* and Δ are defined by (1.3) and (1.4) respectively.

For each n, form the following grid: $0 = x_{n0} < x_{n1} < \cdots < x_{nN} < \beta$, where β and N are both functions of n. For any $\varepsilon > 0$, define for all $x \geqq 0$,

$$A_n(x) = \{\mathbf{x} \mid x\,\Delta F^*(x) > (x-\varepsilon)\bar{f}(x-\varepsilon) + (\bar{F}(x) - \bar{F}(x-\varepsilon)) + \varepsilon/2\},$$
$$B_n(x) = \{\mathbf{x} \mid x\,\Delta F^*(x) < (x+\varepsilon)\bar{f}(x+\varepsilon) - (\bar{F}(x+\varepsilon) - \bar{F}(x)) - \varepsilon/2\}.$$

The following lemma follows immediately from equation (2.1), definition (2.2) and the definitions of $A_n(x)$ and $B_n(x)$.

LEMMA 2.1. *For any $\varepsilon > 0$, for each $x \geqq 0$,*

$$\{\mathbf{x} \mid G_n^*(x) < G_n(x-\varepsilon) - \varepsilon\} \subset \{\mathbf{x} \mid F^*(x) < \bar{F}(x) - \varepsilon/2\} \cup A_n(x),$$
$$\{\mathbf{x} \mid G_n^*(x) > G_n(x+\varepsilon) + \varepsilon\} \subset \{\mathbf{x} \mid F^*(x) > \bar{F}(x) + \varepsilon/2\} \cup B_n(x).$$

LEMMA 2.2. *If $h \to 0$ and $\sum_{n=1}^{\infty} N \exp\{-n(h\varepsilon\beta^{-1})^2/2\} < \infty$, then*

$$\mathbf{F}\{\textstyle\bigcup_{j=0}^{N}(A_n(x_{nj}) \bigcup B_n(x_{nj}))\ \text{i.o.}\} = 0.$$

PROOF. Let $0 < x < \beta$ be fixed and note that $\Delta F^*(x) = n^{-1}\sum_{i=1}^{n} \Delta[x_i \leqq x]$. The variables $\Delta[x_i \leqq x]$ have expectations $\Delta F_i(x)$ and since $f_i(x)$ is a nonincreasing function for $x > 0$, $\Delta F_i(x) \leqq f_i(x)$. It then follows that the expectation of $x\Delta F^*(x)$ is bounded above by $x\bar{f}(x)$. Noting that $(\bar{F}(x) - \bar{F}(x-\varepsilon))$ is bounded below by $\varepsilon\bar{f}(x)$ when $x \geqq \varepsilon$ and by $x\bar{f}(x)$ when $x < \varepsilon$, if we subtract this upper bound, $x\bar{f}(x)$, on the expectation of $x\Delta F^*(x)$ from the right-hand side of the inequality defining $A_n(x)$, then the resulting quantity is bounded below by $\varepsilon/2$. Hence by Theorem 2 of

Hoeffding (1963), since $0 < x < \beta$, $\mathbf{F}(A_n(x)) \leqq \exp\{-n(h\varepsilon\beta^{-1})^2/2\}$. Thus, since $\mathbf{F}(A_n(0)) = 0$ for all n,

$$\mathbf{F}(\textstyle\bigcup_{j=0}^{N} A_n(x_{nj})) \leqq N\exp\{-n(h\varepsilon\beta^{-1})^2/2\}. \tag{2.3}$$

Let n be sufficiently large so that $h \leqq \varepsilon$. Then, $f_i(x+\varepsilon) > 0$ implies $\Delta F_i(x) = f_i(x+\varepsilon)$ and it follows that for all x, $x\Delta F_i(x) \geqq xf_i(x+\varepsilon)$. Hence, the expectation of $x\Delta F^*(x)$ is bounded below by $x\bar{f}(x+\varepsilon)$. Thus, since $(\bar{F}(x+\varepsilon)-\bar{F}(x)) \geqq \varepsilon\bar{f}(x+\varepsilon)$, subtracting $x\bar{f}(x+\varepsilon)$ from the right-hand side of the inequality defining $B_n(x)$ yields a quantity bounded above by $-\varepsilon/2$. Again applying Hoeffding's Theorem 2, since $0 < x < \beta$ and $\mathbf{F}(B_n(0)) = 0$ for all n, we obtain the bound of the right-hand side of (2.3) for $\mathbf{F}(\bigcup_{j=0}^{N} B_n(x_{nj}))$. The infinite series formed by summing this bound over n converges by assumption. Hence, by the Borel–Cantelli Lemma, the proof is complete.

Define the distribution function $\hat{G}_n$ by:

$$\begin{aligned} \hat{G}_n(x) &= 0 && \text{for } x < 0, \\ &= \max\{G_n{}^*(x_{nj}) \mid 0 \leqq x_{nj} \leqq x\} && \text{for } 0 \leqq x < \beta, \\ &= 1 && \text{for } x \geqq \beta. \end{aligned}$$

Note that $G_n{}^*(0) = 0$. Also, let δ be the maximum of the distances between consecutive grid points.

THEOREM 2.1. *If* $\delta \to 0$, $h \to 0$, $G_n(\beta) \to 1$ *and if for every* $\varepsilon > 0$,

$$\textstyle\sum_{n=1}^{\infty} N\exp\{-n(h\varepsilon\beta^{-1})^2\} < \infty,$$

then $d(\hat{G}_n, G_n) \to 0$ a.s. $\mathbf{F}$.

PROOF. Let $\varepsilon > 0$ be arbitrary. By the extension of the Glivenko–Cantelli Theorem to non-identically distributed independent random variables, see Theorem 4.1 of Wolfowitz (1953), $\bar{F}(x) - F^*(x) \to 0$ uniformly in x a.s. $\mathbf{F}$. It then follows from Lemma 2.1 and Lemma 2.2 that

$$\mathbf{F}\{\textstyle\bigcup_{j=0}^{N}(G_n(x_{nj}-\varepsilon)-\varepsilon \leqq G_n{}^*(x_{nj}) \leqq G_n(x_{nj}+\varepsilon)+\varepsilon)' \text{ i.o.}\} = 0, \tag{2.4}$$

where the prime notation is to denote the complement. Since for $0 \leqq x < \beta$, $\hat{G}_n(x) = \hat{G}_n(\tilde{x})$ where $\tilde{x}$ is the largest x_{nj} which is not larger than x,

$$\textstyle\bigcup_{0 \leqq x < \beta}\{\mathbf{x} \mid \hat{G}_n(x) > G_n(x+\varepsilon)+\varepsilon\} \subset \bigcup_{j=0}^{N}\{\mathbf{x} \mid G_n{}^*(x_{nj}) > G_n(x_{nj}+\varepsilon)+\varepsilon\}$$

and if $\delta \leqq \varepsilon$,

$$\textstyle\bigcup_{0 \leqq x < \beta}\{\mathbf{x} \mid \hat{G}_n(x) < G_n(x-2\varepsilon)-\varepsilon\} \subset \bigcup_{j=0}^{N}\{\mathbf{x} \mid G_n{}^*(x_{nj}) < G_n(x_{nj}-\varepsilon)-\varepsilon\}.$$

Since $\delta \to 0$ and $G_n(\beta) \to 1$, by (2.4) and the fact that $\hat{G}_n = G_n$ for $x < 0$, a.s. $\mathbf{F}$, for n sufficiently large, $d(\hat{G}_n, G_n) \leqq 2\varepsilon$, which completes the proof.

3. Uniform $[\theta, \theta+1]$ **case.** We now consider the following family of distributions. For $\theta \in \Omega = (-\infty, +\infty)$ let $f_\theta(x) = [\theta \leqq x < \theta+1]$. It then follows from (1.2) that for all x, $\bar{f}(x) = \int [\theta \leqq x < \theta + 1]\, dG_n(\theta)$. Hence, we have the following relationship which leads us to an estimator of G_n:

$$\bar{f}(x) = G_n(x) - G_n(x-1). \tag{3.1}$$

By (3.1),

$$G_n(x) = \sum_{r=0}^{\infty} \bar{f}(x-r). \tag{3.2}$$

Since $F^*(x) \leqq G_n(x) \leqq F^*(x+1)$, we estimate G_n at a point x by $G_n{}^*(x)$ which is the truncation to the interval $[F^*(x), F^*(x+1)]$ of $\sum_{r=0}^{\infty} \Delta F^*(x-r)$, i.e.

$$G_n{}^*(x) = \{((\sum_{r=0}^{\infty} \Delta F^*(x-r)) \vee F^*(x)) \wedge F^*(x+1)\}. \tag{3.3}$$

For convenience we assume that $h \leqq 1$.

LEMMA 3.1. *For any* $\varepsilon > 0$, *if* $h \leqq \varepsilon$, *then for all* x,

$$\mathbf{F}(\{\mathbf{x} \mid G_n(x-\varepsilon) - \varepsilon \leqq G_n{}^*(x) \leqq G_n(x+\varepsilon) + \varepsilon\}') \leqq 2\exp(-2nh^2\varepsilon^2).$$

PROOF. Since the truncation involved in the definition of $G_n{}^*$ can only improve the estimator, it suffices to prove the lemma for the estimator T_n, defined for all x by $T_n(x) = \sum_{r=0}^{\infty} \Delta F^*(x-r) = n^{-1}\sum_{i=1}^{n} \sum_{r=0}^{\infty} \Delta[x_i \leqq x-r]$. Let x be fixed. Recalling the definition of $\bar{F}$ in (3.1), it is easily seen that

$$\int T_n(x)\, d\mathbf{F} = h^{-1}\sum_{r=0}^{\infty} (\bar{F}(x+h-r) - \bar{F}(x-r)).$$

By (3.1),

$$\sum_{r=0}^{\infty} (\bar{F}(x+h-r) - \bar{F}(x-r)) = \sum_{r=0}^{\infty} (\int_{x-r}^{x+h-r} (G_n(t) - G_n(t-1))\, dt).$$

Writing $\int_{x-r}^{x+h-r} G_n(t-1)\, dt$ as $\int_{x-r-1}^{x+h-r-1} G_n(t)\, dt$ we see that the right-hand side of the equality displayed immediately above is a telescopic series and we obtain

$$\sum_{r=0}^{\infty} (\bar{F}(x+h-r) - \bar{F}(x-r)) = \int_{x}^{x+h} G_n(t)\, dt. \tag{3.4}$$

By (3.4), $\int T_n(x)\, d\mathbf{F} \geqq G_n(x)$. It then follows by Theorem 2 of Hoeffding (1963) that $\mathbf{F}\{\mathbf{x} \mid T_n(x) < G_n(x-\varepsilon) - \varepsilon\} \leqq \exp(-2nh^2\varepsilon^2)$. Similarly, if $h \leqq \varepsilon$, then by (3.4), $\int T_n(x)\, d\mathbf{F} \leqq G_n(x+\varepsilon)$ and applying Hoeffding's bounds again we see that $\mathbf{F}\{\mathbf{x} \mid T_n(x) > G_n(x+\varepsilon) + \varepsilon\} \leqq \exp(-2nh^2\varepsilon^2)$, which completes the proof.

Let $\delta = N^{-1}$, N being a positive integer depending on n and consider the following grid on the real line: $\cdots < -2\delta < -\delta < 0 < \delta < 2\delta < \cdots$. Consider the following distribution function.

$$\hat{G}_n(x) = \sup\{G_n{}^*(j\delta) \mid j\delta \leqq x, j = 0, \pm 1, \pm 2, \cdots\}. \tag{3.5}$$

We now proceed to show that the Lévy distance of $\hat{G}_n$ from G_n converges to zero a.s. $\mathbf{F}$ uniformly in $\boldsymbol{\theta}$, i.e. for each $\varepsilon > 0$, $\exists m(\varepsilon) \ni \mathbf{F}\{\mathbf{x} \mid d(\hat{G}_n, G_n) > \varepsilon$ for some $n \geqq m(\varepsilon)\} < \varepsilon$ for all $\boldsymbol{\theta}$.

THEOREM 3.1. *If* $\sum_{n=1}^{\infty} N\exp(-2nh^2\varepsilon^2) < \infty$ *for all* $\varepsilon > 0$, $\hat{G}_n$ *is defined by equation* (3.5) *and if* $N \to \infty$ *and* $h \to 0$, *then* $d(\hat{G}_n, G_n) \to 0$ a.s. $\mathbf{F}$ *uniformly in* $\boldsymbol{\theta}$.

PROOF. Let $\varepsilon > 0$ be arbitrary. Let q be an integer sufficiently large so that $n \geqq q$ implies that $h \leqq \varepsilon$ and $\delta \leqq \varepsilon$ and let $n \geqq q$. Define J to be the largest integer such that $F^*(J\delta+1) \leqq \varepsilon$. Define the following subset of the real line, letting $\mathscr{J} = \{j \mid (F^*((j+1)\delta+1) - F^*(j\delta)) > \varepsilon, j \geqq J, j = 0 \pm 1, \pm 2, \cdots\}$,

$$A_n = \bigcup_{j \in \mathscr{J}} [j\delta, (j+1)\delta).$$

Note that there are at most $L = (N+1)K$ grid points in A_n, where K is the smallest integer greater than or equal to ε^{-1}. Also note that A_n may be empty. Since $F^*(J\delta+1) \leqq \varepsilon$ implies that for all $x < J\delta$ both $G_n(x) \leqq \varepsilon$ and $\hat{G}_n(x) \leqq \varepsilon$, it follows that if $x < J\delta$, then $G_n(x-\varepsilon) - \varepsilon \leqq \hat{G}_n(x) \leqq G_n(x+\varepsilon) + \varepsilon$. For $m \geqq J$, if $(F^*((m+1)\delta+1) - F^*(m\delta)) \leqq \varepsilon$, then for $x \in [m\delta, (m+1)\delta)$, since both $G_n(x)$ and $\hat{G}_n(x)$ are in the interval $[F^*(m\delta), F^*((m+1)\delta+1)]$, it follows that $G_n(x-\varepsilon) - \varepsilon \leqq \hat{G}_n(x) \leqq G_n(x+\varepsilon) + \varepsilon$.

Let $[m\delta, (m+1)\delta) \subset A_n$. For all x in this interval $\hat{G}_n(x) = \hat{G}_n(m\delta)$. Thus, since $x \notin A_n$ implies $\hat{G}_n(x) \leqq G_n(x+\varepsilon) + \varepsilon$,

$$\bigcup_{x \in A_n} \{\mathbf{x} \mid \hat{G}_n(x) > G_n(x+\varepsilon) + \varepsilon\} \subset \bigcup_{j\delta \in A_n} \{\mathbf{x} \mid G_n^*(j\delta) > G_n(j\delta+\varepsilon) + \varepsilon\}$$

and since $\delta \leqq \varepsilon$,

$$\bigcup_{x \in A_n} \{\mathbf{x} \mid \hat{G}_n(x) < G_n(x-2\varepsilon) - \varepsilon\} \subset \bigcup_{j\delta \in A_n} \{\mathbf{x} \mid G_n^*(j\delta) < G_n(j\delta-\varepsilon) - \varepsilon\}.$$

The $\mathbf{F}$-measure of the union of the two right-hand sides of the above inclusions, by Lemma 3.1, is no larger than $2L\exp(-2nh^2\varepsilon^2)$. Hence, $\mathbf{F}\{\mathbf{x} \mid d(\hat{G}_n, G_n) > 2\varepsilon$ for some $n \geqq q\} \leqq \sum_{n=q}^{\infty} 2L\exp(-2nh^2\varepsilon^2)$ and since $\sum_{n=1}^{\infty} N\exp(-2nh^2\varepsilon^2) < \infty$ by assumption, the right-hand side of this last inequality can be made less than or equal to 2ε for all $\boldsymbol{\theta}$ by choosing q sufficiently large and the proof is complete.

It can easily be shown that for all $a > 0$, $\alpha > 0$ and all c, $\sum_{n=1}^{\infty} n^c \exp\{-an^{\alpha}\} < \infty$. Hence, if we let $N = n^c$, c being a positive integer, $h = n^{-\alpha}$, $\beta = n^{\gamma}$ with $\alpha, \gamma > 0$ and $\alpha + \gamma < \frac{1}{2}$, then the series of the hypothesis of Theorem 2.1 converges. Also, if we let $N = n^c$, c again being a positive integer and $h = n^{-\alpha}$, $0 < \alpha < \frac{1}{2}$, then the series of the hypothesis of Theorem 3.1 converges.

4. Estimating the prior distribution. Let $(\mathscr{X}, \mathscr{B})$ be the measurable space consisting of the real line and the Borel field. Assume that Ω is a Borel subset of the real line and define the σ-field on Ω to be the restriction of the Borel field to Ω. Let the σ-field on Ω^{∞} be the usual product σ-field. We now drop the assumption that F_{θ} is absolutely continuous with respect to μ for each $\theta \in \Omega$ and refer the reader to page 137 of Loève (1963) for a brief discussion of regular conditional probability. For a proof of the following result, see Lemma 3.5 of Fox (1968).

LEMMA 4.1. *If F_{θ}, $\theta \in \Omega$ is a regular conditional probability measure on $(\mathscr{X}, \mathscr{B})$, then* $\mathbf{F}$, $\boldsymbol{\theta} \in \Omega^{\infty}$, *is a regular conditional probability measure on* $(\mathscr{X}^{\infty}, \mathscr{B}^{\infty})$.

Let P be a probability measure defined on the σ-field of subsets of Ω and F_{θ} be a regular conditional probability. We then have a measure, say H, defined on the product space $\Omega \times \mathscr{X}$, resulting from P on Ω and F_{θ} on $\mathscr{X}$. Let H^{∞} be the usual product measure on the space $(\Omega \times \mathscr{X})^\infty$. Consider the one to one mapping of $(\Omega \times \mathscr{X})^\infty$ onto $\Omega^\infty \times \mathscr{X}^\infty$ which associates the point $(\theta_1, \theta_2, \cdots, x_1, x_2, \cdots)$ with the point $((\theta_1, x_1), (\theta_2, x_2), \cdots)$. The measure induced on $\Omega^\infty \times \mathscr{X}^\infty$ by this mapping and H^∞ on $(\Omega \times \mathscr{X})^\infty$ is the measure resulting from P^∞ on Ω^∞ and the regular conditional probability measure $\mathbf{F}$ (see Lemma 4.1) on $\mathscr{X}^\infty$. Denote the marginal probability measure of $\mathbf{x}$ of the pair $(\boldsymbol{\theta}, \mathbf{x})$ by $P^\infty(\mathbf{F})$.

Consider the identity map from Ω into the real line and let G be the distribution function corresponding to the measure induced on the Borel field of the real line by this mapping and P on Ω. We are interested in estimating G. The following theorem establishes sufficient condition for obtaining a Lévy convergent estimate of G. The subsequent corollaries are applications to the specific families discussed in this paper.

THEOREM 4.1. *If P is a probability measure on the σ-field of Borel subsets of Ω and G is the resulting distribution function and if $\hat{G}_n$ is an estimator of G_n based on $(x_1, x_2, \cdots, x_n)$ such that $d(\hat{G}_n, G_n)$ is a jointly measurable in $(\boldsymbol{\theta}, \mathbf{x})$ for each n and if $P^\infty\{\boldsymbol{\theta} \mid d(\hat{G}_n, G_n) \to 0 \text{ a.s. } \mathbf{F}\} = 1$ and F_θ, $\theta \in \Omega$, is a regular conditional probaiblity measure, then $d(\hat{G}_n, G) \to 0$ a.s. $P^\infty(\mathbf{F})$.*

PROOF. By the triangle inequality, $d(\hat{G}_n, G) \leqq d(\hat{G}_n, G_n) + d(G_n, G)$. By the Glivenko–Cantelli Theorem, page 20 of Loève (1963), $d(G_n, G) \to 0$ a.s. P^∞. Let C be the set of pairs $(\boldsymbol{\theta}, \mathbf{x})$ such that $d(\hat{G}_n, G_n) \nrightarrow 0$ and note that C is jointly measurable in $(\boldsymbol{\theta}, \mathbf{x})$. Since by Lemma 4.1, $\mathbf{F}$ is regular, the measure of C is $\int \mathbf{F}(C)\, dP^\infty$. Since $\mathbf{F}(C) = 0$ a.s. P^∞, the proof is complete.

In the proofs of Corollary 3.1 and Corollary 3.2 of Fox (1968), it is shown that the estimators of Section 2 and Section 3 of this paper are such that for each n, $d(\hat{G}_n, G_n)$ is a measurable function on $(\mathscr{X}^\infty, \mathscr{B}^\infty)$ as was tacitly assumed and then that $d(\hat{G}_n, G_n)$ is jointly measurable in $(\boldsymbol{\theta}, \mathbf{x})$ for each n. Hence, we ignore the joint measurability hypothesis of Theorem 4.1 in the proofs of the following corollaries.

COROLLARY 4.1. *If F_θ corresponds to the uniform distribution on the interval $(0, \theta)$, $\theta \in \Omega = (0, \infty)$ and if P is a probability measure on Ω and G is the resulting distribution function and if $\hat{G}_n$ is defined as in Section 2 and the hypotheses of Theorem 2.1 are satisfied with $\beta \to \infty$ replacing $G_n(\beta) \to 1$, then $d(\hat{G}_n, G) \to 0$ a.s. $P^\infty(\mathbf{F})$.*

PROOF. Let B be a Borel set. $F_\theta(B) = \theta^{-1}\mu(B \cap (0, \theta))$ which is a continuous function of $\theta > 0$. Thus, F_θ, $\theta \in \Omega$, is a regular conditional probability measure. Since $\beta \to \infty$, by the Glivenko–Cantelli Theorem, $G_n(\beta) \to 1$ a.s. P^∞. Thus by Theorem 2.1, $P^\infty(\boldsymbol{\theta} \mid d(\hat{G}_n, G_n) \to 0 \text{ a.s. } \mathbf{F}\} = 1$ and it follows by Theorem 4.1 that $d(\hat{G}_n, G) \to 0$ a.s. $P^\infty(\mathbf{F})$.

COROLLARY 4.2. *If F_θ corresponds to the uniform distribution on the interval $[\theta, \theta+1)$, $\theta \in \Omega = (-\infty, +\infty)$ and P is a probability measure on Ω and G is the resulting distribution function and if $\hat{G}_n$ is defined by* (3.5) *and the conditions of Theorem* 3.1 *are satisfied, then $d(\hat{G}_n, G) \to 0$ a.s. $P^\infty(\mathbf{F})$.*

PROOF. Let B be a Borel set. $F_\theta(B) = \mu(B \cap [\theta, \theta + 1))$ which is a continuous function of θ and hence F_θ is a regular conditional probability. Applying Theorem 3.1 and Theorem 4.1 completes the proof.

Acknowledgments. This paper forms part of a Michigan State University doctoral thesis and I wish to thank Professor James Hannan for his guidance throughout the development of this work. Also my thanks go to the National Science Foundation whose grant financed this research and to the referee for his helpful comments.

REFERENCES

[1] DEELY, J. J. and KRUSE, R. L. (1968). Construction of sequences estimating the mixing distribution. *Ann. Math. Statist.* **39** 286–288.

[2] FOX, R. (1968). Contributions to compound decision theory and empirical Bayes squared error loss estimation. Research memorandum RM-214, Department of Statistics and Probability, Michigan State Univ.

[3] HANNAN, J. F. and ROBBINS, H. (1955). Asymptotic solutions of the compound decision problem for two completely specified distributions. *Ann. Math. Statist.*, **26** 37–51.

[4] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

[5] LOÈVE, MICHEL (1963). *Probability Theory* 3rd ed. Van Nostrand, Princeton.

[6] ROBBINS, H. (1951). Asymptotically subminimax solutions of compound statistical decision problems. *Proc. Second Berkeley Symp. Math. Statist. Prob.* 131–148. Univ. of California Press.

[7] ROBBINS, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.*, **35** 1–20.

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