SEQUENTIAL PROBABILITY FORECASTS AND THE PROBABILITY INTEGRAL TRANSFORM

Françoise Seillier
University of California, Berkeley

Technical Reports No. 150
April 1988

Department of Statistics
University of California
Berkeley, California

SEQUENTIAL PROBABILITY FORECASTS AND THE PROBABILITY INTEGRAL TRANSFORM

Françoise Seillier

University of California, Berkeley

ABSTRACT

The probability integral transform and its conditional version provide us with means of assessing the performance of statistical forecasting systems. The distributional characteristics of the residuals are considered under both the assumed sampling model and the probability model induced by the forecasting system. The CPIT can also serve as a tool to generate predictive distributions.

1. Introduction

The generic forecasting system, under our consideration, attaches a predictive distribution to each quantity in an ordered sequence of continuous random variables, A_1, A_2, ... It does so successively for each of them individually. And, at each stage, it makes use of the data up to that instance, i.e. the history of the process so far as well as any relevant covariate information. This we refer to as $\boldsymbol{B}_i$. It generates thus distributions of the form $P(a_{i+1} | \boldsymbol{B}_i)$.

Weather forecasting provides a typical instance of this situation: predictive distributions for future temperatures are elicited on the basis of records up to date as well as information required by the model embodying the forecaster's understanding of the underlying process. Another example: in the context of software reliability, one is interested in predictive distributions for successive interfailure times of a particular computer program. In both cases, when the next outcome is known, the prediction will be assessed. This new piece of information about the process (i.e. this outcome plus values of the covariables for the next predictand) will then be added to the database for producing subsequent ones.

Generally, such forecasting systems (SFS) consist of a parametric family of distributions, call it $\boldsymbol{P} = \{P_{\underline{\theta}} , \underline{\theta} \in \Theta\}$ (Θ denoting a finite dimensional parameter space), together with a prediction rule. The latter serves in disposing of the parameter, thus yielding at each stage a probability forecasting system (PFS). This elimination can be achieved by substituting an estimate based on $\boldsymbol{B}_i$ [i.e. $P_{\hat{\underline{\theta}}_i}(a_{i+1} | \boldsymbol{B}_i)$] or, in a Bayesian framework, by integrating it with respect to some prior distribution [i.e. $\int_\Theta P_{\underline{\theta}}(a_{i+1} | \boldsymbol{B}_i) \, \pi(\underline{\theta} | \boldsymbol{B}_i) \, d\underline{\theta}$], or also, when a fiducial argument is used, by pivotal inversion (Dawid, 1984).

If they are to be taken into account in the validation procedure, both the sequential and probabilistic elements of the forecasting process will greatly restrict us in the choice of suitable tools. Moreover, these should aim at judging the entire distributions rather than assessing some of their specific characteristics (as opposed to criteria like the mean squared error of prediction). It is also hoped that they will come close to depending solely on the observations at hand (thus excluding mathematical constructs part of the SFS). The search for an appropriate test, further complicated by the heavy dependence between forecast distributions, will lead to the probability integral transform (PIT).

Rosenblatt (1952) defines the latter as the transformation leading to residuals $\underline{u}^{(n)}=(u_1,\ldots,u_n)$:

$$u_1 \equiv \text{Prob}(A_1 \le a_1) = F_1(a_1)$$

$$u_2 \equiv \text{Prob}(A_2 \le a_2 \mid A_1 = a_1) = F_2(a_2 \mid a_1)$$

$$\vdots$$

$$u_n \equiv \text{Prob}(A_n \le a_n \mid A_1 = a_1, \ldots, A_{n-1} = a_{n-1}) = F_n(a_n \mid \underline{a}^{(n-1)})$$

where F_n denotes the joint distribution function of $\underline{A}^{(n)} = (A_1, A_2, \ldots, A_n)$.

This multivariate transformation and its conditional version (the CPIT, conditioning on sufficient statistics) indeed produce independent residuals expressed exclusively in terms of observables with a reference distribution unrelated to the realizations or the forecaster's assumed model (section 2). Furthermore, the CPIT provides us with yet another means of constructing forecasts (section 3). In the presence of group structure, this prediction rule is in fact identical to the Bayesian rule with a right-invariant prior distribution.

2. Some Diagnostics for Predictive Performance

Predictive performance can be viewed from two different perspectives. Firstly, in the spirit of the Bayesian philosophy, one can investigate how the residuals would behave under the overall probability model used by the forecaster. On the other hand, one would also like to know the distribution of these statistical constructs under the assumption that the data is actually generated from a member of the parametric family $\boldsymbol{P}$.

2.1 Assessing Predictive Performance from the Forecaster's Viewpoint

Let F_n represent the joint distribution (assumed absolutely continuous) of $\underline{A}^{(n)}$ under the fully-specified probability model arisen from the forecaster's SFS :

$$F_n(a_1, \ldots, a_n) = \int_{-\infty}^{a_1} \cdots \int_{-\infty}^{a_n} P_n(a_1, \ldots, a_n) \; da_1 \ldots da_n$$

where

$$\begin{aligned} P_n(a_1, a_2, \ldots, a_n) &= P_n(a_n \mid \underline{a}^{(n-1)}) \,.\, P_n(\underline{a}^{(n-1)}) \\ &= P_n(a_n \mid \underline{a}^{(n-1)}) \,.\, P_{n-1}(\underline{a}^{(n-1)}) \\ &= P_n(a_n \mid \underline{a}^{(n-1)}) \,.\, P_{n-1}(a_{n-1} \mid \underline{a}^{(n-2)}) \ \ldots \ P_2(a_2 \mid a_1) \,.\, P_1(a_1) \end{aligned}$$

Proposition 1 (Rosenblatt, 1952)

If $\underline{A}^{(n)}$ is indeed an F_n–distributed random vector, then $\underline{U}^{(n)} = (F_1(A_1), F_2(A_2 \mid A_1), \ldots, F_n(A_n \mid \underline{A}^{(n-1)}))$ are independently uniformly distributed on $[0,1]$.

With this result in hand, the data will be declared inconsistent with the PFS under study whenever relevant test statistics evidence departures from the hypotheses of independence and uniformity. One is indeed not only interested in detecting significant deviations from uniformity (which indicate that, on the whole, the shape of distributions included into $\boldsymbol{P}$ does not fit reality), but also in testing whether one has suitably captured in the postulated family of distributions, the correlation structure between successive outcomes. In the context of criteria for evaluating probabilistic forecasts, uniformity implies *calibration* (the lack of it being easily remedied by *recalibration* (Dawid, 1984)), whereas independence, involving a deeper understanding of the underlying mechanism and an ability to process past information, is related to *resolution* and *complete calibration* (Dawid, 1986).

The evaluation is thus carried out under the induced PFS, i.e. for instance, in the context of a Bayesian SFS, it is performed relatively to the model $\int_{\Theta} P_{\underline{\theta}}\,\pi(\underline{\theta})\,d\underline{\theta}$ and in the case of a *plug-in* (Dawid, 1984) under the PFS generatedby regarding $\hat{\underline{\theta}}_{n-1}$ as the true parameter at instance n . Under a particular distribution $P_{\underline{\theta}}$ of the sampling model, however, the residuals $U_1, U_2, \ldots$ will no longer exhibit uniformity nor independence, even if the A_i's are independent and identically distributed (David & Johnson, 1948). Indeed, their probability law then formally reflects the original density. Nonetheless, their marginal and joint distributions are independent of the unknown parameters whenever these specify scale and location.

Example 1: Suppose $\underline{A}^{(n)}$ has a multivariate normal distribution with mean $\underline{\mu}_n$ and variance-covariance matrix Σ_n , both $\underline{\mu}_n' = (\mu_1, \ldots, \mu_n)$ and $\Sigma_n = (\sigma_{ij})$ $(i,j=1,\ldots,n)$ being specified :

$$U_1 \equiv F_1(A_1) = \Phi\left[\frac{A_1 - \mu_1}{(\sigma_{11})^{1/2}}\right]$$

$$U_2 \equiv F_2(A_2 \mid a_1) = \Phi\left[\frac{A_2 - \mu_2 - \frac{\sigma_{21}}{\sigma_{11}}(a_1 - \mu_1)}{\left(\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}\right)^{1/2}}\right]$$

$$\vdots$$

$$U_n \equiv F_n(A_n \mid \underline{a}^{(n-1)}) = \Phi\left[\frac{A_n - \mu_n - (\underline{\sigma}_n^{n-1})' \, \Sigma_{n-1}^{-1} \, (\underline{a}^{(n-1)} - \underline{\mu}_{n-1})}{\left(\sigma_{nn} - (\underline{\sigma}_n^{n-1})' \, \Sigma_{n-1}^{-1} \, \underline{\sigma}_n^{n-1}\right)^{1/2}}\right]$$

where $(\underline{\sigma}_n^{n-1})' = (\sigma_{n1}\,\sigma_{n2} \cdots \sigma_{nn-1})$ and Φ denotes the standard normal distribution function.

Example 2: Consider the PFS constructed from the exponential model with parameter θ by substituting at each stage the maximum likelihood estimate of the parameter, i.e. $A_i \sim \boldsymbol{E}(\theta)$ independently for $i = 1, \ldots, n$, then $P_i(A_i = a_i \mid \underline{a}^{(i-1)}) = \hat{\theta}_{i-1} e^{-\hat{\theta}_{i-1} a_i}$ with $\hat{\theta}_{i-1} = \bar{a}_{i-1}^{-1}$. Then $e^{-\hat{\theta}_{i-1} A_i} \sim U[0,1]$ independent .

2.2 Evaluation with respect to Sampling Models

For the exponential family of distributions, by conditioning on sufficient statistics, the CPIT circumvents the problem described by David & Johnson (1948) and leads to results, similar to those enunciated in Proposition 1, under the sampling distribution for finite samples. Further, in the case of group-structural models, we can also achieve such exact assessments without having recourse to this conditioning.

A. Group-structural Models

Specify a generic structural model by regarding the parameter as an unknown transformation acting on some unobservable error variable i.e.

$$\underline{A} = (A_1, A_2, \ldots) = (\theta \circ E_1, \theta \circ E_2, \ldots) \equiv \theta \circ \underline{E}$$

where θ belongs to a well-defined group of transformations, G say, and $\{E_i\}$ have a known sampling distribution P_0 independent of $\underline{\theta}$, the element of the parameter space Θ characterizing the distribution of a single observable. By assumption, any g in G induces a transformation $\overline{g}$ on the parameter space Θ and the collection $\overline{G}$ of all such $\overline{g}$ not only has a group structure but is also both exactly transitive on Θ and isomorphic to G. Consequently, by fixing a reference element in Θ , any $\underline{\theta}$ is uniquely characterized by a single g in G (which we called θ in the previous model definition): let $\underline{e}$ denote this reference element, $\underline{\theta} = \overline{\theta} \circ \underline{e}$ for some $\overline{\theta}$ in $\overline{G}$, identified with some member of G, θ say. Under the further assumption that G is exact on the sample space, the isomorphism can be extended to the latter and $\underline{A}^{(n)}$ can be expressed as $\eta(\underline{A}^{(n)}) \circ \zeta(\underline{A}^{(n)})$ where $\eta(\underline{A}^{(n)})$ is an equivariant function and $\zeta(\underline{A}^{(n)})$ is the (reference) maximal invariant statistic which labels the orbit of $\underline{A}^{(n)}$ under G i.e. $\{g \circ \underline{a}^{(n)} : g \in G\}$. The model can now be reformulated as

$$\eta(\underline{A}^{(n)}) = \theta\, \eta(\underline{E}^{(n)})$$

$$\zeta(\underline{A}^{(n)}) = \zeta(\underline{E}^{(n)})$$

with $\theta, \eta(\underline{E}^{(n)}) \in G$. In the sequel, denote $\eta(\underline{A}^{(n)})$ by T_n , $\eta(\underline{E}^{(n)})$ by H_n .

The fiducial predictive distribution is then arrived at via the predictive pivot

$$V_{n+1} \equiv T_n^{-1} \circ A_{n+1} = H_n^{-1} \circ E_{n+1}$$

which is obtained by eliminating the parameter θ from

$$A_{n+1} = \theta \circ E_{n+1} \quad \text{and} \quad T_n = \theta\, H_n$$

After inverting the pivotal quantity, the future outcome

$$A_{n+1} = (t_n H_n^{-1}) \circ E_{n+1}$$

is assigned the sampling distribution of the variable on the right of the equality, conditional on t_n and $\zeta(\underline{a}^{(n)})$. This structural predictive distribution can be viewed as a posterior distribution based on the structural distribution of θ (i.e. that of $\theta \mid t_n, \zeta(\underline{a}^{(n)})$) as prior and the sampling distribution of A_{n+1} given t_n and $\zeta(\underline{a}^{(n)})$ (i.e. that of $A_{n+1} \mid t_n, \zeta(\underline{a}^{(n)}), \theta$). As $A_{n+1} = T_n \circ V_{n+1}$, this sampling distribution is dependent on the unknown value of H_n. Since, in this framework, the distribution of H_n conditional on $\zeta(\underline{a}^{(n)})$ is not adjusted after observing $(t_n, \zeta(\underline{a}^{(n)}))$, it is perfectly legitimate to transfer to V_{n+1} the sampling distribution of $H_n^{-1} \circ E_{n+1}$. Furthermore, the ancillary statistic can be discarded whenever T_n is adequate for A_{n+1} with respect to $\underline{A}^{(n)}$. Clearly, $(H_n, E_{n+1}) \perp\!\!\!\perp \zeta(\underline{A}^{(n)})$ implies

$$(1) \quad H_n \perp\!\!\!\perp \zeta(\underline{A}^{(n)}) \quad \text{i.e.} \quad T_n = \theta H_n \perp\!\!\!\perp \zeta(\underline{A}^{(n)}) \mid \theta$$

$$\text{and} \quad (2) \quad E_{n+1} \perp\!\!\!\perp \zeta(\underline{A}^{(n)}) \mid H_n \quad \text{i.e.} \quad A_{n+1} = \theta \circ E_{n+1} \perp\!\!\!\perp \zeta(\underline{A}^{(n)}) \mid \theta, T_n$$

Example 3: If ***P*** consists of a family of location models, the structural equation is given by

$$A_i = \theta + E_i \qquad i = 1, \ldots$$

where E_i denotes an unobervable random variable with a known sampling distribution (independent of the parameter θ), and thus defines a pivotal quantity. This model can also be reexpressed as

$$T_n = \theta + \eta(\underline{E}^{(n)})$$

$$\zeta(\underline{A}^{(n)}) = \zeta(\underline{E}^{(n)})$$

where the fixed joint distribution of $(\eta(\underline{E}^{(n)}), \zeta(\underline{E}^{(n)}))$ is independent of the parameter value. Also, this family of models allows

$$\zeta(A_1, \ldots, A_n) = (0, A_2 - A_1, \ldots, A_n - A_1) \quad , \quad \eta(A_1, \ldots, A_n) = A_1$$

$$\zeta'(A_1, \ldots, A_n) = (A_1 - \overline{A}_n, A_2 - \overline{A}_n, \ldots, A_n - \overline{A}_n) \quad , \quad \eta'(A_1, \ldots, A_n) = \overline{A}_n$$

from which it transpires that ζ is essentially unique while η is not (Dawid & Guttman, 1981). Let $f_{E_i}(e_i)$ denote the density of the unobservable E_i, then with respect to the original model, conditionally on θ

$$f_{A_i}(a_i \mid \theta) = f_{E_i}(a_i - \theta)$$

whereas the reduced model implies

$$f_{\zeta(\underline{A}^{(n)})}(\zeta(\underline{a}^{(n)}) \mid \theta) = f_{\zeta(\underline{E}^{(n)})}(\zeta(\underline{a}^{(n)}))$$

$$\text{and} \quad f_{T_n \mid \zeta(\underline{A}^{(n)})}(t_n \mid \zeta(\underline{a}^{(n)}), \theta) = f_{\eta(\underline{E}^{(n)}) \mid \zeta(\underline{E}^{(n)})}(t_n - \theta \mid \zeta(\underline{a}^{(n)}))$$

The predictive pivot is obtained by substituting for θ using

$$\overline{A}_n = \theta + H_n$$

in an expression involving the future event:

$$A_{n+1} - \overline{A}_n = E_{n+1} - H_n$$

$$\text{i.e.} \qquad A_{n+1} = \overline{A}_n + E_{n+1} - H_n$$

Crucially, the sampling distribution of the $E_{n+1} - H_n$ given the ancillary remains valid even when knowing $\overline{A}_n$.

Applying the PIT to the fiducial predictive distribution is in fact equivalent to considering the value of the fiducial distribution function. And, this leads to tests of consistency of the realizations with the forecaster's model under the sampling distibution.

Proposition 2

Letting $P_i(.)$ denote the structural predictive distribution conditional on the first i-1 observations, $\{P_i(A_i \le a_i), i=1,...,n\}$ are independent uniformly distributed random variables on $[0,1]$ under $P_{\underline{\theta}}$.

Proof : The structural residuals can indeed be equated to those based on the pivots, i.e.

$$P_i(A_i \le a_i) = P_i(V_i \le v_i)$$

provided V_i is, for each value of T_{i-1} , a monotonic increasing function of A_i (a similar argument would apply were V_i monotonic decreasing). As mentioned before, the structural distribution of V_i is obtained by conditioning on $\zeta(\underline{E}^{(i-1)}) = \zeta(\underline{a}^{(i-1)})$ under P_0 :

$$P_i(V_i \le v_i) = P_0(V_i \le v_i \mid \zeta(\underline{E}^{(i-1)}) = \zeta(\underline{a}^{(i-1)}))$$

which is independent in form and distribution of the parameter. These residuals will automatically be i.i.d. $U[0,1]$-distributed if the right-hand side is expressible as $P_0(V_i \le v_i \mid \underline{v}^{(i-1)})$, which is the case when $(\zeta(\underline{E}^{(i-1)}), V_i)$ is equivalent to $\zeta(\underline{E}^{(i)})$ (repeated application of this fact would thus lead to the desired result). One must therefore show that the former is also maximal invariant. V_i being a pivotal quantity and $\zeta(\underline{E}^{(i)})$ an ancillary statistic, it certainly is invariant . Now consider any invariant function $Z(\underline{E}^{(i)})$. Then

$$\begin{aligned} Z(\underline{E}^{(i)}) &= Z(H_{i-1}^{-1} \circ \underline{E}^{(i)}) \\ &= Z(H_{i-1}^{-1} \circ \underline{E}^{(i-1)}, H_{i-1}^{-1} \circ E_i) \\ &= Z(\zeta(\underline{E}^{(i-1)}), V_i) \end{aligned}$$

since $Z(H_{i-1}^{-1} \circ \underline{E}^{(i-1)})$ is invariant (and hence a function of $\zeta(\underline{E}^{(i-1)})$) .

In view of the equivalence relationship between the structural predictive distribution and that constructed via Bayes'theorem applied to a right-invariant prior (Hora & Buehler, 1967), the residuals derived from the latter will also be independent uniforms on $[0,1]$.

Example 1 (continued): Suppose now that $\underline{\mu}_n' = (\mu, \ldots, \mu)$, $\Sigma_n = (\sigma_{ij})$ with $\sigma_{ii} = \sigma$ and for $i \ne j$ $\sigma_{ij} = 0$ (i,j=1,...,n), both μ and σ being unknown. From a structural viewpoint

$$A_i = \mu + \sigma E_i \quad \text{with} \quad E_i \sim N(0,1)$$

First, eliminate μ between

$$(n-1)^{\frac{1}{2}} \sigma^{-1}(\overline{A}_{n-1} - \mu) \sim N(0,1) \quad \text{and} \quad \sigma^{-1}(A_n - \mu)$$

Since these are independent, $\sigma^{-1}\left(1-\frac{1}{n}\right)^{\frac{1}{2}}(A_n - \overline{A}_{n-1}) \sim N(0,1)$. Now, substitute for σ from

$$\sigma^{-2}\sum_{i=1}^{n-1}(A_i - \overline{A}_{n-1})^2 \sim \chi^2_{n-2}$$

The last two expressions being independent,

$$\frac{(n-2)(A_n - \overline{A}_{n-1})}{\left(\frac{n}{n-1}\sum_{i=1}^{n-1}(A_i - \overline{A}_{n-1})^2\right)^{\frac{1}{2}}} \sim t_{n-2}$$

The fiducial predictive distribution is thus

$$A_n \mid \underline{a}^{(n-1)} \sim \frac{1}{(n-2)^{\frac{1}{2}}}\left\{\overline{a}_{n-1} + \left(\frac{n}{n-1}\sum_{i=1}^{n-1}(a_i - \overline{a}_{n-1})^2\right)^{\frac{1}{2}} t_{n-2}\right\}$$

Example 2 (continued): Expressing the observables in terms of a structural model

$$A_i = \theta^{-1} \cdot E_i \qquad \text{where } E_i \sim \mathcal{E}(1) \ \text{ for } i = 1,\dots,n$$

the sufficient statistic becomes

$$T_n = \sum_{i=1}^{n} A_i = \theta^{-1}\sum_{i=1}^{n} E_i$$

Eliminating θ between

$$u_{n-1} = \theta \cdot T_{n-1} \quad \text{and} \quad p_n = \theta \cdot A_n$$

$$T_{n-1}^{-1} A_n + 1 = u_{n-1}^{-1}(u_{n-1} + p_n) \equiv Z$$

Since $u_{n-1} \sim \Gamma(n-1,1)$, $p_n \sim \Gamma(1,1)$ and $u_{n-1} \perp\!\!\!\perp p_n$, $Z^{-1} \sim \beta(n-1,1)$ and hence $T_{n-1}^{-1} A_n + 1 \sim Par(1,n-1)$. The fiducial predictive distribution is therefore shown to be

$$A_n \mid \underline{a}^{(n-1)} \sim \left(\sum_{i=1}^{n-1} a_i\right) Par(1,n-1) - \sum_{i=1}^{n-1} a_i$$

B. The Conditional Probability Integral Transform

For absolutely continuous distribution functions, by conditioning on sufficient statistics, it is possible to avoid the problem referred to earlier, due to the reliance on estimates for the parameters, and achieve exact independence (O'Reilly & Quesenberry, 1973). Assume that the sample $A_1, A_2, \dots, A_n$ is defined on the probability space $(\boldsymbol{R}^n, \boldsymbol{B}_n, P^n_{\underline{\theta}})$ and that $P_{\underline{\theta}}$ belongs to a class $\boldsymbol{P}$ of univariate distributions with absolutely continuous distribution functions. Let

$$\widetilde{F}_n(a_1,\dots,a_\alpha) \equiv P_\theta(A_1 \le a_1,\dots,A_\alpha \le a_\alpha \mid T_n) \qquad \alpha \in \{1,\dots,n\}$$

where $T_n : \boldsymbol{R}^n \to \boldsymbol{R}^{k_n}$ ($k_n \in \{1,\dots,n\}$) is a minimal sufficient statistic for the family $\boldsymbol{P}$. Assuming that $F_n(a)$ is almost surely absolutely continuous, given t_n , then $\widetilde{F}_n(A_i) \sim U[0,1]$ ($i \in \{1,\dots,\alpha\}$). As this distribution does not involve T_n , $\widetilde{F}_n(A_i)$ is thus independent of the sufficient statistic. Moreover, since

$$\widetilde{F}_n(a_j \mid a_1,\dots,a_{j-1}) = P_\theta(A_j \le a_j \mid T_n = t_n, A_1 = a_1,\dots, A_{j-1} = a_{j-1}) \quad \text{a.s.} ,$$

the next result follows immediately from Proposition 1 .

Corollary 3

Provided $\widetilde{F}_n(a_1,\dots,a_\alpha)$ is dominated by the α–dimensional Lebesgue measure,

$$\widetilde{F}_n(A_1), \widetilde{F}_n(A_2 \mid A_1), \dots, \widetilde{F}_n(A_\alpha \mid A_1,\dots,A_{\alpha-1})$$

are α independent uniformly distributed random variables on $[0,1]$ under P_θ .

The maximum value of α such that $\widetilde{F}_n(a_1,\dots,a_\alpha)$ is absolutely continuous, called the absolute continuity rank of $\boldsymbol{P}$ with respect to T_n is in general a function of n . For the exponential family of distributions, it is of the form $n - c$ where c is the number of components in the vector of minimally sufficient statistics, i.e. the number of constraints imposed on the sample space when one is conditioning on any particular value of T_n .

Clearly, in general, each variable in the above corollary involves $(A_1,\dots,A_n)$; consequently, any additional information would change the form of previous residuals. However, as apparent in the following proposition, expressions suitable for a truly sequential assessment can be achieved by considering the sequence backwards and imposing further restrictions on the underlying process. Assuringly, tackling the sequence in reverse order leads to similar results: for $j \in \{n-1,\dots,n-\alpha\}$ $\widetilde{F}_n(A_j \mid A_n,\dots,A_{j+1})$ defines a set of α i.i.d. $U[0,1]$ random variables.

Proposition 4 (O'Reilly & Quesenberry, 1973)

Provided

(1) double transitivity holds, i.e. for each $n \ge 1$, given A_{n+1} , there exists a one-to-one relationship between T_n and T_{n+1}

(2) for each $j \ge 1$, T_j is adequate for A_j with respect to $(A_{j+1},\dots,A_n)$, i.e. $A_j \perp\!\!\!\perp (A_{j+1},\dots,A_n) \mid T_j$,

$$\widetilde{F}_{n-\alpha+1}(A_{n-\alpha+1}), \dots, \widetilde{F}_n(A_n)$$

generate α i.i.d. $U[0,1]$ random variables.

Note that now each residual is expressed solely in terms of the data at hand at that particular time. Condition (1) entails that the information included in T_n and A_{n+1} is equivalent to that in T_{n+1} and A_{n+1} . On the other hand, (2) requires the future to depend on the present only through the sufficient statistic (Dawid, 1979). This is certainly the case when the observations are independent and, is true more generally whenever each distribution P_θ for $(A_j,\dots,A_n)$, in the class $\boldsymbol{P}$, has a density of the form $h(a_j)\, f_\theta(t_j, a_{j+1},\dots, a_n)$ (Takada, 1981) .

In an obvious way, all preceding statements generalize to multivariate classes of distributions. Letting $\underline{A}_1,\dots,\underline{A}_n$ denote k-dimensional vectors and α the absolute continuity rank of $\boldsymbol{P}$ with respect to $\underline{T}_n$, the vector of minimally sufficient statistics,

$$\widetilde{F}_n(\underline{A}_1),\ \widetilde{F}_n(\underline{A}_2 \mid \underline{A}_1),\ \dots,\ \widetilde{F}_n(\underline{A}_\alpha \mid \underline{A}_1,\dots,\underline{A}_{\alpha-1})$$

are i.i.d. U[0,1] . If $\underline{T}_n$ is doubly transitive and

$$\underline{A}_j \perp\!\!\!\perp (\underline{A}_{j+1},\dots,\underline{A}_n) \quad \mid \ \underline{T}_j \qquad i,j$$

not only are $\widetilde{F}_j(A_{ij})$ $(i=1,\dots,k$ and $j=n-\alpha+1,\dots,n)$ independently uniformly distributed but $\widetilde{F}_j(A_{ij} \mid A_{1j},\dots,A_{(i-1)j})$ also defines a set of $k\alpha$ i.i.d. U[0,1] random variables.

Example 1 (continued) : Let

$$P_i = \frac{(i-2)^{1/2}\,(A_i - \overline{A}_{i-1})}{\left(\frac{i}{i-1}\sum_{j=1}^{i-1}(A_j - \overline{A}_{i-1})^2\right)^{1/2}}$$

$$\begin{aligned}\widetilde{F}_i(a_i) &= P_{\mu,\sigma}\Big(P_i \le p_i \ \Big|\ \sum_{j=1}^{i} A_j,\ \sum_{j=1}^{i} A_j^2\Big)\\ &= P_{\mu,\sigma}(P_i \le p_i)\\ &= G_{i-2}(p_n)\end{aligned}$$

where G_k denotes the distribution function of the t-distribution with k degrees of freedom. Hence, since the minimally sufficient statistics $\sum_{j=1}^{i} A_j$ and $\sum_{j=1}^{i} A_j^2$ are doubly transitive, $\widetilde{F}_i(A_i) \sim U[0,1]$ independent for $i = 3,\dots,n$.

Example 2 (continued) : $T_n = \sum_{i=1}^{n} A_i$

$$\begin{aligned}\widetilde{F}_n(a_n) &= P_\theta(A_n T_n^{-1} \le a_n t_n^{-1} \mid T_n = t_n) = P_\theta(A_n T_n^{-1} \le a_n t_n^{-1})\\ &= \int_0^{a_n t_n^{-1}} \frac{\Gamma(n)}{\Gamma(1)\,\Gamma(n-1)}(1-u)^{n-2}\, du\\ &= 1 - (1 - a_n t_n^{-1})^{n-1}\end{aligned}$$

Consequently, $\left(\sum_{j=1}^{i-1} A_j\right)^{i-1} \left(\sum_{j=1}^{i} A_j\right)^{-i+1} \sim U[0,1]$ independent for $i = 2,\dots,n$.

It should be noted that, unless it has been accounted for in the formulation of the model, any systematic ordering of the observations would jeopardize the above distribution-theoretic results described above and the validity of any test performed on the residuals (these being based on the assumption that they are i.i.d. variables). Furthermore, this multivariate transformation lacks invariance under permutation of the observations: a different labelling would lead to other values for $\{\widetilde{F}_i : i=n-\alpha+1,\dots,n\}$. To remedy this defect, O'Reilly & Stephens (1982) consider the distribution of the ordered sample, given the minimal sufficient statistics. However, when the data are generated sequentially, this lack of invariance is somewhat of a side-issue: one is constrained to the order of realization.

The above theory vitally relies on the availability of minimal sufficient statistics of fixed size. For non-regular families, where the range features among the parameters, though such statistics exist, the conditional distributions are not absolutely continuous. For example, if $A_1,\dots,A_n \sim U[0,\theta]$, then T_n is given by $\max\{A_1,\dots,A_n\}$ and $\widetilde{F}_n(A_n)$ cannot be absolutely continuous as the corresponding density allocates a probability mass of $1/n$ to the event $A_n = T_n$, a null set under Lebesgue measure (double transitivity does not hold in this particular instance: T_n cannot be recovered from T_{n+1} and A_{n+1} when these are equal). One is therefore restricted to the Koopman-Darmois family of densities, the largest class, for which such statistics exist, leading to absolutely continuous distribution functions. At each stage, one is limited to consider at most α residuals. Hence, one would wish this number to be an increasing function of n , which is certainly the case for this family.

3. C.P.I.T.- Based Prediction Rule

As mentioned before, the predictive distributions $P(A_i \mid \underline{a}^{(i-1)})$ are constructed from a parametric family of joint distributions for $(A_1, A_2, \dots)$, $\boldsymbol{P}$, via a statistical procedure so that the predictions are expressed solely in terms of past data. The residuals emerging from the C. P. I. T. , when coupled to a fiducial argument, can also serve as a way of eliminating the parameter. This, however, may only be carried out in the somewhat restrictive case where the absolutely continuous distribution function belongs to a parametric family with doubly transitive and adequate minimal sufficient statistics. $\{\widetilde{F}_n(A_n)\}$ are indeed pivotal quantities: their distribution is independent of the parameters. From a fiducial view-point, $\widetilde{F}_n(A_n)$ is now regarded as a random variable in A_n with $\underline{a}^{(n-1)}$ fixed. The predictive density is then produced by differentiating with respect to A_n , i.e.

$$P_n(A_n = a_n \mid a_1, \dots, a_{n-1}) = \frac{\partial}{\partial a_n} \widetilde{F}_n(a_n) \Big|_{\underline{a}^{(n-1)}}$$

$$= \frac{\partial}{\partial a_n} \widetilde{F}_n(a_n) \Big|_{t_n} + \frac{\partial}{\partial t_n} \widetilde{F}_n(a_n) \Big|_{a_n} \frac{\partial t_n}{\partial a_n} \Big|_{\underline{a}^{(n-1)}}$$

Letting $\tilde{f}_n(a_n)$ denote the density corresponding to $\tilde{F}_n(a_n)$

$$\text{i.e.} \quad \tilde{F}_n(a_n) = \int_{-\infty}^{a_n} \tilde{f}_n(z_n)\, dz_n ,$$

the C.P.I.T. method thus generates a predictive density of the form

$$\tilde{f}_n(a_n) + \frac{\partial t_n}{\partial a_n}\Big|_{\underline{a}^{(n-1)}} \int_{-\infty}^{a_n} \left[\frac{\partial}{\partial t_n}\tilde{f}_n(z_n)\right]_{z_n} dz_n$$

For structural models within this class (i.e., for scalar observables, the gamma-scale and normal models), this in fact produces the same results as the usual fiducial distributions. Indeed, the predictive pivot V_{n+1} is assigned the known sampling distribution of $H_n^{-1} \circ E_{n+1}$ and is independent of ancillaries (in the dependent case, provided T_n is adequate for A_{n+1} with respect to $\underline{A}^{(n)}$, which is the very condition for applying the C.P.I.T. usefully in a sequential situation). Hence, a structural forecast distribution is transferred to A_{n+1} by regarding the sampling distribution of $H_n^{-1} \circ E_{n+1}$ still valid when assuming T_n to be known. On the other hand, the C.P.I.T. gives a distribution for A_{n+1} conditional on T_{n+1}. Now, provided the conditions of Proposition 4 hold, $T_n^{-1} \circ A_{n+1}$ is a function of A_{n+1} and T_{n+1}, which is clearly one-to-one in A_{n+1} for fixed T_{n+1}. So, equivalently, one can consider the distribution of V_{n+1} given T_{n+1}. Hence, the C.P.I.T. attaches to V_{n+1} its conditional distribution and, given T_n, induces a predictive distribution for A_{n+1}. Therefore, the two approaches will agree so long as V_{n+1} is independent of T_{n+1}, which is established in the following proposition.

Proposition 5 (Seillier, 1986)

$$V_{n+1} \perp\!\!\!\perp T_{n+1}$$

Proof : The family $\boldsymbol{P}=\{P_\theta : \theta \in G\}$ (where we tacitly identify θ with $\underline{\theta}$ and G with Θ) brought about by the structural model, is G-equivariant, G being a group : for $g \in G$, if A_i has a distribution in $\boldsymbol{P}$ so does $g \circ A_i$

$$\begin{aligned} g \circ A_i &= g \circ (\theta \circ E_i) \\ &= (g\theta) \circ E_i \end{aligned}$$

and $g\theta \in G$. Now, V_{n+1}, a function of $A_1, \ldots, A_{n+1}$, is invariant under G : for all $g \in G$

$$\begin{aligned} g \circ V_{n+1} &\equiv (gT_n)^{-1} \circ (g \circ A_{n+1}) \\ &= (g\theta H_n)^{-1} \circ ((g\theta) \circ E_{n+1}) \\ &= \left(H_n^{-1}(g\theta)^{-1}(g\theta)\right) \circ E_{n+1} \\ &= H_n^{-1} \circ E_{n+1} \\ &= V_{n+1} \end{aligned}$$

By virtue of the fact that $\boldsymbol{P}$ is G-equivariant, T_{n+1} , the minimal sufficient statistic based on $(A_1,...,A_{n+1})$, is found to be G-equivariant i.e. if $t_{n+1}(\underline{a}^{(n+1)}) = t_{n+1}(\underline{b}^{(n+1)})$ then, for any $g \in G$, $t_{n+1}(g \circ \underline{a}^{(n+1)}) = t_{n+1}(g \circ \underline{b}^{(n+1)})$ (using lemma 2.2 of Dawid (1982)). In fact, G acts transitively on T , i.e. for any two values of the sufficient statistic t_{n+1} and t'_{n+1}, one can find an element of G , h say, which transforms one into the other : let

$$t_{n+1} = \theta h_{n+1} \quad \text{and} \quad t'_{n+1} = \mu h_{n+1}$$

$$\text{then} \qquad \theta^{-1} t_{n+1} = \mu^{-1} t'_{n+1} \quad \text{i.e.} \quad t_{n+1} = (\theta\mu^{-1}) t'_{n+1}$$

G being a group, $\theta\mu^{-1} \in G$. Hence there is a unique orbit, implying that any G-invariant statistic must be constant. After considering theorem 3.4 in (Dawid, 1982), we can conclude that

$$V_{n+1} \perp\!\!\!\perp T_{n+1} \mid U$$

U being a maximal G-invariant function of T_{n+1} , and therefore

$$V_{n+1} \perp\!\!\!\perp T_{n+1}$$

Moreover, the fiducial predictive distribution is equivalent to the Bayes posterior distribution obtained from the G-invariant prior (Hora & Buehler, 1967), and hence so is that generated by the C.P.I.T. Implicit in structural prediction is the construction of a prior distribution for the parameters which we elicit not via some prior distributional information but by considering the group structure of the parameter space. By contrast, with the C.P.I.T.-based rule, we merely assume the existence of sufficient statistics (with the gamma-shape family as a typical example).

4. Discussion

In the context of sequential probability forecasts, by generating ancillaries for examining whether these were derived from a model consistent with the data, the P.I.T. and its conditional version provide a means of evaluating predictive ability which is independent in form and distribution of both the forecaster's model and that underlying the process. As a result, they can help not only in assessing models used for prediction purposes but also in selecting them *prequentially* (Dawid, 1984) i.e. with regards to their ability to produce sensible probabilistic forecasts and not to their goodness of fit to past realizations (as a safeguard against the danger of overparametrization, for instance).

Among their attractive features, they are naturally suited to sequential situations (i.e., at any particular time, they do not refer to yet unknown quantities) and yield independent residuals with the same distribution whatever the underlying F . These transforms also bypass any problem which arises from the lack of independence among the data and the forecasting distributions. Unlike most criteria commonly used to measure the suitability of posited models, they do not focus on a single characteristic of the predictive distribution (such as its mean, median or mode) and thus allow us to check the adequacy of the whole distribution. Moreover, by applying such tests to subsequences of the residuals, one could investigate further the nature of the assumed relationship between the events and correct it if need be.

It should be stressed that, whilst for situations where the C.P.I.T. is suitable, the residuals are independent uniform random variables under the sampling distribution, the test based on the P.I.T. is performed under the probability model induced by the statistical forecasting system. For structural models, however, its application to the fiducial SFS, produces the same result under the sampling distribution, given the ancillary statistics. The domain of validity of results under the sampling model therefore extends beyond the exponential family as T_n is no longer required to be sufficient, when one is conditioning on ancillaries. With respect to structural models within the range of application of the C.P.I.T. (e.g., in one-dimensional sample spaces, the normal and gamma-scale families), in view of the equivalence relationship between the predictive distributions constructed from the C.P.I.T., the fiducial argument and Bayes' theorem in association with a right-invariant prior, the residuals calculated by any one of these methods will also be regarded as i.i.d. uniform quantities by the others as well as under the sampling distribution. While we established this equivalence by showing that the C.P.I.T. when coupled to a fiducial argument yields the same transformed variable as the usual pivotal argument and using a well-known identity (Hora & Buehler, 1967), O'Reilly and Villegas (1987) proved directly that the C.P.I.T. and its bayesian counterpart, the predictive probability integral transform, lead to the same residuals.

Summarizing, the PIT provides us with a means of constructing tests of predictive performance under the assumption that probability model induced by the SFS is consistent with the data. Further, the transformed variables will also be independent uniforms under the sampling model in the presence of appropriate group structure (conditionally on ancillary statistics). This will also be the case asymptotically in the more general situation where *prequential* consistency (Dawid, 1984) holds. For the natural exponential family, one can also obtain residuals with the same properties under the sampling model , when conditioning on sufficient statistics. On the other hand, as a tool to generate forecasts, the C.P.I.T. gives rise for structural models to the same predictive density, and hence the same tests as the usual fiducial method and therefore the Bayesian rule with a right-invariant prior.

5. References

David, F.N. & Johnson, N.L. (1948). The probability integral transform when the parameters are estimated from the sample. *Biometrika*, **35**, 182-190.

Dawid, A.P. (1979). Conditional independence in statistical theory (with Discussion). *Journal of the Royal Statistical Society*, B, **41**, 1-31.

Dawid, A. P. (1982). Invariance and independence in multivariate distribution theory. Research report no.9, University College London.

Dawid, A. P. (1984). Statistical theory : The prequential approach (with Discussion). *Journal of the Royal Statistial Society*, A, **147**, 278-292.

Dawid, A.P. (1986). Probability forecasting. In: *Encyclopedia of Statistical Sciences*, Kotz, S., Johnson, N.L. & Read, C.B., editors, volume 7. Wiley Intersciences, New York.

Dawid, A.P. & Guttman, I. (1981). Conjugate Bayesian inference for structural models. *Communications in Satistics - Theory and Methods*, A, **10**(8), 739-748.

Hora, R.B. & Buehler, R.J. (1967). Fiducial theory and invariant prediction. *The Annals of Mathematical Statistics*, **38**, 795-801.

Littlewood, B., Abdel Ghaly, A.A. & Chan, P.Y. (1985). Evaluation of competing software reliability predictions. CSR Research Report, Centre for Software Reliability, The City University, London.

O'Reilly, F.J. & Quesenberry, C.P. (1973). The conditional probability integral transform and its applications to obtain composite chi-square goodness-of-fit tests. *The* Annals of Statistics, **1**, 74-83.

O'Reilly, F.J. & Stephens, M.A. (1982). Characterizations and goodness-of-fit tests. *Journal of the Royal Statistical Society*, B, **44**, 353-360.

O'Reilly, F.J. & Villegas, C. (1987). Conditional and predictive probability integral transforms. *The Canadian Journal of Statistics*, **15**, 151-158.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, **23**, 470-472.

Seillier, F. (1986). Assessment of sequential probabilistic forecasting procedures. Ph. D. Thesis, University of London, England.

Takada, Y. (1981). Invariant prediction rules and an adequate statistics. *The Annals of the Institute of Statistical Mathematics*, **33**, A, 91-100.

TECHNICAL REPORTS

Statistics Department

University of California, Berkeley

1. BREIMAN, L. and FREEDMAN, D. (Nov. 1981, revised Feb. 1982). How many variables should be entered in a regression equation? Jour. Amer. Statist. Assoc., March 1983, 78, No. 381, 131-136.

2. BRILLINGER, D. R. (Jan. 1982). Some contrasting examples of the time and frequency domain approaches to time series analysis. Time Series Methods in Hydrosciences, (A. H. El-Shaarawi and S. R. Esterby, eds.) Elsevier Scientific Publishing Co., Amsterdam, 1982, pp. 1-15.

3. DOKSUM, K. A. (Jan. 1982). On the performance of estimates in proportional hazard and log-linear models. Survival Analysis, (John Crowley and Richard A. Johnson, eds.) IMS Lecture Notes - Monograph Series, (Shanti S. Gupta, series ed.) 1982, 74-84.

4. BICKEL, P. J. and BREIMAN, L. (Feb. 1982). Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. Ann. Prob., Feb. 1982, 11. No. 1, 185-214.

5. BRILLINGER, D. R. and TUKEY, J. W. (March 1982). Spectrum estimation and system identification relying on a Fourier transform. The Collected Works of J. W. Tukey, vol. 2, Wadsworth, 1985, 1001-1141.

6. BERAN, R. (May 1982). Jackknife approximation to bootstrap estimates. Ann. Statist., March 1984, 12 No. 1, 101-118.

7. BICKEL, P. J. and FREEDMAN, D. A. (June 1982). Bootstrapping regression models with many parameters. Lehmann Festschrift, (P. J. Bickel, K. Doksum and J. L. Hodges, Jr., eds.) Wadsworth Press, Belmont, 1983, 28-48.

8. BICKEL, P. J. and COLLINS, J. (March 1982). Minimizing Fisher information over mixtures of distributions. Sankhyā, 1983, 45, Series A, Pt. 1, 1-19.

9. BREIMAN, L. and FRIEDMAN, J. (July 1982). Estimating optimal transformations for multiple regression and correlation.

10. FREEDMAN, D. A. and PETERS, S. (July 1982, revised Aug. 1983). Bootstrapping a regression equation: some empirical results. JASA, 1984, 79, 97-106.

11. EATON, M. L. and FREEDMAN, D. A. (Sept. 1982). A remark on adjusting for covariates in multiple regression.

12. BICKEL, P. J. (April 1982). Minimax estimation of the mean of a mean of a normal distribution subject to doing well at a point. Recent Advances in Statistics, Academic Press, 1983.

14. FREEDMAN, D. A., ROTHENBERG, T. and SUTCH, R. (Oct. 1982). A review of a residential energy end use model.

15. BRILLINGER, D. and PREISLER, H. (Nov. 1982). Maximum likelihood estimation in a latent variable problem. Studies in Econometrics, Time Series, and Multivariate Statistics, (eds. S. Karlin, T. Amemiya, L. A. Goodman). Academic Press, New York, 1983, pp. 31-65.

16. BICKEL, P. J. (Nov. 1982). Robust regression based on infinitesimal neighborhoods. Ann. Statist., Dec. 1984, 12, 1349-1368.

17. DRAPER, D. C. (Feb. 1983). Rank-based robust analysis of linear models. I. Exposition and review.

18. DRAPER, D. C. (Feb 1983). Rank-based robust inference in regression models with several observations per cell.

19. FREEDMAN, D. A. and FIENBERG, S. (Feb. 1983, revised April 1983). Statistics and the scientific method, Comments on and reactions to Freedman, A rejoinder to Fienberg's comments. Springer New York 1985 Cohort Analysis in Social Research, (W. M. Mason and S. E. Fienberg, eds.).

20. FREEDMAN, D. A. and PETERS, S. C. (March 1983, revised Jan. 1984). Using the bootstrap to evaluate forecasting equations. J. of Forecasting, 1985, Vol. 4, 251-262.

21. FREEDMAN, D. A. and PETERS, S. C. (March 1983, revised Aug. 1983). Bootstrapping an econometric model: some empirical results. JBES, 1985, 2, 150-158.

22. FREEDMAN, D. A. (March 1983). Structural-equation models: a case study.

23. DAGGETT, R. S. and FREEDMAN, D. (April 1983, revised Sept. 1983). Econometrics and the law: a case study in the proof of antitrust damages. Proc. of the Berkeley Conference, in honor of Jerzy Neyman and Jack Kiefer. Vol I pp. 123-172. (L. Le Cam, R. Olshen eds.) Wadsworth, 1985.

24. DOKSUM, K. and YANDELL, B. (April 1983). Tests for exponentiality. Handbook of Statistics. (P. R. Krishnaiah and P. K. Sen, eds.) 4, 1984.

25. FREEDMAN, D. A. (May 1983). Comments on a paper by Markus.

26. FREEDMAN, D. (Oct. 1983, revised March 1984). On bootstrapping two-stage least-squares estimates in stationary linear models. Ann. Statist., 1984, 12, 827-842.

27. DOKSUM, K. A. (Dec. 1983). An extension of partial likelihood methods for proportional hazard models to general transformation models. Ann. Statist., 1987, 15, 325-345.

28. BICKEL, P. J., GOETZE, F. and VAN ZWET, W. R. (Jan. 1984). A simple analysis of third order efficiency of estimate Proc. of the Neyman-Kiefer Conference. (L. Le Cam, ed.) Wadsworth, 1985.

29. BICKEL, P. J. and FREEDMAN, D. A. Asymptotic normality and the bootstrap in stratified sampling. Ann. Statist. 12 470-482.

30. FREEDMAN, D. A. (Jan. 1984). The mean vs. the median: a case study in 4-R Act litigation. JBES. 1985 Vol 3 pp. 1-13.

31. STONE, C. J. (Feb. 1984). An asymptotically optimal window selection rule for kernel density estimates. Ann. Statist., Dec. 1984, 12, 1285-1297.

32. BREIMAN, L. (May 1984). Nail finders, edifices, and Oz.

33. STONE, C. J. (Oct. 1984). Additive regression and other nonparametric models. Ann. Statist., 1985, 13, 689-705.

34. STONE, C. J. (June 1984). An asymptotically optimal histogram selection rule. Proc. of the Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer (L. Le Cam and R. A. Olshen, eds.), II, 513-520.

35. FREEDMAN, D. A. and NAVIDI, W. C. (Sept. 1984, revised Jan. 1985). Regression models for adjusting the 1980 Census. Statistical Science. Feb 1986, Vol. 1, No. 1, 3-39.

36. FREEDMAN, D. A. (Sept. 1984, revised Nov. 1984). De Finetti's theorem in continuous time.

37. DIACONIS, P. and FREEDMAN, D. (Oct. 1984). An elementary proof of Stirling's formula. Amer. Math Monthly. Feb 1986, Vol. 93, No. 2, 123-125.

38. LE CAM, L. (Nov. 1984). Sur l'approximation de familles de mesures par des familles Gaussiennes. Ann. Inst. Henri Poincaré. 1985, 21, 225-287.

39. DIACONIS, P. and FREEDMAN, D. A. (Nov. 1984). A note on weak star uniformities.

40. BREIMAN, L. and IHAKA, R. (Dec. 1984). Nonlinear discriminant analysis via SCALING and ACE.

41. STONE, C. J. (Jan. 1985). The dimensionality reduction principle for generalized additive models.

42. LE CAM, L. (Jan. 1985). On the normal approximation for sums of independent variables.

43. BICKEL, P. J. and YAHAV, J. A. (1985). On estimating the number of unseen species: how many executions were there?

44. BRILLINGER, D. R. (1985). The natural variability of vital rates and associated statistics. Biometrics, to appear.

45. BRILLINGER, D. R. (1985). Fourier inference: some methods for the analysis of array and nonGaussian series data. Water Resources Bulletin, 1985, 21, 743-756.

46. BREIMAN, L. and STONE, C. J. (1985). Broad spectrum estimates and confidence intervals for tail quantiles.

47. DABROWSKA, D. M. and DOKSUM, K. A. (1985, revised March 1987). Partial likelihood in transformation models with censored data.

48. HAYCOCK, K. A. and BRILLINGER, D. R. (November 1985). LIBDRB: A subroutine library for elementary time series analysis.

49. BRILLINGER, D. R. (October 1985). Fitting cosines: some procedures and some physical examples. Joshi Festschrift, 1986. D. Reidel.

50. BRILLINGER, D. R. (November 1985). What do seismology and neurophysiology have in common? - Statistics! Comptes Rendus Math. Rep. Acad. Sci. Canada. January, 1986.

51. COX, D. D. and O'SULLIVAN, F. (October 1985). Analysis of penalized likelihood-type estimators with application to generalized smoothing in Sobolev Spaces.

52. O'SULLIVAN, F. (November 1985). A practical perspective on ill-posed inverse problems: A review with some new developments. To appear in Journal of Statistical Science.

53. LE CAM, L. and YANG, G. L. (November 1985, revised March 1987). On the preservation of local asymptotic normality under information loss.

54. BLACKWELL, D. (November 1985). Approximate normality of large products.

55. FREEDMAN, D. A. (June 1987). As others see us: A case study in path analysis. Journal of Educational Statistics. 12, 101-128.

56. LE CAM, L. and YANG, G. L. (January 1986). Distinguished Statistics, Loss of information and a theorem of Robert B. Davies.

57. LE CAM, L. (February 1986). On the Bernstein - von Mises theorem.

58. O'SULLIVAN, F. (January 1986). Estimation of Densities and Hazards by the Method of Penalized likelihood.

59. ALDOUS, D. and DIACONIS, P. (February 1986). Strong Uniform Times and Finite Random Walks.

60. ALDOUS, D. (March 1986). On the Markov Chain simulation Method for Uniform Combinatorial Distributions and Simulated Annealing.

61. CHENG, C-S. (April 1986). An Optimization Problem with Applications to Optimal Design Theory.

62. CHENG, C-S., MAJUMDAR, D., STUFKEN, J. & TURE, T. E. (May 1986, revised Jan 1987). Optimal step type design for comparing test treatments with a control.

63. CHENG, C-S. (May 1986, revised Jan. 1987). An Application of the Kiefer-Wolfowitz Equivalence Theorem.

64. O'SULLIVAN, F. (May 1986). Nonparametric Estimation in the Cox Proportional Hazards Model.

65. ALDOUS, D. (JUNE 1986). Finite-Time Implications of Relaxation Times for Stochastically Monotone Processes.

66. PITMAN, J. (JULY 1986, revised November 1986). Stationary Excursions.

67. DABROWSKA, D. and DOKSUM, K. (July 1986, revised November 1986). Estimates and confidence intervals for median and mean life in the proportional hazard model with censored data.

68. LE CAM, L. and YANG, G.L. (July 1986). Distinguished Statistics, Loss of information and a theorem of Robert B. Davies (Fourth edition).

69. STONE, C.J. (July 1986). Asymptotic properties of logspline density estimation.

71. BICKEL, P.J. and YAHAV, J.A. (July 1986). Richardson Extrapolation and the Bootstrap.

72. LEHMANN, E.L. (July 1986). Statistics - an overview.

73. STONE, C.J. (August 1986). A nonparametric framework for statistical modelling.

74. BIANE, PH. and YOR, M. (August 1986). A relation between Lévy's stochastic area formula, Legendre polynomial, and some continued fractions of Gauss.

75. LEHMANN, E.L. (August 1986, revised July 1987). Comparing Location Experiments.

76. O'SULLIVAN, F. (September 1986). Relative risk estimation.

77. O'SULLIVAN, F. (September 1986). Deconvolution of episodic hormone data.

78. PITMAN, J. & YOR, M. (September 1987). Further asymptotic laws of planar Brownian motion.

79. FREEDMAN, D.A. & ZEISEL, H. (November 1986). From mouse to man: The quantitative assessment of cancer risks. To appear in Statistical Science.

80. BRILLINGER, D.R. (October 1986). Maximum likelihood analysis of spike trains of interacting nerve cells.

81. DABROWSKA, D.M. (November 1986). Nonparametric regression with censored survival time data.

82. DOKSUM, K.J. and LO, A.Y. (November 1986). Consistent and robust Bayes Procedures for Location based on Partial Information.

83. DABROWSKA, D.M., DOKSUM, K.A. and MIURA, R. (November 1986). Rank estimates in a class of semiparametric two-sample models.

84. BRILLINGER, D. (December 1986). Some statistical methods for random process data from seismology and neurophysiology.

85. DIACONIS, P. and FREEDMAN, D. (December 1986). A dozen de Finetti-style results in search of a theory.

86. DABROWSKA, D.M. (January 1987). Uniform consistency of nearest neighbour and kernel conditional Kaplan - Meier estimates.

87. FREEDMAN, D.A., NAVIDI, W. and PETERS, S.C. (February 1987). On the impact of variable selection in fitting regression equations.

88. ALDOUS, D. (February 1987, revised April 1987). Hashing with linear probing, under non-uniform probabilities.

89. DABROWSKA, D.M. and DOKSUM, K.A. (March 1987, revised January 1988). Estimating and testing in a two sample generalized odds rate model.

90. DABROWSKA, D.M. (March 1987). Rank tests for matched pair experiments with censored data.

91. DIACONIS, P and FREEDMAN, D.A. (April 1988). Conditional limit theorems for exponential families and finite versions of de Finetti's theorem.

92. DABROWSKA, D.M. (April 1987, revised September 1987). Kaplan-Meier estimate on the plane.

92a. ALDOUS, D. (April 1987). The Harmonic mean formula for probabilities of Unions: Applications to sparse random graphs.

93. DABROWSKA, D.M. (June 1987, revised Feb 1988). Nonparametric quantile regression with censored data.

94. DONOHO, D.L. & STARK, P.B. (June 1987). Uncertainty principles and signal recovery.

95. RIZZARDI, F. (Aug 1987). Two-Sample t-tests where one population SD is known.

96. BRILLINGER, D.R. (June 1987). Some examples of the statistical analysis of seismological data. To appear in *Proceedings, Centennial Anniversary Symposium, Seismographic Stations, University of California, Berkeley.*

97. FREEDMAN, D.A. and NAVIDI, W. (June 1987). On the multi-stage model for carcinogenesis.

98. O'SULLIVAN, F. and WONG, T. (June 1987). Determining a function diffusion coefficient in the heat equation.

99. O'SULLIVAN, F. (June 1987). Constrained non-linear regularization with application to some system identification problems.

100. LE CAM, L. (July 1987, revised Nov 1987). On the standard asymptotic confidence ellipsoids of Wald.

101. DONOHO, D.L. and LIU, R.C. (July 1987). Pathologies of some minimum distance estimators.

102. BRILLINGER, D.R., DOWNING, K.H. and GLAESER, R.M. (July 1987). Some statistical aspects of low-dose electron imaging of crystals.

103. LE CAM, L. (August 1987). Harald Cramér and sums of independent random variables.

104. DONOHO, A.W., DONOHO, D.L. and GASKO, M. (August 1987). Macspin: Dynamic graphics on a desktop computer.

105. DONOHO, D.L. and LIU, R.C. (August 1987). On minimax estimation of linear functionals.

106. DABROWSKA, D.M. (August 1987). Kaplan-Meier estimate on the plane: weak convergence, LIL and the bootstrap.

107. CHENG, C-S. (August 1987). Some orthogonal main-effect plans for asymmetrical factorials.

108. CHENG, C-S. and JACROUX, M. (August 1987). On the construction of trend-free run orders of two-level factorial designs.

109. KLASS, M.J. (August 1987). Maximizing $E \max_{1 \le k \le n} S_k^+ / ES_n^+$: A prophet inequality for sums of I.I.D. mean zero variates.

110. DONOHO, D.L. and LIU, R.C. (August 1987). The "automatic" robustness of minimum distance functionals.

111. BICKEL, P.J. and GHOSH, J.K. (August 1987). A decomposition for the likelihood ratio statistic and the Bartlett correction — a Bayesian argument.

112. BURDZY, K., PITMAN, J.W. and YOR, M. (September 1987). Some asymptotic laws for crossings and excursions.

113. ADHIKARI, A. and PITMAN, J. (September 1987). The shortest planar arc of width 1.

114. RITOV, Y. (September 1987). Estimation in a linear regression model with censored data.

115. BICKEL, P.J. and RITOV, Y. (September 1987). Large sample theory of estimation in biased sampling regression models I.

116. RITOV, Y. and BICKEL, P.J. (September 1987). Unachievable information bounds in non and semiparametric models.

117. RITOV, Y. (October 1987). On the convergence of a maximal correlation algorithm with alternating projections.

118. ALDOUS, D.J. (October 1987). Meeting times for independent Markov chains.

119. HESSE, C.H. (October 1987). An asymptotic expansion for the mean of the passage-time distribution of integrated Brownian Motion.

120. DONOHO, D. and LIU, R. (October 1987, revised March 1988). Geometrizing rates of convergence, II.

121. BRILLINGER, D.R. (October 1987). Estimating the chances of large earthquakes by radiocarbon dating and statistical modelling. To appear in *Statistics a Guide to the Unknown.*

122. ALDOUS, D., FLANNERY, B. and PALACIOS, J.L. (November 1987). Two applications of urn processes: The fringe analysis of search trees and the simulation of quasi-stationary distributions of Markov chains.

123. DONOHO, D.L. and MACGIBBON, B. (November 1987). Minimax risk for hyperrectangles.

124. ALDOUS, D. (November 1987). Stopping times and tightness II.

125. HESSE, C.H. (November 1987). The present state of a stochastic model for sedimentation.

126. DALANG, R.C. (December 1987). Optimal stopping of two-parameter processes on hyperfinite probability spaces.

127. Same as No. 133.

128. DONOHO, D. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean II.

129. SMITH, D.L. (December 1987). Exponential bounds in Vapnik-Červonenkis classes of index 1.

130. STONE, C.J. (November 1987). Uniform error bounds involving logspline models.

131. Same as No. 140

132. HESSE, C.H. (December 1987). A Bahadur - Type representation for empirical quantiles of a large class of stationary, possibly infinite - variance, linear processes

133. DONOHO, D.L. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean, I.

134. DUBINS, L.E. and SCHWARZ, G. (December 1987). A sharp inequality for martingales and stopping-times.

135. FREEDMAN, D.A. and NAVIDI, W. (December 1987). On the risk of lung cancer for ex-smokers.

136. LE CAM, L. (January 1988). On some stochastic models of the effects of radiation on cell survival.

137. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On the uniform consistency of Bayes estimates for multinomial probabilities.

138. DONOHO, D.L. and LIU, R.C. (January 1988). Geometrizing rates of convergence, III.

139. BERAN, R. (January 1988). Refining simultaneous confidence sets.

140. HESSE, C.H. (December 1987). Numerical and statistical aspects of neural networks.

141. BRILLINGER, D.R. (January 1988). Two reports on trend analysis: a) An Elementary Trend Analysis of Rio Negro Levels at Manaus, 1903-1985 b) Consistent Detection of a Monotonic Trend Superposed on a Stationary Time Series

142. DONOHO, D.L. (Jan. 1985, revised Jan. 1988). One-sided inference about functionals of a density.

143. DALANG, R.C. (February 1988). Randomization in the two-armed bandit problem.

144. DABROWSKA, D.M. and DOKSUM, K.A. (February 1988). Graphical comparisons of cumulative hazards for two populations.

145. ALDOUS, D.J. (February 1988). Lower bounds for covering times for reversible Markov Chains and random walks on graphs.

146. BICKEL, P.J. and RITOV, Y. (February 1988). Estimating integrated squared density derivatives.

147. STARK, P.B. (March 1988). Strict bounds and applications.

148. DONOHO, D.L. and STARK, P.B. (March 1988). Rearrangements and smoothing.

149. NOLAN, D. (March 1988). Asymptotics for a multivariate location estimator.

150. SEILLIER, F. (March 1988). Sequential probability forecasts and the probability integral transform.

151. NOLAN, D. (March 1988). Limit theorems for a random convex set.

152. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On a theorem of Kuchler and Lauritzen.

153. SEILLIER, F. (April 13). Sequential probability forecasts and the probability integral transform.

Copies of these Reports plus the most recent additions to the Technical Report series are available from the Statistics Department technical typist in room 379 Evans Hall or may be requested by mail from:

Department of Statistics
University of California
Berkeley, California 94720

Cost: $1 per copy.