Proceedings of the AMS Centennial Symposium
August 8–12, 1988

Sufficiency as Statistical Symmetry

PERSI DIACONIS

Abstract. Sufficiency is a theoretical tool that has grown up in mathematical statistics. It may be described crudely as the theory of how much data can be thrown away. This paper reviews the basic achievements of the theory in statistical problems and sketches applications in other areas of mathematics. It is shown how the idea gives a suitable framework for exchangeability (an important piece of the Bayesian theory of statistics) and Gibbs states (the rigorous theory of phase transitions in statistical mechanics). In these last settings, sufficiency may be seen as a sweeping generalization of group invariance.

1. Introduction to sufficiency. One of the basic problems of statistics is this: one begins with a space $\mathscr{X}$ and a family of probability measures $\mathscr{P}$ on $\mathscr{X}$. It is assumed that an observation $x \in \mathscr{X}$ is drawn from a fixed, unknown $P \in \mathscr{P}$. We are shown x and required to guess P. For example, the usual formulation for n flips of a coin takes x as the space of binary n-tuples. For each $\theta \in [0, 1]$, a probability P_θ is defined on x by $P_\theta(x) = \theta^t(1-\theta)^{n-t}$ where $t = t(x) = x_1 + \cdots + x_n$. The family $\mathscr{P}$ is taken as $\{P_\theta\}_{\theta \in [0,1]}$. We are shown x and required to guess θ.

In the example, the observation consists of the binary n-tuple x. It is natural to ask if all of this is required or if x can be compressed to $t = x_1 + \cdots + x_n$ without essential loss. This is the subject matter of sufficiency.

In the general set-up a function $T : \mathscr{X} \to \mathscr{Y}$ is called *sufficient for the family* $\mathscr{P}$ if the conditional probability

$$P(x|T(x) = t) \tag{1.1}$$

is the same for each $P \in \mathscr{P}$. In (1.1) the definition of conditional probability is the natural extension of the elementary notion $P(A|B) = P(A \cap B)/P(B)$. Thus, $P(x|T(x) = t)$ is defined as zero unless $T(x) = t$. It is taken as proportional to $P(x)$ if $T(x) = t$ with normalizing constant making it a probability distribution.

1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 60J05.

0-8218-0167-8 \$1.00 + \$.25 per page

This leaves aside technical fine points which can be found in any standard graduate text in probability (e.g., Billingsley [5]).

EXAMPLE 1: COIN TOSSING. For coin tossing, the sum $T(x) = x_1 + \cdots + x_n$ is a sufficient statistic. Indeed,

$$P_\theta\{x|T(x) = t\} = \frac{P_\theta\{x \text{ and } T(x) = t\}}{P_\theta\{T(x) = t\}} = \frac{\theta^t(1-\theta)^{n-t}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} = \frac{1}{\binom{n}{t}}.$$

The right side does not depend on θ. This can also be seen from the following symmetry argument: $P_\theta(x|T(x) = t)$ is the chance of observing the sequence $x = (x_1 \cdots x_n)$ given $T(x) = t$. Imagine someone flipping a weird, biased coin. They announce that there have been two heads out of the first ten tosses. Whatever the bias, those two heads are equally likely to have appeared in any of $\binom{10}{2}$ possible places.

Here is a different interpretation of sufficiency for coin tossing as a fact about symmetric functions. Let $e_i(x_1, x_2, \ldots, x_n)$ be the ith elementary symmetric function in variables $x_1, x_2, \ldots, x_n$. Thus $e_1 = \sum x_i$, $e_2 = \sum_{i<j} x_i x_j$, etc. The generating function for e_i is

$$\sum_{i=0}^{n} e_i t^i = \prod_{i=1}^{n} (1 + x_i t).$$

The factorization of this generating function is equivalent to the sum being sufficient for coin tossing. To see this, divide both sides of the identity above by $(1+\theta)^n$, and multiply and divide e_i by $\binom{n}{i}$:

$$\sum_{i=0}^{n} \frac{1}{\binom{n}{i}} e_i \binom{n}{i} \frac{\theta^i}{(1+\theta)^n} = \prod_{i=1}^{n} \frac{(1+x_i\theta)}{(1+\theta)}.$$

On the right is the generating function for n flips of a coin with probability of heads $\theta/(1+\theta)$. On the left, $\binom{n}{i}\theta^i/(1+\theta)^n$ is the chance that n flips of such a coin lead to i heads. The term $e_i/\binom{n}{i}$ is the generating function for n flips given that i of them are heads. In the language of random variables the identity appears

$$E_t \prod x_j^{X_j} = EE\left(\prod x_j^{X_j} | \sum x_j = t\right).$$

The inner expectation is free of θ because $\sum x_j$ is sufficient for θ.

Many of the identities of symmetric function theory can be put into similar language. There is much of interest to do in fitting Schur functions into this picture. See, e.g., Macdonald [39].

Often, sufficiency is clear via symmetry. The point is that the notion is useful without an underlying group. As an example, consider n binary outcomes in which the chance of 1 increases over time. If the chance of a 1 in place i is taken as $e^{\eta i}/(1 + e^{\eta i})$ with $\eta \in [0, \infty)$, this gives a family of probabilities $\mathscr{P} = \{p_\eta\}_{\eta \in [0,\infty)}$ on binary n-tuples. The statistic $T(x) = \sum_{i=1}^{n} i x_i$ is easily seen to be sufficient for $\mathscr{P}$.

The next example shows sufficiency in a continuous setting.

EXAMPLE 2. Take $X = \mathbb{R}^n$ and $\mathscr{P}$ the family of all probability measures on $\mathbb{R}^n$ invariant under the orthogonal group O_n. Thus $P \in \mathscr{P}$ satisfies

$$P(A) = P(\Gamma A)$$

for every Borel set A and orthogonal matrix Γ.

The sum of squares $T(x) = x_1^2 + \cdots + x_n^2$ is sufficient for $\mathscr{P}$. Indeed $P\{x|T(x) = t\}$ is uniform on the sphere of radius $\sqrt{t}$ for every $P \in \mathscr{P}$. This example will reappear several times in later sections. The final example shows sufficiency in a less standard setting.

EXAMPLE 3: CONVEX SETS. Let $\mathscr{C}$ be the class of compact convex subsets in $\mathbb{R}^d$. For $c \in \mathscr{C}$, define a probability P_c as the uniform measure inside c. Define P_c^n as n-fold product measure. Take

$$x = \mathbb{R}^{nd}, \qquad \mathscr{P} = \{P_c^n\}_{c \in \mathscr{C}}.$$

This is a mathematical model for: "pick n points at random from inside an unknown convex, compact subset." This problem arises in estimating volumes of convex polyhedra. See, e.g., Deyer, Freize, and Kannen **[10]**. It is natural to ask what aspects of the data $x_1 \cdots x_n$ are required to learn about c. It is not hard to see that only the extreme points $T(x)$ of the convex hull are required. Indeed, given $T(x)$, the rest of the data is uniformly distributed inside the convex hull, no matter what convex set c underlies the selection process. It follows that $T(x)$ is sufficient.

The next section reviews the history and main mathematical results of sufficiency. Section 3 introduces exchangeability as part of the Bayesian view of statistics. Section 4 shows how sufficiency ideas give a natural foundation for exchangeability, allowing a theory where there is no natural symmetry. The final section contains pointers to open problems and related subjects.

2. Basic results of sufficiency. Sufficiency began, as with so much else in mathematical statistics, with a paper of R. A. Fisher **[18]**. Fisher was comparing two different estimates for the scale parameter of the normal curve. The estimators were appropriate multiples of

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2} \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n}|X_i - \bar{X}|.$$

Here the observation consists of $X = (x_1, \ldots, x_n)$ and $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Fisher showed that the first estimator was 15% more accurate and indeed that any estimate based on the sum of the absolute deviations would loose some of the information in the full observation. Fisher's argument introduced the ideas of sufficiency which were evident due to the invariance of the normal distribution under the orthogonal group. Later, Fisher **[19]** abstracted the idea away from invariance and outlined a general theory. This history is discussed by Stigler [48] who also reports how an earlier giant, Laplace, missed the idea of sufficiency in his work on a very similar problem.

Fisher and Jerzy Neyman [42] developed techniques for finding sufficient statistics and quantifying in what sense a sufficient statistic contains all of the information in a sample. Basically, given $T(x) = t$, with no knowledge of which $P \in \mathscr{P}$ generated x, a new observation x^* distributed just like the original x can be created by independent randomization. Of course, the distribution of T depends on the underlying P, but that is all.

Another sense in which a sufficient statistic captures the information is given by the Rao-Blackwell theorem. This considers an estimator $\widehat{P}(x)$ of the measure P. If $\widehat{P}(x)$ does not depend on x through a sufficient statistic, then a more accurate estimator can be found, no matter what notion of accuracy is being used. This necessarily vague statement is made precise in any of the standard graduate texts on mathematical statistics of which Lehmann [38] is recommended.

Modern work on the mathematics of sufficiency began with Halmos-Savage [25] and Bahadur [2]. They developed a rigorous general framework using σ-algebras and the Radon-Nikodým theorem. They began the love affair that mathematical statistics has had with refined measure theory. This continues to the present day.

Group theory was also being employed to reduce the dimensionality of statistical problems. If a problem is invariant under a group, the data can be reduced to a so-called maximal invariant (a report of which orbit of the group contains the data point). It might also be possible to reduce by sufficiency and the question of whether these reduction operations commute is natural. Charles Stein gave natural conditions for commutation which were expanded in Hall, Wijsman, and Ghosh [24].

Sufficient statistics arise easily in connection with so-called exponential families of measures. These have densities proportional to $e^{\theta T(x)}$ with respect to a dominating measure which does not depend on θ. For such a family, given a sample of size n, $T(x_1) + T(x_2) + \cdots + T(x_n)$ is a sufficient statistic. Conversely, if a family of measures admits a lower-dimensional sufficient statistic B. O. Koopman, E. J. G. Pitman, and G. Darmois gave conditions under which the family is exponential. To appreciate the problem, consider $\mathscr{P}$ as the set of all measures on $\mathbb{R} \times \mathbb{R}$. There are 1-1 continuous functions from $\mathbb{R} \times \mathbb{R}$ into $\mathbb{R}$. Any of these gives a sufficient statistic for P, which is not any sort of exponential form. To rule out such behavior, some notion of smoothness must be assumed. The best modern version due to Hipp [26] proves a theorem assuming T is locally Lipshitz.

Exponential families constitute convenient families which include most of the classically studied examples. A unified theory is summarized in Lehman [37, 38], Barndorff-Neilson [3], or Johanson [28].

Exponential families are quite a restricted family of measures. Modern statistics deals with far richer classes of probabilities. This suggests a kind of paradox. If statistics is to be of any real use it must provide ways of boiling down great masses of data to a few humanly interpretable numbers. The Koopman-Pitman-Darmois theorem suggests this is impossible unless nature follows highly specialized laws which no one really believes.

There are two ways out of this conundrum. First, the Koopman-Pitman-Darmois theorem depends on reduction to fixed dimension. If the dimension of the reduction is allowed to grow with n a theory may be possible. As an illustration, in the convex set example of §1, the extremal points of the sample were a sufficient statistic. As the sample size grows, a polyhedral convex set has order $(\log n)$ extremal points. See Gröenboom **[23]** for recent work. I do not know of a theory that uses these ideas.

The second way around the conundrum uses the idea of approximate sufficiency. This idea has been developed in a comprehensive fashion by Lucian Lecam. As an example, a statistic T is approximately sufficient for a family $\mathscr{P}$ if

$$\sup_{P,Q\in\mathscr{P}} d(P(\cdot|T=t),\ Q(\cdot|T=t))$$

is small, where d is a metric on measures such as Hellinger's distance or total variation. Le Cam has shown that if a family admits an approximately sufficient statistic, then the best one can do using all of the data is only a small bit better than what is achievable using only the statistic. This is a small part of a dazzling body of work. Le Cam and Yang **[36]** is an accessible introduction.

There are several interesting aspects of sufficiency not described in this brief review. The elegant theory of completeness and sufficiency connects the analytic properties of a family of measure with the distribution of "what's left over after a sufficient reduction." See Lehmann **[38]** for a recent review. The theory of minimal sufficiency asks about the existence of smallest reductions. There are still fascinating open problems here. See Landers and Rogge **[30]**.

Of course, one need not throw away what is left over. These "ancillary statistics" can be used to investigate if the family of measures under consideration is really a reasonable match to the data being considered. This is apparent in Fisher's early work. Diaconis and Smith **[15]** give examples and a review of the literature.

3. Introduction to exchangeability and equivalence of ensembles.

A. *de Finetti's theorem.* Let $\mathbb{Z}_2 = \{0, 1\}$. Let $\mathbb{Z}_2^\infty$ be the infinite product space. A probability P on $\mathbb{Z}_2^\infty$ is *exchangeable* if it is permutation invariant: $P(0, 1, **\cdots) = P(1, 0, **\cdots)$, etc. An example is coin tossing measure with parameter $\theta : P_\theta(t) = \theta^t(1-\theta)^{n-t}$, $t = x_1 + \cdots + x_n$. Here and above $\{x_1, x_2 \cdots x_n, **\cdots\}$ denotes the cylinder set in $\mathbb{Z}_2^\infty$ which begins $x_1, x_2 \cdots x_n$, where x_i are binary digits.

One version of de Finetti's basic result is the following theorem.

THEOREM (de Finetti). *The set of all exchangeable probabilities on $\mathbb{Z}_2^\infty$ is a convex simplex with extreme points the coin tossing measures $\{P_\theta\}_{\theta\in[0,1]}$.*

The theorem says that for each exchangeable P there is a unique probability μ on $[0, 1]$ such that the following integral representation holds:

$$(3.1) \qquad P\{x_1, x_2 \cdots x_n\} = \int \theta^t (1-\theta)^{n-t} \mu(d\theta), \qquad t = x_1 + \cdots + x_n.$$

This holds for every n and binary sequence $x_1 \cdots x_n$ with the same μ.

de Finetti's motivation was philosophical. Statisticians have used expressions like the right-hand side of (3.1) since Bayes and Laplace. The term $\theta^t(1-\theta)^{n-t}$ is the likelihood of observing $x_1 \cdots x_n$. The measure $\mu(d\theta)$ is the prior distribution. The integral represents the probability of observing $x_1 \cdots x_n$ averaging over different values of θ.

Subjective Bayesians like de Finetti prefer not to focus on unobservable parameters like θ. They are perfectly willing to assign probabilities to observable outcomes like the next n flips of a coin. de Finetti's theorem shows that a simple invariance condition characterizes the classical assignments. The theorem does more: starting from an exchangeable measure on observables, the theorem builds a "parameter space" $[0, 1]$, and the likelihood and prior as part of its representation.

A clear, readable introduction to de Finetti's point of view appears in de Finetti [**9**]. Exchangeability is of interest in many areas of probability. de Finetti's theorem can be shown to be easily equivalent to Hausdorff's moment problem. See Feller [**17**]. The survey by Aldous [**1**] gives a splendid treatment with many other applications.

It is natural to try to develop parallel characterizations of the classical parametric models of statistics. As will be seen, symmetry can only go part of the way. The next section uses sufficiency to build a satisfactory general theory. We begin by changing the space and group.

B. *Freedman's theorem.* In 1962, David Freedman gave a version of de Finetti's theorem suitable for the normal distribution. Call a probability P on $\mathbb{R}^\infty$ *orthogonally invariant* if

$$(3.2) \qquad P(A * * \cdots) = P(\Gamma A * * \cdots)$$

for every cylinder set $A * \cdots *$ with $A \subset \mathbb{R}^n$ for some n and Γ in the orthogonal group $O(n)$.

THEOREM (Freedman). *The orthogonally invariant probabilities on $\mathbb{R}^\infty$ are a convex simplex with extreme points $\{P_\sigma\}_{\sigma\in[0,\infty)}$, where P_σ is the product measure on $\mathbb{R}^\infty$ of a mean 0, variance σ^2 Gaussian measure.*

The theorem says for every orthogonally invariant P on $\mathbb{R}^\infty$ there is a unique probability μ on $[0, \infty)$ such that

$$P(A * * \cdots) = \int_A \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-(x_1^2+\cdots+x_n^2)/2\sigma} \mu(d\sigma).$$

The present version of the theorem arose in Bayesian statistics. Earlier, equivalent versions arose in Schoenberg's **[47]** answer to a question in functional analysis: When can a metric space be isometrically imbedded in $\mathscr{L}^2$? Berg, Christensen, and Ressell **[4]** and Graham **[22]** give recent surveys of this line of work. The theorem can also be phrased as a description of all natural measures on ℓ^2—this space is too big to have translation invariant measures but orthogonally invariant measures are widely used as a substitute. Choquet **[7]** contains an extensive discussion.

Perhaps the oldest version in widespread use is a theorem in geometry. This result goes back at least to Mehler **[41]**:

Let $S_{n-1} = \{(x_1 \cdots x_n) \in \mathbb{R}^n : x_1^2 + \cdots + x_n^2 = n\}$. Pick a point from the uniform distribution U on S_{n-1}. The theorem says that the first coordinate of such a point has an approximate Gaussian distribution: for every real $a < b$, as n tends to infinity

$$U\{x \in S_{n-1} : a < x_1 < b\} \sim \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

A proof is easy by calculus. One is required to calculate the surface area of a sphere between a pair of parallel planes. Mehler derived the result while looking at orthogonal expansions on high-dimensional spheres.

An extension of the result implies Freedman's theorem: Indeed, the orthogonally invariant probabilities on $\mathbb{R}^n$ form a convex set. The extreme points are the uniform distribution on spheres.

$$U\{a_1 \le x_1 \le b_1 \cdots a_k \le x_k \le b_k\}$$
$$\sim \int_{a_1}^{b_1} \cdots \int_{a_k}^{b_k} \frac{1}{(\sqrt{2\pi})^k} e^{-(x_1^2 + \cdots x_k)^2/2} dx_1 \cdots dx_n.$$

This shows that the extreme points are approximately products of Gaussian measures and, for measures arising from orthogonally invariant probabilities P on $\mathbb{R}^\infty$, must be exactly product Gaussian.

A careful version of this argument with error estimates appears in Diaconis and Freedman **[13]** or in Diaconis, Eaton, and Lauritzen **[11]**. The latter authors discuss the following variant: pick Γ at random (Haar measure) in $O(n)$. The joint distribution of Γ_{ij}, $i, j \ll n^{1/3}$, are approximately independent product normal variables.

As a final variant, the result appears in the statistical mechanics literature phrased as a simple example of the equivalence of ensembles. Here, a system is constrained to move on a constant energy hypersurface in $6n$-dimensional space. In the easiest case (with no interaction) this surface can be taken as the sphere:

$$x_1^2 + \cdots + x_{6n}^2 = c.$$

In statistical mechanics, the chance of finding the system in some portion of phase space is given by the uniform distribution (the microcanonical ensemble). Physicists routinely calculate with a different measure (the macrocanonical ensemble) supported on all of $\mathbb{R}^{6d}$. In the simple example considered here, this has density proportional to $e^{-(x_1^2+x_2^2+\cdots+x_{6n}^2)/2\sigma^2}$ with σ^2 chosen to make the average energy equal to c. The equivalence of ensembles says that for certain sets the calculation under the macrocanonical ensemble is approximately equal to the calculation under the microcanonical distribution. Usually the bounds are fairly crude—enough to show that sets which are small under one measure are small under the second. In this simple setting, the quantitative versions of Freedman's theorem give more precise results. The microcanonical ensemble is approximately product normal for sets which only depend on $o(n)$ coordinates. See Diaconis and Freedman **[13]** for a precise statement.

The equivalence of ensembles holds for very general energy functions. Lanford **[31]** or Ruelle **[46]** give further details. The general set-up is closely related to the general versions of de Finetti's theorem explained in the next section.

4. Sufficiency and exchangeability. The work on de Finetti's theorem described in §3 can be summarized as the study of measures invariant under a group. In the examples, the extreme points were identified and parametrized by a nice set: $[0, 1]$ for exchangeable binary sequences and $[0, \infty)$ for orthogonally invariant processes. These are special situations. In contrast, the basic set-up of ergodic theory considers processes indexed by $\mathbb{Z}$, with $\mathbb{Z}$ acting by translation. Now there is no neat description of the extreme points—instead they are dense in the space of all invariant measures.

The problem of finding a generalization of the examples which would handle the standard families of mathematical statistics was solved using the language of sufficiency. To explain, observe that the exchangeable processes can either be characterized as measures invariant under the permutation group or as measures for which the sum is a sufficient statistic. Thus a measure is exchangeable if and only if, for each n,

$$P(x_1 \cdots x_n | x_1 + \cdots + x_n = t)$$

is uniform on all binary n-tuples with t ones.

Similarly, a measure is orthogonally invariant if and only if

$$P\{\cdot | x_1^2 + x_2^2 + \cdots + x_n^2 = t\}$$

is uniform on the $\sqrt{t}$ sphere. The following abstraction covers most cases of interest in statistics.

For each i, there is a space Ω_i (usually taken as a Polish space with its Borel σ-algebra). Let $\Omega = \prod_{i=1}^{\infty} \Omega_i$. For each n, there is a "sufficient statistic" $T_n : \prod_{i=1}^{n} \Omega_i \to W_n$, where W_n is some range space. The analog of the uniform distribution on the inverse image of T_n is played by a family of pre-specified measures $Q_{n,t}$ on $\prod_{i=1}^{n} \Omega_i$.

Given T_n and $Q_{n,t}$, define the class of partially exchangeable processes $M_{Q,T}$ as all P on Ω such that

$$P\{\cdot|T_n(x_1 \cdots x_n) = t\} = Q_{n,t}(\cdot).$$

More technically, a regular conditional distribution for P on the first n coordinates given $T_n = t$ is $Q_{n,t}$.

The Q's and T's are required to fit together as follows:

(1) $Q_{n,t}\{T_n^{-1}(t)\} = 1$.

(2) If

$$T_n(x_1 \cdots x_n) = T_n(x_1' \cdots x_n'),$$

then

$$T_{n+1}(x_1 \cdots x_n, y) = T_{n+1}(x_1' \cdots x_n', y).$$

(3) For each $s \in W_n$, $t \in W_{n+1}$,

$$Q_{n+1,t}(x_1 \cdots x_n | T_n(x_1 \cdots x_n) = s, x_{n+1}) = Q_{n,s}(x_1 \cdots x_n).$$

As an example, for coin tossing, $\Omega_i = \{0, 1\}$, $T_n(x_1, \dots, x_n) = x_1 + \cdots + x_n$, and $Q_{n,t}$ is taken as uniform over all $x_1, \dots, x_n$ with $X_1 + \cdots + x_n = t$. Conditions (1)–(3) are easy to check. For example, (3) says that if one is told there are s ones in the first n places and told x_{n+1}, then Q_{n+1} assigns equal conditional probability to all compatible strings.

It is easy to see that the partially exchangeable processes $M_{Q,T}$ form a convex set. The first problem is to find a description of the extreme points. This involves an excursion to infinity. Let $\Sigma = \bigcap_{n=1}^{\infty} \Sigma_n$ with Σ_n the σ-algebra generated by $T_n(X_1 \cdots X_n), X_{n+1}, X_{n+2}, \dots$. This Σ is called the partially exchangeable σ-algebra. The first result is the following abstract version of de Finetti's theorem due to Diaconis and Freedman [**12**].

THEOREM. *If Q_n and T_n satisfy* (1–3) *above, then there is an $E \in \Sigma$ such that $P(E) = 1$ for each $E \in M_{Q,T}$ and such that*

(a) *$Q_{n,T_n(X_1 \cdots X_n)}$ converges weak-star to a limit $Q_{(\omega)}$ as $n \to \infty$, for each $\omega \in E$.*

(b) *$\{Q_\omega\}_{\omega \in E}$ ranges over the extreme points of the convex set $M_{Q,T}$.*

(c) *For each $P \in M_{Q,T}$, there is a unique μ on E such that*

$$P(\cdot) = \int_E Q_\omega(\cdot)\mu(d\omega).$$

The theorem evolved over generations. It begins in the group invariant case with Krylov and Bogulyov. See Oxtoby [**43**] and Farrell [**16**]. Hunts' [**27**] axiomatic treatment of the Martin Boundary of a Markov chain is very close to giving the full result. The crucial conditions (2) and (3) were abstracted in early work of Freedman [**20**] and Bahadhur [**2**].

A general version in rather different language was sketched by Martin-Löf [**40**] and Lauritzen [**32–34**] in Denmark. These authors worked in a more general setting of projective limits rather than with the product description of Ω.

In developing the modern approach to statistical mechanics, Dobrushin, Lanford, and Ruelle developed a similar theory and conditions (1), (2), and (3) are known as the D-L-R conditions in statistical mechanics. Preston **[44]** or Georgii **[21]** contain recent presentations.

The theorem presents the extreme points in a rather abstracted form and further work is required to massage this presentation into a classical mold. Diaconis and Freedman **[12]** present dozens of examples which have occupied researchers in Bayesian statistics for the past thirty years. Aldous **[1]**, Lauritzen **[34]** and Ressell **[45]** also present unified pictures from different points of view. The latter is interesting in presenting a large class of examples where the sufficient statistics are sums with values in a semigroup and the extreme points are indexed by the dual semigroup.

As one example of recent progress, here is a result of Küchler-Lauritzen **[29]** and Diaconis-Freedman **[14]**: Suppose one begins with an exponential family through a sufficient statistic T. One can then form the $Q_{n,T}$ as the conditional laws determined by the family. This gives the ingredients of the general set-up and one can ask if the extreme points of $M_{Q,T}$ correspond with the original exponential family. While it is easy to construct counterexamples, a natural sufficient condition has been found which gives the answer "yes" for any reasonable continuous or discrete family. The argument involves a delicate measure-theoretic extension of Cauchy's functional equation to partially defined functions. It gives infinitely many natural examples of Q's and T's where the extreme points have a simple description.

BIBLIOGRAPHY

1. D. Aldous, *Exchangeability and related topics*, Lecture Notes in Math., vol. 1117, Springer, Berlin, 1984, pp. 1–198.
2. R. R. Bahadur, *Sufficiency and statistical decision functions*, Ann. Math. Statist. **25** (1954), 423–462.
3. O. Barndorff-Nielsen, *Information and exponential families in statistical theory*, Wiley, New York, 1978.
4. C. Berg, J. P. R. Christensen, and P. Ressell, *Harmonic analysis on semi-groups*, Springer, New York, 1984.
5. P. Billingsley, *Probability and measure*, Wiley, New York, 1979.
6. L. Brown, *Fundamentals of statistical exponential families*, Institute of Mathematical Statistics, Hayward, CA, 1987.
7. G. Choquet, *Lectures in analysis, Vol.* III, Benjamin, Reading, MA, 1969.
8. B. de Finetti, *Funzione caratteristica di un fenomeno aleatoria*, Atti Accad. Naz. Lincei Mem. Cl. Sci. Fis. Mat. Natur. (6) **4** (1931), 251–299.
9. ——, *La prévision: ses lois logiques, ses sources subjectives*, Ann. Inst. H. Poincaré **7** (1937), 1–68; translated in *Studies in subjective probability* (H. Kyburg and H. Smokler, eds.), Wiley, New York, 1964.
10. M. E. Deyer, A. M. Frieze, and R. Kannan, *A random polynomial time algorithm for approximating the volume of convex bodies*, Proc. 21st ACM Sympos. on Theory of Computing, 1989, pp. 375–381.
11. P. Diaconis, J. Eaton, and S. Lauritzan, *Finite de Finetti theorems in linear models and multivariate analysis*, Scand. J. Statist. (to appear).

12. P. Diaconis and D. Freedman, *Partial exchangeability and sufficiency* (G. K. Ghosh and F. Roy, eds.), Statistics Applications and New Directions: Proc. Indian Stat. Inst. Golden Jubilee Internat. Conf. Stat., Indian Statistical Institute, Calcutta, India, 1984, pp. 205–236.
13. ——, *A dozen de Finetti-style results in search of a theory*, Ann. Inst. H. Poincaré **23** (1987), 397–443.
14. ——, *Cauchy's equation and de Finetti's theorem*, Scand. J. Statist. **17** (1990), 235–250.
15. P. Diaconis and L. Smith, *Residual analysis of discrete longitudinal data*, Technical Report, Department of Statistics, Stanford University, 1989.
16. R. H. Farrell, *Representation of invariant measures*, Illinois J. Math. **6** (1962), 447–467.
17. W. Feller, *An introduction to probability and its applications, Vol.* II, 2nd ed., Wiley, New York, 1971.
18. R. A. Fisher, *A mathematical examination of the methods of determining the accuracy of an observation by the mean error and by the mean squared error*, Monthly Notices Roy. Astronom. Soc. **80** (1920), 758–770.
19. ——, *On the mathematical foundations of theoretical statistics*, Philos. Trans. Roy. Soc. London Ser. A **222** (1922), 309–368.
20. D. Freedman, *Invariants under mixing which generalize de Finetti's Theorem*, Ann. Math. Statist. **33** (1962), 916–923.
21. H. O. Georgii, *Gibbs measures and phase transitions*, de Gruyter, Berlin, 1988.
22. R. Graham, *Isometric embedding of graphs*, Graph Theory III (L. Beineke and R. Wilson, eds.), Academic Press, London, 1988.
23. P. Gröenboom, *Limit theorems for convex hulls*, Probab. Theory Related Fields **79** (1988), 327–368.
24. W. J. Hall, R. A. Wijsman, and J. K. Ghosh, *The relation between sufficiency and invariance with applications in sequential analysis*, Ann. Math. Statist. **36** (1965), 575–614.
25. P. R. Halmos and L. J. Savage, *Application of the Radon-Nikodym Theorem to the theory of sufficient statistics*, Ann. Math. Statist. **20** (1949), 225–241.
26. C. Hipp, *Sufficient statistics and exponential families*, Ann. Statist. **2** (1974), 1283–1292.
27. G. Hunt, *Markoff chains and Martin boundaries*, Illinois J. Math. **4** (1960), 313–340.
28. S. Johanson, *Introduction to the theory of regular exponential families*, Lecture Notes Series #3, Dept. of Statistics, Copenhagen University, 1979.
29. U. Küchler and S. Lauritzen, *Exponential families, extreme point models, and minimal space-time invariant functions for stochastic processes with stationary independent increments*, Scand. J. Statist. **15** (1989), 237–261.
30. D. Landers and L. Rogge, *Minimal sufficient σ-fields and minimal sufficient statistics. Two counterexamples*, Ann. Math. Statist. **43** (1972), 2045–2049.
31. O. Lanford, *Entropy and equilibrium states in classical statistical mechanics*, Lecture Notes in Physics, vol. 20, Springer, Berlin, 1973, pp. 1–113.
32. S. L. Lauritzen, *On the interrelationships among sufficiency and some related concepts*, Technical Report, Dept. of Statistics, University of Copenhagen, 1974.
33. ——, *Sufficiency, prediction, and extreme models*, Scand. J. Statist. **1** (1974), 128–134.
34. ——, *Extremal families and systems of sufficient statistics*, Lecture Notes in Statistics, vol. 49, Springer, Berlin, 1988.
35. L. Le Cam, *Sufficiency and approximate sufficiency*, Ann. Math. Statist. **35** (1964), 1419–1455.
36. L. Le Cam and G. Yang, *Asymptotic methods in statistical decision theory*, Springer, New York, 1990.
37. E. Lehmann, *Theory of point estimation*, 2nd ed., Wiley, New York, 1983.
38. ——, *Testing statistical hypotheses*, 2nd ed., Wiley, New York, 1986.
39. I. G. MacDonald, *Symmetric functions and Hall polynomials*, Oxford Univ. Press, Oxford, 1979.
40. P. Martin-Löf, *Repetitive structures and the relation between canonical and microcanonical distributions in statistics and statistical mechanics*, Proc. Conf. Foundational Questions in Statistical Inference (O. Barndorf-Nielsen, P. Plaesild, G. Schon, eds.), Aarhus Univ. Press, Aarhus, 1974.

41. F. G. Mehler, *Ueber die Entwicklung einer function von beliebig vielen variablen nach Laplaceschen functionen höheren Ordnung*, J. für Math. (Crelle's J.) **66** (1866), 161–176.
42. J. Neyman, *Sur un teorema concennente le cosidette statiche sufficienti*, Giorn. Ist. Ital. Att. G (1935), 320–334.
43. J. C. Oxtoby, *Ergodic sets*, Bull. Amer. Math. Soc. **58** (1952), 116–136.
44. C. Preston, *Canonical and microcanonical Gibbs states*, Z. Wahrsch. Verw. Gebiete **46** (1979), 125–158.
45. P. Ressel, *de Finetti-type theorems. An analytical approach*, Ann. Probab. **13** (1985), 898–922.
46. D. Ruelle, *Thermodynamic formalism*, Addison-Wesley, Reading, MA, 1978.
47. I. J. Schoenberg, *Metric spaces and positive definite functions*, Trans. Amer. Math. Soc. **44** (1938), 522–536.
48. S. M. Stigler, *Laplace, Fisher, and the discovery of sufficiency*, Biometrica **60** (1973), 439–445.

Department of Mathematics, Harvard University, Cambridge, Massachusetts 02138