Sufficiency, Partial Exchangeability, and Exponential Families

Steffen Lauritzen
University of Oxford

May 17, 2007

Overview of Lectures

1. Exchangeability and de Finetti's Theorem
2. Sufficiency, Partial Exchangeability, and Exponential Families
3. Exchangeable Arrays and Random Networks

Basic references for the series of lectures include Aldous (1985) and Lauritzen (1988). Other references will be given as we go along. For this lecture, the latter reference is particularly relevant.

$X_1, \ldots, X_n, \ldots$ is *exchangeable* if for all $n = 2, 3, \ldots,$, $\pi \in S(n)$

$$X_1, \ldots, X_n \stackrel{\mathcal{D}}{=} X_{\pi(1)}, \ldots, X_{\pi(n)}.$$

de Finetti (1931):
A binary sequence $X_1, \ldots, X_n, \ldots$ is exchangeable if and only if there exists a distribution function F on $[0, 1]$ such that for all n

$$p(x_1, \ldots, x_n) = \int_0^1 \theta^{t_n}(1-\theta)^{n-t_n}\, dF(\theta),$$

where $t_n = \sum_{i=1}^n x_i$. Further, *F is distribution function of $Y = \bar{X}_\infty$ and, conditionally on $Y = \theta$, $X_1, \ldots, X_n, \ldots$ are i.i.d. with expectation θ.*

Hewitt–Savage

Hewitt and Savage (1955):
If $X_1, \ldots, X_n, \ldots$ are exchangeable with values in $\mathcal{X}$, there exists a probability measure μ on $\mathcal{P}(\mathcal{X})$ on $\mathcal{X}$, such that

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = \int Q(A_1) \cdots Q(A_n)\, \mu(dQ),$$

Further, *μ is the distribution function of the empirical measure:*

$$M(A) = \lim_{n\to\infty} M_n(A) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \chi_A(X_i), \quad M \sim \mu.$$

and, *conditionally on $M = Q$, $X_1, \ldots, X_n, \ldots$ are i.i.d. with distribution Q:*

$$P(X_1 \in A_1, \ldots, X_n \in A_n \mid M = Q) = Q(A_1) \cdots Q(A_n).$$

Finite versions of de Finetti's Theorem

$X_1, \ldots, X_n$ are *n-exchangeable* if for fixed n:

$$X_1, \ldots, X_n \stackrel{\mathcal{D}}{=} X_{\pi(1)}, \ldots, X_{\pi(n)} \text{ for all } \pi \in S(n).$$

Diaconis and Freedman (1980b):

If $X_1, \ldots, X_n$ are n-exchangeable and P_k is distribution of $X_1, \ldots, X_k$, P_k can be approximated with the k-marginal $P_{\mu k}$ of an infinitely exchangeable P_μ. For $|\mathcal{X}| = c < \infty$ the bound is

$$||P_k - P_{\mu k}|| \leq \frac{2ck}{n}.$$

For general $\mathcal{X}$ the bound is

$$||P_k - P_{\mu k}|| \leq \frac{k(k-1)}{n}.$$

$||P - Q|| = 2 \sup_A |P(A) - Q(A)|$ is the total variation norm.

If P_0 and P_1 are both exchangeable (finitely or infinitely):

$$P_i(X_1 \in A_1, \ldots, X_n \in A_n) = P_i(X_{\pi(1)} \in A_1, \ldots, X_{\pi(n)} \in A_n), i = 0, 1$$

this also holds for any convex combination

$$P_\alpha = \alpha P_0 + (1-\alpha) P_1, 0 \leq \alpha \leq 1.$$

Thus, *the set of exchangeable measures is convex*. A point P of a convex set $\mathcal{P}$ is an *extreme point* if

$$P = (P_1 + P_2)/2 \text{ and } P, P_1, P_2 \in \mathcal{P} \text{ implies } P = P_1 = P_2.$$

Any point in a compact convex set can be represented as a barycenter (centre of gravity) of a measure concentrated on the extreme points.

The integral representation

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = \int Q(A_1) \cdots Q(A_n) \, \mu(dQ),$$

expresses an arbitrary exchangeable P as *barycenter* of a *unique* measure μ concentrated on the *extreme exchangeable distributions*, which correspond to i.i.d.r.v.

A compact and convex set, where the representing measure μ is uniquely determined by P is a *simplex*.

Extreme points and asymptotic behaviour

Consider the following σ-fields:

The *tail* σ-field of events that do not depend on the first finite number of coordinates:

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma\{X_n, X_{n+1}, \dots, \}.$$

The *exchangeable* σ-field $\mathcal{E}$, of all events A that are not affected by any finite permutation $\pi \in S(n)$.

The *sufficient* σ-field $\mathcal{M}$, generated by the limiting empirical measure M_∞. It clearly holds that

$$\mathcal{M} \subseteq \mathcal{T} \subseteq \mathcal{E}.$$

If P is exchangeable all three σ-fields coincide as measure algebras (Olshen, 1971).

An exchangeable distribution is an extreme exchangeable disribution if and only if it has trivial tail, i.e. $\mathcal{T}$ only contains sets of probability one or zero:

$$A \in \mathcal{T} \implies P(A) \in \{0, 1\}.$$

Necessity is easy. For if $A \in \mathcal{T}$, we could write

$$P(\cdot) = P(\cdot \mid A)P(A) + P(\cdot \mid \neg A)(1 - P(A)).$$

Since $P(\cdot \mid A)$ and $P(\cdot \mid \neg A)$ are both exchangeable, P cannot be an extreme point if $0 < P(A) < 1$.

The converse is a bit more subtle and needs a reverse martingale argument (or de Finetti's theorem) to deduce that if $X_1, \ldots, X_n, \ldots$ are exchangeable, they are conditionally i.i.d. given $\mathcal{T}$.

Same statement would be true if $\mathcal{T}$ were replaced by $\mathcal{M}$ or $\mathcal{E}$, and with essentially the same proof.

For binary variables, $X_1, \ldots, X_n, \ldots$ is exchangeable if and only if for all n

$$P(X_1 = x_1, \ldots, X_n = x_n) = \phi_n(\textstyle\sum_i x_i).$$

Because *$S(n)$ acts transitively on binary n-vectors with fixed sum,* i.e. if x and y are two such vectors, there is a permutation which sends x into y.
So, in the binary case, *exchangeability is equivalent to $t_n = \sum_i x_i$ being sufficient and*

$$p(x_1, \ldots, x_n \mid t_n) = \binom{n}{t_n}^{-1}.$$

In general,the basic sufficient statistic is the *empirical measure* M_n, or for $\mathcal{X} = \mathcal{R}$ the order statistic $(x_{(1)}, x_{(2)}, \ldots, x_{(n)})$.

We say that $t(x)$ is *summarizing* a distribution p if for some ϕ

$$p(x) = \phi(t(x)).$$

Note that *if $t(x)$ is summarizing, it is sufficient* for any family of distributions that it summarizes.
In addition it holds that

$$p(x \mid t) \text{ is uniform on } \{x : t(x) = t\}$$

Exchangeability is equivalent to $t_n = \sum_i x_i$ summarizing the distribution of $X_1, \ldots, X_n$.
If the distribution of a sequence of random variables is summarized by a sequence $t_n, n = 1, 2, \ldots$ of statistics, it is also known as a *partially exchangeable* sequence. This is not necessarily an appropriate term.

Rephrasing de Finetti–Hewitt–Savage

If a family of distributions for a sequence $X_1, \dots, X_n, \dots$ is summarized by the empirical measure, then every distribution in the family is conditionally i.i.d. given the infinitely remote future $\mathcal{T}$ or, equivalently, given the limiting empirical measure M_∞.

Geometric distribution

Let $X_1, X_2, \ldots,$ be i.i.d. with a geometric distribution so

$$p(x_i) = (1-\theta)\theta^{x_i}, \quad x = 0, 1, 2, \ldots$$

Then

$$p(x_1 \ldots, x_n) = (1-\theta)^n \theta^{\sum_i x_i}$$

so $\sum_i x_i$ is summarizing.

Question: What is the family of distributions on $\{0, 1, \ldots\}$ summarized by $\sum_i x_i$?

Answer: Mixtures of distributions which are conditionally i.i.d. and geometric given the tail.

Note this 'partially exchangeable' sequence is in fact also exchangeable.

Uniform distribution

Let $X_1, X_2, \ldots,$ be i.i.d. uniform on $]0, \theta]$:

$$p(x_i) = \theta^{-1}\chi_{]0,\theta]}(x_i), \quad 0 < x < \infty.$$

Then

$$p(x_1 \ldots, x_n) = \theta^{-n}\chi_{]0,\theta]}(\max_i x_i)$$

so $\max_i x_i$ is summarizing.
Could be considered both for x integer or real-valued.

Question: What is the family of distributions on $]0, \theta]$ summarized by $\max_i x_i$?

Answer: Mixtures of distributions which are conditionally i.i.d. and uniform given the tail.

Normal distribution

Let $X_1, X_2, \ldots,$ be i.i.d. $\mathcal{N}(0, \sigma^2)$:

$$p(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

Then

$$p(x_1 \ldots, x_n) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{\sum_i x_i^2}{2\sigma^2} + \frac{\mu \sum_i x_i}{\sigma^2} - \frac{n\mu^2}{2\sigma^2}}$$

so $(\sum_i x_i^2, \sum_i x_i)$ is summarizing.

Question: What is the family of distributions on $\mathcal{R}$ summarized by $(\sum_i x_i^2, \sum_i x_i)$?

Answer: Mixtures of distributions which are conditionally i.i.d. and normally distributed given the tail.

We apparently have a way of generating models corresponding to given summary statistics. *Which statistics are possible?* Clearly $t(x) = x$ is always summarizing.

Question: What is the family of distributions on $\mathcal{R}$ summarized by $\mathtt{median}(x_1, \ldots, x_n)$?

Answer: None.

In fact, *any minimal, summarizing sequence of statistics is recursively computable:*

$$t_{n+1}(x_1, \ldots, x_n, x_{n+1}) = \phi_n\{t_n(x_1, \ldots, x_n), x_{n+1}\}.$$

This property of sufficient statistics was observed by Fisher (1925), see also Freedman (1962); Lauritzen (1988).
So the median can never be a minimal sufficient statistic.

If a sequence of statistics $t_n, n = 1, 2, \ldots$ are all *symmetric*, i.e.

$$t_n(x_1, \ldots, x_n) = t_n(x_{\pi(1)}, \ldots x_{\pi(n)}), \pi \in S(n)$$

and recursively computable, it must be of the form

$$t_n(x_1, \ldots, x_n) = t(x_1) \oplus \cdots \oplus t(x_n),$$

where t takes values in an *Abelian semigroup* i.e. $\oplus$ satisfies

$$a \oplus b = b \oplus a, \quad (a \oplus b) \oplus c = a \oplus (b \oplus c).$$

Conversely, any such statistic is recursively computable and symmetric.

Examples of semigroup statistics are

$$t(x) = x, \quad x \oplus y = x + y,$$

$$t(x) = x, \quad x \oplus y = max(x, y)$$

and

$$t(x) = \delta_x, \quad \delta_x \oplus \delta_y = \delta_x + \delta_y,$$

where δ_x is the distribution with point mass in x.

These correspond to the sum, the maximum, and the empirical distribution as summarizing statistics.

de Finetti's Theorem for semigroups

Let $t : \mathcal{X} \to \mathcal{S}$ be a semigroup valued statistic.

The distribution of $X_1, \ldots, X_n$ of is summarized by $t_n(x_1, \ldots, x_n) = t(x_1) \oplus \cdots \oplus t(x_n)$ for all n if and only if $X_1, \ldots, X_n, \ldots$ are conditionally i.i.d. given the tail $\mathcal{T}$ and

$$P(X_i = x \mid \mathcal{T}) = p(x) = p(x \mid \theta) = c(\theta)^{-1}\rho_\theta\{t(x)\}$$

where ρ_θ is a *character* on the semigroup generated by $t(\mathcal{X})$, i.e. an 'exponential function', satisfying

$$\rho_\theta(u)\rho_\theta(v) = \rho(u \oplus v), \quad \rho_\theta(u) \geq 0.$$

Shown in Lauritzen (1982), see also Lauritzen (1984, 1988); Ressel (1985).

Recall that

$$p(x \mid \theta) = c(\theta)^{-1}\rho_\theta\{t(x)\}.$$

For $x \oplus y = x + y$, the characters are

$$\rho_\theta(x) = \theta^x$$

corresponding to the geometric distribution as before.

For $x \oplus y = max(x, y)$, the characters are

$$\rho_\theta(x) = \chi_{]0,\theta]}(x),$$

corresponding to the geometric distribution.

For the empirical measures $\delta_x \oplus \delta_y = \delta_x + \delta_y$, the characters are

$$\rho_\theta(x) = \theta_x, \quad \theta = \{\theta_x, x \in \mathcal{X}\}.$$

A non-standard example

For distributions on the integers $\mathcal{X} = 1, 2, \ldots$ and $t(x) = x$ with $x \oplus y = xy$ we get

$$p_\theta(x) = \prod_{\nu \in \Pi} \theta_\nu^{n_\nu(x)}, \quad \theta = \{\theta_\nu, \nu \in \Pi\},$$

where Π are the prime numbers and $n_\nu(x)$ the number of times ν divides x.
If X is distributed according to $p(x \mid \theta)$, the multiplicities $n_\nu(X)$ of its prime factors are independent and geometrically distributed with parameter θ_ν (Lauritzen, 1988).

de Finetti's Theorem for Finite Markov chains

Diaconis and Freedman (1980a) show for countable $\mathcal{X}$ that *if the distribution of $X_1, \ldots, X_n$ is for all n summarized by*

$$t_n(x_1, \ldots, x_n) = (x_1, \{n_{xy}\}_{x,y \in \mathcal{X}})$$

where n_{xy} are the transition counts:

$$n_{xy} = \#\{i : (x_i, x_{i+1}) = (x, y)\}$$

and the process is recurrent, then it is a mixture of stationary Markov chains.

Similar results true for

$$t_n(x_1, \dots, x_n) = \{x_1, \oplus_i t(x_i, x_{i+1})\}$$

where $t : \mathcal{X} \times \mathcal{X} \to \mathcal{S}$ is semigroup valued:

The extreme recurrent processes are Markov chains with

$$P(X_{n+1} = y \,|\, X_n = x) = \rho_\theta\{t(x, y)\} \frac{c_\theta(y)}{c_\theta(x)},$$

where c_θ are eigenvectors with eigenvalue 1 for the matrix $m_{xy} = \rho_\theta\{t(x, y)\}$; see Ressel (1988) for full details.

Clearly, then

$$p(x_1, \dots, x_n) = p(x_1)\rho_\theta\{\oplus_i t(x_i, x_{i+1})\} \frac{c_\theta(x_n)}{c_\theta(x_1)}.$$

Finite versions of deFinetti's Theorem for semigroups have been given by Diaconis and Freedman (1988).

Things get more complex with statistics of the form

$$t_n(x_1, \ldots, x_n) = \oplus_i t_i(x_i),$$

for example for $t_n(x_1, \ldots, x_n) = \sum_i ix_i$ (Lauritzen, 1984, 1988).

Next time on to arrays and random graphs!

Aldous, D. (1985). Exchangeability and related topics. In Hennequin, P., editor, *École d'Été de Probabilités de Saint–Flour XIII — 1983*, pages 1–198. Springer-Verlag, Heidelberg. Lecture Notes in Mathematics 1117.

de Finetti, B. (1931). Funzione caratteristica di un fenomeno aleatorio. *Atti della R. Academia Nazionale dei Lincei, Serie 6. Memorie, Classe di Scienze Fisiche, Mathematice e Naturale*, 4:251–299.

Diaconis, P. and Freedman, D. (1980a). de Finetti's theorem for Markov chains. *Annals of Probability*, 8:115–130.

Diaconis, P. and Freedman, D. (1980b). Finite exchangeable sequences. *Annals of Probability*, 8:745–764.

Diaconis, P. and Freedman, D. (1988). Conditional limit theorems for exponential families and finite versions of de finetti's theorem. *Journal of Theoretical Probability*, 1:381–410.

Fisher, R. A. (1925). Theory of statistical estimation. *Proceedings of the Cambridge Philosophical Society*, 22:700–725.

Freedman, D. (1962). Invariants under mixing which generalize de Finetti's theorem. *Annals of Mathematical Statistics*, 33:916–923.

Hewitt, E. and Savage, L. J. (1955). Symmetric measures on Cartesian products. *Transactions of the American Mathematical Society*, 80:470–501.

Lauritzen, S. L. (1982). *Statistical Models as Extremal Families and Systems of Sufficient Statistics*. Aalborg University Press, Aalborg, Denmark.

Lauritzen, S. L. (1984). Extreme point models in statistics (with discussion). *Scandinavian Journal of Statistics*, 11:65–91.

Lauritzen, S. L. (1988). *Extremal Families and Systems of Sufficient Statistics*. Springer-Verlag, Heidelberg. Lecture Notes in Statistics 49.

Olshen, R. A. (1971). The coincidence of measure algebras under an exchangeable probability. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 18:153–158.

Ressel, P. (1985). de Finetti-type theorems: An analytical approach. *Annals of Probability*, 13:898–922.

Ressel, P. (1988). Integral representations for distributions of symmetric stochastic processes. *Probability Theory and Related Fields*, 79:451–467.