Exchangeability and de Finetti's Theorem

Steffen Lauritzen
University of Oxford

April 26, 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 particular lecture, the article by Kingman (1978) is strongly recommended.

Definition

An infinite sequence $X_1, \ldots, X_n, \ldots$ of random variables is said to be *exchangeable* if for all $n = 2, 3, \ldots,$

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

where $S(n)$ is the group of permutations of $\{1, \ldots, n\}$.
For example, for a binary sequence, we may have:

$$p(1,1,0,0,0,1,1,0) = p(1,0,1,0,1,0,0,1).$$

If $X_1, \ldots, X_n, \ldots$ are independent and identically distributed, they are exchangeable, but *not conversely.*

Polya's Urn

Consider an urn with b black balls and w white balls. Draw a ball at random and note its colour. Replace the ball together with a balls of the same colour. Repeat the procedure *ad infinitum*. Let $X_i = 1$ if the i-th draw yields a black ball and $X_i = 0$ otherwise.

The sequence $X_1, \ldots, X_n, \ldots$ is exchangeable:

$$\begin{aligned} p(1,1,0,1) &= \frac{b}{b+w}\frac{b+a}{b+w+a}\frac{w}{b+w+2a}\frac{b+2a}{b+w+3a} \\ &= \frac{b}{b+w}\frac{w}{b+w+a}\frac{b+a}{b+w+2a}\frac{b+2a}{b+w+3a} = p(1,0,1,1). \end{aligned}$$

$X_1, \ldots, X_n, \ldots$ are not independent (nor even a Markov process).

A Gaussian example

Consider a sequence of Gaussian random variables with $X_1, \ldots, X_n$ multivariate Gaussian $\mathcal{N}_n(0, \Sigma)$ with mean zero and covariance matrix determined as

$$\mathbf{E}(X_i^2) = 1, \quad \mathbf{E}(X_i X_j) = \rho \text{ for } i \neq j.$$

For the matrix to be positive definite so the distribution is well defined, we must have

$$\rho \geq -\frac{1}{n-1}$$

and since this should hold for all n, we further conclude $\rho \geq 0$. Since a Gaussian distribution is determined by its mean and covariance, this defines an exchangeable sequence.

de Finetti's Theorem

de Finetti (1931) shows that all exchangeable binary sequences are mixtures of Bernoulli sequences:

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 $p(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$ and $t_n = \sum_{i=1}^n x_i$.

More about de Finetti's Theorem

It further holds that *F is the distribution function of the limiting frequency:*

$$Y = \bar{X}_\infty = \lim_{n\to\infty} \sum_i X_i/n, \quad P(Y \le y) = F(y)$$

and *the Bernoulli distribution is obtained by conditioning with $Y = \theta$:*

$$P(X_1 = x_1, \dots, X_n = x_n \mid Y = \theta) = \theta^{t_n}(1-\theta)^{n-t_n}.$$

Polya's Urn

Urn has initially b black balls and w white balls. Draw a ball at random and note its colour. Replace the ball together with a balls of the same colour. Repeat the procedure *ad infinitum*.

Let $X_i = 1$ if the i-th draw yields a black ball and $X_i = 0$ otherwise.

The average $\bar{X}_n$ converges to a random value Y which has a Beta distribution

$$\lim_{n\to\infty} \bar{X}_n = Y \sim \mathcal{B}(b/a, w/a).$$

Conditionally on $Y = \theta$, $X_1 \ldots, X_n, \ldots$ are independent and Bernoulli with parameter θ.

This has been known as folklore for a long time. First published proof seems to be in Blackwell and Kendall (1964), see also Freedman (1965).

Hewitt–Savage

Hewitt and Savage (1955) have generalized de Finetti's result: *Let $X_1, \ldots, X_n, \ldots$ be an exchangeable sequence of random variables with values in $\mathcal{X}$. Then there exists a probability measure μ on the set of probability measures $\mathcal{P}(\mathcal{X})$ on $\mathcal{X}$, such that*

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

It further holds that *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 *$Q^{\otimes\infty}$ is the distribution obtained by conditioning with M:*

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

Gaussian example

$X_1, \ldots, X_n$ multivariate Gaussian $\mathcal{N}_n(0, \Sigma)$ with mean zero and covariance

$$\mathbf{E}(X_i^2) = 1, \quad \mathbf{E}(X_i X_j) = \rho \text{ for } i \neq j, \rho > 0,$$

the empirical distribution function converges to a Gaussian distribution

$$F_\infty(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \chi_{]-\infty,x]}(X_i) = \Phi\{(x - Y)/\sqrt{1-\rho}\},$$

where Φ is the Gaussian distribution function and

$$Y = \lim_{n\to\infty} \bar{X}_n \sim \mathcal{N}(0, \rho).$$

Conditionally on $Y = \theta$, $X_1 \ldots, X_n, \ldots$ are independent and distributed as $\mathcal{N}(\theta, 1-\rho)$.

Simplest Bayesian model:
Subjective *exchangeable* probability distribution P representing Your expectations for behaviour of $X_1, \ldots, X_n, \ldots$.

Simplest frequentist model:
There is an unknown distribution Q, so that $X_1, \ldots, X_n, \ldots$ are independent and identically distributed with distribution Q, $Q(A)$ being *defined* as the limiting proportion of X's in A.

de Finetti–Hewitt–Savage Theorem provides *bridge* between the two model types:

In P, the distribution Q exists as a random object, *also determined* by the limiting frequency. The distribution, μ, of Q is the *Bayesian prior* distribution:

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

The empirical measure M_n ($\bar{X}_n$ in the binary case) is *sufficient* for the unknown 'parameter' Q.

Summary of de Finetti–Hewitt-Savage

For an exchangeable sequence $X_1, \ldots, X_n, \ldots$ with joint distribution P we have for most, (but not all) measure spaces $\mathcal{X}$:

- An integral representation: $P = \int Q^{\otimes\infty} \, \mu(dQ)$;

Summary of de Finetti–Hewitt-Savage

For an exchangeable sequence $X_1, \ldots, X_n, \ldots$ with joint distribution P we have for most, (but not all) measure spaces $\mathcal{X}$:

- An integral representation: $P = \int Q^{\otimes\infty} \mu(dQ)$;
- The empirical distribution M_n converges to a random distribution M;

Summary of de Finetti–Hewitt-Savage

For an exchangeable sequence $X_1, \dots, X_n, \dots$ with joint distribution P we have for most, (but not all) measure spaces $\mathcal{X}$:

- An integral representation: $P = \int Q^{\otimes\infty} \mu(dQ)$;
- The empirical distribution M_n converges to a random distribution M;
- μ is uniquely determined by P as the distribution of the limiting empirical measure M, so we can write $P = P_\mu$.

$X_1, \dots, X_n$ is *n-exchangeable* if

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

Every n-subset of an exchangeable sequence is n-exchangeable for every n, but not conversely.
Two binary variables X_1, X_2 are 2-exchangeable (pairwise exchangeable) if and only if

$$p(0,1) = p(1,0).$$

A k-exchangeable $X_1, \dots, X_k$ is *n-extendable* if it has the same distribution as the k first of an n-exchangeable variables $X_1, \dots, X_n$

Finite versions of de Finetti's Theorem

Diaconis and Freedman (1980) show for $X_1, \ldots, X_k$ k-exchangeable and n-extendable, where P_k denotes distribution of $X_1, \ldots, X_k$. *The distribution P_k can be approximated with the k-marginal $P_{\mu k}$ of an exchangeable distribution P_μ, with bound satisfying*

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

if $|\mathcal{X}| = c < \infty$, and in the general case

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

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

Proof of classical theorem

Most proofs of the de Finetti–Hewitt–Savage Theorem are based on martingale arguments, considering quantities such as

$$Z_{nk} = \mathbf{E}\{\phi_1(X_1)\phi_2(X_2)\cdots\phi_k(X_k)\,|\,X_n, X_{n+1}, \ldots\}$$

which is a *reverse martingale* w.r.t. the decreasing family of σ-algebras $\mathcal{S}_n = \sigma\{X_n, X_{n+1}, \ldots\}$.
See for example (Kingman, 1978) for a simple and illuminating proof.

Proof of finite theorem

The proof of the finite version is essentially based on a combinatorial exercise, establishing sharp bounds for the approximations of sampling with and without replacement. The finite version yields the integral representation of an exchangeable probability distribution as a corollary.

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.

Blackwell, D. and Kendall, D. (1964). The Martin boundary for Pólya's urn scheme, and an application to stochastic population growth. *Journal of Applied Probability*, 1:284–296.

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. (1980). Finite exchangeable sequences. *Annals of Probability*, 8:745–764.

Freedman, D. A. (1965). Bernard Friedman's urn. *Annals of Mathematical Statistics*, 36:956–970.

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

Kingman, J. F. C. (1978). Uses of exchangeability. *Annals of Probability*, 6:183–197.

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