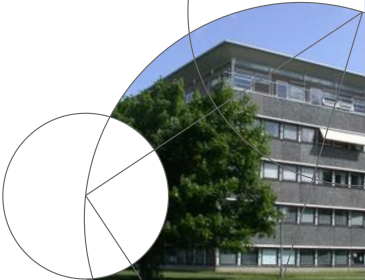
De Finetti's theorem and beyond: Extreme point models revisited

Steffen Lauritzen



Overview

1. Statistical models
2. de Finetti and Hewitt–Savage theorems
3. Sufficient and summarizing statistics
4. Binary arrays
5. Repetitive structures
6. Random graphs

Statistical models

Statistical models

The term 'model' has many different meanings:

Statistical models

The term 'model' has many different meanings:

- **True model** $P_0 \in \mathbb{P}(\mathcal{X})$, with frequentist interpretation;

Statistical models

The term 'model' has many different meanings:

- **True model** $P_0 \in \mathbb{P}(\mathcal{X})$, with frequentist interpretation;
- **Fisherian statistical model:** P_0 assumed to belong to $\mathcal{P} = \{P_\theta : \theta \in \Theta\} \subseteq \mathbb{P}(\mathcal{X})$, but otherwise unknown;

Statistical models

The term 'model' has many different meanings:

- **True model** $P_0 \in \mathbb{P}(\mathcal{X})$, with frequentist interpretation;
- **Fisherian statistical model:** P_0 assumed to belong to $\mathcal{P} = \{P_\theta : \theta \in \Theta\} \subseteq \mathbb{P}(\mathcal{X})$, but otherwise unknown;
- **common Bayesian model,** adding a prior $\mu \in \mathbb{P}(\Theta)$ to the Fisherian;

Statistical models

The term 'model' has many different meanings:

- **True model** $P_0 \in \mathbb{P}(\mathcal{X})$, with frequentist interpretation;
- **Fisherian statistical model:** P_0 assumed to belong to $\mathcal{P} = \{P_\theta : \theta \in \Theta\} \subseteq \mathbb{P}(\mathcal{X})$, but otherwise unknown;
- **common Bayesian model,** adding a prior $\mu \in \mathbb{P}(\Theta)$ to the Fisherian;
- **Bayesian subjective model**, expressing individual beliefs for observables through a single $P_0 \in \mathbb{P}(\mathcal{X} \times \mathcal{Y})$.

Statistical models

The term ‘model’ has many different meanings:

- **True model** $P_0 \in \mathbb{P}(\mathcal{X})$, with frequentist interpretation;
- **Fisherian statistical model:** P_0 assumed to belong to $\mathcal{P} = \{P_\theta : \theta \in \Theta\} \subseteq \mathbb{P}(\mathcal{X})$, but otherwise unknown;
- **common Bayesian model,** adding a prior $\mu \in \mathbb{P}(\Theta)$ to the Fisherian;
- **Bayesian subjective model**, expressing individual beliefs for observables through a single $P_0 \in \mathbb{P}(\mathcal{X} \times \mathcal{Y})$.

Most scientists have a quite different understanding of the term.

Statistical models

The term 'model' has many different meanings:

- **True model** $P_0 \in \mathbb{P}(\mathcal{X})$, with frequentist interpretation;
- **Fisherian statistical model:** P_0 assumed to belong to $\mathcal{P} = \{P_\theta : \theta \in \Theta\} \subseteq \mathbb{P}(\mathcal{X})$, but otherwise unknown;
- **common Bayesian model,** adding a prior $\mu \in \mathbb{P}(\Theta)$ to the Fisherian;
- **Bayesian subjective model**, expressing individual beliefs for observables through a single $P_0 \in \mathbb{P}(\mathcal{X} \times \mathcal{Y})$.

Most scientists have a quite different understanding of the term.

For example, a model could be a functional relationship $y = f(x)$ which is 'fitted' to data $(\tilde{x}, \tilde{y})$ with 'error' $\epsilon = \tilde{y} - f(\tilde{x})$.

de Finetti and Hewitt–Savage theorems

Exchangeability

Definition

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

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

where $S(n)$ is the group of permutations of $\{1, \ldots, n\}$.

Exchangeability

Definition

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

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

where $S(n)$ is the group of permutations of $\{1, \ldots, n\}$.

For example, for a binary sequence, we may have:

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

Exchangeability

Definition

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

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

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.*

de Finetti's Theorem

Theorem (de Finetti (1931))

A binary sequence is exchangeable iff 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 $t_n=\sum_{i=1}^n x_i$.

de Finetti's Theorem

Theorem (de Finetti (1931))

A binary sequence is exchangeable iff 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 the distribution of the limiting frequency:

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

de Finetti's Theorem

Theorem (de Finetti (1931))

A binary sequence is exchangeable iff there exists a distribution function F on $[0,1]$ such that for all n

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

where $t_n = \sum_{i=1}^n x_i$. Further, F is the distribution of the limiting frequency:

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

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

The Hewitt–Savage theorem

Theorem (Hewitt and Savage (1955))

Let $X_1, \ldots, X_n, \ldots$ be exchangeable random variables with values in $\mathcal{X}$. Then there is a probability μ on $\mathbb{P}(\mathcal{X})$ such that

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

Furhter, μ is the distribution 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 obtained by conditioning on M:

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

Contrasting model types

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

Contrasting model types

Simplest Bayesian model:
Subjective *exchangeable* distribution P representing Your expectation 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.

Bridging model types

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

Bridging model types

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

In P, the distribution Q — *i.e. the parameter* — exists as a random object, *also determined* by the limiting frequency.

Bridging model types

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

In P, the distribution Q — *i.e. the parameter* — exists as a random object, *also determined* by the limiting frequency.

The distribution μ of Q is the *Bayesian prior* distribution:

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

Bridging model types

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

In P, the distribution Q — *i.e. the parameter* — exists as a random object, *also determined* by the limiting frequency.

The distribution μ of Q is the *Bayesian prior* distribution:

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

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

A convexity perspective

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

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

A convexity perspective

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

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

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

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

Integral representation

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

Integral representation

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.

Integral representation

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:

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

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:

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

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

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:

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

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).

Characterizing extreme points

Theorem

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

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

Characterizing extreme points

Theorem

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

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

Same statement would be true if $\mathcal{T}$ were replaced by $\mathcal{M}$ or $\mathcal{E}$.

Sufficient and summarizing statistics

Exchangeability and summarizing statistics

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

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

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.

Exchangeability and summarizing statistics

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

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

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*

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

In general, the basic sufficient statistic is the *empirical measure* M_n.

Summarizing statistics

Following Freedman (1962) 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.

Summarizing statistics

Following Freedman (1962) 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 $X_1, \dots, X_n$.

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.

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 $\sum_i x_i$?

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.

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.

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.

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

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.

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.

Properties of summarizing statistics

We apparently have a way of generating models corresponding to given summary statistics. *Which statistics are possible?*

Properties of summarizing statistics

We apparently have a way of generating models corresponding to given summary statistics. *Which statistics are possible?*

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

Properties of summarizing statistics

We apparently have a way of generating models corresponding to given summary statistics. *Which statistics are possible?*

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

Answer: None.

Properties of summarizing statistics

We apparently have a way of generating models corresponding to given summary statistics. *Which statistics are possible?*

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

Answer: None.

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

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

Properties of summarizing statistics

We apparently have a way of generating models corresponding to given summary statistics. *Which statistics are possible?*

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

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 minimal sufficient statistics was observed by Fisher (1925), see also Freedman (1962); Lauritzen (1988).

Properties of summarizing statistics

We apparently have a way of generating models corresponding to given summary statistics. *Which statistics are possible?*

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

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 minimal sufficient statistics was observed by Fisher (1925), see also Freedman (1962); Lauritzen (1988).

So the median can never be a minimal sufficient statistic.

Semigroup statistics

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

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

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

Shown e.g. in Lauritzen (1988); Ressel (1985).

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

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

where Π are the prime numbers and $n_\nu(x)$ the number of times ν divides x.

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

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

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).

Binary arrays

Beyond sequences

Rasch model (Rasch, 1960):
Problem i attempted by person j. There are 'easinesses' $\alpha = (\alpha_i)_{i=1,\dots}$ and 'abilities' $\beta = (\beta_j)_{j=1,\dots}$ so that binary responses X_{ij} are conditionally independent given (α, β) and

$$P(X_{ij} = 1 \mid \alpha, \beta) = 1 - P(X_{ij} = 0 \mid \alpha, \beta) = \frac{\alpha_i \beta_j}{1 + \alpha_i \beta_j}.$$

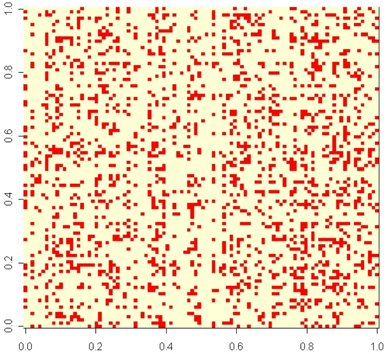
Beyond sequences

Rasch model (Rasch, 1960):
Problem i attempted by person j. There are 'easinesses' $\alpha = (\alpha_i)_{i=1,\ldots}$ and 'abilities' $\beta = (\beta_j)_{j=1,\ldots}$ so that binary responses X_{ij} are conditionally independent given (α, β) and

$$P(X_{ij} = 1 \,|\, \alpha, \beta) = 1 - P(X_{ij} = 0 \,|\, \alpha, \beta) = \frac{\alpha_i \beta_j}{1 + \alpha_i \beta_j}.$$

A *random Rasch matrix* has (α_i) i.i.d. with distribution A and (β_j) i.i.d. B.
Also potential model for *hit of batter i against pitcher j, occurrence of species i on island j,* etc.

Example of random Rasch matrix



Row- and column-exchangeable matrices

A doubly infinite matrix $X = \{X_{ij}\}_{1,1}^{\infty,\infty}$ is said to be

- *row–column exchangeable* (RCE-matrix) if for all m, n, $\pi \in S(m), \rho \in S(n)$
$$\{X_{ij}\}_{1,1}^{m,n} \stackrel{\mathcal{D}}{=} \{X_{\pi(i)\rho(j)}\}_{1,1}^{m,n}.$$
- *weakly exchangeable* (WE-matrix) if for all n and $\pi \in S(n)$
$$\{X_{ij}\}_{1,1}^{n,n} \stackrel{\mathcal{D}}{=} \{X_{\pi(i)\pi(j)}\}_{1,1}^{n,n}.$$

Summarized matrices

A doubly infinite (binary) matrix $X = \{X_{ij}\}_{1,1}^{\infty,\infty}$ is said to be *row-column summarized* (RCS-matrix) if for all m, n

$$p(\{x_{ij}\}_{1,1}^{m,n}) = \phi_{m,n}\{R_1, \ldots, R_m; C_1, \ldots, C_n\},$$

with $R_i = \sum_j x_{ij}$ and $C_j = \sum_j x_{ij}$ the row- and column sums.

Summarized matrices

A doubly infinite (binary) matrix $X = \{X_{ij}\}_{1,1}^{\infty,\infty}$ is said to be *row-column summarized* (RCS-matrix) if for all m, n

$$p(\{x_{ij}\}_{1,1}^{m,n}) = \phi_{m,n}\{R_1, \ldots, R_m; C_1, \ldots, C_n\},$$

with $R_i = \sum_j x_{ij}$ and $C_j = \sum_j x_{ij}$ the row- and column sums.

Note that, in contrast to the case of binary sequences, *RCE-matrices are generally not RCS-matrices and vice versa* because group G_{RC} of row and column permutations does *not* act transitively on matrices with fixed row- and column sums:

Summarized matrices

A doubly infinite (binary) matrix $X = \{X_{ij}\}_{1,1}^{\infty,\infty}$ is said to be *row-column summarized* (RCS-matrix) if for all m, n

$$p(\{x_{ij}\}_{1,1}^{m,n}) = \phi_{m,n}\{R_1, \ldots, R_m; C_1, \ldots, C_n\},$$

with $R_i = \sum_j x_{ij}$ and $C_j = \sum_j x_{ij}$ the row- and column sums.

Note that, in contrast to the case of binary sequences, *RCE-matrices are generally not RCS-matrices and vice versa* because group G_{RC} of row and column permutations does *not* act transitively on matrices with fixed row- and column sums:

If a matrix is both RCE and RCS, it is an *RCES-matrix.*

Convexity formulation

The set of distributions $\mathcal{P}_{RCE}$ is a convex simplex.
In particular, *every $P \in \mathcal{P}_{RCE}$* has a unique representation as a mixture of extreme points $\mathcal{E}_{RCE}$, i.e.

$$P(A) = \int_{\mathcal{E}} Q(A)\mu_P(Q).$$

The same holds if RCE is replaced by RCS, RCES, WE, SWE, SWES, etc. In addition, it can be shown that

$$\mathcal{E}_{RCES} = \mathcal{E}_{RCE} \cap \mathcal{P}_{RCS}, \quad \mathcal{E}_{WES} = \mathcal{E}_{WE} \cap \mathcal{P}_{WS},$$

etc.

Features of extreme measures

Aldous (1981): *for any $P \in \mathcal{P}_{RCE}$ the following are equivalent:*

- $P \in \mathcal{E}_{RCE}$
- *The tail σ-field $\mathcal{T}$ is trivial*
- *The corresponding RCE-matrix X is dissociated.*

Here the *tail* $\mathcal{T}$ is $\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma\{X_{ij}, \min(i,j) \geqslant n\}$ and a matrix is *dissociated* (Silverman, 1976) if for all A_1, A_2, B_1, B_2 with $A_1 \cap A_2 = B_1 \cap B_2 = \emptyset$

$$\{X_{ij}\}_{i \in A_1, j \in B_1} \perp\!\!\!\perp \{X_{ij}\}_{i \in A_2, j \in B_2}.$$

Repetitive structures

Repetitive structure

Let $(\mathcal{I}, \prec)$ be a *partially ordered and directed* index set, i.e.

$$i \in \mathcal{I} \vee j \in \mathcal{I} \implies \exists k \in \mathcal{I} : i \prec k, j \prec k$$

and assume it has a *cofinal* sequence $(i_n)_{n\in\mathbb{N}} \subseteq \mathcal{I}$, i.e.

$$\forall i \in \mathcal{I} : \exists n \in \mathbb{N} : i \prec i_n.$$

Repetitive structure

Let $(\mathcal{I}, \prec)$ be a *partially ordered and directed* index set, i.e.

$$i \in \mathcal{I} \vee j \in \mathcal{I} \implies \exists k \in \mathcal{I} : i \prec k, j \prec k$$

and assume it has a *cofinal* sequence $(i_n)_{n \in \mathbb{N}} \subseteq \mathcal{I}$, i.e.

$$\forall i \in \mathcal{I} : \exists n \in \mathbb{N} : i \prec i_n.$$

Further, assume that with every $i \in \mathcal{I}$ we have a *sample space* $\mathcal{X}_i$ and a system P_{ij} for all $i \prec j$ where P_{ij} is a Markov kernel from $\mathcal{X}_j$ to $\mathcal{X}_i$ that is *consistent*, meaning that if $i \prec j \prec k$ it holds that

$$P_{ik} = P_{ij} \circledast P_{jk}.$$

Repetitive structure

Let $(\mathcal{I}, \prec)$ be a *partially ordered and directed* index set, i.e.

$$i \in \mathcal{I} \vee j \in \mathcal{I} \implies \exists k \in \mathcal{I} : i \prec k, j \prec k$$

and assume it has a *cofinal* sequence $(i_n)_{n \in \mathbb{N}} \subseteq \mathcal{I}$, i.e.

$$\forall i \in \mathcal{I} : \exists n \in \mathbb{N} : i \prec i_n.$$

Further, assume that with every $i \in \mathcal{I}$ we have a *sample space* $\mathcal{X}_i$ and a system P_{ij} for all $i \prec j$ where P_{ij} is a Markov kernel from $\mathcal{X}_j$ to $\mathcal{X}_i$ that is *consistent*, meaning that if $i \prec j \prec k$ it holds that

$$P_{ik} = P_{ij} \circledast P_{jk}.$$

We shall term these *Markov projections.* and the combination of the above items is a *repetitive structure*; see also Martin-Löf (1974) and Lauritzen (1988) for details.

Examples

For example, $\mathcal{I}$ could be the set of finite subsets of the integers and for $A \in \mathcal{I}$ let $\mathcal{X}_A = \mathcal{X}^{\otimes A}$ be the product space.

Then let P_{AB} simply be the coordinate projection: for $x_B = (x_\beta)_{\beta \in B}$ let

$$P_{AB}(x_B, \cdot) = \delta_{x_A}$$

where $x_A = (x_\beta)_{\beta \in A}$.

Examples

For example, $\mathcal{I}$ could be the set of finite subsets of the integers and for $A \in \mathcal{I}$ let $\mathcal{X}_A = \mathcal{X}^{\otimes A}$ be the product space.

Then let P_{AB} simply be the coordinate projection: for $x_B = (x_\beta)_{\beta \in B}$ let

$$P_{AB}(x_B, \cdot) = \delta_{x_A}$$

where $x_A = (x_\beta)_{\beta \in A}$.

Or — slightly more sophisticated — let $P_{AB}(x_B, \cdot)$ be the random distribution over $\mathcal{X}_A$ obtained by taking a random subsample of size $|A|$ of elements in x_B.

Projective limits

We shall be interested in the *projective limit*

$$\mathcal{X} = \operatorname*{proj\,lim}_{i \in \mathcal{I}} \mathcal{X}_i$$

which in our case can be identified as the *extreme points* of the set of consistent sequences of probability measures

$$\operatorname*{proj\,lim}_{i \in \mathcal{I}} \mathcal{X}_i = \{\mu = (\mu_i)_{i \in \mathcal{I}} \mid \mu_i = P_{ij}\mu_j, i \prec j\}.$$

Projective limits

We shall be interested in the *projective limit*

$$\mathcal{X} = \operatorname*{proj\,lim}_{i \in \mathcal{I}} \mathcal{X}_i$$

which in our case can be identified as the *extreme points* of the set of consistent sequences of probability measures

$$\operatorname*{proj\,lim}_{i \in \mathcal{I}} \mathcal{X}_i = \{\mu = (\mu_i)_{i \in \mathcal{I}} \,|\, \mu_i = P_{ij}\mu_j, i \prec j\}.$$

If the Markov projections $P_{ij} = \pi_{ij}$ are coordinate projections, it may be identified with the set of *consistent sequences*

$$\mathcal{X} = \operatorname*{proj\,lim}_{i \in \mathcal{I}} \mathcal{X}_i = \{x = (x_i)_{i \in \mathcal{I}} \,|\, x_i = \pi_{ij}(x_j), i \prec j\}$$

and then there is a unique system of projections $\pi_i : \mathcal{X} \mapsto \mathcal{X}_i, i \in \mathcal{I}$ such that $\pi_{ij} \circ \pi_j = \pi_i$ for $i < j$.

Consistency theorem

Theorem (Kolmogorov)

Let $(\mathcal{X}_i, i \in \mathcal{I}, (\pi_{ij})_{i<j})$ *be a projective system as above,* $\mathcal{X}$ *the projective limit, and* $\mu = (\mu_i)_{i \in \mathcal{I}}$ *a consistent system of probability measures.*

Then there is a unique probability measure $\tilde{\mu}$ *on* $\mathcal{X}$ *such that* $\mu_i = \pi_i(\mu)$ *for all* $i \in \mathcal{I}$*.*

Consistency theorem

Theorem (Kolmogorov)

Let $(\mathcal{X}_i, i \in \mathcal{I}, (\pi_{ij})_{i<j})$ be a projective system as above, $\mathcal{X}$ the projective limit, and $\mu = (\mu_i)_{i\in\mathcal{I}}$ a consistent system of probability measures.

Then there is a unique probability measure $\tilde{\mu}$ on $\mathcal{X}$ such that $\mu_i = \pi_i(\mu)$ for all $i \in \mathcal{I}$.

Clearly, this is only true under regularity conditions (Borel spaces etc.) as the theorem above subsumes the standard version of Kolmogorov's consistency theorem.

Consistency theorem

Theorem (Kolmogorov)

Let $(\mathcal{X}_i, i \in \mathcal{I}, (\pi_{ij})_{i<j})$ be a projective system as above, $\mathcal{X}$ the projective limit, and $\mu = (\mu_i)_{i \in \mathcal{I}}$ a consistent system of probability measures.

Then there is a unique probability measure $\tilde{\mu}$ on $\mathcal{X}$ such that $\mu_i = \pi_i(\mu)$ for all $i \in \mathcal{I}$.

Clearly, this is only true under regularity conditions (Borel spaces etc.) as the theorem above subsumes the standard version of Kolmogorov's consistency theorem.

We shall not distinguish between the sequence μ and the distribution $\tilde{\mu}$ it defines.

Framework for general de Finetti theorems

Consider now a projective system $((\mathcal{X}_i)_{i\in\mathcal{I}}, (\pi_{ij})_{i\prec j})$ as above and a similar system of 'sufficient' or 'summarizing' statistics

$$t_i :\mapsto \mathcal{Y}_i, i \in \mathcal{I}$$

with Markov kernels $Q_i : \mathcal{Y}_i \to \mathcal{X}_i$ representing *potential conditional distributions* of $X_i = \pi_i(X)$ given $Y_i = t_i(X_i)$.

Framework for general de Finetti theorems

Consider now a projective system $((\mathcal{X}_i)_{i\in\mathcal{I}}, (\pi_{ij})_{i\prec j})$ as above and a similar system of 'sufficient' or 'summarizing' statistics

$$t_i :\mapsto \mathcal{Y}_i, i \in \mathcal{I}$$

with Markov kernels $Q_i : \mathcal{Y}_i \to \mathcal{X}_i$ representing *potential conditional distributions* of $X_i = \pi_i(X)$ given $Y_i = t_i(X_i)$.

Define Markov kernels $Q_{ij} : \mathcal{Y}_j \to \mathcal{Y}_i$ for $i \prec j$ as

$$Q_{ij} = t_i(\pi_{ij}(Q_j)).$$

Framework for general de Finetti theorems

Consider now a projective system $((\mathcal{X}_i)_{i\in\mathcal{I}}, (\pi_{ij})_{i\prec j})$ as above and a similar system of 'sufficient' or 'summarizing' statistics

$$t_i :\mapsto \mathcal{Y}_i, i \in \mathcal{I}$$

with Markov kernels $Q_i : \mathcal{Y}_i \to \mathcal{X}_i$ representing *potential conditional distributions* of $X_i = \pi_i(X)$ given $Y_i = t_i(X_i)$.

Define Markov kernels $Q_{ij} : \mathcal{Y}_j \to \mathcal{Y}_i$ for $i \prec j$ as

$$Q_{ij} = t_i(\pi_{ij}(Q_j)).$$

Then Q_{ij} represents conditional distribution of Y_i given Y_j.

Framework for general de Finetti theorems

Consider now a projective system $((\mathcal{X}_i)_{i\in\mathcal{I}}, (\pi_{ij})_{i\prec j})$ as above and a similar system of 'sufficient' or 'summarizing' statistics

$$t_i :\mapsto \mathcal{Y}_i, i \in \mathcal{I}$$

with Markov kernels $Q_i : \mathcal{Y}_i \to \mathcal{X}_i$ representing *potential conditional distributions* of $X_i = \pi_i(X)$ given $Y_i = t_i(X_i)$.

Define Markov kernels $Q_{ij} : \mathcal{Y}_j \to \mathcal{Y}_i$ for $i \prec j$ as

$$Q_{ij} = t_i(\pi_{ij}(Q_j)).$$

Then Q_{ij} represents conditional distribution of Y_i given Y_j.

Say that the system of statistics is *consistent* if for $i \prec j \prec k$

$$Q_{ij} \circledast Q_{jk} = Q_{ik}.$$

Framework for general de Finetti theorems

Consider now a projective system $((\mathcal{X}_i)_{i\in\mathcal{I}}, (\pi_{ij})_{i\prec j})$ as above and a similar system of 'sufficient' or 'summarizing' statistics

$$t_i :\mapsto \mathcal{Y}_i, i \in \mathcal{I}$$

with Markov kernels $Q_i : \mathcal{Y}_i \to \mathcal{X}_i$ representing *potential conditional distributions* of $X_i = \pi_i(X)$ given $Y_i = t_i(X_i)$.

Define Markov kernels $Q_{ij} : \mathcal{Y}_j \to \mathcal{Y}_i$ for $i \prec j$ as

$$Q_{ij} = t_i(\pi_{ij}(Q_j)).$$

Then Q_{ij} represents conditional distribution of Y_i given Y_j.

Say that the system of statistics is *consistent* if for $i \prec j \prec k$

$$Q_{ij} \circledast Q_{jk} = Q_{ik}.$$

Fulfilled for suitable summarizing statistics, not necessarily recursively computable. For example maximal RCE invariants.

de Finetti's theorem

Theorem (Lauritzen (1988))

Let $(\mathcal{X}_i, i \in \mathcal{I}, (\pi_{ij})_{i<j})$ *be a projective system as above with* $\mathcal{X}$ *the projective limit and let* $t_i : \mathcal{X}_i \mapsto \mathcal{Y}_i, i \in \mathcal{I}$ *a system of statistics with associated consistent Markov kernels* Q_i *and* Q_{ij} *as above.*

de Finetti's theorem

Theorem (Lauritzen (1988))

Let $(\mathcal{X}_i, i \in \mathcal{I}, (\pi_{ij})_{i<j})$ *be a projective system as above with* $\mathcal{X}$ *the projective limit and let* $t_i : \mathcal{X}_i \mapsto \mathcal{Y}_i, i \in \mathcal{I}$ *a system of statistics with associated consistent Markov kernels* Q_i *and* Q_{ij} *as above.*

The set M *of systems* $\mu = (\mu_i)_{i\in\mathcal{I}}$ *of probabiities on* $\mathcal{Y}_i$ *which are consistent w.r.t.* Q_{ij} *form a convex simplex.*

de Finetti's theorem

Theorem (Lauritzen (1988))

Let $(\mathcal{X}_i, i \in \mathcal{I}, (\pi_{ij})_{i<j})$ be a projective system as above with $\mathcal{X}$ the projective limit and let $t_i : \mathcal{X}_i \mapsto \mathcal{Y}_i, i \in \mathcal{I}$ a system of statistics with associated consistent Markov kernels Q_i and Q_{ij} as above.

The set M of systems $\mu = (\mu_i)_{i \in \mathcal{I}}$ of probabiities on $\mathcal{Y}_i$ which are consistent w.r.t. Q_{ij} form a convex simplex.

Each such system induces a consistent system $\nu = (\nu_i) = (Q_i \mu_i)$ of probabilty measures on $\mathcal{X}_i$ and thus a probability measure $\tilde{\nu}$ on $\mathcal{X}$ as previously.

de Finetti's theorem – continued

Theorem (Lauritzen (1988))

Q_i are valid conditional distributions for $X_i = \pi_i(X)$ given $Y_i = t_i(X_i)$ w.r.t. $\tilde{\nu}$.

de Finetti's theorem – continued

Theorem (Lauritzen (1988))

Q_i are valid conditional distributions for $X_i = \pi_i(X)$ given $Y_i = t_i(X_i)$ w.r.t. $\tilde{\nu}$.

Every element $\mu \in M$ is the barycentre of a unique probability measure π on extreme points $\mathcal{E}$ of M.

de Finetti's theorem – continued

Theorem (Lauritzen (1988))

Q_i are valid conditional distributions for $X_i = \pi_i(X)$ given $Y_i = t_i(X_i)$ w.r.t. $\tilde{\nu}$.

Every element $\mu \in M$ is the barycentre of a unique probability measure π on extreme points $\mathcal{E}$ of M.

Any $\mu \in M$ is an extreme point if and only if the tail of $X_i, i \in \mathcal{I}$ is trivial wrt $\tilde{\nu}$.

de Finetti's theorem – continued

Theorem (Lauritzen (1988))

Q_i are valid conditional distributions for $X_i = \pi_i(X)$ given $Y_i = t_i(X_i)$ w.r.t. $\tilde{\nu}$.

Every element $\mu \in M$ is the barycentre of a unique probability measure π on extreme points $\mathcal{E}$ of M.

Any $\mu \in M$ is an extreme point if and only if the tail of $X_i, i \in \mathcal{I}$ is trivial wrt $\tilde{\nu}$.

For every $\mu \in M$ and $i \in \mathcal{I}$ it holds that

$$\lim_{j\to\infty} Q_{ij}(\cdot \mid Y_j) = \gamma_i(\cdot) a.s.\ \tilde{\nu},$$

where $\gamma = (\gamma_i)_{i\in\mathcal{I}} \in \mathcal{E}$.

de Finetti's theorem – continued

Theorem (Lauritzen (1988))

Q_i are valid conditional distributions for $X_i = \pi_i(X)$ given $Y_i = t_i(X_i)$ w.r.t. $\tilde{\nu}$.

Every element $\mu \in M$ is the barycentre of a unique probability measure π on extreme points $\mathcal{E}$ of M.

Any $\mu \in M$ is an extreme point if and only if the tail of $X_i, i \in \mathcal{I}$ is trivial wrt $\tilde{\nu}$.

For every $\mu \in M$ and $i \in \mathcal{I}$ it holds that

$$\lim_{j \to \infty} Q_{ij}(\cdot \mid Y_j) = \gamma_i(\cdot) a.s.\ \tilde{\nu},$$

where $\gamma = (\gamma_i)_{i \in \mathcal{I}} \in \mathcal{E}$.

Finally, $\mu \in \mathcal{E}$ if and only if $\gamma = \mu$ almost surely.

Random graphs

Positive definite functions on semigroups

Let $(S, +)$ be an Abelian semigroup with neutral element 0.

Positive definite functions on semigroups

Let $(S, +)$ be an Abelian semigroup with neutral element 0.

Recall that a *character* is a function $\rho : S \to \mathbb{R}$ satisfying

$$\rho(s+t) = \rho(s)\rho(t), \quad \rho(0) = 1.$$

The set of characters form itself a semigroup under multiplication. This is the *dual* semigroup $(\hat{S}, \cdot)$.

Positive definite functions on semigroups

Let $(S, +)$ be an Abelian semigroup with neutral element 0.

Recall that a *character* is a function $\rho : S \to \mathbb{R}$ satisfying

$$\rho(s+t) = \rho(s)\rho(t), \quad \rho(0) = 1.$$

The set of characters form itself a semigroup under multiplication. This is the *dual* semigroup $(\hat{S}, \cdot)$.

A function $\phi : S \to \mathbb{R}$ is *positive definite* if and only if

$$\sum_{j,k=1}^{n} c_j c_k \phi(s_j + s_k) \geqslant 0, \quad \forall n \in \mathbb{N}, c_j \in \mathbb{R}, s_j \in S;$$

Positive definite functions on semigroups

Let $(S, +)$ be an Abelian semigroup with neutral element 0.

Recall that a *character* is a function $\rho : S \to \mathbb{R}$ satisfying

$$\rho(s+t) = \rho(s)\rho(t), \quad \rho(0) = 1.$$

The set of characters form itself a semigroup under multiplication. This is the *dual* semigroup $(\hat{S}, \cdot)$.

A function $\phi : S \to \mathbb{R}$ is *positive definite* if and only if

$$\sum_{j,k=1}^{n} c_j c_k \phi(s_j + s_k) \geqslant 0, \quad \forall n \in \mathbb{N}, c_j \in \mathbb{R}, s_j \in S;$$

in other words, a function is positive definite if and only if all matrices of the form $\{m_{jk} = \phi(s_j + s_k)\}$ are positive semidefinite.

Integral representation

We shall be using the following result from Berg et al. (1976).

Theorem (Berg, Christensen, Ressel)

The set of normalized positive definite functions $\mathcal{P}_1^b(S)$ is a simplex with the bounded characters as extreme points. In particular, any $\phi \in \mathcal{P}_1^b(S)$ has a unique representation as barycentre of a probability measure μ on $\hat{S}$:

$$\phi(s) = \int_{\hat{S}} \rho(s)\, \mathrm{d}\mu(\rho)$$

and the bounded characters $\hat{S}$ form a closed subset of $\mathcal{P}_1^b(S)$.

Graphs

For $A \subseteq \mathbb{N}$, $\mathcal{L}_A$ denotes the set of simple labeled (undirected) graphs with labels in A.

Graphs

For $A \subseteq \mathbb{N}$, $\mathcal{L}_A$ denotes the set of simple labeled (undirected) graphs with labels in A.

For $A \subseteq B$ and $G = (B, E)$, $\pi_{AB}(G) = G_A = (A, E_A)$ is the subgraph induced by A.

Graphs

For $A \subseteq \mathbb{N}$, $\mathcal{L}_A$ denotes the set of simple labeled (undirected) graphs with labels in A.

For $A \subseteq B$ and $G = (B, E)$, $\pi_{AB}(G) = G_A = (A, E_A)$ is the subgraph induced by A.

Further, let $\mathcal{U}_n$ denote the set of simple unlabeled graphs with n vertices.

Graphs

For $A \subseteq \mathbb{N}$, $\mathcal{L}_A$ denotes the set of simple labeled (undirected) graphs with labels in A.

For $A \subseteq B$ and $G = (B, E)$, $\pi_{AB}(G) = G_A = (A, E_A)$ is the subgraph induced by A.

Further, let $\mathcal{U}_n$ denote the set of simple unlabeled graphs with n vertices.

We also let $\mathcal{L} = \cup_{A \subseteq \mathbb{N}} \mathcal{L}_A$ and similarly for $\mathcal{U}$.

Graphs

For $A \subseteq \mathbb{N}$, $\mathcal{L}_A$ denotes the set of simple labeled (undirected) graphs with labels in A.

For $A \subseteq B$ and $G = (B, E)$, $\pi_{AB}(G) = G_A = (A, E_A)$ is the subgraph induced by A.

Further, let $\mathcal{U}_n$ denote the set of simple unlabeled graphs with n vertices.

We also let $\mathcal{L} = \cup_{A \subseteq \mathbb{N}} \mathcal{L}_A$ and similarly for $\mathcal{U}$.

For G and G' in $\mathcal{L}_A$, $G \sim G'$ signifies that they are isomorphic and $[G]$ is equivalence class of graphs isomorphic to G.

Graphs

For $A \subseteq \mathbb{N}$, $\mathcal{L}_A$ denotes the set of simple labeled (undirected) graphs with labels in A.

For $A \subseteq B$ and $G = (B, E)$, $\pi_{AB}(G) = G_A = (A, E_A)$ is the subgraph induced by A.

Further, let $\mathcal{U}_n$ denote the set of simple unlabeled graphs with n vertices.

We also let $\mathcal{L} = \cup_{A \subseteq \mathbb{N}} \mathcal{L}_A$ and similarly for $\mathcal{U}$.

For G and G' in $\mathcal{L}_A$, $G \sim G'$ signifies that they are isomorphic and $[G]$ is equivalence class of graphs isomorphic to G.

We may without ambiguity think of $[G]$ as an unlabeled graph.

Graphs

For $A \subseteq \mathbb{N}$, $\mathcal{L}_A$ denotes the set of simple labeled (undirected) graphs with labels in A.

For $A \subseteq B$ and $G = (B, E)$, $\pi_{AB}(G) = G_A = (A, E_A)$ is the subgraph induced by A.

Further, let $\mathcal{U}_n$ denote the set of simple unlabeled graphs with n vertices.

We also let $\mathcal{L} = \cup_{A \subseteq \mathbb{N}} \mathcal{L}_A$ and similarly for $\mathcal{U}$.

For G and G' in $\mathcal{L}_A$, $G \sim G'$ signifies that they are isomorphic and $[G]$ is equivalence class of graphs isomorphic to G.

We may without ambiguity think of $[G]$ as an unlabeled graph.

For $A \subseteq \mathbb{N}$, $G \in \mathcal{L}_A$, and $\sigma \in \mathcal{S}(A)$, the group of permutations of elements of A, we will let G_σ be the graph obtained from G by relabeling its nodes according to σ.

Symmetric random graphs

We consider a probability distribution P on $\mathcal{L}_A$ and say this is *symmetric* or *exchangeable* if and only if

$$P\{G = H\} = P\{G = H_\sigma\}, \quad \forall H \in \mathcal{L}_A, \sigma \in S(A).$$

Symmetric random graphs

We consider a probability distribution P on $\mathcal{L}_A$ and say this is *symmetric* or *exchangeable* if and only if

$$P\{G = H\} = P\{G = H_\sigma\}, \quad \forall H \in \mathcal{L}_A, \sigma \in S(A).$$

It is practical to represent the distribution P through its *Möbius parameters*

$$Z(F) = P(G : F \subseteq G).$$

Symmetric random graphs

We consider a probability distribution P on $\mathcal{L}_A$ and say this is *symmetric* or *exchangeable* if and only if

$$P\{G = H\} = P\{G = H_\sigma\}, \quad \forall H \in \mathcal{L}_A, \sigma \in S(A).$$

It is practical to represent the distribution P through its *Möbius parameters*

$$Z(F) = P(G : F \subseteq G).$$

Clearly, P is symmetric if and only if Z is symmetric, or, equivalently, if there is a function $\phi : \mathcal{U} \to \mathbb{R}$ such that

$$Z(F) = \phi([F]) \tag{1}$$

or, in other words the map t_A

$$t_A(G) = [G]$$

is summarizing.

Positive definiteness of Möbius transform

We also note that $(\mathcal{U}, +)$ — where $U + V = U \cup V$ is (node disjoint) graph union — is an Abelian semigroup with the empty graph as its neutral element.

Positive definiteness of Möbius transform

We also note that $(\mathcal{U}, +)$ — where $U + V = U \cup V$ is (node disjoint) graph union — is an Abelian semigroup with the empty graph as its neutral element.

Clearly $\phi(\emptyset) = 1$ and ϕ is bounded.

Positive definiteness of Möbius transform

We also note that $(\mathcal{U}, +)$ — where $U + V = U \cup V$ is (node disjoint) graph union — is an Abelian semigroup with the empty graph as its neutral element.

Clearly $\phi(\emptyset) = 1$ and ϕ is bounded. Moreover we have

Lemma

Let G be a symmetric random graph with Möbius parameter Z given as above. Then the function ϕ is bounded and positive definite on $(\mathcal{U}, +)$; in other words, $\phi \in \mathcal{P}_1^b(\mathcal{U})$.

de Finetti's theorem for graphs

Corollary

Let P be the distribution of a random graph with Möbius parameter Z. Then P is symmetric if and only if there is a unique probability measure μ on $\hat{\mathcal{U}}$ such that for all $F \in \mathcal{L}$

$$Z(F) = \int_{\hat{\mathcal{U}}} \rho([F]) \, \mathrm{d}\mu(\rho).$$

de Finetti's theorem for graphs

Corollary

Let P be the distribution of a random graph with Möbius parameter Z. Then P is symmetric if and only if there is a unique probability measure μ on $\hat{\mathcal{U}}$ such that for all $F \in \mathcal{L}$

$$Z(F) = \int_{\hat{\mathcal{U}}} \rho([F])\,\mathrm{d}\mu(\rho).$$

Note that symmetric measures corresponding to the pure characters are *dissociated* (Silverman, 1976): if $F = F_1 \cup F_2$ and F_1 and F_2 are disjoint subgraphs of F it holds that

$$Z(F) = \rho([F]) = \rho([F_1])\rho([F_2]) = Z(F_1)Z(F_2)$$

or, in other words, we have for node-disjoint F_1, F_2:

$$P_\rho(F_1 \cup F_2 \subseteq G) = P_\rho(F_1 \subseteq G)P_\rho(F_2 \subseteq G).$$

Graphons as characters

Theorem

The bounded characters $\hat{\mathcal{U}}$ *on* $(\mathcal{U}, +)$ *are exactly the functions* ρ *satisfying for* $F \in \mathcal{L}_n^*$

$$\rho([F]) = \int_{[0,1]^n} \prod_{ij: i \sim j \in F} W(u_i, u_j)\, \mathrm{d}u$$

for some measurable, symmetric function $W : [0,1]^2 \rightarrow [0,1]$. *The function* W *is unique up to measure-preserving transformations of the unit interval.*

Graphons as characters

Theorem

The bounded characters $\hat{\mathcal{U}}$ *on* $(\mathcal{U}, +)$ *are exactly the functions* ρ *satisfying for* $F \in \mathcal{L}_n^*$

$$\rho([F]) = \int_{[0,1]^n} \prod_{ij:i\sim j\in F} W(u_i, u_j)\,\mathrm{d}u$$

for some measurable, symmetric function $W : [0,1]^2 \to [0,1]$. *The function* W *is unique up to measure-preserving transformations of the unit interval.*

The equivalence class of functions W is known as a *graphon* (Lovász and Szegedy, 2006) or *limit* of an unlabeled graph. For further details, see Lauritzen (2020).

Final summary

- de Finetti theorems form a bridge between Fisherian and Bayesian models;

Final summary

- de Finetti theorems form a bridge between Fisherian and Bayesian models;
- Unifying concept is a *repetitive structure* that leads to parameter being defined as limit of observables;

Final summary

- de Finetti theorems form a bridge between Fisherian and Bayesian models;
- Unifying concept is a *repetitive structure* that leads to parameter being defined as limit of observables;
- Conditioning on tail field provides the Fisherian model 'inside' the Bayesian.

Final summary

- de Finetti theorems form a bridge between Fisherian and Bayesian models;
- Unifying concept is a *repetitive structure* that leads to parameter being defined as limit of observables;
- Conditioning on tail field provides the Fisherian model 'inside' the Bayesian.
- Examples include semigroup exponential families, arrays, random graphs.

Final summary

- de Finetti theorems form a bridge between Fisherian and Bayesian models;
- Unifying concept is a *repetitive structure* that leads to parameter being defined as limit of observables;
- Conditioning on tail field provides the Fisherian model 'inside' the Bayesian.
- Examples include semigroup exponential families, arrays, random graphs.
- Still far too few examples that are properly understood.

References

Aldous, D. (1981). Representations for partially exchangeable random variables. *Journal of Multivariate Analysis*, 11:581–598.

Berg, C., Christensen, J. P. R., and Ressel, P. (1976). Positive definite functions on Abelian semigroups. *Mathematische Annalen*, 259:253–274.

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.

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. *The 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. (2020). Harmonic analysis of symmetric random graphs. *Kybernetika*, 56:1081–1089.

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

Lovász, L. and Szegedy, B. (2006). Limits of dense graph sequences. *Journal of Combinatorial Theory, Series B*, 96(6):933–957.

Martin-Löf, P. (1974). Repetitive structures. In Barndorff-Nielsen, O., Blæsild, P., and Schou, G., editors, *Proceedings of Conference on Foundational Questions in Statistical Inference, Aarhus 1973*, pages 271–293, Aarhus, Denmark. Department of Theoretical Statistics, Aarhus University.

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

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*, volume 1 of *Studies in Mathematical Psychology*. Danmarks Pædagogiske Institut, Copenhagen.

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

Silverman, B. (1976). Limit theorems for dissociated random variables. *Advances of Applied Probability*, 8:806–819.