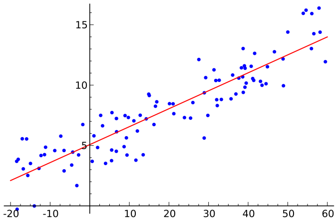
Exchangeability, symmetry and sufficiency

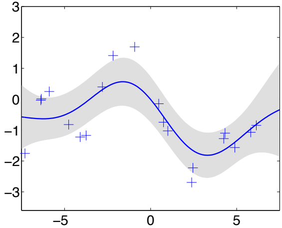
Peter Orbanz

Parameters and Patterns

Parameterized models

$$P[X \in \ .\ |\Theta] \qquad = \qquad \text{Probability[data|pattern]}$$





Inference idea

data = independent “randomness” + underlying pattern

OVERVIEW

General theme

P satisfies invariance property, then

$$\mathbb{P}(X \in \,.\,) = \int_{\mathcal{T}} P[X \in \,.\,|\theta] Q(d\theta)$$

Types of invariance:

1. Symmetry (invariance under group action)
2. Existence of sufficient statistics

Relevance for Bayesian statistics

- Does a random parameter exist?
- When does the likelihood factorize?

$$Q[d\theta | X_1 = x_1, \dots, X_n = x_n] = \frac{\prod_{i=1}^{n} p(x_i|\theta)}{p(x_1, \dots, x_n)} Q(d\theta)$$

For *sequences* of random variables, the answer is given by de Finetti's theorem.

Exchangeability

Random sequence $X_1, X_2, \ldots$ *exchangeable*

$$\mathbb{P}(X_1, X_2, \ldots \in A) = \mathbb{P}(X_{\pi(1)}, X_{\pi(2)}, \ldots \in A) \qquad \text{for every permutation } \pi \in \mathbb{S}_\infty \ .$$

Theorem (de Finetti)

A sequence is exchangeable if and only if it is a mixture of iid sequences, that is:

$$\mathbb{P}(X_1, X_2, \ldots) = \int_{M(\mathcal{X})} \Big(\prod_{n=1}^{\infty} P(X_n) \Big) Q(dP) = \int_{\mathcal{T}} \Big(\prod_{n=1}^{\infty} P[X_n | \Theta = \theta] \Big) Q_{\mathcal{T}}(d\theta)$$

- Θ random measure (since $\Theta(\omega) \in M(\mathcal{X})$)
- Convergence: $F_n \xrightarrow{n \to \infty} \theta$

Modification 1: Rotation invariance

Rotatable sequence

$$\mathbb{P}_n(X_1,\dots,X_n) = \mathbb{P}_n(R_n(X_1,\dots,X_n)) \quad \text{for all} \quad R_n \in O(n)$$

Infinite case

$$X_1, X_2, \dots \text{ rotatable} \quad :\Leftrightarrow \quad X_1,\dots,X_n \text{ rotatable for all } n$$

Theorem (Freedman; Dawid)

Infinite sequence rotatable iff

$$\mathbb{P}(X_1, X_2, \dots) = \int_{\mathbb{R}_+} \Big(\prod_{n=1}^{\infty} N_\sigma(X_n)\Big) Q_{\mathbb{R}_+}(d\sigma) \; .$$

N_σ denotes $(0,\sigma)$-Gaussian.

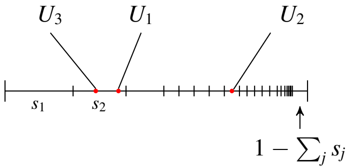
MODIFICATION 2: RANDOM PARTITIONS

Random partition of $\mathbb{N}$

$$\Pi = \{B_1, B_2, \ldots\} \quad \text{e.g.} \quad \{\{1,3,5,\ldots\},\{2,4\},\{10\},\ldots\}$$

Paint-box distribution

- Weights $s_1, s_2, \ldots \geq 0$ with $\sum s_j \leq 1$
- $U_1, U_2, \ldots \sim \text{Uniform}[0,1]$



Sampling $\Pi \sim \beta[\,.\,|\mathbf{s}]$:

$$i, j \in \mathbb{N} \text{ in same block} \quad \Leftrightarrow \quad U_i, U_j \text{ in same interval}$$

$$\{i\} \text{ separate block} \quad \Leftrightarrow \quad U_i \text{ in interval } 1 - \sum s_j$$

Theorem (Kingman)

$$\Pi \text{ exchangeable} \quad \Leftrightarrow \quad \mathbb{P}(\Pi \in \,.\,) = \int \beta[\Pi \in \,.\,|\mathbf{s}]Q(d\mathbf{s})$$

Ergodic Decomposition Theorem (Varadarajan)

- G nice group on space $\mathcal{X}$
- Call measure μ ergodic if $\mu(A) \in \{0, 1\}$ for all G-invariant sets A.
- $\mathcal{E} := \{\text{ergodic probability measures}\}$

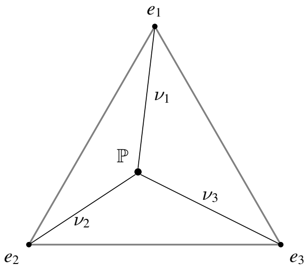
$$\mathbb{P} \in M(\mathcal{X}) \quad G\text{-invariant} \quad \Leftrightarrow \quad \mathbb{P}(A) = \int_{\mathcal{E}} e(A)\nu(de) \quad \text{for unique } \nu \in M(\mathcal{E})$$

Special case: De Finetti

- $G = \mathbb{S}_\infty$ and $\mathcal{Y} = \mathcal{X}^\infty$
- G-invariant sets = exchangeable events
- $\mathcal{E}$ = factorial distributions (“Hewitt-Savage 0-1 law”)

Finite case

$$\mathbb{P} = \sum_{e_i \in \mathcal{E}} \nu_i e_i$$



- $\mathcal{E} = \{e_1, e_2, e_3\}$
- (ν_1, ν_2, ν_3) barycentric coordinates

Infinite/continuous case

$$\mathbb{P}(\,.\,) = \int_{\mathcal{E}} e(\,.\,)\nu(de) = \int_{\mathcal{T}} \mathbf{p}(\,.\,,\theta)\nu_{\mathcal{T}}(d\theta)$$

- $\mathbf{p} : \mathcal{T} \to \mathcal{E} \subset M(\mathcal{X})$ Markov kernel
- $\mathbf{p}$ is random measure with values $\mathbf{p}(\,.\,,\theta) \in \mathcal{E}$
- de Finetti case: $\mathbf{p}(\,.\,,\theta) = \prod_{n \in \mathbb{N}} Q[\,.\,|\Theta = \theta]$ and $\mathcal{T} = M(\mathcal{X})$

SUFFICIENT STATISTICS

Recall definition

Given: Parameterized model $\mathcal{P} = \{P[\,.\,|\Theta = \theta], \theta \in \mathcal{T}\}$.

A statistic $S : \mathcal{X} \to \mathbf{S}$ is *sufficient* for a set $\mathcal{P}$ of distributions if all $P \in \mathcal{P}$ have identical conditional probability given S:

$$P[\,.\,|\Theta = \theta, S = s] \quad = \quad P[\,.\,|S = s] \qquad \text{for all } \theta \;.$$

Intuitively: S contains all relevant information in X for estimation of Θ.

Sufficiency and symmetry

Both concepts formalize which aspects of data carry no information on parameter.

Notation

$$\begin{aligned}
\mathcal{P}_{S,\mathbf{v}} &= \{ \text{ largest set of measures for which } (S, \mathbf{v}) \text{ is sufficient } \} \\
\mathcal{E}_{S,\mathbf{v}} &= \{ P \in \mathcal{P}_{S,\mathbf{v}} \,|\, P(A) \in \{0,1\} \text{ for all } A \in \tau(S) \} \\
\Delta_{S,\mathbf{v}} &= \{ P \in \mathcal{P}_{S,\mathbf{v}} \,|\, S(P) = \delta_s \text{ for some } s \in \mathbf{S} \}
\end{aligned}$$

Theorem (Lauritzen)

Let $S : \mathcal{X} \to \mathbf{S}$ be a statistic and $\mathbf{v} : \mathbf{S} \to M(\mathcal{X})$ a Markov kernel. Then $\mathcal{P}_{S,\mathbf{v}}$ is a convex set and

$$\text{ex}(\mathcal{P}_{S,\mathbf{v}}) = \mathcal{E}_{S,\mathbf{v}} = \Delta_{S,\mathbf{v}} \ .$$

If $\mathcal{P}_{S,\mathbf{v}}$ is weakly closed: For every $P \in \mathcal{P}_{S,\mathbf{v}}$ exists unique $\nu \in M(\mathcal{E}_{S,\mathbf{v}})$ with

$$P = \int_{M(\mathcal{E}_{S,\mathbf{v}})} e(\,.\,)\nu(de) = \int_{\mathbf{S}} \mathbf{v}(\,.\,,s)Q(ds) \ .$$

Induced operator

$$(T_{\mathbf{v}}^{*}P)(A) := \int_{\mathcal{X}} \mathbf{v}(A, S(x))P(dx)$$

Technically speaking: $T_{\mathbf{v}}^{*}$ is the norm adjoint of the Markov operator induced by $\mathbf{v}$.

Fact 1

Probability measure P is $T_{\mathbf{v}}^{*}$ invariant $\quad\Leftrightarrow\quad$ $\mathbf{v}$ is conditional probability of P given S

Fact 2 (standard result in functional analysis)

The set $\mathcal{P}_{\mathrm{S},\mathbf{v}}$ of $T_{\mathbf{v}}^{*}$-invariant measures is a convex set. Its extreme points are the $T_{\mathbf{v}}^{*}$-ergodic measures.

Fact 3 (von Weizsäcker and Winkler)

If $\mathcal{P}_{\mathrm{S},\mathbf{v}}$ is weakly closed, then for every $P \in \mathcal{P}_{\mathrm{S},\mathbf{v}}$, there exists a measure Q such that

$$P = \int_{\mathbf{S}} \mathbf{v}(\,.\,, s)Q(ds)\ .$$

EXAMPLES

de Finetti's theorem

$$\mathbb{P} \text{ exchangeable} \quad \Leftrightarrow \quad S_n(x_1,\ldots,x_n) = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_n} \text{ sufficient}$$

Rotation invariance

$$\mathbb{P} \text{ rotatable} \quad \Leftrightarrow \quad S_n(x_1,\ldots,x_n) = \|(x_1,\ldots,x_n)\|_2 \text{ sufficient}$$

Kingman's theorem

$$\Pi \text{ exchangeable} \quad \Leftrightarrow \quad \text{asymptotic block sizes are sufficient statistic}$$

Exponential families (Küchler and Lauritzen)

Choose $\mathcal{X} = \mathbb{R}^\infty$. Under suitable regularity conditions:

$$S_n \quad \text{additive, i.e.} \quad S_n(x_1,\ldots,x_n) = \frac{1}{n}\sum_{i=1}^{n} S_0(x_i) \quad \Leftrightarrow \quad \mathcal{E}_{\mathrm{S},\mathbf{v}} \text{ is exponential family}$$

Exchangeability: Random Graphs

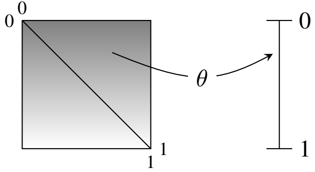
Random graph with independent edges

Given: $\theta : [0,1]^2 \to [0,1]$ symmetric function

- $U_1, U_2, \ldots \sim \text{Uniform}[0,1]$
- Edge (i,j) present:

$$(i,j) \sim \text{Bernoulli}(\theta(U_i, U_j))$$

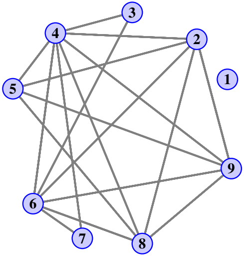
Call this distribution $\Gamma(\mathcal{G} \in \,.\, | \theta)$.



Theorem (Aldous; Hoover)

A random (dense) graph $\mathcal{G}$ is exchangeable iff

$$\mathbb{P}(\mathcal{G} \in \,.\,) = \int_{\mathcal{T}} \Gamma(\mathcal{G} \in \,.\, | \theta) Q(d\theta)$$



EXCHANGEABILITY: RANDOM GRAPHS

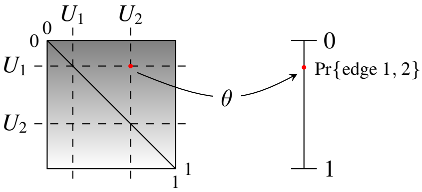
Random graph with independent edges

Given: $\theta : [0,1]^2 \to [0,1]$ symmetric function

- $U_1, U_2, \ldots \sim \text{Uniform}[0,1]$
- Edge (i,j) present:

$$(i,j) \sim \text{Bernoulli}(\theta(U_i, U_j))$$

Call this distribution $\Gamma(\mathcal{G} \in \,.\,|\theta)$.



Theorem (Aldous; Hoover)

A random (dense) graph $\mathcal{G}$ is exchangeable iff

$$\mathbb{P}(\mathcal{G} \in \,.\,) = \int_{\mathcal{T}} \Gamma(\mathcal{G} \in \,.\,|\theta) Q(d\theta)$$