Exchangeability: how Bruno de Finetti's ideas thrive in indeterminate soil

Gert de Cooman

Ghent University
SYSTeMS

ISIPTA '11, 26th July 2011

ENRIQUE MIRANDA

ERIK QUAEGHEBEUR

DE FINETTI'S EXCHANGEABILITY

Bruno de Finetti's exchangeability result

Informal definition

Consider an infinite sequence

$$X_1, X_2, \ldots, X_n, \ldots$$

of random variables assuming values in a finite set $\mathscr{X}$.

This sequence is exchangeable

if the mass function for any finite subset of these is invariant under any permutation of the indices.

Bruno de Finetti's exchangeability result

More formally

Consider any permutation π of the set of indices $\{1, 2, \ldots, n\}$.

For any $x = (x_1, x_2, \ldots, x_n)$ in $\mathscr{X}^n$, we let

$$\pi x := (x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)}).$$

Exchangeability:

If p^n is the mass function of the variables $X_1, \ldots, X_n$, then we require that:

$$p^n(x) = p^n(\pi x),$$

or in other words

$$p^n(x_1, x_2, \ldots, x_n) = p^n(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)}).$$

Bruno de Finetti's exchangeability result

Count vectors

For any $x \in \mathscr{X}^n$, consider the corresponding count vector $T(x)$, where for all $z \in \mathscr{X}$:

$$T_z(x) := |\{k \in \{1, \ldots, n\} : x_k = z\}|.$$

Example

For $\mathscr{X} = \{a, b\}$ and $x = (a, a, b, b, a, b, b, a, a, a, b, b, b)$, we have

$$T_a(x) = 6 \text{ and } T_b(x) = 7.$$

Observe that

$$T(x) \in \mathscr{N}^n := \left\{ m \in \mathbb{N}^{\mathscr{X}} : \sum_{x \in \mathscr{X}} m_x = n \right\}.$$

Bruno de Finetti's exchangeability result

Multiple hypergeometric distribution

There is some π such that $y = \pi x$ iff $T(x) = T(y)$.

Let $m = T(x)$ and consider the permutation invariant atom

$$[m] := \{y \in \mathscr{X}^n \colon T(y) = m\}.$$

This atom has how many elements?

$$\binom{n}{m} = \frac{n!}{\prod_{x \in \mathscr{X}} m_x!}$$

Let $\mathrm{HypGeo}^n(\cdot|m)$ be the expectation operator associated with the uniform distribution on $[m]$:

$$\mathrm{HypGeo}^n(f|m) := \binom{n}{m}^{-1} \sum_{x \in [m]} f(x) \text{ for all } f \colon \mathscr{X}^n \to \mathbb{R}$$

Bruno de Finetti's exchangeability result

The simplex of limiting frequency vectors

Consider the simplex:

$$\Sigma := \left\{\theta \in \mathbb{R}^{\mathscr{X}} : (\forall x \in \mathscr{X})\theta_x \geq 0 \text{ and } \sum_{x \in \mathscr{X}} \theta_x = 1\right\}.$$

Every (multivariate) polynomial $p \in \mathscr{V}^n(\Sigma)$ on Σ of degree at most n has a unique Bernstein expansion

$$p(\theta) = \sum_{m \in \mathscr{N}^n} b_p^n(m) B_m(\theta)$$

in terms of the Bernstein basis polynomials B_m of degree n:

$$B_m(\theta) := \binom{n}{m} \prod_{x \in \mathscr{X}} \theta_x^{m_x}.$$

Bruno de Finetti's exchangeability result

The infinite representation theorem

Infinite Representation Theorem:

The sequence $X_1, \dots, X_n, \dots$ of random variables in the finite set $\mathscr{X}$ is exchangeable iff there is a (unique) coherent prevision H on the linear space $\mathscr{V}(\Sigma)$ of all polynomials on Σ such that for all $n \in \mathbb{N}$ and all $f\colon \mathscr{X}^n \to \mathbb{R}$:

$$E_{p^n}(f) := \sum_{x \in \mathscr{X}} p^n(x) f(x) = H\left(\sum_{m \in \mathscr{N}} \mathrm{HypGeo}^n(f|m) B_m\right).$$

Observe that

$$B_m(\theta) = \mathrm{MultiNom}^n([m]|\theta) = \binom{n}{m} \prod_{x \in \mathscr{X}} \theta_x^{m_x}$$

$$\sum_{m \in \mathscr{N}^n} \mathrm{HypGeo}^n(f|m) B_m(\theta) = \mathrm{MultiNom}^n(f|\theta)$$

WOULD DE FINETTI HAVE LIKED THIS VERSION?

Would de Finetti have liked this version?

My reasons for thinking so:

1. Emphasis on linear (expectation or prevision) operators and linear spaces:

   random variables rather than events

Would de Finetti have liked this version?

My reasons for thinking so:

La prévision : ses lois logiques, ses sources subjectives

par

Bruno de FINETTI.

Donner la loi limite Φ, ou la fonction caractéristique ψ, équivaut donc, comme on le voit, à donner la suite des ω_n; cela suffit par conséquent pour déterminer la probabilité de tout problème relatif à des événements équivalents. Tout problème se ramène, en effet, dans le cas des événements équivalents, aux probabilités $\omega_r^{(n)}$ pour que, sur n épreuves, un nombre r quelconque soient favorables; on a (en posant $s = n - r$)

$$\omega_r^{(n)} = (-1)^s \Delta^s \omega_r = \binom{n}{r} \int_0^1 \xi^r (1-\xi)^s \, d\Phi(\xi), \tag{18}$$

Would de Finetti have liked this version?

My reasons for thinking so:

$$\begin{tikzcd}
\mathscr{G}(\mathscr{X}^n) \arrow[rr, "E_{p^n} = H \circ \mathrm{MultiNom}^n"] \arrow[d, "\mathrm{MultiNom}^n"'] & & \mathbb{R} \\
\mathscr{V}^n(\Sigma) \arrow[urr, "H"'] & &
\end{tikzcd}$$

for all n

Would de Finetti have liked this version?

My reasons for thinking so:

1. Emphasis on linear (expectation or prevision) operators and linear spaces:

   random variables rather than events

2. Emphasis on finitary events and random variables:

   H is only defined on the linear space $\mathscr{V}(\Sigma)$ of all polynomials on Σ, and need only be finitely additive there

Would de Finetti have liked this version?

My reasons for thinking so:

Coherent prevision:

A real functional P on a linear space $\mathscr{K}$ is a coherent prevision if for all $f,g \in \mathscr{K}$ and all real λ:

(i) $\inf f \leq P(f) \leq \sup f$ [boundedness]

(ii) $P(\lambda f) = \lambda \underline{P}(f)$ [homogeneity]

(iii) $P(f+g) = P(f) + P(g)$ [finite additivity]

A coherent prevision always satisfies uniform convergence:

$$\sup|f_n - f| \to 0 \Rightarrow P(f_n) \to P(f)$$

but not necessarily monotone convergence (countable additvity):

$$f_n \downarrow f \text{ pointwise } \Rightarrow P(f_n) \downarrow P(f)$$

Would de Finetti have liked this version?

My reasons for thinking so:

1. Emphasis on linear (expectation or prevision) operators and linear spaces:

   random variables rather than events

2. Emphasis on finitary events and random variables:

   H is only defined on the linear space $\mathscr{V}(\Sigma)$ of all polynomials on Σ, and need only be finitely additive there

3. Preference for assessments on continuous random quantities rather than discontinuous events

Would de Finetti have liked this version?

My reasons for thinking so:

Stone–Weierstraß Theorem (Bernstein's constructive version)

Every continuous function f on Σ is a uniform limit of polynomials:

$$\text{Bernstein approximant } \sum_{m \in \mathcal{N}^n} f\left(\frac{m}{n}\right) B_m \to f \text{ uniformly.}$$

As a corollary, H can be uniquely extended from $\mathscr{V}(\Sigma)$ to a coherent prevision on the linear space $\mathscr{C}(\Sigma)$ of all continuous functions on Σ:

$$H(f) := \lim_{n \to \infty} \sum_{m \in \mathcal{N}^n} f\left(\frac{m}{n}\right) H(B_m) \text{ for all } f \in \mathscr{C}(\Sigma).$$

This H satisfies monotone convergence trivially!

Would de Finetti have liked this version?

My reasons for thinking so:

1. Emphasis on linear (expectation or prevision) operators and linear spaces:

   random variables rather than events

2. Emphasis on finitary events and random variables:

   H is only defined on the linear space $\mathscr{V}(\Sigma)$ of all polynomials on Σ, and need only be finitely additive there

3. Preference for assessments on continuous random quantities rather than discontinuous events

4. Interesting relation with Fundamental Theorem of Prevision

Would de Finetti have liked this version?

My reasons for thinking so:

Infinitary events and random variables:

- strong (as opposed to weak) laws of large numbers
- zero–one laws
- results about unbounded stopping times
- …

$\Rightarrow$ we need to extend H beyond $\mathscr{C}(\Sigma)$

Would de Finetti have liked this version?

My reasons for thinking so:

Infinitary events and random variables:

- strong (as opposed to weak) laws of large numbers
- zero–one laws
- results about unbounded stopping times
- …

$\Rightarrow$ we need to extend H beyond $\mathscr{C}(\Sigma)$

Problem:

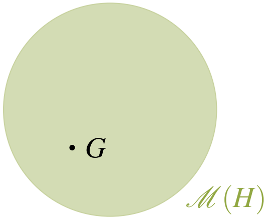
1. Are there coherent previsions G that extend H from $\mathscr{C}(\Sigma)$ to the set $\mathscr{G}(\Sigma)$ of all bounded functions?
2. If so, how can we characterise them?

Would de Finetti have liked this version?

My reasons for thinking so:

Hahn–Banach Theorem:

$$\mathscr{M}(H) := \{G \colon G \text{ extends } H \text{ to } \mathscr{G}(\Sigma)\} \neq \emptyset.$$



Would de Finetti have liked this version?

My reasons for thinking so:

F. Riesz Extension Theorem:

The coherent prevision H on $\mathscr{C}(\Sigma)$ can be uniquely extended to a coherent prevision E_H defined on the set $\mathscr{B}(\Sigma)$ of Borel measurable gambles that moreover satisfies monotone convergence:

$$E_H(f) = \int_\Sigma f(\theta)\,\mathrm{d}\mu_H(\theta) \text{ for all } f \text{ in } \mathscr{B}(\Sigma).$$

where μ_H is a countably additive probability measure on the Borel sets.

Monotone convergence:

$$f_n \downarrow f \text{ point-wise } \Rightarrow E_H(f_n) \downarrow E_H(f)$$

Would de Finetti have liked this version?

My reasons for thinking so:

De Finetti's Theorem on Exchangeable Variables

DAVID HEATH* AND WILLIAM SUDDERTH**

Theorem. To every infinite sequence of exchangeable random variables (X_n) having values in $\{0, 1\}$, there corresponds a probability distribution F concentrated on $[0, 1]$ such that

$$P\{X_1 = 1, \ldots, X_k = 1, X_{k+1} = 0, \ldots, X_n = 0\} = \int_0^1 \theta^k (1-\theta)^{n-k} F(d\theta) \tag{1}$$

for all n and $0 \leq k \leq n$.

Would de Finetti have liked this version?

My reasons for thinking so:

The natural extension $\underline{E}_H$ of H:

For any gamble f on Σ:

$$\underline{E}_H(f) := \min\{G(f) \colon G \in \mathscr{M}(H)\} = \min\{G(f) \colon G \text{ extends } H\}$$

(i) $\underline{E}_H$ is the point-wise smallest coherent lower prevision on $\mathscr{G}(\Sigma)$ that extends H.

(ii) $G \in \mathscr{M}(H) \Leftrightarrow G \geq \underline{E}_H$, and the bounds in de Finetti's Fundamental Theorem of Prevision are given by $[\underline{E}_H(f), \overline{E}_H(f)]$.

(iii) $\underline{E}_H$ is constructible:

$$\underline{E}_H(f) = \sup\{H(g) \colon g \in \mathscr{C}(\Sigma) \text{ and } g \leq f\}$$

(iv) $\underline{E}_H$ is a completely monotone lower prevision.

Would de Finetti have liked this version?

My reasons for thinking so:

1. Emphasis on linear (expectation or prevision) operators and linear spaces:

   random variables rather than events

2. Emphasis on finitary events and random variables:

   H is only defined on the linear space $\mathscr{V}(\Sigma)$ of all polynomials on Σ, and need only be finitely additive there

3. Preference for assessments on continuous random quantities rather than discontinuous events

4. Interesting relation with Fundamental Theorem of Prevision
5. Interesting relation with the notion of adherent probability mass

Would de Finetti have liked this version?

My reasons for thinking so:

```
@ARTICLE{cooman2008,
  author = {{d}e Cooman, Gert and Miranda, Enrique},
  title = {The {F}. {R}iesz {R}epresentation {T}heorem and finite additivity},
  booktitle = {Soft Methods for Handling Variability and Imprecision},
  year = 2008,
  pages = {243--252},
  publisher = {Springer},
  editor = {Dubois, Didier and Lubiano, María Asunción
             and Prade, Henri and Gil, María Ángeles
             and Grzegorzewski, Przemysław and Hryniewicz, Olgierd}
}
```

The F. Riesz Representation Theorem and finite additivity

Gert de Cooman and Enrique Miranda

1 Introduction

Let K be any compact metric space, and consider the linear space $\mathscr{L}(K)$ of all bounded real-valued maps on K. We provide this set with the topology of uniform convergence, which turns $\mathscr{L}(K)$ into a Banach space.

We call *gamble* any bounded real function on K, and *linear prevision* any positive, normalised (i.e., with operator norm 1) real linear functional on a linear subspace of $\mathscr{L}(K)$ that contains the constant gambles. We explain our reasons for this and other terminology in Sec. 3, which is intended to give background information and further discussion of the importance of the problem addressed here.

To set the stage, consider a positive, normalised real linear functional π on the set $\mathscr{C}(K)$ of all continuous bounded real functions on K. The F. Riesz Representation Theorem [14, Thm. 2.22] tells us that there is a *unique* (σ-additive) probability measure μ_π on the Borel sets of K such that for all continuous gambles f

$$\pi(f) = (L)\int f \,\mathrm{d}\mu_\pi,$$

where the integral is a Lebesgue integral associated with the probability measure μ_π. In other words, the linear prevision π on $\mathscr{C}(K)$ extends *uniquely* to a linear prevision L_π to the linear space $\mathscr{B}(K)$ of all Borel-measurable gambles on K that furthermore satisfies the *monotone convergence* requirement: if the increasing sequence of gambles f_n, $n \geq 0$ converges point-wise to a gamble f, then $L_\pi(f_n) \to L_\pi(f)$. The original linear prevision π on $\mathscr{C}(K)$ satisfies this extra monotone convergence con-

Gert de Cooman
Ghent University, SYSTeMS Research Group, Technologiepark–Zwijnaarde 914, 9052 Zwijnaarde, Belgium. e-mail: gert.decooman@ugent.be

Enrique Miranda
Rey Juan Carlos University, Dept. of Statistics and Operations Research, C-Tulipán, s/n, 28933 Móstoles, Spain. e-mail: enrique.miranda@urjc.es

Would de Finetti have liked this version?

My reasons for thinking so:

Consider for θ in Σ, the filter $\mathscr{N}_\theta$ of all neighbourhouds of θ:

$$\mathscr{N}_\theta = \{N_\theta \subseteq \Sigma \colon (\exists O_\theta \in \mathscr{T}_\theta)(O_\theta \subseteq N_\theta)\}$$

Lower oscillation:

For all gambles f on Σ, the lower oscillation $\underline{\mathrm{osc}}(f)$ is the point-wise greatest lower semi-continuous gamble that is dominated by f, and for all $\theta \in \Sigma$:

$$\underline{\mathrm{osc}}_\theta(f) = \sup\{g(\theta) \colon g \in \mathscr{C}(\Sigma) \text{ and } g \leq f\} = \sup_{N_\theta \in \mathscr{N}_\theta} \inf_{\vartheta \in N_\theta} f(\vartheta).$$

is the limit inferior of f for the convergence associated with the neighbourhood filter $\mathscr{N}_\theta$.

Would de Finetti have liked this version?

My reasons for thinking so:

For every $\theta \in \Sigma$, the completely monotone coherent lower prevision $\underline{\mathrm{osc}}_\theta$ is the natural extension of the lower prevision $f \to f(\theta)$ on $\mathscr{C}(\Sigma)$.

What does it represent?

All probability mass lies within any neighbourhood of θ:

$$P(N_\theta) = 1 \text{ for all } N_\theta \in \mathscr{N}_\theta.$$

Oscillation of f in θ:

$$\mathrm{osc}_\theta(f) = \overline{\mathrm{osc}}_\theta(f) - \underline{\mathrm{osc}}_\theta(f) = \inf_{N_\theta \in \mathscr{N}_\theta} \sup_{\vartheta, \vartheta' \in N_\theta} |f(\vartheta) - f(\vartheta')|$$

and f is continuous in θ iff $\mathrm{osc}_\theta(f) = 0$.

Would de Finetti have liked this version?

My reasons for thinking so:

Our simple result:

$$\underline{E}_H(f) = E_H(\underline{\mathrm{osc}}(f)) = \int_\Sigma \underline{\mathrm{osc}}_\theta(f)\,\mathrm{d}\mu_H(\theta) \text{ for all gambles } f \text{ on } \Sigma.$$

Would de Finetti have liked this version?

My reasons for thinking so:

Our simple result:

$$\underline{E}_H(f) = E_H(\underline{\mathrm{osc}}(f)) = \int_\Sigma \underline{\mathrm{osc}}_\theta(f)\,\mathrm{d}\mu_H(\theta) \text{ for all gambles } f \text{ on } \Sigma.$$

Maaß–Choquet Representation Theorem (2003)

Every real functional that satisfies a collection of inequalities $\mathscr{S}$ can be written uniquely as a σ-additive convex combination of the extreme points of the convex closed set of all real functionals that satisfy the system $\mathscr{S}$.

Would de Finetti have liked this version?

My reasons for thinking so:

Our simple result:

$$\underline{E}_H(f) = E_H(\underline{\mathrm{osc}}(f)) = \int_{\Sigma} \underline{\mathrm{osc}}_{\theta}(f)\,\mathrm{d}\mu_H(\theta) \text{ for all gambles } f \text{ on } \Sigma.$$

Maaß–Choquet Representation Theorem (2003)

Every real functional that satisfies a collection of inequalities $\mathscr{S}$ can be written uniquely as a σ-additive convex combination of the extreme points of the convex closed set of all real functionals that satisfy the system $\mathscr{S}$.

Choquet Representation Theorem (1953)

For completely monotone coherent lower probabilities (previsions), the extreme points are the lower probabilities (previsions) $\underline{P}_{\mathscr{F}}$ that assume the value 1 on some filter of sets $\mathscr{F}$, and zero elsewhere:

$$\underline{P}_{\mathscr{F}}(A) = \begin{cases} 1 & A \in \mathscr{F} \\ 0 & A \notin \mathscr{F} \end{cases} \text{ and } \underline{P}_{\mathscr{F}}(f) = \sup_{A \in \mathscr{F}} \inf_{z \in A} f(z).$$

EXCHANGEABILITY FOR COHERENT LOWER PREVISIONS

Exchangeability for coherent lower previsions

Bernoulli **15**(3), 2009, 721–735
DOI: 10.3150/09-BEJ182

Exchangeable lower previsions

GERT DE COOMAN[1,*], ERIK QUAEGHEBEUR[1,**] and ENRIQUE MIRANDA[2]

[1] *Ghent University, SYSTeMS Research Group, Technologiepark–Zwijnaarde 914, 9052 Zwijnaarde, Belgium. E-mails:* [*] *gert.decooman@ugent.be;* [**] *erik.quaeghebeur@ugent.be*
[2] *University of Oviedo, Dept. of Statistics and Operations Research. C-Calvo Sotelo, s/n, 33007, Oviedo, Spain. E-mail: mirandaenrique@uniovi.es*

We extend de Finetti's [*Ann. Inst. H. Poincaré* **7** (1937) 1–68] notion of exchangeability to finite and countable sequences of variables, when a subject's beliefs about them are modelled using coherent lower previsions rather than (linear) previsions. We derive representation theorems in both the finite and countable cases, in terms of sampling without and with replacement, respectively.

Keywords: Bernstein polynomials; coherence; convergence in distribution; exchangeability; imprecise probability; lower prevision; multinomial sampling; representation theorem; sampling without replacement

```
@ARTICLE{cooman2009,
  author = {{d}e Cooman, Gert and Quaeghebeur, Erik and Miranda, Enrique},
  title = {Exchangeable lower previsions },
  journal = {Bernoulli},
  year = 2009,
  pages = {721--735},
  volume = 15,
  doi = {10.3150/09-BEJ182}
}
```

Exchangeability for coherent lower previsions

Definition

Define, for any permutation π of $\{1,\ldots,n\}$ and any gamble f on $\mathscr{X}^n$:

$$\pi^t f := f \circ \pi$$

or equivalently

$$(\pi^t f)(x_1, x_2, \ldots, x_n) := f(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)}).$$

Exchangeability: (Walley, 1991)

If $\underline{P}^n$ is the lower prevision for the variables $X_1, \ldots, X_n$, then we require that:

$$\underline{P}^n(f - \pi^t f) \geq 0 \text{ for all permutations } \pi \text{ and all gambles } f \in \mathscr{G}(\mathscr{X}^n).$$

Equivalently:

$$\underline{P}^n(f - \pi^t f) = \overline{P}^n(f - \pi^t f) = 0.$$

Infinite Representation Theorem: precise case

The sequence $X_1, \ldots, X_n, \ldots$ of random variables in the finite set $\mathscr{X}$ is exchangeable iff there is a (unique) coherent prevision H on the linear space $\mathscr{V}(\Sigma)$ of all polynomials on Σ such that for all $n \in \mathbb{N}$ and all $f\colon \mathscr{X}^n \to \mathbb{R}$:

$$E_{p^n}(f) := \sum_{x \in \mathscr{X}} p^n(x) f(x) = H\left(\sum_{m \in \mathscr{N}} \mathrm{HypGeo}^n(f|m) B_m\right).$$

Infinite Representation Theorem: imprecise case

The sequence $X_1, \ldots, X_n, \ldots$ of random variables in the finite set $\mathscr{X}$ is exchangeable iff there is a (unique) coherent lower prevision $\underline{H}$ on the linear space $\mathscr{V}(\Sigma)$ of all polynomials on Σ such that for all $n \in \mathbb{N}$ and all $f\colon \mathscr{X}^n \to \mathbb{R}$:

$$\underline{P}^n(f) = \underline{H}\left(\sum_{m \in \mathscr{N}} \mathrm{HypGeo}^n(f|m) B_m\right).$$

Exchangeability for coherent lower previsions

Graphically:

$$\begin{tikzcd}
\mathscr{G}(\mathscr{X}^n) \arrow[rr, "\underline{P}^n = \underline{H} \circ \mathrm{MultiNom}^n"] \arrow[d, "\mathrm{MultiNom}^n"'] & & \mathbb{R} \\
\mathscr{V}^n(\Sigma) \arrow[urr, "\underline{H}"'] & &
\end{tikzcd}$$

for all n

EXCHANGEABILITY FOR SETS OF DESIRABLE GAMBLES

Exchangeability for sets of desirable gambles

ARTICLE IN PRESS

International Journal of Approximate Reasoning xxx (2011) xxx–xxx

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Exchangeability and sets of desirable gambles

Gert de Cooman, Erik Quaeghebeur *

Ghent University, SYSTeMS Research Group, Technologiepark – Zwijnaarde 914, 9052 Zwijnaarde, Belgium

A R T I C L E I N F O

Article history:
Available online xxxx

Keywords:
Sets of desirable gambles
Exchangeability
Representation
Natural extension
Updating
Extending an exchangeable sequence

A B S T R A C T

Sets of desirable gambles constitute a quite general type of uncertainty model with an interesting geometrical interpretation. We give a general discussion of such models and their rationality criteria. We study exchangeability assessments for them, and prove counterparts of de Finetti's Finite and Infinite Representation Theorems. We show that the finite representation in terms of count vectors has a very nice geometrical interpretation, and that the representation in terms of frequency vectors is tied up with multivariate Bernstein (basis) polynomials. We also lay bare the relationships between the representations of updated exchangeable models, and discuss conservative inference (natural extension) under exchangeability and the extension of exchangeable sequences.


```
@ARTICLE{cooman2011,
  author = {{d}e Cooman, Gert and Quaeghebeur, Erik},
  title = {Exchangeability and sets of desirable gambles},
  journal = {International Journal of Approximate Reasoning},
  year = 2011,
  note = {In the press.}
}
```

Exchangeability for sets of desirable gambles

Coherent sets of desirable gambles

A subject specifies a set $\mathscr{D} \subseteq \mathscr{G}(\mathscr{X}^n)$ of gambles he strictly accepts, his set of desirable gambles.

Coherence:

$\mathscr{D}$ is called coherent if it satisfies the following rationality requirements:

D1. if $f \leq 0$ then $f \notin \mathscr{D}$ [avoiding partial loss]

D2. if $f > 0$ then $f \in \mathscr{D}$ [accepting partial gain]

D3. if $f_1 \in \mathscr{D}$ and $f_2 \in \mathscr{D}$ then $f_1 + f_2 \in \mathscr{D}$ [combination]

D4. if $f \in \mathscr{D}$ then $\lambda f \in \mathscr{D}$ for all positive real numbers λ [scaling]

Here '$f > 0$' means '$f \geq 0$ and $f \neq 0$'.

Exchangeability for sets of desirable gambles

ELSEVIER International Journal of Approximate Reasoning 24 (2000) 125–148

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING

www.elsevier.com/locate/ijar

Towards a unified theory of imprecise probability

Peter Walley

```
@ARTICLE{walley2000,
   author = {Walley, Peter},
   title = {Towards a unified theory of imprecise probability},
   journal = {International Journal of Approximate Reasoning},
   year = 2000,
   volume = 24,
   pages = {125--148}
}
```

Exchangeability for sets of desirable gambles

Relation with lower previsions:

$$\underline{P}(f) = \sup\{\mu \in \mathbb{R} \colon f - \mu \in \mathscr{D}\}$$

Reasons for preferring sets of desirable gambles are legion:

- more general
- more intuitive and direct
- much simpler conditioning
- nice geometrical interpretation of coherence
- much simpler coherence arguments
- …

Exchangeability for sets of desirable gambles

Definition of exchangeability

Consider random variables $X_1, \dots, X_n$ in $\mathscr{X}$, and a coherent set of desirable gambles

$$\mathscr{D}^n \subseteq \mathscr{G}(\mathscr{X}^n).$$

For any gamble f on $\mathscr{X}^n$ and permutation π of $\{1, \dots, n\}$, exchangeability means that f and $\pi^t f$ are considered to be indifferent:

Exchangeability of $\underline{P}^n$:

For all $f \in \mathscr{G}(\mathscr{X}^n)$ and all permutations π:

$$\underline{P}^n(f - \pi^t f) \geq 0.$$

Exchangeability for sets of desirable gambles

Definition of exchangeability

Consider random variables $X_1, \ldots, X_n$ in $\mathscr{X}$, and a coherent set of desirable gambles

$$\mathscr{D}^n \subseteq \mathscr{G}(\mathscr{X}^n).$$

For any gamble f on $\mathscr{X}^n$ and permutation π of $\{1, \ldots, n\}$, exchangeability means that f and $\pi^t f$ are considered to be indifferent:

Exchangeability of $\mathscr{D}^n$:

For all $f \in \mathscr{G}(\mathscr{X}^n)$, all positive $\mu \in \mathbb{R}$ and all permutations π:

$$\underbrace{f - \pi^t f}_{\text{indifferent}} + \mu \in \mathscr{D}^n.$$

Exchangeability for sets of desirable gambles

Definition of exchangeability

Consider random variables $X_1, \ldots, X_n$ in $\mathscr{X}$, and a coherent set of desirable gambles

$$\mathscr{D}^n \subseteq \mathscr{G}(\mathscr{X}^n).$$

For any gamble f on $\mathscr{X}^n$ and permutation π of $\{1, \ldots, n\}$, exchangeability means that f and $\pi^t f$ are considered to be indifferent:

Exchangeability of $\mathscr{D}^n$:

For all $f \in \mathscr{G}(\mathscr{X}^n)$, all $g \in \mathscr{D}^n$ and all permutations π:

$$\underbrace{f - \pi^t f}_{\text{indifferent}} + g \in \mathscr{D}^n.$$

Rationality axiom for combining desirability with indifference:

DI. $\mathscr{D} + \mathscr{I} \subseteq \mathscr{D}$ [desiring sweetened deals]

Representation Theorem

A sequence $\mathscr{D}^1, \ldots, \mathscr{D}^n, \ldots$ of coherent sets of desirable gambles is exchangeable iff there is some (unique) Bernstein coherent $\mathscr{H} \subseteq \mathscr{V}(\Sigma)$ such that:

$$f \in \mathscr{D}^n \Leftrightarrow \mathrm{MultiNom}^n(f|\cdot) \in \mathscr{H} \text{ for all } n \in \mathbb{N} \text{ and } f \in \mathscr{G}(\mathscr{X}^n).$$

Recall that

$$\mathrm{MultiNom}^n(f|\theta) = \sum_{m \in \mathscr{N}^n} \mathrm{HypGeo}^n(f|m) B_m(\theta)$$

$$\mathrm{HypGeo}^n(f|m) = \binom{n}{m}^{-1} \sum_{x \in [m]} f(x)$$

$$B_m(\theta) = \binom{n}{m} \prod_{x \in \mathscr{X}} \theta_x^{m_x}.$$

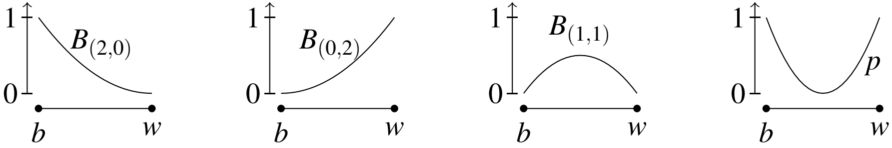
Exchangeability for sets of desirable gambles

Bernstein coherence

A set $\mathscr{H}$ of polynomials on Σ is Bernstein coherent if:

- B1. if p has some non-positive Bernstein expansion then $p \notin \mathscr{H}$
- B2. if p has some positive Bernstein expansion then $p \in \mathscr{H}$
- B3. if $p_1 \in \mathscr{H}$ and $p_2 \in \mathscr{H}$ then $p_1 + p_2 \in \mathscr{H}$
- B4. if $p \in \mathscr{H}$ then $\lambda p \in \mathscr{H}$ for all positive real numbers λ.

There are positive (negative) p with no positive (negative) Bernstein expansion of any degree!



Exchangeability for sets of desirable gambles

Conditioning

Suppose we observe the first n variables, with count vector $m = T(x)$:

$$(X_1, \ldots, X_n) = (x_1, \ldots, x_n) = x.$$

Then the remaining variables

$$X_{n+1}, \ldots, X_{n+k}, \ldots$$

are still exchangeable, with representation $\mathscr{H}\rfloor x = \mathscr{H}\rfloor m$ given by:

$$p \in \mathscr{H}\rfloor m \Leftrightarrow B_m p \in \mathscr{H}.$$

Conclusion:

A Bernstein coherent set of polynomials $\mathscr{H}$ completely characterises all predictive inferences about an exchangeable sequence.

Exchangeability for sets of desirable gambles

Imprecise-probabilistic IID processes

An exchangeable process $X_1, \ldots, X_n, \ldots$ with represening set of polynomials $\mathscr{H}$ is IID when no observation has any influence:

$$\mathscr{H}\rfloor m = \mathscr{H} \text{ for all count vectors } m.$$

Equivalent condition on $\mathscr{H}$:

$$\big(\forall p \in \mathscr{V}(\Sigma)\big)(\forall m)(p \in \mathscr{H} \Leftrightarrow B_m p \in \mathscr{H}).$$

PREDICTIVE INFERENCE SYSTEMS

Predictive inference systems

What are they?

Formal definition:

A predictive inference system is a map Ψ that associates with every finite set of categories $\mathscr{X}$ a Bernstein coherent set of polynomials on $\Sigma_{\mathscr{X}}$:

$$\Psi(\mathscr{X}) = \mathscr{H}_{\mathscr{X}}$$

Basic idea:

Once the set of possible observations $\mathscr{X}$ is determined, then all predictive inferences about successive observations $X_1, \ldots, X_n, \ldots$ in $\mathscr{X}$ are completely fixed by $\Psi(\mathscr{X}) = \mathscr{H}_{\mathscr{X}}$.

Predictive inference systems

Inference principles

Even if (when) you don't like this idea, you might want to concede the following:

Using inference principles to constrain Ψ:

We can use general inference principles to impose conditions on Ψ, or in other words to constrain:

- the values $\mathscr{H}_{\mathscr{X}}$ can assume for different $\mathscr{X}$
- the relation between $\mathscr{H}_{\mathscr{X}}$ and $\mathscr{H}_{\mathscr{Y}}$ for different $\mathscr{X}$ and $\mathscr{Y}$

Predictive inference systems

Inference principles

Even if (when) you don't like this idea, you might want to concede the following:

Using inference principles to constrain Ψ:

We can use general inference principles to impose conditions on Ψ, or in other words to constrain:

- the values $\mathscr{H}_{\mathscr{X}}$ can assume for different $\mathscr{X}$
- the relation between $\mathscr{H}_{\mathscr{X}}$ and $\mathscr{H}_{\mathscr{Y}}$ for different $\mathscr{X}$ and $\mathscr{Y}$

Taken to (what might be called) extremes (Carnap, Walley, ...):

Impose so many constraints (principles) that you end up with a single Ψ, or a parametrised family of them, e.g.:

the λ system, the IDM family

Renaming Invariance

Inferences should not be influenced by what names we give to the categories.

For any onto and one-to-one $\pi\colon \mathscr{X} \to \mathscr{Y}$

$$B_{C_\pi(m)} p \in \mathscr{H}_{\mathscr{Y}} \Leftrightarrow B_m(p \circ C_\pi) \in \mathscr{H}_{\mathscr{X}} \text{ for all } p \in \mathscr{V}(\Sigma_{\mathscr{Y}}) \text{ and } m \in \mathscr{N}_{\mathscr{X}}$$

Pooling Invariance

For gambles that do not differentiate between pooled categories, it should not matter whether we consider predictive inferences for the set of original categories $\mathscr{X}$, or for the set of pooled categories $\mathscr{Y}$.

For any onto $\rho\colon \mathscr{X} \to \mathscr{Y}$

$$B_{C_\rho(m)}p \in \mathscr{H}_{\mathscr{Y}} \Leftrightarrow B_m(p \circ C_\rho) \in \mathscr{H}_{\mathscr{X}} \text{ for all } p \in \mathscr{V}(\Sigma_{\mathscr{Y}}) \text{ and } m \in \mathscr{N}_{\mathscr{X}}$$

Specificity

If, after making a number of observations in $\mathscr{X}$, we decide that in the future we are only going to consider outcomes in a subset $\mathscr{Y}$ of $\mathscr{X}$, we can discard from the past observations those outcomes not in $\mathscr{Y}$.

For any $\mathscr{Y} \subseteq \mathscr{X}$

$$B_{m|_{\mathscr{Y}}}\, p \in \mathscr{H}_{\mathscr{Y}} \Leftrightarrow B_m\,(p \circ \cdot|_{\mathscr{Y}}) \in \mathscr{H}_{\mathscr{X}} \text{ for all } p \in \mathscr{V}\,(\Sigma_{\mathscr{Y}}) \text{ and } m \in \mathscr{N}_{\mathscr{X}}$$

Predictive inference systems

Look how nice!

Renaming Invariance

For any onto and one-to-one $\pi\colon \mathscr{X} \to \mathscr{Y}$

$$B_{C_\pi(m)}\, p \in \mathscr{H}_{\mathscr{Y}} \Leftrightarrow B_m\,(p \circ C_\pi) \in \mathscr{H}_{\mathscr{X}} \text{ for all } p \in \mathscr{V}\,(\Sigma_{\mathscr{Y}}) \text{ and } m \in \mathscr{N}_{\mathscr{X}}$$

Pooling Invariance

For any onto $\rho\colon \mathscr{X} \to \mathscr{Y}$

$$B_{C_\rho(m)}\, p \in \mathscr{H}_{\mathscr{Y}} \Leftrightarrow B_m\,(p \circ C_\rho) \in \mathscr{H}_{\mathscr{X}} \text{ for all } p \in \mathscr{V}\,(\Sigma_{\mathscr{Y}}) \text{ and } m \in \mathscr{N}_{\mathscr{X}}$$

Specificity

For any $\mathscr{Y} \subseteq \mathscr{X}$

$$B_{m|_{\mathscr{Y}}}\, p \in \mathscr{H}_{\mathscr{Y}} \Leftrightarrow B_m\,(p \circ \cdot|_{\mathscr{Y}}) \in \mathscr{H}_{\mathscr{X}} \text{ for all } p \in \mathscr{V}\,(\Sigma_{\mathscr{Y}}) \text{ and } m \in \mathscr{N}_{\mathscr{X}}$$