Exchangeability, Dependence, Inequalities and Games

Yosi Rinott, Hebrew University

Rome, November 16 2006,
Bruno de Finetti Centenary Conference

Structure of talk:

• Main part: dependence properties and some inequalities related to representations of exchangeable random variables. Such constructions arise as (asymptotic) distribution in statistics, reliability theory models, etc.

• A game example.

• A short historical part: games and de Finetti's definition of probability.

de Finetti's definition of probability:

inconsistency in the **odds** (given by the bookmaker) means that a combination of bets can be devised in such a way that someone is certain to win whatever happens. This is called Dutch Book. (?)

Example of **inconsistency**: $E_1 \cup E_2 = \Omega$, and a bookmaker sets odds $P(E_1) = P(E_2) = 1/4$; if I bet \$1 on each event, I pay $1/4 + 1/4 = 1/2$ and I win at least 1 with certainty.

"The condition of **consistency** [of odds or probabilities] is the sole basis on which the whole theory of probability rests ... consists in allowing no chance of a Dutch Book occurring...."

(Intransitive preferences ⇒ possibility of Dutch Booking.)

A more formal definition:

The real-valued function P on $\mathbf{E}$ is said to be a probability on a class of events $\mathbf{E}$ if, for any finite subclass $\{E_1, \ldots, E_n\}$ of $\mathbf{E}$ and any choice of real numbers $c_1, \ldots, c_n$ the payoff $G(\omega) = G(E_1, \ldots, E_n; c_1, \ldots, c_n) =: \sum_{k=1}^{n} c_k[\mathbf{1}_{E_k}(\omega) - P(E_k)]$ satisfies

$$\inf_{\omega \in \{E_1, \ldots, E_n\}} G(\omega) \leq 0 \leq \sup_{\omega \in \{E_1, \ldots, E_n\}} G(\omega).$$

(see Cifarelli and Regazzini 1996).

Example: $E_1 \bigcap E_2 = \emptyset$, $P(E_1) = P(E_2) = 3/4$
$\Rightarrow \inf G(E_1, E_2 : -1, -1) = 1/2$

Using probabilities to define a simple game:

Colonel (B)lotto games: players A and B choose a *marginal distribution* $X \in \mathcal{F}$ and $Y \in \mathcal{G}$ respectively; values are drawn (independently or ...) and the winner is the larger. (Borel 1921, Robertson, 2006, Hart 2006).

Payoff=Probability of win $= P(X > Y)$.

Pure strategy: choice of a distribution.

de Finetti defined probability by payoffs; here we reverse his thinking and define payoff as the probability $P(X > Y)$; it is a zero (fixed) sum game.

RIDDLE:

Consider the following game (Kaminsky, Luks, Nelson, 1984).

Two teams of gladiators $A = \{a_i\}$, and $B = \{b_j\}$. Each round a gladiator is selected from each team with strengths a_i and b_j, and the first wins with probability $\frac{a_i}{a_i+b_j}$. The winner stays on his team with the same strength.

A team loses when all members are dead.

Question: What order is optimal? Should you send your stronger gladiators first? If you manage team A, how does your order depend on $\mathbf{b}$?

Main part

Four types of exchangeable constructions: (א), (ב), (ג), (ד), and related notions of dependence and inequalities

(Finitely) exchangeable random variables appear as models in statistics, reliability (e.g., Spizzichino, 2001), game theory, etc. Below are some representation formulas for such variables, and examples of related inequalities and dependence properties.

Let (א) $h(x_1, \ldots, x_n) = \int f(x_1, \theta) \cdots f(x_n, \theta) d\mu(\theta)$ be the density function of $\mathbf{X} = X_1, \ldots, X_n$. Then the components of $\mathbf{X}$ are positively dependent in the sense that $Cov(X_i, X_j) \geq 0$, $Cov(h(X_i), h(X_j)) \geq 0$ for h increasing, but $P(X_1 > c_1, \ldots, X_n > c_n) \geq \prod P(X_i > c_i)$ only if $c_1 = \ldots = c_n$ no Positive Quadrant Dependence.

Definitions:

A function $f(\mathbf{x}), \mathbf{x} \in \mathbb{R}^n$ is supermodular if $f(\mathbf{x} \vee \mathbf{y}) + f(\mathbf{x} \wedge \mathbf{y}) \geq f(\mathbf{x}) + f(\mathbf{y})$.
Then $g(\mathbf{x}) = e^{f(\mathbf{x})}$ is FKG (TP_2 if $n = 2$, MTP_2, Affiliation).

$(\aleph)$ $\Rightarrow$ if $f(x, \theta)$ is TP_2 then h is MTP_2 (FKG) $\Rightarrow$ very strong positive dependence of $\mathbf{X}$.
Example : $f_\theta(x)$ corresponds to Bernoulli or Binomial(m, θ), Poisson(θ), $\Gamma(\theta, \beta)$, or $\Gamma(\alpha, \theta)$ with density $f_\theta(x) \sim x^{\alpha-1} e^{-x/\theta}$;

so h MTP_2 also if $f_\theta(x) \sim x^{\alpha-1} e^{-x\theta}$.

A strong notion of positive dependence: the components of $\mathbf{X}$ are Associated if $\text{Cov}(\varphi(\mathbf{X}), \psi(\mathbf{X})) \geq 0$ for any increasing φ and ψ, and weakly Associated if $\text{Cov}(\varphi(\mathbf{X}_A), \psi(\mathbf{X}_B)) \geq 0$ for disjoint sets A and $B \subseteq \{1, \ldots, n\}$. Negative Association (Joag-Dev and Proschan (1983)), if the last $\geq$ is reversed.

MTP$_2$ $\Rightarrow$ Association.

Majorization: We say $\mathbf{a} \prec \mathbf{b}$ if $a_{(k)} + \ldots + a_{(n)} \leq b_{(k)} + \ldots + b_{(n)}$ $\forall k$ with equality for $k = 1$.

If $\mathbf{a} \prec \mathbf{b} \Rightarrow \psi(\mathbf{a}) \leq \psi(\mathbf{b})$ we say that ψ is Schur convex. $\Rightarrow \psi$ is symmetric (permutation invariant)

Theorem (Christofides and Vaggelatou, 2002) $\mathbf{X}$ weakly associated and the components of $\mathbf{X}^*$ are independent having the same marginals as $\mathbf{X}$, then $Ef(\mathbf{X}^*) \le Ef(\mathbf{X})$ for any supermodular f.

Example: If $f(\mathbf{X}) = \mathbb{I}_{X_{(k)}>t} + \ldots + \mathbb{I}_{X_{(n)}>t}$, then $-f$ is supermodular, and therefore

$$F_{X_{(k)}}(t) + \ldots + F_{X_{(n)}}(t) \le F^*_{X_{(k)}}(t) + \ldots + F^*_{X_{(n)}}(t).$$

$$\Downarrow$$

$$EX_{(k)} + \ldots + EX_{(n)} \le EX^*_{(k)} + \ldots + EX^*_{(n)}.$$

$$\Downarrow$$

$$(EX_{(1)}, \ldots, EX_{(n)}) \prec (EX^*_{(1)}, \ldots, EX^*_{(n)}).$$

Theorem (Shaked 1977) The latter majorization relation is implied by $(\aleph)$.

Remark: $(\aleph)$ does not imply (weak) Association nor supermodular order relative to iid (the function $h(x_1)g(x_2)$ with increasing h, g is supermodular, so supermodular order implies PQD).

A more general exchangeable density: Conditional independence of non iid variables.

Let ($\beth$) $h(x_1, \ldots, x_n)$
$= \int f(x_1, \theta_1) \cdots f(x_n, \theta_n) g(\theta_1, \ldots, \theta_n) d\boldsymbol{\theta}$
be the density function of $\mathbf{X} = X_1, \ldots, X_n$, where g is a symmetric (exchangeable) multivariate density.
Does not imply positive correlations.
If all $f(x, \theta_i)$ TP_2 and $g(\boldsymbol{\theta})$ MTP_2, then so is h.

Example:
$f_\theta(x) = e^{-\theta} \frac{\theta^x}{x!} \sim$ Poisson $= \varphi_x(\theta) \sim \Gamma(x+1, 1)$.
Note: $\int_o^\infty \varphi_{x_1}(\theta) \varphi_{x_1}(\vartheta - \theta) d\theta \sim \Gamma(x_1 + x_2 + 2, 1)$ - semigroup property.

Claim: with TP_2 and the semigroup property, if g in ($\beth$) is Schur convex, so is h.

Semigroup property: A function $\varphi(\lambda, x)$ defined on $(0, \infty) \times [0, \infty)$ is said to satisfy the semigroup property in λ if $\varphi(\lambda_1 + \lambda_2, x)$ $= \int_0^\infty \varphi(\lambda_1, x-y)\varphi(\lambda_2, y)\, d\mu(y)$, where μ is either the Lebesgue measure or the counting measure on non-negative integers.

Theorem (Proschan and Sethuraman, 1977) Let $\varphi(\lambda, x)$ be TP_2 with the semigroup property, and let $g(x)$ be Schur-convex [Schur-concave]; then $h(\boldsymbol{\lambda}) = \int_0^\infty \cdots \int_0^\infty g(\mathbf{x}) \prod_{i=1}^n \varphi(\lambda_i, x_i) d\mu(x_i)$ is Schur-convex [Schur-concave] whenever the integral exists.

Symmetrization: a special (extreme) case of ($\beth$) where g is a permutation distribution

Let $X_i \sim \phi(x, \theta_i)$ independent. Consider probabilities of permutation-invariant events like $P(\sum_i \psi(X_i) > t)$, or the distribution of order statistics or sums, may consider exchangeable random variables

($\gimel$) $\mathbf{Y} = (Y_1, \ldots Y_n) \sim \frac{1}{n!} \sum_\pi \prod_i \phi(y_i, \theta_{\pi(i)})$.

(Marshall-Olkin, Spizzichino, Pemantle)

[A general symmetrization: $\frac{1}{n!} \sum_\pi \phi(y_{\pi(1)}, \ldots, y_{\pi(n)})$]

($\gimel$) $\Rightarrow$ $\mathbf{Y}$ has dependent - negatively correlated - components.

NO Negative Association (NA), so NA is not closed under symmetrization (a question of Pemantle, 1999). Moreover, no Negative Quadrant Dependence even for $n = 2$. Independent r.v. satisfy all notions of Negative dependence.

Symmetrization has negative correlations, but not even NQD. Like (א).

similar to the case of (א) we can prove that

(ב) $\Rightarrow$ $(F_{X_{(1)}}(t), \ldots, F_{X_{(n)}}(t)) \succ (F^*_{X_{(1)}}(t), \ldots, F^*_{X_{(n)}}(t))$

More generally:

Let (ד) $h(x_1, \ldots, x_n)$
$= \int\int f(x_1; \theta_1, \eta_1) \cdots f(x_n; \theta_n, \eta_n) g(\boldsymbol{\theta}; \boldsymbol{\eta}) d\boldsymbol{\theta} d\boldsymbol{\eta}$
where $g(\pi\boldsymbol{\theta}; \pi'\boldsymbol{\eta}) = g(\boldsymbol{\theta}; \boldsymbol{\eta})$ for all permutations π, π'.

More on Majorization and Schur functions

Theorem (Boland et al 1994): Let $X_1, \ldots, X_n$ be iid Exp(1), then $g(\boldsymbol{\lambda}) = P(\sum_i X_i/\lambda_i > t)$ is Schur convex in $\boldsymbol{\lambda} > 0$. Note λ_i are hazard rates. $P(\sum_i X_i/\lambda_i > t) = \int \mathbb{I}_{\sum_i X_i > t} \prod e^{-\lambda_i x_i} d\mathbf{x}$,($\sqsupset$)

Life is not always simple:

Diaconis (1970's): For iid Exp(1) variables, $P(a_1 X_1 + a_2 X_2 > t)$ is decreasing (increasing) in the partial order $a \succ a'$ for small (large) t.

Diaconis, Perlman + ...(1987) $P(a_1 X_1 + \ldots + a_n X_n > t)$ is Schur concave for iid $X_i \sim \Gamma$ and small t and Schur concave for large t. Conjecture: for each t it is a Schur function.

A "generalization" :

consider $\psi(\mathbf{a}) = P(a_1X_1 + \ldots + a_nX_n > Y)$ where $Y \sim Exp$, X_i independent.

$$\psi(\mathbf{a}) = \int F_Y(\textstyle\sum_i a_i x_i) \prod_i f(x_i) d\mathbf{x} \qquad (*)$$
$$= \int F_Y(\textstyle\sum_i x_i) \prod_i \frac{1}{a_i} f(x_i/a_i) d\mathbf{x}$$

this has the form (ב).

F_Y concave and (*) $\Rightarrow$ $\psi(\mathbf{a})$ is symmetric and concave and hence Schur concave.

Answer to Riddle: **All orders are equally good**. For all orders of gladiators P(Team A wins) is the same.

$X_i, Y_j \sim Exp(1)$ iid $\Rightarrow P(aX_i > bY_j) = \frac{a}{a+b}$

$$\Downarrow$$

Exchangeability: For any order of sending gladiators to fight, P(Team A wins)
$= P(\sum_{i=1}^{m} a_i X_i > \sum_{j=1}^{n} b_j Y_j) = G_{m,n}(\boldsymbol{a}, \mathbf{b})$
$= \frac{a_1}{a_1+b_1} G_{m,n-1}(\boldsymbol{a}, \mathbf{b} \setminus \{b_1\}) + \frac{b_1}{a_1+b_1} G_{m-1,n}(\boldsymbol{a} \setminus \{a_1\}, \mathbf{b})$

Colonel (B)lotto: What is the optimal choice of $\mathbf{a}, \mathbf{b}$: subject to $\sum_i a_i = \alpha$, $\sum_j b_j = \beta$, perhaps $\alpha = \beta$ (equal means), perhaps equal medians, perhaps constraints on $\sum_i 1/a_i$ and $\sum_j 1/b_j$ (sums of hazard rates).

Majorization : let $a_{(1)} \le \cdots \le a_{(n)}$ be the ordered $a_1, \ldots, a_n$. We say $\mathbf{a} \succ \mathbf{a}'$ if $\forall\ k$,
$\sum_{i=k}^n a_{(i)} \ge \sum_{i=k}^n a'_{(i)}$ and $\sum_{i=1}^n a_{(i)} = \sum_{i=1}^n a'_{(i)}$.
The vector $(\bar{a}, \ldots, \bar{a})$ is minimal in the order $\succ$.

If $\mathbf{a} \succ \mathbf{a}' \Rightarrow \psi(\mathbf{a}) \le \psi(\mathbf{a}')$ we say that ψ is Schur concave. $\Rightarrow \psi$ is symmetric (permutation invariant)

A symmetric concave or log-concave function is Schur concave.

Diaconis +: $P(\sum_i a_i X_i > b)$ is decreasing in the partial order $\succ$ =Schur concave for b small, and increasing=Schur convex for b large.
$\overset{?}{\Rightarrow}$ $P(\sum_{i=1}^m a_i X_i > bY_1)$ same behavior?

NO! $P(\sum_{i=1}^{m} a_i X_i > tY_1) = 1 - \prod_i \frac{t}{t+a_i}$, Schur concave. For given $\alpha = \sum_i a_i$ the probability is max when $a_1 = \ldots = a_m$, so against one player ($n = 1$), weak or strong, the best is always equal power.

What happens when there is more than one player in team B, $n \geq 2$?

Here the phenomenon of Diaconis does take place: with $\sum_i a_i = \alpha$ fixed and small b_j's, $P(\text{Team A wins}) = P(\sum_i a_i X_i > \sum_j b_j Y_j)$ is maximal when a_i's are equal, and for large b_j's when there is one strong gladiator.

There seems to be "phase transition": for each $\mathbf{b} = (b_1, \ldots, b_n)$ one of the above two choices of $\mathbf{a}$ is optimal (but above probability is not always Schur)

No transitivity: $\exists\ \mathbf{a}, \mathbf{b}, \mathbf{c}$ such that $P(A \text{ beats } B)$, $P(B \text{ beats } C)$, and $P(C \text{ beats } A) > 1/2$, Kaminsky et al (1984).

Now assume $\sum_i a_i = \sum_j b_j$ (fair game ?)
Claim: Best reply to $b_1 = \ldots = b_m$ is $a_1 = \ldots = a_n$ (equal strengths)

$$\Downarrow$$

Equal strengths is unique Nash (0-sum game, hence also min-max and max-min. value = 1/2).

Conjecture 1: when $\sum_i a_i = \sum_j b_j$
$P(\sum_i a_i X_i > \sum_j b_j Y_j) > 1/2 \Leftrightarrow b \underset{\neq}{\succ} a$.

$$\Downarrow$$

Above claim.

Conjecture 2: $\psi(\mathbf{a}, \mathbf{b}) = P(\sum_i a_i X_i > \sum_j b_j Y_j)$ is Schur convex in $\mathbf{b}$ and Schur concave in $\mathbf{a}$ when $\sum_i a_i = \sum_j b_j$ (not true without such a condition)

$$\Downarrow$$

Conjecture 1 (take $\mathbf{a} = \mathbf{b}$...)

When $\sum_i 1/a_i = \sum_j 1/b_j$ we have $1/a \succ 1/a' \Rightarrow \sum_i a_i \geq \sum_i a'_i$ and indeed
P(Team A wins) is increasing in $\succ$, Boland et al (1994).

Remark (Rotar, Galambos?) Order statistics, or sums of random variables keep their distribution after symmetrization, and it should be possible to study them for exchangeable r.v.'s.

There are CLT's or normal approximations for sums of variables which form a Markov chain, Martingale, Mixing conditions, all depending on order and no exchangeability. But the sum does not depend on the order!

Symmetrization leave the distribution of the sum unchanged. Why are conditions which involve order so relevant? The fact that after symmetrization there are negative correlations implies that this exchangeable distribution is not embeddable in an infinite exchangeable sequence for which CLT is known (Blum et al).

The fact of no NA prevents using Newman's CLT's for NA variables.