Imprecise probabilities: a tool to generalize statistical preference

Ignacio Montes Gutiérrez
imontes@spi.uniovi.es

University of Oviedo

Durham, 6^{th} september of 2010

Overview

Overview

Comparison of random variables

- In some many real situations, we have to take decisions.
- Mathematically, the problem is to compare between random variables.
- The most usual methods are:
  - Stochastic Dominance
  - To choose the random variable whose mean is greater.
- A new way to compare random variables: statistical preference.

Overview

Probabilistic relation

Let A be a set of alternatives. A probabilistic relation is a map

$$Q : A \times A \to [0,1]$$

such that

$$Q(a,b) + Q(b,a) = 1 \quad \forall (a,b) \in A^2.$$

Probabilistic relation *[De Schuymer et al. (2003)]*

Let A be a set of random variables defined on the same probabilistic space. The following probabilistic relation can be defined:

$$Q : A \times A \to [0, 1]$$

such that

$$Q(X, Y) = P(X > Y) + \frac{1}{2} P(X = Y).$$

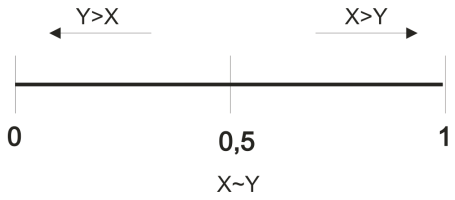
Probabilistic relation *[De Schuymer et al. (2003)]*

Let A be a set of random variables defined on the same probabilistic space. The following probabilistic relation can be defined:

$$Q : A \times A \to [0, 1]$$

such that

$$Q(X, Y) = P(X > Y) + \frac{1}{2} P(X = Y).$$



Statistical preference *[De Schuymer et al. (2003)]*

Let X and Y be two random variables defined on the same probability space. It is said that:

- X is statistically preferred to Y if $Q(X,Y) \geq \frac{1}{2}$. We denote it by $X \geq_{SP} Y$.

Statistical preference *[De Schuymer et al. (2003)]*

Let X and Y be two random variables defined on the same probability space. It is said that:

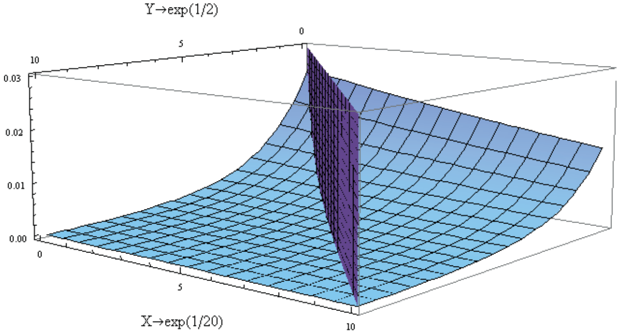
- X is statistically preferred to Y if $Q(X,Y) \geq \frac{1}{2}$. We denote it by $X \geq_{SP} Y$.
- X and Y are statistically indifferent if $Q(X,Y) = \frac{1}{2}$.

Statistical preference *[De Schuymer et al. (2003)]*

Let X and Y be two random variables defined on the same probability space. It is said that:

- X is statistically preferred to Y if $Q(X,Y) \geq \frac{1}{2}$. We denote it by $X \geq_{SP} Y$.
- X and Y are statistically indifferent if $Q(X,Y) = \frac{1}{2}$.
- X is statistically preferred strongly over Y if $Q(X,Y) > \frac{1}{2}$.

Interpretation of statistical preference



Characterizations of statistical preference

Theorem

Let X and Y be two independent random variables, and let X' be identically distributed as X and independent of X and Y. Then, $X \geq_{SP} Y$ if and only if

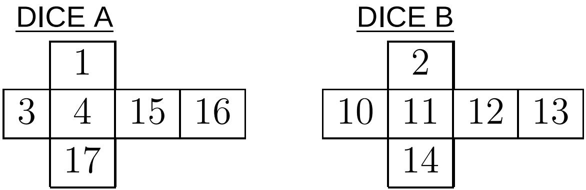
$$E(F_Y(X)) - E(F_X(X)) \geq \frac{1}{2}\big(P(X = Y) - P(X = X')\big).$$

Theorem

Let X and Y be two continuous and independent random variables. Then it holds that

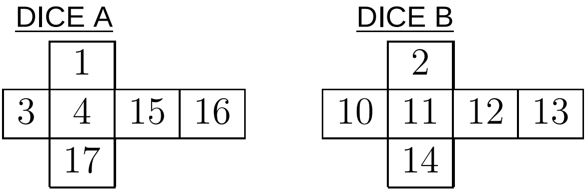
$$X \geq_{SP} Y \Leftrightarrow Me(X - Y) \geq 0.$$

Example



Consider this game: the dice whose number is greater wins.

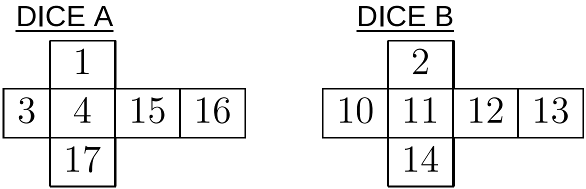
Example



Consider this game: the dice whose number is greater wins.

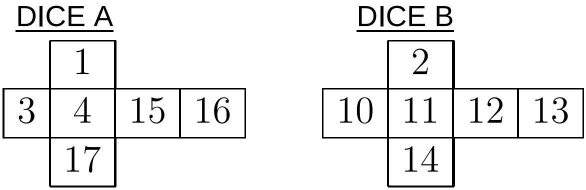
Which one we should choose?

Example



- $E(A) < E(B)$.

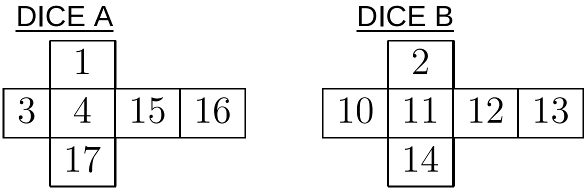
Example



- $E(A) < E(B)$.
- $Q(A,B) > \frac{1}{2}$.

Introduction
An overview on statistical preference
Basic notions
Relationship with stochastic dominance
Statistical preference in normal distributions
Future works
A statistical test
Imprecise probabilities
Final remarks

Example



- $E(A) < E(B)$.
- $Q(A, B) > \frac{1}{2}$.

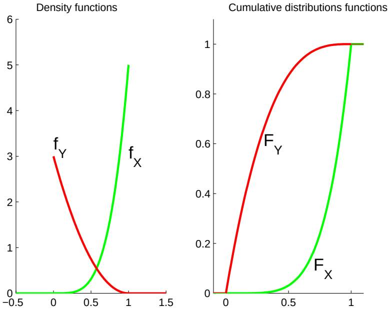
It seems to be more coherent to choose the dice A.

First degree stochastic dominance

Let X and Y be two random variables, X stochastically dominates Y by first degree if

$$F_X(t) \leq F_Y(t) \quad \text{for every } t.$$

It is denoted by $X \geq_{FSD} Y$.

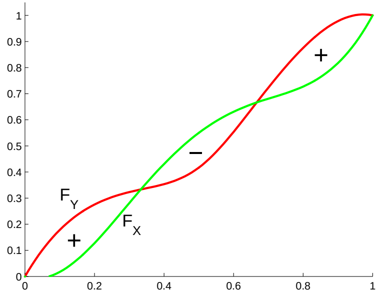


Second degree stochastic dominance

Let X and Y be two random variables, X stochastically dominates Y by second degree if

$$\int_{-\infty}^{t} F_X(x)\mathrm{dx} \leq \int_{-\infty}^{\mathrm{t}} \mathrm{F_Y(y)dy} \quad \text{for every } \mathrm{t}.$$

It is denoted by $X \geq_{SSD} Y$.



Statistical preference Vs stochastic dominance

Statistical preference has the following advantages over stochastic dominance:

- It takes into account the relationship between the random variables.

Statistical preference Vs stochastic dominance

Statistical preference has the following advantages over stochastic dominance:

- It takes into account the relationship between the random variables.
- It always gives us a solution.

Statistical preference Vs stochastic dominance

Statistical preference has the following advantages over stochastic dominance:

- It takes into account the relationship between the random variables.
- It always gives us a solution.
- It gives us a degree of preference of one of the variables over the other.

Non independent random variables

It is known that for every pair of random variables X and Y there exists a copula C such that

$$F_{X,Y}(x,y) = C(F_X(x), F_Y(y)),$$

where a copula is an application

$$C : [0,1]^2 \rightarrow [0,1] \quad \text{such that}$$

- $C(x,0) = C(0,x) = 0$ and $C(x,1) = C(1,x) = x$ for every $x \in [0,1]$.
- $C(x_1,y_1) + C(x_2,y_2) \geq C(x_1,y_2) + C(x_2,y_1)$, for every $x_1, x_2, y_1, y_2 \in [0,1]$.

Non independent random variables

It is known that for every pair of random variables X and Y there exists a copula C such that

$$F_{X,Y}(x,y) = C(F_X(x), F_Y(y)),$$

where a copula is an application

$$C : [0,1]^2 \to [0,1] \quad \text{such that}$$

- $C(x,0) = C(0,x) = 0$ and $C(x,1) = C(1,x) = x$ for every $x \in [0,1]$.
- $C(x_1,y_1) + C(x_2,y_2) \geq C(x_1,y_2) + C(x_2,y_1)$, for every $x_1, x_2, y_1, y_2 \in [0,1]$.

$$\boxed{\max\{ x+y-1,0 \} \leq C(x,y) \leq \min\{x,y\} \quad \forall (x,y) \in [0,1]^2.}$$

Non independent random variables

A particular type of copulas are the archimedean copulas. For them there exists a function $\varphi : [0,1] \to [0,1]$ such that $\varphi(1) = 0$, $\varphi'(x) < 0$ and $\varphi(x)'' > 0$, and then:

$$C(x,y) = \varphi^{[-1]}\left(\varphi(x) + \varphi(y)\right).$$

Non independent random variables

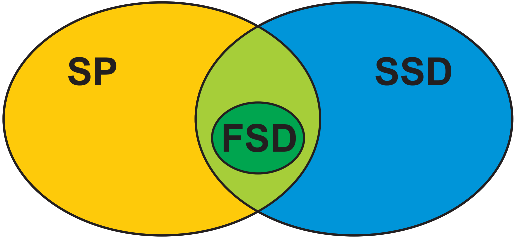
A particular type of copulas are the archimedean copulas. For them there exists a function $\varphi : [0,1] \to [0,1]$ such that $\varphi(1) = 0$, $\varphi'(x) < 0$ and $\varphi(x)'' > 0$, and then:

$$C(x,y) = \varphi^{[-1]}\left(\varphi(x) + \varphi(y)\right).$$

Example: $C(x,y) = x \cdot y$ is an archimedean copula:

$$\varphi(x) = -\log x$$

Statistical preference Vs stochastic dominance



This figure holds in the following situations:

- If X and Y are independent.
- If X and Y are continuous and they are coupled by an archimedean copula.

Normal distributions

- Normal distribution is one of the most important distributions in statistics.
- It appears in some usual real experiments.
- What is the behavior of statistical preference for normal distribution?

Unidimensional normal distribution

Theorem (De Schuymer et al. (2005))

Let X and Y two random variables with normal distributions, $X \equiv \mathcal{N}(\mu_1, \sigma_1)$ and $Y \equiv \mathcal{N}(\mu_2, \sigma_2)$. Then

- $Q(X,Y) = \Phi\left(\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right)$, *where Φ is the cumulative distribution function of a standard normal distribution.*
- $X \geq_{SP} Y \Leftrightarrow \mu_1 \geq \mu_2$.

Bidimensional normal distribution

Theorem

Let $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ *be a bidimensional random vector with normal distribution:*

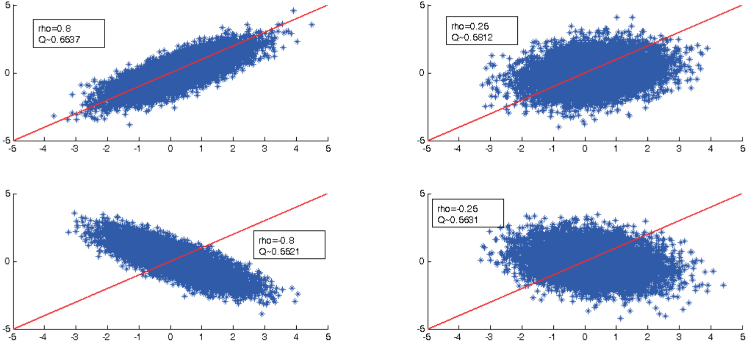
$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \equiv \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}\right).$$

Then:

- $Q(X_1, X_2) = \Phi\left(\frac{\mu_1-\mu_2}{\sqrt{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}}\right)$.
- $X_1 \geq_{SP} X_2 \Leftrightarrow \mu_1 \geq \mu_2$.

Bidimensional normal distribution

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \equiv \mathcal{N}\left(\begin{pmatrix} 0{,}25 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right)$$



Overview

An statistical test

Let X and Y be two independent random variables.

$$X \to x_1, \ldots, x_n$$
$$Y \to y_1, \ldots, y_m$$

An statistical test

Let X and Y be two independent random variables.

$$X \to x_1, \ldots, x_n$$
$$Y \to y_1, \ldots, y_m$$

$$X \stackrel{d}{=} Y??$$

An statistical test

If $X \stackrel{d}{=} Y$, then $Q(X,Y) = \frac{1}{2}$. In particular, it should happen that $Q(\overrightarrow{x}, \overrightarrow{y}) \approx \frac{1}{2}$.

We can define a critical region:

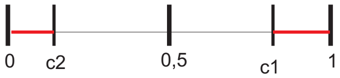
$$\left\{ (\overrightarrow{x}, \overrightarrow{y}) \mid \frac{1}{2} \leq c_1 \leq Q(\overrightarrow{x}, \overrightarrow{y}) \text{ or } Q(\overrightarrow{x}, \overrightarrow{y}) \geq c_2 \geq \frac{1}{2} \right\}$$

An statistical test

If $X \overset{d}{=} Y$, then $Q(X,Y) = \frac{1}{2}$. In particular, it should happen that $Q(\overrightarrow{x}, \overrightarrow{y}) \approx \frac{1}{2}$.

We can define a critical region:

$$\left\{ (\overrightarrow{x}, \overrightarrow{y}) \mid \frac{1}{2} \leq c_1 \leq Q(\overrightarrow{x}, \overrightarrow{y}) \text{ or } Q(\overrightarrow{x}, \overrightarrow{y}) \geq c_2 \geq \frac{1}{2} \right\}$$

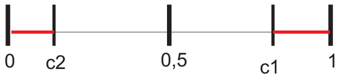


An statistical test

If $X \stackrel{d}{=} Y$, then $Q(X,Y) = \frac{1}{2}$. In particular, it should happen that $Q(\overrightarrow{x}, \overrightarrow{y}) \approx \frac{1}{2}$.

We can define a critical region:

$$\left\{ (\overrightarrow{x}, \overrightarrow{y}) \mid \frac{1}{2} \leq c_1 \leq Q(\overrightarrow{x}, \overrightarrow{y}) \text{ or } Q(\overrightarrow{x}, \overrightarrow{y}) \geq c_2 \geq \frac{1}{2} \right\}$$



What is the distribution of Q??

Imprecise data

- In some situations, there can be some loss of information.
- For example, it is possible that the distributions of the random variables are not well defined.
- How can we compare two random variables in that situation?

Imprecise probabilities

We hope to...

- Translate the concept of statistical preference to an imprecision context.
- Consider models like Choquet capacities, belief functions,...

Overview

Remarks

Our developments in the concept of statistical preference have been made in collaboration with D.Martinetti and supervised by professors S.Díaz and S.Montes.

http://eio.epv.uniovi.es/diedra/index.htm

Bibliography

Stochastic Dominance:

H. Levy, *Stochastic Dominance*, Kluwer Academic Publishers, 1998.

A. Müller and D. Stoyan, *Comparison Methods for Stochastic Models and Risk*, John Wiley & Sons, LTD, 2002.

Bibliography

The origin of statistical preference:

- B. De Schuymer, H. De Meyer, B. De Baets and S. Jenei, *On the cycle-transitivity of the dice model*, Theory and decision **54**, 261–285, 2003.
- B. De Schuymer, H. De Meyer and B. De Baets, *A fuzzy approach to stochastic dominance of random variables*, Lecture notes in Artificial Intelligence **2715**, 253–260, 2003.

Bibliography

Our research on statistical preference:

- I. Montes, D. Martinetti, S. Díaz and S. Montes, *Comparison of random variables coupled by archimedean copulas*, In: SMPS 2010.
- I. Montes, D. Martinetti, S. Díaz and S. Montes, *Study of the statistical preference in two-dimensional normal distributions*, In: SEIO 2010.
- I. Montes, D. Martinetti and S. Díaz, *Statistical preference as a tool in Consensus Process*, In: Consensual Processes, Springer 2010.