On the moments and the distribution of the Choquet integral

Ivan Kojadinovic Jean-Luc Marichal

Kindly presented by Michel Grabisch

Department of Statistics, University of Auckland

Applied Mathematics Unit, University of Luxembourg

EUSFLAT 2007

The Choquet integral

The **Choquet integral** of $x \in \mathbb{R}^n$ w.r.t. a capacity ν on N is defined by

$$C_\nu(x) := \sum_{i=1}^{n} p_i^{\nu,\sigma}\, x_{\sigma(i)},$$

where σ is a permutation on N such that $x_{\sigma(1)} \geqslant \cdots \geqslant x_{\sigma(n)}$, where

$$p_i^{\nu,\sigma} := \nu_i^\sigma - \nu_{i-1}^\sigma, \qquad \forall\, i \in N,$$

and where $\nu_i^\sigma := \nu\big(\{\sigma(1), \ldots, \sigma(i)\}\big)$ for any $i = 0, \ldots, n$. In particular, $\nu_0^\sigma := 0$.

The Choquet integral

Let $\mathfrak{S}_n$ denote the set of permutations on N.

The Choquet integral can therefore be regarded as a **piecewise linear function** that coincides with a weighted arithmetic mean on each n-dimensional region

$$R_\sigma := \{x \in \mathbb{R}^n \mid x_{\sigma(1)} \geqslant \cdots \geqslant x_{\sigma(n)}\} \qquad (\sigma \in \mathfrak{S}_n),$$

whose union covers $\mathbb{R}^n$.

We state hereafter immediate **relationships** between the moments (resp. the c.d.f.) of the **Choquet integral** and the moments (resp. the c.d.f.) of **linear combination of order statistics**.

Distributional relationships with linear combination of order statistics

Let $X_1, \ldots, X_n$ be a random sample from a continuous distribution with p.d.f. f.

Let $X_{1:n} \leqslant \cdots \leqslant X_{n:n}$ denote the corresponding order statistics.

Furthermore, let $Y_\nu := C_\nu(X_1, \ldots, X_n)$ and let h be any function.

By definition of the expectation, we have

$$\begin{aligned}
\mathbf{E}[h(Y_\nu)] &= \int_{\mathbb{R}^n} h(C_\nu(x_1, \ldots, x_n)) \prod_{i=1}^{n} f(x_i)\mathrm{d}x_i \\
&= \sum_{\sigma \in \mathfrak{S}_n} \int_{R_\sigma} h\left(\sum_{i=1}^{n} p_i^{\nu,\sigma}\, x_{\sigma(i)}\right) \prod_{i=1}^{n} f(x_i)\mathrm{d}x_i
\end{aligned}$$

Distributional relationships with linear combination of order statistics

Using the well-known fact that the joint p.d.f. of $X_{n:n} \geqslant \cdots \geqslant X_{1:n}$ is

$$n! \prod_{i=1}^{n} f(x_i), \qquad x_n \geqslant \cdots \geqslant x_1,$$

we obtain

$$\mathbf{E}[h(Y_\nu)] = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \mathbf{E}\left[h\left(\sum_{i=1}^{n} p_i^{\nu,\sigma} X_{n-i+1:n} \right) \right] = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \mathbf{E}[h(Y_\nu^\sigma)].$$

where

$$Y_\nu^\sigma := \sum_{i=1}^{n} p_i^{\nu,\sigma} X_{n-i+1:n}$$

are **linear combinations of order statistics**.

Distributional relationships with linear combination of order statistics

The special cases

$$h(x) = x^r,\ [x - \mathbf{E}(Y_\nu)]^r, \text{ and } e^{tx}$$

provide similar relationships, respectively, for **raw moments**, **central moments**, and **moment-generating functions**.

Now, consider the **minus** (resp. **plus**) **truncated power function** x_-^n (resp. x_+^n) defined to be x^n if $x < 0$ (resp. $x > 0$) and zero otherwise.

Distributional relationships with linear combination of order statistics

Given $y \in \mathbb{R}$, taking $h(x) = (x - y)^0_-$ provides a relationship between the c.d.f. F_ν of Y_ν and those of the random variables $\sum_{i=1}^n p_i^{\nu,\sigma} X_{n-i+1:n}$. Indeed, we clearly have

$$F_\nu(y) := \mathbf{Pr}[Y_\nu \leqslant y] = \mathbf{E}[(Y_\nu - y)^0_-],$$

and, denoting by F_ν^σ the c.d.f. of Y_ν^σ,

$$F_\nu^\sigma(y) := \mathbf{Pr}\left[\sum_{i=1}^n p_i^{\nu,\sigma} X_{n-i+1:n} \leqslant y\right] = \mathbf{E}\left[\left(\sum_{i=1}^n p_i^{\nu,\sigma} X_{n-i+1:n} - y\right)^0_-\right].$$

Distributional relationships with linear combination of order statistics

This immediately gives

$$F_\nu(y) = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} F_\nu^\sigma(y).$$

As one could have expected from the definition of the Choquet integral, the determination of the moments and the distribution functions of the Choquet integral is closely related to the determination of the moments and the distribution functions of linear combinations of order statistics.

The uniform case: main results

Mathematical tools: divided differences, Hermite-Genocchi formula.

There holds

$$F_\nu(y) = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \Delta[(\cdot - y)_-^n : \nu_0^\sigma, \ldots, \nu_n^\sigma]$$

where $\Delta[(\cdot - y)_-^n : \nu_0^\sigma, \ldots, \nu_n^\sigma]$ is the divided difference of $x \mapsto (x-y)_-^n$ at points $\nu_0^\sigma, \ldots, \nu_n^\sigma$.

From a practical perspective: there exists $O(n^2)$ algorithms for computing divided differences. Implemented in **kappalab**.

Also, a result giving the **momemts of the Choquet integral in closed form**.

The non-uniform case: main results

Mathematical tools: approximations of moments of order statistics

Main result: approximations of the 2 first moments of Choquet integral.

From a practical perspective: efficient algorithms for computing the **moments** in the **normal case** and in the **exponential case**. Implemented in **kappalab**.

Thanks very much Michel!