Exchangeability[1]

[1]Based on Migon and Gamerman's (1999) *Statistical Inference: An Integrated Approach*.

Outline

Exchangeability

Exchangeability is a very important concept introduced by de Finetti(1973).

Let $\mathcal{K} = \{k_1, \ldots, k_n\}$ be a permutation of $\{1, \ldots, n\}$. Random quantities $X_1, \ldots, X_n$ assuming only the values 0-1 are exchangeable if the n! permutations $(X_{k_1}, \ldots, X_{k_n})$ have the same n-dimensional probability distribution.

The random quantities $x_1, \ldots, x_n$ are said to be judged (finitely) exchangeable under a probability measure P if the implied joint degree of belief distribution satisfies

$$P(x_1, \ldots, x_n) = P(x_{\pi(1)}, \ldots, x_{\pi(n)})$$

for all permutations π defined on the set $\{1, \ldots, n\}$.

Definition: An infinite sequence of random quantities is exchangeable if any finite sub-sequence is exchangeable.

$\Rightarrow$ It is weaker than but as useful as independence.

$\Rightarrow$ All marginal distributions must be the same!

$\Rightarrow$ A sequence (finite or not) of iid random quantities is trivially exchangeable, although the reciprocal is not true, in general.

Example 1

Consider an urn with m balls:

- r with number 1
- $m - r$ with number 0.

Balls are drawn from the urn, one at time, without replacement.

Let X_k denote the number associated with the k-th ball selected. Then, $X_1, \ldots, X_n$, $n \leq m$ is an exchangeable sequence, but the X_i's are not independent.

Example 2

Let $X_1, X_2, \ldots$ be a sequence of Bernoulli trials with unknown success probability θ.

The classical assumption is that the $X_k's$ are iid.

For a Bayesian, if θ is unknown, the information content of the j-th observation can modify your belief about the distribution of X_k.

If the experiments are judged similar in some sense, the hypothesis of exchangeability is acceptable, while marginal independence is not.

De Finetti's (1937) Theorem

The relevance of the concept of exchangeability is due to the fact that, although it is based on weak assumptions, it allows one to state a very powerful result, known as the De Finetti's Representation Theorem.

Theorem: To all infinite sequences of exchangeable random quantities $\{X_n, n = 1, 2, \ldots\}$ assuming values $\{0, 1\}$ there corresponds a distribution F in (0,1) such that for all n and $k \leq n$,

$$P(Y_n = k) = \int_0^1 \theta^k (1-\theta)^{n-k} dF(\theta)$$

where $Y_n = \sum_{i=1}^n X_i$.

Note that, due to the exchangeability assumption, any sequence with k 1's and $n-k$ 0's also admits a representation according to the above theorem.

The theorem does not say anything about F

1. A degenerate distribution: $P(\theta = \theta_0) = 1$, for some θ_0:
$$P(Y_n = k) = \theta_0^k (1-\theta_0)^{n-k}$$

2. A discrete distribution: $P(\theta = \theta_i) = p_i$, if $\theta = \theta_i$, $i = 1, \dots, s$ with $p_i \geq 0$ and $\sum p_i = 1$:
$$P(Y_n = k) = \sum_{i=1}^{s} p_i \theta_i^k (1-\theta_i)^{n-k}$$

3. A continuous distribution: $\theta \sim Beta(a, b)$:
$$\begin{aligned} P(Y_n = k) &= \int_0^1 \theta^k (1-\theta)^{n-k} \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)} d\theta \\ &= \frac{1}{B(a,b)} \int_0^1 \theta^{a+k-1}(1-\theta)^{b+n-k-1} d\theta \\ &= \frac{B(a+k, b+n-k)}{B(a,b)} \end{aligned}$$

Symmetry + invariance under linear transformations

⇒ The exchangeability concept has already been extended to many other distributions with the inclusion of some additional hypotheses; see Bernardo and Smith (1994) for a review of the main results.

⇒ Symmetry + invariance under linear transformations:

$$p(x_1, \cdots, x_n) = \int_0^\infty \int_{-\infty}^\infty \prod p_N(x_i; \theta, \sigma^2) dF(\theta, \sigma^2)$$

where θ varies in $\Re$, σ varies in $\Re^+$.

Partial exchangeability

In this case the exchangeability holds only under certain conditions.

For example, we can define some groups of variables where exchangeability is valid only within each group.

Definition: Let $\{X_i, i = 1, 2, \ldots, n\}$ be any sequence of random quantities and $\mathcal{K}$ be any permutation of $\{1, 2, \ldots, n\}$. We say that $\boldsymbol{X}$ is partially exchangeable if there are quantities $\{Z_i, i = 1, 2, \ldots, n\}$ such that the distribution of $(\boldsymbol{X} \mid \boldsymbol{Z})$ is the same as that of $(\boldsymbol{X}_{\mathcal{K}} \mid \boldsymbol{Z}_{\mathcal{K}})$,, for any permutation $\mathcal{K}$, where for any vector $\boldsymbol{c} = (c_1, \cdots, c_n)$, $\boldsymbol{c}_{\mathcal{K}} = (c_{k_1}, \cdots, c_{k_n})$.

$\Rightarrow$ When the indexes of the $X_i's$ are permuted, the resulting vector will have the same conditional distribution, as far as the same permutation is applied to the Z_i's indexes.

$\Rightarrow$ This is weaker than exchangeability.

Corollary

Given the conditions of Proposition 4.1 (B&S,1994),

$$p(x_1 + \ldots + x_n = y_n) = \int_0^1 \binom{n}{y_n} \theta^{y_n}(1-\theta)^{n-y_n} dQ(\theta)$$

Corollary: If $x_1, x_2, \ldots$ is an infinitely exchangeable sequence of 0-1 random quantities with probability measure P, the conditional probability function of $x_{m+1}, \ldots, x_n$ given $x_1, \ldots, x_m$, denoted by $p(x_{m+1}, \ldots, x_n | x_1, \ldots, x_m)$, has the form

$$\int_0^1 \prod_{i=m+1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} dQ(\theta|x_1, \ldots, x_m), \quad 1 \leq m < n,$$

where

$$dQ(\theta|x_1, \ldots, x_m) = \frac{\prod_{i=1}^m \theta^{x_i}(1-\theta)^{1-x_i} dQ(\theta)}{\int_0^1 \prod_{i=1}^m \theta^{x_i}(1-\theta)^{1-x_i} dQ(\theta)}$$

and

$$Q(\theta) = \lim_{n\to\infty} P\left(\frac{y_n}{n} \leq \theta\right).$$