Exchangeability

In probability theory, the **random variables** Y_1, ..., Y_N are said to be *exchangeable* (or *permutable* or *symmetric*) if their joint distribution $F(y_1, \ldots, y_N)$ is symmetric; that is, if F is invariant under permutation of its arguments, so that

$$F(z_1, \ldots, z_N) = F(y_1, \ldots, y_N)$$

whenever $z_1, \ldots, z_N$ is a permutation of $y_1, \ldots, y_N$. There is a related epidemiologic usage which is described in the article on **confounding**. In many ways, sequences of exchangeable random variables play a role in subjective **Bayesian** theory analogous to that played by independent identically distributed (iid) sequences in classical frequentist theory. In particular, the assumption that a sequence of random variables is exchangeable allows the development of inductive statistical procedures for inference from observed to unobserved members of the sequence [1–3, 5, 6, 9].

Exchangeable random variables are identically distributed, and iid variables are exchangeable. Now suppose that $Y_1, \ldots, Y_N$ are iid given an unknown parameter θ that indexes their joint distribution (*see* **Identifiability**). Such variables will not be unconditionally independent when θ is a random variable, but will be exchangeable. Consider, for example, the case in which $Y_1, \ldots, Y_N$ have a joint density. The unconditional density of $Y_1, \ldots, Y_N$ will be

$$\begin{aligned} f(y_1, \ldots, y_N) &= \int_\theta f(y_1, \ldots, y_N|\theta)\ \mathrm{d}F(\theta) \\ &= \int \prod_i f(y_i|\theta)\ \mathrm{d}F(\theta). \end{aligned}$$

Exchangeability of $Y_1, \ldots, Y_N$ follows from the identity of the marginal densities in the product. However, given that these densities depend on θ, the integral and product cannot be interchanged, so that $f(y_1, \ldots, y_N) \neq \prod_i f(y_i)$. We thus have that a mixture of iid sequences is an exchangeable sequence, but not iid except in trivial cases.

One consequence of this result is that the usual procedures for generating a sequence $Y_1, \ldots, Y_N$ of iid random variables for inference on an unknown parameter (such as Bernoulli trials of **binary data** with unknown success probability) generate only an exchangeable sequence when the parameter is generated randomly and the sequence is considered unconditionally. From a Bayesian perspective, this means that, when your uncertainty about the parameter is integrated with your uncertainty about the realizations of $Y_1, \ldots, Y_N$, the latter are (for you) exchangeable but dependent. This subjective dependence is immediately clear if you consider (say) tossing a coin $N = 99$ times, with Y_i the indicator of heads on toss i. Starting from a **uniform prior** for the chance of heads, you should have $\Pr(Y_{99} = 1) = 1/2$ before seeing any toss, but

$$\Pr\left(Y_{99} = 1 \middle| \sum_{i=1}^{98} Y_i = 98\right) = 0.99$$

after seeing the first 98 tosses come up heads [8].

A generalization important for statistical modeling is *partial* or *conditional* exchangeability [2, 3]. For example, suppose that the sequence $Y_1, \ldots, Y_N$ is partitioned into disjoint subsequences. Then the sequence is said to be partially exchangeable given the partition if each subsequence can be permuted without changing the joint distribution. If the Y_i represent survival times within a cohort of male stroke patients, then a judgment of unconditional exchangeability of the Y_i would be unreasonable if the patient ages were known, because age is a known predictor of survival time. Nonetheless, one might regard the survival times as partially exchangeable, given age, if no further prognostically relevant partitioning was possible based on the available data.

While exchangeability is weaker than iid, de Finetti [[3], Chapter 11] proved that finite subsequences of an infinite exchangeable sequence of Bernoulli (binary) variates must have representations as mixtures of iid Bernoulli sequences – a partial converse of the fact that any mixture of iid sequences is an exchangeable sequence. More precisely, suppose that $Y_1, Y_2, \ldots$ is an infinite sequence of exchangeable Bernoulli variates (that is, every finite subsequence of the sequence is exchangeable), and that θ is the limit of $(Y_1 + \cdots + Y_n)/n$ as n goes to infinity. De Finetti showed that there exists a distribution function $P(\theta)$ for θ such that, for all n,

$$\begin{aligned} \Pr(Y_1 = y_1, \ldots, Y_n = y_n) &\equiv \Pr(y_1, \ldots, y_n) \\ &= \int_0^1 \theta^s (1-\theta)^{n-s}\ \mathrm{d}P(\theta), \end{aligned} \qquad (1)$$

where $s = y_1 + \cdots + y_n$. Many Bayesian statisticians find this theorem helpful, because it partially specifies the form of the predictive probability $\Pr(y_1, \ldots, y_n)$ when $Y_1, \ldots, Y_n$ can be considered a subsequence of an infinite exchangeable sequence.

In the representation shown in (1), $P(\theta)$ is recognizable as the *prior distribution* for θ, a distribution that may be developed from what is known about θ before the Y_i are observed. As noted in [7], however, the strength of the theorem's conclusion is easy to overstate: it does *not* imply that all binary data must be analyzed using the representation shown in (1); it merely says that if you judge $Y_1, Y_2, \ldots$ to be an exchangeable sequence, then there is a $P(\theta)$ that allows you to use (1) to specify $\Pr(y_1, \ldots, y_n)$.

Finite versions of the theorem [4] show that, if $Y_1, \ldots, Y_n$ is the start of an exchangeable Bernoulli sequence $Y_1, \ldots, Y_N$ and n/N is small enough, then $\Pr(y_1, \ldots, y_n)$ may be approximately expressed as in (1), with the approximation improving as n/N approaches zero. There are further generalizations to exchangeable sequences of **polytomous** variates, as well as exchangeable sequences of continuous variates [4]. The latter generalization requires a prior distribution on the space of continuous distributions, however, which can be much harder to specify than a prior for a vector of **multinomial** parameters, and which may lead to intractable computational problems [5].

References

[1] Bernardo, J.M. & Smith, A.F.M. (1994). *Bayesian Theory*. Wiley, New York.

[2] De Finetti, B. (1937). Foresight: its logical laws, its subjective sources, reprinted in *Studies in Subjective Probability*, H.E. Kyburg & H.E. Smokler, eds. Wiley, New York, 1964.

[3] De Finetti, B. (1974). *Theory of Probability* (two vols). Wiley, New York.

[4] Diaconis, P. & Freedman, D. (1980). Finite exchangeable sequences, *Annals of Probability* **8**, 745–764.

[5] Draper, D. (1995). Assessment and propagation of model uncertainty (with discussion), *Journal of the Royal Statistical Society, Series B* **57**, 45–97.

[6] Draper, D., Hodges, J., Mallows, C. & Pregibon, D. (1993). Exchangeability and data analysis (with discussion), *Journal of the Royal Statistical Society, Series A* **196**, 9–37.

[7] Freedman, D.A. (1995). Some issues in the foundations of statistics (with discussion), *Foundations of Science* **1**, 19–83.

[8] Good, I.J. (1983). *Good Thinking*. University of Minnesota Press, Minneapolis.

[9] Lindley, D.V. & Novick, M.R. (1981). The role of exchangeability in inference, *Annals of Statistics* **9**, 45–58.

(*See also* **Foundations of Probability**; **Subjective Probability**)

SANDER GREENLAND & DAVID DRAPER