Exchangeable Sequences, Polya's Urn and De-Finetti's Theorem

Mohammad Emtiyaz Khan
CS,UBC

October 6, 2007

Abstract

We discuss the notion of exchangeability and give an example to demonstrate the fact that the condition of exchangeability is weaker than the condition of identical and independent distribution. We discuss implications of exchangeability using De-Finetti's theorem.

1 Exchangeability

An infinite sequence $\{x_1, x_2, x_3, \dots\}$ of random variables is said to be **exchangeable** if for any finite cardinal number n and any two finite sequences $i_1, \dots, i_n$ and $j_1, \dots, j_n$, the two sequences

$$x_{i_1}, \dots, x_{i_n} \text{ and } x_{j_1}, \dots, x_{j_n}$$

have the same probability distribution [1]. Exchangeability defines one kind of "similarity" between measurements $\{x_n\}$ [2]. For example, when the subscripts (or 'labels') are uninformative, the order of measurements doesn't matter and it is reasonable to assume that the distribution is invariant to permutation. We can see that this condition may not always hold (e.g. time-series), however there are many cases where exchangeability can be safely assumed.

The condition of exchangeability is stronger than the assumption of identical marginal distributions (the definition with $n = 1$ turns into identical marginal distribution condition). Also exchangeability is a weaker condition than i.i.d. assumption, which we show next with an example.

2 Polya's Urn

First we establish that Polya's urn give rise to exchangeable sequences. Consider an urn containing r red balls and b blue balls. Now consider the following experiment: we draw a ball, note its color, and replace the ball back in the urn along with c additional balls of the same color. Let us denote the event of observing a red ball at i^{th} trial by R_i, and similarly observing a blue ball by B_i. Then observing two red balls in three draws is same regardless of the sequence it is observed:

$$P(B_1, R_2, R_3) = P(B_1)P(R_2|B_1)P(R_3|R_2, B_1) = \frac{b}{r+b}\frac{r}{r+b+c}\frac{r+c}{r+b+c}$$
$$P(R_1, B_2, R_3) = P(R_1)P(B_2|R_1)P(R_3|B_2, R_1) = \frac{r}{r+b}\frac{b}{r+b+c}\frac{r+c}{r+b+c}$$
$$P(R_1, R_2, B_3) = P(R_1)P(R_2|R_1)P(B_3|R_2, R_1) = \frac{r}{r+b}\frac{r+c}{r+b+c}\frac{b}{r+b+c}$$

The probability only depends on the number of events observed. It is easy to extend this result to any finite number of events and show that the joint probability of those events depends only on number of events of observing red (or blue) balls. It is clear that the draws from this experiment form an exchangeable sequence.

Now we show that the marginal distributions are identical but events are not independent. The later is obvious as the joint probability can not be factorized. The marginal distribution of observing a red ball and blue ball in any draw is identical, for example, when $n = 2$,

$$\begin{aligned} P(R_2) &= P(R_1)P(R_2|R_1) + P(B_1)P(R_2|B_1) \\ &= \frac{r}{b+r}\frac{r+c}{r+b+c} + \frac{b}{b+r}\frac{r}{r+b+c} \\ &= \frac{r(r+c)+br}{(b+r)(r+b+c)} \\ &= \frac{r}{b+r} \end{aligned} \quad (1)$$

and similarly,

$$P(B_2) = \frac{b}{b+r} \quad (2)$$

In general with induction for any n, $P(R_n) = r/(b+r)$ and $P(B_n) = b/(b+r)$. Hence the sequences are identical, but not independent. We conclude that the notion of exchangeability is stronger than the notion of identical distribution but weaker than i.i.d.

3 De-Finetti's Theorem

For a more precise and detailed discussion see [2] and [3]. According to De-Finetti's theorem, if $\{x_1, x_2, \ldots\}$ is an exchangeable sequence, then there exists a parametric model $p(x_i|\theta)$ with some parameter θ which is the limit of a function of the x_i's as $n \to \infty$. Also there exist a probability distribution for θ with which the joint probability of the sequence can be derived as follows:

$$p(x_1, \ldots, x_n) = \int_\Theta \prod_{i=1}^{n} p(x_i|\theta)p(\theta)d\theta \quad (3)$$

In other words, De-Finetti's theorem (roughly) says that if a sequence is exchangeable then any finite subset can be considered as a random sample of a model $p(x_i|\theta)$ and there exist a prior distribution for θ which justifies a Bayesian approach. Examples and concept of partial exchangeability, which lead to hierarchical Bayesian model, are discussed in [2].

Acknowledgement Thanks to Kevin Murphy for encouraging me to write this document.

References

[1] http://en.wikipedia.org/wiki/De_Finetti's_theorem

[2] Jose M. Bernardo, The Concept of Exchangeability and its Applications, Far East J. Math. Sci., available at http://www.uv.es/ bernardo/Exchangeability.pdf

[3] http://www-stat.stanford.edu/c̃gates/PERSI/courses/stat_121/lectures/exch/