1 Comments on De Finetti Theorem

Theorem 1 Exchangeability: \( p(x_1, \ldots, x_n) = p(x_{\pi(1)}, \ldots, x_{\pi(n)}) \) for any permutation \( \pi \) and for all \( n \), iff

\[
p(x_1, \ldots, x_n) = \int dp(\theta) \prod_{i=1}^{n} p(x_i|\theta)
\]

for some underlying distribution \( p(\theta) \), and \( p(x_i|\theta) \).

Notes:

- The underlying distribution \( p(\theta) \) could be, for example, the Dirichlet process, \( G \sim DP(\cdot) \). \( G \) is a measure on \( \theta \) but also a random variable. First pick \( G \), and then for each fixed \( G \) sample the \( x_i \)'s according to \( p(x_i|G) \). So an observer observing \( x \) will not know the underlying \( G \) or which sample indices correspond to which \( x_i \)'s.
- Exchangeability is not equivalent to independent and identically distributed (iid). The graphical model representing exchangeability shows that \( \{x_1, \ldots, x_n\} \) are not independent.

\[p(\theta) \quad \Theta \quad p(x_i|\theta) \quad \begin{array}{c}x_i \quad n \end{array} \]

\[\equiv \quad \Theta \quad \begin{array}{c}x_1 \quad x_2 \quad \ldots \quad x_n \end{array}\]

Figure 1: Exchangeability \( \neq \) iid.

2 Dirichlet Process Mixtures

We have already discussed three representations of the Dirichlet Process: the stick breaking representation, the chinese restaurant process, and the urn model.
Figure 2: Dirichlet Process mixture.

- In the stick breaking representation, \( G \sim DP(\alpha, G_0) \) can be written as an infinite sum of spikes \( G = \sum_{i=0}^{\infty}\pi_i\delta_{\theta_i} \), where \( \theta_i|\alpha, G_0 \sim G_0, \pi'_i|\alpha, G_0 \sim Beta(1, \alpha) \), and \( \pi_i = \pi'_i\prod_{j=1}^{i-1}(1 - \pi_j) \). The \( \pi_i \)'s can be represented as the remaining length of a stick after \( i - 1 \) pieces have been broken off each with size according to the Beta distribution. Here \( G_0 \) is a measure on \( \theta_i \) and \( \alpha \) determines how closely the histogram of spikes represents \( G_0 \).

- In the chinese restaurant process, \( x_1 \) starts a new table and picks \((\mu, \sigma^2)\) from the prior, and each subsequent \( x_i \) joins a table if it is close to \((\mu, \sigma^2)\), or else starts a new table with a new \((\mu, \sigma^2)\) chosen from the prior. Here \( G_0 \) is a measure on \((\mu, \sigma^2)\).

- In the urn model,

\[
x_i|\theta_i \sim F(\theta_i)
\]

\[
\theta_i|G \sim G
\]

\[
G \sim DP(\alpha, G_0)
\]

\[
\theta_i|\theta_1, \ldots, \theta_{i-1} \sim \frac{\alpha}{i - 1 + \alpha}G_0 + \frac{1}{i - 1 + \alpha}\sum_{j=1}^{i-1}\delta(\theta_j)
\]

We now introduce a fourth representation of Dirichlet Processes as the infinite limit of a finite mixture (see the Radford Neal paper).

- 

\[
x_i|c, \phi \sim F(\phi_{c_i})
\]

\[
c_i|p \sim Multinomial(p_1, \ldots, p_k)
\]

\[
\phi_c \sim G_0
\]

\[
p_1, \ldots, p_k \sim Dirichlet(\frac{\alpha}{k}, \ldots, \frac{\alpha}{k})
\]
Figure 3: Finite mixture.