Statistical inference: De Finetti’s Theorem

Timo Koski

20.01.2010
The lecture deals with personalistic probability modelling by De Finetti’s theorem.


Definition

An infinite sequence of random variables \((X_1, X_2, \ldots, X_n, \ldots)\) is said to be **infinitely exchangeable** under probability measure \(p\), if the joint probability of every finite subsequence \((X_{n_1}, X_{n_2}, \ldots, X_{n_k})\) satisfies

\[
(X_{n_1}, X_{n_2}, \ldots, X_{n_k}) \overset{d}{=} (X_{\tau(n_1)}, X_{\tau(n_2)}, \ldots, X_{\tau(n_k)})
\]

for all permutations \(\tau\) defined on the set \(\{1, 2, 3, \ldots, k\}\).
Even the probability $p$ is called exchangeable.

The notion of exchangeability involves a judgement of complete symmetry among all the observables (‘potentially infinite number of observables’) under consideration. An infinite sequence of random variables $(X_1, X_2, \ldots, X_n, \ldots)$ is judged to be infinitely exchangeable.
Next we state and prove a famous representation theorem due to Bruno de Finetti. We prove it for a binary process. The proof below is due to Heath & Sudderth. There are several completely general proofs, see, e.g., (Schervish, Theory of Statistics, 1995). In a latter part of the lecture we use a key result proved found in R. Durrett: *Probability: Theory and Examples. Second Edition*. Duxbury Press, 1996, by a technique of reverse martingales, then completed by an more abstract measure theory argument from Schervish. de Finetti’s theorem is, as such, a result in probability theory. We include this in a course on statistical inference, because the theorem is a cornerstone of Bayesian statistical inference, and is a critique of objectivistic modes of statistical inference.
Proposition

Let \((X_1, X_2, \ldots, X_n, \ldots)\) be an infinitely exchangeable sequence of binary random variables with probability measure \(p\).

Then there exists a distribution function \(\Pi\) such that the joint probability \(p(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)\) has the form

\[
p(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \int_0^1 \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} \Pi(d\theta),
\]

where \(\Pi\) is the distribution function

\[
\Pi(\theta) = \lim_{n \to \infty} P\left(\frac{1}{n} \sum_{i=1}^n x_i \leq \theta\right)
\]
Proof of the Representation Theorem

We set for ease of writing

\[ p(x_1, x_2, \ldots, x_n) = p(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n). \]

Suppose that \( x_1 + \ldots + x_n = y_n \in \{0, 1, \ldots, n\} \). Then exchangeability gives

\[ p(x_1 + \ldots + x_n = y_n) = \binom{n}{y_n} p(x_{\tau(1)}, x_{\tau(2)}, \ldots, x_{\tau(n)}) \]

for any permutation \( \tau \) of \( \{1, 2, \ldots, n\} \), i.e.,

\[ p(x_{\tau(1)}, x_{\tau(2)}, \ldots, x_{\tau(n)}) = \frac{1}{\binom{n}{y_n}} p(x_1 + \ldots + x_n = y_n). \quad (1) \]
Proof of the Representation Theorem

For arbitrary $N \geq n \geq y_n \geq 0$ we have by marginalization

$$p(x_1 + \ldots + x_n = y_n)$$

$$= \sum_{y_N = y_n}^{N-(n-y_n)} p(x_1 + \ldots + x_n = y_n | x_1 + x_2 + \ldots + x_N = y_N) p(x_1 + \ldots + x_N = y_N)$$

$$= \sum_{y_N = y_n}^{N-(n-y_n)} \binom{y_N}{y_n} \binom{N - y_N}{n - y_n} p(x_1 + \ldots + x_N = y_N)$$
Proof of the Representation Theorem

The expression above for

\[ \sum_{y_N = y_n}^{N-(n-y_n)} p(x_1 + \ldots + x_n = y_n|x_1 + \ldots + x_N = y_N) p(x_1 + \ldots + x_N = y_N) \]

can be argued as follows. It follows by the assumption of exchangeability that given the event \(\{x_1 + \ldots + x_N = y_N\}\), all possible rearrangements of the \(y_n\) cases of ones in among the \(n\) places are equally likely.
Proof of the Representation Theorem

Thus we can think of an urn containing $N$ items, of which $y_N$ are ones, and $N - y_N$ are zeros. We pick $n$ items without replacement. Then

$$\binom{y_N}{y_n} \binom{N - y_N}{n - y_n} \binom{N}{n}$$

is the probability of obtaining $y_n$ ones and $n - y_n$ zeros, and the probability function of a hypergeometric distribution $\text{Hyp}(N, n, y_N)$, c.f. G. Blom, G. Englund et.al. kap. 7.3.)
Proof of the Representation Theorem

\[
\binom{y_N}{y_n} \binom{N - y_N}{n - y_n} = \binom{n}{y_n} \cdot \frac{(y_N)_{y_n} (N - y_N)_{n - y_n}}{(N)_n},
\]

where

\[(N)_n = N \cdot (N - 1) \cdot (N - 2) \cdots (N - (n - 1)) = \frac{N!}{(N - n)!}\]

so that \((y_N)_{y_n} = y_N (y_N - 1) \cdots (y_N - (y_n - 1))\) etc.
In other words, by (1) we have

\[ p\left(x_{\tau(1)}, x_{\tau(2)}, \ldots, x_{\tau(n)}\right) = \frac{1}{\binom{n}{y_n}} p\left(x_1 + \ldots + x_n = y_n\right) \]

\[ = \frac{1}{\binom{n}{y_n}} \sum_{y_N = y_n} \binom{n}{y_n} \cdot \frac{(y_N)^{y_n} (N - y_N)^{n-y_n}}{(N)^n} p\left(x_1 + \ldots + x_N = y_N\right) \]
Proof of the Representation Theorem

We define the function $\Pi_N (\theta)$, as a step function which is zero for $\theta < 0$ and has jumps of size $p \left( x_1 + x_2 + \ldots + x_N = y_N \right)$ at $\theta = y_N / N$, $y_N = 0, 1, \ldots, N$. 
Proof of the Representation Theorem

Thus we have

\[
p(x_1 + \ldots + x_n = y_n) = \binom{n}{y_n} \int_0^1 \frac{(\theta N)^{y_n} ((1 - \theta) N)^{n-y_n}}{(N)^n} d\Pi_N(\theta)
\]
Proof of the Representation Theorem: basic probability

As $N \to \infty$

$$\frac{(\theta N)_y}{(N)_n} \frac{((1 - \theta)N)_{n-y}}{n-y_n} \to \theta^n (1 - \theta)^{n-y_n}$$

uniformly in $\theta$. (Approximation of a hypergeometric probability by a binomial probability, c.f., G. Blom, G. Englund et.al. kap. 7.3.)
Approximation, Hyp($N, n, p$) (an argument due to Jan Grandell)

\[
\frac{\binom{N\theta}{k} \binom{N(1-\theta)}{n-k}}{\binom{N}{n}} = \frac{N\theta!}{k!(N\theta - k)!} \frac{N(1 - \theta)!}{(n - k)!\lbrack N(1 - \theta) - (n - k)\rbrack!} \frac{n!(N - n)!}{N!} \\
= \frac{n!}{k!(n - k)!} \frac{N\theta!(N(1 - \theta)!(N - n)!}{(N\theta - k)!\lbrack N(1 - \theta) - (n - k)\rbrack!N!} \\
\approx \frac{n!}{k!(n - k)!} \frac{(N\theta)^k (N(1 - \theta))^{n-k}}{N^n} = \binom{n}{k} \theta^k (1 - \theta)^{n-k}.
\]
Proposition

If $X$ is $\text{Hyp}(N, n, p)$-distributed with $n/N \leq 0.1$ then $X$ is approximatively $\text{Bin}(n, p)$. 

Proof of the Representation Theorem

By Helly’s theorem (see e.g. R. Ash: Real Analysis and Probability) there exists a subsequence $\Pi_{N_1}, \Pi_{N_2}, \ldots$ such that

$$\lim_{j \to \infty} \Pi_{N_j} = \Pi$$

where $\pi$ is a distribution function. We have proved the assertion as claimed.
The interpretation of this representation theorem is of profound significance from the point of view of subjectivistic modelling. It is as if:

- the $x_i$ are judged independent $\text{Be}(\theta)$ conditional on the random quantity $\theta$.
- $\theta$ itself is assigned the probability distribution $\Pi$ by the law of large numbers

$$\theta = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} x_i,$$

so that $\Pi$ can be seen as 'beliefs about the limiting frequency of ones'.
A Corollary: Predictive Probability

Assume \( n > m \)

\[
p(x_{m+1}, x_{m+2}, \ldots, x_n | x_1, x_2, \ldots, x_m) = \frac{p(x_1, x_2, \ldots, x_n)}{p(x_1, x_2, \ldots, x_m)}
\]

\[
= \int_0^1 \prod_{i=m+1}^n \theta^{x_i} (1 - \theta)^{1-x_i} d\Pi(\theta | x_1, x_2, \ldots, x_m),
\]

where

\[
d\Pi(\theta | x_1, x_2, \ldots, x_m) = \frac{\prod_{i=1}^m \theta^{x_i} (1 - \theta)^{1-x_i} d\Pi(\theta)}{\int_0^1 \prod_{i=1}^m \theta^{x_i} (1 - \theta)^{1-x_i} d\Pi(\theta)}
\]

which again shows the role of Bayes formula in predictive probability.
The interpretation of this representation theorem is of profound significance from the point of view of subjectivistic modelling. It is as if:

- there is no true parameter, only data and judgement