M3S3/M4S3 STATISTICAL THEORY II

THE DE FINETTI 0-1 REPRESENTATION THEOREM

Definition : Exchangeability
A *finite* sequence of random variables $X_1, X_2, \ldots, X_n$ is *(finitely) exchangeable* with (joint) probability measure P, if, for any permutation π of indices

$$P(X_1, X_2, \ldots, X_n) = P(X_{\pi(1)}, X_{\pi(2)}, \ldots, X_{\pi(n)})$$

For example, the random variables (X_1, X_2, X_3, X_4) are exchangeable if

$$P(X_1, X_2, X_3, X_4) = P(X_2, X_4, X_1, X_3) = P(X_1, X_3, X_2, X_4) = \cdots$$

An *infinite* sequence, $X_1, X_2, \ldots$, is *infinitely exchangeable* if any finite subset of the sequence is finitely exchangeable.

Theorem 3.1 ***(The De Finetti 0-1 Representation Theorem)***
If $X_1, X_2, ...$ is an infinitely exchangeable sequence of 0-1 variables with probability measure P, then there exists a distribution function Q such that the joint mass function of $(X_1, X_2, ..., X_n)$ has the form

$$p\left(X_1, X_2, ..., X_n\right) = \int_0^1 \left\{ \prod_{i=1}^n \theta^{X_i} \left(1-\theta\right)^{1-X_i} \right\} dQ\left(\theta\right)$$

where

$$Q\left(t\right) = \lim_{n\to\infty} P\left[\frac{Y_n}{n} \leq t\right]$$

and $Y_n = \sum_{i=1}^n X_i$, *and*

$$\theta \stackrel{def}{=} \lim_{n\to\infty} Y_n/n \qquad \because \qquad Y_n/n \xrightarrow{a.s.} \theta$$

is the (strong-law) limiting relative frequency of 1s.

PROOF By exchangeability, for $0 \leq y_n \leq n$

$$P\left[Y_n = y_n\right] = \binom{n}{y_n} p\left(x_1, x_2, ..., x_n\right) = \binom{n}{y_n} p\left(x_{\pi(1)}, x_{\pi(2)}, ..., x_{\pi(n)}\right) \tag{1}$$

where $X_i = x_i$ and

$$y_n = \sum_{i=1}^n x_i$$

and $\pi\left(\right)$ is any permutation of the indices. For finite N, let $N \geq n \geq y_n \geq 0$. Then, by exchangeability

$$P\left[Y_n = y_n\right] = \sum P\left[Y_n = y_n | Y_N = y_N\right] P\left[Y_N = y_N\right] \tag{2}$$

where the summation extends over $\left(y_n, ..., N - (n - y_n)\right)$. Now the conditional probability for Y_n, given $Y_N = y_N$, denoted $P\left[Y_n = y_n | Y_N = y_N\right]$, is a **hypergeometric** mass function

$$P\left[Y_n = y_n | Y_N = y_N\right] = \frac{\binom{y_N}{y_n}\binom{N - y_N}{n - y_n}}{\binom{N}{n}} \qquad 0 \leq y_n \leq n.$$

Rewriting the binomial coefficients, we have

$$P\left[Y_n = y_n\right] = \binom{n}{y_n} \sum \frac{\left(y_N\right)_{y_n} \left(N - y_N\right)_{n-y_n}}{\left(N\right)_n} P\left[Y_N = y_N\right] \tag{3}$$

where

$$(x)_r = x(x-1)(x-2)\ldots(x-r+1).$$

Define function $Q_N(\theta)$ on $\mathbb{R}$ as the step function which is zero for $\theta < 0$, and has steps of size $P[Y_N = y_N]$ at $\theta = y_N/N$ for $y_N = 0, 1, 2, ..., N$. Hence, utilizing the Lebesgue integral notation, we can re-write

$$P[Y_n = y_n] = \binom{n}{y_n} \int_0^1 \frac{(\theta N)_{y_n} ((1-\theta)N)_{n-y_n}}{(N)_n} dQ_N(\theta). \tag{4}$$

This result holds for any finite N, but in equation (2) we need to consider $N \to \infty$. In the limit,

$$\frac{(\theta N)_{y_n} ((1-\theta)N)_{n-y_n}}{(N)_n} \to \theta^{y_n} (1-\theta)^{n-y_n} = \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i}$$

as $(x)_r \to x^r$ if $x \to \infty$ with r fixed. Now, the function $Q_N(t)$ is a step function, starting at zero and ending at one, with N steps of varying sizes at particular values of t. Now, there exists a result (*the Helly Theorem*) proving that the sequence $\{Q_N(\theta); N = 1, 2, \ldots\}$ has a convergent subsequence $\{Q_{N_j}(\theta)\}$ such that, for some distribution function Q,

$$\lim_{j\to\infty} Q_{N_j}(\theta) = Q(\theta)$$

Thus the result follows comparing equation (1) and the limiting form of equation (4) as $N \longrightarrow \infty$.

Corollary : Posterior Predictive Distributions For $1 \leq m \leq n$

$$\begin{aligned} p(X_{m+1}, X_{m+2}, ..., X_n | X_1, X_2, ..., X_m) &= \frac{p(X_1, X_2, ..., X_n)}{p(X_1, X_2, ..., X_m)} \\ &= \int_0^1 \left\{ \prod_{i=m+1}^{n} \theta^{X_i} (1-\theta)^{1-X_i} \right\} dQ(\theta | X_1, ..., X_m) \end{aligned} \tag{5}$$

where, if

$$Q(\theta) = \int_0^\theta dQ(t)$$

we have

$$dQ(\theta | X_1, ..., X_m) = \frac{\prod_{i=1}^{m} \theta^{X_i} (1-\theta)^{1-X_i} dQ(\theta)}{\int_0^1 \prod_{i=1}^{m} \theta^{X_i} (1-\theta)^{1-X_i} dQ(\theta)}$$

as the updated "prior" measure. Hence, if $Y_{n-m} = \sum_{i=m+1}^{n} X_i$, we have from equation (5)

$$p(Y_{n-m} | X_1, ..., X_m) = \int_0^1 \binom{n-m}{y_{n-m}} \theta^{Y_{n-m}} (1-\theta)^{(n-m)-Y_{n-m}} dQ(\theta | X_1, ..., X_m)$$

which identifies $Q(\theta | X_1, ..., X_m)$ as the *limiting posterior predictive distribution* as $n \longrightarrow \infty$ with m fixed, as from equation (5) and the representation theorem itself, we have

$$\lim_{n\to\infty} P\left[\frac{Y_{n-m}}{n-m} \leq \theta\right] = Q(\theta | X_1, ..., X_m).$$

Interpretation: The De Finetti Representation

$$p\left(X_1, X_2, ..., X_n\right) = \int_0^1 \left\{\prod_{i=1}^{n} \theta^{X_i} \left(1-\theta\right)^{1-X_i}\right\} dQ\left(\theta\right)$$

can be interpreted in the following way;

- The joint distribution of the observable quantities $X_1, X_2, ..., X_n$ can be represented via a conditional/marginal decomposition.
  - The conditional distribution is
  $$\left\{\prod_{i=1}^{n} \theta^{X_i} \left(1-\theta\right)^{1-X_i}\right\}$$
  formed **as if it were a likelihood** for data $X_1, X_2, ..., X_n$ **conditional** on a quantity θ.
  - The marginal distribution is determined by the probability measure $Q\left(\theta\right)$, which may admit a density (wrt Lebesgue measure) p_θ, and leave the representation as
  $$p\left(X_1, X_2, ..., X_n\right) = \int_0^1 \left\{\prod_{i=1}^{n} \theta^{X_i} \left(1-\theta\right)^{1-X_i}\right\} p_\theta(\theta)\, d\theta$$
- θ is a quantity **defined by**
$$Y_n/n \xrightarrow{a.s.} \theta$$
that is, a strong law limit of observable quantities.
- Q defines a probability measure for θ which we may term the **prior** probability measure.
- In the corollary,
$$dQ\left(\theta|X_1, ..., X_m\right) = p_{\theta|X_1,...,X_m}(\theta|X_1, ..., X_m) = \frac{\prod\limits_{i=1}^{m} \theta^{X_i} \left(1-\theta\right)^{1-X_i} dQ\left(\theta\right)}{\int_0^1 \prod\limits_{i=1}^{m} \theta^{X_i} \left(1-\theta\right)^{1-X_i} dQ\left(\theta\right)}$$
defines the **updated** prior formed in light of the data $X_1, ..., X_m$; this is the **posterior** distribution for θ.

Thus, from a very simple and natural assumption (exchangeability) about observable random quantities, we have a theoretical justification for using Bayesian methods, and a natural interpretation of parameters as limiting quantities. The theorem can be extended from the simple 0-1 case to very general situations

Theorem 3.2 ***The De Finetti General Representation Theorem***
If $X_1, X_2, ...$ is an infinitely exchangeable sequence of variables with probability measure P, then there exists a distribution function Q on $\mathcal{F}$, the set of all distribution functions on $\mathbb{R}$, such that the joint distribution of $(X_1, X_2, ..., X_n)$ has the form

$$p\left(X_1, X_2, ..., X_n\right) = \int_{\mathcal{F}} \prod_{i=1}^{n} F(X_i)\, dQ\left(F\right)$$

where F is an unknown/unobservable distribution function

$$Q\left(F\right) = \lim_{n\to\infty} P_n(\widehat{F}_n)$$

is a probability measure on the space of functions $\mathcal{F}$, defined as a limiting measure as $n \longrightarrow \infty$ on the ***empirical distribution function*** *$\widehat{F}_n$.*