Belief Functions, Probability and Exchangeability

A *belief function* is a function that assigns numbers to statements such that, the larger the number, the higher the degree of belief in the statement.

The axioms of belief functions parallel the axioms of probability. In fact, *any legitimate probability function may be used as a belief function*.

Independence and *conditional independence* play key roles in statistical models.

Two events A and B are conditionally independent given event C if

$$P(A \cap B|C) = P(A|C)P(B|C).$$

The simple argument on p. 17 of Hoff shows that conditional independence of A and B given C implies that

$$P(A|C \cap B) = P(A|C).$$

In words, *knowing C and B gives no more information about A than does C by itself*.

Random variables $Y_1, \ldots, Y_n$ are conditionally independent given a parameter $\boldsymbol{\theta}$ if, for every collection $A_1, \ldots, A_n$ of subsets,

$$P(Y_1 \in A_1, \ldots, Y_n \in A_n | \boldsymbol{\theta}) = \prod_{i=1}^{n} P(Y_i \in A_i | \boldsymbol{\theta}).$$

Exchangeable Random Variables

Students who have had a math stat course have certainly encountered the notion of independent random variables. However, not all these students learn about *exchangeability*.

Suppose random variables $Y_1, \ldots, Y_n$ have joint density $p(y_1, \ldots, y_n)$. If for all vectors $(y_1, \ldots, y_n)$ and all permutations $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)$ of $\{1, \ldots, n\}$,

$$p(y_1, \ldots, y_n) = p(y_{\pi_1}, \ldots, y_{\pi_n}),$$

then $Y_1, \ldots, Y_n$ are said to be exchangeable.

So, for example, if $n = 2$, Y_1 and Y_2 are exchangeable if

$$p(y_1, y_2) = p(y_2, y_1)$$

for all (y_1, y_2).

If we think of $Y_1, \ldots, Y_n$ as data, exchangeability says that the *ordering* of the data conveys no extra information than that in the observations themselves.

Suppose that $Y_1, \ldots, Y_n$ are independent and identically distributed (i.i.d.) with common density f. Then the joint density p of $Y_1, \ldots, Y_n$ is

$$p(y_1, \ldots, y_n) = \prod_{i=1}^{n} f(y_i).$$

Now, for any permutation $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)$ of $\{1, \ldots, n\}$,

$$p(y_{\pi_1}, \ldots, y_{\pi_n}) = \prod_{i=1}^{n} f(y_{\pi_i}) = \prod_{i=1}^{n} f(y_i),$$

and so $Y_1, \ldots, Y_n$ are exchangeable.

So, if $Y_1, \ldots, Y_n$ are identically distributed and unconditionally independent, then they are exchangeable, but exchangeability does not imply **unconditional** independence.

It turns out, though, that a certain type of **conditional** i.i.d. assumption *is* equivalent to exchangeability. This is *de Finetti's theorem*, which we'll get to shortly.

Suppose $\boldsymbol{\theta}$ takes on values in a set Θ, and for each $\boldsymbol{\theta} \in \Theta$, assume that $Y_1, \ldots, Y_n$ are conditionally independent and identically distributed given $\boldsymbol{\theta}$. Suppose, furthermore, that $\boldsymbol{\theta}$ has probability density p. What is the unconditional distribution m of $Y_1, \ldots, Y_n$?

We have

$$
\begin{aligned}
m(y_1, \ldots, y_n) &= \int_{\Theta} p(y_1, \ldots, y_n | \boldsymbol{\theta}) p(\boldsymbol{\theta})\, d\boldsymbol{\theta} \\
&= \int_{\Theta} \prod_{i=1}^{n} p(y_i | \boldsymbol{\theta}) p(\boldsymbol{\theta})\, d\boldsymbol{\theta}.
\end{aligned}
$$

The integrand remains unchanged if we compute $m(y_{\pi_1}, \ldots, y_{\pi_n})$ for any permutation of $(y_1, \ldots, y_n)$, and so $Y_1, \ldots, Y_n$ are exchangeable.

The interesting thing is that, if we add a condition to exchangeability, *the converse of the last result is also true*. This is de Finetti's theorem, which says that $Y_1, \ldots, Y_n$ are exchangeable for all n if and only if for all n $Y_1, \ldots, Y_n$ are conditionally i.i.d. given $\boldsymbol{\theta}$ for some conditional density $p(y|\boldsymbol{\theta})$ and some density p for $\boldsymbol{\theta}$.

The extra condition is *for all n*. If $Y_1, \ldots, Y_n$ are exchangeable for all n, then there exist $p(y|\boldsymbol{\theta})$ and p such that

$$m(y_1, \ldots, y_n) = \int_{\Theta} \prod_{i=1}^{n} p(y_i|\boldsymbol{\theta}) p(\boldsymbol{\theta})\, d\boldsymbol{\theta}$$

for each n.

While exchangeability is an interesting concept, it is not always a reasonable model for data. There are cases where the indices on the observations convey information.

For example, suppose the data are observed over time and the index i on Y_i denotes time.