Understanding the statement of De-Finetti's Theorem (1930)

Theorem 0.1: If $x_1, x_2, \ldots$ is an infinitely exchangeable sequence of 0-1 random quantities with probability measure P, there exists a distribution function Q such that the joint mass function $p(x_1, \ldots, x_n)$ for $x_1, \ldots, x_n$ has the form

$$p(x_1, \ldots, x_n) = \int_0^1 \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} dQ(\theta),$$

where

$$Q(\theta) = \lim_{n\to\infty} P[Y_n/n \leq \theta],$$

where Y_n/n is a random variable that takes on the value $y_n = x_1 + \ldots + x_n$ and $\theta = \lim_{n\to\infty} y_n/n$.

Now let's walk through the different components of the theorem so that we understand what it actually says with regards to the prior distribution of θ.

1. **$Q(\theta)$ is a distribution *function* not a density**.
   Notice that we have $dQ(\theta) \equiv Q'(\theta)d\theta$ unlike the (wrong) form I used in class $Q(\theta)d\theta$. Thus, $Q(\theta)$ is a *cumulative distribution function* (or simply a distribution function) and not the density function $Q'(\theta)$. This is the source of the main mixup in class since we tried to think of $Q(\theta)$ as the density function.

2. **Form of $P[Y_n/n \leq \theta]$ for finite n.**
   Y_n, the number of x's that have a value of 1, is a random variable that can take on the values $0, 1, \ldots, n$. In here I used y_n to denote specific values so as to avoid confusion between the random variable and its instantiations. Thus, $P[Y_n/n \leq \theta]$ is a step function that is 0 for every $\theta < 0$ and has jumps of $p(x_1 + \ldots + x_n = y_n)$ at $\theta = y_n/n$ for $y_n = 0, \ldots, n$. This is the place where its easy to get confused. By the law of large numbers $\theta = \lim_{n\to\infty} y_n/n$. This just means that the value that the random variable will take tends to the true value of θ as n grows. Given θ, even if n is very large, Y_n the random variable can still take on the random of values and thus $P[Y_n/n \leq \theta]$ is simply a (cumulative) distribution function and true to the requirements $\lim_{\theta\to-\infty} P[Y_n/n \leq \theta] = 0$ and $\lim_{\theta\to+\infty} P[Y_n/n \leq \theta] = 1$.

3. **The limit of** $P[Y_n/n \leq \theta]$.
   As we discussed above, for a given n, $P[Y_n/n \leq \theta]$ defines a distribution function that is a series of steps and thus matches a discrete distribution. Now we take n to ∞. Note that only n is taken to the limit and Y_n is still a random variable that can receive the values $0, 1, \ldots, n$. You now need to envision how, as n grows, the steps in $P[Y_n/n \leq \theta]$ become finer grained until it becomes a continuous distribution function (not a density!). It is this limit to which $Q(\theta)$, the distribution function of the prior density equals.

If you understand the above you are actually most of the way to understanding the full proof of the theorem. If you are interested, it appears in pages $172-173$ in the book title 'Bayesian Theory' by Bernardo and Smith. That chapter also has the statement and proof outline for more general cases than the binary one.