Exchangeability and deFinetti's Theorem

Definition:

The random variables $X_1, X_2, \ldots, X_n$ are said to be **exchangeable** if the distribution of the random vector $(X_1, X_2, \ldots, X_n)$ is the same as that of $(X_{\pi_1}, X_{\pi_2}, \ldots, X_{\pi_n})$ for any permutation $(\pi_1, \pi_2, \ldots, \pi_n)$ of the indices $\{1, 2, \ldots, n\}$.

We write

$$(X_1, X_2, \ldots, X_n) \stackrel{d}{=} (X_{\pi_1}, X_{\pi_2}, \ldots, X_{\pi_n}).$$

This means that the pdf or cdf for $(X_1, X_2, \ldots, X_n)$ must be the same as that for $(X_{\pi_1}, X_{\pi_2}, \ldots, X_{\pi_n})$.

(If you'd allow me to be overly pedantic, the **joint pdf** of separate random variables $X_1, X_2, \ldots, X_n$ is **<u>the</u> pdf** of the vector valued $\vec{X} = (X_1, X_2, \ldots, X_n)$. We probably shouldn't talk about the joint pdf of $(X_1, X_2, \ldots, X_n)$, and yet we probably will!)

Let us denote the joint pdf of $X_1, X_2, \ldots, X_n$ as $f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n)$.

For the X_i to be exchangeable, we want

$$f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = f_{X_{\pi_1}, X_{\pi_2}, \ldots, X_{\pi_n}}(x_1, x_2, \ldots, x_n)$$

to be the same.

For example, if $n = 5$, and using a particular permutation $\pi = (3, 1, 5, 2, 4)$, we want

$$f_{X_1, X_2, X_3, X_4, X_5}(x_1, x_2, x_3, x_4, x_5) = f_{X_3, X_1, X_5, X_2, X_4}(x_1, x_2, x_3, x_4, x_5).$$

If the X_i are discrete, this is equivalent to

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4, X_5 = x_5) = P(X_3 = x_1, X_1 = x_2, X_5 = x_3, X_2 = x_4, X_4 = x_5).$$

Since the right-hand side is the same as $P(X_1 = x_2, X_2 = x_4, X_3 = x_1, X_4 = x_5, X_5 = x_3)$, we could phrase exchangeability in terms of the arguments of the pdf and require that

$$f_{X_1, X_2, X_3, X_4, X_5}(x_1, x_2, x_3, x_4, x_5) = f_{X_1, X_2, X_3, X_4, X_5}(x_2, x_4, x_1, x_5, x_3).$$

This is true for continous random variables as well even though we can't express the joint pdf as a probability.

$\boxed{\textbf{Example 1:}}$

Suppose you have an urn containing 1 red balls and 2 white balls. Draw out balls, one at a time and without replacement, and note the color.

Define

$$X_i = \begin{cases} 1 & , \text{ if the } i\text{th ball is red} \\ 0 & , \text{ otherwise} \end{cases}$$

The random variables X_1, X_2, X_3 are exchangeable.

Proof: If the arguments for $P(X_1 = x_1, X_2 = x_2, X_3 = x_3)$ are anything other than two 0's and one 1, regardless of the order, the probability is zero. So, we must only check arguments that are permutations of $(1, 0, 0)$.

$$P(X_1 = 1, X_2 = 0, X_2 = 0) = \frac{1}{3} \cdot 1 \cdot 1 = \frac{1}{3}$$

$$P(X_1 = 0, X_2 = 1, X_3 = 0) = \frac{2}{3} \cdot \frac{1}{2} \cdot 1 = \frac{1}{3}$$

$$P(X_1 = 0, X_2 = 0, X_3 = 1) = \frac{2}{3} \cdot \frac{1}{2} \cdot 1 = \frac{1}{3}$$

Since these are all the same, the random variables X_1, X_2, and X_3 are exchangeable.

$\boxed{\textbf{Example 2:}}$ **iid $\Rightarrow$ exchangeable**

Suppose that $X_1, X_2, \ldots, X_n$ are independent and identically distributed (iid). Then $X_1, X_2, \ldots, X_n$ are exchangeable.

Proof: Let $f(x)$ be the pdf for any one of the X_i. They are identically distributed, so they all follow one pdf.

The joint pdf (discrete or continuous) is

$$f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) \stackrel{iid}{=} f(x_1) \cdot f(x_2) \cdots f(x_n)$$

Since you can multiply the terms on the right-hand side in any order, the left-hand side is clearly symmetric in its arguments. Thus, $X_1, X_2, \ldots, X_n$ are exchangeable.

$\boxed{\textbf{Example 3:}}$ **exchangeable $\Rightarrow$ identically distributed**

Note that we already know, from Example 1, that **exchangeable random variables are not necessarily iid**. If we got a red ball on the first draw, we definitely won't get one on the subsequent two draws and if we don't get a red on the first draw, we have a higher probability of getting one in the second two draws. So, there is dependence between the outcomes of the draws.

However, exchangeable random variables must be identically distributed. In the context of Example 1, this means that $P(X_i = j)$ will be the same for all $i = 1, 2, 3$ and for all j.

Proof: (Continuous Case) Suppose that $X_1, X_2, \ldots, X_n$ are exchangeable random variables. Consider any two indices $i \neq j$. Without loss of generality, we can assume that $i < j$.

The marginal pdf for X_i is

$$f_{X_i}(x_i) = \int\int\cdots\int f_{X_1,X_2,\ldots,X_n}(x_1,\ldots,x_i,\ldots,x_j,\ldots,x_n)\,dx_1\cdots dx_{i-1}\,dx_{i+1}\cdots dx_n$$

We can exchange the ith and jth arguments to get

$$f_{X_i}(x_i) = \int\int\cdots\int f_{X_1,X_2,\ldots,X_n}(x_1,\ldots,x_j,\ldots,x_i,\ldots,x_n)\,dx_1\cdots dx_{i-1}\,dx_{i+1}\cdots dx_n$$

This doesn't change the fact that we will be integrating x_j out and leaving x_i in. Since x_i is in the jth position, the resulting integral is $fX_j(x_i)$.

Thus, we have shown that $f_{X_i}(x_i) = f_{X_j}(x_i)$ and so X_i and X_j are identically distributed. Since i and j were arbitrarily chosen, we have that all of $X_1, X_2, \ldots, X_n$ have the same distribution.

Definition:

The random variables $X_1, X_2, \ldots,$ in an **infinite sequence** are said to be **exchangeable** if the finite collection $X_1, X_2, \ldots, X_n$ are exchangeable for any finite $n \geq 1$.

Example 4: **Pólya's Urn**

Suppose you have an urn containing R_0 red balls and W_0 white balls. Let $c \geq 0$ be a fixed integer.

Draw a ball, note the color, replace the ball and put an additional c balls of that color in the urn as well. Rinse and repeat.

Define

$$X_i = \begin{cases} 1 & , \text{ if the } i\text{th ball is red} \\ 0 & , \text{ otherwise} \end{cases}$$

The random variables in the infinite sequence $X_1, X_2, \ldots$ are exchangeable.

Proof: We begin with an illustration that we will generalize. Note that

$$P(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 1, X_5 = 0)$$

$$= \frac{R_0}{R_0+W_0} \cdot \frac{R_0+c}{R_0+W_0+c} \cdot \frac{W_0}{R_0+W_0+2c} \cdot \frac{R_0+2c}{R_0+W_0+3c} \cdot \frac{W_0+c}{R_0+W_0+4c}$$

Fix any positive integer n and consider the sequence $X_1, X_2, \ldots, X_n$. We want to give an expression for $P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$. From the particular case illustrated above, it is easy to see that the denominator for this more general case will be

$$(R_0+W_0)(R_0+W_0+c)(R_0+W_0+2c)\cdots(R_0+W_0+(n-1)c).$$

Note that $\sum_{i=1}^n x_i$ is the number of red balls chosen in n draws from the urn and $n - \sum_{i=1}^n x_i$ is the number of white balls.

If $\sum x_i = n$ (all balls drawn are red), we have that

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$
$$= \frac{R_0(R_0+c)(R_0+2c)\cdots(R_0+c(\sum x_i - 1))}{(R_0+W_0)(R_0+W_0+c)(R_0+W_0+2c)\cdots(R_0+W_0+(n-1)c)}.$$

If $\sum x_i = 0$ (all balls drawn are white), we have that

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$
$$= \frac{W_0(W_0+c)(W_0+2c)\cdots(W_0+c(n-\sum x_i - 1))}{(R_0+W_0)(R_0+W_0+c)(R_0+W_0+2c)\cdots(R_0+W_0+(n-1)c)}.$$

If $0 < \sum x_i < n$, we have that

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$
$$= \frac{R_0(R_0+c)(R_0+2c)\cdots(R_0+c(\sum x_i - 1)) \cdot W_0(W_0+c)(W_0+2c)\cdots(W_0+c(n-\sum x_i - 1))}{(R_0+W_0)(R_0+W_0+c)(R_0+W_0+2c)\cdots(R_0+W_0+(n-1)c)}.$$

In all cases, the probability is a function of $\sum x_i$ and not the individual positions of the 1's and 0's that make up the x_i.

Thus, $X_1, X_2, \ldots, X_n$ are exchangeable.

Since this is true for any finite $n \geq 1$. The infinite sequence of random variables is exchangeable.

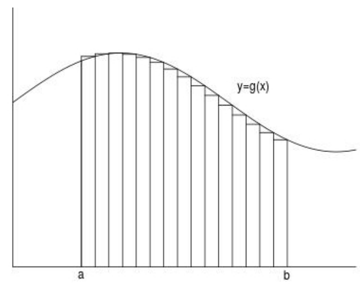
The Riemann-Stieltjes Integral

Recall the definition of the "usual" Riemann integral of a function g over the interval $[a, b]$, depicted here for a non-negative g.

One partitions up the interval $[a, b]$ into a sequence of points

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b,$$

and then defines rectangles with heights given by g evaluated at some point in each subinterval $[x_i, x_{i+1}]$, depicted as follows.



The area under the curve is approximated by the area of the rectangles which is given by

$$\sum_{i=0}^{n-1} g(c_i)(x_{i+1} - x_i)$$

for some $c_i \in [x_i, x_{i+1}]$.

If we use Γ to denote a generic partition of $[a, b]$ and $|\Gamma|$ to denote the length of the longest subinterval in Γ, we define the Riemann integral of g as

$$\int_a^b g(x)\,dx = \lim_{|\Gamma|\to 0} \sum_{i=0}^{n-1} g(c_i)(x_{i+1} - x_i).$$

For a **Riemann-Stiltjes integral**, we measure the base of those rectangles through a non-decreasing real-valued function F over $[a, b]$, and replace $x_{i+1} - x_i$ with $F(x_{i+1}) - F(x_i)$. Our notation for this will be

$$\int_a^b g(x)\,dF(x)$$

and the definition is

$$\int_a^b g(x)\,dF(x) = \lim_{|\Gamma|\to 0} \sum_{i=0}^{n-1} g(c_i)[F(x_{i+1}) - F(x_i)] \tag{1}$$

for some $c_i \in [x_i, x_{i+1}]$.

Note that, we may rewrite (1) as

$$\int_a^b g(x)\,dF(x) = \lim_{|\Gamma|\to 0} \sum_{i=0}^{n-1} g(c_i)\frac{F(x_{i+1}) - F(x_i)}{x_{i+1} - x_i} \cdot (x_{i+1} - x_i). \tag{2}$$

This now looks like a usual Riemann integral. In fact, if F is differentiable, we have that

$$\lim_{|\Gamma|\to 0} \frac{F(x_{i+1}) - F(x_i)}{x_{i+1} - x_i} = F'(x_i)$$

and the Riemann-Stieltjes integral can be written as

$$\int_a^b g(x)\,dF(x) = \int_a^b g(x)F'(x)\,dx. \tag{3}$$

Although it was not required in the definition of the Riemann-Stiltjes integral, for us F will be a cdf. The derivative, if it exists, will be a pdf f. In this case, the integral can be written as

$$\int_a^b g(x)\, dF(x) = \int_a^b g(x)F'(x)\, dx = \int_a^b g(x)f(x)\, dx.$$

The nice thing about Riemann-Stieltjes integration is that it allows us to unify our treatment of discrete and continuous random variables.

Suppose that X is a discrete random variable with distribution given by

x	1	2	3
$P(X = x)$	1/4	1/4	1/2

The cdf is then the step function

$$F(x) = P(X \leq x) = \begin{cases} 0 & , \quad x < 1 \\ 1/4 & , \quad 1 \leq x < 2 \\ 1/2 & , \quad 2 \leq x < 3 \\ 1 & , \quad x \geq 3 \end{cases}$$

(* A sketch of the step function here might be helpful for what is to follow. I just want to get these notes out without spending time on the graphic right now!)

Note that F is differentiable on the intervals $(-\infty, 1)$, $(1, 2)$, $(2, 3)$, and $(3, \infty)$. In these places, it is flat and constant and the derivative is zero. By (3) we then have that the Riemann-Stieltjes integral $\int_{-\infty}^{\infty} g(x)\, dF(x)$ is "mostly zero".

However, if you consider the definition in terms of a shrinking partition, when you evaluate $F(x_{i+1}) - F(x_i)$, you will eventually catch all of the "jumps" of F and you will have to evaluate g at the points where these jumps occur.

The evaluation of the integral is

$$\begin{aligned} \int_{-\infty}^{\infty} g(x)\, dF(x) &= \int_{-\infty}^{1} g(x)F'(x)\, dx + g(1) \cdot (1/4) + \int_1^2 g(x)F'(x)\, dx + g(2) \cdot (1/4) \\ &\quad + \int_2^3 g(x)F'(x) + g(3) \cdot (1/2) + \int_3^{\infty} g(x)F'(x) \\ &= \tfrac{1}{4}g(1) + \tfrac{1}{4}g(2) + \tfrac{1}{2}g(3) \end{aligned}$$

since F' is zero in all of those Riemann integrals.

If you look at the original discrete distribution, you'll notice that we have computed the expected value of $g(X)$. Furthermore, we wrote it as an integral! Indeed we have

$$\mathsf{E}[X] = \int_{-\infty}^{\infty} g(x)\, dF(x)$$

for both discrete and continuous random variables. (Also for random variables with "mixed" distributions where there are jumps but some non-trivially differentiable parts!)

We also have that the property that

$$\int_{-\infty}^{\infty} dF(x) = 1$$

for the discrete, continuous, and mixed cases. (Verify for yourself if it is not clear to you!)

$\boxed{\textbf{de Finetti's Theorem:}}$

An infinite sequence of binary (0/1) random variables, $\{X_n\}_{n=1}^{\infty}$, is exchangeable if and only if there exists a cdf F on $[0,1]$ such that

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \int_0^1 \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}\, dF(\theta).$$