## Independence of random variables

We all have an intuitive sense of what it means for two events to be independent of each other. And sometimes mathematics captures intuitive notions by constructing intuitive formal definitions. But sometimes it does not, and independence is one of them. Formally, independence of random variable is defined as:

> **Definition.** Let $(X, \mathcal{F}, \mu)$ be a probability space and consider a family of random variables $\{f_i\}_{i \in J}$ where each $f_i : X \to \mathbb{R}$. We will say that the random variables $\{f_i\}_{i \in J}$ are *independent* if and only if for every choice of measurable sets $\{A_i\}_{i \in J}$ in $\mathbb{R}$, the family of sets $\{f_i^{-1}(A_i)\}_{i \in J}$ is independent.

I would like to spend a little time explaining what this definition means, perhaps with some pictures. To begin, the definition of independence for random variables relies on the notion of independence for measurable sets. Let us recall that definition first:

> **Definition.** Let $(X, \mathcal{F}, \mu)$ be a probability space and $\{B_i\}_{i \in J}$ be a family of measurable sets, that is, each $B_i \in \mathcal{F}$. We will say that the sets $\{B_i\}_{i \in J}$ are *independent* if and only if for any finite subset $I \subset J$, we have
>
> $$\mu\Big(\bigcap_{i \in I} B_i\Big) = \prod_{i \in I} \mu(B_i).$$

It is easiest to understand what independence means when thinking just about two sets. The condition for $B_1$ and $B_2$ to be independent requires that $\mu(B_1 \cap B_2) = \mu(B_1) \cdot \mu(B_2)$, but it may be easier to see what is going on by writing this equation as:

$$\frac{\mu(B_1 \cap B_2)}{\mu(B_1)} = \mu(B_2).$$

The right side can be thought of as the proportion of points in $X$ that lie in $B_2$. The left side is the proportion of points in $B_1$ that lie in $B_2$. Independence means that if we restrict our world from $X$ to $B_1$, the proportion of points belonging $B_2$ stays the same.

**Example.** As an example of sets that are not independent, let $X$ be the set of all people in the world, $B_1$ be the set of mathematicians, and $B_2$ be the set of toddlers. For the sets $B_1$ and $B_2$ to be independent, we would need the proportion of $B_2$ in $X$ to equal the proportion of $B_2$ in $B_1$. The former is around 2%, while the latter is very close to 0%.

If you are willing to think of measure geometrically as something that measures area, we can draw a diagram to explain independence. See Figure 1 below. Note that you have to draw the sets $B_1$ and $B_2$ somewhat carefully so that the measure of their intersection is indeed the product of their respective measures. The sets in the left diagram are independent, while the ones on the right are not!

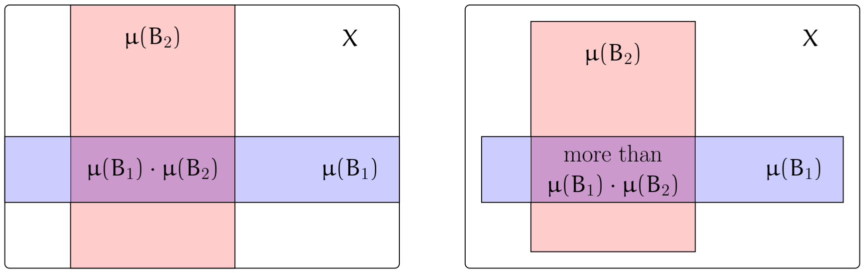


FIGURE 1. For independent sets in a probability space $X$, the measure of their intersection is exactly the product of their measures.

There is a slightly different way to think about independent sets, which is really just another way of saying what we have said above. The question that independence is trying to answer is whether knowing whether an element lies in the set $B_1$ influences the probability of it lying in the set $B_2$. For instance, if you are selecting a person at random from the entire world population, are you just as likely to select a toddler if you are choosing from among mathematicians?

We are now ready to talk about independent random variables. For a random variable $f : X \to \mathbb{R}$, the definition starts with a measurable subset $A$ of $\mathbb{R}$, and then examines its preimage $f^{-1}(A)$. Using pictures, we are looking at something like Figure 2:

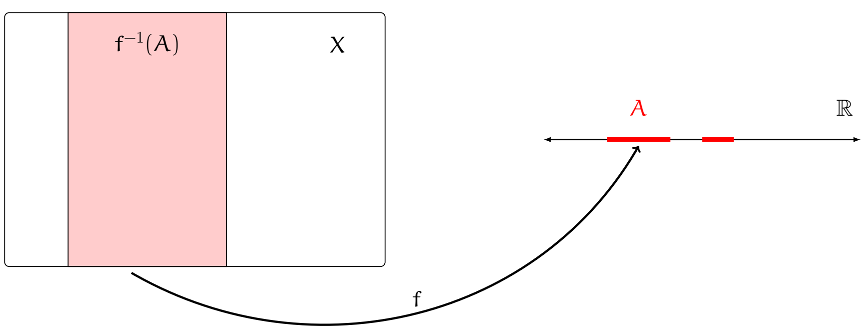


FIGURE 2. Preimage of random variable $f$ in probability space $X$.

For two random variables $f_1$ and $f_2$ to be independent, the picture is a little more complicated. The sets $f_1^{-1}(A_1)$ and $f_2^{-1}(A_2)$ need to be independent for all possible choices of $A_1$ and $A_2$. That is, we need

$$\mu(f_1^{-1}(A_1) \cap \mu(f_2^{-1}(A_2)) = \mu(f_1^{-1}(A_1)) \cdot \mu(f_2^{-1}(A_2))$$

to always hold. In Figure 3, this is illustrated using sets:

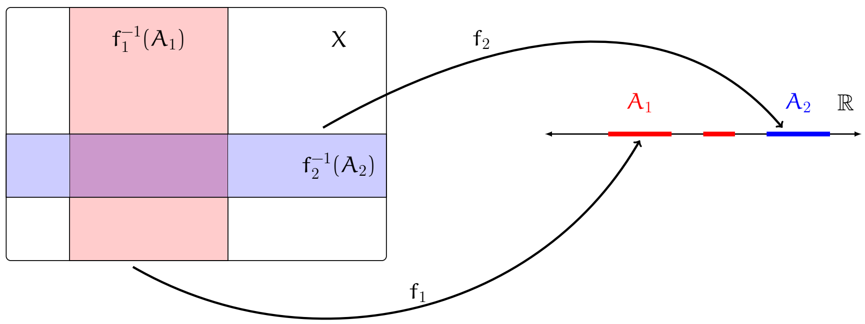


FIGURE 3. Independence of the random variables $f_1$ and $f_2$ depends on the measure of the intersection of the preimages for any two measurable subsets of $\mathbb{R}$.

**Example.** Suppose that $X$ is a set of people, $f_1$ assigns each person their weight, and $f_2$ does the same with height. Now choose a range of weights $A_1$ and a range of heights $A_2$. For these random variables to be independent, we would need the set of people whose weight lies in $A_1$, that is, $f_1^{-1}(A_1)$, to be independent from the set of people whose height lies in $A_2$, that is, $f_2^{-1}(A_2)$, for all possible choices of $A_1$ and $A_2$.

The independence condition for random variables is practically tricky to verify as one has to show something is true for *all* choices of subsets $A_1$ and $A_2$. But this condition is important. The goal is to make sure that there is *no* set of values $A_1$ of $f_1$ that can give us extra information about when $f_2$ takes values on another set $A_2$.