about $\Omega$ if we know the “joint distribution”

$$Pr(X=x \text{ and } Y=y)$$

for each $x$ in the range of $X$ and each $y$ in the range of $Y$. We say that $X$ and $Y$ are independent random variables if

$$Pr(X=x \text{ and } Y=y) = Pr(X=x) \cdot Pr(Y=y)$$ (8.5)

for all $x$ and $y$. Intuitively, this means that the value of $X$ has no effect on the value of $Y$.

For example, if $\Omega$ is the set of dice rolls $D^2$, we can let $S_1$ be the number of spots on the first die and $S_2$ the number of spots on the second. Then the random variables $S_1$ and $S_2$ are independent with respect to each of the probability distributions $Pr_{00}, Pr_{11},$ and $Pr_{01}$ discussed earlier, because we defined the dice probability for each elementary event dd’ as a product of a probability for $S_1 = d$ multiplied by a probability for $S_2 = d'$. We could have defined probabilities differently so that, say,

$$Pr(S_2=5)/Pr(S_2=6) \neq Pr(S_1=2)/Pr(S_1=1);$$

but we didn’t do that, because different dice aren’t supposed to influence each other. With our definitions, both of these ratios are $Pr(S_2=5)/Pr(S_2=6)$.

We have defined $S$ to be the sum of the two spot values, $S_1 + S_2$. Let’s consider another random variable $P$, the product $S_1S_2$. Are $S$ and $P$ independent? Informally, no; if we are told that $S = 2$, we know that $P$ must be 1. Formally, no again, because the independence condition (8.5) fails spectacularly (at least in the case of fair dice): For all legal values of $s$ and $p$, we have $0 < Pr_{00}[S=s] \cdot Pr_{00}[P=p] \leq \frac{1}{6} \cdot \frac{1}{6}$; this can’t equal $Pr_{00}[S=s \text{ and } P=p]$, which is a multiple of $\frac{1}{36}$.

If we want to understand the typical behavior of a given random variable, we often ask about its “average” value. But the notion of “average” is ambiguous; people generally speak about three different kinds of averages when a sequence of numbers is given:

- the mean (which is the sum of all values, divided by the number of values);
- the median (which is the middle value, numerically);
- the mode (which is the value that occurs most often).

For example, the mean of $\{3, 1, 4, 1, 5\}$ is $\frac{3+1+4+1+5}{5} = 2.8$; the median is 3; the mode is 1.

But probability theorists usually work with random variables instead of with sequences of numbers, so we want to define the notion of an “average” for random variables too. Suppose we repeat an experiment over and over again,