# Topic 2: Random Variables and their Distributions

**Rohini Somanathan**

**Course 003, 2018**

# Sample spaces and random variables

**The outcomes of some experiments inherently take the form of real numbers:**

- **crop yields with the application of a new type of fertiliser**
- **students scores on an exam**
- **miles per litre of an automobile**

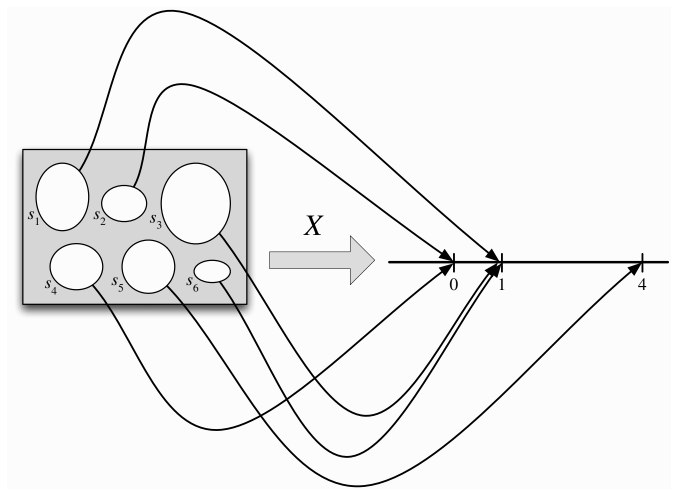
**Other experiments have a sample space that is not inherently a subset of Euclidean space**

- **Outcomes from a series of coin tosses**
- **The character of a politician**
- **The modes of transport taken by a city's population**
- **The degree of satisfaction respondents report for a service provider -patients in a hospital may be asked whether they are very satisfied, satisfied or dissatisfied with the quality of treatment. Our sample space would consist of arrays of the form $(VS, S, S, DS, ....)$**
- **The caste composition of elected politicians.**
- **The gender composition of children attending school.**

**A random variable is a function that assigns a real number to each possible outcome $s \in S$.**

# Random variables

**Definition (random variable):** *Given an experiment with sample space* $\mathbf{S}$*, a random variable is a function from the sample space* $\mathbf{S}$ *to the real numbers,* $\mathbf{X} : \mathbf{S} \rightarrow \mathbb{R}$.




**The mapping itself is deterministic, the randomness comes from the sample space.**

**X is a function that takes us from a probability space** $(S, \mathcal{F}, \mathbb{P})$ **to an induced probability space** $(R(X), \mathcal{B}, P_X(A))$.

**As long as every set** $A \in R(X)$ **is associated with an event in our original sample space** $S$, $P_X(A)$ **is just the probability assigned to that event by** $\mathbb{P}$

**We usually omit the X in** $P_X(A)$ **unless there is ambiguity.**

# Random variables..examples

1. **Tossing a coin ten times.**

   The sample space consists of the $2^{10}$ possible sequences of heads and tails.

   There are many different random variables that could be associated with this experiment: $X_1$ could be the number of heads, $X_2$ the longest run of heads divided by the longest run of tails, $X_3$ the number of times we get two heads immediately before a tail, etc...

   For $s = HTTHHHHTTH$, what are the values of these random variables?

2. **Choosing a point in a rectangle within a plane**

   An experiment involves choosing a point $s = (x, y)$ at random from the rectangle $S = \{(x, y) : 0 \leq x \leq 2, 0 \leq y \leq 1/2\}$

   The random variable $X$ could be the $x-$coordinate of the point and an event is $X$ taking values in $[1, 2]$

   Another random variable $Z$ would be the distance of the point from the origin, $Z(s) = \sqrt{x^2 + y^2}$

**Heights, weights, distances, temperature, scores, incomes...** In these cases, we can have

$X(s) = s$ since these are already expressed as real numbers.

# Discrete random variables

**Random variables can be discrete, continuous or a mixture of these two.**

**Definition ( discrete random variable) :** *A random variable $X$ is said to be discrete if there is a finite list of values $a_1, a_2, \dots a_n$ or an infinite list of values $a_1, a_2, \dots$ such that $P(X = a_j$ for some $j) = 1$. If $X$ is a discrete r.v., then the finite or countably infinite set of values $x$ such that $P(X = x) > 0$ is called the support of $X$.*

**Definition ( probability mass function): The** *probability mass function (PMF) of a discrete r.v. is given by the function $f_X(x) = P(X = x)$. This is positive if $x$ is in the support of $X$ and $0$ otherwise.*

**Note that the event $\{X = x\}$ is defined as $\{s \in S : X(s) = x\}$ so $P(X = x)$ is meaningful, but $P(X)$ is not since probabilities are defined for events, not for random variables.**

**Some probability distributions are so common and useful that they have names and we study them thoroughly. We now introduce some of these.**

# Bernoulli distribution

**Definition (Bernoulli distribution) :** *A random variable* $\mathbf{X}$ *is said to have the Bernoulli distribution with parameter* $\mathbf{p}$ *if* $\mathbf{P(X = 1) = p}$ *and* $\mathbf{P(X = 0) = 1 - p}$*, where* $\mathbf{0 < p < 1}$*. We write this as* $\mathbf{X \sim Bern(p)}$

**Any experiment that can result in only a success or failure is called a Bernoulli trial**

**The parameter $\mathbf{p}$ is called the success probability of the $\mathbf{Bern(p)}$ distribution.**

**There is therefore a whole family of Bernoulli distributions and we specify a particular one when we pin down this parameter.**

**Definition (Indicator random variable) :** *The indicator random variable of an event* $\mathbf{A}$ *is the r.v. which equals 1* $\mathbf{A}$ *occurs and zero otherwise. The indicator r.v. of event* $\mathbf{A}$ *is denoted by* $\mathbf{I_A}$*. Note that*

$$\mathbf{I_A \sim Bern(P(A))}$$

# Binomial distribution

**Definition (Binomial distribution) :** *Let* $X$ *be the number of successes in* $n$ *independent Bernoulli trials with success probability* $p$*. Then* $X$ *has a Binomial distribution with parameters* $n$ *and* $p$*. We write*

$$X \sim \text{Bin}(n, p)$$

**Definition (Binomial PMF) :** *If* $X \sim \text{Bin}(n, p)$, *then the PMF of X is*

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

*for* $k = 0, 1, 2, \dots n$ *and zero otherwise.*

**Proof: The probability of any specific sequence of $k$ successes and $n-k$ failures is $p^k(1-p)^{n-k}$. There are $\binom{n}{k}$ such sequences since we just need to select the position of the successes The above expression is non-negative and sums to 1 by the binomial theorem, so we have a valid PMF.**

**Note: If $X \sim \text{Bin}(n, p)$, then $n - X \sim \text{Bin}(n, q)$, where $q = 1 - p$. Also, with $n$ even, $\text{Bin}(n, \frac{1}{2})$ is symmetric around $\frac{n}{2}$. Try to prove both these results.**

# The hypergeometric distribution

**Think of an urn with $w$ white and $b$ black balls. If we draw $n$ balls with replacement, the number of white balls drawn $X \sim \text{Bin}(n, \frac{w}{w+b})$.**

**If we draw without replacement, $X$ follows a hypergeometric distribution, with parameters $w, b$ and $n$. We denote this by $\text{HGeom}(w, b, n)$.**

**Definition (Hypergeometric PMF) :** *If* $X \sim \mathbf{HGeom}(w, b, n)$, *then the PMF of X is*

$$P(X = k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$$

**for integers $0 \leq k \leq w$ and $0 \leq n - k \leq b$, and zero otherwise.**

**The number of white balls $k$ can take values from $0$ to $n$, the sum of these is $\binom{w+b}{n}$ so we have a valid PMF.**

**Applications: All experiments where there are two types of items in the population, each of which may or may not be in the sample (double-tagged items: type $w$ or $b$ is the first tag, in and out of the sample is the second )**

# Discrete uniform distribution

**Think of a finite, non-empty set $C$ of numbers. Choose one of these numbers at random and call the chosen number $X$. Then $X \sim DUnif(C)$**

**Definition (Discrete Uniform PMF) :** *If* $X \sim DUnif(C)$, *then the PMF of X is*

$$P(X = x) = \frac{1}{|C|}$$

**for $x \in C$, and zero otherwise.**

**For any $A \subseteq C$,**

$$P(X \in A) = \frac{|A|}{|C|}$$

# The cumulative distribution function

**All random variables can be described completely by their cumulative distribution function**

**Definition (cumulative distribution function):** *The cumulative distribution function of a random variable* $X$ *is the function* $F_X$ *given by* $F_X(x) = P(X \leq x)$ *for* $-\infty < x < \infty$

**The X subscript denotes the r.v. in question and is usually dropped when there is no ambiguity.**

**If there are a finite number of elements $w$ in $A$, this probability can be computed as**

$$F(x) = \sum_{w \leq x} f(w)$$

**In this case, the distribution function will be a step function, jumping at all points $x$ in $R(X)$ which are assigned positive probability.**

**Practice:** **Derive and plot CDFs for the four discrete distributions we have studied so far.**

# Properties of the CDF

**Any CDF has the following three properties:**

1. **Increasing:** If $x_1 \leq x_2$ then $F(x_1) \leq F(x_2)$.

   ( The occurrence of the event $\{X \leq x_1\}$ implies the occurrence of $\{X \leq x_2\}$ so $P(X \leq x_1) \leq P(X \leq x_2)$)

2. **Right continuous:** $F(a) = \lim_{x \to a+} F(x)$

   For continuous r.v.s it will be continuous throughout, for discrete r.v.s it will jump at all points with positive probability.

3. **Convergence to 0 and 1 in the limits:** $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$

   ( $\{x : x \leq \infty\}$ is the entire sample space and $\{x : x \leq -\infty\}$ is the null set. )

**An advantage of a distribution function is that it is defined in the same way for all types of random variables and also has an empirical counterpart.**

**Definition (Empirical distribution function):** *Let* $X_1, X_2, \ldots X_n$ *be i.i.d random variables with CDF* $F$*. The empirical CDF of* $X_1, X_2, \ldots X_n$ *is defined as* $\hat{F}_X(x) = \frac{R_n(x)}{n}$ *where* $R_n(x) = \sum_{j=1}^{n} I(X_j \leq x)$

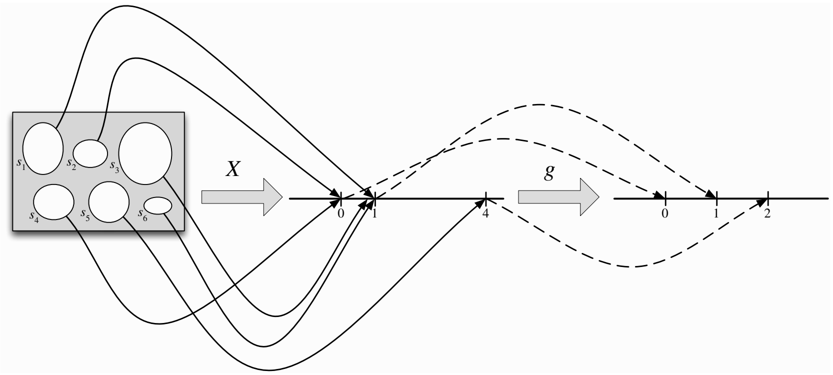
**So $\hat{F}_X(x)$ gives us the fraction of sample values below $x$. This is a central object for all nonparametric statistics since we make no assumptions about the family of distributions we sample.**

# Functions of a random variable

**If $X$ is an r.v., is it meaningful to talk of $X^2$ or $\log(X)$? In what sense?**

**Definition (Function of a random variable):** *For an experiment with sample space $S$, an r.v. $X$, and a function $g : \mathbb{R} \to \mathbb{R}$, $g(X)$ is the r.v. that maps $s$ to $g(X(s))$ for all $s \in S$.*

**From this definition, we see that $g(X)$ is the composition of functions $X$ and $g$. We first apply $X$ to $s$ and then $g$. The PMF of $Y = g(X)$ is easily derived if $g$ is one-to-one since the support of $Y$ is just the set of all $g(x)$ with $x$ in the support of $X$.**




**If $g$ is not one-to-one, there may be multiple values of $x$, such that $g(x) = y$, and to compute the PMF, we sum the probabilities of $X$ taking on any of these values.**

# Functions of two random variables

**Definition (Function of two random variables):** *Given an experiment with sample space* $S$, *if* $X$ *and* $Y$ *are two r.v.s that map* $s \in S$ *to* $X(s)$ *and* $Y(s)$ *respectively, then* $g(X, Y)$ *is the r.v. that maps* $s$ *to* $g(X(s), Y(s))$.

**Examples:**

$$Z = X + Y$$

$$Z = \max(X, Y)$$

$$Z = X + X^2$$

**etc...**

**Draw yourself a picture of the sample space for two fair coin tosses, showing how $X$ (no. heads), $Y$ (no. tails) and $X + Y$ map elements of $S$ to $\mathbb{R}$.**

# Functions of r.v.s : an example

**A random walk (story): Suppose that a particle takes $n$ steps on a number line, starting at zero. At each step, it moves right or left with equal probabilities and each step is independent. Let $X$ be the number of steps to the right, and $Y$ be the particle's position after $n$ steps. Find the PMFs of these r.v.s**

**Solution: Clearly $X \sim \text{Bin}(n, .5)$ and if $X = j$, $Y = j - n + j = 2j - n$. So $Y = 2X - n$ and since $X$ takes values in $\{(0, 1, 2, \ldots n)\}$, $Y$ takes values in $\{(-n, 2 - n, 4 - n, \ldots, n)\}$**

**The PMF of $Y$ can be found from the PMF of $X$:**

$$P(Y = k) = P(2X - n = k) = P(X = \frac{n + k}{2}) = \binom{n}{\frac{n+k}{2}}\left(\frac{1}{2}\right)^n$$

# Random variables vs. their distributions

**Random variables with identical distributions are not the same! Two random variables can have the same distribution, yet never take the same values**

**Examples:**

**X = head on a fair coin flipped once, Y = tail on a fair coin flipped once**

**X = head on a fair coin flipped once, Y = even no. on the roll of a fair die**

**This distinction becomes especially important later in the course when we discuss the convergence of sequences of random variables.**

# Independence of two random variables

**Definition (Independence of two r.v.s) :** *The two random variables* $X$ *and* $Y$ *are said to be independent if, for any two sets* $A$ *and* $B$ *of real numbers,*

$$P(X \in A \text{ and } Y \in B) = P(X \in A)P(Y \in B)$$

**In words, if $A$ is an event whose occurrence depends only values taken by $X$ and B's occurrence depends only on values taken by $Y$, then the random variables $X$ and $Y$ are independent only if the events $A$ and $B$ are independent, for all such events $A$ and $B$.**

- **The condition for independence can be alternatively stated in terms of the joint and marginal distribution functions of $X$ and $Y$ by letting the sets $A$ and $B$ be the intervals $(-\infty, x)$ and $(-\infty, y)$ respectively.**

  $$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$$

  **Or** $F(x, y) = F_1(x)F_2(y)$

- **For discrete distributions, we simply define the sets $A$ and $B$ as the points $x$ and $y$ and require $f(x, y) = f_1(x)f_2(y)$.**

- **In terms of the density functions, we say that $X$ and $Y$ are independent if it is possible to choose functions $f_1$ and $f_2$ such that the following factorization holds for $(-\infty < x < \infty$ and $-\infty < y < \infty)$**

  $$f(x, y) = f_1(x)f_2(y)$$

# Independence of several random variables

**Definition (Independence of several r.v.s) :** *The random variables* $X_1, \dots X_n$ *are said to be independent if*

$$P(X_1 \leq x_1, X_2 \leq x_2, \dots X_n \leq x_n) = P(X_1 \leq x_1)P(X_2 \leq x_2)\dots P(X_n \leq x_n)$$

*for all* $x_1, \dots, x_n$ *in* $\mathbb{R}$.

*For infinitely many r.v.s, we say that they are independent if every finite subset of the r.v.s is independent.*