37262 Mathematical Statistics

Lecture 1

Random Experiments

- A random experiment is one whose outcome is not determined in advance.
- Any action which may have more than one possible outcome can be considered to be a random experiment.
- The set of all possible outcomes of a random experiment is called its **sample space**, usually denoted by S or Ω .

Random Experiments

- For example, if the random experiment consists of rolling one die and recording the number shown, the sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- If the random experiment consists of rolling three dice and recording if the numbers shown are odd or even, the sample space is odd or even, the sample space is
  $\Omega = \{OOO, OOE, OEO, OEE, EOO, EOE, EEO, EEE\}$

  where, for example, OOE represents the first two dice showing an odd number and the third an even.

Discrete and Continuous Sample Spaces

- The examples on the previous slide (coin flipping and die rolling) have **discrete** sample spaces.
- That is, we can write a list of possible outcomes as separate points.
- For other experiments, this is not possible as we have a **continuous** sample space.

- For example, if the experiment consists of measuring t, the time in seconds until the next train arrives is simply all non-negative values of t , $\Omega = \{t : t \geq 0\}$.

Discrete and Continuous Sample Spaces

- Both continuous and discrete sample spaces can be either finite or infinite.

- A discrete sample space is finite if it contains a finite number of points.

- An example of an infinite sample space would result from an experiment recording the mass of an object to the nearest gram $\Omega = \{0,1,2,3,4,\ldots\} = \mathbb{R}^{\geq 0}$.

- A continuous finite sample space needs to have both an upper and lower bound. For example, if we wanted to measure the lifespan, *L*, to date or until failure (in years) of a lightbulb manufactured 10 years ago, then $\Omega = \{L : 0 \leq L \leq 10\}$.

Events

- A subset of a sample space is called an **event**.

- We say that event A occurs if and only if the outcome of a random experiment is one of the points in A.

- For example if an experiment consists of rolling two dice, the sample space contains 36 elements:

$$\Omega = \{11, 12, 13, 14, 15, 16, 21, 22, \ldots, 64, 65, 66\}$$

11	21	31	41	51	61
12	**22**	32	42	52	62
13	23	**33**	43	53	63
14	24	34	**44**	54	64
15	25	35	45	**55**	65
16	26	36	46	56	**66**

- If we define the event A to correspond to "the two dice show the same number" then this is denoted by the subset of Ω given by $A = \{11, 22, 33, 44, 55, 66\}$.

Events

- If an experiment consists of measuring how long, t (in minutes), until the next train arrives, the sample space is an infinite continuous interval $\Omega = \{t : t \geq 0\}$.

- If we define the event B to correspond to "a train arrives within the next ten minutes", this is denoted by the subset of Ω given by $B = \{t : 0 \leq t < 10\}$.

- Note that since sample spaces can be discrete or continuous and finite or infinite, subsets of these (i.e. events) can also be classified in the same way.
- There can sometimes be possible outcomes of a random experiment which fall into more than one event

Probability

- In order to fairly assess the chance of random events occurring, we need some absolute measure of chance.

- We can define a probability through relative frequencies. That is, if we can look at a large number of "identical situations", $\text{Probability of } A = P(A) \approx \dfrac{\text{Number of times event } A \text{ occurs}}{\text{Number of trials}}$.

- If we flipped a coin a million times and observed Tails on half a million flips, we might estimate $P(Tails) \approx \dfrac{1}{2}$.

- This is known as the **frequentist** interpretation of probability.

Random Variables

- A **random variable** is a function which maps all possible outputs Ω of a random experiment to some subset of $\mathbb{R}$.

- In other words, it takes the outcome of a given experiment (which could be numerical, or could be a category e.g. “seven of clubs” or “Tails”) and assigns a real number to it.

- For experiments whose sample space is numerical values, a random variable can simply defined as the number of the event in the sample space.

Random Variables

- By convention, random variables tend to be denoted with capital letters.

- We can then define events corresponding to whether or not the variable takes a given numerical value.

- For example, if X is the number shown when rolling a regular fair six-sided die, we could have the event $X = 3$ which would correspond to the die showing the number 3.

- Where we are considering the random variable taking a value from a (non-random) sample space, this is usually denoted with a lower case letter. For example,

$$P(Y = k) = \begin{cases} \dfrac{k}{21} & k \in \{1,2,3,4,5,6\} \\ 0 & \text{otherwise} \end{cases}$$ could define a random variable Y.

Probability Mass Functions

- For a discrete random variable, the probability mass function gives, for all real numbers, the probability that the random variable gives that numerical value.
- For example, if we are rolling one regular fair six-sided die and defining the random variable X to be the number shown, then we have that all values 1,2,…,6 occur with probability 1/6 and that no other values can occur.
- This is $P(X = k) = \begin{cases} 1/6 & k = 1 \\ 1/6 & k = 2 \\ 1/6 & k = 3 \\ 1/6 & k = 4 \\ 1/6 & k = 5 \\ 1/6 & k = 6 \\ 0 & \text{otherwise} \end{cases}$

Probability Mass Functions

- Consider now rolling one regular fair six-sided die and defining the random variable Y to be the number of letters in the name of the numbers shown (ONE, TWO, THREE, FOUR, FIVE or SIX).

- Three of the (equally probable outcomes) have 3 letters, so there is a 3/6 or 1/2 chance that Y will take the value 3.

- Similarly, there is a 1/3 chance of getting a number with four letters and a 1/6 chance of getting a five letter number.

- Overall, we get $P(Y = k) = \begin{cases} 1/2 & k = 3 \\ 1/3 & k = 4 \\ 1/6 & k = 5 \\ 0 & \text{otherwise} \end{cases}$

Probability Density Functions

- Recall that, for a continuous variable X. instead of a probability mass function, we define a **probability density function** to be $f(x)$ such that $P(a < X < b) = \int_a^b f(x)dx$.
- The probability density function $f(x)$ gives a relative measure of how likely the random variable is to take a value in a given region. It is not, though, itself a probability.
- An intuitive interpretation of the density function is that, for very small $\varepsilon > 0$,

$$P\left(a - \frac{\varepsilon}{2} < X < a + \frac{\varepsilon}{2}\right) = \int_{a-\frac{\varepsilon}{2}}^{a+\frac{\varepsilon}{2}} f(x)dx \approx \varepsilon f(a)$$

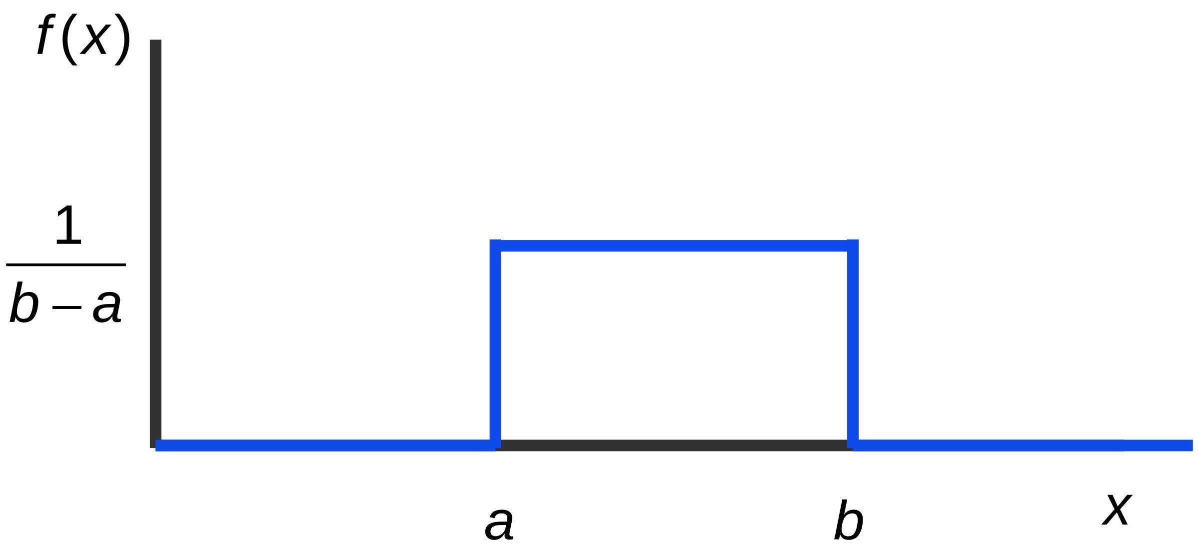
- In other words, for very small $\varepsilon > 0$, the probability that $X = a$ (with a margin of error no more than ε centred around a is approximately $\varepsilon f(a)$.

Uniform Random Variable

- Maybe the simplest type of continuous random variable is the **uniform** variable.

- For $a < b$ if $X \sim U[a,b]$ then X takes a value between a and b such that X is equally likely to be any two intervals of equal width.

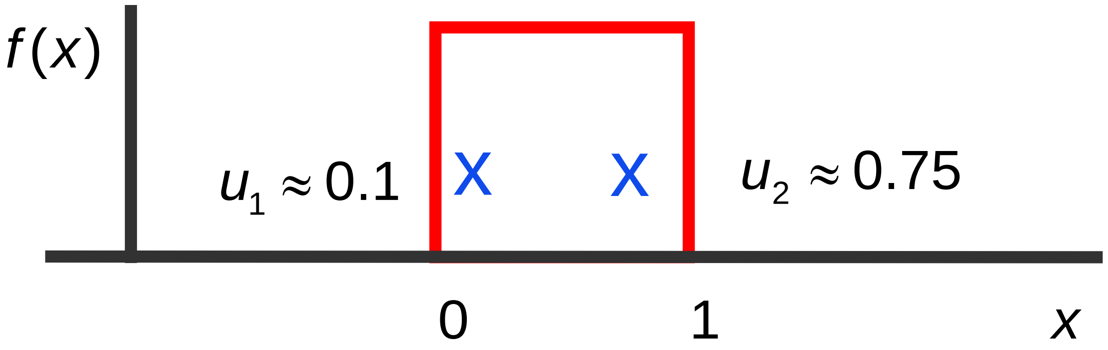
- That is $f(x) = \begin{cases} \dfrac{1}{b-a} & x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$



Simulating Discrete Events

- All major mathematical computer packages contain a random number generator which can generate realisations of a uniform random variable between 0 and 1.

- That is, it can be visualised as picking a point at random from the “block” of probability density such that any two regions of equal area are equally likely to be chosen.

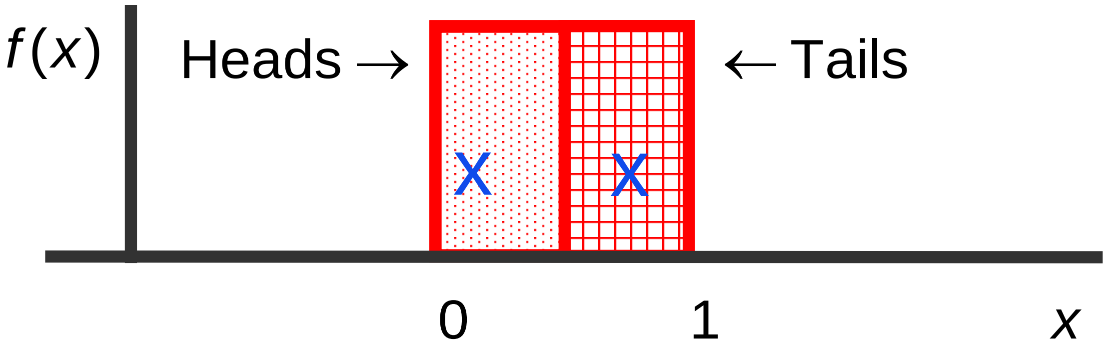


- The realisation of $X \sim U[0,1]$ can then be generated by measure how far along this interval the randomly selected point was chosen.

Simulating Discrete Events

- Realisations of the outcome ofa random experiment with a discrete sample space can easily be generated using realisations of a $X \sim U[0,1]$ variable.

- For example, if we wish to simulate flipping a fair coin once, then there are two possible outcomes (Head, Tails), each of which occurs with probability 0.5.



- We can do this by generating realisations of $X \sim U[0,1]$ and calling Heads if the random number ≤ 0.5 and calling Tails if the number is > 0.5.

Bernoulli Distribution

- Any random variable whose probability mass function can

be written as $P(X=k)=\begin{cases} 1-p & k=0 \\ p & k=1 \\ 0 & \text{otherwise} \end{cases}$

is known as a **Bernoulli** variable.

- In this case, we write $X \sim Bern(p)$.

- This distribution has range $\{0,1\}$.

Jacob Bernoulli

(1655-1705)

Bernoulli Distribution

- The distribution depends on one parameter, p, which gives the probability of obtaining a 1, rather than a zero.

- For example, the number of Tails from a single fair coin flip $\sim Bern(0.5)$ or, when selecting one person at random, the number of selected people born on a Saturday $\sim Bern\left(\frac{1}{7}\right)$

- The expectation and variance of $X \sim Bern(p)$ can easily be calculated.

- $E(X) = \sum k \times p(X = k) \;= [0 \times (1 - p)] + [1 \times p] \;= p$

Jacob Bernoulli

(*1655-1705*)

Bernoulli Distribution

- Similarly, $E(X^2) = \sum k^2 \times p(X = k) \; = \left[0^2 \times (1 - p)\right] + \left[1^2 \times p\right] = p$
- This therefore gives $Var(X) = E(X^2) - E(X)^2 = p - p^2 = p(1 - p)$.
- These values are perhaps intuitive. If we expect half of our experiments to give a 1 then, on average, each experiment gives the value 0.5.
- The variance is zero if $p = 0$ or $p = 1$. This is because there is no variability between realisations of this experiment – we already knew the outcome would either certainly happen (1) or certainly not happen (0).

Jacob Bernoulli

(1655-1705)

Bernoulli Distribution

- Given realisations $\{u_1, u_2, u_3, \ldots\}$ of a $U[0,1]$ variable, we can easily simulate realisations $\{x_1, x_2, x_3, \ldots\}$ of $X \sim Bern(p)$ through

the rule $x_i = \begin{cases} 1 & \text{if } u_i \le p \\ 0 & \text{if } u_i > p \end{cases}$

Jacob Bernoulli

(1655-1705)

Binomial Distribution

- A generalisation of Bernoulli variables gives rise to another commonly seen variable.

- Adding the outcomes of *n* identical independent Bernoulli variables gives a **Binomial** variable.

- If $X_1 \sim Bern(p), X_2 \sim Bern(p), \ldots, X_n \sim Bern(p)$ then $\left[X_1 + X_2 + \ldots + X_n\right] \sim Bin(n,p)$.

- Clearly $Y \sim Bern(p)$ and $Y \sim Bin(1,p)$ mean exactly the same thing.
- A binomial random variable requires two parameters:

- *n*: The number of independent Bernoulli variables
- *p*: The probability of a 1 from each Bernoulli variable

Binomial Distribution

- For $X \sim Bin(n,p)$, the probability mass function is

$$P(X=k)=\begin{cases}\dfrac{n!}{(n-k)!k!}p^k(1-p)^{n-k} & k\in\{0,1,2,\ldots,n\}\\ 0 & \text{otherwise}\end{cases}$$

- The easiest way to calculate the expectation or variance of this is through Bernoulli variables.

- We have that $E(X)=np$ and $Var(X)=np(1-p)$

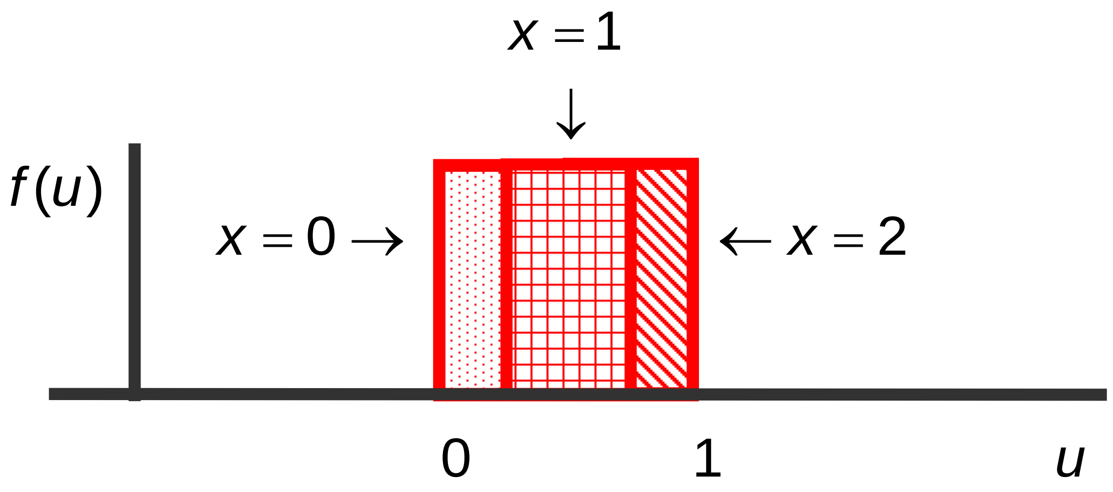
Binomial Distribution

- Consider the problem of simulating realisations of $X \sim Bin(2, 0.5)$
- We can do this "directly" by noting that $P(X = k) = \begin{cases} 0.25 & k = 0 \\ 0.5 & k = 1 \\ 0.25 & k = 2 \\ 0 & \text{otherwise} \end{cases}$

Binomial Distribution

- One rule for simulation from $X \sim Bin(2,0.5)$ would therefore be to use realisations

$\{u_1,u_2,u_3,\ldots\}$ of a $U[0,1]$ variable and set $x_i = \begin{cases} 0 & \text{if } u_i \le 0.25 \\ 1 & \text{if } 0.25 < u_i \le 0.75 \\ 2 & \text{if } u_i > 0.75 \end{cases}$



Simulating from a General Probability Mass Function

- Consider a random variable Z with probability mass function

$$P(Z=k)=\begin{cases} p_1 & k=s_1 \\ p_2 & k=s_2 \\ p_3 & k=s_3 \\ \vdots & \vdots \\ p_n & k=s_n \end{cases} \text{ where } \sum_{i=1}^{n} p_n = 1$$

- Realisations can be obtained from realisations u_i of a uniform variable by the rule

$$z_i=\begin{cases} s_1 & \text{if } u_i \le p_1 \\ s_2 & \text{if } p_1 < u_i \le p_1+p_2 \\ s_3 & \text{if } p_1+p_2 < u_i \le p_1+p_2+p_3 \\ \vdots & \vdots \\ p_n & \text{if } p_1+p_2+\ldots+p_{n-1} < u_i \end{cases}$$

Convolutions

- An alternative (and probably simpler) way to simulate from a binomial distribution is through **convolutions**.

- We know that a $Bin(n,p)$ can be obtained by summing n independent $Bern(p)$ variables.

- Instead of working out a decision rule from a relatively complicated probability mass function

$$P(X=k)=\begin{cases}\dfrac{n!}{(n-k)!k!}p^k(1-p)^{n-k} & k\in\{0,1,2,\ldots,n\}\\ 0 & \text{otherwise}\end{cases}$$

we can instead simply apply the much simpler Bernoulli decision rule n times and sum these.

Convolutions

- For example, we can use the realisations

$\{u_1, u_2, \ldots, u_5\} = \{0.114, 0.763, 0.906, 0.300, 0.476\}$ and $\{u_6, u_7, \ldots, u_{10}\} = \{0.887, 0.03, 0.531, 0.617, 0.297\}$

to generate two realisations of $X \sim Bin(5, 0.5)$.

- We can convert the uniform decimals into Bernoulli outcomes by the rule $y_i = \begin{cases} 0 & \text{if } u_i \le 0.5 \\ 1 & \text{if } u_i > 0.5 \end{cases}$.

- $\{u_1, u_2, \ldots, u_5\} = \{0.114, 0.763, 0.906, 0.300, 0.476\}$ hence $\{y_1, y_2, \ldots, y_5\} = \{0, 1, 1, 0, 0\}$ and $\{u_6, u_7, \ldots, u_{10}\} = \{0.887, 0.03, 0.531, 0.617, 0.297\}$ hence $\{y_6, y_7, \ldots, y_{10}\} = \{1, 0, 1, 1, 0\}$

- Summing these gives two realisations of $X \sim Bin(5, 0.5)$, $\begin{aligned} x_1 &= y_1 + y_2 + \ldots + y_5 = 2 \\ x_2 &= y_6 + y_7 + \ldots + y_{10} = 3 \end{aligned}$

Geometric Distribution

- One other common random variable which can arise from independent Bernoulli trials is a **Geometric** variable.

- We write $X \sim Geo(p)$ if X is the number of successive independent identical Bernoulli variables until the first 1 is obtained.

- For example, when flipping a fair coin repeatedly, the number of flips until the first Heads $\sim Geo(0.5)$.
- The range of $X \sim Geo(p)$ is easily seen to be $\{1, 2, 3, \ldots\}$.

Geometric Distribution

- When considering a number of independent $Bern(p)$ variables, we obtain the first 1 on the kth variable if and only if the first $(k-1)$ are 0s and the kth is a 1.

- That is, $P(X=k)=\begin{cases}(1-p)^{k-1}p & k=1,2,3,\ldots\\ 0 & \text{otherwise}\end{cases}$

- We can verify that this is a valid probability mass function since

$$\sum_{k=1}^{\infty}P(X=k)=\sum_{k=1}^{\infty}(1-p)^{k-1}p \quad =p+(1-p)p+(1-p)^2p+(1-p)^3p+(1-p)^4p+$$

- This is a geometric series, first term p, common ratio $(1-p)$.
- The infinite sum is therefore $\sum_{k=1}^{\infty}P(X=k)=\dfrac{p}{1-(1-p)}=1.$

Geometric Distribution

- Again, rather than simulating from a geometric distribution directly, we can use its relation to a Bernoulli distribution for an easier method of simulating realisations.

- We do not need to calculate an assignment rule from $P(X=k)=\begin{cases}(1-p)^{k-1}p & k=1,2,3,\ldots \\ 0 & \text{otherwise}\end{cases}$

- Instead, we can simply generate Bernoulli (0 or 1) variables and then count consecutively how many variables we observe until the first 1.

- Knowing the relationships between common standard distributions can make many simulations or calculations considerably easier.