STAT20060 - Statistics and Probability Handout 4 - Random Variables

Damien McParland

Random Variables

A random variable is a variable which assumes numerical values associated with the random outcomes of an experiment, where one (an only one) numerical value is assigned to each sample point.

For example,

1. Number of defective items in a batch.
2. Number of cars crossing a bridge in a day.
3. Highest daily temperature in Dublin.
4. Trading price of a gold bullion each day.

Types of random variables

Two types

There are two types or random variables:

- A random variable is said to be discrete if it can assume only a countable number of values.
- A random variable that can assume values corresponding to any of the points contained in one or more intervals is called continuous.

- Examples 1 and 2 on the previous slide are discrete random variables.
- Examples 3 and 4 on the previous slide are continuous random variables.

Probability distribution (discrete case)

Discrete probability distribution

The probability distribution for a discrete random variable Y can be represented by a formula, a table, or a graph, which provides the probabilities $p(y)$ corresponding to each and every value of y.

For any discrete probability distribution the following must be true:

1. $0 \leq p(y) \leq 1$ for all y.
2. $\sum_y p(y) = 1$ where the summation is over all possible values of y.

Example: Probability distribution (discrete)

- X – Number observed after rolling a fair die:

	$X = 1$	$X = 2$	$X = 3$	$X = 4$	$X = 5$	$X = 6$
$p(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

- Y – Number of heads observed in two coin tosses:

	$Y = 0$	$Y = 1$	$Y = 2$
$p(y)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Expected value of discrete random variables

Expected Value

The mean or expected value of a discrete random variable is:

$$\mu = \mathbb{E}(X) = \sum_x xp(x)$$

For example, consider rolling a die again and let X be the number observed:

$$\begin{aligned}
\mathbb{E}(X) &= \sum_x xp(x) = 1.\left(\frac{1}{6}\right) + 2.\left(\frac{1}{6}\right) + \ldots + 6.\left(\frac{1}{6}\right) = 3.5 \\
\mathbb{E}(Y) &= \sum_y yp(y) = 0.\left(\frac{1}{4}\right) + 1.\left(\frac{1}{2}\right) + 2.\left(\frac{1}{4}\right) = 1
\end{aligned}$$

Expected value of discrete random variables

- Note: A random variable may never be equal to its expected value. The expected value is the mean of the probability distribution or a measure of central tendency.

- The expected value of a function of a random variable $g(X)$ can be found similarly:

$$\mathbb{E}(g(X)) = \sum_x g(x)p(x)$$

e.g. From the previous example, let $g(X) = X^2$:

$$\mathbb{E}(X^2) = \sum_x x^2 p(x) = 1^2.\left(\frac{1}{6}\right) + 2^2.\left(\frac{1}{6}\right) + \ldots + 6^2.\left(\frac{1}{6}\right) = 15.17$$

Variance of a random variable

Variance

The variance of a discrete random variable X is:

$$\sigma^2 = \mathbb{E}\left[(x-\mu)^2\right] = \sum (x-\mu)^2 p(x) = \sum x^2 p(x) - \mu^2$$

- The variance is a measure of the spread of the probability distribution.

The standard deviation of a discrete random variable is the square root of the variance:

$$\sigma = \sqrt{\sigma^2}$$

- A sometimes more convenient formula for variance is:

$$\sigma^2 = \mathbb{E}(X^2) - \mu^2$$

Example: Dice

- Let X be the number displayed after rolling one die.
- From previous calculations we know that:
  - $\mathbb{E}(X) = 3.5$
  - $\mathbb{E}(X^2) = 15.17$
- What is the standard deviation of X?

Example: Car Ratings

Cars are rated on a scale of 1-5 stars with regard to the level of protection they offer in a head-on collision. 98 cars were tested and the star allocations are summarised in the following table:

$*$	$**$	$***$	$****$	$*****$
0	4	17	59	18

- Calculate the probability distribution for Y, the number of stars obtained by one of the cars selected at random.
- Calculate $\mathbb{P}(Y \leq 3)$.
- Hence calculate $\mathbb{P}(Y > 3)$.
- What is the expected value of Y?
- Calculate the standard deviation of Y.

Properties of Expectation and Variance:

Expectation:

1. $\mathbb{E}(c) = c$ where c is a constant.

2. $\mathbb{E}[cg(X)] = c\mathbb{E}[g(X)]$,
   in particular $\mathbb{E}[cX] = c\mathbb{E}[X]$

3. If $g_1(X)$, $g_2(X), \ldots, g_k(X)$ be k functions of X. Then

$$\mathbb{E}[g_1(X) + \ldots + g_k(X)] = \mathbb{E}[g_1(X)] + \ldots \mathbb{E}[g_k(X)]$$

Variance:

1. $Var(c) = 0$ where c is a constant.

2. $Var(a_1X_1 + \ldots + a_mX_m) = a_1^2 Var(X_1) + \ldots a_m^2 Var(X_m)$
   if $X_1, \ldots, X_m$ are **independent** random variables.

Example:

The number of trades on which a trader makes a loss per day, Y, is assumed to have a mean and variance of 4. If the profit made by the trader per day is given by $X = 1000 - 10Y - 5Y^2$. Find the expected profit per day.

Let O be a random variable denoting overhead costs, which has standard deviation 50. If the profit made by a trading team was given by $P = 7X - 1.5O - 10$, what is the standard deviation of P? Assume that O and X are independent and $\mathbb{E}(X^2) = 800,000$.

Binomial Distribution

- Suppose an experiment is conducted where there are a fixed number (n) of independent trials.

- For each trial we observe either a success or a failure.

- Further, suppose that the probability $\mathbb{P}(\text{Success}) = p$ for all trials. Also, $\mathbb{P}(\text{Failure}) = q = 1 - p$.

- Then, the number of observed successes, X, has a probability mass function (pmf) equal to

$$p(x) = \mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ for } x = 0, 1, \ldots, n.$$

- We write $X \sim \text{binomial}(n, p)$.

Examples

Binomial

- Roll a die 10 times, Y is the number of 6s observed.
- 100 consumers are selected at random and given a blind taste test. Each states which of two brands (A and B) of chocolate they prefer. Y is the number who prefer brand A.

Not Binomial

- In the run up to an election every person in 5 city neighbourhoods are asked which of two candidates they will vote for. This is not binomial since the people are not chosen at random and so are unlikely to be independent. It is likely that people who live in the same neighbourhood will have similar income levels, education etc. and so will vote in a similar way.

Binomial Distribution: Explained

- p^x is the probability of x successes.
- $(1-p)^{n-x}$ is the probability of $n-x$ failures.
- $\binom{n}{x}$ is the number of ways of observing x successes in n trials.

Binomial Distribution: Properties

- If $X \sim \text{binomial}(n, p)$ then:

$$\begin{aligned}
\mathbb{E}(X) &= np \\
\mathbb{V}\text{ar}(X) &= np(1-p) \\
\mathbb{P}\{X \leq x\} &= \sum_{y=0}^{x} \binom{n}{y} p^y (1-p)^{n-y}, \text{ for } x = 0, 1, \ldots, n.
\end{aligned}$$

Binomial Distribution: Example

- A financial broker sells products to customers.
- Suppose he sells twenty products.
- Each product has probability $p = 0.9$ of making money for the customer.
- What is the probability that at least 18 of the products make money?

Binomial Distribution: Example (Revisited)

- What is the probability that at most 2 products lose money?

- **Result:** If $X \sim \text{binomial}(n, p)$ and if $Y = n - X$.
  Then,
  $$Y \sim \text{binomial}(n, 1 - p).$$

Hypergeometric Random Variables

Characteristics of a Hypergeometric Random Variable

1. The experiment consists of randomly drawing n elements without replacement from a set of N elements, s of which are 'special' and $(N - s)$ are regular.
2. The hypergeometric random variable X is the number of special items in the draw of n elements.

Examples:

- The lottery – 49 numbers in total (N), 6 special numbers (s), you pick 6 (n) and X is the number of special numbers you match on your ticket.
- The government inspect bridges each year. Out of a total of 20 bridges 5 are not up to code. If a sample of 8 of the bridges are inspected then the number of defective bridges in that sample is a hypergeometric random variable.

The Hypergeometric Distribution

Hypergeometric Distribution

$$\mathbb{P}(X = x) = \frac{\binom{s}{x}\binom{N-s}{n-x}}{\binom{N}{n}}$$

$$\mu = \frac{ns}{N} \qquad\qquad \sigma^2 = n\left(\frac{s}{N}\right)\left(\frac{N-s}{N}\right)\left(\frac{N-n}{N-1}\right)$$

where,

- $N =$ Total number of elements.
- $s =$ Number of special items in N elements.
- $n =$ Number of elements drawn.
- $x =$ Number of special items in the n elements.

Example: Catalysts

An experiment is conducted to select a suitable catalyst for the commercial production of ethylenediamine (EDA), a product used in soaps. Suppose a chemical engineer randomly selects 3 catalysts for testing from a group of 10 catalysts, 6 of which have low acidity and 4 of which have high acidity.

- Find the probability that no highly acidic catalyst is selected.
- Find the probability that exactly one highly-acidic catalyst is selected.
- How many highly acidic acids would you expect to be chosen?
- What is the standard deviation of the number of highly acidic catalysts chosen?

Poisson Distribution

- Suppose an experiment is conducted where the number of events are counted for:
  - a fixed time period
  - a given length
  - a given area
  - a given volume
  - ...
- The probability that even occurs in a given unit of time/length/area/volume is the same for all units.
- The number of events that occur in disjoint units of time/length/area/volume are independent.

Poisson Random Variables

Examples

- The number of vehicles crossing a bridge in a day.
- The number of jobs sent to a computer processor in an hour.
- The number of particles of a pollutant in a litre of river water.

The Poisson probability distribution

The probability distribution for a Poisson random variable Y is given by:

$$\mathbb{P}(Y = y) = \frac{\lambda^y e^{-\lambda}}{y!}$$

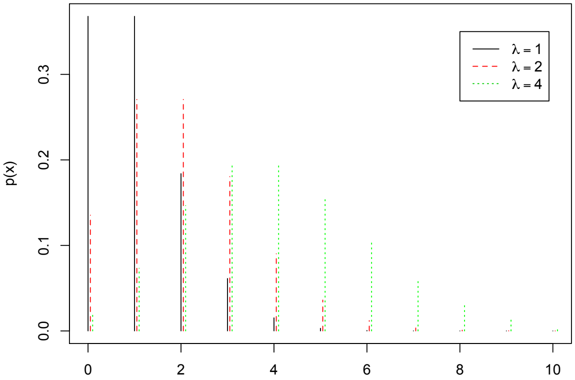
where

- $\lambda =$The rate events occur during a given unit of time, area or volume.
- The mean and variance of a Poisson random variable are:
  - $\mathbb{E}(Y) = \mu = \lambda$
  - $\text{Var}(Y) = \sigma^2 = \lambda$

Poisson Distribution: Properties

- The parameter λ is called the rate parameter.
- It turns out that

$$\mathbb{E}(X) = \lambda = \mu \text{ and } \mathbb{V}\mathrm{ar}(X) = \lambda = \sigma^2.$$



Poisson Distribution: Property

- Suppose we sum two Poisson random variables, then the sum is also Poisson.

- That is, if
$$X \sim \text{Poisson}(\lambda) \text{ and } Y \sim \text{Poisson}(\mu),$$
then
$$X + Y \sim \text{Poisson}(\lambda + \mu).$$

Poisson Distribution: Class Example

- Suppose that faults occurring in cables being manufactured follow a Poisson process with a rate of 2 faults per 100 m.
  - What is the probability of no faults in 100 m of cable?
  - What is the probability of one fault in 100 m of cable?
  - What is the probability of two or more faults in 100 m of cable?
  - What is the probability of no faults in 200 m of cable?

Continuous Random Variables

- Recall: A continuous random variable is one which can assume values corresponding to any of the points contained in one or more intervals.

- The graphical form of the probability distribution for a continuous random variable X is a smooth curve called a probability density function (pdf) and is denoted by $f(x)$.

For $f(x)$ to be a valid density we must have:

$$f(x) \geq 0 \ \forall \ x \text{ and } \int_{-\infty}^{+\infty} f(x) = 1$$

- Picture:

Continuous Random Variables

- The probability that X takes a value between two points a and b is given by the area under the density curve between these points.
- Hence the total area under the curve must be 1 in order to satisfy the second probability axiom. ($\mathbb{P}(S) = 1$)
- Since there is no area over a single point:

$$\mathbb{P}(X = x) = 0 \text{ for all } x$$

Thus, $\mathbb{P}(a < x < b) = \mathbb{P}(a \leq x \leq b)$.

Cumulative Distribution Function

- The cumulative distribution function, $F(x)$ gives the area under the density curve to the left of a point x.

- It is found by integrating the density function between the lower limit of the range of the random variable and x.

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(x)dx$$

- Note: The lower limit of the range of a random variable is not always $-\infty$.

Mean and Variance

- Expected value of X:

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x\, f(x) dx$$

- Expected value of $g(x)$:

$$\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x)\, f(x) dx$$

- As in the discrete case:

$$\text{Var}(X) = \sigma^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

- Note that the integration above should take place over the range of X. This may not always be $(-\infty, +\infty)$

The Uniform Distribution

- Continuous random variables that have equally likely outcomes over their range of possible values possess a uniform probability distribution.

- Picture:

The Uniform Distribution

Uniform Distribution

$X \sim U(a, b)$

- pdf: $f(x) = \frac{1}{b-a}$
- Mean: $\mu = \frac{a+b}{2}$
- Variance: $\sigma^2 = \frac{(b-a)^2}{12}$
- $\mathbb{P}(c < x < d) = (d - c)/(b - a)$, where $a \leq c < d \leq b$

Since the areas under the density function are simple rectangles there is no need to integrate. However, integrating will provide the same answers (CHECK!).

Example: Uniform

A company's rolling machine is producing sheets of steel of varying thickness. The thickness Y is a uniform random variable taking values between 1.5 and 2 milimeters. Any sheets less than 1.6 milimeters thick must be scrapped.

1. Calculate the mean and standard deviation of Y. Graph the probability density function and mark the mean as well as 1 and 2 standard deviation intervals around the mean on the horizontal axis.
2. If this distribution models this situation correctly, what proportion of sheets would you expect to be scrapped.

Exponential Distribution

- The exponential distribution is an example of a waiting time distribution.
- Time between machine breakdowns, time between natural disasters, time between cars arriving at a toll booth etc. are all well modelled by an exponential distribution.

The exponential distribution

- Probability density function : $f(x) = \lambda e^{-\lambda x}$.
- Expected value: $\mu = \frac{1}{\lambda}$.
- Variance: $\sigma^2 = \frac{1}{\lambda^2}$

- Picture:

Probabilities and the Exponential Distribution

- Note that the standard deviation of an exponentially distributed random variable is equal to its mean.
- As usual with continuous random variables, probabilities are found by integrating the density function over the interval of interest.
- It can be convenient to use the cumulative distribution function instead.

CDF for an exponential random variable

If $X \sim Exp(\lambda)$ where λ is the rate then,

- $\mathbb{P}(X \leq x) = 1 - e^{-\lambda x}$.
- Hence, $\mathbb{P}(X > x) = e^{-\lambda x}$

Example: Exponential – Machine Breakdown

Suppose the the length of time (in days) between breakdowns for a particular machine in a manufacturing plant is exponentially distributed with mean 18. What is the probability that at least 3 weeks pass without a breakdown?

The Memoryless Property

- The exponential distribution is said to be memoryless since:

$$\mathbb{P}(X > a + b | X > a) = \mathbb{P}(X > b)$$

- This has important implications for the exponential's role as a waiting time distribution.
- Only the duration of the future waiting time affects the probability it will be surpassed, not the amount of time which has already elapsed.
    - For machine breakdown this is realistic.
    - For human lifetimes it is not.

Poisson and Exponential Distributions

- The Poisson distributions is a discrete counting process for the number of events that occur in an interval.

- If events occur according to a Poisson process with rate λ then the probability of k events in an interval of duration t is:

$$\mathbb{P}(X = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

- Consider the random variable t_0, the time until the first event in a Poisson process.

- If we are interested in the probability that t_0 exceeds some given time t^* then we require 0 events to occur in the interval $[0, t^*]$

Poisson and Exponential Distributions

- From above this probability is:

$$\mathbb{P}(X = 0) = \frac{(\lambda t^*)^0 e^{-\lambda t^*}}{0!} = e^{-\lambda t^*}$$

- But for an exponential random variable Y with rate parameter λ:

$$\mathbb{P}(Y > t^*) = e^{-\lambda t^*}$$

- Thus the time until the first event in a Poisson process with rate parameter λ per unit time follows an exponential distribution with rate parameter λ.

- The time between consecutive events also follows an exponential distribution with rate parameter λ.

Example:

The number of ice creams sold at a seaside stall follows a Poisson distribution with a rate of 20 per day. What is the probability it takes less than half an hour to sell the first ice-cream of the day? (Assume the working day is 8 hours.)

The Weibull Distribution

- The Weibull is another waiting time distribution.

If $X \sim \text{Weibull}(\alpha, \beta)$ then the density function is:

$$f_X(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^{\alpha}}$$

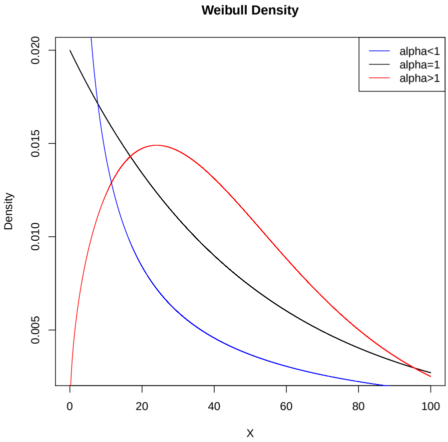
- The shape parameter, α, describes the failure rate.

  $\alpha < 1$: High failure rate to begin with. Decreasing over time.
  e.g. High infant mortality.

  $\alpha = 1$: Failure rate is constant over time.
  e.g. Random external events are causing mortality or failure.

  $\alpha > 1$: Failure rate increases with time.
  e.g. Systems which are affected by aging, parts which are more likely to fail as time goes on.

The Weibull Distribution



- A Weibull random variable with $\alpha = 1$ is an exponential random variable with rate parameter $\frac{1}{\beta}$. Hence the exponential distribution is a special case of the Weibull distribution.

Example:

A company requires a new ventilation system to be installed to meet government regulations. If it is installed within 100 days the cost will be €10,000. If it is installed after the 100 day mark the cost is reduced to €8,000 but there is a 60% chance the government will find out and fine them €1,000. Calculate the expected cost of installing the system if the installation time is t where:

$$t \sim \text{Weibull}(\alpha = 1/4, \beta = 4)$$

The Normal Distribution

- The normal distribution is one of the most useful and widely used distributions in statistics.

The Normal Distribution

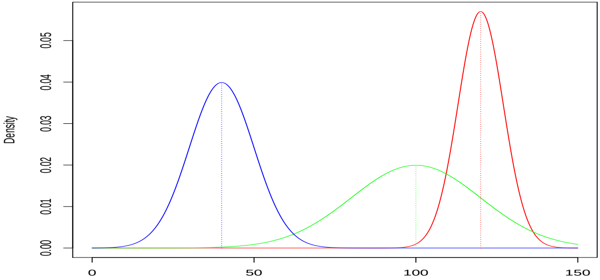
The density function of a normal random variable Y is given by:

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/(2\sigma^2)}$$

The parameters μ and σ^2 are the mean and variance of the normal random variable Y.

- Picture:

The Normal Distribution



$(\mu = 40,\ \sigma = 10)(\mu = 100,\ \sigma = 20)\ (\mu = 120,\ \sigma = 7)$

- The location is governed by μ.
- The spread is governed by σ.
- The height is also governed by σ since the area under each curve must be 1.

The Normal Distribution

- Picture:

- $\mu \pm \sigma$ includes 68% of observations.
- $\mu \pm 1.96\sigma$ includes 95% of observations.
- $\mu \pm 2.58\sigma$ includes 99% of observations.

The Standard Normal Distribution

- There is no closed form expression for the integral of the normal density function. Approximations are obtained using numerical methods and these are tabulated.

A normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$ is called a standard normal distribution. It is often denoted by Z.

- Note: Since the normal distribution is symmetric. If $X \sim N(0, \sigma^2)$ then

$$\mathbb{P}(X < -a) = \mathbb{P}(X > a)$$

- Picture:

The Normal Distribution

If Y is a normal random variable with mean μ and variance σ^2,

$$Y \sim N(\mu, \sigma^2)$$

then,

$$Z = \frac{Y - \mu}{\sigma}$$

is a standard normal random variable.

$$Z \sim N(0, 1)$$

- This fact will be used to find probabilities from any normal distribution.

Example:

Suppose $X \sim N(\mu = 5, \sigma^2 = 3)$.

- Calculate $\mathbb{P}(X \leq 7)$.
- Calculate $\mathbb{P}(X \geq 6)$.
- Calculate $\mathbb{P}(X \leq 2)$.
- Calculate $\mathbb{P}(2.5 \leq X \leq 7)$

Example: Paper Friction

In a study to investigate the paper feeding process in a photocopier the coefficient of friction is a proportion which measures the degree of friction between adjacent sheets of paper in the paper stack. This coefficient is assumed to be normally distributed with mean 0.55 and standard deviation 0.013. During system operation the friction coefficient is measured at randomly selected times.

1. Find the probability that the friction coefficient falls between 0.53 and 0.56.
2. Is it likely to observe a friction coefficient below 0.52? Explain.

The Normal Distribution

- It can also be useful to use the normal tables in reverse.
- Suppose $X \sim N(2,2)$, for what value a is $\mathbb{P}(X \leq a) = 0.1$?

$$\mathbb{P}(X \leq a) = 0.1$$
$$\mathbb{P}\left(\frac{X-2}{\sqrt{2}} \leq \frac{a-2}{\sqrt{2}}\right) = 0.1$$

From tables $\mathbb{P}(Z \geq 1.28) = 0.1$ and since the standard normal distribution is symmetric this means that $\mathbb{P}(Z \leq -1.28) = 0.1$. So,

$$\begin{aligned} \frac{a-2}{\sqrt{2}} &= -1.28 \\ \Rightarrow a &= (-1.28)(\sqrt{2}) + 2 \\ \Rightarrow a &= 0.19 \end{aligned}$$

Example: Exam Results

Scores on an examination are assumed to be normally distributed with a mean of 78 and a variance of 36.

- Suppose that students scoring in the top 10% of this distribution are to receive an A grade. What is the minimum score a student must achieve to obtain an A grade?
- What must be the cutoff for passing the exam if the examiner wants only the lowest 25% of all scores to fail?
- Find, approximately, what proportion of students have scores 5 or more points above the passing cutoff.

Example: Skyscraper

The lifetime of steel joists in a skyscraper are independent and normally distributed with a mean of 50 years and a standard deviation of 6 years.

1. What is the probability that a steel joist lasts more than 55 years?
2. What proportion of the steel joists would you expect to last between 42 and 58 years?
3. In a random sample of 4 joists what is the probability that 2 last less than 55 years?