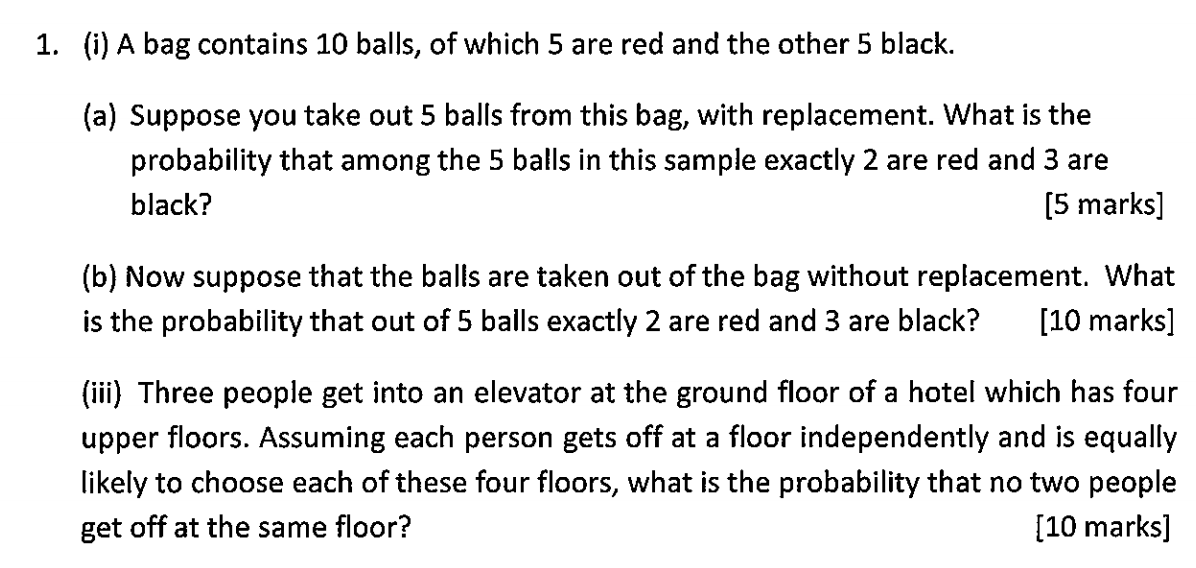
**ST3009 – Statistics**

**2017 Exam Solutions**

**Question 1**



i)

Probability of red = 0.5

Probability of black = 0.5

ii)

There are ways of drawing 5 balls from 10.

There are ways of drawing 2 red balls from 5.

There are ways of drawing 3 black balls from 5.

Therefore, the probability is:

iii) 4 floors

3 guests

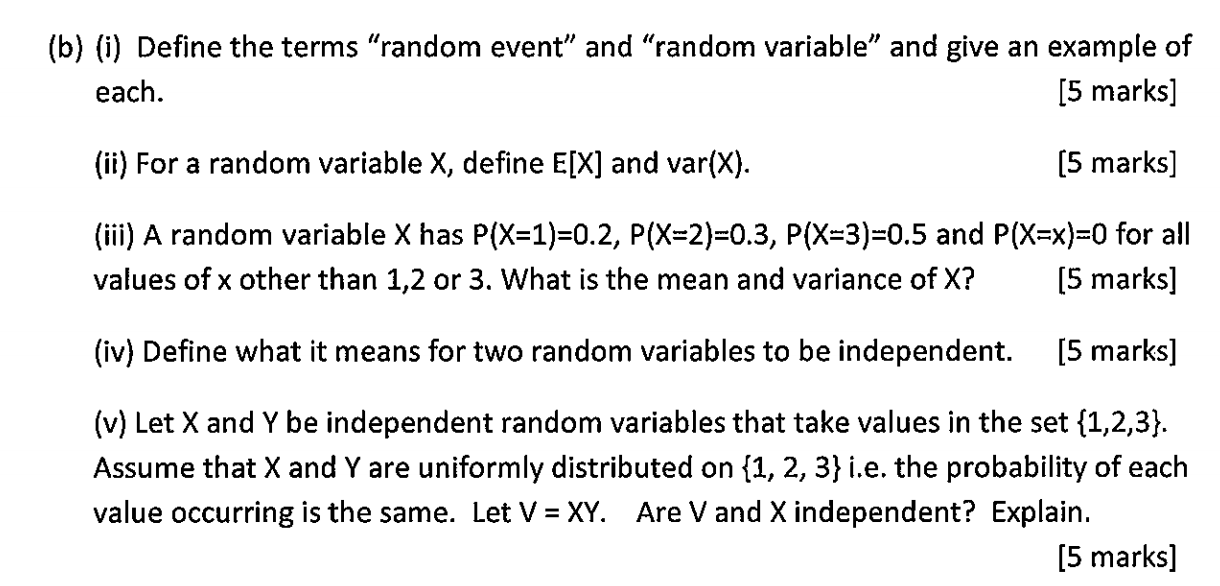
P(floorX) = 0.25

First guest has 4 floors to choose from, second has 3, third has 2.

Total combinations = 4 \* 3 \* 2 = 24

There are 4^3 ways of 3 guests choosing from 4 floors, therefore probability is:

**Question 2**

****

i)

**Random Event:** A random event is a subset of the sample space. Consider the tossing of two coins. The event {H, H} is a random event which is a subset of the sample space.

**Random Variable:** A random variable maps a random event to a real number. Consider the tossing of a six-sided die. Let the random variable X denote the number tossed by the die. X can take the values [1,6].

ii)

iii)

iv) Two random variables X and Y are independent iff:

Holds for all values of x and y that variables X and Y can take.

v) X and Y can take values {1, 2, 3}

P(X/Y = 1/2/3) = p

V = XY

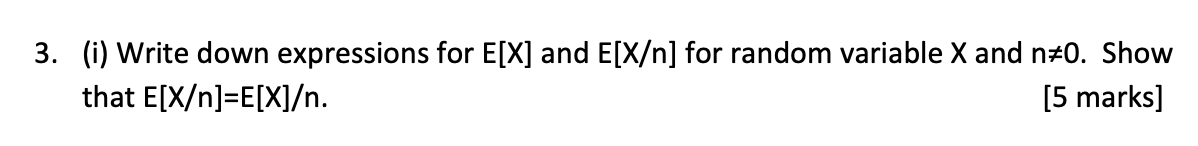
To verify they are dependent consider the example P(V=1 and X=2).

P(V=1) = P(X=1 and Y=1) = (1/3)(1/3) = 1/9

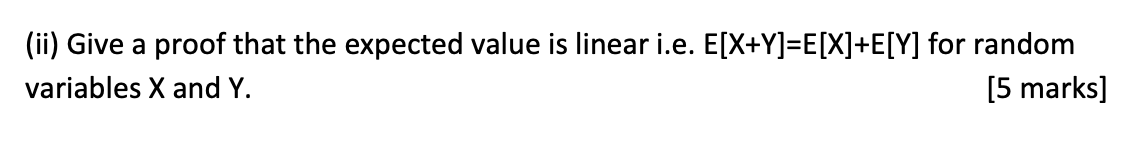
P(X=2) = 1/3

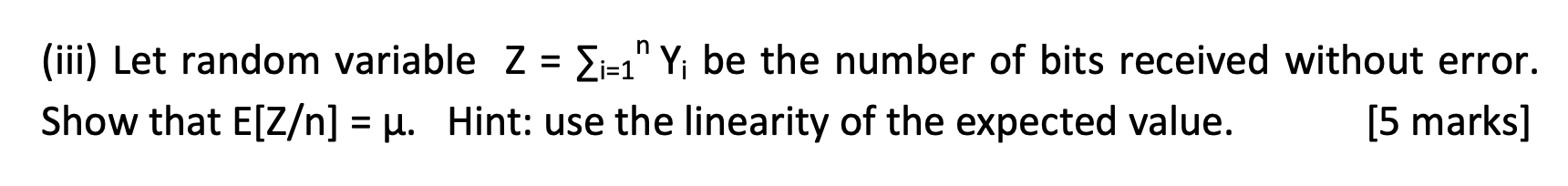
P(V=1 and X=2) = 0 since there is no value of Y for which V=XY=1 when X=2. Therefore V and X are not independent.

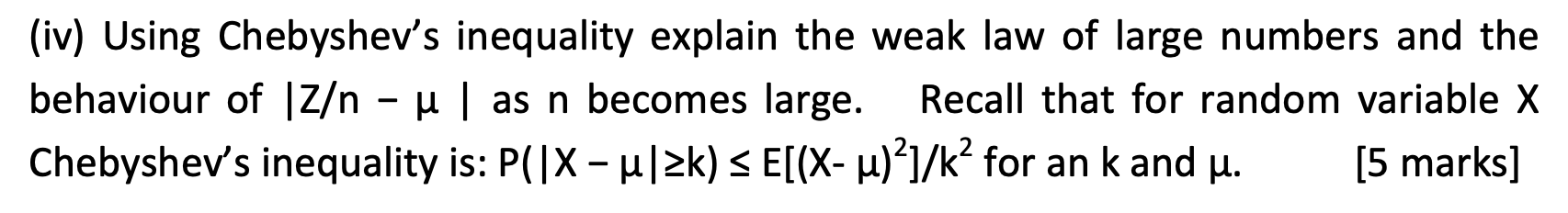
**Question 3**

****

Therefore, it is clear that E[X/n] = E[X]/n.



****

****

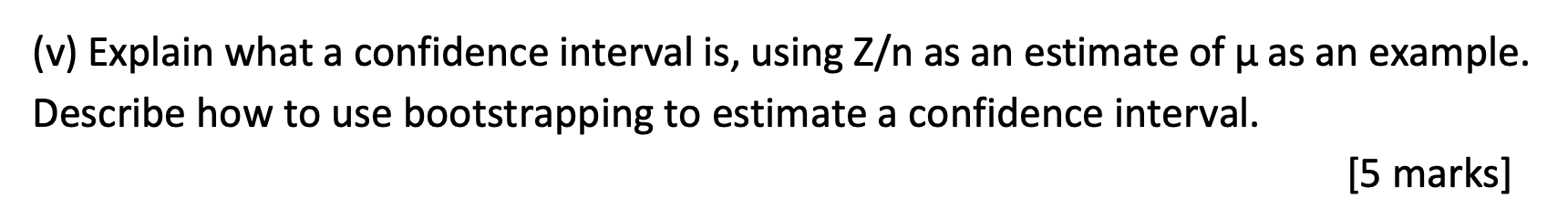
First, calculate the variance. Since our Yi’s are independent we can represent that variance as:

represents a bit being an error or not, thus is an indicator variable and can only take values 0 or 1. As a result:

Using this, we can then calculate Var(Yi):

Then plugging this into Chebyshev’s inequality:

From this, we can see that as n goes to infinity, goes to 0.



A confidence interval is typically a statement of the form:

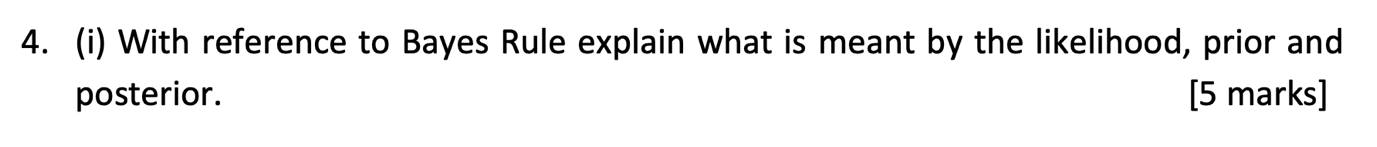
Where c represents the confidence that a random variable X lies between a and b. For example c might be 0.95.

is an example of the following confidence interval:

Bootstrapping can be used to estimate a confidence interval as follows. Suppose we have observed n values of . In bootstrapping we re-sample (with replacement) from these observed values. Letting S be the indices of the values sampled, we then calculate:

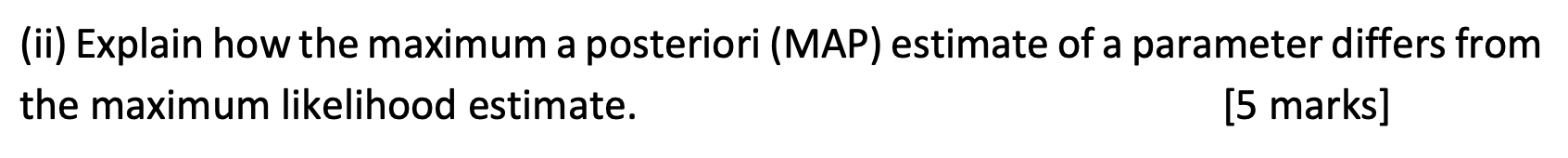
Repeating this we obtain a sequence of estimates for from which we can estimate the distribution of (from the fraction of times each value appears). Using this estimated distribution we can now either calculate the value c for a confidence interval by just summing up the fraction of values lying in the interval of interest or for a specified value of c we can calculate an interval over which the sum of the fractions is greater than or equal to c.

**Question 4**



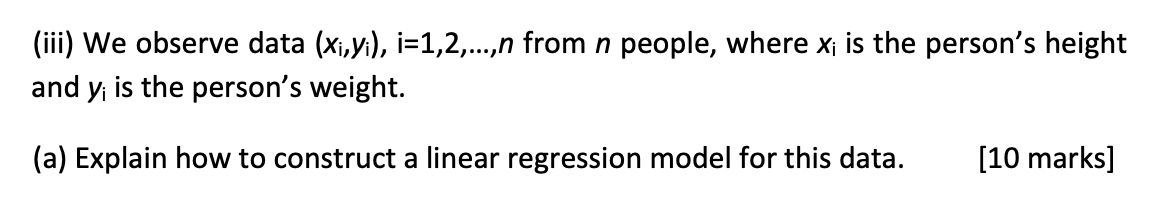
For random events E and F, Bayes Rule states:

* is the Likelihood
* is the Prior
* is the Posterior



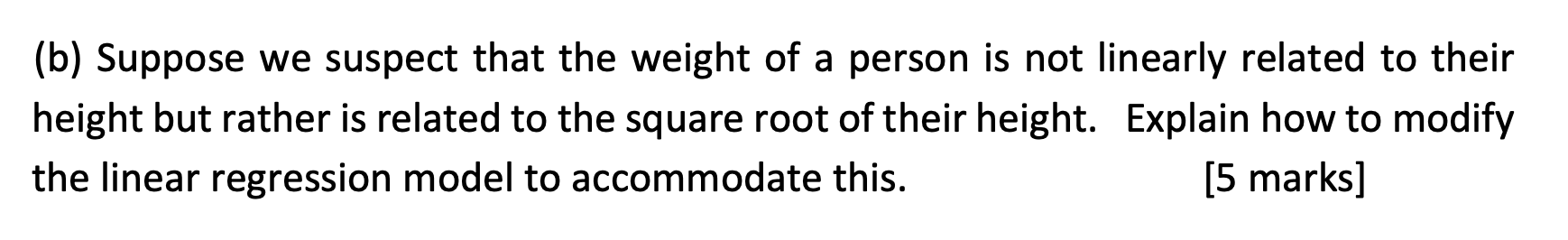
A MAP maximises the posterior P(E|F)

A ML maximises the likelihood P(F|E)



We model each value as the sum of an underlying linear function plus zero-mean Gaussian noise i.e as the following (where ni is Gaussian noise):

We then typically select the value for that maximises the likelihood, or equivalently maximises the log-likelihood:



We can change the model to be as:

And now select that maximises: