# MATH/STAT 3360, Probability
FALL 2012

Toby Kenney

In Class Examples

**question**

A statistics textbook has 8 chapters. Each chapter has 50 questions. How many questions are there in total in the book?

**question**

How many 9-digit numbers are there that contain each of the digits 1–9 once.

## question

A band are planning a tour of Canada. They are planning to perform in 15 venues across Canada, before returning to their home in Halifax. They want to find the cheapest way to arrange their tour. The only way to do this is to work out the cost of all possible orderings of the venues. They have a computer, which can calculate the cost of 100,000,000 orderings of venues every second.

(a) How long will it take the computer to calculate the cost of all possible orderings of the venues?

(b) What if they want to extend the tour to 18 venues?

**question**

How many 3-element subsets are there in a 12-element set?

## question

Show that for any $m, n$ $\sum_{m=0}^{n} \binom{2n}{2m} = \sum_{m=0}^{n-1} \binom{2n}{2m+1}$ [Hint: use the binomial theorem to expand $(a+b)^n$ for some choice of $a$ and $b$.]

## question

You are revising for an exam. You need to memorise 1,000 pages from the text book. On the day you start revising, you can memorise 30 pages from the book. However, every day that you spend revising leaves you more tired the next day, and the number of pages that you can memorise decreases by 1 (so that on the second day, you can only memorise 29 pages, and on the third day 28, and so on). You can increase the number of pages that you can memorise in a day back up to 30 by taking a two day break. How many days before the exam do you need to start revising in order to memorise all 1,000 pages?

**question**

How many distinct ways can the letters of the word "PROBABILITY" be arranged.

## question

A professor decides to award a fixed number of each grade to his students. He decides that among his class of 10 students, he will award one A, two B, three C, three D and one fail. How many different results are possible in the course.

## question

In the polynomial $(x + y + z)^6$, what is the coefficient of $x^2y^3z$?

## question

For events $A$, $B$, $C$ and $D$, how are the events $(A \cap B) \cup (CD)$ and $(A \cup C)(B \cup D)$ related?

## question

For an experiment with sample space $\{1,2,3,4,5\}$, is there a probability $P$ such that $P(\{1,2\}) = 0.4$, $P(\{1,3\}) = 0.6$ and $P(\{2,3\}) = 0.1$?

## question

For events $A$, $B$ and $C$, we know that $P(A) = 0.6$, $P(B) = 0.5$ and $P(C) = 0.4$, and that $P(AB) = 0.4$, $P(AC) = 0.3$ and $P(BC) = 0.2$, and $P(ABC) = 0.1$. What is $P(A \cup B \cup C)$.

## question

A fair coin is tossed 5 times.
(a) What is the probability that the sequence HHH occurs somewhere in the 5 tosses?
(b) What is the probability that the sequence THT occurs somewhere in the 5 tosses?

## question

What is the probability that a number chosen uniformly at random in the range 1–1,000,000 is divisible by at least one of 2, 3, or 5?

## question

You are planning to learn some languages. You want to maximise the number of people with whom you will be able to speak, or equivalently, the probability that you will be able to speak with a randomly chosen person. You perform some preliminary research and find out the following:

| Language | Proportion of people who speak it |
|---|---|
| Chinese | 19.4% |
| English | 7.3% |
| Hindi | 7.1% |
| Spanish | 6.0% |
| Arabic | 4.0% |

. . .

## question (continued)

You also investigate what proportion of people speak each pair of languages, as summarised in the following table.

| | Chinese | English | Hindi | Spanish | Arabic |
|---|---|---|---|---|---|
| Chinese | | 0.2% | 0.1% | 0.0% | 0.1% |
| English | 0.2% | | 1.8% | 1.6% | 0.7% |
| Hindi | 0.1% | 1.8% | | 0.0% | 0.0% |
| Spanish | 0.0% | 1.6% | 0.0% | | 0.1% |
| Arabic | 0.1% | 0.7% | 0.0% | 0.1% | |

If you assume that the number of people speaking three languages is small enough to be neglected, which three languages should you learn in order to maximise the number of people to whom you can speak?

**question**

Two players play a game: they toss a coin repeatedly until one of the sequences `TTH` or `THT` comes up. If `TTH` comes up first, the first player wins. Otherwise the second player wins. What is the probability that the first player wins?

## question

What is the probability that a poker hand (with five cards in total) contains exactly three Aces?

## question

Two fair 6-sided dice are rolled.
(a) What is the probability that the larger of the two numbers obtained is 5, given that the smaller is 3. [By “the smaller number is 3” I mean that one of the numbers is 3, and the other is at least 3, so that if the roll is (3,3), the smaller number is 3.]
(b) One of the dice is red, and the other is blue. What is the probability that the number rolled on the red die is 5 given that the number rolled on the blue die is 3 and the number rolled on the red die is at least 3?
(c) Why are the answers to (a) and (b) different?

**question**

What is the probability that the sum of two dice is at least 7 conditional on it being odd?

## question

Two of your friends often talk to you about the books they like. From experience, you have found that if Andrew likes a book, the probability that you will also like it is 70%, while if Brian likes a book, the probability that you will also like it is 40%. You are shopping with another fried, who tells you that she remembers Andrew telling her that he liked the book you are considering buying, but she's not completely sure that it was Andrew who told her, rather than Brian. If the probability that it was Andrew who told her is 80%, what is the probability that you will like the book?

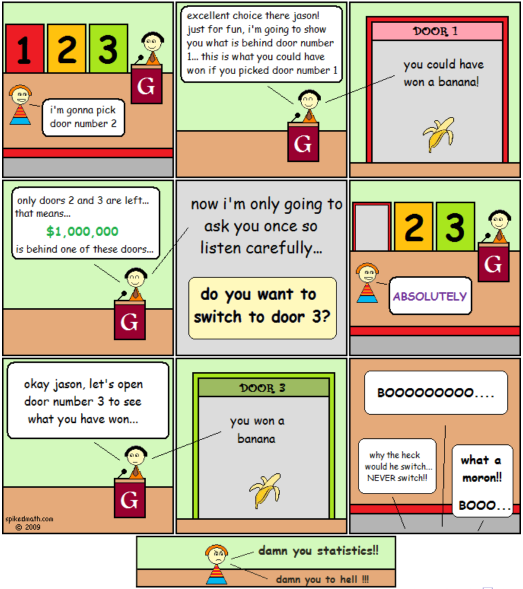
## question

You are a contestant in a game-show. There are three doors, numbered 1, 2, and 3, one of which has a prize behind it. You are to choose a door, then the host will open another door to show that the prize is not behind it (if the door you chose has the prize behind it, he chooses which door to open at random). You then have the opportunity to switch to the other unopened door. In this case, you have cheated and know that the prize is not behind door 2. You choose door 1, and the host opens door 2.

(a) Should you switch to door 3?

(b) What is your probability of winning (assuming you make the best choice of whether or not to switch)?

# Warning


1
2
3
G
i'm gonna pick door number 2
excellent choice there jason! just for fun, i'm going to show you what is behind door number 1... this is what you could have won if you picked door number 1
DOOR 1
you could have won a banana!
only doors 2 and 3 are left... that means...
$1,000,000
is behind one of these doors...
now i'm only going to ask you once so listen carefully...
do you want to switch to door 3?
ABSOLUTELY
okay jason, let's open door number 3 to see what you have won...
spikedmath.com
© 2009
DOOR 3
you won a banana
BOOOOOOOOO....
why the heck would he switch... NEVER switch!!
what a moron!!
BOOO...
damn you statistics!!
damn you to hell !!!

## question

A test for a particular disease has probability 0.005 of giving a positive result if the patient does not have the disease (and always gives a positive result if the patient has the disease). The disease is rare, and the probability that an individual has the disease is 0.0001.
(a) An individual is tested at random, and the test gives a positive result. What is the probability that the individual actually has the disease?
(b) A patient goes to the doctor with a rash, which is a symptom of this rare disease, but can also be caused by other problems. The probability that a person without the disease has this rash is 0.01. If the patients test comes back positive, what is the probability that he has the disease?

## question

A civil servant wants to conduct a survey on drug abuse. She is concerned that people may not admit to using drugs when asked. She therefore asks the participants to toss a coin (so that she cannot see the result) and if the toss is heads, to answer the question “YES”, regardless of the true answer. If the toss is tails, they should give the true answer. That way, if they answer “YES” to the question, she will not know whether it is a true answer. Suppose that 5% of the survey participants actually use drugs. What is the probability that an individual uses drugs, given that they answer “YES” to the question on the survey?

## question

4 fair coins are tossed. Which of the following pairs of events are independent?

| | | |
|---|---|---|
| (a) | The first coin is a head. | There are exactly two heads among the four tosses. |
| (b) | There are exactly two heads among the first three tosses | There are exactly two heads among the last three tosses. |
| (c) | The sequence `TTH` occurs in the 4 tosses. | The sequence `THT` occurs. |

## question

For three events $A$, $B$ and $C$, we know that $P(A|C) = 0.4$, $P(B|C) = 0.3$, $P(ABC) = 0.1$, $P(C) = 0.8$. Find $P(A \cup B|C)$.

### question

What is the cumulative distribution function for the sum of two fair dice?

**question**

What is the expected value of the product of the numbers rolled on two fair 6-sided dice.

### question

Two fair 6-sided dice are rolled, and the sum is taken. What is the variance of this sum?

## question

For a particular course, the probability of a student passing is 0.8. There are 15 students taking the course. (a) What is the probability that exactly 10 students pass?
(b) If we use the Poisson approximation to the Binomial, what would we calculate as the probability of exactly 10 students passing?

**question**

100 fair 6-sided dice are rolled, what is the expected value of the total of all 100 dice.

## question

An insurance company sets it's premium for car insurance at \$500. They estimate that the probability of a customer making a claim is $\frac{1}{200}$, in which case they will pay out \$5,000. They sell policies to 100 customers. After costs, the premiums are enough for them to pay out up to 5 claims. What is the probability that the insurance company will have to pay out more than 5 claims?

## question

You are bidding for an object in an auction. You do not know the other bids. You believe that the highest other bid (in dollars) will be unifomly distributed in the interval $(100, 150)$. You can sell the item for \$135. What should you bid in order to maximise expected profit?

## question

You are considering an investment. You would originally invest \$1,000, and every year, the investment will either increase by 60% or decrease by 60%, with equal probability. You plan to use the investment after 20 years.

(a) What is the expected value of the investment after 20 years?

(b) What is the probability that the investment will be worth more than \$1,000 after 20 years?

**question**

What is the expected value of the product of the numbers rolled on two fair 6-sided dice.

## question

You are bidding for an object in an auction. You do not know the other bids. You believe that the highest other bid (in dollars) will be unifomly distributed in the interval $(100, 150)$. You can sell the item for \$135. What should you bid in order to maximise expected profit?

### question

Two fair 6-sided dice are rolled, and the sum is taken. What is the variance of this sum?

## question

An insurance company sets it's premium for car insurance at $500. They estimate that the probability of a customer making a claim is $\frac{1}{200}$, in which case they will pay out $5,000. They sell policies to 100 customers. After costs, the premiums are enough for them to pay out up to 5 claims. What is the probability that the insurance company will have to pay out more than 5 claims?

## question

For a particular course, the probability of a student passing is 0.8. There are 15 students taking the course. (a) What is the probability that exactly 10 students pass?
(b) If we use the Poisson approximation to the Binomial, what would we calculate as the probability of exactly 10 students passing?

## question

100 fair 6-sided dice are rolled, what is the expected value of the total of all 100 dice.

## question

A gambler starts with \$5, and continues to play a game where he either wins or loses his stake, with equal probability, until he either has \$25 or has \$0. He is allowed to choose how much to bet each time, but is not allowed to bet more than he has, and does not bet more than he would need to win to increase his total to \$25. He decides how much to bet each time by rolling a fair (6-sided) die, and betting the minimum of the number shown, the amount he has, and the amount he needs to increase his total to \$25. What is the probability that he quits with \$25? [You may assume that he will certainly reach either \$25 or \$0 eventually. Hint: some of the information in this question is not necessary.]

## question

You are considering an investment. You would originally invest \$1,000, and every year, the investment will either increase by 60% or decrease by 60%, with equal probability. You plan to use the investment after 20 years.

(a) What is the expected value of the investment after 20 years?

(b) What is the probability that the investment will be worth more than \$1,000 after 20 years?

## question

What is the cumulative distribution function for the sum of two fair dice?

## question

A random variable $X$ has probability density function $f_X(x) = \frac{2}{\sqrt{3}\pi(1+\frac{x^2}{3})^2}$.

(a) What is the expectation $\mathbb{E}(X)$?

(b) What is the variance $\mathrm{Var}(X)$? [Hint: differentiate $\frac{1}{1+\frac{x^2}{3}}$, and use this to integrate by parts. Recall that $\frac{d \arctan\left(\frac{x}{\sqrt{3}}\right)}{dx} = \frac{1}{\sqrt{3}\left(1+\frac{x^2}{3}\right)}$, and that $\arctan(-\infty) = -\frac{\pi}{2}$ and $\arctan(\infty) = \frac{\pi}{2}$.]

## question

Which of the following distributions for $X$ gives the larger probability that $X > 3$?
(i) $X$ is uniformly distributed on $(-2, 7)$.
(ii) $X$ is uniformly distributed on $(0, 5)$?

## question

Mr. Wilson is driving to the garage to buy some winter tyres. His old tyres have a large worn patch, covering 20% of the tyre. If any part of this worn patch is on an icy patch in the road, the car will skid. There is an icy patch 30cm long on the road in front of him. If the circumference of his tyres is 1m, what is the probability that his car will go into a skid on this patch of ice?

## question

Let $X$ be normally distributed with mean 1 and standard deviation 1. What is the probability that $X^2 > 1$?

## question

An online banking website asks for its customers date of birth as a security question. Assuming that the age of customers who have an online banking account is normally distributed with mean 35 years and standard deviation 10 years:

(a) how many guesses would a criminal need to make in order to have a 50% chance of correctly guessing the date of birth of a particular customer?

(b) If the criminal estimates that he can safely make up to 200 guesses without being caught, what is the probability of guessing correctly.

## question

A scientist, Dr. Jones believes he has found a cure for aging. If he is right, people will no longer die of old age (but will still die of disease and accidents at the same rate). In this case, what would life expectancy be for people who do not grow old? [You may assume that people currently die from accidents and disease at a roughly constant rate before the age of about 50, and that after Dr. Jones' cure, they will continue die at this rate forever. Currently 93% of people live to age 50.]

**question**

If $X$ is an exponential random variable with parameter $\lambda$, what is the distribution function of $X^2$?

**question**

Calculate the probability density function of the square of a normal random variable with mean 0 and standard deviation 1.

**question**

A fair (6-sided) die is rolled 10 times. What is the probability of getting the same number of '5's and '6's?

## question

A fair (6-sided) die is rolled 50 times. What is the probability of getting exactly 7 sixes and 6 fives?

## question

Give an example of a joint distribution function for two continuous random variables $X$ and $Y$ such that for any $x$ for which $f_X(x) > 0$, $\mathbb{E}(Y|X = x) = 0$ and for any $y$ for which $f_Y(y) > 0$, $\mathbb{E}(X|Y = y) = 0$, but such that $X$ and $Y$ are not independent.

## question

Show that the minimum of two independent exponential random variables with parameters $\lambda_1$ and $\lambda_2$ is another exponential random variable with parameter $\lambda_1 + \lambda_2$.

## question

If $X$ is uniformly distributed on $(0, 2)$ and $Y$ is uniformly distributed on $(-1, 3)$, what is the probability density function of $X + Y$?

## question

The number of claims paid out by an insurance company in a year is a Poisson distribution with parameter $\lambda$. The number of claims it can afford to pay is $n$. It is considering merging with another insurance company which can afford to pay $m$ claims and which will have to pay out $Y$ claims, where $Y$ is a poisson distribution with parameter $\mu$. Under what conditions on $m$,$n$,$\lambda$ and $\mu$ will the considered merger reduce the company's risk of bankrupcy?[You may use the normal approximation to the Poisson for estimating the risk of bankrupcy.]

## question

$X$ is normally distributed with mean 2 and standard deviation 1. $Y$ is normally distributed with mean 5 and standard deviation 3.
(a) What is the distribution of $X + Y$?
(b) What is the probability that $Y - X > 1$?

## question

(a) What is the joint distribution function of the sum of two fair dice, and the (unsigned) difference [e.g., for the roll $(3, 6)$ or $(6, 3)$ the difference is 3]?
(b) What is the conditional distribution of the difference, conditional on the sum being 7?

## question

$X$ is a normal random variable with mean 0 and variance 1. $Y$ is given as $Y = X^3 + N(0, 2^2)$. What is the conditional distribution of $X$ given $Y = 2$? [You do not need to calculate the normalisation constant.]

## question

If the point $(X, Y)$ is uniformly distributed inside the set $\{(x, y)|x^2 + 2y^2 - xy < 5\}$, what is the conditional distribution of $X$ given that $y = 1$.

**question**

Consider the following experiment: Roll a fair (6-sided) die. If the result $n$ is odd, toss $n$ fair coins and count the number of heads. If $n$ is even, roll $n$ fair dice and take the sum of the numbers. What is the expected outcome?

## question

A company has 50 employees, and handles a number of projects. The time a project takes (in days) follows an exponential random variable with parameter $\frac{k}{100}$, where $k$ is the number of employees working on that project. The company receives $m$ projects simultaneously, and has no other projects at that time. What is the expected time until completion of a randomly chosen project if:

(a) The company divides its workers equally between the $m$ projects, and starts all projects at once? [Ignore any requirements that the number of employees working on a given project should be an integer.]

(b) The company orders the projects at random, and assigns all its workers to project 1, then when project 1 is finished, assigns all its workers to project 2, and so on?

## question

A government is trying to demonstrate that speed cameras prevent accidents. It conducts the following study:

(i) Pick 10 locations.
(ii) Count the number of accidents at each location in the past 3 years.
(iii) Choose the location $x$ with the most accidents, and install a speed camera there.
(iv) Count the number of accidents at location $x$ in the following 3 years.
(v) The number of accidents "prevented" is the number of accidents at location $x$ during the 3 years before installing the speed camera, minus the number of accidents at location $x$ in the 3 years after installing the speed camera.

Assume that the number of accidents at each site is an independent Poisson random variable with parameter 0.5, and that installing a speed camera at a given location has no effect on the number of

## question

In a particular area in a city, the roads follow a rectangular grid pattern. You want to get from a location $A$ to a location $B$ that is $m$ blocks East and $n$ blocks North of $A$.

(a) How many shortest paths are there from $A$ to $B$?

(b) After a storm, each road segment independently has probability $p$ of being flooded. What is the expected number of shortest paths from $A$ to $B$ that are still useable?

(d) [Homework sheet 8] Use Markov's inequality to find a bound on the probability that there is still a useable shortest path from $A$ to $B$.

## question

A random graph is generated on $n$ vertices where each edge has probability $p$ of being in the graph. What is
(a) The expected number of triangles?
(b) The variance of the number of triangles? [Hint: there are two types of pairs of triangles to consider — those that share an edge, and those that don't.]

## question

An ecologist is studying diversity of butterfly population in a region. She wants to identify the species of butterfly that are common in the population. She plans to collect a large sample of butterflys, and identify all the species of butterfly present in the sample. She wants to collect enough species that 90% of the butterfly population belongs to species she has collected. That is, supposed that a randomly collected butterfly is of type $i$ with probability $p_i$. She wants to collect types $i_1, i_2, \ldots, i_m$ such that $\sum_{j=1}^{m} p_{i_j} \geqslant 0.90$. She wants to determine how large a sample she needs to collect so that her probability of achieving this is at least 95%.

Let $X$ be the proportion of the population covered by species she has sampled. That is, let $X = \sum_{j=1}^{m} p_{i_j}$. Suppose that she collects a sample of size $n$, and that there are a total of $N$ species of butterfly, with probability $p_1, \ldots, p_N$.

(a) Find $E(X)$.

(b) Find $Var(X)$.

(c) [Sheet 8] Use Chebyshev's inequality to obtain a lower bound on the probability $X \geq .90$ [For suitably large $n$.]

## question

Show that $\mathrm{Cov}(X, Y) \leqslant \sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}$. [Hint: consider $\mathrm{Var}(\lambda X - Y)$ for suitable $\lambda$.]

### question

An outbreak of a contagious disease starts with a single infected person. Each infected person $i$ infects $N_i$ other people, where each $N_i$ is a random variable with expectation $\mu < 1$. If the $N_i$ are all independent and identically distributed, what is the expectation of the total number of people infected?

## question

If $X$ is normally distributed with mean $\mu$ and variance $\sigma^2$, and $Y$ is exponentially distributed with parameter $\lambda$:
(a) Calculate the moment generating function of $X - Y$. [You may use the formulae for moment generating functions of $X$ and $Y$ from the book.]
(b) [Homework Sheet 8] Use the Chernoff bound with $t = \frac{-\mu}{\sigma^2}$ to obtain a bound on the probability that $X$ is more than $Y$.
(c) [Homework Sheet 8] If $\mu = 1$, what variance $\sigma^2$ gives the largest bound on the probability that $X > Y$?

### question

Use Markov's inequality to obtain a bound on the probability that the product of two independent exponentially distributed random variables with parameters $\lambda$ and $\mu$ is at least 1.

## question

[Sheet 7 Q.7] In a particular area in a city, the roads follow a rectangular grid pattern. You want to get from a location $A$ to a location $B$ that is $m$ blocks East and $n$ blocks North of $A$.

(d) Use Markov's inequality to find a bound on the probability that there is still a useable shortest path from $A$ to $B$.

## question

[Sheet 7 Q.6] An ecologist is studying diversity of the butterfly population in a region. She wants to identify the species of butterfly that are common in the population. She plans to collect a large sample of butterflys, and identify all the species of butterfly present in the sample. She wants to collect enough species that 90% of the butterfly population belongs to species she has collected. That is, supposed that a randomly collected butterfly is of type $i$ with probability $p_i$. She wants to collect types $i_1, i_2, \ldots, i_m$ such that $\sum_{j=1}^{m} p_{i_j} \geqslant 0.90$. She wants to determine how large a sample she needs to collect so that her probability of achieving this is at least 95%.

Let $X$ be the proportion of the population covered by species she has sampled. That is, let $X = \sum_{j=1}^{m} p_{i_j}$. Suppose that she collects a sample of size $n$, and that there are a total of $N$ species of butterfly, with probability $p_1, \ldots, p_N$.

(c) [Sheet 8] Use Chebyshev's inequality to obtain a lower bound on the probability $X \geqslant .90$ [For suitably large $n$.]

## question

A car insurance company is trying to decide what premium to charge. It estimates that it can attract 20,000 customers. Each customer has a probability $\frac{1}{100}$ of making a claim for \$3,000. The company has to cover \$1,000,000 in costs. What premium should it set in order to have a probability 0.00001 of being unable to cover its costs?

**question**

A gambler is playing roulette in a casino. The gambler continues to bet \$1 on the number where the ball will land. There are 37 numbers. If the gambler guesses correctly, the casino pays him \$35 (and he keeps his \$1). Otherwise, he loses his \$1. If he keeps playing, how long is it before the probability that he has more money than he started with is less than 0.0001?

## question

We take $n$ random samples from a uniform distribution on the interval $[0, 1]$.
(a) What will the approximate distribution of the mean of these samples be for large $n$?
(b) What is the probability density of the limiting distribution of the mean at 0.95 [This is a rough approximation for the probability that the mean is at least 0.95 using the CLT approximation]?
(c) What is the probability that all $n$ samples are larger than 0.95?
(d) How do the answers to (b) and (c) compare for large $n$?

## question

[Sheet 7 Q.1] If $X$ is normally distributed with mean $\mu$ and variance $\sigma^2$, and $Y$ is exponentially distributed with parameter $\lambda$:
(b) Use the Chernoff bound with $t = \frac{-\mu}{\sigma^2}$ to obtain a bound on the probability that $X$ is more than $Y$.
(c) If $\mu = 1$, what variance $\sigma^2$ gives the largest bound on the probability that $X > Y$?

## question

You are investing your money in the stock market. In year $i$, the value of your investment increases by $X_i\%$ where the $X_i$ are i.i.d. random variables with mean 2. That is, if the value of your investment at the start of year $i$ is $I_i$, then the value at the end of year $i$ is given by $I_{i+1} = I_i\left(1 + \frac{X_i}{100}\right)$. After $n$ years, your *rate of return* is given by $r_n = \left(\left(\frac{I_n}{I_0}\right)^{\frac{1}{n}} - 1\right) \times 100\%$ where $I_n$ is the value of your investment after $n$ years (so $I_0$ is your initial investment). [This $r$ gives the annual percentage interest that would result in a total $I_n$ after $n$ years, starting from $I_0$.]

(a) What can you say about the behaviour of $r_n$ as $n \to \infty$? [Hint: think about $Y_i = \log\left(1 + \frac{X_i}{100}\right)$, and $J_n = \log\left(\frac{I_n}{I_0}\right)$.]

...

## question (continued)

(b) A friend tells you that if the amount you have after $i$ years is $I_i$, then the expected amount that you have after $i+1$ years is $1.02 \times I_i$. Therefore, since the amount by which the value of your investment changes each year is independent, after $n$ years, the expected value of your investment will be $(1.02)^n \times I_0$. Explain the relationship between this equation and your answer to (a).