

A page of formulas, and a Normal probability table are provided.

Explain/justify your answers.

This exam has five questions: it continues on the back.

1. (10 points) A disease is determined by a genetic factor G . Disease onset can be prevented, but the preventive treatment is expensive. Only 1% of people in the population carry the genetic factor G that will cause the disease.

Doctors propose to test everyone for the disease factor G . To make this possible, they must use a cheap and easy test, which unfortunately has a 5% error rate:

$$P(\text{test } G\text{-positive} \mid G) = P(\text{test } G\text{-negative} \mid G^c) = 0.95.$$

where G^c is the event that the person does **not** carry G .

(a) (3 points) Show that the probability a random person from the population will test positive for G is 5.9%.

(b) (4 points) Given a person tests positive for G , show the probability he/she actually carries the disease factor G is 0.161, while the probability he/she does not (G^c) is 0.839.

(c) (3 points) Given a person tests negative for G , show the probability he/she actually carries the disease factor G is 0.00053.

2. (12 points: 4 each part) This question continues question 1: use the results given to you in question 1.

In fact, the disease is only serious for a person who also carries another genetic factor H . For people who carry G but not H treatment is cheap and effective. Only people who carry both G and H need the expensive preventive treatment.

Scientists discover that people who carry G also carry H with probability $3/4$ ($P(H \mid G) = 0.75$),

but that people who do not carry G carry H only with probability $1/4$ ($P(H \mid G^c) = 0.25$).

(Note these probabilities are for the true G or G^c state; not for the result of the cheap test.)

The test for H is expensive, but has no error.

Doctors propose therefore to test only those who tested positive for G for the severity factor H , and to give the preventive treatment to those who test positive for both tests.

(a) Given that a person tested positive for G , show that the probability he/she also tests positive for H is about $1/3$.

(b) What is the probability that a random person from the population ends up getting the expensive preventive treatment **unnecessarily**? (i.e. They tested positive for both tests, but do not have G .)

(c) What is the probability that a random person from the population will develop the severe form of the disease, because they were not given the preventive treatment?

(i.e. They have G and H , but were not tested for H .)

3. (10 points) Widgets coming off the production line are faulty with probability $p = 0.1$. All widgets are independent. Each set of ten widgets from the production line is immediately assembled into a gadget. Unfortunately it is not possible to test a gadget without destroying it, but it is known that a gadget will work (is *good*) so long as it has at most 1 faulty widget. It is possible to test each of the ten widgets in a gadget.

(a) (2 points) Show that the probability an assembled gadget will **not** work (i.e. is *bad*) is 0.264.

(b) (4 points) One option is to test the widgets one by one. The gadget will be discarded as soon as two faulty widgets are found. In this way, a gadget is sold if and only if it is *good*. Note that, if none of the first 9 widgets is faulty, it will not be necessary to test the tenth one. What is the probability a gadget is discarded after testing exactly 9 widgets? Given that exactly 9 widgets are tested, what is the probability the gadget is sold and not discarded?

(c) (4 points) Another option is to test only the first five widgets in a gadget. If none of these five are faulty, the gadget will be sold. If one or more is faulty, the gadget will be discarded. Under this scheme, given a gadget is *bad*, what is the probability it is sold? Also, given a gadget is *good*, what is the probability it is discarded?

(Hint for (c): Note that in 5 widgets, the probabilities of 0, 1, or ≥ 2 faulty are

$$0.9^5 = .59, 5 \times 0.1 \times 0.9^4 = 0.33, \text{ and } 1 - .59 - 0.33 = 0.08 \text{ respectively.})$$

4. (10 points) Every week 40,000 adults in the small city of Zgum take part in a national lottery. For each individual, each week, the probability of a win is 1 in 20,000 ($\frac{1}{2} \times 10^{-4}$), and wins are independent over individuals and over weeks.

(a) (3 points) Use the Poisson approximation to show that, in any one week, the probability that more than two citizens of Zgum (Zgumians) win is approximately 0.323.

(b) (3 points) Over the 4 years 2001-2004 (208 weeks), what are the mean and variance of the number of weeks in which more than two Zgumians win?

(c) (4 points) Use the Normal approximation to find the probability that the number of weeks in 2001-2004 in which more than two Zgumians win is at least 60.

5. (12 points: 3 each part) The Statistics Department of a major University has a new Copier machine, which has two identical components which unfortunately fail repeatedly. The components fail independently, and each has a time to failure which is an exponential random variable with rate parameter 1 per week.

When both components are working, the machine can be used for email, fax etc., but when only one component is working it will function as a traditional copier only. To save money, the department only has the machine repaired when both components have failed.

(a) Preliminary hint: show that if random variable T has the exponential probability density function $f_T(t) = \lambda \exp(-\lambda t)$ on $0 < t < \infty$, then $P(T > t) = \exp(-\lambda t)$.

(b) Show that the probability that both components are still working after t weeks is $\exp(-2t)$, and hence that the time to failure of either component is an exponential random variable with rate parameter 2 per week.

(c) Suppose one component fails after s weeks. Show that the probability that the other component has still not failed t weeks later is $\exp(-t)$, and hence that the machine continues to function as a regular copier for an additional time which is an exponential random variable with rate parameter 1.

(d) Given the machine is functioning as a copier, the probability it can also be used for email, fax etc. is $1/3$. Explain why this statement is true.