**UNIVERSITY OF TORONTO**
**FACULTY OF APPLIED SCIENCE AND ENGINEERING**

**Second Year — Engineering Science**

**STA286H1 S — Probability and Statistics**

**Term Test**

**Monday, March 27, 2017**

**POSSIBLE SOLUTIONS**

Note: far more significant digits are printed here than would have been reasonable or necessary.

Also, the question numbering and sub-lettering is strange due to a document rendering error that produced the original test itself.

### The Story

A gas distribution company knows that a gas meter will have a serious and dangerous defect with probability $p = 0.01$. It decides to send work crews out to inspect all gas meters to try to find the defective ones. A work crew is able to inspect 60 meters in one 8 hour shift. You are the supervisor all the work crews.

1. **(20 marks total)** Assume the defect status of one gas meter does not affect the probability that any other gas meter is defective.

Page 2 (a) **(7 marks)** What is the probability that a work crew will find at least 1 defective gas meter during one shift?

Denote by $X$ the number of defective meters found in one shift.

$$P(X \geq 1) = 1 - P(X = 0) = 1 - 0.99^{60} = 0.4528434$$

Page 2 (b) **(3 marks)** Work Crew Alpha inspects 100 gas meters without finding a single one that is defective. They phone you to complain that the whole project is a waste of time and that they should have found at least one defective meter by now. What can you say to motivate them to continue with the project?

Many possibilities, including, for example:

- the expected number of inspections before finding the first defective is $1/0.01 = 100$, so it is not surprising they haven't found one yet.
- the probability of not finding one in 100 inspections is $0.99^{100} = 0.3660323$.
- the defect is "serious and dangerous", so it is worth the effort to find defective meters even though they are rare.

Page 3 (a) **(3 marks)** Work Crew Alpha, the same work crew as in (b), goes two full shifts (120 gas meters) and still haven't found one that is defective. As they go out for their third shift, they tell you "We've done two shifts without finding one defective meter. Surely today our luck will change!" Is Work Crew Alpha correct in their assessment of the situation?

No. The number of inspections until the first defect found is Geometric, which is memoryless. There is no reason to think their "luck" will change.

Page 3 (b) **(7 marks)** You decide to motivate the work crews by offering a prize to the first work crew to find a defective gas meter in 10 of its shifts. Work Crew Beta wins the prize after its 12th shift. In response, Work Crew Alpha files a union grievance claiming that there's less than a 1 in 1000 chance that Work Crew Beta could have won the prize so quickly. Assess Work Crew Alpha's complaint.

Denote by $X$ the number of shifts it takes to have 10 shifts in which a defective gas meter is found. The probability of finding one or more defects being found is (from (a)) $p_s = 0.4528434$. So $X \sim \text{NegBin}(10, p_s)$.

There are a few ways to evaluate the complaint. The best, in principle, is probably to calculate $P(X \leq 12)$, which is:

$$
\begin{aligned}
P(X \leq 12) &= \sum_{k=10}^{12} \binom{11}{k} (1-p_s)^{11-k} p_s^k \\
&= 0.00036 + 0.00198 + 0.00597
\end{aligned}
$$

which is far above "1 in 1000". It would have been sufficient, in this case, just to calculate $P(X = 12)$, say.

**More commentary:** The question was unambiguously a negative binomial distribution question based on shifts. The language used in the question had all the hallmarks of "negaitve binomial". There is no "other" interpretation. If your answer was not somehow based on a $\text{NegBin}(10, p_s)$ probability calculation, you were wrong. In particular, I am not aware of any inspection-based calculation that can be correct. The "unit" here is shift, not inspection. **In other words, don't try to argue this with me.**

2. **(20 marks total)** Continuing with the same scenario, since a work crew can do 60 inspections in an 8 hour shift, it inspections occur at a rate of one per 8 minutes on average. This means that a work crew finds defective gas meters at a rate of 0.00125 per minute.

Page 4 (a) **(7 marks)** Calculate the probability that the 1st defective meter is found within the first 200 minutes.

Denote by $X$ the time the first defective meter is found. $X \sim \text{Exp}(\lambda)$ with $\lambda = 0.00125$.

$$
\int_0^{200} \lambda e^{-\lambda x}\, dx = 1 - e^{-\lambda 200} = 0.2211992
$$

Page 4 (b) **(3 marks)** By the time 200 minutes have elapsed, a work crew has found only 1 defective meter. What is the probability that the defective meter was just discovered in the previous 15 minutes?

Denote by $U$ the time when the defective unit was found, given that 1 had been found by the time 200 minutes had elapsed. $U \sim \text{Uniform}[0, 200]$. So $P(185 < U < 200) = 15/200 = 0.075$.

Page 5 (a) **(7 marks)** Calculate the probability that 3 defective meters will be found within 500 minutes.

Denote by $Y$ the amount of time before the 3rd defective meter is found. Then $Y \sim \text{Gamma}(3, \lambda)$. The easiest way to calculate this probability, however, is to use the Poisson process $N(t)$ directly, as follows:

$$
\begin{aligned}
P(Y < 500) &= 1 - P(N(500) \leq 2) \\
&= 1 - \sum_{k=0}^{2} \frac{(500\lambda)^k e^{-(500\lambda)}}{k!} \\
&= 1 - 0.9743431
\end{aligned}
$$

It is also possible to suffer through two integrations-by-parts to get the same answer directly using the Gamma density.

Apparently calculators these days can do numerical integration. Some people just did this. On the one hand, I suppose I cannot tease you incessantly about integration not being all about anti-derivatives and really being a numerical problem, and then get upset when people get a numerical solution. On the other hand, seriously? Are you kidding me? Anyway, it was mostly about identifying the Gamma model. Now I know what calculators can do and will govern myself accordingly.

Some people may have gone directly to the Poisson process approach (essentially reading "3" as "3 or more", which is fine), skipping the Gamma model entirely, which is also fine.

Page 5 (b) **(3 marks)** 225 minutes have elapsed since the work crew last found a defective meter. What is the expected amount of time remaining until they find their next defective meter?

The inter-arrival time of defective meters has an $\text{Exp}(\lambda)$ distribution, which is memoryless. So given 225 minutes have elapsed since the previous defect, the expected time to the next is still $1/\lambda = 800$.

3. **(10 marks total)** Still continuing with the same scenario. Six months have elapsed. Work Crew Alpha has completed 7000 gas meter inspections. Work Crew Beta has independently completed 7800 gas meter inspections. (The difference is because Alpha had to spend time on other projects from time to time.)

Give an approximation to the probability that Work Crew Alpha discovered more defective gas meters than Work Crew Beta did.

(Suggested setup: denote by X and Y the numbers of defective meters discovered and find an approximation to $P(X > Y) = P(X - Y > 0)$.)

Let $X \sim \text{Binomial}(7000, 0.01)$ and $Y \sim \text{Binomial}(7800, 0.01)$.

Both have approximate normal distributions, so $X - Y$ will also have an approximate normal distribution with mean

$$E(X - Y) = (7000)(0.01) - (7800)(0.01) = -8$$

and variance

$$\text{Var}\,(X - Y) = \text{Var}\,(X) + \text{Var}\,(Y) = (7000)(0.01)(0.99) + (7800)(0.01)(0.99) = 146.52$$

So the approximate probability will be (with $Z \sim N(0, 1)$):

$$P(X - Y > 0) \approx P(Z > (0 - (-8))/\sqrt{146.52}) = P(Z > 0.6609088) = 0.2543354$$

4. **(10 marks total)** Suppose $X_1, X_2$, and $X_3$ are random variables, $r_1, r_2$, and $r_3$ are positive integers, and $0 < p < 1$. Also, the random variables have these properties:

- they are independent.
- $X_1 \sim \text{NegBin}(r_1, p)$ and $X_2 \sim \text{NegBin}(r_2, p)$
- $X_1 + X_2 + X_3 \sim \text{NegBin}(r_1 + r_2 + r_3, p)$

What is the distribution of $X_3$? Prove your result.

If you cannot give a rigourous mathematical proof, you may give a casual proof. (Please do not do both.)

Use moment generating functions.

$$\begin{aligned} M_{X_1+X_2+X_3}(t) &= M_{X_1}(t)M_{X_2}(t)M_{X_3}(t) \\ \left(\frac{pe^t}{q+pe^t}\right)^{r_1+r_2+r_3} &= \left(\frac{pe^t}{q+pe^t}\right)^{r_1}\left(\frac{pe^t}{q+pe^t}\right)^{r_2} M_{X_3}(t) \end{aligned}$$

So $M_{X_3}(t) = \left(\frac{pe^t}{q+pe^t}\right)^{r_3}$. Therefore $X_3 \sim \text{NegBin}(r_3, p)$.

A casual proof would involve describing an independent sequence of Bernoulli($p$) trials with at least $r_1+r_2+r_3$ trials, and arguing the difference $X_3$ between $X_1 + X_2 + X_3$ and $X_1 + X_2$ is just the number of trials it took to go from $r_1 + r_2 + 1$ to $r_1 + r_2 + r_3$ successes. Due to independence of trials, this is the same as waiting for $r_3$ successes no matter what happened before. So $X_3 \sim \text{NegBin}(r_3, p)$.