Question 1. Answer the following short questions.

(a) What is the following expression equal to? Use definitions and results that we have studied during the quarter to answer. You do not have to prove those results we have found earlier, just use them.

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\theta}} e^{tx - \frac{1}{2\theta}(x-\gamma)^2} dx$$

The Pdf of a normal distribution is

$f(x) = \frac{1}{\sqrt{2\pi}\,6} e^{-(x-\mu)^2/26^2}$

let $\gamma = \mu$, $\theta = 6^2$, $f(x)$ becomes $\frac{1}{\sqrt{2\pi\theta}} e^{-\frac{1}{2\theta}(x-\gamma)^2}$

since MGF is defined as $E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx$ (since normal distribution is defined from $-\infty$ to ∞)

$\Rightarrow$ this expression is the MGF of a normal distribution with value $e^{\mu t + \frac{6t^2}{2}}$

(b) A student said that the following is a density function. Is there something wrong with the student's claim? Explain.

$$f(x) = \frac{1}{6}x - \frac{11}{15}, \qquad 0 \le x \le 10$$

$\frac{1}{6}x - \frac{11}{15} < 0$

$\frac{1}{6}x < \frac{11}{15}$

$x < 4.4$

Hence, when $0 \le x < 4.4$, $f(x)$ has a negative value. Since for a pdf, $f(x)$ must have a non-negative value for any x in the domain, the student's claim of domain is wrong since there will be negative $f(x)$ for some x in the domain

(c) A random variable has the following density function. Find the 36th percentile.

$$f(x) = \frac{2}{25}(5 - x), \qquad 0 \le x \le 5$$

$\int_0^c \frac{2}{25}(5 - x)\,dx = 0.36$

$\int_0^c \frac{2}{5} - \frac{2}{25}x\,dx = 0.36$

$\left[\frac{2}{5}x - \frac{x^2}{25}\right]_0^c = 0.36$

$\frac{2}{5}c - \frac{c^2}{25} - 0.36 = 0$

$c_1 = 9 \quad c_2 = 1$

Since 9 is outside the domain of x, 9 is rejected, $c = 1$. The 36th percentile is when $x = 1$

31

(d) Prove, using the definition of expectation of a function of a continuous random variable that

$$E[(X-a)^2] = Var(X) + (E(X) - a)^2$$

where a is a constant.

$$E[(x-a)^2] = \int_x (x-a)^2 f(x)\,dx$$
$$= \int_x (x^2 - 2ax + a^2) f(x)\,dx$$
$$= \int_x x^2 f(x)\,dx - 2a\int_x x f(x)\,dx + a^2 \int_x f(x)\,dx$$
$$= E(x^2) - 2a E(x) + a^2 \cdot 1$$
Since $E(x^2) = Var(x) + [E(x)]^2$
$$= Var(x) + \{[E(x)]^2 - 2aE(x) + a^2\}$$
$$= Var(x) + [E(x) - a]^2 \quad \text{(proven)}$$

(e) The probability that a randomly chosen person in a town has completed a Bachelor's degree is 0.3. We want to hire someone with a Bachelor's degree to do a desk job. If we have already interviewed 10 candidates and have not found one with a Bachelor's degree, what is the probability that the first candidate with a Bachelor's degree is the 15th one? Show work.

Let x denote the number of candidates interviewed before finding one with a Bachelor's degree

This is a geometric distribution

$$P(x=15 \mid x>10) = \frac{P(x=15 \cap x>10)}{P(x>10)} = \frac{P(x=15)}{P(x>10)} \quad \text{since } 15>10$$
$$= \frac{0.3 \cdot (1-0.3)^{14}}{(1-0.3)^{10}}$$
$$= 0.3 \cdot (1-0.3)^4$$
$$= 0.07203$$

(f) A motel owner has bought 7 television sets from Ultra View TV Corporation. It is known that 26% of Ultra-View's televisions have to be returned for repairs during the first year of operation. The motel owner puts 4 of the 7 televisions in private rooms and 3 in the lobby. If 3 of them have to be returned for repairs during the first year, what is the probability that 2 of the 3 to be repaired are the ones in the lobby? Show work and summarize explicitly what assumptions you are making about this question.

Assuming these televisions have independent probability of being repaired in the first year and they have the same quality

Let X denote the total number of televisions being repaired, L denote the number in the lobby, R denote the number in the private room

$$P(L=2 \mid X=3) = \frac{P(L=2 \cap X=3)}{P(X=3)} = \frac{P(L=2 \cap R=1)}{P(X=3)} = \frac{\binom{3}{2}\binom{4}{1}}{\binom{7}{3}} = 0.342857 \quad \text{(Hypergeometric)}$$

2.7

Question 2. The length of time spent interviewing a candidate for a job in a high tech company, in minutes, is a continuous random variable with the following density function:

$$f(x) = \frac{2}{100}x, \quad 0 \le x \le 10$$

(a) If someone enters the interview room ahead of you, and you are next in line, what is the probability that you will have to wait less then 5 minutes to be interviewed? Show work.

$P(\text{wait less then 5 minutes}) = F(5) = \int_0^5 \frac{2}{100}x \, dx$

$= \left[\frac{x^2}{100}\right]_0^5$

$= \frac{25}{100} - 0$

$= 0.25$

(b) Does this random variable, X, have the memoryless (or Markov) property? Show work.

Memoryless property is $P(X>s+t \mid X>t) = P(X>s)$, s and t are constants in the domain

$P(X>s+t \mid X>t) = \frac{P(X>s+t \cap X>t)}{P(X>t)} = \frac{P(X>s+t)}{P(X>t)}$ since if $X>s+t$, $X>t$ for nonnegative s and t

$\Rightarrow = \frac{\bar{F}(s+t)}{\bar{F}(t)} = \frac{1-\int_0^{s+t}\frac{2}{100}x\,dx}{1-\int_0^t \frac{2}{100}x\,dx}$

$= \frac{1-\left[\frac{x^2}{100}\right]_0^{s+t}}{1-\left[\frac{x^2}{100}\right]_0^t} = \frac{1-\frac{(s+t)^2}{100}}{1-\frac{t^2}{100}} \neq 1-\frac{s^2}{100} = P(X>s)$

Thus, it does not have memoryless property

(c) Consider 4 randomly chosen interviews taking place at this high tech company. What is the probability that 2 or less of those interviews will last less than 5 minutes? Show work.

Let Y denote the number of interviews will take less than 5 minutes
This is a binomial distribution with 4 trials and $p=0.25$

$P(Y \le 2) = P(Y=0) + P(Y=1) + P(Y=2)$

$= \binom{4}{0}(1-0.25)^4 + \binom{4}{1}(1-0.25)^3 0.25 + \binom{4}{2}(1-0.25)^2 0.25^2$

$= 0.75^4 + 4 \times 0.75^3 \times 0.25 + 6 \cdot 0.75^2 \cdot 0.25^2$

$= 0.9492$

less than

(d) What is the probability that 50 of the next 100 interviews will last less than 5 minutes? Show work.

This is a binomial distribution with $np = 100 \cdot 0.25 = 25 > 10$, $n(1-p) = 100 \cdot (1-0.25) = 75 > 10$

Hence we can use a normal distribution to approximate

This is a normal distribution with $\mu = np = 25$, $\sigma^2 = np(1-p) = 25 \cdot 0.75 = 18.75$ $\sigma = 4.33$

Let K denote the number of interviews that last less than 5 minutes in the next 100 interviews

Z is the standard normal variable

$P(K < 50) = P(Z < \frac{50-\mu}{\sigma}) = P(Z < \frac{50-25}{4.33}) = P(Z < 5.77)$

$= 1$ according to table provided

Question 3. The length of time taken by a worker to complete a certain key task in house construction is a gamma random variable, with parameters $\alpha = 50$ and $\lambda = 5$.

(a) The cost C of completing the task is related to the square of the time required for completion by the formula

$$C = 100 + 40X + 3X^2$$

Find the expected cost. Show work.

$E(C) = E(100 + 40X + 3X^2) = \int_x (100 + 40x + 3x^2) f(x) dx$

$= 100 \int_x f(x) dx + 40 \int x f(x) dx + 3 \int_x x^2 f(x) dx$

$= 100 + 40 E(X) + 3E(X^2)$

for a gamma distribution, $\mu = \frac{\alpha}{\lambda} = \frac{50}{5} = 10$, $\sigma^2 = \frac{\alpha}{\lambda^2} = \frac{50}{25} = 2$

Since $E(X^2) = \sigma^2 + \mu^2 = 2 + 10^2 = 102$, $E(C) = 100 + 40 \times 10 + 3 \times 102$

$E(X) = \mu = 10$

$= 806$

(b) What is the probability that the total amount of time spent by 100 workers doing similar key task at 100 similar house constructions is larger than 800? Show work.

Let X_i denote the amount of time spent by worker i doing the key task

$E(X_1 + X_2 + \dots + X_{100}) = 100\mu = 100 \times 10 = 1000$

$\sigma(X_1 + X_2 + \dots + X_{100}) = 100\sigma^2 = 100 \times 2 = 200$, $\sigma = \sqrt{200} = 14.142$

Since 100 is large, we can use normal distribution to approximate by CLT

$P(X_1 + \dots + X_{100} > 800) = 1 - P(X_1 + \dots + X_{100} < 800) = 1 - P(Z < \frac{800-1000}{14.142}) = 1 - P(Z < -14.142)$

$= \cancel{1-0}$ $1 - 0$

$= 1$

/3