1. Let $X_1, X_2, ..., X_n$ be a sequence of independent random variables where $X_i \sim Exp(\theta_i)$

(a) What is the mgf of $S_n = \sum_{i=0}^{n} X_i$? (2)

$$M_{S_n}(t) = \prod_{i=1}^{n} M_{X_i}(t) = \prod_{i=1}^{n} \frac{1}{1-\theta_i t}$$

1pt   1pt

1pt if assumed were i.i.d

(b) From here on assume that $\theta_i = \theta$ so the sequence becomes iid, and $M_{S_n}(t) = (1-\theta t)^{-n}$. What is the distribution (name and parameters) of S_n? (2)

$$S_n \sim Gamma(\alpha = n,\ \theta = \theta)$$

1pt   1pt for parameters

(c) Let $\bar{X}_n = \frac{S_n}{n}$. What is the distribution (name and parameters) of $\bar{X}_n$? (2)

$$M_{\bar{X}_n}(t) = M_{S_n}(t/n) = \left(1 - \frac{\theta t}{n}\right)^{-n}$$ 1pt

$$\bar{X}_n \sim Gamma(\alpha = n,\ \theta = \theta/n)$$ 1pt

(d) What does the LLN imply will happen to the variance of $\bar{X}_n$ as $n \to \infty$? What happens to the variance for this specific distribution of $\bar{X}_n$? (2)

$$LLN \Rightarrow Var(\bar{X}_n) \longrightarrow 0$$ 1pt

$$Var(\bar{X}_n) = \underbrace{n\left(\frac{\theta}{n}\right)^2}_{\text{variance of gamma}} = \frac{\theta^2}{n} \longrightarrow 0$$ 1pt

(e) How would you transform $\bar{X}_n$ to get Z_n so that the CLT applies? Can you write down the mgf of Z_n? (2)

$$CLT \quad \frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{Var(\bar{X}_n)}} = \frac{\bar{X}_n - \theta}{\theta/\sqrt{n}} = \sqrt{n}\,\frac{\bar{X}_n - \theta}{\theta} = Z_n$$

1pt
½ pt if get CLT right

$$Y_n = \bar{X}_n - \theta \quad M_{Y_n}(t) = E(e^{Yt}) = E(e^{\bar{X}_n - \theta}) = e^{\theta} M_{\bar{X}_n}(t)$$

$$M_{Z_n}(t) = M_{Y_n}\left(\frac{\sqrt{n}}{\theta}t\right) = e^{\theta}\left(1 - \frac{\theta\left(\frac{\sqrt{n}}{\theta}t\right)}{n}\right)^{-n} = e^{\theta}\left(1 - \frac{t}{\sqrt{n}}\right)^{-n}$$

1pt
A+ question

(c) When considering a transformation (univariate or bivariate), give three factors to consider when selecting a technique to find the distribution of the transformed random variable. (3)

* is there an inverse?
* is the transformation the same as the cdf?
* if bivariate, distribution f. technique not v useful.
* does cdf have closed form?
* or any other reasonable consideration

3. I'm interested in the weight of Rice squirrels. To collect some data, I have been walking around the inner loop and then whenever I see a squirrel, I race after it, wrestle it to the ground and then weight it with my pocket scales. Let $R_1, R_2, \ldots, R_n$ the be the random variables the represent the weights of squirrels captured using this process.

(a) Would you be comfortable describing this sequence as iid? Why/why not? (2)

1pt No - each squirrel not equally likely to be caught

* weak/fat squirrels slow
* scare squirrels away
* may catch same squirrel twice
* I get tired
* only squirrels on inner loop
* ...

1pt for valid reason

(b) For the rest of the question, assume that the sequence is iid. What can you say about the distribution of W_i? What can you say about the distribution of $\sqrt(n)W_n = \sqrt(n)\sum_{i=0}^{n} W_i/n$? (4)

Can't say much about w_i - know it is positive and within reasonable bounds

1pt for any thing that is true

We know:

CLT $\Rightarrow$ $\sqrt{n}\,\dfrac{W_i - \mu}{\sigma} \sim N(0,1)$

2pt for some invocation of the CLT.

$\Rightarrow$ $W_n \sim N(\mu, \sigma^2/n)$ 1pt

2. Let $A \sim Normal(0, 1)$, and $B = c + dA$, where $-\infty < c < \infty$ and $d > 0$ are real numbers.

(a) Find the distribution of B using the change of variables technique. (5)

$u(x) = c + dx$

Range of B : $\mathbb{R}$ 1pt

$v(x) = \frac{x-c}{d}$ 1pt $\quad \frac{dv}{dx} = 1/d$ 1pt

$$f(a) = \frac{1}{\sqrt{2\pi}} e^{-a^2/2}$$

$$f(b) = \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{b-c}{d}\right)^2/2} \frac{1}{d} = \frac{1}{\sqrt{2\pi d^2}} e^{-\frac{(b-c)^2}{2d^2}}$$ 1pt

$\Rightarrow \quad B \sim Normal(c, d^2)$ 1pt

(b) Find the distribution of B using the mgf. (2)

$M_B(t) = E(e^{B}) = E(e^{(c+dA)t})$ 1pt

$= e^{ct} E(e^{dA})$

$= e^{ct} M_A(dt)$

$= e^{ct} e^{\frac{1}{2}d^2t^2} = e^{ct + \frac{1}{2}d^2}$ 1pt

$\Rightarrow \quad B \sim Normal(c, d^2)$

(c) Estimate that probability that $\bar{X}_n$ is more than two standard deviations away from the population mean $\mu = E(X_i)$. Why is your estimate only approximate? (Hint: if $Var(X_i) = \sigma^2$, what is $Var(\bar{X}_n)$?) (3)

Chebychev 1pt

$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$

$P(|\bar{X}_n - \mu| \geq k\sigma/\sqrt{n}) \leq 1/k^2$

$\Rightarrow P(|\bar{X}_n - \mu| \geq 2\sigma/\sqrt{n}) \leq 1/4$ 1pt

Approximate because Chebychev applies to any dist 1pt

Normal/CLT 1pt

$P(|\bar{X}_n - \mu| \geq 2\sigma/\sqrt{n}) = 0.05$

(d) What other technique could you have used to answer the part (c)? What's the advantage of the technique you chose? (1)

CLT

Independent of actual dist.

Chebychev

Tighter bound, but dist. is only approximation

4. Imagine that I also measure the height of each squirrel, and I find that joint distribution of height in cm (H) and weight in g (W) is bivariate normal.

(a) Approximately what value do you think ρ will be? Why? (2)

$\rho \approx 0.5$ or similar (1pt)

Expect some relationship but not that strong 1pt

(b) If ρ was zero, what would that imply about W and H? (1)

$\rho = 0$ + bivariate normal $\Rightarrow$ independence 1pt

no linear relationship ½pt

Assume that the parameters of the bivariate normal distribution are $mu_w = 700$, $\sigma_w^2 = 400$, $\mu_h = 20$, $\sigma_h^2 = 9$ and $\rho = 0.5$.

(c) Compute $Cov(W, H)$. (2)

$$\rho = \frac{\sigma_{WH}}{\sigma_W \sigma_H} \Rightarrow \sigma_{WH} = \rho_{WH}\, \sigma_H \sigma_W = 0.5\sqrt{400}\sqrt{9} = 30$$

1pt 1pt

(d) How would you prove that the marginal distributions are $W \sim Normal(\mu_w, \sigma_w^2)$ and $H\ Normal(\mu_h, \sigma_h^2)$? (Just outline the steps) (2)

Integrate to get marginals 1pt

+ Compare marginal pdfs to normal pdf 1pt

(e) What's the probability that a randomly selected squirrel weighs more than 740g? (2)

$$P(W > 740) = P\left(Z > \frac{740-700}{20}\right) = P(Z > 2) = 0.025$$

1pt 1pt

(f) How would you compute the probability that a randomly selected squirrel weighs more than 740g **or** is taller than 26cm? (1)

$P(W > 740 \cup H > 26) = 1 - P(W < 740 \cap H < 26)$ 1/2

OR $= P(W > 741) + P(H > 26) - P(W > 741 \cap H > 26)$

+ integration 1/2

Question	Points	Score
1	10	
2	10	
3	10	
4	10	
Total:	40	