UNIVERSITY OF TORONTO SCARBOROUGH

Department of Computer and Mathematical Sciences

STAB57H3F: An Introduction to Statistics

Midterm, Winter 2024 (Date: March 01, 2024)

Duration: 1 hour and 40 minutes

NAME (Please print): ____

Signature: ____

Student ID:

Aids allowed: A calculator (No phone calculators are allowed).

If you do not find an exact table value in the formula sheet, mention in your answer the one you are looking for and continue with the calculation by using the closest one from the formula sheet.

All your work must be presented clearly in order to get credit. Answer alone (even though correct) will only qualify for ZERO credit. Show your answer strictly in the space provided.

There are 12 pages including this page. Please check to see if you have all the pages. Use page 12 for rough work.

Good Luck!

Question-1 [6 points]: Suppose $X_1, X_2, ..., X_n$ are independently drawn from a Poisson(λ) with probability mass function

$$P[X_i = x_i] = \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} \quad \text{where } x_i = 0, 1, 2, ... \text{ and } \lambda > 0$$

Assume λ to be the **unknown parameter**.

a.[3 points] By showing detailed calculation, find a maximum likelihood estimator **(MLE)** of λ. [no need to check the sign of the second derivative]

b.[2 points] Suppose you are told that $E[X_i] = V[X_i] = \lambda$. By showing detailed calculation, find the **MSE** of your estimator that you have found in part(a).

c.[1 points] Name the **approximate distribution** of your estimator along with the associated parameters.

Question-2 [6 points]: Suppose $X_1, X_2, ..., X_n$ are independently drawn from a $Gamma(\alpha, \theta)$ distribution with pdf

$$f(x) = \frac{1}{\theta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\theta}}; \qquad \alpha > 0 \ ; \ \theta > 0$$

a.[3 points] You are told that $E[X] = \alpha\theta$ and $V[X] = E[X^2] - (E[X])^2 = \alpha\theta^2$.

Find the **method of moments estimates** of α and θ.

b.[3 points] Suppose $\alpha = 2$ and the pdf of $Gamma(2, \theta)$ is

$$f(x) = \frac{1}{\theta^2}\, x\, e^{-\frac{x}{\theta}}$$

where, θ is the only unknown parameter. Find the maximum likelihood estimate **(MLE)** of θ. [no need to check the second derivative]

Question-3 [7 points]: Suppose, a company is interested in estimating the average amount of time customers spend on their website per visit. They take a random sample of website visits and record the time spent on each visit (in minutes). The numbers are given below.

7.6, 6.0, 9.2, 8.4, 7.6, 7.2, 5.2, 6.1, 6.5, 6.3, 10.0, 8.2, 9.4, 8.3, 9.4

You are told that for this sample, $\sum x_i = 115.4$ and $\sum(x_i - \bar{x})^2 = 29.8$

a.[3 points] Calculate a **90% confidence interval for the true average** time spent by all customers on the website per visit based on this sample data.

b.[1 points] Interpret the confidence interval.

c.[3 points] Suppose you are helping this company in calculating sample size for a larger study. You were told that population standard deviation can be assumed to be 1.5. What should be the **sample size** if they want the width of the 99% confidence interval for the true average to be less than 0.5?

Question-4 [6 points]: A company claims that the average amount of time customers spend on their website per visit is **more** than 8 minutes. To test their claim, they take a random sample of website visits and record the time spent on each visit (in minutes). The numbers are given below.

$$7.6,\ 6.0,\ 9.2,\ 8.4,\ 7.6,\ 7.2,\ 5.2,\ 6.1,\ 6.5,\ 6.3,\ 10.0,\ 8.2,\ 9.4,\ 8.3,\ 9.4$$

You are told that for this sample, $\sum x_i = 115.4$ and $\sum (x_i - \bar{x})^2 = 29.8$

a.[3 points] At 5% level of significance, by stating the null and alternative hypothesis, **test** their hypothesis and write a complete **conclusion**. [*hint:* take $\mu = 8$ as H_0]

b.[3 points] [unrelated to part(a)] Suppose you want to test $H_0 : \mu = 8$ vs. $H_a : \mu = 4$ based on 16 observations at 5% level of significance. The population standard deviation can be assumed to be 1.5. By deriving the rejection rule in terms of $\bar{X}$, **calculate the power** of this test. Leave your final answer in terms of *pnorm()* or the CDF of standard normal [i.e. $\Phi(\ )$]

Question-5 [5 points]: Suppose X is a single observation from a $Gamma(3, \beta)$ density.

It can be shown that $\frac{2X}{\beta}$ follows a χ^2 distribution with $df = 6$.

a.[2 points] Explain briefly why $\frac{2X}{\beta}$ can be used as a **pivotal quantity**.

b.[3 points] Using $\frac{2X}{\beta}$ as a pivotal quantity, **construct a 90% confidence interval (CI)** for β. (Constructing a CI means deriving the expressions of the lower and upper bound).

Formula Sheet

1. Distributions:

Distribution	pdf or pmf	mean	variance	MGF
Bernoulli(θ)	$\theta^x(1-\theta)^{1-x}$	θ	$\theta(1-\theta)$	$(1-\theta)+\theta e^t$
Binomial (m,θ)	$\binom{m}{x}\theta^x(1-\theta)^{m-x}$	$m\theta$	$m\theta(1-\theta)$	$[(1-\theta)+\theta e^t]^m$
Poisson(λ)	$\frac{e^{-\lambda}\lambda^x}{x!}$	λ	λ	$exp[\lambda(e^t-1)]$
Uniform$[a,b]$	$1/(b-a)$	$(a+b)/2$	$(b-a)^2/12$	$\begin{cases}(e^{tb}-e^{ta})/t(b-a) & ,t\neq 0\\ 1 & ,t=0\end{cases}$
Normal (μ,σ^2)	$(2\pi\sigma^2)^{-1/2}exp[-\frac{1}{2\sigma^2}(x-\mu)^2]$	μ	σ^2	$exp[\mu t+\sigma^2t^2/2]$
Exponential(β)	$\beta e^{-\beta x}$	$1/\beta$	$1/\beta^2$	$(1-t/\beta)^{-1}$
Exponential(θ)	$\frac{1}{\theta}e^{-\frac{x}{\theta}}$	θ	θ^2	$(1-t\theta)^{-1}$
Gamma(α,β)	$\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$	α/β	α/β^2	$(1-t/\beta)^{-\alpha}\,,t<\beta$
Beta(a,b)	$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}$	$a/(a+b)$	$\frac{ab}{(a+b)^2(a+b+1)}$	$1+\sum_{k=1}^{\infty}\left(\prod_{r=0}^{k-1}\frac{\alpha+r}{\alpha+\beta+r}\right)\frac{t^k}{k!}$

2. LLN: $\frac{1}{n}\sum_i X_i \xrightarrow{P} E[X_i]$; CLT: $\bar{X}_n \xrightarrow{D} N(\mu,\frac{\sigma^2}{n})$; $\frac{Z}{\sqrt{\chi^2_{(m)}/m}} \sim t_{(m)}$; $\frac{\chi^2_{(m)}/m}{\chi^2_{(n)}/n} \sim F_{(m,n)}$; $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$

3. Method of Moments: $\frac{1}{n}\sum_i X_i^k$ is an estimator of $E[X^k]$

4. Maximum Likelihood Estimation:
   - Log-likelihood, $l(\theta) = \log\prod_i f_\theta(x_i) = \sum_i \log f_\theta(x_i)$
   - Score, $l'(\theta) = \frac{\partial}{\partial\theta}l(\theta)$
   - At mle ($\hat{\theta}$), $l'(\theta) = 0$
   - Fisher Info for single obs, $I(\theta_0) = E[\{\frac{\partial}{\partial\theta}\log f(X|\theta)\big|_{\theta=\theta_0}\}^2] = -E[\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\big|_{\theta=\theta_0}]$
   - Fisher info for n obs, $nI(\theta_0) = E[\{\frac{\partial}{\partial\theta}\log f(X_1,X_2,...,X_n|\theta)\big|_{\theta=\theta_0}\}^2] = -E[\frac{\partial^2}{\partial\theta^2}\log f(X_1,X_2,...,X_n|\theta)\big|_{\theta=\theta_0}]$
   - $l'(\theta_0) \xrightarrow{D} N(0,nI(\theta_0))$; $\quad \frac{1}{n}l''(\theta_0) \xrightarrow{P} -I(\theta_0)$; $\quad \hat{\theta} \xrightarrow{D} N(\theta_0,\frac{1}{nI(\theta_0)})$
   - Plug-in est. of Fisher Info, $nI(\hat{\theta}) = -E[\frac{\partial^2}{\partial\theta^2}\log f(X_1,X_2,...,X_n|\theta)]\big|_{\theta=\hat{\theta}}$

5. $Bias[T] = E[T]-\theta$; $\quad MSE[T] = Var[T]+(Bias[T])^2$; $eff[T_1,T_2] = \frac{V[T_2]}{V[T_1]}$; C-R Inequality: $V[T] \geq \frac{1}{nI(\theta_0)}$

6. Consistency in probability, $T_n \xrightarrow{P} \theta$ as $n\to\infty$

7. Factorization theorem: $T(X_1,X_2,...X_n)$ is said to be sufficient for θ if the joint probability function factors in the form, $f(x_1,x_2,...x_n|\theta) = g[T(x_1,x_2,...x_n),\theta]*h(x_1,x_2,...x_n)$

8. Formula related to confidence intervals:

$\bar{X} \pm z_{\frac{(1+\gamma)}{2}}\frac{\sigma}{\sqrt{n}}$; $\quad \bar{X} \pm t_{\frac{(1+\gamma)}{2}(df=n-1)} * \frac{S}{\sqrt{n}}$; $\quad \left(\frac{(n-1)S^2}{\chi^2_{\frac{1+\gamma}{2}(n-1)}}, \frac{(n-1)S^2}{\chi^2_{\frac{1-\gamma}{2}(n-1)}}\right)$; $\quad \bar{X} \pm z_{\frac{(1+\gamma)}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}$

9. Margin of Error, $ME = z_{\frac{(1+\gamma)}{2}}\frac{\sigma}{\sqrt{n}}$; $\quad width = 2*ME$; $\quad$ for proportion, $ME = z_{\frac{(1+\gamma)}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}$

10. Some quantiles from z, t and χ^2 table:

$z_{0.05} = -1.644854$	$z_{0.975} = 1.959964$	$z_{0.99} = 2.326348$	$z_{0.995} = 2.575829$
$\chi^2_{0.025(29)} = 16.04707$	$\chi^2_{0.05(6)} = 1.635383$	$\chi^2_{0.95(6)} = 12.59159$	$\chi^2_{0.975(29)} = 45.72229$
$t_{0.05(14)} = -1.76131$	$t_{0.95(14)} = 1.76131$	$t_{0.95(29)} = 1.699127$	$t_{0.975(29)} = 2.04523$

This page is for **rough work**. It will not be graded.