August 2022 Name:______________________

COMPREHENSIVE WRITTEN EXAM – STOR654 MATHEMATICAL STATISTICS

Unless otherwise noted, all problem parts have equal weight. In budgeting your time expect that some part will take longer than others. When solving multi-part problems feel free to use results of earlier parts even if you cannot solve them in proving later parts.

1. Let $X_1, X_2, \dots$ be i.i.d. sample from a Pareto(γ), i.e.,

$$f(x|\gamma) = \frac{\gamma}{(1+x)^{\gamma+1}} I_{(0,\infty)}(x), \quad \gamma > 0.$$

In some of the problem parts below you might need to make extra assumptions? If you make extra assumptions, state them clearly.

(a) Does Pareto(γ) form an exponential family?

(b) Find the method of moments (MM) estimator $\hat{\gamma}$ based on the first moment.

(c) Recall that KL divergence between two densities f and g is

$$KL(f\|g) = \int \log\left(\frac{f(x)}{g(x)}\right) f(x)\, dx.$$

Prove $KL(f\|g) \geq 0$ and $KL(f\|g) = 0$ if and only if $f = g$ a.s.

(d) Calculate the $KL(f\|g)$ divergence between f the true Pareto(γ) and g the estimated Pareto($\hat{\gamma}$). (Hint: $\int_0^\infty \frac{\log(1+x)}{(1+x)^{\gamma+1}}\, dx = \gamma^{-2}$.)

(e) Assume now that the data is analyzed using a misspecified model under which the observations $X_1, X_2, \dots$ are assumed to be i.i.d. Exponential(λ), i.e., the data was generated from a Pareto but analyzed as if it was exponential. Find the MLE $\hat{\lambda}$ under the misspecified model.

(f) Calculate the $KL(f\|g)$ divergence between f the true Pareto(γ) and g the estimated misspecified Exponential($\hat{\lambda}$).

(g) Consider the following sample from Pareto(2): $0.41, 0.39, 0.21, 2.57, 0.40$. For this sample compare the quality of the two estimated distributions, Pareto($\hat{\gamma}$) and Exponential($\hat{\lambda}$), using $KL(f\|g)$ divergence between the distribution that generated data (f) and each of the estimated distributions (g).

2. Let $X_1, X_2, \ldots$ be i.i.d. sample from a Log-normal(μ, σ^2) distribution, i.e., the sequence $Y_i = \log X_i,\ i = 1, 2, \ldots$ is i.i.d. $\mathrm{N}(\mu.\sigma^2)$, with $\mu \in \mathbb{R}$ and $\sigma > 0$.

   (a) What is the density and cdf of X_i? Also find the density of $X_{(1)} = \min(X_1, \ldots, X_n)$.

   (b) Find a minimal sufficient statistic. Is it complete?

   (c) Compute the expectation $\eta = EX_i$.

   (d) Find the MLE $\hat{\eta}$ of η.

   (e) Is $\hat{\eta}$ unbiased estimator of η?

   (f) How would you test $\mathcal{H}_0 : \mu = 0$ vs $\mathcal{H}_1 : \mu \neq 0$. What p-value would you use?

   (g) Propose a test for testing $\mathcal{H}_0 : \eta = 1$ vs $\mathcal{H}_1 : \eta \neq 1$.

   (h) Assume that $\sigma^2 = 1$ is known and that the prior on μ is $\mathrm{N}(0, 1)$. What is the posterior distribution of μ based given the observations $X_1, \ldots, X_n$?