1. Let $Y_i = e^{X_i}$ where $X_1, \ldots, X_n \sim$ i.i.d. $N(\mu, \sigma^2)$. Let $\phi = \mathrm{E}[Y_i]$.

   (a) Find the expectation and variance of Y_i, and the expectation and variance of $\bar{Y}$, in terms of μ and σ^2.

   (b) Find the MLE $\hat{\phi}$ of ϕ based on $Y_1, \ldots, Y_n$, and find an approximation to the variance of $\hat{\phi}$. Discuss the magnitude of $\mathrm{Var}[\hat{\phi}]$ relative to $\mathrm{Var}[\bar{Y}]$.

   (*) Perform a simulation study where you compare $\hat{\phi}$ and $\bar{Y}$ in terms of bias, variance and MSE.

2. Let $Y_1, \ldots, Y_n \sim$ i.i.d. gamma(a, b), parameterized so that $\mathrm{E}[Y_i] = a/b$.

   (a) Write down the log-likelihood and obtain the likelihood equations, that is, the equations that the MLE is the solution to.

   (b) Compute the Fisher information and use this to obtain a joint asymptotic distribution for $(\hat{a}_{MLE}, \hat{b}_{MLE})$.

   (c) Let $\mu = a/b$. Obtain the asymptotic distribution of $\hat{\mu}_{MLE} = \hat{a}_{MLE}/\hat{b}_{MLE}$. Compare the approximate variance of $\hat{\mu}_{MLE}$ to that of $\bar{Y}$.

3. Suppose $Y_1, \ldots, Y_n \sim$ i.i.d. gamma(a, b), but the statistician thinks that $Y_1, \ldots, Y_n \sim$ i.i.d. $N(\mu, \sigma^2)$ for some unknown values of μ, σ^2.

   (a) What values of (μ, σ^2) will maximize the expected log likelihood, $\mathrm{E}[\log p(Y|\mu, \sigma^2)]$? Here, the expectation is with respect to the true gamma distribution for Y, and your answer should depend on (a, b).

   (b) Make an argument that $(\hat{\mu}_{MLE}, \hat{\sigma}^2_{MLE})$ converges in probability to something, say what that something is and explain your reasoning.

   (c) What is the standard error of $\hat{\mu}_{MLE}$ for the statistician who assumes normality? In other words, what does this statistician use as an estimate of the standard deviation of $\hat{\mu}_{MLE}$? How does this compare to the standard error of $\hat{\mu}_{MLE}$ for a statistician who correctly assumes the gamma model (as in the previous problem)?

(d) Discuss the consequences of model misspecification in this case.

4. Suppose researcher A obtains an i.i.d. sample $X_1, \ldots, X_n$ from some population that has density known to be in the set $\{f(x|\theta) : \theta \in \Theta\}$. Also suppose researcher B wants to analyze the experiment, but prefers working with $Y_1, \ldots, Y_n$ where $Y_i = g(X_i)$ for some monotonic function g.

   (a) Write out the likelihood for researcher A in terms of f and the X_i's, and write out the likelihood for researcher B in terms of f, g and the Y_i's.

   (b) Relate the MLE of θ and its asymptotic variance obtained by researcher A from the X-likelihood to that obtained by researcher B from the Y-likelihood.

5. Let $Y_1, \ldots, Y_n \sim$ i.i.d. Uniform$(0, \theta)$ for some $\theta > 0$.

   (a) Find an estimator of θ based on $\bar{Y}$, and compute its variance.

   (b) Write out the likelihood function, and find the MLE of θ.

   (c) Using CDFs or otherwise, find the distribution of the MLE.

   (d) Find the expectation, variance and MSE of the MLE and compare to the estimator from (a).

   (*) Can you get an approximate variance for the MLE via the Fisher information?

   (*) Try to find an unbiased estimator of θ based on the MLE. Can you improve upon the MSE?