1. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d. random variables with $\Pr(Y_i = 1|X_i = x_i, \alpha, \beta) = e^{\alpha+\beta x_i}/(1 + e^{\alpha+\beta x_i})$.

   (a) Find a formula for the log-likelihood and the score function, and obtain equations that determine the MLE (i.e., the "likelihood equations"). Try to write things in terms of $p_i = e^{\alpha+\beta x_i}/(1 + e^{\alpha+\beta x_i})$ as much as possible.

   (b) Write out an expression for the expected information and find the asymptotic distribution of $\hat{\theta}_{MLE}$. Describe how the asymptotic variance of the MLE depends on the distribution of the X_i's.

   (c) Write out the observed information, and write out a step of a Newton-Raphson algorithm for finding the zeros of the likelihood equations (and maximizing the log likelihood).

   (*) Implement your optimization in algorithm in `R` and compare to what you get using the `glm` command. Try to reproduce the standard errors from the `summary` table using your formula for the observed information (Hint: The standard errors in the table should be the square roots of n^{-1} times the diagonal of the estimated asymptotic variance).

2. Let $Y_1, \ldots, Y_n \sim$ i.i.d. P_{θ_0} where $P_{\theta_0} \in \mathcal{P} = \{P_\theta : \theta \in \Theta\}$ and P_θ has density $p(y|\theta) = \theta y^{\theta-1}$ for $y \in (0, 1)$.

   (a) Find the expectation of Y in terms of θ, then rewrite the density in terms of $\mu = E[Y]$ (solve for θ as a function of μ, then plug-in to the density).

   (b) Find the variance of the sample mean $\bar{Y}$ in terms of μ.

   (c) Compare the Crameŕ-Rao lower bound for unbiased estimators of $E[Y]$ to the to the variance of $\bar{Y}$.

   (d) Find the MLE of μ. Via simulation or otherwise, compare the variance of the MLE of $E[Y]$ to the variance of $\bar{Y}$ and comment.

3. Information inequalities for biased estimators:

   (a) For a parametric family $\mathcal{P} = \{P_\theta : \theta \in \Theta \subset \mathbb{R}\}$ where the densities of P_θ are differentiable, derive a more general information inequality that provides a lower bound on the variance of any function $t(Y_1, \ldots, Y_n)$, where $Y_1, \ldots, Y_n \sim$ i.i.d. P_θ.

   (b) For the model $Y_1, \ldots, Y_n \sim$ i.i.d. $N(\mu, 1)$, the Bayesian posterior mean estimator $\hat{\mu}$ of μ under the prior $\mu \sim N(0, \tau^2)$ is $\bar{y} \times n/(n + 1/\tau^2)$. Use (a) to obtain a lower bound on the variance of $\hat{\mu}$ and compare to the actual (frequentist) variance, $\text{Var}[\hat{\mu}|\mu]$.

4. For this problem, you may use the following facts: (a) if $Z_1, \ldots, Z_n$ are i.i.d. $N(0, 1)$ then $\sum(Z_i - \bar{Z})^2$ has a χ^2_{n-1} distribution and is independent of $\bar{Z}$, and (b) if $Z \sim N(0, 1)$ and $X^2 \sim \chi^2_\nu$ are independent then $Z/\sqrt{X^2/\nu}$ has a t-distribution with ν degrees of freedom. Now suppose $Y_1, \ldots, Y_n \sim$ i.i.d. $N(\mu, \sigma^2)$. Let $\bar{Y}$ be the sample mean and let $S^2 = \sum(Y_i - \bar{Y})^2/(n - 1)$ be the sample variance.

   (a) What is the distribution of $\sqrt{n}(\bar{Y} - \mu)/\sigma$, and why?

   (b) What is the distribution of $(n - 1)S^2/\sigma^2$, and why?

   (c) What is the distribution of $\sqrt{n}(\bar{Y} - \mu)/S$, and why?

   (d) Construct a level-α test of $H : \mu = \mu_0$, where μ_0 is some number. In other words, find a test statistic (that can depend on the data and μ_0) and an acceptance region so that the probability that the test statistic is inside the acceptance region is $1 - \alpha$ if $\mu = \mu_0$.