Notes on Homework 3

1. **Profile log-likelihood.** Suppose \( Y = (Y_1, \ldots, Y_n) \) is a vector of independent, identically distributed random variables from the density \( f(y; \psi, \lambda) \), where \( \psi \in \mathbb{R} \) is the parameter of interest and \( \lambda \in \mathbb{R} \) is a nuisance parameter. The profile log-likelihood is defined as \( \ell_p(\psi) = \ell(\psi, \hat{\lambda}_\psi) \), where \( \hat{\lambda}_\psi \) is assumed to satisfy the score equation \( \partial \ell(\psi, \lambda)/\partial \lambda = 0 \).

   (a) Show that the estimator of \( \psi \) that satisfies the profile score equation \( \partial \ell_p(\psi)/\partial \psi = 0 \) is the same as the maximum likelihood estimator of \( \psi \).

   (b) Show that the profile information function \( j_p(\psi) = -\partial \ell_p(\psi)/\partial \psi \partial \psi^T \) satisfies

   \[
   \{j_p(\psi)\}^{-1} = j^{\psi \psi}(\psi, \hat{\lambda}_\psi),
   \]

   where \( j^{\psi \psi}(\theta) \) is the \((\psi, \psi)\) block of \( j^{-1}(\theta) \), the inverse of the observed Fisher information from the log-likelihood function \( \ell(\psi, \lambda) \).

   (c) Use Taylor series expansion to show that

   \[
   \hat{\lambda}_\psi - \hat{\lambda} = -j^{-1}_{\lambda \lambda}(\hat{\psi}, \hat{\lambda})j_{\lambda \psi}(\hat{\psi}, \hat{\lambda})(\psi - \hat{\psi}) + O_p(n^{-1}).
   \]

   (d) Expand \( \ell_p(\psi) \) about \( \hat{\psi} \) and use the results of (b) and (c) to show that

   \[
   w_p(\psi) = 2\{\ell_p(\hat{\psi}) - \ell_p(\psi)\} = (\psi - \hat{\psi})^2 j_p(\hat{\psi}) + o_p(1),
   \]

   and hence that the limiting distribution of \( w_p(\psi) \) is \( \chi_1^2 \), under the model.

I don’t think any Taylor series are needed for (a) or (b), just the score equations. In (d) the expansion is about \( \hat{\psi} \), not \( \psi \) as stated in an earlier version.
2. **BNC, Exercise 3.6.** Based on observations $y_1, \ldots, y_n$ independently normally distributed with unknown mean and variance, obtain the profile log-likelihood for $\Pr(Y > a)$, where $a$ is an arbitrary constant, and compare inference based on this with the exact answer from the non-central $t$-distribution.

The “compare inference ... distribution” is rather cryptic. The following will hopefully get you started.

First, if $Z_1 \sim N(\delta, 1)$, and independently $Z_2 \sim \chi^2_f$, then $Z_1/\sqrt{Z_2/f}$ follows a non-central $t$-distribution, with non-centrality parameter $\delta$ and degrees of freedom $n - 1$. This density is available in R, using the `ncp` argument to `pt, dt, qt, rt`.

Let $\hat{\psi} = \Phi((\bar{y} - a)/s)$ be the maximum likelihood estimate of the parameter of interest $\psi = \Phi((\mu - a)/\sigma)$, where $\bar{y} = \Sigma y_i/n, s^2 = \Sigma(y_i - \bar{y})^2/(n - 1)$.\(^1\) Consider finding a value $\psi_L \in \mathbb{R}$, say, for which $\Pr(\hat{\psi} > \psi_L) = 1 - \alpha$; then $\psi_L$ is a lower confidence bound for $\psi$. If we used the Wald statistic to compute this, then the solution is simply $\psi_L = \psi - z_{\alpha}j_p(\psi)^{1/2}$. For the solution based on the non-central $t$, we write

$$
\Pr(\hat{\psi} > \psi_L) = \Pr\{\Phi((\bar{y} - a)/s) > \psi_L\} = \Pr\{(\bar{y} - a)/s > \Phi^{-1}(\psi_L)\} = \Pr\{(\bar{y} - a)/s > Z_L\},
$$

say, and this last equation has an expression in terms of the non-central $t$ distribution, with non-centrality parameter (I think) $\sqrt{n\Phi^{-1}(\psi)}$.

3. **Adapted from BNC, Ex. 2.24.**

(a) Suppose $Y_1, \ldots, Y_n$ are independent, identically distributed as Poisson with mean $\theta$. Show that the conditional distribution of $Y_1, \ldots, Y_n$, given $S = \Sigma Y_i$, is Multinomial$(S, \pi)$ where $\pi = (1/n, \ldots, 1/n)$.

This distribution can in principle be used to assess goodness of fit of the Poisson model, but if $n$ is much bigger than 2 or 3 it will be difficult to determine which directions in the sample space to examine.

\(^1\)Strictly speaking, this is not the m.l.e., because the m.l.e. of $\sigma^2$ has divisor $n - 1$. Let’s ignore that complication for now.
(b) A summary statistic that could be used to see whether data are consistent with the moment properties of the Poisson is \( T = \Sigma (Y_i - \bar{Y})^2 / ((n - 1)\bar{Y}) \). Show that

\[
E(T \mid S = s) = 1, \quad \text{var}(T \mid S = s) = \frac{2(1 - 1/s)}{n - 1},
\]

and thus that, conditionally on \( S = s, (n - 1)sT/(s - 1) \) has the same first two moments as a \( \chi^2_{(n-1)s/(s-1)} \).

The question came up on Friday about a faster way to compute the variance than grinding it through the multinomial. I haven’t tried this, but it might be a little simpler to use the result that the marginal distribution of any component of a multinomial is a binomial, and the joint distribution of any pair of multinomials is a trinomial.

(c) Explore the extension of this to assessing goodness of fit for a Poisson regression, where \( y_i \sim \text{Po}(\theta_i) \), and \( \log \theta_i = \alpha + \beta x_i \).

4. SM, Problem 4.9.1. The logistic density is a location-scale family with density function

\[
f(y; \mu, \sigma) = \frac{\exp\{(y - \mu)/\sigma\}}{\sigma[1 + \exp\{(y - \mu)/\sigma\}]}, \quad -\infty < y < \infty, -\infty < \mu < \infty, \sigma > 0.
\]

(a) When \( \sigma = 1 \), show that the expected Fisher information about \( \mu \) in \( y \) is 1/3.

(b) If instead of observing \( y \), we observe \( z = 1 \) if \( y > 0 \), otherwise \( z = 0 \). When \( \sigma = 1 \) show that the maximum expected Fisher information about \( \mu \) in \( z \) is 1/4, achieved at \( \mu = 0 \), so that the maximum relative efficiency is 3/4.

Corrected from earlier statement.

5. Saddlepoint approximation. Suppose \( X_1, \ldots, X_n \) are independent and identically distributed on \( \mathbb{R} \), with density function \( f(x) \) and moment generating function \( M_X(t) = E\{e^{tX}\} \) assumed to exist for \( t \) in an open interval about 0, and cumulant generating function \( K_X(t) = \log M_X(t) \). The saddlepoint approximation to the density of \( \bar{X} = n^{-1}\Sigma X_i \) is given by

\[
f_X(\bar{x}) \doteq \frac{1}{\sqrt{2\pi}} \left( \frac{n}{K_X''(\hat{\phi})} \right)^{1/2} \exp\{nK_X(\hat{\phi}) - n\hat{\phi}\bar{x}\},
\]
where \( \hat{\phi} = \hat{\phi}(\bar{x}) \) satisfies the equation \( K'_X(\hat{\phi}) = \bar{x} \).

(a) Show that if \( Y_1, \ldots, Y_n \) are independent and identically distributed from a scalar parameter exponential family

\[
f(y; \theta) = \exp\{\theta y - c(\theta) - d(y)\}
\]

that the saddlepoint approximation to the density of \( \hat{\theta} \) is given by

\[
f_{\hat{\theta}}(\hat{\theta}; \theta) \doteq \frac{1}{\sqrt{2\pi}} j^{1/2}(\hat{\theta}) \exp\{\ell(\theta) - \ell(\hat{\theta})\}.
\]

(b) If \( y_1, \ldots, y_n \) are independent and identically distributed from a scalar parameter location family

\[
f(y; \theta) = f_0(y - \theta),
\]

then we showed in class that the exact density of the maximum likelihood estimator \( \hat{\theta} \), given \( a \), where \( a_i = y_i - \hat{\theta}, i = 1, \ldots, n \), is

\[
f_{\hat{\theta} | A}(\hat{\theta} | a; \theta) = \frac{\exp\{\ell(\theta; y)\}}{\int \exp\{\ell(\theta; y)\} d\theta},
\]

where in the right hand side we recall that \( y_i = \hat{\theta} + a_i \). By expanding \( \ell(\theta) \) in the denominator in a Taylor series about \( \hat{\theta} \), show that the exact conditional density can be approximated by

\[
f_{\hat{\theta} | A}(\hat{\theta} | a; \theta) \doteq \frac{1}{\sqrt{2\pi}} j^{1/2}(\hat{\theta}) \exp\{\ell(\theta) - \ell(\hat{\theta})\}.
\]

Both these approximations have similar versions for \( p \)-dimensional parametric models, with slight changes in notation. Both approximations have relative error \( O(n^{-1}) \), and when re-normalized to integrate to 1 have relative error \( O(n^{-3/2}) \).

You are not required to show these last two statements, but bonus marks if you do.