

# Stat 5421 (Geyer) Fall 2020

## Homework Assignment 2

Due Sunday, October 11, 2020, at midnight

Solve each problem. Explain your reasoning. No credit for answers with no explanation.

**2.1.** Agresti, problem 1.10. Calculate both likelihood ratio and Pearson chi-square test statistics and the corresponding  $P$ -values. For the former, use  $0 \cdot \log(0) = 0$ , which makes sense because  $x \log(x) \rightarrow 0$  as  $x \rightarrow 0$ . Should we expect good asymptotic approximation here?

**2.2.** For the Poisson distribution with data  $x$  and mean  $\mu$ , the log likelihood is

$$l(\mu) = x \log(\mu) - \mu,$$

the score (first derivative of log likelihood) is

$$l'(\mu) = \frac{x}{\mu} - 1,$$

the MLE is

$$\hat{\mu} = x,$$

observed Fisher information is

$$J(\mu) = \frac{x}{\mu^2},$$

and expected Fisher information is

$$I(\mu) = \frac{1}{\mu}$$

(all of these were derived in class but are not in the likelihood handout). There is no  $n$  but we can think of any Poisson as the sum of IID Poissons (because sum of independent Poisson is Poisson). This tells us that the normal approximation for  $x$  is good when  $\mu$  is large and bad when  $\mu$  is small, and similarly for the chi-square approximation for the null distribution of Wald, Wilks, and Rao test statistics.

What are the Wilks, Rao, and Wald two-tailed tests for testing  $H_0 : \mu = \mu_0$  versus  $H_1 : \mu \neq \mu_0$ ? Does it make a difference whether observed or expected Fisher information is used? If it does, give both forms. Apply these to do the hypothesis test for data  $x = 123$  and null hypothesis  $\mu_0 = 100$ . (If observed or expected Fisher information makes a difference, there may be 5 tests. If not, 3 tests. If it makes a difference for one but not the other, 4 tests.)

**2.3.** What are the confidence intervals obtained by inverting each of the tests found in the 2.2? Apply them to the same data as in 2.2. Do 95% confidence intervals.

**2.4.** The negative binomial distribution is one model for overdispersed Poisson. The negative binomial has PMF in its usual parameterization

$$f(x) = \binom{r+x-1}{x} p^r (1-p)^x, \quad x = 0, 1, \dots \quad (1)$$

where  $r > 0$  and  $0 < p < 1$  are parameters and where

$$\binom{r}{k} = \frac{r \cdot (r-1) \cdot \dots \cdot (r-k+1)}{k!}$$

(this generalizes the usual definition of binomial coefficient so it makes sense for any nonnegative integer  $k$  and any positive real number  $r$ ). The R function that evaluates the negative binomial PMF is `dnbinom`, and the expression that evaluates (1) is

`dnbinom(x, r, p)`

The mean and variance of a random variable with PMF (1) are

$$E(X) = \frac{r(1-p)}{p}$$

$$\text{var}(X) = \frac{r(1-p)}{p^2}$$

Writing  $E(X) = \mu$  and  $\text{var}(X) = \nu$  solving for the other parameters gives

$$p = \frac{\mu}{\nu}$$

$$r = \frac{\mu^2}{\nu - \mu}$$

Since  $r > 0$  we always have  $\nu > \mu$ .

For comparison the Poisson distribution has PMF

$$f(x) = \frac{\mu^x}{x!} e^{-\mu}, \quad x = 0, 1, \dots$$

and mean and variance

$$E(X) = \mu$$

$$\text{var}(X) = \mu$$

Hence the variance of a negative binomial with mean  $\mu$  is always greater than the variance of a Poisson with mean  $\mu$ , so in this sense the negative binomial is a model of overdispersed Poisson. But the connection between the two distributions is stronger than that. It can be shown (it is one of my 5101 homework problems) that, if the conditional distribution of  $X$  given  $Y$  is  $\text{Poisson}(Y)$  and  $Y$  has a gamma distribution, then the marginal distribution of  $X$  is negative binomial. So the negative binomial can arise as a mixture of Poissons with different means. Conversely, if we take negative binomial distributions with parameters  $r$  and  $\mu$  and let  $r$  go to infinity holding  $\mu$  fixed, they converge to a  $\text{Poisson}(\mu)$  distribution.

Thus negative binomial with parameters  $r$  and  $\mu$  is a two-parameter model, and Poisson with parameter  $\mu$  is a one-parameter submodel, but it is a funny submodel. It is nested within negative binomial but not nested as a submanifold. Testing negative binomial (alternative) versus Poisson (null) does not give quite the same asymptotics because this is really a one-tailed test. There is no way to let  $r$  go past Poisson (which we can think of as  $r = \infty$ ) to some sort of underdispersed Poisson. This does not satisfy the “usual regularity conditions” for MLE, Wald, Wilks, Rao. If we use the  $(\mu, \nu)$  parameterization, we might have one-sided partial derivatives with respect to  $\nu$  at  $\nu = \mu$ , but this is not obvious. Nor do the “usual regularity conditions” talk about one-sided derivatives. Thus it is unclear whether Rao tests or likelihood ratio tests make sense.

The Wald test however does make sense, since it requires only the “usual regularity conditions” for the big model. Even the Wald test is a bit strange, however, because it is a one-sided multivariate test. In this model it is impossible to have  $\nu < \mu$ , so we must be testing

$$H_0 : \nu = \mu$$

$$H_1 : \nu > \mu$$

and this means the  $P$ -value is half the “usual”  $P$ -value, which assumes a two-tailed null hypothesis.

- (a) Perform this Wald test on the data

<http://www.stat.umn.edu/geyer/5421/mydata/hw2-4.txt>

For the  $g$  function in the the definition of the Wald test use  $g(\mu, \nu) = \nu - \mu$ . What does the test say about whether the data are Poisson or negative binomial?

- (b) Assuming the likelihood ratio test is OK (even though we are not sure about this), also perform a likelihood ratio test (also one-sided) and interpret that. Compare with the Wald test. Does it look like these may be asymptotically equivalent in this situation?