Theory of Statistics.

Homework III

Updated February 15, 2002. MT

8.1. From the description and statement of Berkson's analysis, the average number of emissions per second was 0.8392. This is taken as the estimate of the Poisson rate parameter λ. The problem gives the observed number of counts in each cell. You need to compute the expected number of counts in each cell, under the assumption that the Poisson model holds. The total count is 12169, and you find the expected counts per cell using

$$12169 \times \pi_k = 12169 \times \frac{0.8392^k e^{-0.8393}}{k!}$$

for $k = 0, 1, 2, 3, 4$. For $k \geq 5$, the last cell, you get the cell probability by complementation:

$$12169 \times (1 - \sum_{k=0}^{4} \pi_k).$$

Obtain the following table:

	O	E
0	5267	5257.6900
1	4436	4412.2530
2	1800	1851.3810
3	534	517.8931
4	111	108.6540
5+	21	21.1283

Pearson's chi-square is calculated as $X^2 = 2.1226$ on 4 degrees of freedom. You can find the p-value from a chi-square table as 0.71322, which means the Poisson model cannot be rejected based on X^2.

8.6 The average number of hops X between flights is estimated from the sample as

$$\bar{X} = 1 \times (48/130) + 2 \times (31/130) + \cdots + 12 \times (1/130) = 2.79.$$

(a) Under the assumption that X follows a geometric distribution with parameter p –

$$\mathbb{P}[X = k] = p(1-p)^{k-1} \quad k = 1, 2, \cdots, \tag{1}$$

both methods of moments and maximum likelihood give the estimator $\hat{p} = 1/\bar{X}$ which leads to the estimate 0.36 for p. (b) You can check that the asymptotic variance of the MLE $\hat{p} = 1/\bar{X}$ is $p^2(1-p)/130$. Plug-in the estimate 0.36 for p to obtain an approximate standard deviation of 0.025. An approximate 95% confidence interval is then $0.36 \pm 1.96 \times 0.025$, where 1.96 is the "benchmark" quantile from the standard Normal. (c) To test the goodness of fit of your geometric model, compute the expected frequencies in each cell under that model using (1) with p set to 0.36. You obtain the following table:

	O	E
1	48	46.5565
2	31	29.8834
3	20	19.1813
4	9	12.3120
5	6	7.9027
6	5	5.0726
7	4	3.2559
8	2	2.0899
9	1	1.3414
10	1	0.8610
11	2	0.5527
12	1	0.3547

You can now do the calculation for Pearson's χ^2 resulting in a value of 6.7186 on $12-1-1 = 10$ degrees of freedom (p-value is 0.7517). However, it is a good rule of thumb (and a good habit) to ensure the cells aren't expected to be empty, as the asymptotic distribution of the Pearson statistic is a poor approximation in that case. If you combine the last 6 groups, say, into a 7+ group with $O = 11$ and $E \approx 9$, then no cell has less than 5 counts, and recomputing the χ^2 gives a value of 1.8017 on $7 - 1 - 1 = 5$ degrees of freedom. Note the p-value is then 0.8758, which confirms a better fit.

8.5 (a) The expectation of X under the geometric model is $\mathbb{E}X = 1/p$. Thus $p = 1/\mathbb{E}X$ and the method of moments sets $\hat{p} = 1/\bar{X}$. (b) For the MLE, the log-likelihood for each observation X is given by

$$\log(p) + (x - 1)\log(1 - p).$$

So the $1/n$-log-likelihood is the sample average of those, that is

$$\log(p) + (\bar{x} - 1)\log(1 - p).$$

Taking derivatives, the first order condition for the maximum is given by

$$0 = \frac{1}{\hat{p}} - (\bar{x} - 1)\frac{1}{1 - \hat{p}},$$

which gives the estimator $\hat{p} = 1/\bar{X}$. The second order condition

$$0 > -\frac{1}{p^2} - (\bar{x} - 1)\frac{1}{(1-p)^2}$$

ensures the solution is indeed a maximum.

8.10 Even though the MLE cannot be obtained in closed form here, which might make you think the MME is preferable, you will see later in the course that the asymptotic variance (dispersion) of the MLE is closest to optimal, as it attains the "gold standard" set by the inverse of the Fisher information for any unbiased estimator. But the MLE is very much model-dependent, which can be a great disadvantage if the model is grossly misspecified.

8.14 (a) The variance of $X \sim f(x \mid \sigma)$ is given by $\text{Var}(X) = \mathbb{E}X^2 = 2\sigma^2$. (The mean is zero here). The MME equates this with the sample second moment:

$$2\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

for an IID sample of size n. (b) For the MLE, the first order condition from maximizing the log-likelihood is

$$0 = -\frac{n}{\hat{\sigma}} + \frac{\sum_{i=1}^n |x_i|}{\hat{\sigma}^2}$$

which gives the estimator $\hat{\sigma} = \frac{1}{n}\sum_{i=1}^n |X_i|$. The second order condition

$$0 > 1 - 2 \times \frac{\sum_{i=1}^n |x_i|}{n\sigma}$$

is indeed satisfied for $\sigma = \hat{\sigma}$. (d) To find a sufficient statistic for σ, note that the likelihood factorizes into a part involving $\sum_{i=1}^n |X_i|$ and σ, call it

$$g(\sum |x_i|, \sigma) = \frac{1}{(2\sigma)^n}\exp\left(-\frac{\sum_{i=1}^n |x_i|}{\sigma}\right),$$

and a part not dependent on σ, namely the trivial part

$$h(x_1, \cdots, x_n) \equiv 1.$$

Thus the statistic $\sum_{i=1}^n |X_i|$ is sufficient.

8.17 (a) When μ is known, the first order condition from maximizing the log-likelihood with respect to σ^2 (NB!) is

$$0 = -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2/\hat{\sigma}^4.$$

The resulting ML estimator (check second order condition) is

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2.$$

(b) When σ^2 is known, the first order condition from maximizing the log-likelihood with respect to μ is

$$0 = \sum_{i=1}^{n}(x_i - \hat{\mu})/\sigma^2$$

which gives $\hat{\mu} = \bar{X}$. (Again, check second order condition!)

8.19 (a) For $X \sim f(x \mid \theta)$, the method of moments equates $\mathbb{E}X = \theta + 1$ with the sample moment $\bar{X}$. Thus the MME of θ is given by $\hat{\theta} = \bar{X} - 1$. (b) For the MLE, notice that the negative exponential is strictly decreasing with a maximum value of 1 at θ. When sample data $X_1, \cdots, X_n$ is obtained from f, and subsequently ordered as $\theta \le X_{(1)} \le \cdots \le X_{(n)}$, the sample point closest to θ is $X_{(1)} = \min(X_1, \cdots, X_n)$. $X_{(1)}$ thus maximizes the likelihood. (c) The domain of definition of f depends on the parameter of interest here. Note that after reordering the sample from smallest to largest, it is enough ("sufficient") to ensure $X_{(1)} \ge \theta$. You can then decompose the likelihood into a parameter-free part, namely $h \equiv e^{-\sum x_i}$ and a part involving the sufficient statistic and the parameter, namely $g \equiv e^{n\theta} 1_{\{x_{(1)} \ge \theta\}}$.

8.54 (a) The joint distribution of $X_1, \cdots, X_n$ is given by

$$f_{X_1,\cdots,X_n}(x_1, \cdots, x_n) = \frac{\lambda^{\sum_i x_i} e^{-n\lambda}}{\prod_i x_i!}.$$

The distribution of $T = \sum_i X_i$ is a Poisson$(n\lambda)$ whose density is given by

$$f_T(t) = \frac{(n\lambda)^t e^{-n\lambda}}{t!}.$$

The conditional distribution of $X_1, \cdots, X_n$ given T is therefore

$$\frac{f_{X_1,\cdots,X_n,T}(x_1, \cdots, x_n, t)}{f_T(t)} = \frac{f_{X_1,\cdots,X_n}(x_1, \cdots, x_n)}{f_T(t)} = \frac{t!}{n^t \times \prod_i x_i!}$$

which does not depend on λ. By definition of sufficiency, this says that T is sufficient. (b) The conditional distribution of $X_1, \cdots, X_n$ given X_1 is

$$\frac{f_{X_1,\cdots,X_n}(x_1, \cdots, x_n)}{f_{X_1}(x_1)} = \frac{\lambda^{\sum_{i \ne 1} x_i} e^{-(n-1)\lambda}}{\prod_{i \ne 1} x_i!}$$

which does depend on λ. Hence X_1 cannot be sufficient. (c) Factorize the likelihood using $g(\sum_i x_i, \lambda) = \lambda^{\sum_i x_i} e^{-n\lambda}$ and $h(x_1, \cdot, x_n) = 1/\prod_i x_i!$.