May 4, 2004

Solutions: Chapter 8 Problems 3, 4, 28, 45, 47

Problem 3. Number of yeast cells on a square X. Test the Poisson model $X \in \text{Pois}(\lambda)$.

Concentration 1. $\bar{X} = 0.6825$, $\overline{X^2} = 1.2775$, $s^2 = 0.8137$, $s = 0.9021$, $s_{\bar{X}} = 0.0451$. Approximate 95% CI for μ: 0.6825 ± 0.0884.

Pearson's chi-square test based on $\hat{\lambda} = 0.6825$:

x	0	1	2	3	4+	Total
Observed	213	128	37	18	4	400
Expected	202.14	137.96	47.08	10.71	2.12	400

Observed test statistic $X^2 = 10.12$, df $= 5 - 1 - 1 = 3$, $P < 0.025$. Reject the model.

Concentration 2. $\bar{X} = 1.3225$, $\overline{X^2} = 3.0325$, $s = 1.1345$, $s_{\bar{X}} = 0.0567$. Approximate 95% CI for μ: 1.3225 ± 0.1112.

Pearson's chi-square test: observed test statistic $X^2 = 3.16$, df $= 4$, $P > 0.10$. Accept the model.

Concentration 3. $\bar{X} = 1.8000$, $s = 1.1408$, $s_{\bar{X}} = 0.0701$. Approximate 95% CI for μ: 1.8000 ± 0.1374.

Pearson's chi-square test: observed test statistic $X^2 = 7.79$, df $= 5$, $P > 0.10$. Accept the model.

Concentration 4. $n = 410$, $\bar{X} = 4.5659$, $s^2 = 4.8820$, $s_{\bar{X}} = 0.1091$. Approximate 95% CI for μ: 4.566 ± 0.214.

Pearson's chi-square test: observed test statistic $X^2 = 13.17$, df $= 10$, $P > 0.10$. Accept the model.

Problem 4. Likelihood function of $X \in \text{Bin}(n,p)$ for a given n and $X = x$ is
$L(p) = \binom{n}{x} p^x (1-p)^{n-x} \propto p^x (1-p)^{n-x}$.

a) Minimize $\ln p^x (1-p)^{n-x}) = x \ln p + (n-x)\ln(1-p)$. Since
$\frac{\partial}{\partial p}(x \ln p + (n-x)\ln(1-p)) = \frac{x}{p} - \frac{n-x}{1-p}$
solve $\frac{x}{p} = \frac{n-x}{1-p}$ to find MLE $\hat{p} = \frac{x}{n}$.

b) Cramer-Rao: if $\tilde{p}$ is unbiased, then $\text{Var}(\tilde{p}) \geq (kI(p))^{-1} = 1/I(p)$ since the sample size is $k = 1$. Here
$I(p) = -\text{E}(\frac{d^2}{dp^2} \ln f(X|p)$ and $f(X|p) = \binom{n}{X} p^X (1-p)^{n-X}$.
Compute
$\ln f(X|p) = const + X \ln p + (n-X)\ln(1-p)$, $\frac{d}{dp} \ln f(X|p) = \frac{X}{p} - \frac{n-X}{1-p}$,
$\frac{d^2}{dp^2} \ln f(X|p) = -\frac{X}{p^2} - \frac{n-X}{(1-p)^2}$, and $I(p) = \text{E}(\frac{X}{p^2} + \frac{n-X}{(1-p)^2}) = \frac{n}{p(1-p)}$.

Thus $\mathrm{Var}(\tilde{p}) \geq 1/I(p) = \frac{p(1-p)}{n} = \mathrm{Var}(\hat{p})$.

c) Plot $L(p) = 252p^5(1-p)^5$.

Problem 28.
a) $\bar{X} = 3.6109$ and $s^2 = 3.4181$. Here $n = 16$ and $s_{\bar{X}} = 0.4622$.

b) and c) Exact CI

	90%	95%	99%
μ	3.61 ± 0.81	3.61 ± 0.98	3.61 ± 1.36
σ^2	(2.05; 7.06)	(1.87; 8.19)	(1.56; 11.15)
σ	(1.43; 2.66)	(1.37; 2.86)	(1.25; 3.34)

d) Find sample size x to halve the CI length:
$t_{15}(\alpha/2) \cdot \frac{s}{\sqrt{16}} = 2 \cdot t_{x-1}(\alpha/2) \cdot \frac{s'}{\sqrt{x}}$ implies $x \approx (2 \cdot 4)^2 = 64$.
Further adjustment for 95% CI:
$t_{15}(\alpha/2) = 2.13$, $t_{x-1}(\alpha/2) \approx 2$, therefore $x \approx (2 \cdot 4 \cdot \frac{2}{2.13})^2 = 56.4$.

Problem 45. Uniform distribution $\mathrm{U}(0, \theta)$:
$f(x) = \frac{1}{\theta}$ for $0 \leq x \leq \theta$, $f(x) = 0$ otherwise.
a) $\mu = \theta/2$, $\tilde{\theta} = 2\bar{X}$, $\mathrm{E}(\tilde{\theta}) = \theta$, $\mathrm{Var}(\tilde{\theta}) = \frac{4\sigma^2}{n} = \frac{\theta^2}{3n}$.

b) $L(\theta) = \frac{1}{\theta^n}$ for $\theta \geq X_{(n)}$, $L(\theta) = 0$ otherwise. MLE $\hat{\theta} = X_{(n)} = \max(X_1, \ldots, X_n)$.

c) Sampling distribution of the MLE $\hat{\theta} = X_{(n)}$:
$\mathrm{P}(X_{(n)} \leq x) = (\frac{x}{\theta})^n$ with pdf $f_{\hat{\theta}} = \frac{n}{\theta^n} \cdot x^{n-1}$, $0 \leq x \leq \theta$
biased $\mathrm{E}(\hat{\theta}) = \frac{n}{n+1}\theta$, $\mathrm{E}(\hat{\theta}^2) = \frac{n}{n+2}\theta^2$, $\mathrm{Var}(\hat{\theta}) = \frac{\theta^2}{(n+1)^2(n+2)}$.
Mean square errors:
$\mathrm{MSE}(\hat{\theta}) = (-\frac{\theta}{n+1})^2 + \frac{\theta^2}{(n+1)^2(n+2)} = \frac{n+3}{n+2} \cdot \frac{\theta^2}{(n+1)^2}$
$\mathrm{MSE}(\tilde{\theta}) = \frac{\theta^2}{3n}$

d) Corrected MLE $\hat{\theta}_c = \frac{n+1}{n} \cdot X_{(n)}$ with $\mathrm{E}(\hat{\theta}_c) = \theta$, $\mathrm{Var}(\hat{\theta}_c) = \frac{\theta^2}{n^2(n+2)}$.

Problem 47. Genetic model: $p_1 = \frac{2+\theta}{4}$, $p_2 = \frac{1-\theta}{4}$, $p_3 = \frac{1-\theta}{4}$, $p_4 = \frac{\theta}{4}$, where $0 < \theta < 1$. In particular, if $\theta = 0.25$, then the genes are unlinked and the genotype frequencies are

	Green	White	Total
Starchy	$\frac{9}{16}$	$\frac{3}{16}$	$\frac{3}{4}$
Sugary	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{1}{4}$
Total	$\frac{3}{4}$	$\frac{1}{4}$	1

a) Sample counts $(X_1, X_2, X_3, X_4) \in \mathrm{Mn}(n, p_1, p_2, p_3, p_4)$ with $n = 3839$.
Likelihood

$L(\theta) = \binom{n}{x_1,x_2,x_3,x_4}(2+\theta)^{x_1}(1-\theta)^{x_2+x_3}\theta^{x_4}4^{-n}$

$\frac{d}{d\theta}\ln L(\theta) = \frac{x_1}{2+\theta} - \frac{x_2+x_3}{1-\theta} + \frac{x_4}{\theta}$.

Solve

$\frac{x_1}{2+\theta} + \frac{x_4}{\theta} = \frac{x_2+x_3}{1-\theta}$ or $\theta^2 n + \theta(2x_2 + 2x_3 + x_4 - x_1) - 2x_4 = 0$

to find MLE $\hat{\theta} = \frac{-u+\sqrt{u^2+8nx_4}}{2n} = 0.0357$, where $u = 2x_2 + 2x_3 + x_4 - x_1$.

Asymptotic variance $\mathrm{Var}(\hat{\theta}) \approx \frac{1}{I(\theta)}$, where $I(\theta) = -\mathrm{E}(\frac{d^2}{d\theta^2}\ln f(X_1, X_2, X_3, X_4|\theta))$.

Since

$\frac{d^2}{d\theta^2}\ln L(\theta) = -\frac{x_1}{(2+\theta)^2} - \frac{x_2+x_3}{(1-\theta)^2} - \frac{x_4}{\theta^2}$

$I(\theta) = \frac{n}{4(2+\theta)} + \frac{2n}{4(1-\theta)} + \frac{n}{4\theta} = \frac{n(1+2\theta)}{2\theta(2+\theta)(1-\theta)}$ and $I(\hat{\theta}) = 29345.8$

we have $s_{\hat{\theta}} = 0.0058$.

b) $0.0357 \pm 1.96 \cdot 0.0058 = 0.0357 \pm 0.0114$

c) Parametric bootstrap using Matlab:

```
p1=0.5089, p2=0.2411, p3=0.2411, p4=0.0089
n=3839; B=1000; b=ones(B,1);
x1=binornd(n,p1,B,1);
x2=binornd(n*b-x1,p2/(1-p1));
x3=binornd(n*b-x1-x2,p3/(1-p1-p2));
x4=n*b-x1-x2-x3;
u=2*x2+2*x3+x4-x1;
t=(-u+sqrt(u.^2+8*n*x4))/(2*n);
std(t)
histfit(t)
```

gives std(t)=0.0058.

d) Two ends of interval covering 95% of the components of the vector t produced by bootstrapping:

```
c1=prctile(t,2.5)
c2=prctile(t,97.5)
```

are c1=0.0250 and c2=0.0473. 95% CI for θ: $(2\hat{\theta}-c_2, 2\hat{\theta}-c_1) = (0.0241, 0.0464)$