8.3 Lecture 10 Monday 02/07/01

Homework and Labs. see the logictics sectionhttp://www.stanford.edu/class/stat200/node1.html

8.4 Last time :Fitting Gamma densities

I did more Monte Carlo studies and with a simulation of size 10,000, I found

```
mean(out1(:,2))
1.79
var(out1(:,2))
    0.1256
var(out1(:,1))
    0.0042
mean(out1(:,1))
0.3958
```

About simulation size and the amount of error in Monte Carlo studies.

8.4.1 Poisson Example -where we don't need the simulation

Practical example:
23 grid square, measurements of abestos counts.

Call $S = \sum X_i$ our estimate is provided by $\hat{\lambda} = \frac{S}{n}$, its distribution is what we call the sampling distribution of thes estimate, what is it?

S is a sum of Poisson independent var. so it is Poisson $(n\lambda_o)$

$$P(\hat{\lambda} = v) = P(S = nv) = \frac{(n\lambda_o)^{(nv)} e^{-n\lambda_o}}{(nv)!}$$

This can be proved used the mgf:

$$e^{\lambda(e^t-1)} e^{\mu(e^t-1)} = e^{(\lambda+\mu)(e^t-1)}$$

because

$$M_{X+Y} = M_X M_Y \text{ when } X \text{ and } Y \text{ are independent}$$

Since S is Poisson, both its mean and variance are $n\lambda_o$, so

$$E(\hat{\lambda}) = \frac{1}{n} E(S) = \lambda_o$$

and

$$Var(\hat{\lambda}) = \frac{1}{n^2} Var(S) = \frac{\lambda_o}{n}$$

So that $\hat{\lambda}$ is an unbiased estimate of λ. The standard devaition of the sampling distribution is always called the standard error of $\hat{\lambda}$ and it is

$$\sigma_{\hat{\lambda}} = \sqrt{\frac{\lambda_0}{n}}$$

as usual this is function of an unknown parameter, we will plug in the best estimate we can:

$$s_{\hat{\lambda}} = \sqrt{\frac{\hat{\lambda}}{n}}$$

The sampling distribution is also know because for $n\lambda_o$ big enough, we have the Poisson tends to Normal, see page 167, it has a typo:

$$Z_n = \frac{X_n - E(X_n)}{\sqrt{Var(X_n)}} = \frac{X_n - \lambda_n}{\sqrt{\lambda_n}}$$

$$logM_{Z_n}(t) = -t\sqrt{\lambda_n} + \lambda_n(e^{\frac{t}{\sqrt{\lambda_n}}} - 1)$$

$$logM_{Z_n}(t) \longrightarrow \frac{t^2}{2}$$

In our example, $n = 23$, and $\lambda \sim 25$, so the parameter for the sum is larger than 500, we are safe with the normal approximation.

8.4.2 An angular distribution

We study the angle θ at which electrons are emitted in muon decay, its cosine has a density ($x = cos\theta$):

$$f(x|\alpha) = \frac{1 + \alpha x}{2}$$

α is the parameter we are interested in, it is related to polarization.

$$\mu = \int_{-1}^{1} x(\frac{1 + \alpha x}{2})dx = \left[\frac{x^2}{4} + \frac{\alpha x^3}{6}\right]_{-1}^{1} = \frac{\alpha}{3}$$

Thus the method of moments tells us to estimate α by $\hat{\alpha} = 3\bar{X}$. Summary of the method:

1. Write the moments as function of the parameters.
2. Write the parameters as a function of the moments.
3. Plug in the moment estimates computedd from the sample in the functions.

8.4.3 Consistency

Under certain reasonable conditions, when the sample size increases the method of moments estimates thus constructed approach the true parameters more and more; the estimates are said to be **consistent**.

Definition:
Let $\hat{\theta}_n$ be an estimate of a parameter θ,(the little n signifies it is based on a sample of size n), the estimate $\hat{\theta}_n$ is said to be consistent in probability if $\hat{\theta}_n$ converges in probability to θ when $n \longrightarrow \infty$;

$$\forall \epsilon \; P(|\hat{\theta}_n - \theta| > \epsilon) \longrightarrow 0 \quad when \quad n \longrightarrow \infty$$

Weak Law of Large Numbers (WLLN,page 164 in Rice, and see pages 49-52, Lehmann) says that the $\hat{\mu}_k$ converge in probability to μ.

$$(P|\hat{\mu_k} - \mu_k| > \epsilon) \longrightarrow 0 \qquad as \qquad n \longrightarrow \infty$$

If the functions relating moments and parameters are continuous functions the estimates converge to the parameters, this justifies the estimation of the standard errors we have been using:

$$\sigma_{\hat{\theta}} = \frac{1}{\sqrt{n}}\sigma(\theta_0) \text{ has been replaced by } s_{\hat{\theta}} = \frac{1}{\sqrt{n}}\sigma(\theta)$$

9 Maximum Likelihood Estimation

$X_1, X_2, X_3, \dots X_n$ have joint density denoted

$$f_\theta(x_1, x_2, \dots, x_n) = f((x_1, x_2, \dots, x_n|\theta)$$

Given observed values $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$, the liklihood of θ is the function

$$lik(\theta) = f(x_1, x_2, \dots, x_n|\theta)$$

considered as a function of θ.

If the distribution is discrete, f will be the frequency distribution function.

In words: $\boxed{\text{lik}(\theta)\text{=probability of observing the given data as a function of } \theta.}$

Definition:
The maximum likelihood estimate (mle) of θ is that value of θ that maximises $lik(\theta)$: it is the value that makes the observed data the "most probable".

If the X_i are iid, then the liklihood simplifies to

$$lik(\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

Rather than maximising this product which can be quite tedious, we often use the fact that the logarithm is an increasing function so it will be equivalent to maximise the log liklihood:

$$l(\theta) = \sum_{i=1}^{n} \log(f(x_i|\theta))$$