STAT 341

Format:
Lectures Monday and first hour Friday
Section Wednesday
Problem session second hour Friday
(except this week)

Midterm on Friday February 7
Final as scheduled Wednesday March 19

Chapter 5-6 in the text

More information on class webpage

Climate change

Climate change [is] changes in long-term averages of daily weather.
NASA: Climate and weather web site

Climate is what you expect; weather is what you get.
Heinlein: Notebooks of Lazarus Long (1978)

Climate is the distribution of weather.
AMSTAT News (June 2010)

Some evidence

Sea level rise

Arctic ice

Global Historical Climatology Network

Global Climate Network Temperature Stations
Global temperature

Estimate temperature everywhere (by solving a stochastic partial differential equation numerically)

Comparison of global averages

Ranking temperatures

“Global temperatures in 2003 were 0.56°C (1.01°F) above the long-term (1880-2003) average, ranking 2003 the second warmest year on record, which tied 2002.” (NOAA 2012)

What is the uncertainty in such a statement?
Statistics vs. probability

A probability model is a fixed, fully determined distribution
\[ P(X=k) = 3^k \frac{e^{-3}}{k!} \]

A statistical model is not fully determined:
\[ P(X=k) = \lambda^k \frac{e^{-\lambda}}{k!}, \lambda > 0 \]

We use data from a sample (often independent observations from this distribution) to estimate \( \lambda \), determine whether \( \lambda < 3 \), etc.

Probability is the science of randomness – regularity in irregularity

Build probability models to describe observed phenomena

Statistics is a system for formal inference based on probability models

Estimate parameters of models
Test hypotheses
Assess uncertainty

Lund, Sweden, annual maximum temperatures

A possible model

The generalized extreme value distribution has cdf
\[ F(x; \mu, \sigma, \xi) = \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^\frac{1}{\xi} \right\} \]

The parameters describe location, spread and shape, respectively. If we knew the values of the parameters we could draw the histogram of the data and overlay the corresponding density function to see if it is a good fit.
Sampling distribution

Since \( \hat{\theta}(X_1, X_2, ..., X_n) \) is a random variable we can figure out its distribution. Since it comes from a sample it is called a sampling distribution. Its cdf is (as usual) \( P(\hat{\theta}(X_1, X_2, ..., X_n) \leq y) \), and we can determine, e.g., its mean

\[
E_{\theta} [\hat{\theta}(X_1, X_2, ..., X_n)] = \int \cdots \int \hat{\theta}(x_1, ..., x_n) f(x_1, ..., x_n; \theta) \, dx_1 \cdots dx_n
\]

Note that it is not a number, but a function of the unknown parameter \( \theta \).

What estimation is all about

In 1918 R. A. Fisher proposed estimating parameters by considering how likely are the data if \( \theta \) is the true parameter? Choosing the parameter that makes the observations most likely is formalized using the likelihood function

\[
L(\theta) = \begin{cases} 
  f_{x_1, ..., x_n}(x_1, ..., x_n; \theta), & \text{continuous case} \\
  p_{x_1, ..., x_n}(x_1, ..., x_n; \theta), & \text{discrete case}
\end{cases}
\]

The data are fixed
The parameter is varying
The method of maximum likelihood

Define the mle $\hat{\theta} = \arg \max (L(\theta))$

We compute it by setting $L'(\theta) = 0$ and checking that $L''(\theta) < 0$, or that $L'$ has sign change $+ 0 -$ about the maximum. Alternatively, plot $L'(\theta)$ as a function of $\theta$ and find the maximum numerically.

Computational trick: maximize the log likelihood $\ell(\theta) = \log(L(\theta))$

### Exponential, cont.

$$L(\lambda) = \lambda^n \exp(-\lambda \sum x_i)$$

$$\ell(\lambda) = \log L(\lambda) = n \log \lambda - \lambda \sum x_i$$

$$\ell'(\lambda) = \frac{n}{\lambda} - \sum x_i$$

$$\ell'(|\lambda) = 0 \Rightarrow \hat{\lambda} = \frac{n}{\sum x_i} = \frac{1}{\bar{x}}$$

$$\ell''(\lambda) = -\frac{n}{\lambda^2} < 0$$

### Normal example

$$L(\mu) = \left(\frac{1}{2\pi}\right)^{n/2} e^{-\frac{1}{2} \sum (x_i - \mu)^2}$$
Stopping

Flip a coin until the first heads. Suppose it takes 6 tries. The likelihood is

Now suppose we were going to flip the coin 6 times, and happened to get one head. The likelihood is

How are the mles different?

Fact: Changing the likelihood by a constant does not change the mle.

However, the standard errors (sd of sampling distribution) would be different.

Approximate standard errors for the mle

An approximate formula for the estimated standard error of an mle (valid for large samples and for smooth likelihoods) is

$$\text{ese}(\hat{\theta}) = \left( -\frac{d^2}{d\theta^2} l(\hat{\theta}) \right)^{-1/2}$$

The method of moments

Karl Pearson (1900) proposed this way of estimating parameters. Suppose we have a distribution with $E_\theta(X)=h(\theta)$. By the law of large numbers,

$$E_\hat{\theta}(X) \sim \bar{X}$$

so we let our estimate $\hat{\theta}(x_1, ..., x_n)$ solve the equation

$$h(\theta) = \bar{X}$$

The exponential case

Let $X_1, ..., X_n$ be iid $\text{Exp}(\lambda)$, so $E_\lambda(X)=1/\lambda$. Below is $h(\lambda)$ plotted against $\lambda$, with three samples of size 50 for $\lambda=2.5$ marked.

Karl Pearson, 1857-1936
Bias calculation

We have $Y = \sum X_i \sim \Gamma(n, \lambda)$, the gamma distribution with density

$$f(y) = \frac{\lambda^n y^{n-1} e^{-\lambda y}}{(n-1)!}, \quad y > 0$$

$$E(\tilde{\lambda}) = nE(1/Y) =$$

$$\text{bias}(\tilde{\lambda}; \lambda) =$$

The hypergeometric distribution

Consider $X \sim \text{Hyp}(N, n, w)$ and assume that we observe $X=x$. How can we estimate $p=w/N$?

$$E(X) =$$

Muon decay

The cosine $X$ of the emission angle of electrons in muon decay has density

$$f_x(x; \alpha) = \frac{1+\alpha x}{2}, \quad -1 \leq x \leq 1, \quad -1 \leq \alpha \leq 1$$

The mean of the density is

$$h(\alpha) =$$

and the resulting mom estimator is

Some properties

Unbiased
Minimum variance
Minimum variance unbiased
Minimum mean squared error:

$$\text{Mse}(\theta^*; \theta) = E_\theta [(\theta^* - \theta)^2]$$

For unbiased estimators $\text{Mse} = \text{Var}$

$$\text{Mse}(\theta^*; \theta) = \text{Var}(\theta^*) + [\text{Bias}(\theta^*; \theta)]^2$$
Important distinctions

$X_i$ is a random variable
$x_i$ is the value of the random variable after we observe it (a number)

An estimator is a function of random variables. It is a random variable.

An estimate (the observed value of the estimator) is a function of observed values. It is a number.

Friday’s lecture

Maximum likelihood
Method of moments
Comparison of estimators
Mean squared error

Normal case again

Sample from $N(\mu,1)$, but we want to estimate $\theta = \mu^2$.

Invariance of mle

Suppose we are interested in estimating a (nice) function $h(\theta)$, using a sample from $f(x;\theta)$. If $\hat{\theta}$ is the mle of $\theta$, then $h(\hat{\theta})$ is the mle of $h(\theta)$. We say that the mle is invariant under reparametrization.
Standard errors

Sample from Bin(1, p)
Mle?
Bias?
Variance?
Standard error?
Estimated standard error?

Binomial distribution
Consider a binomial experiment with four trials, and outcomes \( x_1, \ldots, x_4 \). Here are three possible estimators of the success probability \( p \):
\[
\hat{p}_1 = \bar{X} \\
\hat{p}_2 = X_2 \\
\hat{p}_3 = X_1 + 2X_2 - 4X_3 + 2X_4
\]
Which of these are unbiased?

Computing mean squared error

Sample from Exp(\( \lambda \))

Sample variance

The mle of the variance for the normal distribution is
\[
\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2
\]
A tedious calculation (p. 384) shows that
\[
E(\hat{\sigma}^2) = \frac{n-1}{n} \sigma^2
\]
This is true for any distribution with a variance. Often one defines the sample variance as the unbiased estimator
\[
S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2
\]
Efficiency

Define the relative efficiency of two unbiased estimators by the ratio of their variances

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\text{Var}(\hat{\theta}_2)}{\text{Var}(\hat{\theta}_1)}$$

In the binomial case

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) =$$

and

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_3) =$$

Uniform case

$$\text{Mse}(\hat{\theta}) =$$

$$\text{Mse}(\tilde{\theta}) =$$

Uniform case

Recall that $E(\hat{\theta}) = \frac{n\theta}{n+1}$, so $\theta^* = \frac{n+1}{n} \theta$ is unbiased.

$$\text{eff}(\hat{\theta}, \theta^*) =$$

Finding the best unbiased estimator

Cramér-Rao

Harald Cramér 1893-1985
C. R. Rao 1920-

Important contributors to mathematical statistics.
Cramér-Rao inequality: $X_1, ..., X_n$ iid from smooth density, unbiased estimator $\theta^* = \theta^* (X_1, ..., X_n)$

$$\text{Var}(\theta^*) = \left[ -nE\left( \frac{\partial^2}{\partial \theta^2} \ln f(Y; \theta) \right) \right]^{-1}$$
Exponential case

\[ \frac{n-1}{n} \lambda \] is unbiased, with variance \[ \frac{\lambda^2}{n} \]. The lower bound is obtained from

\[ \ln(f(x; \lambda)) = \ln \lambda - \lambda x \]

\[ \frac{\partial}{\partial \lambda} \ln(f(x; \lambda)) = \frac{1}{\lambda} - x \]

\[ \frac{\partial^2}{\partial \lambda^2} \ln(f(x; \lambda)) = -\frac{1}{\lambda^2} \]

Hence the bound is \( \lambda^2/n \) which is smaller than the variance of the biascorrected mle/mome.

Parametrization matters!

Let \( \theta = 1/\lambda \), so \( E(X) = \theta \). Then the mle of \( \theta \) is \( \bar{X} \) which is unbiased with variance

\[ \text{Var}(X) / n = \theta^2 / n. \]

The log density is \( -\ln(\theta) - x/\theta \)
with second derivative \( 1/\theta^2 - 2x/\theta^3 \).
Thus the lower bound is

\[ (n(2 E(X) / \theta^3 - 1 / \theta^3))^{-1} = \theta^2/n. \]

The \( \lambda \)-parametrization is called the natural parametrization, while the \( \theta \)-parametrization is the mean-value parametrization.

Another best unbiased estimator

Let \( X \sim \text{Po}(\lambda) \), and try to estimate \( P(X=0) \).

If \( h(x) \) is unbiased we have

\[ \sum_{x=0}^{\infty} \frac{h(x) \lambda^x}{x!} e^{-\lambda} = e^{-\lambda} \]

\[ \sum_{x=0}^{\infty} h(x) \lambda^x = 1 \]

By identifying terms in the power series we must have

\[ h(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{if } x \neq 0 \end{cases} \]

What is the mle?

Is it unbiased?