Note Title 1/12/2005

Stat 543 1-12-05

The most basic tool of statistical theory is the "likelihood function" — a (random) function of θ that says how

$$\text{"}P_\theta[X = \text{data in hand}]\text{"}$$

I can't be precise about this in general w/o Stat 642 background — I can say exactly what I mean for discrete and cont$^{\underline{s}}$ cases

Discrete Cases: If X is discrete with (joint) pmf $f(x|\theta)$ then

$$L(\theta) = f(x|\theta)$$

data in hand

(treated as a function of θ)

Cont$^{\underline{s}}$ Cases: If X is cont$^{\underline{s}}$ with (joint) pdf $f(x|\theta)$ then

$$L(\theta) = f(x|\theta)$$

data in hand

(treated as a function of θ)

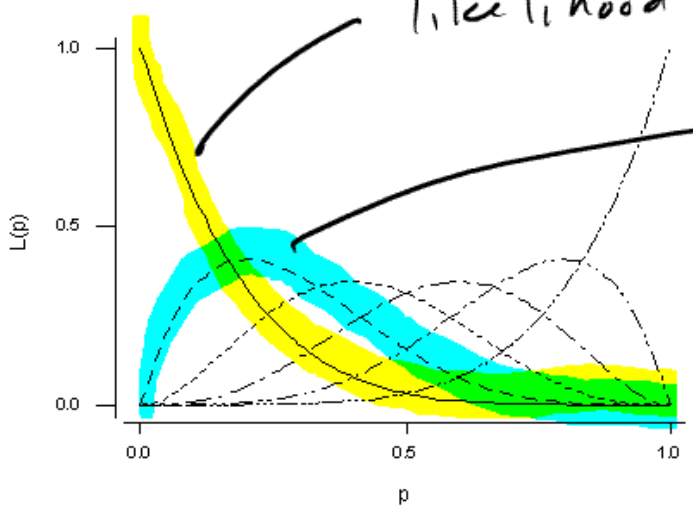
Example $X \sim Bi(5, p)$

$$f(x \mid p) = \binom{5}{x} p^x (1-p)^{5-x} \quad \text{for}$$

$x = 0, 1, 2, \ldots, 5$

treat this as a function of p

likelihood function if $x = 0$



likelihood function if $x = 1$

small p will typically produce small x's ... likelihoods that are big for small

Perhaps "where the likelihood is big" is a reliable indicator of what p's are plausible based on the data

<u>Example</u> $X_1, X_2, \ldots, X_n$ iid $N(\mu, \sigma^2)$

$$L(\mu,\sigma)_{(x_1, x_2, \ldots, x_n)} = f(x \mid \mu, \sigma) = \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right)\right)$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sigma^n} \exp\left(-\frac{1}{2\sigma^2} \sum (x_i-\mu)^2\right)$$

A (random) function of (μ, σ) that hopefully is a reliable indicator about (μ, σ)

Classical/frequentist inference relies completely on the likelihood function –

So called "Bayes" inference uses the likelihood function ($L(\theta)$) <u>plus</u> a so-called "prior" dsn for θ – in a sense Bayes inference turns a statistical problem back into a probability problem by using the machinery of joint and conditional dsns – the idea is:

$$L(\theta) = f(x \mid \theta)$$

at least in disrete or conts cases

a "prior" dsn

$\rightarrow$ G some dsn on Θ specified by $g(\theta)$

$$f(x|\theta) + g(\theta) \Rightarrow \text{jt dsn for } X, \theta$$

$$\Downarrow$$

$$\underbrace{\text{conditional dsn for } \theta | X = x}$$

specified by

$$g(\theta|x) = \frac{f(x,\theta)}{\int f(x,\theta)\, d\theta}$$

$f(x|\theta)g(\theta)$

could be a sum

So as a function of θ

specifies a "posterior dsn" for θ

$$g(\theta|x) \propto f(x,\theta) = g(\theta)\underbrace{f(x|\theta)}_{L(\theta)}$$

and a Bayesian uses this product of $L(\theta)$ and $g(\theta)$ as the basis for inference

<u>Example</u> $X \sim Bi(5, p)$

$$L(p) = f(x|p) = \binom{5}{x} p^x (1-p)^{5-x}$$

Suppose the prior dsn of p is $U(0,1)$

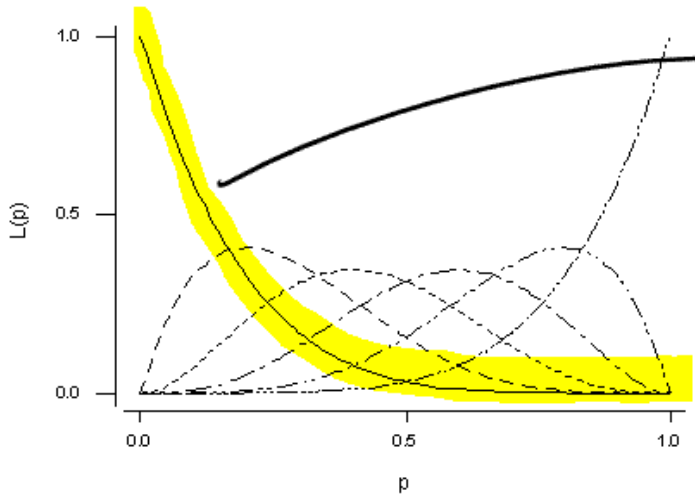
i.e. $g(p) = I[0 < p < 1]$

$$g(p|x) \propto g(p)\, f(x|p)$$

0 or 1

posterior $p|x$ is Beta $(x+1, 5-x+1)$

$$\binom{5}{x} p^{x} (1-p)^{5-x}$$



both $L(p)$ if $x=0$ and proportional to $g(p|x)$ if $x=0$

<u>Example</u> $X_1, X_2, \ldots, X_n$ iid $N(\mu, 1)$

$$\underset{L(\mu)}{f(x|\mu)} = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2}\sum(x_i-\mu)^2\right]$$

Suppose a prior dsn for μ is $N(0, \gamma^2)$, i.e.

$$g(\mu) = \frac{1}{\sqrt{2\pi}\,\gamma} \exp\left[-\frac{1}{2}\frac{\mu^2}{\gamma^2}\right]$$

$$g(\mu|x) \propto \exp\left[\frac{-1}{2}\left(-2\mu\sum x_i + \mu^2\left(n+\frac{1}{\gamma^2}\right)\right)\right]$$

$$\vdots$$

$$\exp\left[-\frac{1}{2\,\frac{\gamma^2}{n\gamma^2+1}}\left(\mu^2 - 2\mu\bar{x}\frac{n\gamma^2}{n\gamma^2+1}\right)\right]$$

i.e. the posterior dsn of μ is

$$N\left(\bar{x}\left(\frac{n\gamma^2}{n\gamma^2+1}\right), \frac{\gamma^2}{n\gamma^2+1}\right)$$

so, e.g. there is 95% posterior probability

$$\bar{x}\left(\frac{n\gamma^2}{n\gamma^2+1}\right) - 1.96\sqrt{\frac{\gamma^2}{n\gamma^2+1}} < \mu < \qquad +$$

In general the $g(\theta|x) \propto L(\theta)\, g(\theta)$
That doesn't come out in clean analytic terms