Math 152

3/3/09

The following plausible argument supports our observation that the posterior is approximately normal with mean equal to the maximum likelihood estimate $\hat{\theta}$

and variance approximately equal to

$[-l''(\hat{\theta})]^{-1}$!

$$f_{\Theta|X}(\theta|x) \propto f_\Theta(\theta)\, f_{X|\Theta}(x|\theta).$$

$$= e^{\log f_\Theta(\theta)}\, e^{\log f_{X|\Theta}(x|\theta)}$$

$$= e^{\log f_\Theta(\theta)}\, e^{l(\theta)}$$

We then approximate,

$$f_{\Theta|X}(\theta|x) \propto e^{[l(\hat{\theta}) + (\theta-\hat{\theta})l'(\hat{\theta}) + \frac{1}{2}(\theta-\hat{\theta})^2 l''(\hat{\theta})]}$$

the justification being that in the region where the likelihood is significant, the prior is $\approx$ constant and when the likelihood is insignificant so is the posterior.

Since $l'(\hat{\theta})=0$, this is

$$f_{\theta|x}(\theta|x) \quad \propto \quad e^{\frac{1}{2}(\theta-\hat{\theta})^2 l''(\hat{\theta})}$$

which is proportional to the normal density with mean $\hat{\theta}$ and variance $[-l''(\hat{\theta})]^{-1}$

Another Bayesian estimation example

$$f(x \mid \theta, \xi) = \left(\frac{\xi}{2\pi}\right)^{1/2} e^{-\frac{1}{2}\xi(x-\theta)^2}$$

Normal with mean θ, precision ξ.

$$\boxed{\xi = \frac{1}{\sigma^2}}$$

Unknown mean and known variance

$\xi = \xi_0$ is known.

θ is unknown.

which we treat as a R.V. Θ.

We take $\Theta \sim N(\theta_0, \xi_{prior}^{-1})$.

which will be "flat" (uninformative). if ξ_{prior} is small.

If $X = (X_1, \ldots, X_n)$ are independent given θ.

$$f_{\theta|X}(\theta|x) \propto f_{X|\theta}(x|\theta)\, f_\theta(\theta).$$

$$= \left(\frac{\xi_0}{2\pi}\right)^{n/2} \prod_{i=1}^{n} e^{-\frac{\xi_0}{2}(x_i-\theta)^2} \cdot \sqrt{\frac{\xi_{prior}}{2\pi}}\, e^{-\left(\frac{\xi_{prior}}{2}(\theta-\theta_0)^2\right)}$$

$$(*) \quad \propto \quad e^{-\frac{1}{2}\left[\xi_0 \sum_{i=1}^{n}(x_i-\theta)^2 + \xi_{prior}(\theta-\theta_0)^2\right]}.$$

constants being absorbed.

Now $\sum_{i=1}^{n}(x_i-\theta)^2 = \sum_{i=1}^{n}(x_i - \bar{x} + \bar{x} - \theta)^2$

$$= \sum_{i=1}^{n}(x_i-\bar{x})^2 + 2(x_i-\bar{x})(\bar{x}-\theta) + (\bar{x}-\theta)^2$$

$$= \left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right) + n(\bar{x}-\theta)^2.$$

So

$$(*) \quad \text{is} \quad \propto \quad e^{-\frac{1}{2}\left[n\,\xi_0(\bar{x}-\theta)^2 + \xi_{prior}(\theta-\theta_0)^2\right]}$$

$$= e^{-1/2\left[a\theta^2 + b\theta + c\right]}$$

$$= e^{-1/2\left[a\left(\theta+\frac{b}{2a}\right)^2 + c - \frac{b^2}{4a}\right]}.$$

So we have a normal density for the posterior with

$$\xi_{post} = a$$

$$\theta_{post} = -\frac{b}{2a}.$$

or $b = -2a\,\theta_{post}$

Now

$$n\xi_0(\bar{x}-\theta)^2 + \xi_{prior}(\theta-\theta_0)^2$$

$$= n\xi_0\left[\bar{x}^2 - 2\bar{x}\theta + \theta^2\right] + \xi_{prior}\left(\theta^2 - 2\theta\theta_0 + \theta_0^2\right)$$

$$= (n\xi_0 + \xi_{prior})\,\theta^2 + \theta\left(-2\bar{x}n\xi_0 - 2\theta_0\xi_{prior}\right)$$

$$+ (\text{const}).$$

so

$$\xi_{post} = n\xi_0 + \xi_{prior}.$$

$$\theta_{post} = \frac{\bar{x}n\xi_0 + \theta_0\xi_{prior}}{n\xi_0 + \xi_{prior}}$$

$$= \bar{x}\,\frac{n\xi_0}{n\xi_0 + \xi_{prior}} + \theta_0\,\frac{\xi_{prior}}{n\xi_0 + \xi_{prior}}$$

precision is increased.

θ_{post} is a weighted average of sample mean and prior mean.

If $\xi_{prior} << n\xi_0$.

(n large or prior very flat).

then

$\theta_{post} \approx \bar{x}$

$\xi_{post} \approx n\xi_0$

So if n is suff large or ξ_{prior} is small.

the posterior density of θ is $\approx$ normal with mean $\bar{x}$ and variance $\frac{1}{n\xi_0} = \frac{\sigma_0^2}{n}$.

Sufficiency

$X_1, \ldots, X_n$ a sample

from $f(x|\theta)$

Is there a statistic

$T(X_1, \ldots, X_n)$

that contains all information in the sample about θ?.

Def: $T(X_1, \ldots, X_n)$ is

a *sufficient* statistic for θ

iff

$$f_{X_1, \ldots X_n | T}(x_1, \ldots, x_n | t)$$

does not depend on θ for any value of $T = t$.

i.e. given a value $T = t$ the conditional dist. of $X_1, \ldots, X_n$ contains no further information on θ.

e.g. Bernoulli Trials. (n)

unknown prob success $= \theta$.

$T =$ total # of successes

turns out to be sufficient.

Can we throw away all other info on X_i and just record T?.

formally: yes.

realistically: no.

e.g. $X_1, \ldots X_n$ ind. Bernoulli

$$P(X_i = 1) = \theta.$$

$$T = \sum_{i=1}^{n} X_i$$

$$P(X_1 = x_1, \ldots X_n = x_n \mid T = t)$$

$$= \frac{P(X_1 = x_1, \ldots X_n = x_n, T = t)}{P(T = t)}.$$

$$= \frac{\theta^t (1-\theta)^{n-t}}{\binom{n}{t} \theta^t (1-\theta)^{n-t}} = \frac{1}{\binom{n}{t}}$$

Given that the numbers of successes is t the probability that they occur on any particular set of t outcomes is the same for any value of θ.

So T is <u>sufficient for θ</u>.

<u>Thm:</u> $T(X_1, \ldots, X_n)$ is sufficient for θ

iff

$$(*)\; f(x_1, \ldots, x_n \mid \theta) = g\big(T(x_1, \ldots, x_n), \theta\big) \cdot h(x_1, \ldots, x_n)$$

<u>Pf:</u> (Discrete Case).

let $\vec{X} = (X_1, \ldots, X_n)$ $\quad \vec{x} = (x_1, \ldots, x_n)$

and suppose $f(x_1, \ldots x_n \mid \theta)$ factors as in $(*)$

then

$$P(T=t) = \sum_{T(\vec{x})=t} P(\vec{X}=\vec{x}).$$

$$= g(t,\theta) \sum_{T(\vec{x})=t} h(\vec{x})$$

and $\therefore$

$$P(\vec{X}=\vec{x} \mid T=t) = \frac{P(\vec{X}=\vec{x}, T=t)}{P(T=t)}$$

$$= \frac{\cancel{g(t,\theta)}\, h(\vec{x})}{\cancel{g(t,\theta)} \sum_{T(\vec{x})=t} h(\vec{x})}$$

$$= \frac{h(\vec{x})}{\sum_{T(\vec{x})=t} h(\vec{x})}.$$

which does not depend on θ.

<u>Conversely</u>, Suppose $f_{X|T}(x|t)$

has no dependence on θ.

let $g(t,\theta) = P(T=t \mid \theta)$

$h(\vec{x}) = P(\vec{X}=\vec{x} \mid T=t)$

Then

$$P(\vec{X}=\vec{x} \mid \theta) = P(\vec{X}=\vec{x} \mid T=t)\, P(T=t \mid \theta)$$

$$= g(t,\theta)\, \underbrace{h(\vec{x})}_{\text{ind. of } \theta}$$

$\blacksquare$

#18 $X_1, \ldots X_n$ i.i.d. on $[0,1]$

$$f(x \mid \alpha) = \frac{\Gamma(3\alpha)}{\Gamma(\alpha)\Gamma(2\alpha)} x^{\alpha-1}(1-x)^{2\alpha-1}$$

$\alpha > 0$.

We know $E(X) = 1/3$

$\mathrm{Var}(X) = \dfrac{2}{9(3\alpha+1)}$.

method of moments for α

a) $\hat{\mu}_1 = \frac{1}{n} \sum_{i=1}^{n} X_i$

$\hat{\mu}_2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2$

Set $\hat{\mu}_2 - \left(\frac{1}{3}\right)^2 = \frac{2}{9(3\alpha+1)}$

and solve for α.

$$\frac{1}{\frac{9}{2}\left(\hat{\mu}_2 - 1/9\right)} = 3\hat{\alpha} + 1$$

$$\hat{\alpha} = \frac{1}{3}\left(\frac{2}{9\hat{\mu}_2 - 1} - 1\right)$$

$$= \frac{2}{27\hat{\mu}_2 - 3} - \frac{1}{3}.$$

b). m.l.e for α.

$$l(\alpha) = \sum_{i=1}^{n} \Big(\log \Gamma(3\alpha) - \log\Gamma(2\alpha) - \log \Gamma(\alpha)$$
$$+ (\alpha-1)\log X_i + (2\alpha-1)\log(1-X_i)\Big)$$

$$l'(\alpha) = 3n \frac{\Gamma'(3\alpha)}{\Gamma(3\alpha)} - \frac{2n\Gamma'(2\alpha)}{\Gamma(2\alpha)} - n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$$

$$+ \sum_{i=1}^{n} \log X_i + 2\sum_{i=1}^{n} \log(1-X_i).$$

$$l'(\alpha) = 0$$

$$\Leftrightarrow \quad 3\frac{\Gamma'(3\alpha)}{\Gamma(3\alpha)} - 2\frac{\Gamma'(2\alpha)}{\Gamma(2\alpha)} - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$$

$$\underbrace{+ \frac{1}{n}\sum_{i=1}^{n} \log X_i + 2\left(\frac{1}{n}\sum_{i=1}^{n}\log(1-X_i)\right)}_{+\frac{1}{n}\left(\sum_{i=1}^{n}\log(X_i(1-X_i))\right)}$$

$$= 0.$$

c) $l''(\alpha) = \ldots$

let $g(x) = \frac{\Gamma'(x)}{\Gamma(x)}$.

$$l''(\alpha) = 9 \cdot g'(3\alpha) - 4g'(2\alpha) - g'(\alpha)$$

$$-\frac{1}{E(l''(\alpha))} = \frac{1}{9g'(3\alpha) - 4g'(2\alpha) - g'(\alpha)}$$

and take $\alpha = \hat{\alpha}$.

d) $f(x_1, \dots, x_n \mid \alpha)$

$$= \prod_{i=1}^{n} \frac{\Gamma(3\alpha)}{\Gamma(\alpha)\Gamma(2\alpha)} x_i^{\alpha-1} (1-x_i)^{2\alpha-1}$$

$$= \left(\frac{\Gamma(3\alpha)}{\Gamma(\alpha)\Gamma(2\alpha)} \right)^n \frac{\prod_{i=1}^{n} x_i^{\alpha} (1-x_i)^{2\alpha}}{\prod_{i=1}^{n} x_i (1-x_i)}.$$

$$= \left(\frac{\Gamma(3\alpha)}{\Gamma(\alpha)\Gamma(2\alpha)} \right)^n g(T(\vec{x}), \alpha) \cdot h(\vec{x}).$$

where $T(\vec{x}) = \prod_{i=1}^{n} x_i (1-x_i)^2.$

$$g(\vec{T}, \alpha) = \vec{T}^{\alpha}$$

$$h(\vec{x}) = \frac{1}{\prod_{i=1}^{n} x_i (1-x_i)}.$$