Stat 543 4-25-05

Recall: Kullback-Leibler Information

$$K(\theta, \theta') = E_\theta \log \frac{f(X|\theta)}{f(X|\theta')}$$

Lemma If X has pdf or pmf $f(x|\theta)$ then

$$K(\theta, \theta') \geq 0$$

and if θ and θ' produce different dsns for X, the inequality is strict

pf:

$$K(\theta, \theta') = E_\theta \log \frac{f(X|\theta)}{f(X|\theta')}$$

$$= E_\theta - \log \frac{f(X|\theta')}{f(X|\theta)}$$

Jensen's Inequality $\longrightarrow$ $\geq - \log \left(E_\theta \dfrac{f(X|\theta')}{f(X|\theta)} \right)$

$-\log(\cdot)$ is a convex function of $\cdot$

$= -\log 1$

$= 0$

e.g. in cont's case This is
$\int \dfrac{f(x|\theta')}{\cancel{f(x|\theta)}} \cancel{f(x|\theta)}\, dx$

The strict inequality comes from the strict convexity of $-\log(\cdot)$ and Jensen's inequality in that context. $\geq$

Back to our iid model — $X_1, X_2, \ldots, X_n$ iid
$f(x|\theta)$ marginal pmf or pdf — as usual

$$L_n(\theta) = \prod_{i=1}^{n} f(X_i \mid \theta)$$

Lemma In this iid model if the θ and θ' dsns for X_1 are different

$$\frac{L_n(\theta')}{L_n(\theta)} \xrightarrow{P_\theta} 0$$

(This is actually what stands behind Thm 1 on ML handout)

Pf:
$$\frac{L_n(\theta')}{L_n(\theta)} = \exp\left(-\log \frac{L_n(\theta)}{L_n(\theta')}\right)$$
$$= \exp\left(-(n)\frac{1}{n}\log \frac{L_n(\theta)}{L_n(\theta')}\right)$$

$$= \exp\left(-n\left(\underbrace{\frac{1}{n}\sum_{i=1}^{n}\log\frac{f(X_i|\theta)}{f(X_i|\theta')}}\right)\right)$$

by WLLN under θ
This $\Longrightarrow K(\theta,\theta') > 0$
in probability

So if $\epsilon > 0$ I can choose m large enough so that if $n \geq m$

$$P_\theta\left[\frac{1}{n}\sum_{i=1}^{n}\log\frac{f(X_i|\theta)}{f(X_i|\theta')} > \frac{1}{2}K(\theta,\theta')\right] \geq 1-\epsilon$$

Then, if $n \geq m$

$$P_\theta\left[\frac{L_n(\theta')}{L_n(\theta)} \leq \exp\left(-n\,\frac{1}{2}K(\theta,\theta')\right)\right] \geq 1-\epsilon$$

Since $\exp\left(-\frac{n K(\theta,\theta')}{2}\right)$ is eventually $\leq$ any positive #, we're done. ∎

<u>Corollary</u> In an iid model if Θ is finite, $g(\theta) > 0$ for each θ and no two dsns for X_1 (specified by $f(x|\theta)$) are the same, then the posterior dsn $g(\cdot|X)$ is "consistent" in the sense that

$$g(\theta|X) \xrightarrow{P_\theta} 1$$

while

$$g(\theta'|X) \xrightarrow{P_\theta} 0 \quad \text{for any } \theta' \neq \theta$$

<u>Pf</u>:

$$g(\theta|X) = \frac{g(\theta)\,L_n(\theta)}{g(\theta)\,L_n(\theta) + \sum_{\theta' \neq \theta} g(\theta')\,L_n(\theta')}$$

$$= \frac{1}{1 + \sum_{\theta \neq \theta'} \frac{g(\theta')\,L_n(\theta')}{g(\theta)\,L_n(\theta)}}$$

there are a finite # of terms in this sum, all converging in θ probability to 0 and so some continuity Thm implies that

$$g(\theta|X) \xrightarrow{P_\theta} \frac{1}{1+0} = 1$$

Since $g(\cdot|X)$ specifies a discrete dsn over Θ we then must have $g(\theta'|X) \xrightarrow{P_\theta} 0$ if $\theta' \neq \theta$. ∎

There are more complicated non-finite Θ versions of this result that say that under appropriate conditions in an iid model with a "diffuse" prior G, for $A \subset \Theta$ and $\theta \in A$

r.v.'s

(G) posterior probability assigned to A $\xrightarrow{P_\theta}$ 1

(G) posterior probability assigned to A^c $\xrightarrow{P_\theta}$ 0

There are also Thms that allow me to make large sample approximations to the form of a posterior — i.e. there are results like

Result In an iid model where a prior for $\theta \in \mathbb{R}^1$ has a density that is cont$^{\underline{s}}$ and positive at θ_0 and regularity conditions hold, if $\delta_n(X)$ is the "MLE" of θ, then under the θ_0 dsn for X the posterior density ___ "random pdf" of

$$\sqrt{-\ell_n''(\delta_n(X))}\left(\theta - \delta_n(X)\right)$$

converges to the std normal density $\left(\frac{1}{\sqrt{2\pi}} \exp{-\frac{\theta^2}{2}}\right)$