TOPIC: Missing Information

DATE:

FILE UNDER:

PAGE:

- $f(y|\theta)$: observed data density
- $g(y,z|\theta)$: complete data density
- $h(z|y,\theta) \triangleq \dfrac{g(y,z|\theta)}{f(y|\theta)}$ missing data density

$$f(y|\theta) = \frac{g(y,z|\theta)}{h(z|\theta,y)}$$

$$-\log f(y|\theta) = -\log g(y,z|\theta) - \left[-\log h(z|y,\theta)\right]$$

$\frac{\partial}{\partial\theta\partial\theta^T}$ ↓

$$-\frac{\partial}{\partial\theta\partial\theta^T}\log f(y|\theta) = -\frac{\partial}{\partial\theta\partial\theta^T}\log g(y,z|\theta) - \left[-\frac{\partial}{\partial\theta\partial\theta^T}\log h(z|y,\theta)\right]$$

$E_{z|y}$ ↓

$$\underbrace{I_y(\theta)}_{\text{obs}} = \underbrace{I_{y,z}(\theta)}_{\text{complete}} - \underbrace{I_{z|y}(\theta)}_{\text{missing}}$$

- How to compute $I_{y|z}(\theta)$ and $I_{z|y}(\theta)$?

Let $S(y|\theta) = \frac{\partial}{\partial\theta}\log f(y|\theta)$

$$S(y,z|\theta) = \frac{\partial}{\partial\theta}\log g(y,z|\theta)$$

TOPIC:

DATE:

FILE UNDER:

PAGE:

Louis 1982 showed:

$$I_{Z|Y}(\theta) = E_{Z|Y}\left[S(Y,Z|\theta)S(Y,Z|\theta)^T\right] - S(Y|\theta)S(Y|\theta)^T$$

with expectation taken w.r.t $h(Z|Y,\theta)$

$$\Rightarrow I_Y(\theta) = I_{Y,Z}(\theta) - E_{Z|Y}\left[S(Y,Z|\theta)S(Y,Z|\theta)^T\right] - S(Y|\theta)S(Y|\theta)^T$$

Computation only on complete data

$= 0$ for $\theta = \hat{\theta}$

Note $I_{Y,Z}(\hat{\theta}) = -E\left[\frac{\partial^2}{\partial\theta\partial\theta'}\log g(Y,Z|\theta) \,\middle|\, \theta_n, Y\right]$

$= -Q''(\hat{\theta}|\hat{\theta})$

Meilijson (89) showed that if Y's are iid, then

~~$I_Y(\theta) = Var(S(Y|\theta))$~~

$$S(Y|\theta) = \sum_{i=1}^{n} S(Y_i|\theta)$$

TOPIC:

DATE:

FILE UNDER:

PAGE:

$$I_Y(\theta) = \mathrm{Var}(s(Y|\theta))$$

$$= \frac{1}{n}\sum s(y_i|\theta)s(y_i|\theta)^T - \underbrace{\left[\frac{1}{n}\sum s(y_i|\theta)\right]\left[\frac{1}{n}\sum s(y_i|\theta)\right]^T}_{0 \text{ for } \theta=\hat{\theta}}$$

How?

Louis '82 showed

$$s(Y|\theta) = E_{Z|Y}\left[s(Y,Z|\theta)\right]$$

$$I_Y(\hat{\theta}) = \frac{1}{n}\sum E_{Z|Y}\left[s(y_i,z_i|\hat{\theta})\right] E_{Z}\left[s(y_i,z_i|\hat{\theta})\right]^T$$

Ex: Mixture models

$$\ell_i(\theta) = z_i \log \phi(y_i|\mu_1,\sigma_1^2) + (1-z_i)\log(y_i|\mu_2\sigma_2) + z_i\log\lambda + (1-z_i)\log(1-\lambda)$$

$$s(y_i,z_i|\theta) = \begin{pmatrix} \frac{y_i-\mu_1}{\sigma_1^2} z_i \\ \frac{y-\mu_2}{\sigma_2^2}(1-z_i) \\ \left(-\frac{1}{\sigma_1} + \frac{1}{\sigma_1^3}(y_i-\mu_1)\right) z_i \\ \left(-\frac{1}{\sigma_2} + \frac{1}{\sigma_2^3}(y_i-\mu_2)\right)(1-z_i) \\ \frac{z_i}{\lambda} + \frac{1-z_i}{1-\lambda} \end{pmatrix}$$

TOPIC:

DATE:

FILE UNDER:

PAGE:

[EX FILE [X|Y]]

It is MCMC, but not Gibbs. exactly

Data Augmentation (Tanner + Wong '87)

We have some data $y_1, \dots, y_n$, and we propose a model

$$y \sim p(y|\theta)$$

$$\theta \sim \pi(\theta)$$

So we have a likelihood $p(y|\theta)$ and a prior $\pi(\theta)$.

We want the posterior $p(\theta|y)$.

$$p(\theta|y) \propto p(y|\theta)\pi(\theta)$$

BUT because of some missing data z, it is difficult to evaluate $p(\theta|y)$.

Let (y, z) be the "complete data" and we want to know

$$p(\theta|y) = \int p(\theta|y,z)\, p(z|y)\, dz \qquad (1)$$

$p(\theta|y,z)$: "complete data posterior" → easy to calculate

$p(z|y)$: predictive density for z.

Notice also that we have

$$p(z|y) = \int p(z|y,\theta)\, p(\theta|y)\, d\theta \qquad (2)$$

($p(z|y,\theta)$: easy to calculate)

Substituting, (2) into (1), we get

$$p(\theta|y) = \int p(\theta|y,z) \left[\int p(z|y,\theta')\, p(\theta'|y)\, d\theta' \right] dz$$

$$= \int \left[\int p(\theta|y,z)\, p(z|y,\theta')\, dz \right] p(\theta'|y)\, d\theta'$$

System of integral equations

TOPIC:

DATE:

FILE UNDER:

PAGE:

Let $K(\theta, \theta') = \int p(\theta|y,z)\, p(z|y,\theta')\, dz$

$$p(\theta|y) = \int K(\theta,\theta')\, p(\theta'|y)\, d\theta'$$

(1) (2)

(fixed point system) If $p(\theta'|y)$ is the true posterior, then it will map to itself.

Let T be a functional which maps a function $g(\theta)$ to

$$Tg(\theta) = \int K(\theta,\theta')\, g(\theta')\, d\theta'$$

If we take some initial value $p_0(\theta|y)$, what happens when we iterate $T p_0(\theta|y)$ i.e.

$$p_1(\theta|y) = T\, p_0(\theta|y) = \int K(\theta,\theta')\, p_0(\theta'|y)\, d\theta'.$$

Does the sequence $\{p_i(\theta|y)\}$ converge to anything?

(1) $\{p_i(\theta|y)\} \longrightarrow p(\theta|y)$ monotonically (always getting closer)
true posterior

(2) $p(\theta|y)$ is the unique solution to system of integral equations

(3) $p_i(\theta|y) \longrightarrow p(\theta|y)$ linearly.

TOPIC:

DATE:

FILE UNDER:

PAGE:

DA algorithm

Roughly: At iteration i

(1) Sample $z_1, \ldots, z_m \sim p_i(z|y)$

(2) Estimate $p_{i+1}(\theta|y) = \frac{1}{m}\sum_{i=1}^{m} p(\theta|y, z_i)$

complete data posterior

Repeat until $\|p_{i+1}(\theta|y) - p_i(\theta|y)\| < \varepsilon$.

---

(1) Pick some initial $p_0(\theta|y)$. At step i

(2a) Generate $\theta \sim p_i(\theta|y)$

(2b) Generate $z \sim p(z|y, \theta)$

Repeat (2a) – (2b) m times to get $z_1, \ldots, z_m$.

(3) Let $p_{i+1}(\theta|y) = \frac{1}{m}\sum p(\theta|y, z_i)$

mixture of conditional densities.

Monte Carlo estimate of

$$\int p(\theta|y, z)\, p(z|y)\, dz$$

DA not so straightforward when there are more than 2 "missing" components.

Use Gibbs Sampling for more complicated problems

TOPIC:

DATE:

FILE UNDER:

PAGE:

Ex. 3-stage hierarchical model

$Y|\alpha \sim N(\alpha, 1)$

$\alpha|\theta \sim N(\theta, 1)$

$\theta \sim \pi(\theta) = N(0,1)$

$p(Y, \alpha, \theta) = p(Y|\alpha,\theta)\, p(\alpha|\theta)\, \pi(\theta)$
$N(\alpha,1) \quad N(\theta,1) \quad N(0,1)$

$p(\alpha|Y,\theta) \propto p(Y|\alpha,\theta)\, p(\alpha|\theta)$

$= N\left(Y + \frac{1}{2}(\theta - Y), \frac{1}{2}\right)$

$p(\theta|Y,\alpha) \propto p(Y|\alpha,\theta)\, p(\alpha|\theta)\, \pi(\theta)$

$N(\quad)$

(1) find some $p_i(\theta|y)$ [Normal?]

(2a) sample $\theta \sim p_i(\theta|y)$

(2b) Sample $\alpha \sim p(\alpha|y,\theta)$

Repeat m times to get $\alpha_1, \ldots, \alpha_m$

(3) Let $p_{iq}(\theta|y) = \frac{1}{m}\sum p_i(\theta|y,\alpha)$

TOPIC:

DATE:

FILE UNDER:

PAGE:

	EM	DA
Paradigm	likelihood	Bayesian
Target	MLE	Posterior dist.
Method to Handle Z	Average (conditional)	sample (conditional)
Need $p(Z\|Y)$	Compute E	sample from
Std. errors	No, by default	Yes, have entire posterior (model error)
Rate of convergence	linear	linear
Monotonicity	✓	✓

Both methods "impute" missing data via the specification of the complete data model.

Once $p(Y, Z)$ is specified, everything else is determined.