International Conference in Honor of P.L. Hsu's 100th Brithday

Peking University, Beijing

Session Stat3 on 7 July 2010, 2:00-2:30 pm

Bayesian Inference Is a Two-way Street

Kai Wang Ng (kaing@hku.hk)

Department of Statistics & Actuarial Science

The University of Hong Kong

э l G ɣ X>Fa®-¥´¢¨´²>² ¨£G>
>
ʁ П ˜ Ł A>y Ḯ ⊐ A Ÿ Ł ƈ{>
Э П ˜ Ł Ш>y Ḯ˜ Ш Ÿ ŵ ɔ Ụ ғ {>

An old-time scholar learns for [satisfying] oneself. [Scholarship]
A later-day scholar learns for [satisfying] others. [Entrepreneurship]

本人意見：

先為己、然後為人之學者，
至善者也。

My academic idol is one who puts scholarship before entrepreneurship.

Professor P.L. Hsu was well know for his

- Spirit: scholarship before entrepreneurship
- Practice: go through the statement and proof of every theorem he needed or quoted, and provide a simpler derivation if possible

Reading and hearing about Professor Hsu's virtues, I must confess that I have not been able to follow. Only occasionally have I tried his practice.

In honor of Professor Hsu in this occasion, thanks to the organizers for inviting me, I would like to share one of my experiences of trying out his practice, which led to

an unexpected journey — *reversing the Bayes' process of inference.*

1. What's the problem?

Missing (augmented) Data Problem:
Complete data $(Y; Z)$, but Z not available — either missing or strategically augmented.

Frequentist: MLE of μ by EM algorithm

- **Folklore version: (may not ascend likelihood)**
  For an initial $\mu^{(0)}$, the "E-step" finds the conditional mean $E(Z|y; \mu^{(0)}) = z^{(0)}$ and the "M-step" finds the new MLE

$$\mu^{(1)} = arg\ max\{log f(y; z^{(0)}|\mu)\}:$$

- **Dempster *et al.* (1977) (ascending):**
  For an initial $\mu^{(0)}$, determine function

$$Q(\mu|\mu^{(0)}) = E(log f(y; Z|\mu)|y; \mu^{(0)}):$$

  Then find the new MLE (M-step)

$$\mu^{(1)} = arg\ max\ Q(\mu|\mu^{(0)}):$$

- **Meng & Rubin (1993) ECM: [Bio'ka]**
  The M-step replaced by a number of conditional maximization steps.

In Bayesian analysis of $(Y; Z)$, where Z is not available, the question is to find $¼(\mu|y)$, based on $p(\mu|y; z)$ & $f(z|y; \mu)$.

1. **Tanner and Wong (1987, *JASA*, 5 discussants)**

   ***Data Augmentation (DA)* by successive substitution**

   with discussions respectively by Dempster, Morris, Rubin, Haberman, and O'Hagan

2. **Schervish & Carlin (1992, *JCGS*)**

   Convergence rate of *DA* algorithm

3. **Gelfand & Smith (1990, *JASA*)**

   - **construction of pdf using *Gibbs sampler***
   - **"Extend" without convergence conditions *DA* algorithm to 3 groups**

4. **Liu, Wong and Kong (1994, *Biometrika*. 1995, *JRSSB*)**

   Covariance structure of *Gibbs and DA*

5. **M. Tanner (3rd Printing of 1st Ed. in 1993; 2nd Ed. in 1994 or 1995; 3rd Ed. in 1996.)**

   "Tools for Statistical Inference ", *Springer*

Tanner & Wong, 1987, *JASA*; Tanner, 1993-1996: Successive substitution

To find $\pi(\mu|y)$ based on $p(\mu|y;z)$, $f(z|\mu;y)$:

$$\underbrace{\pi(\mu|y)}_{g(\mu)} = \int p(\mu|y;z)\, h(z|y)dz$$

$$= \int \underbrace{p(\mu|y;z)}_{f_1(\mu|z)} \left\{ \int \underbrace{f(z|\acute{\alpha};y)}_{f_2(z|\acute{\alpha})} \underbrace{\pi(\acute{\alpha}|y)}_{g(\acute{\alpha})} d\acute{\alpha} \right\} dz$$

$$g(\mu) = \int \left\{ \int f_1(\mu|z) f_2(z|\acute{\alpha})dz \right\} g(\acute{\alpha})d\acute{\alpha}$$

$$= \int K(\mu;\acute{\alpha})g(\acute{\alpha})d\acute{\alpha} :$$

Successive substitution to approximate $g(\mu)$:

$$g_{k+1}(\mu) = \int K(\mu;\acute{\alpha})g_k(\acute{\alpha})d\acute{\alpha}$$

Sufficient conditions for $g_k \to g$ as $k \to \infty$:

(i) $K(\mu;\acute{\alpha})$ is uniformly bounded

(ii) $K(\mu;\acute{\alpha})$ is equi-continuous in μ

(iii) $\forall\, \mu_0 \in \Theta$ $\exists$ open neigh. U s.t.

$$K(\mu;\acute{\alpha}) > 0 \quad \forall\ \mu;\acute{\alpha} \in U :$$

(iv) Starting with g_0 satisfying: $\sup_{\mu} \dfrac{g_0(\mu)}{g(\mu)} < \infty$.

Data Augmentation by Monte Carlo

Each substitution of

$$g_{k+1}(\mu) = \int \left[\int f_1(\mu|z)\, f_2(z|\acute{A})dz \right] g_k(\acute{A})d\acute{A}$$

is done the by Monte Carlo integration:

(a) "Imputation Step"

(a1) Draw $\mu_1, \ldots, \mu_M$ from $g_k(\mu)$

(a2) For each μ_i, draw $z_{ij} \sim f_2(z|\mu_i)$, $j = 1, \ldots, N$.

(b) "Posterior Step"

$$g_{k+1}(\mu) = \frac{1}{NM} \sum_{i=1}^{M} \sum_{j=1}^{N} f_1(\mu|z_{ij}) :$$

When the two functions, $g_k(\cdot)$ and $g_{k+1}(\cdot)$, are close enough, the procedure terminates.

2. How did I get interested in?

- **1994: Prof. Wing H. Wong (COPSS Award Winner 1993) was Head of Statistics, CUHK**
- **Tanner's 1st Ed. (3rd printing) was a reference in a course at HKU (3rd Ed. in 1996)**
- **Prof. Per Mykland from Chicago told me that the topic was the hottest in USA then**
- ***First* local conference ever by HKSS, mainly due to efforts of Prof. Wing Wong.**
- **My past interest in Fixed-Point Theorems:**

  (1969) "Generalizations of Some Fixed Point Theorems In Metric Spaces." M.Sc. Thesis, University of Alberta.

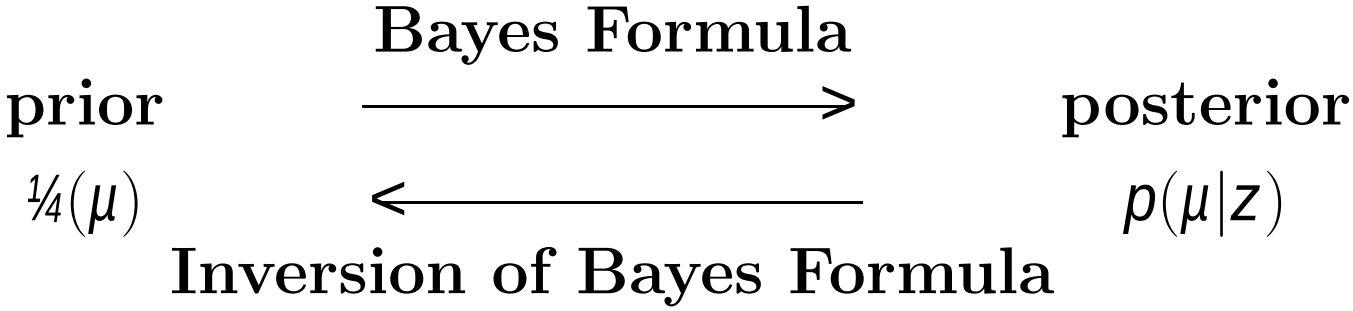
  (1970) "A Remark on Contractive Mappings", *Canadian Mathematical Bulletin* **13, 111–113.**

- **Aim:** *To replace the sufficient conditions of convergence by practicable ones.*
- **Approach: Suggest convenient distance measures, so that the integral operator is a contractive mapping on the space of densities in a particular application.**

3. Why to reverse Bayes' process?

In all Bayesian Missing Data Problems, where Z is the missing part, the aim is to find $\pi(\mu|y)$, given $p(\mu|y;z)$ & $f(z|y;\mu)$. Since y is a given constant throughout, like a parameter indexing the family of joint distributions for $(Z;\mu)$, we may drop y in all 3 densities in Tanner & Wong's integral equation. In other words, the aim is equivalent to reversing the Bayes process:

[Given Likelihood $f(z|\mu)$]



"*We Bayesians don't need IBF .*"

"*You've missed the forest for the trees.*"

中國俗語有云：

當局者迷 旁觀者清

Stakeholders are entrenched

Bystanders are enlightened

蘇東坡〈題西林壁〉：

橫看成嶺側成峰，
遠近高低各不同。
不識廬山真面目，
只緣身在此山中。

Being inside the mountains
Can't tell the outset for certain

4. Back to basics (1995)

So I went back to the basic identity

$$f_Z(z)/\pi(\mu) = f(z|\mu)/p(\mu|z) \qquad (1)$$

where $(\mu; z)$ is in joint support $\mathcal{S}(\mu; Z)$. For any μ where $\mathcal{S}(Z|\mu) = \mathcal{S}(Z)$, integration on both sides and re-arranging reciprocals give

$$\pi(\mu) = \left(\int_{\mathcal{S}(Z)} f(z|\mu)/p(\mu|z)\, dz \right)^{-1} \qquad (2)$$

which is called the point-wise IBF for μ. If there exists z_0 where $\mathcal{S}(\mu|z_0) = \mathcal{S}(\mu)$, we have

$$1/f_Z(z_0) = \int_{\mathcal{S}(\mu)} p(\mu|z_0)/f(z_0|\mu)\, d\mu; \qquad (3)$$

hence the function-wise IBF for all μ,

$$\pi(\mu) = \left(\int_{\mathcal{S}(\mu)} \frac{p(\mu|z_0)}{f(z_0|\mu)} d\mu \right)^{-1} \frac{p(\mu|z_0)}{f(z_0|\mu)}; \qquad (4)$$

which is the same for any such z_0.

Meaning to Tanner & Wong integral equ.

$$\pi(\mu|y) = \int p(\mu|y;z) \left[\int f(z|y;\acute{A})\pi(\acute{A}|y)d\acute{A} \right] dz \quad (*)$$

As $\mathcal{S}(Z|\mu) = \mathcal{S}(Z)\ \forall\mu$ and $\mathcal{S}(\mu|Z) = \mathcal{S}(\mu)\ \forall z_0$, it has explicit solution in dual forms,

$$\pi(\mu|y) = \left[\int \frac{f(z|y;\mu)}{p(\mu|y;z)}\, dz \right]^{-1} \quad (**)$$

$$= \left[\int \frac{p(\mu|y;z)}{f(z|y;\mu)}\, d\mu \right]^{-1} \frac{p(\mu|y;z)}{f(z|y;\mu)}; \quad (***)$$

where in the second form $(***)$, z can be any value of Z. Check: substitute $\pi(\acute{A}|y)$ into RHT of $(*)$ with the second form $(***)$, then the inside integral gives the predictive density $f_Z(z|y)$, and hence the outside integral returns $\pi(\mu|y)$.

"*All things are DIFFICULT*
before they are EASY."
Thomas Fuller : *Gnomologia*

Thomas Fuller in *Gnomologia*:

"All things are DIFFICULT before they are EASY."

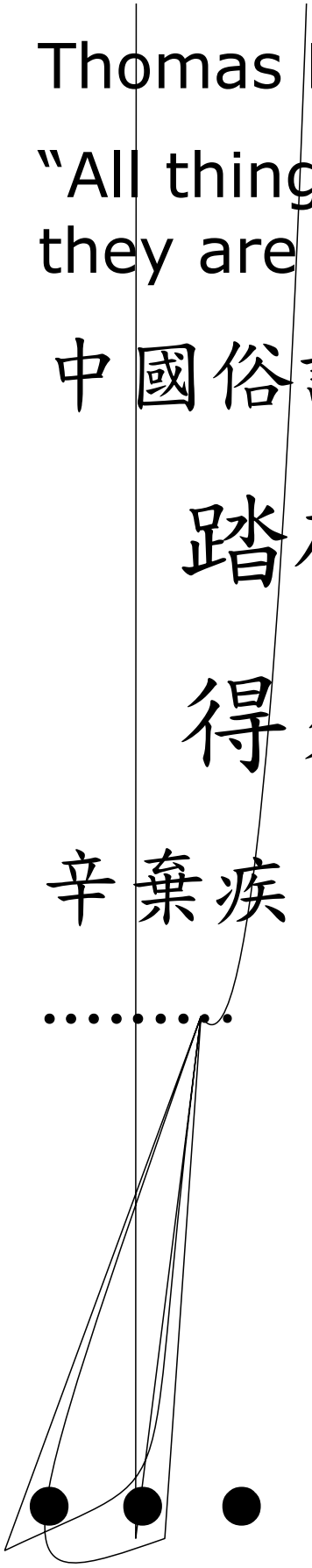
中國俗語有云：

踏破鐵鞋無覓處

得來全不費功夫

辛棄疾〈青玉案〉：

.....



Point-wise meaning of prior density

For a given value of μ,

(1) either $\mathcal{S}(Z|\mu) = \mathcal{S}(Z)$,
(2) or $\mathcal{S}(Z|\mu) \subset \mathcal{S}(Z)$, having unconditional probability $P_Z(\mathcal{S}(Z|\mu)) = ® < 1$.

Under repeated sampling of data given a μ, the prior density at the μ equals

- *the harmonic mean of posterior density at μ* **in case of (1)**
- *® times the harmonic mean of posterior density at μ* **in case of (2)**

G3TD6

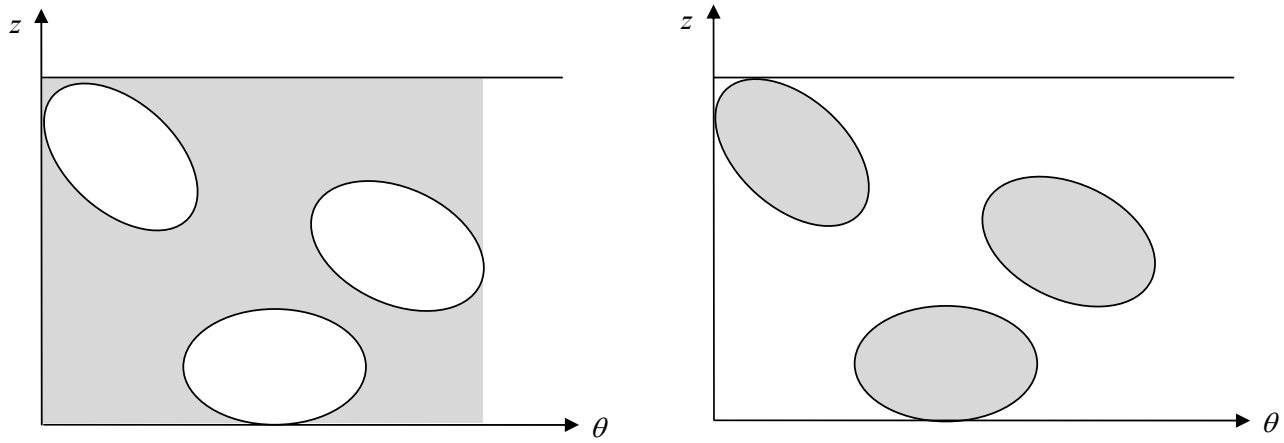
Function-wise meanings of prior density

$$\pi(\mu) = \left(\int_{\mathcal{S}(\mu)} \frac{p(\mu|z_0)}{f(z_0|\mu)} d\mu \right)^{-1} \frac{p(\mu|z_0)}{f(z_0|\mu)}; \quad (4)$$

- **No need for $\mathcal{S}(Z;\mu) = \mathcal{S}(Z) \times \mathcal{S}(\mu)$, only one z_0 such that $\mathcal{S}(\mu|z_0) = \mathcal{S}(\mu)$ can determine the whole function $\pi(\cdot)$.**
- **If numerical integration is needed, it's done only on $\mathcal{S}(\mu|z_0) = \mathcal{S}(\mu)$.**
- **Without the integral, sampling from $\pi(\cdot)$ can be done by Rejection Sampling, Adaptive Rejection Sampling (Gilks & Wild, 1992), Metropolis Sampling, Metropolis-Hastings Sampling, and Rubin's SIR method (1987,1988).**
- **It can assist the construction of highest density regions/intervals of μ**
- **It can serve as a benchmark for checking convergence of MCMC sampling by comparing $log(\pi(\cdot))$ with the simulated counterpart.**

5. Further back to basics (1996)

What about haphazard patterns of positivity (like the shaded areas below), where even the function-wise IBF cannot handle?



Bayes' Formula originated in <u>event form</u>, right? Go back further. <u>Other forms shall follow</u>.

Bayes' formula:

Let $\{H_1; H_2; \cdots; H_m\}$ and $\{A_1; A_2; \cdots; A_n\}$ be two distinct partitions of the sample space. For $j = 1; \cdots; m;\ i = 1; \cdots; n$,

$$P(H_j|A_i) = \frac{P(A_i|H_j)P(H_j)}{\sum_{k=1}^{m} P(A_i|H_k)P(H_k)}$$

Put all probabilities in a combined two-way table, where $P_{ij} = P(H_j|A_i)$, $L_{ij} = P(A_i|H_j)$, $p_j = P(H_j)$ and $q_i = P(A_i)$:

$P(H_j\|A_i)$ / $P(A_i\|H_j)$	H_1	$\cdots$	H_j	$\cdots$	H_m	$P(A_i)$
A_1	P_{11} / L_{11}	$\cdots$	P_{1j} / L_{1j}	$\cdots$	P_{1m} / L_{1m}	q_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_i	P_{i1} / L_{i1}	$\cdots$	P_{ij} / L_{ij}	$\cdots$	P_{im} / L_{im}	q_i
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_n	P_{n1} / L_{n1}	$\cdots$	P_{nj} / L_{nj}	$\cdots$	P_{nm} / L_{nm}	q_n
$P(H_j)$	p_1	$\cdots$	p_j	$\cdots$	p_m	1

IBF is to find the values on one margin (and as a result, the values on the other), given all pairs of values in the cells.

P_{ij} and L_{ij} are simultaneously both zero if and only if $A_i \cap H_j = \emptyset$ and both positive if and only if $A_i \cap H_j \neq \emptyset$. Thus the ratio $r_{ij} = P_{ij}/L_{ij}$ is well-defined in at least one cell in each row & column, resulting in a system of equations,

$$p_j/q_i = r_{ij} \equiv P_{ij}/L_{ij}; \quad i = 1; 2; \cdots; n; \ j = 1; \cdots; m:$$

for the $m + n - 2$ effective unknowns, p_j and q_i such that $\sum_{j=1}^{m} p_j = 1$ and $\sum_{i=1}^{n} q_i = 1$. The number of equations is just the number of defined r_{ij}. In view of the previous IBF, we need only

consider the case where r_{ij} is undefined in at least one cell in each row & column.

This can be done numerically, But we aim at an algebraic solution for inspiration to PDF setting. The underlying idea is the simple fact:

- **A set of positive numbers are uniquely determined by their relative proportions and their total. If only the proportions are known, but not the total, there are infinitely many solutions.**

In the first condition: the numbers are *completely proportionable*; **and in the second condition, the numbers are** *proportionable*.

The following are facts for $\{p_j\}$ (also for $\{q_i\}$):

(a) Given p_{j^*}; $(p_1/p_{j^*}; \cdots; p_{(j^*-1)}/p_{j^*}; 1; p_{(j^*+1)}/p_{j^*}; \cdots; p_m/p_{j^*})$, we have

$$p_j = \frac{p_j}{p_{j^*}} \left\{ \sum_{j=1}^{m} \frac{p_j}{p_{j^*}} \right\}^{-1}; \qquad j = 1; 2; \cdots; m.$$

(b) Given all the consecutive ratios $(p_1/p_2; p_2/p_3; \cdots; p_{m-1}/p_m)$, we can obtain the proportions vs. a common denominator by chained multiplications; e.g.,

$$\frac{p_1}{p_m} = \prod_{j=1}^{m-1} \frac{p_j}{p_{j+1}}; \; \frac{p_2}{p_m} = \prod_{j=2}^{m-1} \frac{p_j}{p_{j+1}}; \cdots; \frac{p_{m-1}}{p_m} = \prod_{j=m-1}^{m-1} \frac{p_j}{p_{j+1}}.$$

(c) If in a particular row i, a subset of the ratios, $r_{ij_1}; r_{ij_2}; \cdots; r_{ij_k}$, are defined, then $p_{j_1}; p_{j_2}; \cdots$, and p_{j_k} are proportionable in row i; for example, the proportions relative to p_{j_k} are:

$$\frac{p_{j_1}}{p_{j_k}} = \frac{r_{ij_1}}{r_{ij_k}}; \frac{p_{j_2}}{p_{j_k}} = \frac{r_{ij_2}}{r_{ij_k}}; \cdots; \frac{p_{j_{k-1}}}{p_{j_k}} = \frac{r_{ij_{k-1}}}{r_{ij_k}}.$$

Notation: $[p_{j_1}(i)p_{j_2}(i)\cdots(i)p_{j_k}]$; $[j_1(i)j_2(i)\cdots(i)j_k]$.

Conclusions:

- **Let Let $(p_{j_1}; p_{j_2}; \cdots; p_{j_m})$ be any permutation of $\{p_j\}$. If p_{j_1} and p_{j_2} are proportionable in row i_1, p_{j_2} and p_{j_3} proportionable in row $i_2; \cdots; p_{j_{(m-1)}}$ and p_{j_m} proportionable in row $i_{(m-1)}$, or in notation,**

$$[p_{j_1}(i_1)p_{j_2}(i_2)p_3(i_3)\cdots p_{j_{(m-1)}}(i_{(m-1)})p_{j_m}];$$

  then $\{p_j\}$ are completely proportionable.

- **The solution for $\{p_j\}$ and $\{q_i\}$ is not unique if the following situation happens for both $\{p_j\}$ and $\{q_i\}$:**
  There are two or more subsets of $\{p_j\}$ (or $\{q_i\}$) whose union equals the whole set of $\{p_j\}$ (or $\{q_i\}$) and which satisfy two conditions: (i) each member of one subset is found not proportionable with any member of another subset and (ii) members within the same subset are proportionable unless the subset is a singleton.

Example (IBF for haphazard patterns)

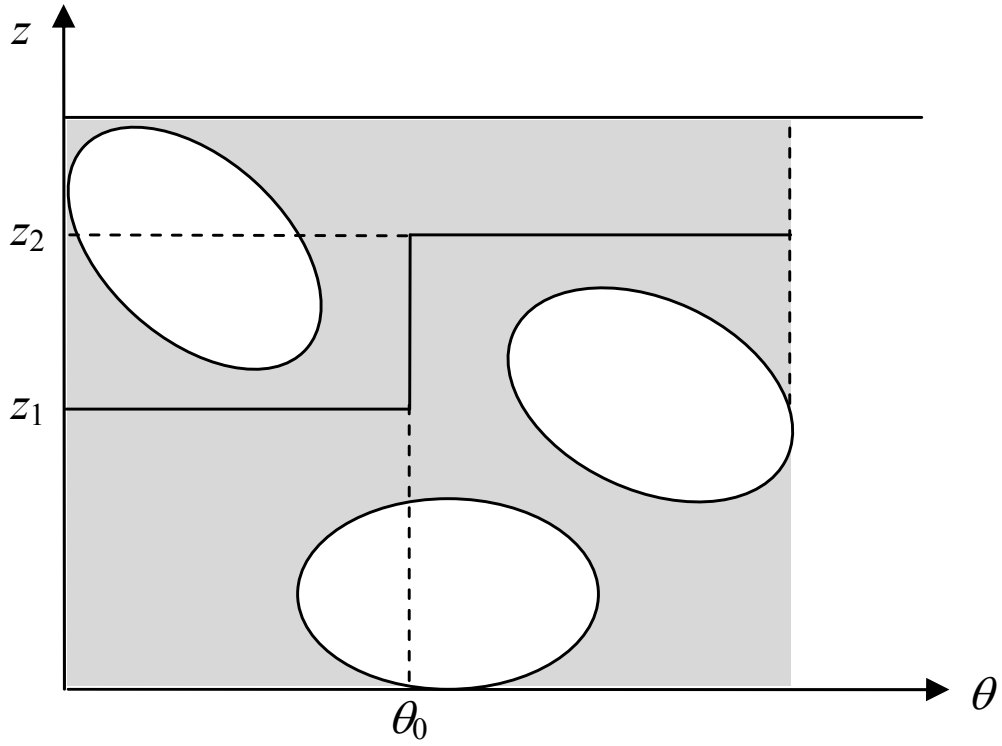
Only well-defined $r_{ij} = P_{ij}/L_{ij}$ are shown. We need only demonstrate finding $\{p_j\}$.

	p_1	p_2	p_3	p_4	p_5	p_6	Notation	Available proportions
q_1				r_{14}		r_{16}	[4(1)6]	$p_6/p_4 = r_{16}/r_{14}$
q_2		r_{22}	r_{23}		r_{25}		[2(2)3(2)5]	$p_2/p_5 = r_{22}/r_{25}$,
q_3						r_{36}		$p_3/p_5 = r_{23}/r_{25}$
q_4	r_{41}				r_{45}		[1(4)5]	$p_1/p_5 = r_{41}/r_{45}$
q_5			r_{53}					
q_6		r_{62}			r_{65}		[2(6)5]	**Done in row 2**
q_7	r_{71}			r_{74}			[1(7)4]	$p_4/p_5 = (r_{74}/r_{71})(p_1/p_5)$;
								$p_6/p_5 = (p_6/p_4)(p_4/p_5)$.

Note: Consider the table without the last row. Then $\{p_2; p_3; p_5; p_1\}$ and $\{p_4; p_6\}$ are two distinct proportionable subsets, each having no member proportionable with any member of the other subset. There is one solution for $\{p_j\}$ associated with each a where $0 < a < 1$:

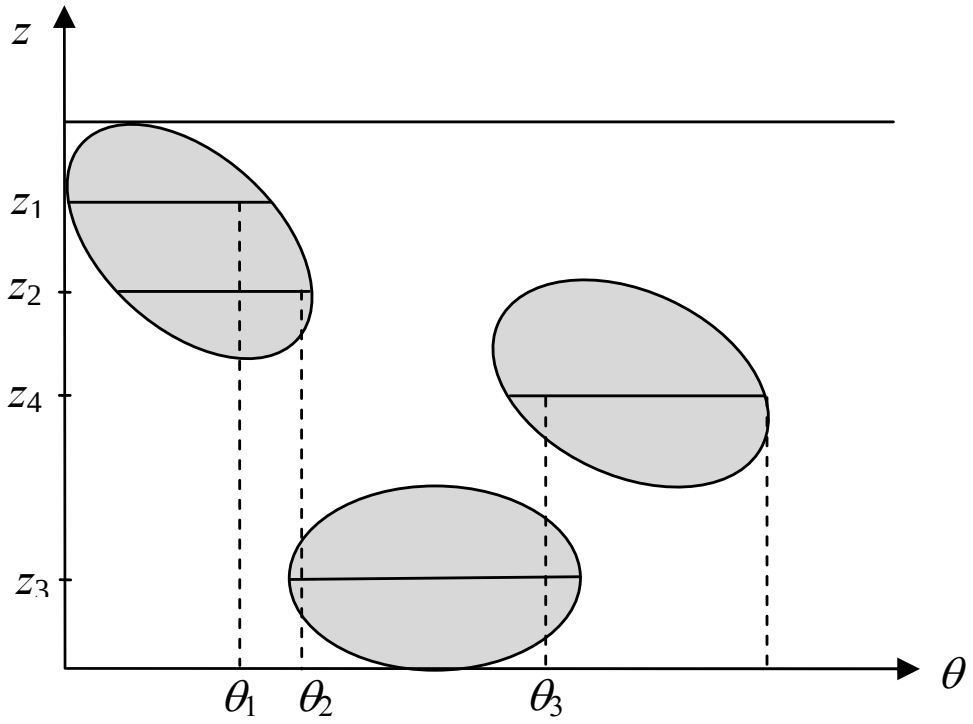
$$p_2 + p_3 + p_5 + p_1 = a; \quad p_4 + p_6 = 1 - a:$$

A similar situation for $\{q_2; q_4; q_5; q_1\}$ and $\{q_1; q_3\}$. Hence IBF has infinitely many solutions here.



Let $r(\mu; z) = p(\mu|z)/f(z|\mu)$.

$\forall \mu \in \mathcal{S}(\mu|z_1) : \pi(\mu)/\pi(\mu_0) = r(\mu; z_1)/r(\mu_0$



Let $Q(\mu;\mu_j;z_i) = r(\mu;z_i)/r(\mu_j;z_i)$**. Then**

$\forall \mu \in \mathcal{S}(\mu|z_i); i = 1;2 : \pi(\mu)/\pi(\mu_1) = Q(\mu;\mu_1;z_i)$:

$\forall \mu \in \mathcal{S}(\mu|z_i); i = 2;3 : \pi(\mu)/\pi(\mu_2) = Q(\mu;\mu_2;z_i)$:

So $\forall \mu \in \mathcal{S}(\mu|z_3)$:

$\frac{\pi(\mu)}{\pi(\mu_1)} = \frac{\pi(\mu)}{\pi(\mu_2)} \times \frac{\pi(\mu_2)}{\pi(\mu_1)} = Q(\mu;\mu_2;z_3) \times [Q(\mu;\mu_1;z_2)/Q(\mu;\mu_2;z_2)]$:

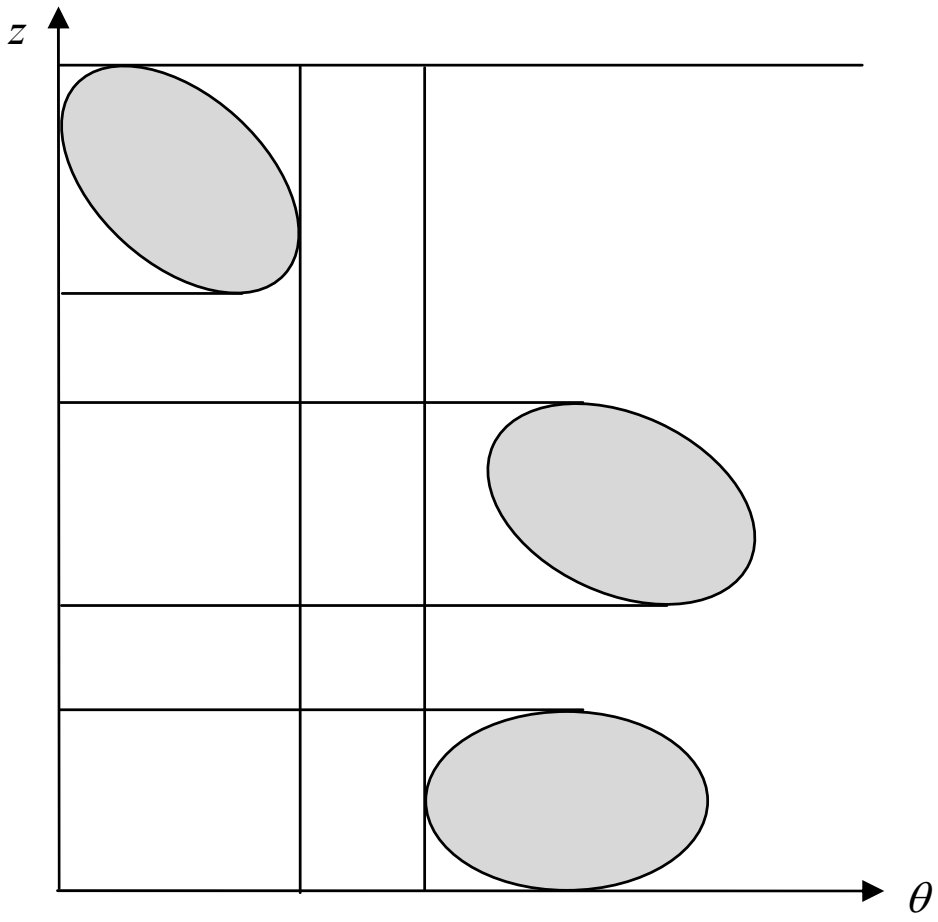
$\forall \mu \in \mathcal{S}(\mu|z_i); i = 3;4 : \pi(\mu)/\pi(\mu_3) = Q(\mu;\mu_3;z_i)$:

And $\forall \mu \in \mathcal{S}(\mu|z_4)$: $\frac{\pi(\mu)}{\pi(\mu_1)} = \frac{\pi(\mu)}{\pi(\mu_3)} \times \frac{\pi(\mu_3)}{\pi(\mu_2)} \times \frac{\pi(\mu_2)}{\pi(\mu_1)}$

$= Q(\mu;\mu_3;z_4) \times \frac{Q(\mu;\mu_2;z_3)}{Q(\mu;\mu_3;z_3)} \times \frac{Q(\mu;\mu_1;z_2)}{Q(\mu;\mu_2;z_2)}$:

Hence we need only find $\pi(\mu_1)$**, which can be obtained by integrating** $\pi(\mu)/\pi(\mu_1)$ **over the 4 subsets of conditional supports,** $\mathcal{S}^*(\mu|z_i) \subset \mathcal{S}(\mu|z_i)$ **so that** $\cup_{i=1}^{4} \mathcal{S}^*(\mu|z_i) = \mathcal{S}(\mu)$**:**

$$\frac{1}{\pi(\mu_1)} = \int_{\mathcal{S}^*(\mu|z_1)} + \int_{\mathcal{S}^*(\mu|z_2)} + \int_{\mathcal{S}^*(\mu|z_3)} + \int_{\mathcal{S}^*(\mu|z_4)} :$$



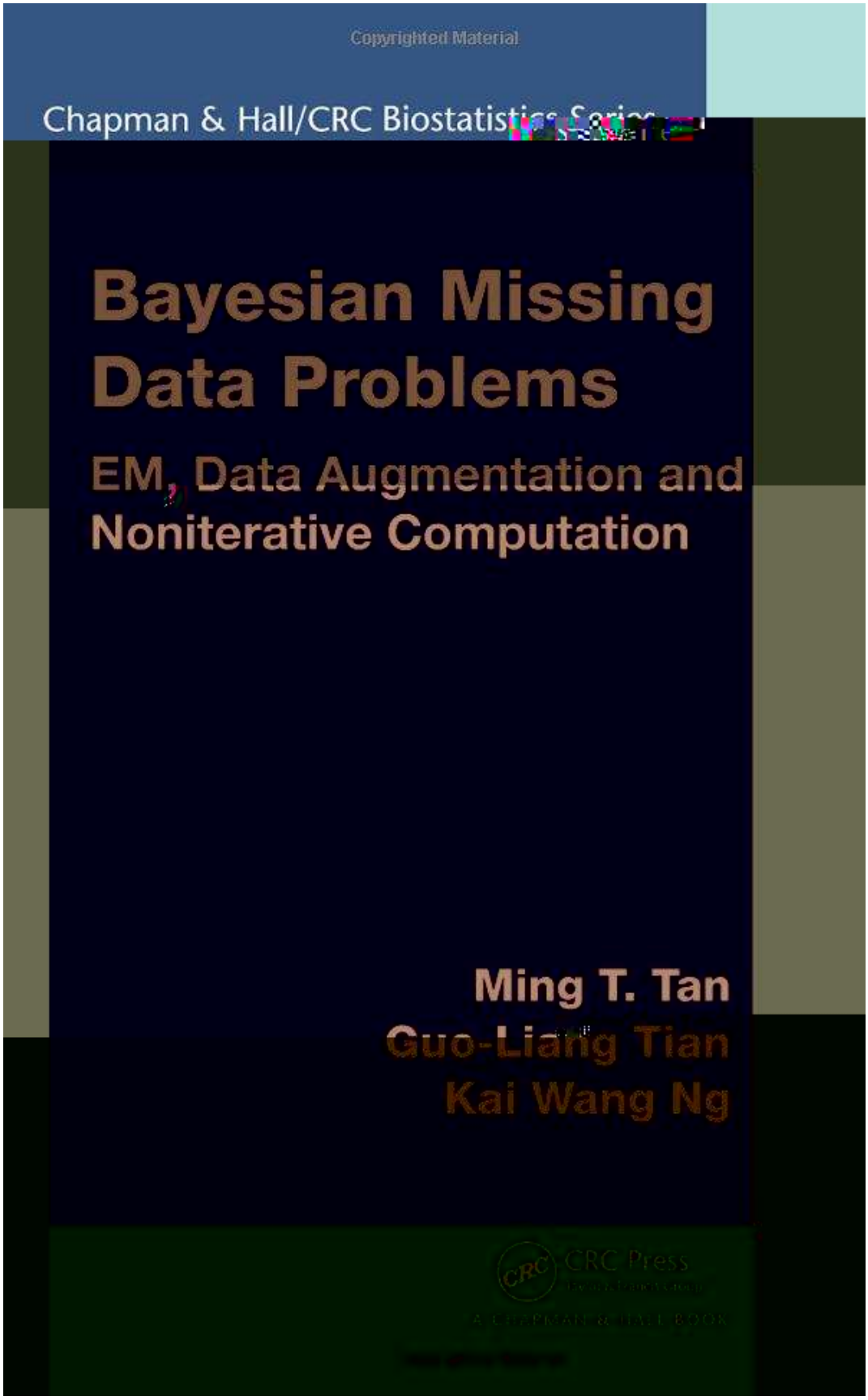
The figure is obtained from the previous one by moving the three constituent regions of joint support away from each other, so that they are no longer *projection-connected*. In this case there are infinitely many IBF solutions. Each solution corresponds to a possible set of unconditional probabilities on the 3 regions, the information about which cannot be recovered from the two families of conditional PDF.

Many thanks to those who quietly showed encouragements and support in the last 15 years and to those who drew my attention to the story of The Emperor's New Coat and to the following sayings:

"All great truths begin as blasphemies."
– George Bernard Shaw

"It requires a very unusual mind to undertake the analysis of the obvious."
– Alfred N. Whitehead

"A scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die and a new generation grows up that is familiar with it."
– Max Planck


Copyrighted Material
Chapman & Hall/CRC Biostatistics Series
Bayesian Missing Data Problems
EM, Data Augmentation and Noniterative Computation
Ming T. Tan
Guo-Liang Tian
Kai Wang Ng
CRC Press

Statistics

Bayesian Missing Data Problems: EM, Data Augmentation and Noniterative Computation presents solutions to missing data problems through explicit or noniterative sampling calculation of Bayesian posteriors. The methods are based on the inverse Bayes formulae discovered by one of the authors in 1995. Applying the Bayesian approach to important real-world problems, the authors focus on exact numerical solutions, a conditional sampling approach via data augmentation, and a noniterative sampling approach via EM-type algorithms.

After introducing the missing data problems, Bayesian approach, and posterior computation, the book succinctly describes EM-ty