Improper mixtures and Bayes's theorem

Peter McCullagh and Han Han

Department of Statistics
University of Chicago

DASF-III conference
Toronto, March 2010

Outline

1. Bayes's theorem
    - Improper mixtures
2. Poisson processes
    - Bayes's theorem for PP
    - Gaussian sequences

Bayes's theorem

Non-Bayesian model: Domain $\mathcal{S} = \mathcal{R}^\infty$

Parameter space Θ; elements $\theta \in \Theta$:

$\{P_\theta : \theta \in \Theta\}$ family of prob distributions on $\mathcal{S}$

Observation space $\mathcal{S}_n = \mathcal{R}^n$; event $A \subset \mathcal{S}$; elements $y \in \mathcal{S}_n$

Bayesian model: above structure plus

π: prob distn on Θ implies

$P_\pi(A \times d\theta) = P_\theta(A)\pi(d\theta)$ joint distn on $\mathcal{S} \times \Theta$

$Q_\pi(A) = P_\pi(A \times \Theta)$ mixture distribution on $\mathcal{S}$

Bayes's theorem: conditional probability given $Y = y$

associates with each $y \in \mathcal{S}_n$ a probability distribution on Θ

$y \mapsto P_\pi(d\theta \,|\, Y = y) = P_\pi(dy \times d\theta)/Q_\pi(dy)$

Bayes/non-Bayes remarks

Non-Bayesian model:
family of distributions $\{P_\theta\}$ on $\mathcal{S}$

Bayesian model is a *single* distribution/process:
Joint distribution $P_\theta(A)\,\pi(d\theta)$ on $\mathcal{S} \times \Theta$, or
Mixture distribution $P_\pi(A)$ on $\mathcal{S}$

Parametric inference:
Use the joint density to get a posterior distn $P_\pi(d\theta \,|\, y)$
e.g. $P_\pi(2.3 < \theta_1 < 3.7 \,|\, y)$

Nonparametric inference (sample-space inference):
$\mathcal{S} = \mathcal{R}^\infty = \mathcal{R}^n \times \mathcal{R}^\infty$: $\quad Y^{(n)} : \mathcal{R}^\infty \to \mathcal{R}^n$
obs $y^{(n)} \in \mathcal{R}^n \mapsto Q_\pi(A \,|\, y^{(n)})$ for $A \subset \mathcal{S}$
e.g. $Q_\pi(Y_{n+1} < 3.7 \,|\, y^{(n)})$ or $Q_\pi(2.3 < \bar{Y}_\infty < 3.7 \,|\, y^{(n)})$

Improper mixtures: $\nu(\Theta) = \infty$

$P_\theta(A)\nu(d\theta)$ not a probability distribution on $\mathcal{S} \times \Theta$
— No theorem of conditional probability

Nonetheless ....
If $Q_\nu(dy) = \int P_\theta(dy;\theta)\,\nu(d\theta) < \infty$, the formal Bayes ratio $P_\theta(dy)\,\nu(d\theta)/Q_\nu(dy)$ is a probability distribution on Θ

Distinction: Bayes *calculus* versus Bayes's *theorem*

If there is a theorem here, what is its nature?
(i) conditional distn is associated with some y-values and not others
(ii) what about DSZ marginalization paradoxes?

Marginalization paradoxes (Dawid, Stone and Zidek)

Exponential ratio model:

$Y \sim \phi e^{-\phi y}, \quad X \sim \theta\phi e^{-\theta\phi x}$ indep

Parameter of interest: $\theta = E(Y)/E(X)$.

Prior: $\pi(\theta) d\theta\, d\phi$

Analysis I:

Joint density: $\theta\phi^2 e^{-\phi(\theta x + y)}\, dx\, dy$

Marginal posterior: $\pi(\theta \mid x, y) \propto \frac{\theta\,\pi(\theta)}{(\theta+z)^3}$

where $z = y/x$.

Analysis II: based on Z alone

$p(z \mid \theta) = \frac{\theta}{(\theta+z)^2} \qquad \pi(\theta \mid z) \propto \frac{\theta\,\pi(\theta)}{(\theta+z)^2}$

Apparent contradiction or paradox

Prior $\pi(\theta) d\theta\, d\phi/\phi$ gives same answer both ways.

What does Bayes's theorem do?

Conventional proper Bayes:
Begins with the family $\{P_\theta : \theta \in \Theta\}$ and $\pi(d\theta)$
Creates a random element (Y, T) in $\mathcal{S} \times \Theta$
with distribution $P_\theta(dy)\,\pi(d\theta)$
Computes the conditional distribution given $Y = y$

Can we do something similar with an improper mixture ν?

(i) Associate with the family $\{P_\theta\}$ and measure ν some sort of random object in $\mathcal{S} \times \Theta$
(ii) Observe a piece of this object, (projection onto $\mathcal{S}$ or $\mathcal{S}_n$)
(iii) Compute the conditional distribution given the observation

What sort of random object?

Improper mixtures and Bayes's theorem

(i) Bayes's theorem is just conditional probability;
joint distribution on $\mathcal{S} \times \Theta \mapsto$ conditional distribution

(ii) Bayes's theorem needs joint probability distribution (Lindley)
— but not necessarily on $\mathcal{S} \times \Theta$

(iii) Poisson process converts a measure ν on Θ into
a prob distn on the power set $\mathrm{Pow}(\Theta)$

(iv) Prob distn π generates a random element $T \in \Theta$
measure ν generates a random subset $T \subset \Theta$

(v) Sampling: how do we observe a random set?

(vi) Can Bayes's theorem now be used?

Poisson process in $\mathcal{S} = \mathcal{S}_0 \times \mathcal{S}_1$

Domain $\mathcal{S}$, measure space with measure μ
Countability condition (Kingman 1993)

$$\mu = \sum_{n=1}^{\infty} \mu_n \qquad \mu_n(\mathcal{S}) < \infty.$$

$Z \subset \mathcal{S}$ a Poisson process with mean measure μ: $Z \sim \mathrm{PP}(\mu)$
$\#(Z \cap A) \sim \mathrm{Po}(\mu(A))$ independently for $A \cap A' = \emptyset$

Product structure $\mathcal{S} = \mathcal{S}_0 \times \mathcal{S}_1$ gives

$$Z = (Y, X) = \{(Y_i, X_i) : i = 1, 2, \ldots\} \quad (Y_i \in \mathcal{S}_0, \quad X_i \in \mathcal{S}_1)$$

Projection $Y = Z[n] \subset \mathcal{S}_0$ is $\mathrm{PP}(\mu_0)$ where $\mu_0(A) = \mu(A \times S_1)$

Observation on a point process: $\mathcal{S} = \mathcal{R}^\infty$

Point process $Z \subset \mathcal{R}^\infty$: countable set of infinite sequences

$$Z_1 = (\overbrace{Z_{11}, Z_{12}, \ldots, Z_{1n}}^{Y_1 = Z_1[n]}, Z_{1,n+1}, \ldots)$$

$$Z_m = (\overbrace{Z_{m1}, Z_{m2}, \ldots, Z_{mn}}^{Y_m = Z_m[n]}, Z_{m,n+1}, \ldots)$$

$Z \subset \mathcal{R}^\infty \sim \mathrm{PP}(\mu)$; $\quad Y = Z[n] \subset \mathcal{R}^n$; $\quad Y \sim \mathrm{PP}(\mu_0)$

Sampling region $A \subset \mathcal{R}^n$ such that $\mu_0(A) = \mu(A \times \mathcal{R}^\infty) < \infty$;

Observation $\mathbf{y} = Y \cap A$; $\quad \#\mathbf{y} < \infty$

Inference for sequences $Z[A] = \{Z_i : Y_i \in A\}$

Observation on a point process: $\mathcal{S} = \mathcal{R}^\infty$

Point process $Z \subset \mathcal{R}^\infty$: countable set of infinite sequences
PP events: $Z = \{Z_1, Z_2, \ldots\}$
one PP event $Z_i = (Z_{i1}, Z_{i2}, \ldots)$ an infinite sequence
$Z_i = (Y_i, X_i)$: $Y_i = (Z_{i1}, \ldots, Z_{in})$; $X_i = (Z_{i,n+1}, Z_{i,n+2}, \ldots)$
$Y_i = Z_i[n]$ initial segment of Z_i; X_i subsequent trajectory

Observation space $\mathcal{S}_0 = \mathcal{R}^n$:
Sampling protocol: test set $A \subset \mathcal{S}_0$ such that $\mu_0(A) < \infty$
Observation: $Y \cap A$ a *finite* subset of $\mathcal{S}_0$

Inference for what?
for the subsequent trajectories $X[A] = \{X_i : Y_i \in A\}$, if any.

Bayes's theorem for PPP

Test set $A \subset \mathcal{R}^n$ such that $\mu_0(A) = \mu(A \times \mathcal{R}^\infty) < \infty$
Observation $\mathbf{y} = Y \cap A \subset \mathcal{R}^n$;
Subsequent trajectories $\mathbf{x} = X[A] = \{X_i : Y_i \in A\}$

(i) Finiteness: $\mu_0(A) < \infty$ implies $\#\mathbf{y} < \infty$ w.p.1
(ii) Trivial case: If $\mathbf{y}$ is empty $\mathbf{x} = \emptyset$
(iii) Assume $\mu_0(A) > 0$ and $m = \#\mathbf{y} > 0$
(iv) Label the events $Y_1, \ldots, Y_m$ independently of Z.
(v) Given m, $Y_1, \ldots, Y_m$ are iid $\mu_0(dy)/\mu_0(A)$
(vi) $(Y_1, X_1), \ldots, (Y_m, X_m)$ are iid with density $\mu(dx, dy)/\mu_0(A)$
(vii) Conditional distribution

$$p(d\mathbf{x} \mid \mathbf{y}) = \prod_{i=1}^{m} \frac{\mu(dx_i\, dy_i)}{\mu_0(dy_i)} = \prod_{i=1}^{n} \mu(dx_i \mid y_i)$$

Remarks on the conditional distribution

$$p(d\mathbf{x} \mid \mathbf{y}) = \prod_{i=1}^{m} \frac{\mu(dx_i\, dy_i)}{\mu_0(dy_i)} = \prod_{i=1}^{n} \mu(dx_i \mid y_i)$$

(i) Finiteness assumption: $\mu_0(A) < \infty$
given $\#(Y \cap A) = m < \infty$, the values $Y_1, \ldots, Y_m$ are iid

(ii) Conditional independence of trajectories:
$X_1, \ldots, X_m$ are conditionally independent given $Y \cap A = \mathbf{y}$

(iii) Lack of interference:
Conditional distn of X_i given $Y \cap A = \mathbf{y}$ depends only on y_i
– unaffected by m or by configuration of other events

(iv) Role of test set A
no guarantee that a test set exists such that $0 < \mu_0(A) < \infty$!
if $y \in \mathcal{S}_0$ has a nbd A s.t. $0 < \mu_0(A) < \infty$ then the test set has no effect.

Improper parametric mixtures

$\mathcal{S} = \mathcal{R}^n \times \Theta$ product space
$\{P_\theta(dy) \colon \theta \in \Theta\}$ a family of prob distns
$\nu(d\theta)$ a countable measure on Θ: $\nu(\Theta) = \infty$
$\Rightarrow \mu = P_\theta(dy)\nu(d\theta)$ countable on $\mathcal{S} = \mathcal{R}^n \times \Theta$
$\mu_0(A) = \mu(A \times \Theta) = \int_\Theta P_\theta(A)\, \nu(d\theta)$ on $\mathcal{R}^n$

The process:
$Z \sim \mathrm{PP}(\mu)$ a random subset of $\mathcal{S}$
$Z = \{(Y_1, X_1), (Y_2, X_2), \ldots\}$ (countability)
$Y \sim \mathrm{PP}(\mu_0)$ in $\mathcal{R}^n$ and $X \sim \mathrm{PP}(\nu)$ in Θ

Infinite number of sequences $Y \subset \mathcal{R}^n$
one parameter $X_i \in \Theta$ for each $Y_i \in Y$

Improper parameteric mixture (contd.)

Observation:

a test set $A \subset \mathcal{R}^n$ such that $\mu_0(A) < \infty$
the subset $\mathbf{y} = Y \cap A$ (finite but could be empty)
but $\#\mathbf{y} > 0$ implies $\mu_0(A) > 0$

The inferential goal:

$X[A] : Y_i in A$ a finite random subset of Θ

Elements (parameters) in $X[A]$ are conditionally independent with distribution $\nu(d\theta)P_\theta(dy)/\mu_0(dy)$

Vindication of the formal Bayes calculus!

Summary of assumptions

The Poisson process:

Countability: $\nu(A) = \sum_{j=0}^{\infty} \nu_j(A) \quad (\nu_j(\mathcal{S}) < \infty)$;

includes nearly every imaginable improper mixture!

implies that μ_0 is countable on $\mathcal{S}_0 = \mathcal{R}^n$

σ-finiteness, local finiteness,.. sufficient but not needed

Observation space and sampling protocol:

need $\mathcal{S}_0$ and $A \subset \mathcal{S}_0$ such that $\mu_0(A) < \infty$

— not guaranteed by countability condition

— may be satisfied even if μ_0 not σ-finite

— may require $n \geq 2$ or $n \geq 3$

— may exclude certain points s.t. $\mu_0(\{y\}) = \infty$

Gaussian sequences: parametric formulation

$\mathcal{S} = \mathcal{R}^n \times \Theta, \qquad P_{\theta,\sigma}$ iid $N(\theta, \sigma^2)$
$\nu(d\theta) = d\theta\, d\sigma/\sigma^p$ on $\mathcal{R} \times \mathcal{R}^+$ (improper on Θ)
$\mu(dy\, d\theta) = N_n(\theta, \sigma^2)(dy)\, d\theta\, d\sigma/\sigma$ (joint measure on $\mathcal{S}$)
$Z \subset \mathcal{R}^n \times \Theta$ is a PP with mean measure μ
Marginal process $Y \subset \mathcal{R}^n$ is Poisson with mean measure

$$\mu_0(dy) = \frac{\Gamma((n+p-2)/2)2^{(p-3)/2}\pi^{-(n-1)/2}n^{-1/2}\, dy}{(\sum_{i=1}^n (y_i - \bar{y})^2)^{(n+p-2)/2}}.$$

Test sets $A \subset \mathcal{R}^n$ such that $0 < \mu_0(A) < \infty$

does not exist unless $n \geq 2$ and $n > 2 - p$

For each test set A, finite subset $\mathbf{y} \subset A$ and for each $y \in \mathbf{y}$

$$p(\theta, \sigma \mid \mathbf{y}, y \in \mathbf{y}) = \phi_n(dy; \theta, \sigma)\sigma^{-p}/\mu_0(dy)$$

Formal Bayes inferential statements

Given a proper mixture π on Θ, what does Bayes's theorem do?
associates with each integer $n \geq 0$, and almost every $y \in \mathcal{R}^n$ a distribution

$$P_n(d\theta\, d\sigma \,|\, y) \propto \phi_n(y; \theta, \sigma)\, \pi(d\theta\, d\sigma)$$

This holds in particular for $n = 0$ and $y = 0$ in $\mathcal{R}^0$.

Given an improper mixture ν on Θ, Bayes's theorem
associates with each test set $A \subset \mathcal{R}^n$, with each finite subset $\mathbf{y} \subset A$, and with almost every $y \in \mathbf{y}$

$$P_n(\theta, \sigma \,|\, \mathbf{y}, y \in \mathbf{y}) = \phi_n(dy; \theta, \sigma)\sigma^{-p}/\mu_0(dy)$$

independently for $y_1, \ldots$ in $\mathbf{y}$.

The first statement is not correct for improper mixtures.

Nonparametric version I

$T \subset \mathcal{R} \times \mathcal{R}^+$ Poisson with mean measure $d\theta\, d\sigma/\sigma^p$
To each $t = (t_1, t_2)$ in T associate an iid $N(t_1, t_2^2)$ sequence Z_t
The set $Z \subset \mathcal{R}^\infty$ of sequences $Z \sim \mathrm{PP}(\mu)$

$$\mu_n(dz) = \frac{\Gamma((n+p-2)/2) 2^{(p-3)/2} \pi^{-(n-1)/2} n^{-1/2}\, dz}{(\sum_{i=1}^n (z_i - \bar{z}_n)^2)^{(n+p-2)/2}}$$

such that $\mu_n(A) = \mu_{n+1}(A \times \mathcal{R})$.
Projection: $Z[n] \sim \mathrm{PP}(\mu_n)$ in $\mathcal{R}^n$.

Observation: test set A plus $Z[n] \cap A = \mathbf{z}$
For $z \in \mathbf{z}$ the subsequent trajectory $z_{n+1}, \ldots$ is distributed as

$$\mu_{n+k}(z_1, \ldots, z_n, z_{n+1}, \ldots, z_{n+k})/\mu_n(z)$$

exchangeable Student t on $n+p-2$ d.f.

Nonparametric version II

$Y[2] \subset \mathcal{R}^2$ is Poisson with mean measure

$$dy_1\, dy_2/|y_1 - y_2|^p$$

Extend each $y \in Y[2]$ by the Gosset rule

$$y_{n+1} = \bar{y}_n + s_n \epsilon_n \sqrt{(n^2 - 1)/(n(n+p-2))}$$

where $\epsilon_n \sim t_{n+p-2}$ indep.

Then $Y \subset \mathcal{R}^\infty$ is the same point process as Z
$Y \sim Z \sim \mathrm{PP}(\mu)$ in $\mathcal{R}^\infty$.

Bernoulli sequences

$Y_1, \ldots, Y_n, \ldots$ iid Bernoulli(θ)
Take $\mathcal{S}_0 = \{0, 1\}^n$ as observation space
$\nu(d\theta) = d\theta/(\theta(1-\theta))$.

Product measure $\mu(y, d\theta) = d\theta\, \theta^{n_1(y)-1}(1-\theta)^{n_0(y)-1}$
Marginal measure on $\{0, 1\}^n$

$$\mu_0(y) = \begin{cases} \Gamma(n_0(y))\Gamma(n_1(y))/\Gamma(n) & n_0(y), n_1(y) > 0 \\ \infty & y = 0^n \text{ or } 1^n \end{cases}$$

Test set $A \subset \{0, 1\}^n$ excludes $0^n, 1^n$, so $n \geq 2$

$$P_\nu(\theta \mid Y \cap A = \mathbf{y}, y \in \mathbf{y}) = \text{Beta}_{n_1(y), n_0(y)}(\theta)$$

Marginalization paradox revisited

Exponential ratio model:
$Y \sim \phi e^{-\phi y}$, $\quad X \sim \theta\phi e^{-\theta\phi x}$ indep $\quad$ Prior: $\pi(\theta) d\theta\, d\phi$

Joint measure $\phi^2 \theta e^{-\phi(y+\theta x)} \pi(\theta)$ on $\mathcal{R}^2 \times \Theta$
Bivariate marginal measure has a density in $\mathcal{R}^2$

$$\lambda(x, y) = 2 \int_0^\infty \frac{\theta\pi(\theta)\, d\theta}{(\theta x + y)^3}$$

Locally finite, so observable on small test sets $A \subset \mathcal{R}^2$

Bayes's PP theorem gives

$$p(\theta \mid Y \cap A = \mathbf{y},\ (x, y) \in \mathbf{y}) \propto \frac{\theta\, \pi(\theta)}{(\theta + z)^3}$$

where $z = y/x$.
Correct standard conclusion derived from the Bayes calculus.

Marginalization paradox contd.

Conclusion depends on $z = y/x$ alone
Induced marginal measure on $\mathcal{R}^+$ for the ratios $z = y/x$

$$\Lambda_Z(A) = \begin{cases} 0 & \mathrm{Leb}(A) = 0 \\ \infty & \mathrm{Leb}(A) > 0 \end{cases}$$

No test set such that $0 < \Lambda(A) < \infty$
Bayes PP theorem does not support the formal Bayes calculus

Could adjust the mixture measure: $\pi(\theta)\, d\theta\, d\phi/\phi$
Two versions of Bayes calculus give $\theta\pi(\theta)/(z+\theta)^2$

But there is no PP theorem to support version II

Conclusions

(i) Is there a version of Bayes's theorem for improper mixtures?
— Yes.

(ii) Is it the same theorem?
— No. Sampling scheme is different
— Conclusions are of a different structure

(iii) Are the conclusions compatible with the formal Bayes calculus?
— To a certain extent, yes.

(iv) How do the conclusions differ from proper Bayes?
— Nature of the sampling schemes:
proper Bayes: sample $\{1, 2, \ldots, n\}$, observation $Y \in \mathcal{R}^n$
improper Bayes: sample $A \subset \mathcal{R}^n$, observation $Y \cap A$
— Finiteness condition on sampling region A:
— usually $A \subset \mathcal{R}^n$ does not exist unless $n \geq k$

(v) Admissibility of estimates?