Bayes Methods

Mock 2 2018
1. A field is divided into an \( m \times n \) grid of square cells. The number \( y_{i,j} \) of buttercups in each cell \( (i, j) \) is counted. The observation model is

\[
y_{i,j} \sim \text{Poisson}(\mu_{i,j}), \ i = 1, \ldots, m, \ j = 1, \ldots, n.
\]

Suppose we parameterise \( \mu = \phi \psi^T \) as \( m \times n \) outer product, where \( \phi = (\phi_1, \ldots, \phi_m)^T, \phi \in [0, \infty)^m \) and \( \psi = (\psi_1, \ldots, \psi_n)^T, \psi \in [0, \infty)^n \) are column vectors of positive parameters.

Thinking of the cells in the matrix \( \mu \) as a lattice, denote by \( N_{i,j} \) the set of labels \( (a, b) \), \( a \in 1 : m, \ b \in 1 : n \) for cells adjoining cell \( (i, j) \) (interior cells have 4 neighbors, and boundary cells 3, except at the corners where each cell has 2 neighbors). Consider the joint prior density

\[
\pi(\phi, \psi) \propto \exp(-\alpha \bar{\phi} - \beta \bar{\psi}) \exp(-\theta \sum_{i=1}^m \sum_{j=1}^n \sum_{(a,b) \in N_{i,j}} |\phi_a \psi_b - \phi_i \psi_j|),
\]

where \( \bar{\phi} = m^{-1} \sum_{i=1}^m \phi_i, \bar{\psi} = n^{-1} \sum_{j=1}^n \psi_j \) and \( \alpha, \beta \) and \( \theta \) are fixed positive prior hyperparameters.

(a)

(i) Show that \( \pi(\phi, \psi) \) is a proper prior.

(ii) Write down the posterior distribution \( \pi(\phi, \psi|y) \) for \( \phi \) and \( \psi \) in terms of the model elements.

(iii) Write down a simple Metropolis Hastings algorithm targeting \( \pi(\phi, \psi|y) \).

(iv) Let \( \chi = (\phi, \psi) \) and \( \chi = (\phi', \psi') \). Suppose the transition density for your algorithm is \( K(\chi, \chi') \). Show that \( K \) satisfies detailed balance with respect to \( \pi \). Why is this important?

(b)

(i) Let \( u \sim g(u) \) be a random variable with density \( g \) and suppose \( (\chi, u) \in \Omega \). Let \( f : \Omega \rightarrow \Omega \) be an invertible differential function and suppose that if \( (\chi', u') = f(\chi, u) \) then \( (\chi, u) = f(\chi', u') \). Given \( \chi \) we generate an MCMC proposal by simulating \( u \sim g(u) \), computing \( (\chi', u') = f(\chi, u) \) and taking \( \chi' \) as our proposal. Write down the acceptance probability for this proposal.

(ii) Suppose \( u_1 \sim U(0.5, 2), u_2 \sim U(0.5, 2) \) and \( (\chi', u') = f(\chi, u) \) with \( u = (u_1, u_2) \) and

\[
\begin{align*}
\phi' &= \bar{\phi} + u_1(\phi - \bar{\phi}) \\
\psi' &= \bar{\psi} + u_2(\psi - \bar{\psi}).
\end{align*}
\]

Show that \( (\chi, u) = f(\chi', u') \) with \( u' = (u'_1, u'_2) = (1/u_1, 1/u_2) \).

(iii) Show that the acceptance probability is

\[
\alpha(\phi', \psi'|\phi, \psi) = \min \left\{ \frac{\exp(-\theta \sum_{i,j} \sum_{(a,b) \in N_{i,j}} |\phi'_a \psi'_b - \phi'_i \psi'_j|)}{\exp(-\theta \sum_{i,j} \sum_{(a,b) \in N_{i,j}} |\phi_a \psi_b - \phi_i \psi_j|)} \right\} u_1^{-3} u_2^{-3}.
\]

noting any further conditions on \( \phi', \psi' \).

*Hint: Suppose \( A \) is an invertible square matrix and \( u, v \) are column vectors.*

\[
\det(A + uu^T) = (1 + v^T A^{-1} u) \det(A).
\]

This acceptance probability should read min\{1, ...\}. Also, the likelihood of the data is missing - a factor

\[
py(\phi, \psi, psf')py(\phi, \psi, psi)
\]

missing before \( u_1 \).
2. (a) Let \( S = (S_1, \ldots, S_K) \) be a partition of \( 1 : n = \{1, 2, \ldots, n\} \).

(i) Specify the Chinese Restaurant process for arrivals \( 1, 2, \ldots, n \).

(ii) Let \( P_{\text{CRP}}(S) \) equal the probability the outcome of the CRP is the random partition \( S \). Show that

\[
P_{\text{CRP}}(S) = \frac{\Gamma(\alpha) \alpha^K}{\Gamma(\alpha + n)} \prod_{k=1}^{K} \Gamma(n_k)
\]

[Note that \( \Gamma(\alpha + n) = \Gamma(\alpha) \prod_{i=1}^{n} (\alpha + i - 1) \).]

(iii) Show that \( P_{\text{CRP}}(S) \) does not depend on the order of the arrivals.

(iv) Let \( i_1, i_2, i_3 \in 1 : n \) be three fixed labels. What is the probability that \( i_1, i_2, i_3 \) are in the same partition set?

(b) Consider a mixture of normal densities with a fixed number \( M \) of components, Dirichlet distributed mixture component weights \( w = (w_1, \ldots, w_M) \), and a prior \( \pi(\theta^*) \) for the mixture component parameters \( \theta^* = (\theta^*_1, \ldots, \theta^*_M) \):

\[
\begin{align*}
    w & \sim \text{Dirichlet}(\alpha_1, \ldots, \alpha_M) \quad \text{with} \quad \alpha_M = \alpha/M \text{ for } \alpha > 0 \text{ fixed}; \\
    z_i & \sim \text{Multinom}(w), \ i = 1, \ldots, n; \\
    \theta^*_{m} & \sim \pi(\theta^*_{m}), \ m = 1, \ldots, M.
\end{align*}
\]

Here \( z_i \sim \text{Multinom}(w), \ i = 1, \ldots, n \) means that \( z_i = m, m \in \{1, \ldots, M\} \) with probability \( w_m \). In this model \( z_i \in \{1, \ldots, M\} \) is the label of the cluster to which \( y_i \) belongs. The observation model is

\[
y_i \sim f(y_i; \theta^*_z), \ i = 1, \ldots, n.
\]

Suppose the list \( z = (z_1, \ldots, z_n) \) of cluster labels contains \( K \leq M \) unique values \( m_1, \ldots, m_K \). For \( k = 1, \ldots, K \) let \( S_k = \{i : z_i = m_k, i = 1, \ldots, n\} \). Let \( S = (S_1, \ldots, S_K) \).

We write \( S = S(z) \) for the partition determined from \( z \) in this way.

(i) Write down the posterior for \( \theta^*, z, w|y \) in terms of the model elements.

(ii) Calculate the marginal prior probability \( \pi_z(z) = \int \pi_z(w|z) \pi_w(w) dw \) for a set of cluster labels.

(iii) For \( k = 1, \ldots, K \), let \( n_k = |S_k| \) give the number of items in cluster \( k \). Let

\[
P(S) = \sum_{z : S(z) = S} \pi_z(z)
\]

denote the prior distribution over partitions. Show that

\[
P(S) = \frac{\Gamma(\alpha)}{\Gamma(\alpha/M)^K (M - K)!} \prod_{k=1}^{K} \frac{\Gamma(\alpha/M + n_k)}{\Gamma(\alpha + n)}.
\]

(iv) Show that the prior distribution over partitions converges to a CRP as \( M \to \infty \).

(v) Write down the distribution to which the marginal posterior distribution of \( \theta^*, S|y \) converges in the limit as \( M \to \infty \).
3. Let $A$ be the points in the unit square $[0,1]^2$. For $i = 1, \ldots, n$, let $x_i = (x_{i,1}, x_{i,2})$ be points $x_i \in A$ and let $x = \{x_1, \ldots, x_n\}$ be a set of $n$ points. The number and locations of the points are distributed according to a Strauss point process with parameters $\lambda > 0$, $0 \leq \gamma \leq 1$ and $R \geq 0$. In particular, let $c(x; R)$ be the function

$$c(x; R) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} I_{|x_i-x_j| < R}$$

counting the number of $R$-close pairs of points in $x$. Let $\Omega_n$ be the set of distinguishable point patterns $x = \{x_1, \ldots, x_n\}$ in $A$ and let $\Omega = \bigcup_{n=0}^{\infty} \Omega_n$. The distribution of $x|\lambda, \gamma, R$ is

$$p(x|\lambda, \gamma, R) = \xi(\lambda, \gamma, R)^c(x; R) \lambda^n,$$

where $\xi(\lambda, \gamma, R)$ is a normalising constant ensuring the density is normalised to one over $\Omega$.

(a) [5 marks]

(i) Give an expression for the normalising constant $\xi(\lambda, \gamma, R)$ for $p(x|\lambda, \gamma, R)$.

(ii) Suppose $\gamma = 1$. Show, by integrating $p(x|\lambda, \gamma, R)$ over $x \in \Omega_n$ at fixed $n$, that the number of points $n$ in the random set $x$ has a Poisson distribution.

(b) [8 marks] Suppose $x$ is observed and we wish to estimate $\lambda, \gamma, R$.

(i) Write down the posterior density $\pi(\lambda, \gamma, R|y)$ giving the density in as much detail as you can but without specifying priors.

(ii) Suppose an ABC algorithm approximately targeting $\pi(\lambda, \gamma, R|y)$ and based on summary statistics $S(x)$ is available. The ABC algorithm yields samples $(\lambda^{(t)}, \gamma^{(t)}, R^{(t)}), t = 1, 2, \ldots, T$ with associated statistics $S^{(t)}$. Explain how to make a regression adjustment to the sampled $\gamma^{(t)}$-values. What is the purpose of this adjustment and how is it justified?

(iii) What exact distribution does your ABC algorithm target? Briefly interpret the approximation being made in ABC.

(c) Consider a reversible jump algorithm simulating the Strauss process $x \sim p(x|\lambda, \gamma, R)$. The algorithm is structured as follows. Suppose $X_t = x$ with $x = (x_1, \ldots, x_n)$ so point pattern $x$ has $n$ points in it.

To add a point pick a new point $z \sim A$ (throw a dart at $A$). Let $x' = x \cup z$ and $n' = n + 1$. The number of $R$-close pairs goes up to $c(x')$ and the number of points goes up to $n + 1$.

To delete a point, choose a point $x_i \in x$ at random from $i = 1, 2, \ldots, n$. Let $x' = x \setminus x_i$ and $n' = n - 1$. The number of $R$-close pairs goes down to $c(x')$ and the number of points goes down to $n - 1$.

(i) Calculate the acceptance probability $\alpha(x'|x)$ when $x'$ differs from $x$ by the addition of one point.

(ii) Calculate the acceptance probability $\alpha(x'|x)$ when $x'$ differs from $x$ by the deletion of one point.

(iii) Outline a reversible jump MCMC algorithm targeting $p(x|\lambda, \gamma, R)$.
Show

\[ \int_{\mathbb{R}_+^d} \pi (\phi, \psi) \rho(\phi) \rho(\psi) \, d\phi \, d\psi < \infty \]

but \( \exp (-\Theta) \leq 1 \) and

\[ \int \rho(\phi) \rho(\psi) \exp \left( -\frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} \phi_{i,j}^2 + \psi_{i,j}^2 \right) \, d\phi \, d\psi = (\frac{\sigma_m}{\alpha})^m (\frac{\sigma_n}{\beta})^n < \infty \]

so integral is finite so prior is proper.

(ii) \( \pi_i (\phi, \psi | y) \propto L(\phi, \psi | y) \pi_i (\phi, \psi) \)

\[ L(\phi, \psi | y) \propto \prod_{i=1}^{m} \prod_{j=1}^{n} \phi_{i,j} \psi_{i,j} e^{-\phi_{i,j} y_{i,j}} \]

(iii) \( \phi_i^{(0)} = \phi_i, \psi_i^{(0)} = \psi_i \) (simple initialization)

(iiv) \( \phi_i \in U(\phi_i - \delta, \phi_i + \delta) \quad i = 1 \ldots m \)

\( \psi_{i,j} \in U(\psi_{i,j} - \delta, \psi_{i,j} + \delta) \quad j = 1 \ldots n \)

\[ I^+ = \prod_{i=1}^{m} \prod_{j=1}^{n} \phi_{i,j} \psi_{i,j} > 0 \quad \prod_{i=1}^{m} \psi_{i,j} > 0 \]

\( \omega_{i,j} \propto \omega \left( \phi^{(0)} | \phi^{(0)}, \psi^{(0)} \right) = \min \left\{ 1, \frac{\pi_i (\phi, \psi | y) \pi_i (\phi, \psi)}{\pi_i (\phi^{(0)}, \psi^{(0)} | y)} \right\} \)

Set \( \phi^{(0)} = \phi, \psi^{(0)} = \psi \) else \( \phi^{(0)} = \phi^{(0)}, \psi^{(0)} = \psi^{(0)} \)

(\text{Remark: this may be inefficient but we don't have enough info to know depends on data - also the testing could be done before evaluating \( \omega \), but above is min acceptable})

(iiv) Above has \( \pi(x | x') = q(x' | x) \pi(x | x') \)

\[ l(x, x') = q(x' | x) \pi(x | x') \]
Detailed balance:

\[ \Pi(x' | y) \kappa(x' | x) = \Pi(x | y) \kappa(x | x') \]

\[ \Rightarrow \Pi(x' | y) \kappa(x' | x) = \Pi(x' | y) \kappa(x' | x') \]

WLOG assume \( \Pi(x' | y) \geq \Pi(x | y) \kappa(x' | x) \rightarrow 1 \)

\[ \Rightarrow \Pi(x | y) = \Pi(x' | y) \times \min \left( 1, \frac{\Pi(x' | y)}{\Pi(x | y)} \right) \]

\[ = \Pi(x | y) \]

so DB held. Important for a very different form / stationary DBn. With irreducibility leads to ergodicity.

2.

(i) "Write down" \( \kappa(x' | x) = \min \left\{ 1, \frac{\Pi(x' | y) g(u, \cdot)}{\Pi(x | y) g(u, \cdot)} \right\} \)

\[ \text{such that } \Pi(x' | y) g(u, \cdot) \kappa(x', \cdot | x) = \Pi(x | y) g(u, \cdot) \]

\[ = \Pi(x' | y) g(\kappa(u, \cdot), \cdot) \left| \frac{\delta(x, u)}{\delta(x', u)} \right| \kappa(x, u | x', u') \]

(ii) Must show if \( u_i = \frac{1}{n_x}, u'_i = \frac{1}{n'_x} \) then

\[ \phi = \phi' + u'_i (\phi' - \phi) \]

\[ \psi = \psi' + u_i (\psi - \psi') \]

But \( \phi = \phi' (\xi(\psi - \phi) = 0 \text{ etc.) } \) so rearrange

\[ \phi = \phi' + u'_i (\phi' - \phi) \]

\[ = \phi' + u'_i (\phi' - \phi) \text{ similarly } \psi. \]
The likelihood ratio is missing in Eqn (*) - a factor \( L(\phi', \psi'; y) / L(\phi, \psi; y) = p(y|\phi', \psi') / p(y|\phi, \psi) \) - present in the first line but dropped in the second. Can add \( -\ell(\phi', \psi'; y) - \ell(\phi, \psi; y) \) to \( H \), with \( \ell(\cdot) \) the log-likelihood.

The expression for \( L(\phi, \psi; y) \) is given in 1a(ii) and does not need to be repeated.

\[
1. (v) \quad \varsigma = \frac{\tilde{\phi}'}{\tilde{\phi}}' \quad \tilde{\psi} = \tilde{\psi}' = e^{-\alpha \tilde{\phi} - \alpha \tilde{\psi}} \quad \text{cancels},
\]

\[
\alpha(\phi', \psi', \phi, \psi) = \min \left\{ 1, \frac{\prod (\phi', \psi') L(\phi', \psi'; y) \mathcal{g}_m(\phi', \psi')}{\prod (\phi, \psi) L(\phi, \psi; y) \mathcal{g}_m(\phi, \psi)} \right\}
\]

\[
H'(\phi, \psi) = \sum_{i,j} \sum_{(k, b) \in \mathcal{N}_D} \frac{1}{n} \psi_j - \phi_a \psi_0
\]

\[
J = \begin{bmatrix}
\phi' & \psi' & u_{11} & u_{12} \\
\phi & A & 0 & 0 \\
\psi & 0 & B & 0 \\
\phi - \phi' & 0 & -\frac{1}{n} & 0 \\
u_{12} & 0 & \psi - \chi & 0 & -\frac{1}{n}
\end{bmatrix}
\]

\[
\phi_i' = \phi + u_{11}(\phi_i - \phi)
\]

\[
A = \begin{bmatrix}
\begin{array}{c}
\delta_i \\
o_{ji} \\
0_{1j}
\end{array}
\end{bmatrix}
\]

\[
\text{diag } d_{ij} = \frac{1}{n} u_{11} (1 - \frac{1}{m})
\]

\[
\text{offd. } o_{ij} = \frac{1}{m} - \frac{u_{11}}{m} = \delta
\]

\[
\Delta = \begin{bmatrix}
u_{11} + \Delta \\
\Delta
\end{bmatrix}
\]

\[
\frac{\delta_i}{\delta_i'} = \frac{1}{m} - \frac{u_{11}}{m}
\]
\[ A = u_1 I_m + \Delta 11^T \]

\[ (u_1, 0) + (\frac{1}{n_m}, \ldots, \frac{1}{n_m}) \times \Delta \]

\[ \det(A) = (1 + \Delta 11^T I_m) u_1^m \]

\[ = (1 + \frac{m}{u_1} \Delta) u_1^m \]

\[ = (1 + \frac{m}{u_1} \left( 1 - \frac{u_1}{m} \right)) u_1^m \]

\[ = (1 + \frac{1}{u_1} - 1) u_1^m \]

\[ = u_1^{m-1} \quad (\det(B) \equiv u_2^{n-1}) \]

\[ \det(F) = \det(A) \det(B) \frac{1}{u_1^2} \frac{1}{u_2} \]

\[ = u_1^{m-3} u_2^{n-3} \]

Substitute values and get answer.
Mech 2

2. 1st customer sets at \( t_i \), \( i = 1 \) to \( 5 \).

(i) After his arrival task \( C_t \), \( c = 1 \) to \( 5 \)

occupied \( n_j^{(c)} = \# \) at table \( j \), \( j = 1 \) to \( k \).

if 1st assigned table \( j \) up \( n_j^{(c)} \)

new table up \( \frac{c}{x+c} \)

Outcome is partition \( S_1, \ldots, S_m \) of \( 1 \) to \( n \)

2(a) (iii) Popp 51 does not depend on
labels only only on totals \( n_p \) so

if we rearrange order in which customers
arrive permuted outcome occurs with
same prob.

(A1) \( \Pr ( i_1, i_2, i_3 \in \text{some set} ) \)

\( = \Pr (1, 2, 3 \in \text{a set} ) \)

\( = \frac{1}{x+1} \times \frac{2}{x+2} \)
\( (b) (i) \)

\[
\pi_i (u^*, y) \propto \pi_i (w) \pi_i (z_i | w) \prod_{m=1}^{M} \pi_i (\theta^*_{m}) \\
\times \prod_{i=1}^{n} f(w_i, \theta^*_i)
\]

\[
\pi_i (w) = \frac{P(y_i | \xi_i)}{P(\xi_i)} \prod_{i=1}^{n} \frac{\sum_{m=1}^{M} \pi_i (\theta^*_{m})}{\pi_i (\theta_i)}
\]

\[
\pi_i (z_i | w) = \frac{\sum_{m=1}^{M} \pi_i (\theta^*_{m})}{\prod_{i=1}^{n} \pi_i (\theta_i)}
\]

\[
\tilde{\theta}_m = \sum_{i=1}^{n} \prod_{i \in m} z_i = m \quad \text{if in group } m.
\]

\( (ii) \) PS4, QS

\( (iii) \) u

\( (iv) \) u

\( (v) \) u.
Math 2 Q3

Q3 (a) (i).

\[ E(\lambda, \gamma, R) = \sum_{n=0}^{\infty} \gamma \left( \frac{\lambda \gamma}{R} \right)^n \frac{d\gamma}{\gamma} \]

(ii) \( \lambda \in \mathbb{R}^1, \gamma \in [0, 1], R \in \mathbb{R}^+ \)

\[ \pi(\lambda, \gamma, R | p) = \pi(\lambda, \gamma, R) \frac{p(\gamma | \lambda, \gamma, R)}{Z} \]

\[ Z = \int_0^\infty \int_0^\infty \int_0^\infty \pi(\lambda, \gamma, R) p(\lambda | \lambda, \gamma, R) d\lambda d\gamma dR \]

(iii) \( p(\gamma \in \mathbb{R}^1 | x) \sim p(\gamma \in \mathbb{R}^1 | x) \)

\[ \text{if} \quad |1/\gamma| \in \mathcal{D} \text{ and } x \in \mathcal{D} \]

\[ \int_{\mathcal{D}} \text{p}(x) \text{d}x = \int_{\mathcal{D}} \text{p}(x) \lambda \text{ d}x \]

\[ = E(\lambda, 1, R) \int_{\mathcal{D}} \lambda^n \text{d}x \]

\[ = \lambda^n E(\lambda, 1, R) \int_{\mathcal{D}} \text{d}x \cdots \text{d}x \]

\[ = 1 \]

So \( p(\gamma \in \mathbb{R}^1 | x) \propto \frac{x^n}{n!} \quad n = 0, 1, \ldots \)

\[ \boxed{\text{or } 1^{n} \gamma^n} \]

\[ \text{for } \pi(\lambda, \gamma, R) \]
\( y(i+1) \) for \( i=1, \ldots, 7 \) ABC sample.

Suppose \( Y/S \sim \mu(S) + \epsilon \)

don't expect any change on \( S \) (no shift in data just mean shift).

Suppose \( \mu(S') = \beta_1 + \beta_2 (S' - S) \)

for \( d(S, S) \) small.

\( \Rightarrow \ y(i) = \beta_1 + \beta_2 (S'(y(i)) - S(y)) \)

\( \Rightarrow \ n = Y(i) - \hat{\beta}_1 - \beta_2 (S'(y(i)) - S(y)) \)

\( \Rightarrow \ E(Y(i)) = Y(i) - \hat{\beta}_1 - \beta_2 (S'(y(i)) - S(y)) \)

now \( \mu(S'(y(i)) = \hat{\beta}_1 \)

\( \Rightarrow \ \mu(S'(y(i)) + E(i) = Y(i) \)

\( \Rightarrow \ \beta_1 + Y(i) = \beta_1 - \beta_2 (S'(y(i)) - S(y)) \)

so \( \beta_1 + Y(i) - \hat{\beta}_1 - \beta_2 (S'(y(i)) - S(y)) \approx Y(S'(y(i)) \) approx

so \( Y_{new}(i) = Y(i) - \hat{\beta}_2 (S'(y(i)) - S(y)) \)

is better distribution \( S \ Y(S'(y(i)) \).
(b) \[ X, Y \in \mathbb{N}(x, y) \mid d(X, x) < r \]

where \( X \) is "unknown true" data and \( n \) is observed data.

It is as if we only knew the data up to some uncertainty.

\[ c(n') = c(n) - c(1) \]

\[ \lambda = \frac{\lambda^{(n')}}{\lambda^{(1)}} \]

\[ \lambda = \frac{1}{n+1} \]

\[ X = \frac{1}{n} \frac{\lambda^{(n')}}{\lambda^{(1)}} \]

for \( \alpha' = x \lambda^{(n')} \), \( \alpha = x \lambda^{(1)} \)

\[ X' = x' \lambda^{n'} \]

\[ Y' = x' \lambda^{n'} \]

\[ wp_{1/2} (\alpha', x') = \sum_{\alpha} \lambda^{n'} \]

\( wp_{1/2} \) (if add) or \( x'_{(ii)} \) (if delete) accept

\( x' \) (set \( x + x' = x' \)) else reject \( (x' + x = x') \).