CMPE 58N - Lecture 5
Monte Carlo methods

Introduction to model selection

Department of Computer Engineering,

Boğaziçi University, Istanbul, Turkey

Instructor: A. Taylan Cemgil

Fall 2009

Example: Poisson change-point model revisited

- We have a sequence of counts x_j.
- Is the count distribution changing or not ?

Generative model

- We model the counts with Poisson random variables x_j, $j = 1 \dots M$

$$\begin{aligned} \mathcal{PO}(x; \lambda) &= e^{-\lambda} \frac{\lambda^x}{x!} \\ &= \exp\left(+x \log \lambda - \lambda - \log(x!)\right) \end{aligned}$$

- The unknown intensity can be modelled by a Gamma random variable

$$\begin{aligned} \mathcal{G}(\lambda; a, b) &\equiv \frac{1}{\Gamma(a)} b^a \lambda^{a-1} \exp(-b\lambda) \\ &= \exp\left((a-1) \log \lambda - b\lambda - \log \Gamma(a) + a \log b\right) \end{aligned}$$

- The unknown intensity λ_1 is jumping to a new unknown value λ_2 after an unknown index m

Generative model

$$
\begin{aligned}
m &\sim \mathcal{U}\{1 \dots M\} \\
\lambda_i &\sim \mathcal{G}(\lambda_i; a, b) \\
x_j &\sim \begin{cases} \mathcal{PO}(x_j; \lambda_1) & 1 \leq j \leq m \\ \mathcal{PO}(x_j; \lambda_2) & m < j \leq M \end{cases}
\end{aligned}
$$

Goal: Compute $p(m|x_{1:M})$

Full Joint Density

$$
\begin{aligned}
p(x_{1:M}, \lambda_1, \lambda_2, m) &= p(x_{1:M}|\lambda_1, \lambda_2, m)p(\lambda_1)p(\lambda_2)p(m) \\
&= \left(\prod_{j=1}^{m} p(x_j|\lambda_1)\right)\left(\prod_{j=m+1}^{M} p(x_j|\lambda_2)\right)p(\lambda_1)p(\lambda_2)p(m)
\end{aligned}
$$

Integrating out λ_1 and λ_2

$$p(x_{1:M}, m) = \int d\lambda_1 d\lambda_2 p(x_{1:M}|\lambda_1, \lambda_2, m) p(\lambda_1) p(\lambda_2) p(m)$$

Compound Poisson

$$\mathcal{CP}(x; a, b) = \int d\lambda \mathcal{PO}(x; \lambda)\mathcal{G}(\lambda; a, b)$$

$$\begin{aligned}
\mathcal{CP}(x; a, b) &= \int d\lambda \exp((a + x - 1)\log\lambda - (b + 1)\lambda \\
&\quad - \log(x!) - \log\Gamma(a) + a\log b) \\
&= \int d\lambda \mathcal{G}(\lambda; a + x, b + 1)\exp(\log\Gamma(a + x) - (a + x)\log(b + 1) \\
&\quad - \log(x!) - \log\Gamma(a) + a\log b) \\
&= \frac{\Gamma(a + x)}{\Gamma(x + 1)\Gamma(a)}\frac{b^a}{(b + 1)^{(a+x)}}
\end{aligned}$$

Compound Poisson

$$
\begin{aligned}
p(x)p(\lambda|x) &= p(x|\lambda)p(\lambda) \\
\mathcal{CP}(x;a,b)\mathcal{G}(\lambda;a+x,b+1) &= \mathcal{PO}(x;\lambda)\mathcal{G}(\lambda;a,b)
\end{aligned}
$$

Full Joint Distribution

$$
\begin{aligned}
\mathcal{L} &= \log p(x_{1:M}|\lambda_1, \lambda_2, m) + \log p(\lambda_1) + \log p(\lambda_2) + \log p(m) \\
&= \sum_{j=1}^{m} \left(+x_j \log \lambda_1 - \lambda_1 - \log(x_j!)\right) \\
&\quad + \sum_{j=m+1}^{M} \left(+x_j \log \lambda_2 - \lambda_2 - \log(x_j!)\right) \\
&\quad +(a-1)\log \lambda_1 - b\lambda_1 - \log \Gamma(a) + a \log b \\
&\quad +(a-1)\log \lambda_2 - b\lambda_2 - \log \Gamma(a) + a \log b - \log M
\end{aligned}
$$

Full Joint Distribution

$$
\begin{aligned}
p(x_{1:M}, m) &= \frac{b^a \Gamma(a + \sum_{j=1}^{m} x_j)}{\Gamma(a)(m+b)^{(a+\sum_{j=1}^{m} x_j)} \prod_{j=1}^{m} \Gamma(x_j+1)} \\
&\quad \times \frac{b^a \Gamma(a + \sum_{j=m+1}^{M} x_j)}{\Gamma(a)(M-m+b)^{(a+\sum_{j=m+1}^{M} x_j)} \prod_{j=m+1}^{M} \Gamma(x_j+1)} \\
&\propto \frac{\Gamma(a + \sum_{j=1}^{m} x_j)\Gamma(a + \sum_{j=m+1}^{M} x_j)}{(m+b)^{(a+\sum_{j=1}^{m} x_j)}(M-m+b)^{(a+\sum_{j=m+1}^{M} x_j)}}
\end{aligned}
$$

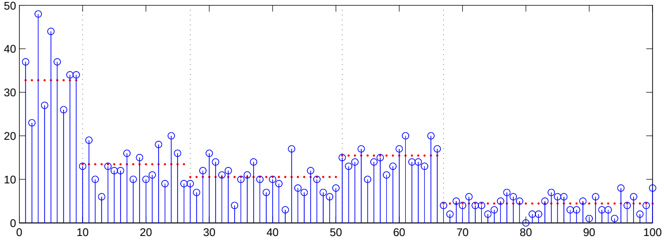
Implementation

```
cx = cumsum(data.x);
cxi = cx(end) - cx;
tlp = gammaln(data.a + cx) +  gammaln(data.a + cxi) ...
 - (data.a + cx).*log((1:data.M) + data.b) ...
 - (data.a + cxi).*log(data.M - (1:data.M) + data.b);

bar(normalize_exp(tlp, 2));
set(gca, 'xlim', [0 data.M+1])
xlabel('m')
```

$$
\begin{aligned}
p(x_{1:M}, m = 0) &= \frac{b^a \Gamma(a)}{\Gamma(a) b^a} \\
&\times \frac{b^a \Gamma(a + \sum_{j=1}^{M} x_j)}{\Gamma(a)(M + b)^{(a + \sum_{j=1}^{M} x_j)} \prod_{j=1}^{M} \Gamma(x_j + 1)}
\end{aligned}
$$

Multiple Change Points



A Multiple Change Point model

$$
\begin{aligned}
c_1 &= 1 \\
j &\geq 2 \\
c_j &= c_{j-1} + \mathcal{B}(r_j; w) \\
\lambda_c &\sim \mathcal{G}(\lambda_c; a, b) \\
x_j | \lambda_{c_j} &\sim \mathcal{PO}(x_j; \lambda_{c_j})
\end{aligned}
$$

Multiple Change Points – Indicator representation

$$
\begin{aligned}
\lambda_0 &\sim \mathcal{G}(\lambda_0; a, b) \\
r_j &\sim \mathcal{B}(r_j; w) \\
\lambda_j | r_j, \lambda_{j-1} &\sim [r_j = 0]\, \delta(\lambda_j - \lambda_{j-1}) + [r_j = 1]\, \mathcal{G}(\lambda_j; a, b) \\
x_j | \lambda_j &\sim \mathcal{PO}(x_j; \lambda_j)
\end{aligned}
$$

Observations

- This change point model is actually a hidden Markov model with the latent chain defined on joint variables (r_j, λ_j)
- Hence, we could in principle run the forward backward algorithm to find the posterior marginals
- Take home messages:
  - Complicated looking models can be entirely tractable (simple looking models can be entirely intractable)
  - Equivalent models can have substantially different looking specifications
  - One specification can be easier to work with than others

Assignment 3

1. Derive and Implement a MH algorithm for the single changepoint model to sample from $p(m|x_{1:M})$
2. Derive and Implement a Gibbs sampler for the multiple change point problem to sample from $p(r_{1:M}|x_{1:M})$
3. (Optional) Derive and Implement an exact algorithm for the multiple change point problem to compute $p(r_j|x_{1:M})$ for $j = 1 \ldots M$.
4. (Reading) pp122-124 from Liu
5. (Reading) Section 6 pp63-66 from Bristol notes