Gibbs sampler:

Poisson with a change point

Hedibert Freitas Lopes

February 2014

Coal mining disasters in Great Britain

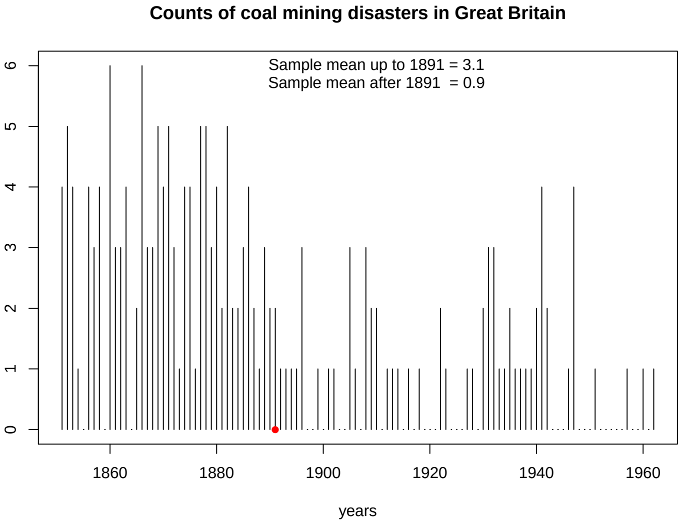
Annual counts of coal mining disasters from 1851 to 1962[1].

4	5	4	1	0	4	3	4	0	6
3	3	4	0	2	6	3	3	5	4
5	3	1	4	4	1	5	5	3	4
2	5	2	2	3	**4**	**2**	**1**	**3**	**2**
2	**1**	**1**	**1**	**1**	**3**	0	0	1	0
1	1	0	0	3	1	0	3	2	2
0	1	1	1	0	1	0	1	0	0
0	2	1	0	0	0	1	1	0	2
3	3	1	1	2	1	1	1	1	2
4	2	0	0	0	1	4	0	0	0
1	0	0	0	0	0	1	0	0	1
0	1								

Scientific question: There is a suspicion of a change point m along the observation process where the count means change, $m = 1, \ldots, n$.

[1] Jarret (1979) A note on the intervals between coal-mining disasters. *Biometrika*, **66**, 191-193.

Time series



Counts of coal mining disasters in Great Britain
Sample mean up to 1891 = 3.1
Sample mean after 1891 = 0.9
0
1
2
3
4
5
6
1860
1880
1900
1920
1940
1960
years

Model[2]

Model: For a given change point m, the observations follow

$$\begin{aligned} y_t|\lambda, m &\sim Poi(\lambda) \quad t = 1, \ldots, m \\ y_t|\phi, m &\sim Poi(\phi) \quad t = m+1, \ldots, n. \end{aligned}$$

Recall that $X \sim Poi(\theta)$ when

$$Pr(X = k|\theta) = \frac{e^{-\theta}\theta^k}{k!} \qquad k = 0, 1, 2, \ldots,$$

with $E(X|\theta) = \theta$.

---

[2] Carlin, Gelfand and Smith (1992) Hierarchical Bayesian analysis of change-point problems. *Applied Statistics*, **41**, 389-405. Tanner (1996, page 147) *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, 3rd edition, Springer-Verlag: New York.

Prior specification

The model is completed with independent prior distributions $\lambda \sim G(\alpha, \beta)$, $\phi \sim G(\gamma, \delta)$ and m uniformly distributed over $\{1, \ldots, n\}$ where α, β, γ and δ are known constants.

More precisely,

$$
\begin{aligned}
p(\lambda|\alpha, \beta) &= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \qquad \text{for } \lambda, \alpha, \beta > 0 \\
p(\phi|\gamma, \delta) &= \frac{\delta^{\gamma}}{\Gamma(\gamma)} \phi^{\gamma-1} e^{-\delta\phi} \qquad \text{for } \phi, \gamma, \delta > 0 \\
Pr(m) &= \frac{1}{n} \qquad \text{for } m = 1, \ldots, n.
\end{aligned}
$$

Posterior

$$
\begin{aligned}
p(\lambda, \phi, m|y) &\propto \left(e^{-m\lambda}\lambda^{\sum_{t=1}^{m} y_t}\right)\left(e^{-(n-m)\phi}\phi^{\sum_{t=m+1}^{n} y_t}\right) \\
&\times \left(\lambda^{\alpha-1}e^{-\beta\lambda}\right)\left(\phi^{\gamma-1}e^{-\delta\phi}\right) \\
&\propto \left(\lambda^{(\alpha+s_m)-1}e^{-(\beta+m)\lambda}\right)\left(\phi^{(\gamma+s_n-s_m)-1}e^{-(\delta+n-m)\phi}\right)
\end{aligned}
$$

where $s_m = \sum_{t=1}^{m} y_t$ for $m = 1, \ldots, n$.

The full conditional densities of λ and ϕ are

$$
\begin{aligned}
(\lambda|m, y) &= G(\alpha + s_m, \beta + m) \\
(\phi|m, y) &= G(\gamma + s_n - s_m, \delta + n - m)
\end{aligned}
$$

while

$$
Pr(m|\lambda, \phi, y) \propto \lambda^{\alpha+s_m-1}e^{-(\beta+m)\lambda}\phi^{\gamma+s_n-s_m-1}e^{-(\delta+n-m)\phi}
$$

for $m = 1, \ldots, n$.

Exact calculations

It is possible to obtain $Pr(m|y)$ analytically:

$$Pr(m|y) \propto \frac{\Gamma(\alpha + s_m)\Gamma(\gamma + s_n - s_m)}{(m+\beta)^{\alpha+s_m}(n-m+\delta)^{\gamma+s_n-s_m}} \ ,$$

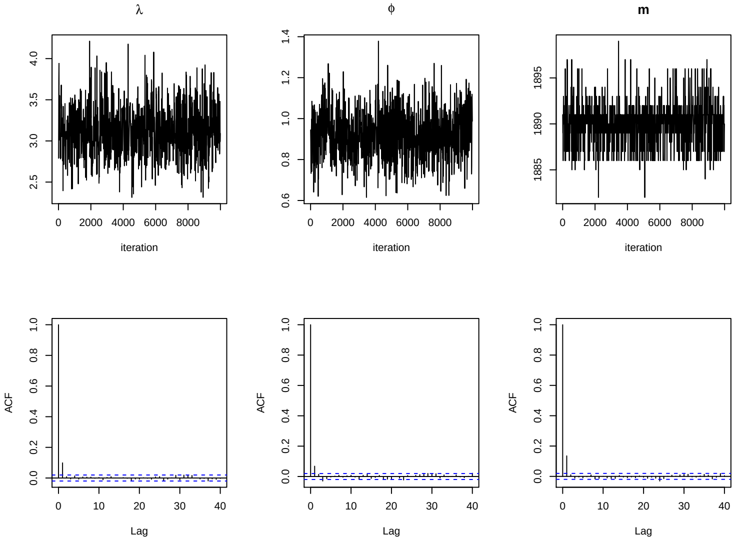
for $m = 1, \ldots, n$, Therefore,

$$\begin{aligned} p(\lambda|y) &= \sum_{m=1}^{n} p(\lambda|y,m)Pr(m|y) \\ p(\phi|y) &= \sum_{m=1}^{n} p(\phi|y,m)Pr(m|y) \end{aligned}$$

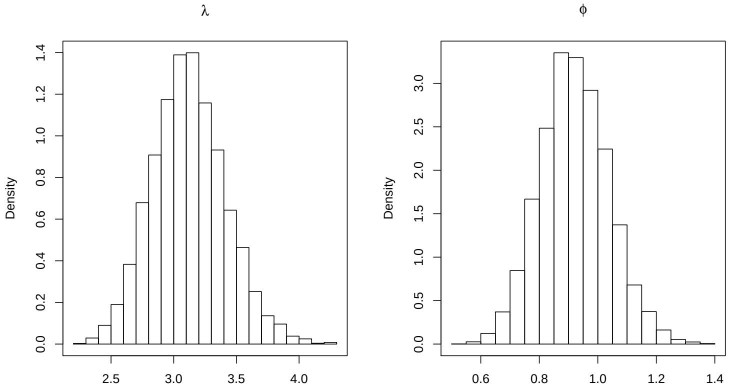
For instance

$$E(\lambda|y) = \sum_{m=1}^{n} E(\lambda|m,y)Pr(m|y)$$

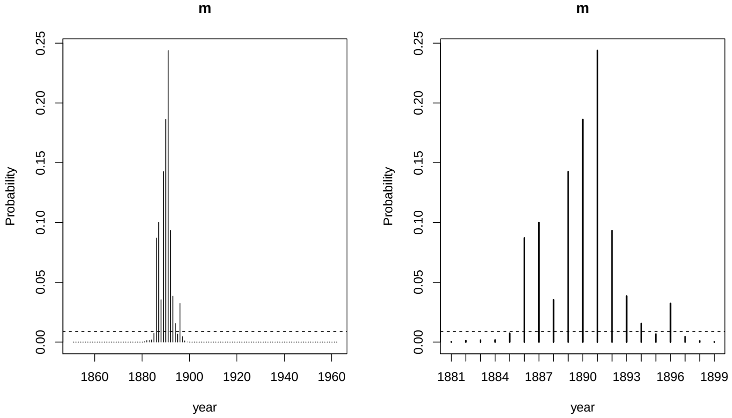
Coal mining disasters: MCMC output



Coal mining disasters: Posterior inference of λ and ϕ



Coal mining disasters: Posterior inference of m



Dashed line is the prior $Pr(m) = 1/n \quad \forall m$.

Coal mining disasters: Posterior summaries

Gibbs approximations are based on 10,000 draws, starting at $m^{(0)} = 1891$ and after discarding the first 10,000 draws.

	Exact			Gibbs sampler		
Par.	Mean	St.Dev.	95% C.I.	Mean	St.Dev.	95% C.I.
λ	3.12	0.28	(2.57,3.72)	3.12	0.29	(2.58,3.73)
ϕ	0.92	0.11	(0.68,0.96)	0.92	0.12	(0.70,1.16)
m	1890	2.42	(1886,1895)	1890	2.42	(1886,1896)

C.I. stands for credibility interval.