Stat 225 - Winter 2003

Exam Solutions.

3/2003

(1) (a) i. $p(\lambda_1,\lambda_2|Y) \propto p(Y|\lambda_1,\lambda_2)\,p(\lambda_1,\lambda_2) \propto \left[\prod_{t=1}^{k} \frac{\lambda_1^{Y_t} e^{-\lambda_1}}{Y_t!}\right]\left[\prod_{t=k+1}^{T} \frac{\lambda_2^{Y_t} e^{-\lambda_2}}{Y_t!}\right] \lambda_1^{a_1-1} e^{-\lambda_1 b_1}\, \lambda_2^{a_2-1} e^{-\lambda_2 b_2}$

Since $p(\lambda_1,\lambda_2|Y) = g(\lambda_1)\,g(\lambda_2)$ (can be factored) we know λ_1, λ_2 are indep.

ii. $p(\lambda_1,\lambda_2|Y) \propto \lambda_1^{a_1+\sum_{t=1}^{k} Y_t} e^{-\lambda_1 b_1 - \lambda_1 k}\, \lambda_2^{a_2+\sum_{t=k+1}^{T} Y_t} e^{-\lambda_2 b_2 - \lambda_2(T-k)}$

Recognize as gamma density.

$$p(\lambda_1|Y) = \text{Gamma}\left(\lambda_1 \,\middle|\, a_1+\sum_{t=1}^{k} Y_t,\ b_1+k\right)$$
$$p(\lambda_2|Y) = \text{Gamma}\left(\lambda_2 \,\middle|\, a_2+\sum_{t=k+1}^{T} Y_t,\ b_2+(T-k)\right)$$

Note: Several people tried to do i. by finding $p(\lambda_1|Y) \propto p(Y|\lambda_1)\,p(\lambda_1)$ but Y's dist^n depends on both λ_1, λ_2 so you can't do this calculation. $p(\lambda_1|Y)$ can be found from the joint dist^n given above.

(b) i. $p(\lambda_1,\lambda_2,k|Y) \propto p(Y|\lambda_1,\lambda_2,k)\,p(\lambda_1)\,p(\lambda_2)\,p(k)$

$$\propto \left(\prod_{t=1}^{k} \lambda_1^{Y_t} e^{-\lambda_1}\right)\left(\prod_{t=k+1}^{T} \lambda_2^{Y_t} e^{-\lambda_2}\right) \lambda_1^{a_1-1} e^{-\lambda_1 b_1}\, \lambda_2^{a_2-1} e^{-\lambda_2 b_2}$$

(same as before since $p(k)$ is uniform but now k is a r.v.)

ii. Gibbs sampling: $p(\lambda_1,\lambda_2|Y,k)$ is the posterior dist^n from (a) hence known

$p(k|\lambda_1,\lambda_2,Y)$ is not a known dist^n but it is a discrete dist^n taking values in $\{1,2,\cdots T-1\}$. Just compute above at each value and renormalize to create distribution

Note: A grid here would be 3-dimensional. Probably not practical if T is large.

(c) i. $p(Y_t|Y_{-t}) = \int p(Y_t|\lambda_1,Y_{-t})\,p(\lambda_1|Y_{-t})\,d\lambda_1$

$= \int p(Y_t|\lambda_1)\,p(\lambda_1|Y_{-t})\,d\lambda$ (Y_t is indep of Y_{-t} given λ_1)

ii. Compute $p(Y_t|Y_{-t})$. If this is small then Y_t is unusual given the other data. Of course it can be hard to judge "small" if order of magnitude of Y's is large because then every value is unlikely. In such a case we might sample Y_t^{rep} from $p(Y_t|Y_{-t})$ and see where observed Y_t falls in the distribution.

(2) (a) In this case Y_{ij} are not iid given θ_i since they also depend on M_{ij}. They are <u>not</u> exchangeable. If all of the M_{ij}'s for a given i are equal, then the Y_{ij} will be exchangeable.

② ⓑ $p(\underline{\theta}, \alpha, \beta \mid Y) \propto \left(\prod_{i=1}^{I} \prod_{j=1}^{n_i} \theta_i m_{ij} e^{-\theta_i m_{ij} Y_{ij}}\right)\left(\prod_{i=1}^{I} \theta_i^{\alpha-1} e^{-\theta_i \beta}\right) p(\alpha, \beta)$

ⓒ $p(\underline{\theta} \mid \alpha, \beta, Y) \propto \prod_{i=1}^{I} \theta_i^{\alpha + n_i - 1} e^{-\theta_i(\beta + \sum_j m_{ij} Y_{ij})}$ (only include terms depending on $\underline{\theta}$)

$\Rightarrow \theta_i,\ i=1, \dots I$ are indep. $\text{Gamma}(\alpha + n_i,\ \beta + \sum_j m_{ij} Y_{ij})$ r.v.s

ⓓ The conditional posterior dist for θ_i takes the prior info (α, β) and updates to include the data yielding $(\alpha + n_i,\ \beta + \sum_j m_{ij} Y_{ij})$. Another way to see this is to look at prior and posterior mean:

- prior mean of $\theta_i = \frac{\alpha}{\beta}$

posterior mean $= \dfrac{\alpha + n_i}{\beta + \sum_j m_{ij} Y_{ij}} = \dfrac{\frac{\alpha}{\beta} \cdot \beta + \frac{n_i}{\sum_j m_{ij} Y_{ij}} \sum m_{ij} Y_{ij}}{\beta + \sum m_{ij} Y_{ij}}$

which combines prior mean $\frac{\alpha}{\beta}$ and MLE for θ_i $\frac{n_i}{\sum_j Y_{ij} m_{ij}}$

ⓔ This is just like the HW and example in class.

Approach 1: Find $p(\alpha, \beta \mid Y) = \dfrac{p(\alpha, \beta, \theta \mid Y) \text{ from ⓑ}}{p(\theta \mid \alpha, \beta, Y) \text{ from ⓒ}}$

- Use 2-D grid approx to $p(\alpha, \beta \mid Y)$ to generate (α, β) sample
- Sample θ from $p(\theta \mid \alpha, \beta, Y)$

Approach 2: MCMC — We know $p(\theta \mid \alpha, \beta, Y)$

— Need to say something about $p(\alpha, \beta \mid \theta, Y)$ (HARD!)

ⓕ Given posterior sample, for θ_i, say $\theta_i^{(1)}, \theta_i^{(2)}, \dots \theta_i^{(M)}$ (from part ⓔ), we just sample $Y^{(pred,k)} \sim \text{Exp}(\theta_i^{(k)} \cdot 2)$ $k = 1, \dots M$.

ⓖ Here we need θ_{new}. To simulate this predictive distribution.

For $k = 1, \dots M$

- Generate $\theta_{new} \sim p(\alpha, \beta \mid Y)$ (a new job type parameter from population dist^n)
- Generate $Y^{new} \sim \text{Exp}(\theta_{new} \cdot 2)$