No.

Date

Stat 602 Exam 1 — Shangyang Huang, sh389,

1. $Y_i \mid \alpha \overset{iid}{\sim} \text{Pareto}(\alpha, c = 1000)$, $i = 1 \ldots n$

$\alpha \sim \text{Gamma}(a, b = \text{rate})$

(a) Posterior distribution of α

Likelihood: $p(y_{1:n} \mid \alpha) = \prod_{i=1}^{n} \frac{\alpha c^{\alpha}}{y_i^{\alpha+1}} \mathbb{1}(y_i > c = 1000)$

$$\underset{\alpha}{\propto} \frac{\alpha^n c^{n\alpha}}{\left(\prod_{i=1}^{n} y_i\right)^{\alpha+1}}$$

prior: $p(\alpha) = \frac{b^a}{\Gamma(a)} \alpha^{a-1} e^{-b\alpha} \quad (\alpha > 0)$

$$\underset{\alpha}{\propto} \alpha^{a-1} e^{-b\alpha}$$

posterior. $p(\alpha \mid y_{1:n}) \underset{\alpha}{\propto} p(y_{1:n} \mid \alpha)\, p(\alpha)$

$$= \frac{\alpha^n c^{n\alpha}}{\left(\prod_{i=1}^{n} y_i\right)^{\alpha+1}} \cdot \alpha^{a-1} e^{-b\alpha}$$

$$\underset{\alpha}{\propto} \alpha^{(a+n-1)} \left(\frac{c^n}{\prod_{i=1}^{n} y_i}\right)^{\alpha} e^{-b\alpha}$$

$$= \alpha^{(a+n-1)} e^{(-\alpha(b - n\ln c + \ln \prod_{i=1}^{n} y_i)}$$

$\Rightarrow$ posterior $\alpha \sim \text{Gamma}\left(a+n,\ b + \sum_{i=1}^{n} \ln y_i - n \ln 1000\right)$

$$\sim \text{Gamma}\left(a+n,\ b + \sum_{i=1}^{n} \ln\left(\frac{y_i}{1000}\right)\right)$$

No.

Date

(b) MLE of α, $\hat{\alpha}_n = \dfrac{n}{\sum_{i=1}^{n} \ln\left(\frac{y_i}{1000}\right)}$

Show posterior mean = weighted average of prior mean and MLE

prior mean of $\alpha = \dfrac{a}{b} = \mathbb{E}(\alpha)$

posterior mean of $\alpha = \dfrac{a+n}{b+\sum_{i=1}^{n} \ln\left(\frac{y_i}{1000}\right)}$

$$\Rightarrow \frac{a+n}{b+\sum_{i=1}^{n} \ln\left(\frac{y_i}{1000}\right)} = \frac{a}{b+\sum_{i=1}^{n} \ln\left(\frac{y_i}{1000}\right)} + \frac{n}{b+\sum_{i=1}^{n} \ln\left(\frac{y_i}{1000}\right)}$$

$$= \frac{b}{b+\sum_{i=1}^{n} \ln\left(\frac{y_i}{1000}\right)} \cdot \frac{a}{b} + \frac{\sum_{i=1}^{n} \ln\left(\frac{y_i}{1000}\right)}{b+\sum_{i=1}^{n} \ln\left(\frac{y_i}{1000}\right)} \cdot \frac{n}{\sum_{i=1}^{n} \ln\left(\frac{y_i}{1000}\right)}$$

$$= \frac{b}{b+\sum_{i=1}^{n} \ln\left(\frac{y_i}{1000}\right)} \cdot E(\alpha) + \frac{\sum_{i=1}^{n} \ln\left(\frac{y_i}{1000}\right)}{b+\sum_{i=1}^{n} \ln\left(\frac{y_i}{1000}\right)} \cdot \hat{\alpha}_n$$

10) $n = 0$, No data on insurance claims,

$n \to \infty$, posterior mean ~~used~~ approaches the $\hat{\alpha}$

and the prior is outweighted.

2. Conjugate distribution

Benefit: It is computationally convenient.

Limitation: It can be overly restrictive / inappropriate for the practical scenario.

3. Normal Distribution $Y_i | \theta \overset{iid}{\sim} \text{Normal}(\theta, \sigma^2)$, $i = 1, \dots, n$

(a)(i) A conjugate prior for θ is

$$\theta \mid \mu_0, \tau_0^2 \sim \text{Normal}(\mu_0, \tau_0^2)$$

(ii) Posterior distribution of $\theta \mid y_{1:n} \sim \text{Normal}(M, L)$

where $$M = \frac{\mu_0/\tau_0^2 + \sum_{i=1}^n y_i/\sigma^2}{1/\tau_0^2 + n/\sigma^2}$$

$$L = \left(1/\tau_0^2 + n/\sigma^2\right)^{-1}$$

(b) Posterior predictive distribution $p(\tilde{y} \mid y_{1:n})$

$$p(\tilde{y} \mid y_{1:n}) = \int p(\tilde{y}, \theta \mid y_{1:n})\, d\theta$$

$$= \int p(\tilde{y} \mid \theta)\, p(\theta \mid y_{1:n})\, d\theta$$

$$\cancel{= \int \frac{1}{\sqrt{2\pi L}} e^{-\frac{1}{2L}(\tilde{y}-\theta)^2} \cdot}$$

$$= \int \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\tilde{y}-\theta)^2} \cdot \frac{1}{\sqrt{2\pi L}} e^{-\frac{1}{2L}(\theta-M)^2}\, d\theta$$

$$\underset{\theta}{\propto} \int e^{-\frac{1}{2\sigma^2}(\tilde{y}-\theta)^2 - \frac{1}{2L}(\theta-M)^2}\, d\theta$$

No.

Date

(c) (i) It makes sense to fit model in equation 0.1 when we have historical data to estimate the prior variance σ^2 of θ, e.g., when we study the height of a population we already have some background data on how variable heights are.

(ii) We'd not know σ^2 if there's scarce data for us to estimate prior variance well, like in the sprinter's experiment.

(d) Normal-Gamma model

No.

Date

4. $Y_i | \gamma \overset{iid}{\sim} \text{Weibull}(\gamma, k)$, $i = 1, \dots, k$

(a) Set $k = 5$. The laptop's time-to-failure should center around ~~a time that~~ its designed life expectancy, and have tails on the left and right.

(b) $p(y_i | \gamma = \frac{1}{\lambda}, k) = k\lambda (y_i \lambda)^{k-1} \exp(-(y_i\lambda)^k)$, $y_i \geq 0$

$$\Rightarrow p(y_{1:n} | \lambda) = \prod_{i=1}^{n} k\lambda (y_i\lambda)^{k-1} \exp(-(y_i\lambda)^k)$$

$$\underset{\lambda}{\propto} \lambda^n \lambda^{(k-1)n} \prod_{i=1}^{n} \exp(-(y_i\lambda)^k)$$

$$= \lambda^n \lambda^{n(k-1)} \exp\left(-\sum_{i=1}^{n} (y_i\lambda)^k\right)$$

(c) Suppose $k=2$. consider $p(\lambda) \propto \lambda^{a-1} \exp(-b\lambda)$

posterior $p(\lambda | y_i) = p(y_i|\lambda) p(\lambda) \underset{\lambda}{\propto} 2\lambda (y_i\lambda)^{2-1} \exp(-(y_i\lambda)^2) \cdot \lambda^{a-1} \exp(-b\lambda)$

$$\underset{\lambda}{\propto} \lambda^{a+1} \exp(-\lambda(b + y_i^2\lambda))$$

Cannot be written in the form of a Gamma distribution because of the λ^2 in the power. $\Rightarrow \lambda \sim \text{Gamma}$ is not a conjugate for the likelihood.

No.

Date

(d) $\lambda = \theta^{\frac{1}{k}}$.

Then,

$$p(y_{1:n}|\theta) \underset{\theta}{\propto} (\theta^{1/k})^n \, (\theta^{1/k})^{n(k-1)} \exp\left(-\sum_{i=1}^{n} (y_i \theta^{1/k})^k\right)$$

$$= \theta^{\left(\frac{n}{k} + \frac{nk-n}{k}\right)} \exp\left(-\sum_{i=1}^{n} y_i^k \theta\right)$$

$$= \theta^n \exp\left(-\theta \cdot \sum_{i=1}^{n} y_i^k\right)$$

Prior $\quad p(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} \underset{\theta}{\propto} \theta^{a-1} e^{-b\theta}$

$$\Rightarrow p(\theta|y_{1:n}) \underset{\theta}{\propto} \theta^n \exp\left(-\theta \sum_{i=1}^{n} y_i^k\right) \cdot \theta^{a-1} e^{-b\theta}$$

$$= \theta^{a+n-1} \exp\left(-\theta \left(b + \sum_{i=1}^{n} y_i^k\right)\right)$$

$\Rightarrow \theta | y_{1:n} \sim \text{Gamma}\left(a+n, \; b + \sum_{i=1}^{n} y_i^k\right)$ for all k.

(e) Posterior mean

$$E(\theta \mid y_{1:n}) = \frac{a+n}{b+\sum_{i=1}^{n} y_i^k}$$ by definition of Gamma.

$$\hat{\theta} = \frac{n}{\sum_{i=1}^{n} y_i^k}$$

prior mean $E(\theta) = \frac{a}{b}$

$$\Rightarrow E(\theta \mid y_{1:n}) = \frac{a+n}{b+\sum_{i=1}^{n} y_i^k} = \frac{a}{b+\sum_{i=1}^{n} y_i^k} + \frac{n}{b+\sum_{i=1}^{n} y_i^k}$$

$$= \frac{a}{b} \cdot \frac{b}{b+\sum_{i=1}^{n} y_i^k} + \frac{n}{\sum_{i=1}^{n} y_i^k} \cdot \frac{\sum_{i=1}^{n} y_i^k}{b+\sum_{i=1}^{n} y_i^k}$$

$$= \frac{b}{b+\sum_{i=1}^{n} y_i^k} \cdot E(\theta) + \frac{\sum_{i=1}^{n} y_i^k}{b+\sum_{i=1}^{n} y_i^k} \cdot \hat{\theta}$$

When one would place higher weight on the prior mean:

1) lots of historical data leading to strong prior belief i.e., large b.

2) ~~scarce~~ small sample size leading to small $\sum_{i=1}^{n} y_i^k$

No.

Date

(f) Set $k=1$ in (d)

$$\cancel{Z} \quad Z|\theta \sim \text{Weibull}(\theta^{1/k}, k) = \text{Weibull}(\theta, 1)$$

Predicative distribution $p(Z|y_{1:n})$.

From (d). $\cancel{p(y_i|\theta_i)}$ $p(\theta|y_{1:n}) \sim \text{Gamma}\left(a+n,\ b+\sum_{i=1}^{n} y_i\right)$

When $k=1$, $p(Z|\theta, 1) = \theta \exp(-z\theta)$

$$\Rightarrow \quad p(Z|y_{1:n}) = \int \theta \exp(-z\theta) \cdot \theta^{a+n-1} \exp\left(-\left(b+\sum_{i=1}^{n} y_i\right)\theta\right) \cdot \frac{\left(b+\sum_{i=1}^{n} y_i\right)^{a+n}}{\Gamma(a+n)} \, d\theta$$

$$= \frac{\left(b+\sum_{i=1}^{n} y_i\right)^{a+n}}{\Gamma(a+n)} \int \theta^{(a+n+1)-1} \exp\left(-\theta\left(b+z+\sum_{i=1}^{n} y_i\right)\right) d\theta$$

$$= \frac{\left(b+\sum_{i=1}^{n} y_i\right)^{a+n}}{\Gamma(a+n)} \cdot \frac{\Gamma(a+n+1)}{\left(b+z+\sum_{i=1}^{n} y_i\right)^{a+n+1}} \int \text{Gamma}\left(a+n+1,\ b+z+\sum_{i=1}^{n} y_i\right) d\theta$$

$$= \frac{(a+n)\left(b+\sum_{i=1}^{n} y_i\right)^{a+n}}{\left(b+z+\sum_{i=1}^{n} y_i\right)^{a+n+1}}$$

No.

Date

(c) $n=0$, No data on insurance claims, weight on $\hat{\alpha}$ is 0

completely determined by prior

$n \to \infty$, ~~posterior mean~~ approaches ~~the~~ $\hat{\alpha}$

~~and the prior is outweighted.~~

the weight on $\hat{\alpha}$ approaches 1

completely ignore the prior.