Final practice questions

Daniel Worrall

October 2018

1 Flour-milling

A miller is grinding wheat into flour. She models the size of flour particles using a Weibull distribution, which is

$$p(x|\beta,\theta) = \frac{\beta}{\theta} x^{\beta-1} e^{-\frac{x^\beta}{\theta}}, \qquad x, \beta, \theta > 0. \tag{1}$$

a) She collects a dataset of particle sizes $\mathcal{D} = \{x_1, x_2, ..., x_N\}$. Show that the likelihood of the parameters given $\mathcal{D}$ can be written in the form

$$p(\mathcal{D}|\beta,\theta) = \frac{\beta^A}{\theta^B} \left(\prod_{i=1}^N x_i^{\beta-1} \right) \exp\left\{ -\frac{N\bar{x}_\beta}{\theta} \right\} \tag{2}$$

where $\bar{x}_\beta = \frac{1}{N}\sum_{i=1}^N x_i^\beta$, and A and B are constants, which you need to fill in.

Answer

$$\begin{aligned}
p(\mathcal{D}|\beta,\theta) &= \prod_{i=1}^N p(x_i|\beta,\theta) \\
&= \prod_{i=1}^N \frac{\beta}{\theta} x_i^{\beta-1} e^{-\frac{x_i^\beta}{\theta}} \\
&= \left(\prod_{i=1}^N \frac{\beta}{\theta} \right) \left(\prod_{i=1}^N x_i^{\beta-1} \right) \left(\prod_{i=1}^N e^{-\frac{x_i^\beta}{\theta}} \right) \\
&= \frac{\beta^N}{\theta^N} \left(\prod_{i=1}^N x_i^{\beta-1} \right) e^{-\frac{N \frac{1}{N} \sum_{i=1}^N x_i^\beta}{\theta}} \\
&= \frac{\beta^N}{\theta^N} \left(\prod_{i=1}^N x_i^{\beta-1} \right) \exp\left\{ -\frac{N\bar{x}_\beta}{\theta} \right\}
\end{aligned}$$

b) The miller wishes to find the posterior over the parameter θ, so she uses the inverse-Gamma conjugate prior

$$p(\theta|a,b) = \frac{1}{Z(a,b)}\theta^{-a}e^{-\frac{b}{\theta}}. \tag{3}$$

Using your answer from part a) find the posterior distribution over λ, exploiting the conjugacy property. Note that you do not need to compute the normalizing factor.

Answer

$$\begin{aligned}
p(\theta|\mathcal{D},\beta,a,b) &\propto p(\mathcal{D}|\beta,\theta)p(\theta|a,b) \\
&= \frac{\beta^N}{\theta^N}\left(\prod_{i=1}^{N} x_i^{\beta-1}\right)\exp\left\{-\frac{N\bar{x}_\beta}{\theta}\right\}\frac{1}{Z(a,b)}\theta^{-a}e^{-\frac{b}{\theta}} \\
&\propto \theta^{-N}\theta^{-a}\exp\left\{-\frac{N\bar{x}_\beta}{\theta}\right\}e^{-\frac{b}{\theta}} \\
&= \theta^{-(a+N)}\exp\left\{-\frac{b+N\bar{x}_\beta}{\theta}\right\}
\end{aligned}$$

So the posterior is

$$p(\theta|\mathcal{D},\beta,a,b) = \frac{1}{Z(a+N, b+N\bar{x}_\beta)}\theta^{-(a+N)}\exp\left\{-\frac{b+N\bar{x}_\beta}{\theta}\right\}$$

2 Internet sociology

A sociologist is modeling the lengths $x > 0$ of comments posted in internet discussion forums. It is known that these comments tend to follow a log-normal distribution

$$p(x|\mu,\rho) = \frac{1}{x}\sqrt{\frac{\rho}{2\pi}}\exp\left\{-\frac{\rho}{2}\left(\log x - \mu\right)^2\right\}. \tag{4}$$

a) A collection of comments is analyzed and their lengths are stored in the dataset $\mathcal{D} = \{x_1, x_2, ..., x_N\}$. Write down the log-likelihood of the parameters $\{\mu, \rho\}$ of the log-normal given this dataset.

Answer

$$
\begin{aligned}
\mathcal{L}(\mu,\rho) &= \log p(\mathcal{D}|\mu,\rho) \\
&= \sum_{i=1}^{N} \log \frac{1}{x_i}\sqrt{\frac{\rho}{2\pi}} \exp\left\{-\frac{\rho}{2}\left(\log x_i-\mu\right)^2\right\} \\
&= \sum_{i=1}^{N} \log \frac{1}{x_i} + \sum_{i=1}^{N} \log \sqrt{\frac{\rho}{2\pi}} + \sum_{i=1}^{N} \log \exp\left\{-\frac{\rho}{2}\left(\log x_i-\mu\right)^2\right\} \\
&= -\sum_{i=1}^{N} \log x_i + \frac{1}{2}\sum_{i=1}^{N} \log \frac{\rho}{2\pi} - \sum_{i=1}^{N} \frac{\rho}{2}\left(\log x_i-\mu\right)^2 \\
&= -\sum_{i=1}^{N} \log x_i + \frac{N}{2} \log \frac{\rho}{2\pi} - \frac{\rho}{2}\sum_{i=1}^{N} \left(\log x_i-\mu\right)^2
\end{aligned}
$$

b) What is the maximum likelihood estimator of μ?

Answer

$$
\begin{aligned}
\frac{\partial}{\partial\mu}\mathcal{L}(\mu,\rho) &= \frac{\partial}{\partial\mu}\left[-\sum_{i=1}^{N}\log x_i + \frac{N}{2}\log\frac{\rho}{2\pi} - \frac{\rho}{2}\sum_{i=1}^{N}\left(\log x_i-\mu\right)^2\right] \\
&= -\frac{\partial}{\partial\mu}\left[\frac{\rho}{2}\sum_{i=1}^{N}\left(\log x_i-\mu\right)^2\right] \\
&= -\frac{\rho}{2}\sum_{i=1}^{N}\frac{\partial}{\partial\mu}\left(\log x_i-\mu\right)^2 \\
&= \rho\sum_{i=1}^{N}\left(\log x_i-\mu\right) = 0 \\
\implies N\mu &= \sum_{i=1}^{N}\log x_i \\
\mu_{\mathrm{ML}} &= \frac{1}{N}\sum_{i=1}^{N}\log x_i
\end{aligned}
$$

c) What is the maximum likelihood estimator of ρ?

Answer

$$
\begin{aligned}
\frac{\partial}{\partial \rho}\mathcal{L}(\mu,\rho) &= \frac{\partial}{\partial \rho}\left[-\sum_{i=1}^{N}\log x_i + \frac{N}{2}\log\frac{\rho}{2\pi} - \frac{\rho}{2}\sum_{i=1}^{N}\left(\log x_i - \mu\right)^2\right] \\
&= \frac{\partial}{\partial \rho}\left[\frac{N}{2}\log\rho - \frac{N}{2}\log\pi - \frac{\rho}{2}\sum_{i=1}^{N}\left(\log x_i - \mu\right)^2\right] \\
&= \frac{N}{2\rho} - \frac{1}{2}\sum_{i=1}^{N}\left(\log x_i - \mu\right)^2 = 0 \\
\implies \rho &= \frac{N}{\sum_{i=1}^{N}\left(\log x_i - \mu\right)^2}
\end{aligned}
$$

3 Phylogenetics

A phylogenist is studying the number of descendants of a new species of synthetic bacterium. The species has been genetically engineered so that on average each bacterium has less than 1 offspring. The result of this is that the species will eventually die out completely. The phylogenist is interested in modeling the total number $k \geq 0$ of this new bacterium that exist before they all die out.

a) The phylogenist proposes two models:

Model $\mathcal{M}_1$ is a simple Poisson distribution

$$
p(k|\lambda,\mathcal{M}_1) = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad \lambda > 0 \tag{5}
$$

Model $\mathcal{M}_2$ is a more sophisticated Borel distribution

$$
p(k|\mu,\mathcal{M}_2) = \frac{e^{-\mu k}(\mu k)^{k-1}}{k!}, \qquad 0 \leq \mu \leq 1. \tag{6}
$$

Given the dataset $\mathcal{D} = \{x_1, x_2, ..., x_N\}$, and given the fact that $\lambda = \mu = 1$, show that the log Bayes factor can be written as

$$
\log\frac{p(\mathcal{D}|\lambda=1,\mathcal{M}_1)}{p(\mathcal{D}|\mu=1,\mathcal{M}_2)} = N(\bar{k}-1) - \sum_{i=1}^{N}(k_i - 1)\log k_i \tag{7}
$$

where $\bar{k} = \frac{1}{N}\sum_{i=1}^{N} x_i$.

Answer

$$
\begin{aligned}
\log \frac{p(\mathcal{D}|\lambda = 1, \mathcal{M}_1)}{p(\mathcal{D}|\mu = 1, \mathcal{M}_2)} &= \log p(\mathcal{D}|\lambda = 1, \mathcal{M}_1) - \log p(\mathcal{D}|\mu = 1, \mathcal{M}_2) \\
&= \sum_{i=1}^{N} \log e^{-1} \frac{1_i^k}{k_i!} - \sum_{i=1}^{N} \log \frac{e^{-1k_i}(1k_i)^{k_i-1}}{k_i!} \\
&= \sum_{i=1}^{N} \log \frac{e^{-1}}{k_i!} - \sum_{i=1}^{N} \log \frac{e^{-k_i}(k_i)^{k_i-1}}{k_i!} \\
&= \sum_{i=1}^{N} \log e^{-1} \cancel{- \log k_i!} - \log e^{-k_i} - \log(k_i)^{k_i-1} + \cancel{\log k_i!} \\
&= \sum_{i=1}^{N} -1 + k_i - (k_i - 1) \log k_i \\
&= -N + N \frac{1}{N} \sum_{i=1}^{N} k_i - \sum_{i=1}^{N} (k_i - 1) \log k_i \\
&= -N + N\bar{k} - \sum_{i=1}^{N} (k_i - 1) \log k_i \\
&= N(\bar{k} - 1) - \sum_{i=1}^{N} (k_i - 1) \log k_i
\end{aligned}
$$

b) If the phylogenist finds that

$$
\bar{k} < 1 + \frac{1}{N} \sum_{i=1}^{N} (k_i - 1) \log k_i \tag{8}
$$

which model explains the data $\mathcal{D}$ better?

Answer

$$\bar{k} < 1 + \frac{1}{N}\sum_{i=1}^{N}(k_i - 1)\log k_i$$

$$N(\bar{k} - 1) < \sum_{i=1}^{N}(k_i - 1)\log k_i$$

$$N(\bar{k} - 1) - \sum_{i=1}^{N}(k_i - 1)\log k_i < 0$$

$$\implies \log\frac{p(\mathcal{D}|\lambda = 1, \mathcal{M}_1)}{p(\mathcal{D}|\mu = 1, \mathcal{M}_2)} = N(\bar{k} - 1) - \sum_{i=1}^{N}(k_i - 1)\log k_i < 0$$

Thus model $\mathcal{M}_2$ explains the data better.

4 The evidence

The exponential distribution is

$$p(x|\lambda) = \lambda e^{-\lambda x}, \qquad x \geq 0, \lambda > 0. \tag{9}$$

a) Given a dataset $\mathcal{D} = \{x_1, x_2, ..., x_N\}$, write down the likelihood of λ.

Answer

$$\begin{aligned}\prod_{i=1}^{N} p(x_i|\lambda) &= \prod_{i=1}^{N} \lambda e^{-\lambda x_i} \\ &= \lambda^N \prod_{i=1}^{N} e^{-\lambda x_i} \\ &= \lambda^N e^{-\lambda \sum_{i=1}^{N} x_i}\end{aligned}$$

b) The Gamma distribution is used as a prior where

$$p(\lambda|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda} \tag{10}$$

Using the fact that

$$\int_0^\infty z^{\alpha-1}e^{-\beta z}\,\mathrm{d}z = \frac{\Gamma(\alpha)}{\beta^\alpha} \tag{11}$$

and using your answer to part a), find the evidence $p(\mathcal{D}|\alpha, \beta)$.

Answer

$$\begin{aligned}
p(\mathcal{D}|\alpha, \beta) &= \int_0^\infty p(\mathcal{D}|\lambda)p(\lambda|\alpha, \beta)\,\mathrm{d}x \\
&= \int_0^\infty \lambda^N e^{-\lambda \sum_{i=1}^N x_i} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}\,\mathrm{d}\lambda \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \lambda^{N+\alpha-1} e^{-\lambda \sum_{i=1}^N x_i - \beta\lambda}\,\mathrm{d}\lambda \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \lambda^{N+\alpha-1} e^{-\lambda(\beta + \sum_{i=1}^N x_i)}\,\mathrm{d}\lambda \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(N+\alpha)}{(\beta + \sum_{i=1}^N x_i)^{N+\alpha}}
\end{aligned}$$