Machine Learning and Computational Statistics

Ilias Katsampalos

December 2, 2020

Exercise 1

a) we know that

$$p(x) = \theta^2 x exp(-\theta x) u(x)$$

The likelihood function can be written as

$$L(x|\theta) = \prod_{i=1}^{N} p(x_i) = \prod_{i=1}^{N} \theta^2 x_i exp(-\theta x_i) u(x_i) = \theta^{2N} exp(-\theta \sum_{i=1}^{N} x_i) \prod_{i=1}^{N} x_i u(x_i)$$

We need to maximize the likelihood function which is the same as maximizing the log-likelihood. The log-likelihood function can be written as:

$$\ln L(x|\theta) = 2N \ln \theta - \theta \sum_{i=1}^{N} x_i + \sum_{i=1}^{N} x_i + \sum_{i=1}^{N} u(x_i)$$

Taking the partial derivative of the log-likelihood and setting to 0 we obtain:

$$\frac{\partial \ln L(x|\theta)}{\partial \theta} = \frac{2N}{\theta} - \sum_{i=1}^{N} x_i = 0 \Longrightarrow \theta_{MLE} = \frac{2N}{\sum_{i=1}^{N} x_i}$$

b)
i) $\theta_{MLE} = 0.84$
ii) $\hat{p}(2.3) = 0.235$
iii) $\hat{p}(2.9) = 0.179$

Exercise 2

According to Bays' Theorem:

$$p(\theta|x) = \frac{p(\theta)p(x|\theta)}{p(x)}$$

In order to find the Maximum A-posterior Probability we will need to maximize $p(\theta)p(x|\theta)$

$$
\begin{aligned}
&argmax p(\theta)p(x|\theta) \\
&\Rightarrow argmax \ln\left(p(\theta)p(x|\theta)\right) \\
&\Rightarrow argmax \ln\left[\frac{1}{\sqrt{2\pi\sigma_0^2}} exp(-\frac{(\theta-\theta_0)^2}{2\sigma_0^2})\theta^{2N} exp(-\theta\sum_{i=1}^{N} x_i)\prod_{i=1}^{N} x_i u(x_i)\right] \\
&\Rightarrow argmax -\frac{1}{2}\ln 2\pi - \ln\sigma_0 + \frac{(\theta-\theta_0)^2}{2\sigma_0^2} + 2N\ln\theta - \theta\sum_{i=1}^{N} x_i + \sum_{i=1}^{N} x_i + \sum_{i=1}^{N} u(x_i)
\end{aligned}
$$

Taking the partial derivative w.r.t θ and setting to 0 we have

$$
\begin{aligned}
\frac{\partial p(\theta)p(x|\theta)}{\partial\theta} &= 0 \\
-\frac{(\theta-\theta_0)}{\sigma_0^2} + \frac{2N}{\theta} - \sum_{i=1}^{N} x_i &= 0 \\
-\theta(\theta-\theta_0) + 2N\sigma_0^2 - \theta\sigma_0^2\sum_{i=1}^{N} x_i &= 0 \\
-\theta^2 + \theta\theta_0 - \theta\sigma_0^2\sum_{i=1}^{N} x_i + 2N\sigma_0^2 &= 0 \\
\theta^2 - \theta\theta_0 + \theta\sigma_0^2\sum_{i=1}^{N} x_i - 2N\sigma_0^2 &= 0 \\
\theta^2 - \theta(\theta_0 - \sigma_0^2\sum_{i=1}^{N} x_i) - 2N\sigma_0^2 &= 0
\end{aligned}
$$

We have a 2nd order polynomial with

$$\Delta = (\theta_0 - \sigma_0^2\sum_{i=1}^{N} x_i)^2 + 8N\sigma_0^2 > 0$$

Then

$$\theta_{1,2} = \frac{\theta_0 - \sigma_0^2\sum_{i=1}^{N} x_i \pm \sqrt{(\theta_0 - \sigma_0^2\sum_{i=1}^{N} x_i)^2 + 8N\sigma_0^2}}{2}$$

We have $\theta_2 < 0$ since

$$\theta_0 - \sigma_0^2\sum_{i=1}^{N} x_i - \sqrt{(\theta_0 - \sigma_0^2\sum_{i=1}^{N} x_i)^2 + k} < 0$$

where $k > 0$

so

$$\theta_{MAP} = \frac{\theta_0 - \sigma_0^2\sum_{i=1}^{N} x_i + \sqrt{(\theta_0 - \sigma_0^2\sum_{i=1}^{N} x_i)^2 + 8N\sigma_0^2}}{2}$$

b)
i) N $\longrightarrow$ inf $then \theta_{MAP} = \theta_0$ since we have sampled the whole population
ii) σ^2 >> then $\theta_{MLE} = \frac{2N}{\sum_{i=1}^{N} x_i}$ because the first term of the following equation becomes 0

$$-\frac{(\theta - \theta_0)}{\sigma_0^2} + \frac{2N}{\theta} - \sum_{i=1}^{N} x_i = 0$$

iii) σ^2 << then $\theta_{MAP} = \theta_0$ because $\sigma_0^2 \sum_{i=1}^{N} x_i \simeq 0$ and $8N\sigma_0^2 \simeq 0$

Exercise 3

$$\begin{aligned}
L(\mu) &= \sum_{i=1}^{N}(x_n - \mu)^2 + \lambda((\mu - \mu_0)^2 - \rho) = \\
&= \sum_{i=1}^{N} x_n^2 - 2\mu \sum_{i=1}^{N} x_n + \sum_{i=1}^{N} \mu^2 + \lambda[\mu^2 - 2\mu\mu_0 + \mu_0^2 - \rho] = \\
&= \sum_{i=1}^{N} x_n^2 - 2\mu \sum_{i=1}^{N} x_n + \sum_{i=1}^{N} \mu^2 + \lambda\mu^2 - 2\lambda\mu\mu_0 + \lambda\mu_0^2 - \lambda\rho = \\
&= \sum_{i=1}^{N} x_n^2 - 2\mu \sum_{i=1}^{N} x_n + N\mu^2 + \lambda\mu^2 - 2\lambda\mu\mu_0 + \lambda\mu_0^2 - \lambda\rho
\end{aligned}$$

Taking the partial derivative w.r.t μ and setting to 0 we have:

$$\begin{aligned}
&\frac{\partial L}{\partial \mu} = -2 \sum_{i=1}^{N} x_n + 2N\mu + 2\lambda\mu - 2\lambda\mu_0 = 0 \\
&\Longrightarrow -\sum_{i=1}^{N} x_n + N\mu + \lambda\mu - \lambda\mu_0 = 0 \\
&\Longrightarrow -\sum_{i=1}^{N} x_n + \mu(N + \lambda) - \lambda\mu_0 = 0 \\
&\Longrightarrow \mu_{RR} = \frac{\sum_{i=1}^{N} x_n + \lambda\mu_0}{N + \lambda}
\end{aligned}$$

Exercise 4

a)

$$p(Y|\lambda) = \prod_{i=1}^{N} \lambda e^{-\lambda x_i} = \lambda^N e^{-\lambda \sum_{i=1}^{N} x_i}$$

b)

$$
\begin{aligned}
& argmax P(\lambda)p(Y|\lambda) \\
&= argmax \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-b\lambda} \lambda^{N} e^{-\lambda \sum_{i=1}^{N} x_i} \\
&= argmax \ln \left[\frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-b\lambda} \lambda^{N} e^{-\lambda \sum_{i=1}^{N} x_i} \right] \\
&= argmax \alpha \ln b - \ln \Gamma(\alpha) + (\alpha - 1) \ln \lambda - b\lambda + N \ln \lambda - \lambda \sum_{i=1}^{N} x_i
\end{aligned}
$$

Taking the partial derivative w.r.t λ and setting to 0 we have:

$$
\begin{aligned}
& \frac{\partial P(\lambda)p(Y|\lambda)}{\partial \lambda} = 0 \\
& \implies \frac{\alpha - 1}{\lambda} - b + \frac{N}{\lambda} - \sum_{i=1}^{N} x_i = 0 \\
& \implies \alpha - 1 - b\lambda + N - \lambda \sum_{i=1}^{N} x_i = 0 \\
& \implies \lambda_{MAP} = \frac{\alpha - 1 + N}{b + \sum_{i=1}^{N} x_i}
\end{aligned}
$$

c)

$$
p(x) = \frac{\alpha - 1 + N}{b + \sum_{i=1}^{N} x_i} exp\left(- \frac{\alpha - 1 + N}{b + \sum_{i=1}^{N} x_i} x_i \right)
$$

d)

$$
P(\lambda|Y) = \frac{P(\lambda)P(Y|\lambda)}{P(Y)} = \frac{1}{P(Y)} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-b\lambda} \lambda^{N} e^{-\lambda \sum_{i=1}^{N} x_i} = \frac{P(\lambda)P(Y|\lambda)}{P(Y)} = \frac{1}{P(Y)} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{N+\alpha-1} e^{-\lambda(b + \sum_{i=1}^{N} x_i)}
$$

setting $\frac{\beta^{\alpha}}{P(Y)\Gamma(\alpha)} = C$ we have $C\lambda^{N+\alpha-1} e^{-\lambda(b+\sum_{i=1}^{N} x_i)}$ which is in a form of a gamma distribution $C\lambda^{r} e^{-s\lambda}$ where $r = N + a - 1$ and $s = b + \sum_{i=1}^{N} x_i$

e)

$$
\frac{Posterior}{Prior} = \frac{\frac{1}{P(Y)} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{N+\alpha-1} e^{-\lambda(b+\sum_{i=1}^{N} x_i)}}{\frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-b\lambda}} = \frac{\lambda^{N} e^{-\lambda \sum_{i=1}^{N} x_i}}{P(Y)} \implies Prior \cdot \frac{\lambda^{N} e^{-\lambda \sum_{i=1}^{N} x_i}}{P(Y)} = Posterior
$$

We can see how the prior was scaled by a factor of $\frac{\lambda^{N} e^{-\lambda \sum_{i=1}^{N} x_i}}{P(Y)}$

f) Since $p(\lambda|y)$ is a gamma distribution, we express $p(\lambda|y)$ as:

$$p(\lambda|y) = \frac{(b + \sum_{i=1}^{N} x_i)^{N+a}}{\Gamma(N + a)} \lambda^{N+a-1} e^{-\lambda(b+\sum_{i=1}^{N} x_i)}$$

Then

$$\begin{aligned}
p(x|y) &= \int p(x|\lambda)p(\lambda|y)d\lambda \\
&= \int \lambda e^{-\lambda x} \frac{(b + \sum_{i=1}^{N} x_i)^{N+a}}{\Gamma(N + a)} \lambda^{N+a-1} e^{-\lambda(b+\sum_{i=1}^{N} x_i)} \\
&= \frac{(b + \sum_{i=1}^{N} x_i)^{N+a}}{\Gamma(N + a)} \int \lambda e^{-\lambda x} \lambda^{N+a-1} e^{-\lambda(b+\sum_{i=1}^{N} x_i)} \\
&= \frac{(b + \sum_{i=1}^{N} x_i)^{N+a}}{\Gamma(N + a)} \int \lambda^{N+a} e^{-\lambda(b+\sum_{i=1}^{N} x_i+x)} \\
&= \frac{(b + \sum_{i=1}^{N} x_i)^{N+a}}{\Gamma(N + a)} \frac{\Gamma(N + a + 1)}{(b + \sum_{i=1}^{N} x_i + x)^{N+a+1}} \\
&= \frac{(b + \sum_{i=1}^{N} x_i)^{N+a}}{\Gamma(N + a)} \frac{(N + a)\Gamma(N + a)}{(b + \sum_{i=1}^{N} x_i + x)^{N+a+1}} \\
&= \frac{(N + a)(b + \sum_{i=1}^{N} x_i)^{N+a}}{(b + \sum_{i=1}^{N} x_i + x)^{N+a+1}}
\end{aligned}$$

which is a lomax distribution with $d = b + \sum_{i=1}^{N} x_i$ and $c = N + a$

g) $p(x) = 0.411e^{-0.411x}$

h) $p(x|y) = \frac{55{,}806{,}059{,}528}{(17+x)^9}$

i) $P(X = 2) = 0.14709$ and $P(X = 2|Y) = 0.13688$