

# Statistical Computing for Scientists and Engineers

## Homework 1

Jiale Shi

September/11/2018

# 1 Problem 1

Consider  $n$  samples  $D_n = (x_1, \dots, x_n)$ , independently drawn from normal distribution,  $N_{\mu, \sigma^2}$ , with known variance and unknown mean,  $\sigma^2$  and  $\mu$ , respectively. Assume that the mean has a prior given by the normal distribution:

$$p(\mu|m, v^2) = N(\mu|m, v^2) \propto \exp\left(\frac{1}{2v^2}(\mu - m)^2\right) \quad (1)$$

Derive the posterior:

$$p(\mu|D) \sim N(\mu_n, \sigma_n^2) \quad (2)$$

$$\mu_n = \sigma^2 \left( \frac{m}{v^2} + \frac{n\bar{x}}{\sigma^2} \right), \sigma_n^2 = \frac{1}{\frac{n\bar{x}}{\sigma^2} + \frac{1}{v^2}} \quad (3)$$

in which  $\bar{x}$  is the empirical mean ( $\bar{x} = \sum_{i=1}^n x_i$ ).

Answer:

$$\begin{aligned} p(\mu|D) &= \prod_{i=1}^n f(x_i|\mu)p(\mu) \\ &\propto \exp\left\{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2} - \frac{(\mu - m)^2}{2v^2}\right\} \\ &\propto \exp\left\{-\frac{\mu^2}{2}\left(\frac{n}{\sigma^2} + \frac{1}{v^2}\right) + \mu\left(\frac{\sum_{i=1}^n x_i}{\sigma^2} + \frac{m}{v^2}\right)\right\} \\ &\propto \exp\left\{-\frac{\mu^2}{2}\left(\frac{n}{\sigma^2} + \frac{1}{v^2}\right) + \mu\left(\frac{n\bar{x}}{\sigma^2} + \frac{m}{v^2}\right)\right\} \\ &\propto \exp\left\{-\frac{1}{2\sigma_n^2}(\mu - \mu_n)^2\right\} \end{aligned} \quad (4)$$

$$\frac{1}{\sigma_n^2} = \frac{1}{v^2} + \frac{n}{\sigma^2} \Rightarrow \sigma_n^2 = \frac{1}{\frac{1}{v^2} + \frac{n}{\sigma^2}} \quad (5)$$

$$\mu_n = \sigma_n^2 \left( \frac{m}{v^2} + \frac{n\bar{x}}{\sigma^2} \right) = \left( \frac{1}{\frac{1}{v^2} + \frac{n}{\sigma^2}} \right) \left( \frac{m}{v^2} + \frac{n\bar{x}}{\sigma^2} \right) \quad (6)$$

## 2 Problem 2

Consider a univariate Gaussian  $N(X|\mu, \lambda^{-1})$  and a dataset  $X = \{x_1, x_2, x_3, \dots, x_N\}$ ,  $x_i \sim N(X|\mu, \lambda^{-1})$  of i.i.d observations. Write down the likelihood for this model and show that a conjugate prior is of the form:

$$p(\mu, \lambda) \propto \left[ \lambda^{1/2} \exp\left(-\frac{\lambda\mu^2}{2}\right) \right]^\beta \exp(c\lambda\mu - d\lambda) \quad (7)$$

where  $c, d, \beta$  are constants. Given  $p(\mu, \lambda) = p(\mu | \lambda)p(\lambda)$ , we can recognize that the above prior is the product of a Gaussian  $p(\mu | \lambda)$  whose precision is a linear function of  $\lambda$  and of the gamma distribution  $p(\lambda)$ . As a result the normalized prior take the form:

$$p(\mu, \lambda) = N(\mu | \mu_0, (\beta\lambda)^{-1}) \text{Gam}(\lambda | a, b) \quad (8)$$

where  $\mu_0 = c/\beta, \alpha = 1 + \beta/2$  and  $b = d - c^2/2\beta$ . Show that the posterior  $p(\mu, \lambda | X)$  is also a Gaussian-gamma distribution of the same functional form as the prior:

$$N(\mu | \mu_N, ((N + \beta)\lambda)^{-1}) \text{Gam}(\lambda | a_N, b_N) \quad (9)$$

write down expressions for the parameters of this posterior distribution  $\mu_N, a_N$  and  $b_N$

Answer: the likelihood takes the form:

$$\begin{aligned} p(X|\mu, \lambda) &= \prod_{n=1}^N f(x_n|\mu) \propto \lambda^{N/2} \exp\left\{-\frac{1}{2}\lambda \sum_{n=1}^N (x_n - \mu)^2\right\} \\ &\propto \left[ \lambda^{1/2} \exp\left\{-\frac{\lambda\mu^2}{2}\right\} \right]^N \exp\left\{\lambda\mu \sum_{n=1}^N x_n - \frac{1}{2}\lambda \sum_{n=1}^N x_n^2\right\} \end{aligned} \quad (10)$$

we need a prior that has a similar functional form in terms of  $\lambda$  and  $\mu$ .

$$\begin{aligned} \beta &= N \\ c &= \sum_{n=1}^N x_n \\ d &= \frac{1}{2} \sum_{n=1}^N x_n^2 \\ p(\mu, \lambda) &\propto \left[ \lambda^{1/2} \exp\left(\frac{\lambda\mu^2}{2}\right) \right]^\beta \exp(c\lambda\mu - d\lambda) \end{aligned} \quad (11)$$

$$\begin{aligned}
p(\mu, \lambda) &\propto \left[ \lambda^{1/2} \exp\left(\frac{\lambda \mu^2}{2}\right) \right]^\beta \exp(c\lambda\mu - d\lambda) \\
&= (\beta\lambda)^{1/2} \exp\left\{-\frac{\beta\lambda}{2} \left(\mu - \frac{c}{\beta}\right)^2\right\} \lambda^{(\beta-1)/2} \exp\left\{-\left(d - \frac{c^2}{2\beta}\right) \lambda\right\} \\
&= p(\mu|\lambda)p(\lambda) \\
&= N\left(\mu|\mu_0 = \frac{c}{d}, \sigma = (\beta\lambda)^{-1}\right) \text{Gamma}\left(\lambda|a = \frac{1+\beta}{2}, b = d - \frac{c^2}{2\beta}\right)
\end{aligned} \tag{12}$$

$$\begin{aligned}
p(\mu, \lambda|X) &\propto \lambda^{N/2} \lambda^{a-1} \exp\left[-\left(b + \frac{1}{2} \sum_{n=1}^N x_n^2 + \frac{\beta}{2} \mu_0^2\right) \lambda\right] \times \\
&\quad [\lambda(N + \beta)]^{1/2} \exp\left\{-\frac{\lambda(N + \beta)}{2} \left[\mu^2 - \frac{2}{N + \beta} (\beta\mu_0 + \sum_{n=1}^N x_n) \mu\right]\right\} \\
&= N(\mu|\mu_N, [\lambda(N + \beta)]^{-1}) \text{Gam}(\lambda|a_N, b_N)
\end{aligned} \tag{13}$$

$$\begin{aligned}
\mu_N &= -\frac{\beta\mu_0 + \sum_{n=1}^N x_n}{N + \beta} \\
a_N &= \frac{N}{2} + a \\
b_N &= -\left(b + \frac{1}{2} \sum_{n=1}^N x_n^2 + \frac{\beta}{2} \mu_0^2 - \frac{(\beta\mu_0 + \sum_{n=1}^N x_n)^2}{2(N + \beta)}\right)
\end{aligned} \tag{14}$$

### 3 Problem 3

The Wishart distribution distribution  $W_m(n, \Sigma)$ .

1. To prove  $E[X|n, \Sigma] = n\Sigma$ .

$X = \sum_{i=1}^n z_i z_i^T \sim W_m(n, \Sigma)$ , where  $z_1, \dots, z_n \sim N_m(0, \Sigma)$ , and  $z_1, \dots, z_n$  are independent with each other.

$$E[X|n, \Sigma] = E\left[\sum_{i=1}^n z_i z_i^T\right] = \sum_{i=1}^n E[z_i z_i^T] = \sum_{i=1}^n \text{Var}(z_i) = n\Sigma \quad (15)$$

Therefore,  $E[X|n, \Sigma] = n\Sigma$ .

2. To prove  $\text{Cov}[X] = 2n\Sigma \otimes \Sigma$ .

$$\begin{aligned} \text{Cov}(X) &= \text{Cov}\left(\sum_{i=1}^n z_i z_i^T\right) \\ &= \sum_{i=1}^n \text{Cov}(z_i z_i^T) \\ &= n\text{Cov}(z_1 z_1^T) \\ &= n\text{Cov}(z_1 \otimes z_1) \\ &= n\text{Cov}((\Sigma \otimes \Sigma)(x \otimes x)) \\ &= 2n\Sigma \otimes \Sigma \end{aligned} \quad (16)$$

## 4 Problem 4

(a) Derive the log Gamma distribution  $Y \sim \log[X]$ ,  $X = G(\alpha, \beta)$  and show that it belongs to both families above.

Answer:  $Y \sim \log[X]$ ,  $X = G(\alpha, \beta)$ , therefore,  $X \sim \exp[Y]$ .

$$\begin{aligned} f_Y(x) &= f_{X=e^Y}(e^x) \frac{de^x}{dx} = \frac{1}{\Gamma(\alpha)} e^{-\beta e^x} (\beta e^x)^\alpha \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta e^x} (e^x)^\alpha \end{aligned} \quad (17)$$

Check whether it belongs to exponential family.

$$\begin{aligned} \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta e^x} (e^x)^\alpha &= \frac{\beta^\alpha}{\Gamma(\alpha)} \exp\{-\beta e^x + x\alpha\} \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \exp\{(-\beta, \alpha)(e^x, x)^T\} \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \exp\{\eta(\alpha, \beta)^T u(x)\} \end{aligned} \quad (18)$$

where  $u(x) = (e^x, x)^T$ ,  $\eta(\alpha, \beta) = (-\beta, \alpha)^T$ ,  $h(x) = 1$ ,  $g(\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}$ . Therefore, the log Gamma distribution  $Y$  belongs to the exponential family.

(b) Provide examples of two distributions that part of the exponential family.

Answer: Consider the Bernoulli Distribution

$$\begin{aligned} p(x|\mu) &= \mu^x (1 - \mu)^{1-x} \\ &= \exp\{x \ln \mu + (1 - x) \ln(1 - \mu)\} \\ &= (1 - \mu) \exp\left\{\ln\left(\frac{\mu}{1 - \mu}\right) x\right\} \end{aligned} \quad (19)$$

$$\begin{aligned} \eta &= \ln\left(\frac{\mu}{1 - \mu}\right) \\ \mu &= \sigma(\eta) = \frac{1}{1 + e^{-\eta}} \\ g(\eta) &= 1 - \mu = \frac{1}{1 + e^\eta} \\ p(x|\mu) &= g(\eta) \exp\{\eta x\}, \mu(x) = x, h(x) = 1, g(\eta) = 1 - \mu = \frac{1}{1 + e^\eta} \end{aligned} \quad (20)$$

Consider the Beta distribution

$$\begin{aligned} p(x|a, b) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} \\ &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \exp[(a-1)\ln(x) + (b-1)\ln(1-x)] \\ &= h(x)g(\eta) \exp\{\eta^T u(x)\} \end{aligned} \quad (21)$$

$$\begin{aligned}
\eta &= (a-1, b-1)^T \\
\mu(x) &= (\ln(x), \ln(1-x))^T \\
g(a, b) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \\
h(x) &= 1
\end{aligned} \tag{22}$$

(c) Provide examples of two distributions that are part of the local-scale family.

Answer:

Consider the Cauchy distribution: take  $\Psi(x) = \frac{1}{\pi} \frac{1}{1+x^2}$

$$p(x|a, b) = \frac{1}{b} \Psi\left(\frac{x+a}{b}\right) = \frac{1}{\pi b} \frac{1}{1 + \left(\frac{x+a}{b}\right)^2} \tag{23}$$

defines the Cauchy L-S family.

Consider the Uniform distribution: take  $\Psi(x) = I_{(0,1)}(x)$

$$I_{(0,1)}(x) = 1, \text{ if } 0 < x < 1, I_{(0,1)}(x) = 0, \text{ otherwise} \tag{24}$$

$\Psi(x)$  is the pdf of the Unif(0,1) distribution

$$p(x|a, b) = \frac{1}{b} \Psi\left(\frac{x+a}{b}\right) = \frac{1}{b} I_{(0,1)}\left(\frac{x+a}{b}\right) = \frac{1}{b} I_{(-a, -a+b)}(x) \tag{25}$$

This is the pdf of the Unif(-a, -a+b)

## 5 Problem 5

(a) Start by plotting a histogram of the camera's max resolution. Fit a normal distribution to this data and plot the fitted distribution on top of histogram. Repeat for the camera weight.

Answer: the camera's max resolution

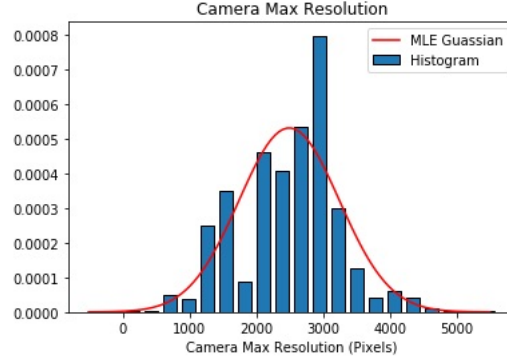


Figure 1: Histogram and normal distribution fitting of the camera's max resolution

the camera's weight

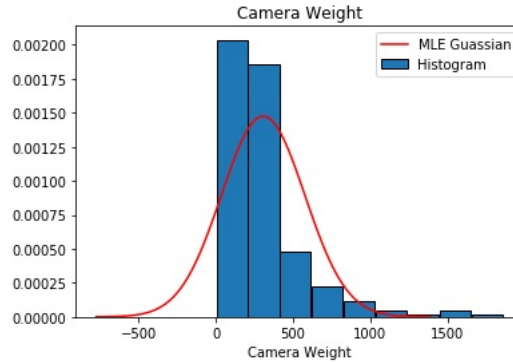


Figure 2: Histogram and normal distribution fitting of the camera's weight

(b) Standardize the camera max resolution data using the fitted mean,  $\bar{x}_i$ , and standard deviation,  $\sigma_i$ . This can be completed by computing  $(x_{ij} - \bar{x}_i)/\sigma_i$ . Plot a standard Gaussian  $N(0, 1)$  on top of the normalized data. Comment on why we would be interested in doing this.

Answer: from (a) the normal distribution fitting for the max resolution  $\bar{x}_i = 2491.7618497109806$ ,  $\sigma_i = 752.5256308818603$ . Then, we standardize the camera max resolution data by computing  $(x_i - \bar{x}_i)/\sigma_i$ .



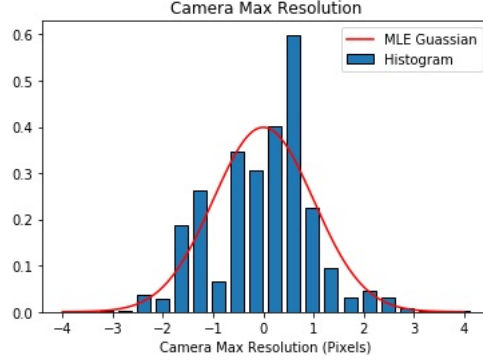


Figure 3: Standard Gaussian  $N(0,1)$  fitting of the normalized camera's max resolution

The reason that we would be interested in doing this is that the standardized data is easier to compare with each other.

(c) Fit a Gamma,  $G(\alpha, \beta)$ , distribution to the histogram of the camera's weight through MLE. Plot both fitted distributions on top of the histogram. Comment on which distribution provides a better fit.

start by finding the log-likelihood of Gamma distribution

$$f = \ln(T) = (\alpha - 1) \sum \ln(x) - n \ln(\Gamma(\alpha)) - n\alpha \ln(\beta) - \frac{1}{\beta} \sum x \quad (26)$$

$$\frac{d \ln(T)}{d\beta} = -\frac{n\alpha}{\beta} + \frac{1}{\beta^2} \sum x = 0 \quad (27)$$

$$\beta = \frac{\sum x}{n\alpha} = \frac{\bar{x}}{\alpha} \quad (28)$$

$$f(\alpha) = (\alpha - 1)n \ln(\bar{x}) - n \ln(\Gamma(\alpha)) - n\alpha \ln\left(\frac{\bar{x}}{\alpha}\right) - n\alpha \quad (29)$$

Using the minka-newton method from the paper provided.

$$f'(\alpha) = n \ln(\bar{x}) - n\Psi(\alpha) - n \ln(\bar{x}) + n \ln(\alpha) \quad (30)$$

$$f''(\alpha) = -n\Psi'(\alpha) + \frac{n}{\alpha} \quad (31)$$

$$\frac{1}{\alpha^{new}} = \frac{1}{\alpha^{old}} + \frac{f'(\alpha)}{\alpha^2 f''(\alpha)} = \frac{1}{\alpha^{old}} + \frac{\ln(\bar{x}) - \Psi(\alpha) - \ln(\bar{x}) + \ln(\alpha)}{-\alpha^2 \Psi'(\alpha) + \alpha} \quad (32)$$

From Figure 4, we find that the value of  $\alpha$  converges very quickly.  $\alpha \approx 1.84$  and  $\beta \approx 166.6387422858897$ .

Then we use the  $\alpha$  and  $\beta$  to plot the fitted Gamma distribution and compare with the Gaussian in Figure 2. After comparing with the Gaussian distribution

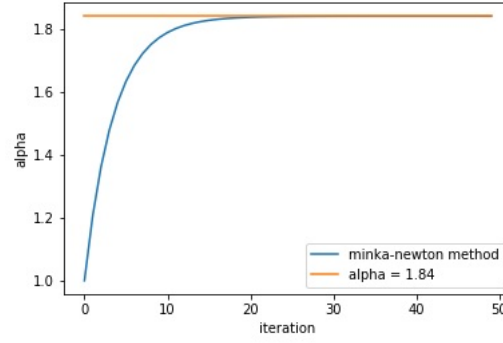


Figure 4: alpha value with minka-newton method

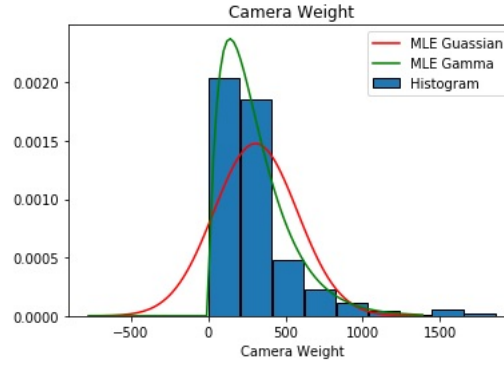


Figure 5: Gamma distribution and Gaussian distribution camera's weight

fitting, we find that the Gamma distribution is better than the Gaussian distribution for the camera's weight.

(d) Plot the camera's release year vs. max-resolution. Fit a 2D Gaussian distribution to this data through MLE. Plot the fitted Gaussian on top of the data.

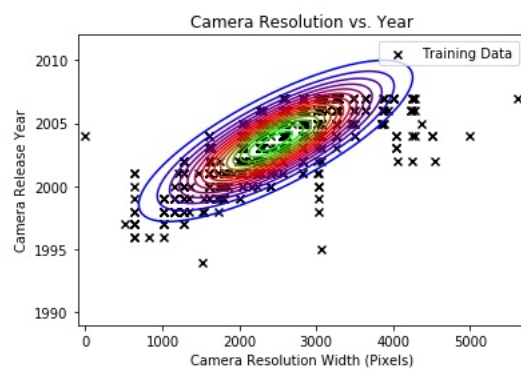


Figure 6: 2D Gaussian distribution camera's release year vs. max-resolution