TMA4295 Statistical inference
Exercise 8 - solution

Problem 1

$X \sim gamma(\alpha, \beta)$. Since

$$f(x|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} = \frac{1}{\Gamma(\alpha)\beta^\alpha} exp\left((\alpha-1)log(x) - \frac{1}{\beta}x\right)$$

it belongs to the exonential family with $h(x) = 1$, $c(\theta) = \frac{1}{\Gamma(\alpha)\beta^\alpha}$, $t_1(x) = log(x)$, $\omega_1 = \alpha - 1$, $t_2 = -x$, $\omega_2(\theta) = 1/\beta$. Using theorem 3.4.2 we have that

$$\begin{aligned} E(log(x)) &= -\frac{d}{d\alpha} log\left(\frac{1}{\Gamma(\alpha)\beta^\alpha}\right) \\ &= \frac{\Gamma(\alpha)'}{\Gamma(\alpha)} + log(\beta). \end{aligned}$$

Problem 2

$X_1, ..., X_n$ i.i.d. uniformly distributed on $[0, \theta]$.

a) The moment estimator is $\hat{\theta}_M = 2\bar{X}$ and it can't be written as a function of $T(X)$.

b) If $n = 3$ the moment estimator of θ is $\hat{\theta}_M = 6$. It is not reasonable since we have an observation with value 8.

c) We first derive the MLE for θ.

$$L(\theta|\mathbf{X}) = \prod_i f(X_i|\theta) = \frac{1}{\theta^n} \prod_i I_{[0.\theta]}(X_i) = \frac{1}{\theta^n} I_{[0.\theta]}(\max_i X_i)$$

We can observe that $L(\theta|\mathbf{X})$ is a decreasing function for $\theta > \max_i X_i$, so $L(\theta|\mathbf{X})$ is maximized at $\theta = \max_i X_i$. Hence $\hat{\theta}_{MLE} = \max_i X_i$. To compute the mean, variance and MSE, we first have to find the pdf of $T = max_i X_i$. Let's first look at the cdf

$$F_T(t) = P(T \leq t) = P(X_1 \leq t, ..., X_n \leq t) = \prod_i P(X_i \leq t) = \begin{cases} 0 & t < 0 \\ \left(\frac{t}{\theta}\right)^n & 0 \leq t \leq \theta \\ 1 & t > 1 \end{cases} \tag{1}$$

and so the pdf is the derivative of 1

$$f_T(t) = \frac{nt^{n-1}}{\theta^n} \quad \text{if } 0 \leq t \leq \theta.$$

Then we easily get

$$\begin{aligned} E(T) &= \frac{n}{1+n}\theta \\ Var(T) &= \frac{n}{(1+n)^2(2+n)}\theta^2 \\ MSE(T) = Var(T) + Bias(T)^2 &= \frac{2}{(1+n)(2+n)}\theta^2. \end{aligned}$$

d) The unbiased estimator is given by $\hat{\theta} = \frac{1+n}{n} T(\mathbf{X})$, with variance $Var(\hat{\theta}) = \frac{\theta^2}{n(n+1)}$ which is also equal to the mean squares error for this estimator.

The moment estimator is unbiased with variance equal to $\frac{\theta^2}{3n}$ which is also the mean squared error.

For $n = 1$ all the estimators are the same.

For $n = 2$ $MSE(\hat{\theta}_{MLE}) = MSE(\hat{\theta}_M) > MSE(\hat{\theta})$.

For $n = 3$ $MSE(\hat{\theta}_M) > MSE(\hat{\theta}_{MLE}) > MSE(\hat{\theta})$.

Problem 3

$X_1, ..., X_n$ are i.i.d. $N(\mu, \mu^2)$, where μ need to be estimated.

Moment method:

Using the first moment gives $\hat{\mu}_M = \bar{\mathbf{X}}$. Using the second moment gives $\hat{\mu}_M = \sqrt{\frac{1}{2n}\sum_i X_i^2}$. Combination of these two estimators can give better estimator.

Maximum likelihood method:

Maximum likelihood estimator is given by $\hat{\mu}_{MLE} = \frac{-\sum_i X_i \pm \sqrt{\left(\sum_i X_i\right)^2 + 4\sum_i X_i^2}}{2n}$. There are two maximums, one global and one local, and it is necessary to check the values of the likelihood to decide which one is the global maximum.

Maximum likelihood method for exponential families:

The assumptions for use of the general results for MLE in exponential families are not fulfilled.

Problem 4

a) We first find the pdf of $\mathbf{X}|\theta$

$$f(\mathbf{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left(\frac{-\sum_i x_i^2 - n\theta^2 + 2n\theta\bar{x}}{2\sigma^2}\right).$$

Then using the fact that $f(\theta|\mathbf{X}) \propto f(\mathbf{X}|\theta)f(\theta)$ and trying to form the exponent of the form of the normal distribution we get the result.

b) The conjugate prior for the normal distribution is the normal distribution.

c) The Bayes estimator of θ is $E(\theta|\mathbf{X}) = \frac{\sigma^2}{\sigma^2+n\tau^2}m + \frac{n\tau^2}{\sigma^2+n\tau^2}\bar{\mathbf{x}}$, which is a linear combination of the prior and sample means.

From the form of the Bayes estimator it can be seen that if the prior information is unsure, i.e. τ^2 is big, then the influence of m is weak (the influence of the prior is weak). If the variance of the sample is big, i.e. σ^2 is big, the influence of the sample is weak.