University of Stavanger

EXAM IN: STA500 INTRODUCTION TO PROBABILITY AND STATISTICS 2

DURATION: 4 HOURS DATE: February 17, 2016

PERMITTED AIDS: Approved simple calculator (HP30S, Casio FX82, TI-30, Citizen SR-270X, Texas BA II Plus or HP17bII+).

THE EXAM CONSISTS OF 3 PROBLEMS ON 2 PAGES, 9 PAGES OF ENCLOSURES.

COURSE RESPONSIBLE: Tore Selland Kleppe PHONE: 51 83 17 17

Problem 1: A seismologist is monitoring the location of earth quakes in a 100 km times 100 km square region in a mountain range. A location in the region is given in (x, y) coordinates so that $(0,0), (0,1), (1,0)$ and $(1,1)$ defines the four corners of the square region and distance is measured in unit 100 km. The location (X, Y) of each earth quake is assumed to be a bivariate random variable with probability density function given by

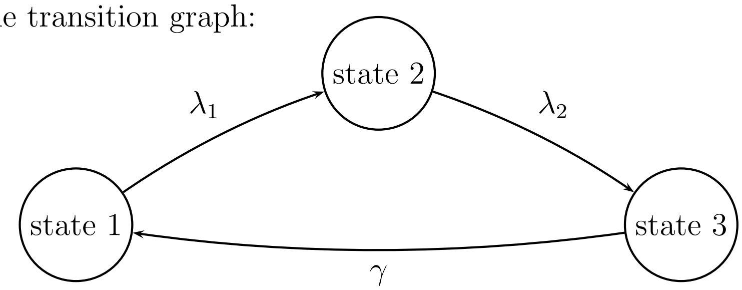
$$f(x,y) = \frac{200}{151}\left(1 - (x-0.3)^2 - \frac{(y-0.9)^2}{2}\right),\ 0 \le x, y \le 1.$$

a) Show that the marginal distribution of the x-coordinate has the density $f(x) = \frac{527}{453} - \frac{200}{151}(x-0.3)^2,\ 0 \le x \le 1.$

The seismologist has a chain of seismometers running in the the x-direction that can measure exactly the x-coordinate of any given earthquake. Suppose an earth quake occurred with x-coordinate equal to 0.3.

b) Find the density of $Y|X = 0.3$.
Are X and Y independent?

Problem 2: Consider the three state continuous time Markov chain $\{X(t), t \ge 0\}$ specified by the transition graph:



where the specific transition rates are $0 < \lambda_1, \lambda_2, \gamma < \infty$.

a) Why does this model admit unique steady state probabilities?
Find the steady state probabilities.

Assume that $X(0) = 1$.

b) What is the expected time until the chain first visits state 3?
   At what time $\tau(u)$ is there a probability u that the chain has left state 1 for the first time?
   Suppose now that it is known that the chain remained in state 1 until time 1. What is the expected time after time 1 the chain will remain in state 1?

Problem 3: Let $X_1, X_2, \ldots, X_n$ be independent data from a distribution with probability density function

$$f(x) = \sqrt{\frac{\tau}{2\pi}} \exp\left(-\frac{\tau}{2}x^2\right), \; -\infty < x < \infty,$$

i.e. X_i has a normal distribution with $E(X_i) = 0$ and $Var(X_i) = \frac{1}{\tau}$. The parameter τ is known as the precision parameter.

a) Write down the log-likelihood function for the parameter τ based on data $X_1, X_2, \ldots, X_n$.
   Show that $\hat{\tau} = \frac{n}{\sum_{i=1}^n X_i^2}$ is the maximum likelihood estimator for τ.
   Comment on the maximum likelihood estimator in light of the invariance principle.
   Find a 95% Wald approximate confidence interval for τ.

b) Argue for why $h(\hat{\tau}, \tau) = \frac{n\tau}{\hat{\tau}}$ has a χ_n^2 distribution.

c) Find an exact $(1-\alpha)100\%$ confidence interval for τ.
   What is the interpretation of a $(1-\alpha)100\%$ confidence interval?

Now consider Bayesian estimation of τ (in the same situation as above) using a gamma(a, b) prior.

d) Find the posterior distribution of τ.
   Find the Bayes estimator $\hat{\tau}_{Bayes}$.

e) Find a $(1-\alpha)100\%$ credible interval (Bayes interval) for τ.
   Compare the interpretation of the credible interval to that of the confidence interval found in c).

Suppose now that we have $n = 6$ observations $x_1, x_2, \ldots, x_6$ and that $\sum_{i=1}^n x_i^2 = 4.451109$.

f) Compute the Wald and exact confidence intervals based on the data mentioned above with $\alpha = 0.05$.
   Compute the credible interval based on the data mentioned above, a gamma(10,0.1) prior and $\alpha = 0.05$.
   Comment on the differences between the different intervals.

Solutions

1 a)
To find the x-marginal, we integrate the joint density wrt y, which h,i.e

$$f(x) = \int f(x,y)dy = (200/151)\left[(17/60) + \left(119/200 - (x-3/10)^2\right)\right]$$
$$= 527/453 - (200/151)(x-3/10)^2.$$

1 b)
The conditional density is given in terms of the joint and marginal-x densities:

$$f(y|x) = f_{X,Y}(x,y)/f_X(x) = \frac{\frac{200}{151}\left(1-(x-0.3)^2 - \frac{(y-0.9)^2}{2}\right)}{527/453 - (200/151)(x-3/10)^2.}$$

Specifically conditioning on $X = 0.3$ gives us that

$$f(y|x=0.3) = \frac{\frac{200}{151}\left(1-\frac{(y-0.9)^2}{2}\right)}{527/453.} = 600/527 - (300/527)(y-0.9)^2.$$

The two variables are dependent as $f(y|x)$ depends on x.
2 a)
The model is Irreducible as all states communicate and Positive recurrent as each state is revisited with probability 1 in finite expected time.
The steady state probabilities solve the equations

$$\begin{aligned}
\gamma\pi_3 - \lambda_1\pi_1 &= 0\\
\lambda_1\pi_1 - \lambda_2\pi_2 &= 0\\
\lambda_2\pi_2 - \gamma\pi_3 &= 0\\
\pi_1 + \pi_2 + \pi_3 &= 1
\end{aligned}$$

where one of the upper three is superfluous. Using the upper two an the latter we obtain that

$$\begin{aligned}
\pi_3 &= (\lambda_1/\gamma)\pi_1\\
\pi_2 &= (\lambda_1/\lambda_2)\pi_1\\
1 &= (1 + (\lambda_1/\lambda_2) + (\lambda_1/\gamma))\pi_1\\
&\Downarrow\\
\pi_1 &= \frac{1}{1 + (\lambda_1/\lambda_2) + (\lambda_1/\gamma)}\\
&= \frac{\gamma\lambda_2}{\gamma\lambda_2 + \gamma\lambda_1 + \lambda_1\lambda_2}
\end{aligned}$$

Then plugging into the first two equations:

$$\begin{aligned}
\pi_2 &= (\lambda_1/\lambda_2)\pi_1 = \frac{\gamma\lambda_1}{\gamma\lambda_2 + \gamma\lambda_1 + \lambda_1\lambda_2}\\
\pi_3 &= (\lambda_1/\gamma)\pi_1 = \frac{\lambda_1\lambda_2}{\gamma\lambda_2 + \gamma\lambda_1 + \lambda_1\lambda_2}
\end{aligned}$$

2 b)
The time until the chain first enters state 3 is distributed as the sum of independent exponential($1/\lambda_1$) and exponential($1/\lambda_2$) random variables, and therefore the expected time until the chain enters state 3 is $1/\lambda_1 + 1/\lambda_2$.

The time T until the chain jumps from state 1 has an exponential($1/\lambda_1$) distribution, and therefore we have that $F_T(t) = 1 - \exp(-t\lambda_1)$. Now, we wish to solve for t so that

$$P(T > \tau) = 1 - F_T(\tau) = \exp(-\tau\lambda_1) = u$$

which gives us $\tau(u) = -\log(u)/\lambda_1$.
The answer to the last question is still $1/\lambda_1$ due to the memoryless-property of the exponential distribution.
3 a)
Likelihood

$$L(\tau) = \prod_{i=1}^{n} \sqrt{\frac{\tau}{2\pi}} \exp\left(-\frac{\tau}{2}X_i^2\right) = \left(\frac{\tau}{2\pi}\right)^{\frac{n}{2}} \exp\left(-\frac{\tau}{2}\sum_{i=1}^{n} X_i^2\right).$$

Log-likelihood

$$l(\tau) = \frac{n}{2}\log(\tau) - \frac{n}{2}\log(2\pi) - \frac{\tau}{2}\sum_{i=1}^{n} X_i^2.$$

First derivative (score)

$$\frac{\partial}{\partial\tau} l(\tau) = \frac{n}{2\tau} - \frac{1}{2}\sum_{i=1}^{n} X_i^2.$$

Critical point and MLE:

$$\begin{aligned} 0 &= \frac{n}{2\hat{\tau}} - \frac{1}{2}\sum_{i=1}^{n} X_i^2 \\ &\Downarrow \\ \hat{\tau} &= \frac{n}{\sum_{i=1}^{n} X_i^2}. \end{aligned}$$

Second derivative at (candidate) MLE:

$$\frac{\partial^2}{\partial\tau^2} l(\hat{\tau}) = -\frac{n}{2\hat{\tau}^2} = -\frac{\left(\sum_{i=1}^{n} X_i^2\right)^2}{2n} < 0$$

I.e. the second derivative at the MLE is negative, and therefore $\hat{\tau}$ is a maximizer.
This estimator is the inverse of the MLE for σ^2 in a $N(0, \sigma^2)$ population and therefore in line with the invariance principle.
The 95 % Wald interval is given as

$$\left[\hat{\tau} \mp 1.96\sqrt{\frac{2\hat{\tau}^2}{n}}\right] = \left[\hat{\tau} \mp 1.96\sqrt{\frac{2n}{\left(\sum_{i=1}^{n} X_i^2\right)^2}}\right].$$

3 b)

Let $\sigma = \tau^{-1/2}$ be the standard deviation of X_i. Then observe that we can rewrite h as

$$h(\hat{\tau}, \tau) = \frac{n\tau}{\hat{\tau}} = \frac{n\tau \sum_{i=1}^{n} X_i^2}{n} = \frac{1}{\sigma^2} \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} \underbrace{\left(\underbrace{\frac{X_i}{\sigma}}_{\sim N(0,1)} \right)^2}_{\sim \chi_1^2}.$$

Thus $h(\hat{\tau}, \tau)$ is a sum of n independent (as the X_is are independent) χ_1^2 variables and must therefore have a χ_n^2-distribution.

3 c)

Using the known distribution of $h(\hat{\tau}, \tau)$, we have that

$$\begin{aligned} P\left(\chi^2_{1-\alpha/2,n} < \frac{n\tau}{\hat{\tau}} < \chi^2_{\alpha/2,n}\right) &= 1 - \alpha \\ &\Downarrow \\ P\left(\chi^2_{1-\alpha/2,n} \frac{\hat{\tau}}{n} < \tau < \chi^2_{\alpha/2,n} \frac{\hat{\tau}}{n}\right) &= 1 - \alpha \end{aligned}$$

I.e. the confidence interval is given by

$$\left[\chi^2_{1-\alpha/2,n} \frac{\hat{\tau}}{n},\ \chi^2_{\alpha/2,n} \frac{\hat{\tau}}{n}\right].$$

The interpretation is that the confidence interval covers the true parameter in a fraction $(1-\alpha)$ of repeated experiments. I.e. it is not meaningful to talk about e.g. the probability of the true parameter being in the interval is $(1-\alpha)$ in this non-Bayesian setting.

3 d)

The likelihood function was found in 3 a) and we therefore have that the posterior kernel can be written as

$$\begin{aligned} p(\tau|\text{data}) &\propto \underbrace{\left(\frac{\tau}{2\pi}\right)^{\frac{n}{2}} \exp\left(-\frac{\tau}{2}\sum_{i=1}^{n} X_i^2\right)}_{L(\tau)} \underbrace{\tau^{a-1} \exp\left(-\frac{\tau}{b}\right)}_{\propto p(\tau)}, \\ &\propto \tau^{n/2+a-1} \exp\left(-\tau\left[\frac{1}{2}\sum_{i=1}^{n} X_i^2 + \frac{1}{b}\right]\right) \end{aligned}$$

The posterior is recognized to be a gamma(a^*, b^*) distribution where

$$\begin{aligned} a^* &= \frac{n}{2} + a, \\ b^* &= \left[\frac{1}{2}\sum_{i=1}^{n} X_i^2 + \frac{1}{b}\right]^{-1}. \end{aligned}$$

Due to the gamma-distributed posterior, the Bayes estimator is given as

$$\hat{\tau}_{Bayes} = a^* b^* = \frac{\frac{n}{2} + a}{\frac{1}{2}\sum_{i=1}^{n} X_i^2 + \frac{1}{b}}.$$

3 e)
The credible interval is found using the relationship between the gamma and χ^2 distributions:

$$
\begin{aligned}
P\left(\gamma_{1-\alpha/2,a^*,b^*} < \tau < \gamma_{1-\alpha/2,a^*,b^*}\right) &= 1-\alpha, \\
&\Downarrow \\
P\left(\underbrace{\frac{2}{b^*}\gamma_{1-\alpha/2,a^*,b^*}}_{=\chi^2_{1-\alpha/2,2a^*}} < \underbrace{\frac{2}{b^*}\tau}_{\sim\chi^2_{2a^*}} < \underbrace{\frac{2}{b^*}\gamma_{\alpha/2,a^*,b^*}}_{=\chi^2_{\alpha/2,2a^*}}\right) &= 1-\alpha.
\end{aligned}
$$

Thus $\gamma_{1-\alpha/2,a^*,b^*} = b^*\chi^2_{1-\alpha/2,2a^*}/2$ and $\gamma_{\alpha/2,a^*,b^*} = b^*\chi^2_{\alpha/2,2a^*}/2$ and the credible interval can be written as

$$
\left[\frac{b^*\chi^2_{1-\alpha/2,2a^*}}{2},\ \frac{b^*\chi^2_{\alpha/2,2a^*}}{2}\right].
$$

3 f)
Maximum likelihood estimator:

$$
\hat{\tau} = \frac{n}{\sum_{i=1}^n x_i^2} = 1.347979
$$

Wald interval:

$$
\left[\hat{\tau} \mp 1.96\sqrt{\frac{2\hat{\tau}^2}{n}}\right] = [-0.1774028, 2.8733601].
$$

Exact confidence interval:

$$
\left[\chi^2_{1-\alpha/2,n}\frac{\hat{\tau}}{n},\ \chi^2_{\alpha/2,n}\frac{\hat{\tau}}{n}\right] = [0.2779856, 3.2462416].
$$

Credible interval ($a^* = 13,\ b^* = 0.08179588$)

$$
\left[\frac{b^*\chi^2_{1-\alpha/2,2a^*}}{2},\ \frac{b^*\chi^2_{\alpha/2,2a^*}}{2}\right] = [0.5661872, 1.7145713]
$$

It is seen that the Wald interval contains negative values, which is undesirable and show that it is a rather poor approximation for small n.
Comparing the exact confidence interval and the credible interval we see that the credible interval is much narrower. This is a consequence of the fact that both data and prior information is used for the credible interval, whereas only data is used for the confidence interval. In particular, the prior more informative than the data, as the posterior standard deviation is only approximately 7 % smaller than the prior standard deviation.