University of Stavanger

EXAM IN: STA500 INTRODUCTION TO PROBABILITY AND STATISTICS 2

DURATION: 4 HOURS DATE: February 13th, 2019.

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

THE EXAM CONSISTS OF 4 PROBLEMS ON 3 PAGES, 18 PAGES OF ENCLOSURES.

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

Note: Throughout this exam, all logarithms are natural logarithms, so that $\log(x) = \ln(x)$, and 10-based logarithms are not used.

Problem 1:

Suppose $X \sim N(0,1)$ and define $Y = \exp(-2(X+1))$.

a) Find the probability density function of Y. Which distribution does Y have?
Find $E(Y)$ and $Var(Y)$.
Find also $P(Y > 1)$.

Consider a situation where $X \sim N(0,1)$ and $Y = X + \eta$, $\eta \sim N(0,1)$, where X and η are independent. The joint probability density function of X and Y is given as:

$$f_{X,Y}(x,y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y-x)^2}{2}\right).$$

You might find useful that $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X,Y}(x,y) dx dy = 1$.

b) Argue for why the marginal distribution of Y is $N(0,2)$.
Find the covariance $Cov(X,Y)$ and correlation $corr(X,Y)$.
Are X and Y independent?

Problem 2: Consider a situation where we have n independent observations $Y_i,\ i = 1, \dots, n$. Each observation Y_i has a gamma distribution with shape parameter $\alpha = 4$ and scale parameter $\beta = \theta x_i$, where $x_i > 0$ are a known, non-random quantities and $\theta > 0$ is the parameter of interest.[1] The probability density function of Y_i is given as

$$f_{Y_i}(y_i) = \frac{1}{(\theta x_i)^4 \Gamma(4)} y_i^3 \exp\left(-\frac{y_i}{\theta x_i}\right).$$

a) Find the log-likelihood function for θ and show that the maximum likelihood estimator is given as $\hat{\theta} = \frac{1}{4n} \sum_{i=1}^{n} \frac{Y_i}{x_i}$

b) Verify that $\hat{\theta}$ given above is a consistent estimator for θ as $n \to \infty$.

c) Which distribution does $\frac{8n\hat{\theta}}{\theta}$ have?

Problem 3: Consider a discrete time Markov chain with state space $\{0, 1\}$ with transition probability matrix

$$P = \begin{pmatrix} 0 & 1 \\ (1-\alpha) & \alpha \end{pmatrix}$$

where $0 < \alpha < 1$ is a parameter.

a) Explain why this Markov chain admit steady state probabilities.
Find the steady state probabilities.

Assuming that we have available a realisation $X_1, X_2, \dots, X_T$ of the discrete time Markov chain with transition probability matrix P above, and wish to estimate α. I.e. X_t is the state the Markov chain was in at time t, for $t = 1, \dots, T$. The likelihood function for α may be written as

$$L(\alpha | X_1, \dots, X_T) = \prod_{t=2}^{T} \left(\alpha^{X_t} (1-\alpha)^{1-X_t}\right)^{X_{t-1}}$$

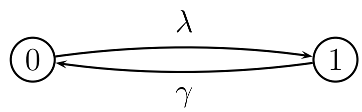
b) Find the maximum likelihood estimator for α.
Comment on why the maximum likelihood estimator you obtain makes sense.

Now, consider Bayesian estimation of α using a uniform (i.e. Beta(1,1))-prior.

c) Find the posterior distribution of α given $X_1, \dots, X_T$.
Find the Bayes estimator, and compare it to the maximum likelihood estimator.

[1] E.g. you may think of Y_i as the life time of component i, when component i is subject to some stress level x_i.

Problem 4: Consider the continuous time Markov chain $X(t)$ with state space $\mathcal{S} = \{0, 1\}$ specified in terms of the graph



a) Write down specific transition rates (q_{01}, q_{10}) and total rates out (ν_0, ν_1)?
Find the steady state probabilities for this process.

It can be shown that the full set of transition probabilities $p_{ij}(t)$ for this process have the form

$$\begin{aligned} p_{00}(t) &= \frac{\gamma}{\lambda+\gamma} + \frac{\lambda}{\lambda+\gamma} \exp(-(\lambda+\gamma)t), \\ p_{10}(t) &= \frac{\gamma}{\lambda+\gamma} \left[1 - \exp(-(\lambda+\gamma)t)\right], \\ p_{01}(t) &= \frac{\lambda}{\lambda+\gamma} \left[1 - \exp(-(\lambda+\gamma)t)\right], \\ p_{11}(t) &= \frac{\lambda}{\lambda+\gamma} + \frac{\gamma}{\lambda+\gamma} \exp(-(\lambda+\gamma)t). \end{aligned}$$

b) For $\lambda = 1$, $\gamma = 2$, calculate the following probabilities:
$P(X(1) = 0)$ (i.e. the unconditional probability of being in state 0).
$P(X(1) = 0 | X(0) = 0)$.
$P(X(3) = 0, X(1) = 1 | X(0) = 1)$.
$P(X(t) = 0 \; \forall \, t \in (0,1) | X(0) = 0)$.
(The last probability may be interpreted as the probability that the process remains in state 0 all times between time $t = 0$ and $t = 1$, given that it is in state 0 at time $t = 0$.)

Solutions

1,a)
From first principles using transformation formula; $y = u(x) = \exp(-2(x+1)) \Rightarrow -1 - \log(y)/2 = w(y) = x$. $|w'(y)| = 1/(2y)$, which gives the density

$$f_Y(y) = \frac{1}{\sqrt{2\pi}2y} \exp\left(-\frac{1}{2}(-\log(y)/2 - 1)^2\right)$$

This is recognized to be a log-normal distribution with parameters $\sigma = 2$, $\mu = -2$. Alternatively, it is OK to recognise that $-2(x+1) \sim N(-2, 2^2)$ and use knowledge of the relation between normal and log-normal distributions.
The mean and variance obtain from the formulas for the log-normal distribution
$E(Y) = \exp(\mu + \sigma^2/2) = \exp(-2 + 4/2) = 1$
$Var(Y) = \exp(2(\mu + \sigma^2)) - \exp(2\mu + \sigma^2) = \exp(4) - 1$
To find the sought probability, we have that $P(Y > 1) = P(-2(X+1) > log(1))$ $= P(-2X - 2 > 0) = P(-2X > 2) = P(X < -1) \approx 0.1587$, (numerical answer from standard normal table).
1,b)
The random variable Y obtains as the sum of two Normal random variables and is therefore also Normal. The distribution of Y is therefore fully characterised by the mean and variance:
$E(Y) = E(X) + E(\eta) = 0$; $Var(X) = Var(X) + Var(\eta) = 2 \Rightarrow Y \sim N(0, 2)$.
The covariance is defined as $(E(X - E(X))(Y - E(Y)))$. From the information, it is clear that both marginal expectations are zero, thus $Cov(X, Y) = E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{X,Y}(x, y)dxdy = 1$. The correlation is given as $\rho = Cov(X, Y)/\sqrt{Var(X)Var(Y)} = 1/\sqrt{2}$. Since the variables have non-zero covariance, they are dependent. (alternatively, it is OK to show that the joint density does not factorise as the product of marginal densities.)
2,a)
Likelihood function:

$$L(\theta|Y_1, \dots, Y_n) = \prod_{i=1}^{n} \frac{1}{(\theta x_i)^4\Gamma(4)} Y_i^3 \exp\left(-\frac{Y_i}{\theta x_i}\right).$$

$$\propto \theta^{-4n} \exp\left(-\frac{1}{\theta}\sum_{i=1}^{n}\frac{Y_i}{x_i}\right)$$

Log likelihood function:

$$l(\theta) = \text{constant} - 4n\log(\theta) - \frac{1}{\theta}\sum_{i=1}^{n}\frac{Y_i}{x_i}$$

Derivative (score function)

$$\frac{\partial}{\partial\theta}l(\theta) = -\frac{4n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n}\frac{Y_i}{x_i}$$

Multiply score by θ^2 and set equal to zero:

$$-4n\theta + \sum_{i=1}^{n}\frac{Y_i}{x_i} = 0$$

$$\Downarrow$$

Potential MLE:

$$\hat{\theta} = \frac{1}{4n}\sum_{i=1}^{n}\frac{Y_i}{x_i}$$

Check second derivative of log-likelihood

$$\frac{\partial^2}{\partial\theta^2}l(\theta) = \frac{4n}{\theta^2} - \frac{2}{\theta^3}\sum_{i=1}^{n}\frac{Y_i}{x_i}$$

(not obvious that this is uniformly negative, thus) Evaluating second derivative at critical point (via some straight forward algebra):

$$\frac{\partial^2}{\partial\theta^2}l(\hat{\theta}) = -\frac{64n^3}{\left(\sum_{i=1}^{n}\frac{Y_i}{x_i}\right)^2}$$

I.e. negative curvature around critical point $\hat{\theta} \Rightarrow \hat{\theta}$ is MLE.
2,b)
First, it is clear that $E(Y_i) = \alpha\beta = 4x_i\theta$ and $Var(Y_i) = \alpha\beta^2 = 4x_i^2\theta^2$. Thus,

$$E(\hat{\theta}) = E\left(\frac{1}{4n}\sum_{i=1}^{n}\frac{Y_i}{x_i}\right) = \frac{1}{4n}\sum_{i=1}^{n}\underbrace{E(\frac{Y_i}{x_i})}_{=4\theta} = \theta.$$

I.e. the estimator is unbiased. Now, to show consistency, the variance must vanish as $n \to \infty$:

$$Var(\hat{\theta}) = \frac{1}{16n^2}Var\left(\sum_{i=1}^{n}\frac{Y_i}{x_i}\right)\underbrace{=}_{indep.}\frac{1}{16n^2}\sum_{i=1}^{n}\underbrace{\frac{Var(Y_i)}{x_i^2}}_{=4\theta^2} = \frac{\theta^2}{4n}.$$

Since the estimator is unbiased and the variance vanishes as $n \to \infty$ the estimator is consistent.
2,c)
First notice that

$$\frac{8n\hat{\theta}}{\theta} = \frac{8n}{\theta}\frac{1}{4n}\sum_{i=1}^{n}\frac{Y_i}{x_i} = \frac{2}{\theta}\sum_{i=1}^{n}(Y_i/x_i).$$

From the text, it is clear that Y_i has a gamma distribution with shape parameter $\alpha = 4$ and scale parameter $\beta = x_i\theta$, thus, via transformation from general gamma to χ^2 that $2Y_i/\beta = (2/\theta)(Y_i/x_i) \sim \chi^2_8$. From this information it is clear (via sum of independent χ^2 RVs) that $\frac{8n\hat{\theta}}{\theta} = (2/\theta)\sum_{i=1}^{n}(Y_i/x_i) \sim \chi^2_{8n}$.
3,a)
The process is irreducible (both states communicate with each other) and is aperiodic (as $P_{11} > 0$). The steady state probs π obtain by solving $P^T\pi = \pi$ along with $\sum_i \pi_i = 1$. Taking the first equation and the normalisation condition, we obtain

$$(1-\alpha)\pi_1 = \pi_0,\ \pi_0 + \pi_1 = 1\ \Downarrow$$

$$(1-\alpha)\pi_1 + \pi_1 = 1\ \Rightarrow \pi_1 = 1/(2-\alpha)$$

$$\pi_0 = 1 - 1/(2-\alpha) = (1-\alpha)/(2-\alpha)$$

3,b)
Likelihood function rewritten

$$L(\alpha) = \prod_{t=2}^{T} (\alpha^{X_t}(1-\alpha)^{(1-X_t)})^{X_{t-1}}$$

Log likelihood function

$$l(\alpha) = \sum_{t=2}^{T} X_{t-1}\left(X_t \log(\alpha) + (1-X_t)\log(1-\alpha)\right)$$

Derivative (score function)

$$\frac{\partial}{\partial\alpha} l(\alpha) = \sum_{t=2}^{T} X_{t-1}\left(\frac{X_t}{\alpha} - \frac{1-X_t}{1-\alpha}\right)$$

Multiply score by $\alpha(1-\alpha)$ and set equal to zero:

$$\alpha \sum_{t=2}^{T} X_{t-1}(1-X_t) = (1-\alpha)\sum_{t=2}^{T} X_{t-1}X_t$$

$$\Downarrow$$

$$\hat{\alpha} = \frac{\sum_{t=2}^{T} X_{t-1}X_t}{\sum_{t=2}^{T} X_{t-1}}$$

Now check second derivative

$$\frac{\partial^2}{\partial\alpha^2} l(\alpha) = \sum_{t=2}^{T} X_{t-1}\left(-\frac{X_t}{\alpha^2} - \frac{1-X_t}{(1-\alpha)^2}\right)$$

Since X_t only take values 0 or 1, the second derivative is uniformly negative: $\hat{\alpha}$ above is MLE.
Comment: We see that the MLE count the proportion of times the process remains in state 1 (i.e. $X_{t-1}X_t = 1$), given that the process was in state 1 in time $t-1$. In the theoretical model, the probability of this occurring is α, thus we are estimating a probability with the corresponding observed frequency.
3,c)
The prior is flat, so the shape of the posterior is that of the likelihood function, slightly rewritten:

$$p(\alpha|X_1,\ldots,X_T) \propto \prod_{t=2}^{T} \alpha^{X_{t-1}X_t}(1-\alpha)^{X_{t-1}(1-X_t)} = \alpha^{\sum_{t=2}^{T} X_{t-1}X_t}(1-\alpha)^{\sum_{t=2}^{T} X_{t-1}(1-X_t)}$$

This is recognised to be a Beta$(1+\sum_{t=2}^{T} X_{t-1}X_t, 1+\sum_{t=2}^{T} X_{t-1}(1-X_t))$ kernel, so the posterior distribution is Beta with parameters given above.
The posterior mean/Bayes estimator is found, using the formula for the Beta distribution mean, to be

$$\hat{\alpha}_{\text{Bayes}} = \frac{1+\sum_{t=2}^{T} X_{t-1}X_t}{1+\sum_{t=2}^{T} X_{t-1}X_t + 1 + \sum_{t=2}^{T} X_{t-1}(1-X_t)} = \frac{1+\sum_{t=2}^{T} X_{t-1}X_t}{2+\sum_{t=2}^{T} X_{t-1}}$$

It is seen that the Bayes estimator is similar to the MLE for large samples, and converges to the MLE when $T \to \infty$. For small samples, the Bayes estimator is "pulled" towards the prior mean, namely 0.5.

4,a)

First, $q_{01} = \nu_0 = \lambda$, $q_{10} = \nu_1 = \gamma$.

To get the steady state probabilities we consider the "flow" conservation equation for state 0 and the normalising relation:

$$0 = q_{10}\pi_1 - \nu_0\pi_0 = \gamma\pi_1 - \lambda\pi_0 \text{ and } \pi_0 + \pi_1 = 1$$

These results in

$$\pi_1 = \frac{\lambda}{\gamma}\pi_0 \;\Rightarrow (1 + \frac{\lambda}{\gamma})\pi_0 = 1 \;\Rightarrow\; \pi_0 = \frac{\gamma}{\lambda + \gamma}$$

$$\pi_1 = 1 - \pi_0 = \frac{\lambda}{\lambda + \gamma}$$

4,b)

First, marginal / steady state prob $P(X_1 = 0) = \pi_0 = 1/3 = 0.333333333$

Second; $P(X_1 = 0|X_0 = 0) = p_{00}(1) = 2/3 + 1/3\exp(-3) = 0.6832624$

Third; $P(X_3 = 0, X_1 = 1|X_0 = 1) = p_{11}(1) \times p_{10}(2)$

$= (1/3 + 2/3\exp(-3)) \times 2/3(1 - \exp(-6)) = 0.2437441$

Fourth; The sought probability is 1 minus the probability that the process leaves state 0 in $t \in (0, 1)$. The time the process spends in state 0 (until it leaves the first time) is exponentially distributed with mean $1/\lambda = 1$, thus, the sought probability is $1 - \int_0^1 \exp(-t)dt = \exp(-1) = 0.36787944$.