University of Stavanger

EXAM IN: STA500 INTRODUCTION TO PROBABILITY AND STATISTICS 2

DURATION: 4 HOURS DATE: NOVEMBER 30th, 2018

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

THE EXAM CONSISTS OF 5 PROBLEMS ON 3 PAGES, 18 PAGES OF ENCLOSURES.

EXAM RESPONSIBLE: Jan Terje Kvaløy PHONE: 51 83 22 55

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

Problem 1: Consider the "Gaussian mixture" distribution for random variable X, with probability density function given as

$$f_X(x) = \frac{1}{2}\mathcal{N}(x; 1, 1) + \frac{1}{2}\mathcal{N}(x; -1, 1),$$

where $\mathcal{N}(x; \mu, \sigma^2)$ is the $N(\mu, \sigma^2)$ probability density function evaluated at x, i.e.

$$\mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

a) Find $E(x)$ and $Var(X)$.

Now, consider the bivariate random variable (X, Y) with joint probability density:

$$f_{X,Y}(x, y) = \exp(-y - xy)y,\ x > 0,\ y > 0.$$

b) Find the marginal probability density of Y and the conditional probability density of $X|Y$.
Are X and Y independent?

Now consider a random variable X with a Weibull distribution with cumulative distribution function

$$F_X(x) = 1 - \exp\left(-\frac{x^2}{2}\right),\ x > 0.$$

c) Find the corresponding probability density function.
Which parameters α and β does this Weibull distribution have?
Find $P(X > 1)$.

Problem 2: Consider a situation where we conduct N independent experiments, where the outcome of each experiment, $X_j,\ j = 1, \ldots, N$, has a Binomial distribution with $n_j \geq 1$ trials and a common success probability p (where $0 < p < 1$). I.e.

$$X_j \sim \text{Binomial}(n_j, p), j = 1, \ldots, N,$$

(and it is assumed that n_j are fixed quantities).

a) Find the likelihood function for p and show that maximum likelihood estimator $\hat{p}$ is given as

$$\hat{p} = \frac{\sum_{j=1}^{N} X_j}{\sum_{j=1}^{N} n_j}.$$

b) Which distribution does $Y = \sum_{j=1}^{N} X_j$ have?
Find $E(\hat{p})$ and $Var(\hat{p})$.
Is $\hat{p}$ a consistent estimator for p as the number of experiments N goes to infinity?

Problem 3: Consider the linear regression situation where we have independent observations given as

$$Y_i \sim N(\beta x_i, \sigma^2),\ i = 1, \ldots, n,$$

where β is a parameter, $x_i,\ i = 1, \ldots, n$ are known covariates, and σ^2 is assumed known. Notice that the constant/intercept term in the regression line is set to 0. The log-likelihood function for β is given by

$$l(\beta) = \text{constant} - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (Y_i - \beta x_i)^2.$$

a) Show that the maximum likelihood estimator $\hat{\beta}$ is given by

$$\hat{\beta} = \frac{\sum_{i=1}^{n} Y_i x_i}{\sum_{i=1}^{n} x_i^2}.$$

Which distribution does $\hat{\beta}$ have? Use this information to construct a 95% confidence interval for β.

b) Now notice that $l(\beta)$ can be written as

$$l(\beta) = \text{constant} - \frac{\sum_i x_i^2}{2\sigma^2} (\beta - \hat{\beta})^2,$$

(you do not need to show this). Consider a $N(\beta_0, \sigma_0^2)$-prior for β. Find the posterior distribution of β, i.e. $p(\beta | Y_1, \ldots, Y_n)$. Hint: You may find the equality

$$-\frac{(x-a)^2}{2b} - \frac{(x-c)^2}{2d} = -\frac{1}{2\left(\frac{bd}{b+d}\right)} \left(x - \frac{ad+bc}{b+d}\right)^2 + C(a, b, c, d)$$

(where $C(a, b, c, d)$ does not depend on x) useful.

Problem 4: Consider a discrete time Markov chain $\{X_n\}$ with state space $\mathcal{S} = \{0, 1, 2\}$ and transition probability matrix P given as

$$P = \begin{pmatrix} 0 & \frac{9}{10} & \frac{1}{10} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

a) Draw a transition graph of the Markov chain (remember to label all the arrows).
Find the two step transition probability matrix P^2.
Find $P(X_3 = 0 | X_1 = 0)$ and $P(X_4 = 0 | X_2 = 0, X_1 = 1)$.

b) Why does this Markov chain admit steady state probabilities?
Find the steady state probabilities.
Calculate $E(X_n)$.

Problem 5: Consider a Poisson process X_t with intensity λ (and $X_0 = 0$).

a) Which distribution does the time until the nth event in the Poisson process have?
Suppose $\lambda = 5$, find $P(X_1 > 5)$.
Suppose $\lambda = 5$, find the expected time until the 5th event occur.

Solutions

1,a)
First question: $E(X) = \frac{1}{2} \times 1 + \frac{1}{2} \times -1 = 0$.
Second question: $Var(X) = E(X^2) = \frac{1}{2}(1 + 1^2) + \frac{1}{2}(1 + (-1)^2) = 2$.

1,b)
The density admit the factorisation $f_{X,Y}(x, y) = exp(-y) \times exp(-xy)y$ where the former is the marginal of Y (i.e. exponential with mean 1) and the latter is the $X|Y$ density (exponential with mean $1/y$).
Alternatively, from first principles:

$$f_Y(y) = \int_0^\infty f_{X,Y}(x, y)dx = \exp(-y) \underbrace{\int \exp(-xy)ydx}_{=1} = \exp(-y),$$

i.e. exponential with mean 1. The conditional obtains as

$$f_{X|Y}(x|y) = \frac{f_{(X,Y)}(x, y)}{f_Y(y)} = \frac{\exp(-y - xy)y}{\exp(-y)} = \exp(-xy)y,$$

i.e. exponential with mean $1/y$. As the distribution of $X|Y$ depends on y, X and Y are dependent.

1,c)
Density;

$$f_X(x) = \frac{d}{dx} F_X(x) = x \exp\left(-\frac{x^2}{2}\right),$$

thus, this corresponds to a Weibull distr. with parameters $\alpha = 1/2$ and $\beta = 2$. The sought probability obtains as

$$P(X > 1) = 1 - F_X(1) = \exp(-1/2) \approx 0.6065306597.$$

2,a)
Likelihood function:

$$L(p; X_1, \dots, X_N) = \prod_{j=1}^N f_{X_j}(X_j) \propto \prod_{j=1}^N p^{X_j}(1-p)^{n_j - X_j} = p^{\sum_j X_j}(1-p)^{(\sum_j n_j)-(\sum_j X_j)}.$$

Log-likelihood function

$$l(p) = \text{constan} + (\sum_j X_j)\log(p) + ((\sum_j n_j) - (\sum_j X_j))\log(1-p)$$

Derivative wrt p equal to zero

$$\begin{aligned}
\frac{d}{dp}l(p) &= \frac{\sum_j X_j}{p} - \frac{(\sum_j n_j) - (\sum_j X_j)}{1-p} = 0 && (1)\\
&\Downarrow && (2)\\
0 &= (1-p)\sum_j X_j - p((\sum_j n_j) - (\sum_j X_j)) && (3)\\
&\Downarrow && (4)\\
\hat{p} &= \frac{\sum_j X_j}{\sum_j n_j}. && (5)
\end{aligned}$$

Check second derivative:

$$\frac{d^2}{dp^2}l(p) = -\frac{\sum_j X_j}{p^2} - \frac{(\sum_j n_j) - (\sum_j X_j)}{(1-p)^2},$$

which must be negative since $\sum_j X_j \geq 0$ and $(\sum_j n_j) - (\sum_j X_j) \geq 0$.

2,b)
Y is the sum of N independent Binomially distributed variables with common p and is therefore also Binomially distributed with parameters $\sum_j n_j$ and p.
Moreover, $E(Y) = p\sum_j n_j$ and $Var(Y) = p(1-p)\sum_j n_j$, and therefore

$$E(\hat{p}) = \frac{1}{\sum_j n_j}E(Y) = p,\ Var(\hat{p}) = \frac{1}{(\sum_j n_j)^2}p(1-p)\sum_j n_j = \frac{p(1-p)}{\sum_j n_j}.$$

$\hat{p}$ is unbiased and the variance vanishes as N goes to infinity (as $n_j \geq 1$). Thus the estimator is consistent.
3,a)

Derivative of log-likelihood equal to 0

$$\begin{aligned}
\frac{d}{d\beta}l(\beta) &= \frac{1}{\sigma^2}\sum_i x_i(y_i - \beta x_i), && (6)\\
&\Downarrow && (7)\\
\sum_i x_i y_i &= \beta\sum_i x_i^2 && (8)\\
&\Downarrow && (9)\\
\hat{\beta} &= \frac{\sum_i x_i y_i}{\sum x_i^2}. && (10)
\end{aligned}$$

Second derivative:

$$\frac{d^2}{d\beta^2} = -\frac{1}{\sigma^2}\sum_i x_i^2.$$

which must be negative.

Notice that $\hat{\beta}$ is a linear combination of Gaussian random variable and therefore also Gaussian. Thus it remains to find the mean and variance of $\hat{\beta}$:

$$E(\hat{\beta}) = \frac{1}{\sum_i x_i^2}\sum_i x_i E(Y_i) = \beta.$$

$$Var(\hat{\beta}) = \frac{1}{(\sum_i x_i^2)^2} \sum_i x_i^2 Var(Y_i) = \frac{\sigma^2}{\sum_i x_i^2}.$$

I.e. $\hat{\beta} \sim N(\beta, \sigma^2 / \sum_i x_i^2)$.

Based on this information, it is clear that

$$g(\beta, \hat{\beta}) = \frac{\hat{\beta} - \beta}{\sigma / \sqrt{\sum_i x_i^2}} \sim N(0, 1).$$

Thus,

$$0.95 = P\left(-1.96 < \frac{\hat{\beta} - \beta}{\sigma / \sqrt{\sum_i x_i^2}} < 1.96\right) \quad (11)$$

$$0.95 = P\left(\hat{\beta} - 1.96\sigma / \sqrt{\sum_i x_i^2} < \beta < \hat{\beta} + 1.96\sigma / \sqrt{\sum_i x_i^2}\right) \quad (12)$$

3,b)

The posterior obtains as $p(\beta|\text{data}) \propto L(\beta)p(\beta)$. Thus, in this case

$$p(\beta|\text{data}) \propto \exp\left(-\frac{\sum_i x_i^2}{2\sigma^2}(\beta - \hat{\beta})^2 - \frac{(\beta - \beta_0)^2}{2\sigma_0^2}\right).$$

Now, using the hint (with $x = \beta, a = \hat{\beta}, b = \sigma^2 / \sum_i x_i^2, c = \beta_0, d = \sigma_0^2$) one obtains that

$$p(\beta|\text{data}) \propto \exp\left(-\frac{(\beta - B)^2}{2V}\right),$$

where

$$B = \frac{\hat{\beta}\sigma_0^2 + \beta_0 \sigma^2 / \sum_i x_i^2}{\sigma_0^2 + \sigma^2 / \sum_i x_i^2},$$

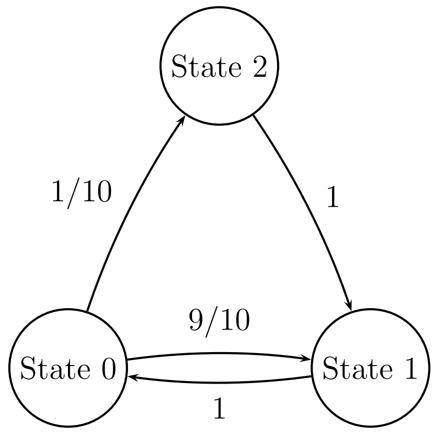
and

$$V = \frac{\sigma_0^2 \sigma^2 / \sum_i x_i^2}{\sigma_0^2 + \sigma^2 / \sum_i x_i^2}.$$

Thus, recognizing the latter representation of the posterior as a Gaussian kernel, the posterior distribution of β is $N(B, V)$.

4,a)

The transition graph looks something like



The two step transition matrix obtains as

$$P^2 = P \cdot P = \begin{pmatrix} \frac{9}{10} & \frac{1}{10} & 0 \\ 0 & \frac{9}{10} & \frac{1}{10} \\ 1 & 0 & 0 \end{pmatrix}$$

$P(X_3 = 0|X_1 = 0) = P(X_4 = 0|X_2 = 0, X_1 = 1) = 9/10.$

4,b)

The chain irreducible (all states communicate), has a finite state space, and is aperiodic. The latter can be seen e.g. as the chain may return to state 0 in $2, 3, 4, 5, \ldots$ transitions, and therefore the period must be one.
The steady state probabilities π obtain as (choosing two of the equations of)

$$\pi = P^T \pi = \begin{pmatrix} 0 & 1 & 0 \\ 0.9 & 0 & 1 \\ 0.1 & 0 & 0 \end{pmatrix} \pi \tag{13}$$

and $\sum_i \pi_i = 1$. Choosing the former and latter of the equations of (13), we obtain

$$\pi_0 = \pi_1,\ \pi_2 = 0.1\pi_0, \pi_0 + \pi_1 + \pi_2 = 1$$

Solving for π_0 first (based on last equation):

$$\pi_0 + \pi_0 + 0.1\pi_0 = 1\ \Rightarrow \pi_0 = 1/(2 + 0.1) = 10/21 \approx 0.4761904762$$

And thus: $\pi_1 = \pi_0 = 10/21$, $\pi_2 = \pi_0/10 = 1/21 \approx 0.04761904762$.
The sought expectation is $E(X_n) = 0 \times 10/21 + 1 \times 10/21 + 2 \times 1/21 = 4/7 \approx 0.5714285714$.
5,a)

First part; time until nth event has a Gamma distribution with shape parameter $\alpha = n$ and scale parameter $\beta = 1\lambda$.
Second part; X_1 has a Poisson distribution with mean 5; from the table we have that $P(X_1 > 5) = 1 - P(X_1 \leq 5) = 1 - 0.6160 = 0.384$ Third part; based on first part, expected time until 5th event is the expectation in the gamma distribution $\alpha\beta = 5/\lambda = 1$.