ST3009 – Statistics
2016 Exam Solutions

Question 1

1. A bag contains 4 red balls and 4 white balls. One ball is drawn from the bag and put to one side. A second ball is now drawn from the bag. What is the probability that the first ball is red and the second ball is white? Are the events of drawing are a red ball and then a white ball independent? [15 marks]

Sample Space = { R, R, R, R, W, W, W, W }

Balls removed without replacement.

$$P(First\ Red) = \frac{4}{8} = 0.5$$

$$P(Second\ White) = \frac{4}{7} = 0.5714$$

$$P(First\ Red\ and\ Second\ White) = 0.5 * 0.5714 = 0.2857$$

For two events to be independent:

$$P(X \cap Y) = P(X) * P(Y)$$

$$P(Red\ First) = \frac{1}{2}$$

$$P(White\ Second) = P(White\ Second|Red\ First)P(Red\ First) + P(White\ Second|WhiteFirst)P(White\ First)$$

$$P(White\ Second) = \left(\frac{4}{7}\right)\left(\frac{4}{8}\right) + \left(\frac{3}{7}\right)\left(\frac{4}{8}\right) = 0.5$$

$$P(Red\ First)P(White\ Second) = 0.5 * 0.5 = 0.25$$

But:

$$P(Red\ First)P(White\ Second) \neq P(First\ Red\ and\ Second\ White)$$

Therefore, the events are not independent.

Question 2

2. We transmit a bit of information which is "0" with probability 1-p and "1" with probability p. Because of noise on the channel, each transmitted bit is received correctly with probability 1-q where q<1/2.

a) Suppose we observe a "1" at the receiver. What is the probability that the transmitted bit was a "1"? [15 marks]

Probability of receiving a 0 → 1-p
Probability of receiving a 1 → p

Probability of correct result → 1-q

$$\boldsymbol{P(E|F) = \frac{P(F|E)P(E)}{P(F)}}$$

$$\boldsymbol{P(F) = P(F|E)P(E) + P(F|E')P(E')}$$

$$P(1\,transmitted|1\,observed) = \frac{P(1\,observed|1\,transmitted)P(1\,transmitted)}{P(1\,observed)}$$

$$P(1\,transmitted|1\,observed) = \frac{(1-q)p}{P(1\,observed)}$$

$$P(1\,observed) = [(1-q)p] + [(1-p)q]$$

$$P(1\,transmitted|1\,observed) = \frac{(1-q)p}{[(1-q)p] + [(1-p)q]}$$

b) Suppose we transmit each information bit n times over the channel e.g. if n=3 and the information bit is a "1" then we send "111". What is the probability that the information bit is "1" given that you have observed n "1"s at the receiver. What happens when n becomes large? [15 marks]

Let F be the event that n 1's are observed at the receiver and E be the event that a 1 is sent:

$$\boldsymbol{P(E|F) = \frac{P(F|E)P(E)}{P(F)}}$$

$$\boldsymbol{P(F) = P(F|E)P(E) + P(F|E')P(E')}$$

$$P(F|E) = (1-q)^n$$

$$P(E) = p$$

$$P(F) = (1-q)^n(p) + q^n(1-p)$$

$$\boldsymbol{P(E|F)} = \frac{(1-q)^n(p)}{(1-q)^n(p) + q^n(1-p)}$$

$$\boldsymbol{P(E|F)} = \frac{p}{p + \frac{q^n(1-p)}{(1-q)^n}}$$

As n becomes large, $\frac{q^n(1-p)}{(1-q)^n}$ goes to 0 since q < 0.5 and so, as a result P(E|F) tends to 1.

c) For the setup in part (b), what is the probability that the information bit is "1" given that you have observed m "1"s (and n-m "0"s) at the receiver, m≤n. [10 marks]

Let F be the event that m 1's are observed at the receiver and E be the event the information bit is a 1:

$$P(F|E) = \binom{n}{m}(1-q)^m(q)^{m-n}$$

$$P(F|E') = \binom{n}{m}(q)^m(1-q)^{m-n}$$

And so:

$$\boldsymbol{P(E|F)} = \frac{\binom{n}{m}(1-q)^m(q)^{m-n}(p)}{\binom{n}{m}(1-q)^m(q)^{m-n} + \binom{n}{m}(q)^m(1-q)^{m-n}}$$

Question 3

3. A server has 32GB of memory. We are interested in the probability that the server is overloaded, meaning the memory usage by all of the running jobs exceeds 32GB. Suppose the memory usage of a job is 0.5GB with probability 0.5 and 1GB with probability 0.5, and that the memory usage of different jobs are independent.

a) Suppose exactly 32 jobs are running. Using Markov's inequality, compute an upper bound on the probability that the server is overloaded. [10 marks]

Overloaded → Memory of all running jobs > 32GB
Memory of a job → 0.5GB w/ probability of 0.5
→ 1GB w/ probability of 0.5

Let X_i be the memory used by job i:

$$E[X_i] = 0.5 * 0.5 + 1 * 0.5 = 0.75$$

The total memory usage is:

$$U = \sum_{i=1}^{32} X_i$$

Therefore, the expected memory usage is:

$$E[U] = \sum_{i=1}^{32} \mathrm{E}[X_i] = 32 * 0.75 = 24$$

By Markov's inequality:

$$P(X \geq a) \leq \frac{E[X]}{a}$$

$$P(U \geq 32) \leq \frac{24}{32}$$

$$P(U \geq 32) \leq 0.75$$

Therefore, the upper bound on the probability that the server is overloaded is 0.75 or 75%.

b) Suppose now that a random number N of jobs are running, with $P(N=n)=p(1-p)^{(n-1)}$, where p is a parameter. Using Markov's inequality, compute an upper bound on the probability that the server is overloaded. What value of p should we choose to ensure that the probability of overload is less than 0.1 (based on Markov's inequality). Useful fact: $\sum_{n=0}^{\infty} np(1-p)^{(n-1)} = \frac{1}{p}$ [15 marks]

Random number N of jobs running at any one time with:

$$P(N=n) = p(1-p)^{n-1}$$

Now the memory usage, U can be defined as:

$$U = \sum_{i=1}^{N} X_i$$

The new expected value is then:

$$E[U] = \sum_{n=0}^{\infty} E\left[\sum_{i=1}^{n} X_i \,\middle|\, N=n\right] P(N=n) = \sum_{n=0}^{\infty}\sum_{i=1}^{n} E[X_i|N=n] * P(N=n)$$

$$E[U] = \sum_{n=0}^{\infty}\sum_{i=1}^{n} E[X_i] * P(N=n)$$

$$E[U] = \sum_{n=0}^{\infty} n * E[X_i] * P(N=n)$$

$$E[U] = 0.75 * \sum_{n=0}^{\infty} n * P(N=n)$$

$$E[U] = 0.75 * E[N]$$

We can also calculate E[N] as follows:

$$E[N] = \sum_{n=0}^{\infty} n * P(N=n)$$

$$E[N] = \sum_{n=0}^{\infty} n * p(1-p)^{n-1}$$

$$E[N] = \frac{1}{p}$$

Therefore:

$$E[U] = \frac{0.75}{p}$$

Using this we can then calculate an upper bound of Markov's inequality as follows:

$$P(X \geq a) \leq \frac{E[X]}{a}$$

$$P(U \geq 32) \leq \frac{E[U]}{32}$$

$$P(U \geq 32) \leq \frac{0.75}{32p}$$

In order to ensure the probability of overloading the server has an upper bound of 0.1 solve:

$$\frac{0.75}{32p} = 0.1$$

$$0.75 = 3.2p$$

$$p = 0.2343$$

Question 4

4. Consider a linear regression model in which random variable Y is the sum of a deterministic linear function of input x plus random noise M. That is, Y=θx+M, where θ is a parameter we would like to estimate. Suppose noise M normally distributed with mean 0 and variance 1.

a) Write down an expression for the probability distribution function of Y given θ and x.

[5 marks]

Noise is a number M normally distributed with mean 0 and variance of 1:

$$M \sim N(0,1)$$

So we can then express the probability distribution function f of y given θ and x as:

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

b) You are given n independent and identically distributed training examples d={(x_1,y_1), ..., (x_n,y_n)}. Write an expression for the likelihood of this training data. [5 marks]

$$f_{D|\theta}(d|\theta) = \prod_{i=1}^{n} f_{Y|\theta,x}(y_i,|\theta,x_i)$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \prod_{i=1}^{n} \exp\left(-\frac{(y_i-\theta x_i)^2}{2}\right)$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{\sum_{i=1}^{n}(y_i-\theta x_i)^2}{2}\right)$$

c) Now write an expression for the log-posterior probability density function for θ assuming a Gaussian prior over θ with mean 0 and standard deviation σ. How is this used to obtain a MAP estimate for θ? [10 marks]

By Bayes rule the posterior is:

$$f_{\theta|D}(\theta|d) = \frac{f_{D|\theta}(d|\theta) f_\theta(\theta)}{f_D(d)}$$

We know $f_{D|\theta}(d|\theta)$ from part (b):

$$f_{D|\theta}(d|\theta) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{\sum_{i=1}^n (y_i - \theta x_i)^2}{2}\right)$$

The prior ($f_\theta(\theta)$) is as follows (Gaussian with mean 0 and standard deviation σ:

$$f_\theta(\theta) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\theta^2}{2\sigma^2}\right)$$

Putting these together and taking logs we get:

$$\log f_{\theta|D}(\theta|d) \propto -\frac{\sum_{i=1}^n (y_i - \theta x_i)^2}{2} - \frac{\theta^2}{2\sigma^2}$$

The MAP estimate for θ is the value that maximises this.