Exam

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 red ball and then a white ball independent ?

Solution

There are 8 balls in the bag initially, 4 of which are red. All balls are equally likely to be drawn from the bag so the probability of a red ball is 4/8. After drawing the red ball there are now 7 balls in the bag, 3 red and 4 white. The probability of now drawing a white ball is therefore 4/7 and P(red then white) = 4/8 x 4/7 = 2/7.

P(red first) = 4/8=1/2

P(white first) = 4/8=12

P(white second) = P(white second | red first)P(red first) + P(white second|white first) P(white first) = 4/7x4/8 + 3/7x4/8 = 1/2

P(red first)P(white second) = 1/2 x 1/2 = 1/4 = 0.25

But P(red then white)=2/7=0.285 which is not equal to P(red first)P(white second)=0.21, so the events are not independent.

2. We transmit a bit of information which is 0 with probability 1-p and 1 with 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” ?
b) Suppose we transmit the same information bit n times over the channel. 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 ?
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.

Solution

a) Let E be the event that a 1 is transmitted and F the event that a 1 is observed. Using Bayes Rule:

   $P(E|F) = P(F|E)P(E)/P(F) = P(F|E)P(E)/ (P(F|E)P(E)+ P(F|E^c)(1-P(E)))$

   Now $P(F|E)=1-q$, $P(E)=p$, $P(F|E^c)=q$. So

   $P(E|F) = (1-q)p/((1-q)p + q(1-p))$

b) Let F be the event that n 1’s are observed at the receiver. By Bayes Rule:

$P(E|F) = P(F|E)P(E)/P(F)$

Now $P(F|E) = (1-q)^n$, $P(E)=p$ and

$P(F) = P(F|E)P(E) + P(F|E^c)(1-P(E))$
$= (1-q)^n p + q^n(1-p)$
So,

$P(E|F) = (1-q)^n p/((1-q)^n p + q^n(1-p)) = p/(p + (1-p)\, q^n/(1-q)^n)$

As n becomes large, $q^n/(1-q)^n$ goes to 0 since $q<1/2$, and so $P(E|F_1,F_2,\ldots F_n)$ goes to 1.

c) Let F be the event that m 1's are observed at the receiver. Now

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

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

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 is independent.

a) Suppose exactly 32 jobs are running. Using Markov's inequality, compute an upper bound on the probability that the server is overloaded.
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: $\sum_{n=0}^{\infty} np(1-p)^{(n-1)} = \frac{1}{p}$.

Solution

a) Let X_i be the memory used by job i. $E[X_i] = 0.5\text{x}1/2 + 1\text{x}1/2 = 0.75$. The total memory usage is $S=\sum_{i=1}^{32} X_i$ and so $E[S] = \sum_{i=1}^{32} E[X_i] = 32\times 0.75 = 24$. By Markov's inequality,

$P(S\geq 32) \leq E[S]/32 = 24/32 = 0.75$

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

$$E[S] = \sum_{n=0}^{\infty}\sum_{i=1}^{n} E[X_i]P(N = n) = 0.75\sum_{n=0}^{\infty} nP(N = n) = 0.75E[N]$$

We have that

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

and so E[S]=0.75/p. By Markov's inequality,

$P(S \geq 32) \leq E[S]/32 = 0.75/32p$

For 0.75/32p = 0.1 we need p=7.5/32=0.23

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~N(0,1). That is, Y=θx+M, where θ is a parameter we would like to estimate.

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

<u>Solution</u>

a) Noise M~N(0,1) so $f_{Y|\theta,x}(y,|\theta,x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y-\theta x)^2}{2}\right)$
b) $f_{D|\theta}(d|\,\theta) =$
$\prod_{i=1}^{n} f_{Y|\theta,x}(y_i,|\theta,x_i) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \prod_{i=1}^{n} \exp\left(-\frac{(y-\theta x)^2}{2}\right) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{\sum_{i=1}^{n}(y-\theta x)^2}{2}\right)$

c) By Bayes Rule the posterior is $f_{\theta|D}(\theta|d) = f_{D|\theta}(d|\,\theta)f_{\theta}(\theta)/f_D(d)$. $f_{\theta|D}(\theta|d)$ is given in (b) and the prior $f_{\theta}(\theta)= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\theta^2}{2\sigma^2}\right)$ (Gaussian with mean 0 and standard deviation σ). Taking logs,

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

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