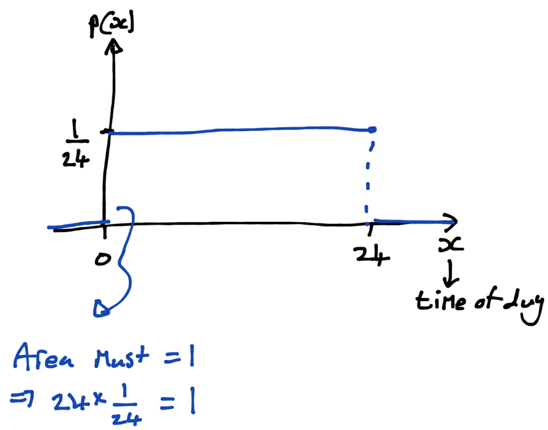
COM2004/3004 – Exercise Sheet 1 – Solutions

Question 1

X is a continuous random variable representing the time of day at which an event occurs, measured in hours 0 to 24. The event is equally likely to occur at any time.

(i) Sketch the probability density function, $p(x)$



(ii) Write out the equation for $p(x)$.

$$p(x) = \begin{cases} \frac{1}{24} & 0 \leq x \leq 24 \\ 0 & \text{otherwise} \end{cases}$$

(iii) What is the probability that the event happening between 8:00 and 12:00?
This is the probability

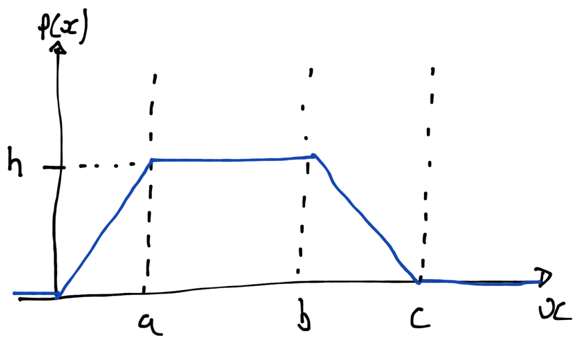
$$\begin{aligned} P(X \geq 8:00 \cap X \leq 12:00) &= \int_8^{12} p(x)dx \\ &= \frac{1}{24}(12-8) \\ &= \frac{4}{24}. \end{aligned}$$

. . . .

Question 2

A pdf, $p(x)$, has a value that linearly increases from $x = 0$ to $x = a$, remains constant from $x = a$ to $x = b$ and then linearly decreases from $x = b$ to $x = c$, where $0 < a < b < c$. $p(x) = 0$ for $x < 0$ and for $x > c$. The pdf has no discontinuities.

(i) Sketch $p(x)$



(ii) What is the value of $p(x)$ for $x = (a + b)/2$
This is just asking us to find the value of $p(x)$ for x at the midpoint between a and b. To find this we need to remember that the area under the curve is 1. From the plot we can see that the area is composed of two triangles and a rectangle, all of height h. So

$$\frac{1}{2}ha + (b - a)h + \frac{1}{2}h(c - b) = 1,$$

rearranging we find that

$$h = \frac{2}{b - a + c}.$$

(iii) What is the probability that x has a value less than a?
This is just the area of the fist triangle,

$$\frac{1}{2}ah = \frac{a}{b - a + c}$$

(iv) What is the probability that x has a value greater than b?
This is just the area of the 2nd triangle,

$$\frac{1}{2}(c - b)h = \frac{c - b}{b - a + c}$$

....

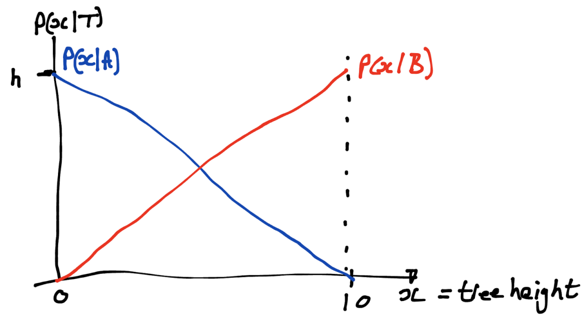
Question 3

This is the problem that we started in the lecture but with some extra questions

On an alien planet there are two types of tree, type A and type B. The heights of type A have a pdf that linearly decreases from a maximum at a height of 0 m down to a value of 0 at a height of 10 m. The heights of type B have a pdf that is 0 at 0 m and linearly increases to a maximum value at 10 m. There are no trees taller than 10 m. Assume also that there are an equal number of A and Bs on the planet.

(i) Sketch the pdfs of the heights for type A and type B, ie., $p(x|A)$ and $p(x|B)$

The area under the pdfs should both (individually) equal 1. Since they are triangles the area is $\frac{1}{2}h10$



so $h = \frac{1}{5}$ to make sure the area is 1. It is the same height for type B trees. We know that both functions are linear so we can write the functions as:

$$p(x|T = A) = \begin{cases} \frac{1}{5}\frac{10-x}{10} & 0 \leq x \leq 10 \\ 0 & otherwise \end{cases}$$

$$p(x|T = B) = \begin{cases} \frac{1}{5}\frac{x}{10} & 0 \leq x \leq 10 \\ 0 & otherwise \end{cases}$$

(ii) What is the probability of a type A tree being between 3 and 4 meters tall?

This is area under the pdf between 3 and 4 meters. We can geometrically calculate this by using a trapezoidal rule (note the width of this region is 1, and this is part of the conversion from pdf to probability) so that

$$\begin{aligned} P(3 \leq X \leq 4|A) &= \frac{1}{2}\left(p(x = 3|T = A) + p(x = 4|T = A)\right) \\ &= \frac{1}{2}\left(\frac{1}{5}\frac{10-3}{10} + \frac{1}{5}\frac{4}{10}\right) \\ &= \frac{1}{10}\left(\frac{7}{10} + \frac{6}{10}\right) \\ &= \frac{13}{100} = 0.13 \end{aligned}$$

(iii) A tree is found and observed to be 6 meters tall. What is the probability that it is of type A?

Let T be the type of tree. Using Bayes theorem,

$$p(T = A|x = 6) = \frac{p(x = 6|T = A)p(T = A)}{p(x = 6)} = \frac{p(x = 6|T = A)}{p(x = 6)} = \frac{0.4 \times 0.2}{p(x = 6)}$$

and

$$P(T=B|x=6) = \frac{p(x=6|T=B)p(T=B)}{p(x=6)} = \frac{p(x=6|T=B)}{p(x=6)} = \frac{0.6 \times 0.2}{p(x=6)}$$

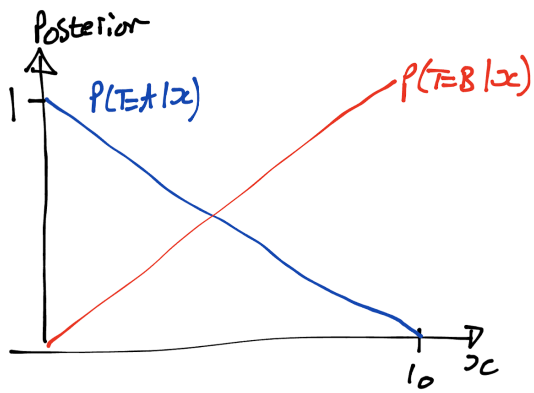
To find $p(x=6)$ we can use the fact that $P(T=A|x=6) + P(T=B|x=6) = 1$ so

$$p(x=6) = 0.4x0.2 + 0.6x0.2 = 0.2$$

Substituting this into the above we have

$$p(T=A|x=6) = 0.4 \times 0.2/p(x=6)0.4 \times 0.2/0.2 = 0.4$$

(iv) Sketch the posterior probability $p(A|x)$ as a function of x.



$$p(T=A|x) = \begin{cases} \frac{10-x}{10} & 0 \leq x \leq 10 \\ 0 & otherwise \end{cases}$$

(v) Using Bayes decision rule, what would be the threshold height that a classifier would use to separate type A and type B?
This is the value of x where our decision between type A and type B will change. Mathematically this is given by $P(T=A|x) = P(T=B|x)$. Inserting our equations for the posteriors gives:

$$\begin{aligned} P(T=A|x) &= P(T=B|x) \\ \frac{10-x}{10} &= \frac{x}{10} \\ 10-x &= x \\ x &= 5 \end{aligned}$$

So the threshold height will be 5m.

(vi) What would be the probability of classification error for this classifier?
This is the area under the likelihood distributions where the wrong class could be predicted. e.g for type A, this is the area above the threshold of 5m but below the likelihood function as this represents the probability that it truly is a type A tree but through our classifier we would assign it type B. The same is also true for type B but below the threshold. In this example we can simply consider the two triangles that form these regions.
For a misclassification of type A, this is the region $5 \leq x \leq 10$ and the probability density at $x = 5$ is $1/10$. Therefore the area is

$$\frac{1}{2}(10-5)\frac{1}{10} = \frac{5}{20} = \frac{1}{4}.$$

The same can be done for type B trees but both are weighted by $P(T) = 1/2$ so the total classification error will be 0.25.

(vii) If it is now discovered that there are in fact twice as many type B trees as type A trees, i.e. the prior probabilities are not equal. What would be the new best decision threshold?
Find x where $p(T = A|x) = P(T = B|x)$

$$\begin{aligned} p(x|T = A)p(T = A) &= p(x|T = B)p(T = B) \\ p(x|T = A)1/3 &= p(x|T = B)2/3 \\ 1/50(10 - x)1/3 &= 1/50x2/3 \\ x &= 10/3 \end{aligned}$$

(viii) What would be the probability of classification error in this unequal prior scenario?

$$\begin{aligned} P(T = A)P(x > 10/3|T = A) + P(T = B)P(x < 10/3|T = B) = \\ 1/3 \times (10 - 10/3) \times 1/2 \times 2/3 \times 1/5 + 2/3 \times 10/3 \times 1/2 \times 1/3 \times 1/5 = \\ 2/9 \end{aligned}$$

(ix) What is the average height of the type A trees? Of the type B trees?
For type B trees,

$$\begin{aligned} E(x) &= \int_0^{10} xp(x)dx \\ &= \int_0^{10} x^2/50dx \\ &= [x^3/150]_0^{10} \\ 10 * 10 * 10/150 &= 20/3 \end{aligned}$$

So the average height for type B trees is 20/3 m. By symmetry type A trees will have an average height of 10/3 m.

(x) (Harder) What is the probability that a type A tree will be taller than a type B tree?
Let A be the height of a randomly selected Type A tree and B be the height of a randomly selected Type B tree.

This will require integrating $P(A > B) = \int P(A > B, B = x)dx = \int P(A > x)p(B = x)dx$

The question will have the same answer regardless of the maximum height that was chosen for the trees. So let's make the algebra a bit simpler by considering a max height of 1 m.

The equations for the pdfs between 0 and 1 are now

$$\begin{aligned} P(A = x) =& 2 - 2x \\ p(B = x) =& 2x \end{aligned}$$

$P(A > x^*)$ is the area under $P(A = x)$ from some point x^*. This is a triangle with base $(1 - x^*)$ and height $P(A = x^*) = 2 - 2x^*$ so the area is $(1 - x^*)(1 - x^*)$

$$
\begin{aligned}
P(A > B) &= \int_0^1 P(A > x)p(B = x)dx \\
&= \int_0^1 (1 - x)(1 - x)2xdx \\
&= \int_0^1 2x - 4x^2 - 2x^3 dx \\
&= \int_0^1 2x - 4x^2 - 2x^3 dx \\
&= \left[x^2 - \frac{4}{3}x^3 - \frac{1}{2}x^4\right]_0^1 \\
&=1 - \frac{4}{3} - \frac{1}{2} \\
&=\frac{1}{6}
\end{aligned}
$$