Martin Lundfall, Malte Siemers, Henri Bunting

1.1 Distributions and expected values

a)

In order for $p(x)$ to be a probability density function, its integral over the whole domain should be one. Thus we derive the value $c = \frac{1}{2}$.

$$1 = \int_0^\pi c \cdot sin(x) = [-c \cdot cos(x)]_0^\pi \implies c = \frac{1}{2}$$

b)

To compute the expected value we simply take the integral of $x \cdot p(x)$ over the whole domain:

$$< X >_p = \int_0^\pi \frac{xsin(x)}{2} dx = \left[\frac{-xcos(x) + sin(x)}{2} \right]_0^\pi = \frac{\pi}{2} = 0$$

c)

The variation is given by $Var(x) = \int x^2 p(x)dx - < X >^2$. We get:

$$Var(x) = \int_0^\pi \left(\frac{x^2 sin(x)}{2} \right) dx - \frac{\pi^2}{4} = \left[\frac{-x^2 cos(x)}{2} + cos(x) + xsin(x) \right]_0^\pi - \frac{\pi^2}{4}$$
$$= \frac{\pi^2}{2} - 2 - \frac{\pi^2}{4} = \frac{\pi^2 - 8}{4}$$

1.2 Marginal densitites

a)

The marginal densities are given by integrating over the other variable:

$$p_x(x) = \int_0^1 \frac{3}{7}(2-x)(x+y)dy = \frac{3}{14}(-2x^2 + 3x + 2)$$
$$p_y(y) = \int_0^2 \frac{3}{7}(2-x)(x+y)dy = \frac{2}{7}(3y + 2)$$

b)

For independence, we check whether $p(x)p(y) = p(x, y)$.

$$p_x(x)p_y(y) = \frac{3}{49}(-2x^2 + 3x + 2)(3y + 2) \neq p_{x,y}(x, y) = \frac{3}{7}(2-x)(x+y)$$

The variables are dependent.

1.3 Taylor expansion

The taylor expansion up to third degree of $\sqrt{1+x}$ around zero is:

$$1 + \frac{x}{2} - \frac{x^2}{8} + \frac{x^3}{16}$$

1.4 Determinant of a matrix

The matrix:

$$A = \begin{pmatrix} 5 & 8 & 16 \\ 4 & 1 & 8 \\ -4 & -4 & -11 \end{pmatrix}$$

has determinant:

$$det(A) = 5 \cdot 1 \cdot (-11) + 8 \cdot 8 \cdot (-4) + 16 \cdot 4 \cdot (-4)$$
$$-(16 \cdot 1 \cdot (-4)) - (8 \cdot 4 \cdot (-11)) - (5 \cdot 8 \cdot (-4)) = 9$$

and trace:

$$Tr(A) = 5 + 1 - 11 = -5$$

1.5 Critical points

For f(x,y) one can see its a rotated parabola shifted in z by c with a critical point, its minimum, at (0,0). For g(x,y) we get the partial derivatives:

$$\frac{\partial g}{\partial x} = 2x$$
$$\frac{\partial g}{\partial y} = -2y$$

Both vanish at (0,0), so we have a critical point. The mixed second partial derivatives are also equal to zero at (0,0) and

$$\frac{\partial^2 g}{\partial x^2} = 2$$
$$\frac{\partial^2 g}{\partial y^2} = -2$$

are positive and negative, so we have a saddle point at (0,0).

1.6 Bayes rule

Bayes rule gives us:

$$
\begin{aligned}
P(D|+) &= \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|\bar{D})P(\bar{D})} \\
P(D|+) &= \frac{0.95 \cdot 0.01}{0.95 \cdot 0.01 + 0.001 \cdot 0.99} \approx 0.90562 \\
P(\bar{D}|+) &= \frac{P(+|\bar{D})P(\bar{D})}{P(+|D)P(D) + P(+|\bar{D})P(\bar{D})} \\
P(\bar{D}|+) &= \frac{0.001 \cdot 0.99}{0.95 \cdot 0.01 + 0.001 \cdot 0.99} \approx 0.09437 \\
P(\bar{D}|-) &= \frac{P(-|\bar{D})P(\bar{D})}{P(-|D)P(D) + P(-|\bar{D})P(\bar{D})} \\
P(\bar{D}|-) &= \frac{0.999 \cdot 0.99}{0.001 \cdot 0.01 + 0.999 \cdot 0.99} \approx 0.99999 \\
P(D|-) &= \frac{P(-|D)P(D)}{P(-|D)P(D) + P(-|\bar{D})P(\bar{D})} \\
P(D|-) &= \frac{0.001 \cdot 0.01}{0.001 \cdot 0.01 + 0.999 \cdot 0.99} \approx 0.00001
\end{aligned}
$$

We can also see that $1 - P(D|+) = P(\bar{D}|+)$ and $1 - P(D|-) = P(\bar{D}|-)$ respectively, which is valid.

1.7 Learning paradigms

a)

Unsupervised learning requires no additional information whereas supervised learning has training data and reinforcement learning has rating data.

Supervised learning aims to predict new answers from previous data.

Unsupervised learning aims to organize data.

Reinforcement learning aims to find the correct action or strategy.

b)

	type	observations	labels	rewards
Teach a dog to catch a ball	Reinforcement	ball returned nothing returned something else returned		kiss / ”good” none / ”bad” none / ”bad”
read hand-written address	Supervised	single letter image	digitized value	
identify groups of users with the same music taste	Unsupervised	user and music taste		