1 Math Primer

1.1 Distributions and expected values

a) Requirement for $p(x)$ to be probability density function (PDF):

$$\int_{-\infty}^{\infty} p(x)\,dx = 1$$

Calculate antiderivative $P(x)$:

$$P(x) = \begin{cases} -c \cdot cos(x) + const. & x \in [0, \pi] \\ const. & else \end{cases}$$

Choose $const.$ to be 0.

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} p(x)\,dx \\ &= \int_{0}^{\pi} p(x)\,dx \\ &= \Big[P(x)\Big]_0^\pi \\ &= -c \cdot cos(\pi) + c \cdot cos(0) \\ &= 2c \\ 0.5 &= c \end{aligned}$$

b) Expected value of PDF $p(x)$:

$$\langle X \rangle_p = \int_{-\infty}^{\infty} x \cdot p(x)\,dx$$

Calculate antiderivative $Q(x)$ of $x \cdot p(x)$:

$$Q(x) = \begin{cases} 0.5\,(sin(x) - x \cdot cos(x)) & x \in [0, \pi] \\ const. & else \end{cases}$$

Choose $const.$ to be 0.

$$\begin{aligned} \langle X \rangle_p &= \int_{-\infty}^{\infty} x \cdot p(x)\,dx \\ &= \int_{0}^{\pi} x \cdot p(x)\,dx \\ &= \Big[Q(x)\Big]_0^\pi \\ &= 0.5\,(sin(\pi) - \pi \cdot cos(\pi)) - 0.5\,(sin(0) - 0 \cdot cos(0)) \\ &= \frac{\pi}{2} \end{aligned}$$

c) Calculate antiderivative $R(x)$ of $\left(x - \langle X\rangle_p\right)^2 \cdot p(x)$:

$$\begin{aligned}\left(x - \langle X\rangle_p\right)^2 \cdot p(x) &= \left(x - \frac{\pi}{2}\right)^2 \cdot \frac{1}{2} \cdot sin(x) \\ &= \left(x^2 - \pi x + \frac{\pi^2}{4}\right) \cdot \frac{1}{2} \cdot sin(x)\end{aligned}$$

$$R(x) = \begin{cases} \frac{1}{2}\left(-x^2 + \pi x - \frac{1}{4}\pi^2 - 2\right) \cdot cos(x) + \left(x - \frac{1}{2}\pi\right) \cdot sin(x) + const. & x \in [0, \pi] \\ const. & else \end{cases}$$

Choose *const.* to be 0.

$$\begin{aligned}\left\langle X^2\right\rangle_p - \langle X\rangle_p^2 &= \int_{-\infty}^{\infty} \left(x - \langle X\rangle_p\right)^2 \cdot p(x)\, dx \\ &= \int_0^{\pi} \left(x - \frac{\pi}{2}\right)^2 \cdot \frac{1}{2} \cdot sin(x)\, dx \\ &= \Big[R(x)\Big]_0^{\pi} \\ &= \frac{1}{2}\left(-\pi^2 + \pi^2 - \frac{1}{4}\pi^2 - 2\right) \cdot cos(\pi) + \left(\pi - \frac{1}{2}\pi\right) \cdot sin(\pi) - \frac{1}{2}(-2) \cdot cos(0) - 0 \\ &= \frac{1}{8}\pi^2 + 1 + 1 \\ &= 2 + \frac{1}{8}\pi^2 \\ &\approx 3.2337\end{aligned}$$

c)

- calculation of the variance $\langle X^2\rangle_p - \langle X\rangle_p^2$:

$$
\begin{aligned}
\langle X^2\rangle_p \;-\; \langle X\rangle_p^2 \quad &= \quad \int_{-\infty}^{\infty} (x - \langle X\rangle_p)^2 \; p(x)dx &(1)\\
&= \quad \frac{1}{2}\int_0^{\pi} (x - \frac{\pi}{2})^2 \; sin(x)dx &(2)\\
&= \quad \frac{1}{2}[\int_0^{\pi} x^2 \; sin(x) \; dx \;-\; \int_0^{\pi} \pi \, x \, sin(x) \, dx \;+\; \int_0^{\pi} \frac{\pi^2}{4} \; sin(x) \, dx \;] &(3)\\
&= \quad \frac{1}{2}[\int_0^{\pi} x^2 \; sin(x) \; dx \;-\pi^2 + \int_0^{\pi} \frac{\pi^2}{4} \; sin(x) \, dx \;] &(4)\\
&= \quad \frac{1}{2}[\int_0^{\pi} x^2 \; sin(x) \; dx \;-\pi^2 - \frac{\pi^2}{4} \; cos(x) \Big|_0^{\pi} \;] &(5)\\
&= \quad \frac{1}{2}[\; -x^2 \, cos(x) \Big|_0^{\pi} \;+\; \int_0^{\pi} x \, cos(x) \, dx \;-\pi^2 + \frac{\pi^2}{2}] &(6)\\
&= \quad \frac{1}{2}[\; \pi^2 \;+\; 2\int_0^{\pi} x \, cos(x) \, dx \;-\; \frac{\pi^2}{2}] &(7)\\
&= \quad \frac{1}{2}[\, 2x \, sin(x) \Big|_0^{\pi} \;-\; 2\int_0^{\pi} sin(x) \, dx \;+\; \frac{\pi^2}{2}] &(8)\\
&= \quad \frac{1}{2}[\, -2\int_0^{\pi} sin(x) \, dx \;+\; \frac{\pi^2}{2}] &(9)\\
&= \quad \frac{1}{2}[\, -4 \;+\; \frac{\pi^2}{2}] &(10)\\
&= \quad \frac{\pi^2}{4} \;-\; 2 &(11)\\
&\approx \quad 0.4674 &(12)
\end{aligned}
$$

1.2 Marginal densities

a) to calculate the marginal densities, we need to solve the following integrals:

$$
\begin{aligned}
p_x(x) \quad &= \quad \int_{-\infty}^{\infty} p_{x,y}(x, \tilde{y}) \; d\tilde{y} &(13)\\
p_y(y) \quad &= \quad \int_{-\infty}^{\infty} p_{x,y}(\tilde{x}, y) \; d\tilde{x} &(14)
\end{aligned}
$$

$$
\begin{aligned}
p_x(x) \quad &= \quad \frac{3}{7}\int_0^1 \; 2x - x^2 + \tilde{y}(2 - x) \; d\tilde{y} &(15)\\
&= \quad \frac{3}{7}\,(2x - x^2)\tilde{y} \;+\; \tilde{y}^2 \frac{3(2-x)}{14}\Big|_0^1 &(16)\\
&= \quad \frac{3}{7}\,(2x - x^2) + \frac{6 - 3x}{14} &(17)\\
&= \quad -\frac{3}{7}x^2 + \frac{9}{14}x + \frac{6}{14} &(18)
\end{aligned}
$$

$$p_y(y) = \frac{3}{7}\int_0^2 2\tilde{x} + 2y - \tilde{x}^2 - \tilde{x}y\, d\tilde{x} \tag{19}$$
$$= \frac{3}{7}\Big(\ \tilde{x}^2 + 2\tilde{x}y - \frac{1}{3}\tilde{x}^3 - \frac{1}{2}\tilde{x}^2 y\ \Big)\Big|_0^2 \tag{20}$$
$$= \frac{3}{7}\Big(\ 2y + \frac{4}{3}\ \Big) \tag{21}$$
$$= \frac{6}{7}y\ +\ \frac{12}{21} \tag{22}$$

b) if x and y are statistical independent, the probability density function needs to be writable as $p_{x,y}(x,y)\ =\ p(x)\,p(y)$

- the roots of x need to be independent of y

$$p_{x,y}(x,y)\ =\ \frac{3}{7}(2-x)(x+y)\ =\ -\frac{3}{7}(x-2)(x+y)\ =\ -\frac{3}{7}(x-x_{0,1})(x-x_{0,2}) \tag{23}$$

- since the roots of x, $x_{0,1} = 2$ and $x_{0,2} = -y$, depend on y, the variables are not independent

1.3 Taylor expansion

General form of the taylor series:

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n$$

Calculate derivatives up to $n = 3$:

$$f'(x) = \frac{1}{2\sqrt{x+1}}$$
$$f''(x) = -\frac{1}{4\,(x+1)^{\frac{3}{2}}}$$
$$f'''(x) = \frac{3}{8\,(x+1)^{\frac{5}{2}}}$$
$$\sum_{n=0}^{3} \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n = 1 + \frac{x}{2} - \frac{x^2}{8} + \frac{x^3}{16}$$

1.4 Determinant of a matrix

$$A = \begin{pmatrix} 5 & 8 & 16 \\ 4 & 1 & 8 \\ -4 & -4 & -11 \end{pmatrix}$$
$$\begin{aligned}
det(A) &= a_{11}a_{22}a_{33} + a_{12}a_{23} + a_{31} + a_{13}a_{21}a_{32} - a_{31}a_{22}a_{13} - a_{32}a_{23}a_{11} - a_{33}a_{21}a_{12} \\
&= 5 \cdot 1 \cdot (-11) + 8 \cdot 8 \cdot (-4) + 16 \cdot 4 \cdot (-4) - (-4) \cdot 1 \cdot 16 - (-4) \cdot 8 \cdot 5 - (-11) \cdot 4 \cdot 8 \\
&= -55 - 256 - 256 + 64 + 160 + 352 \\
&= 9 \\
tr(A) &= a_{11} + a_{22} + a_{33} \\
&= 5 + 1 - 11 \\
&= -5
\end{aligned}$$

1.5 Critical points

a) Check if $a = (0, 0)$ is a critical point for f and g.

   Calculate Jabobimatrix (first derivatives):

$$\begin{aligned} Jf &= (2x, 2y) \\ Jg &= (2x, -2y) \end{aligned}$$

   A critical point requires all first derivatives to be zero:

$$\begin{aligned} 2x = 0 &\Leftrightarrow x = 0 \\ 2y = 0 &\Leftrightarrow y = 0 \end{aligned}$$

   This means $a = (0, 0)$ is a critical point for f. Now we do the same for g:

$$\begin{aligned} 2x = 0 &\Leftrightarrow x = 0 \\ -2y = 0 &\Leftrightarrow y = 0 \end{aligned}$$

   We get that a is a critical point for g, too.

b) Check if a is an extremum.

   Calculate Hessian matrix:

$$\begin{aligned} Hf &= \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \\ Hg &= \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} \end{aligned}$$

   Obviously, $det(Hf(a)) \neq 0$ and $det(Hg(a)) \neq 0$.

   Now we calculate the Eigenvalues:

$$det(Hf(a) - \lambda I) = 0 \Leftrightarrow det \begin{pmatrix} 2-\lambda & 0 \\ 0 & 2-\lambda \end{pmatrix} = 0 \Leftrightarrow (2-\lambda)^2 = 0 \Leftrightarrow \lambda = 2$$

   All Eigenvalues are positive, so $Hf(a)$ is positive definite, which means a is a minimum of f.

   For $Hg(a)$ we observe positive and negative Eigenvalues:

$$det(Hg(a) - \lambda I) = 0 \Leftrightarrow det \begin{pmatrix} 2-\lambda & 0 \\ 0 & -2-\lambda \end{pmatrix} = 0 \Leftrightarrow (2-\lambda)(-2-\lambda) = 0 \Leftrightarrow \lambda \in \{2, -2\}$$

   This means Hg(a) is neither positive nor negative definite, so a is not an extremum of g.

1.6 Bayes rule

From the description we get

$$\begin{aligned}
P(D) &= 0.01 \\
P(\bar{D}) &= 0.99 \\
P(+|D) &= 0.95 \\
P(-|D) &= 0.05 \\
P(+|\bar{D}) &= 0.001 \\
P(-|\bar{D}) &= 0.999
\end{aligned}$$

from which we can calculate the remaining probabilities using Bayes rule:

$$\begin{aligned}
&P(D|+) = \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|\bar{D})P(\bar{D})} = \frac{0.95 \cdot 0.01}{0.95 \cdot 0.01 + 0.001 \cdot 0.99} \approx 0.9056 \\
&P(\bar{D}|+) = 1 - P(D|+) \approx 0.0944 \\
&P(\bar{D}|-) = \frac{P(-|\bar{D})P(\bar{D})}{P(-|\bar{D})P(\bar{D}) + P(-|D)P(D)} = \frac{0.999 \cdot 0.99}{0.999 \cdot 0.99 + 0.05 \cdot 0.01} \approx 0.9995 \\
&P(D|-) = 1 - P(\bar{D}|-) \approx 0.0005
\end{aligned}$$

1.7 Learning paradigms

a)
- In supervised learning we provide control labels to the machine during training, like a teacher would do. We can measure the performance by calculating the Error between targets and actual results.
- In unsupervised learning the machine has to learn without outside help, so its only inputs are the observations. This yields us the opportunity to discover hidden structures in our data.
- In reinforcement learning, additionally to providing the machine with observations like always, we give feedback to its classifications of our data (reinforcement signal, ranging anywhere between "good" and "bad"). We can measure the performance by calculating the cumulative reward.

b)
- To teach a (robot?) dog to catch a ball

  Reinforcement learning seems appropriate. As observations we would provide the trajectory of the ball. As feedback to the dog's movements we would provide the distance between dog and ball (lower is better), whether the dog has lifted the ball, and once he has, the distance between trainer and dog (lower is better).
- To read hand written addresses from letters

  Here we would employ reinforcement learning with a training phase, as we cannot train each persons handwriting to the machine right from the start. As observations one image per word seems appropriate. As feedback one could provide a combination of count of mistakes for some proof-read samples, and for each letter wether it could actually be delivered or not.

  If the user set was limited, providing huge amounts of handwritten address samples (again one image per word) with their manually digitized versions as control labels might suffice.
- identify groups of users with the same taste of music

  Here we would employ unsupervised learning to discover clusters in our data. As observations we could provide (pre-calculated) metadata of each user's music-collection, like user-id, bpm per track, genre-tags etc. .