CDS: machine learning

Jaqueline Zeitler Sander Meis Wieke Bergers

August 2020

Week 1

Example 2.10 Mackay

The exercise asks to calculate the probability that we selected urn A given that the ball we draw is black (b).

$$P(A \mid b) = \frac{P(b \mid A)P(A)}{P(b)} = \frac{P(b \mid A)P(A)}{P(b \mid A)P(A) + P(b \mid B)P(B)}$$

$$P(A \mid b) = \frac{\frac{1}{3} * \frac{1}{2}}{\frac{1}{3} * \frac{1}{2} + \frac{2}{3} * \frac{1}{2}} = \frac{1}{3} \tag{1}$$

Exercise 2.1

The exercise asks to calculate the probability that we selected urn u given that we draw an even number of black balls in ten times (with replacement). We know

$$P(b \mid u) = \frac{u}{10} = f_u$$

$$P(N_B = i \mid u) = \binom{10}{i} P(b \mid u)^i (1 - P(b \mid u))^{10-i} = \binom{10}{i} {f_u}^i (1 - f_u)^{10-i}$$

$$P(N_B \in \{0, 2, 4, 6, 8, 10\} \mid u) = (1 - f_u)^{10} + 45 f_u^2 (1 - f_u)^8 + 210 f_u^4 (1 - f_u)^6$$
$$+210 f_u^6 (1 - f_u)^4 + 45 f_u^8 (1 - f_u)^2 + f_u^{10}$$

$$P(u) = \frac{1}{11}$$

As all urns are equally likely and all urns have the same amount (10) of balls,

$$P(b) = \frac{\#black}{\#balls} = \frac{55}{110} = \frac{1}{2}$$

$$P(N_B = i) = \binom{10}{i} P(b)^i (1 - P(b))^{10-i} = \binom{10}{i} \frac{1}{2^{10}}$$

$$P(N_B \in \{0,2,4,6,8,10\}) = (1+45+210+210+45+1)\frac{1}{2^{10}} = \frac{512}{1024} = \frac{1}{2}$$

Using Bayes' theorem

$$P(u \mid N_B) = \frac{P(N_B \mid u)P(u)}{P(N_B)}$$

$$P(u \mid N_B) = ((1-f_u)^{10} + 45f_u^2(1-f_u)^8 + 210f_u^4(1-f_u)^6$$
$$+210f_u^6(1-f_u)^4 + 45f_u^8(1-f_u)^2 + f_u^{10})\frac{1}{11}\frac{1}{\frac{1}{2}}$$

$$P(u \mid N_B) = \frac{2}{11}((1-f_u)^{10} + 45f_u^2(1-f_u)^8 + 210f_u^4(1-f_u)^6$$
$$+210f_u^6(1-f_u)^4 + 45f_u^8(1-f_u)^2 + f_u^{10})$$

$$P(u \mid N_B) = \frac{2}{11}((1-\frac{u}{10})^{10} + 45\Big(\frac{u}{10}\Big)^2(1-\frac{u}{10})^8 + 210\Big(\frac{u}{10}\Big)^4(1-f_u)^6$$
$$+210\Big(\frac{u}{10}\Big)^6(1-\frac{u}{10})^4 + 45\Big(\frac{u}{10}\Big)^8(1-\frac{u}{10})^2 + \Big(\frac{u}{10}\Big)^{10})$$

$$P(u \mid N_B) = \frac{2}{11*2^{10}}((1-(10-u)^{10} + 45u^2(10-u)^8 + 210u^4(10-u)^6$$
$$+210u^6(10-u)^4 + 45u^8(10-u)^2 + u^{10})$$

Exercise 2.2

a) We can now infer that μ is likely 'close' to x, since the probability of getting an x more than 2 standard deviations from μ is only 4.6%, and getting one more than 3 standard deviations away is only 0.3%.

To be precise we can use Bayes' rule.

$$p(\mu|x) = \frac{p(x|\mu)p(\mu)}{p(x)}$$

every μ equally likely, so also every x equally likely before setting conditions

$$\begin{aligned}
&= p(x|\mu) \\
&= \frac{1}{\sqrt{2\pi}\sigma}\exp\Big\{-\frac{(x-\mu)^2}{2\sigma^2}\Big\} \\
&= \frac{1}{\sqrt{2\pi}\sigma}\exp\Big\{-\frac{(\mu-x)^2}{2\sigma^2}\Big\}
\end{aligned}$$

So you get a normal in μ with x as mean.

b) Because the data points are generated independently, the probability that this dataset is drawn from the oracle is the product of probabilities that the data points are drawn from the oracle.

$$p(x_1, x_2, ..., x_N|\mu, \sigma) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)\right\}$$

$$=(\frac{1}{\sqrt{2\pi}\sigma})^N \exp\left\{\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2\right)\right\}$$

Keeping in mind that we know $\sigma = \sigma^2 = 1$ we can remove the σ's.

$$=(\frac{1}{\sqrt{2\pi}})^N \exp\left\{\left(-\frac{1}{2}\left(\sum_{i=1}^{N}x_i^2 - 2\sum_{i=1}^{N}x_i\mu + N\mu^2\right)\right)\right\}$$

$$=(\frac{1}{\sqrt{2\pi}})^N \exp\left\{\left(-\frac{1}{2}\left(\sum_{i=1}^{N}x_i^2 - 2N\bar{x}\mu + N\mu^2\right)\right)\right\}$$

$$=(\frac{1}{\sqrt{2\pi}})^N \exp\left\{\left(-\frac{1}{2}\left(N^2\bar{x}^2 - 2N\bar{x}\mu + N\mu^2 - 2\sum_{i\neq j}x_ix_j\right)\right)\right\}$$

$$=(\frac{1}{\sqrt{2\pi}})^N \exp\left\{\left(-\frac{1}{2}\left(N(N\bar{x}^2 - 2\bar{x}\mu + \mu^2) - 2\sum_{i\neq j}x_ix_j\right)\right)\right\}$$

$$=(\frac{1}{\sqrt{2\pi}})^N \exp\left\{\left(-\frac{1}{2}\left(N(N\bar{x}^2 - 2\bar{x}\mu + \mu^2)\right) - \sum_{i\neq j}x_ix_j\right)\right\}$$

$$=(\frac{1}{\sqrt{2\pi}})^N \exp\left\{\left(-\frac{1}{2}\left(N(N\bar{x}^2 - 2\bar{x}\mu + \mu^2)\right)\right) \exp\left(-\sum_{i\neq j}x_ix_j\right)\right\}$$

$$=(\frac{1}{\sqrt{2\pi}})^N \exp\left\{(-\sum_{i\neq j}x_ix_j)\right\} \exp\left\{\left(-\frac{1}{2/N}(N\bar{x}^2 - 2\bar{x}\mu + \mu^2)\right)\right\}$$

$$=(\frac{1}{\sqrt{2\pi}})^N \exp\left\{(-\sum_{i\neq j}x_ix_j)\right\} \exp\left\{\left(-\frac{1}{2/N}(\bar{x}^2 - 2\bar{x}\mu + \mu^2) - \frac{(N-1)\bar{x}^2}{2/N}\right)\right\}$$

$$=(\frac{1}{\sqrt{2\pi}})^N \exp\left\{(-\sum_{i\neq j}x_ix_j)\right\} \exp\left\{\left(-\frac{1}{2/N}(\bar{x}^2 - 2\bar{x}\mu + \mu^2)\right)\right\} \exp\left\{\left(-\frac{(N-1)\bar{x}^2}{2/N}\right)\right\}$$

$$=(\frac{1}{\sqrt{2\pi}})^N \exp\left\{(-\sum_{i\neq j}x_ix_j)\right\} \exp\left\{\left(-\frac{(N-1)\bar{x}^2}{2/N}\right)\right\} \exp\left\{\left(-\frac{(\mu-\bar{x})^2}{2/N}\right)\right\}$$

Now we see a bunch of constants multiplied by a Gaussian distribution in μ with variance $\frac{1}{N}$, and all the constants can be normalised away, so we have the desired results.

Exercise 2.3

a) The n-dimensional multivariate gaussian distribution is proportional to

$$\begin{aligned}
p(x|\mu,\Sigma) =& \frac{1}{\sqrt{(w\pi)^n det\Sigma}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right) \\
\propto& \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right) \\
=& \exp\left(\frac{1}{2}((x^Tx+\mu^T\mu)\Sigma^{-1} - 2\mu\cdot x\Sigma^{-1})\right) \\
\propto& \exp\left(\frac{1}{2}(x^Tx\Sigma^{-1} - \mu\cdot x\Sigma^{-1}\right) \\
=& \exp\left(\frac{1}{2}\langle x^Tx, \Sigma^{-1}\rangle - \langle\mu\Sigma^{-1}, x\rangle\right)
\end{aligned}$$

Now it is in the shape of an exponential family according to the exponential families pdf.

b) In the slides we see on slide 30 how the univariate Gaussian distribution has is the max entropy distribution. Generalizing this the the multivariate case as given above we see that $\phi_1(x) = x^Tx$ and $\lambda_1 = -\Sigma^{-1}$ and $\phi_2(x) = x$ and $\lambda_2 = \mu\Sigma^{-1}$

Exercise 2.4

a) show that $\int q_1(z_1)\log q_1(z_1)dz_1 = -\log\sqrt{2\pi\sigma_1^2} - \frac{1}{2}$ given q_1 is a normalized Gaussian distribution:

$$q_1(z_1) = \frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left\{-\frac{z_1^2}{2\sigma_1^2}\right\}$$

$$\int q_1(z_1)\,dz_1 = 1$$

$$\int_{-\infty}^{\infty} x^2\exp\{-ax^2\}\,dx = \frac{1}{2}\sqrt{\frac{\pi}{a^3}}$$

$$
\begin{aligned}
&\int q_1(z_1)\log q_1(z_1)dz_1 = \\
&= \int \frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left\{-\frac{z_1^2}{2\sigma_1^2}\right\}\log\left(\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left\{-\frac{z_1^2}{2\sigma_1^2}\right\}\right)dz_1 \\
&= \frac{1}{\sqrt{2\pi\sigma_1^2}}\int \exp\left\{-\frac{z_1^2}{2\sigma_1^2}\right\}\log\left(\frac{1}{\sqrt{2\pi\sigma_1^2}}\right)dz_1 + \frac{1}{\sqrt{2\pi\sigma_1^2}}\int \exp\left\{-\frac{z_1^2}{2\sigma_1^2}\right\}\log\left(\exp\left\{-\frac{z_1^2}{2\sigma_1^2}\right\}\right)dz_1 \\
&= \log\left(\frac{1}{\sqrt{2\pi\sigma_1^2}}\right)\frac{1}{\sqrt{2\pi\sigma_1^2}}\int \exp\left\{-\frac{z_1^2}{2\sigma_1^2}\right\}dz_1 - \frac{1}{2\sigma_1^2}\frac{1}{\sqrt{2\pi\sigma_1^2}}\int z_1^2\exp\left\{-\frac{z_1^2}{2\sigma_1^2}\right\}dz_1 \\
&= -\log\left(\sqrt{2\pi\sigma_1^2}\right)\cdot 1 - \frac{1}{2\sigma_1^2}\frac{1}{\sqrt{2\pi\sigma_1^2}}\int z_1^2\exp\left\{-\frac{z_1^2}{2\sigma_1^2}\right\}dz_1 \\
&= -\log\sqrt{\pi\sigma_1^2} - \frac{1}{2}\frac{1}{\sqrt{2\pi\sigma_1^6}}\cdot\frac{1}{2}\sqrt{\frac{\pi}{\frac{1}{2\sigma_1^2}^3}} \\
&= -\log\sqrt{2\pi\sigma_1^2} - \frac{1}{2}
\end{aligned}
$$

□

Exactly the same procedure can be followed for q_2.

b) Use this result to show that

$$
KL(q||p) = -1 - \log\sqrt{2\pi\sigma_1^2} - \log\sqrt{2\pi\sigma_2^2} + \frac{1}{2}a\left(\sigma_1^2 + \sigma_2^2\right)
$$

$$
\begin{aligned}
KL(q||p) &= -\int q(\mathbf{z})\ln\frac{p(\mathbf{z})}{q(\mathbf{z})}d\mathbf{z} \\
&= \int q(\mathbf{z})\ln q(\mathbf{z})d\mathbf{z} - \int q(\mathbf{z})\ln p(\mathbf{z})d\mathbf{z}
\end{aligned}
$$

the first part

$$
\begin{aligned}
A &= \int q(\mathbf{z})\ln q(\mathbf{z})d\mathbf{z} \\
&= \int\int q_1(z_1)q_2(z_2)\ln q_1(z_1)q_2(z_2)dz_1dz_2 \\
&= \int q_2(z_2)dz_2\int q_1(z_1)\ln q_1(z_1)dz_1 + \int q_1(z_1)dz_1\int q_2(z_2)\ln q_2(z_2)dz_2 \\
&= 1\cdot\int q_1(z_1)\ln q_1(z_1)dz_1 + 1\cdot\int q_2(z_2)\ln q_2(z_2)dz_2 \\
&= -\log\sqrt{2\pi\sigma_1^2} - \frac{1}{2} - \log\sqrt{2\pi\sigma_2^2} - \frac{1}{2} \\
&= -1 - \log\sqrt{2\pi\sigma_1^2} - \log\sqrt{2\pi\sigma_2^2}
\end{aligned}
$$

second part E(x) is 0 so cross terms disappear.
$E(x^2)$:

$$
\begin{aligned}
\int z_i^2 q_i(z_i) dz_i &= \frac{1}{\sqrt{2\pi\sigma_i^2}} \int z_i^2 \exp\left\{-\frac{z_i^2}{2\sigma_i^2}\right\} dz_i \\
&= \frac{1}{\sqrt{2\pi\sigma_i^2}} \sqrt{2\pi\sigma_i^6} \\
&= \sigma_i^2
\end{aligned}
$$

$$
\begin{aligned}
B &= \int q(\mathbf{z}) \ln p(\mathbf{z}) d\mathbf{z} \\
&= -\frac{1}{2} \int q_1(z_1) q_2(z_2) \cdot \left(a z_1^2 + a z_2^2 + 2b z_1 z_2\right) dz_1 dz_2 \\
&= -\frac{1}{2} \int q_1(z_1) q_2(z_2) a z_1^2 dz_1 dz_2 - \frac{1}{2} \int q_1(z_1) q_2(z_2) a z_2^2 dz_1 dz_2 \\
&= -\frac{1}{2} \int q_2(z_2) dz_2 \int q_1(z_1) a z_1^2 dz_1 - \frac{1}{2} \int q_1(z_1) dz_1 \int q_2(z_2) a z_2^2 dz_2 \\
&= -\frac{a}{2} \left(\int q_1(z_1) z_1^2 dz_1 + \int q_2(z_2) z_2^2 dz_2 \right) \\
&= -\frac{a}{2} \left(\sigma_1^2 + \sigma_2^2\right)
\end{aligned}
$$

So

$$
KL(q||p) = A - B = -1 - \log\sqrt{2\pi\sigma_1^2} - \log\sqrt{2\pi\sigma_2^2} + \frac{1}{2} a \left(\sigma_1^2 + \sigma_2^2\right)
$$

$\square$

c)

$$
\begin{aligned}
\frac{dKL(q||p)}{d\sigma_1} &= -\frac{1}{\sigma_1} + a\sigma_1 = 0 \rightarrow \sigma_1^2 = \frac{1}{a} \\
\frac{dKL(q||p)}{d\sigma_2} &= -\frac{1}{\sigma_2} + a\sigma_2 = 0 \rightarrow \sigma_2^2 = \frac{1}{a}
\end{aligned}
$$

d)

$$
\begin{aligned}
KL(p||q) &= -\int p(\mathbf{z}) \ln \frac{q(\mathbf{z})}{p(\mathbf{z})} d\mathbf{z} \\
&= \int p(\mathbf{z}) \ln p(\mathbf{z}) d\mathbf{z} - \int p(\mathbf{z}) \ln q(\mathbf{z}) d\mathbf{z} \\
A &= \int p(\mathbf{z}) \ln p(\mathbf{z}) d\mathbf{z} \\
&= \int -\frac{1}{2} \left(a z_1^2 + a z_2^2 + 2b z_1 z_2 \right) p(\mathbf{z}) d\mathbf{z} \\
&= -\frac{a}{2} \int \left(z_1^2 + z_2^2 \right) p(\mathbf{z}) d\mathbf{z} \\
B &= \int p(\mathbf{z}) \ln q(\mathbf{z}) d\mathbf{z} \\
&= -\ln 2\pi\sigma_1\sigma_2 \int p(\mathbf{z}) d\mathbf{z} - \frac{1}{2} \int \left(\frac{z_1^2}{\sigma_1^2} + \frac{z_2^2}{\sigma_2^2} \right) p(\mathbf{z}) d\mathbf{z} \\
&= -\ln 2\pi\sigma_1\sigma_2 \int p(\mathbf{z}) d\mathbf{z} - \frac{1}{2\sigma_1^2} \int z_1^2 p(\mathbf{z}) d\mathbf{z} - \frac{1}{2\sigma_2^2} \int z_2^2 p(\mathbf{z}) d\mathbf{z}
\end{aligned}
$$

$$
\begin{aligned}
KL(p||q) &= A - B \\
&= -\frac{a}{2} \int \left(z_1^2 + z_2^2 \right) p(\mathbf{z}) d\mathbf{z} + \ln 2\pi\sigma_1\sigma_2 \int p(\mathbf{z}) d\mathbf{z} + \frac{1}{2\sigma_1^2} \int z_1^2 p(\mathbf{z}) d\mathbf{z} + \frac{1}{2\sigma_2^2} \int z_2^2 p(\mathbf{z}) d\mathbf{z} \\
&= \ln 2\pi\sigma_1\sigma_2 \int p(\mathbf{z}) d\mathbf{z} + (\frac{1}{2\sigma_1^2} - \frac{a}{2}) \int z_1^2 p(\mathbf{z}) d\mathbf{z} + (\frac{1}{2\sigma_2^2} - \frac{a}{2}) \int z_2^2 p(\mathbf{z}) d\mathbf{z} \\
&= \ln 2\pi\sigma_1\sigma_2 \cdot 1 + (\frac{1}{2\sigma_1^2} - \frac{a}{2}) \mathbb{E}(z_1^2) + (\frac{1}{2\sigma_2^2} - \frac{a}{2}) \mathbb{E}(z_2^2)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial KL(p||q)}{\partial \sigma_1} &= \frac{1}{\sigma_1} - \frac{1}{\sigma_1^3} \mathbb{E}(z_1^2) = 0 \\
\sigma_1^2 &= \mathbb{E}(z_1^2) \\
\frac{\partial KL(p||q)}{\partial \sigma_2} &= \frac{1}{\sigma_2} - \frac{1}{\sigma_2^3} \mathbb{E}(z_2^2) = 0 \\
\sigma_2^2 &= \mathbb{E}(z_2^2)
\end{aligned}
$$

e) $\mathbb{E}_p(z_i z_j) = \Lambda_{ij}^{-1}$ compute inverse of $\Lambda = \begin{pmatrix} a & b \\ b & a \end{pmatrix}$

$\Lambda^{-1} = \frac{1}{det(\Lambda)} \begin{pmatrix} a & -b \\ -b & a \end{pmatrix}$ which is $\begin{pmatrix} \frac{a}{a^2-b^2} & \frac{-b}{a^2-b^2} \\ \frac{-b}{a^2-b^2} & \frac{a}{a^2-b^2} \end{pmatrix}$ so $\mathbb{E}_p(z_1^2) = \sigma_{z_1}^2 = \mathbb{E}_p(z_2^2) = \sigma_{z_2}^2 = \frac{a}{a^2-b^2}$

Week 2

Exercise 3.12 MacKay

If the original counter was white, then P(D,w)=1 If the original counter was black, then P(D,b)=0.5 assume equal priors: P(w)=P(b)=0.5

$$\begin{aligned}
P(w \mid D) &= \frac{P(D \mid w)P(w)}{P(D)} \\
&= \frac{P(D \mid w)P(w)}{P(D \mid w)P(w) + P(D \mid b)P(b)} = \frac{2}{3}
\end{aligned}$$

Exercise 28.1 MacKay

$$\begin{aligned}
P(x \mid \mathbb{H}_0) &= \frac{1}{2} && x \in (-1, 1) \\
P(x \mid m, \mathbb{H}_1) &= \frac{1}{2}(1 + mx) && x \in (-1, 1)
\end{aligned}$$

equal priors:

$$\begin{aligned}
P(\mathbb{H}_0) &= P(\mathbb{H}_1) \\
P(\mathbb{H}_0 \mid D) &= \frac{P(D \mid \mathbb{H}_0)P(\mathbb{H}_0)}{P(D)} \\
P(\mathbb{H}_1 \mid D) &= \frac{P(D \mid \mathbb{H}_1)P(\mathbb{H}_1)}{P(D)} \\
\frac{P(\mathbb{H}_1 \mid D)}{P(\mathbb{H}_0 \mid D)} &= \frac{P(D \mid \mathbb{H}_1)}{P(D \mid \mathbb{H}_0)}
\end{aligned}$$

We have $N = 5$ datapoints $D = \{0.3, 0.5, 0.7, 0.8, 0.9\}$. So the evidencec for $\mathbb{H}_0$ and $\mathbb{H}_1$ are:

$$\begin{aligned}
P(D \mid \mathbb{H}_0) &= \prod_{n=1}^{N} P(x_n \mid \mathbb{H}_0) = \frac{1}{2^5} \\
P(D \mid \mathbb{H}_1, m) &= \prod_{n=1}^{N} P(x_n \mid \mathbb{H}_1) \\
&= \frac{1}{2^5}(1 + 0.3m)(1 + 0.5m)(1 + 0.7m)(1 + 0.8m)(1 + 0.9m) \\
&= \frac{1}{2^5}\left(\frac{189m^5}{2500} + \frac{6897m^4}{10000} + \frac{299m^3}{125} + \frac{199m^2}{50} + \frac{16m}{5} + 1\right)
\end{aligned}$$

Assuming m is uniformly distributed: $P(m, \mathbb{H}_1) = \frac{1}{2}$. So the evidence for $\mathbb{H}_1$ is:

$$P(D \mid \mathbb{H}_1) = \int_{-1}^{1} P(D \mid m, \mathbb{H}_1) P(m, \mathbb{H}_1) dm = \frac{1}{2^5} \frac{369691}{150000}$$

$$\frac{P(\mathbb{H}_1 \mid D)}{P(\mathbb{H}_0 \mid D)} = \frac{P(D \mid \mathbb{H}_1)}{P(D \mid \mathbb{H}_0)} = \frac{369691}{150000} \approx 2.46$$

Exercise 3.1

a) We have to compare the model $\mathcal{H}_0$, where the coin lands on heads with probability $\frac{1}{2}$ to model $\mathcal{H}_1$ where it lands on head with probability p_h which can be anything between 0 and 1. A priori we assume both models to be equally likely.

First we work out the case where the number of heads is zero out of two.

$$p(\mathcal{H}_0 | TT, 2 \text{ throws}) = \frac{p(TT | 2 \text{ throws}, p_h = \frac{1}{2}) p(\mathcal{H}_0)}{p(TT | 2 \text{ throws})} = \frac{\frac{1}{2}^2 \cdot \frac{1}{2}}{\frac{1}{8} + \frac{1}{6}}$$

$$p(\mathcal{H}_1 | TT, 2 \text{ throws}) = \frac{p(TT | 2 \text{ throws}, p_h \sim U(0,1)) p(\mathcal{H}_1)}{p(TT | 2 \text{ throws})} = \frac{\int_0^1 (1-p_h)^2 dp_h \cdot \frac{1}{2}}{\frac{1}{8} + \frac{1}{6}} = \frac{\frac{1}{3 \cdot 2}}{\frac{1}{8} + \frac{1}{6}}$$

So the probability of $\mathcal{H}_0$ is $\frac{3}{7}$, while the probability of $\mathcal{H}_1$ is $\frac{4}{7}$.

Now to to this exercise for when the number of heads is 1:

$$p(\mathcal{H}_0 | HT \vee TH, 2 \text{ throws}) = \frac{p(HT \vee TH | 2 \text{ throws}, p_h = \frac{1}{2}) p(\mathcal{H}_0)}{p(HT \vee TH | 2 \text{ throws})} = \frac{2 \cdot \frac{1}{2}^2 \cdot \frac{1}{2}}{\frac{1}{4} + \frac{1}{6}}$$

$$p(\mathcal{H}_1 | HT \vee TH, 2 \text{ throws}) = \frac{p(HT \vee TH | 2 \text{ throws}, p_h \sim U(0,1)) p(\mathcal{H}_1)}{p(HT \vee TH | 2 \text{ throws})} = \frac{\int_0^1 2(1-p_h) p_h dp_h \cdot \frac{1}{2}}{\frac{25}{72} + \frac{1}{6}} = \frac{\frac{1}{6}}{\frac{1}{4} + \frac{1}{6}}$$

Here we see that the probability of $\mathcal{H}_0$ is $\frac{3}{5}$, while the probability of $\mathcal{H}_1$ is $\frac{2}{5}$.

Lastly we look at this for the case that we get two heads:

$$p(\mathcal{H}_0 | HH, 2 \text{ throws}) = \frac{p(HH | 2 \text{ throws}, p_h = \frac{1}{2}) p(\mathcal{H}_0)}{p(HH | 2 \text{ throws})} = \frac{\frac{1}{2}^2 \cdot \frac{1}{2}}{\frac{1}{8} + \frac{1}{6}} =$$

$$p(\mathcal{H}_1 | HH, 2 \text{ throws}) = \frac{p(HH | 2 \text{ throws}, p_h \sim U(0,1)) p(\mathcal{H}_1)}{p(HH | 2 \text{ throws})} = \frac{\int_0^1 p_h^2 dp_h \cdot \frac{1}{2}}{\frac{1}{8} + \frac{1}{6}} = \frac{\frac{1}{6}}{\frac{1}{8} + \frac{1}{6}}$$

So the probability of $\mathcal{H}_0$ is $\frac{3}{7}$, while the probability of $\mathcal{H}_1$ is $\frac{4}{7}$.

b) Model $\mathcal{H}_0$ is very good at explaining results with about $\frac{1}{2}$ heads, but not at explaining other results. The other model can explain everything, since it can have any p_h.

Model $\mathcal{H}_0$ is the small but high peak, but $\mathcal{H}_1$ is the broad, flat distribution.

Exercise 27.1

a)

We want to infer λ the rate of photons arriving at our photon counter. We collect r photons, where r follows a Poisson distribution with rate λ. So to find the posterior distribution of λ we use Bayes' Rule:

$$p(\lambda|r) = \frac{p(r|\lambda)p(\lambda)}{p(r)} = \frac{\exp(-\lambda)\frac{\lambda^r}{r!}\frac{1}{\lambda}}{p(r)} = \frac{\exp(-\lambda)\frac{\lambda^{r-1}}{r!}}{p(r)}$$

Note that we haven't calculated the normalising constant, as that is what we are going to approximate with Laplace's method. We differentiate the numerator with regards to λ and set this derivative to zero to find the peak of this distribution.

$$\frac{\partial p(\lambda|r)}{\partial \lambda} = -\exp(-\lambda)\frac{\lambda^{r-1}}{r!} + \exp(-\lambda)(r-1)\frac{\lambda^{r-2}}{r!} = 0$$

This holds if $\lambda = 0$, but this is not a valid parameter for the Poisson distribution, or if $-\frac{\lambda^{r-1}}{r!} + (r-1)\frac{\lambda^{r-2}}{r!} = 0$ which we simplify to

$$\begin{aligned}
-\frac{\lambda^{r-1}}{r!} + (r-1)\frac{\lambda^{r-2}}{r!} &= 0 \\
\frac{\lambda^{r-1}}{r!} &= (r-1)\frac{\lambda^{r-2}}{r!} \\
\lambda^{r-1} &= (r-1)\lambda^{r-2} \\
\lambda &= (r-1)
\end{aligned}$$

So we expand around $r-1$.

$$\ln p(\lambda|r) \simeq \ln p(\lambda = (r-1)|r) - \frac{c}{2}(\lambda - (r-1))^2$$

where

$$
\begin{aligned}
c &= -\frac{\partial^2}{\partial\lambda^2} \ln p(\lambda|r)\Big|_{\lambda=r-1} \\
c &= -\frac{\partial^2}{\partial\lambda^2} \ln\left(\exp(-\lambda)\frac{\lambda^{r-1}}{r!}\right)\Big|_{\lambda=r-1} \\
c &= -\frac{\partial^2}{\partial\lambda^2}\left(-\lambda + \ln\left(\frac{\lambda^{r-1}}{r!}\right)\right)\Big|_{\lambda=r-1} \\
c &= -\frac{\partial}{\partial\lambda}\left(-1 + \left(\frac{1}{\frac{\lambda^{r-1}}{r!}} \cdot (r-1) \cdot \frac{\lambda^{r-2}}{r!}\right)\right)\Big|_{\lambda=r-1} \\
c &= -\frac{\partial}{\partial\lambda}\left(-1 + \left(\frac{r!}{\lambda^{r-1}} \cdot (r-1) \cdot \frac{\lambda^{r-2}}{r!}\right)\right)\Big|_{\lambda=r-1} \\
c &= -\frac{\partial}{\partial\lambda}\left(-1 + \frac{r-1}{\lambda}\right)\Big|_{\lambda=r-1} \\
c &= -\left(-\frac{r-1}{\lambda^2}\right)\Big|_{\lambda=r-1} \\
c &= \frac{1}{r-1}
\end{aligned}
$$

So we can approximate $p(\lambda|r)$ with the Gaussian

$$
\begin{aligned}
Q(\lambda|r) \equiv& p(r-1|r)\exp\left(-\frac{c}{2}(\lambda-(r-1))^2\right) \\
=& \exp(1-r)\frac{(r-1)^{(r-1)}}{r!}\exp\left(-\frac{1}{2(r-1)}(\lambda-(r-1))^2\right)
\end{aligned}
$$

with normalizing constant $Z_Q = p(\lambda = r-1|r)\sqrt{\frac{2\pi}{c}} = \exp(1-r)\frac{(r-1)^{(r-1)}}{r!}\sqrt{2\pi(r-1)}$

Now we plug in this Z_Q as the approximate normalizing constant for our Gaussian and we get

$$
\begin{aligned}
p_1(\lambda|r) &= \frac{Q(\lambda|r)}{Z_Q} \\
&= \frac{\exp(1-r)\frac{(r-1)^{(r-1)}}{r!}\exp\left(-\frac{1}{2(r-1)}(\lambda-(r-1))^2\right)}{\exp(1-r)\frac{(r-1)^{(r-1)}}{r!}\sqrt{2\pi(r-1)}} \\
&= \frac{\exp\left(-\frac{1}{2(r-1)}(\lambda-(r-1))^2\right)}{\sqrt{2\pi(r-1)}}
\end{aligned}
$$

b)

We have our original prior $p(\lambda) = \frac{1}{\lambda}$ and we consider the transformation $y = \log(\lambda)$ where log is the natural logarithm. This means that $\lambda = \exp(y)$ and $\frac{d\lambda}{dy}$ is also $\exp(y)$.

Now we use the hint that $p(\lambda)d\lambda = p(y)dy$, and rewrite that into $p(y) = p(\lambda)\frac{d\lambda}{dy}$. Filling in all the values we get $p(y) = \frac{1}{\lambda}\exp(y) = \frac{1}{\exp(y)}\exp(y) = 1$.

c)

We have the unnormalized posterior distribution $p(\lambda|r) = \exp(-\lambda)\frac{\lambda^{r-1}}{r!}$, and we go about transforming it. We use the transformation $y = \log(\lambda)$. So we replace all λs with $\exp(y)$ and get

$$p(y|r) = \exp(-\exp(y))\frac{\exp(y)^{r-1}}{r!}$$

Now we will do the Laplace approximation again, starting by calculating the point y_0 around which to expand.

$$\frac{dp(y|r)}{dy} = \frac{\exp(-\exp(y) + (-1+r)y)(-1-\exp(y)+r)}{r!} = 0$$

Since the exp is never zero, this only happens when $(-1 - \exp(y) + r) = 0$, so if $y_0 = \log(r-1)$. Of course, we could have known this already because $y = \log(\lambda)$, so $y_0 = \log(\lambda_0)$.

Now we Taylor-expand the logarithm of $p(y|r)$ around $y_0 = \log(r-1)$

$$\begin{aligned}\log(p(y|r)) \simeq &\log\left(\exp(-\exp(\log(r-1)))\frac{\exp(\log(r-1))^{r-1}}{r!}\right) - \frac{c}{2}(y - \log(r-1))^2\\ =&(1-r) + \log\left(\frac{(r-1)^{r-1}}{r!}\right) - \frac{c}{2}(y - \log(r-1))^2\end{aligned}$$

Where

$$\begin{aligned}c &= -\frac{\partial^2}{\partial y^2}\log p(y|r)\Big|_{y=\log(r-1)}\\ c &= -\frac{\partial^2}{\partial y^2}\ln\left(\exp(-\exp(y))\frac{\exp(y)^{r-1}}{r!}\right)\Bigg|_{y=\log(r-1)}\\ c &= -\frac{\partial^2}{\partial y^2} - \exp(y) + y(r-1) - \log(r!)\Big|_{y=\log(r-1)}\\ c &= - - \exp(y)\Big|_{y=\log(r-1)}\\ c &= r-1\end{aligned}$$

So we can approximate $p(y|r)$ with a Gaussian

$$\begin{aligned}Q(y|r) &\equiv p(\log(r-1)|r)\exp\left(-\frac{c}{2}(y-\log(r-1))^2\right)\\ &= \exp(1-r)\frac{(r-1)^{(r-1)}}{r!}\exp\left(-\frac{r-1}{2}(y-\log(r-1))^2\right)\end{aligned}$$

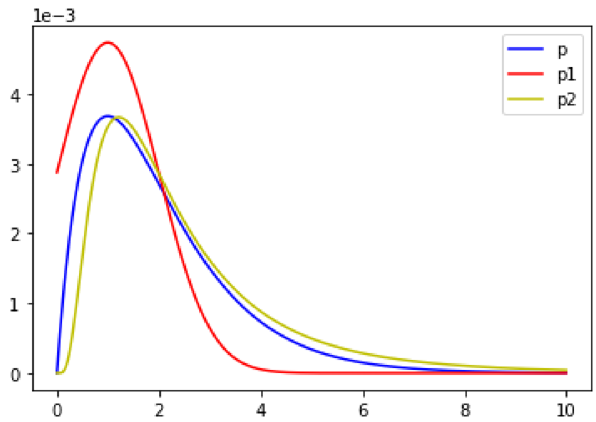

Figure 1: plot of p and its two Laplace approximations for r=2

with normalizing constant

$$Z_Q = p(y_0|r)\sqrt{\frac{2\pi}{c}} = \exp(1-r)\frac{(r-1)^{(r-1)}}{r!}\sqrt{\frac{2\pi}{r-1}}$$

And this is then approximately equal to the normalizing constant Z_p of $p(y|r)$.

So now the second laplace approximation

$$p_2(y|r) = \frac{\exp\left(-\frac{r-1}{2}(y-\log(r-1))^2\right)}{\sqrt{\frac{2\pi}{r-1}}}$$

d)

Now we substitute $y = \log(\lambda)$ and multiply by our original prior again to go from $p_2(y|r)$ to $p_2(\lambda|r)$ so we get

$$p_2(\lambda|r) = \frac{1}{\lambda}\frac{\exp\left(-\frac{r-1}{2}(\log(\lambda)-\log(r-1))^2\right)}{\sqrt{\frac{2\pi}{r-1}}} = \frac{\exp\left(-\frac{r-1}{2}(\log\left(\frac{\lambda}{r-1}\right))^2\right)}{\lambda\sqrt{\frac{2\pi}{r-1}}}$$

The plots for p, p_1, p_2 are shown on the next page.

Exercise 28.1

a

$$P(t_i|w_0,w_1) = N(t_i \mid w_0 + w_1 x_i, \sigma^2)$$
$$P(t_i|w_0,w_1) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-(t_i - w_0 - w_1 x_i)^2}{2\sigma^2}\right)$$

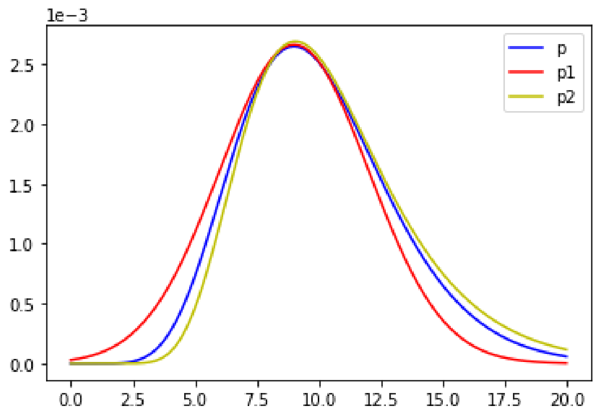

Figure 2: plot of p and its two Laplace approximations for r=10

As the data points are independent, the probability of the data is just the product of the probability of the individual data points.

$$\begin{aligned}
P(D|w_0, w_1) =& \prod_{i=1}^{N} (t_i|w_0, w_1) \\
=& \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(t_i - w_0 - w_1 x_i)^2}{2\sigma^2} \right) \\
=& \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(\sum_{i=1}^{N} \frac{-(t_i - w_0 - w_1 x_i)^2}{2\sigma^2} \right) \\
=& \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(\sum_{i=1}^{N} \frac{-(t_i^2 - 2w_0 t_i - 2w_1 x_i t_i + w_0^2 - w_0 w_1 x_i + w_1^2 x_i^2)}{2\sigma^2} \right)
\end{aligned}$$

Now using that x_i is centered so the sum is 0:

$$\begin{aligned}
P(D|w_0, w_1) =& \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{\sum_{i=1}^{N} t_i^2}{2\sigma^2} \right) \exp\left(\frac{\sum_{i=1}^{N} 2w_0 t_i + N w_0^2}{2\sigma^2} \right) \exp\left(\frac{\sum_{i=1}^{N} (2w_1 x_i t_i - w_1^2 x_i^2)}{2\sigma^2} \right) \\
=& \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{N} t_i^2 \right) \exp\left(\frac{w_0}{\sigma^2} \sum_{i=1}^{N} t_i - \frac{N w_0^2}{2\sigma^2} \right) \exp\left(\frac{w_1}{\sigma^2} \sum_{i=1}^{N} x_i t_i - \frac{w_1^2}{2\sigma^2} \sum_{i=1}^{N} x_i^2 \right)
\end{aligned}$$

So the probability is a product of a term that does not depend on either w_0 or w_1, one that purely depends on w_0 and one that only depends on w_1:

$$P(D|w_0, w_1) = C f(w_0) g(w_1)$$

b

$$P(D|\mathcal{H}_i) = \int \int P(D|w_0, w_1) P(w_0) P(w_1) \mathrm{d}w_0 \mathrm{d}w_1$$

$$
\begin{aligned}
P(D|\mathcal{H}_1) =& \int\int C f(w_0) g(w_1) N(w_0|0,1)\delta(w_1)\mathrm{d}w_0\mathrm{d}w_1 \\
=& \frac{Cg(0)}{\sqrt{2\pi}}\int f(w_0)\exp\left(-\frac{w_0^2}{2}\right)\mathrm{d}w_0
\end{aligned}
$$

$$
\begin{aligned}
P(D|\mathcal{H}_2) =& \int\int C f(w_0) g(w_1) N(w_0|0,1) N(w_1|0,1)\mathrm{d}w_0\mathrm{d}w_1 \\
=& \frac{C}{2\pi}\int\int f(w_0) g(w_1)\exp\left(-\frac{w_0^2}{2}\right)\exp\left(-\frac{w_1^2}{2}\right)\mathrm{d}w_0\mathrm{d}w_1 \\
=& \frac{C}{2\pi}\int f(w_0)\exp\left(-\frac{w_0^2}{2}\right)\mathrm{d}w_0\int g(w_1)\exp\left(-\frac{w_1^2}{2}\right)\mathrm{d}w_1
\end{aligned}
$$

So the ratio between the two is given by

$$
\begin{aligned}
P(D|\mathcal{H}_2)/P(D|\mathcal{H}_1) =& \frac{\frac{C}{2\pi}\int f(w_0)\exp\left(-\frac{w_0^2}{2}\right)\mathrm{d}w_0\int g(w_1)\exp\left(-\frac{w_1^2}{2}\right)\mathrm{d}w_1}{\frac{Cg(0)}{\sqrt{2\pi}}\int f(w_0)\exp\left(-\frac{w_0^2}{2}\right)\mathrm{d}w_0} \\
=& \frac{1}{\sqrt{2\pi}}\int g(w_1)\exp\left(-\frac{w_1^2}{2}\right)\mathrm{d}w_1 \\
=& \frac{1}{\sqrt{2\pi}}\int \exp\left(\frac{w_1}{\sigma^2}\sum_{i=1}^{N} x_i t_i - \frac{w_1^2}{2\sigma^2}\sum_{i=1}^{N} x_i^2\right)\exp\left(-\frac{w_1^2}{2}\right)\mathrm{d}w_1 \\
=& \frac{1}{\sqrt{2\pi}}\int \exp\left(\frac{N\left\langle xt\right\rangle}{\sigma^2}w_1 - \left(\frac{N\left\langle x^2\right\rangle}{2\sigma^2}+\frac{1}{2}\right)w_1^2\right)\mathrm{d}w_1
\end{aligned}
$$

Using the standard integral ($a > 0$)

$$
\int_{-\infty}^{\infty}\exp(-ax^2 - bx)\mathrm{d}x = \sqrt{\frac{\pi}{a}}\exp\left(\frac{b^2}{4a}\right)
$$

this can be calculated:

$$
\begin{aligned}
P(D|\mathcal{H}_2)/P(D|\mathcal{H}_1) =& \frac{1}{\sqrt{2\pi}} \int \exp\left(\frac{N\left\langle xt\right\rangle}{\sigma^2} w_1 - \left(\frac{N\left\langle x^2\right\rangle}{2\sigma^2} + \frac{1}{2}\right) w_1^2\right) \mathrm{d}w_1 \\
=& \frac{1}{\sqrt{2\pi}} \sqrt{\frac{2\pi}{\frac{N\left\langle x^2\right\rangle}{\sigma^2} + 1}} \exp\left(\frac{\left(\frac{N\left\langle xt\right\rangle}{\sigma^2}\right)^2}{\frac{2N\left\langle x^2\right\rangle}{\sigma^2} + 2}\right) \\
=& \sqrt{\frac{\sigma^2}{N\left\langle x^2\right\rangle + \sigma^2}} \exp\left(\frac{N^2 \left\langle xt\right\rangle^2}{\sigma^4} \frac{\sigma^2}{2N\left\langle x^2\right\rangle + 2\sigma^2}\right) \\
=& \frac{1}{\sqrt{2\pi}} \sqrt{\frac{2\pi}{\frac{N\left\langle x^2\right\rangle}{\sigma^2} + 1}} \exp\left(\frac{\left(\frac{N\left\langle xt\right\rangle}{\sigma^2}\right)^2}{\frac{2N\left\langle x^2\right\rangle}{\sigma^2} + 2}\right) \\
=& \sqrt{\frac{\sigma^2}{N\left\langle x^2\right\rangle + \sigma^2}} \exp\left(\frac{N^2 \left\langle xt\right\rangle^2}{2N\sigma^2\left\langle x^2\right\rangle + 2\sigma^4}\right)
\end{aligned}
$$

c

$\sigma^2 = \left\langle x^2\right\rangle = 1$:

$$
P(D|\mathcal{H}_2)/P(D|\mathcal{H}_1) = \sqrt{\frac{1}{N+1}} \exp\left(\frac{N^2\left\langle xt\right\rangle^2}{2N+2}\right)
$$

In the limit for $N \to \infty$ and $\left\langle xt\right\rangle^2 \leq \frac{\log N}{N}$

$$
\begin{aligned}
P(D|\mathcal{H}_2)/P(D|\mathcal{H}_1) \leq& \sqrt{\frac{1}{N}} \exp\left(\frac{N^2 \frac{\log N}{N}}{2N}\right) \\
=& \sqrt{\frac{1}{N}} \exp\left(\frac{\log N}{2}\right) = \sqrt{\frac{1}{N}} \exp\left(\log\sqrt{N}\right) = 1
\end{aligned}
$$

So if $\left\langle xt\right\rangle^2 \leq \frac{\log N}{N}$, $\mathcal{H}_2$ is preferred. Naturally, otherwise $\mathcal{H}_1$ is preferred.

Exercise 28.2

a

The probability is

$$
P(D|\mathcal{H}_0) = \frac{n!}{n_1! \dots n_k!} \frac{1}{k^n}
$$

Or, written in another way:

$$
P(D|\mathcal{H}_0) = \frac{n!}{k^n} \prod_{i=1}^{k} \frac{1}{n_i!}
$$

b

In a similar vein as a, if we know all of the p_i we can calculate the probabiltity of the data given that $\{p_i\}$:

$$P(D|\{p_i\}, \mathcal{H}_1) = n! \prod_{i=1}^{k} \frac{1}{n_i!} p_i^{n_i}$$

Now, to get the probability for all possible $\{p_i\}$ we have to integrate over all of them while satisfying the constraint that their sum is 1.

$$P(D|\mathcal{H}_1) = \int_0^1 \mathrm{d}p_1 \ldots \int_0^1 \mathrm{d}p_k n! \prod_{i=1}^{k} \frac{1}{n_i!} p_i^{n_i} \delta\left(\sum_i p_i - 1\right)$$

This is equal to $n!$ times the integral given in the hint with $f(p_i) = \frac{1}{n_i!} p_i^{n_i}$ so we can use it to write:

$$P(D|\mathcal{H}_1) = (k-1)!n! \int_{simplex} \mathrm{d}\vec{\mathbf{p}} n! \prod_{i=1}^{k} \frac{1}{n_i!} p_i^{n_i} \delta\left(\sum_i p_i - 1\right)$$

This integral can be written in terms or the Dirichlet distribution:

$$P(D|\mathcal{H}_1) = \frac{(k-1)!n!}{\prod_{i=1}^{k} n_i} (k-1)!n! \int_{simplex} \mathrm{d}\vec{\mathbf{p}} Z(\alpha\mathbf{m}) * \mathrm{Dirichlet}^I(\mathbf{p}|\alpha\mathbf{m})$$

For this last step we use the substitution

$$\begin{aligned}
\alpha m_i - 1 =& n_i \\
\sum_i \alpha m_i - k =& \sum_i n_i \\
\alpha \sum_i m_1 =& n + k \\
\alpha =& n + k
\end{aligned}$$

The integral of the probability distribution is one, and Z does not depend on $\vec{\mathbf{p}}$ so:

$$\begin{aligned}
P(D|\mathcal{H}_1) =& \frac{(k-1)!n!}{\prod_i n_i!} Z(\alpha\mathbf{m}) \\
=& (k-1)!n! \prod_i \frac{1}{n_i!} \frac{\Gamma(\alpha m_i)}{\Gamma(\alpha)} \\
=& (k-1)!n! \prod_i \frac{\Gamma(n_i+1)}{n_i!} \frac{1}{\Gamma(n+k)} \\
=& \frac{(k-1)!n!}{(n+k-1)!^k}
\end{aligned}$$

c

$$P(\mathcal{H}_0|D) = \frac{P(D|\mathcal{H}_0)P(\mathcal{H}_0)}{P(D)}$$
$$P(\mathcal{H}_1|D) = \frac{P(D|\mathcal{H}_1)P(\mathcal{H}_1)}{P(D)}$$

$$P(\mathcal{H}_0) = P(\mathcal{H}_1) = \frac{1}{2}$$

Because the hypotheses are equally likely and the data is the same, just comparing the priors would do. The exercise asks for the posteriors though...

In the first case:

$$P(D|\mathcal{H}_0) = \frac{30!}{6^{30}}\frac{1}{3!}\frac{1}{3!}\frac{1}{2!}\frac{1}{2!}\frac{1}{9!}\frac{1}{11!} = \frac{30!}{3!3!2!2!9!11!6^{30}} \approx 5.75e-7$$

$$P(D|\mathcal{H}_1) = \frac{5!30!}{35!^6} \approx 3.08e-6$$

$$P(D) = \frac{1}{2}\frac{30!}{3!3!2!2!9!11!6^{30}} + \frac{1}{2}\frac{5!30!}{35!^6} \approx 1.83e-6$$

$$P(\mathcal{H}_0|D) = \frac{\frac{30!}{3!3!2!2!9!11!6^{30}}\frac{1}{2}}{\frac{1}{2}\frac{30!}{3!3!2!2!9!11!6^{30}} + \frac{1}{2}\frac{5!30!}{35!^6}} \approx 0.157$$

$$P(\mathcal{H}_1|D) = \frac{\frac{5!30!}{35!^6}\frac{1}{2}}{\frac{1}{2}\frac{30!}{3!3!2!2!9!11!6^{30}} + \frac{1}{2}\frac{5!30!}{35!^6}} \approx 0.843$$

So here hypothesis 1 is more likely, as it should be.

In the second case:

$$P(D|\mathcal{H}_0) = \frac{30!}{6^{30}}\frac{1}{5!^6} = \frac{30!}{5!^6 6^{30}} \approx 4.01e-4$$

$$P(D|\mathcal{H}_1) = \frac{5!30!}{35!^6} \approx 3.08e-6$$

$$P(D) = \frac{1}{2}\frac{30!}{5!^6 6^{30}} + \frac{1}{2}\frac{5!30!}{35!^6} \approx 2.02e-4$$

$$P(\mathcal{H}_0|D) = \frac{\frac{30!}{5!^6 6^{30}}\frac{1}{2}}{\frac{1}{2}\frac{30!}{5!^6 6^{30}} + \frac{1}{2}\frac{5!30!}{35!^6}} \approx 0.992$$

$$P(\mathcal{H}_1|D) = \frac{\frac{5!30!}{35!^6}\frac{1}{2}}{\frac{1}{2}\frac{30!}{5!^6 6^{30}} + \frac{1}{2}\frac{5!30!}{35!^6}} \approx 0.008$$

So here hypothesis 0 is more likely, as it should be.

Week 3

Exercise 1

$$C(P,N) = 2\sum_{i=0}^{N-1}\binom{P-1}{i} \tag{2}$$

There are 2^P possible patterns. So the fraction $R = \frac{C(P,N)}{2^P}$ of the problems is linearly separable.

a

In the case $P \leq N$

$$\begin{aligned} R =& \frac{1}{2^P}\left(2\sum_{i=0}^{P-1}\binom{P-1}{i} + 2\sum_{i=P}^{N-1}\binom{P-1}{i}\right) \\ =& \frac{1}{2^{P-1}}2^{P-1} = 1 \end{aligned}$$

So all problems are linearly separable if $P \leq N$.

b

In the case $P = 2N$

$$\begin{aligned} R =& \frac{2}{2^{2N}}\sum_{i=0}^{N-1}\binom{2N-1}{i} \\ =& \frac{1}{2^{2N-1}}\sum_{i=0}^{N-1}\binom{2N-1}{i} \end{aligned}$$

The sum is evaluated as:

$$\sum_{i=0}^{N-1}\binom{2N-1}{i} = \sum_{i=2N-1}^{N}\binom{2N-1}{2N-1-i} = \frac{1}{2}\sum_{i=0}^{2N-1}\binom{2N-1}{i} = 2^{2N-2}$$

Thus

$$R = \frac{1}{2^{2N-1}}2^{2N-2} = \frac{1}{2}$$

So exactly half of the problems are linearly separable if $P = 2N$.