Machine Learning 2 - Homework 1

Pascal M. Esser / 11642858

April 10, 2018

Collaborators: Sindy Loewe, Alex Geenen, Gabriele Ce, Linda Petrini, Davide Belli

1

Let $\boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{z} \in \mathbb{R}^n$. This gives us:

$$\begin{aligned} p(\boldsymbol{x}) &= \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x) && (1) \\ &= (2\pi)^{-n/2}|\boldsymbol{\Sigma}_x|^{-1/2} exp\left[\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_x)^\top \boldsymbol{\Sigma}_x^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_x)\right] && (2) \\ p(\boldsymbol{z}) &= \mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z) && (3) \\ &= (2\pi)^{-n/2}|\boldsymbol{\Sigma}_z|^{-1/2} exp\left[\frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu}_z)^\top \boldsymbol{\Sigma}_z^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_z)\right] && (4) \end{aligned}$$

Now we have: $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{z}$.

Mean

$$\begin{aligned} \mathbb{E}[\boldsymbol{y}] &= \mathbb{E}[\boldsymbol{x}+\boldsymbol{z}] && (5) \\ &= \mathbb{E}[\boldsymbol{x}] + \mathbb{E}[\boldsymbol{z}] && (6) \\ &= \boldsymbol{\mu}_x + \boldsymbol{\mu}_z && (7) \end{aligned}$$

Cov

$$\begin{aligned} \boldsymbol{\Sigma}_y &= \mathbb{E}[(\boldsymbol{y} - \mathbb{E}[\boldsymbol{y}])^2] && (8) \\ &= \mathbb{E}[(\boldsymbol{x}+\boldsymbol{z} - \mathbb{E}[\boldsymbol{x}+\boldsymbol{z}])^2] && (9) \\ &= \mathbb{E}[(\boldsymbol{x}+\boldsymbol{z} - \boldsymbol{\mu}_x + \boldsymbol{\mu}_z)^2] && (10) \\ &= \mathbb{E}[\boldsymbol{x}^2 + \boldsymbol{z}^2 + 2\boldsymbol{x}\boldsymbol{z} - 2\boldsymbol{x}\boldsymbol{\mu}_x - 2\boldsymbol{x}\boldsymbol{\mu}_z - 2\boldsymbol{z}\boldsymbol{\mu}_x - 2\boldsymbol{z}\boldsymbol{\mu}_z + \boldsymbol{\mu}_x^2 + \boldsymbol{\mu}_z^2 + 2\boldsymbol{\mu}_x\boldsymbol{\mu}_z] && (11) \\ &= \mathbb{E}[(\boldsymbol{x}-\boldsymbol{\mu}_x)^2(\boldsymbol{z}-\boldsymbol{\mu}_z)^2 + 2(\boldsymbol{z}-\boldsymbol{\mu}_z)(\boldsymbol{x}-\boldsymbol{\mu}_x)] && (12) \\ &= \boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_z + 2\mathbb{E}[(\boldsymbol{z}-\boldsymbol{\mu}_z)(\boldsymbol{x}-\boldsymbol{\mu}_x)] && (13) \\ &= \boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_z + 2Cov[\boldsymbol{x}, \boldsymbol{z}] && (14) \end{aligned}$$

assuming $x \perp\!\!\!\perp z$ it follows that $Cov[\boldsymbol{x}, \boldsymbol{z}] = 0$ and therefore $\boldsymbol{\Sigma}_y = \boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_z$.

2

2.1

$$p(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \prod_{n=1}^{N} \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu},\boldsymbol{\Sigma}) \quad = \prod_{n=1}^{N} (2\pi)^{-D/2}|\Sigma|^{-1/2} exp\left[-\frac{1}{2}(\boldsymbol{x}_n-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_n-\boldsymbol{\mu})\right] \tag{15}$$

2.2

$$\begin{aligned}
p(\mu|\boldsymbol{\chi},\boldsymbol{\Sigma},\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0) &= \frac{p(\boldsymbol{\chi}|\boldsymbol{\mu},\boldsymbol{\Sigma})p(\boldsymbol{\mu}|\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)}{p(\boldsymbol{\chi}|\boldsymbol{\Sigma},\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)} && (16)\\
&= \frac{p(\boldsymbol{\chi}|\boldsymbol{\mu},\boldsymbol{\Sigma})p(\boldsymbol{\mu}|\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)}{\int p(\boldsymbol{\chi}|\boldsymbol{\mu}',\boldsymbol{\Sigma})p(\boldsymbol{\mu}'|\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)d\boldsymbol{\mu}'} && (17)\\
&\text{where } \int p(\boldsymbol{\chi}|\boldsymbol{\mu}',\boldsymbol{\Sigma})p(\boldsymbol{\mu}'|\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)d\boldsymbol{\mu}' \text{ is independent of } \boldsymbol{\mu} && (18)\\
&\propto p(\boldsymbol{\chi}|\boldsymbol{\mu},\boldsymbol{\Sigma})p(\boldsymbol{\mu}|\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0) && (19)\\
&= \prod_{n=1}^{N} (2\pi)^{-D/2}|\Sigma|^{-1/2} exp\left[-\frac{1}{2}(\boldsymbol{x}_n-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_n-\boldsymbol{\mu})\right] && (20)\\
&* (2\pi)^{-D/2}|\Sigma|^{-1/2} exp\left[-\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)\right] && (21)
\end{aligned}$$

2.3

starting from the results from 2.2:

$$\begin{aligned}
p(\mu|\boldsymbol{\chi},\boldsymbol{\Sigma},\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0) &= \prod_{n=1}^{N} (2\pi)^{-D/2}|\Sigma|^{-1/2} exp\left[-\frac{1}{2}(\boldsymbol{x}_n-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_n-\boldsymbol{\mu})\right] && (22)\\
&* (2\pi)^{-D/2}|\Sigma|^{-1/2} exp\left[-\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)\right] && (23)\\
&= (2\pi)^{-(ND-D)/2} * |\Sigma^N \Sigma_0|^{-1/2} exp\left[-\frac{1}{2}\left((\boldsymbol{\mu}-\boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0) + (\boldsymbol{x}_n-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_n-\boldsymbol{\mu})\right)\right] && (24)\\
&\text{set } C_1 \equiv (2\pi)^{-(ND-D)/2} && (25)\\
&\text{set } \boldsymbol{K}_1 \equiv -\frac{1}{2}\left((\boldsymbol{\mu}-\boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0) + (\boldsymbol{x}_n-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_n-\boldsymbol{\mu})\right) && (26)\\
p(\mu|\boldsymbol{\chi},\boldsymbol{\Sigma},\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0) &= C_1|\Sigma^N \Sigma_0|^{-1/2} exp\left[\boldsymbol{K}_1\right] && (27)\\
& && (28)
\end{aligned}$$

We can now expend and rewrite $\boldsymbol{K}_1$ under the use of $\boldsymbol{a}_1^\top \boldsymbol{\Sigma} \boldsymbol{a}_2 = \boldsymbol{a}_2^\top \boldsymbol{\Sigma} \boldsymbol{a}_1$ as $\boldsymbol{\Sigma}$ is symmetric.

$$\boldsymbol{K}_1 = -\frac{1}{2}\left((\boldsymbol{\mu}-\boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0) + (\boldsymbol{x}_n-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_n-\boldsymbol{\mu})\right) \tag{29}$$

$$= -\frac{1}{2}\left(\boldsymbol{\mu}^\top \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}^\top \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 - \boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu} + \boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 + \sum_{n=1}^{N}\left(\boldsymbol{x}_n^\top \Sigma^{-1}\boldsymbol{x}_n - \boldsymbol{x}_n^\top \Sigma^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}^\top \Sigma^{-1}\boldsymbol{x}_n + \boldsymbol{\mu}^\top \Sigma^{-1}\boldsymbol{\mu}\right)\right) \tag{30}$$

$$= -\frac{1}{2}\left(\boldsymbol{\mu}^\top \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu} - 2\boldsymbol{\mu}^\top \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 + \boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 + \sum_{n=1}^{N}\left(\boldsymbol{x}_n^\top \Sigma^{-1}\boldsymbol{x}_n - 2\boldsymbol{x}_n^\top \Sigma^{-1}\boldsymbol{\mu} + \boldsymbol{\mu}^\top \Sigma^{-1}\boldsymbol{\mu}\right)\right) \tag{31}$$

$$= -\frac{1}{2}\left(\boldsymbol{\mu}^\top \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu} - 2\boldsymbol{\mu}^\top \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 + \boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 + \sum_{n=1}^{N}\boldsymbol{x}_n^\top \Sigma^{-1}\boldsymbol{x}_n - 2\sum_{n=1}^{N}\boldsymbol{x}_n^\top \Sigma^{-1}\boldsymbol{\mu} + \boldsymbol{\mu}^\top \Sigma^{-1}\boldsymbol{\mu}\right) \tag{32}$$

$$\text{set: } \boldsymbol{C}_2 = \sum_{n=1}^{N}\boldsymbol{x}_n^\top \Sigma^{-1}\boldsymbol{x}_n + \boldsymbol{\mu}_0^\top \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 \tag{33}$$

$$= -\frac{1}{2}\left(\boldsymbol{C}_2 - 2\mu^\top\left(\boldsymbol{\Sigma}^{-1}\sum_{n=1}^{N}\boldsymbol{x}_n + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0\right) + \boldsymbol{\mu}^\top\left(N\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}_0^{-1}\right)\boldsymbol{\mu}\right) \tag{34}$$

With this we can now write:

$$p(\mu|\boldsymbol{\chi},\boldsymbol{\Sigma},\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0) = C_1|\Sigma^N\Sigma_0|^{-1/2}exp\left[-\frac{1}{2}\left(\boldsymbol{C}_2 - 2\boldsymbol{\mu}^\top\left(\boldsymbol{\Sigma}^{-1}\sum_{n=1}^{N}\boldsymbol{x}_n + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0\right) + \boldsymbol{\mu}^\top\left(N\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}_0^{-1}\right)\boldsymbol{\mu}\right)\right] \tag{35}$$

$$\propto |\Sigma^N\Sigma_0|^{-1/2}exp\left[-2\boldsymbol{\mu}^\top \boldsymbol{K}_2 + \boldsymbol{\mu}^\top \boldsymbol{K}_3\boldsymbol{\mu}\right] \tag{36}$$

$$\text{with } \boldsymbol{K}_2 = \boldsymbol{\Sigma}^{-1}\sum_{n=1}^{N}\boldsymbol{x}_n + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 \tag{37}$$

$$\text{with } \boldsymbol{K}_3 = N\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}_0^{-1} \tag{38}$$

This gives us the proof that $p(\mu|\boldsymbol{\chi},\boldsymbol{\Sigma},\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)$ is in a Gaussian as the form displayed above consists of a constant term (C_1) a variance therm ($|\Sigma^N\Sigma_0|^{-1/2}$) and in the exponent again a constant term in addition to a term linear in μ: $\boldsymbol{\Sigma}^{-1}\sum_{n=1}^{N}\boldsymbol{x}_n + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0$ and a term quadratic in μ: $N\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}_0^{-1}$.

In the next step we calculate mean and variance. Let $\boldsymbol{K}_4 = -2\boldsymbol{\mu}^\top \boldsymbol{K}_2 + \boldsymbol{\mu}^\top \boldsymbol{K}_3\boldsymbol{\mu}$

$$\begin{aligned}
\boldsymbol{K}_4 =& -2\boldsymbol{\mu}^\top \boldsymbol{K}_2 + \boldsymbol{\mu}^\top \boldsymbol{K}_3\boldsymbol{\mu} && (39)\\
=& -\boldsymbol{\mu}^\top \boldsymbol{K}_2 - \boldsymbol{\mu}^\top \boldsymbol{K}_2 + \boldsymbol{\mu}^\top \boldsymbol{K}_3\boldsymbol{\mu} && (40)\\
=& -\boldsymbol{\mu}^\top \boldsymbol{K}_3(\boldsymbol{\mu} - \boldsymbol{K}_3^{-1}\boldsymbol{K}_2) - \boldsymbol{\mu} && (41)\\
=& \boldsymbol{\mu}^\top \boldsymbol{K}_3(\boldsymbol{\mu} - \boldsymbol{K}_3^{-1}\boldsymbol{K}_2) - \boldsymbol{\mu}^\top \boldsymbol{K}_2 && (42)\\
=& \boldsymbol{\mu}^\top \boldsymbol{K}_3(\boldsymbol{\mu} - \boldsymbol{K}_3^{-1}\boldsymbol{K}_2) - \boldsymbol{K}_2^\top \boldsymbol{\mu} && (43)\\
& \text{add } -\boldsymbol{K}_2^\top \boldsymbol{K}_3^{-1}\boldsymbol{K}_2 + \boldsymbol{K}_2^\top \boldsymbol{K}_3^{-1}\boldsymbol{K}_2 && (44)\\
=& \boldsymbol{\mu}^\top \boldsymbol{K}_3(\boldsymbol{\mu} - \boldsymbol{K}_3^{-1}\boldsymbol{K}_2) - \boldsymbol{K}_2^\top \boldsymbol{K}_3^{-1}\boldsymbol{K}_2 + \boldsymbol{K}_2^\top \boldsymbol{K}_3^{-1}\boldsymbol{K}_2 - \boldsymbol{K}_2^\top \boldsymbol{\mu} && (45)\\
=& \boldsymbol{\mu}^\top \boldsymbol{K}_3(\boldsymbol{\mu} - \boldsymbol{K}_3^{-1}\boldsymbol{K}_2) - \boldsymbol{K}_2^\top \boldsymbol{K}_3^{-1}\boldsymbol{K}_2 - \boldsymbol{K}_2^\top (\boldsymbol{K}_3^{-1}\boldsymbol{K}_2 - \boldsymbol{\mu}) && (46)\\
& \text{add } \boldsymbol{K}_3\boldsymbol{K}_3^{-1} \text{ and omit } -\boldsymbol{K}_2^\top \boldsymbol{K}_3^{-1}\boldsymbol{K}_2 \text{ as a constant} && (47)\\
\propto& \boldsymbol{\mu}^\top \boldsymbol{K}_3(\boldsymbol{\mu} - \boldsymbol{K}_3^{-1}\boldsymbol{K}_2) - \boldsymbol{K}_2^\top \boldsymbol{K}_3\boldsymbol{K}_3^{-1}(\boldsymbol{K}_3^{-1}\boldsymbol{K}_2 - \boldsymbol{\mu}) && (48)\\
=& \boldsymbol{\mu}^\top \boldsymbol{K}_3\boldsymbol{\mu} - \boldsymbol{\mu}^\top \boldsymbol{K}_3\boldsymbol{K}_3^{-1}\boldsymbol{K}_2 - \boldsymbol{K}_2^\top \boldsymbol{K}_3^{-1}\boldsymbol{K}_3\boldsymbol{K}_3^{-1} + \boldsymbol{K}_2^\top \boldsymbol{K}_3^{-1}\boldsymbol{K}_3\boldsymbol{\mu} && (49)\\
=& (\boldsymbol{\mu}^\top - \boldsymbol{K}_2^\top \boldsymbol{K}_3^{-1})\boldsymbol{K}_3(\boldsymbol{\mu}^\top - \boldsymbol{K}_3^{-1}\boldsymbol{K}_2) && (50)\\
=& (\boldsymbol{\mu}^\top - \boldsymbol{K}_3^{-1}\boldsymbol{K}_2)\boldsymbol{K}_3(\boldsymbol{\mu}^\top - \boldsymbol{K}_3^{-1}\boldsymbol{K}_2) && (51)
\end{aligned}$$

which gives us by substituting back:

$$\boldsymbol{\Sigma}_N = \boldsymbol{K}_3^{-1} = (N\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}_0^{-1})^{-1} \tag{52}$$

$$\boldsymbol{\mu}_N = \boldsymbol{K}_3^{-1}\boldsymbol{K}_2 = \boldsymbol{\Sigma}_N(\boldsymbol{\Sigma}^{-1}\sum_{N}^{n=1}\boldsymbol{x}_n + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0) \tag{53}$$

2.4

$$\begin{aligned}
\log p(\mu|\boldsymbol{\chi}, \boldsymbol{\Sigma}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) &\propto \log exp\left[-\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_N)^\top \boldsymbol{\Sigma}_N^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_N)\right] && (54)\\
&\propto \boldsymbol{\mu}^\top \boldsymbol{\Sigma}_N^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}^\top \boldsymbol{\Sigma}_N^{-1}\boldsymbol{\mu}_N + \boldsymbol{\mu}_N^\top \boldsymbol{\Sigma}_N^{-1}\boldsymbol{\mu}_N && (55)\\
\frac{\partial \log p(\mu|\boldsymbol{\chi}, \boldsymbol{\Sigma}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)}{\partial \boldsymbol{\mu}} &= \boldsymbol{\Sigma}_N^{-1}\boldsymbol{\mu} - \boldsymbol{\Sigma}_N^{-1}\boldsymbol{\mu}_n && (56)\\
&= 0 && (57)\\
&\Rightarrow \boldsymbol{\Sigma}_N^{-1}\boldsymbol{\mu} = \boldsymbol{\Sigma}_N^{-1}\boldsymbol{\mu}_n && (58)\\
&\Rightarrow \boldsymbol{\mu} = \boldsymbol{\mu}_N && (59)
\end{aligned}$$

3

3.1

Setting:

$$m = \sum_{n=1}^{N} x_n \tag{60}$$

$$l = \sum_{n=1}^{N} 1 - x_n \tag{61}$$

$$\chi = (x_1, x_2, x_3)^\top \tag{62}$$

Calculating the MLE for Bernoulli:

$$\mathcal{L}(\chi|\mu, m, l) = \prod^{N} \mu^{x_n}(1 - x_n)^{1-x_n} \tag{63}$$

$$\log(\mathcal{L}(\chi|\mu, m, l)) = \sum^{N} x_n \log \mu - (1 - x_n)\log(1 - \mu) \tag{64}$$

$$\frac{\partial \log(\mathcal{L}(\chi|\mu, m, l))}{\partial \mu} = \frac{1}{\mu}\sum x_n - \frac{1}{1-\mu} = 0 \tag{65}$$

$$\Rightarrow \frac{m}{\mu} = \frac{N - m}{1 - \mu} \tag{66}$$

$$\Rightarrow \mu_{MLE} = \frac{m}{m+l} = \frac{1}{n}\sum x_n \tag{67}$$

for the given coin toss setting this give $\mu_{MLE} = 3/(3 + 0) = 1$.

3.2

$$p(\mu|\chi, a, b) = \frac{p(\chi|\mu)p(\mu|a, b)}{\int p(\chi|\mu')p(\mu'|a, b)d\mu'} \tag{68}$$

$$\propto p(\chi|\mu)p(\mu|a, b) \tag{69}$$

$$\propto \mu^{m+a-1}(1 - \mu)^{l+b-1} \tag{70}$$

$$\log p(\mu|\chi, a, b) = (m + a - 1)\log \mu - (l + b - 1)\log(1 - \mu) \tag{71}$$

$$\frac{\partial \log p(\mu|\chi, a, b)}{\partial \mu} = \frac{m + a - 1}{\mu} - \frac{l + b - 1}{1 - \mu} \tag{72}$$

$$= 0 \tag{73}$$

$$\Rightarrow \mu_{MAP} = \frac{m + a - 1}{m + l + a + b - 2} \tag{74}$$

$$p(x = 1|\chi) = \int_0^1 p(x = 1|\mu)p(\mu|\chi)d\mu \tag{75}$$

$$= \mathbb{E}[\mu|\chi] \tag{76}$$

$$\text{with } \mathbb{E}[\mu] = \frac{a}{a+b} \tag{77}$$

$$p(x = 1|\chi) = \frac{m + a}{m + a + l + b} \tag{78}$$

for the given coin toss setting this give $p(x = 1|\chi) = \frac{3+a}{3+a+b}$

3.3

As a starting point we use the results from 3.1 and 3.2. We can now rewrite $p(x = 1|\chi)$ as

$$
\begin{aligned}
p(x = 1|\chi) &= \frac{m + a}{m + a + l + b} && (79)\\
&= \frac{m}{m + a + l + b} + \frac{a}{m + a + l + b} && (80)\\
&= \frac{m + l}{m + a + l + b}\left(\frac{m}{m + l}\right) + \frac{a + b}{m + a + l + b}\left(\frac{a}{a + b}\right) && (81)\\
& && (82)
\end{aligned}
$$

with $\frac{m+l}{m+a+l+b} \equiv 1 - \lambda$ and $\frac{a+b}{m+a+l+b} \equiv (\lambda)$ we get the decomposition we wanted to proof: $\lambda\mathbb{E}[\mu|\chi] + (1 - \lambda)\mu_{MLE} = p(x = 1|\chi)$. (with $\lambda \in [0, 1]$)

4

4.1

$$Pois(k|\lambda) = \frac{\lambda^k e^{0\lambda}}{k!} \tag{83}$$
$$= \frac{1}{k!} \exp\left[k \log \lambda - \lambda\right] \tag{84}$$
$$h(k) = \frac{1}{k!} \tag{85}$$
$$\eta = \log \lambda \tag{86}$$
$$g(\eta) = \exp(-e^{\eta}) \tag{87}$$
$$u(K) = k \tag{88}$$

$$Gam(\tau|a,b) = \frac{1}{\Gamma(a)} b^a \tau^{a-1} e^{-b\tau} \tag{89}$$
$$= \frac{b^a}{\Gamma(a)} \tau^{-1} \exp\left[a \log \tau - b\tau\right] \tag{90}$$
$$h(\tau) = \frac{1}{\tau} \tag{91}$$
$$\eta = (a,b)^\top \tag{92}$$
$$g(\eta) = \frac{\eta_2^{\eta_1}}{\Gamma(\eta_1)} \tag{93}$$
$$u(\eta) = (\log \tau, -\tau)^\top \tag{94}$$

The Cauchy distribution can not be but into an exponential family form as the sum $(1 + \left(\frac{x-\mu}{\gamma}\right)^2)$ cannot be factorized as it would be necessary.

$$vonMise(x|k,\mu) = \frac{1}{2\pi I_0(k)} e^{k\cos(x-\mu)} \tag{95}$$
$$= \frac{1}{2\pi I_0(k)} e^{k(\cos x \cos \mu + \sin x \sin \mu)} \tag{96}$$
$$= \frac{1}{2\pi I_0(k)} e^{(\cos x, \sin x)(k\cos\mu, k\sin\mu)^\top} \tag{97}$$
$$h(x) = 1 \tag{98}$$
$$\eta = (k\cos\mu, k\sin\mu)^\top \tag{99}$$
$$\Rightarrow k = \sqrt{\eta_1^2 + \eta_2^2} \tag{100}$$
$$g(\eta) = \frac{1}{2\pi I_0\left(\sqrt{\eta_1^2+\eta_2^2}\right)} \tag{101}$$

4.2

For the first momentum of the Poisson we get

$$\mathbb{E}[k] = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} \tag{102}$$
$$= \lambda e^{\lambda(e^0-1)+0} \tag{103}$$
$$= \lambda \tag{104}$$

Calculating the second momentum:

$$\mathbb{E}[k^2] = \sum_{k=0}^{\infty} 2k e^{-\lambda} \frac{\lambda^k}{k!} \tag{105}$$
$$= \lambda \sum_{k=0}^{\infty} (k) e^{-\lambda} \frac{\lambda^{k-1}}{(k-1)!} \tag{106}$$
$$= \lambda \sum_{k=0}^{\infty} (k+1) e^{-\lambda} \frac{\lambda^k}{(k)!} \tag{107}$$
$$= \mathbb{E}[k+1] \tag{108}$$
$$= \lambda(\mathbb{E}[k]+1) \tag{109}$$
$$= \lambda^2 + \lambda \tag{110}$$

this gives us a mean of $\mathbb{E}[k] = \lambda$ and a variance of $\mathbb{E}[k^2] - [\mathbb{E}[k]]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$

First momentum of the gamma

$$\mathbb{E}[\tau] = \int_0^{\infty} \tau \frac{b^a}{\Gamma(a)} \tau^{a-1} e^{-b\tau} d\tau \tag{111}$$
$$= \frac{b^a}{\Gamma(a)} \int_0^{\infty} \tau^a e^{-b\tau} d\tau \tag{112}$$
$$= \frac{\Gamma(a+1)}{b\Gamma(a)} \int_0^{\infty} \frac{b^{a+1}}{\Gamma(a)} \tau^a e^{-b\tau} d\tau \tag{113}$$
$$= \frac{\Gamma(a+1)}{b\Gamma(a)} \tag{114}$$
$$= \frac{a\Gamma(a)}{b\Gamma(a)} \tag{115}$$
$$= \frac{a}{b} \tag{116}$$

Second momentum of the gamma:

$$\mathbb{E}[\tau^2] = \int_0^\infty \tau^2 \frac{b^a}{\Gamma(a)} \tau^{a-1} e^{-b\tau} d\tau \tag{117}$$
$$= \frac{b^a}{\Gamma(a)} \int_0^\infty \tau^{a+1} e^{-b\tau} d\tau \tag{118}$$
$$= \frac{\Gamma(a+2)}{b^2\Gamma(a)} \int_0^\infty \frac{b^{a+1}}{\Gamma(a+1)} \tau^{a+1} e^{-b\tau} d\tau \tag{119}$$
$$= \frac{\Gamma(a+2)}{b^2\Gamma(a)} \tag{120}$$
$$= \frac{(a+1)\Gamma(a+1)}{b\Gamma(a)} \tag{121}$$
$$= \frac{(a+1)!}{b^2(a-1)!} \tag{122}$$
$$= \frac{(a+1)a}{b^2} \tag{123}$$

this gives us a mean of $\mathbb{E}[\tau] = \frac{a}{b}$ and a variance of $\mathbb{E}[\tau^2] - [\mathbb{E}[\tau]]^2 = \frac{(a+1)a}{b^2} - \left(\frac{a}{b}\right)^2 = \frac{a}{b^2}$

4.3

we start from the exponential family form of the Poisson distribution:

$$p(k|\lambda) = exp[-\lambda]\frac{1}{k!}exp[k \log \lambda] \tag{124}$$

this has the prior form (equation taken from the lecture):

$$p(\lambda) \propto \exp[-\lambda]^a \exp[b \log \lambda] \tag{125}$$
$$\propto \lambda^b e^{-a\lambda} \tag{126}$$

This gives us the conjugate prior and we see it matches the core of the Gamma exponential family. More specifically: $\lambda \sim Gam(\alpha, \beta)$ with $\alpha = b + 1, \beta = a$.

5

Starting from Bishop eq 2.161 we get the mean as:

$$St(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Lambda},\boldsymbol{v}) = \int_0^\infty \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu},(\eta\boldsymbol{\Lambda})^{-1})Gam(\eta|v/2,v/2)d\eta \tag{127}$$

$$\mathbb{E}[x] = \int x \int_0^\infty \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu},(\eta\boldsymbol{\Lambda})^{-1})Gam(\eta|v/2,v/2)dxd\eta \tag{128}$$

$$= \int\int_0^\infty x\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu},(\eta\boldsymbol{\Lambda})^{-1})Gam(\eta|v/2,v/2)d\eta dx \tag{129}$$

$$= \int_0^\infty \boldsymbol{\mu}Gam(\eta|v/2,v/2)d\eta \tag{130}$$

$$= \boldsymbol{\mu}\int_0^\infty Gam(\eta|v/2,v/2)d\eta \tag{131}$$

$$= \boldsymbol{\mu} \tag{132}$$

for $v > 2$ we can calculate the covariance as:

$$Cov[x] = \mathbb{E}[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^\top] \tag{133}$$

$$= \int(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^\top \int_0^\infty \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu},(\eta\boldsymbol{\Lambda})^{-1})Gam(\eta|v/2,v/2)d\eta dx \tag{134}$$

$$= \int\int_0^\infty (\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^\top\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu},(\eta\boldsymbol{\Lambda})^{-1})Gam(\eta|v/2,v/2)dxd\eta \tag{135}$$

$$= \int_0^\infty (\eta\boldsymbol{\Lambda})^{-1}Gam(\eta|v/2,v/2)dxd\eta \tag{136}$$

$$= \Lambda^{-1}\int_0^\infty \frac{1}{\eta}\frac{(v/2)^{v/2}}{\Gamma(v/2)}\eta^{v/2-1}exp[-(v/2)\eta]d\eta \tag{137}$$

$$= \Lambda^{-1}\frac{v/2}{v/2-1}\int_0^\infty \frac{(v/2)^{v/2-1}}{\Gamma(v/2-1)}\eta^{v/2-2}exp[-(v/2)\eta]d\eta \tag{138}$$

$$= \Lambda^{-1}\frac{(v/2)}{v/2-1} \tag{139}$$

$$= \Lambda^{-1}\frac{v}{v-2} \tag{140}$$

$$\tag{141}$$

For the mode we look at bishop 2.162

$$St(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Lambda},\boldsymbol{v}) = \frac{\Gamma(D/2+v/2)}{\Gamma(v/2)}\frac{|\Lambda|^{1/2}}{(\pi v)^{D/2}}\left[1+\frac{\Delta^2}{v}\right]^{-D/2-v/2} \tag{142}$$

and as the probability density function is monotonically decreasing in Δ^2, with $\Delta^2 = (\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Lambda}(\boldsymbol{x}-\boldsymbol{\mu})$ we get $mode[x] = \boldsymbol{\mu}$ as the distance goes to zero.