SHRUTI RAO
2636454
s2.rao@student.vu.nl

MAP Solution with Correlated Responses

a

Assuming N training vectors and using the provided likelihood function of the whole dataset:

$$p(\mathbf{t}|\mathbf{\Psi}, \mathbf{w}, \mathbf{\Omega}) = \mathcal{N}(\mathbf{t}|\mathbf{\Psi}\mathbf{w}, \mathbf{\Omega})$$

The probability densify function for a multivariate Gaussian distribution is given by:

$$\frac{1}{2\pi^{N/2}|\Sigma|^{1/2}} exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$$

As the data samples are no longer i.i.d, the likelihood in matrix/vector form is:

$$\begin{aligned} p(\mathcal{D}|\boldsymbol{\theta}) &= \mathcal{N}(\mathbf{t}|\mathbf{\Psi}\mathbf{w}, \mathbf{\Omega}) \\ &= \frac{1}{2\pi^{N/2}|\Omega|^{1/2}} exp\left\{-\frac{1}{2}(\boldsymbol{t}-\mathbf{\Psi}\boldsymbol{w})^T\mathbf{\Omega}^{-1}(\boldsymbol{t}-\mathbf{\Psi}\boldsymbol{w})\right\} \end{aligned}$$

b

Let $\mathbf{\Omega} = \boldsymbol{A^T D A}$ where $\mathbf{D}$ is a diagonal matrix containing eigenvalues of $\mathbf{\Omega}$.
Let $\boldsymbol{A^T = A^{-1}}$. Then following the Spectral Theorem:

$$\begin{aligned} p(\mathcal{D}|\boldsymbol{\theta}) &= \frac{1}{2\pi^{N/2}|\Omega|^{1/2}} exp\left\{-\frac{1}{2}(\boldsymbol{t}-\mathbf{\Psi}\boldsymbol{w})^T\mathbf{\Omega}^{-1}(\boldsymbol{t}-\mathbf{\Psi}\boldsymbol{w})\right\} \\ &= \frac{1}{2\pi^{N/2}|\boldsymbol{A^T D A}|^{1/2}} exp\left\{-\frac{1}{2}(\boldsymbol{t}-\mathbf{\Psi}\boldsymbol{w})^T(\boldsymbol{A^T D A})^{-1}(\boldsymbol{t}-\mathbf{\Psi}\boldsymbol{w})\right\} \\ &= \frac{1}{2\pi^{N/2}|\boldsymbol{A^T D A}|^{1/2}} exp\left\{-\frac{1}{2}(\boldsymbol{t}-\mathbf{\Psi}\boldsymbol{w})^T\boldsymbol{(A^T)^{-1} D^{-1} A^{-1}}(\boldsymbol{t}-\mathbf{\Psi}\boldsymbol{w})\right\} \\ &= \frac{1}{2\pi^{N/2}|\boldsymbol{A^T D A}|^{1/2}} exp\left\{-\frac{1}{2}(\boldsymbol{t}-\mathbf{\Psi}\boldsymbol{w})^T\boldsymbol{(A^{-1})^{-1} D^{-1} A^{T}}(\boldsymbol{t}-\mathbf{\Psi}\boldsymbol{w})\right\} \\ &= \frac{1}{2\pi^{N/2}|\boldsymbol{A^T D A}|^{1/2}} exp\left\{-\frac{1}{2}(\boldsymbol{At}-\boldsymbol{A}\mathbf{\Psi}\boldsymbol{w})^T\boldsymbol{D^{-1}}(\boldsymbol{At}-\boldsymbol{A}\mathbf{\Psi}\boldsymbol{w})\right\} \\ &= \frac{1}{2\pi^{N/2}|\boldsymbol{A^{-1} D A}|^{1/2}} exp\left\{-\frac{1}{2}(\boldsymbol{At}-\boldsymbol{A}\mathbf{\Psi}\boldsymbol{w})^T\boldsymbol{D^{-1}}(\boldsymbol{At}-\boldsymbol{A}\mathbf{\Psi}\boldsymbol{w})\right\} \\ &= \frac{1}{2\pi^{N/2}|\boldsymbol{D}|^{1/2}} exp\left\{-\frac{1}{2}(\boldsymbol{At}-\boldsymbol{A}\mathbf{\Psi}\boldsymbol{w})^T\boldsymbol{D^{-1}}(\boldsymbol{At}-\boldsymbol{A}\mathbf{\Psi}\boldsymbol{w})\right\} \end{aligned}$$

Moreover, $\boldsymbol{\tau} := \boldsymbol{At}$ and $\mathbf{\Phi} := \boldsymbol{A}\mathbf{\Psi}$:

$$= \frac{1}{2\pi^{N/2}|\boldsymbol{D}|^{1/2}} exp\left\{-\frac{1}{2}(\boldsymbol{\tau}-\mathbf{\Phi}\boldsymbol{w})^T\boldsymbol{D^{-1}}(\boldsymbol{\tau}-\mathbf{\Phi}\boldsymbol{w})\right\}$$

c

Factorization of the distribution into a univariate Gaussian Distribution:

$$p(\mathcal{D}|\boldsymbol{\theta}) = \frac{1}{2\pi^{N/2}|\boldsymbol{D}|^{1/2}} exp\left\{-\frac{1}{2}(\boldsymbol{\tau} - \boldsymbol{\Phi w})^T \boldsymbol{D^{-1}}(\boldsymbol{\tau} - \boldsymbol{\Phi w})\right\}$$

D is a diagonal matrix and the determinant of D can be expressed as the product of its elements. Also, also we know that: $(\boldsymbol{\tau} - \boldsymbol{\Phi w})^T(\boldsymbol{\tau} - \boldsymbol{\Phi w}) = ||\boldsymbol{\tau - \Phi w}||^2 = \sum_{i=1}^{N}(\boldsymbol{\tau_i} - \boldsymbol{w^T \Phi_i})^2$

$$\begin{aligned}
&= \frac{1}{2\pi^{N/2}|\boldsymbol{D}|^{1/2}} exp\left\{-\frac{1}{2}\boldsymbol{D^{-1}}\sum_{i=1}^{N}(\boldsymbol{\tau_i} - \boldsymbol{w^T \Phi_i})^2\right\} \\
&= \frac{1}{2\pi^{N/2}|\boldsymbol{D}|^{1/2}} exp\left\{-\frac{\beta}{2}\sum_{i=1}^{N}(\boldsymbol{\tau_i} - \boldsymbol{w^T \Phi_i})^2\right\} \\
&= \frac{1}{2\pi^{N/2}|\beta|^{1/2}} exp\left\{-\frac{\beta}{2}\sum_{i=1}^{N}(\boldsymbol{\tau_i} - \boldsymbol{w^T \Phi_i})^2\right\} \\
&= \frac{1}{2\pi^{N/2}|\beta|^{1/2}} \prod_{i=1}^{N} exp\left\{-\frac{\beta}{2}(\boldsymbol{\tau_i} - \boldsymbol{w^T \Phi_i})^2\right\} \\
&= \prod_{i=1}^{N} \frac{\beta_i^{-1/2}}{2\pi^{1/2}} exp\left\{-\frac{\beta_i}{2}(\boldsymbol{\tau_i} - \boldsymbol{w^T \Phi_i})^2\right\} \\
&= \prod_{i=1}^{N} \mathcal{N}\left(\boldsymbol{\tau_i}|\boldsymbol{w^T \Phi_i}, \frac{1}{\beta}\right)
\end{aligned}$$

d

The prior over **w** is given by: $p(\boldsymbol{w}) = \mathcal{N}(\boldsymbol{w}|\boldsymbol{0}, \alpha \boldsymbol{I})$ where **0** is a vector of 0's and it's explicit form is:

$$\begin{aligned}
p(\boldsymbol{w}) &= \frac{1}{2\pi^{N/2}|\alpha|^{N/2}} exp\left\{-\frac{\alpha}{2}(\boldsymbol{w} - \boldsymbol{0})^T(\boldsymbol{w} - \boldsymbol{0})\right\} \\
&= \frac{-\alpha^{N/2}}{2\pi^{N/2}} exp\left\{-\frac{\alpha}{2}\boldsymbol{w}^T \boldsymbol{w}\right\}
\end{aligned}$$

Taking the logarithm on both sides:

$$\begin{aligned}
ln(p(\boldsymbol{w})) &= ln\left(\frac{-\alpha^{N/2}}{2\pi^{N/2}} exp\{-\frac{\alpha}{2}\boldsymbol{w}^T \boldsymbol{w}\}\right) \\
&= -\frac{N}{2}ln(\alpha) - \frac{N}{2}ln(2\pi) - \frac{\alpha}{2}\boldsymbol{w^T w} \\
&= -\frac{\alpha}{2}\boldsymbol{w^T w} + C \text{ where } \; C = -\frac{N}{2}ln(\alpha) - \frac{N}{2}ln(2\pi)
\end{aligned}$$

e

The posterior for $p(\boldsymbol{w}|\mathcal{D})$ over **w** is given by:

$$
\begin{aligned}
p(\boldsymbol{w}|\mathcal{D}) &= \frac{p(\mathcal{D}|\boldsymbol{w})p(\boldsymbol{w})}{p(\mathcal{D})} \\
&= \frac{\mathcal{N}(\mathbf{t}|\boldsymbol{\Psi}\mathbf{w},\boldsymbol{\Omega})\mathcal{N}(\boldsymbol{w}|\mathbf{0},\alpha\boldsymbol{I})}{\int \mathcal{N}(\mathbf{t}|\boldsymbol{\Psi}\mathbf{w},\boldsymbol{\Omega})\mathcal{N}(\boldsymbol{w}|\mathbf{0},\alpha\boldsymbol{I})dw} \\
&= \frac{\mathcal{N}(\mathbf{t}|\boldsymbol{\Psi}\mathbf{w},\boldsymbol{\Omega})\mathcal{N}(\boldsymbol{w}|\mathbf{0},\alpha\boldsymbol{I})}{p(\boldsymbol{t}|\boldsymbol{\Psi},\boldsymbol{\alpha},\boldsymbol{\Omega})}
\end{aligned}
$$

f

Using the formula for posterior from part (e):

$$
p(\boldsymbol{w}|\mathcal{D}) = \frac{p(\mathcal{D}|\boldsymbol{w})p(\boldsymbol{w})}{p(\mathcal{D})}
$$

As $p(\mathcal{D})$ or $p(\boldsymbol{t}|\boldsymbol{\Psi},\boldsymbol{\alpha},\boldsymbol{\Omega})$ does not depend on w, thus we treat it as a constant $\mathcal{I}$ with respect to **w**. Doing so makes it easier to compute MAP estimator rather than take the full posterior distribution. Taking the integral for $\mathcal{I}$ is usually very challenging and thus gets avoided in this manner.

Matrix Form:

$$
\begin{aligned}
p(\mathcal{D}|\boldsymbol{w}) &= \frac{1}{2\pi^{N/2}|\boldsymbol{D}|^{1/2}} exp\left\{-\frac{\beta}{2}(\boldsymbol{\tau}-\boldsymbol{\Phi w})^T(\boldsymbol{\tau}-\boldsymbol{\Phi w})\right\} \\
ln(p(\mathcal{D}|\boldsymbol{w})) &= ln\left(\frac{-|\boldsymbol{D}|^{1/2}}{2\pi^{N/2}} exp\left\{-\frac{\beta}{2}(\boldsymbol{\tau}-\boldsymbol{\Phi w})^T(\boldsymbol{\tau}-\boldsymbol{\Phi w})\right\}\right) \\
&= -\frac{1}{2}ln(\boldsymbol{D}) - \frac{N}{2}ln(2\pi) - \frac{\beta}{2}(\boldsymbol{\tau}-\boldsymbol{\Phi w})^T(\boldsymbol{\tau}-\boldsymbol{\Phi w}) \\
p(\boldsymbol{w}) &= \frac{-\alpha^{N/2}}{2\pi^{N/2}} exp\left\{-\frac{\alpha}{2}\boldsymbol{w}^T\boldsymbol{w}\right\} \\
ln(p(\boldsymbol{w})) &= ln\left(\frac{-\alpha^{N/2}}{2\pi^{N/2}} exp\left\{-\frac{\alpha}{2}\boldsymbol{w}^T\boldsymbol{w}\right\}\right) \\
&= -\frac{N}{2}ln(\alpha) - \frac{N}{2}ln(2\pi) - \frac{\alpha}{2}(\boldsymbol{w^T w})
\end{aligned}
$$

Applying the results from above we get the log-posterior for the matrix form:

$$
\begin{aligned}
ln(p(\boldsymbol{w}|\mathcal{D})) &= ln(p(\mathcal{D}|\boldsymbol{w})) + ln(p(\boldsymbol{w})) - ln(I) \\
&= -\frac{1}{2}ln(\boldsymbol{D}) - \frac{N}{2}ln(2\pi) - \frac{\beta}{2}(\boldsymbol{\tau}-\boldsymbol{\Phi w})^T(\boldsymbol{\tau}-\boldsymbol{\Phi w}) - \frac{N}{2}ln(\alpha) - \frac{N}{2}ln(2\pi) - \frac{\alpha}{2}(\boldsymbol{w^T w}) - ln(\mathcal{I}) \\
&= -\frac{\alpha}{2}(\boldsymbol{w^T w}) - \frac{\beta}{2}(\boldsymbol{\tau}-\boldsymbol{\Phi w})^T(\boldsymbol{\tau}-\boldsymbol{\Phi w}) + C \\
\text{where } C &= -\frac{1}{2}ln(\boldsymbol{D}) - \frac{N}{2}ln(2\pi) - \frac{N}{2}ln(\alpha) - \frac{N}{2}ln(2\pi) - ln(\mathcal{I})
\end{aligned}
$$

Factored Form:

$$
\begin{aligned}
p(\mathcal{D}|\boldsymbol{w}) &= \frac{1}{2\pi^{N/2}|\boldsymbol{D}|^{1/2}} exp\left\{-\frac{\beta}{2}\sum_{i=1}^{N}(\boldsymbol{\tau_i} - \boldsymbol{w^T\Phi_i})^2\right\} \\
ln(p(\mathcal{D}|\boldsymbol{w})) &= ln\left(\frac{1}{2\pi^{N/2}|\boldsymbol{D}|^{1/2}} exp\left\{-\frac{\beta}{2}\sum_{i=1}^{N}(\boldsymbol{\tau_i} - \boldsymbol{w^T\Phi_i})^2\right\}\right) \\
&= -\frac{1}{2}ln(\boldsymbol{D}) - \frac{N}{2}ln(2\pi) - \frac{\beta}{2}\sum_{i=1}^{N}(\boldsymbol{\tau_i} - \boldsymbol{w^T\Phi_i})^2 \\
p(\boldsymbol{w}) &= \frac{-\alpha^{N/2}}{2\pi^{N/2}} exp\left\{-\frac{\alpha}{2}\sum_{i=0}^{N-1} w_i^2\right\} \\
ln(p(\boldsymbol{w})) &= ln\left(\frac{-\alpha^{N/2}}{2\pi^{N/2}} exp\left\{-\frac{\alpha}{2}\sum_{i=0}^{N-1} w_i^2\right\}\right) \\
&= -\frac{N}{2}ln(\alpha) - \frac{N}{2}ln(2\pi) - \frac{\alpha}{2}\sum_{i=0}^{N-1} w_i^2
\end{aligned}
$$

Applying the results from above, we get the log-posterior for the matrix form for **w**:

$$
\begin{aligned}
&= -\frac{1}{2}ln(\boldsymbol{D}) - \frac{N}{2}ln(2\pi) - \frac{\beta}{2}\sum_{i=1}^{N}(\boldsymbol{\tau_i} - \boldsymbol{w^T\Phi_i})^2 - \frac{N}{2}ln(\alpha) - \frac{N}{2}ln(2\pi) - \frac{\alpha}{2}\sum_{i=0}^{N-1} w_i^2 - ln(\mathcal{I}) \\
&= -\frac{\beta}{2}\sum_{i=1}^{N}(\boldsymbol{\tau_i} - \boldsymbol{w^T\Phi_i})^2 - \frac{\alpha}{2}\sum_{i=0}^{N-1} w_i^2 + C \\
&\text{where } C = -\frac{1}{2}ln(\boldsymbol{D}) - \frac{N}{2}ln(2\pi) - -\frac{N}{2}ln(\alpha) - \frac{N}{2}ln(2\pi) - ln(\mathcal{I})
\end{aligned}
$$

g

We use the log-posterior of the matrix form to solve for the derivative and obtain $\mathbf{w}_{MAP}$

$$
\begin{aligned}
ln(p(\boldsymbol{w}|\mathcal{D})) &= -\frac{\alpha}{2}(\boldsymbol{w^T w}) - \frac{\beta}{2}(\boldsymbol{\tau} - \boldsymbol{\Phi w})^T(\boldsymbol{\tau} - \boldsymbol{\Phi w}) + C \\
\frac{\partial ln(p(\boldsymbol{w}|\mathcal{D}))}{\partial \boldsymbol{w}} &= -\alpha\boldsymbol{w}^T + \beta(\boldsymbol{\tau} - \boldsymbol{\Phi w})^T\boldsymbol{\Phi} = 0 \\
0 &= -\alpha\boldsymbol{w}^T + \beta(\boldsymbol{\tau}^T - \boldsymbol{w^T\Phi^T})\boldsymbol{\Phi} \\
0 &= -\alpha\boldsymbol{w}^T + \beta(\boldsymbol{\tau}^T\boldsymbol{\Phi} - \boldsymbol{w^T\Phi^T\Phi}) \\
0 &= -\alpha\boldsymbol{w}^T + \beta\boldsymbol{\tau}^T\boldsymbol{\Phi} - \beta\boldsymbol{w^T\Phi^T\Phi}
\end{aligned}
$$

Gathering all the w terms together:

$$
\begin{aligned}
\alpha\boldsymbol{w}^T + \beta\boldsymbol{w^T\Phi^T\Phi} &= \frac{\beta}{2}\boldsymbol{\tau}^T\boldsymbol{\Phi} \\
\boldsymbol{w}^T(\alpha\mathcal{I} + \beta\boldsymbol{\Phi^T\Phi}) &= \beta\boldsymbol{\tau}^T\boldsymbol{\Phi}
\end{aligned}
$$

Taking the transpose on both sides:

$$
\begin{aligned}
(\alpha\mathcal{I} + \beta\boldsymbol{\Phi^T\Phi})\boldsymbol{w} &= \beta\boldsymbol{\Phi}^T\boldsymbol{\tau} \\
\boldsymbol{w_{MAP}} &= (\alpha\mathcal{I} + \beta\boldsymbol{\Phi^T\Phi})^{-1}\beta\boldsymbol{\Phi}^T\boldsymbol{\tau}
\end{aligned}
$$

h

We can substitute $\boldsymbol{\tau} := \boldsymbol{At}$ and $\boldsymbol{\Phi} := \boldsymbol{A\Psi}$ into the equation for $\boldsymbol{w_{MAP}}$:

$$
\begin{aligned}
\boldsymbol{w_{MAP}} &= (\alpha\mathcal{I} + \beta\boldsymbol{\Phi^T\Phi})^{-1}\beta\boldsymbol{\Phi}^T\boldsymbol{\tau} \\
&= (\alpha\mathcal{I} + \beta\boldsymbol{(A\Psi)^T A\Psi})^{-1}\beta\boldsymbol{A\Psi}^T\boldsymbol{At} \\
&= (\alpha\mathcal{I} + \beta\boldsymbol{\Psi^T A^T A\Psi})^{-1}\beta\boldsymbol{A\Psi}^T\boldsymbol{At}
\end{aligned}
$$

Since we had also substituted D^{-1} with β, we can add D^{-1} back

$$
\begin{aligned}
&= (\alpha\mathcal{I} + D^{-1}\boldsymbol{\Psi^T A^T A\Psi})^{-1}D^{-1}\boldsymbol{A\Psi}^T\boldsymbol{At} \\
&= (\alpha\mathcal{I} + \boldsymbol{\Psi^T\Omega^{-1}\Psi})^{-1}\boldsymbol{\Psi}^T\boldsymbol{\Omega^{-1}t}
\end{aligned}
$$

ML Estimate of Angle Measurement

We are given that *s, c* are measurements for sine and cosine respectively.
The standard deviation is given to be σ Let $\cos\theta$ and $\sin\theta$ be the mean for *c* and *s* respectively.
Using univariate Gaussian, the following can be modelled:

$$
\begin{aligned}
p(D|\theta) = \mathcal{N}(x|\mu,\sigma) &= \prod_{i=1}^{N}\mathcal{N}(x_i|\theta_i,\sigma) \text{ where } \; x_i \in \{c_i, s_i\}; \theta_i \in \{\cos\theta, \sin\theta\} \\
&= \prod_{i=1}^{N}\frac{1}{\sqrt{2\pi\sigma}}exp\left\{-\frac{1}{2\sigma}(c_i-\cos\theta)^2\right\}\frac{1}{\sqrt{2\pi\sigma}}exp\left\{-\frac{1}{2\sigma}(s_i-\sin\theta)^2\right\} \\
&= \left[\frac{1}{(2\pi\sigma)^{N/2}}exp\left\{-\frac{1}{2\sigma}\sum_{i=1}^{N}(c_i-\cos\theta)^2\right\}\right]\left[\frac{1}{(2\pi\sigma)^{N/2}}exp\left\{-\frac{1}{2\sigma}\sum_{i=1}^{N}(s_i-\sin\theta)^2\right\}\right]
\end{aligned}
$$

Taking the logarithm on both sides and gathering irrelevant constants into C, where $C = -\frac{N}{2}\ln(2\pi\sigma) - \frac{N}{2}\ln(2\pi\sigma)$:

$$
\begin{aligned}
\ln(p(D|\theta)) &= -\frac{N}{2}\ln(2\pi\sigma) - \frac{1}{2\sigma}\sum_{i=1}^{N}(c_i-\cos\theta)^2 - \frac{N}{2}\ln(2\pi\sigma) - \frac{1}{2\sigma}\sum_{i=1}^{N}(s_i-\sin\theta)^2 \\
&= -\frac{1}{2\sigma}\sum_{i=1}^{N}(c_i-\cos\theta)^2 - \frac{1}{2\sigma}\sum_{i=1}^{N}(s_i-\sin\theta)^2 + C
\end{aligned}
$$

Partially differentiating with respect to θ:

$$
\begin{aligned}
\frac{\partial\ln p(D|\theta)}{\partial\theta} &= \frac{\partial(-\frac{1}{2\sigma}\sum_{i=1}^{N}(c_i-\cos\theta)^2 - \frac{1}{2\sigma}\sum_{i=1}^{N}(s_i-\sin\theta)^2 + C)}{\partial\theta} \\
&= -\frac{1}{2\sigma}\sum_{i=1}^{N}(c_i-\cos\theta)(\sin\theta)(2) - \frac{1}{2\sigma}\sum_{i=1}^{N}(s_i-\sin\theta)(\cos\theta)(2) \\
&= -\frac{1}{\sigma}\sum_{i=1}^{N}(c_i\sin\theta-\cos\theta\sin\theta) - \frac{1}{\sigma}\sum_{i=1}^{N}(s_i\cos\theta-\sin\theta\cos\theta) \\
&= -\frac{1}{\sigma}\left[\sum_{i=1}^{N}(c_i\sin\theta-\cos\theta\sin\theta) - \sum_{i=1}^{N}(s_i\cos\theta-\sin\theta\cos\theta)\right] \\
&= -\frac{1}{\sigma}\left[\sum_{i=1}^{N}(c_i\sin\theta-\cos\theta\sin\theta-s_i\cos\theta+\sin\theta\cos\theta)\right] \\
&= -\frac{1}{\sigma}\left[\sum_{i=1}^{N}(c_i\sin\theta-s_i\cos\theta)\right]
\end{aligned}
$$

Setting the equation to 0 and solving for θ:

$$
\begin{aligned}
-\frac{1}{\sigma}\sum_{i=1}^{N}(c_i \sin\theta - s_i \cos\theta) &= 0 \\
-\sum_{i=1}^{N}(c_i \sin\theta - s_i \cos\theta) &= 0 \\
\sum_{i=1}^{N}(s_i \cos\theta) = \sum_{i=1}^{N}(c_i \sin\theta) \\
\cos\theta \sum_{i=1}^{N}(s_i) \ = sin\theta \sum_{i=1}^{N}(c_i) \\
\cos\theta \sum_{i=1}^{N}(s_i) = \sin\theta \sum_{i=1}^{N}(c_i) \\
\frac{\sin\theta}{\cos\theta} = \frac{\sum_{i=1}^{N}(s_i)}{\sum_{i=1}^{N}(c_i)} \\
\tan\theta = \frac{\sum_{i=1}^{N}(s_i)}{\sum_{i=1}^{N}(c_i)} \\
\theta = \arctan\frac{\sum_{i=1}^{N}(s_i)}{\sum_{i=1}^{N}(c_i)}
\end{aligned}
$$

ML and MAP Solution of Poisson Fit

a

The probability density function for a Poission Distribution with a parameter λ is given by:

$$p(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

The parameter λ can then be estimated by the function:

$$\lambda_{ML} = argmax_{\lambda\in(0,\infty)}\{p(D|\lambda)\}$$

Thus the likelihood can be then modeled as:

$$
\begin{aligned}
p(D|\lambda) &= p(\{x_i\}_{i=1}^{N}|\lambda) \\
&= \prod_{i=1}^{N} p(x_i|\lambda) \\
&= \prod_{i=1}^{N} \frac{\lambda^x e^{-\lambda}}{x!} \\
&= \frac{\lambda^{\sum_{i=1}^{N} x_i} e^{-n\lambda}}{\sum_{i=1}^{N} x_i!}
\end{aligned}
$$

Taking the logarithm, followed by derivative on both sides:

$$\begin{aligned}
\ln(p(D|\lambda)) &= \ln\left(\frac{\lambda^{\sum_{i=1}^{N} x_i} e^{-n\lambda}}{\sum_{i=1}^{N} x_i!}\right) \\
&= \sum_{i=1}^{N} x_i \ln(\lambda) - n\lambda - \sum_{i=1}^{N} x_i! \\
\frac{\partial(\ln p(D|\lambda))}{\lambda} &= \frac{1}{\lambda}\sum_{i=1}^{N} x_i - n = 0 \\
\frac{1}{\lambda}\sum_{i=1}^{N} x_i = n \\
\lambda = \frac{1}{n}\sum_{i=1}^{N} x_i
\end{aligned}$$

b

The prior is given to be: $p(\lambda) \propto exp(-\lambda/a)$ and the likelihood is $p(D|\lambda)$. All together:

$$\begin{aligned}
p(\lambda|D) &= \frac{p(D|\lambda)p(\lambda)}{p(D)} \propto p(D|\lambda)p(\lambda) \\
\ln(p(\lambda|D)) &= \ln(p(D|\lambda)) + \ln(p(\lambda)) \\
&= \ln(\lambda)\sum_{i=1}^{N} x_i - n\lambda - \sum_{i=1}^{N} x_i! + ln(e^{-\lambda/a}) \\
&= \frac{1}{\lambda}\sum_{i=1}^{N} x_i - n - \frac{1}{a} \\
\lambda &= \sum_{i=1}^{N} x_i \left(\frac{1}{n} + a\right)
\end{aligned}$$

c

As n increases, both the numerator and denominator approach infinity in prior and likelihood cases. As a approaches infinity, the prior gets larger and larger and approaches 1. Similarly, as a approaches 0, the prior gets smaller and smaller and approaches 0