Homework 1

JD Gallego Posada, *University of Amsterdam* 11/04/2017

Problem 1

Mean

$$\mathbb{E}\left[\mathbf{y}\right] = \mathbb{E}\left[\mathbf{x}+\mathbf{z}\right] = \mathbb{E}\left[\mathbf{x}\right] + \mathbb{E}\left[\mathbf{z}\right] = \boldsymbol{\mu}_{\mathbf{x}} + \boldsymbol{\mu}_{\mathbf{z}}$$

Covariance

$$\begin{aligned} Cov\left[\mathbf{y}\right] &= Cov\left[\mathbf{x}+\mathbf{z}\right] = \mathbb{E}\left[\left(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}}+\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}}\right)\left(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}}+\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}}\right)^T\right] \\ &= \mathbb{E}\left[\left(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}}\right)\left(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}}\right)^T\right] + 2\mathbb{E}\left[\left(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}}\right)\left(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}}\right)^T\right] + \mathbb{E}\left[\left(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}}\right)\left(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}}\right)^T\right] \\ &= \boldsymbol{\Sigma}_{\mathbf{x}} + 2\boldsymbol{\Sigma}_{\mathbf{xz}} + \boldsymbol{\Sigma}_{\mathbf{z}} \end{aligned}$$

Problem 2

1. *Write down the likelihood of the data* $p(\mathcal{X}|\boldsymbol{\mu},\boldsymbol{\Sigma})$

Under i.i.d. assumptions,

$$p(\mathcal{X}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \prod_{n=1}^{N} p(\mathbf{x}_n|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \prod_{n=1}^{N} \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu},\boldsymbol{\Sigma})$$

2. *Write down the posterior* $p(\boldsymbol{\mu}|\mathcal{X},\boldsymbol{\Sigma},\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)$

$$p(\boldsymbol{\mu}|\mathcal{X},\boldsymbol{\Sigma},\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0) = \frac{\mathcal{N}(\boldsymbol{\mu}|\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)\prod_{n=1}^{N}\mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu},\boldsymbol{\Sigma})}{\int_{\Omega}\mathcal{N}(\boldsymbol{\mu}'|\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)\prod_{n=1}^{N}\mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}',\boldsymbol{\Sigma})d\boldsymbol{\mu}'}$$

3. *Show that* $p(\boldsymbol{\mu}|\mathcal{X},\boldsymbol{\Sigma},\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)$ *is a Gaussian distribution* $p(\boldsymbol{\mu}|\boldsymbol{\mu}_N,\boldsymbol{\Sigma}_N)$.

$$\begin{aligned} \log p(\boldsymbol{\mu}|\mathcal{X},\boldsymbol{\Sigma},\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0) &= c_1 - \frac{1}{2}\sum_n(\mathbf{x}_n-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}_n-\boldsymbol{\mu}) - \frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)^T\boldsymbol{\Sigma}_0^{-1}(\mu-\boldsymbol{\mu}_0) \\ &= c_2 - \frac{1}{2}\left(-2\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\sum_n\mathbf{x}_n + N\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{\mu}^T\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu} - 2\boldsymbol{\mu}^T\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0\right) \\ &= c_2 - \frac{1}{2}\left(\boldsymbol{\mu}^T\left(\boldsymbol{\Sigma}_0^{-1}+N\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{\mu} - 2\boldsymbol{\mu}^T\left(\boldsymbol{\Sigma}^{-1}\sum_n\mathbf{x}_n + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0\right)\right) \end{aligned}$$

Thus, $\boldsymbol{\mu}\sim\mathcal{N}(\boldsymbol{\mu}|\boldsymbol{\mu}_N,\boldsymbol{\Sigma}_N)$ with $\boldsymbol{\Sigma}_N^{-1} = \boldsymbol{\Sigma}_0^{-1} + N\boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{\mu}_N = \boldsymbol{\Sigma}_N\left(\boldsymbol{\Sigma}^{-1}\sum_n\mathbf{x}_n + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0\right)$.

4. *Derive the maximum a posterior solution for* $\boldsymbol{\mu}$.

Since the posterior distribution is Gaussian, the mode coincides with the mean. Therefore, the MAP solution for $\boldsymbol{\mu}$ is $\boldsymbol{\mu}_N$.

5. *Derive expressions for sequential update of* μ_N *and* σ_N^2.

$$\sigma_N^{2\,-1} = N\sigma^{2\,-1} + \sigma_0^{2\,-1} \rightarrow \sigma_N^2 = \frac{\sigma_0^2\sigma^2}{N\sigma_0^2+\sigma^2}$$

$$\mu_N = \frac{\sigma_0^2\sigma^2}{N\sigma_0^2+\sigma^2}\left(\frac{N}{\sigma^2}\frac{1}{N}\sum_{n=1}^{N}x_n + \frac{\mu_0}{\sigma_0^2}\right) = \frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\overline{x} + \frac{\sigma^2}{N\sigma_0^2+\sigma^2}\mu_0$$

5. *Derive the same results starting from the posterior distribution $p(\mu|x_1,\ldots,x_{N-1})$ and multiplying by the likelihood function $p(x_N|\mu,\sigma^2) = \mathcal{N}(x_N|\mu,\sigma^2)$.*

We know that $p(\mu|x_1,\ldots,x_{N-1}) = \mathcal{N}(\mu|\mu_{N-1},\sigma_{N-1}^2)$ and that the product of this posterior with $p(x_N|\mu,\sigma^2)$ will yield a new Gaussian $p(\mu|x_1,\ldots,x_N) = \mathcal{N}(\mu|\mu_N,\sigma_N^2)$, where:

$$\sigma_N^{2\,-1} = \sigma^{2\,-1} + \sigma_{N-1}^{2\,-1} = \sigma^{2\,-1} + (N-1)\sigma^{2\,-1} + \sigma_0^{2\,-1} = N\sigma^{2\,-1} + \sigma_0^{2\,-1}$$

$$\begin{aligned}
\mu_N &= \sigma_n^2\left(\frac{x_N}{\sigma^2} + \frac{\mu_{N-1}}{\sigma_{N-1}^2}\right) \\
&= \frac{\sigma_0^2\sigma^2}{N\sigma_0^2+\sigma^2}\frac{x_N}{\sigma^2} + \frac{\sigma_0^2\sigma^2}{N\sigma_0^2+\sigma^2}\frac{(N-1)\sigma_0^2+\sigma^2}{\sigma_0^2\sigma^2}\mu_{N-1} \\
&= \frac{\sigma_0^2}{N\sigma_0^2+\sigma^2}x_N + \frac{(N-1)\sigma_0^2+\sigma^2}{N\sigma_0^2+\sigma^2}\mu_{N-1} \\
&= \frac{\sigma_0^2}{N\sigma_0^2+\sigma^2}x_N + \frac{(N-1)\sigma_0^2+\sigma^2}{N\sigma_0^2+\sigma^2}\frac{\sigma_0^2\sigma^2}{(N-1)\sigma_0^2+\sigma^2}\left(\frac{1}{\sigma^2}\sum_{n=1}^{N-1}x_n + \frac{\mu_0}{\sigma_0^2}\right) \\
&= \frac{\sigma_0^2}{N\sigma_0^2+\sigma^2}x_N + \frac{\sigma_0^2}{N\sigma_0^2+\sigma^2}\sum_{n=1}^{N-1}x_n + \frac{\sigma^2}{N\sigma_0^2+\sigma^2}\mu_0 \\
&= \frac{\sigma_0^2}{N\sigma_0^2+\sigma^2}\sum_{n=1}^{N}x_n + \frac{\sigma^2}{N\sigma_0^2+\sigma^2}\mu_0
\end{aligned}$$

Problem 3

1. *Show that the product of two Gaussians gives another (un-normalized) Gaussian.*

$$\begin{aligned}
\log\mathcal{N}(\mathbf{x}|\mathbf{a},\mathbf{A})\mathcal{N}(\mathbf{x}|\mathbf{b},\mathbf{B}) &= c_1 - \frac{1}{2}(\mathbf{x}-\mathbf{a})^T\mathbf{A}^{-1}(\mathbf{x}-\mathbf{a}) - \frac{1}{2}(\mathbf{x}-\mathbf{b})^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{b}) \\
&= c_1 - \frac{1}{2}\left(\mathbf{x}^T\mathbf{A}^{-1}\mathbf{x} - 2\mathbf{x}\mathbf{A}^{-1}\mathbf{a} + \mathbf{a}^T\mathbf{A}^{-1}\mathbf{a} + \mathbf{x}^T\mathbf{B}^{-1}\mathbf{x} - 2\mathbf{x}\mathbf{B}^{-1}\mathbf{b} + \mathbf{b}^T\mathbf{B}^{-1}\mathbf{b}\right) \\
&= c_2 - \frac{1}{2}\left(\mathbf{x}^T(\mathbf{A}^{-1}+\mathbf{B}^{-1})\mathbf{x} - 2\mathbf{x}\left(\mathbf{A}^{-1}\mathbf{a}+\mathbf{B}^{-1}\mathbf{b}\right)\right)
\end{aligned}$$

Thus, $\mathbf{C}^{-1} = \mathbf{A}^{-1} + \mathbf{B}^{-1}$ and $\mathbf{c} = \mathbf{C}(\mathbf{A}^{-1}\mathbf{a} + \mathbf{B}^{-1}\mathbf{b})$.

2. *Using the Woodbury, Sherman & Morrison formula, prove that $\mathbf{C} = (\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1} = \mathbf{A} - \mathbf{A}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{A} = \mathbf{B} - \mathbf{B}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B}$.*

To prove the first identity, take $\mathbf{Z} = \mathbf{A}^{-1}$, $\mathbf{U} = \mathbf{V} = \mathbb{I}$ and $\mathbf{W} = \mathbf{B}^{-1}$. The second follows by symmetry.

3. *Show that $K^{-1} = (2\pi)^{-D/2}|\mathbf{A}+\mathbf{B}|^{-1/2}\exp\left(-\frac{1}{2}(\mathbf{a}-\mathbf{b})^T(\mathbf{A}+\mathbf{B})^{-1}(\mathbf{a}-\mathbf{b})\right)$.*

We first prove several results.

$$\mathbf{A}^{-1}\mathbf{C}\mathbf{B}^{-1} = \mathbf{A}^{-1}(\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1}\mathbf{B}^{-1} = (\mathbf{B}\mathbf{A}^{-1}\mathbf{A} + \mathbf{B}\mathbf{B}^{-1}\mathbf{A})^{-1} = (\mathbf{A}+\mathbf{B})^{-1}$$

This makes it trivial to see that $|\mathbf{A}\mathbf{C}^{-1}\mathbf{B}| = |\mathbf{A}+\mathbf{B}|$.

$$\begin{aligned}
\mathbf{c}\mathbf{C}^{-1}\mathbf{c} &= (\mathbf{a}^T\mathbf{A}^{-1} + \mathbf{b}^T\mathbf{B}^{-1})\mathbf{C}(\mathbf{A}^{-1}\mathbf{a} + \mathbf{B}^{-1}\mathbf{b}) \\
&= \mathbf{a}^T\mathbf{A}^{-1}\mathbf{C}\mathbf{A}^{-1}\mathbf{a} + \mathbf{b}^T\mathbf{B}^{-1}\mathbf{C}\mathbf{B}^{-1}\mathbf{b} + 2\mathbf{a}^T\mathbf{A}^{-1}\mathbf{C}\mathbf{B}^{-1}\mathbf{b} \\
&= \mathbf{a}^T\mathbf{A}^{-1}(\mathbf{A} - \mathbf{A}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{A})\mathbf{A}^{-1}\mathbf{a} + \mathbf{b}^T\mathbf{B}^{-1}(\mathbf{B} - \mathbf{B}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B})\mathbf{B}^{-1}\mathbf{b} + 2\mathbf{a}^T(\mathbf{A}+\mathbf{B})^{-1}\mathbf{b}
\end{aligned}$$

$$= \mathbf{a}^T\mathbf{A}^{-1}\mathbf{a} - \mathbf{a}^T(\mathbf{A}+\mathbf{B})^{-1}\mathbf{a} + \mathbf{b}^T\mathbf{B}^{-1}\mathbf{b} - \mathbf{b}^T(\mathbf{A}+\mathbf{B})^{-1}\mathbf{b} + 2\mathbf{a}^T(\mathbf{A}+\mathbf{B})^{-1}\mathbf{b}$$
$$= \mathbf{a}^T\mathbf{A}^{-1}\mathbf{a} + \mathbf{b}^T\mathbf{B}^{-1}\mathbf{b} - (\mathbf{a}-\mathbf{b})^T(\mathbf{A}+\mathbf{B})^{-1}(\mathbf{a}-\mathbf{b})$$

Now, consider the original problem $K^{-1}\mathcal{N}(\mathbf{x}|\mathbf{c},\mathbf{C}) = \mathcal{N}(\mathbf{x}|\mathbf{a},\mathbf{A})\mathcal{N}(\mathbf{x}|\mathbf{b},\mathbf{B})$:

$$K^{-1} = (2\pi)^{-D/2}|\mathbf{A}\mathbf{C}^{-1}\mathbf{B}|^{-1/2}\exp\left(-\frac{1}{2}\left[(\mathbf{x}-\mathbf{a})^T\mathbf{A}^{-1}(\mathbf{x}-\mathbf{a}) + (\mathbf{x}-\mathbf{b})^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{b}) - (\mathbf{x}-\mathbf{c})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{c})\right]\right)$$

From part 1, we have already seen that the terms in this equation containing $\mathbf{x}$ vanish. Thus, we are left with the constant terms, and by using the previous results, we have:

$$K^{-1} = (2\pi)^{-D/2}|\mathbf{A}+\mathbf{B}|^{-1/2}\exp\left(-\frac{1}{2}(\mathbf{a}-\mathbf{b})^T(\mathbf{A}+\mathbf{B})^{-1}(\mathbf{a}-\mathbf{b})\right)$$

Problem 4

1. *We toss the coin 3 times and it all comes up with heads. How likely is that in the next toss, the coin comes up with head according to MLE?*
We prove a more general result first. Let m be the number of heads and l the number of tails.

$$\mathcal{L}(m \text{ heads and } l \text{ tails}) = \mu^m(1-\mu)^l \rightarrow \log\mathcal{L} = m\log\mu + l\log(1-\mu) \rightarrow \mu_{\text{MLE}} = \frac{m}{m+l}$$

Given 3 heads and 0 tails, $\mu_{\text{MLE}} = 1$.
2. *Suppose the prior $\mu \sim$ Beta$(\mu|a,b)$. What is the probability that the coin comes up with head in the 4th toss?*

$$p(\mu|m \text{ heads and } l \text{ tails}, a, b) \propto \mu^{m+a-1}(1-\mu)^{l+b-1}$$

From part 1, it is clear that the value of μ that maximizes the posterior probability is

$$\mu_{\text{MAP}} = \frac{m+a-1}{(m+l)+(a+b)-2}$$

Then, given $m=3$ and $l=0$, the probability that the 4th toss is head is $\dfrac{a+2}{a+b-1}$. Besides, note that the posterior distribution is Beta$(\mu|m+a, l+b)$.
3. *Suppose that we observe m times that the coin lands heads and l times that it lands tails. Show that the posterior mean lies between the prior mean and μ_{MLE}.*
The prior and posterior means are $\frac{a}{a+b}$ and $\frac{m+a}{m+l+a+b}$, respectively, since the prior and posterior are Beta distributions.

$$\frac{m+a}{m+l+a+b} = \frac{m}{m+l+a+b} + \frac{a}{m+l+a+b} = \frac{m+l}{m+l+a+b}\frac{m}{m+l} + \frac{a+b}{m+l+a+b}\frac{a}{a+b}$$

This is clearly a linear convex combination of μ_{MLE} and the prior mean.

Problem 5

In this section we use the result $\mathbb{E}\left[u(\mathbf{x})\right] = -\nabla_{\boldsymbol{\eta}}\log g(\boldsymbol{\eta})$, covered in the lectures.

<u>Poisson</u>

$$\text{Poisson}(x|\lambda) = \frac{1}{x!}e^{-\lambda}\lambda^x = \frac{1}{x!}e^{-\lambda+x\log\lambda}$$

Thus, we can express it in exponential family form with $h(x) = \frac{1}{x!}$, $u(x) = x$, $\eta = \log\lambda$ and $g(\eta) = e^{-\lambda} = e^{-e^{\eta}}$.

$$\mathbb{E}[x] = \mathbb{E}[u(x)] = -\frac{\partial \log g(\eta)}{\partial \eta} = -\frac{\partial \log e^{-e^\eta}}{\partial \eta} = \frac{\partial e^\eta}{\partial \eta} = e^\eta = \lambda$$

Gamma

$$\text{Gamma}(x|a,b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-xb} = \frac{b^a}{\Gamma(a)} x^{-1} e^{-xb + a\log x}$$

Thus, we can express it in exponential family form with $h(x) = \frac{1}{x}$, $\mathbf{u}(x) = [-x, \log x]^T$, $\boldsymbol{\eta} = [b,a]^T$ and $g(\boldsymbol{\eta}) = \frac{b^a}{\Gamma(a)}$.

$$\mathbb{E}[x] = \mathbb{E}[-\mathbf{u}_1(x)] = \frac{\partial \log g(\boldsymbol{\eta}_1)}{\boldsymbol{\eta}_1} = \frac{\partial \log \frac{b^a}{\Gamma(a)}}{\partial b} = \frac{\partial a \log b - \log \Gamma(a)}{\partial b} = \frac{a}{b}$$

⋆ Problem 6

For this exercise we consider the following equivalent formulations of the D-dimensional Student's t-distribution.

$$\text{St}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda}, \nu) = \frac{\Gamma(\frac{D+\nu}{2})}{\Gamma(\frac{\nu}{2})} \frac{|\boldsymbol{\Lambda}|^{1/2}}{(\pi\nu)^{D/2}} \left[1 + \frac{1}{\nu}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Lambda}(\mathbf{x}-\boldsymbol{\mu})\right]^{-\frac{D+\nu}{2}} = \int_{\mathbb{R}+} \mathcal{N}\left(\mathbf{x}|\boldsymbol{\mu}, \frac{1}{\eta}\boldsymbol{\Lambda}^{-1}\right) \text{Gamma}\left(\eta|\frac{\nu}{2}, \frac{\nu}{2}\right) d\eta$$

Mean

$$\mathbb{E}[\mathbf{x}-\boldsymbol{\mu}] = k\int_{\mathbb{R}^D} \left[1 + \frac{1}{\nu}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Lambda}(\mathbf{x}-\boldsymbol{\mu})\right]^{-\frac{D+\nu}{2}} (\mathbf{x}-\boldsymbol{\mu})\, d\mathbf{x} = k\int_{\mathbb{R}^D} \left[1 + \frac{1}{\nu}\mathbf{z}^T \boldsymbol{\Lambda}\mathbf{z}\right]^{-\frac{D+\nu}{2}} \mathbf{z}\, d\mathbf{z} = 0$$

This final equality holds, as it is the integral of an odd function. The linearity of the expected value gives the desired result $\mathbb{E}[\mathbf{x}] = \boldsymbol{\mu}$.

Covariance

$$\begin{aligned}
Cov[\mathbf{x}] &= \int_{\mathbb{R}^D} \text{St}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda}, \nu)(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T d\mathbf{x} \\
&= \int_{\mathbb{R}+}\int_{\mathbb{R}^D} \mathcal{N}\left(\mathbf{x}|\boldsymbol{\mu}, \frac{1}{\eta}\boldsymbol{\Lambda}^{-1}\right) \text{Gamma}\left(\eta|\frac{\nu}{2}, \frac{\nu}{2}\right)(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T d\mathbf{x} d\eta \\
&= \int_{\mathbb{R}+} \frac{1}{\eta}\boldsymbol{\Lambda}^{-1} \text{Gamma}\left(\eta|\frac{\nu}{2}, \frac{\nu}{2}\right) d\eta \\
&= \boldsymbol{\Lambda}^{-1}\int_{\mathbb{R}+} \eta^{-1} \frac{1}{\Gamma(\frac{\nu}{2})}\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}} \eta^{\frac{\nu}{2}-1} \exp\left(-\frac{\nu}{2}\eta\right) d\eta \\
&= \boldsymbol{\Lambda}^{-1}\int_{\mathbb{R}+} \frac{1}{(\frac{\nu}{2}-1)\Gamma(\frac{\nu}{2}-1)} \frac{\nu}{2}\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}-1} \eta^{(\frac{\nu}{2}-1)-1} \exp\left(-\frac{\nu}{2}\eta\right) d\eta \\
&= \boldsymbol{\Lambda}^{-1}\frac{\frac{\nu}{2}}{(\frac{\nu}{2}-1)}\int_{\mathbb{R}+} \frac{1}{\Gamma(\frac{\nu}{2}-1)} \left(\frac{\nu}{2}\right)^{\frac{\nu}{2}-1} \eta^{(\frac{\nu}{2}-1)-1} \exp\left(-\frac{\nu}{2}\eta\right) d\eta \\
&= \frac{\nu}{\nu-2}\boldsymbol{\Lambda}^{-1}\int_{\mathbb{R}+} \text{Gamma}\left(\eta|\frac{\nu}{2}-1, \frac{\nu}{2}\right) d\eta = \frac{\nu}{\nu-2}\boldsymbol{\Lambda}^{-1}
\end{aligned}$$

For this final Gamma distribution to be well-defined, it is required that $\nu > 2$.

Mode

Note that $\text{St}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda}, \nu)$ is monotonously decreasing in $(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Lambda}(\mathbf{x}-\boldsymbol{\mu})$. Thus, the mode of $\text{St}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda}, \nu)$ is located at $\boldsymbol{\mu}$.