1. Probability Theory Reminders

$$f_{\mathbf{X}}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{1}$$

a) In order to solve this task we will make use of the following well-known integrals:

$$\int_{-\infty}^{+\infty} f_{\mathbf{X}} dx = 1 \tag{2}$$

$$\int_{-\infty}^{+\infty} u e^{-\frac{u^2}{a}} du = 0 \tag{3}$$

$$\int_{-\infty}^{+\infty} u^2 e^{-au^2} du = \frac{\sqrt{\pi}}{2a^{\frac{3}{2}}} \tag{4}$$

$$\begin{aligned}
\mathbb{E}[\mathbf{X}] &= \int_{-\infty}^{+\infty} x f_{\mathbf{X}} dx \\
&= \int_{-\infty}^{+\infty} (x-\mu) f_{\mathbf{X}} dx + \int_{-\infty}^{+\infty} \mu f_{\mathbf{X}} dx \\
&= \int_{-\infty}^{+\infty} (x-\mu) f_{\mathbf{X}} dx + \mu \int_{-\infty}^{+\infty} f_{\mathbf{X}} dx \\
&= \int_{-\infty}^{+\infty} (x-\mu) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \mu \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (x-\mu) e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \mu \\
&= \mu
\end{aligned} \tag{5}$$

Where in order to get from line 3 to 4 we used Equation 2, and in order to get from line 4 to 5 we used Equation 3 with $u \coloneqq (x-\mu)$.

$$\begin{aligned}
\mathbb{E}[(\mathbf{X}-\mu)^2] &= \int_{-\infty}^{+\infty} (x-\mu)^2 f_{\mathbf{X}} dx \\
&= \int_{-\infty}^{+\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx
\end{aligned} \tag{6}$$

Using $a \coloneqq \frac{1}{2\sigma^2}$ and $u \coloneqq (x-\mu)$ we have:

$$
\begin{aligned}
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} u^2 e^{-au^2} du \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \frac{\sqrt{\pi}}{2a^{3/2}} \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \frac{\sqrt{\pi}}{2(\frac{1}{2\sigma^2})^{3/2}} \\
&= \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{2\sigma^2}} \frac{\sqrt{\pi}}{2(\frac{1}{2\sigma^2})^{3/2}} \\
&= \frac{1}{\sqrt{2\sigma^2}} \frac{1}{2(\frac{1}{2\sigma^2})^{3/2}} \\
&= \frac{1}{\sqrt{2}\sigma} \frac{1}{2(\frac{1}{2\sigma^2})^{3/2}} \\
&= \frac{1}{\sqrt{2}\sigma} \frac{1}{2(\frac{1}{2\sqrt{2}\sigma^3})} \\
&= \frac{1}{\sqrt{2}\sigma} \frac{2\sqrt{2}\sigma^3}{2} \\
&= \frac{1}{\sqrt{2}\sigma} \sqrt{2}\sigma^3 \\
&= \frac{\sigma^3}{\sigma} \\
&= \sigma^2
\end{aligned} \tag{7}
$$

Where we used Equation 4 in order to get from line 2 to 3 in Equation 7.

b)

$$
\begin{aligned}
F_{\mathbf{X}}(x) &= \int_{-\infty}^{x} f_{\mathbf{X}}(z) dz \\
&= \int_{-\infty}^{x} \frac{1}{2\beta} e^{-\frac{|x-\mu|}{\beta}} dx \\
&= \frac{1}{2\beta} \int_{-\infty}^{x} e^{-\frac{|x-\mu|}{\beta}} dx
\end{aligned} \tag{8}
$$

Using $u \coloneqq (x-\mu)$ we have:

$$
= \frac{1}{2\beta} \int_{-\infty}^{x-\mu} e^{-\frac{|u|}{\beta}} du \tag{9}
$$

For $x - \mu < 0$:

$$
\begin{aligned}
&= \frac{1}{2\beta} \int_{-\infty}^{x-\mu} e^{-\frac{|u|}{\beta}} du \\
&= \frac{1}{2\beta} \int_{-\infty}^{x-\mu} e^{\frac{u}{\beta}} du \\
&= \frac{1}{2\beta} \beta e^{\frac{u}{\beta}} \Big|_{-\infty}^{x-\mu} \\
&= \frac{1}{2} e^{\frac{x-\mu}{\beta}}
\end{aligned} \tag{10}
$$

And for $x - \mu \geq 0$ we split the integral into two parts:

$$= \frac{1}{2\beta}\int_{-\infty}^{\mu} e^{-\frac{|x-\mu|}{\beta}} dx + \frac{1}{2\beta}\int_{\mu}^{x} e^{-\frac{|x-\mu|}{\beta}} dx \tag{11}$$

Substituting $t \coloneqq \frac{z-\mu}{\beta}$ and using $z = \beta t + \mu$ in order to get $\frac{dz}{dt} = \frac{d(\beta t+\mu)}{dt} = \beta$:

$$\begin{aligned}
&= \frac{1}{2\beta}\int_{-\infty}^{\mu} e^{t} dt + \frac{1}{2\beta}\int_{\mu}^{x} e^{-t} dt \\
&= \frac{1}{2\beta}\beta e^{t}\Big|_{-\infty}^{\mu} + \frac{1}{2\beta}\beta e^{-t}\Big|_{\mu}^{x} \\
&= \frac{1}{2} + 0 - \frac{1}{2\beta}\beta e^{-\frac{x-\mu}{\beta}} + \frac{1}{2} \\
&= 1 - \frac{1}{2}e^{-\frac{x-\mu}{\beta}}
\end{aligned} \tag{12}$$

Where we resubstituted t and evaluated the expressions in order to get from line 2 to 3.

Therefore, the complete cdf of the Laplacian reads as follows:

$$F_{\mathbf{X}}(x) = \left\{ \begin{array}{ll} 1 - \frac{1}{2}e^{-\frac{x-\mu}{\beta}}, & \text{if } \mu \leq x \\ \frac{1}{2}e^{\frac{x-\mu}{\beta}}, & \text{if } \mu > x \end{array} \right\} \tag{13}$$

The median of the Laplacian is where $\frac{1}{2}$ of the events lie above and below the value, respectively. Therefore, we have:

$$F_{\mathbf{X}}(x) = \int_{-\infty}^{x} F_{\mathbf{X}}(x) dx = \frac{1}{2} \tag{14}$$

Using the results from above we again have a case distinction. For $x - \mu < 0$:

$$\begin{aligned}
\frac{1}{2}e^{\frac{x-\mu}{\beta}} &= \frac{1}{2} && (15)\\
e^{\frac{x-\mu}{\beta}} &= 1 && (16)\\
-(z-\mu) &= \log 1\beta && (17)\\
0 &= -(z-\mu) && (18)\\
z &= \mu && (19)\\
& && (20)
\end{aligned}$$

For $x - \mu \geq 0$:

$$\begin{aligned}
1 - \frac{1}{2}e^{\frac{x-\mu}{\beta}} &= \frac{1}{2} && (21)\\
\frac{1}{2} &= \frac{1}{2}e^{\frac{x-\mu}{\beta}} && (22)\\
1 &= e^{\frac{x-\mu}{\beta}} && (23)\\
\log 1\beta &= -(z-\mu) && (24)\\
0 &= -(z-\mu) && (25)\\
z &= \mu && (26)\\
& && (27)
\end{aligned}$$

Therefore, the median is equal to μ.

c) Since $\mathbf{X}$ and $\mathbf{Y}$ are independent we have $f_{\mathbf{X},\mathbf{Y}} = f_{\mathbf{X}} f_{\mathbf{Y}}$ and therefore:

$$\begin{aligned} f_{\mathbf{X}} f_{\mathbf{Y}} &\sim \mathcal{N}(x \mid 0,\, \sigma_x^2)\mathcal{N}(y \mid 0,\, \sigma_y^2) \\ &= \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{(y-\mu_y)^2}{2\sigma_y^2}} \\ &= \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}} \end{aligned} \tag{28}$$

Where we used that $\mu_x = \mu_y = 0$ by definition in order to get from line 2 to 3.

For $\mathbf{Q} \coloneqq \frac{\mathbf{X}}{\mathbf{Y}}$ we have:

$$\begin{aligned} f_{\mathbf{Q}}(q) &= \int_{-\infty}^{\infty} |x| f_{\mathbf{X},\mathbf{Y}}(qx, x) dx \\ &= 2\int_{0}^{\infty} |x| f_{\mathbf{X},\mathbf{Y}}(qx, x) dx \\ &= \frac{1}{\pi\sigma_x\sigma_y} \int_{0}^{\infty} |x| e^{\frac{-(qx)^2\sigma_y^2 - x^2\sigma_x^2}{2\sigma_x^2\sigma_y^2}} dx \end{aligned} \tag{29}$$

Substituting $a \coloneqq \left(\frac{q^2\sigma_y^2 - \sigma_x^2}{2\sigma_x^2\sigma_y^2}\right)$ and using the standard integral $\int_0^\infty xe^{-ax^2} dx = \frac{1}{2a}$ we get:

$$\begin{aligned} &= \frac{1}{\pi\sigma_x\sigma_y} \frac{1}{2\left(\frac{q^2\sigma_y^2 + \sigma_x^2}{2\sigma_x^2\sigma_y^2}\right)} \\ &= \frac{2\sigma_x^2\sigma_y^2}{2\pi\sigma_x\sigma_y(q^2\sigma_y^2 + \sigma_x^2)} \\ &= \frac{\sigma_x\sigma_y}{\pi(q^2\sigma_y^2 + \sigma_x^2)} \end{aligned} \tag{30}$$

Using $\gamma \coloneqq \frac{\sigma_x}{\sigma_y}$ we get:

$$\begin{aligned} &= \frac{\sigma_x\sigma_y}{\pi(q^2\sigma_y^2 + \sigma_x^2)} \\ &= \frac{\sigma_x\sigma_y}{\pi(q^2\sigma_y^2 + \sigma_x^2)} \frac{\sigma_y}{\sigma_y} \\ &= \gamma \frac{\sigma_y}{\pi(q^2\sigma_y^2 + \sigma_x^2)} \sigma_y \\ &= \gamma \frac{\sigma_y^2}{\pi(q^2\sigma_y^2 + \sigma_x^2)} \\ &= \frac{1}{\pi} \gamma \frac{\sigma_y^2}{q^2\sigma_y^2 + \sigma_x^2} \\ &= \frac{1}{\pi} \frac{\gamma}{q^2 + \gamma^2} \end{aligned} \tag{31}$$

which corresponds exactly to the Cauchy distribution with location parameter $x_0 = 0$ and scale $\gamma = \frac{\sigma_x}{\sigma_y}$.

2. Bayesian Inference

a) Using the premise that the random variables are independent we have:

$$\begin{aligned}\mathcal{L}(\mu) &\coloneqq p(d|\mu) \\ &= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}} e^{\frac{-(d_i-\mu)^2}{2}} \\ &= \frac{1}{(2\pi)^{\frac{N}{2}}} e^{-\frac{\sum_{i=1}^{N}(d_i-\mu)^2}{2}}\end{aligned} \tag{32}$$

Taking the logarithm of Equation 32:

$$\begin{aligned}\log(\mathcal{L}(\mu)) &= \log(p(d|\mu)) \\ &= -\frac{N}{2}\log(2\pi) - \frac{\sum_{i=1}^{N}(d_i-\mu)^2}{2}\end{aligned} \tag{33}$$

b)

$$\begin{aligned}\hat{\mu} &= \underset{\mu}{\operatorname{argmax}}\ \mathcal{L}(\mu) \\ &= \underset{\mu}{\operatorname{argmax}}\ \log(\mathcal{L}(\mu)) \\ &= \underset{\mu}{\operatorname{argmax}}\ -\frac{N}{2}\log(2\pi) - \frac{\sum_{i=1}^{N}(d_i-\mu)^2}{2}\end{aligned} \tag{34}$$

Setting $\frac{d}{d\mu}\log(\mathcal{L}) = 0$ let $\hat{\mu}$ be the optimum, then:

$$\begin{aligned}-\frac{\sum_{i=1}^{N}(d_i-\mu)^2}{2} &= 0 \\ -\hat{\mu}N + \sum_{i=1}^{N} d_i &= 0 \\ \hat{\mu} &= \frac{1}{N}\sum_{i=1}^{N} d_i\end{aligned} \tag{35}$$

c)

$$
\begin{aligned}
p(\mu|d) &= \frac{p(d|\mu)p(\mu)}{p(d)} \propto p(d|\mu)p(\mu) \\
&= p(\mu)p(d_1|\mu)p(d_2|\mu)...p(d_n|\mu) \\
&= \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{\frac{-(\mu-\mu_0)^2}{2\sigma_0^2}} \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}} e^{\frac{-(d_i-\mu)^2}{2}} \\
&= \frac{1}{(2\pi)^{\frac{n+1}{2}}\sqrt{\sigma_0^2}} e^{\frac{-(\mu^2+2\mu\mu_0+\mu_0^2)}{2\sigma_0^2} \sum_{i=1}^{N} \frac{d_1^2-2d_o\mu+\mu^2}{2}} \\
&\propto exp\left(\frac{-\mu^2-\mu^2\sigma_0^2+2\mu(\mu_0+\sigma_0^2 d_1+...+\sigma_0^2 d_N)}{2\sigma_0^2} - \frac{\mu_0^2+\sigma_0^2 d_1+...+\sigma_0^2 d_N^2}{2\sigma_0^2}\right) \\
&= exp\left(\frac{-\mu^2+2\mu\frac{\mu_0+\sum_{i=1}^{N}\sigma_0^2 d_i}{1+N\sigma_0^2} - \left(\frac{\mu_0+\sum_{i=1}^{N}\sigma_0^2 d_i}{1+N\sigma_0^2}\right)^2}{2\frac{\sigma_0^2}{1+N\sigma_0^2}}\right) exp\left(-\frac{\mu_0^2+\sum_{i=1}^{N}\sigma_0^2 d_i^2}{2\sigma_0^2}\right) \\
&\propto exp\left(\frac{-\mu^2+2\mu\frac{\mu_0+\sum_{i=1}^{N}\sigma_0^2 d_i}{1+N\sigma_0^2} - \left(\frac{\mu_0+\sum_{i=1}^{N}\sigma_0^2 d_i}{1+N\sigma_0^2}\right)^2}{2\frac{\sigma_0^2}{1+N\sigma_0^2}}\right)
\end{aligned} \tag{36}
$$

Using the second binomial formula ($(a-b)^2 = a^2 + 2ab + b^2$) we get:

$$
\propto exp\left(\frac{-\left(\mu - \left(\frac{\mu_0+\sum_{i=1}^{N}\sigma_0^2 d_i}{1+N\sigma_0^2}\right)\right)^2}{2\frac{\sigma_0^2}{1+N\sigma_0^2}}\right) \tag{37}
$$

By comparison of coefficients we can now easily extract $\hat{\mu}$ and $\hat{\sigma}$:

$$
\begin{aligned}
\hat{\mu} &= \frac{\mu_0+\sum_{i=1}^{N}\sigma_0^2 d_i}{1+N\sigma_0^2} \\
\hat{\sigma} &= \frac{\sigma_0^2}{1+N\sigma_0^2}
\end{aligned} \tag{38}
$$

We then normalize the posterior:

$$
\begin{aligned}
p(\mu|d) &= C * exp\left(-\frac{1}{2}\frac{(\mu-\hat{\mu})^2}{\sigma^2}\right) \\
&= \frac{1}{\sqrt{2\pi\hat{\sigma}_0^2}} exp\left(-\frac{1}{2}\frac{(\mu-\hat{\mu})^2}{\sigma^2}\right) \\
&= \mathcal{N}(\mu|\hat{\mu}, \hat{\sigma}_0^2)
\end{aligned} \tag{39}
$$

d) Using the results from c) with $\mu_p := \hat{\mu}$, using the notation of μ_p as the mean of the posterior and from now on $\hat{\mu}$ as the MAP estimate of μ:

$$\begin{aligned} \frac{d}{d\mu}\log(p(\mu|d)) &= 0 \\ \frac{d}{d\mu}\left(-\frac{1}{2}\frac{(\mu-\mu_p)^2}{\sigma_p^2}\right) &= 0 \\ \frac{-2(\mu-\mu_p)}{\sigma_p^2} &= 0 \\ \hat{\mu} = \mu_p &= \frac{\mu_0 + \sum_{i=1}^{N}\sigma_0^2 d_i}{1+N\sigma_0^2} \end{aligned} \tag{40}$$

e) The uninformative prior for μ is proportional to 1, i.e. $p(\mu) \propto 1$ and therefore (using the results from a)):

$$p(\mu|d) \propto p(d|\mu)p(\mu) = p(d|\mu) = \frac{1}{(2\pi)^{\frac{N}{2}}} exp\left(-\frac{\sum_{i=1}^{N}(d_i-\mu)^2}{2}\right) \tag{41}$$

Normalizing and calculating the MAP:

$$\begin{aligned} &\underset{\mu}{\text{argmax}}\ \log\left(C\frac{1}{(2\pi)^{\frac{N}{2}}} exp\left(-\frac{\sum_{i=1}^{N}(d_i-\mu)^2}{2}\right)\right) \\ &= \frac{d}{d\mu}\left(\log(C) + \log\left(\frac{1}{(2\pi)^{\frac{N}{2}}}\right) - \frac{\sum_{i=1}^{N}(d_i-\mu)^2}{2}\right) \\ &= -\frac{\sum_{i=1}^{N}(d_i-\mu)^2}{2} = 0 \end{aligned} \tag{42}$$

which yields the same as Equation 35. The MLE and the MAP are the same, except for the uniform prior, which will, however, not have any influence as:

$$\log(C) = \log\left(\int_{-\infty}^{\infty} p(\mu|\mathcal{R})\right) = 1 \tag{43}$$

3. Bayesian Inference: Linear Model

a) The posterior is given by Bayes' theorem:

$$p(\beta|D) = \frac{p(D|\beta)p(\beta)}{p(D)} \tag{44}$$

Using $p(y_0|\beta, I) = \mathcal{N}(y_0|\beta x, \sigma^2)$ and the results from 2)e) (MAP is equal to $\hat{\beta}$):

$$\begin{aligned} \frac{d}{d\beta}\log\left(p(D|\beta)\right) &= 0 \\ \frac{-2\left(\beta - \frac{y_0}{x_0}\right)}{2\frac{\sigma_1^2}{x_0^2}} &= 0 \\ -\frac{x_0^2}{\sigma_1^2}\left(\beta - \frac{y_0}{x_0}\right) &= 0 \\ \beta - \frac{y_0}{x_0} &= 0 \\ \hat{\beta} = \frac{y_0}{x_0} &= \hat{\beta}_{MAP} \end{aligned} \tag{45}$$

The variance follows similarly:

$$\begin{aligned} \frac{-2\left(\beta - \frac{y_0}{x_0}\right)}{2\frac{\sigma_1^2}{x_0^2}} &= 0 \\ \frac{x_0^2}{\sigma_1^2} &= \sigma_p(\beta|D) \end{aligned} \tag{46}$$

and the standard deviation is trivially following as $\frac{\sigma}{x_0}$. The complete posterior of β after observing D is therefore:

$$p(\beta|D) \propto \mathcal{N}\left(\beta|\frac{y_0}{x_0}, \frac{\sigma^2}{x_0^2}\right) \tag{47}$$

b) The setting stays the same, except for the change in the prior $p(\beta) = \mathcal{N}\left(0, \tau^2\right)$. We therefore write our posterior:

$$\begin{aligned} p(\beta|D) &\propto exp\left(-\frac{1}{2\sigma^2}(y_0 - \beta x_0)^2 - \frac{1}{2\tau^2}\beta^2\right) \\ &= exp\left(-\frac{1}{2\sigma^2}\left(y_0^2 - 2y_0\beta x_0 + \beta^2\left(x_0^2 + \frac{\sigma^2}{\tau^2}\right)\right)\right) \\ &= exp\left(-\frac{\left(x_0^2 + \frac{\sigma^2}{\tau^2}\right)}{2\sigma^2}\left(\frac{y_0^2}{\left(x_0^2 + \frac{\sigma^2}{\tau^2}\right)} - \frac{2y_0\beta x_0}{\left(x_0^2 + \frac{\sigma^2}{\tau^2}\right)} + \beta^2\right)\right) \end{aligned} \tag{48}$$

By using the second binomial formula and omitting the term independent of β we get:

$$= exp\left(-\frac{\left(x_0^2 + \frac{\sigma^2}{\tau^2}\right)}{2\sigma^2}\left(\frac{y_0 x_0}{\left(x_0^2 + \frac{\sigma^2}{\tau^2}\right)} - \beta^2\right)\right) \tag{49}$$

Now, by comparison of coefficients we can fix the parameters of our posterior:

$$p(\beta|D) = \mathcal{N}\left(\frac{y_0 x_0}{\left(x_0^2 + \frac{\sigma^2}{\tau^2}\right)}, \frac{\sigma^2}{\left(x_0^2 + \frac{\sigma^2}{\tau^2}\right)}\right) \tag{50}$$