High Performance Computing for Science and Engineering II.

Exercise Set 1

Florian Mahlknecht

2020-03-05

1 Probability Theory Reminders

1.1 Expected value and Variance of the Normal Distribution

Given

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

we observe that

$$\begin{aligned} f_X(\mu - x) &= f_X(\mu + x) \\ \iff (x-\mu)^2\big|_{\mu-x} &= (x-\mu)^2\big|_{\mu+x} = x^2 \end{aligned} \tag{2}$$

Expected value We recall the definition of the expected value:

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f_X(x) dx$$

By exploiting additivity of integrals on intervals, we may split up the integral into

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f(x) dx = \int_{-\infty}^{\mu} x f(x) dx + \int_{\mu}^{+\infty} x f(x) dx$$

We introduce a change of integration variable to $\tilde{x} = \mu - x$, which sets the upper and lower bound respectively to $\mu - \mu = 0$, and switches the sgn. This yields:

$$\begin{aligned} \mathbb{E}[X] &= -\int_{+\infty}^{0} (\mu - x) f(\mu - x) dx - \int_{0}^{-\infty} (\mu - x) f(\mu - x) dx = \\ &= \int_{0}^{+\infty} (\mu - x) f(\mu - x) dx + \int_{0}^{+\infty} (\mu + x) f(\mu + x) dx \end{aligned}$$

where in the last integral we change variable again with $\tilde{x} = -x$. Now we can apply eq. (2), simplifying the integral to:

$$\mathbb{E}[X] = \int_0^{+\infty} (2\mu) f(\mu + x) dx = \mathbb{E}[X] = (2\mu)\frac{1}{2}\int_{-\infty}^{+\infty} f(\mu + x) dx = \frac{2\mu}{2} = \mu$$

By just exploiting the symmetry property (eq. (2)) and the normalization property of (any) probability density function, i.e. $\int_{-\infty}^{+\infty} f_X(x)dx = 1$.

□ *q.e.d.*

Note that one could also show that the integral is indeed one, without relying on the normalization property:

$$\begin{aligned} a &\coloneqq \int_{-\infty}^{+\infty} e^{-t^2} dt \\ a^2 &= \int_{-\infty}^{+\infty} e^{-t^2} dt \int_{-\infty}^{+\infty} e^{-t^2} dt \\ &= \int_{-\infty}^{+\infty} e^{-x^2} dx \int_{-\infty}^{+\infty} e^{-y^2} dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-(x^2+y^2)} dx dy \end{aligned} \tag{3}$$

By switching to polar coordinates on the plane we can integrate over ρ on $[0, +\infty)$ and over θ on $[0, 2\pi]$.

$$\iint_{\mathbb{R}^2} e^{-(x^2+y^2)} dx dy = \int_0^{+\infty} \int_0^{2\pi} e^{-\rho^2} \rho d\rho d\theta = \\ 2\pi \int_0^{+\infty} e^{-\rho^2} \rho d\rho = 2\pi \, \frac{1}{2} e^{-\rho^2}\Big|_0^{+\infty} = \pi$$

$$\Rightarrow a^2 = \pi, \; a > 0 \Rightarrow a = \sqrt{\pi} \Rightarrow \int_{-\infty}^{+\infty} e^{-t^2} dt = \sqrt{\pi}$$

With this knowledge we can now calculate:

$$\begin{aligned} &\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &t^2 \coloneqq \frac{(x-\mu)^2}{2\sigma^2} \\ &t = \frac{x-\mu}{\sqrt{2}\,\sigma} \\ &\sqrt{2}\,\sigma \, dt = dx \\ &\frac{\sqrt{2}\,\sigma}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-t^2} dt = \frac{1}{\sqrt{\pi}}\sqrt{\pi} = 1 \end{aligned}$$

□ *q.e.d.*

Florian Mahlknecht
fmahlknecht@student.ethz.ch
19-945-351

Variance The variance is defined as

$$\mathrm{Var}[X] = \mathbb{E}\left[\left(X-\mu\right)^2\right)\right] \tag{4}$$

which yields:

$$\mathrm{Var}[X] = \mathbb{E}\left[\left(X-\mathbb{E}[X]\right)^2\right] = \mathbb{E}\left[\left(X-\mu\right)^2\right] = \mathbb{E}[X^2] - \mu^2 = \int_{\mathbb{R}} x^2 f_X(x)dx - \mu^2$$

By substituting the Gaussian bell function we get further:

$$\mathrm{Var}[X] = \int_{-\infty}^{+\infty} \frac{x^2}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx - \mu^2 \tag{5}$$

A change in integration variable to $\tilde{x} = x + \mu$ yields:

$$\begin{aligned}
\int_{-\infty}^{+\infty} \frac{x^2 + 2x\mu}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx + \mu^2 - \mu^2 &= \\
\int_{-\infty}^{+\infty} \frac{x^2}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx + 2\mu\,\mathbb{E}\left[\mathcal{N}\left(x|0,\sigma^2\right)\right] &= \\
\int_{-\infty}^{+\infty} \frac{x^2}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx &
\end{aligned}$$

Introducing

$$y \coloneqq \frac{1}{2\sigma^2} > 0 \tag{6}$$

we get

$$\begin{aligned}
\sqrt{\frac{y}{\pi}} \int_{-\infty}^{+\infty} x^2 e^{-yx^2} dx &= \\
\sqrt{\frac{y}{\pi}} \frac{\partial}{\partial y}\left[-\int_{-\infty}^{+\infty} e^{-yx^2} dx\right] &= \\
-\sqrt{\frac{y}{\pi}} \frac{\partial}{\partial y}\sqrt{\frac{\pi}{y}} &= \\
\left(-\frac{y^{\frac{1}{2}}}{\sqrt{\pi}}\right)\left(\frac{\partial}{\partial y}\sqrt{\pi}y^{-\frac{1}{2}}\right) &= \\
\left(-y^{\frac{1}{2}}\right)\left(-\frac{1}{2}y^{-\frac{3}{2}}\right) &= \\
\frac{1}{2y} &= \sigma^2
\end{aligned}$$

□ *q.e.d.*

Note that again we made use of the integral derived from eq. (3), this time with the additional factor $y > 0$, which is very easy to show to appear in the denominator, and again landing under the squared root.

fmahlknecht@student.ethz.ch
19-945-351

1.2 Cumulative distribution function

Given the Laplace distribution parameterized by μ and β:

$$f_X(x) = \frac{1}{2\beta} e^{-\frac{|x-\mu|}{\beta}} \tag{7}$$

To get the cumulative distribution function, we apply the definition:

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(z)\, dz = \\ &= \int_{-\infty}^{x} \frac{1}{2\beta} e^{-\frac{|z-\mu|}{\beta}} dz \end{aligned}$$

To explicitly express the absolute value in the integrand, we may split it up the integral until μ, for now assuming $x > \mu$:

$$F_X(x) = \int_{-\infty}^{\mu} \frac{1}{2\beta} e^{-\frac{|z-\mu|}{\beta}} dz + \int_{\mu}^{+\infty} \frac{1}{2\beta} e^{-\frac{|z-\mu|}{\beta}} dz$$

In the first integral we have $z \leq \mu$ and in the second $z \geq \mu$. Therefore we may simplify the absolute value as follows:

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{\mu} \frac{1}{2\beta} e^{-\frac{-(z-\mu)}{\beta}} dz + \int_{\mu}^{x} \frac{1}{2\beta} e^{-\frac{z-\mu}{\beta}} dz = \\ &= \frac{1}{2} e^{\frac{z-\mu}{\beta}} \Big|_{-\infty}^{\mu} - \frac{1}{2} e^{-\frac{z-\mu}{\beta}} \Big|_{\mu}^{x} \\ &= \frac{1}{2} - \frac{1}{2}\left(e^{-\frac{x-\mu}{\beta}} - 1\right) \\ &= \frac{1}{2} + \frac{1}{2}\left(1 - e^{-\frac{x-\mu}{\beta}}\right) \qquad x \geq \mu \end{aligned} \tag{8}$$

For the case in which $x < \mu$ we have to subtract from the first integral, evaluating to $\frac{1}{2}$, the area which we added in advance, using the very same integrating condition, $z \leq \mu$:

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{\mu} \frac{1}{2\beta} e^{-\frac{-(z-\mu)}{\beta}} dz - \int_{x}^{\mu} \frac{1}{2\beta} e^{-\frac{-(z-\mu)}{\beta}} dz = \\ &= \frac{1}{2} - \frac{1}{2} e^{\frac{z-\mu}{\beta}} \Big|_{x}^{\mu} = \\ &= \frac{1}{2} - \frac{1}{2}\left(1 - e^{\frac{x-\mu}{\beta}}\right) \qquad x \leq \mu \end{aligned} \tag{9}$$

Bringing eqs. (8) and (9) together in the same expression is straight forward by using the absolute value and sign function:

$$F_X(x) = \frac{1}{2} + \operatorname{sgn}(x-\mu)\frac{1}{2}\left(1 - e^{-\frac{|x-\mu|}{\beta}}\right) \tag{10}$$

Florian Mahlknecht
fmahlknecht@student.ethz.ch
19-945-351

To find the median of the Laplace distribution it is enough to impose

$$F_X(M) = \frac{1}{2} \tag{11}$$

since this is exactly where the 50 % quantil is found.

$$\frac{1}{2} + \operatorname{sgn}(M-\mu)\frac{1}{2}\left(1 - e^{-\frac{|M-\mu|}{\beta}}\right) = \frac{1}{2}$$
$$\iff M - \mu = 0 \Rightarrow M = \mu$$

1.3 Quotient probability distribution

The probability distribution of the quotient $Q = \frac{X}{Y}$ is given by:

$$f_Q(q) = \int_{-\infty}^{+\infty} |x| \, f_{X,Y}(px, x) dx \tag{12}$$

We furthermore know that $X \sim \mathcal{N}(0, \sigma_X^2)$ and $Y \sim \mathcal{N}(0, \sigma_Y^2)$.

Joint probability The joint probability is simply given by the product, i.e.:

$$f_{X,Y}(x,y) = \frac{1}{\sqrt{2\pi}\sigma_X} \frac{1}{\sqrt{2\pi}\sigma_Y} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_X^2} + \frac{y^2}{\sigma_Y^2}\right)}$$
$$= \frac{1}{2\pi \, \sigma_X \, \sigma_Y} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_X^2} + \frac{y^2}{\sigma_Y^2}\right)}$$

Cauchy distribution Using this to express in more explicit terms eq. (12), we get:

$$f_Q(q) = \int_{-\infty}^{+\infty} \frac{|x|}{2\pi \, \sigma_X \, \sigma_Y} e^{-\frac{1}{2}\left(\frac{q^2 x^2}{\sigma_X^2} + \frac{x^2}{\sigma_Y^2}\right)} dx$$
$$= \int_{-\infty}^{+\infty} \frac{|x| \, \gamma}{2\pi \, \sigma_X^2} e^{-\frac{x^2}{2\sigma_X^2}(q^2+\gamma^2)} dx$$

with

$$\gamma = \frac{\sigma_X}{\sigma_Y} \tag{13}$$

For simplicity we introduce $t :=$, therefore:

$$f_Q(q) = \frac{\gamma}{2\pi \, \sigma_X^2} \int_{-\infty}^{+\infty} |x| \, e^{-x^2 \frac{q^2+\gamma^2}{2\sigma_X^2}} dx$$

By symmetry we have

fmahlknecht@student.ethz.ch
19-945-351

$$
\begin{aligned}
f_Q(q) &= \frac{\gamma\, 2}{2\pi\, \sigma_X^2} \int_0^{+\infty} |x|\, e^{-x^2 \frac{q^2+\gamma^2}{2\sigma_X^2}}\, dx \\
&= \frac{\gamma}{\pi\, \sigma_X^2} \int_0^{+\infty} x\, e^{-x^2 \frac{q^2+\gamma^2}{2\sigma_X^2}}\, dx \\
&= -\frac{\gamma}{\pi\, \sigma_X^2} \frac{\sigma_X^2}{q^2+\gamma^2} e^{-x^2 \frac{p^2+\gamma^2}{2\sigma_X^2}} \Bigg|_0^{+\infty} \\
&= \frac{\gamma}{\pi\, \sigma_X^2} \frac{\sigma_X^2}{q^2+\gamma^2} \\
&= \frac{1}{\pi} \frac{\gamma}{q^2+\gamma^2}
\end{aligned}
$$

Which is indeed equal to the Chauchy distribution with the introduced scale $\gamma = \frac{\sigma_X}{\sigma_Y}$ and $x_0 = 0$:

$$
\frac{1}{\pi} \frac{\gamma}{(x-x_0)^2+\gamma^2} \Bigg|_{x_0=0} = f_Q(x) = \frac{1}{\pi} \frac{\gamma}{x^2+\gamma^2} \tag{14}
$$

$\square$ $q.e.d.$

2 Bayesian Inference

Likelihood function The likelihood function of μ is defined as:

$$
\mathcal{L}(\mu) := p(\mathbf{d}|\mu) \tag{15}
$$

Using the fact that $\mathbf{d}$ are realizations of *independent* random variables, normally distributed around μ and with $\sigma = 1$, we can extend the expression into:

$$
p(\mathbf{d}|\mu) = \prod_{k=1}^{N} p(d_k|\mu) = \prod_{k=1}^{N} \frac{1}{\sqrt{2\pi}} e^{-\frac{(d_k-\mu)^2}{2}} \tag{16}
$$

Maximum likelihood estimate To find the maximum likelihood estimate for the parameter μ we compute:

$$
\hat{\mu} = \arg\max_{\mu} \mathcal{L}(\mu) = \arg\max_{\mu} \log\left(\mathcal{L}(\mu)\right) \tag{17}
$$

Since the logarithm is a strictly monotonically increasing function. Applying this yields

$$
\begin{aligned}
\hat{\mu} &= \arg\max_{\mu} \log \prod_{k=1}^{N} \frac{1}{\sqrt{2\pi}} e^{-\frac{(d_k-\mu)^2}{2}} \\
&= \arg\max_{\mu} \sum_{k=1}^{N} \log\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{(d_k-\mu)^2}{2}} \right) \\
&= \arg\max_{\mu} \sum_{k=1}^{N} \left[\log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{(d_k-\mu)^2}{2} \right]
\end{aligned}
$$

Florian Mahlknecht
fmahlknecht@student.ethz.ch
19-945-351

For $\hat{\mu}$ to be a maximum we must have:

$$\frac{\partial}{\partial \mu}\mathcal{L}(\hat{\mu}) = 0 \tag{18}$$

$$\frac{\partial^2}{\partial \mu^2}\mathcal{L}(\hat{\mu}) < 0 \tag{19}$$

This yields:

$$\sum_{k=1}^{N}(d_k - \hat{\mu}) = 0 \quad \Rightarrow \quad \sum_{k=1}^{N} d_k - N\,\hat{\mu} = 0 \quad \Rightarrow \quad \hat{\mu} = \frac{1}{N}\sum_{k=1}^{N} d_k$$

$$\sum_{k=1}^{N} -1 = -N < 0$$

This shows that the arithmetic average is the best estimate for the mean of the distribution (as expected).

Posterior distribution Before observing the data we believe that $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$, which leads to the following prior:

$$p(\mu) = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}} \tag{20}$$

By using Bayes' theorem (for simplicity without regularization term) we have:

$$p(\mu|\mathbf{d}) \propto p(\mathbf{d}|\mu)\, p(\mu) \tag{21}$$

Substituting eq. (16) and eq. (20) yields

fmahlknecht@student.ethz.ch
19-945-351

$$
\begin{aligned}
p(\mathbf{d}|\mu) &\propto \left[\prod_{k=1}^{N} \frac{1}{\sqrt{2\pi}} e^{-\frac{(d_k-\mu)^2}{2}}\right] \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}} \qquad (22)\\
&\propto \exp\left(-\frac{1}{2}\left(\frac{\sum_{k=1}^{N}(d_k-\mu)^2}{1} + \frac{(\mu-\mu_0)^2}{\sigma_0^2}\right)\right)\\
&= \exp\left(-\frac{1}{2}\frac{\sum_{k=1}^{N}\sigma_0^2(\mu-d_k)^2 + (\mu-\mu_0)^2}{\sigma_0^2}\right)\\
&= \exp\left(-\frac{1}{2}\frac{(N\sigma_0^2+1)\,\mu^2 - 2\mu\left(\sum_{k=1}^{N}\sigma_0^2\, d_k^2 + \mu_0\right) + \sum_{k=1}^{N}\sigma_0^2 d_k^2 + \mu_0^2}{\sigma_0^2}\right)\\
&= \exp\left(-\frac{1}{2}\frac{\mu^2 - 2\mu\left(\frac{\sum_{k=1}^{N}\sigma_0^2\, d_k^2+\mu_0}{N\sigma_0^2+1}\right) + g(d_k,\mu_0)}{\frac{\sigma_0^2}{N\sigma_0^2+1}}\right)\\
&= \exp\left(-\frac{1}{2}\frac{\left(\mu - \frac{\sum_{k=1}^{N}\sigma_0^2\, d_k^2+\mu_0}{N\sigma_0^2+1}\right)^2 + \tilde{g}(d_k,\mu_0)}{\frac{\sigma_0^2}{N\sigma_0^2+1}}\right)\\
&\propto \exp\left(-\frac{\left(\mu - \frac{\sum_{k=1}^{N}\sigma_0^2\, d_k^2+\mu_0}{N\sigma_0^2+1}\right)^2}{2\,\frac{\sigma_0^2}{N\sigma_0^2+1}}\right)
\end{aligned}
$$

where $g(d_k, \mu_0)$ and $\tilde{g}(d_k, \mu_0)$ are used to complete the square. After reshaping the equation in this way it is easy to see, that we have obtained the following normal distribution:

$$
p(\mu|\mathbf{d}) \sim \mathcal{N}\left(\frac{\sum_{k=1}^{N}\sigma_0^2\, d_k^2 + \mu_0}{N\sigma_0^2+1}, \frac{\sigma_0^2}{N\sigma_0^2+1}\right) \qquad (23)
$$

with the given mean and variance.

The obtained distribution intuitively makes sense, since the more data we collect, the less important becomes prior μ_0, depending on the variance σ_0 we expect.

Maximum a posteriori estimate

$$
\hat{\mu} = \arg\max_{\mu} p(\mu|\mathbf{d}) = \arg\max_{\mu} \log\left(p(\mu|\mathbf{d})\right) \qquad (24)
$$

As before we check

$$
\frac{\partial}{\partial\mu} p(\hat{\mu}|\mathbf{d}) = 0 \qquad (25)
$$

$$
\frac{\partial^2}{\partial\mu^2} p(\hat{\mu}|\mathbf{d}) < 0 \qquad (26)
$$

Applying the logarithm in the same manner as before, we simply get (omitting the positive, constant denominator in the exponent):

Florian Mahlknecht
fmahlknecht@student.ethz.ch
19-945-351

$$-\left(\hat{\mu} - \frac{\sum_{k=1}^{N}\sigma_0^2 d_k^2 + \mu_0}{N\sigma_0^2 + 1}\right) = 0 \quad \Rightarrow \quad \hat{\mu} = \frac{\sum_{k=1}^{N}\sigma_0^2 d_k^2 + \mu_0}{N\sigma_0^2 + 1}$$

$$-1 < 0$$

And we have that the mean is again a maximum and the best estimate for a normal distribution.

Uninformative prior If we use an uninformative prior, i.e. a uniform distribution in $\mathbb{R}$ with all values equally likely, we have:

$$p(\mu) \propto 1 \tag{27}$$

and then the Bayes' theorem from eq. (21) reduces to

$$p(\mu|\mathbf{d}) \propto p(\mathbf{d}|\mu) \tag{28}$$

Recalling the maximum likelihood estimate, and the maximum a posteriori estimate from before:

$$\hat{\mu}_{MLE} = \arg\max_{\mu} \mathcal{L}(\mu) = \arg\max_{\mu} p(\mathbf{d}|\mu)$$
$$\hat{\mu}_{MPE} = \arg\max_{\mu} p(\mu|\mathbf{d})$$

When looking at eq. (28), we can conclude that there is no difference anymore:

$$\hat{\mu}_{MLE} = \hat{\mu}_{MPE} \tag{29}$$

Which intuitively makes sense: if we do not have any guess or assumption about possible parameters for a model of the data, correcting this model with the posteriori estimate, simply means completely relying on the measured data, i.e. adopting the maximum likelihood of the parameters.

3 Bayesian Inference: Linear Model

Given the model

$$y = \beta x + \varepsilon \qquad \varepsilon \sim \mathcal{N}(0, \sigma) \tag{30}$$

we observe

$$\mathbf{D} = (x_0, y_0)$$

The full Bayesian' theorem yields:

$$p(\beta\,|\,\mathbf{D}) = \frac{p(\mathbf{D}\,|\,\beta)\,p(\beta)}{p(\mathbf{D})} \tag{31}$$

fmahlknecht@student.ethz.ch
19-945-351

Uninformative prior As seen before an uninformative prior means:

$$p(\beta) \propto 1 \tag{32}$$

Therefore, eq. (31), starting while omitting the regularization term again, becomes:

$$p(\beta \,|\, \mathbf{D}) \propto p(\mathbf{D} \,|\, \beta) \tag{33}$$

This can be developed as follows:

$$\begin{aligned} p(\mathbf{D} \,|\, \beta) &= p(x_0, y_0 \,|\, \beta) \\ &= p(y_0 \,|\, x_0, \beta) \\ &= p(y_0 \,|\, x_0, \beta)\, p(x_0 \,|\, \beta) \\ &\propto p(y_0 \,|\, x_0, \beta) \\ &= p(\varepsilon = y_0 - \beta\, x_0 \,|\, x_0, \beta) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_0 - \beta\, x_0 - 0)^2}{2\sigma^2}} \end{aligned}$$

In order to get now the posterior probability distribution, we simply rearrange the terms, in order to get a normal distribution in β:

$$p(\beta \,|\, \mathbf{D}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_0 - \beta\, x_0)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\beta - \frac{y_0}{x_0})^2}{2\left(\frac{\sigma}{x_0}\right)^2}} = \mathcal{N}\left(\frac{y_0}{x_0}, \left(\frac{\sigma}{x_0}\right)^2\right) \tag{34}$$

The maximum posterior estimate, again for a normal distribution as shown already before, see section 2, is simply the mean, i.e.:

$$\hat{\beta}_{MPE} = \frac{y_0}{x_0} \tag{35}$$

which again, given that our prior is uninformative (see eq. (32)), is the same as the maximum likelihood estimate. Note that this is not surprising at all, since we have an unbiased model error (with $\mu = 0$), after observing just one datapoint, having no informative prior, it is simply the quotient, making fit perfectly the model to the only information (i.e. datapoint) we have.

Needless to say, the variance of $p(\beta \,|\, \mathbf{D})$, is $\left(\frac{\sigma}{x_0}\right)^2$.

Bayesian linear regression Now assuming as a prior $\beta \sim \mathcal{N}(0, \tau^2)$, we get:

$$p(\beta \,|\, \mathbf{D}) \propto p(\mathbf{D} \,|\, \beta)\, p(\beta) \tag{36}$$

Using the likelihood from before we can express explicitly:

$$p(\beta \,|\, \mathbf{D}) \propto \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\beta - \frac{y_0}{x_0})^2}{2\left(\frac{\sigma}{x_0}\right)^2}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\beta)^2}{2\tau^2}} \tag{37}$$

Which is a simpler version of the case we faced before in eq. (22), with $N = 1$. So from the 'multiplication' of the two distributions:

Florian Mahlknecht
fmahlknecht@student.ethz.ch
19-945-351

$$\mathcal{N}\left(\frac{y_0}{x_0}, \left(\frac{\sigma}{x_0}\right)^2\right) \qquad \mathcal{N}\left(0, \tau^2\right)$$

by using the results from before, illustrated in eq. (23), we get:

$$p(\beta \mid \mathbf{D}) \propto \mathcal{N}\left(\frac{\frac{y_0}{x_0}\,\tau^2}{\tau^2 + \left(\frac{\sigma}{x_0}\right)^2}, \frac{\left(\frac{\sigma}{x_0}\right)^2 \tau^2}{\left(\frac{\sigma}{x_0}\right)^2 + \tau^2}\right) \tag{38}$$

fmahlknecht@student.ethz.ch
19-945-351