Mackay ex 29.15 mathematical background to implementation

<u>Note</u>: The implementation can be found within the Github as "29-15.py"

We want to sample from $p(\mu, \beta | D)$
where $\beta = 1/\sigma^2$, and we assume the following priors:
$p(\mu) = \mathcal{N}(\mu | 0, \sigma_\mu^2)$ (with σ_μ a to be set parameter)
and $p(\beta) = 1/\beta$.
For Gibbs sampling we need to derive:
$P(\mu | \beta, D)$ and $P(\beta | \mu, D)$

$$p(\mu | \beta, D) = \frac{p(\mu)\, p(\beta, D | \mu)}{p(D, \beta)} = \frac{p(\mu)\, p(D | \mu, \beta)\, \overbrace{p(\beta | \mu)}^{= p(\beta)}}{\int d\mu\, p(D | \mu, \beta)\, p(\beta)\, p(\mu)}$$

$$= \frac{p(\mu)\, p(D | \mu, \beta)}{\int d\mu\, p(\mu)\, p(D | \mu, \beta)}$$

This distribution is fully defined by the mean and variance/std of the nominator, since we know that the denominator will take of all the neccesary normalization

$$p(\mu)\, p(D | \mu, \beta) = \underbrace{p(\mu)}_{= \mathcal{N}(\mu | 0, \sigma_\mu) \text{ ("Broad Gaussian")}} \prod_{i=1}^{N} \mathcal{N}(x_i | \mu, \sigma = \tfrac{1}{\sqrt{\beta}})$$

$$\sim \exp\left(\frac{-1}{2\sigma_\mu^2} \mu^2\right) \exp\left(\sum_{i=1}^{N} \frac{-1}{2\sigma^2} (x_i - \mu)^2\right)$$

$$= \exp\left(\frac{-\mu^2}{2\sigma_\mu^2} + \left(\frac{-1}{2\sigma^2} \sum_{i=1} x_i^2\right) + \sum_{i=1}^{N} \frac{2 x_i \mu}{2\sigma^2} - \sum_{i=1}^{N} \frac{\mu^2}{2\sigma^2}\right)$$

We can neglect all terms independent of μ since normalization will take care of those.

$$\sim \exp\left(\mu \sum_{i=1}^{N} \frac{x_i}{\sigma^2} + \mu^2 \left(\frac{-1}{2\sigma_\mu^2} - \frac{N}{2\sigma^2}\right)\right)$$

From here we denote:

$$C_1 = \sum_{i=1}^{N} x_i / \sigma^2 \quad \text{and} \quad C_2 = -\left(\frac{1}{2\sigma_\mu^2} + \frac{N}{2\sigma^2}\right)$$

$$\rightarrow \exp(C_1\mu + C_2\mu^2)$$

$$= \exp\left(C_2\left(\left(\frac{-C_1}{2C_2} - \mu\right)^2 - \left(\frac{C_1^2}{4C_2^2}\right)\right)\right)$$

Once again neglect terms indep. of μ.

$$\sim \exp\left(C_2\left(\frac{-C_1}{2C_2} - \mu\right)^2\right)$$

From this we derive:

$$\mu_s = \frac{-C_1}{2C_2} = \frac{\frac{1}{\sigma^2}\sum_{i=1}^{N} x_i}{\frac{1}{\sigma_\mu^2} + \frac{N}{\sigma^2}} = \frac{\sum_{i=1}^{N} x_i}{\frac{\sigma^2}{\sigma_\mu^2} + N} = \frac{\sum_{i=1}^{N} x_i}{\frac{1}{\beta\sigma_\mu^2} + N}$$

$$\frac{-1}{2\sigma_s^2} = C_2 \rightarrow \sigma_s^2 = \frac{1}{\frac{1}{\sigma_\mu^2} + \beta N}$$

$\Rightarrow$ To sample from $p(\mu \mid \beta, D)$ we can sample from:

$$\mathcal{N}(\mu \mid \mu_s, \sqrt{\sigma_s^2})$$

$p(\beta \mid \mu, D)$:

$$= \frac{p(\beta)\, p(\mu, D \mid \beta)}{p(\mu, D)} \quad \text{we similarly find:} = \frac{p(\beta)\, p(D \mid \mu, \beta)}{\int d\beta\, p(\beta)\, p(D \mid \mu, \beta)}$$

$$p(\beta)\, p(D \mid \mu, \beta) = \frac{1}{\beta} \exp\left(\frac{\beta}{2}\sum_{i=1}^{N}(x_i - \mu)^2\right)\left(\sqrt{\frac{\beta}{2\pi}}\right)^N$$

$$\sim \beta^{N/2 - 1} \exp\left(\beta \sum_{i=1}^{N}(x_i - \mu)^2 / 2\right)$$

Given that, again, the denominator takes care of the needed normalisation we can conclude that

$$p(\beta \mid \mu, D) = \text{Gamma}(\beta \mid k_\beta, \theta_\beta)$$

where $k_\beta = N/2$ and $\theta_\beta = \dfrac{2}{\sum_{i=1}^{N}(x_i - \mu)^2}$

And thus we can now use Gibbs sampling to infere μ and β of the real data.