Example Sheet 1

# 1 Bayes' theorem

The average German woman has 1.36 children and her average lifetime is 82 years. 48.7% of the German population are women.

(a) Give an estimate of $P(\text{pregnant}|\text{female})$ for Germany. Compare to $P(\text{female}|\text{pregnant})$.

(b) Use Bayes' theorem to give an estimate of $P(\text{pregnant})$ and of $P(\text{pregnant}, \text{female})$ (mind the comma!).

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| $y$ | 0 | 1 | 2 | 3 | 5 | 6 | 7 | 8 |
| $\sigma_y$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

Table 1: Example data to be used for this exercise sheet.

# 2 Design matrix

(a) Show that $\frac{1}{ac-b^2}\begin{pmatrix} c & -b \\ -b & a \end{pmatrix}$ is the inverse matrix of $\begin{pmatrix} a & b \\ b & c \end{pmatrix}$.

(b) Let $D \cdot \vec{\theta}$ be a linear model of $\vec{y}$. Show that if all measurement errors are identical and uncorrelated, i.e., $\Sigma = \sigma^2 I$, then the maximum-likelihood solution reduces to the least-squares solution $\hat{\vec{\theta}} = (D^T \cdot D)^{-1} \cdot D^T \cdot \vec{y}$ where all measurement errors cancel out.

(c) For a straight-line model $f(x) = a_0 + a_1 x$ and the data given in Table 1, fill in the missing values of the design matrix:

$$D = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \end{pmatrix}$$

(d) Given the above design matrix, evaluate the least-squares solution $\hat{\vec{\theta}} = (D^T \cdot D)^{-1} \cdot D^T \cdot \vec{y}$ *by hand*. (Do *not* use a computer here! For the matrix inversion, use the result of exercise 2a.) Draw a sketch of the data given in Table 1 and the best-fit model *by hand* (no print-out!).

(e) Interpret what kind of model $f(x)$ corresponds to the following design matrix:

$$D = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 2 \\ 1 & 0 & 3 \\ 0 & 1 & 4 \\ 0 & 1 & 5 \\ 0 & 1 & 6 \\ 0 & 1 & 7 \end{pmatrix}$$

(Keep your discussion short! Try *not* to run for the Nobel Prize in Literature!)

# 3 Tikhonov regularisation

Let the log-posterior with conjugate prior be given by:

$$\log P(\theta|D) = -\frac{1}{2}\left(\vec{y} - X \cdot \vec{\theta}\right)^T \cdot \Sigma^{-1} \cdot \left(\vec{y} - X \cdot \vec{\theta}\right) - \frac{1}{2}\left(\vec{z} - Z \cdot \vec{\theta}\right)^T \cdot \Sigma_P^{-1} \cdot \left(\vec{z} - Z \cdot \vec{\theta}\right)$$

This is the most general form of a conjugate prior for a linear model with Gaussian errors and it is called "Tikhonov regularisation". Let the model be a polynomial of order up to 5, i.e.

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5$$

Furthermore, let

$$\vec{z} = \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} \qquad Z = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix} \qquad \Sigma_P = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0.1 \\ 0 & 0.1 & 1 \end{pmatrix}$$

be given.

(a) Describe in words the prior constraints on all six model parameters $\vec{\theta} = (a_0, a_1, a_2, a_3, a_4, a_5)^T$. Ignore $\Sigma_P$ for the moment!

(b) Now look at $\Sigma_P$ and explain in words how the different prior constraints are related to each other.

(c) Show that the gradient of the log-posterior is given by:

$$\vec{\nabla}_\theta \log P(\theta|D) = X^T \cdot \Sigma^{-1} \cdot \left(\vec{y} - X \cdot \vec{\theta}\right) - Z^T \cdot \Sigma_P^{-1} \cdot \left(\vec{z} - Z \cdot \vec{\theta}\right)$$

(d) Set the gradient to zero and solve for the maximum a-posteriori estimate $\vec{\theta}_{\mathrm{MPE}}$.

# 4 Convolution

(a) Let the random variate $x$ be drawn from $P_X(x)$ and let the random variate $y$ be drawn from $P_Y(y)$. The random variate $z = x + y$ is subject to $P_Z(z)$. Show that $P_Z(z)$ is given by the so-called convolution integral:

$$P_Z(z) = \int P_X(z-y) P_Y(y) dy$$

Convolution is mathematically denoted as $P_Z = P_X * P_Y$.

(b) Remember that the characteristic function is the Fourier transform of the PDF, i.e., $P_X(x) = \int \phi_X(u) e^{iux} du$, $P_Y(y) = \int \phi_Y(u) e^{iuy} du$, and $P_Z(z) = \int \phi_Z(u) e^{iuz} du$. Plug these into the convolution integral and derive the so-called convolution theorem:

$$\phi_Z(u) = \phi_X(u) \phi_Y(u)$$

You may find the identity $\int e^{i(u'-u)y} dy = \delta(u'-u)$ useful. (In words: If $P_Z$ is given by the convolution integral of $P_X$ and $P_Y$, then the characteristic function $\phi_Z$ is given by the simple product of $\phi_X$ and $\phi_Y$.)

(c) The Gaussian with mean $\mu$ and standard deviation $\sigma$ has the characteristic function $\phi_G(u;\mu,\sigma) = \exp\left[i\mu u - \frac{1}{2}\sigma^2 u^2\right]$. Show that the Gaussian is "invariant under convolution", i.e., that the convolution of two Gaussians with means $\mu_1$ and $\mu_2$ and standard deviations $\sigma_1$ and $\sigma_2$ is again a Gaussian. Use the characteristic function to show that and to derive the mean and standard deviation of the resulting convolved Gaussian.

(d) The Cauchy distribution with mean $\mu$ and width $\sigma$ has the characteristic function $\phi_C(u;\mu,\sigma) = \exp\left[i\mu u - \sigma|u|\right]$. Show that the Cauchy distribution is invariant under convolution, too, and also derive mean and width of the resulting convolved Cauchy distribution.

(e) Invariance under convolution is a very useful property of PDFs, since the sums of random variates have PDFs that can be easily computed. Distributions that are invariant under convolution therefore deserve a special name: They are called *stable* distributions. A special class (there are more than one) of stable distributions are the so-called $\alpha$-stable distributions, which have the characteristic function:

$$\phi_S(u;\mu,\sigma,\alpha,\beta) = \exp\left[iu\mu - \sigma^\alpha|u|^\alpha\left(1 - i\beta\,\mathrm{sgn}(t)\Phi(\alpha)\right)\right]$$

Where $\mu \in \mathbb{R}$ is a location, $\sigma > 0$ is a width, $\beta \in [-1,1]$ is the skewness/asymmetry and $\alpha \in (0,2]$ is called the stability parameter. $\Phi(\alpha)$ is a function of $\alpha$ but not of $u$, so we do not care about it here. Show that $\phi_S(u;\mu_1,\sigma_1,\alpha,\beta_1)\phi_S(u;\mu_2,\sigma_2,\alpha,\beta_2)$ for *identical* $\alpha$ (!) is again a stable distribution. Derive expressions for $\mu$, $\sigma$ and $\beta$ of the resulting convolved stable distribution. Give values of $\beta$ and $\alpha$ for the Gaussian distribution and for the Cauchy distribution by comparing to results from 2c and 2d.[1]

[1] From the whole class of stable distributions, there are only 4 choices of $\beta$ and $\alpha$ that result in a characteristic function that has an analytic PDF. Two of these 4 choices are the Gaussian and the Cauchy. (The other two choices with analytic PDF are known as Lévy distribution and inverse Lévy distribution.) In all other cases, the PDF has to be evaluated numerically (by FFT) from the characteristic function. Stable distributions are used, e.g., to model the time evolution of stock prices (with limited success as you may have guessed).