Hierarchical Bayesian Models: Introduction

Václav Šmídl

February 14, 2022

What it means to be a Bayesian?

What it means to be a Bayesian?

Bayesian is someone who uses probability to quantify uncertainty

What it means to be a Bayesian?

Bayesian is someone who uses probability to quantify uncertainty

Uncertainty types:

Aleatoric (random): different outcome for different run of an experiment

- rolling a dice

Epistemic (systematic): something we could know but do not in practice. Due to insufficient data, missing measurement, incomplete model, etc.

- distance of a star in the universe.

What it means to be a Bayesian?

Bayesian is someone who uses probability to quantify uncertainty

Uncertainty types:

Aleatoric (random): different outcome for different run of an experiment

- rolling a dice

Epistemic (systematic): something we could know but do not in practice. Due to insufficient data, missing measurement, incomplete model, etc.

- distance of a star in the universe.

Bayesian is someone who see natural that distance of a star has Normal distribution with mean 10ly and std 1ly.

What it means to be a Bayesian?

Bayesian is someone who uses probability to quantify uncertainty

Uncertainty types:

Aleatoric (random): different outcome for different run of an experiment

- rolling a dice

Epistemic (systematic): something we could know but do not in practice. Due to insufficient data, missing measurement, incomplete model, etc.

- distance of a star in the universe.

Bayesian is someone who see natural that distance of a star has Normal distribution with mean 10ly and std 1ly.

- probability distribution represents a *degree of belief*.

Course layout

Aim: practical use of general methodology

- show a range of models and their specifics
- minimum level of formality – focus on key aspects

Literature: Bishop, Ch.M. Pattern recognition and machine learning. Springer, 2006.

Details: scholar.google.com, wikipedia

Marks: homeworks worth 110 points in total

	points
A	>85
B	>70
C	>55

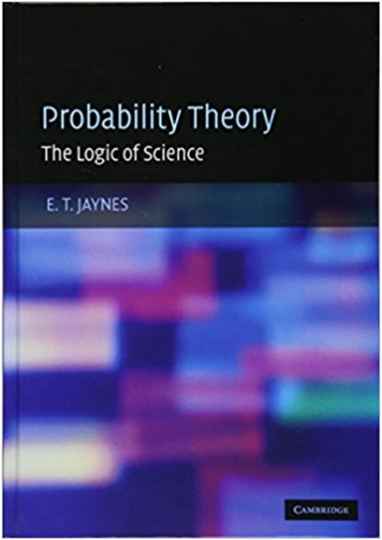
Probability theory

Probability theory as:

- extension of logic
- language
- necessity for making decisions under uncertainty

Alternatives:

- Dempster-Shafer
- fuzzy logic



Probability calculus discrete

Random variables:

$$X \in \{x_1, \dots, x_M\}$$
$$Y \in \{y_1, \dots, y_L\}$$

Joint probability

$$P(X = x_i, Y = y_j) = \frac{n_{i,j}}{N}$$

where N ($N \to \infty$) is the number of realizations and $n_{i,j}$ is the number of trials where $X = x_i$, $Y = y_j$.

Rules:

1. sum rule

$$P(X = x_i) = \sum_{j=1}^{L} P(X = x_i, Y = y_i),$$

2. product rule

$$P(X, Y) = p(Y|X)p(X)$$

Bayes rule

Named after reverend Thomas Bayes, the Bayes rule is a consequence of the product rule:

$$
\begin{aligned}
P(X, Y) &= p(Y|X)p(X), \\
&= p(X|Y)p(Y),
\end{aligned}
$$

yielding:

$$p(Y|X)p(X) = p(X|Y)p(Y),$$

with

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$

Cancer example

- Approximately 1% of women aged 40-50 have breast cancer.
- A woman with breast cancer has a 90% chance of a positive test.
- A woman without cancer has a 10% chance of a false positive result.

What is the probability a woman has breast cancer given that she just had a positive test?

Cancer example

- Approximately 1% of women aged 40-50 have breast cancer.
- A woman with breast cancer has a 90% chance of a positive test.
- A woman without cancer has a 10% chance of a false positive result.

What is the probability a woman has breast cancer given that she just had a positive test?

- X =1 if a woman has cancer
- Y =1 if the test is positive

We want to know

$$P(X = 1|Y = 1) = \frac{P(Y|X)P(X)}{P(Y)}$$

Cancer example

- Approximately 1% of women aged 40-50 have breast cancer.
- A woman with breast cancer has a 90% chance of a positive test.
- A woman without cancer has a 10% chance of a false positive result.

What is the probability a woman has breast cancer given that she just had a positive test?

- X =1 if a woman has cancer
- Y =1 if the test is positive

We want to know

$$P(X=1|Y=1) = \frac{P(Y|X)P(X)}{P(Y)}$$

$$\begin{aligned}
&P(Y=1|X=1) = 0.9,\\
&P(X=1) = 0.01,\\
&P(Y) = \textstyle\sum_X P(Y|X)P(X) = \\
&\qquad P(Y|X=1)P(X=1)+\\
&\qquad P(Y|X=0)P(X=0)\\
&\quad = 0.9 * 0.01 + 0.1 * 0.99 = 0.108
\end{aligned}$$

$$P(X=1|Y=1) = \tfrac{0.009}{0.108} = 8.3\%$$

Probability calculus continuous

Random variable: X, realization x
Probability density function: $p_X(x)$,

Probability calculus continuous

Random variable: X, realization x
Probability density function: $p_X(x)$, $\int p_X(x)dx = 1$

Probability calculus continuous

Random variable: X, realization x
Probability density function: $p_X(x)$, $\int p_X(x)dx = 1$
Cumulative density function: $P(X < x) = \int_{-\infty}^{x} p_X(t)dt$

Probability calculus continuous

Random variable: X, realization x
Probability density function: $p_X(x)$, $\int p_X(x)dx = 1$
Cumulative density function: $P(X < x) = \int_{-\infty}^{x} p_X(t)dt$
Rules:

1. sum rule
$$p(x) = \int p(x,y)dy$$
2. product rule
$$p(x,y) = p(y|x)p(x) = p(x|y)p(y)$$
3. change of variables $y = g(x)$
$$p_Y(y) = p_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

Probability calculus continuous

Random variable: X, realization x
Probability density function: $p_X(x)$, $\int p_X(x)dx = 1$
Cumulative density function: $P(X < x) = \int_{-\infty}^{x} p_X(t)dt$
Rules:

1. sum rule

$$p(x) = \int p(x, y)dy$$

2. product rule

$$p(x, y) = p(y|x)p(x) = p(x|y)p(y)$$

3. change of variables $y = g(x)$

$$p_Y(y) = p_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

Engineering (ML) notation: meaning given by context.

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

Transformation of variables

Given:

$$p(x) = \mathcal{N}(0, 1), \qquad\qquad y = ax + b,$$

find $p(y)$.

Transformation of variables

Given:

$$p(x) = \mathcal{N}(0, 1), \qquad\qquad y = ax + b,$$

find $p(y)$.

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$$

Transformation of variables

Given:

$$p(x) = \mathcal{N}(0,1), \qquad\qquad y = ax + b,$$

find $p(y)$.

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \qquad\qquad \int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}.$$

Transformation of variables

Given:

$$p(x) = \mathcal{N}(0,1), \qquad\qquad y = ax + b,$$

find $p(y)$.

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \qquad\qquad \int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}.$$

$$x = \frac{y-b}{a} \implies \frac{d}{dy} g^{-1}(y) = \frac{1}{a}$$

Transformation of variables

Given:

$$p(x) = \mathcal{N}(0,1), \qquad\qquad y = ax + b,$$

find $p(y)$.

$$\begin{aligned} p(x) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \qquad\qquad \int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}. \\ x &= \frac{y-b}{a} \quad \Longrightarrow \quad \frac{d}{dy} g^{-1}(y) = \frac{1}{a} \\ p(y) &= \frac{1}{\sqrt{2\pi}a} \exp\left(-\frac{1}{2}\left(\frac{y-b}{a}\right)^2\right) \end{aligned}$$

Transformation of variables

Given:

$$p(x) = \mathcal{N}(0,1), \qquad y = ax + b,$$

find $p(y)$.

$$\begin{aligned} p(x) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \qquad \int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}. \\ x &= \frac{y-b}{a} \implies \frac{d}{dy} g^{-1}(y) = \frac{1}{a} \\ p(y) &= \frac{1}{\sqrt{2\pi}a} \exp\left(-\frac{1}{2}\left(\frac{y-b}{a}\right)^2\right) \end{aligned}$$

Gaussian integral

$$\int_{-\infty}^{\infty} e^{-(x-c)^2} dx = \sqrt{\pi}$$

$$\int_{-\infty}^{\infty} e^{-(x-c)^2/a} dx = \sqrt{\pi}\sqrt{a}$$

Transformation of variables

Given:

$$p(x) = \mathcal{N}(0,1), \qquad y = ax + b,$$

find $p(y)$.

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \qquad \int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}.$$

$$x = \frac{y-b}{a} \implies \frac{d}{dy} g^{-1}(y) = \frac{1}{a}$$

$$p(y) = \frac{1}{\sqrt{2\pi}a} \exp\left(-\frac{1}{2}\left(\frac{y-b}{a}\right)^2\right)$$

Gaussian integral

$$\int_{-\infty}^{\infty} e^{-(x-c)^2} dx = \sqrt{\pi}$$

$$\int_{-\infty}^{\infty} e^{-(x-c)^2/a} dx = \sqrt{\pi}\sqrt{a}$$

Proportionality:

$$p(x) \propto \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$$

Multivariate case

Consider

$$p(e_1) = \mathcal{N}(0,1), \qquad p(e_2) = \mathcal{N}(0,1),$$

probability of vector $\boldsymbol{e} = [e_1, e_2]$ for independent e_1, e_2?

Multivariate case

Consider

$$p(e_1) = \mathcal{N}(0,1), \qquad p(e_2) = \mathcal{N}(0,1),$$

probability of vector $\boldsymbol{e} = [e_1, e_2]$ for independent e_1, e_2?

$$p(\boldsymbol{e}) = p(e_1)p(e_2)$$

Multivariate case

Consider

$$p(e_1) = \mathcal{N}(0,1), \qquad p(e_2) = \mathcal{N}(0,1),$$

probability of vector $\boldsymbol{e} = [e_1, e_2]$ for independent e_1, e_2?

$$\begin{aligned} p(\boldsymbol{e}) &= p(e_1)p(e_2) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}e_1^2\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}e_2^2\right) \end{aligned}$$

Multivariate case

Consider

$$p(e_1) = \mathcal{N}(0,1), \qquad p(e_2) = \mathcal{N}(0,1),$$

probability of vector $\boldsymbol{e} = [e_1, e_2]$ for independent e_1, e_2?

$$\begin{aligned}
p(\boldsymbol{e}) &= p(e_1)p(e_2) \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}e_1^2\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}e_2^2\right) \\
&= \frac{1}{\sqrt{2\pi}^2} \exp\left(-\frac{1}{2}(e_1^2 + e_2^2)\right)
\end{aligned}$$

Multivariate case

Consider

$$p(e_1) = \mathcal{N}(0,1), \qquad p(e_2) = \mathcal{N}(0,1),$$

probability of vector $\boldsymbol{e} = [e_1, e_2]$ for independent e_1, e_2?

$$\begin{aligned}
p(\boldsymbol{e}) &= p(e_1)p(e_2) \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}e_1^2\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}e_2^2\right) \\
&= \frac{1}{\sqrt{2\pi}^2} \exp\left(-\frac{1}{2}(e_1^2 + e_2^2)\right) \\
&= \frac{1}{\sqrt{2\pi}^2} \exp\left(-\frac{1}{2}\boldsymbol{e}^\top \boldsymbol{e}\right) = \mathcal{N}(0, I)
\end{aligned}$$

Multivariate case

Consider

$$p(e_1) = \mathcal{N}(0,1), \qquad p(e_2) = \mathcal{N}(0,1),$$

probability of vector $\boldsymbol{e} = [e_1, e_2]$ for independent e_1, e_2?

$$\begin{aligned}
p(\boldsymbol{e}) &= p(e_1)p(e_2) \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}e_1^2\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}e_2^2\right) \\
&= \frac{1}{\sqrt{2\pi}^2} \exp\left(-\frac{1}{2}(e_1^2 + e_2^2)\right) \\
&= \frac{1}{\sqrt{2\pi}^2} \exp\left(-\frac{1}{2}\boldsymbol{e}^\top \boldsymbol{e}\right) = \mathcal{N}(0, I)
\end{aligned}$$

Distribution of $\boldsymbol{x} = A\boldsymbol{e} + \boldsymbol{b}$?

Multivariate case

Consider

$$p(e_1) = \mathcal{N}(0,1), \qquad p(e_2) = \mathcal{N}(0,1),$$

probability of vector $\boldsymbol{e} = [e_1, e_2]$ for independent e_1, e_2?

$$\begin{aligned}
p(\boldsymbol{e}) &= p(e_1)p(e_2) \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}e_1^2\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}e_2^2\right) \\
&= \frac{1}{\sqrt{2\pi}^2} \exp\left(-\frac{1}{2}(e_1^2 + e_2^2)\right) \\
&= \frac{1}{\sqrt{2\pi}^2} \exp\left(-\frac{1}{2}\boldsymbol{e}^\top \boldsymbol{e}\right) = \mathcal{N}(0, I)
\end{aligned}$$

Distribution of $\boldsymbol{x} = A\boldsymbol{e} + \boldsymbol{b}$?

$$p(\boldsymbol{x}) = \frac{1}{\sqrt{2\pi}^d \, |A|} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{b})^\top A^{-\top} A^{-1} (\boldsymbol{x} - \boldsymbol{b})\right) = \mathcal{N}(\boldsymbol{b}, \Sigma)$$

where $\Sigma = A^\top A$

Marginalization

Consider

$$p(x_1, x_2) = \mathcal{N}\left(\begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix}, \begin{bmatrix}\sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22}\end{bmatrix}\right),$$

Compute $p(x_1) = \int p(x_1, x_2) dx_2$.

Marginalization

Consider

$$p(x_1, x_2) = \mathcal{N}\left(\begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix}, \begin{bmatrix}\sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22}\end{bmatrix}\right),$$

Compute $p(x_1) = \int p(x_1, x_2) dx_2$. Denote $V = \begin{bmatrix} v_{11} & v_{12} \\ v_{12} & v_{22}\end{bmatrix} = \Sigma^{-1}$

$$p(x_1, x_2) \quad \propto \quad \exp\left(-\frac{1}{2}[x_1 - \mu_1, x_2 - \mu_2]\begin{bmatrix} v_{11} & v_{12} \\ v_{12} & v_{22}\end{bmatrix}\begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2\end{bmatrix}\right)$$

Marginalization

Consider

$$p(x_1, x_2) = \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \right),$$

Compute $p(x_1) = \int p(x_1, x_2) dx_2$. Denote $V = \begin{bmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{bmatrix} = \Sigma^{-1}$

$$\begin{aligned} p(x_1, x_2) &\propto \exp\left(-\frac{1}{2} [x_1 - \mu_1, x_2 - \mu_2] \begin{bmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix} \right) \\ &\propto \exp\left(-\frac{1}{2} \left[(x_1 - \mu_1)^2 v_{11} + (x_1 - \mu_1)(x_2 - \mu_2) 2 v_{12} + (x_2 - \mu_2)^2 v_{22} \right] \right) \end{aligned}$$

Marginalization

Consider

$$p(x_1, x_2) = \mathcal{N}\left(\begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix}, \begin{bmatrix}\sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22}\end{bmatrix}\right),$$

Compute $p(x_1){=}\int p(x_1, x_2)dx_2$. Denote $V = \begin{bmatrix}v_{11} & v_{12} \\ v_{12} & v_{22}\end{bmatrix} = \Sigma^{-1}$

$$\begin{aligned} p(x_1, x_2) &\propto \exp\left(-\frac{1}{2}[x_1 - \mu_1, x_2 - \mu_2]\begin{bmatrix}v_{11} & v_{12} \\ v_{12} & v_{22}\end{bmatrix}\begin{bmatrix}x_1 - \mu_1 \\ x_2 - \mu_2\end{bmatrix}\right) \\ &\propto \exp\left(-\frac{1}{2}\left[(x_1 - \mu_1)^2 v_{11} + (x_1 - \mu_1)(x_2 - \mu_2)2v_{12} + (x_2 - \mu_2)^2 v_{22}\right]\right) \\ &\quad \left\{(a+b)^2 = a^2 + 2ab + b^2, \quad b^2 = (x_2 - \mu_2)^2 v_{22}, \right\} \end{aligned}$$

Marginalization

Consider

$$p(x_1, x_2) = \mathcal{N}\left(\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}, \begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{bmatrix}\right),$$

Compute $p(x_1){=}\int p(x_1, x_2)dx_2$. Denote $V = \begin{bmatrix}v_{11} & v_{12}\\ v_{12} & v_{22}\end{bmatrix} = \Sigma^{-1}$

$$\begin{aligned}
p(x_1, x_2) &\propto \exp\left(-\frac{1}{2}[x_1 - \mu_1, x_2 - \mu_2]\begin{bmatrix}v_{11} & v_{12}\\ v_{12} & v_{22}\end{bmatrix}\begin{bmatrix}x_1 - \mu_1\\ x_2 - \mu_2\end{bmatrix}\right)\\
&\propto \exp\left(-\frac{1}{2}\left[(x_1 - \mu_1)^2 v_{11} + (x_1 - \mu_1)(x_2 - \mu_2)2v_{12} + (x_2 - \mu_2)^2 v_{22}\right]\right)\\
&\quad \left\{(a+b)^2 = a^2 + 2ab + b^2, \quad b^2 = (x_2 - \mu_2)^2 v_{22}, \right\}\\
&\quad \{2ab = (x_1 - \mu_1)(x_2 - \mu_2)2v_{12} \;\; a = (x_1 - \mu_1)v_{12}/\sqrt{v_{22}}\}
\end{aligned}$$

Marginalization

Consider

$$p(x_1, x_2) = \mathcal{N}\left(\begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix}, \begin{bmatrix}\sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22}\end{bmatrix}\right),$$

Compute $p(x_1) = \int p(x_1, x_2) dx_2$. Denote $V = \begin{bmatrix}v_{11} & v_{12} \\ v_{12} & v_{22}\end{bmatrix} = \Sigma^{-1}$

$$\begin{aligned}
p(x_1, x_2) &\propto \exp\left(-\frac{1}{2}[x_1 - \mu_1, x_2 - \mu_2]\begin{bmatrix}v_{11} & v_{12} \\ v_{12} & v_{22}\end{bmatrix}\begin{bmatrix}x_1 - \mu_1 \\ x_2 - \mu_2\end{bmatrix}\right) \\
&\propto \exp\left(-\frac{1}{2}\left[(x_1 - \mu_1)^2 v_{11} + (x_1 - \mu_1)(x_2 - \mu_2)2v_{12} + (x_2 - \mu_2)^2 v_{22}\right]\right) \\
&\qquad \left\{(a + b)^2 = a^2 + 2ab + b^2, \quad b^2 = (x_2 - \mu_2)^2 v_{22}, \right\} \\
&\qquad \{2ab = (x_1 - \mu_1)(x_2 - \mu_2)2v_{12} \;\; a = (x_1 - \mu_1)v_{12}/\sqrt{v_{22}}\} \\
&\propto \exp\left(-\frac{1}{2}\left[(x_1 - \mu_1)^2 v_{11} - (x_1 - \mu_1)^2 v_{12}^2/v_{22}\right]\right) \times \\
&\quad \exp\left(-\frac{1}{2}[(x_1 - \mu_1)v_{12}/\sqrt{v_{22}} + (x_2 - \mu_2)\sqrt{v_{22}}]^2\right)
\end{aligned}$$

Marginalization

Consider

$$p(x_1, x_2) = \mathcal{N}\left(\begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix}, \begin{bmatrix}\sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22}\end{bmatrix}\right),$$

Compute $p(x_1) = \int p(x_1, x_2) dx_2$. Denote $V = \begin{bmatrix} v_{11} & v_{12} \\ v_{12} & v_{22}\end{bmatrix} = \Sigma^{-1}$

$$\begin{aligned}
p(x_1, x_2) &\propto \exp\left(-\frac{1}{2}[x_1 - \mu_1, x_2 - \mu_2]\begin{bmatrix} v_{11} & v_{12} \\ v_{12} & v_{22}\end{bmatrix}\begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2\end{bmatrix}\right) \\
&\propto \exp\left(-\frac{1}{2}\left[(x_1 - \mu_1)^2 v_{11} + (x_1 - \mu_1)(x_2 - \mu_2)2v_{12} + (x_2 - \mu_2)^2 v_{22}\right]\right) \\
&\qquad \left\{(a+b)^2 = a^2 + 2ab + b^2, \quad b^2 = (x_2 - \mu_2)^2 v_{22}, \right\} \\
&\qquad \{2ab = (x_1 - \mu_1)(x_2 - \mu_2)2v_{12} \;\; a = (x_1 - \mu_1)v_{12}/\sqrt{v_{22}}\} \\
&\propto \exp\left(-\frac{1}{2}\left[(x_1 - \mu_1)^2 v_{11} - (x_1 - \mu_1)^2 v_{12}^2/v_{22}\right]\right) \times \\
&\qquad \exp\left(-\frac{1}{2}[(x_1 - \mu_1)v_{12}/\sqrt{v_{22}} + (x_2 - \mu_2)\sqrt{v_{22}}]^2\right) \\
&\propto \exp\left(-\frac{1}{2}\left[(x_1 - \mu_1)^2 (v_{11}v_{22} - v_{12}^2)/v_{22}\right]\right) \times \\
&\qquad \exp\left(-\frac{1}{2}v_{22}\,[x_2 - \mu_2 + (x_1 - \mu_1)v_{12}/v_{22}]^2\right)
\end{aligned}$$

Marginalization

$$\int p(x_1, x_2) dx_2 \quad \propto \quad \exp\left(-\frac{1}{2}\left[(x_1-\mu_1)^2(v_{11}v_{22}-v_{12}^2)/v_{22}\right]\right) \times$$
$$\int \exp\left(-\frac{1}{2} v_{22}\left[x_2-\mu_2+(x_1-\mu_1)v_{12}/v_{22}\right]^2\right) dx_2$$

Marginalization

$$\begin{aligned}
\int p(x_1, x_2) dx_2 &\propto \exp\left(-\frac{1}{2}\left[(x_1-\mu_1)^2(v_{11}v_{22}-v_{12}^2)/v_{22}\right]\right) \times \\
&\quad \int \exp\left(-\frac{1}{2}v_{22}\left[x_2-\mu_2+(x_1-\mu_1)v_{12}/v_{22}\right]^2\right) dx_2 \\
&\propto \exp\left(-\frac{1}{2}\left[(x_1-\mu_1)^2(v_{11}v_{22}-v_{12}^2)/v_{22}\right]\right)
\end{aligned}$$

Marginalization

$$\int p(x_1, x_2)dx_2 \propto \exp\left(-\frac{1}{2}\left[(x_1-\mu_1)^2(v_{11}v_{22}-v_{12}^2)/v_{22}\right]\right) \times$$

$$\int \exp\left(-\frac{1}{2}v_{22}\left[x_2-\mu_2+(x_1-\mu_1)v_{12}/v_{22}\right]^2\right) dx_2$$

$$\propto \exp\left(-\frac{1}{2}\left[(x_1-\mu_1)^2(v_{11}v_{22}-v_{12}^2)/v_{22}\right]\right)$$

$$\begin{bmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{bmatrix}^{-1} = \frac{1}{v_{11}v_{22}-v_{12}^2}\begin{bmatrix} v_{22} & -v_{12} \\ -v_{12} & v_{11} \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$$

Marginalization

$$\int p(x_1, x_2) dx_2 \propto \exp\left(-\frac{1}{2}\left[(x_1 - \mu_1)^2 (v_{11} v_{22} - v_{12}^2)/v_{22}\right]\right) \times$$

$$\int \exp\left(-\frac{1}{2} v_{22} \left[x_2 - \mu_2 + (x_1 - \mu_1) v_{12}/v_{22}\right]^2\right) dx_2$$

$$\propto \exp\left(-\frac{1}{2}\left[(x_1 - \mu_1)^2 (v_{11} v_{22} - v_{12}^2)/v_{22}\right]\right)$$

$$\begin{bmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{bmatrix}^{-1} = \frac{1}{v_{11} v_{22} - v_{12}^2} \begin{bmatrix} v_{22} & -v_{12} \\ -v_{12} & v_{11} \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$$

Yielding:

$$p(x_1) = \int p(x_1, x_2) dx_2 \propto \exp\left(-\frac{1}{2} \frac{(x_1 - \mu_1)^2}{\sigma_{11}^2}\right) = \mathcal{N}(x_1, \sigma_{11})$$

Completion of squares

Consider

$$\begin{aligned} ax^2 + bx + c \quad => \quad & a(x+y)^2 + d \\ & a(x^2 + 2xy + y^2) + d \\ & ax^2 + 2axy + ay^2 + d \end{aligned}$$

Completion of squares

Consider

$$
\begin{aligned}
ax^2 + bx + c \quad &=> \quad a(x+y)^2 + d \\
&\qquad\; a(x^2 + 2xy + y^2) + d \\
&\qquad\; ax^2 + 2axy + ay^2 + d
\end{aligned}
$$

$$bx = 2axy$$

Completion of squares

Consider

$$
\begin{aligned}
ax^2 + bx + c \quad &=> \quad a(x+y)^2 + d \\
&\quad\quad\;\; a(x^2 + 2xy + y^2) + d \\
&\quad\quad\;\; ax^2 + 2axy + ay^2 + d \\
bx = 2axy \quad &=> \quad y = b/(2a)
\end{aligned}
$$

Completion of squares

Consider

$$\begin{aligned} ax^2 + bx + c &\quad => \quad a(x+y)^2 + d \\ &\quad\quad a(x^2 + 2xy + y^2) + d \\ &\quad\quad ax^2 + 2axy + ay^2 + d \\ bx = 2axy &\quad => \quad y = b/(2a) \\ ax^2 + bx + ay^2 - ay^2 + c &\quad => \quad d = c - ay^2 \end{aligned}$$

Completion of squares

Consider

$$
\begin{aligned}
ax^2 + bx + c \quad &=> \quad a(x+y)^2 + d \\
& a(x^2 + 2xy + y^2) + d \\
& ax^2 + 2axy + ay^2 + d \\
bx = 2axy \quad &=> \quad y = b/(2a) \\
ax^2 + bx + ay^2 - ay^2 + c \quad &=> \quad d = c - ay^2 \\
ax^2 + bx + c \quad &= \quad a(x + b/(sa))^2 + c - b^2/(4a)
\end{aligned}
$$

Multivariate:

$$
\boldsymbol{x}^\top A\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c \quad => \quad (\boldsymbol{x} - \mu)^\top A(\boldsymbol{x} - \mu) + d
$$

Completion of squares

Consider

$$\begin{aligned}
ax^2 + bx + c \quad &=> \quad a(x+y)^2 + d \\
&\qquad\; a(x^2 + 2xy + y^2) + d \\
&\qquad\; ax^2 + 2axy + ay^2 + d \\
bx = 2axy \quad &=> \quad y = b/(2a) \\
ax^2 + bx + ay^2 - ay^2 + c \quad &=> \quad d = c - ay^2 \\
ax^2 + bx + c \quad &= \quad a(x + b/(sa))^2 + c - b^2/(4a)
\end{aligned}$$

Multivariate:

$$\begin{aligned}
\boldsymbol{x}^\top A\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c \quad &=> \quad (\boldsymbol{x} - \mu)^\top A(\boldsymbol{x} - \mu) + d \\
&\qquad\; \boldsymbol{x}^\top A\boldsymbol{x} - \mu^\top A\boldsymbol{x} - \boldsymbol{x}^\top A\mu + \mu^\top \mu + d
\end{aligned}$$

Completion of squares

Consider

$$
\begin{aligned}
ax^2 + bx + c \quad &=> \quad a(x+y)^2 + d \\
& \quad a(x^2 + 2xy + y^2) + d \\
& \quad ax^2 + 2axy + ay^2 + d \\
bx = 2axy \quad &=> \quad y = b/(2a) \\
ax^2 + bx + ay^2 - ay^2 + c \quad &=> \quad d = c - ay^2 \\
ax^2 + bx + c \quad &= \quad a(x + b/(sa))^2 + c - b^2/(4a)
\end{aligned}
$$

Multivariate:

$$
\begin{aligned}
\mathbf{x}^\top A\mathbf{x} + \mathbf{b}^\top \mathbf{x} + c \quad &=> \quad (\mathbf{x} - \mu)^\top A(\mathbf{x} - \mu) + d \\
& \quad \mathbf{x}^\top A\mathbf{x} - \mu^\top A\mathbf{x} - \mathbf{x}^\top A\mu + \mu^\top \mu + d \\
\text{for sym. } A: \quad & \quad \mu^\top A\mathbf{x} = \mathbf{x}^\top A\mu
\end{aligned}
$$

Completion of squares

Consider

$$
\begin{aligned}
ax^2 + bx + c &=> a(x+y)^2 + d \\
&\quad\;\; a(x^2 + 2xy + y^2) + d \\
&\quad\;\; ax^2 + 2axy + ay^2 + d \\
bx = 2axy &=> y = b/(2a) \\
ax^2 + bx + ay^2 - ay^2 + c &=> d = c - ay^2 \\
ax^2 + bx + c &= a(x + b/(sa))^2 + c - b^2/(4a)
\end{aligned}
$$

Multivariate:

$$
\begin{aligned}
\boldsymbol{x}^\top A\boldsymbol{x} + \boldsymbol{b}^\top\boldsymbol{x} + c &=> (\boldsymbol{x} - \mu)^\top A(\boldsymbol{x} - \mu) + d \\
&\quad\;\; \boldsymbol{x}^\top A\boldsymbol{x} - \mu^\top A\boldsymbol{x} - \boldsymbol{x}^\top A\mu + \mu^\top\mu + d \\
\text{for sym. } A: &\quad\;\; \mu^\top A\boldsymbol{x} = \boldsymbol{x}^\top A\mu \\
&=> \boldsymbol{b}^\top\boldsymbol{x} = 2\mu^\top A\boldsymbol{x}
\end{aligned}
$$

Completion of squares

Consider

$$\begin{aligned} ax^2 + bx + c &\quad => \quad a(x+y)^2 + d \\ &\quad\quad a(x^2 + 2xy + y^2) + d \\ &\quad\quad ax^2 + 2axy + ay^2 + d \\ bx = 2axy &\quad => \quad y = b/(2a) \\ ax^2 + bx + ay^2 - ay^2 + c &\quad => \quad d = c - ay^2 \\ ax^2 + bx + c &\quad = \quad a(x + b/(sa))^2 + c - b^2/(4a) \end{aligned}$$

Multivariate:

$$\begin{aligned} \boldsymbol{x}^\top A\boldsymbol{x} + \boldsymbol{b}^\top\boldsymbol{x} + c &\quad => \quad (\boldsymbol{x}-\mu)^\top A(\boldsymbol{x}-\mu) + d \\ &\quad\quad \boldsymbol{x}^\top A\boldsymbol{x} - \mu^\top A\boldsymbol{x} - \boldsymbol{x}^\top A\mu + \mu^\top\mu + d \\ \text{for sym. } A: &\quad\quad \mu^\top A\boldsymbol{x} = \boldsymbol{x}^\top A\mu \\ &\quad => \quad \boldsymbol{b}^\top\boldsymbol{x} = 2\mu^\top A\boldsymbol{x} \\ &\quad\quad \mu^\top = \frac{1}{2}\boldsymbol{b}^\top A^{-1} \quad \mu = \frac{1}{2}A^{-1}\boldsymbol{b} \end{aligned}$$

Conditional distribution

Consider

$$p(x_1, x_2) = \mathcal{N}\left(\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}, \begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{bmatrix}\right),$$

Compute $p(x_2|x_1)$

Conditional distribution

Consider

$$p(x_1, x_2) = \mathcal{N}\left(\begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix}, \begin{bmatrix}\sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22}\end{bmatrix}\right),$$

Compute $p(x_2|x_1) = \frac{p(x_1,x_2)}{p(x_1)}$.

Conditional distribution

Consider

$$p(x_1, x_2) = \mathcal{N}\left(\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}, \begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{bmatrix}\right),$$

Compute $p(x_2|x_1) = \frac{p(x_1,x_2)}{p(x_1)}$.Using previous results:

$$\begin{aligned} p(x_1, x_2) \quad &\propto \quad \exp\left(-\frac{1}{2}\left[(x_1-\mu_1)^2(v_{11}v_{22}-v_{12}^2)/v_{22}\right]\right)\times \\ &\exp\left(-\frac{1}{2}v_{22}\left[x_2-\mu_2+(x_1-\mu_1)v_{12}/v_{22}\right]^2\right)\end{aligned}$$

Conditional distribution

Consider

$$p(x_1, x_2) = \mathcal{N}\left(\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}, \begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{bmatrix}\right),$$

Compute $p(x_2|x_1) = \frac{p(x_1,x_2)}{p(x_1)}$.Using previous results:

$$\begin{aligned} p(x_1, x_2) \quad &\propto \quad \exp\left(-\frac{1}{2}\left[(x_1-\mu_1)^2(v_{11}v_{22}-v_{12}^2)/v_{22}\right]\right)\times \\ &\qquad \exp\left(-\frac{1}{2}v_{22}\left[x_2-\mu_2+(x_1-\mu_1)v_{12}/v_{22}\right]^2\right) \\ p(x_2|x_1) = \frac{p(x_1,x_2)}{p(x_1)} \quad &\propto \quad \exp\left(-\frac{1}{2}v_{22}\left[x_2-\mu_2+(x_1-\mu_1)v_{12}/v_{22}\right]^2\right) \end{aligned}$$

Conditional distribution

Consider

$$p(x_1, x_2) = \mathcal{N}\left(\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}, \begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{bmatrix}\right),$$

Compute $p(x_2|x_1) = \frac{p(x_1,x_2)}{p(x_1)}$.Using previous results:

$$\begin{aligned}
p(x_1, x_2) &\propto \exp\left(-\frac{1}{2}\left[(x_1-\mu_1)^2(v_{11}v_{22}-v_{12}^2)/v_{22}\right]\right) \times \\
&\quad \exp\left(-\frac{1}{2}v_{22}\left[x_2-\mu_2+(x_1-\mu_1)v_{12}/v_{22}\right]^2\right) \\
p(x_2|x_1) = \frac{p(x_1,x_2)}{p(x_1)} &\propto \exp\left(-\frac{1}{2}v_{22}\left[x_2-\mu_2+(x_1-\mu_1)v_{12}/v_{22}\right]^2\right) \\
&= \mathcal{N}(\mu_2-(x_1-\mu_1)v_{12}/v_{22}, v_{22}^{-1})
\end{aligned}$$

Conditional distribution

Consider

$$p(x_1, x_2) = \mathcal{N}\left(\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}, \begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{bmatrix}\right),$$

Compute $p(x_2|x_1) = \frac{p(x_1,x_2)}{p(x_1)}$.Using previous results:

$$\begin{aligned}
p(x_1, x_2) &\propto \exp\left(-\frac{1}{2}\left[(x_1-\mu_1)^2(v_{11}v_{22}-v_{12}^2)/v_{22}\right]\right)\times\\
&\quad \exp\left(-\frac{1}{2}v_{22}\left[x_2-\mu_2+(x_1-\mu_1)v_{12}/v_{22}\right]^2\right)\\
p(x_2|x_1) = \frac{p(x_1,x_2)}{p(x_1)} &\propto \exp\left(-\frac{1}{2}v_{22}\left[x_2-\mu_2+(x_1-\mu_1)v_{12}/v_{22}\right]^2\right)\\
&= \mathcal{N}(\mu_2-(x_1-\mu_1)v_{12}/v_{22}, v_{22}^{-1})\\
\begin{bmatrix}v_{11} & v_{12}\\ v_{12} & v_{22}\end{bmatrix} &= \begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{bmatrix}^{-1} = \frac{1}{\sigma_{11}\sigma_{22}-\sigma_{12}^2}\begin{bmatrix}\sigma_{22} & -\sigma_{12}\\ -\sigma_{12} & \sigma_{11}\end{bmatrix}
\end{aligned}$$

Conditional distribution

Consider

$$p(x_1, x_2) = \mathcal{N}\left(\begin{bmatrix}\mu_1\\\mu_2\end{bmatrix}, \begin{bmatrix}\sigma_{11} & \sigma_{12}\\\sigma_{12} & \sigma_{22}\end{bmatrix}\right),$$

Compute $p(x_2|x_1) = \frac{p(x_1,x_2)}{p(x_1)}$.Using previous results:

$$\begin{aligned}
p(x_1, x_2) &\propto \exp\left(-\frac{1}{2}\left[(x_1-\mu_1)^2(v_{11}v_{22}-v_{12}^2)/v_{22}\right]\right)\times\\
&\qquad \exp\left(-\frac{1}{2}v_{22}\left[x_2-\mu_2+(x_1-\mu_1)v_{12}/v_{22}\right]^2\right)\\
p(x_2|x_1) = \frac{p(x_1,x_2)}{p(x_1)} &\propto \exp\left(-\frac{1}{2}v_{22}\left[x_2-\mu_2+(x_1-\mu_1)v_{12}/v_{22}\right]^2\right)\\
&= \mathcal{N}(\mu_2-(x_1-\mu_1)v_{12}/v_{22}, v_{22}^{-1})\\
\begin{bmatrix}v_{11} & v_{12}\\ v_{12} & v_{22}\end{bmatrix} &= \begin{bmatrix}\sigma_{11} & \sigma_{12}\\\sigma_{12} & \sigma_{22}\end{bmatrix}^{-1} = \frac{1}{\sigma_{11}\sigma_{22}-\sigma_{12}^2}\begin{bmatrix}\sigma_{22} & -\sigma_{12}\\ -\sigma_{12} & \sigma_{11}\end{bmatrix}\\
p(x_2|x_1) &= \mathcal{N}(\mu_2+(x_1-\mu_1)\sigma_{12}/\sigma_{22}, \sigma_{22}-\sigma_{12}^2/\sigma_{11})
\end{aligned}$$

Gaussian likelihood model

Consider observations y_1 and y_2 to be generated independently from a Gaussian distribution with unknown mean and variance:

$$
\begin{aligned}
p(y_1, y_2|m, s) &= p(y_1|m, s)p(y_2|m, s) \\
&= \frac{1}{\sqrt{2\pi s}} \exp\left(-\frac{1}{2}\frac{(y_1 - m)^2}{s}\right) \frac{1}{\sqrt{2\pi s}} \exp\left(-\frac{1}{2}\frac{(y_2 - m)^2}{s}\right)
\end{aligned}
$$

Maximum likelihood (log-likelihood) estimates of m, s:

Gaussian likelihood model

Consider observations y_1 and y_2 to be generated independently from a Gaussian distribution with unknown mean and variance:

$$\begin{aligned} p(y_1, y_2|m, s) &= p(y_1|m, s)p(y_2|m, s) \\ &= \frac{1}{\sqrt{2\pi s}} \exp\left(-\frac{1}{2}\frac{(y_1 - m)^2}{s}\right) \frac{1}{\sqrt{2\pi s}} \exp\left(-\frac{1}{2}\frac{(y_2 - m)^2}{s}\right) \end{aligned}$$

Maximum likelihood (log-likelihood) estimates of m, s:

$$L = \log p(y_1, y_2|m, s) \quad = \quad -\log s - \frac{1}{2}\left[\frac{(y_1 - m)^2}{s} + \frac{(y_2 - m)^2}{s}\right]$$

Gaussian likelihood model

Consider observations y_1 and y_2 to be generated independently from a Gaussian distribution with unknown mean and variance:

$$\begin{aligned} p(y_1, y_2|m, s) &= p(y_1|m, s)p(y_2|m, s) \\ &= \frac{1}{\sqrt{2\pi s}} \exp\left(-\frac{1}{2}\frac{(y_1 - m)^2}{s}\right) \frac{1}{\sqrt{2\pi s}} \exp\left(-\frac{1}{2}\frac{(y_2 - m)^2}{s}\right) \end{aligned}$$

Maximum likelihood (log-likelihood) estimates of m, s:

$$\begin{aligned} L = \log p(y_1, y_2|m, s) &= -\log s - \frac{1}{2}\left[\frac{(y_1 - m)^2}{s} + \frac{(y_2 - m)^2}{s}\right] \\ \frac{dL}{dm} &= \frac{(y_1 - m)}{s} + \frac{(y_2 - m)}{s} \\ \frac{dL}{dm} &= -\frac{1}{s} + \frac{1}{s^2}\left[\frac{(y_1 - m)^2 + (y_2 - m)^2)}{2}\right] \end{aligned}$$

Gaussian likelihood model

Consider observations y_1 and y_2 to be generated independently from a Gaussian distribution with unknown mean and variance:

$$
\begin{aligned}
p(y_1, y_2|m, s) &= p(y_1|m, s)p(y_2|m, s) \\
&= \frac{1}{\sqrt{2\pi s}} \exp\left(-\frac{1}{2}\frac{(y_1 - m)^2}{s}\right) \frac{1}{\sqrt{2\pi s}} \exp\left(-\frac{1}{2}\frac{(y_2 - m)^2}{s}\right)
\end{aligned}
$$

Maximum likelihood (log-likelihood) estimates of m, s:

$$
\begin{aligned}
L = \log p(y_1, y_2|m, s) &= -\log s - \frac{1}{2}\left[\frac{(y_1 - m)^2}{s} + \frac{(y_2 - m)^2}{s}\right] \\
\frac{dL}{dm} &= \frac{(y_1 - m)}{s} + \frac{(y_2 - m)}{s} \\
\frac{dL}{dm} &= -\frac{1}{s} + \frac{1}{s^2}\left[\frac{(y_1 - m)^2 + (y_2 - m)^2)}{2}\right] \\
\hat{m} &= \arg\max_m L = \frac{y_1 + y_2}{2} \\
\hat{s} &= \arg\max_{s,m} L = \frac{(y_1 - \hat{m})^2 + (y_2 - \hat{m})^2)}{2}
\end{aligned}
$$

Bayesian model

Observations: y_1, y_2 that are assumed to be 2 realizations of random variable

$$p(y_i|m,s) = \mathcal{N}(m,s), \quad i = 1,2,$$

where m, s are unknown. We seek

$$p(m,s|y_1,y_2) =$$

Bayesian model

Observations: y_1, y_2 that are assumed to be 2 realizations of random variable

$$p(y_i|m,s) = \mathcal{N}(m,s), \quad i = 1,2,$$

where m, s are unknown. We seek

$$p(m,s|y_1,y_2) = \frac{p(y_1,y_2,m,s)}{p(y_1,y_2)} =$$

Bayesian model

Observations: y_1, y_2 that are assumed to be 2 realizations of random variable

$$p(y_i|m,s) = \mathcal{N}(m,s), \quad i = 1,2,$$

where m, s are unknown. We seek

$$p(m,s|y_1,y_2) = \frac{p(y_1,y_2,m,s)}{p(y_1,y_2)} = \frac{p(y_1|m,s)p(y_2|m,s)p(m|s)p(s)}{p(y_1,y_2)}$$

We choose(!):

$$s \quad \sim \quad iG(\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\left(\frac{1}{s}\right)^{\alpha+1}\exp\left(-\frac{\beta}{s}\right)$$

Bayesian model

Observations: y_1, y_2 that are assumed to be 2 realizations of random variable

$$p(y_i|m,s) = \mathcal{N}(m,s), \quad i = 1,2,$$

where m, s are unknown. We seek

$$p(m,s|y_1,y_2) = \frac{p(y_1,y_2,m,s)}{p(y_1,y_2)} = \frac{p(y_1|m,s)p(y_2|m,s)p(m|s)p(s)}{p(y_1,y_2)}$$

We choose(!):

$$\begin{aligned} s &\sim iG(\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\left(\frac{1}{s}\right)^{\alpha+1}\exp\left(-\frac{\beta}{s}\right) \\ m &\sim \mathcal{N}(0,s) \end{aligned}$$

Nomenclature:
Likelihood: $p(y_i|m,s)$ with parameters m, s
Prior: $p(m,s)$ with hyper-parameters α, β.

Derivation

$$
\begin{aligned}
p(m,s|y_1,y_2) &= \frac{p(y_1|m,s)p(y_2|m,s)p(m|s)p(s)}{p(y_1,y_2)} \\
&\propto p(y_1|m,s)p(y_2|m,s)p(m|s)p(s) \\
&\propto \frac{1}{\sqrt{s}^3} \exp\left(-\frac{1}{2}\frac{(y_1-m)^2}{s} - \frac{1}{2}\frac{(y_1-m)^2}{s} - \frac{1}{2}\frac{m^2}{s} - \frac{\beta}{s}\right)\frac{1}{s^{\alpha+1}} \\
&\propto s^{-\frac{3}{2}-\alpha-1} \exp\left(-\frac{1}{2s}\left[(y_1-m)^2+(y_2-m)^2+m^2\right]-\frac{\beta}{s}\right) \\
&\quad \text{completion of squares} \\
p(m,s|y_1,y_2) &\propto \frac{1}{\sqrt{s}^3} \exp\left(-\frac{1}{2}\frac{3}{s}\left[m-\frac{(y_1+y_2)}{3}\right]^2 - \frac{1}{2s}S - \frac{\beta}{s}\right)\frac{1}{s^{\alpha+1}} \\
S &= -3\left(\frac{y_1+y_2}{3}\right)^2 + (y_1^2+y_2^2)
\end{aligned}
$$

Completion

$$
\begin{aligned}
(y_1-m)^2+(y_2-m)^2+m^2 &= 3m^2 - 2(y_1+y_2)m + y_1^2 + y_2^2 \\
&= 3(m^2 - 2\frac{y_1+y_2}{3}m) + y_1^2 + y_2^2 \\
&= 3m^2 - 2(y_1+y_2)m + 3\left(\frac{y_1+y_2}{3}\right)^2 - 3\left(\frac{y_1+y_2}{3}\right)^2 + (y_1^2+y_2^2)
\end{aligned}
$$

Decomposition of the joint

$$p(m, s|y_1, y_2) \propto \frac{1}{s^{\frac{3}{2}+\alpha+1}} \exp\left(-\frac{1}{2}\frac{3}{s}\left[m - \frac{(y_1+y_2)}{3}\right]^2 - \frac{1}{2s}S - \frac{\beta}{s}\right)$$

integrating over m (Gauss integral, yielding $\sqrt{s}\sqrt{\pi}$)

$$p(s) \propto \frac{1}{s^{\frac{2}{2}+\alpha+1}} \exp\left(-\frac{\beta + S/2}{s}\right)$$

$$= iG\left(\alpha+1, \beta + \frac{S}{2}\right)$$

$$p(m|s) \propto \frac{1}{\sqrt{s}} \exp\left(-\frac{1}{2}\frac{3}{s}\left[m - \frac{(y_1+y_2)}{3}\right]^2\right)$$

$$= \mathcal{N}\left(\frac{y_1+y_2}{3}, \frac{s}{3}\right)$$

Alternative decomposition

$$p(m,s|y_1,y_2) \propto \frac{1}{s^{\frac{3}{2}+\alpha+1}} \exp\left(-\frac{1}{2}\frac{3}{s}\left[m-\frac{(y_1+y_2)}{3}\right]^2 - \frac{1}{2s}S - \frac{\beta}{s}\right)$$

integrating over s (norm.coef of iG)

$$p(m|y_1,y_2) \propto \left(\frac{3}{2}\left[m-\frac{(y_1+y_2)}{3}\right]^2 + \frac{S}{2} + \beta\right)^{-\frac{3}{2}-\alpha}$$

$$= St\left(\frac{(y_1+y_2)}{3}, \frac{S+2\beta}{3}, \frac{3}{2}+\alpha\right)$$

$$p(s|m,y_1,y_2) \propto \frac{1}{s^{\frac{3}{2}+\alpha+1}} \exp\left(-\frac{1}{s}\left\{\frac{3}{2}\left[m-\frac{(y_1+y_2)}{3}\right]^2 + \frac{1}{2}S + \beta\right\}\right)$$

$$= iG\left(\frac{3}{2}+\alpha, \frac{3}{2}\left[m-\frac{(y_1+y_2)}{3}\right]^2 + \frac{1}{2}S + \beta\right)$$

Integration of inverse Gamma

$$\int \left(\frac{1}{s}\right)^{\alpha+1} \exp\left(-\frac{\beta}{s}\right) ds = \frac{\Gamma(\alpha)}{\beta^\alpha}$$

Take home message

- likelihood is typically product of distributions (i.i.d)
- prior is designed
- proportionality is useful (believe that normalization can be done later),
- completion of squares $=$ working with Gaussian distribution
- Marginal and Conditional distributions of a Gaussian is a Gaussian

Homework assignment (5 points each decomposition)

Model:

$$
\begin{aligned}
s &\sim iG(\alpha, \beta) \\
m &\sim \mathcal{N}(0, \tau) \\
x_1 &\sim \mathcal{N}(m, s) \\
x_2 &\sim \mathcal{N}(m, s)
\end{aligned}
$$

Bayes rule:

$$p(m, s|x_1, x_2) = \frac{p(x_1|m, s)p(x_2|m, s)p(m|s)p(s)}{p(x_1, x_2)}$$

Find decompositions:

$$
\begin{aligned}
p(m, s|x_1, x_2) &= p(m|s, x_1, x_2)p(s|x_1, x_2) \\
p(m, s|x_1, x_2) &= p(s|m, x_1, x_2)p(m|x_1, x_2)
\end{aligned}
$$

Hint:

$$p(s|x_1, x_2) \propto \int p(x_1|m, s)p(x_2|m, s)p(m|s)p(s)dm$$

$$p(m|x_1, x_2) \propto \int p(x_1|m, s)p(x_2|m, s)p(m|s)p(s)ds$$