The Bayes Theorem

Laurent Bertino and Geir Evensen

Nansen Environmental and Remote Sensing Center, Bergen, Norway,
Hydro Research Centre, Bergen, Norway

Why Bayes?

- Provides a *framework* for data assimilation.
- All data assimilation methods can be deduced from it.

Principle of Bayesian analysis

"The probability of A given B is not the same as the probability of B given A".

$$P(A|B) \neq P(B|A)$$

- Non-Bayesian probabilities for data knowing the parameters.
- Bayesian probabilities for parameters knowing the data.

The Bayes theorem

By definition of conditional probabilities:

$$
\begin{aligned}
P(A \cap B) &= P(A|B)P(B) \\
&= P(B|A)P(A) \\
\big[&= P(A)P(B) \big]^*
\end{aligned}
$$

Thus the Bayes theorem

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|!B)P(!B)}$$

$$P(B|A) \propto P(A|B)P(B)$$

(* if A and B are independent)

Practical problem

- 1% of women aged 40 have breast cancer
- A mammography test has 80% success rate
- A mammography test has 10% false alarm rate

A woman receives a positive mammography test, what is the probability that she actually has cancer?

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|!B)P(!B)}$$

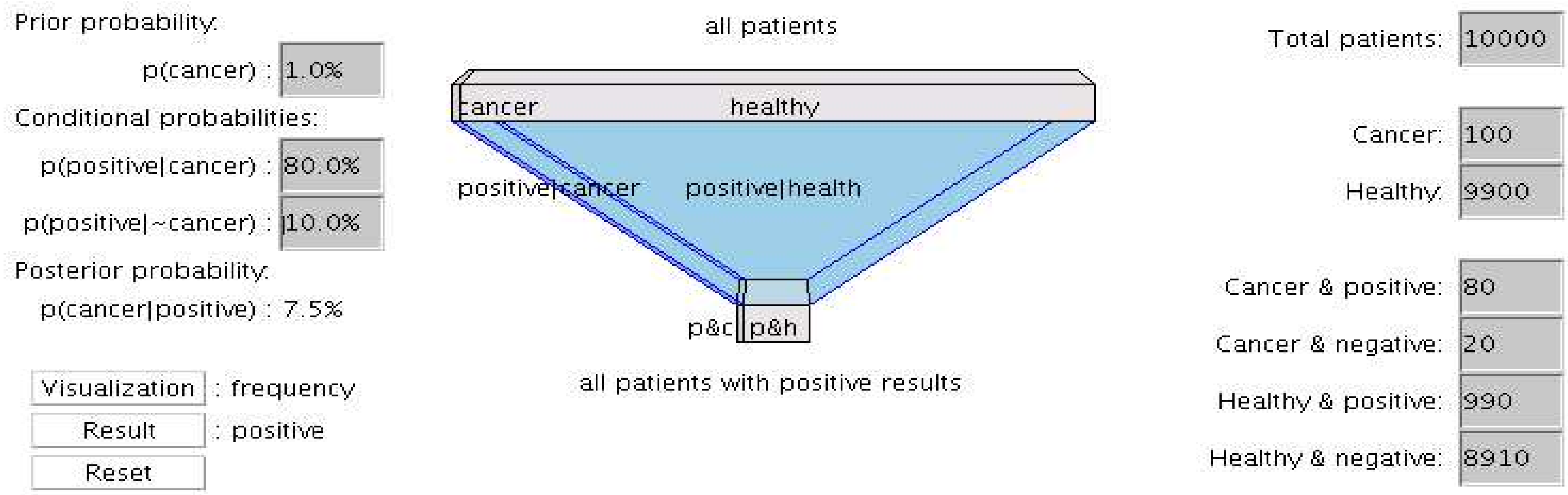
Notations:

A: Mammo+, !A: Mammo-

B: Cancer, !B: Healthy

Solution

$$P(B|A) = \frac{80 \times 1}{80 \times 1 + 10 \times 99} = 7.5\%$$



Conclusion

Worth taking the test?

- The posterior (7.5%) is different from the prior (1%)
- What if the test had been negative?

The test may be inaccurate but still helps *updating* the prior.

Continuous variables

- The state variable ψ and its *pdf*: $f(\boldsymbol{\psi})$
- The observations $\boldsymbol{d}$ and their *pdf*: $f(\boldsymbol{d})$

$$f(\boldsymbol{\psi}|\boldsymbol{d}) = \frac{f(\boldsymbol{d}|\boldsymbol{\psi})f(\boldsymbol{\psi})}{\int f(\boldsymbol{d}|\boldsymbol{\psi})f(\boldsymbol{\psi})d\boldsymbol{\psi}}$$
$$\propto f(\boldsymbol{d}|\boldsymbol{\psi})f(\boldsymbol{\psi})$$

- The denominator is difficult to compute except for very special cases of distributions.
- Gaussian variables...

The Gaussian case

- The Gaussian distribution is perfectly determined by:
  - The mean (or "expectation" E)
  - The variance (and covariance for multivariate problems)
- Conserved by linear combinations.
- The maximum likelihood estimator coincides with the conditional expectation $E(\boldsymbol{\psi}|\boldsymbol{d})$.
- Hilbert space geometry: $E(\boldsymbol{\psi}|\boldsymbol{d})$ is also the least squares estimator.

$$\boldsymbol{\psi}^{\mathrm{a}} = \boldsymbol{\psi}^{\mathrm{f}} + \boldsymbol{C}^{\mathrm{f}}_{\psi\psi}\boldsymbol{\mathcal{M}}^{\mathrm{T}}(\boldsymbol{\mathcal{M}}\boldsymbol{C}^{\mathrm{f}}_{\psi\psi}\boldsymbol{\mathcal{M}}^{\mathrm{T}} + \boldsymbol{C}_{\epsilon\epsilon})^{-1}(\boldsymbol{d} - \boldsymbol{\mathcal{M}}\boldsymbol{\psi}^{\mathrm{f}})$$

$$\boldsymbol{C}^{\mathrm{a}}_{\psi\psi} = \boldsymbol{C}^{\mathrm{f}}_{\psi\psi} - \boldsymbol{C}^{\mathrm{f}}_{\psi\psi}\boldsymbol{\mathcal{M}}^{\mathrm{T}}\left(\boldsymbol{\mathcal{M}}\boldsymbol{C}^{\mathrm{f}}_{\psi\psi}\boldsymbol{\mathcal{M}}^{\mathrm{T}} + \boldsymbol{C}_{\epsilon\epsilon}\right)^{-1}\boldsymbol{\mathcal{M}}\boldsymbol{C}^{\mathrm{f}}_{\psi\psi}.$$

Bayesian formulation

Model equations and measurements:

$$\frac{\partial \boldsymbol{\psi}}{\partial t} = \boldsymbol{g}(\boldsymbol{\psi}) + \boldsymbol{q},$$

$$\boldsymbol{\psi}|_{t_0} = \boldsymbol{\Psi}_0 + \boldsymbol{a},$$

$$\boldsymbol{\mathcal{M}}\boldsymbol{\psi} = \boldsymbol{d} + \boldsymbol{\epsilon}.$$

Bayes theorem becomes:

$$f(\boldsymbol{\psi}, \boldsymbol{\psi}_0|\boldsymbol{d}) \propto f(\boldsymbol{\psi}, \boldsymbol{\psi}_0) f(\boldsymbol{d}|\boldsymbol{\psi}, \boldsymbol{\psi}_0).$$

$$= f(\boldsymbol{\psi}|\boldsymbol{\psi}_0) f(\boldsymbol{\psi}_0) f(\boldsymbol{d}|\boldsymbol{\psi}).$$

Gaussian priors

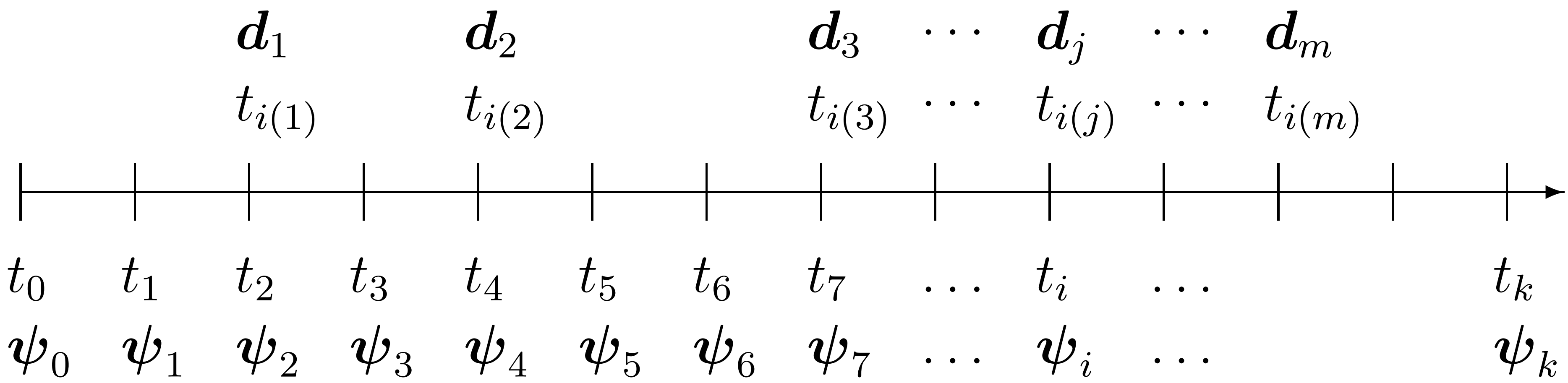
Cost function rederived using Gaussian priors.

$$f(\boldsymbol{\psi}, \boldsymbol{\psi}_0 | \boldsymbol{d}) \propto \exp\left\{-\frac{1}{2}\mathcal{J}[\boldsymbol{\psi}]\right\},$$

$$\mathcal{J}[\psi] = \left(\frac{\partial \boldsymbol{\psi}}{\partial t} - \boldsymbol{g}(\boldsymbol{\psi})\right)^{\mathrm{T}} \bullet \boldsymbol{W}_{qq} \bullet \left(\frac{\partial \boldsymbol{\psi}}{\partial t} - \boldsymbol{g}(\boldsymbol{\psi})\right)$$
$$+ (\boldsymbol{\psi}_0 - \boldsymbol{\Psi}_0)^{\mathrm{T}} \circ \boldsymbol{W}_{aa} \circ (\boldsymbol{\psi}_0 - \boldsymbol{\Psi}_0)$$
$$+ (\boldsymbol{\mathcal{M}}\psi - \boldsymbol{d})^{\mathrm{T}} \boldsymbol{W}_{\epsilon\epsilon} (\boldsymbol{\mathcal{M}}\psi - \boldsymbol{d}).$$

- Complex cost function for nonlinear systems.
- MLH solution, hard to solve and to compute error estimates.

Discretisation in time



Assume

- The recursive idea: "Today's posterior is tomorrow's prior"
- Model is first order Markov process.

$$f(\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_k, \boldsymbol{\psi}_0) \propto f(\boldsymbol{\psi}_0) \prod_{i=1}^{k} f(\boldsymbol{\psi}_i | \boldsymbol{\psi}_{i-1}).$$

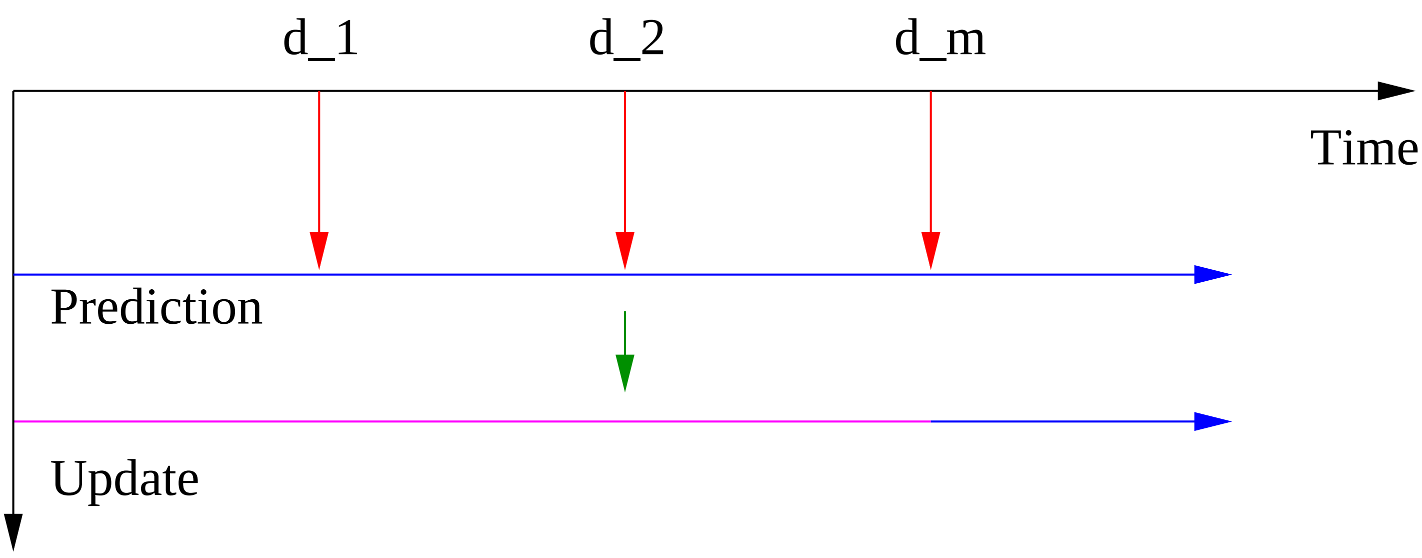
- Measurement errors are independent in time

$$f(\boldsymbol{d}|\boldsymbol{\psi}) = \prod_{j=1}^{m} f(\boldsymbol{d}_j | \boldsymbol{\psi}_{i(j)}).$$

Bayes for discrete state

- Bayes theorem then becomes

$$f(\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_k, \boldsymbol{\psi}_0 | \boldsymbol{d}) \propto f(\boldsymbol{\psi}_0) \prod_{i=1}^{k} f(\boldsymbol{\psi}_i | \boldsymbol{\psi}_{i-1}) \prod_{j=1}^{m} f(\boldsymbol{d}_j | \boldsymbol{\psi}_{i(j)})$$



Rewrite as:

$$f(\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_k, \boldsymbol{\psi}_0 | \boldsymbol{d}) \propto f(\boldsymbol{\psi}_0)$$

$$\prod_{i=1}^{i(1)} f(\boldsymbol{\psi}_i | \boldsymbol{\psi}_{i-1}) f(\boldsymbol{d}_1 | \boldsymbol{\psi}_{i(1)})$$

$$\prod_{i=i(1)+1}^{i(2)} f(\boldsymbol{\psi}_i | \boldsymbol{\psi}_{i-1}) f(\boldsymbol{d}_2 | \boldsymbol{\psi}_{i(2)}) \cdots$$

$$\prod_{i=i(m-1)+1}^{i(m)} f(\boldsymbol{\psi}_i | \boldsymbol{\psi}_{i-1}) f(\boldsymbol{d}_m | \boldsymbol{\psi}_{i(m)})$$

$$\prod_{i=i(m)+1}^{k} f(\boldsymbol{\psi}_i | \boldsymbol{\psi}_{i-1})$$

Sequential processing of measurements

- First update

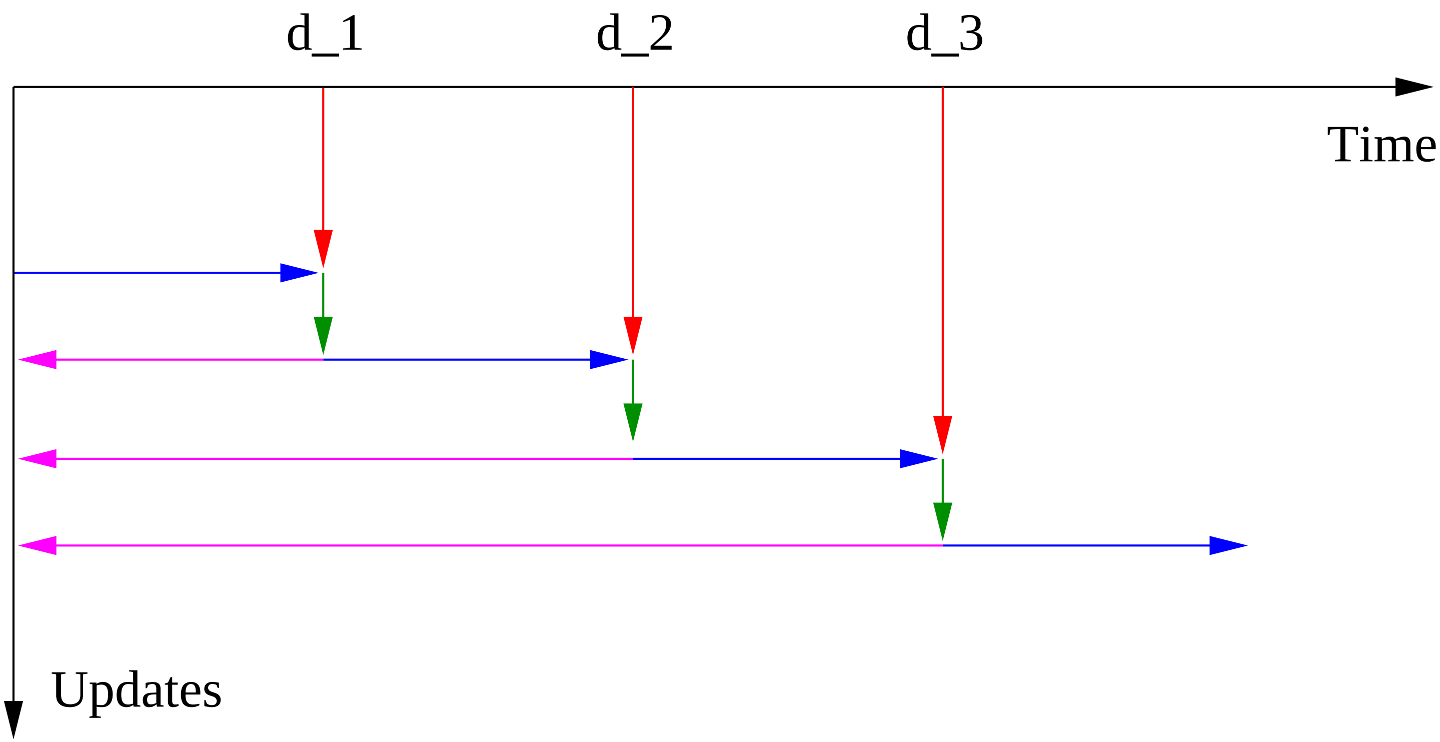
$$f(\psi_1, \ldots, \psi_{i(1)}, \psi_0 | \boldsymbol{d}_1) \propto f(\boldsymbol{\psi}_0) \prod_{i=1}^{i(1)} f(\boldsymbol{\psi}_i | \boldsymbol{\psi}_{i-1}) f(\boldsymbol{d}_1 | \boldsymbol{\psi}_{i(1)})$$

- Second update

$$f(\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_{i(2)}, \boldsymbol{\psi}_0 | \boldsymbol{d}_1, \boldsymbol{d}_2) \propto f(\psi_1, \ldots, \psi_{i(1)}, \psi_0 | \boldsymbol{d}_1) \prod_{i=i(1)+1}^{i(2)} f(\boldsymbol{\psi}_i | \boldsymbol{\psi}_{i-1}) f(\boldsymbol{d}_2 | \boldsymbol{\psi}_{i(2)})$$

Summary

- Independent measurements processed sequentially in time.
- Sequence of inverse problems.
- Solution of one sub-problem is prior for next.
- Hard to solve using traditional variational methods.
- Well suited for ensemble methods.



References

- Funny introduction to Bayesian statistics http://yudkowsky.net/bayes/bayes.html
- Robert and Casella 2004, Monte Carlo Statistical Methods, Springer.