Bayes versus Fisher

Sensor Fusion

Fredrik Gustafsson
fredrik.gustafsson@liu.se

Gustaf Hendeby
gustaf.hendeby@liu.se

Linköping University

Bayes versus Fisher

Thomas Bayes 1701-1761

Ronald Aylmer Fisher 1890-1962

Purpose

Explain the difference of Bayes and Fisher statistics

- There are two schools in statistics: The Bayesian and the frequentist/Fisherian view
- For practitioners, the difference is only philosophical, not big difference
- They are complementary tools: Fisherian methods work well in estimation, Bayesian methods more flexible in filtering.
- There are a few tricks to easily related them to eachother.

Bayes versus Fisher

- Partly philosophical, belief in prior knowledge $p(x)$
- Focus on posterior $p(x|y) = p(y|x)p(x)/p(y)$
- MAP (maximum *a posteriori*) estimate $\hat{x}^{MAP} = \arg\max_x p(x|y)$
- The posteriori distribution gives complete information about the estimation uncertainty, from which e.g., the covariance can be computed.

- Only look at data y for inference about x, everything else is prejudice and gives bias
- Focus on likelihood $p(y|x)$
- ML estimate $\hat{x}^{ML} = \arg\max_x p(y|x)$
- FIM (Fisher Information Matrix) $\mathcal{I}(x)$ is defined in terms of likelihood and can be used to approximate $\mathrm{Cov}(\hat{x}^{ML})$, having the CRLB constraint $\mathrm{Cov}(\hat{x}^{ML}) \geq \mathcal{I}^{-1}(x^0)$.

MAP versus ML

Consider a simple example (special case of a linear state space model):

$$x = v, \qquad \mathrm{Cov}(v) = Q,$$
$$y = x + e, \qquad \mathrm{Cov}(e) = R.$$

The ML estimate is trivial

$$\hat{x}^{ML} = \arg\max_x p(y|x) = y,$$
$$\mathrm{Cov}\left(\hat{x}_k^{ML}\right) = R.$$

Note that the Fisher approach ignores any possible prior that may exists for x.

MAP versus ML

For the MAP estimate, we use the Gaussian distribution and completion of the squares to get

$$\begin{aligned}
\hat{x}^{MAP} &= \arg\max_x p(y|x)p(x) = \arg\min_x -2\log\big(p(y|x)\big) - 2\log\big(p(x)\big)\\
&= \arg\min_x \frac{(y-x)^2}{R} + \frac{x^2}{Q} = \arg\min_x \frac{Qx^2 - 2Qyx + Qy^2 + Rx^2}{QR}\\
&= \cdots = \arg\min_x \frac{Q+R}{QR}\left(x - \frac{Qy}{Q+R}\right)^2 + \frac{y^2}{Q+R}\\
&= \frac{Q}{Q+R}y
\end{aligned}$$

Since $\mathrm{Cov}(y) = R$, the covariance is given by

$$\mathrm{Cov}\left(\hat{x}_k^{MAP}\right) = \frac{Q^2}{(Q+R)^2}R < R$$

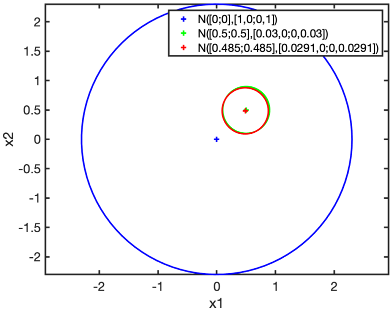
Some Reflections and Generalizations

- The MAP estimate has always smaller covariance than the ML estimate, since the prior adds information.
- Note that $\mathrm{Cov}\left(\hat{x}_k^{MAP}\right) \to \mathrm{Cov}\left(\hat{x}_k^{ML}\right)$ as $Q \to \infty$. That is, MAP will give the same result as ML for a *non-informative prior*.
- The MAP estimate can actually be computed using the sensor fusion formula, where $y_1 = 0 = x - v$ and $y_2 = x + e$ are used. Here, y_1 is seen as a *virtual measurement* of x.

Example

- Consider the simple scalar example repeated in two independent dimensions
- Define a rather uninformative prior (Q large) for x
- Let the measurement noise be much smaller ($R = 0.03$)
- Apply the fusion formula to get the MAP estimate and its covariance
- Illustrate with confidence ellipsoids

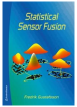
```
x=ndist([0;0],eye(2))
y=ndist([0.5;0.5],0.03*eye(2))
xhat=fusion(x,y)
plot2(x,y,xhat)
axis('equal')
```



Summary

MAP versus ML

- ML is a special case of MAP when using a non-informative prior.
- The prior can be seen as a virtual measurement.
- The MAP estimate can be computed with the sensor fusion formula for the real and virtual measurement.



This is an introduction to Part II