Some Thoughts on Filtering

Dominic Steinitz
@idontgetoutmuch
https://idontgetoutmuch.wordpress.com

Tuesday 9 October 18

Outline

Apollo 8 launched on December 21, 1968

03:17:45:17 (Dec. 25, 1968, 6:36 a.m. UTC)

Jim Lovell (Commander Module Pilot) : Roger. Do you wish me to reinitialize the W-matrix at this time?

03:17:45:26

Ken Mattingly (CAPCOM) : Affirmative, Apollo 8

Setting

Samples

- I can sample from a normal distribution with the mean known only to me and with a publicly known variance.
- Based on your knowledge about me you have a (prior) belief about the mean of this distribution: it's centred at 0 and is probably between -3 and +3.
- At time 1, I give you some information: a sample from my distribution.
- At time 2, I give you more information: another sample from my distribution.
- And so on . . .

Bayes' Theorem

$$\mathbb{P}(\theta \mid D) \triangleq \frac{\mathbb{P}(\theta \cap D)}{\mathbb{P}(D)}$$

Also

$$\mathbb{P}(D \mid \theta) \triangleq \frac{\mathbb{P}(\theta \cap D)}{\mathbb{P}(\theta)}$$

Thus

$$\mathbb{P}(\theta \mid D) \propto \mathbb{P}(D \mid \theta)\mathbb{P}(\theta)$$

An Application of Bayes' Theorem

I have told you that the data are drawn from

$$\mathbb{P}(x \,|\, \mu) \propto \exp -\frac{(x-\mu)^2}{2\sigma^2}$$

And, based on your knowledge about me, you have a prior

$$\mathbb{P}(\mu) \propto \exp -\frac{(\mu-\mu_0)^2}{2\sigma_0^2}$$

This gives

$$\mathbb{P}(\mu \,|\, x) \propto \exp -\frac{(x-\mu)^2}{2\sigma^2} \times \exp -\frac{(\mu-\mu_0)^2}{2\sigma_0^2}$$

An Application of Bayes' Theorem

$$\mathbb{P}(\mu \mid x) \propto \exp -\frac{(x-\mu)^2}{2\sigma^2} \times \exp -\frac{(\mu-\mu_0)^2}{2\sigma_0^2}$$

An Application of Bayes' Theorem

$$
\begin{aligned}
\mathbb{P}(\mu \mid x) &\propto \exp -\frac{(x-\mu)^2}{2\sigma^2} \times \exp -\frac{(\mu-\mu_0)^2}{2\sigma_0^2} \\
&= \exp -[\frac{x^2 - 2\mu x + \mu^2}{2\sigma^2} + \frac{\mu^2 - 2\mu\mu_0 + \mu_0^2}{2\sigma_0^2}]
\end{aligned}
$$

An Application of Bayes' Theorem

$$
\begin{aligned}
\mathbb{P}(\mu \mid x) &\propto \exp - \frac{(x-\mu)^2}{2\sigma^2} \times \exp - \frac{(\mu-\mu_0)^2}{2\sigma_0^2} \\
&= \exp - [\frac{x^2 - 2\mu x + \mu^2}{2\sigma^2} + \frac{\mu^2 - 2\mu\mu_0 + \mu_0^2}{2\sigma_0^2}] \\
&= \exp - [\mu^2(\frac{1}{2\sigma^2} + \frac{1}{2\sigma_0^2}) - 2\mu(\frac{x}{2\sigma^2} + \frac{\mu_0}{2\sigma_0^2}) + (\frac{x^2}{2\sigma^2} + \frac{\mu_0^2}{2\sigma_0^2})]
\end{aligned}
$$

An Application of Bayes' Theorem

$$
\begin{aligned}
\mathbb{P}(\mu \mid x) &\propto \exp -\frac{(x-\mu)^2}{2\sigma^2} \times \exp -\frac{(\mu-\mu_0)^2}{2\sigma_0^2} \\
&= \exp -[\frac{x^2 - 2\mu x + \mu^2}{2\sigma^2} + \frac{\mu^2 - 2\mu\mu_0 + \mu_0^2}{2\sigma_0^2}] \\
&= \exp -[\mu^2(\frac{1}{2\sigma^2} + \frac{1}{2\sigma_0^2}) - 2\mu(\frac{x}{2\sigma^2} + \frac{\mu_0}{2\sigma_0^2}) + (\frac{x^2}{2\sigma^2} + \frac{\mu_0^2}{2\sigma_0^2})] \\
&\triangleq \exp -[\frac{1}{2\sigma_1^2}(\mu^2 - 2\mu\mu_1 + \mu_1^2)]
\end{aligned}
$$

An Application of Bayes' Theorem

$$\exp -[\mu^2(\frac{1}{2\sigma^2} + \frac{1}{2\sigma_0^2}) - 2\mu(\frac{x}{2\sigma^2} + \frac{\mu_0}{2\sigma_0^2}) + (\frac{x^2}{2\sigma^2} + \frac{\mu_0^2}{2\sigma_0^2})] \triangleq \exp -[\frac{1}{2\sigma_1^2}(\mu^2 - 2\mu\mu_1 + \mu_1^2)]$$

Collecting together terms

An Application of Bayes' Theorem

$$\exp -[\mu^2(\frac{1}{2\sigma^2} + \frac{1}{2\sigma_0^2}) - 2\mu(\frac{x}{2\sigma^2} + \frac{\mu_0}{2\sigma_0^2}) + (\frac{x^2}{2\sigma^2} + \frac{\mu_0^2}{2\sigma_0^2})] \triangleq$$
$$\exp -[\frac{1}{2\sigma_1^2}(\mu^2 - 2\mu\mu_1 + \mu_1^2)]$$

Collecting together terms

$$\frac{1}{\sigma_1^2} = \frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}$$

An Application of Bayes' Theorem

$$\exp -[\mu^2(\frac{1}{2\sigma^2} + \frac{1}{2\sigma_0^2}) - 2\mu(\frac{x}{2\sigma^2} + \frac{\mu_0}{2\sigma_0^2}) + (\frac{x^2}{2\sigma^2} + \frac{\mu_0^2}{2\sigma_0^2})] \triangleq \exp -[\frac{1}{2\sigma_1^2}(\mu^2 - 2\mu\mu_1 + \mu_1^2)]$$

Collecting together terms

$$\frac{1}{\sigma_1^2} = \frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}$$
$$\frac{\mu_1}{\sigma_1^2} = \frac{\mu_0}{\sigma_0^2} + \frac{x}{\sigma^2}$$

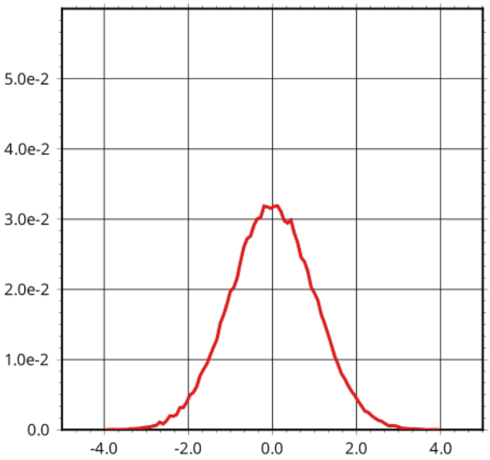
Prior by Sampling

We can also represent the prior as a *histogram*

$\mu_0, \sigma_0 :: \mathit{Double}$
$\mu_0 = 0.0$
$\sigma_0 = 1.00$

$\mathit{priors} :: \mathit{Histogram}\ \mathit{BinD}\ \mathit{Double}$
$\mathit{priors} = \mathit{hist}\ \$\ \mathit{runSampler}\ (\mathit{normal}\ \mu_0\ \sigma_0)\ 42\ 100000$

Prior



Secret, Model and Data

But in reality the data come from a normal distribution with mean and (known) variance

$$\mu :: \textit{Double}$$
$$\mu = 1.00$$

$$\sigma :: \textit{Double}$$
$$\sigma = 0.9$$

Secret, Model and Data

But in reality the data come from a normal distribution with mean and (known) variance

$$\mu :: \mathit{Double}$$
$$\mu = 1.00$$

$$\sigma :: \mathit{Double}$$
$$\sigma = 0.9$$

The function in Haskell for the likelihood $\mathbb{P}(x \mid \mu)$

$$\begin{aligned}
&\mathit{likelihood} :: \mathit{Double} \to \mathit{Double} \to \mathit{Double} \\
&\mathit{likelihood}\ x\ \mu = n \mathbin{/} d \\
&\quad \textbf{where} \\
&\qquad n = \mathit{exp}\ (-(x - \mu) \uparrow 2 \mathbin{/} (2 * \sigma \uparrow 2)) \\
&\qquad d = \mathit{sqrt}\ (2 * \pi * \sigma \uparrow 2)
\end{aligned}$$

Secret, Model and Data

But in reality the data come from a normal distribution with mean and (known) variance

$$\begin{aligned}
&\mu :: \mathit{Double} \\
&\mu = 1.00 \\
&\sigma :: \mathit{Double} \\
&\sigma = 0.9
\end{aligned}$$

The function in Haskell for the likelihood $\mathbb{P}(x \mid \mu)$

$$\begin{aligned}
&\mathit{likelihood} :: \mathit{Double} \to \mathit{Double} \to \mathit{Double} \\
&\mathit{likelihood}\ x\ \mu = n \,/\, d \\
&\quad \textbf{where} \\
&\qquad n = \mathit{exp}\ (-(x - \mu) \uparrow 2 \,/\, (2 * \sigma \uparrow 2)) \\
&\qquad d = \mathit{sqrt}\ (2 * \pi * \sigma \uparrow 2)
\end{aligned}$$

Finally I give you some noisy data

$$\begin{aligned}
&\mathit{ds} :: [\mathit{Double}] \\
&\mathit{ds} = \mathit{runSampler}\ (\mathit{normal}\ \mu\ \sigma)\ 2\ 10
\end{aligned}$$

Apply Bayes'

Now you can use Bayes' Theorem to create a posterior

$$
\begin{aligned}
&\mathit{posteriorize} :: \mathit{Histogram}\ \mathit{BinD}\ \mathit{Double} \rightarrow \\
&\qquad\qquad\quad \mathit{Double} \rightarrow \\
&\qquad\qquad\quad \mathit{Histogram}\ \mathit{BinD}\ \mathit{Double} \\
&\mathit{posteriorize}\ h\ x = H.\mathit{bmap}\ \mathit{bayes}\ h \\
&\quad \textbf{where} \\
&\qquad \mathit{bayes} :: \mathit{Double} \rightarrow \mathit{Double} \rightarrow \mathit{Double} \\
&\qquad \mathit{bayes}\ \theta\ p = p * \mathit{likelihood}\ x\ \theta
\end{aligned}
$$

Apply Bayes'

Now you can use Bayes' Theorem to create a posterior

$$
\begin{aligned}
&\mathit{posteriorize} :: \mathit{Histogram}\ \mathit{BinD}\ \mathit{Double} \to \\
&\qquad\qquad \mathit{Double} \to \\
&\qquad\qquad \mathit{Histogram}\ \mathit{BinD}\ \mathit{Double} \\
&\mathit{posteriorize}\ h\ x = H.\mathit{bmap}\ \mathit{bayes}\ h \\
&\quad \textbf{where} \\
&\qquad \mathit{bayes} :: \mathit{Double} \to \mathit{Double} \to \mathit{Double} \\
&\qquad \mathit{bayes}\ \theta\ p = p * \mathit{likelihood}\ x\ \theta
\end{aligned}
$$

$$
\begin{aligned}
&\mathit{qs} :: [\mathit{Histogram}\ \mathit{BinD}\ \mathit{Double}] \\
&\mathit{qs} = \mathit{scanl}\ \mathit{posteriorize}\ \mathit{priors}\ \mathit{ds}
\end{aligned}
$$

Apply Bayes'

Now you can use Bayes' Theorem to create a posterior

$$
\begin{aligned}
&\mathit{posteriorize} :: \mathit{Histogram}\ \mathit{BinD}\ \mathit{Double} \to \\
&\qquad\qquad\qquad \mathit{Double} \to \\
&\qquad\qquad\qquad \mathit{Histogram}\ \mathit{BinD}\ \mathit{Double} \\
&\mathit{posteriorize}\ h\ x = H.\mathit{bmap}\ \mathit{bayes}\ h \\
&\quad \textbf{where} \\
&\qquad \mathit{bayes} :: \mathit{Double} \to \mathit{Double} \to \mathit{Double} \\
&\qquad \mathit{bayes}\ \theta\ p = p * \mathit{likelihood}\ x\ \theta
\end{aligned}
$$

$$
\begin{aligned}
&\mathit{qs} :: [\mathit{Histogram}\ \mathit{BinD}\ \mathit{Double}] \\
&\mathit{qs} = \mathit{scanl}\ \mathit{posteriorize}\ \mathit{priors}\ \mathit{ds}
\end{aligned}
$$

$$
\begin{aligned}
&\mathit{ss} :: [\mathit{Double}] \\
&\mathit{ss} = \mathit{map}\ H.\mathit{sum}\ \mathit{qs}
\end{aligned}
$$

Apply Bayes'

Now you can use Bayes' Theorem to create a posterior

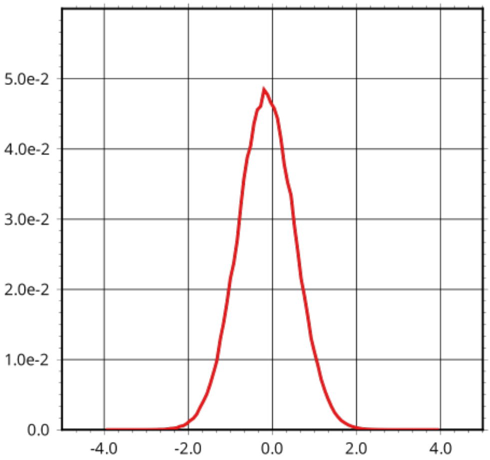
$$
\begin{aligned}
&\mathit{posteriorize} :: \mathit{Histogram}\ \mathit{BinD}\ \mathit{Double} \to \\
&\qquad\qquad\qquad \mathit{Double} \to \\
&\qquad\qquad\qquad \mathit{Histogram}\ \mathit{BinD}\ \mathit{Double} \\
&\mathit{posteriorize}\ h\ x = H.\mathit{bmap}\ \mathit{bayes}\ h \\
&\quad \textbf{where} \\
&\qquad \mathit{bayes} :: \mathit{Double} \to \mathit{Double} \to \mathit{Double} \\
&\qquad \mathit{bayes}\ \theta\ p = p * \mathit{likelihood}\ x\ \theta
\end{aligned}
$$

$$
\begin{aligned}
&\mathit{qs} :: [\mathit{Histogram}\ \mathit{BinD}\ \mathit{Double}] \\
&\mathit{qs} = \mathit{scanl}\ \mathit{posteriorize}\ \mathit{priors}\ \mathit{ds}
\end{aligned}
$$

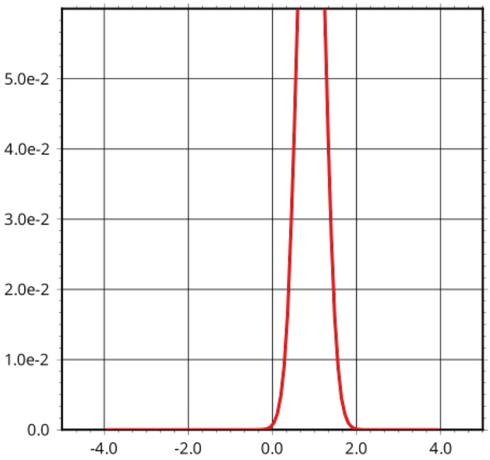
$$
\begin{aligned}
&\mathit{ss} :: [\mathit{Double}] \\
&\mathit{ss} = \mathit{map}\ H.\mathit{sum}\ \mathit{qs}
\end{aligned}
$$

$$
\begin{aligned}
&\mathit{ns} :: [\mathit{Histogram}\ \mathit{BinD}\ \mathit{Double}] \\
&\mathit{ns} = \mathit{zipWith}\ (\lambda s\ q \to H.\mathit{map}\ (/s)\ q)\ \mathit{ss}\ \mathit{qs}
\end{aligned}
$$

After 1 Observation



After 10 Observations



Observing Apollo (in 1D)

We want to track the lunar module

Newton's second law of motion

$$ma = F$$

Observing Apollo (in 1D)

We want to track the lunar module

Newton's second law of motion

$$ma = F$$

As first order equation system

$$\frac{\mathrm{d}v}{\mathrm{d}t} = F$$
$$\frac{\mathrm{d}x}{\mathrm{d}t} = v$$

Mathematical Model

Let's assume there are no forces acting on the lunar module.

Discretizing

$$\frac{v_k - v_{k-1}}{\Delta t} = 0$$
$$\frac{x_k - x_{k-1}}{\Delta t} = v_k$$

Writing x_1 for x and x_2 for v

Mathematical Model

Let's assume there are no forces acting on the lunar module.

Discretizing

$$\frac{v_k - v_{k-1}}{\Delta t} = 0$$
$$\frac{x_k - x_{k-1}}{\Delta t} = v_k$$

Writing x_1 for x and x_2 for v

$$\begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \end{bmatrix} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1^{(k-1)} \\ x_2^{(k-1)} \end{bmatrix}$$

Mathematical Model

Let's assume there are no forces acting on the lunar module.

Discretizing

$$\frac{v_k - v_{k-1}}{\Delta t} = 0$$
$$\frac{x_k - x_{k-1}}{\Delta t} = v_k$$

Writing x_1 for x and x_2 for v

$$\begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \end{bmatrix} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1^{(k-1)} \\ x_2^{(k-1)} \end{bmatrix} + \boldsymbol{Q}_k$$

Can only observe position

Observing

$$\begin{bmatrix} y_1^{(k)} \end{bmatrix} = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \end{bmatrix}$$

Can only observe position

Observing

$$\begin{bmatrix} y_1^{(k)} \end{bmatrix} = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \end{bmatrix} + \boldsymbol{R}_k$$

In vector notation

$$\boldsymbol{x}_k = \boldsymbol{A}\boldsymbol{x}_{k-1} + \boldsymbol{Q}_k$$
$$\boldsymbol{y}_k = \boldsymbol{H}\boldsymbol{x}_k + \boldsymbol{R}_k$$

Sampling

I can sample a whole path from this model from which I will reveal a step at a time.

Sampling

I can sample a whole path from this model from which I will reveal a step at a time.

$$
\begin{aligned}
&\boldsymbol{x}_{\mathrm{init}} :: M.\mathit{Vector}\ \mathit{Double} \\
&\boldsymbol{x}_{\mathrm{init}} = M.\mathit{fromList}\ [0, 1] \\
&\mathit{apolloSamples} :: [(M.\mathit{Vector}\ \mathit{Double}, M.\mathit{Vector}\ \mathit{Double})] \\
&\mathit{apolloSamples} = \mathit{evalState}\ (\mathit{ML.unfoldrM}\ \mathit{apolloSample}\ \boldsymbol{x}_{\mathrm{init}}) \\
&\qquad\qquad\qquad\qquad\qquad (\mathit{pureMT}\ 17) \\
&\quad \textbf{where} \\
&\qquad \mathit{apolloSample}\ \boldsymbol{x}^{\flat} = \textbf{do} \\
&\qquad\quad \boldsymbol{x} \leftarrow \mathit{sample}\ \$\ \mathit{rvar}\ (G.\mathit{Normal}\ (\boldsymbol{A}\ \#{>}\ \boldsymbol{x}^{\flat})\ \boldsymbol{Q}) \\
&\qquad\quad \boldsymbol{y} \leftarrow \mathit{sample}\ \$\ \mathit{rvar}\ (G.\mathit{Normal}\ (\boldsymbol{H}\ \#{>}\ \boldsymbol{x})\ \boldsymbol{R}) \\
&\qquad\quad \mathit{return}\ \$\ \mathit{Just}\ ((\boldsymbol{x}, \boldsymbol{y}), \boldsymbol{x})
\end{aligned}
$$

Prior

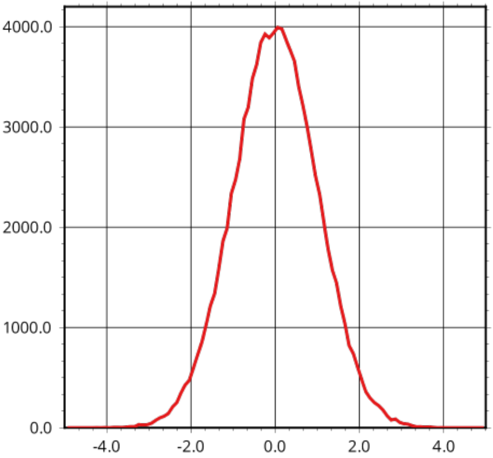
Based on your knowledge about Apollo you have a (prior) belief about the mean of this distribution: it's centred at $(0, 1)$ and both the position and velocity are probably ± 3.

$$
\begin{aligned}
&\boldsymbol{m}_0 :: \mathit{M.Vector\ Double} \\
&\boldsymbol{m}_0 = \mathit{M.fromList}\ [0, 1] \\
&\boldsymbol{\Sigma}_0 :: \mathit{M.Herm\ Double} \\
&\boldsymbol{\Sigma}_0 = \mathit{M.sym}\ \$\ (2 >< 2)\ [1.0, 0.0, \\
&\qquad\qquad\qquad\qquad\quad 0.0, 1.0] \\
&\mathit{priorsApollo} :: \mathit{Histogram}\ (\mathit{Bin2D\ BinD\ BinD})\ \mathit{Double} \\
&\mathit{priorsApollo} = \mathit{hist2}\ \$\ \mathit{conv}\ \$ \\
&\quad \mathit{prePriors}\ 2\ 100000 \\
&\quad \textbf{where}\ \mathit{prePriors\ seed\ n} = \\
&\qquad \mathit{evalState}\ (\mathit{replicateM}\ n\ (\mathit{sample}\ \$\ \mathit{G.Normal}\ \boldsymbol{m}_0\ \boldsymbol{\Sigma}_0)) \\
&\qquad\qquad\qquad (\mathit{pureMT}\ (\mathit{fromIntegral\ seed}))
\end{aligned}
$$

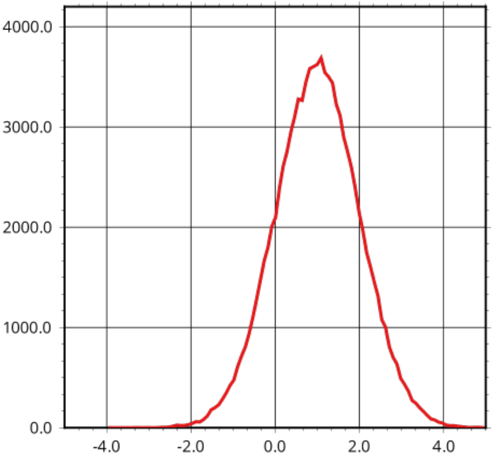
An Aside: Marginals

$$
\begin{aligned}
&marginalX, marginalY :: Histogram\ BinD\ Double \\
&marginalX = H.reduceX\ H.sum\ priorsApollo \\
&marginalY = H.reduceY\ H.sum\ priorsApollo
\end{aligned}
$$

Marginal Prior for Position



Marginal Prior for Velocity



Model

Recall the state update

$$\boldsymbol{x}_k = \boldsymbol{A}\boldsymbol{x}_{k-1} + \boldsymbol{Q}_k$$

Model

Recall the state update

$$\boldsymbol{x}_k = \boldsymbol{A}\boldsymbol{x}_{k-1} + \boldsymbol{Q}_k$$

In Haskell

$$\begin{aligned}
&\mathit{newState} :: \mathit{MonadRandom}\ m \Rightarrow \\
&\quad \mathit{M.Vector}\ \mathit{Double} \to m\ (\mathit{M.Vector}\ \mathit{Double}) \\
&\mathit{newState}\ \boldsymbol{x}^{\flat} = \mathit{sample}\ \$\ \mathit{rvar}\ (\mathit{G.Normal}\ (\boldsymbol{A}\ \#\!>\ \boldsymbol{x}^{\flat})\ \boldsymbol{Q})
\end{aligned}$$

Model

Recall the state update

$$\boldsymbol{x}_k = \boldsymbol{A}\boldsymbol{x}_{k-1} + \boldsymbol{Q}_k$$

In Haskell

$$
\begin{aligned}
&\mathit{newState} :: \mathit{MonadRandom}\ m \Rightarrow \\
&\quad M.\mathit{Vector}\ \mathit{Double} \rightarrow m\ (M.\mathit{Vector}\ \mathit{Double}) \\
&\mathit{newState}\ \boldsymbol{x}^\flat = \mathit{sample}\ \$\ \mathit{rvar}\ (G.\mathit{Normal}\ (\boldsymbol{A}\ \#\!>\ \boldsymbol{x}^\flat)\ \boldsymbol{Q})
\end{aligned}
$$

Recall the observations are given by

$$\boldsymbol{y}_k = \boldsymbol{H}\boldsymbol{x}_k + \boldsymbol{R}_k$$

Model

Recall the state update

$$\boldsymbol{x}_k = \boldsymbol{A}\boldsymbol{x}_{k-1} + \boldsymbol{Q}_k$$

In Haskell

$$\begin{aligned}&\mathit{newState} :: \mathit{MonadRandom}\ m \Rightarrow\\ &\quad \mathit{M.Vector}\ \mathit{Double} \rightarrow m\ (\mathit{M.Vector}\ \mathit{Double})\\ &\mathit{newState}\ \boldsymbol{x}^{\flat} = \mathit{sample}\ \$\ \mathit{rvar}\ (\mathit{G.Normal}\ (\boldsymbol{A}\ \#\!>\ \boldsymbol{x}^{\flat})\ \boldsymbol{Q})\end{aligned}$$

Recall the observations are given by

$$\boldsymbol{y}_k = \boldsymbol{H}\boldsymbol{x}_k + \boldsymbol{R}_k$$

Thus the likelihood of observation b given the state a is given by

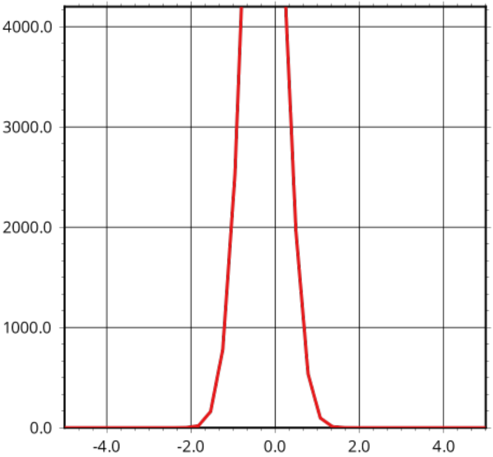
$$\begin{aligned}&\mathit{likelihoodA} :: \mathit{M.Vector}\ \mathit{Double} \rightarrow \mathit{M.Vector}\ \mathit{Double} \rightarrow \mathit{Double}\\ &\mathit{likelihoodA}\ a\ b = \mathit{pdf}\ (\mathit{G.Normal}\ (\boldsymbol{H}\ \#\!>\ a)\ \boldsymbol{R})\ b\end{aligned}$$

Apply Bayes'

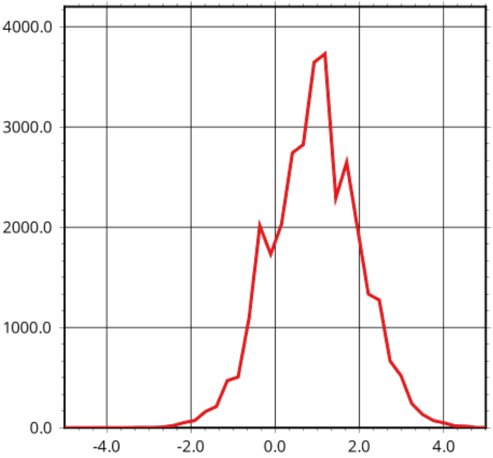
Now you can use Bayes' to track the path of the apollo.

```
posteriorizeA :: MonadRandom m ⇒
  Histogram (Bin2D BinD BinD) Double →
  M.Vector Double →
  m (Histogram (Bin2D BinD BinD) Double)
posteriorizeA q d = do
  let xfs = H.asList q
      xs = map pair2Vec $ map fst xfs
      fs = map snd xfs
  x ← mapM newState xs
  let newQ = hist2' $ zip (map vec2Pair x) fs
  return $ H.bmap (λ(u, v) p → p ∗ likelihoodA (M.vector [u, v]) d) newQ
```

Position Estimate After 1 Observation



Velocity Estimate After 1 Observation



Marginals Again

```
*Test> take 1 apolloSamples
[([9.99586e-2,0.75767],[-0.36189])]
```

$$
\begin{aligned}
&\mathit{marginal1X}, \mathit{marginal1Y} :: \mathit{Histogram\ BinD\ Double}\\
&\mathit{marginal1X} = \mathit{H.reduceX\ H.sum\ test}\\
&\mathit{marginal1Y} = \mathit{H.reduceY\ H.sum\ test}
\end{aligned}
$$

$$
\begin{aligned}
&\mathit{m0X1}, \mathit{m1X1}, \mathit{muX1} :: \mathit{Double}\\
&\mathit{m0X1} = \mathit{H.sum\ marginal1X}\\
&\mathit{m1X1} = \mathit{H.sum}\,\$\,\mathit{H.bmap}\,(\lambda f\ v \to v * f)\ \mathit{marginal1X}
\end{aligned}
$$

$$
\begin{aligned}
&\mathit{muY1} = \mathit{m1Y1}\ /\ \mathit{m0Y1}\\
&\mathit{muX1} = \mathit{m1X1}\ /\ \mathit{m0X1}
\end{aligned}
$$

```
*Test> muX1
0.94108
*Test> muY1
-0.26897
```

Recall the Model

Wait a Minute . . .

Compute power in 1968?

Recall the Model

Wait a Minute ...

Compute power in 1968?

Recall our model:

$$\boldsymbol{x}_i = \boldsymbol{A}_{i-1}\boldsymbol{x}_{i-1} + \boldsymbol{Q}_{i-1}$$
$$\boldsymbol{y}_i = \boldsymbol{H}_i\boldsymbol{x}_i + \boldsymbol{R}_i$$

where

$$\boldsymbol{Q}_i \sim \mathcal{N}\left(0, \boldsymbol{\Sigma}_i^{(x)}\right), \quad \boldsymbol{R}_i \sim \mathcal{N}\left(0, \boldsymbol{\Sigma}_i^{(y)}\right)$$

The model is **linear** and the errors are **Gaussian**

Pólya

One way to solve a problem is to generalize it

$$
\begin{aligned}
x_k &\sim p(x_k \mid x_{k-1}) \\
y_k &\sim p(y_k \mid x_k)
\end{aligned}
$$

- State update is Markovian
- The past is independent of the future given the present
- The measurement given the current state is independent of the past (quasi-Markovian)

Pólya

One way to solve a problem is to generalize it

$$
\begin{aligned}
x_k &\sim p(x_k \,|\, x_{k-1}) \\
y_k &\sim p(y_k \,|\, x_k)
\end{aligned}
$$

- State update is Markovian
- The past is independent of the future given the present
- The measurement given the current state is independent of the past (quasi-Markovian)

Re-writing our linear Gaussian model

$$
\begin{aligned}
p(x_k \,|\, x_{k-1}) &= \mathcal{N}(x_k \,|\, A_{k-1}x_{k-1}, Q_k) \\
p(y_k \,|\, x_k) &= \mathcal{N}(y_k \,|\, H_k x_k, R_k)
\end{aligned}
$$

Bayesian Filtering Equations

What we'd like is

$$p(x_k \mid y_{1:k})$$

$$\begin{aligned} p(x_k, x_{k-1} \mid y_{1:k-1}) &= p(x_k \mid x_{k-1}, y_{1:k-1}) p(x_{k-1} \mid y_{1:k-1}) \\ &= p(x_k \mid x_{k-1}) p(x_{k-1} \mid y_{1:k-1}) \end{aligned}$$

Marginalising

$$p(x_k \mid y_{1:k-1}) = \int p(x_k \mid x_{k-1}) p(x_{k-1} \mid y_{1:k-1}) \mathrm{d}x_{k-1}$$

A Histogram Interpretation

We can think of our histograms as roughly being a distribution

$$G = \sum w_i \delta(x^{(i)})$$

where the x_i are the midpoints of the cells and the w_i are the (normalised) number of observations in the cell and δ is the Dirac delta "function".

$$\begin{aligned} p(x_k \mid y_{1:k-1}) &= \int p(x_k \mid x_{k-1}) p(x_{k-1} \mid y_{1:k-1}) \mathrm{d}x_{k-1} \\ &= \sum p(x_k \mid x_{k-1}) w_i \delta(x_{k-1}^{(i)}) \\ &= \sum w_i \delta(x_k^{(i)}) \end{aligned}$$

where $x_k^{(i)} \sim p(x_k \mid x_k^{(i-1)})$

Bayesian Filtering Equations

$$
\begin{aligned}
p(x_k \mid y_{1:k}) &= \frac{1}{Z_k} p(y_k \mid x_k, y_{1:k-1}) p(x_k \mid y_{1:k-1}) \\
&= \frac{1}{Z_k} p(y_k \mid x_k) p(x_k \mid y_{1:k-1})
\end{aligned}
$$

In histogram terms

$$
\begin{aligned}
p(x_k^{(i)} \mid y_{1:k}) &= \frac{1}{Z_k} p(y_k \mid x_k^{(i-1)}) w_i \delta(x_{k-1}^{(i)}) \\
&= \frac{1}{Z_k} w_i' \delta(x_{k-1}^{(i)})
\end{aligned}
$$

where $w_i' = \frac{1}{Z_k} p(y_k \mid x_k^{(i-1)}) w_i$ and $\frac{1}{Z_k} = \sum w_i'$.

Kalman Itself

A **lot** of algebraic manipulation gives the optimal solution.

Prediction Step

$$\hat{\boldsymbol{x}}_i^\flat = \boldsymbol{A}_{i-1}\hat{\boldsymbol{x}}_{i-1}$$

$$\hat{\boldsymbol{\Sigma}}_i^\flat = \boldsymbol{A}_{i-1}\hat{\boldsymbol{\Sigma}}_{i-1}\boldsymbol{A}_{i-1}^\top + \boldsymbol{Q}_{i-1}$$

Correction Step

$$\boldsymbol{v}_i = \boldsymbol{y}_i - \boldsymbol{H}_i\hat{\boldsymbol{x}}_i^\flat$$

$$\boldsymbol{S}_i = \boldsymbol{H}_i\hat{\boldsymbol{\Sigma}}_i^\flat\boldsymbol{H}_i^\top + \boldsymbol{R}_i$$

$$\boldsymbol{K}_i = \hat{\boldsymbol{\Sigma}}_i^\flat\boldsymbol{H}_i^\top\boldsymbol{S}_i^{-1}$$

$$\hat{\boldsymbol{x}}_i = \hat{\boldsymbol{x}}_i^\flat + \boldsymbol{K}_i\boldsymbol{v}_i$$

$$\hat{\boldsymbol{\Sigma}}_i = \hat{\boldsymbol{\Sigma}}_i^\flat - \boldsymbol{K}_i\boldsymbol{S}_i\boldsymbol{K}_i^\top$$

Original Bayes' Theorem Example

Re-express Original

$$x_k = x_{k-1} + q_{k-1}, \quad q_{k-1} \sim \mathcal{N}(0, 0)$$
$$y_k = x_k + r_k, \quad r_k \sim \mathcal{N}(0, R)$$

Prediction Step

$$\hat{x}_i^\flat = \hat{x}_{i-1} \quad \hat{\Sigma}_i^\flat = \hat{\Sigma}_{i-1}$$

Correction Step

$$v_i = y_i - \hat{x}_i^\flat \quad S_i = \hat{\Sigma}_i^\flat + R_i \quad K_i = \frac{\hat{\Sigma}_i^\flat}{S_i}$$
$$\hat{x}_i = \hat{x}_i^\flat + K_i v_i \quad \hat{\Sigma}_i = \hat{\Sigma}_i^\flat - K_i S_i K$$

Original Bayes' Theorem Example

Prediction Step

$$\hat{x}_i^\flat = \hat{x}_{i-1} \quad \hat{\Sigma}_i^\flat = \hat{\Sigma}_{i-1}$$

$$\hat{x}_1^\flat = \hat{x}_0 = \mu_0 \quad \hat{\Sigma}_1^\flat = \hat{\Sigma}_0 = \sigma_0^2$$

Correction Step

$$v_i = y_i - \hat{x}_i^\flat \quad S_i = \hat{\Sigma}_i^\flat + R_i \quad K_i = \frac{\hat{\Sigma}_i^\flat}{S_i}$$

$$\hat{x}^i = \hat{x}_i^\flat + K_i v_i \quad \sigma_1^2 = \hat{\Sigma}_i^\flat - K_i S_i K$$

Correction Step

$$v = y - \mu_0 \quad S = \sigma_0^2 + \sigma^2 \quad K = \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}$$

$$\mu_1 = \mu_0 + Kv \quad \sigma_1^2 = \sigma_0^2 - KSK$$

Original Bayes' Theorem Example

Correction Step

$$v = y - \mu_0 \quad S = \sigma_0^2 + \sigma^2 \quad K = \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}$$

$$\mu_1 = \mu_0 + Kv \quad \sigma_1^2 = \sigma_0^2 - KSK$$

Correction Step

$$\mu_1 = \mu_0 + \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}(y - \mu_0) \quad \sigma_1^2 = \sigma_0^2 - \frac{(\sigma_0^2)^2}{\sigma_0^2 + \sigma^2}$$

Original Bayes' Theorem Example

Correction Step

$$v = y - \mu_0 \quad S = \sigma_0^2 + \sigma^2 \quad K = \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}$$

$$\mu_1 = \mu_0 + Kv \quad \sigma_1^2 = \sigma_0^2 - KSK$$

Correction Step

$$\mu_1 = \mu_0 + \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}(y - \mu_0) \quad \sigma_1^2 = \sigma_0^2 - \frac{(\sigma_0^2)^2}{\sigma_0^2 + \sigma^2}$$

Correction Step

$$\frac{\mu_1}{\sigma_1^2} = \frac{\mu_0}{\sigma_0^2} + \frac{\mu}{\sigma^2} \quad \frac{1}{\sigma_1^2} = \frac{1}{\sigma^2} + \frac{1}{\sigma_0^2}$$

Further Reading

- Simo Särkkä's book "Bayesian Filtering and Smoothing"
- Libbi: http://libbi.org
- Haskell package `random-fu` for random variables
- Haskell package `histogram-fill` for histograms
- Haskell package `kalman` for kalman, extended kalman, unscented kalman, particle filtering and smoothing
- Same in Python: https://filterpy.readthedocs.io/en/latest
- Haskell package `hmatrix` for statically typed matrices