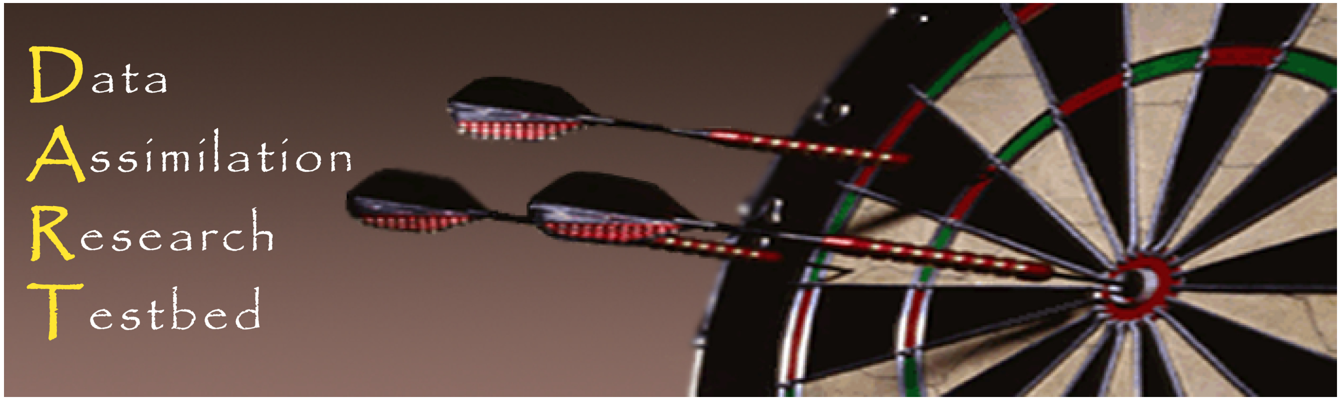
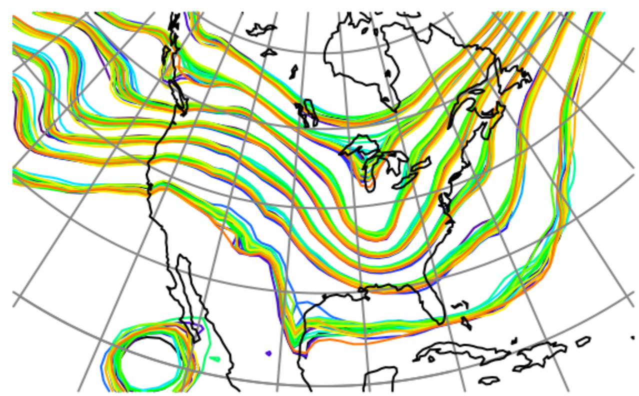


DART Tutorial Part 1:
Filtering For a One Variable System

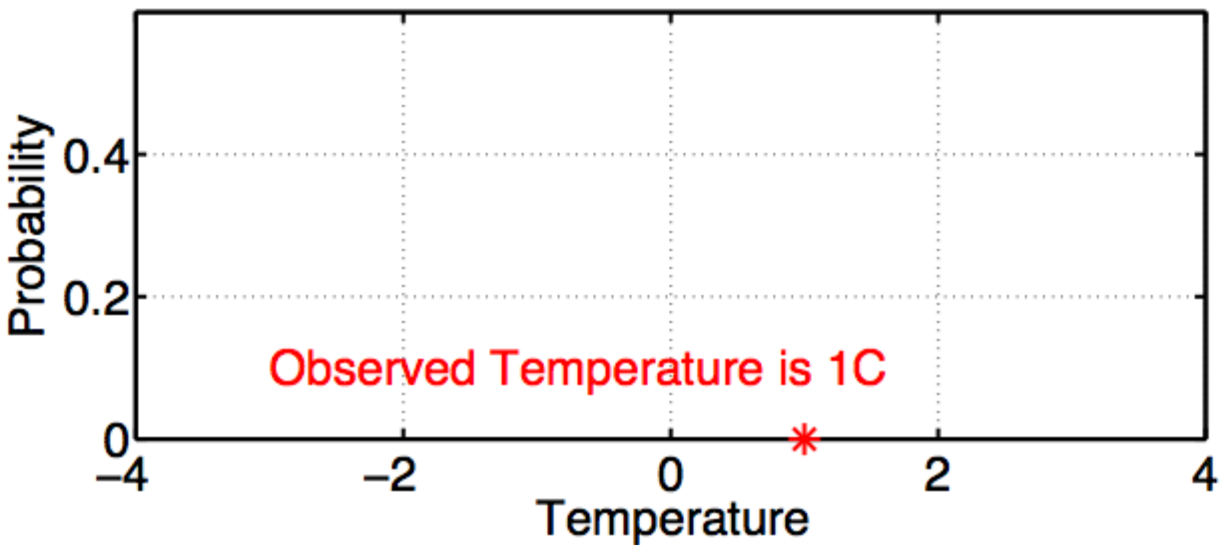


The National Center for Atmospheric Research is sponsored by the National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

NCAR | National Center for
UCAR | Atmospheric Research

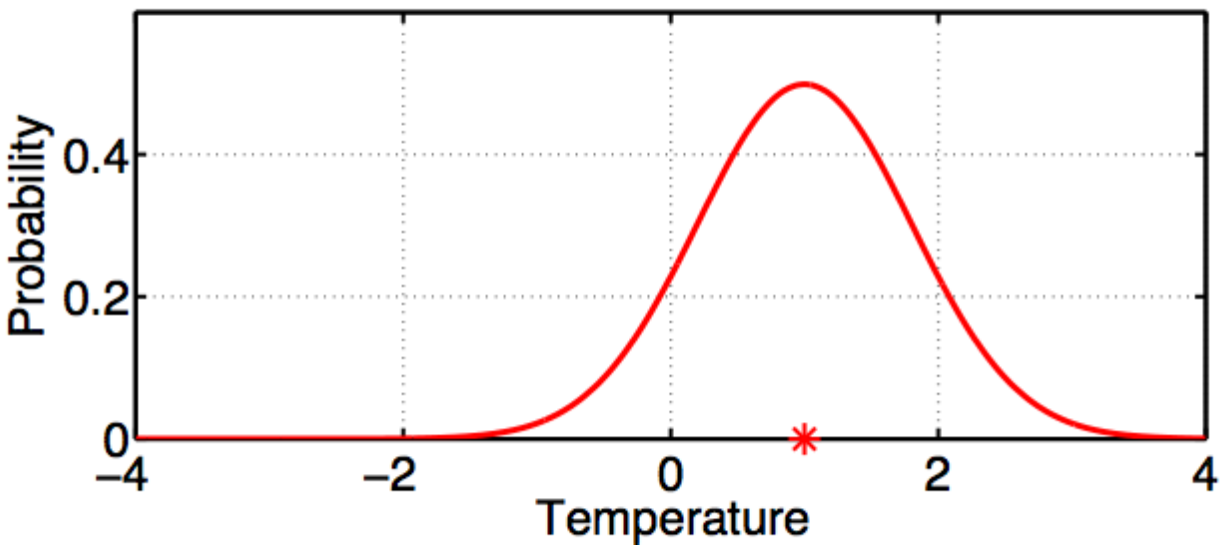
Example: Estimating the Temperature Outside

An observation has a value (*),



Example: Estimating the Temperature Outside

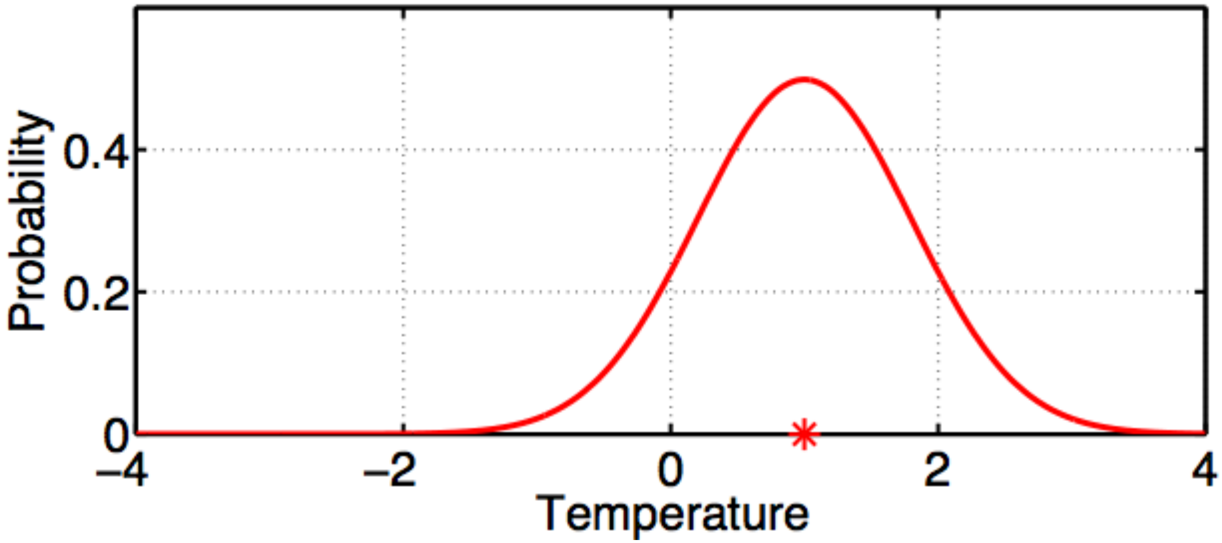
An observation has a value (*),

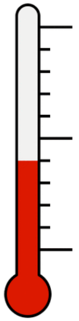


and an error distribution (red curve) that is associated with the instrument.

Example: Estimating the Temperature Outside

Thermometer outside measures 1°C.

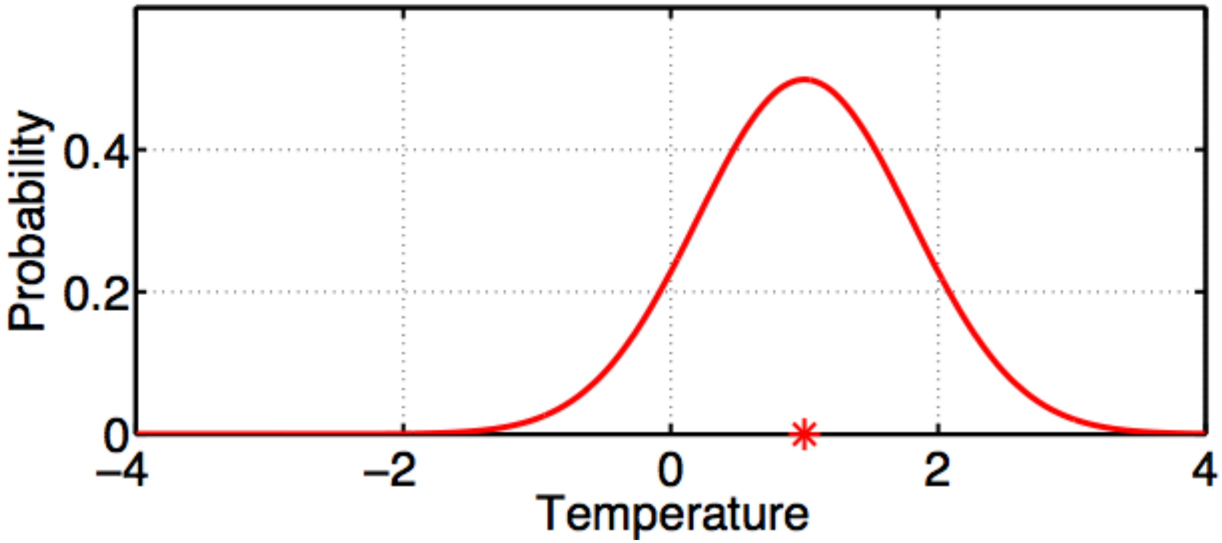


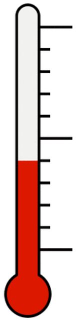


Instrument builder says thermometer
is unbiased with +/- 0.8°C Gaussian error.

Example: Estimating the Temperature Outside

Thermometer outside measures 1°C.



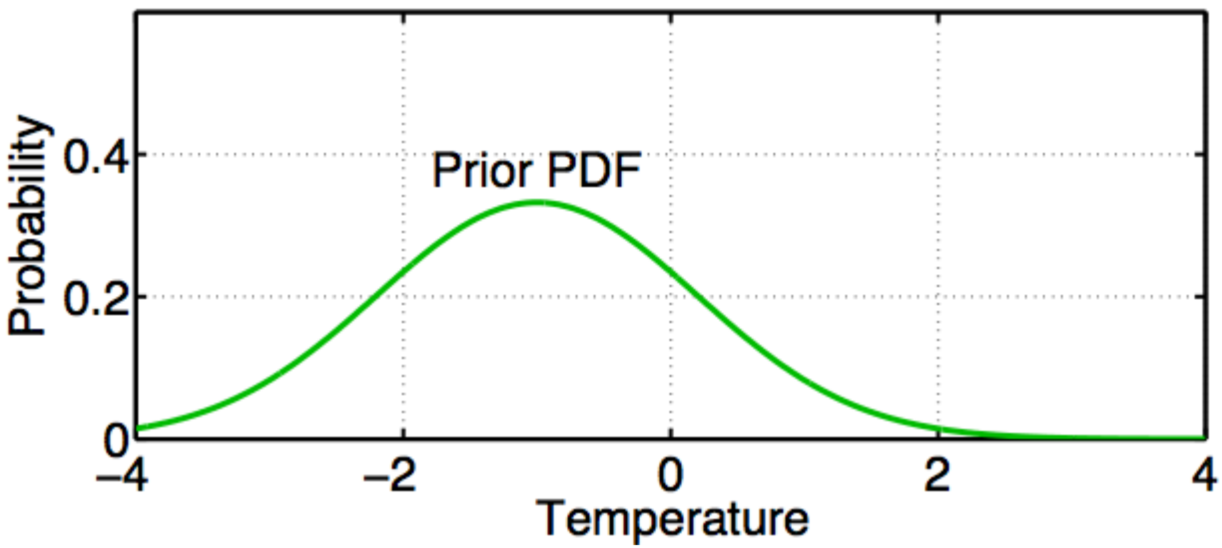


The red plot is $P(T \mid T_0)$;

probability of temperature *given* that T_o was observed.

Example: Estimating the Temperature Outside

We also have a prior estimate of temperature.



The green curve is $P(T \mid C)$;

probability of temperature given all available prior information C.

Example: Estimating the Temperature Outside

Prior information C can include:

1. Observations of things besides T;
2. Model forecast made using observations at earlier times;
3. *a priori* physical constraints ($T > -273.15^oC$);
4. Climatological constraints ($-30^oC < T < 40^oC$).

Combining the Prior Estimate and Observation

Likelihood: Probability that T_o is observed if T is true value and given prior information C.

Bayes Theorem:

$$P(T \mid T_o, C) = \frac{P(T_o \mid T, C)\, P(T \mid C)}{P(T_o \mid C)}$$

Prior

Posterior: Probability of T given observations and Prior. Also called **update** or **analysis**.

Combining the Prior Estimate and Observation

Rewrite Bayes as:

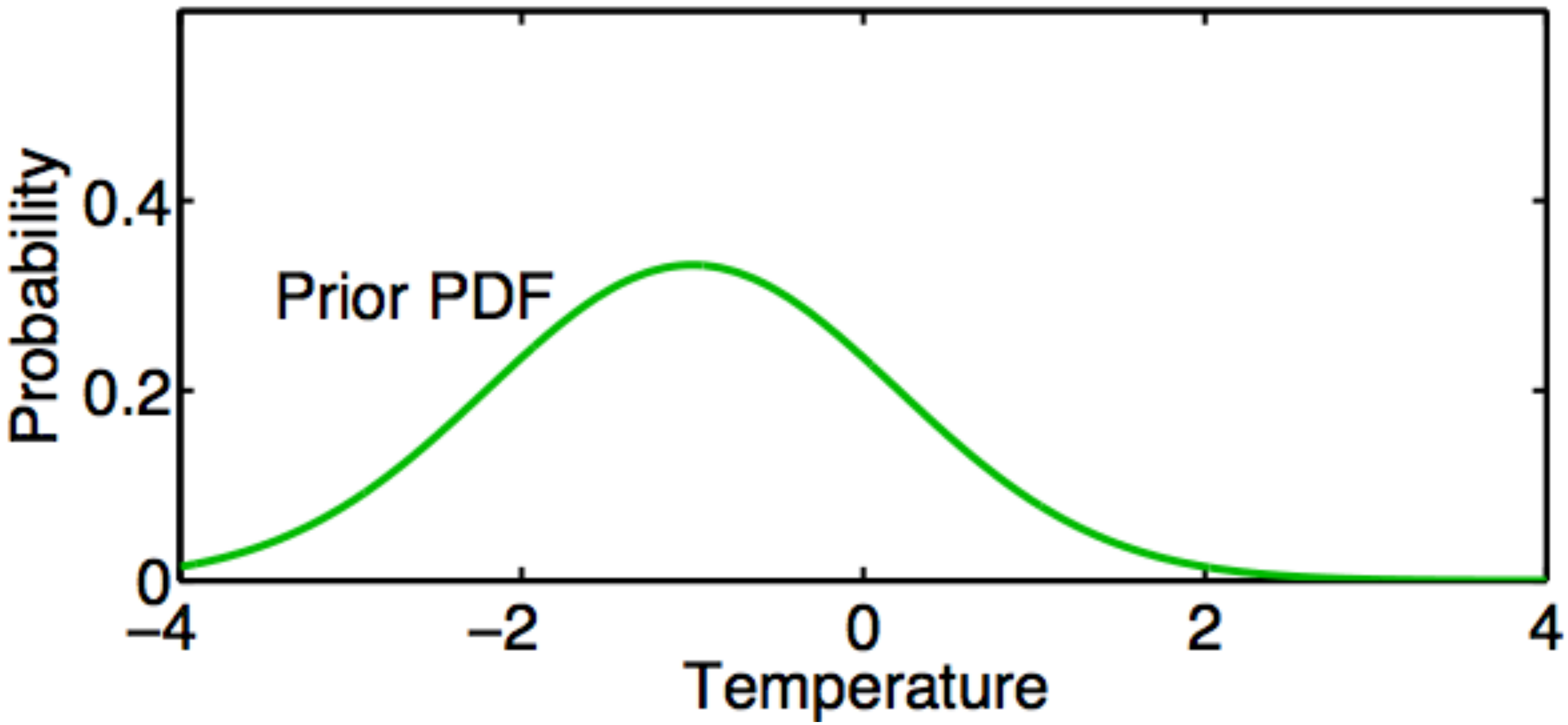
$$\frac{P(T_o \mid T,C)P(T \mid C)}{P(T_o \mid C)} = \frac{P(T_o \mid T,C)P(T \mid C)}{\int P(T_o \mid x)P(x \mid C)dx}$$

$$= \frac{P(T_o \mid T,C)P(T \mid C)}{normalization}$$

Denominator normalizes so Posterior is PDF.

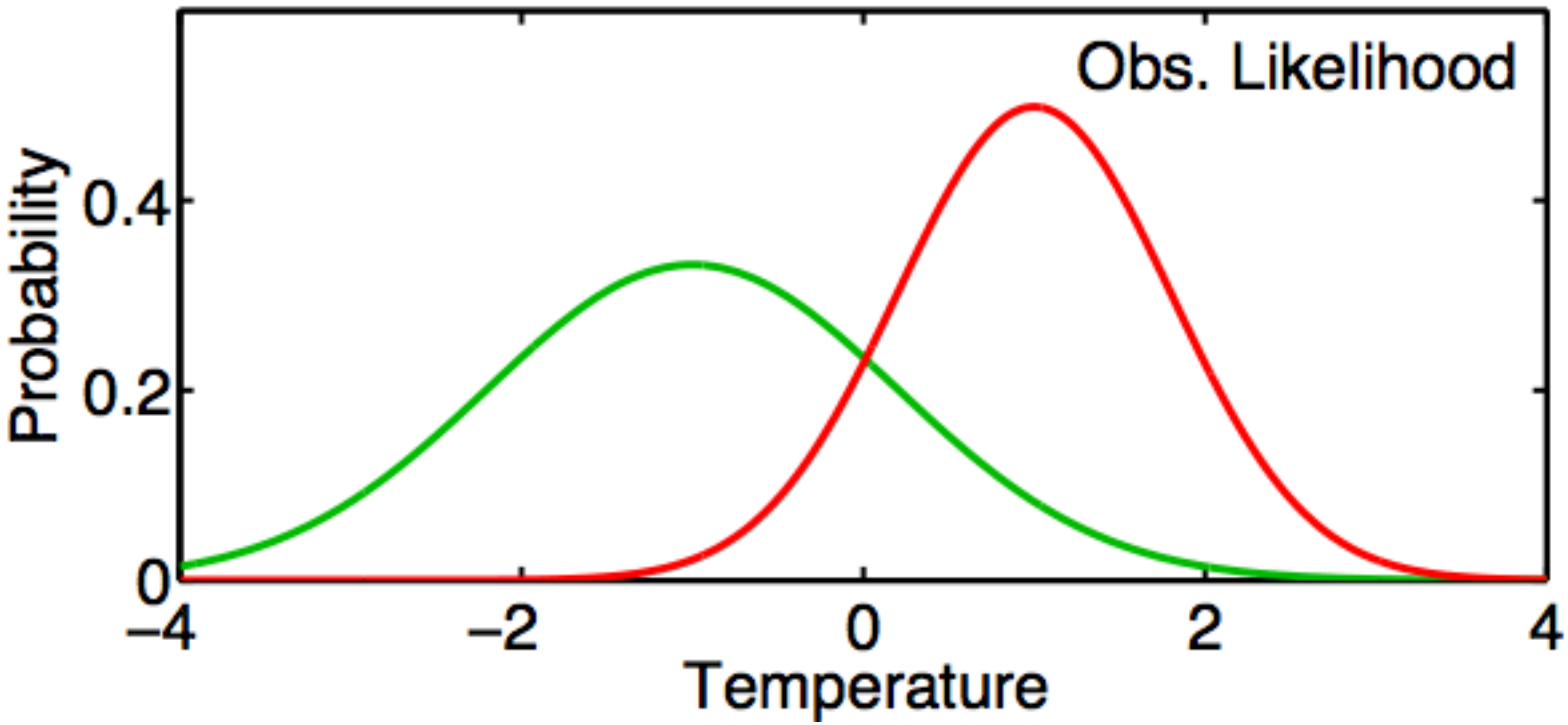
Combining the Prior Estimate and Observation

$$P\left(T \mid T_0, C\right) = \frac{P(T_0 \mid T, C)P(T \mid C)}{normalization}$$



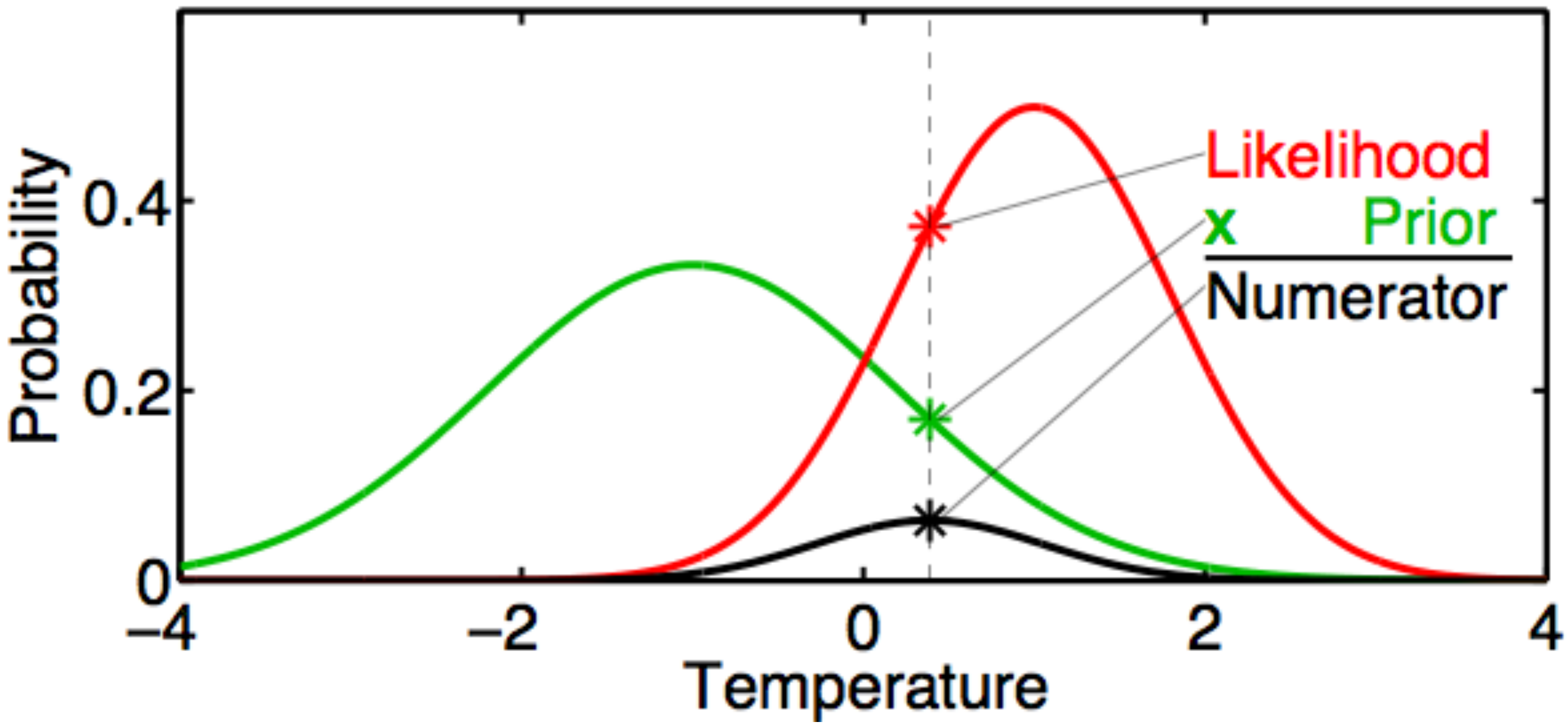
Combining the Prior Estimate and Observation

$$P\left(T \mid T_0, C\right) = \frac{P(T_0 \mid T, C)P(T \mid C)}{normalization}$$



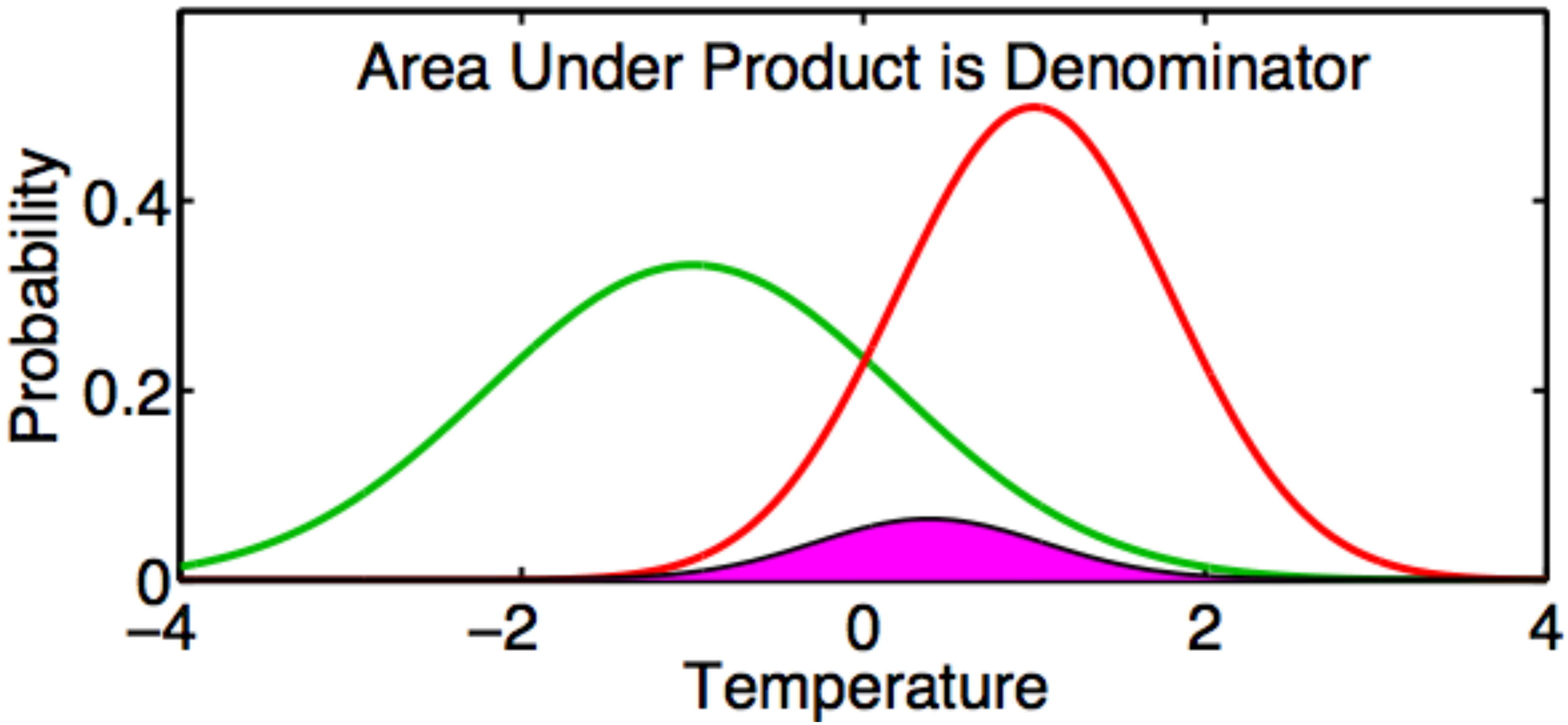
Combining the Prior Estimate and Observation

$$P\left(T \mid T_0, C\right) = \frac{P(T_0 \mid T, C)P(T \mid C)}{normalization}$$



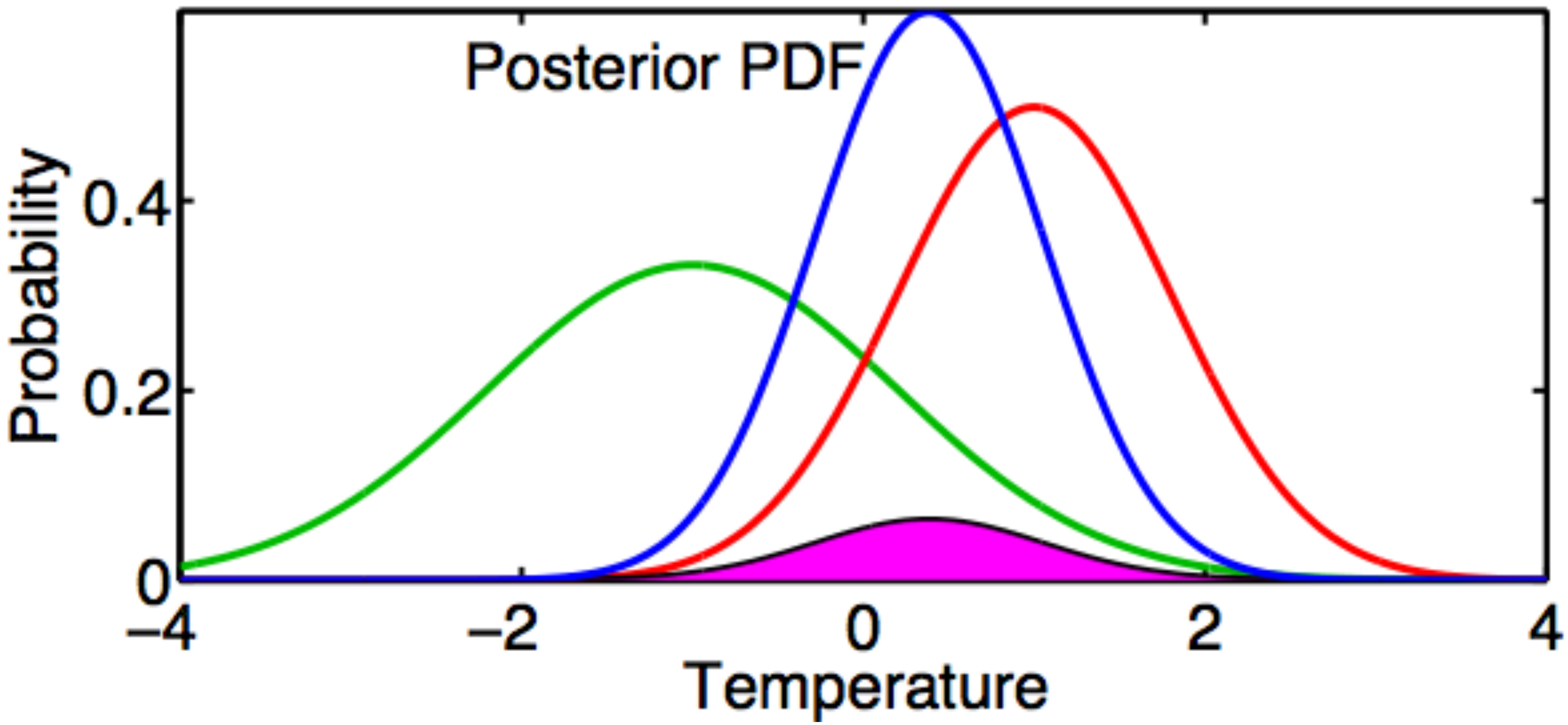
Combining the Prior Estimate and Observation

$$P\left(T \mid T_0, C\right) = \frac{P(T_0 \mid T, C)P(T \mid C)}{normalization}$$



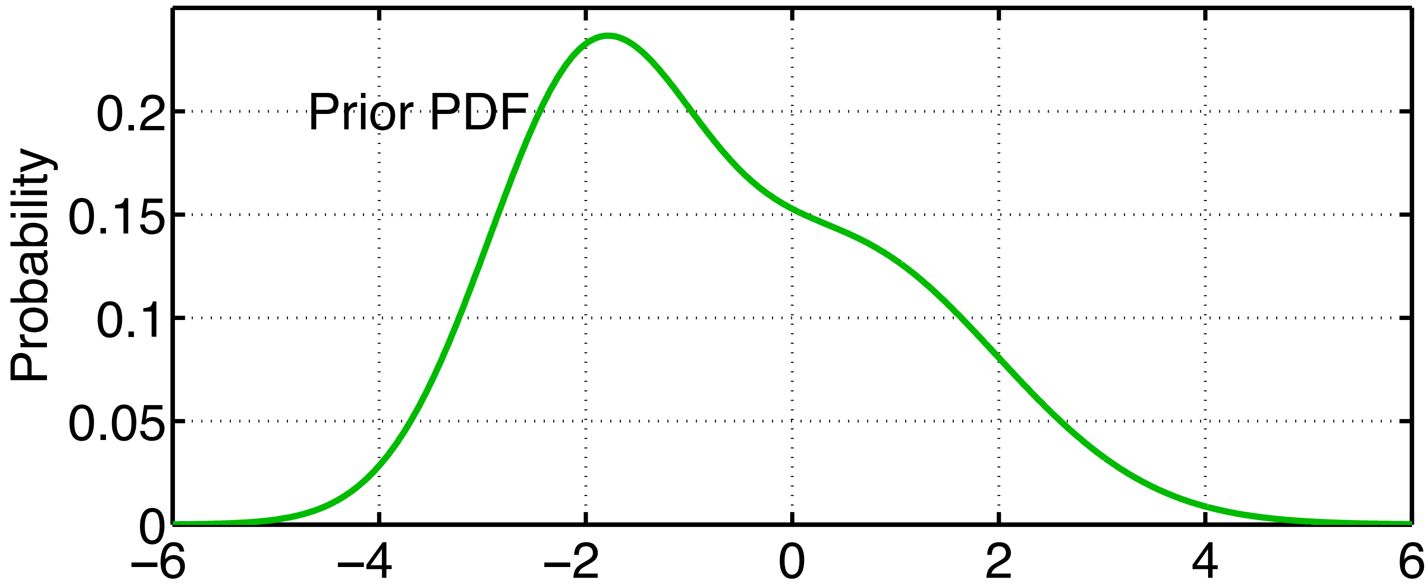
Combining the Prior Estimate and Observation

$$P\left(T \mid T_0, C\right) = \frac{P(T_0 \mid T, C)P(T \mid C)}{normalization}$$



Bayes' Rule

$$p(A|BC) = \frac{p(B|AC)\,p(A|C)}{p(B|C)} = \frac{p(B|AC)\,p(A|C)}{\int p(B|x)\,p(x|C)\,dx}$$



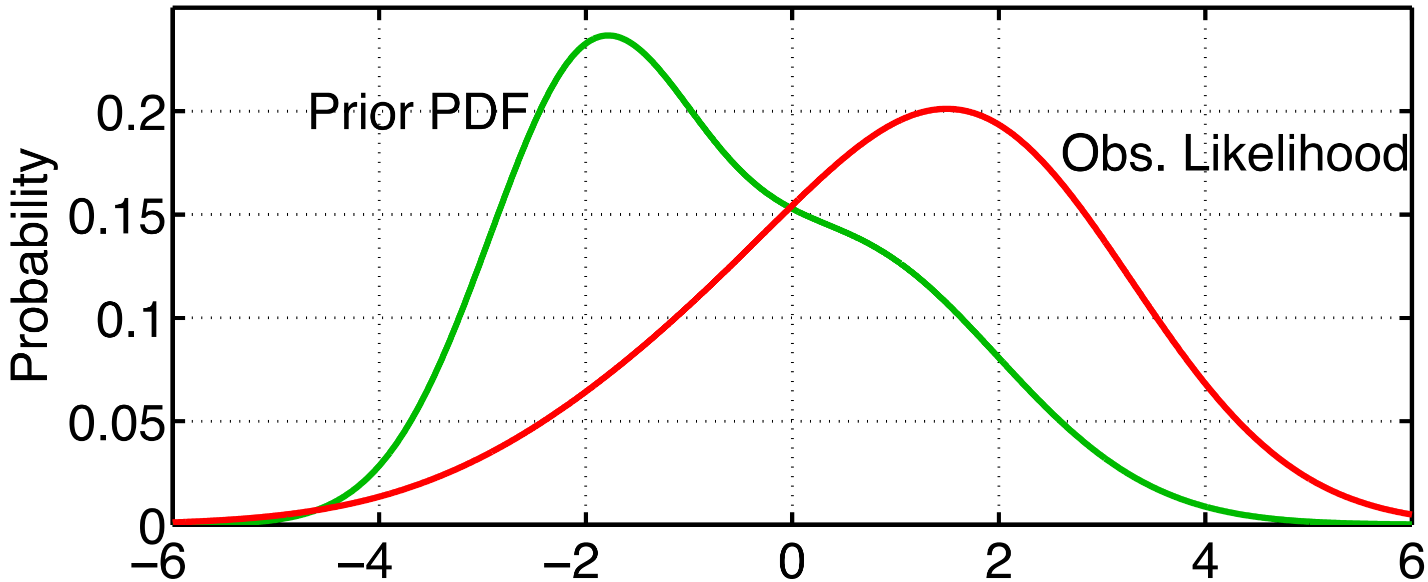
A : Prior Estimate based on all previous information, *C*.

B : An additional observation.

p(A|BC) : Posterior (updated estimate) based on *C* and *B*.

Bayes' Rule

$$p(A|BC) = \frac{p(B|AC)p(A|C)}{p(B|C)} = \frac{p(B|AC)p(A|C)}{\int p(B|x)p(x|C)dx}$$



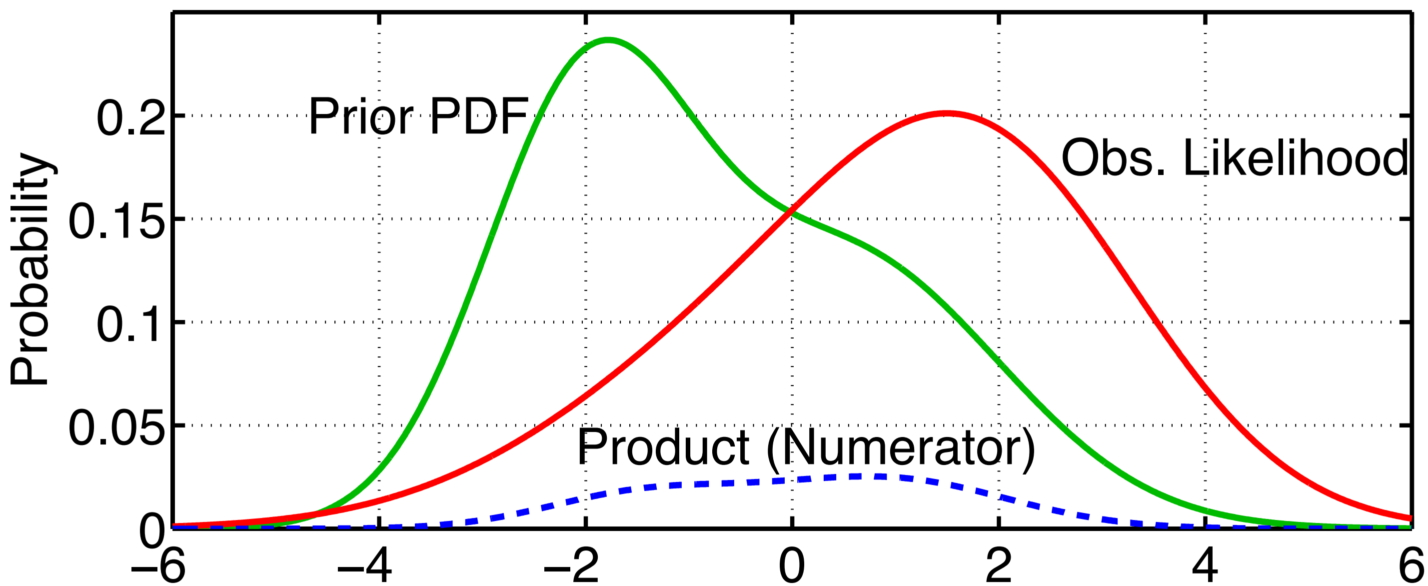
A : Prior Estimate based on all previous information, C.

B : An additional observation.

$p(A/BC)$: Posterior (updated estimate) based on C and B.

Bayes' Rule

$$p(A \mid BC) = \frac{p(B \mid AC)\,p(A \mid C)}{p(B \mid C)} = \frac{p(B \mid AC)\,p(A \mid C)}{\int p(B \mid x)\,p(x \mid C)\,dx}$$



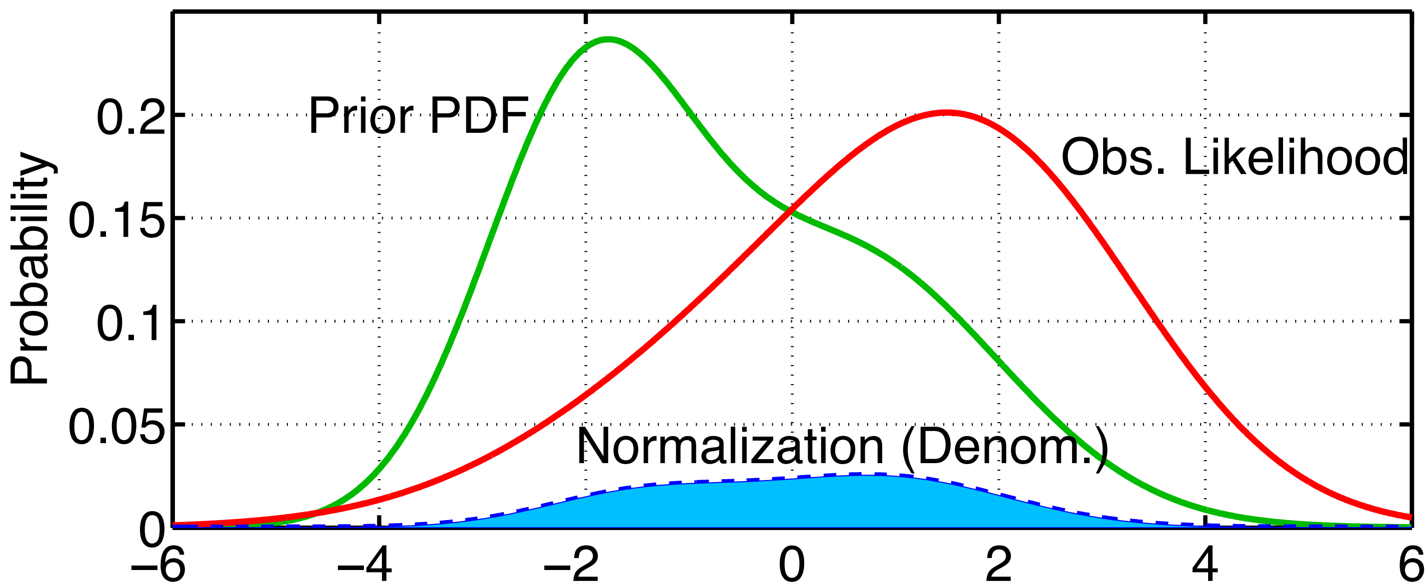
A : Prior Estimate based on all previous information, C.

B : An additional observation.

$p(A/BC)$: Posterior (updated estimate) based on C and B.

Bayes' Rule

$$p(A \mid BC) = \frac{p(B \mid AC)\,p(A \mid C)}{p(B \mid C)} = \frac{p(B \mid AC)\,p(A \mid C)}{\int p(B \mid x)\,p(x \mid C)\,dx}$$



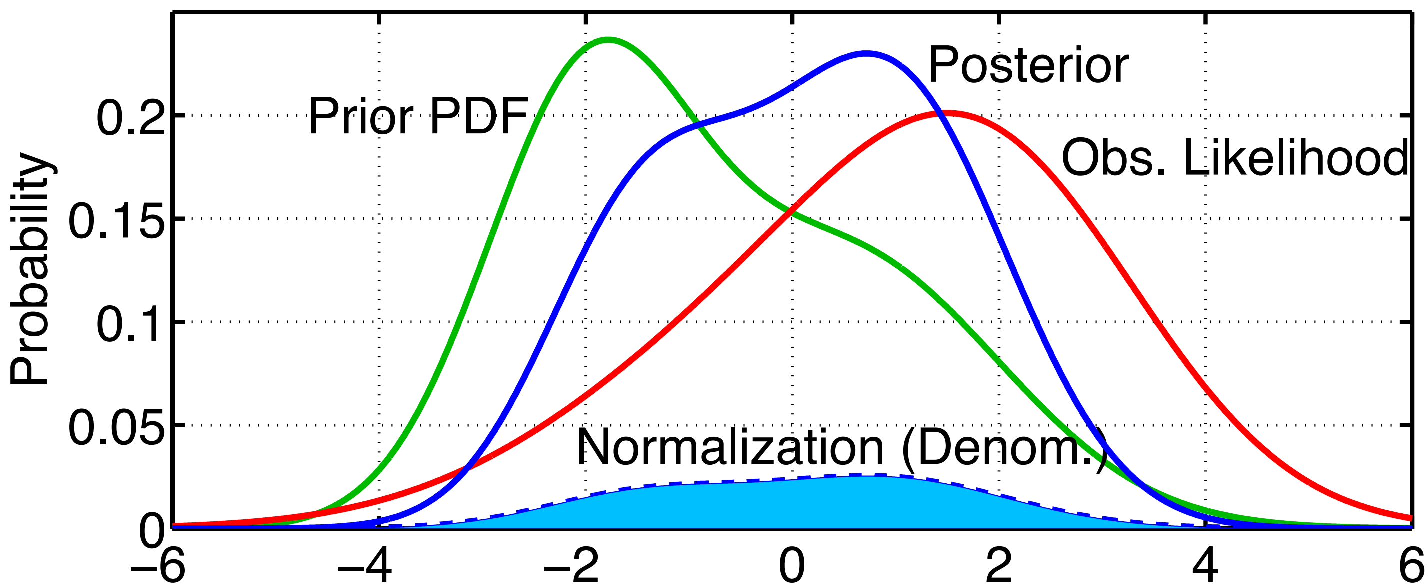
A : Prior Estimate based on all previous information, C.

B : An additional observation.

$p(A|BC)$: Posterior (updated estimate) based on C and B.

Bayes' Rule

$$p(A|BC) = \frac{p(B|AC)\,p(A|C)}{p(B|C)} = \frac{p(B|AC)\,p(A|C)}{\int p(B|x)\,p(x|C)\,dx}$$



A : Prior Estimate based on all previous information, C.

B : An additional observation.

$p(A/BC)$: Posterior (updated estimate) based on C and B.

Color Scheme

Green == Prior

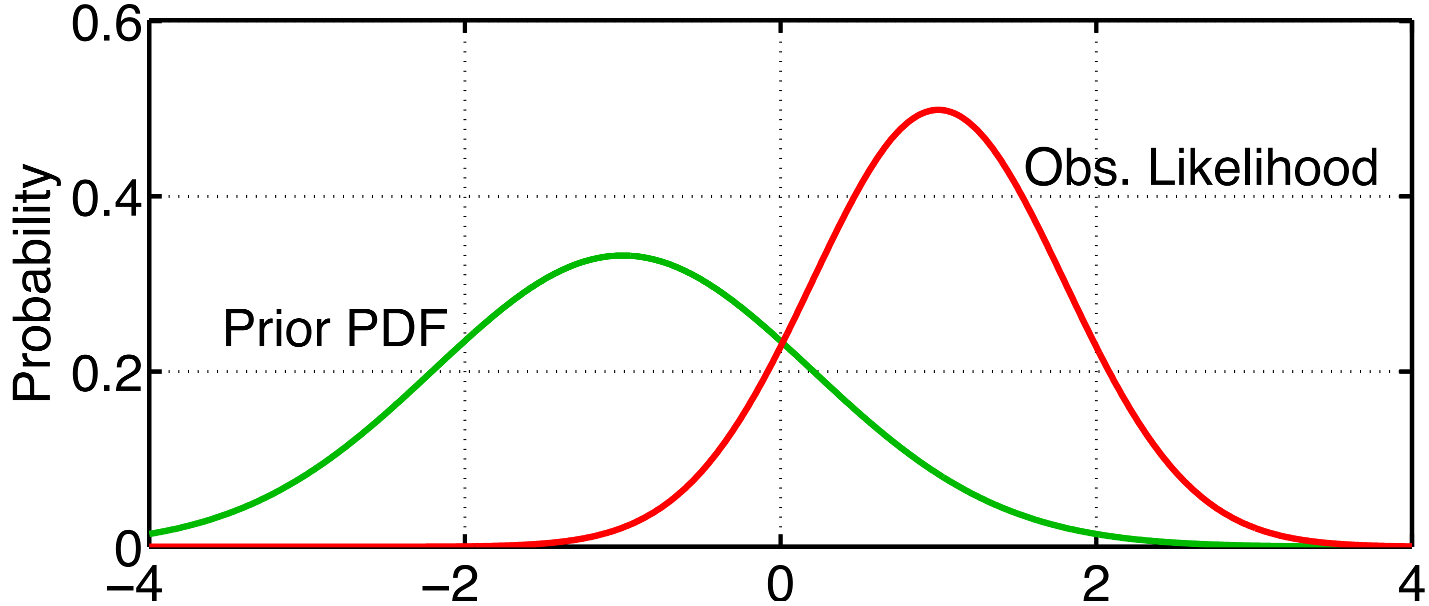
Red == Observation

Blue == Posterior

The same color scheme is used throughout ALL Tutorial materials.

Product of Two Gaussians

$$p(A|BC) = \frac{p(B|AC)\,p(A|C)}{p(B|C)} = \frac{p(B|AC)\,p(A|C)}{\int p(B|x)\,p(x|C)\,dx}$$



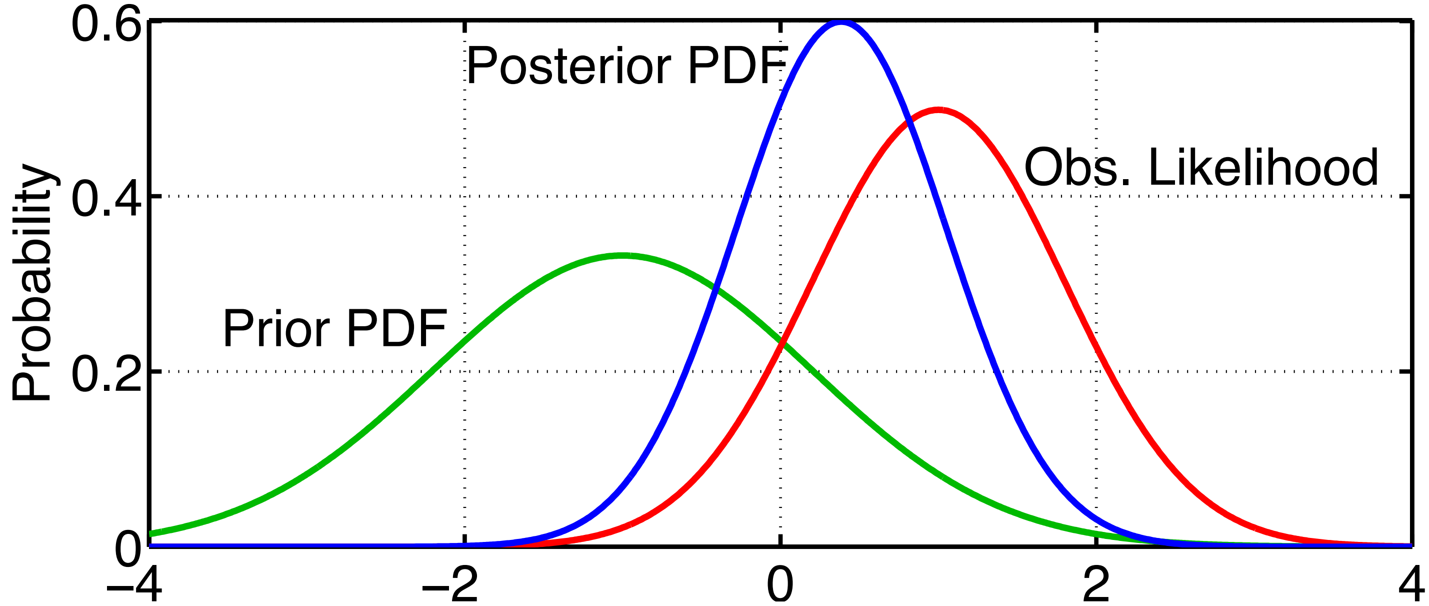
Any 1-D normal distribution can be represented as a PDF:

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

Product of Two Gaussians

$$p(A|BC) = \frac{p(B|AC)p(A|C)}{p(B|C)} = \frac{p(B|AC)p(A|C)}{\int p(B|x)p(x|C)dx}$$

This product is closed for Gaussian distributions.



Product of Two Gaussians

Product of d-dimensional normals with means μ_1 and μ_2 and covariance matrices Σ_1 and Σ_2 is normal.

$$N(\mu_1, \Sigma_1) N(\mu_2, \Sigma_2) = cN(\mu, \Sigma)$$

Covariance: $$\Sigma = (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}$$

Mean: $$\mu = (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1} (\Sigma_1^{-1} \mu_1 + \Sigma_2^{-1} \mu_2)$$

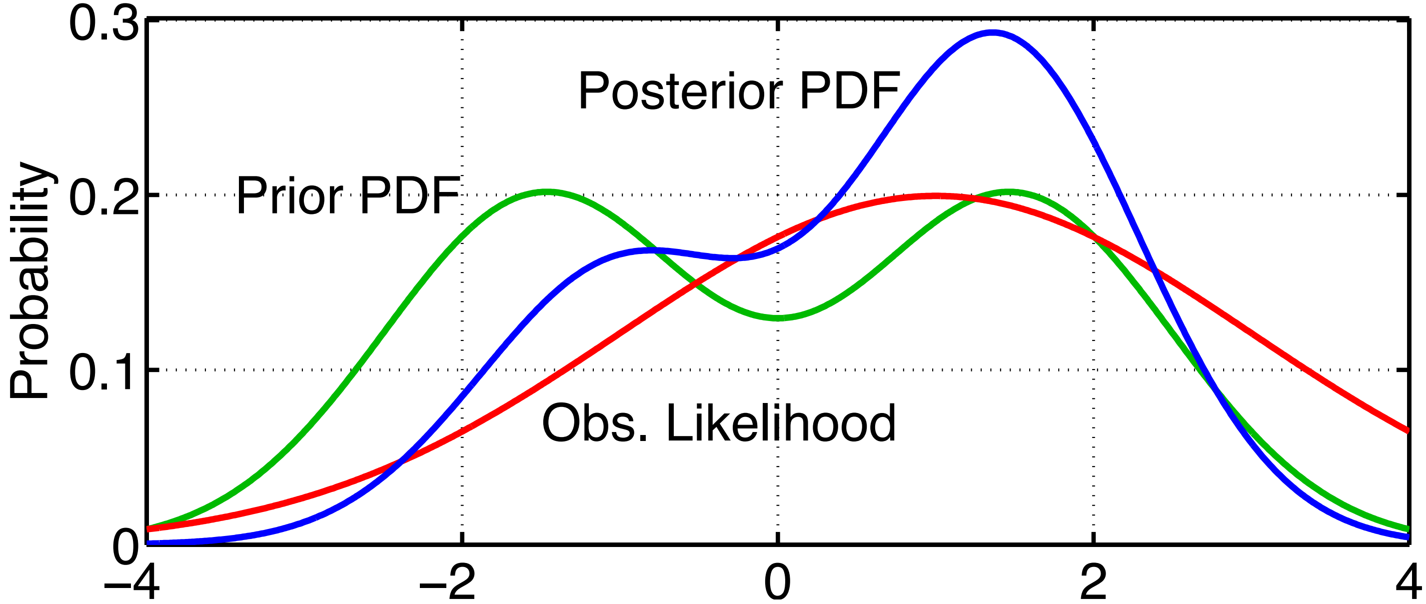
Weight: $$c = \frac{1}{(2\pi)^{d/2} \left| \Sigma_1 + \Sigma_2 \right|^{1/2}} \exp\left\{ -\frac{1}{2} \left[(\mu_2 - \mu_1)^T (\Sigma_1 + \Sigma_2)^{-1} (\mu_2 - \mu_1) \right] \right\}$$

The weight is simply the normalization of the normal distribution defined by the product of the prior and observation likelihood.

Product of Two Gaussians

$$p(A|BC) = \frac{p(B|AC)p(A|C)}{p(B|C)} = \frac{p(B|AC)p(A|C)}{\int p(B|x)p(x|C)dx}$$

This product can be determined analytically for Gaussian distributions.

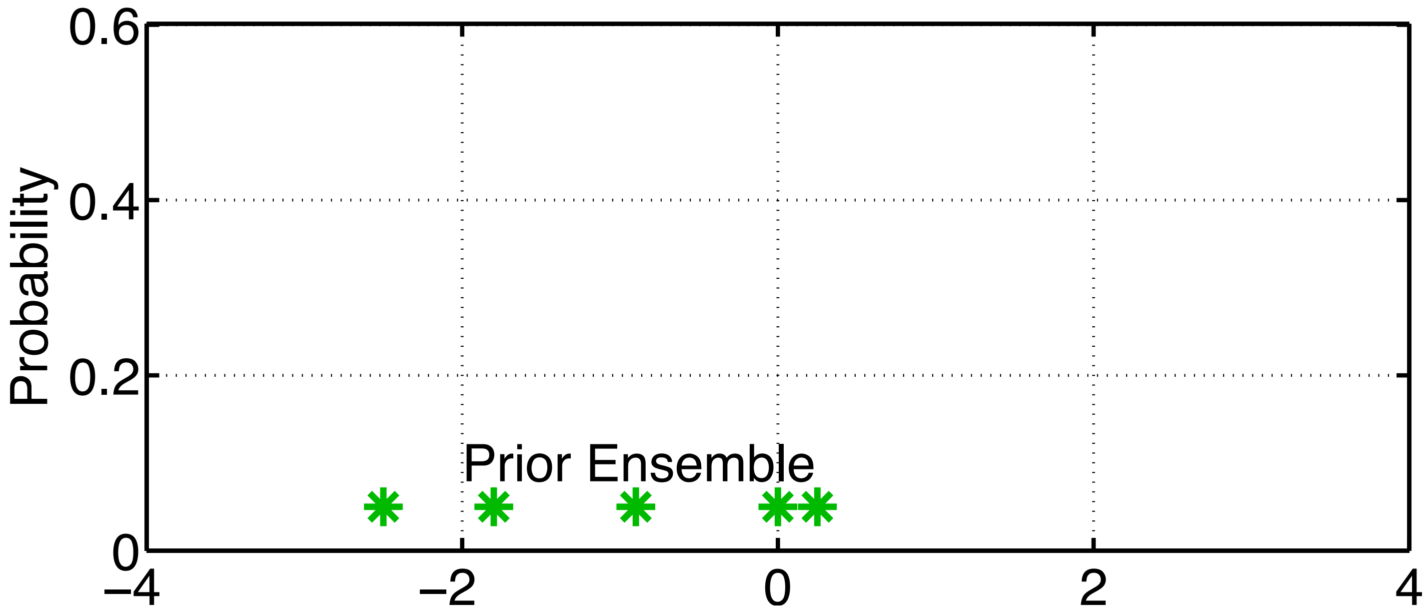


But, for general distributions, there's no analytical product.

Product of Two Gaussians

$$p(A|BC) = \frac{p(B|AC)p(A|C)}{p(B|C)} = \frac{p(B|AC)p(A|C)}{\int p(B|x)p(x|C)dx}$$

Ensemble filters: <u>Prior is available as finite sample.</u>

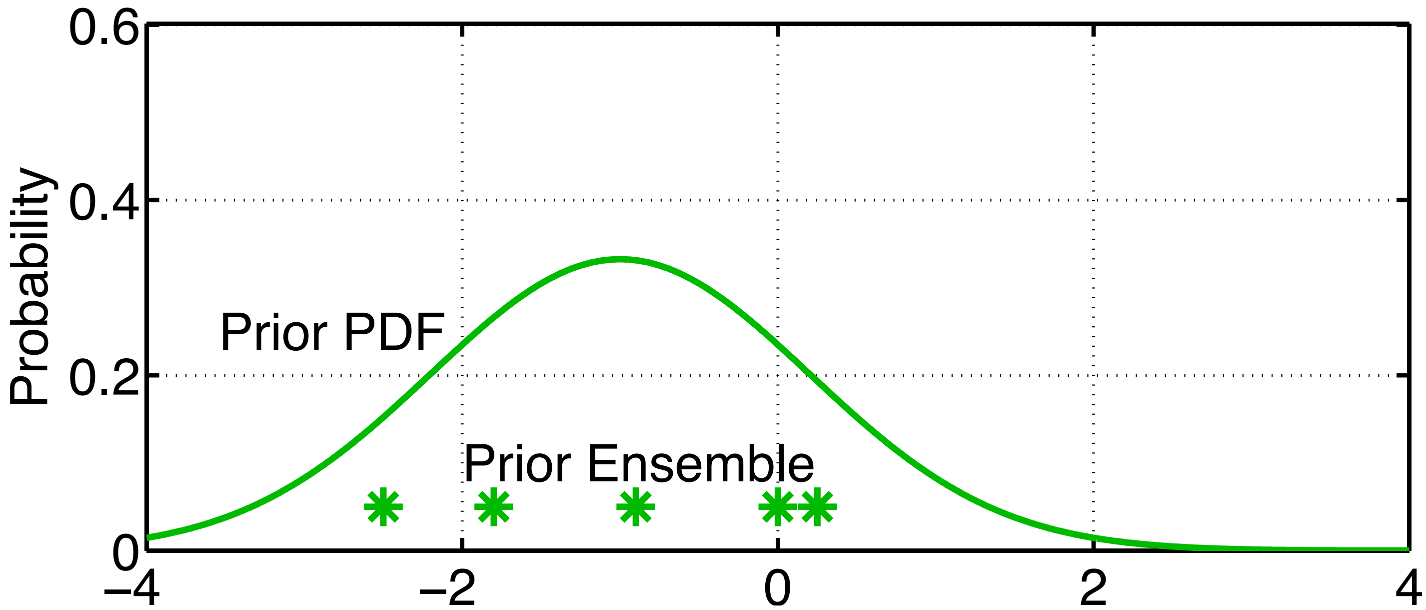


Don't know much about properties of this sample.
May naively assume it is random draw from 'truth'.

Product of Two Gaussians

$$p(A|BC) = \frac{p(B|AC)p(A|C)}{p(B|C)} = \frac{p(B|AC)p(A|C)}{\int p(B|x)p(x|C)dx}$$

How can we take product of sample with continuous likelihood?

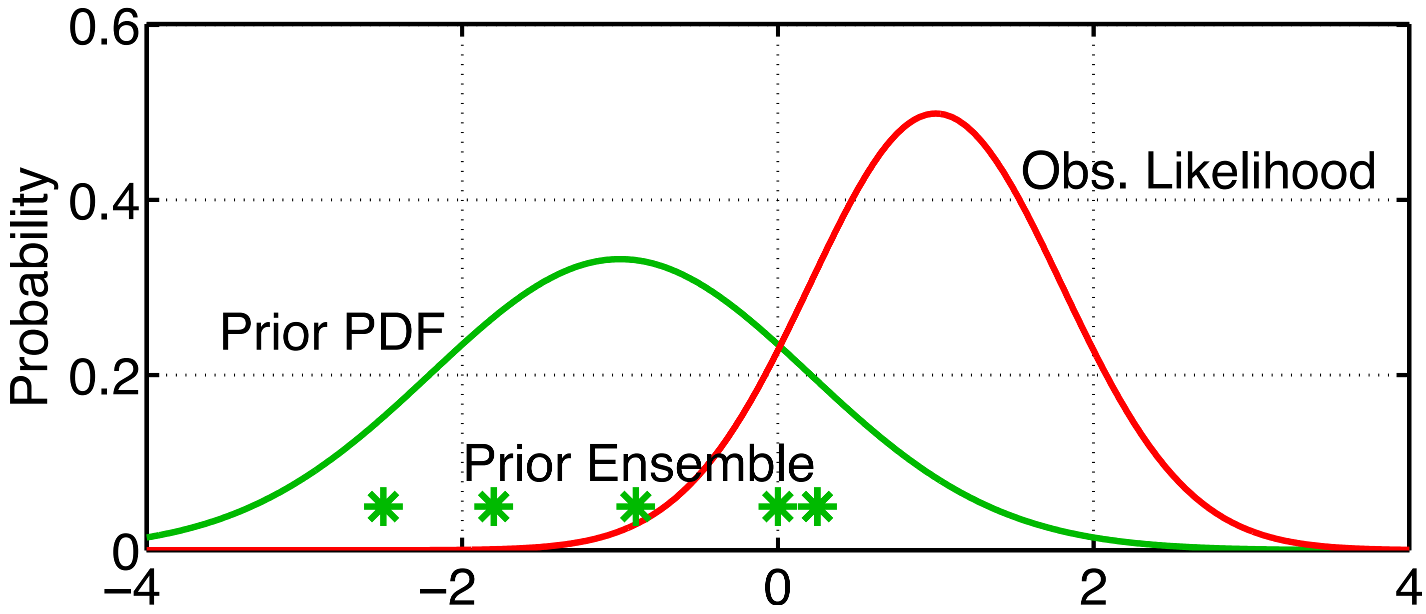


Fit a continuous (Gaussian for now) distribution to sample.

Product of Two Gaussians

$$p(A|BC) = \frac{p(B|AC)p(A|C)}{p(B|C)} = \frac{p(B|AC)p(A|C)}{\int p(B|x)p(x|C)dx}$$

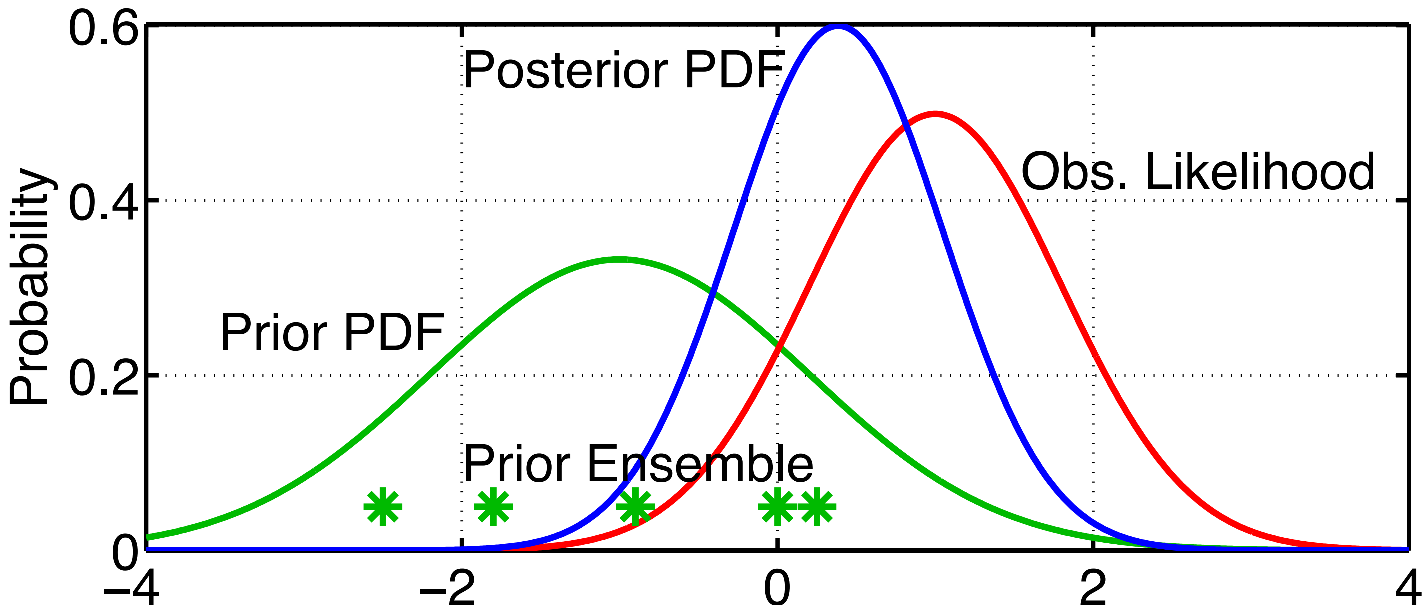
Observation likelihood usually continuous (nearly always Gaussian).



Product of Two Gaussians

$$p(A|BC) = \frac{p(B|AC)p(A|C)}{p(B|C)} = \frac{p(B|AC)p(A|C)}{\int p(B|x)p(x|C)dx}$$

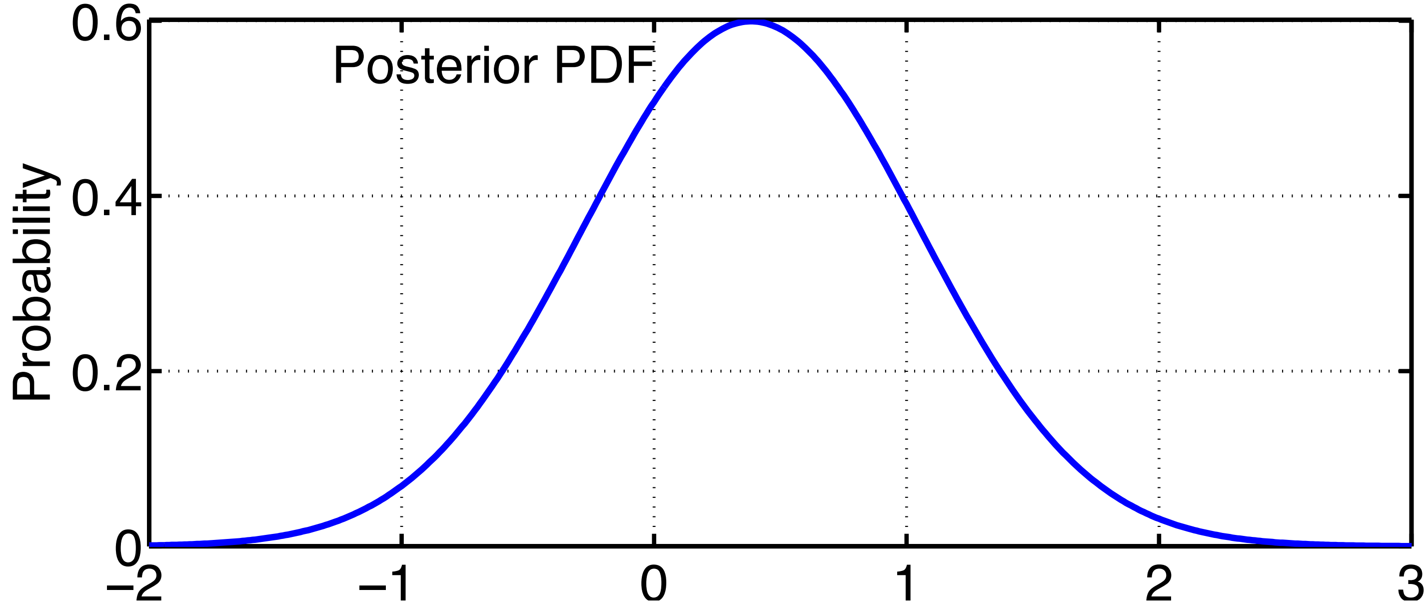
Product of prior Gaussian fit and Obs. likelihood is Gaussian.



Analytically computing continuous posterior is simple. BUT, we need to have a SAMPLE of this PDF...

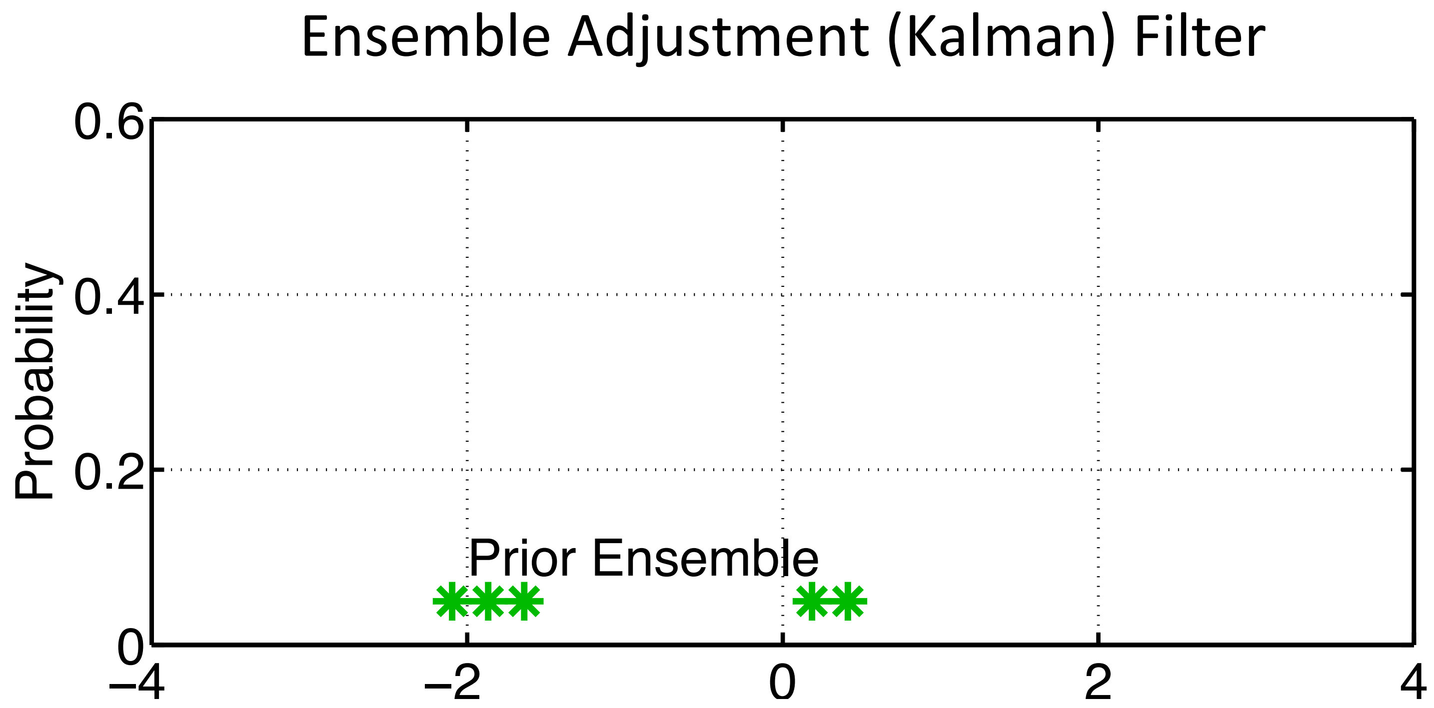
Sampling Posterior PDF

There are many ways to do this.

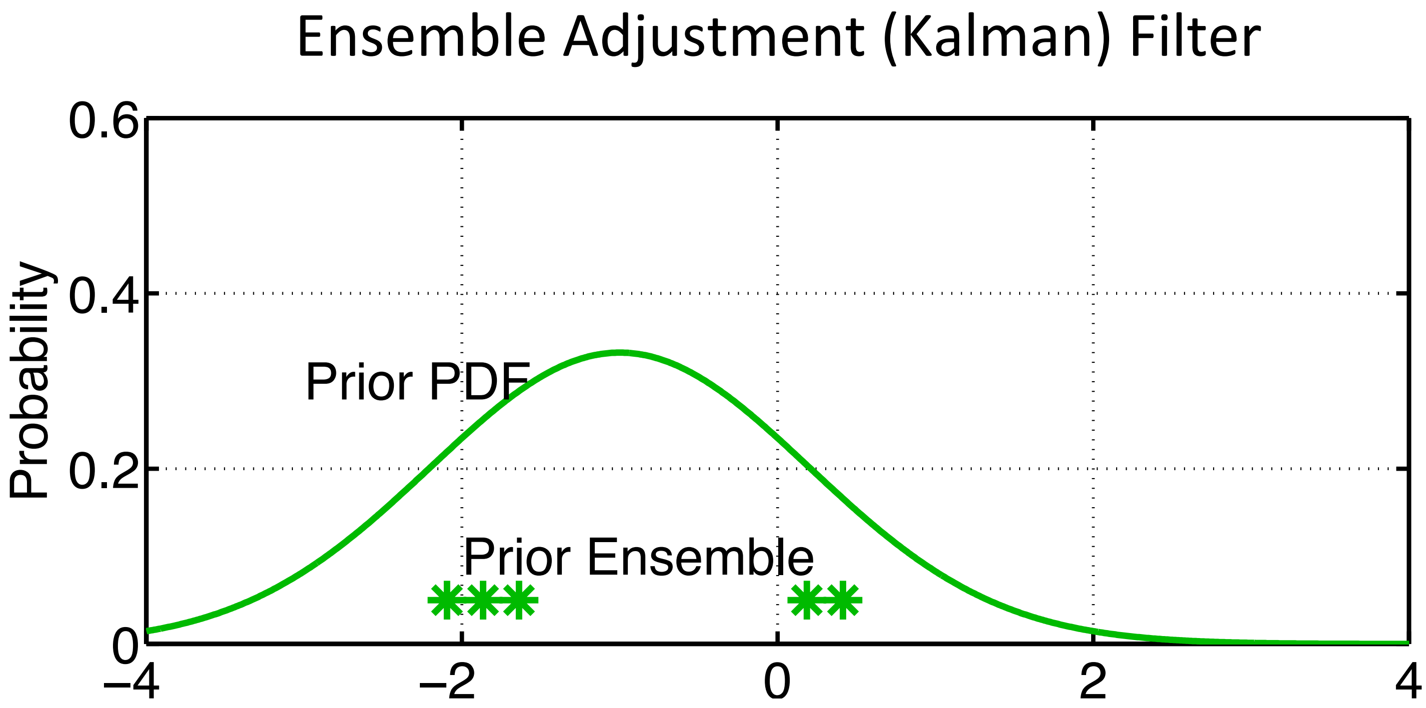


Exact properties of different methods may be unclear.
Trial and error still best way to see how they perform.
Will interact with properties of prediction models, etc.

Sampling Posterior PDF

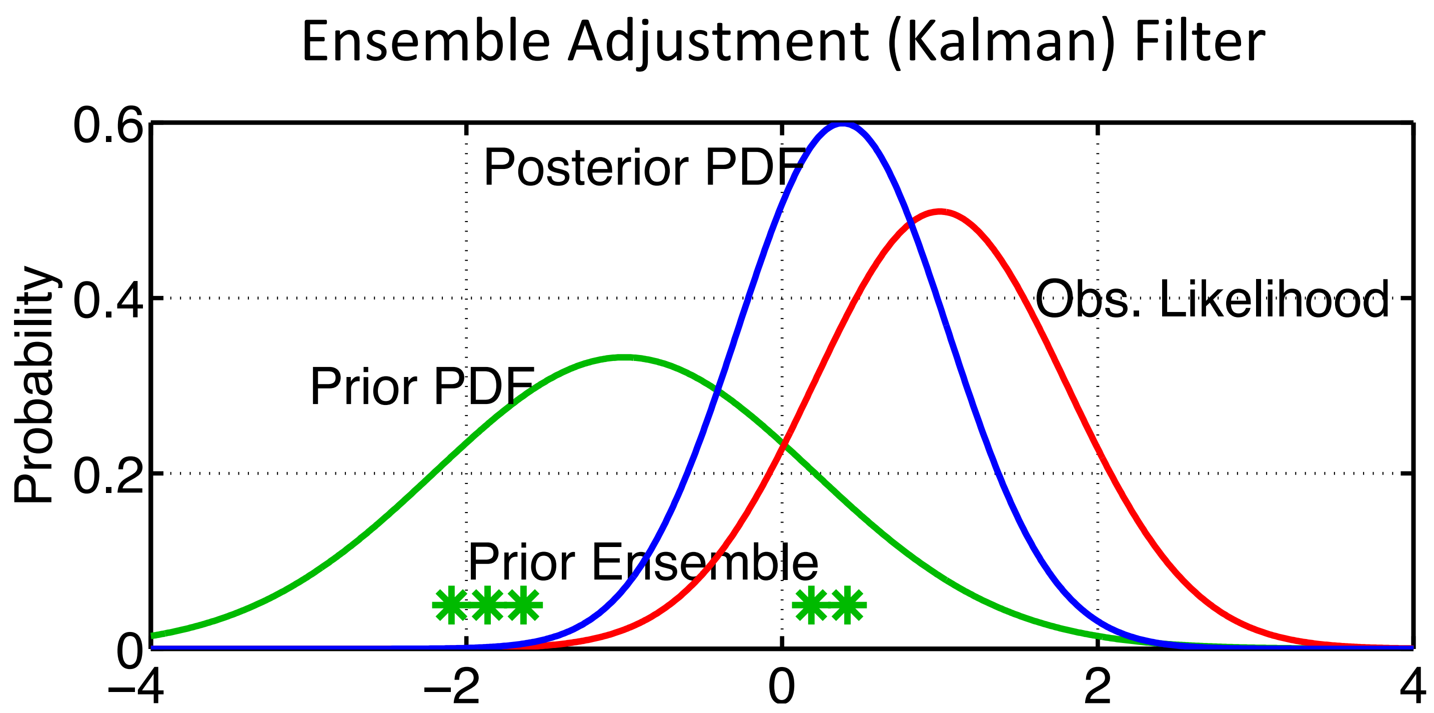


Sampling Posterior PDF



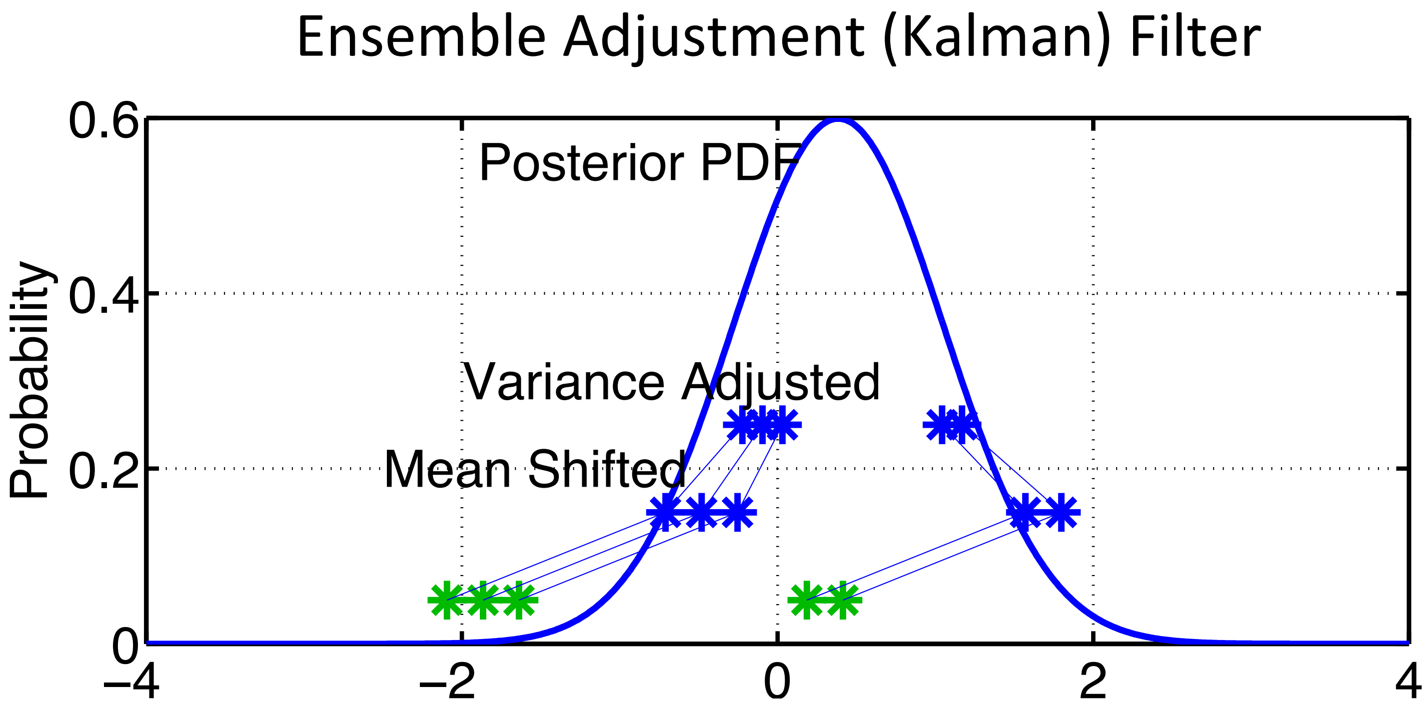
Again, fit a Gaussian to sample.

Sampling Posterior PDF



Compute posterior PDF (same as previous algorithms).

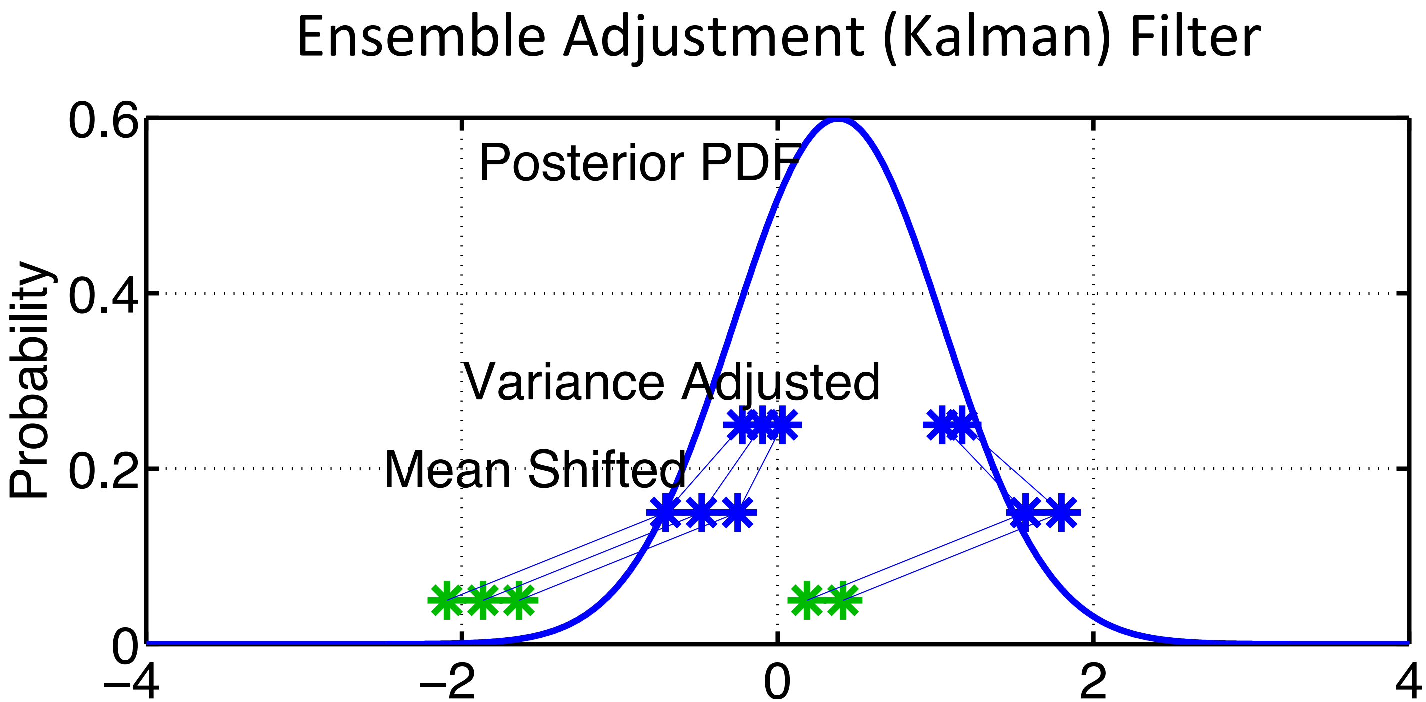
Sampling Posterior PDF



Use deterministic algorithm to 'adjust' ensemble.

1. 'Shift' ensemble to have exact mean of posterior.
2. Use linear contraction to have exact variance of posterior.

Sampling Posterior PDF



$$x_i^u = \left(x_i^p - \overline{x}^p\right) \bullet \left(\sigma^u / \sigma^p\right) + \overline{x}^u$$

i = 1,..., ensemble size.

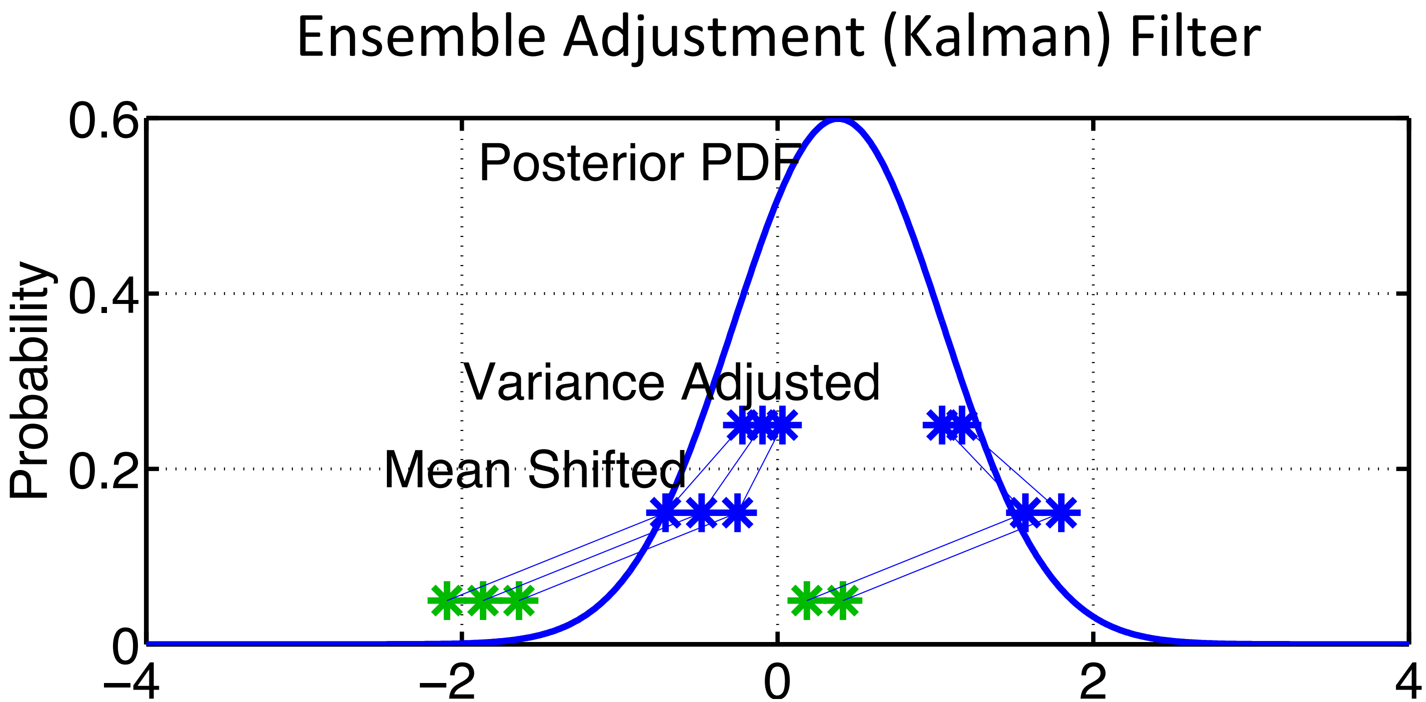
p is prior,

u is update (posterior),

σ is standard deviation,

overbar is ensemble mean.

Sampling Posterior PDF



Bimodality maintained, but not appropriately positioned or weighted. No problem with random outliers.