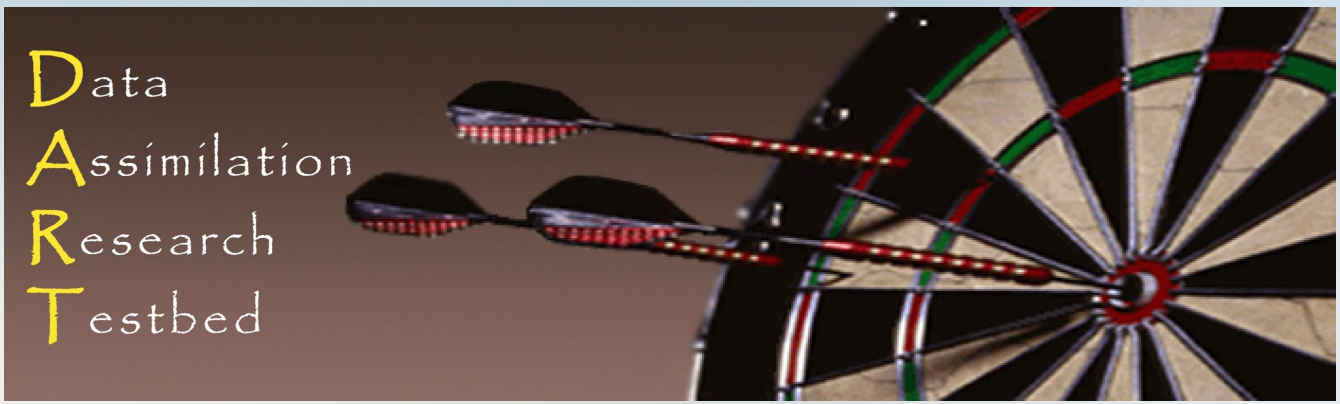
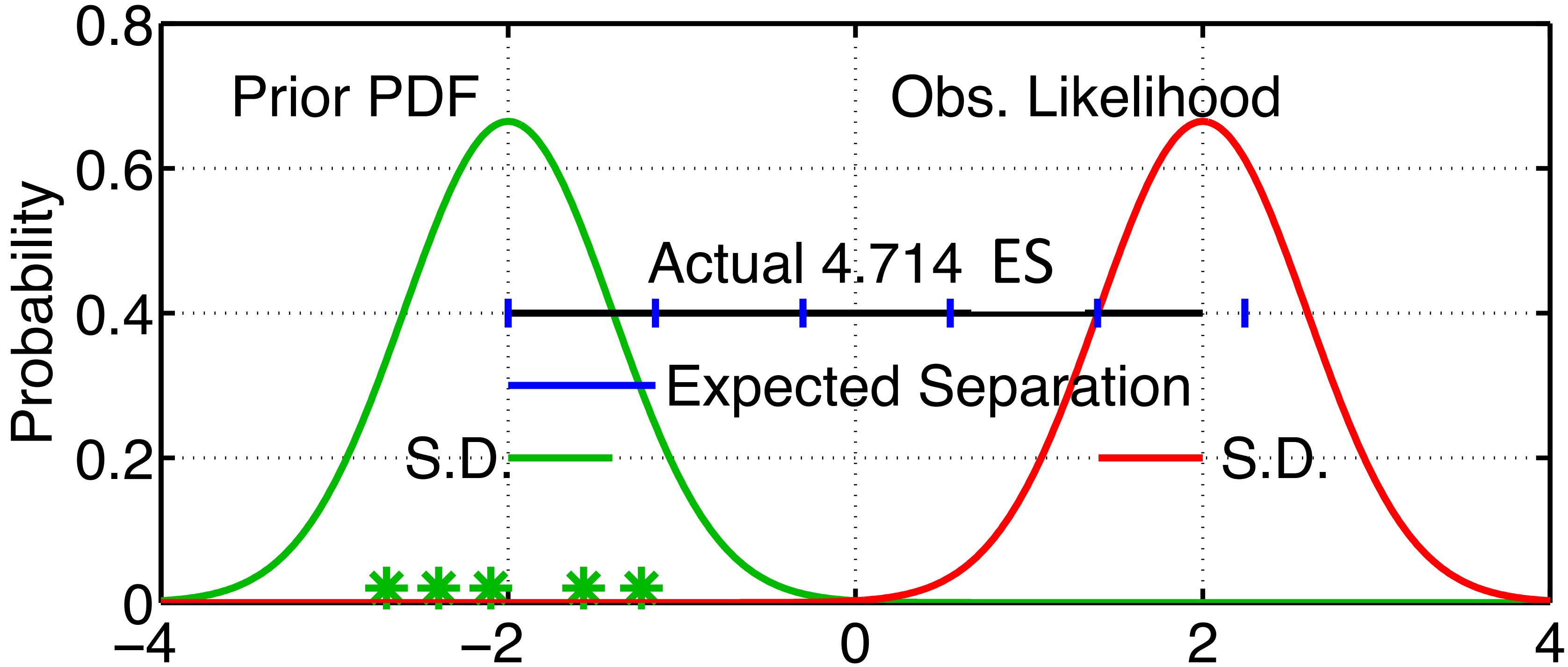


DART_LAB Tutorial Section 5: Adaptive Inflation

This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977.

Variance inflation for observations: An adaptive error-tolerant filter



1. For observed variable, have estimate of prior-observed inconsistency.

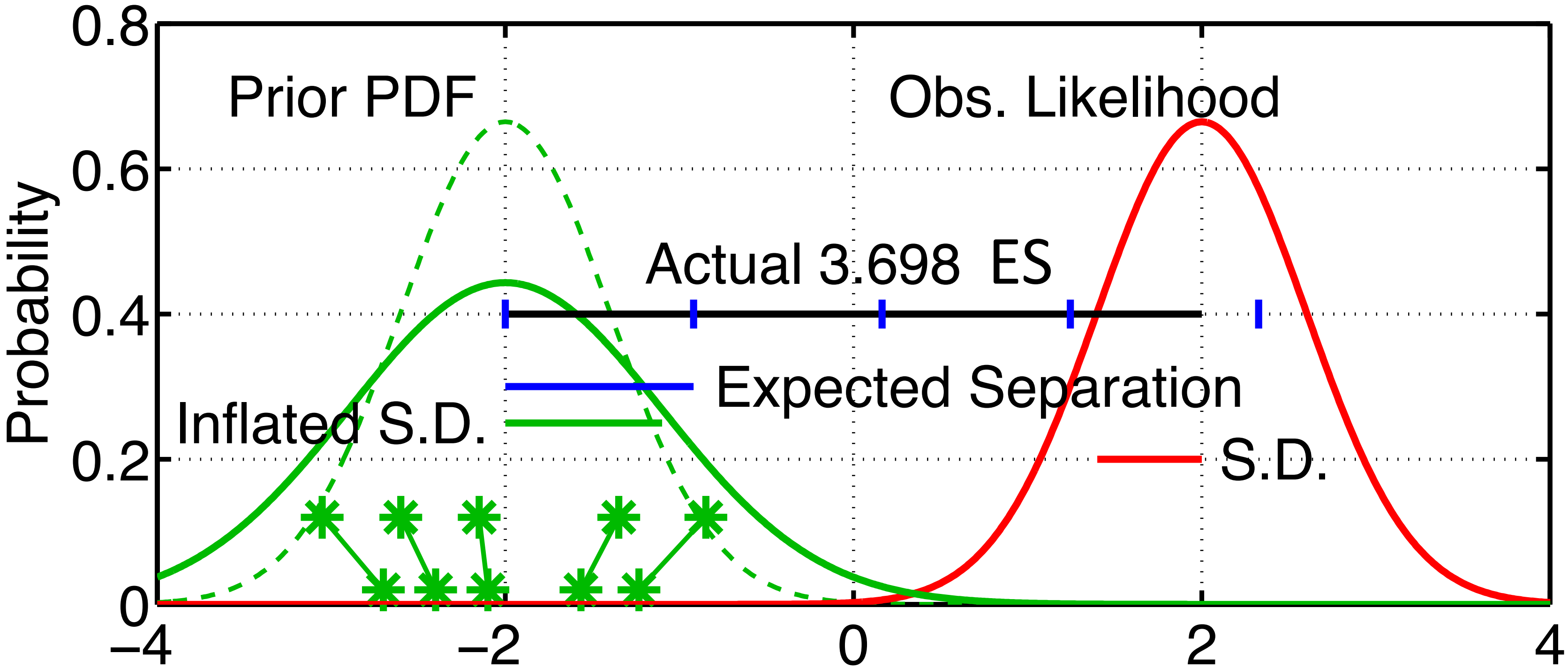
2. Define ES, Expected separation |prior_mean – observation| = $\sqrt{\sigma_p^2 + \sigma_O^2}$

Assumes that prior and observation are supposed to be unbiased.
Is it model error or random chance?

Glossary for this Section

y_O: Observed value for observation y at time t_k
σ_O^2: Observation error variance for observation y at time t_k
$\bar{y}_p$: Mean of prior ensemble estimate of observation y at time t_k
σ_p^2: Variance of prior ensemble estimate of observation y at time t_k
Y_k: All the observations that have been assimilated up to and including the kth observation of y
λ: Multiplicative inflation factor for observation y (see section 3)
λ_k: Estimate of λ at time t_k
$\bar{\lambda}_p$: Mean of prior estimate of λ at time t_k
$\bar{\lambda}_u$: Mean of posterior (updated) estimate of λ at time t_k
$\sigma_{\lambda,p}^2$: Variance of prior estimate of λ at time t_k
$\sigma_{\lambda,u}^2$: Variance of posterior (updated) estimate of λ at time t_k
D: The distance (absolute value of the difference) between y_O and $\bar{y}_p$

Variance inflation for observations: An adaptive error-tolerant filter

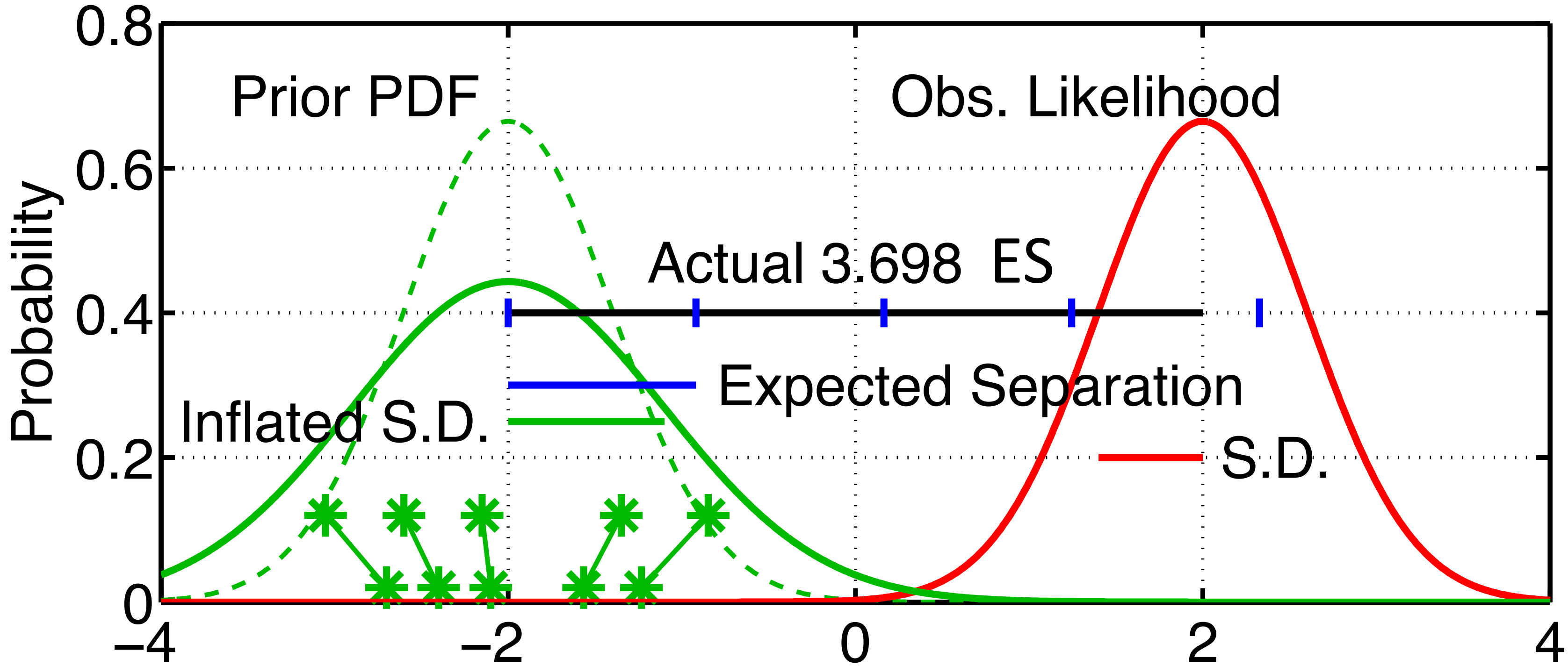


1. For observed variable, have estimate of prior-observed inconsistency.

2. Expected separation |prior_mean – observation| = $\sqrt{\sigma_p^2 + \sigma_o^2}$

3. Inflating increases the 'expected separation':
   Increases 'apparent' consistency between prior and observation.

Variance inflation for observations: An adaptive error-tolerant filter

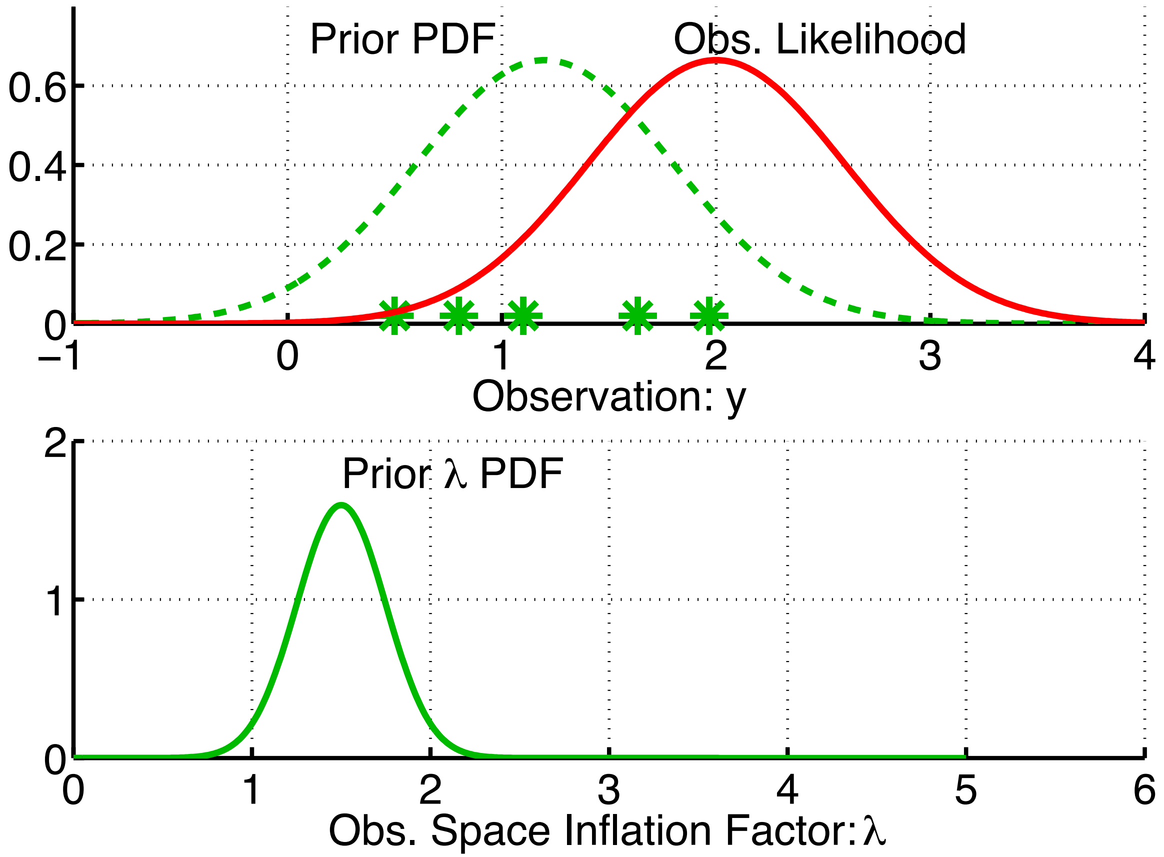


Distance D from prior mean to obs is distributed as: $N\left(0, \sqrt{\lambda\sigma_p^2 + \sigma_O^2}\right) = N(0, \theta)$

Prob y_O is observed given λ: $p(y_O|\lambda) = (2\pi\theta^2)^{-1/2} exp(-D^2/2\theta^2)$

Variance inflation for observations: An adaptive error-tolerant filter

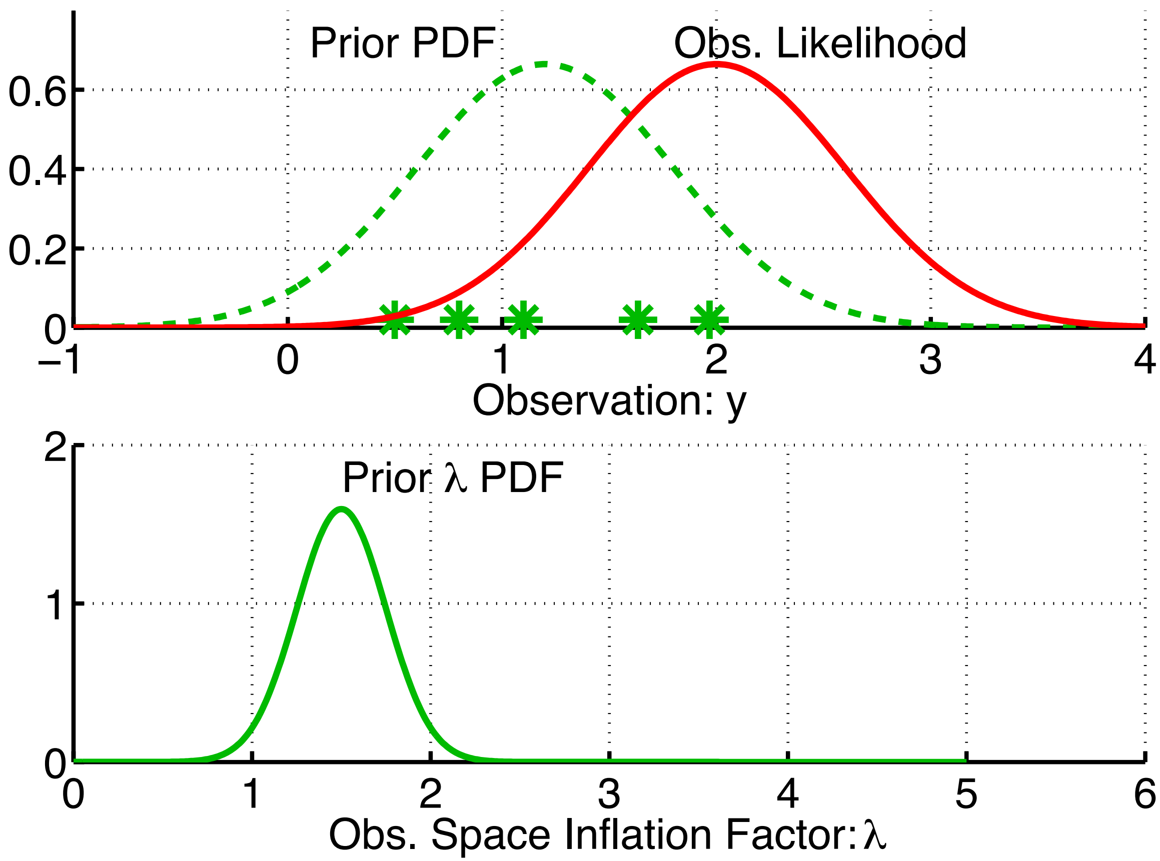
Use Bayesian statistics to get estimate of λ_k, the inflation factor at time t_k.



Assume prior is gaussian: $p(\lambda_k | Y_{k-1}) = N\left(\bar{\lambda}_p, \sigma^2_{\lambda,p}\right)$

Variance inflation for observations: An adaptive error-tolerant filter

Use Bayesian statistics to get estimate of λ_k, the inflation factor at time t_k.



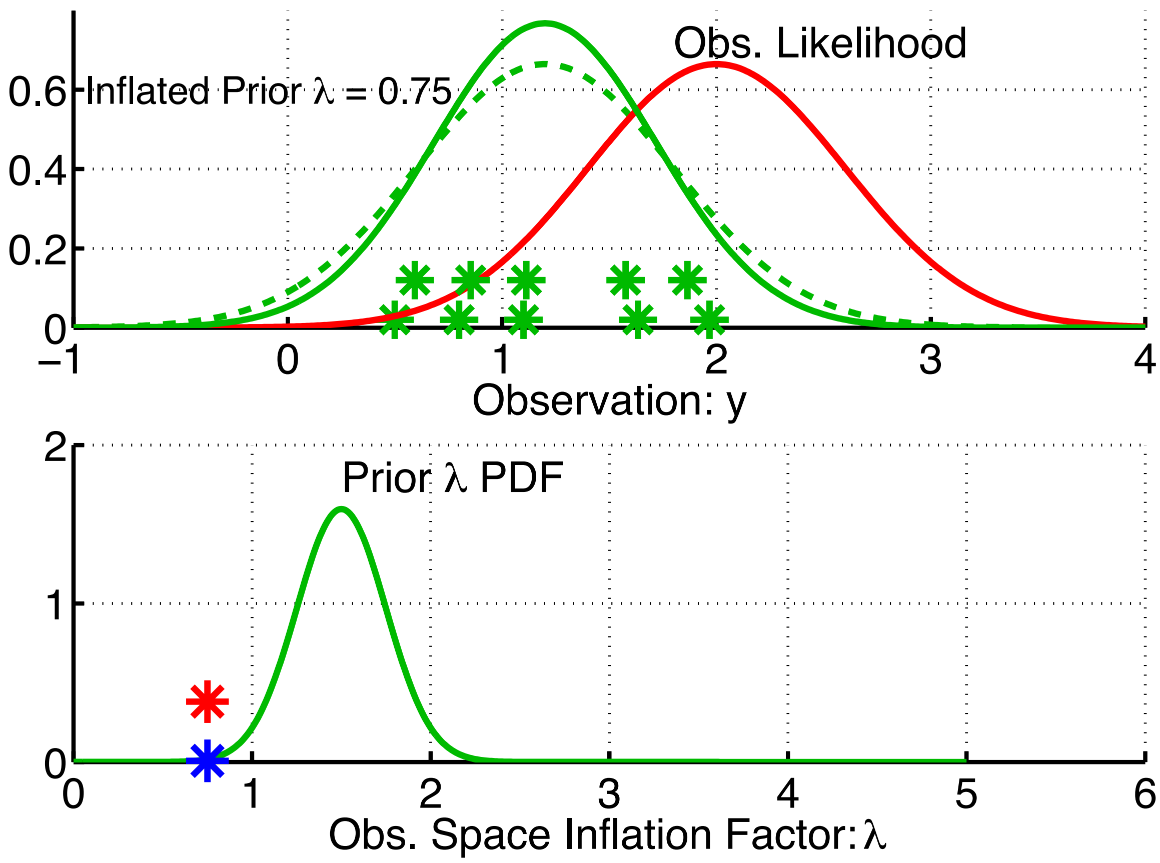
We've assumed a gaussian for prior $p(\lambda_k|Y_{k-1})$.

Recall that the likelihood, $p(y_k|\lambda)$ can be evaluated from normal PDF using the last equation on slide 5.

$$p(\lambda_k, Y_k) = p(y_k|\lambda)p(\lambda_k|Y_{k-1})/\text{normalization}$$

Variance inflation for observations: An adaptive error-tolerant filter

Use Bayesian statistics to get estimate of λ_k, the inflation factor at time t_k.



Get $p(y_k|\lambda = 0.75)$ from normal PDF, last eq. on slide 5; this is red asterisk on lower panel.

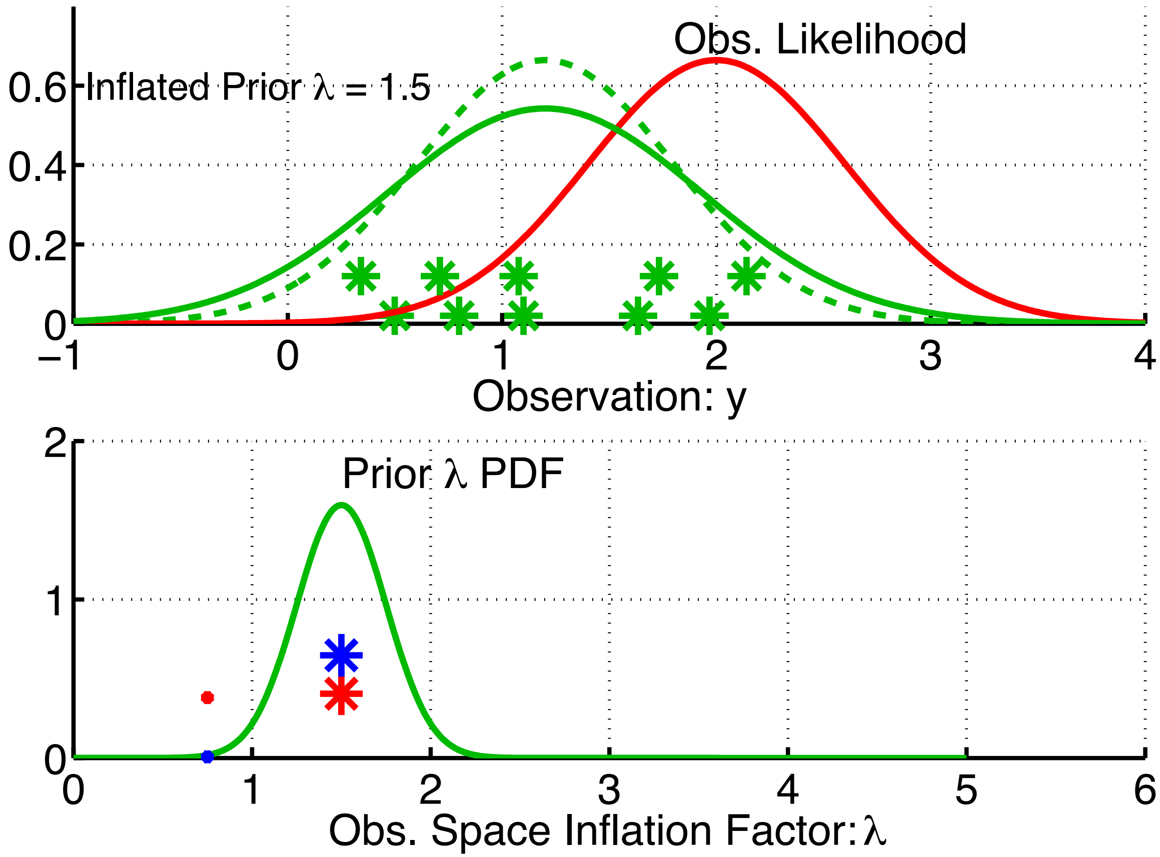
Multiply by $p(\lambda_k = 0.75|Y_{k-1})$, the value of green curve on the lower panel,

to get $p(\lambda_k = 0.75|Y_k)$, the blue asterisk.

$$p(\lambda_k, Y_k) = p(y_k|\lambda)p(\lambda_k|Y_{k-1})/\text{normalization}$$

Variance inflation for observations: An adaptive error-tolerant filter

Use Bayesian statistics to get estimate of λ_k, the inflation factor at time t_k.



Get $p(y_k|\lambda = 1.50)$ from normal PDF, last eq. on slide 5; this is red asterisk on lower panel.

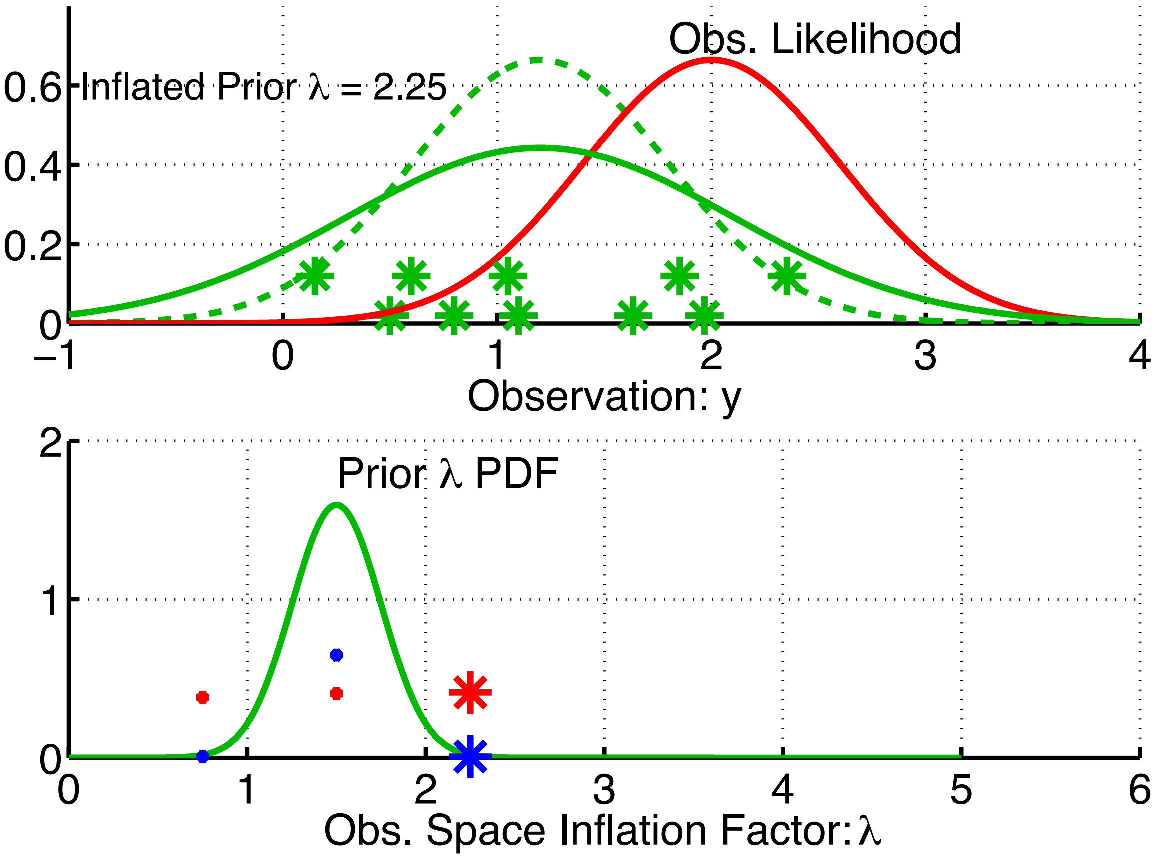
Multiply by $p(\lambda_k = 1.50|Y_{k-1})$, the value of green curve on the lower panel,

to get $p(\lambda_k = 1.50|Y_k)$, the blue asterisk.

$$p(\lambda_k, Y_k) = p(y_k|\lambda)p(\lambda_k|Y_{k-1})/\text{normalization}$$

Variance inflation for observations: An adaptive error-tolerant filter

Use Bayesian statistics to get estimate of λ_k, the inflation factor at time t_k.



Get $p(y_k|\lambda = 2.25)$ from normal PDF, last eq. on slide 5; this is red asterisk on lower panel.

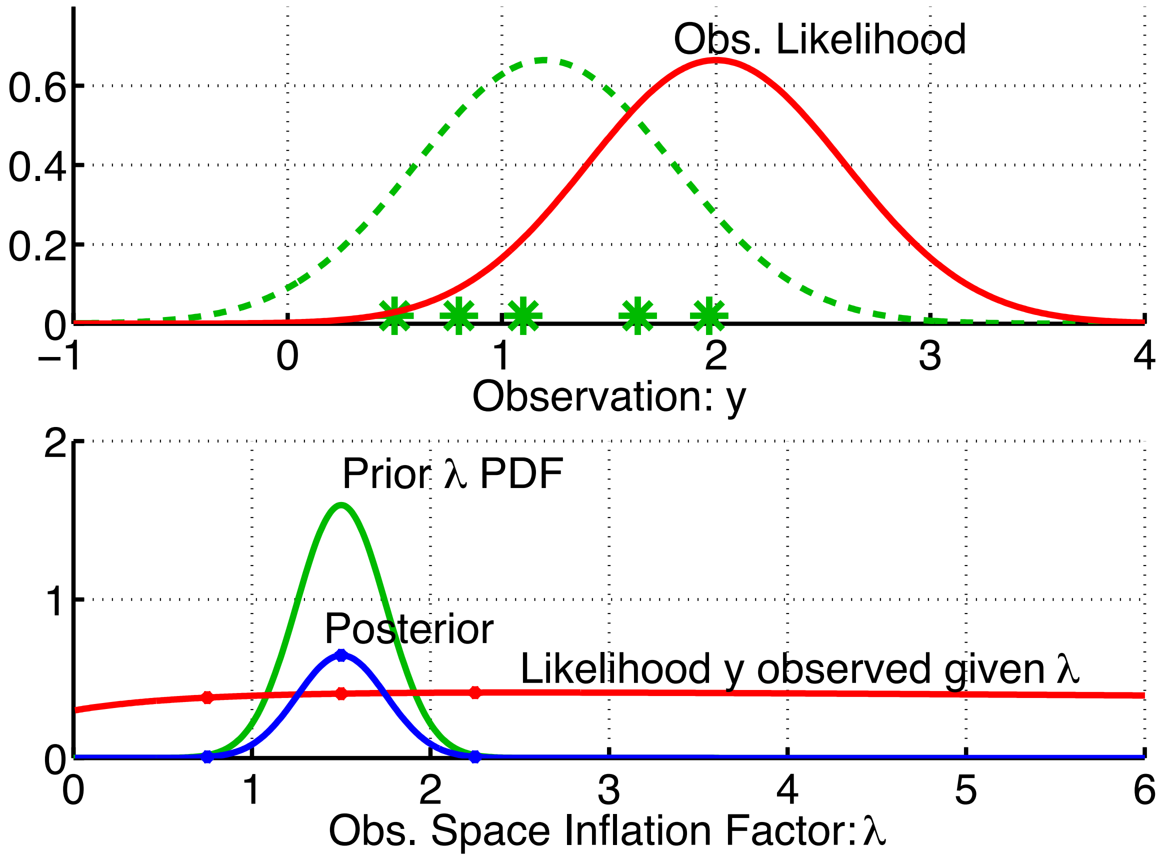
Multiply by $p(\lambda_k = 2.25|Y_{k-1})$, the value of green curve on the lower panel,

to get $p(\lambda_k = 2.25|Y_k)$, the blue asterisk.

$$p(\lambda_k, Y_k) = p(y_k|\lambda)p(\lambda_k|Y_{k-1})/\text{normalization}$$

Variance inflation for observations: An adaptive error-tolerant filter

Use Bayesian statistics to get estimate of λ_k, the inflation factor at time t_k.



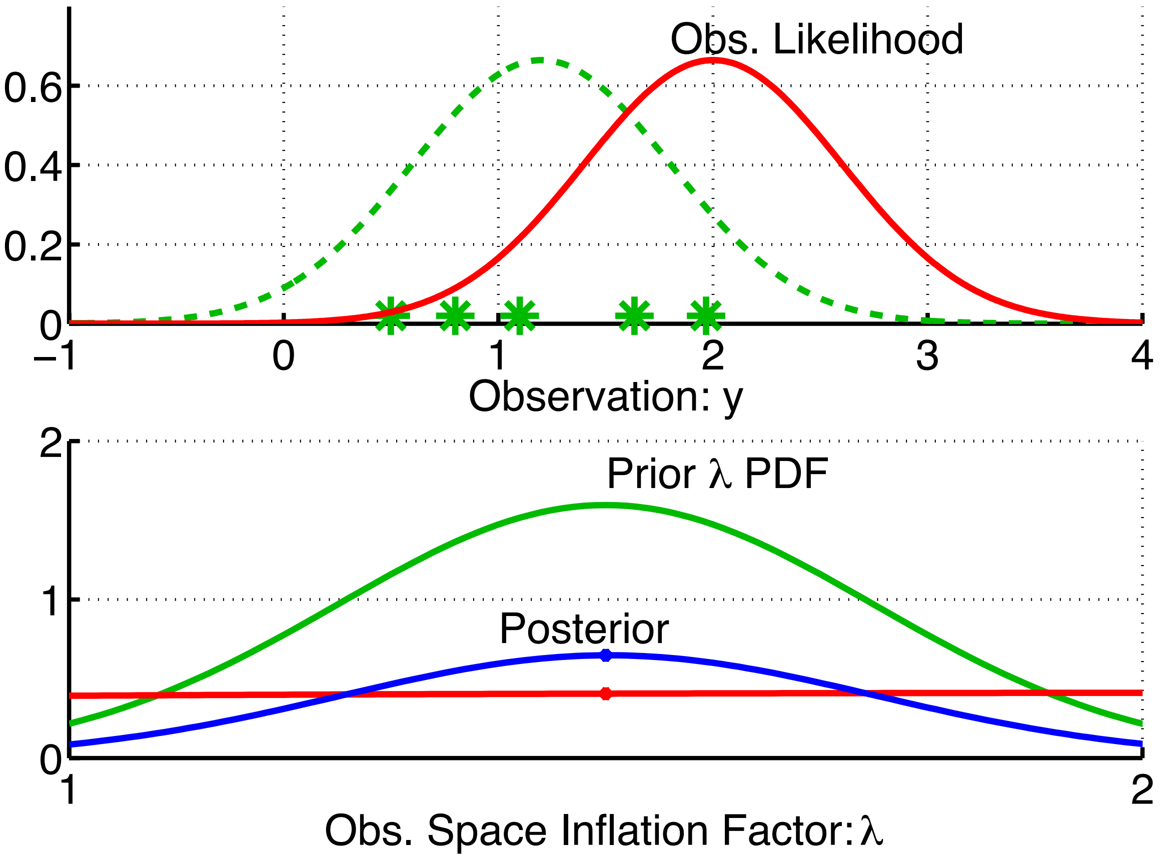
Repeat for a range of values of λ.

Now must get posterior in same form as prior (gaussian).

$$p(\lambda_k, Y_k) = p(y_k|\lambda)p(\lambda_k|Y_{k-1})/\text{normalization}$$

Variance inflation for observations: An adaptive error-tolerant filter

Use Bayesian statistics to get estimate of λ_k, the inflation factor at time t_k.



Very little information about λ in a single observation.

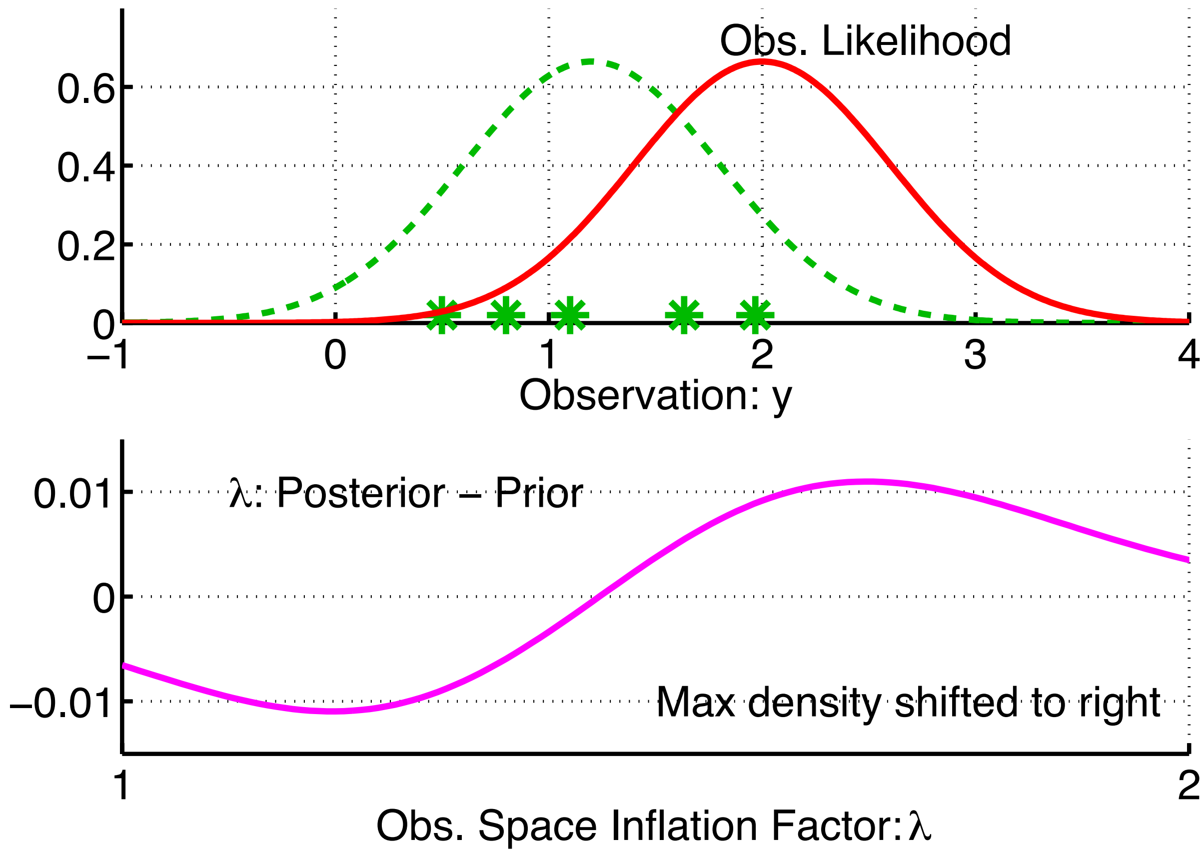
Posterior and prior are very similar.

Normalized posterior indistinguishable from prior.

$$p(\lambda_k, Y_k) = p(y_k|\lambda)p(\lambda_k|Y_{k-1})/\text{normalization}$$

Variance inflation for observations: An adaptive error-tolerant filter

Use Bayesian statistics to get estimate of λ_k, the inflation factor at time t_k.



Very little information about λ in a single observation.

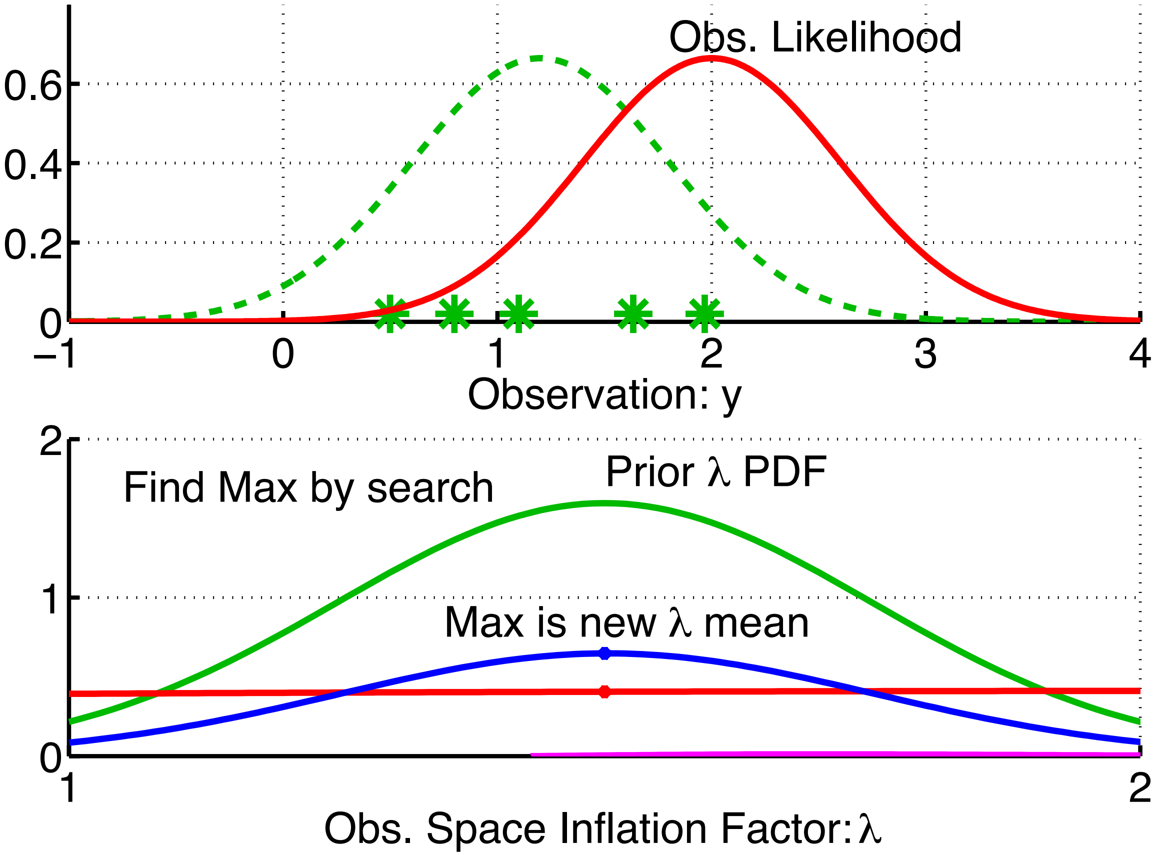
Posterior and prior are very similar.

Difference shows slight shift to larger values of λ.

$$p(\lambda_k, Y_k) = p(y_k|\lambda)p(\lambda_k|Y_{k-1})/\text{normalization}$$

Variance inflation for observations: An adaptive error-tolerant filter

Use Bayesian statistics to get estimate of λ_k, the inflation factor at time t_k.



One option is to use Gaussian posterior for λ.

Select max (mode) of posterior as mean of updated Gaussian.

Do a fit for updated standard deviation.

$$p(\lambda_k, Y_k) = p(y_k|\lambda)p(\lambda_k|Y_{k-1})/\text{normalization}$$

Variance inflation for observations: An adaptive error-tolerant filter

A. Computing updated inflation mean, $\bar{\lambda}_u$.

Mode of $p(y_k|\lambda)p(\lambda_k|Y_{k-1})$ can be found analytically!

Solving of $\frac{\partial[p(y_k|\lambda)p(\lambda_k|Y_{k-1})]}{\partial y} = 0$ leads to 6th order poly in θ.

This can be reduced to a cubic equation and solved to give mode.

New $\bar{\lambda}_u$ is set to the mode.

This is relatively cheap compared to computing regressions.

Variance inflation for observations: An adaptive error-tolerant filter

B. Computing updated inflation variance, $\sigma^2_{\lambda,u}$.

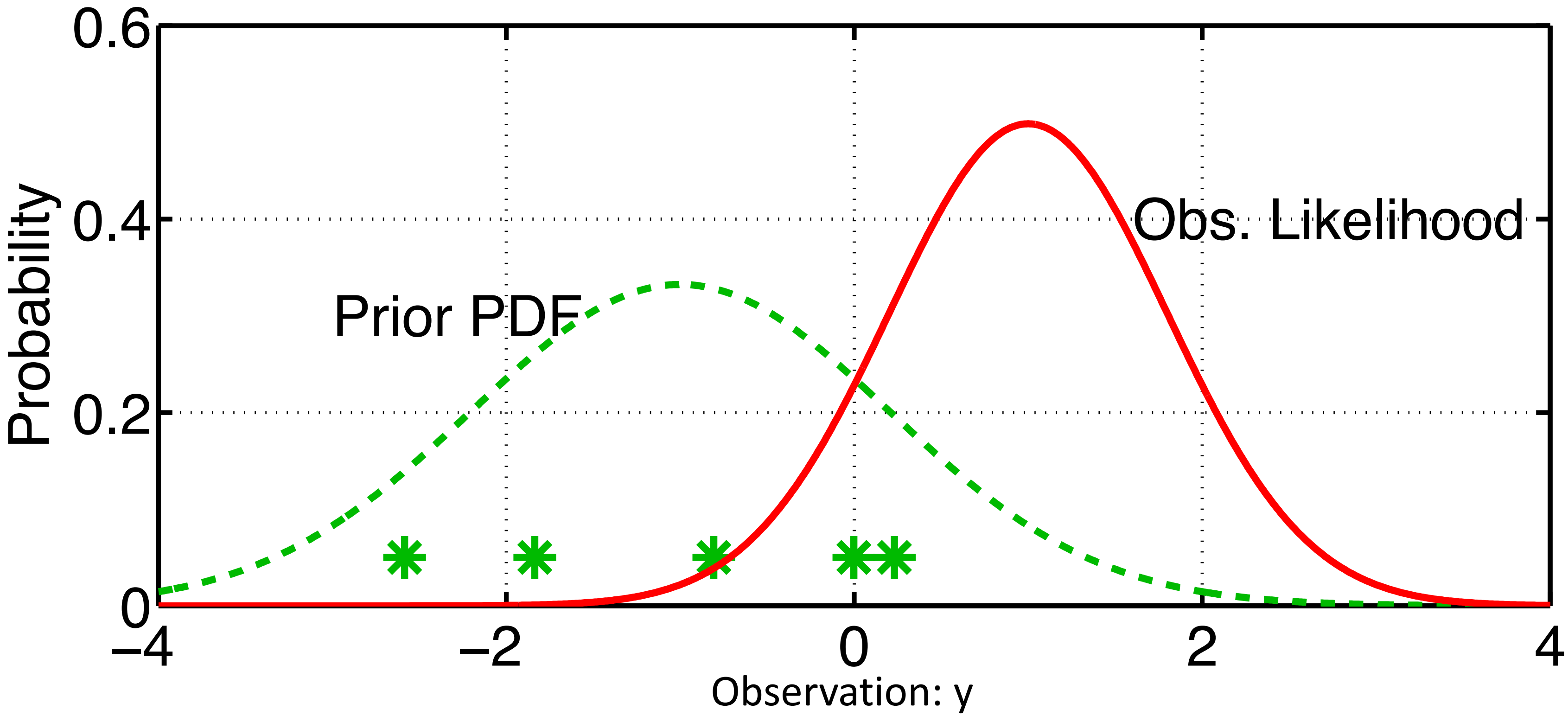
1. Evaluate numerator at mean, $\rho_m = p(\bar{\lambda}_u)$

2. Evaluate numerator at second point, $\rho_\sigma = p(\bar{\lambda}_u + \sigma_{\lambda,p})$

3. Find $\sigma^2_{\lambda,u}$ so $N(\bar{\lambda}_u, \sigma^2_{\lambda,u})$ goes through ρ_m and ρ_σ

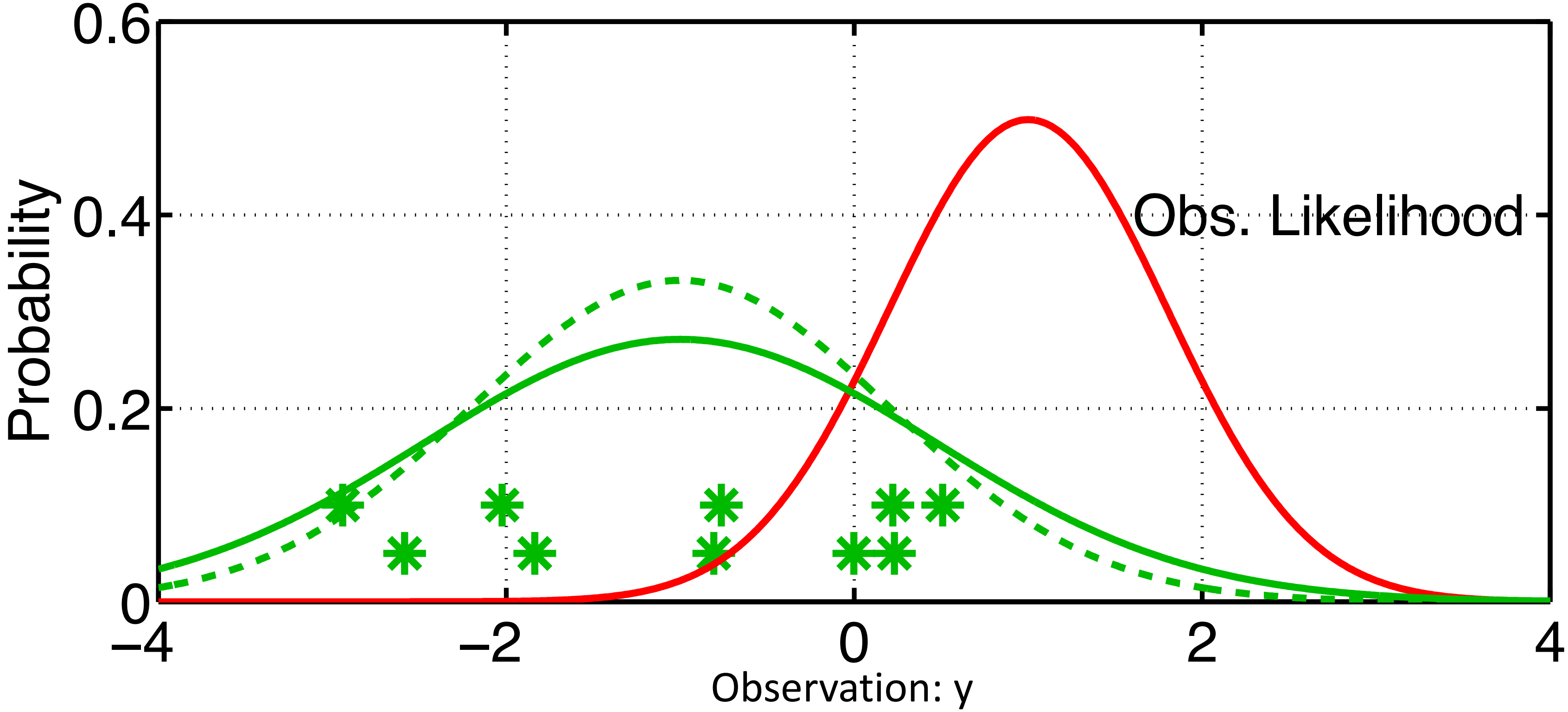
4. Compute as $\sigma^2_{\lambda,u} = -\sigma^2_{\lambda,p}/2ln(r)$ where $r = \rho_\sigma/\rho_m$

Single Variable Computations with Adaptive Error Correction



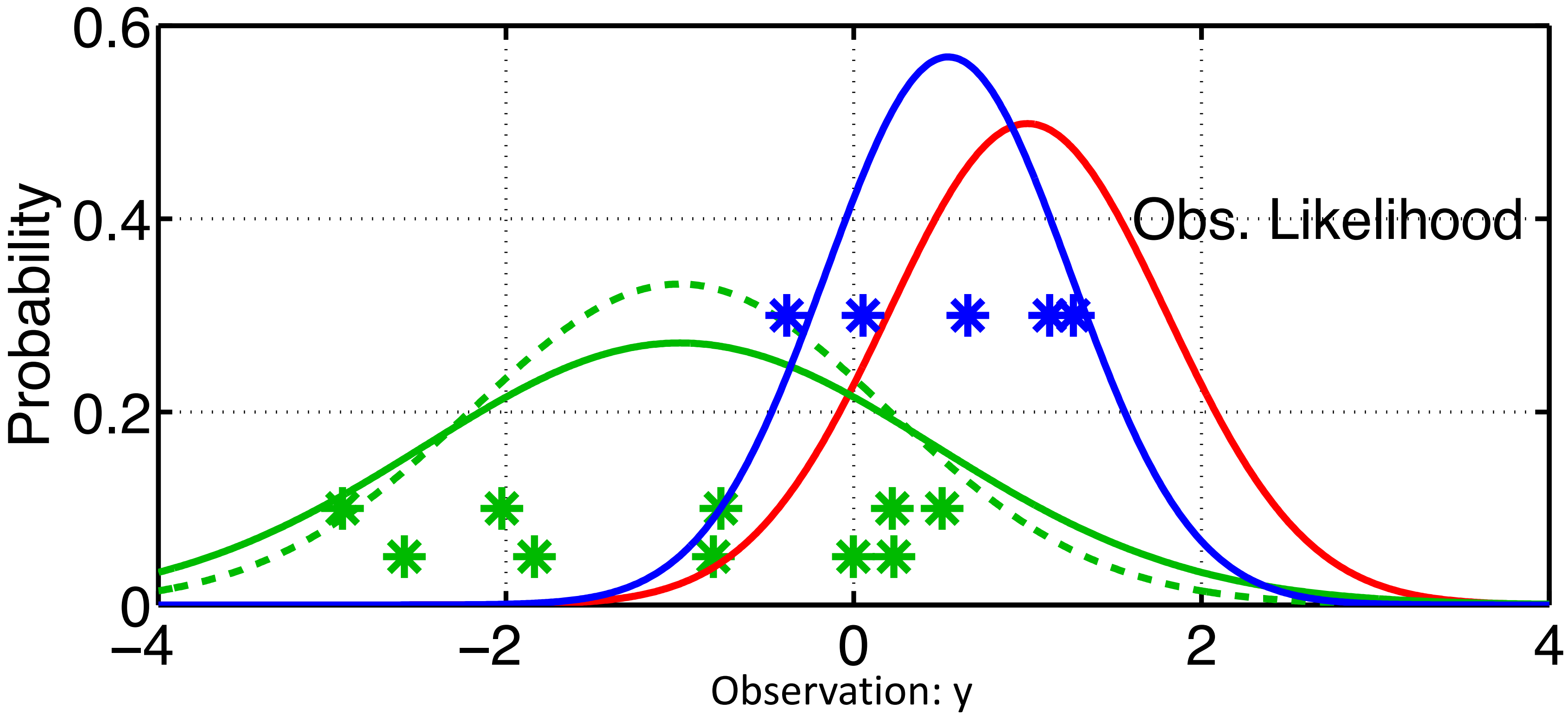
1. Compute updated inflation distribution, $p(\lambda_k|Y_k)$.

Single Variable Computations with Adaptive Error Correction



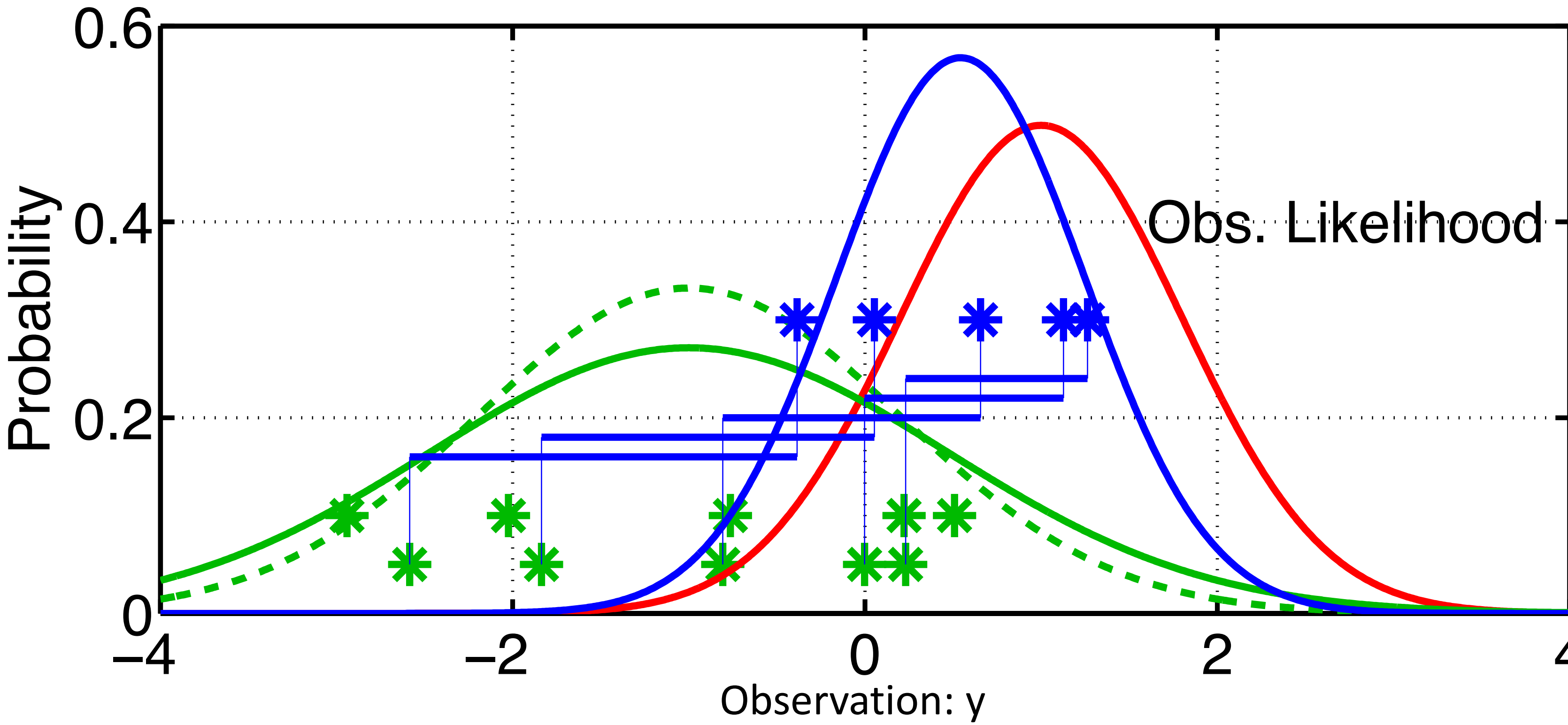
1. Compute updated inflation distribution, $p(\lambda_k|Y_k)$.
2. Inflate ensemble using mean of updated λ distribution.

Single Variable Computations with Adaptive Error Correction



1. Compute updated inflation distribution, $p(\lambda_k|Y_k)$.
2. Inflate ensemble using mean of updated λ distribution.
3. Use inflated prior to compute posterior ensemble of y.

Single Variable Computations with Adaptive Error Correction



1. Compute updated inflation distribution, $p(\lambda_k|Y_k)$.
2. Inflate ensemble using mean of updated λ distribution.
3. Compute posterior for y using inflated prior.
4. Compute increments from ORIGINAL prior ensemble.

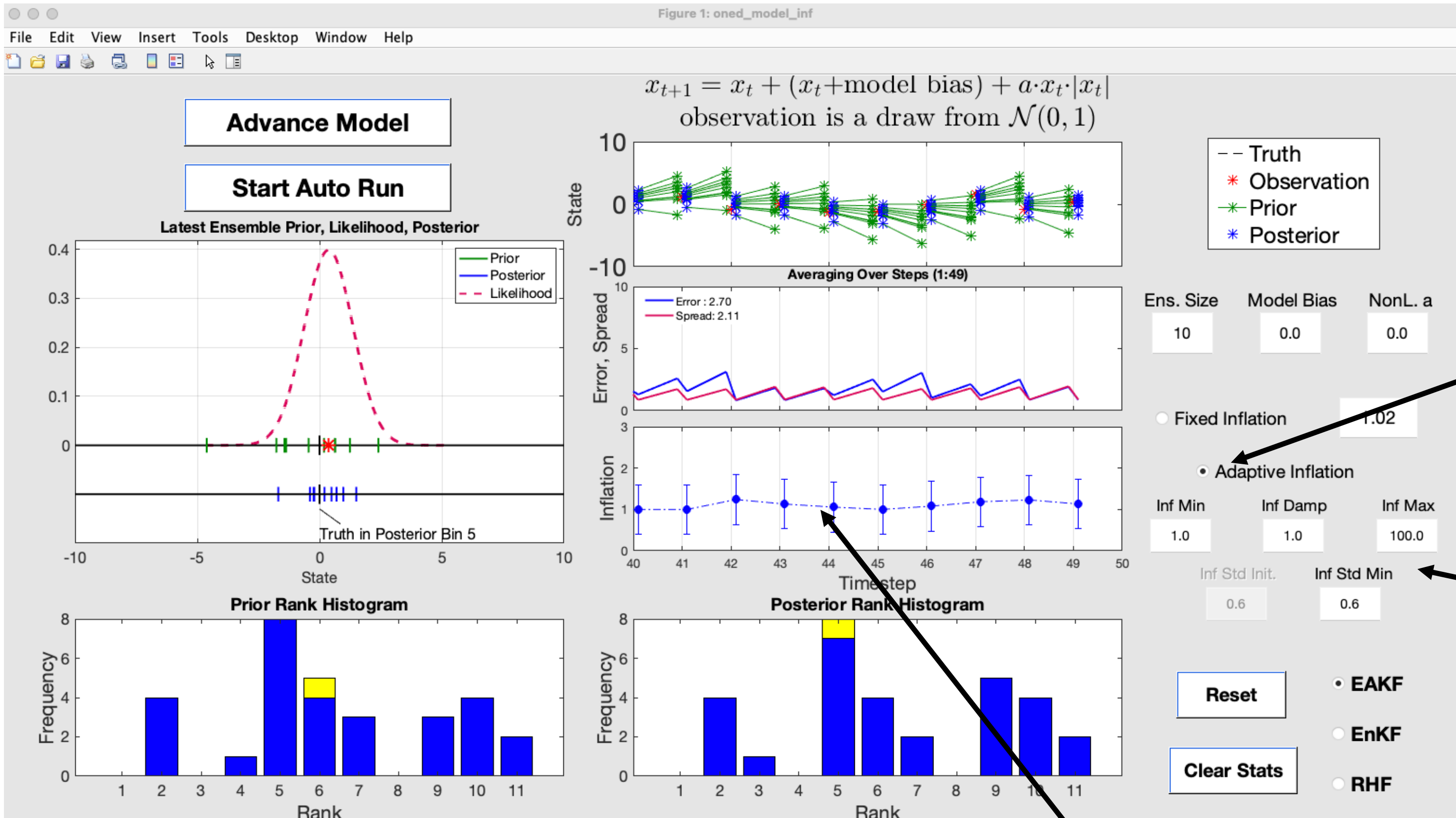
Matlab Hands-on: oned_model_inf

Adaptive inflation can be tested with matlab script *oned_model_inf.m*

Can explore 5 different values that control adaptive inflation:

- Minimum value of inflation mean, often set to 1 (no deflation).
- Inflation damping, more on this later. Value of 1.0 turns it off.
- Maximum value of inflation mean.
- The initial value of the inflation standard deviation.
- A lower bound on inflation standard deviation (it will asymptote to zero if allowed).

Matlab Hands-on: oned_model_inf



Turn on adaptive inflation.

Inflation settings.

Inflation value and its standard deviation.

Matlab Hands-on: oned_model_inf

Oned_model_inf generates a summary file each time statistics are reset.

File *oned_model_inf.log* keeps track of parameter settings and metrics.

```
# Time step: 1 (Initial configuration)
  - Ensemble size = 4
  - Model bias = 0.00
  - Nonlinear `a` parameter = 0.00
  - Inflation value = 1.02
  - (Adaptive) Inflation lower bound = 1.00
  - (Adaptive) Inflation upper bound = 100.00
  - (Adaptive) Inflation damping factor = 1.00
  - (Adaptive) Inflation standard deviation = 0.60
  - (Adaptive) Inflation standard deviation lower bound = 0.60

# Time step: 19
  >> Statistics over period (1:19): avg. RMSE = 3.03, avg. Spread = 2.19
  $$ User input: Ensemble size has been changed from 4 to 10
```

Matlab Hands-on: oned_model_inf

Try adaptive inflation.

Pick a lower value for standard deviation.
Initial lower bound on inflation of 1.0, upper bound large (100).

Try introducing a model bias.

What happens if lower bound is less than 1?

Inflation Damping

Inflation mean damped towards 1 every assimilation time.

- *inf_damping* 0.9: 90% of the inflation difference from 1.0 is retained.

Can be useful in models with heterogeneous observations in time.
For instance, a well-observed hurricane crosses a model domain.
Adaptive inflation increases along hurricane trace.
After hurricane, fewer observations, no longer need so much inflation.

For large earth system models, following values may work:

inf_sd_initial = 0.6,
inf_damping = 0.9,
inf_sd_lower_bound = 0.6.

Adaptive Inflation of State Variables

Suppose we want a global inflation for state variables, λ_s , instead.

Make same least squares assumption that is used in ensemble filter.

Inflation of λ_s for state variables inflates obs. priors by same amount.

Get same likelihood as before: $p(y_O|\lambda) = (2\pi\theta^2)^{-1/2} exp(-D^2/2\theta^2)$

$$\theta = \sqrt{\lambda_s \sigma_p^2 + \sigma_O^2}$$

Compute updated distribution for λ_s exactly as for single observed variable.

Implementation of Adaptive Multivariate Inflation Algorithm

1. Apply inflation to state variables with mean of λ_s distribution.

2. Do following for observations at given time sequentially:
    a. Compute forward operator to get prior ensemble.
    b. Compute updated estimate for λ_s mean and variance.
    c. Compute increments for prior ensemble.
    d. Regress increments onto state variables.

Spatially varying adaptive inflation algorithm

Have a distribution for λ for each state variable i, $\lambda_{s,i.}$

Use prior correlation from ensemble to determine impact of $\lambda_{s,i}$ on prior variance for given observation.

If γ is correlation between state variable i and observation then

$$\theta = \sqrt{\left[1 + \gamma\left(\sqrt{\lambda_{s,i}} - 1\right)\right]^2 \sigma_p^2 + \sigma_O^2}$$

Equation for finding mode of posterior is now full 12th order:

Analytic solution appears unlikely.

Can do Taylor expansion of θ around $\lambda_{s,i}$.

Retaining linear term is normally quite accurate.

There is an analytic solution to find mode of product in this case!

Matlab Hands-on: run_Lorenz_96_inf

Spatially Varying Adaptive inflation can be tested with matlab script *run_lorenz96_inf.m*

Can explore 3 different values that control adaptive inflation:

- Minimum value of inflation, often set to 1 (no deflation).
- Inflation damping. Value of 1.0 turns it off.
- The value of the inflation standard deviation.
    - Lower bound on standard deviation is set to same value.
    - In this case, standard deviation just stays fixed at selected value.

Matlab Hands-on: run_Lorenz_96_inf

Adaptive inflation controls.

Adaptive inflation spatial plot.

Matlab Hands-on: run_Lorenz_96_inf

Explore some of the following:

How does adaptive inflation change as localization is changed?
How does adaptive inflation change for different values of inflation standard deviation?
If the lower bound is smaller than 1, does deflation (inflation < 1) happen?

Matlab Hands-on: run_Lorenz_96_inf

run_Lorenz_96_inf generates a summary file each time statistics are reset.

File *run_lorenz_96_inf.log* keeps track of parameter settings and metrics.

```
# Time step: 63
  >> Statistics over period (35:63): avg. Prior RMSE = 1.25, avg. Prior
Spread = 1.32
  $$ User input: Assimilation Type has been changed from `EAKF` to `RHF`


     Current configuration:
     - Forcing `F` parameter = 8.00
     - Assimilation type is `RHF`
     - Observation Network is `1:40:1`
     - Ensemble size = 20
     - Localization = 0.30
     - Inflation Algorithm is `Gaussian`
     - Adaptive inflation lower bound = 1.00
     - Adaptive inflation upper bound = 5.00
     - Adaptive inflation damping factor = 0.90
     - Adaptive inflation standard deviation = 0.60
     - Adaptive inflation standard deviation lower bound = 0.60
```

Simulating Model Error in 40-Variable Lorenz-96 Model

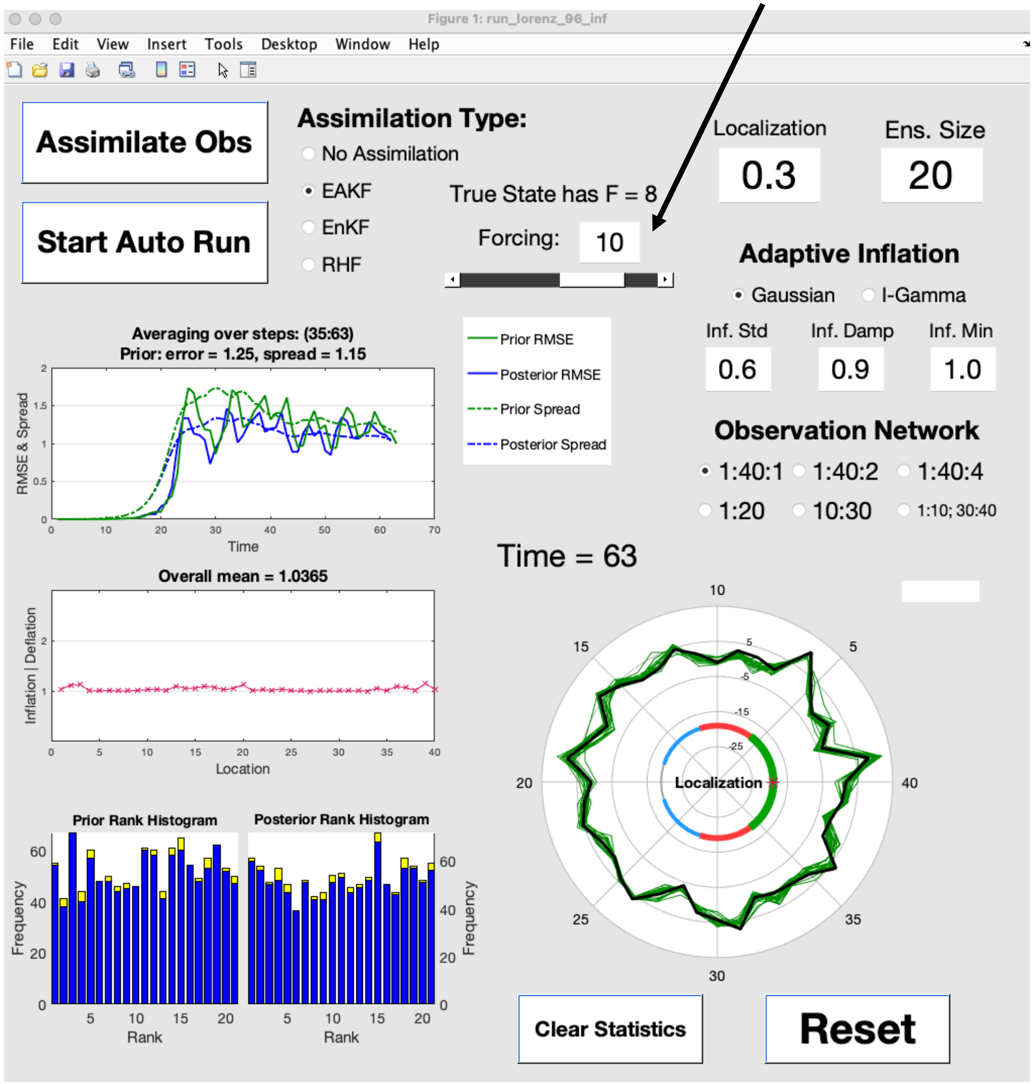
Inflation can deal with all sorts of errors, including model error.

Can simulate model error in Lorenz-96 by changing forcing.
Synthetic observations are from model with forcing = 8.0.

Both *run_lorenz_96* and *run_lorenz_96_inf* allow model error.

Spatially Varying Adaptive Inflation

Model error: Change forcing for assimilating model here.



Spatially Varying Adaptive Inflation with Model Error

Explore some of the following:

Change the model forcing to a larger or smaller value (say 6 or 10).

How does adaptive inflation respond to model error?
Do good values of localization change as model error increases?

Observing Network Exploration

The impact of the observing system can be explored.

Lorenz-96 has 40 state variables.

Following observing systems are available:

- Observe all 40 variables (1:40:1)
- Observe every other variable (1:40:2)
- Observe every 4^{th} variable (1:40:4)
- Observe the first 20 variables only (1:20)
- Observe the 10^{th} to 30^{th} variables only (10:30)
- Observe variables 30 to 40 and 1 to 10 (1:10; 30:40)

Observing Network Exploration

Observing network controls.

Inflation Distribution Family

The inflation for each state variable is a random variable.

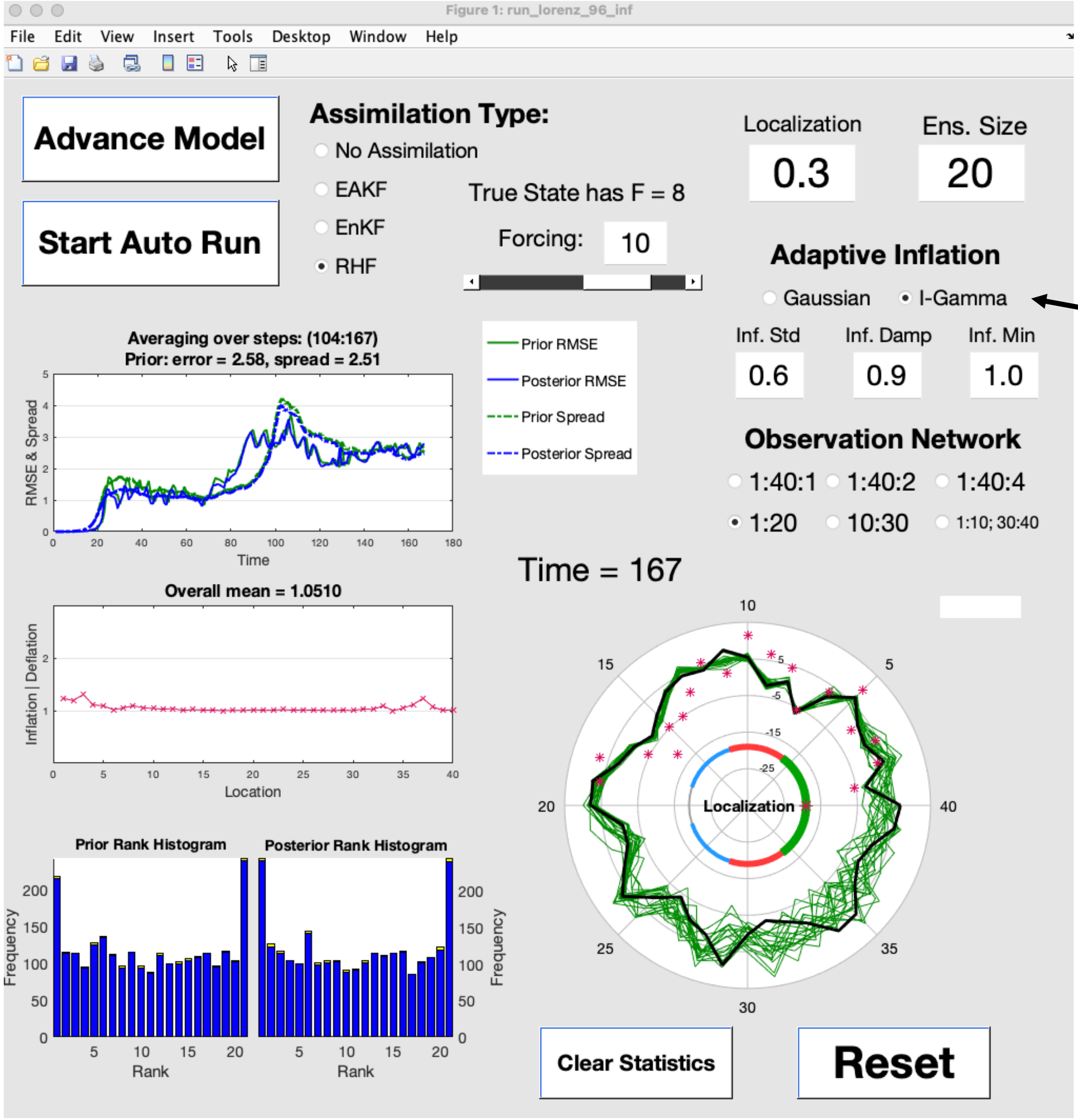
So far, it has been a normal distribution.

But, inflation/deflation is bounded below at zero.

Just like for bounded state variables, we can use a more appropriate distribution.

Inverse gamma is a good choice and is now the DART default.

Inflation Distribution Family



Choose distribution family for inflation.