Bayesian Sequential Sampling Policies and Sufficient Conditions for Convergence to a Global Optimum

Peter Frazier[1] Warren Powell[2]

[1]Operations Research & Information Engineering, Cornell University
[2]Operations Research & Financial Engineering, Princeton University

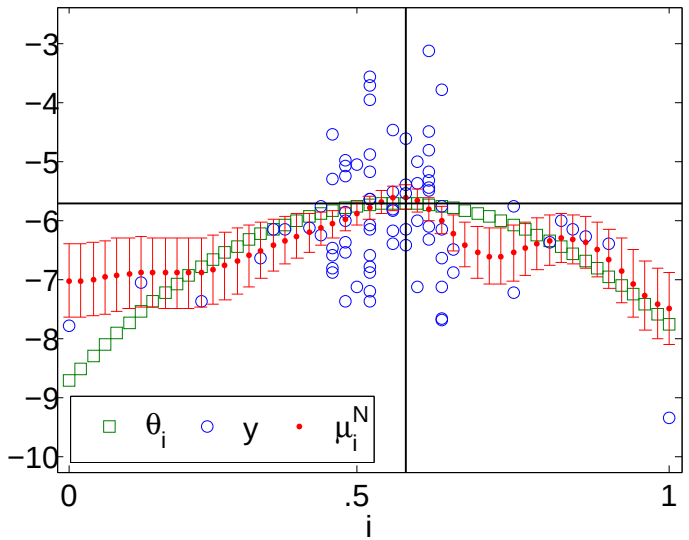
Sunday July 12, 2009
INFORMS Applied Probability Society Conference
Ithaca, NY

Sequential Information Collection Problem

- A **sequential information collection problem** is any problem with the following characteristics:
  - we would like to learn about some unknown through measurement, and then make some real-world decision based on what we learn.
  - The measurements are made sequentially.
  - Call the real-world decision the **implementation decision**, to distinguish it from the **measurement decisions**.
  - A **sequential sampling policy** is an adaptive rule for choosing measurement decisions.
- By making measurement decisions sequentially rather than statically, we can often learn much more efficiently.

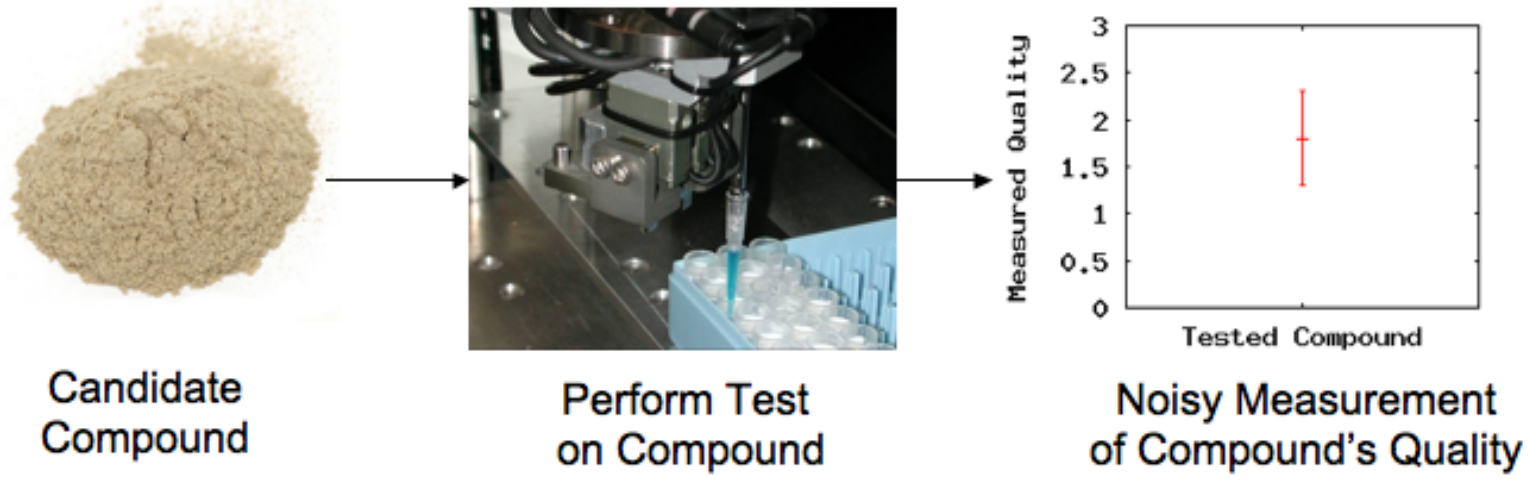
Example: Simulation Optimization

- We have a discrete event simulation with several parameters, and we would like to know which parameter values cause the simulation output to have the largest mean.
  - For example, we would like to choose the design of an emergency room to maximize patient throughput.
- Measurement decision: Which parameter values should we simulate, and for how long?
- Proceed sequentially, first obtaining rough estimates over the space, and then focus on portions of the space in which the maximum is more likely to reside.



Example: Drug Discovery

- A pharmaceutical research group working to develop a new drug begins with a library of millions of compounds, from which they wish to identify those compounds that perform well.
- Robots perform an initial screening, producing noisy estimates of each compound's quality.
- Researchers then decide sequentially which compounds to revisit, and which new chemical compounds to synthesize and test.
- Measurement decision: Which chemical compound should we test, and which laboratory test should we use?



Sequential Bayesian Information Collection Problems

These problems can all be formulated as **sequential Bayesian information collection problems**.

1. We begin with a prior distribution on some unknown truth θ,
   - e.g., θ contains the true mean and variance of simulation output at each set of parameter values.
2. We make a sequence of measurements, deciding which types of measurement to make as we go.
   - e.g., We decide sequentially which set of parameters to test next.
3. After T measurements, we choose an implementation decision i and suffer a loss $R(\theta, i)$.
   - e.g., After measuring, we choose a set of parameters i and build our emergency room according to those parameters. The difference between the patient throughput we get and the best patient throughput we could have gotten is $R(\theta, i)$.

Sequential Bayesian Information Collection Problems

- We try to minimize $R(\theta, i)$ by choosing our implementation decision i as well as we can with the information we have about θ.
- The best expected loss that we can achieve with our measurement policy is

$$\mathbb{E}\left[min_i \mathbb{E}_T\left[R(\theta, i)\right]\right].$$

- The best that we can do with perfect knowledge is

$$\mathbb{E}\left[min_i R(\theta, i)\right].$$

Convergence to a Global Optimum

- We say that a measurement policy **converges to a global optimum** if

$$\lim_{T \to \infty} \mathbb{E}\left[\min_i \mathbb{E}_T\left[R(\theta, i)\right]\right] = \mathbb{E}\left[\min_i R(\theta, i)\right].$$

- This means that, in the limit as we have an infinite number of measurements, we do as well as we would with perfect knowledge.
- This can also be viewed as **asymptotic optimality**: we do as well as the optimal sequential sampling policy in the limit.
- Contribution: we provide general sufficient conditions under which convergence to a global optimum holds.

Convergence to a Global Optimum

- Many simple policies (e.g., equal allocation) converge to global optima, but perform very badly for finitely many samples.
- If we have a more complex policy that performs well in some numerical experiments, we would like to know that it also converges to a global optimum.
- Convergence to a global optimum is **not** a proof of quality, but lack of it suggests there may be something wrong.

Simple Case

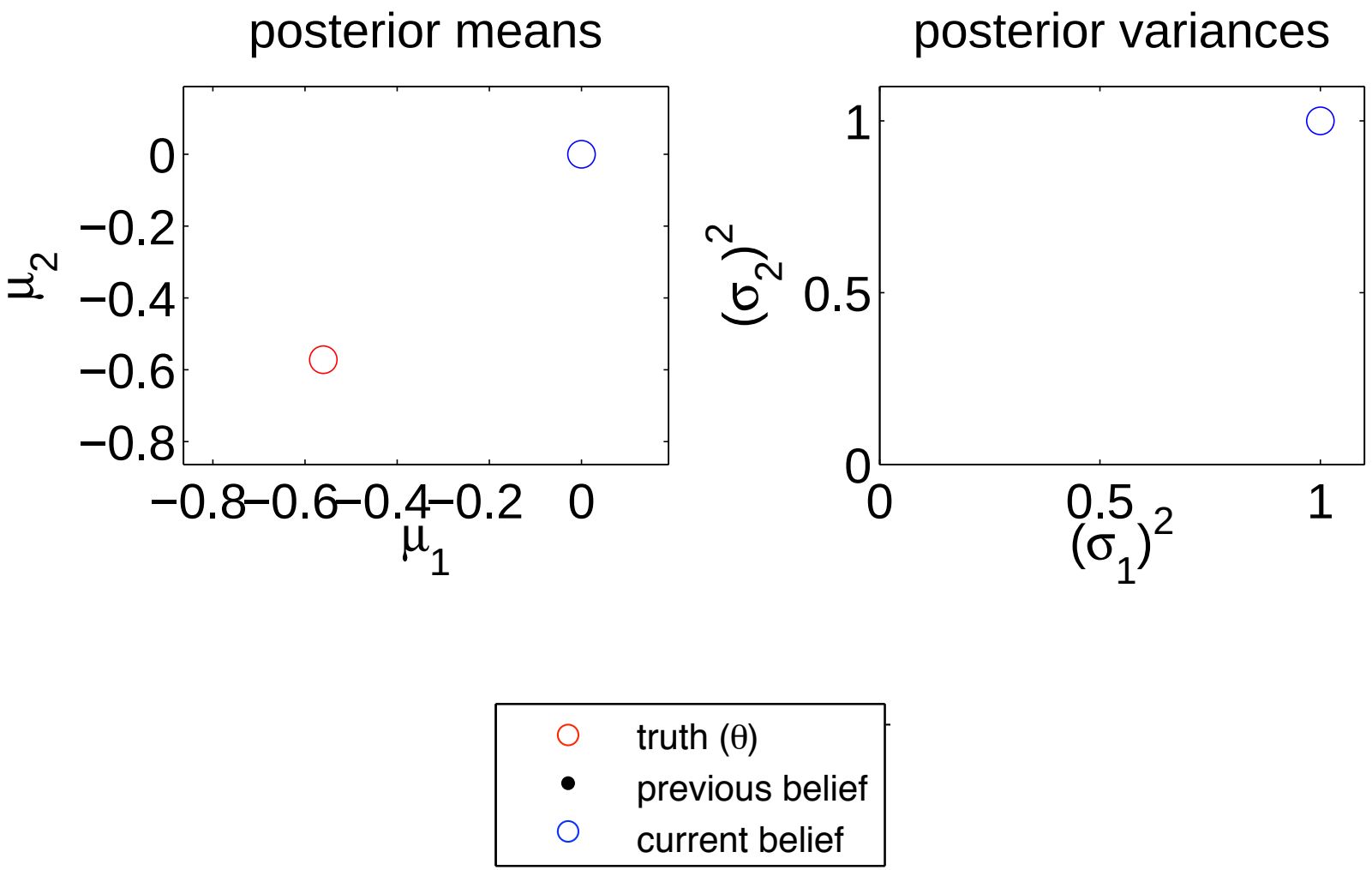
Suppose we have two new chemical compounds for treating a disease, Treatment 1 and Treatment 2.

- θ_i is the unknown quality of Treatment i, $i = 1, 2$.
- Under our prior, the θ_i are independent and $\text{Normal}(0, 1)$.
- Measurements of Treatment i are distributed as $\text{Normal}(\theta_i, 1)$.
- Our posterior at time t on θ_i is $\text{Normal}(\mu_{ti}, \sigma_{ti}^2)$, where

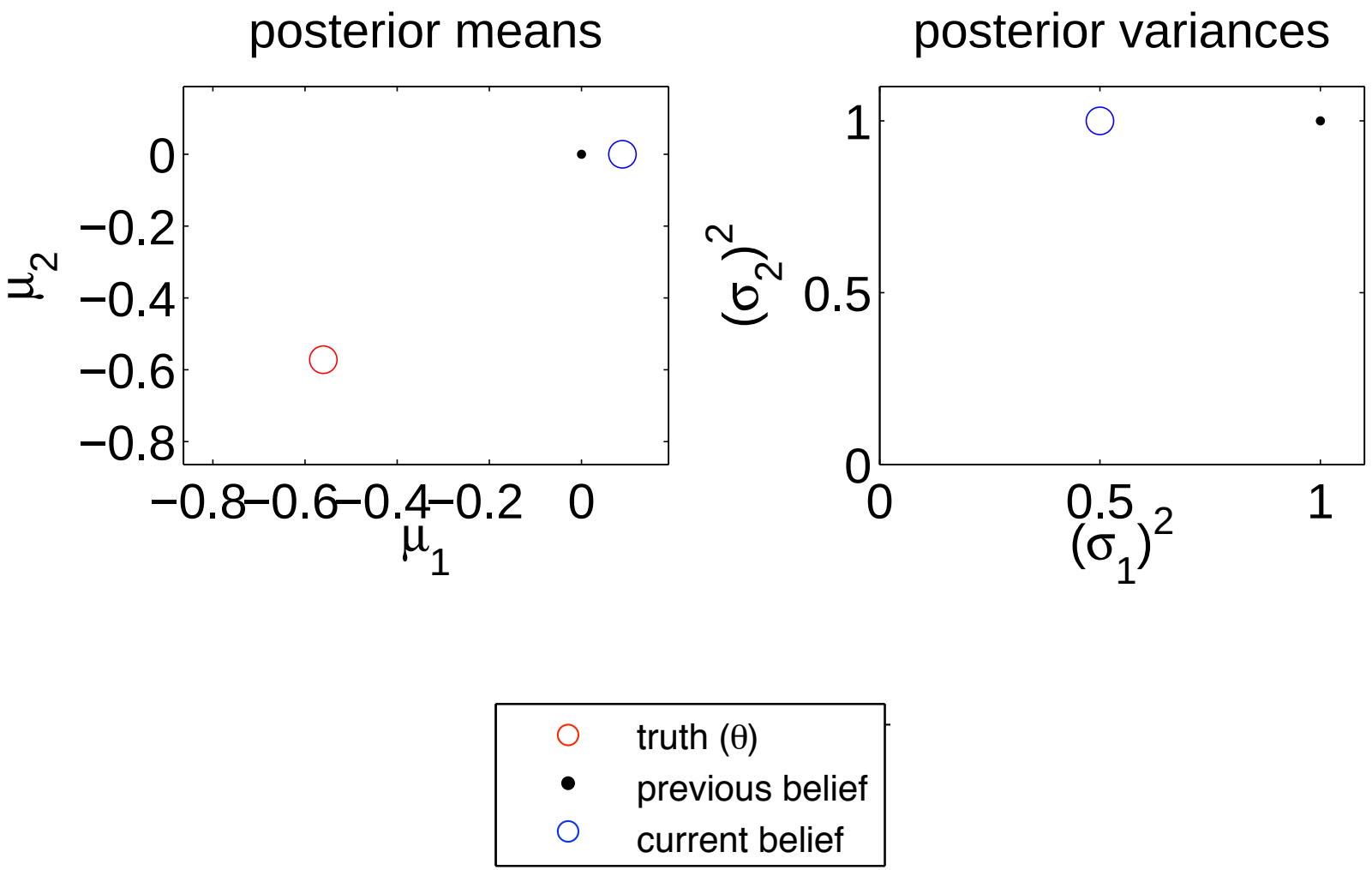
$$\mu_{t+1,i} = \left(\mu_{t,i}\sigma_{ti}^{-2} + \text{observation}\right) / \left(\sigma_{ti}^{-2} + 1\right),$$
$$\sigma_{t+1,i}^2 = 1 / \left(\sigma_{ti}^{-2} + 1\right).$$

  when we measure i at time t.

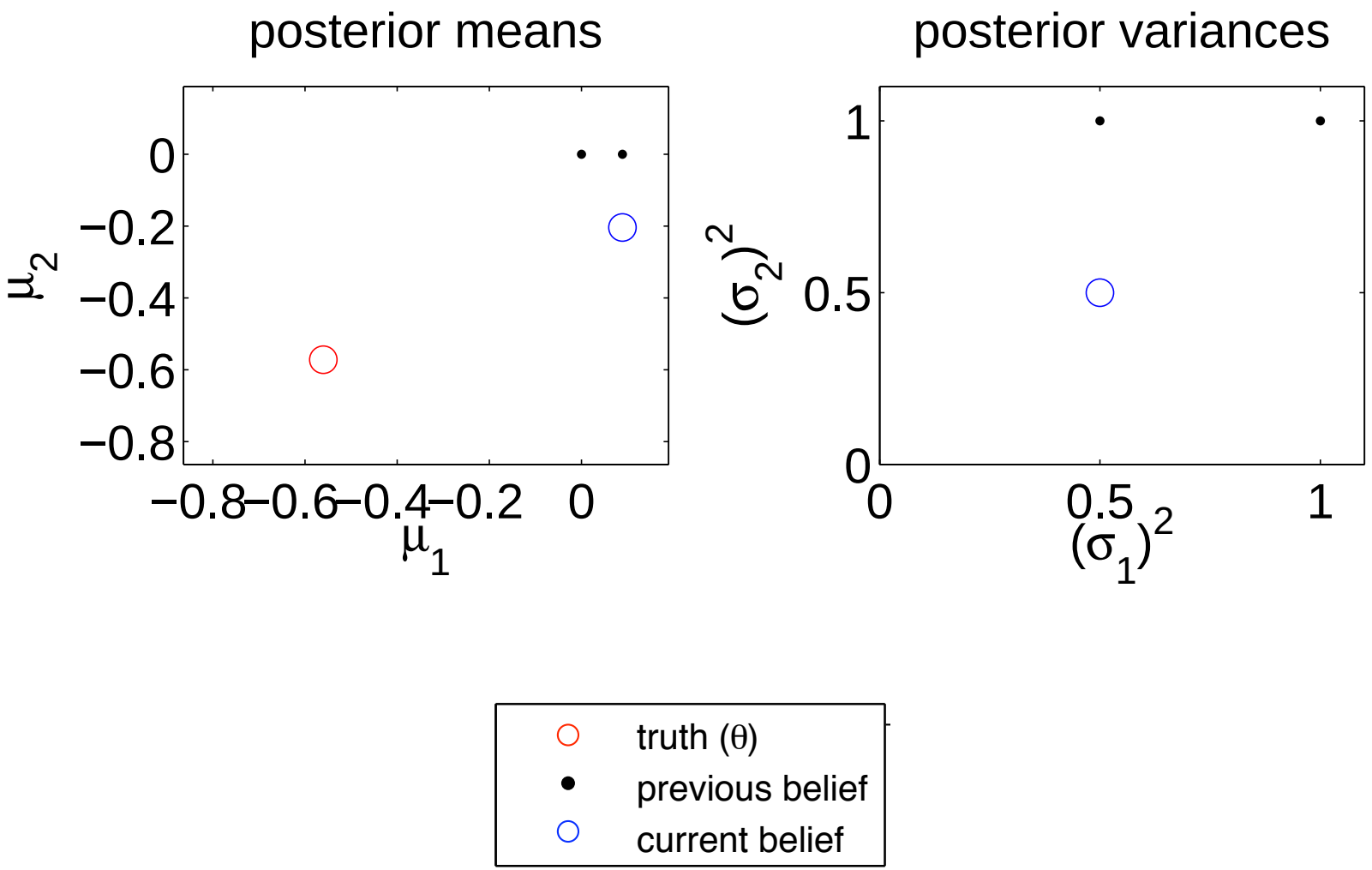
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

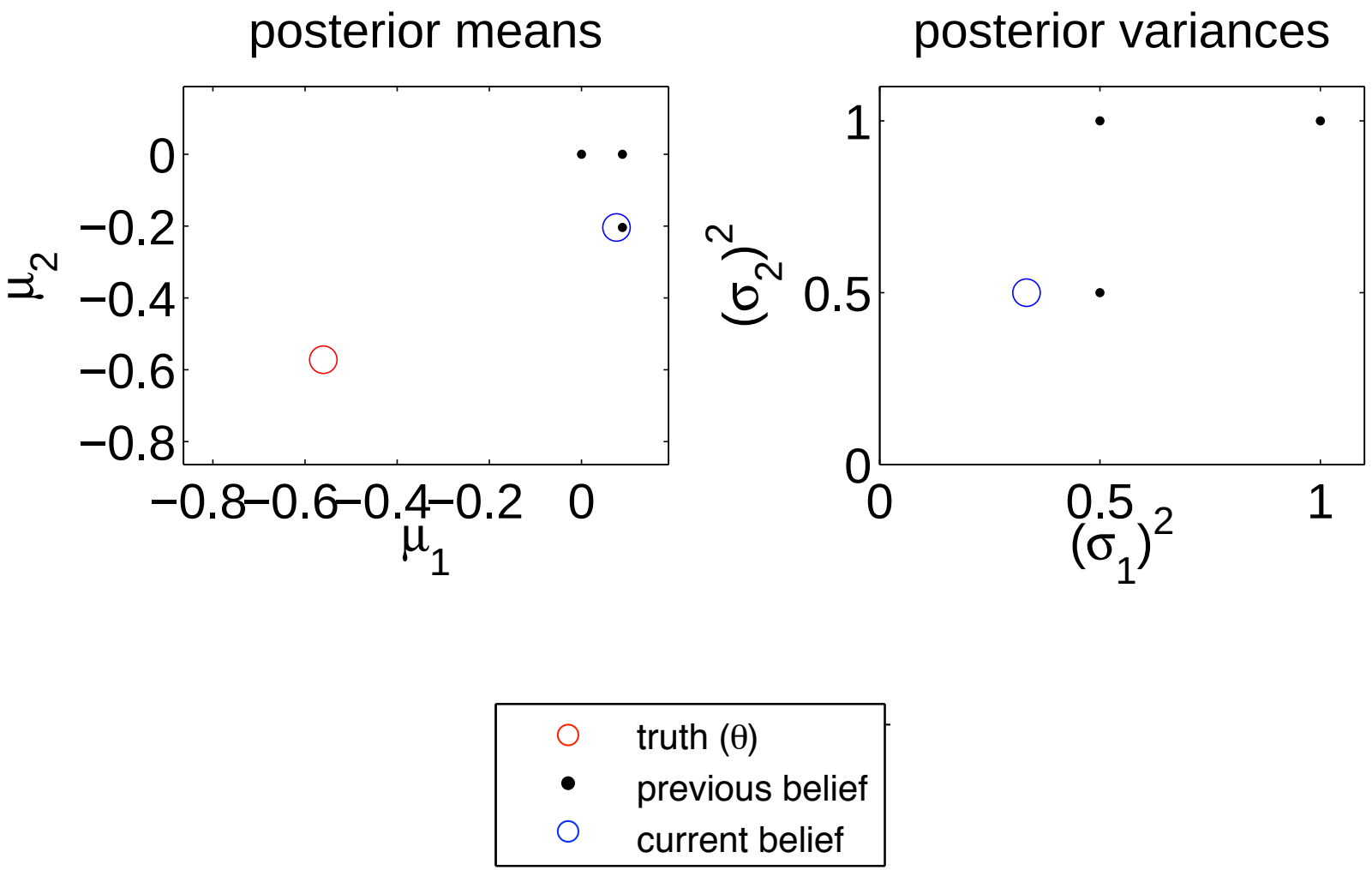
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

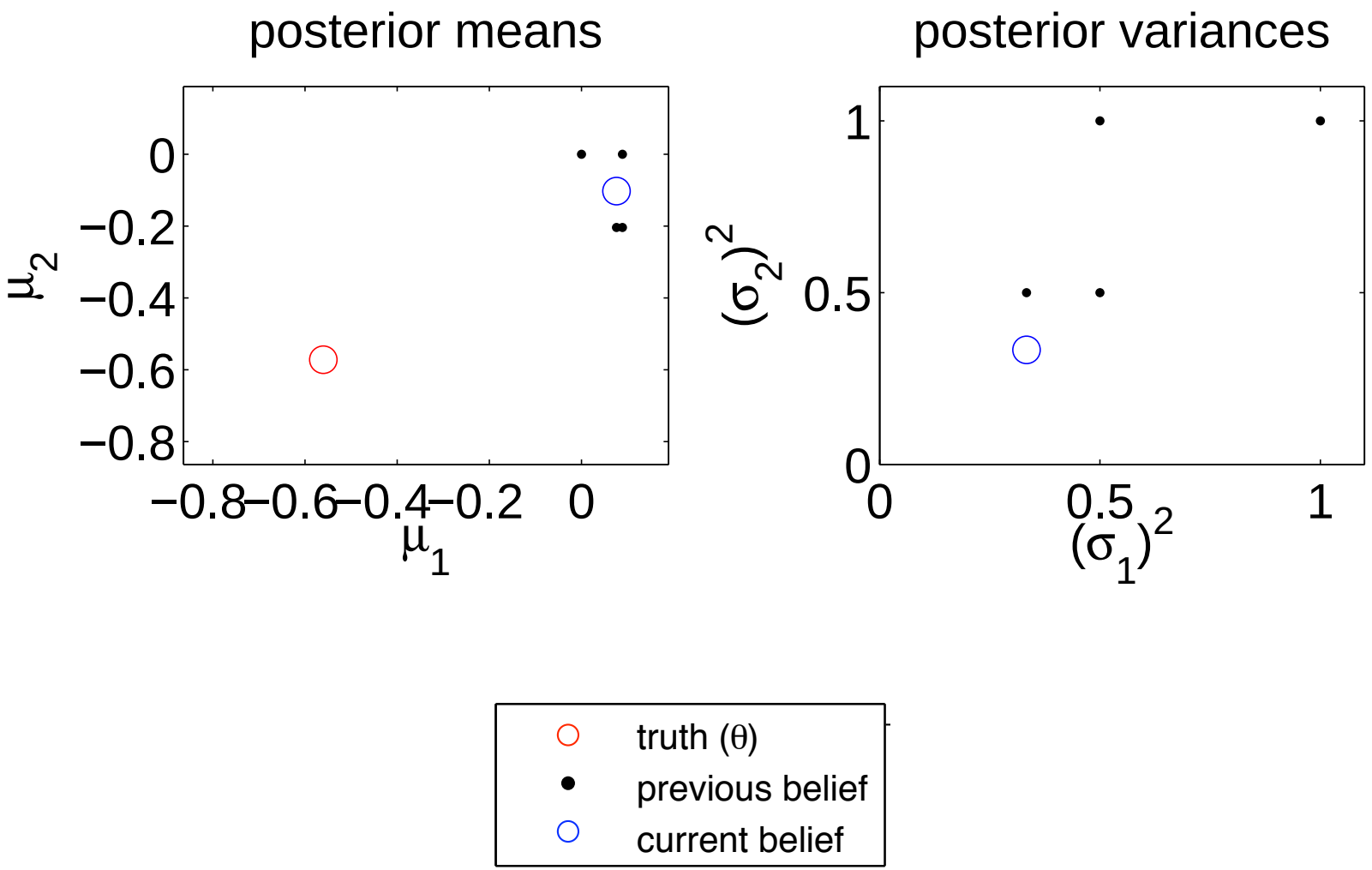
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

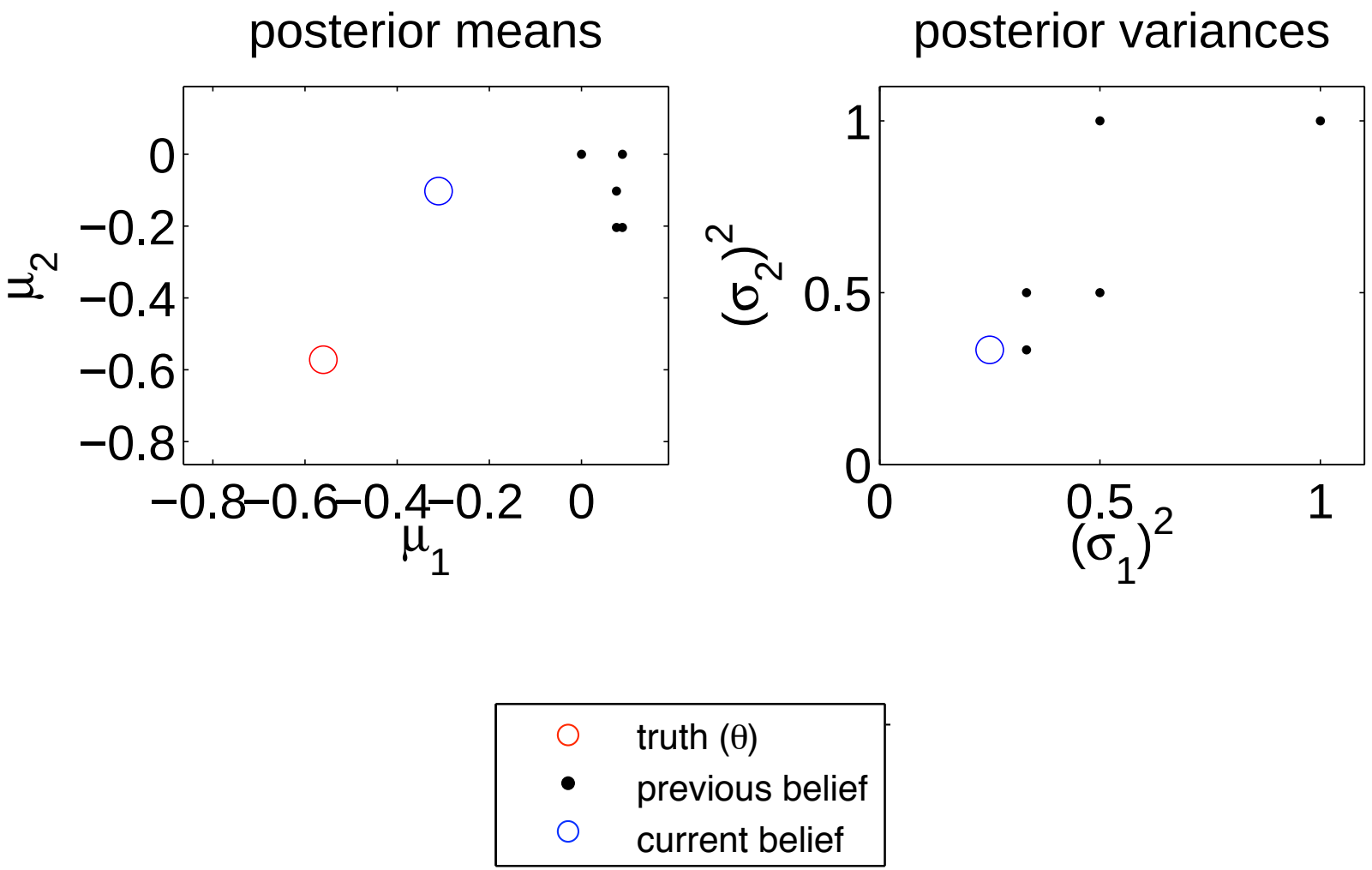
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

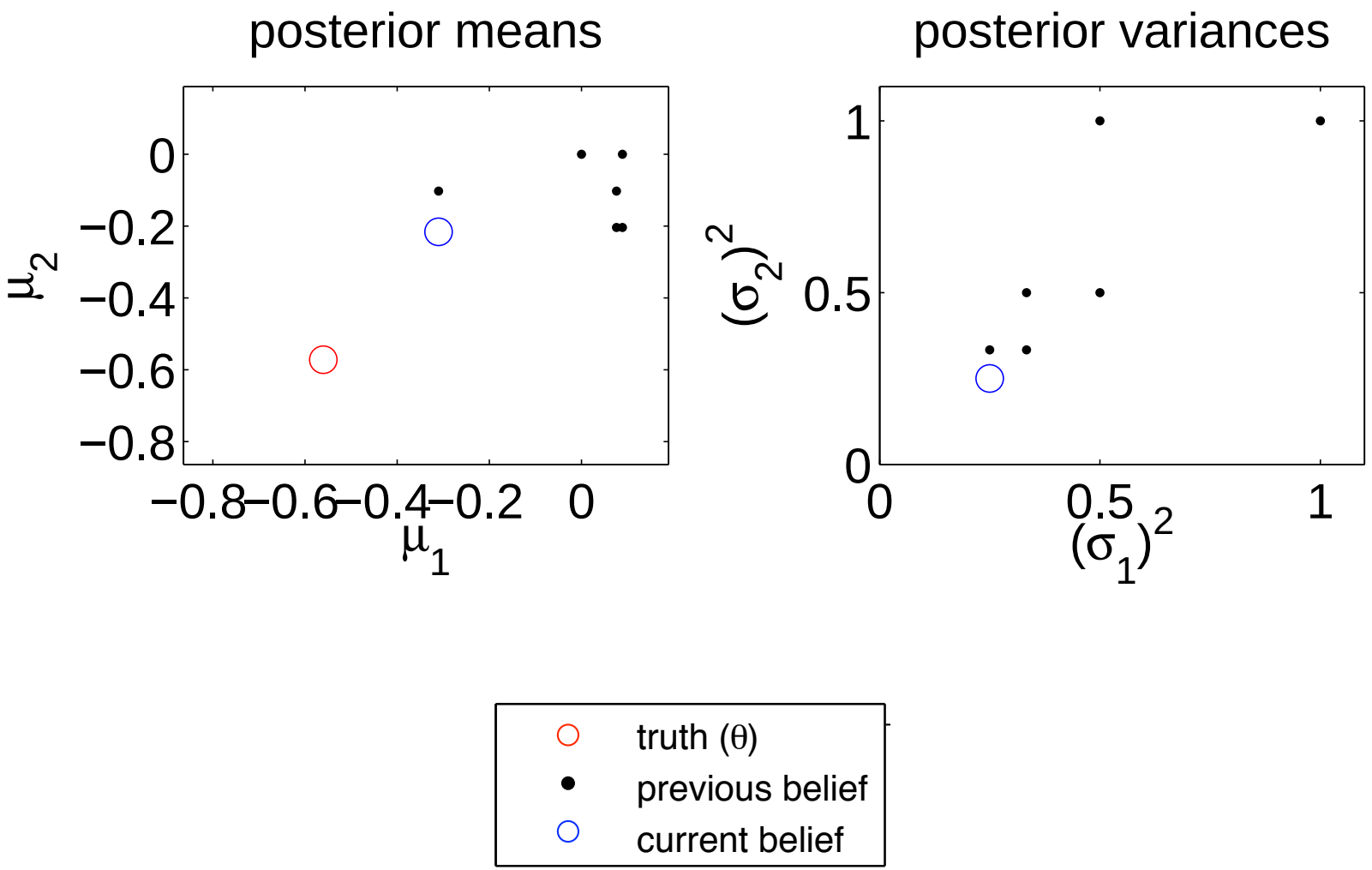
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

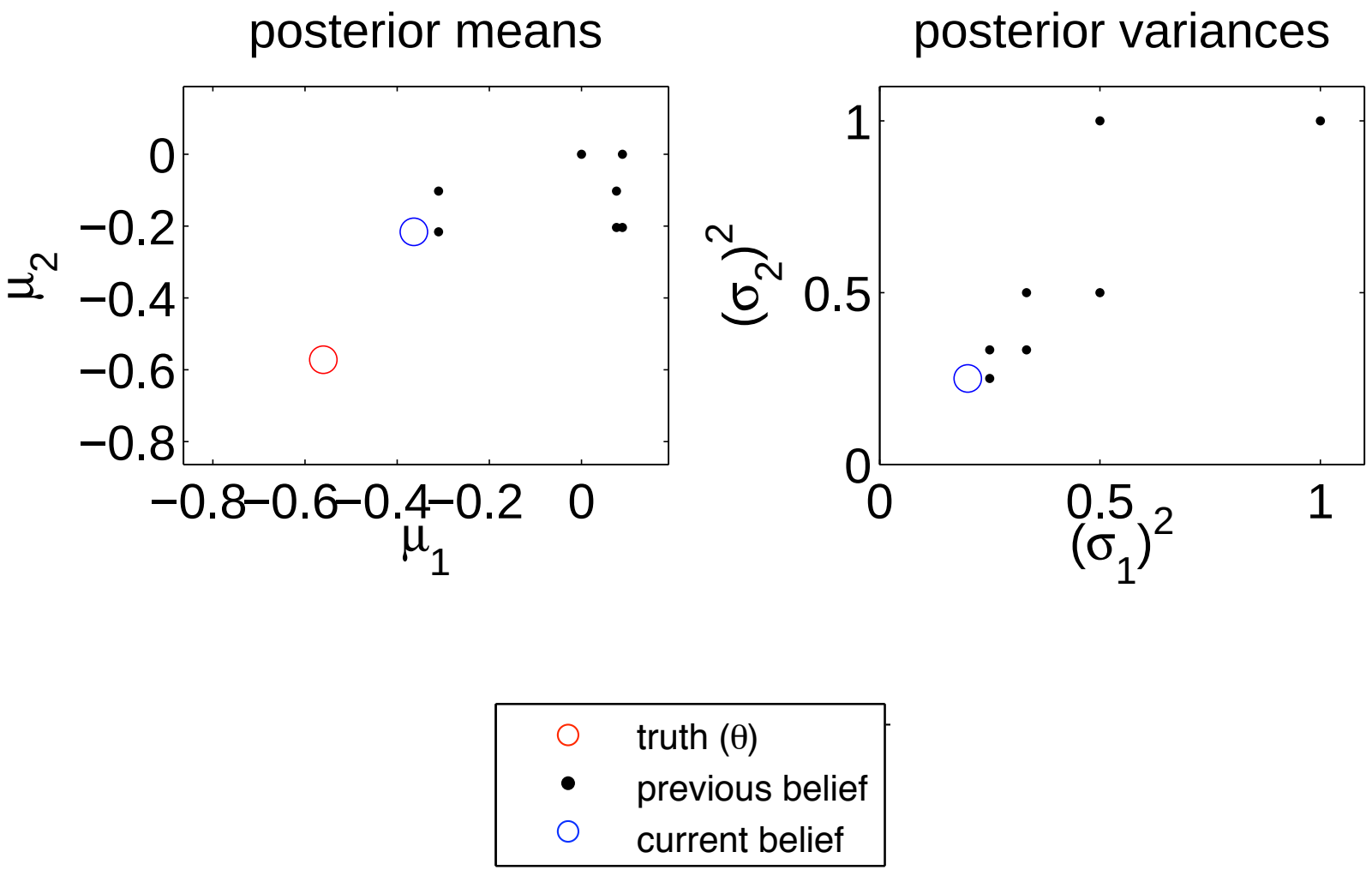
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8
−0.6
−0.4
−0.2
0
μ1
1
0.5
0
(σ2)2
0
0.5
1
(σ1)2
truth (θ)
previous belief
current belief

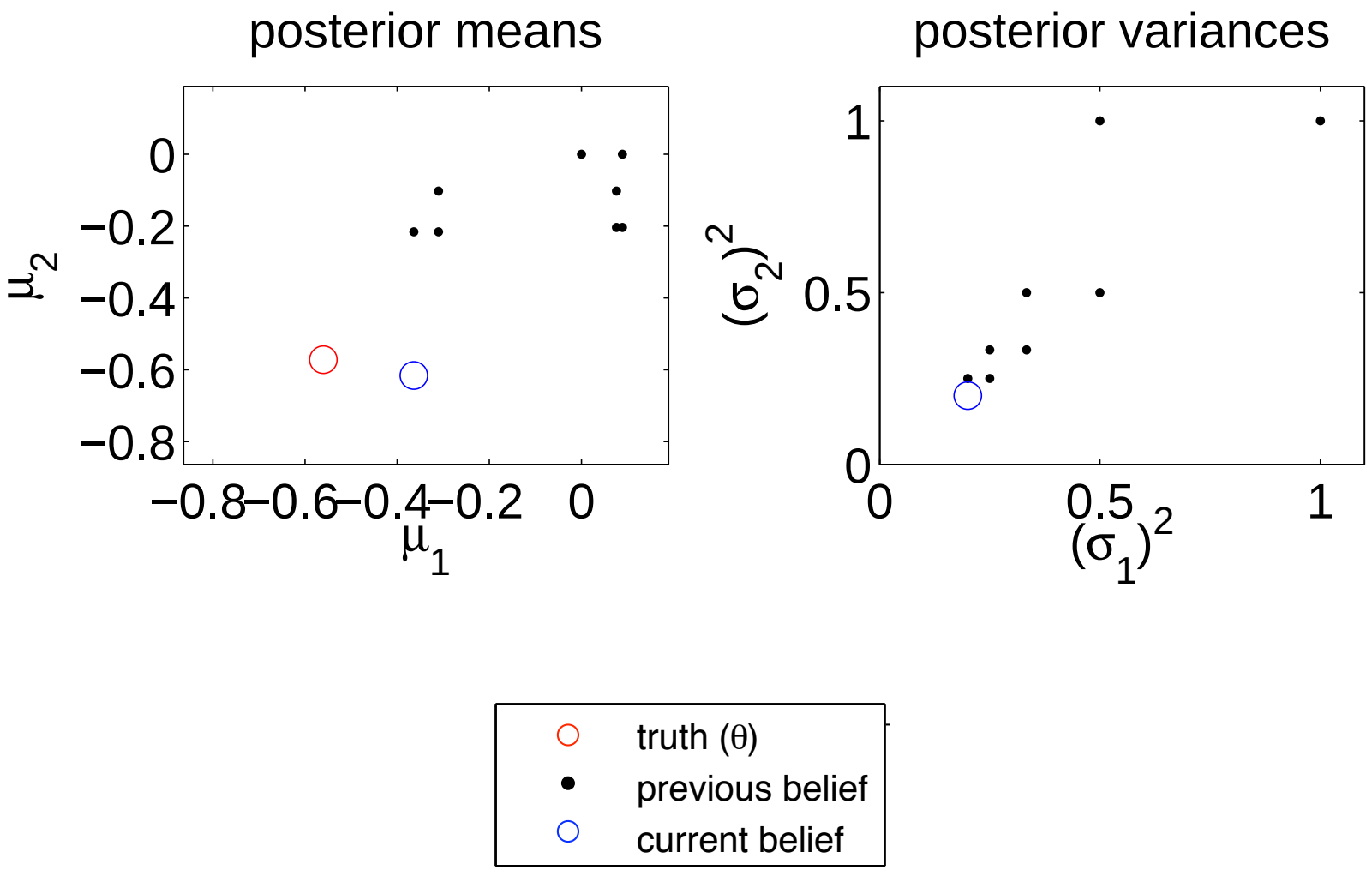
Simple Case


posterior means
posterior variances
μ_2
μ_1
$(\sigma_2)^2$
$(\sigma_1)^2$
0
−0.2
−0.4
−0.6
−0.8
−0.8 −0.6 −0.4 −0.2 0
0.5
1
truth (θ)
previous belief
current belief

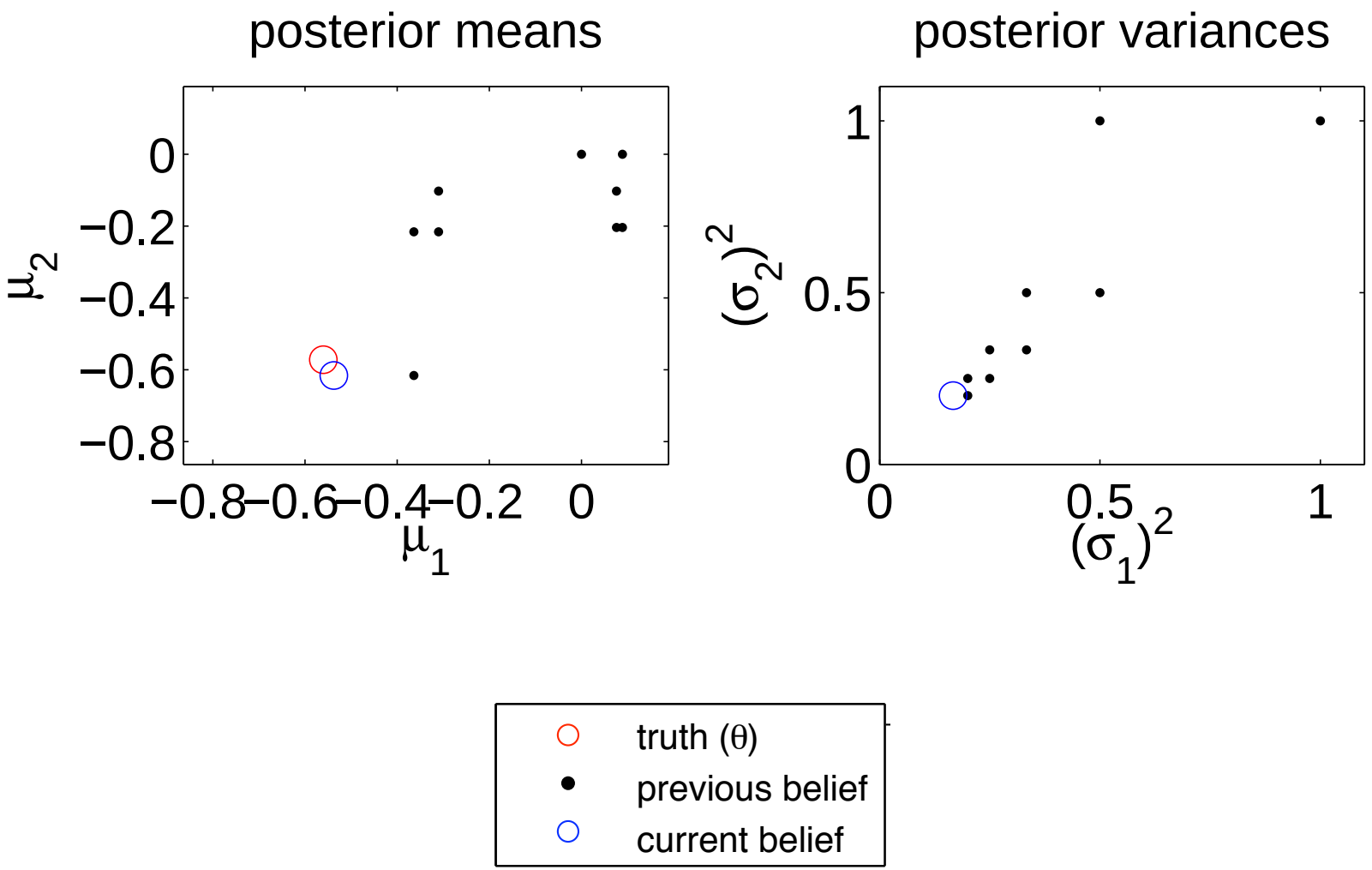
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

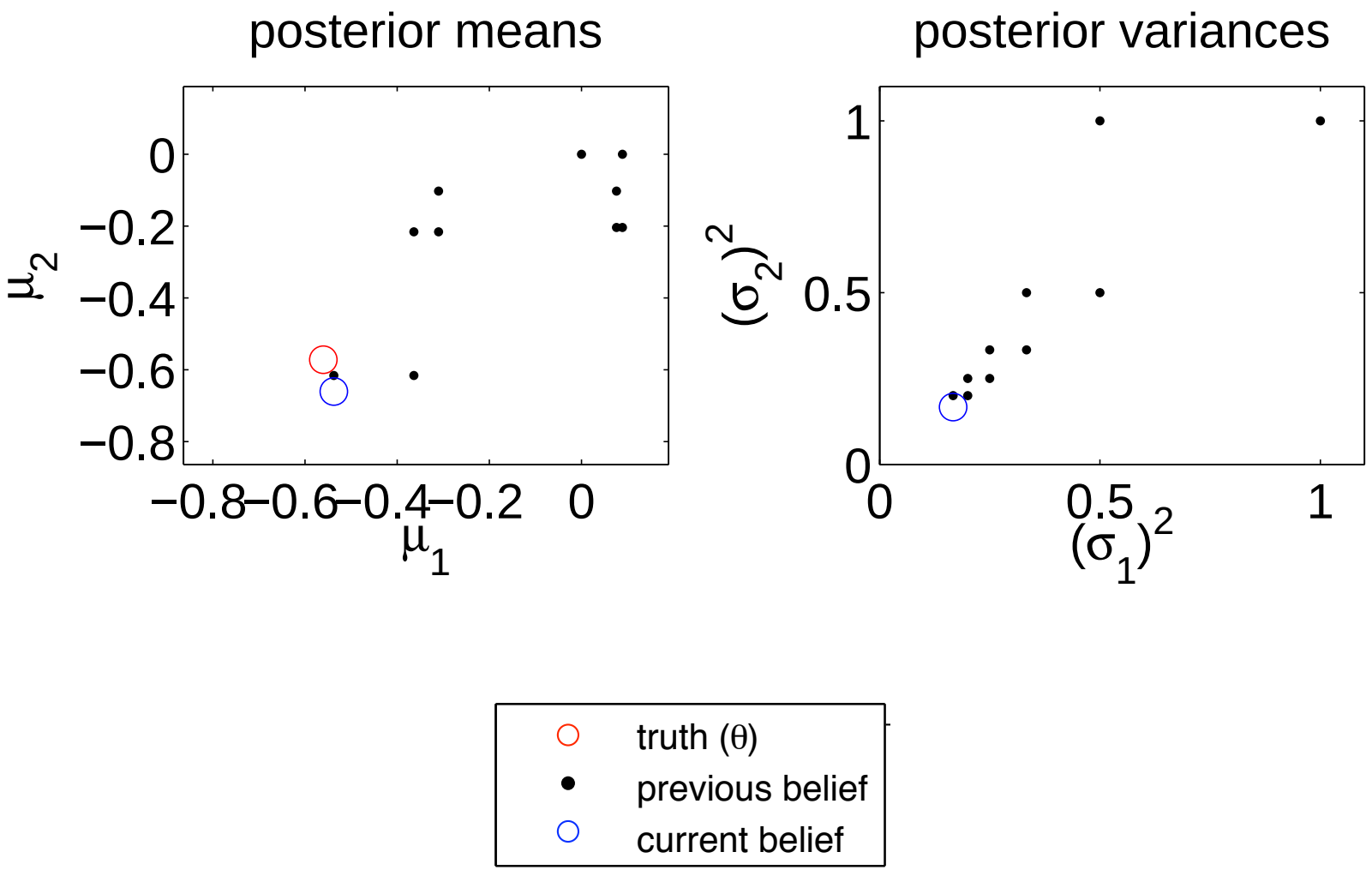
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

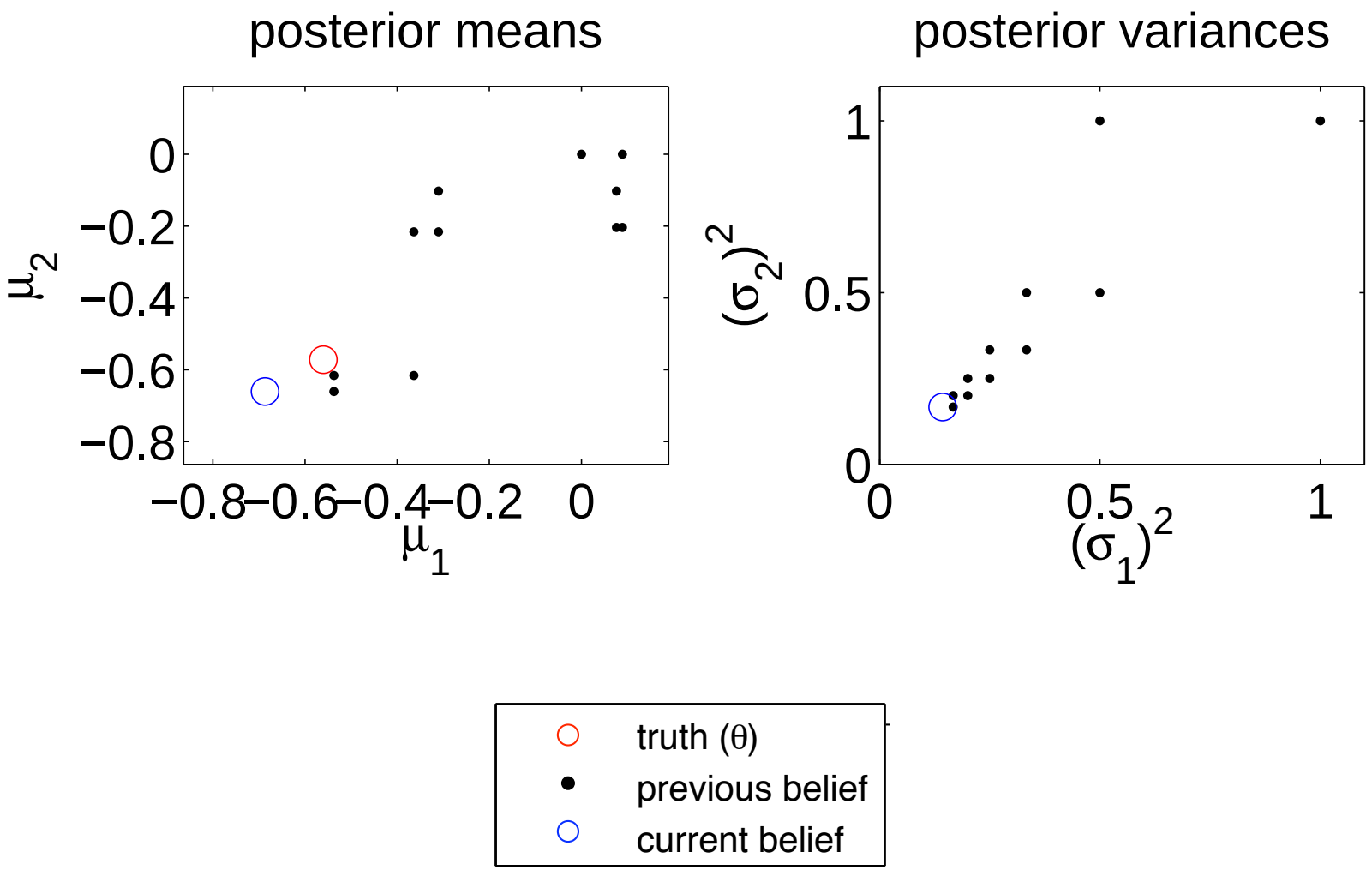
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

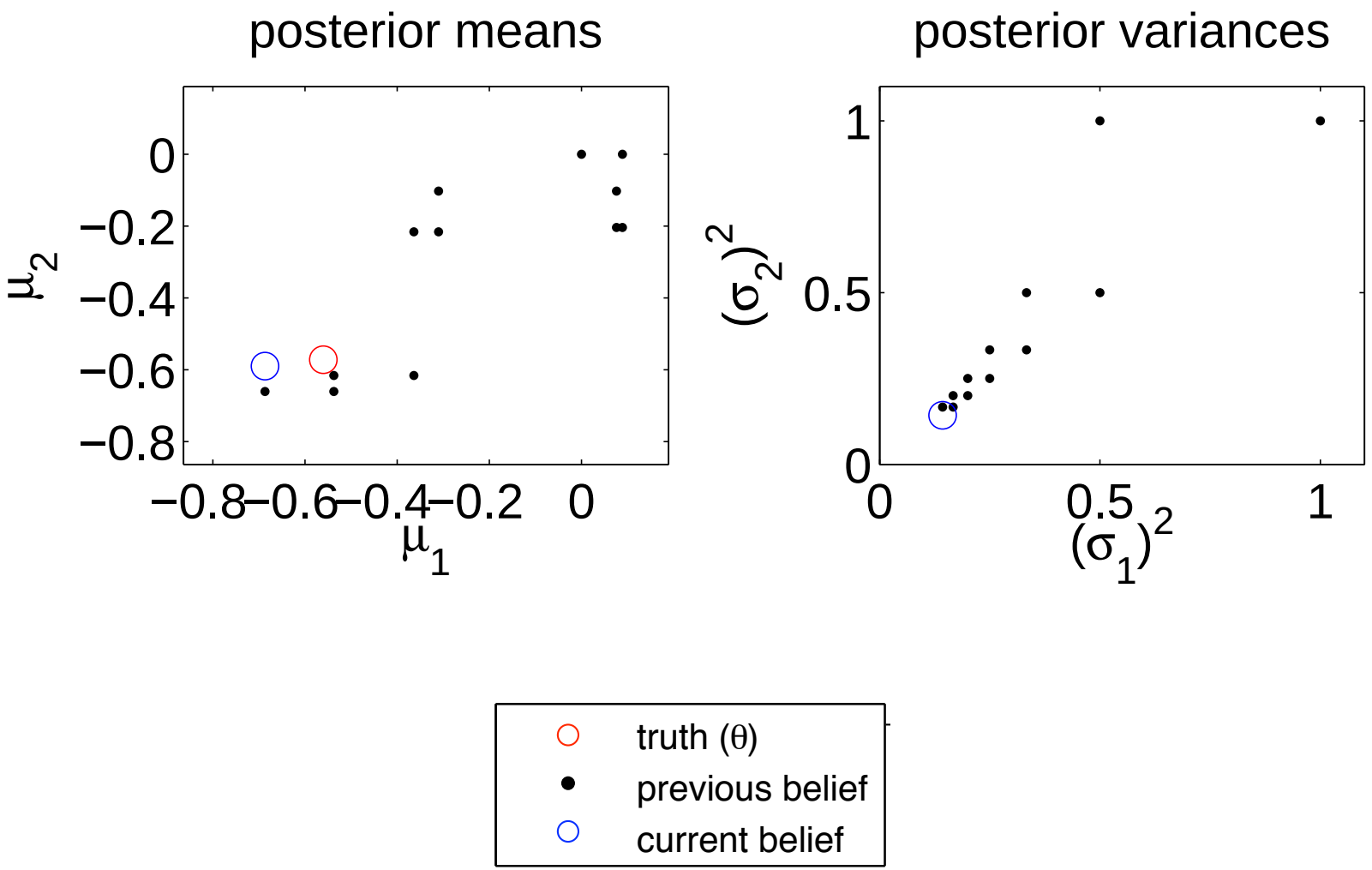
Simple Case


posterior means
posterior variances
μ_2
μ_1
$(\sigma_2)^2$
$(\sigma_1)^2$
0
−0.2
−0.4
−0.6
−0.8
0.5
1
truth (θ)
previous belief
current belief

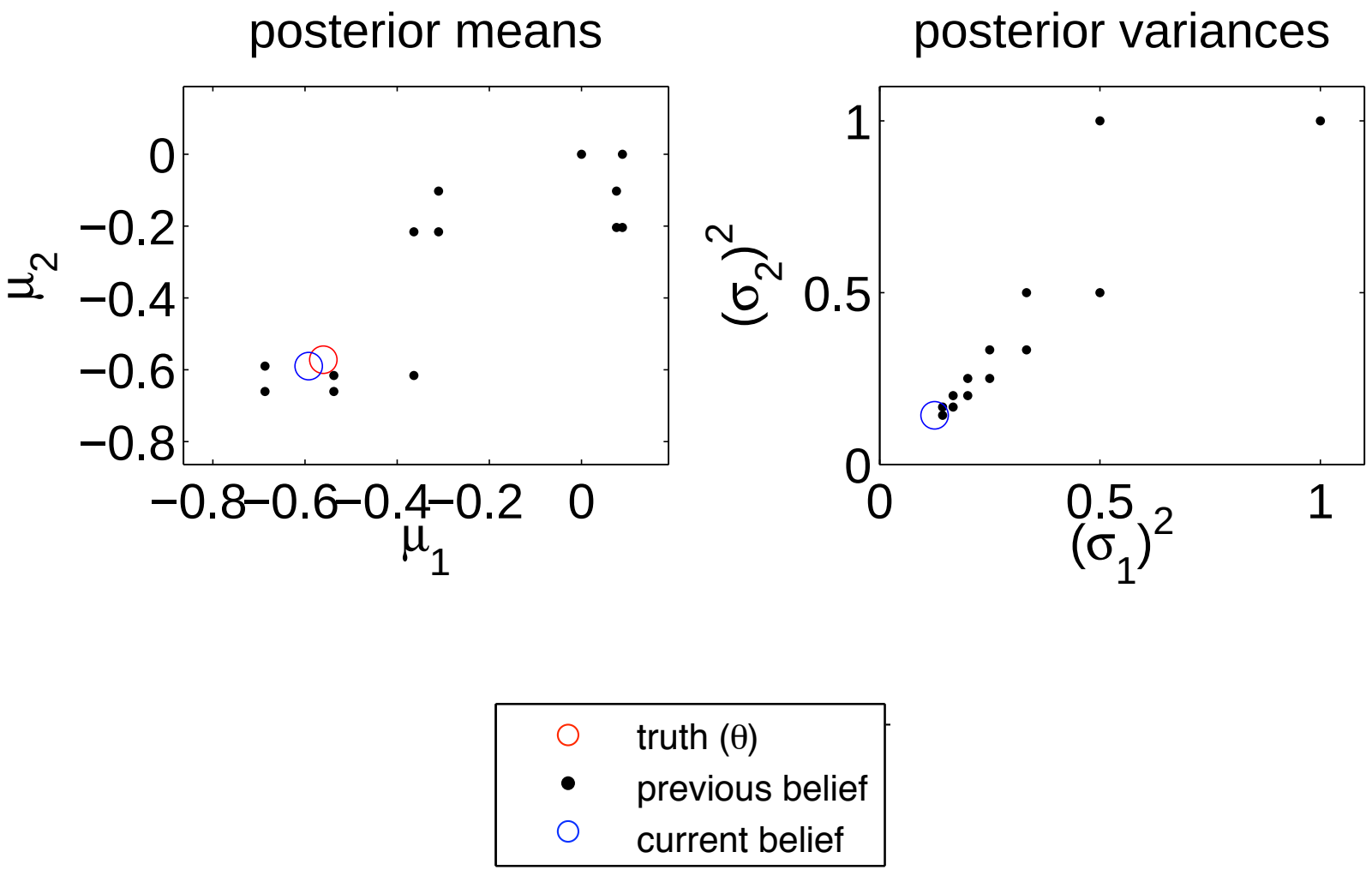
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

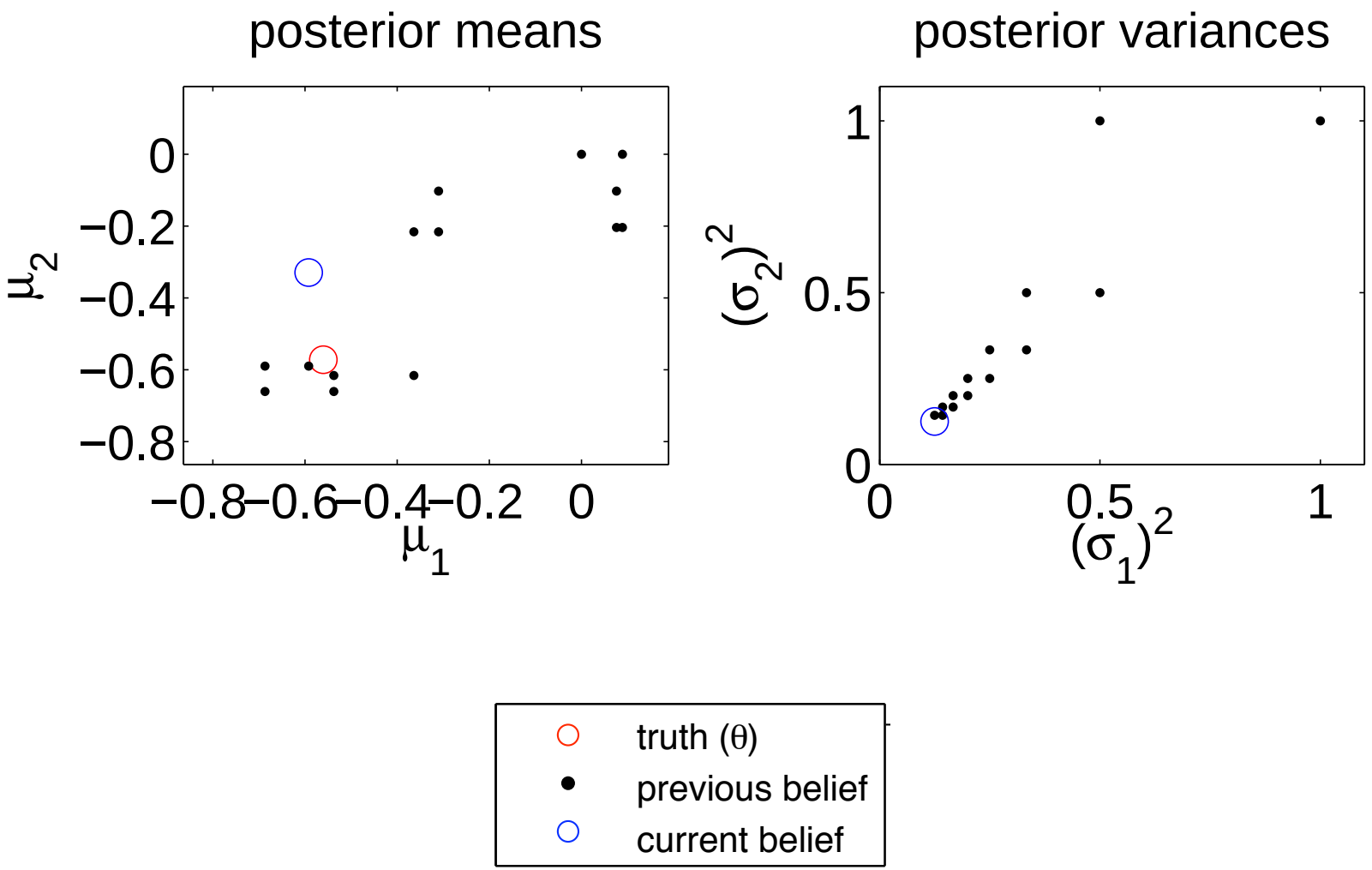
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

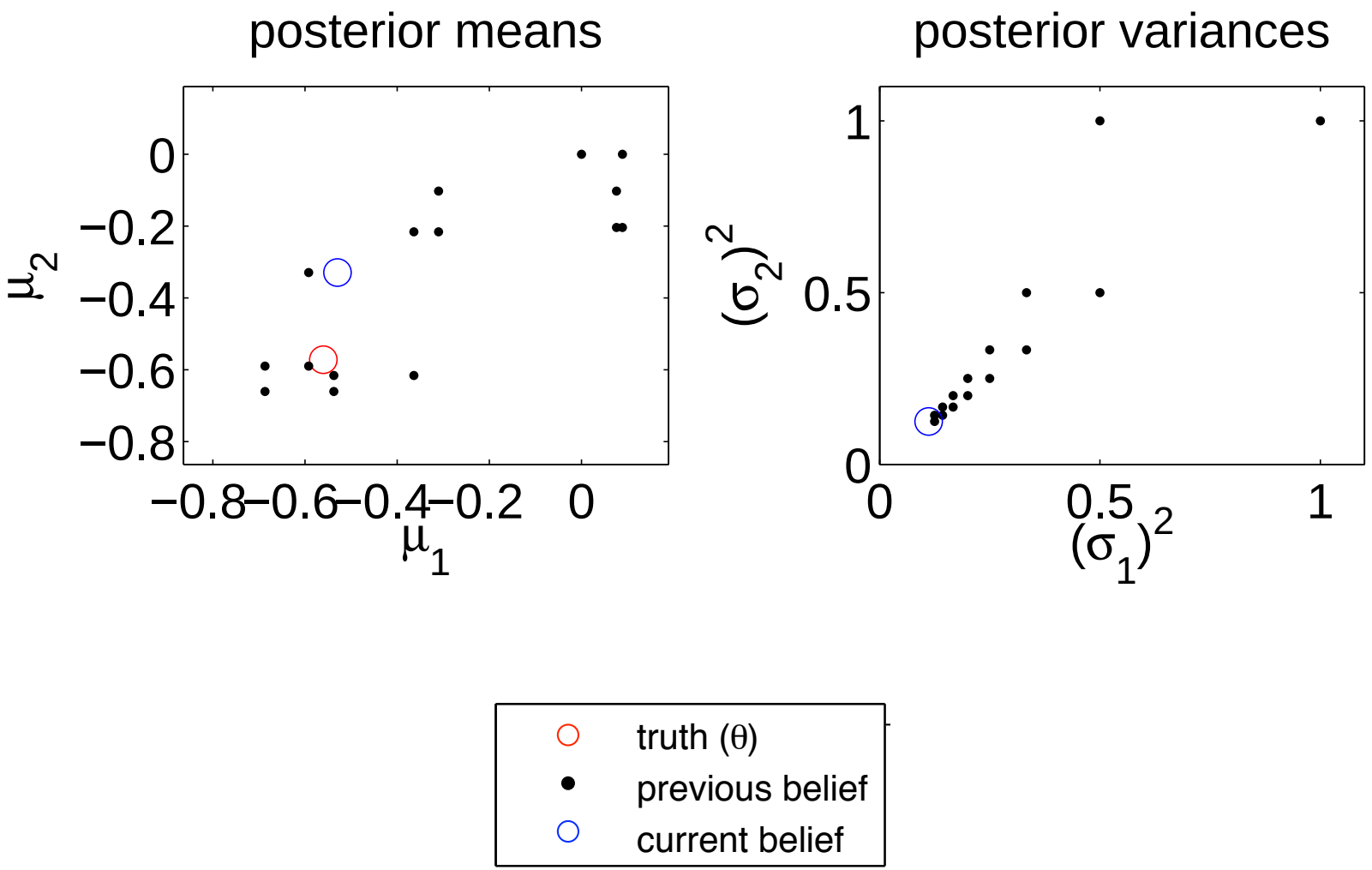
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

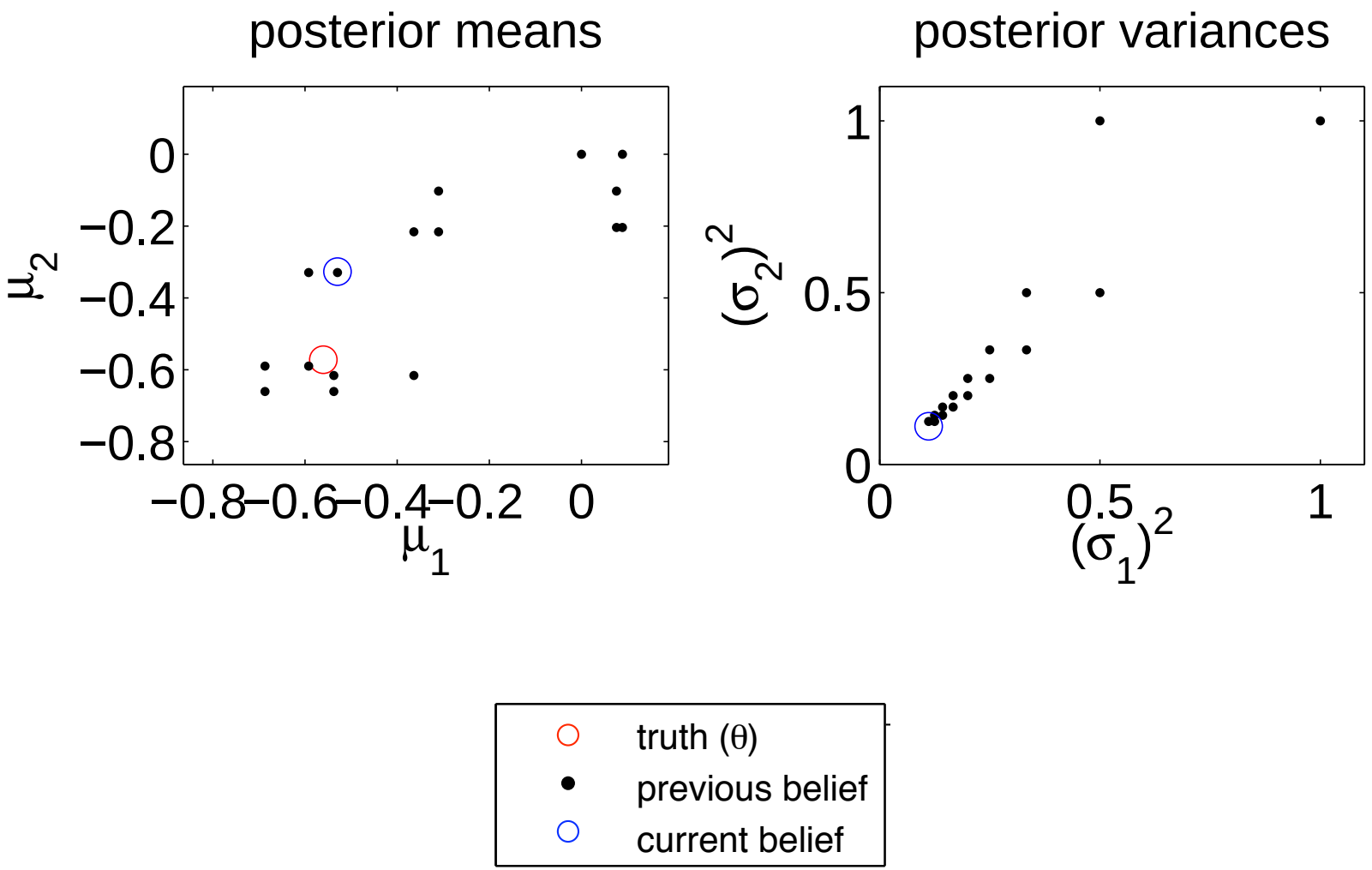
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

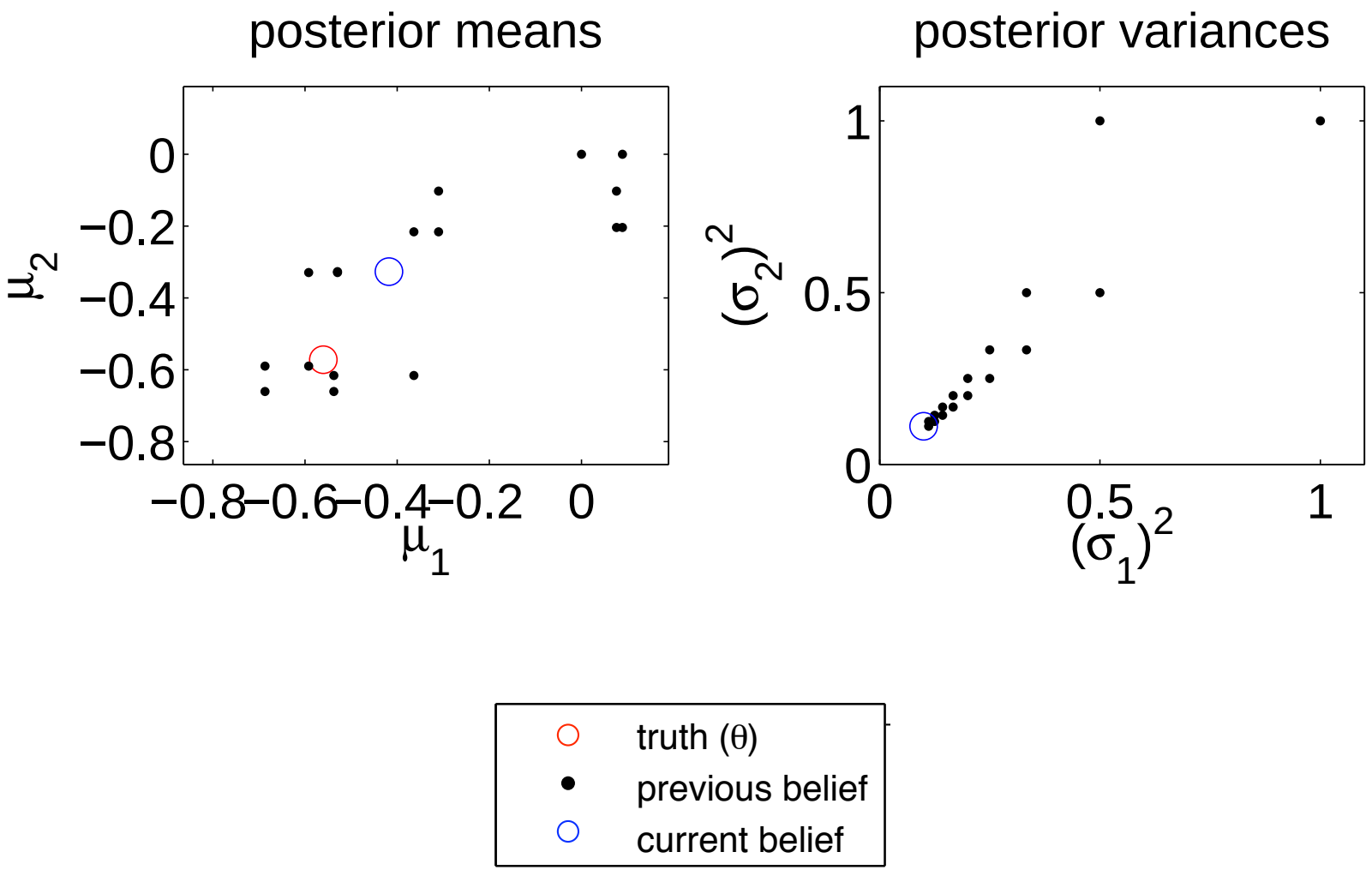
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

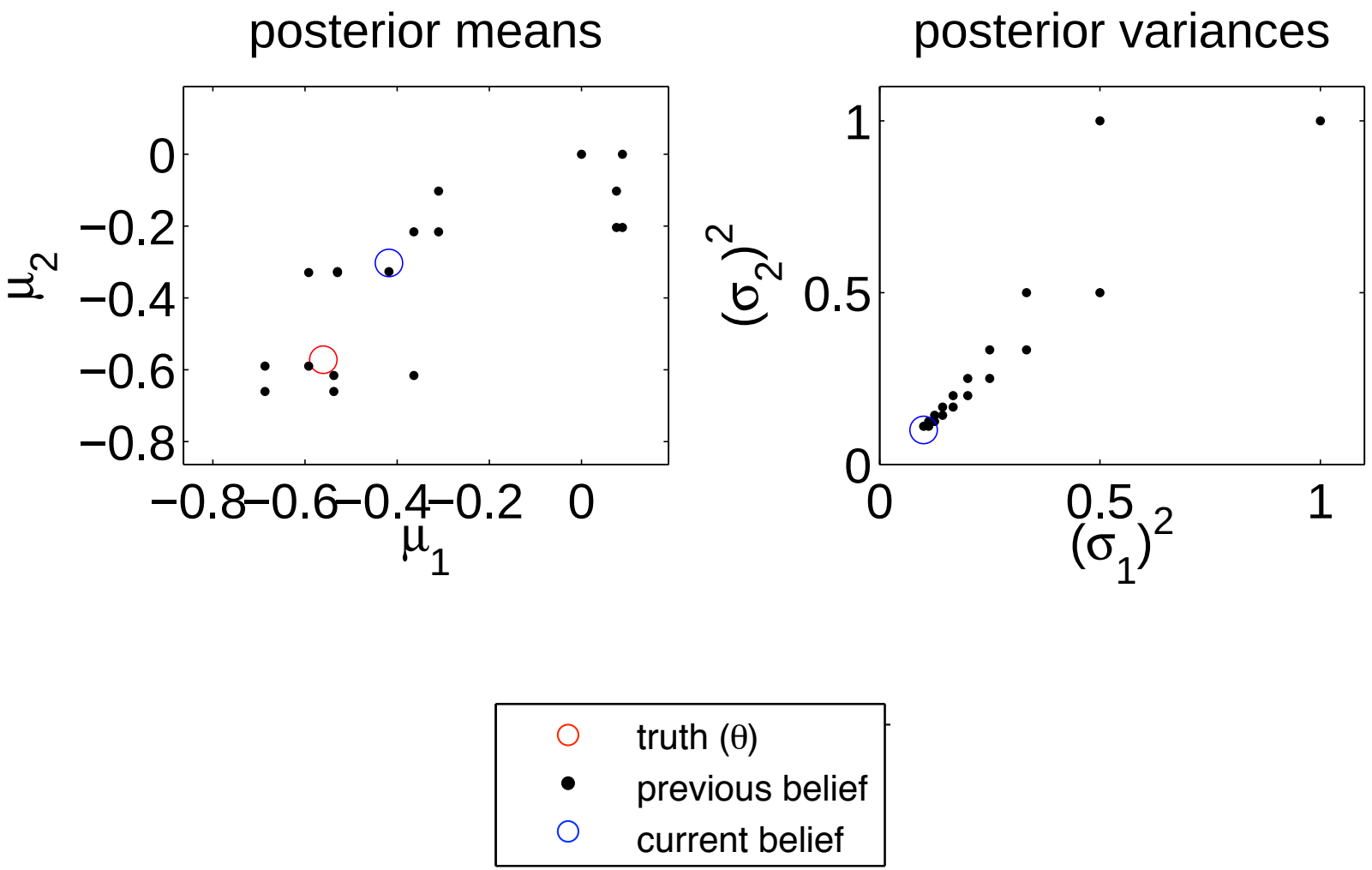
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

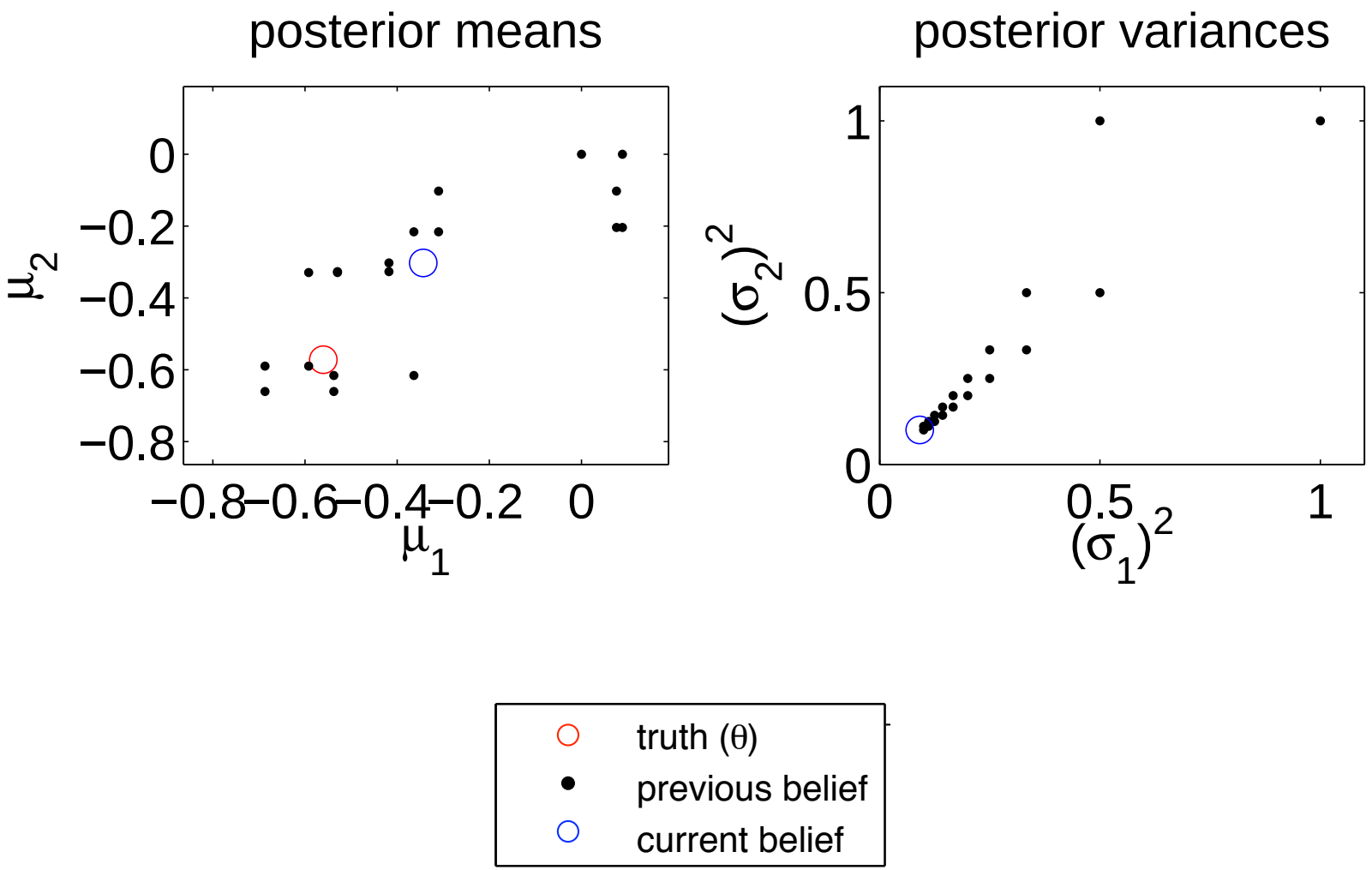
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

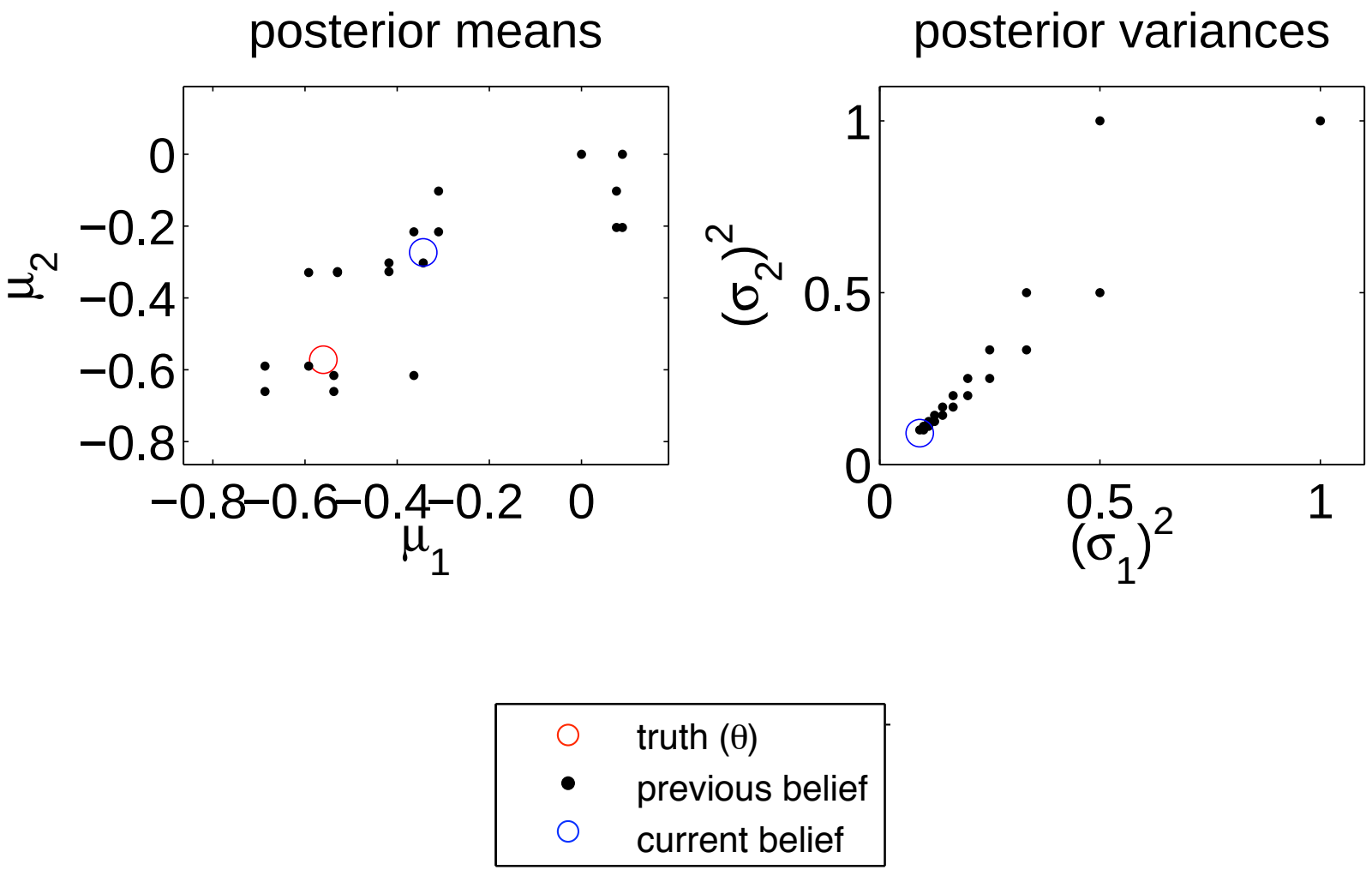
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

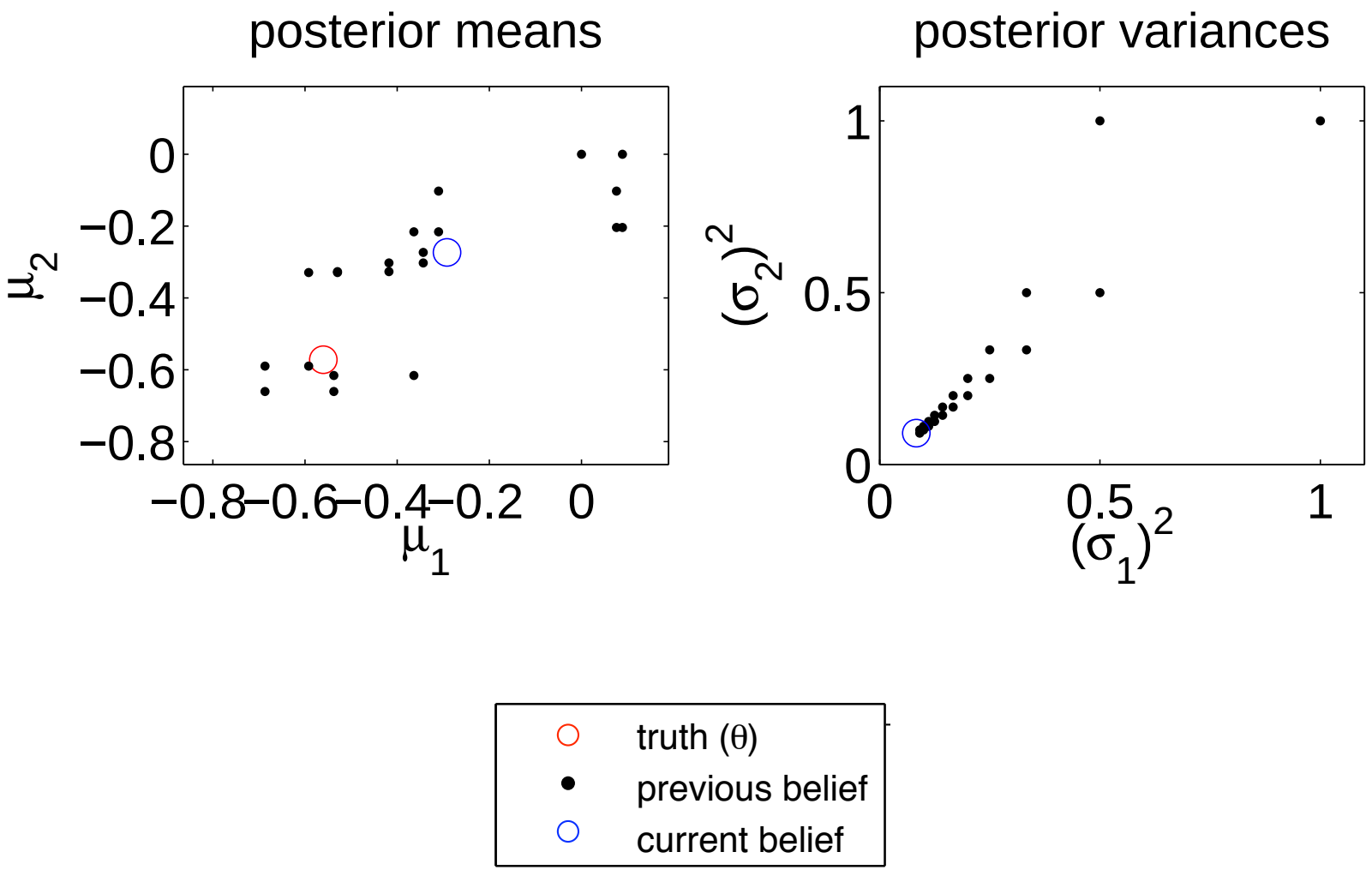
Simple Case


posterior means
posterior variances
μ_2
μ_1
$(\sigma_2)^2$
$(\sigma_1)^2$
0
−0.2
−0.4
−0.6
−0.8
0.5
1
truth (θ)
previous belief
current belief

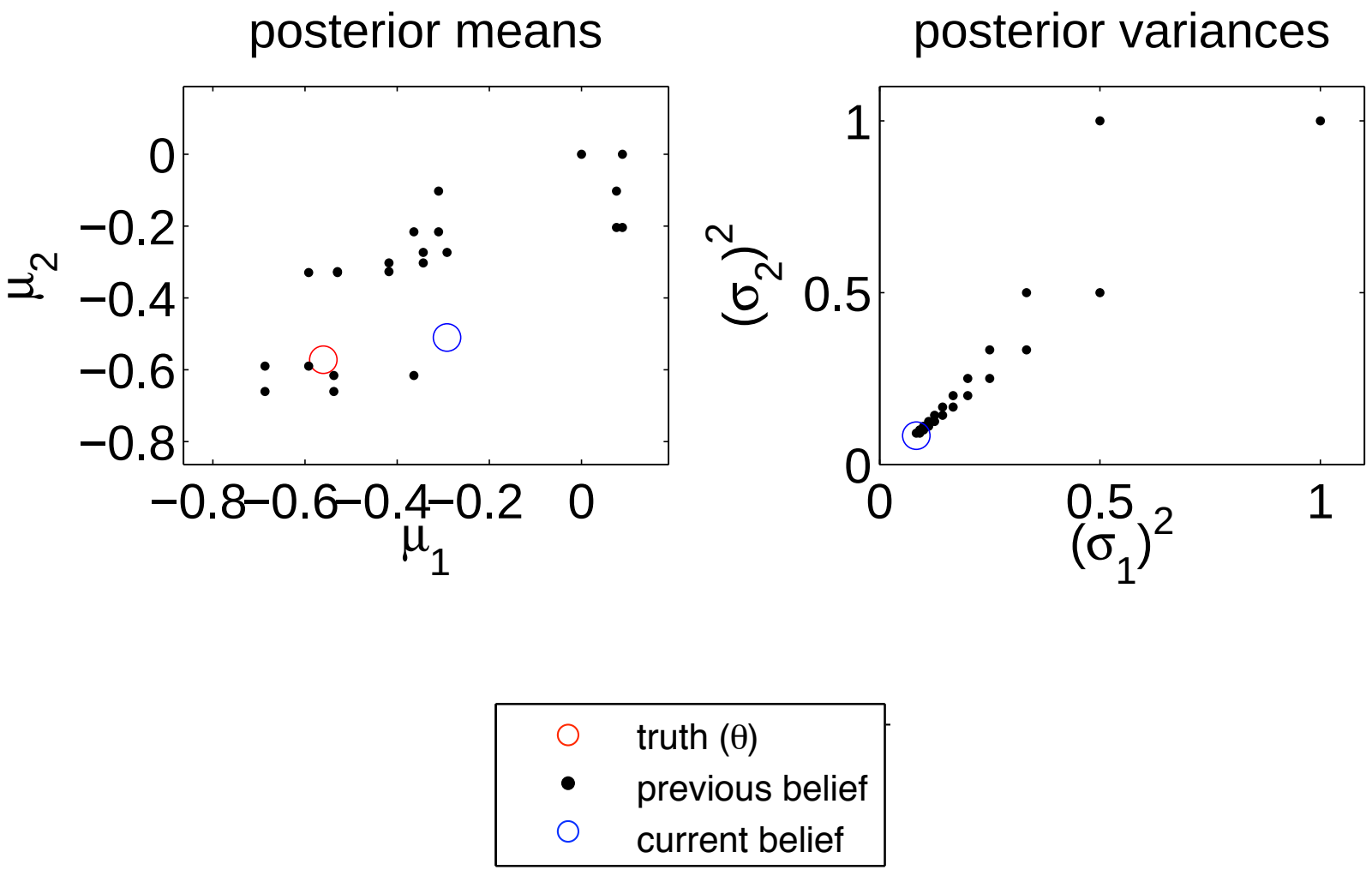
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

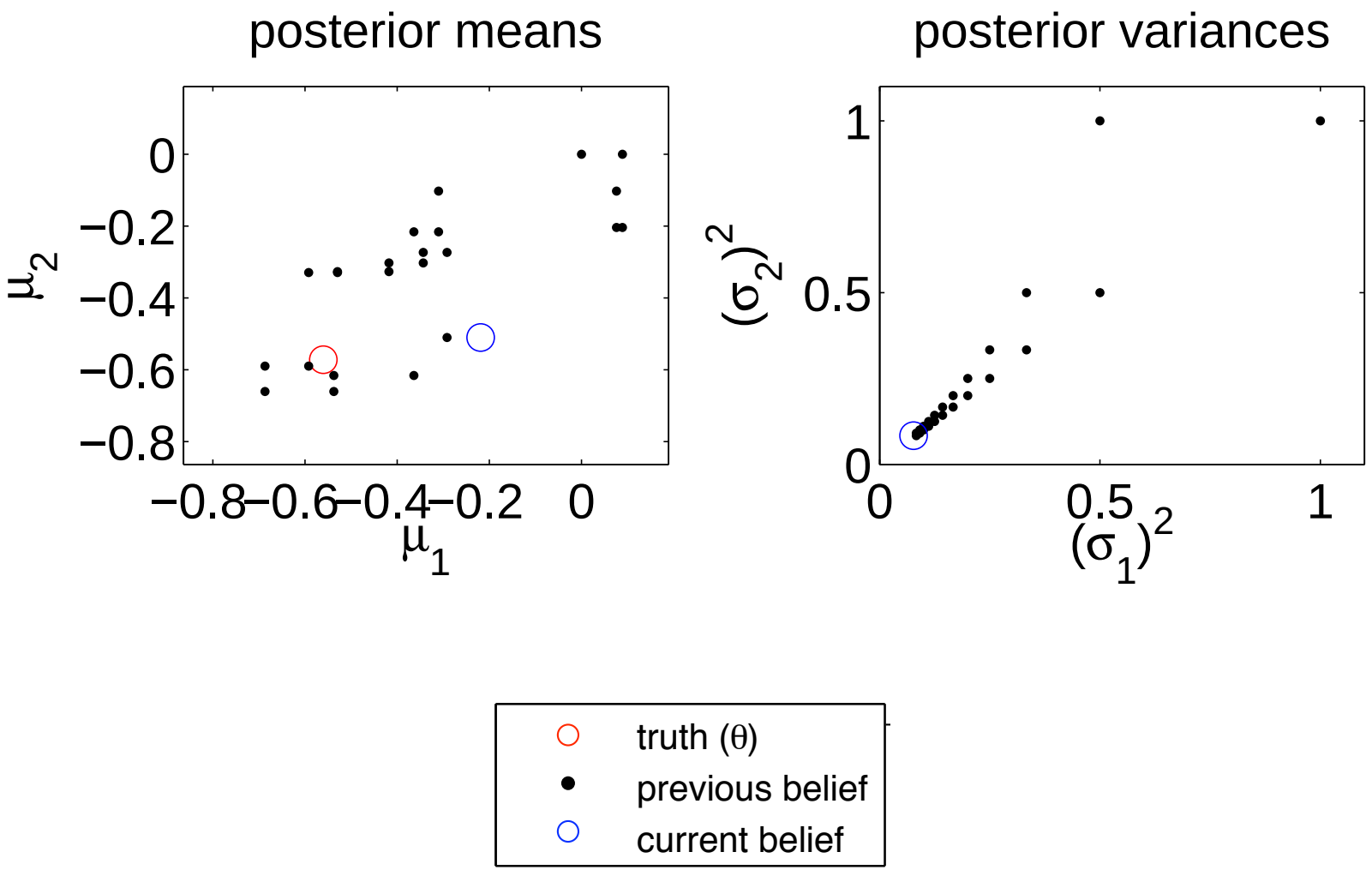
Simple Case


posterior means
posterior variances
μ_2
μ_1
$(\sigma_2)^2$
$(\sigma_1)^2$
0
−0.2
−0.4
−0.6
−0.8
0.5
1
truth (θ)
previous belief
current belief

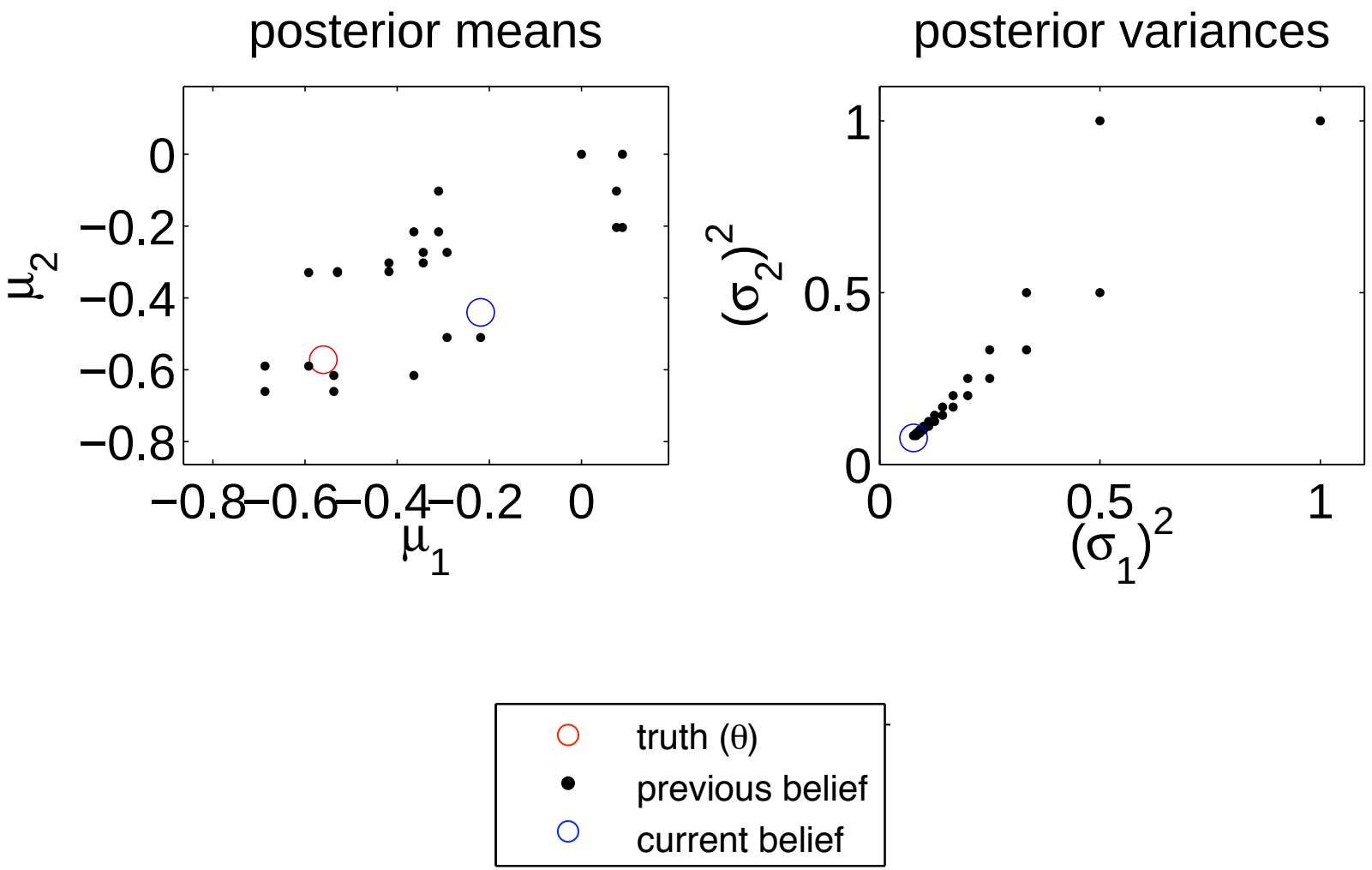
Simple Case


posterior means
posterior variances
μ_2
μ_1
$(\sigma_2)^2$
$(\sigma_1)^2$
0
−0.2
−0.4
−0.6
−0.8
0.5
1
truth (θ)
previous belief
current belief

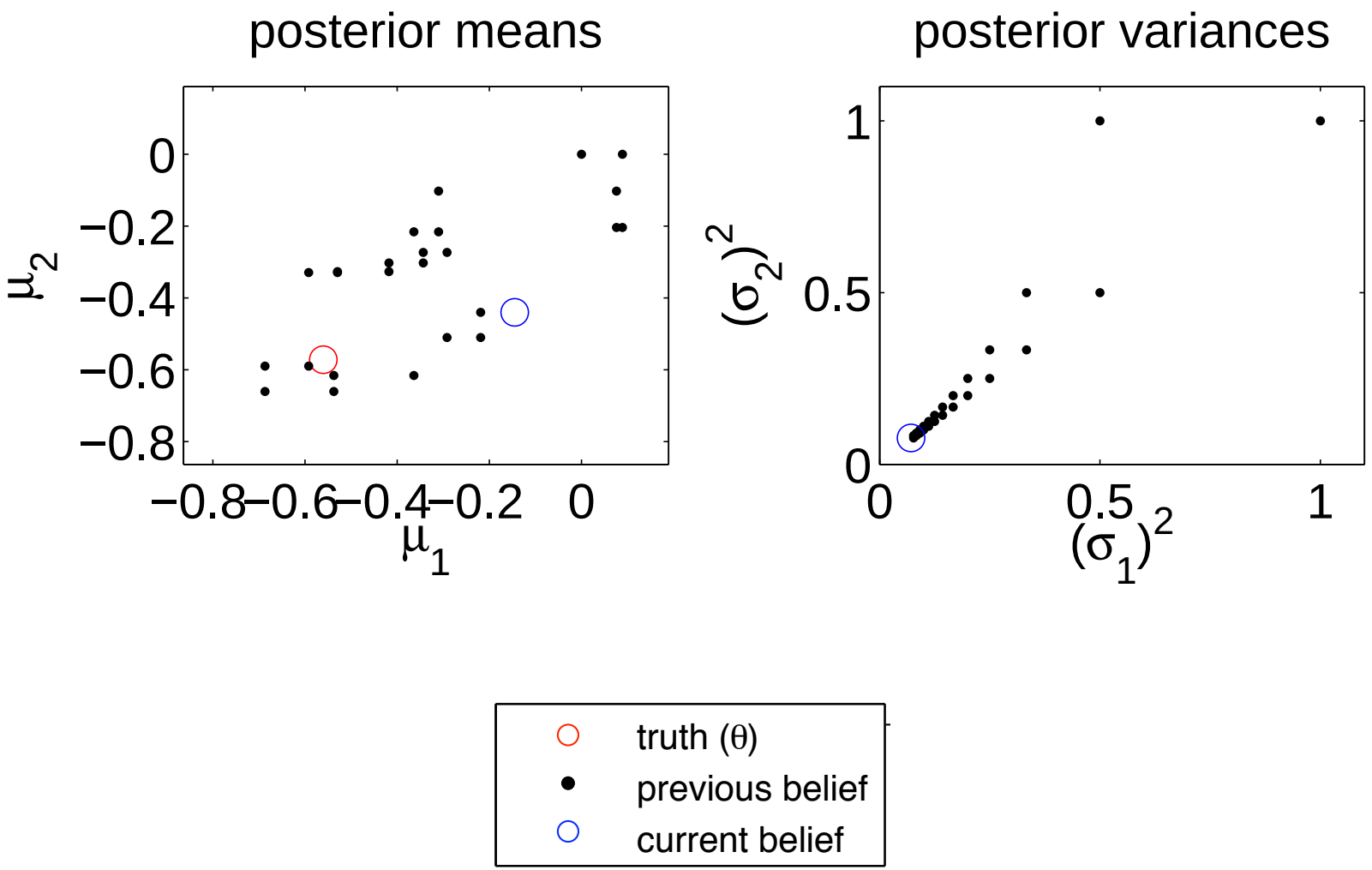
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

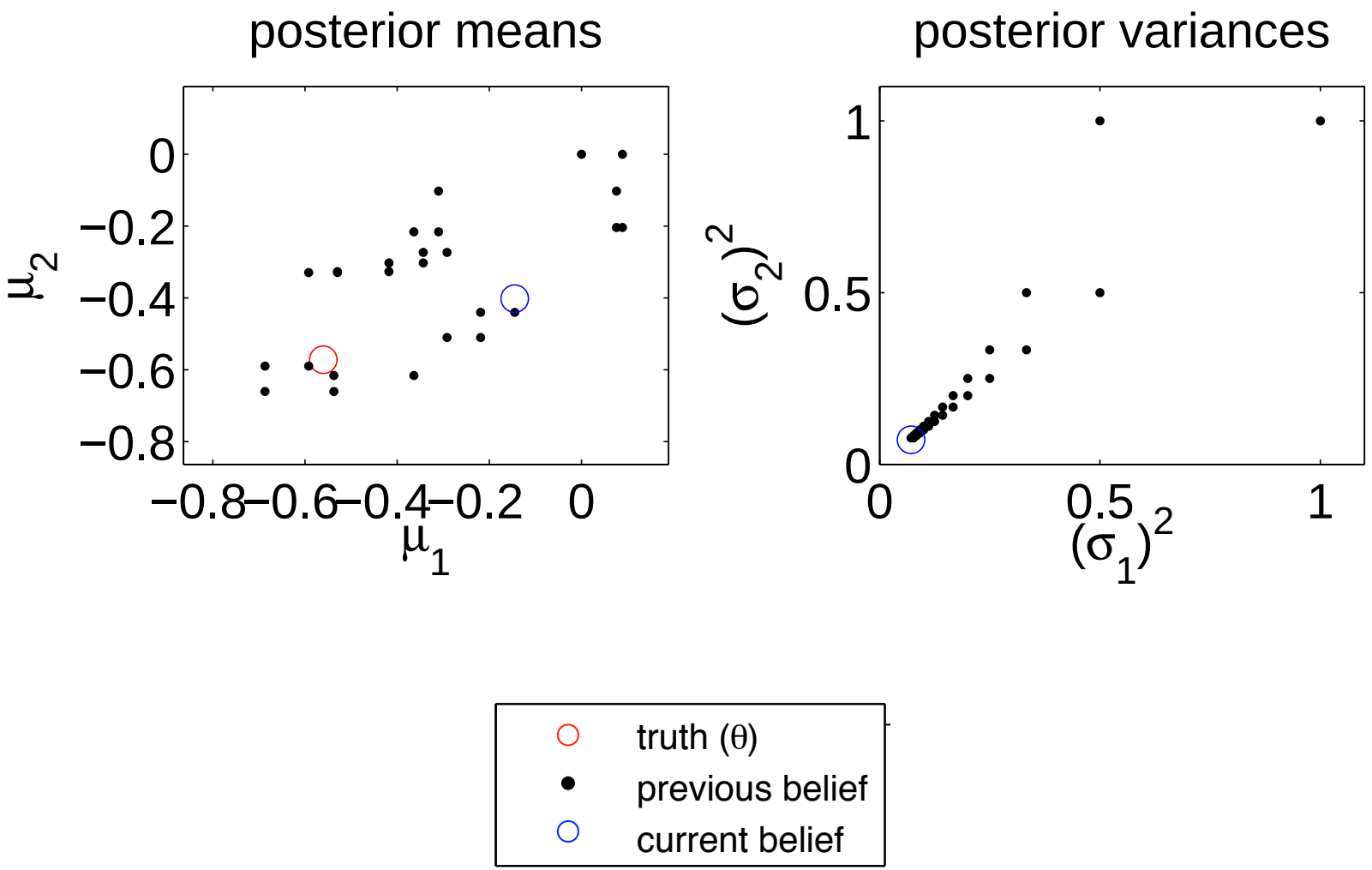
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

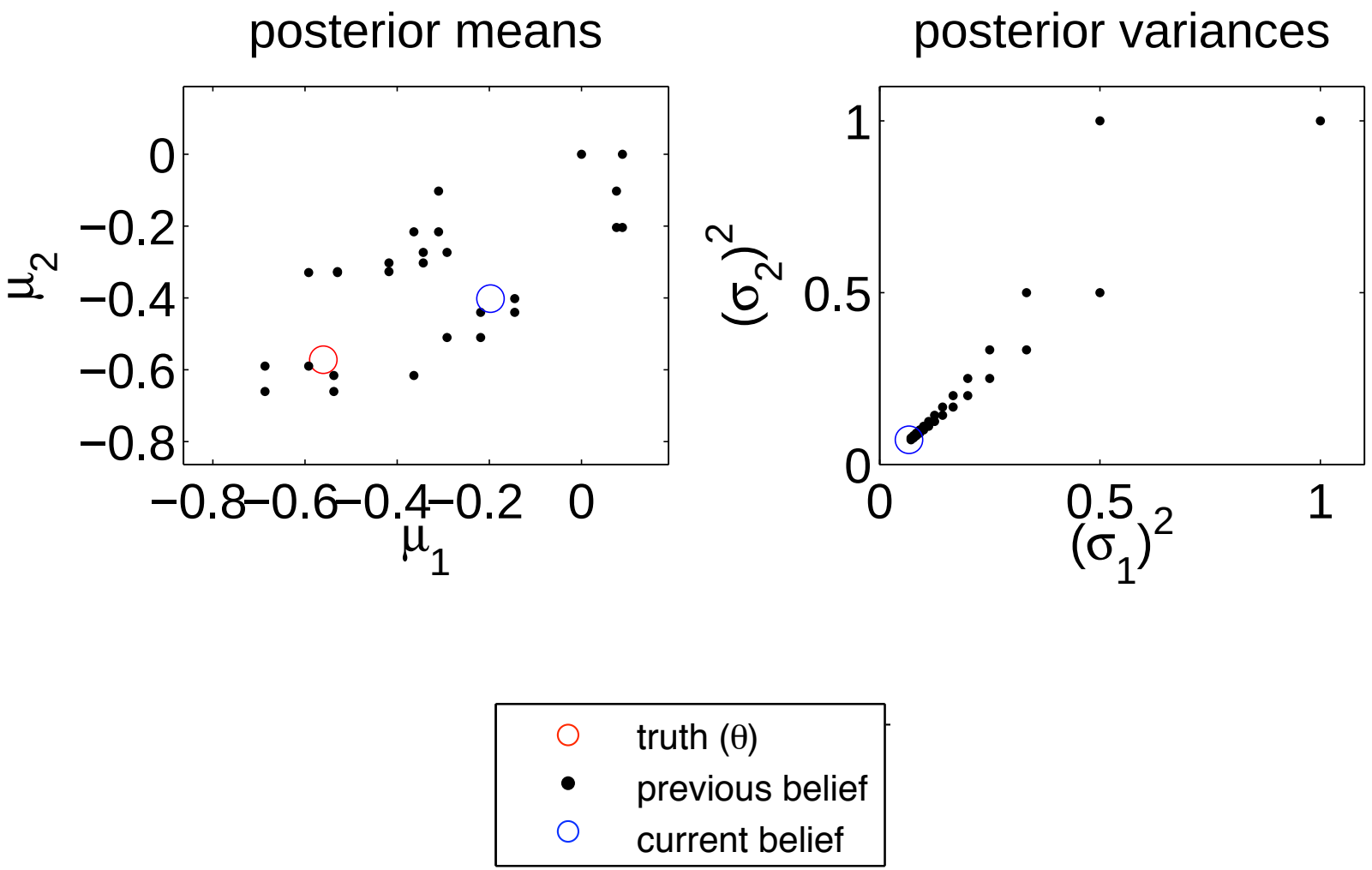
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

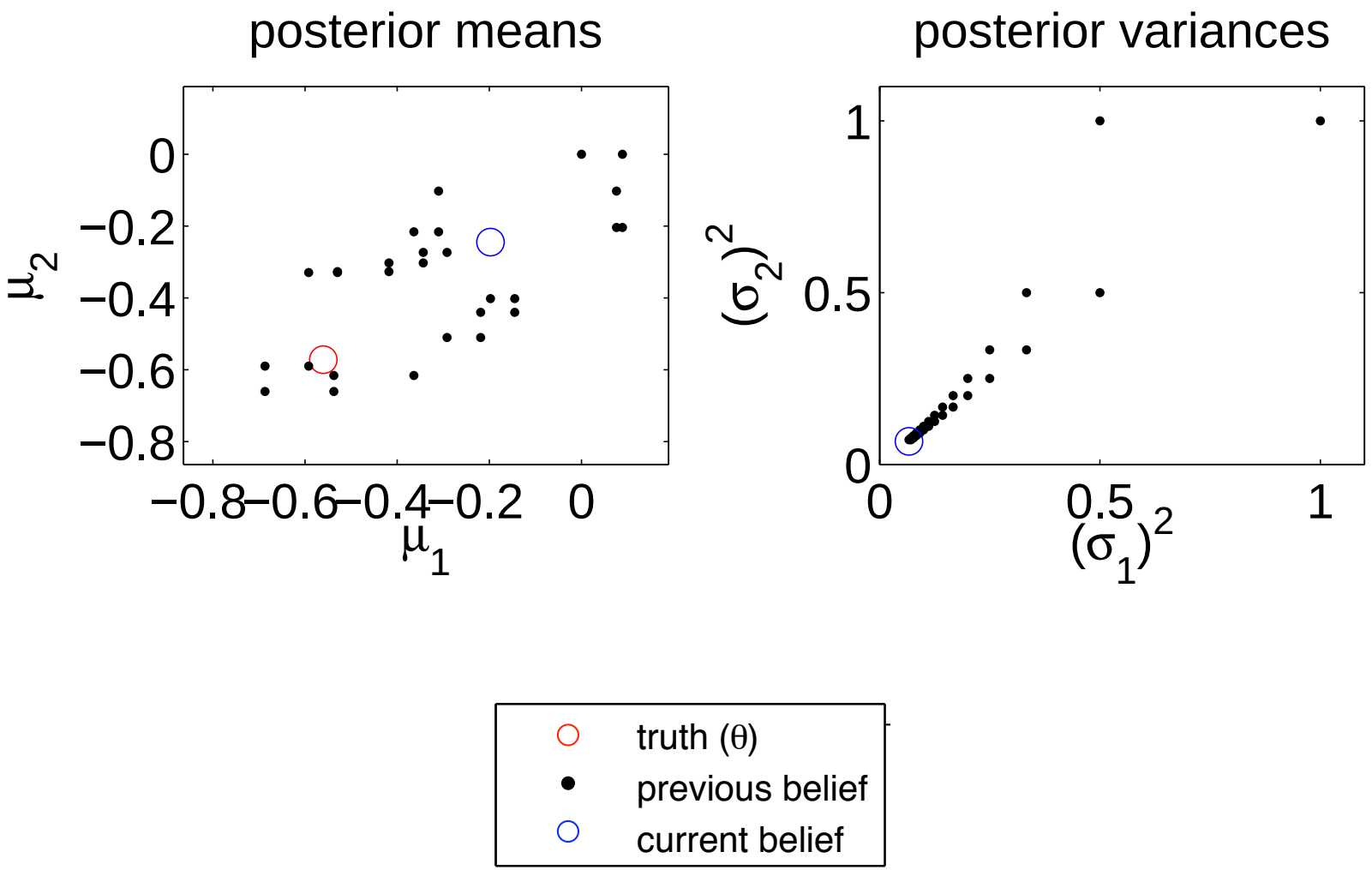
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

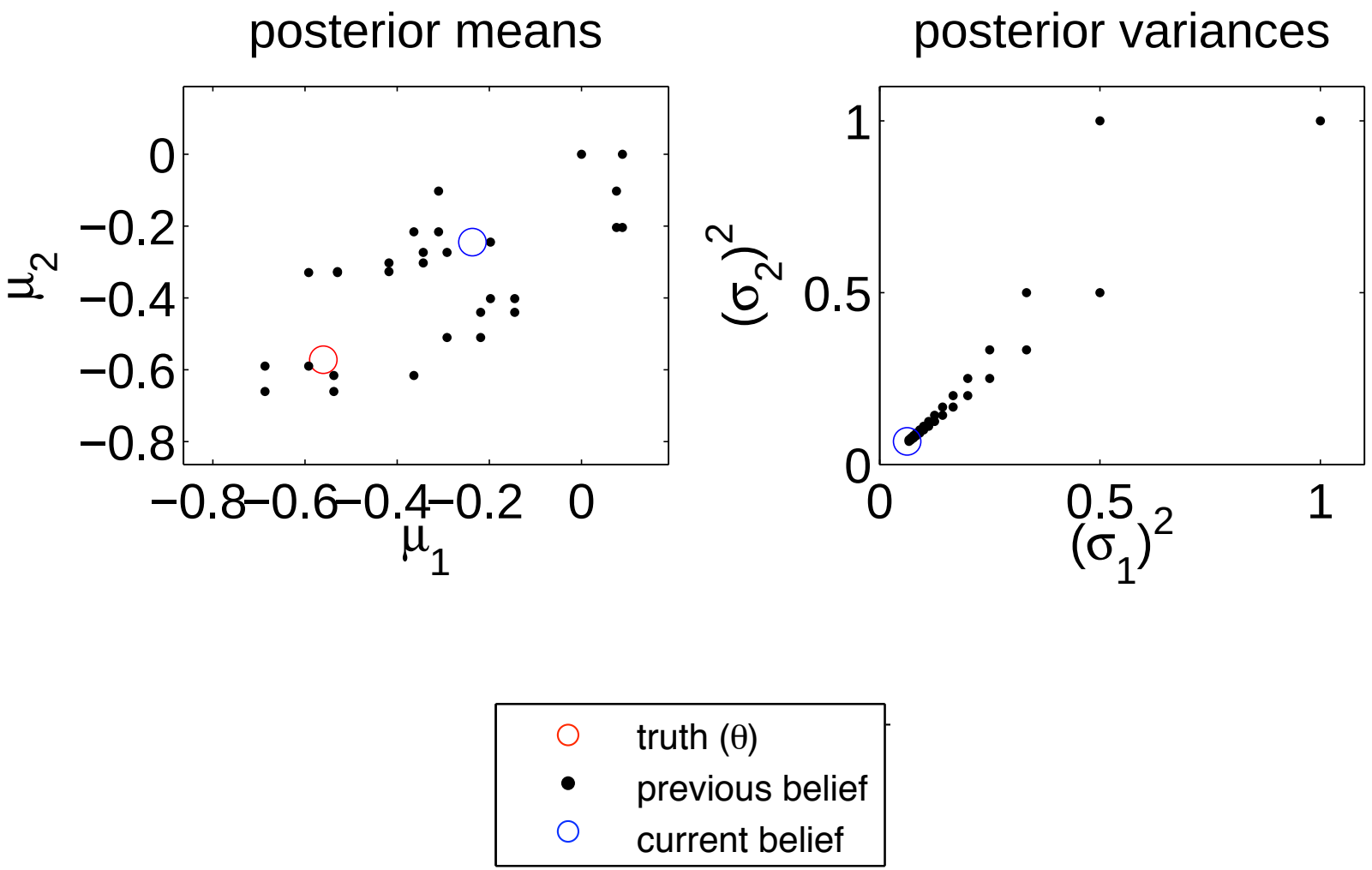
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

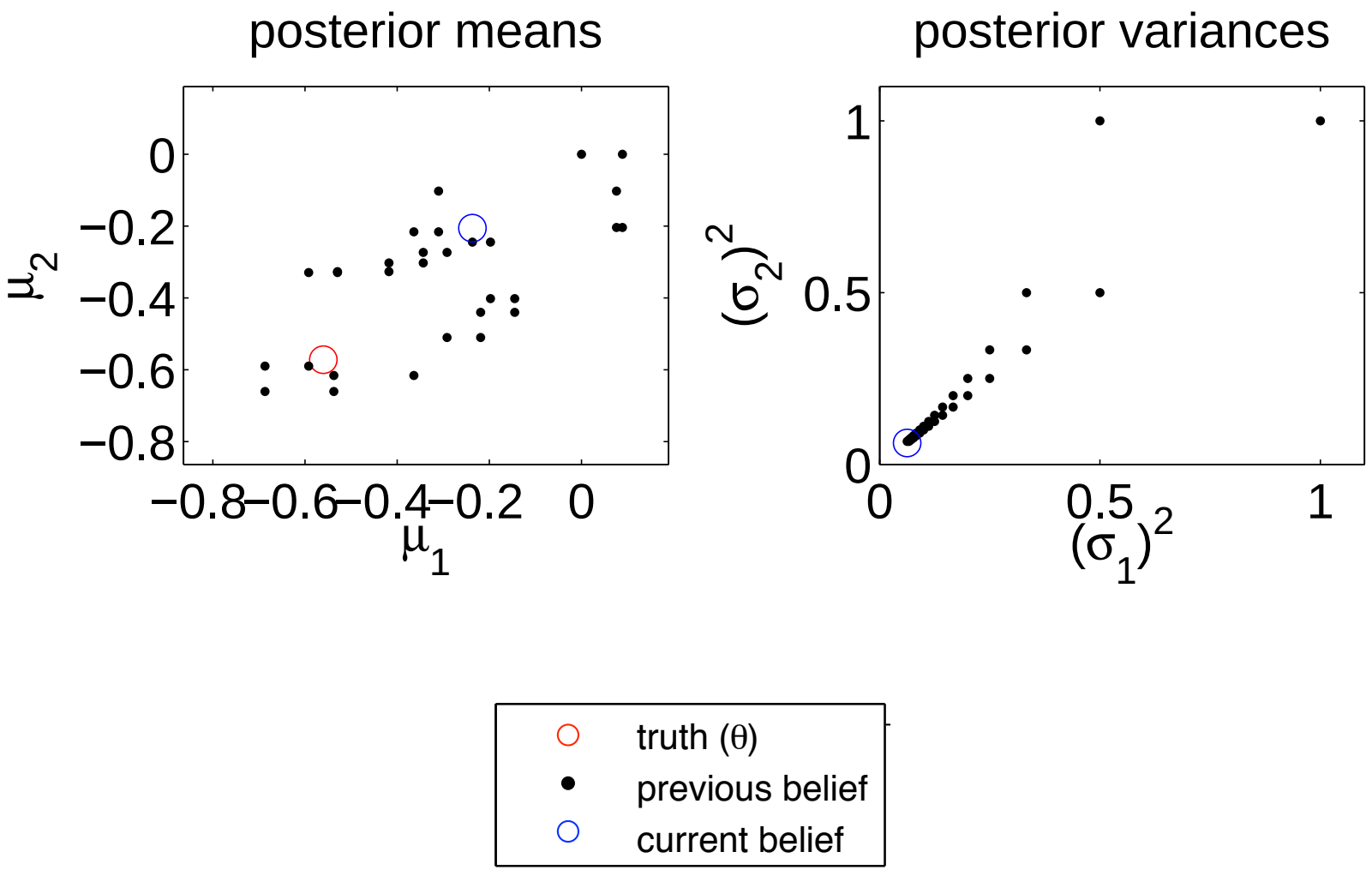
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

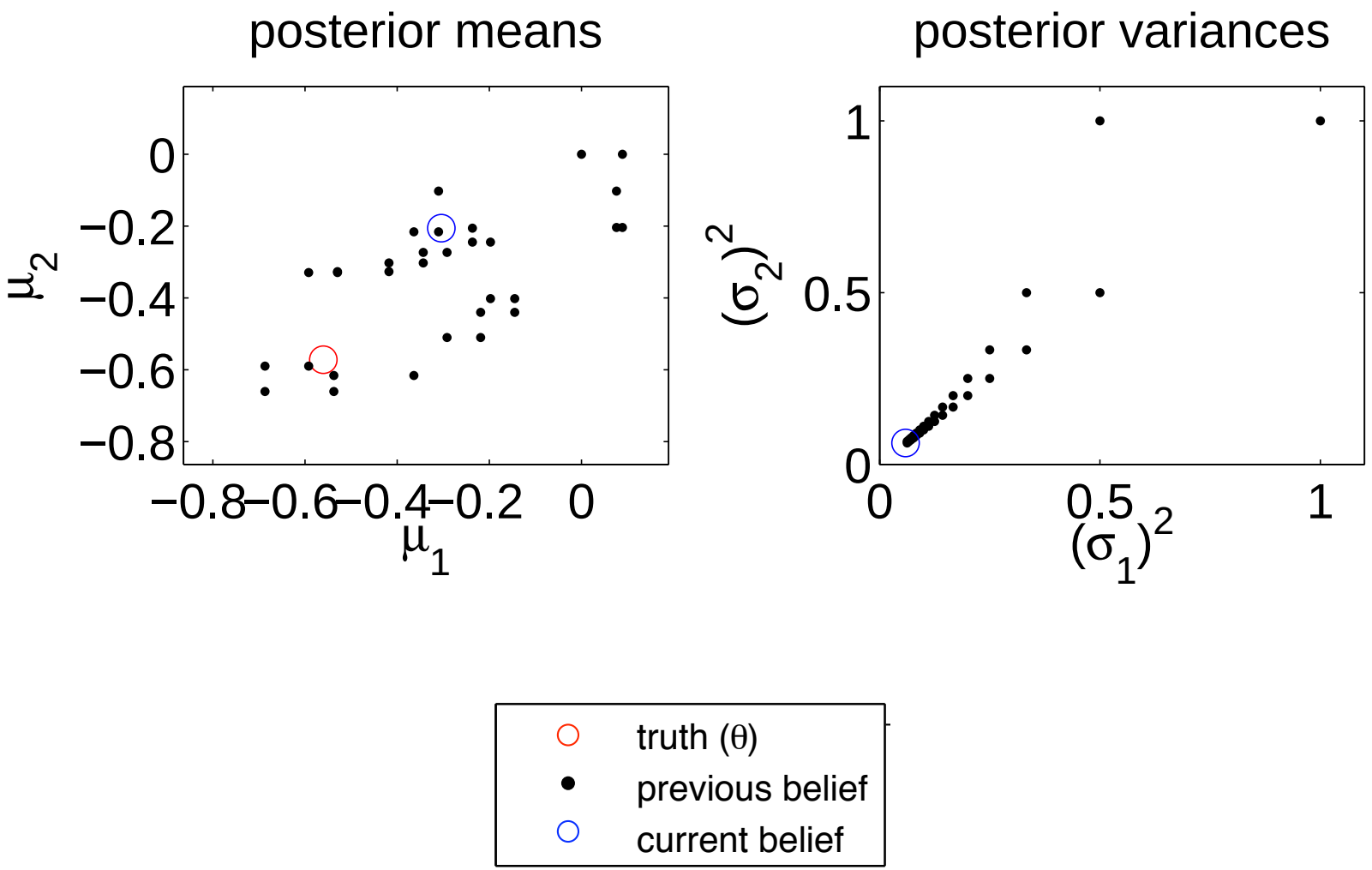
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

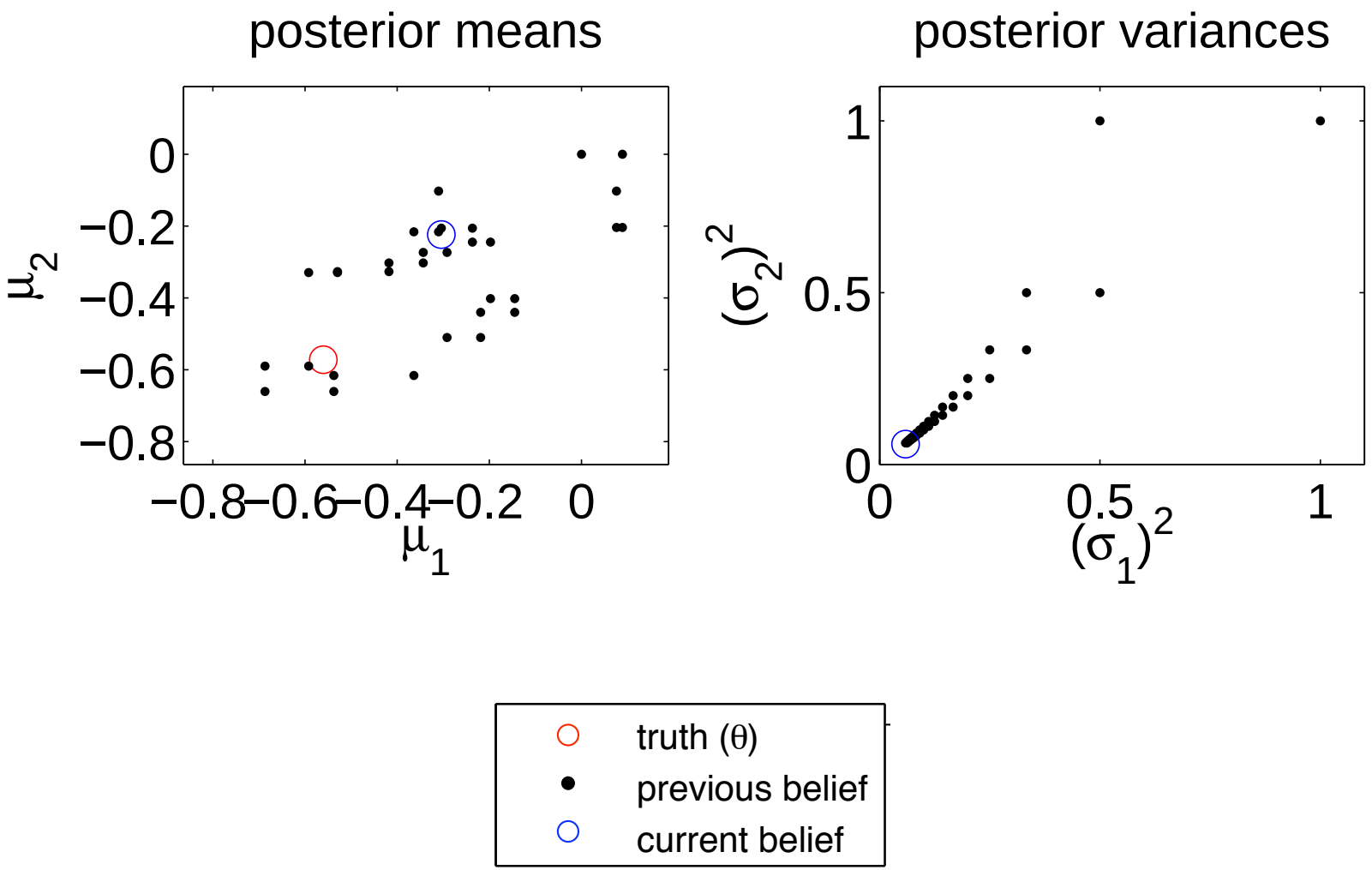
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

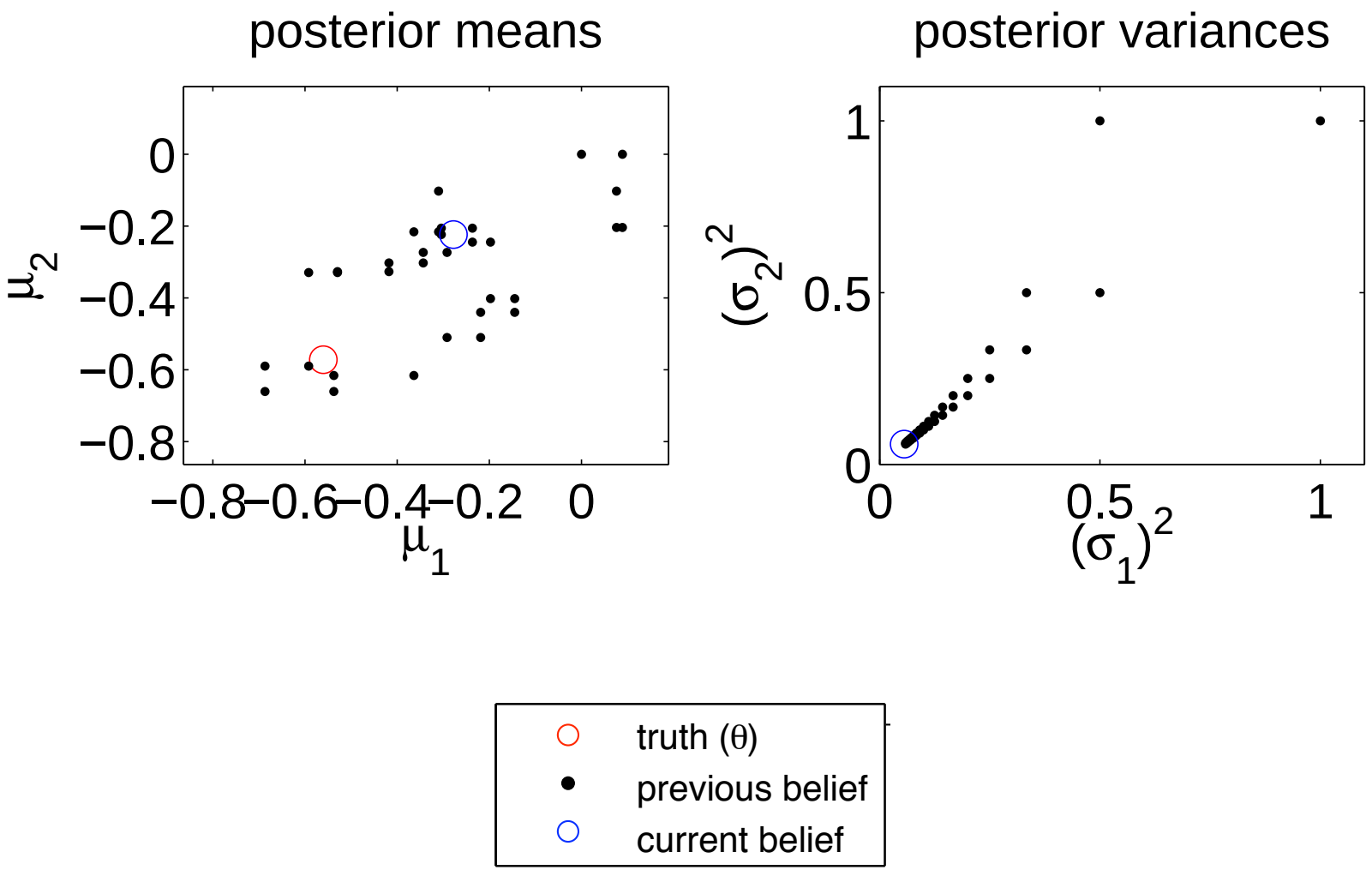
Simple Case


posterior means
posterior variances
μ_2
μ_1
$(\sigma_2)^2$
$(\sigma_1)^2$
0
−0.2
−0.4
−0.6
−0.8
0.5
1
truth (θ)
previous belief
current belief

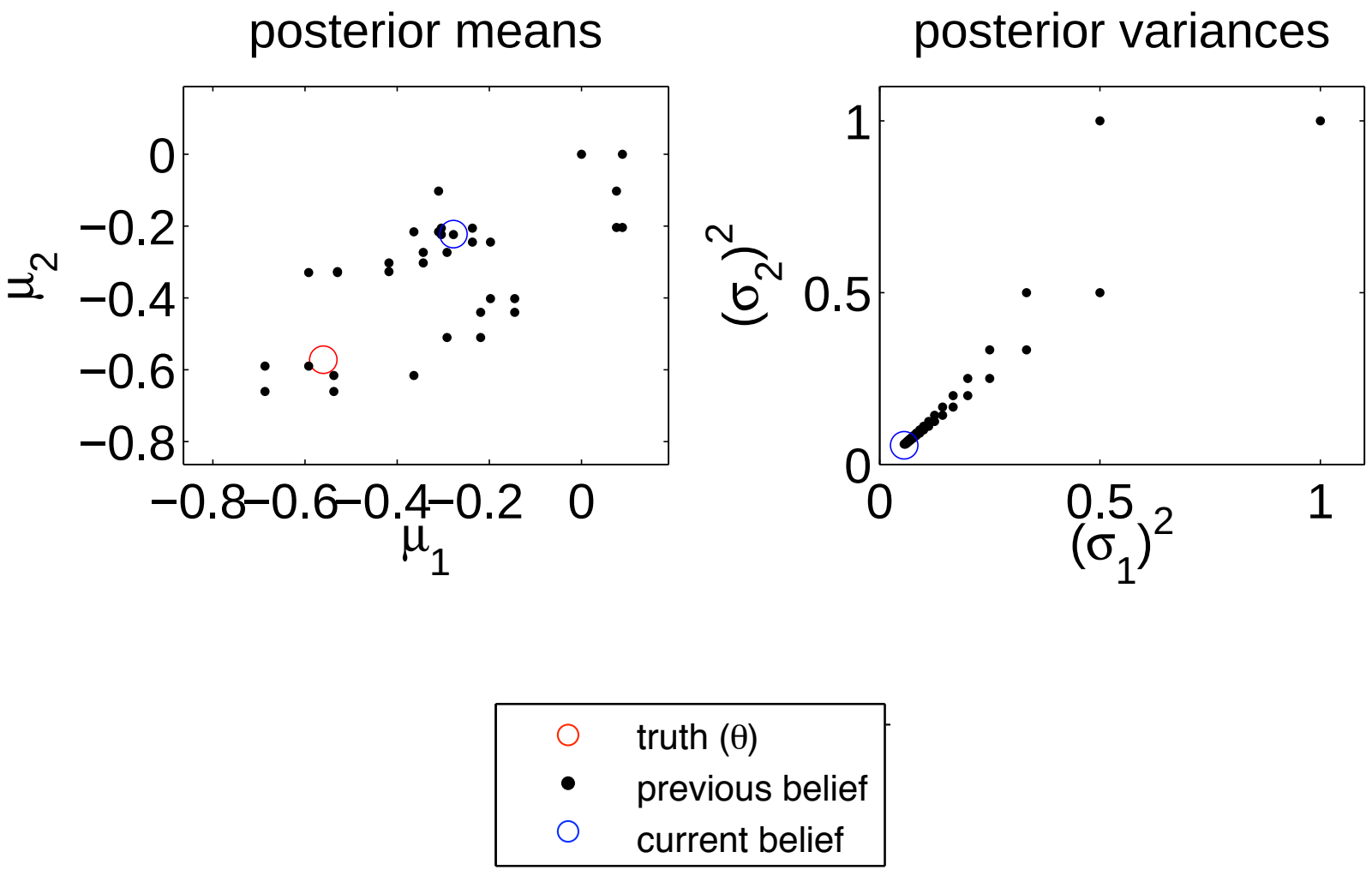
Simple Case


posterior means
posterior variances
μ_2
μ_1
$(\sigma_2)^2$
$(\sigma_1)^2$
0
−0.2
−0.4
−0.6
−0.8
0.5
1
truth (θ)
previous belief
current belief

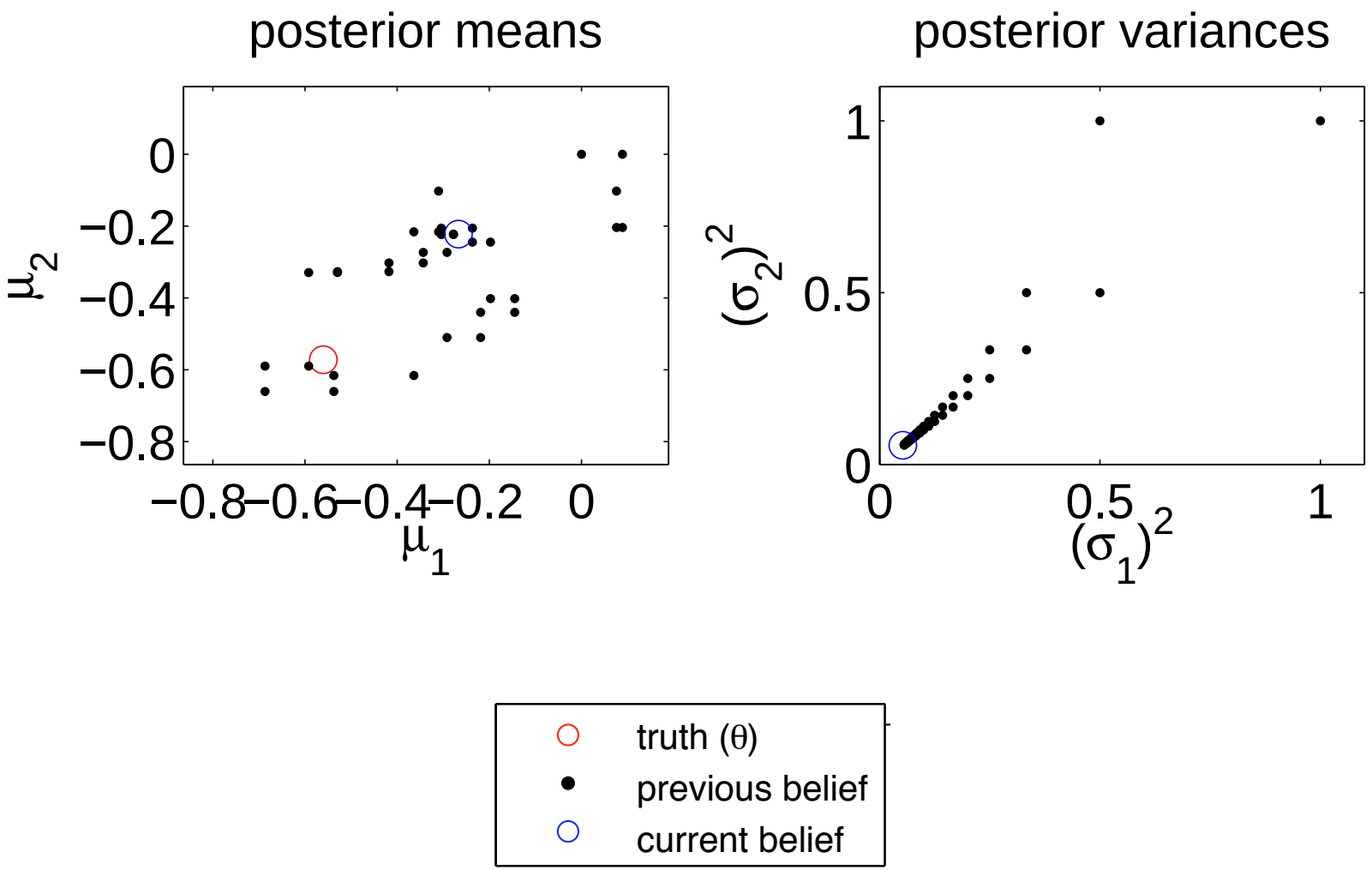
Simple Case


posterior means
posterior variances
μ_2
μ_1
$(\sigma_2)^2$
$(\sigma_1)^2$
0
−0.2
−0.4
−0.6
−0.8
0.5
1
truth (θ)
previous belief
current belief

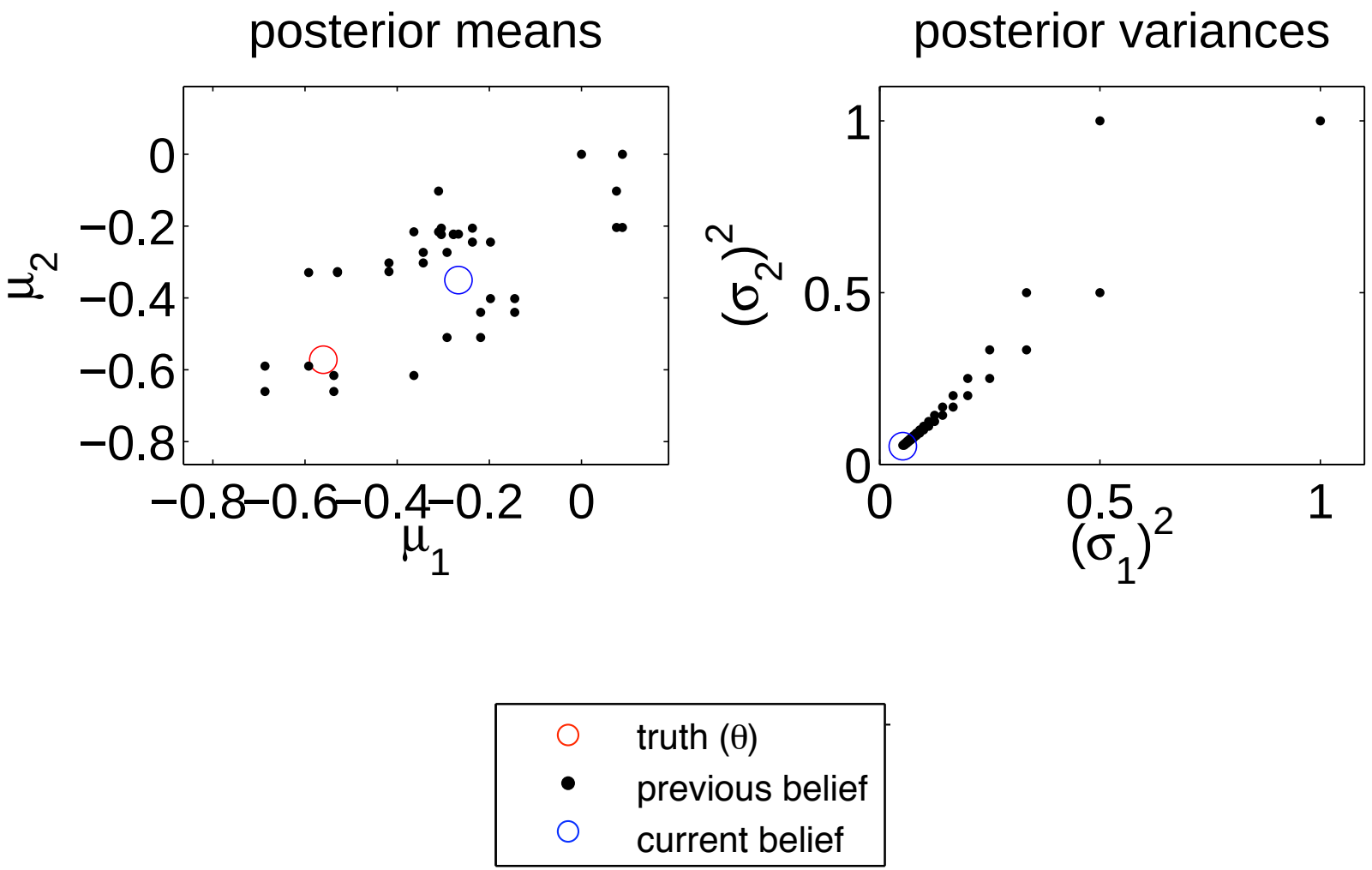
Simple Case


posterior means
posterior variances
μ_2
μ_1
$(\sigma_2)^2$
$(\sigma_1)^2$
0
−0.2
−0.4
−0.6
−0.8
0.5
1
truth (θ)
previous belief
current belief

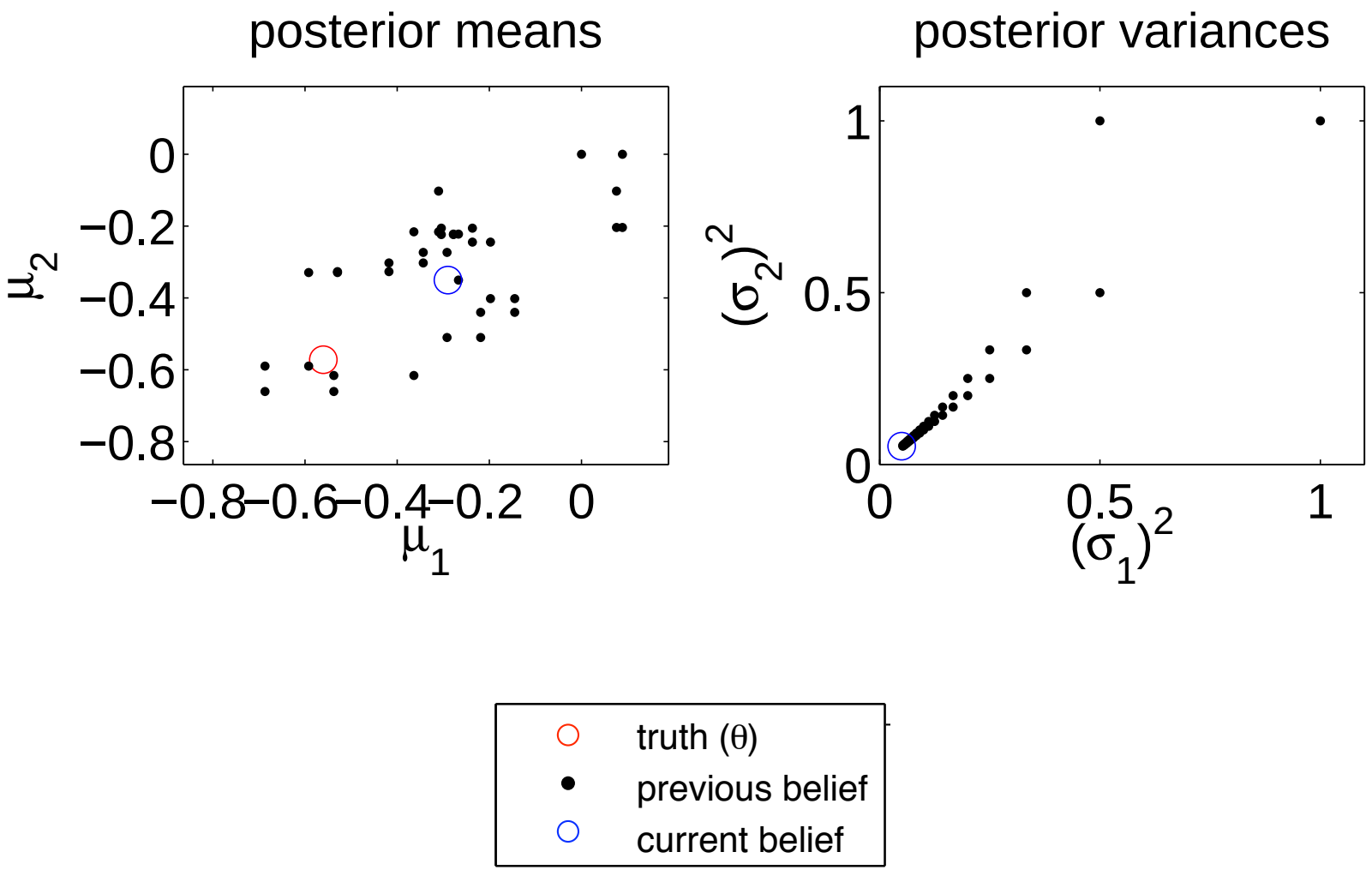
Simple Case


posterior means
posterior variances
μ_2
μ_1
$(\sigma_2)^2$
$(\sigma_1)^2$
0
−0.2
−0.4
−0.6
−0.8
−0.8
−0.6
−0.4
−0.2
0
1
0.5
0
0
0.5
1
truth (θ)
previous belief
current belief

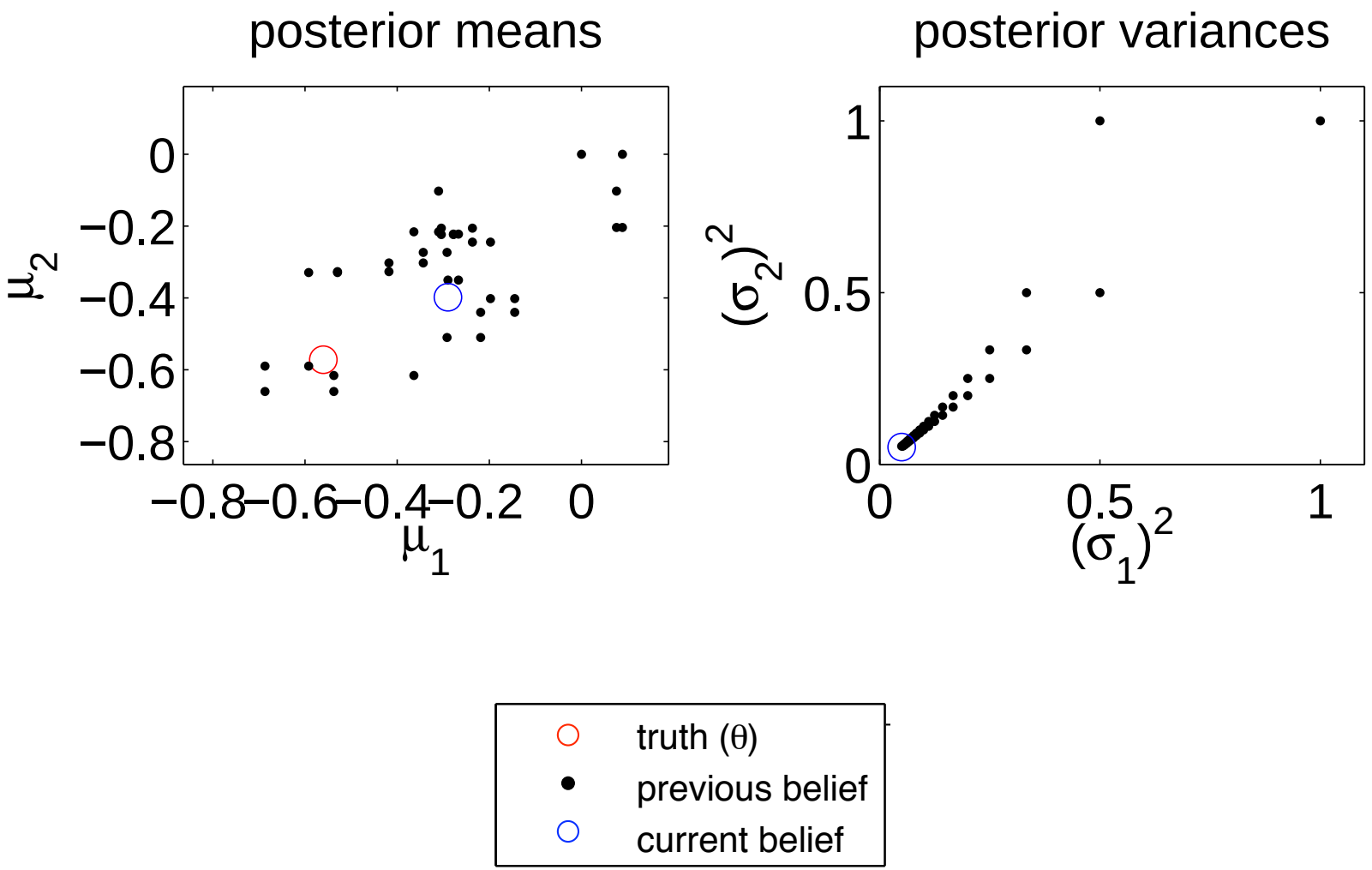
Simple Case


posterior means
posterior variances
μ_2
μ_1
$(\sigma_2)^2$
$(\sigma_1)^2$
0
−0.2
−0.4
−0.6
−0.8
−0.8
−0.6
−0.4
−0.2
0
1
0.5
0
0
0.5
1
truth (θ)
previous belief
current belief

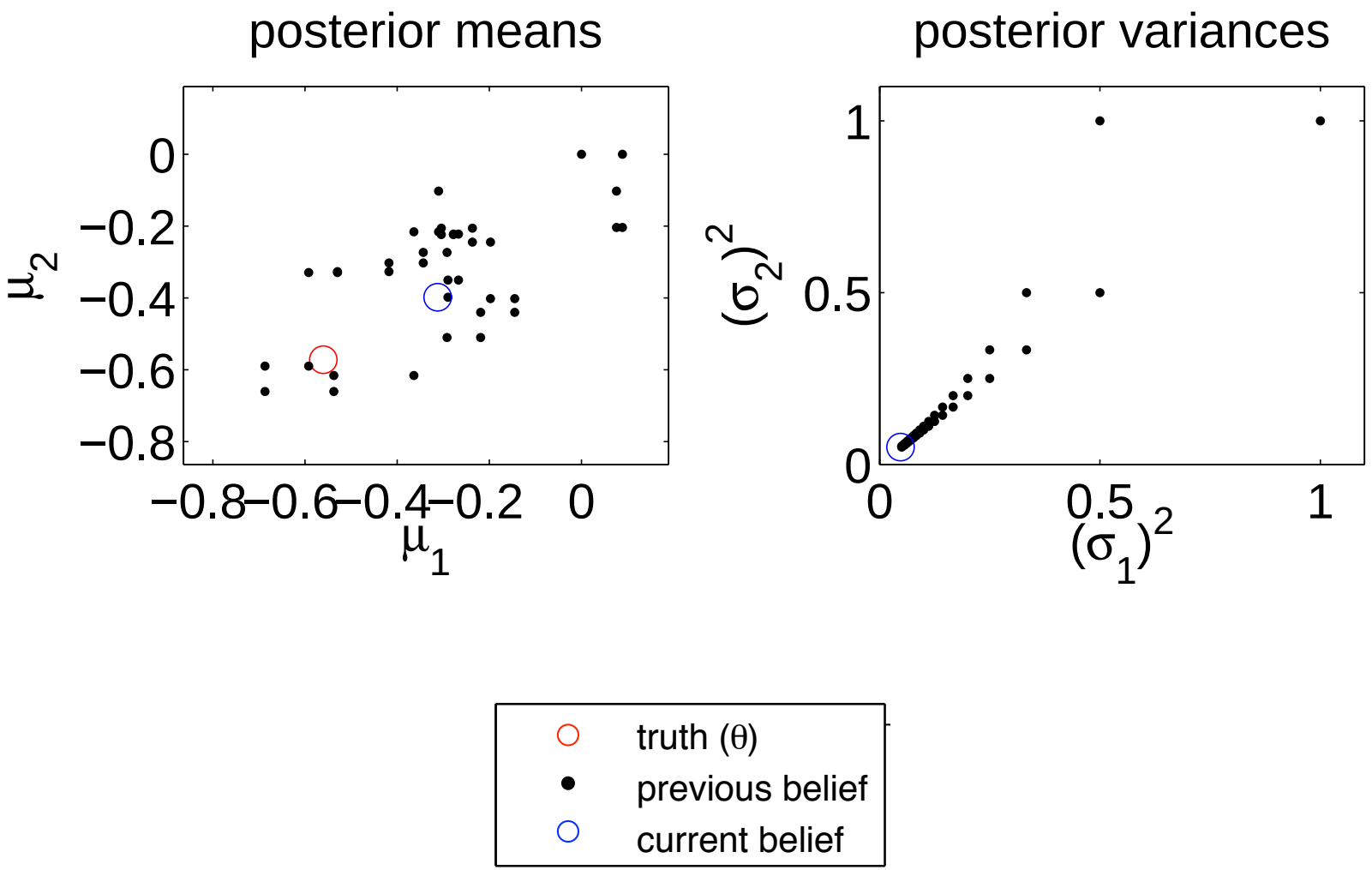
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

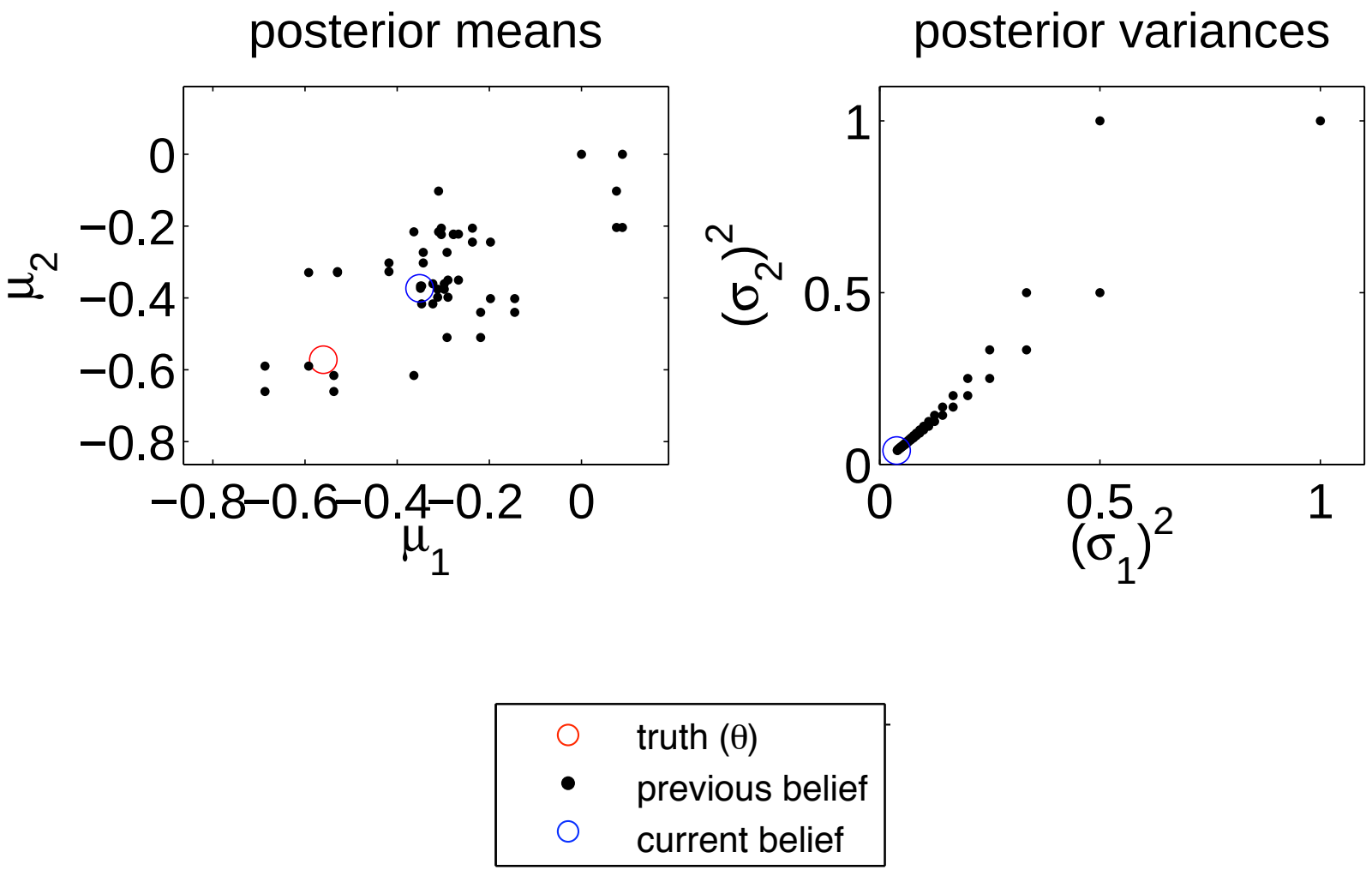
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

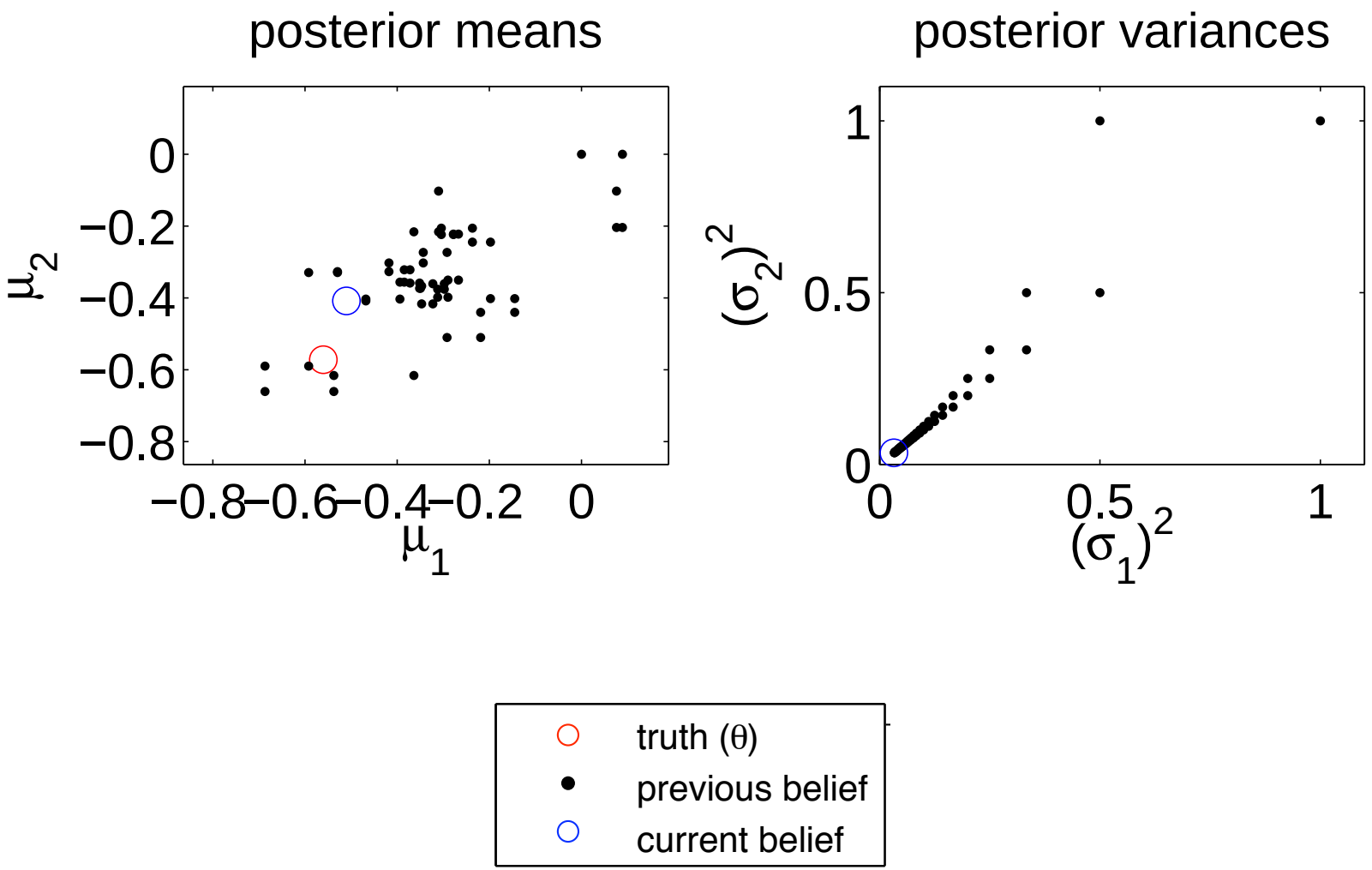
Simple Case


posterior means
posterior variances
μ_2
μ_1
$(\sigma_2)^2$
$(\sigma_1)^2$
0
−0.2
−0.4
−0.6
−0.8
−0.8 −0.6 −0.4 −0.2 0
0.5
1
truth (θ)
previous belief
current belief

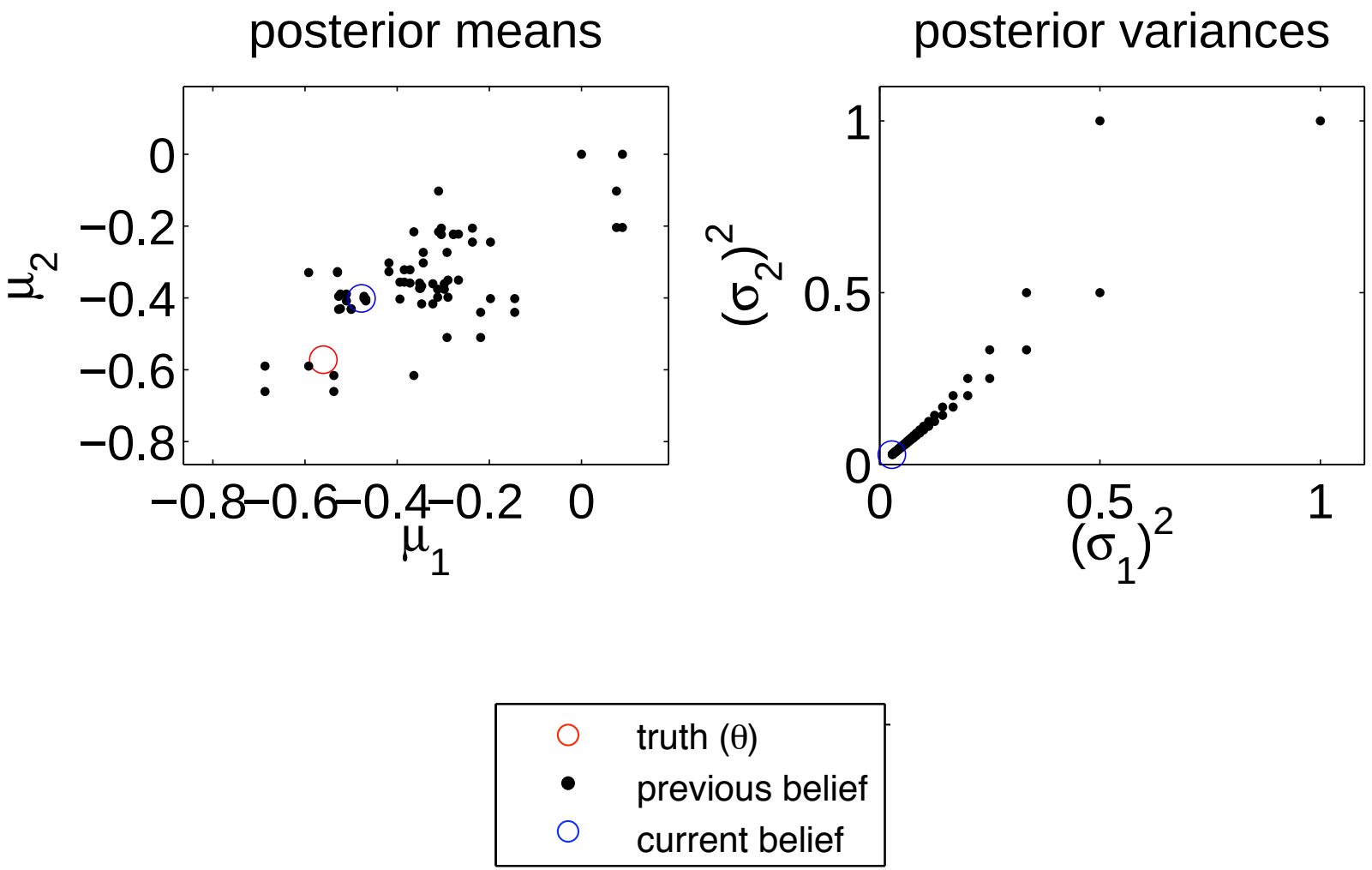
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

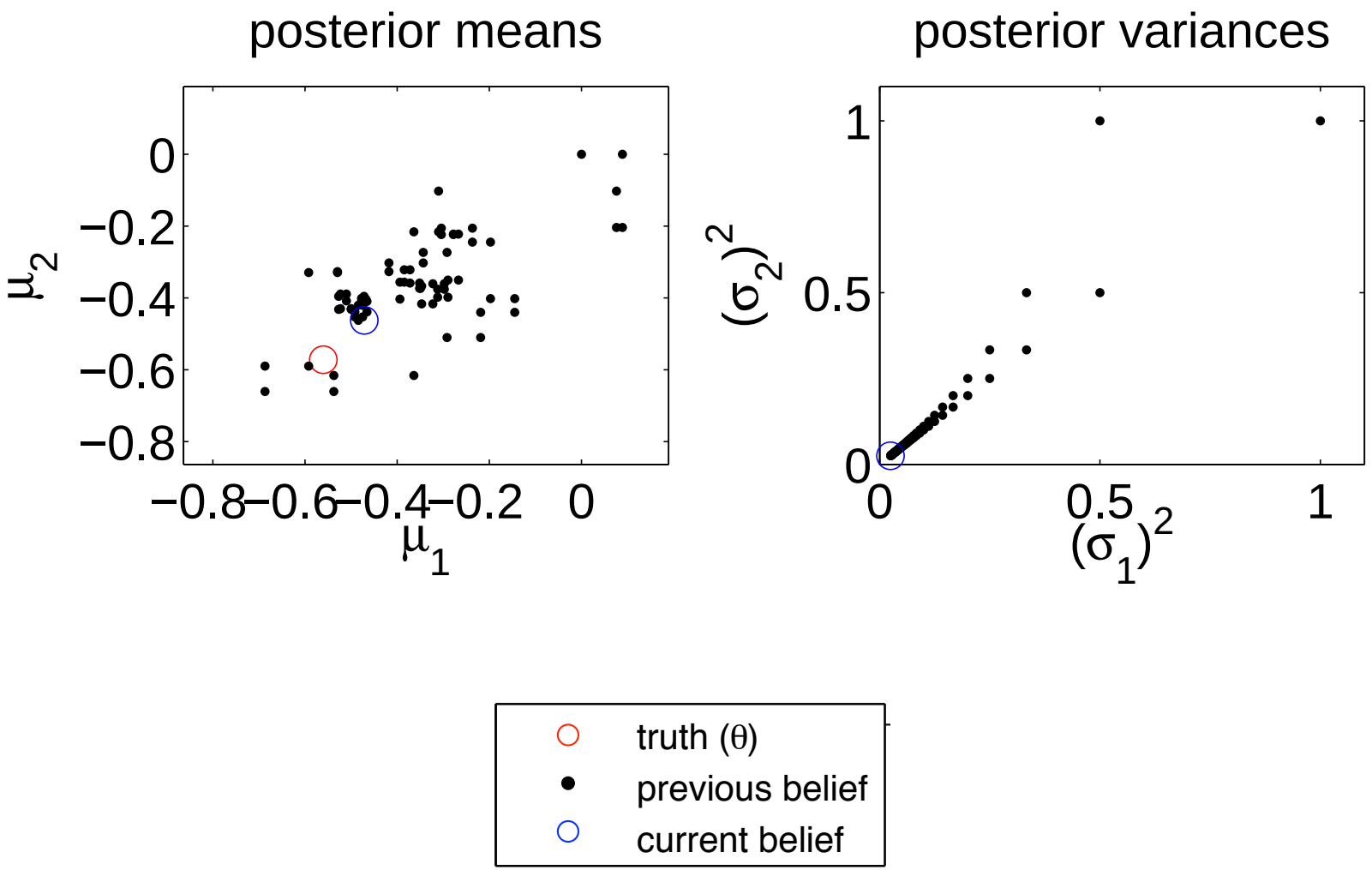
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

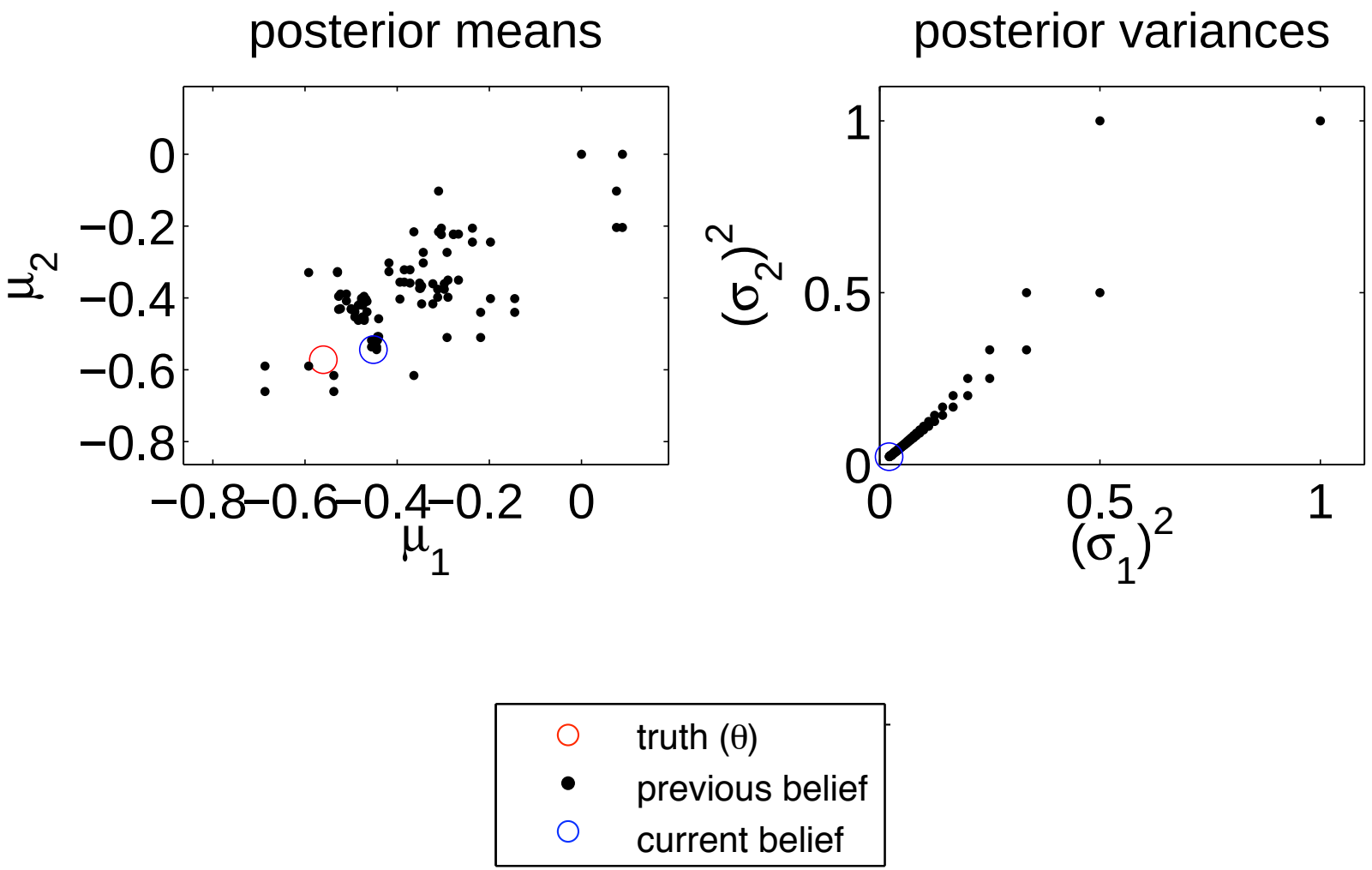
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8
−0.6
−0.4
−0.2
0
μ1
1
0.5
0
(σ2)2
0
0.5
1
(σ1)2
truth (θ)
previous belief
current belief

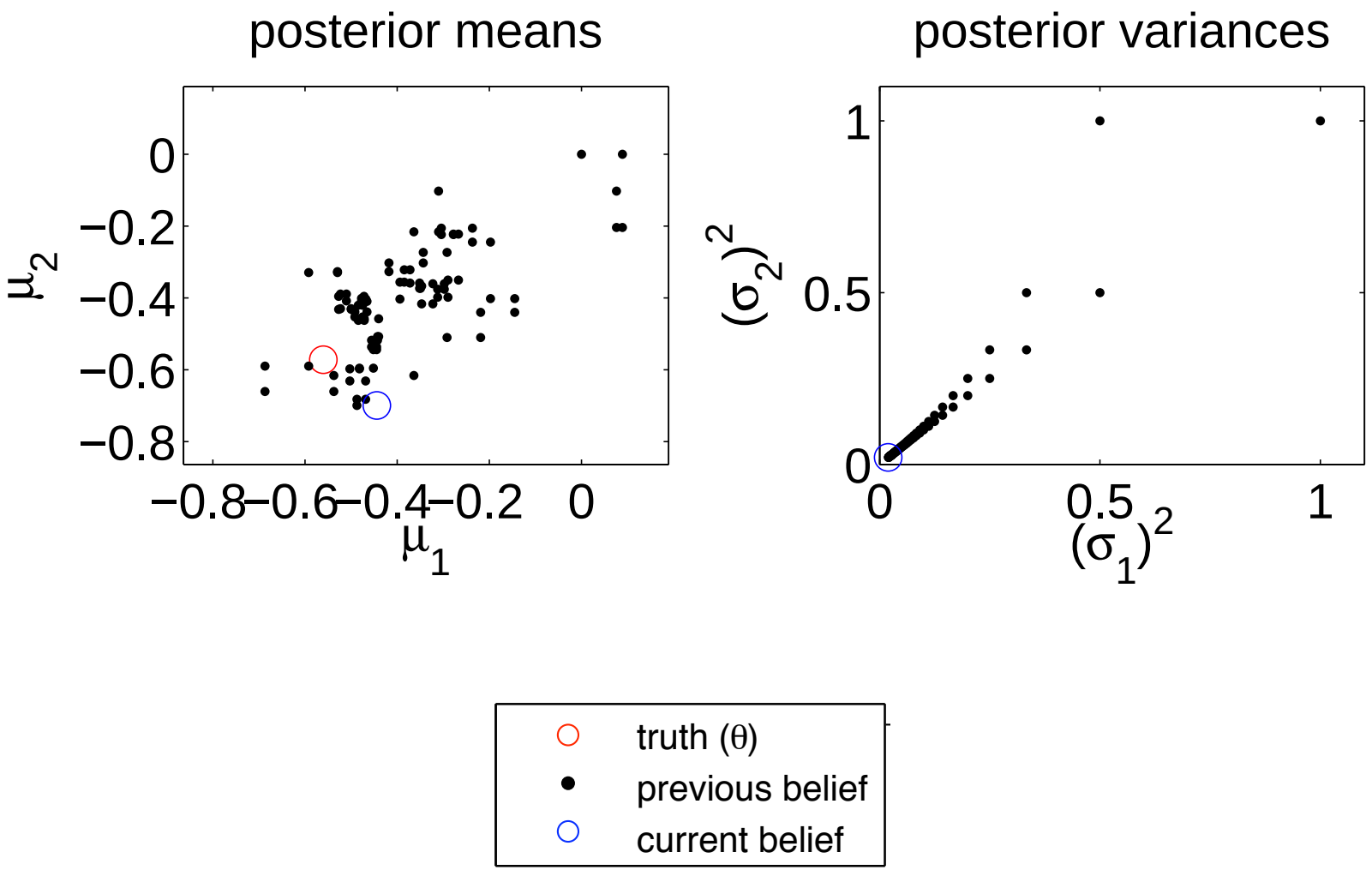
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

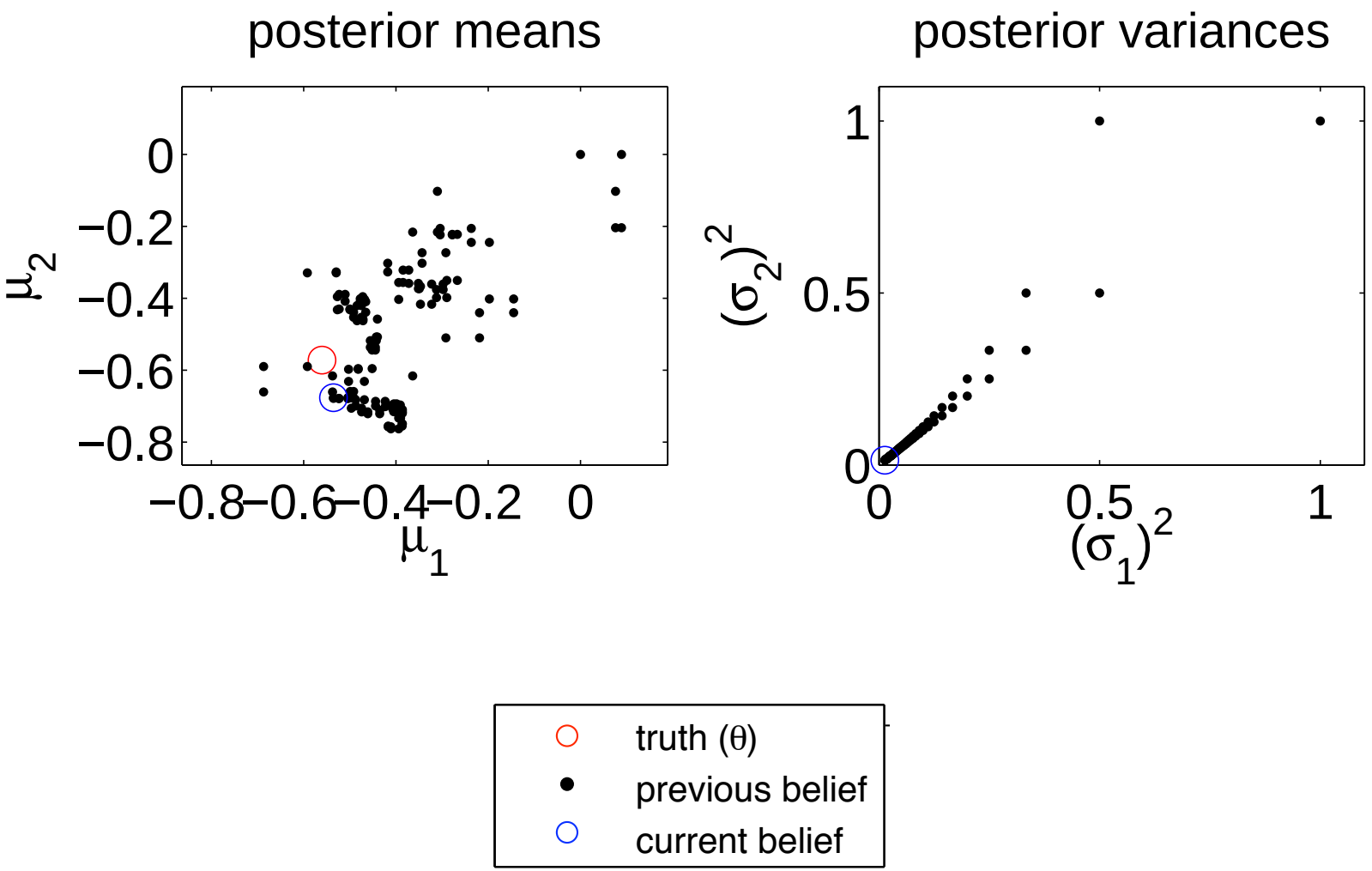
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

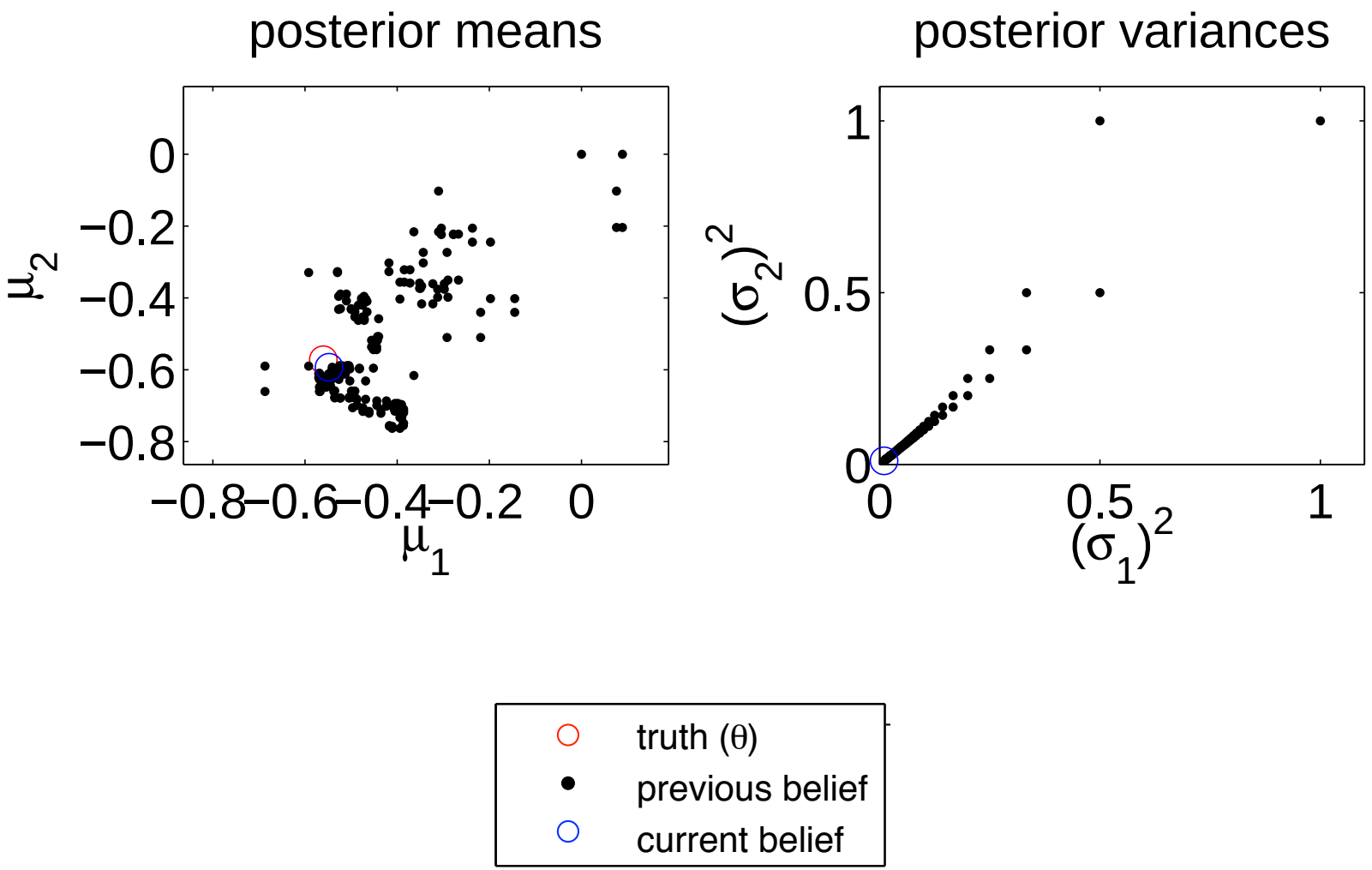
Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

Simple Case


posterior means
posterior variances
0
−0.2
−0.4
−0.6
−0.8
μ2
−0.8 −0.6 −0.4 −0.2 0
μ1
1
0.5
0
(σ2)2
0 0.5 1
(σ1)2
truth (θ)
previous belief
current belief

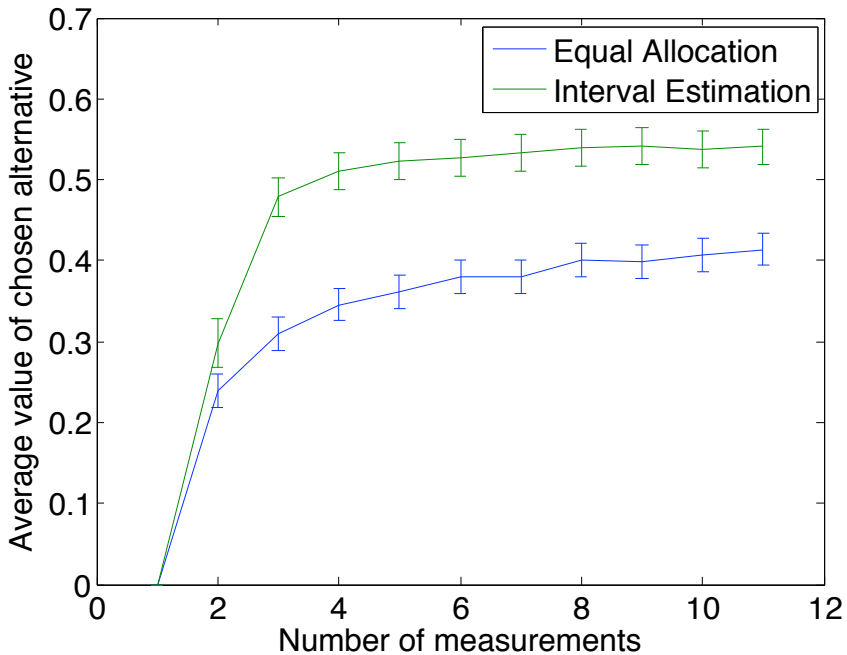
Simple Case

The **Interval Estimation (IE)** policy (Kaelbling 1993) measures

$$\arg\max_i \mu_{ti} + z_{\alpha/2}\sigma_{ti},$$

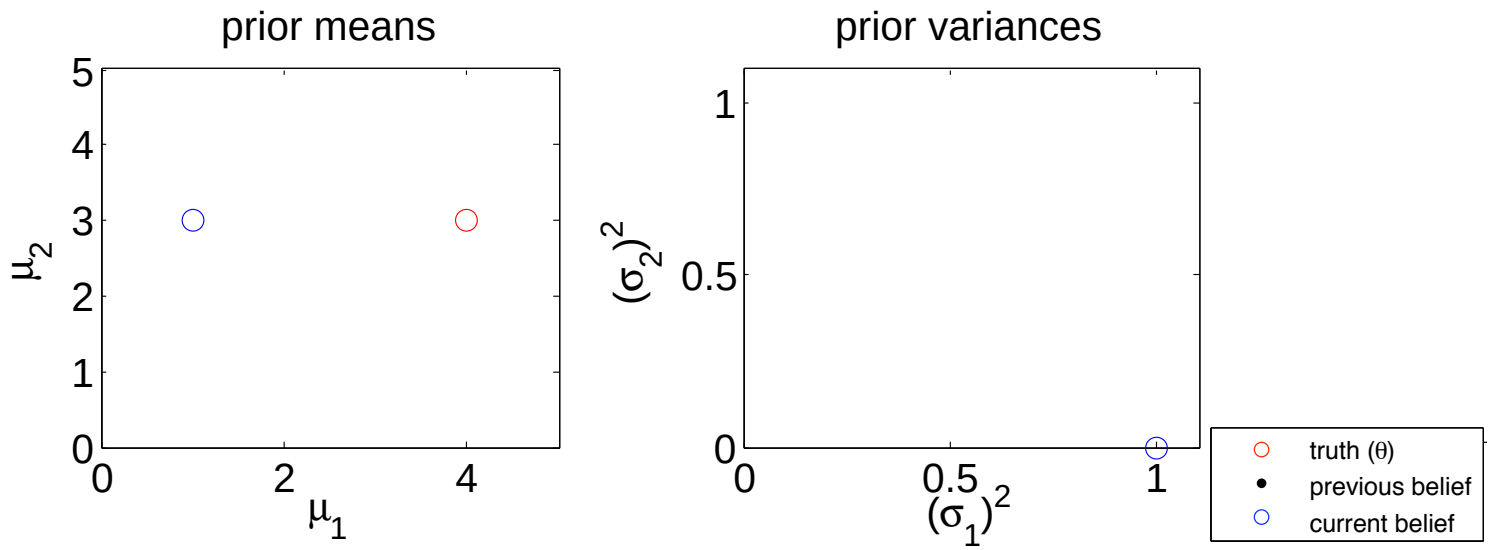
where $z_{\alpha/2}$ is the $\alpha/2$ quantile of the standard normal distribution.

- IE often works well in practice.
- The figure shows results for the risk function $R(\theta, i) = (\max_j \theta_j) - \theta_i$, where $\theta_0 = 0$ is an extra "do nothing treatment" with known value.



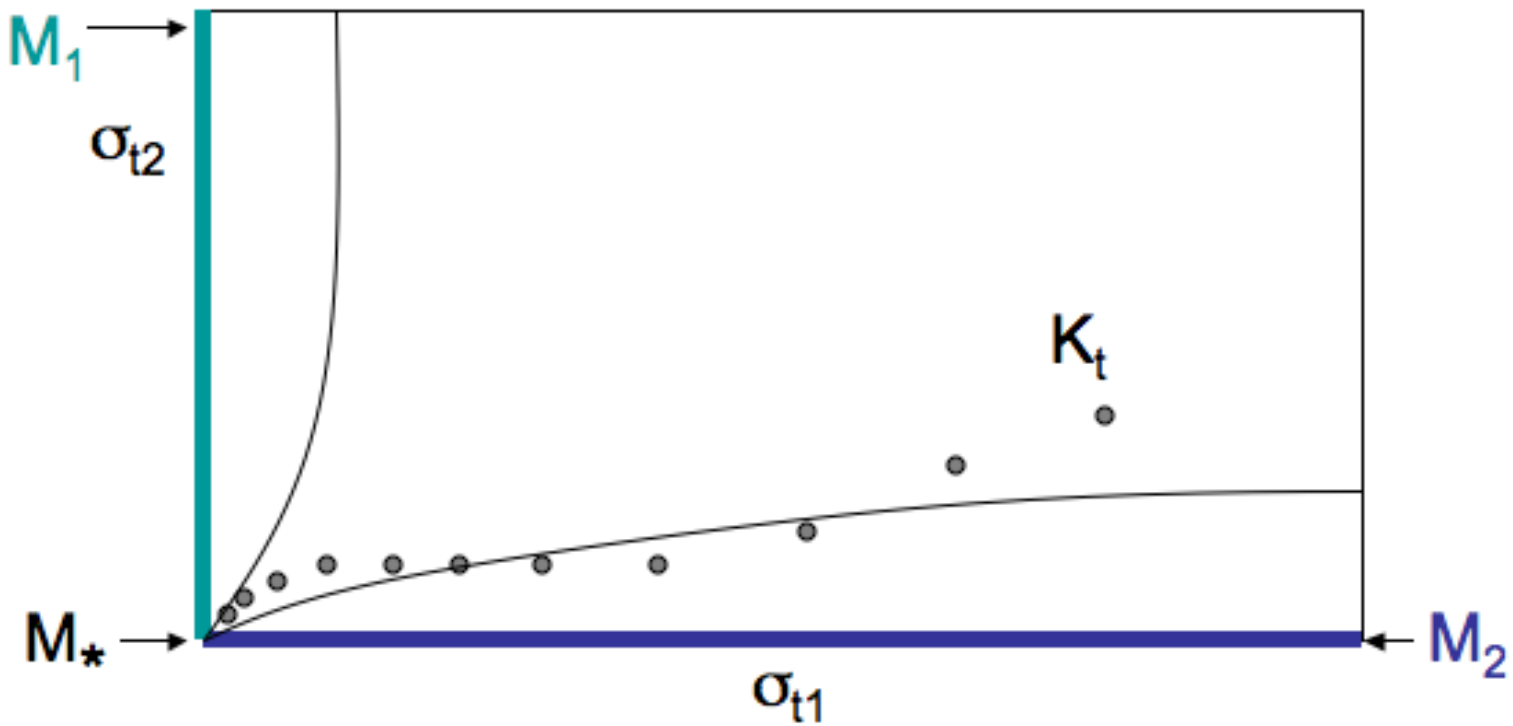
Simple Case

- IE often works well in practice ...but does not always converge to a global optimum
- Consider a prior $\mu_{01} = 1$, $\sigma_{01} = 1$, $\mu_{02} = 1.01 + z_{\alpha/2}$, $\sigma_{02} = 0$.
  - Recall IE measures $\arg\max_i \mu_{ti} + z_{\alpha/2}\sigma_{ti}$.
  - Treatment 1 has IE value $1 + z_{\alpha/2}$.
  - Treatment 2 has IE value $1.01 + z_{\alpha/2}$.
- IE will measure treatment 2 forever without measuring treatment 1, and will never discover if drug 1 is actually better.



Simple Case

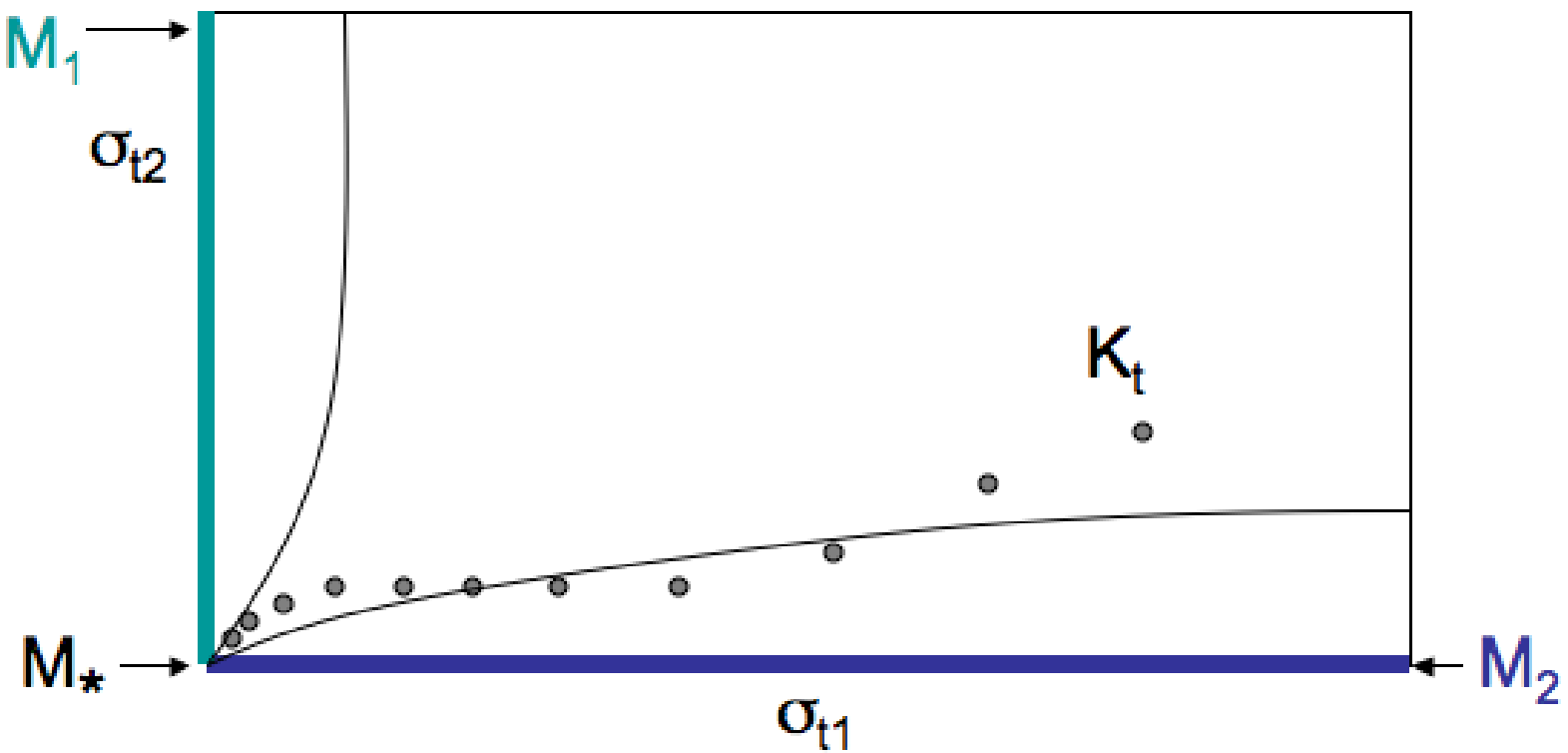
- Let $K_t = (\mu_t, \sigma_t^2)$ denote our posterior belief at time t.
- Let M_i be the set of μ, σ^2 with $\sigma_i^2 = 0$. This is the set of posterior distributions under which alternative i is "stuck" – measuring it does not change the posterior.
- Let $M_* = M_1 \cap M_2$. This is the set of posterior distributions under which every alternative is stuck, and we know θ perfectly.



Simple Case

Theorem

Suppose that every pair μ,σ^2 in $M_1 \cup M_2 \setminus M_*$ has an open set U containing it such that the policy measures an unstuck alternative for every posterior belief in U. Then the posterior belief has a limit in M_* and the policy converges to a global optimum.



General Case

We now generalize, assuming only the following:

- Sampling distributions for each measurement type reside in an exponential family.
  - Essentially, this condition only concerns the measurement noise, and does not restrict the form of θ, its prior distribution, or the form of the implementation decision.
- The number of measurement types and implementation decisions is finite.
- A few other technical conditions.

General Case

Theorem

If each $k \in \mathscr{K} \setminus M_*$ has an open neighborhood U in $\mathrm{cl}(\mathscr{K})$ such that the policy measures outside A_k for every $k' \in U$, then the sampling policy converges to a global optimum.

Notation:

- $\mathscr{K}$ – set of possible posterior distributions.
- k – one particular posterior distribution.
- A_k – measurements that are stuck under k.
- M_* – set of posteriors in which we know θ perfectly.
- U – buffer region surrounding any point at which we could get stuck.

Application to Existing R&S Policies

From this general theorem follows convergence to global optima of the following policies:

1. OCBA for linear loss (He et al. 2007).
2. LL(S) (Chick & Inoue 2001).
3. LL(1) (Chick et al. 2007).
4. (R1,. . . ,R1) (KG for independent normal ranking and selection) (Gupta & Miescke 1996, Frazier et al. 2008).
5. KG on a graph (Ryzhov & Powell 2009).

This theorem is particularly easy to apply for a broad class of policies known as KG policies, of which 3, 4 and 5 are examples.

Thank You

Any questions?