POPULATION FILTER INFERENCE

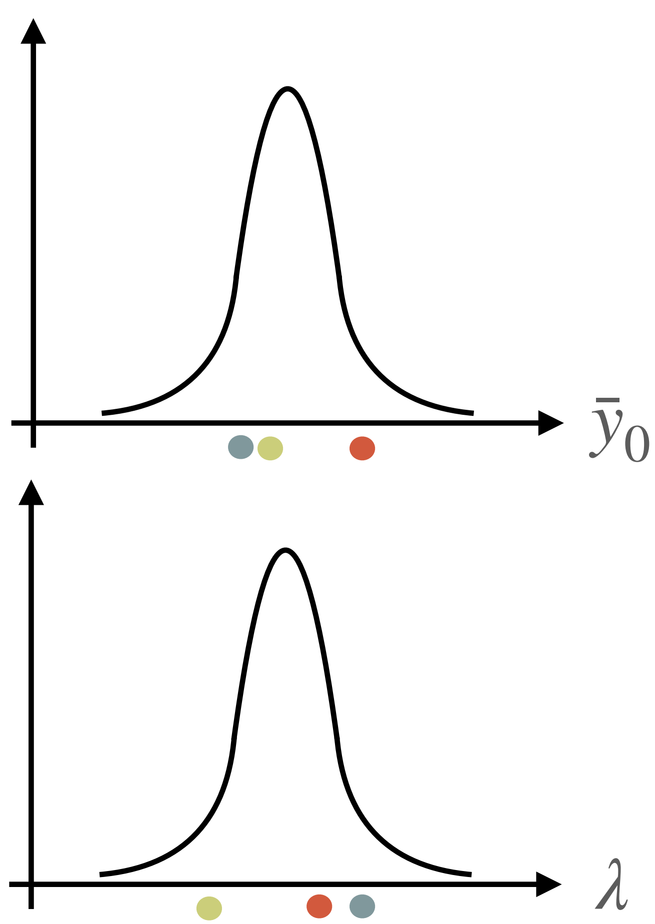
A scalable hierarchical inference framework

EXAMPLE: HIERARCHICAL EXP. GROWTH MODEL

Population model

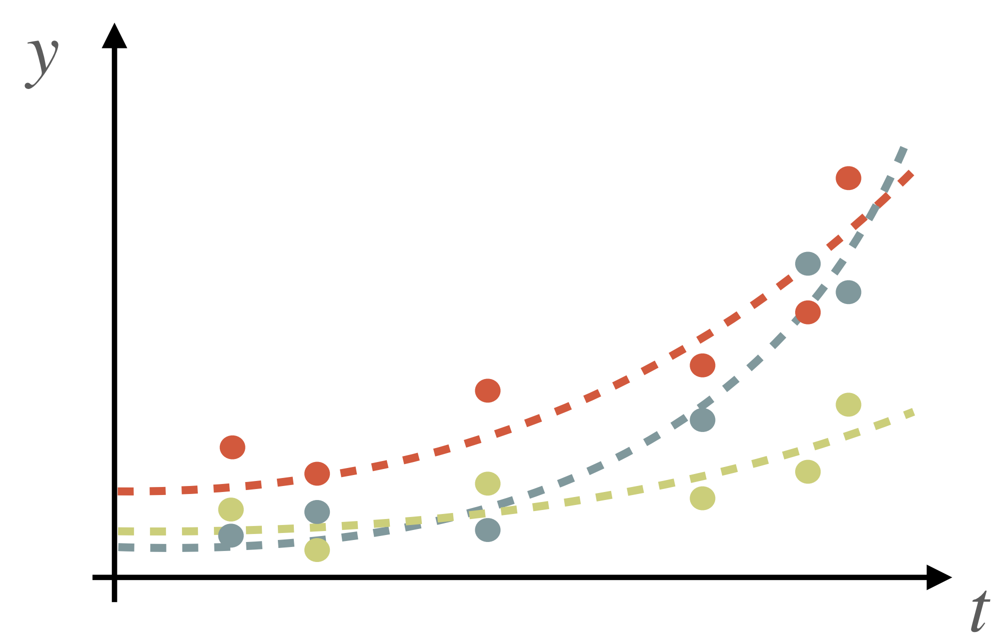
$$\bar{y}_0 \sim \mathcal{N}\left(\mu_{\bar{y}_0}, \sigma^2_{\bar{y}_0}\right)$$

$$\lambda \sim \mathcal{N}\left(\mu_\lambda, \sigma^2_\lambda\right)$$



Individual model

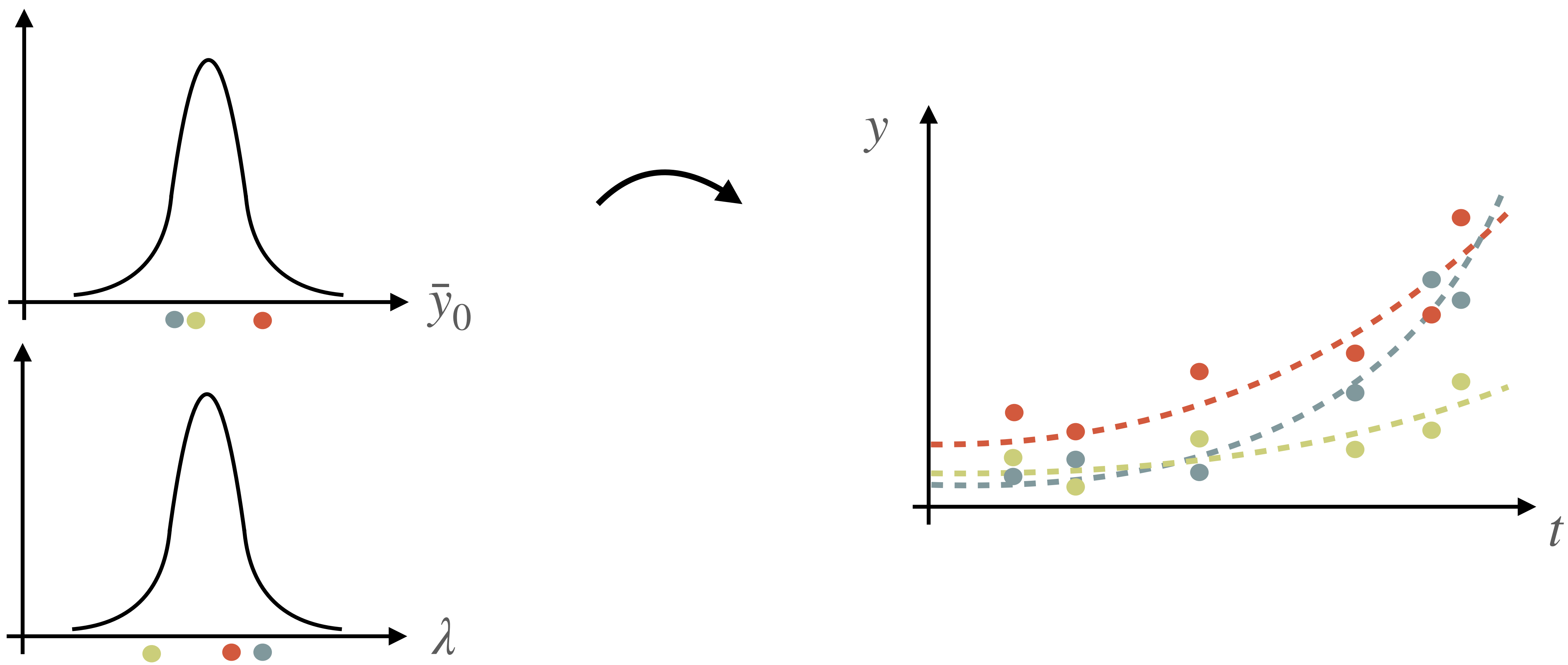
$$y = \bar{y}_0 e^{\lambda t} + \epsilon$$
$$\epsilon \sim \mathcal{N}(0,\sigma^2)$$



EXAMPLE: HIERARCHICAL EXP. GROWTH MODEL

Data-generating model

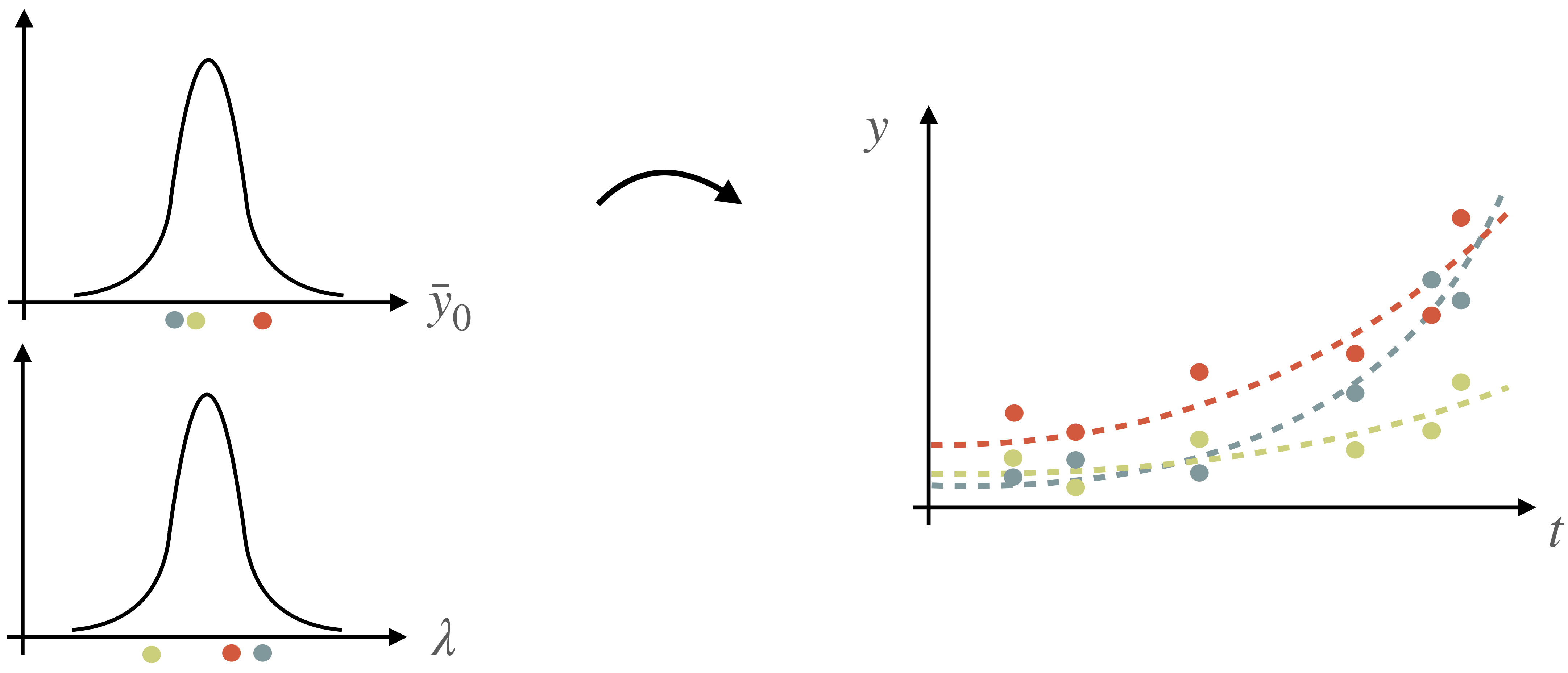
$$p(y \,|\, \mu_{\bar{y}_0}, \sigma_{\bar{y}_0}, \mu_\lambda, \sigma_\lambda, \sigma, t) = \int \mathrm{d}\bar{y}_0 \,\mathrm{d}\lambda\, \mathcal{N}(y \,|\, \bar{y}_0 \mathrm{e}^{\lambda t}, \sigma^2)\, \mathcal{N}(\bar{y}_0 \,|\, \mu_{\bar{y}_0}, \sigma^2_{\bar{y}_0})\, \mathcal{N}(\lambda \,|\, \mu_\lambda, \sigma^2_\lambda)$$



EXAMPLE: HIERARCHICAL EXP. GROWTH MODEL

Data-generating model

$$\underbrace{p(y\,|\,\mu_{\bar{y}_0},\sigma_{\bar{y}_0},\mu_\lambda,\sigma_\lambda,\sigma,t)}_{p(y|\theta,t)} = \int \underbrace{\mathrm{d}\bar{y}_0\,\mathrm{d}\lambda}_{\mathrm{d}\psi}\;\underbrace{\mathcal{N}(y\,|\,\bar{y}_0 e^{\lambda t},\sigma^2)}_{p(y|\psi,t)}\;\underbrace{\mathcal{N}(\bar{y}_0\,|\,\mu_{\bar{y}_0},\sigma^2_{\bar{y}_0})\,\mathcal{N}(\lambda\,|\,\mu_\lambda,\sigma^2_\lambda)}_{p(\psi|\theta)}$$



HOW CAN WE ESTIMATE POPULATION PARAMETERS FROM DATA?

1. NAÏVE APPROACH: POPULATION PARAMETER POSTERIOR

Log-likelihood & log-posterior

LOG-LIKELIHOOD:

$$\log p(\mathscr{D} \,|\, \theta) = \sum_j \log p(y_j^{\mathrm{obs}} \,|\, \theta, t_j) \qquad \text{for data} \qquad \mathscr{D} = \{(y_j^{\mathrm{obs}}, t_j)\}$$

POSTERIOR:

$$p(\theta \,|\, \mathscr{D}) \propto p(\mathscr{D} \,|\, \theta)\, p(\theta)$$

Recall: Hierarchical exp. growth model

$$p(y \,|\, \theta, t) = \int \mathrm{d}\bar{y}_0\, \mathrm{d}\lambda\, \mathscr{N}(y \,|\, \bar{y}_0 \mathrm{e}^{\lambda t}, \sigma^2)\, \mathscr{N}(\bar{y}_0 \,|\, \mu_{\bar{y}_0}, \sigma^2_{\bar{y}_0})\, \mathscr{N}(\lambda \,|\, \mu_\lambda, \sigma^2_\lambda)$$

1. NAÏVE APPROACH: POPULATION PARAMETER POSTERIOR

Log-likelihood & log-posterior

LOG-LIKELIHOOD:

$$\log p(\mathscr{D} \,|\, \theta) = \sum_j \log p(y_j^{\text{obs}} \,|\, \theta, t_j) \qquad \text{for data} \qquad \mathscr{D} = \{(y_j^{\text{obs}}, t_j)\}$$

POSTERIOR:

$$p(\theta \,|\, \mathscr{D}) \propto p(\mathscr{D} \,|\, \theta)\, p(\theta)$$

Recall: Hierarchical exp. growth model

$$p(y \,|\, \theta, t) = \int \mathrm{d}\bar{y}_0\, \mathrm{d}\lambda\, \mathscr{N}(y \,|\, \bar{y}_0 \mathrm{e}^{\lambda t}, \sigma^2)\, \mathscr{N}(\bar{y}_0 \,|\, \mu_{\bar{y}_0}, \sigma^2_{\bar{y}_0})\, \mathscr{N}(\lambda \,|\, \mu_\lambda, \sigma^2_\lambda)$$

VERY HARD TO COMPUTE NUMERICALLY!

2. HIERARCHICAL APPROACH: JOINT POSTERIOR

$$p(y\,|\,\theta, t) = \int \mathrm{d}\bar{y}_0\, \mathrm{d}\lambda\, \mathcal{N}(y\,|\,\bar{y}_0 \mathrm{e}^{\lambda t}, \sigma^2)\, \mathcal{N}(\bar{y}_0\,|\,\mu_{\bar{y}_0}, \sigma^2_{\bar{y}_0})\, \mathcal{N}(\lambda\,|\,\mu_\lambda, \sigma^2_\lambda)$$

DROP INTEGRAL

$$p(y, \psi\,|\,\theta, t) = \mathcal{N}(y\,|\,\bar{y}_0 \mathrm{e}^{\lambda t}, \sigma^2)\, \mathcal{N}(\bar{y}_0\,|\,\mu_{\bar{y}_0}, \sigma^2_{\bar{y}_0})\, \mathcal{N}(\lambda\,|\,\mu_\lambda, \sigma^2_\lambda)$$

Log-likelihood & log-posterior

LOG-LIKELIHOOD FOR AN INDIVIDUAL:

$$\log p(\mathscr{D}_i, \psi_i\,|\,\theta) = \sum_j \log p(y^{\mathrm{obs}}_{ij}, \psi_i\,|\,\theta, t_{ij})$$

for data of individual i

$$\mathscr{D}_i = \{(y^{\mathrm{obs}}_{ij}, t_{ij})\}$$

POPULATION LOG-LIKELIHOOD:

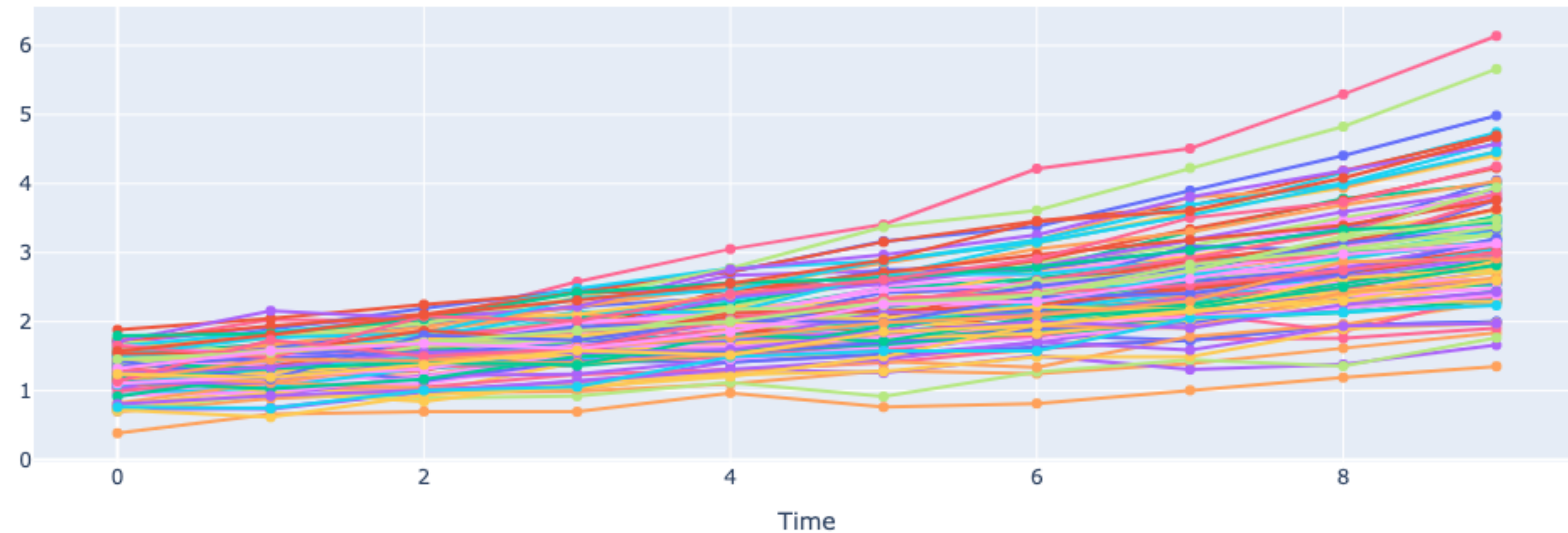
$$\log p(\mathscr{D}, \{\psi_i\}\,|\,\theta) = \sum_i \log p(\mathscr{D}_i, \psi_i\,|\,\theta)$$

POSTERIOR:

$$p(\theta, \{\psi_i\}\,|\,\mathscr{D}) \propto p(\mathscr{D}, \{\psi_i\}\,|\,\theta)\, p(\theta)$$

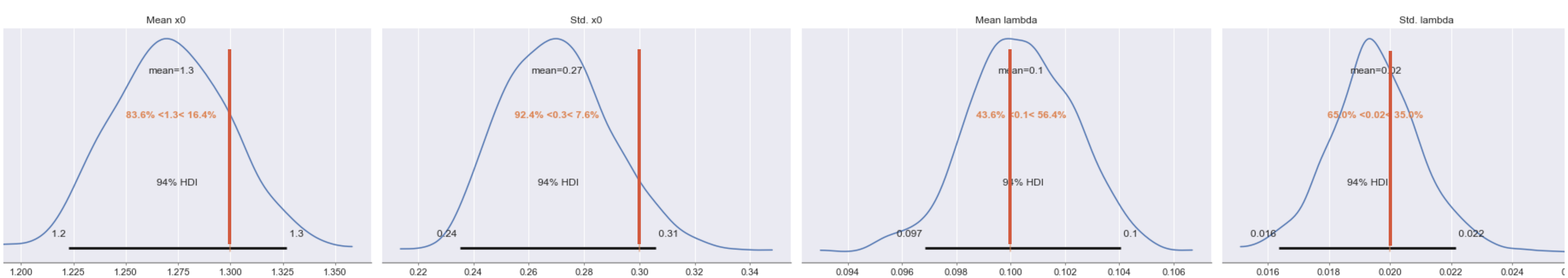
2. HIERARCHICAL APPROACH IN PRACTICE: EXAMPLE 1

DATA: 100 INDIVIDUALS WITH 10 OBSERVATIONS



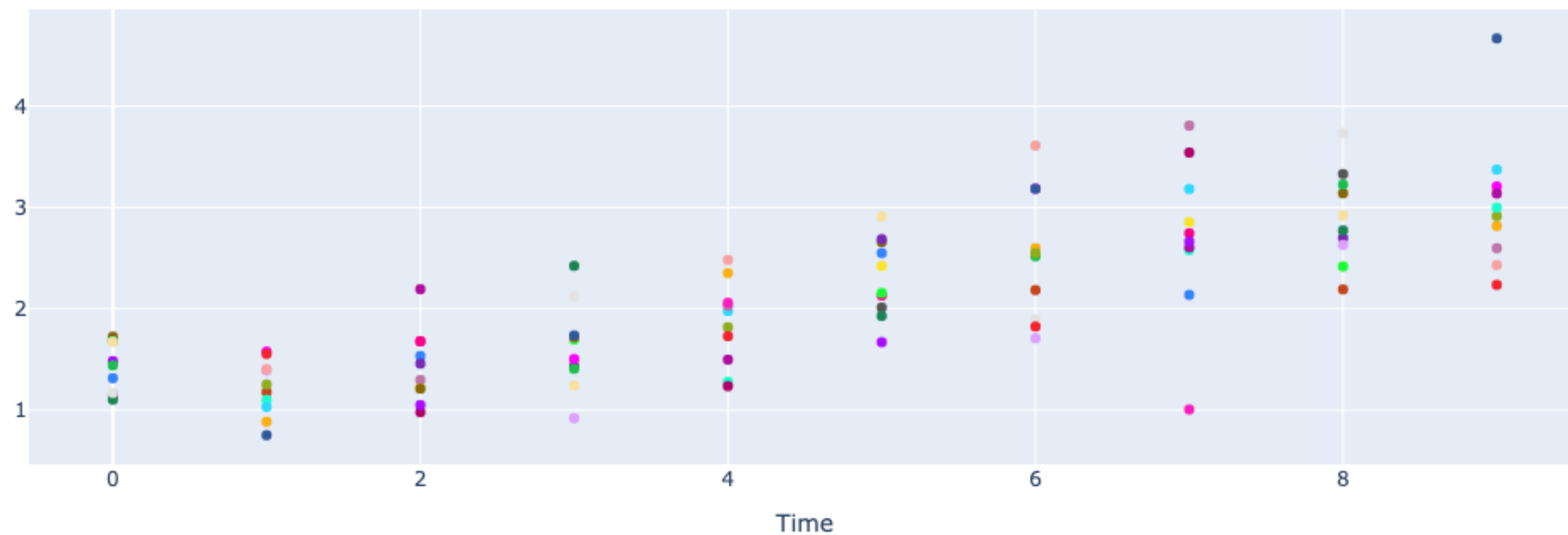
INFERENCE RESULTS:

RUN TIME ~ 13 MINUTES



2. HIERARCHICAL APPROACH IN PRACTICE: EXAMPLE 2

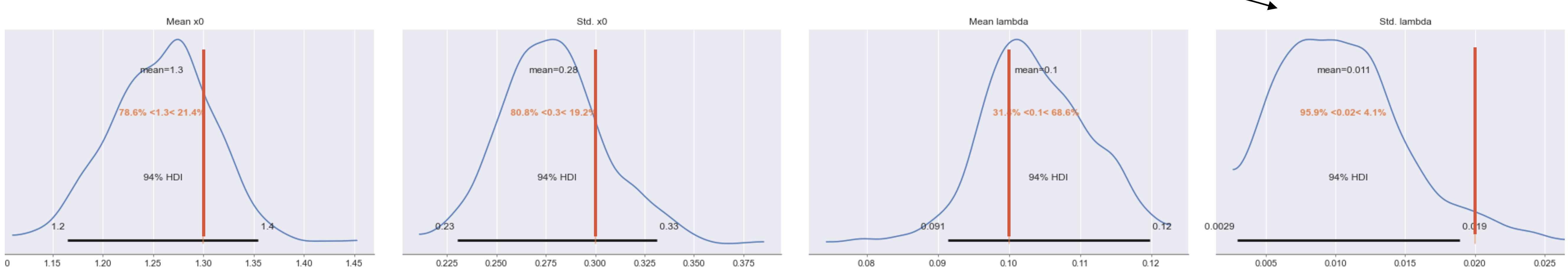
DATA: 100 INDIVIDUALS WITH 1 OBSERVATION EACH – SNAPSHOT DATA



INFERENCE RESULTS:

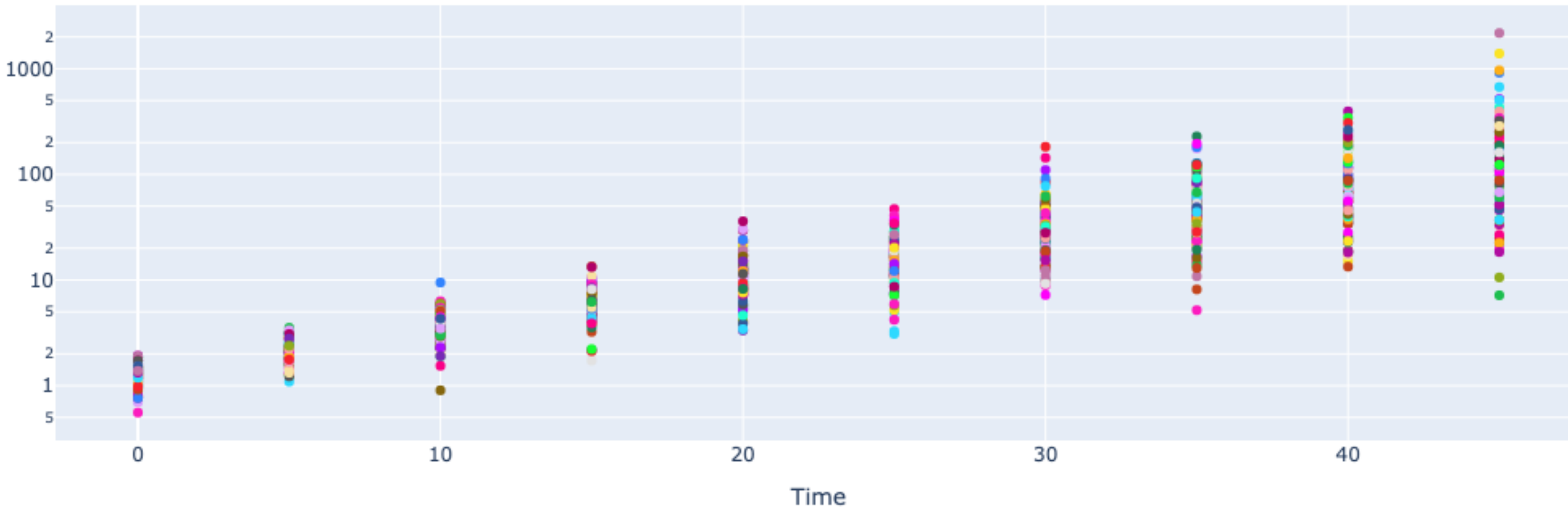
SIGNS OF SHRINKAGE

RUN TIME ~ 23 MINUTES



2. HIERARCHICAL APPROACH IN PRACTICE: EXAMPLE 3

DATA: 1000 INDIVIDUALS WITH 1 OBSERVATION EACH – SNAPSHOT DATA



INFERENCE RESULTS:

RUN TIME ~ DAYS

- HIERARCHICAL INFERENCE BECOMES INTRACTABLE.
- POPULATION VARIATION WILL BE SIGNIFICANTLY UNDERESTIMATED (SHRINKAGE).

WHAT ARE THE CHALLENGES WITH SNAPSHOT DATA?

WHAT ARE THE CHALLENGES WITH SNAPSHOT DATA?

Recall: Hierarchical exp. growth model

SOME INTUITION:

$$p(y, \psi \,|\, \theta, t) = p_n(y \,|\, \bar{y}_0 \mathrm{e}^{\lambda t}, \sigma^2)\, p_n(\bar{y}_0 \,|\, \mu_{\bar{y}_0}, \sigma^2_{\bar{y}_0})\, p_n(\lambda \,|\, \mu_\lambda, \sigma^2_\lambda)$$

1. **WITHOUT POPULATION MODEL: DATA IS FITTED EXACTLY.**

$$\bar{y}_{0,i} \mathrm{e}^{\lambda_i t} = y_i^{\mathrm{obs}} \quad \text{and} \quad \mathcal{N}(y_i^{\mathrm{obs}} \,|\, \bar{y}_{0,i} \mathrm{e}^{\lambda_i t}, \sigma^2) = 1 \quad \text{for} \quad \sigma \to 0$$

DEFINES A SET OF PARAMETERS FOR EACH INDIVIDUAL $\{\psi_i\} = \{(\bar{y}_{0,i}, \lambda_i)\}$.

2. **WITH POPULATION MODEL: VARIATION IN** $\{\psi_i\}$ **IS PENALISED.**

$$\underbrace{\prod_{i=1}^{N} \mathcal{N}(y_i^{\mathrm{obs}} \,|\, \bar{y}_{0,i} \mathrm{e}^{\lambda_i t}, \sigma^2)}_{N \text{ factors}} \quad \text{COMPETES AGAINST} \quad \underbrace{\prod_{i=1}^{N} \mathcal{N}(\bar{y}_{0,i} \,|\, \mu_{\bar{y}_0}, \sigma^2_{\bar{y}_0})\, \mathcal{N}(\lambda_i \,|\, \mu_\lambda, \sigma^2_\lambda)}_{2N \text{ factors}}$$

➤ **IT'S BENEFICIAL TO INCREASE NOISE AND REDUCE POPULATION VARIATION.**

BUT: IMBALANCE (N VS. 2N) LEADS TO TOO MUCH REDUCTION IN POPULATION VARIATION (I.E. SHRINKAGE).

WHAT ARE THE PROBLEMS WITH SNAPSHOT DATA?

In general

$$\underbrace{\prod_{i=1}^{N} p(y_i^{\text{obs}} \,|\, \psi_i)}_{N \text{ factors}} \quad \textbf{COMPETES AGAINST} \quad \underbrace{\prod_{i=1}^{N} \prod_{k=1}^{K} p(\psi_{ik} \,|\, \theta_k)}_{KN \text{ factors}}$$

➤ THE MORE PARAMETERS THE MORE SHRINKAGE OF POPULATION VARIATION.

RESOLUTIONS:

1. STRONG PRIOR ON NOISE (OR FIX NOISE).
   - DOES NOT REMOVE SHRINKAGE.
   + REDUCES MAGNITUDE OF SHRINKAGE.

2. IF POSSIBLE, AVOID FITTING TO INDIVIDUALS AND JUST FIT ON POPULATION LEVEL.
   + REMOVES SHRINKAGE
   + MAKES INFERENCE TIME INDEPENDENT OF NUMBER OF OBSERVED INDIVIDUALS
   - MAKES NOISE NON-IDENTIFIABLE
   - NEEDS APPROXIMATION OF POPULATION DISTRIBUTION

3. POPULATION FILTER APPROACH: MOTIVATION 1

MOTIVATION 1: WANT TO FIT POPULATION DISTRIBUTION DIRECTLY TO DATA.

PROBLEM: CANNOT COMPUTE POPULATION DISTRIBUTION.

$\longrightarrow$ APPROXIMATE POPULATION DISTRIBUTION, E.G. WITH A NORMAL DISTRIBUTION!

$$p(y \,|\, \theta, t) = \int \mathrm{d}\psi \, p(y \,|\, \psi, t) \, p(\psi \,|\, \theta) \approx \mathcal{N}(y \,|\, \mu(\theta, t), \sigma^2(\theta, t))$$

WHERE

$$\mu(\theta, t) = \mathbb{E}[y \,|\, \theta, t] \quad \text{and} \quad \sigma^2(\theta, t) = \mathrm{Var}[y \,|\, \theta, t]$$

AT FIRST GLANCE THIS LOOKS LIKE WE NOT ONLY HAD TO INTRODUCE AN APPROXIMATION, BUT WE ALSO INTRODUCED TWO FURTHER INTEGRALS THAT WE CANNOT SOLVE.

3. POPULATION FILTER APPROACH: MOTIVATION 1

MOTIVATION 1: WANT TO FIT POPULATION DISTRIBUTION DIRECTLY TO DATA.

PROBLEM: CANNOT COMPUTE POPULATION DISTRIBUTION.

$\longrightarrow$ **APPROXIMATE POPULATION DISTRIBUTION, E.G. WITH A NORMAL DISTRIBUTION!**

$$p(y \,|\, \theta, t) = \int \mathrm{d}\psi \, p(y \,|\, \psi, t) \, p(\psi \,|\, \theta) \approx \mathcal{N}(y \,|\, \mu(\theta, t), \sigma^2(\theta, t))$$

WHERE

$$\mu(\theta, t) = \mathbb{E}[y \,|\, \theta, t] \quad \text{and} \quad \sigma^2(\theta, t) = \mathrm{Var}[y \,|\, \theta, t]$$

BUT WE CAN USE SAMPLING TO ESTIMATE THEM: $y_i(\theta, t_j) \sim p(\,\cdot\,|\, \psi_i, t_j), \quad \psi_i \sim p(\,\cdot\,|\, \theta)$

$$\hat{\mu}(\theta, t_j) = \frac{1}{N_s} \sum_{i=1}^{N_s} y_i(\theta, t_j) \quad \text{and} \quad \hat{\sigma}^2(\theta, t_j) = \frac{1}{N_s - 1} \sum_{i=1}^{N_s} \Big(y_i(\theta, t_j) - \hat{\mu}(\theta, t_j) \Big)^2$$

3. POPULATION FILTER APPROACH: MOTIVATION 1

WE CAN NOW CONSTRUCT AN APPROXIMATE POPULATION POSTERIOR:

APPROXIMATE LOG-LIKELIHOOD:

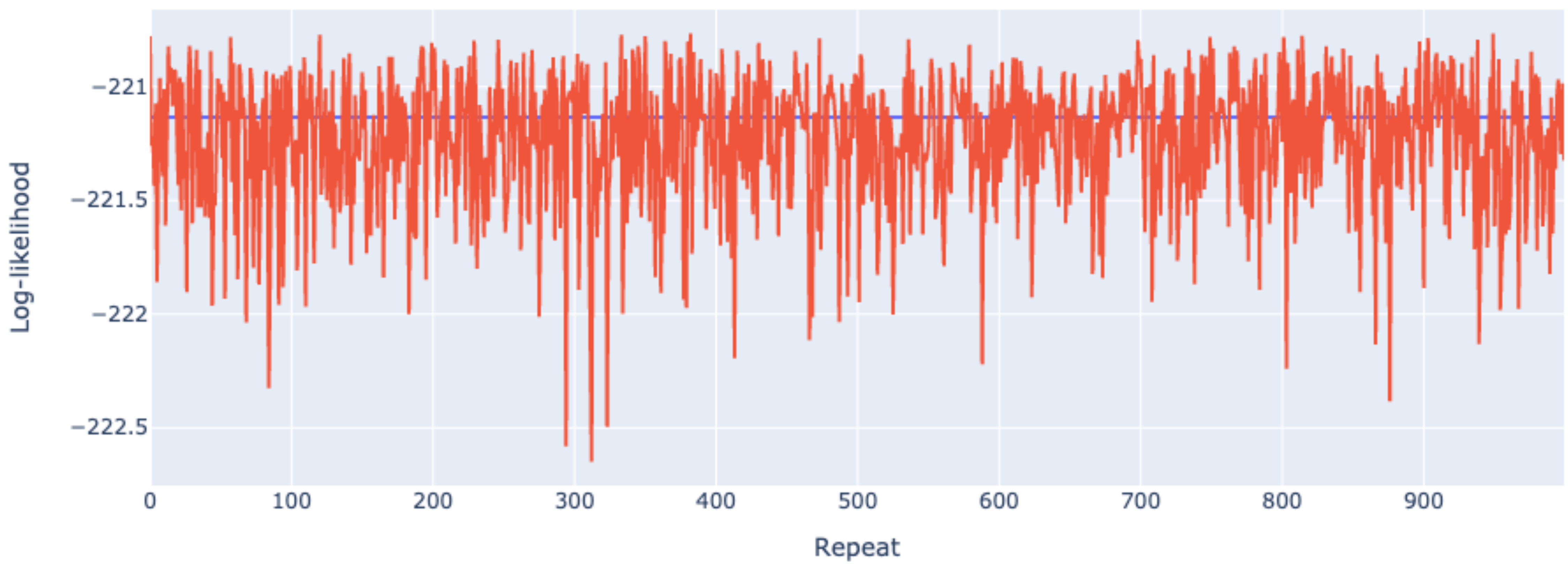
$$\log p_{\mathrm{KDE}}(\mathscr{D} \mid \theta) = \sum_j \log p_n(y_j^{\mathrm{obs}} \mid \hat{\mu}(\theta, t_j), \hat{\sigma}^2(\theta, t_j)) \qquad \text{for data} \qquad \mathscr{D} = \{(y_j^{\mathrm{obs}}, t_j)\}$$

POSTERIOR:

$$p_{\mathrm{KDE}}(\theta \mid \mathscr{D}) \propto p_{\mathrm{KDE}}(\mathscr{D} \mid \theta)\, p(\theta)$$

THAT'S GREAT! – EXCEPT IT DOESN'T WORK…

3. POPULATION FILTER APPROACH: MOTIVATION 1



THE LOG-LIKELIHOOD IS NOW A STOCHASTIC FUNCTION OF THE POPULATION PARAMETERS.

THIS BREAKS ADAPTIVE AND GRADIENT-BASED MCMC ALGORITHMS (LIKE ACMC, NUTS)!

⟶ WE NEED TO CONTROL THE STOCHASTICITY WHILST SAMPLING.

3. POPULATION FILTER APPROACH: MOTIVATION 1

LET'S MAKE THE REALISATIONS OF THE RANDOM SAMPLES KNOWN TO THE SAMPLER:

$$\hat{\mu}(\{\epsilon_{ij}\}, \{\psi_i\}, t_j) = \frac{1}{N_s} \sum_{i=1}^{N_s} y(\epsilon_{ij}, \psi_i, t_j), \quad \text{where} \quad \begin{aligned} y(\epsilon_{ij}, \psi_i, t_j) &= \bar{y}(\psi_i, t_j) + \epsilon_{ij}, \\ \psi_i &\sim p(\,\cdot\,|\,\theta), \\ \epsilon_{ij} &\sim p(\,\cdot\,|\,\sigma) \end{aligned}$$

$$\hat{\sigma}^2(\{\epsilon_{ij}\}, \{\psi_i\}, t_j) = \frac{1}{N_s - 1} \sum_{i=1}^{N_s} \Big(y(\epsilon_{ij}, \psi_i, t_j) - \hat{\mu}(\{\epsilon_{ij}\}, \{\psi_i\}, t_j) \Big)^2$$

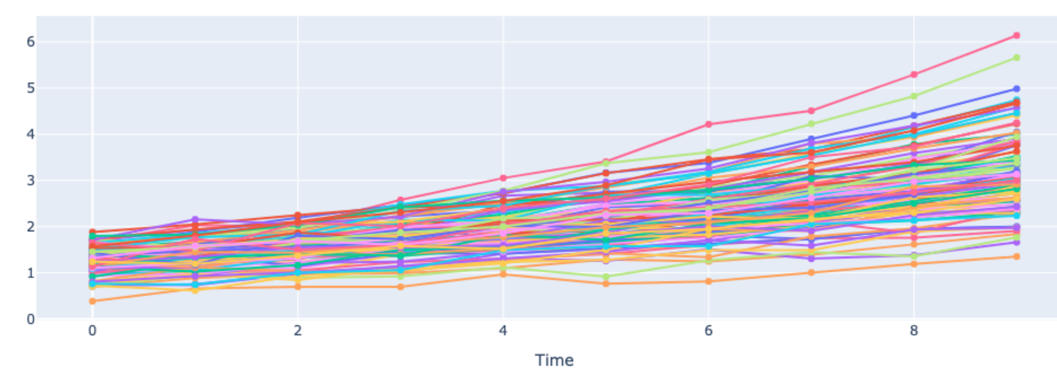
THIS RESULTS IN A JOINT POSTERIOR FOR THE POPULATION PARAMETERS AND THE AUXILIARY SAMPLES

$$\log p_{\text{KDE}}(\mathscr{D}, \{\epsilon_{ij}\}, \{\psi_i\}\,|\,\theta) = \sum_{ij} \log \mathscr{N}(y_j^{\text{obs}}\,|\,\{\epsilon_{ij}\}, \{\psi_i\}, t_j) + \log p(\psi_i\,|\,\theta) + \log p(\epsilon_{ij}\,|\,\sigma)$$

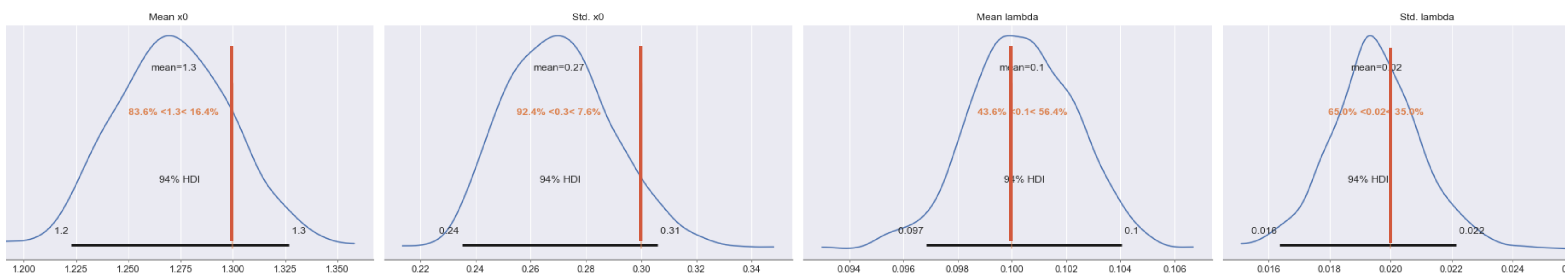
$$p_{\text{KDE}}(\theta, \{\epsilon_{ij}\}, \{\psi_i\}\,|\,\mathscr{D}) \propto p_{\text{KDE}}(\mathscr{D}, \{\epsilon_{ij}\}, \{\psi_i\}\,|\,\theta)\, p(\theta)$$

3. POPULATION FILTER APPROACH IN PRACTICE: EXAMPLE 1

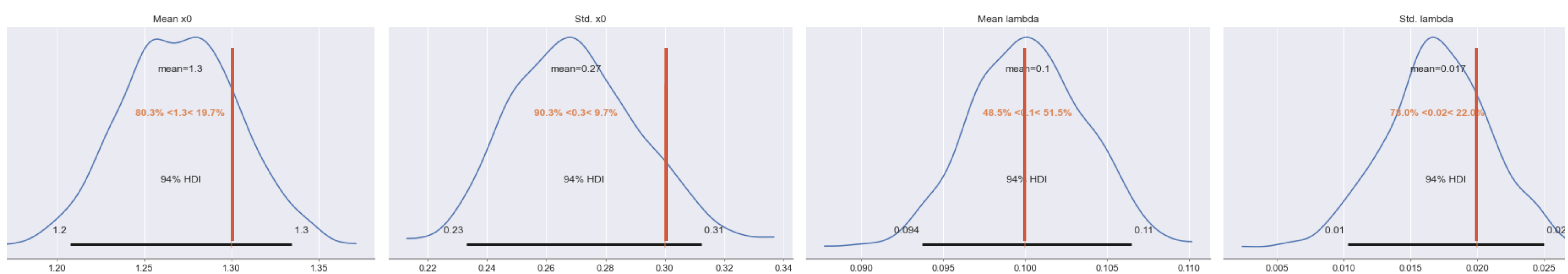
DATA: 100 INDIVIDUALS WITH 10 OBSERVATIONS



INFERENCE RESULTS FROM HIERARCHICAL INFERENCE (SLIDE 9): RUN TIME ~ 13 MINUTES

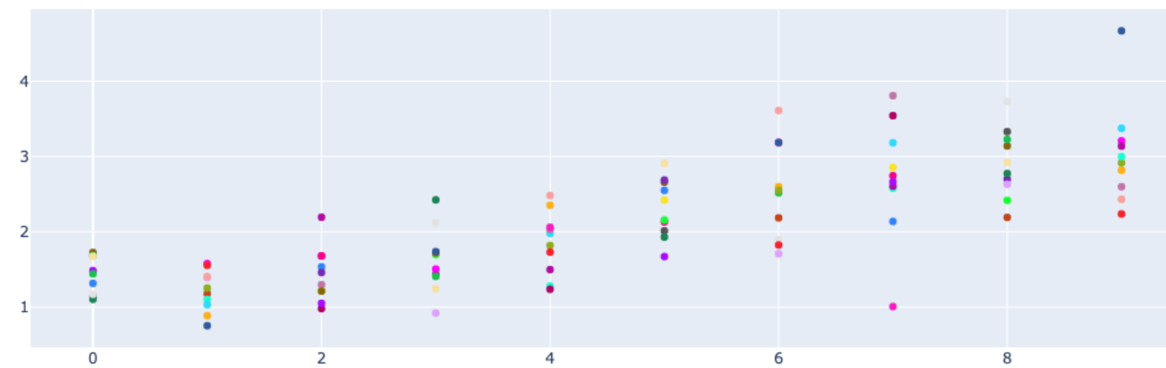


INFERENCE RESULTS FROM POPULATION KDE INFERENCE: RUN TIME ~ 13 MINUTES

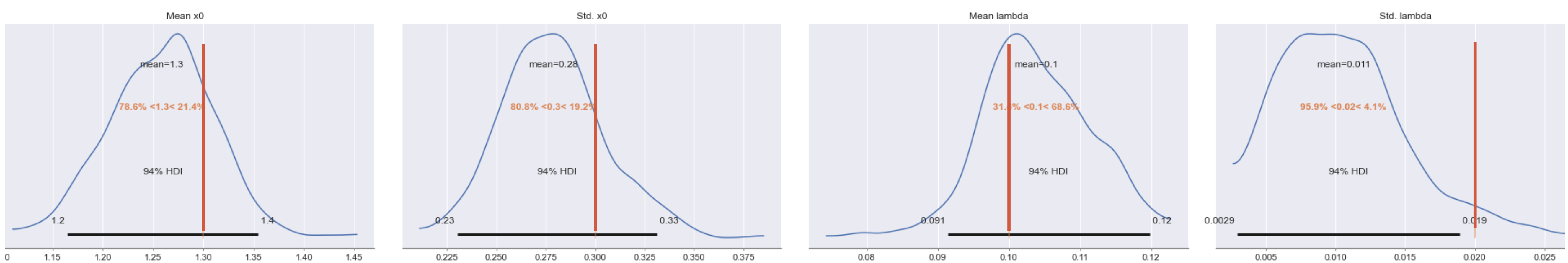


3. POPULATION FILTER APPROACH IN PRACTICE: EXAMPLE 1

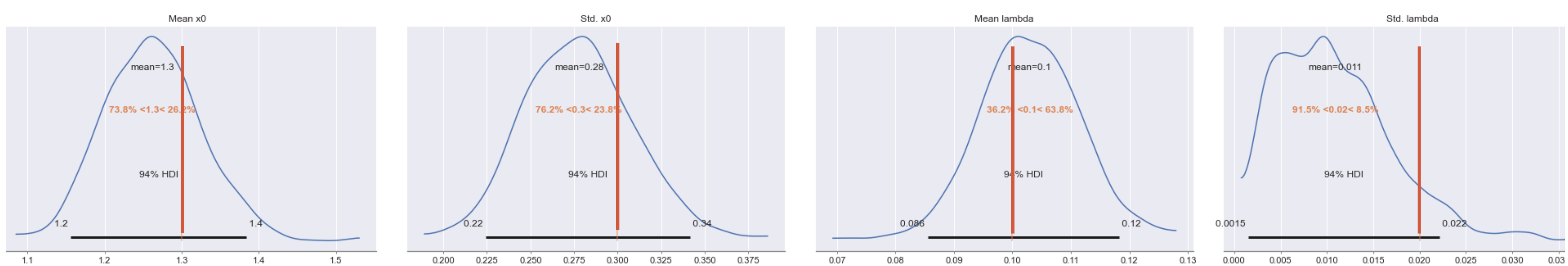
DATA: 100 INDIVIDUALS WITH 1 OBSERVATION EACH – SNAPSHOT DATA



INFERENCE RESULTS FROM HIERARCHICAL INFERENCE (SLIDE 10): RUN TIME ~ 23 MINUTES

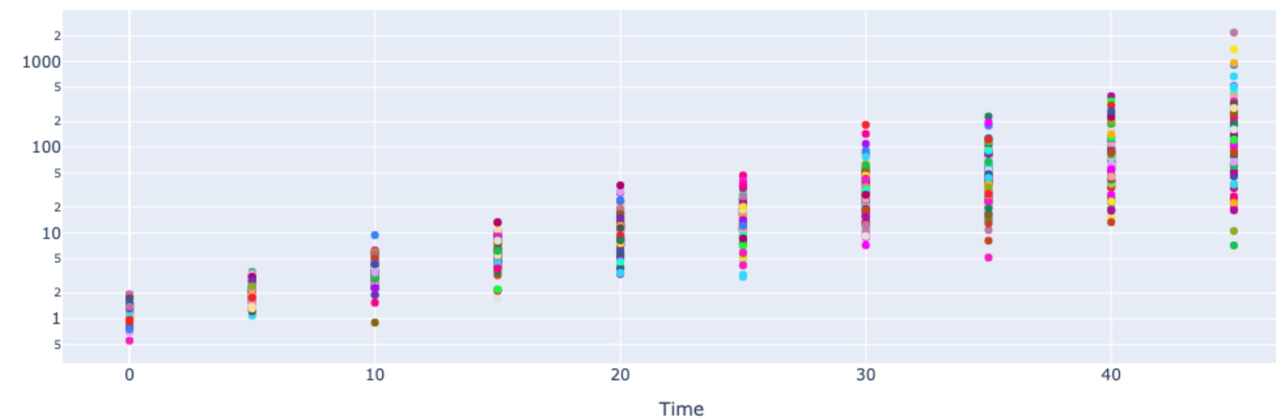


INFERENCE RESULTS FROM POPULATION KDE INFERENCE: RUN TIME ~ 8 MINUTES



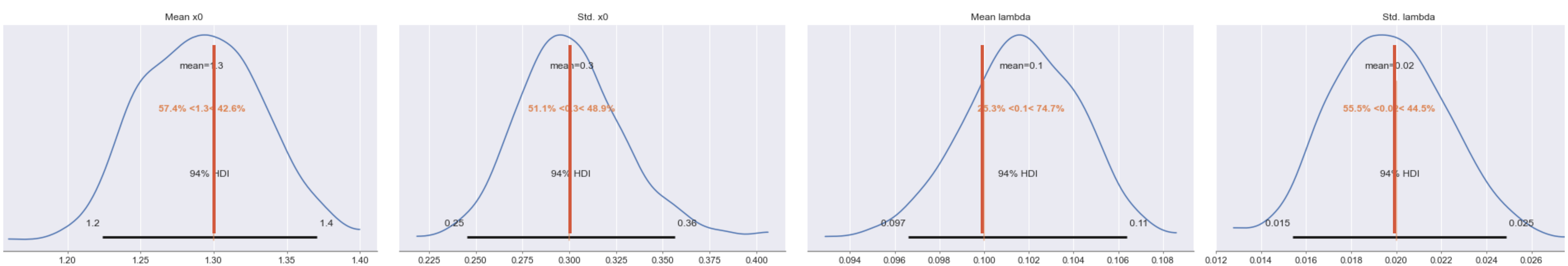
3. POPULATION FILTER APPROACH IN PRACTICE: EXAMPLE 1

DATA: 1000 INDIVIDUALS WITH 1 OBSERVATION EACH – SNAPSHOT DATA



INFERENCE RESULTS FROM HIERARCHICAL INFERENCE (SLIDE 11): RUN TIME ~ DAYS

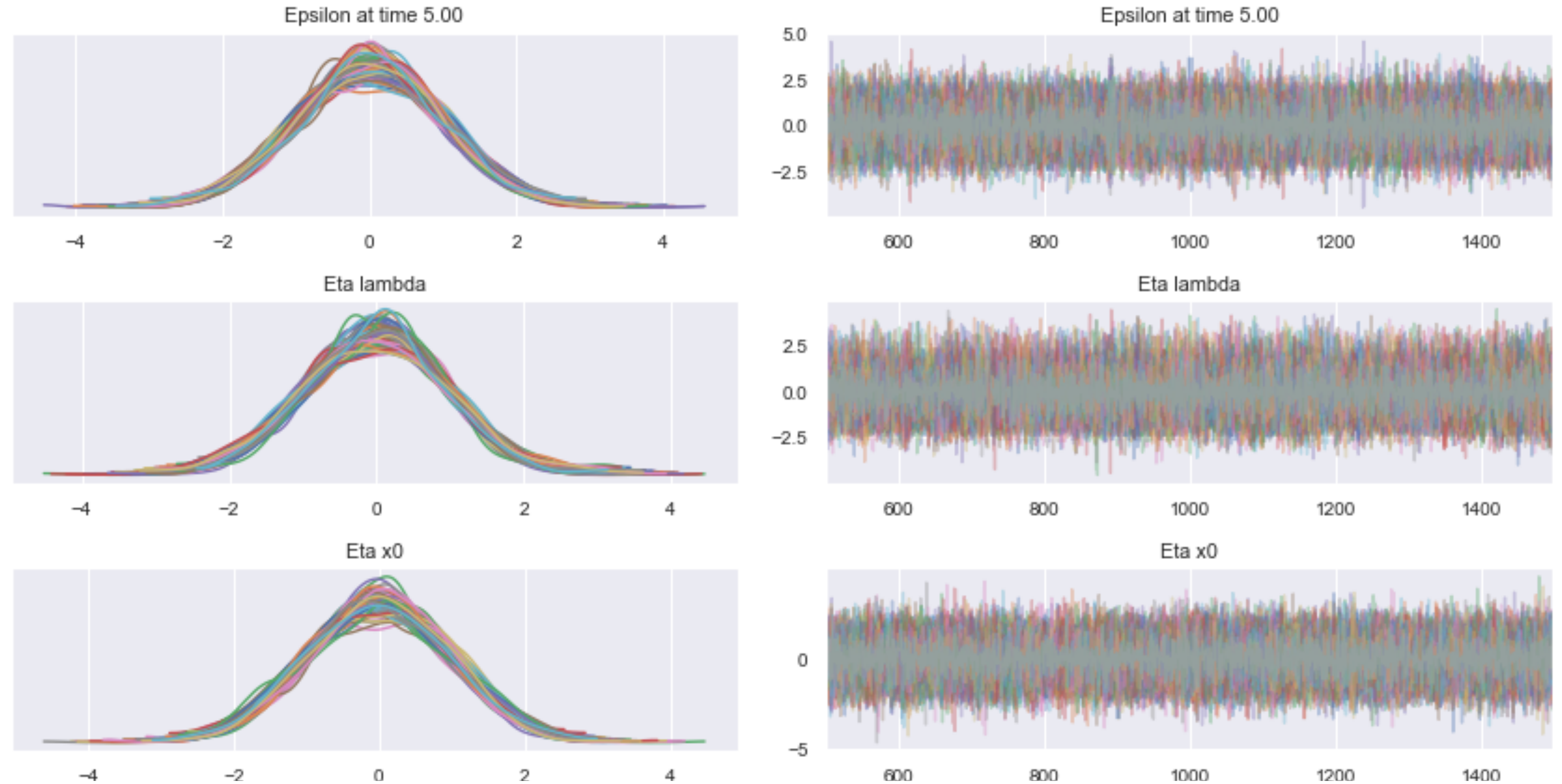
INFERENCE RESULTS FROM POPULATION KDE INFERENCE: RUN TIME ~ 30 MINUTES



→ NO SIGNS OF SHRINKAGE AND SIGNIFICANTLY SUPERIOR EXECUTION TIME.

WHY DOES THIS SCALE SO WELL?

A RELATIVELY SMALL NUMBER OF SAMPLES FROM THE POPULATION DISTRIBUTION ARE SUFFICIENT TO APPROXIMATE THE POPULATION DISTRIBUTION WELL (~ 100 SAMPLES), BECAUSE THE MCMC SAMPLER EFFECTIVELY INTEGRATES OVER POPULATION DISTRIBUTION FOR US:



→ DIMENSIONALITY OF POSTERIOR DOES NOT SCALE WITH NUMBER OF OBSERVED INDIVIDUALS.

OPEN QUESTIONS

1. HOW OFTEN IS THE ASSUMPTION OF A GAUSSIAN POPULATION DISTRIBUTION SUFFICIENT FOR GOOD INFERENCE RESULTS?

2. HOW EASY CAN WE FIND GOOD KDE APPROXIMATIONS WHEN A GAUSSIAN KDE ISN'T A GOOD CHOICE?

3. IS IT STRAIGHTFORWARD TO VALIDATE WHETHER A GIVEN CHOICE OF POPULATION KDE IS SUFFICIENT?

4. IF A GOOD KDE CANNOT BE FOUND, HOW WELL DOES A MOMENT BASED APPROACH WORK THAT DOESN'T MAKE ASSUMPTIONS ABOUT THE DISTRIBUTIONAL SHAPE (LIKELY NEEDS AT LEAST 100 OBSERVATIONS PER TIME POINT TO BE ABLE TO GO BEYOND 2. MOMENT).

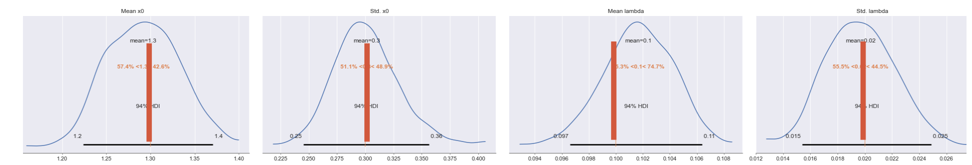
ADDRESSING Q1:

HOW OFTEN IS THE ASSUMPTION OF A GAUSSIAN POPULATION DISTRIBUTION SUFFICIENT FOR GOOD INFERENCE RESULTS?

HOW JUSTIFIED IS GAUSSIAN POPULATION KDE?

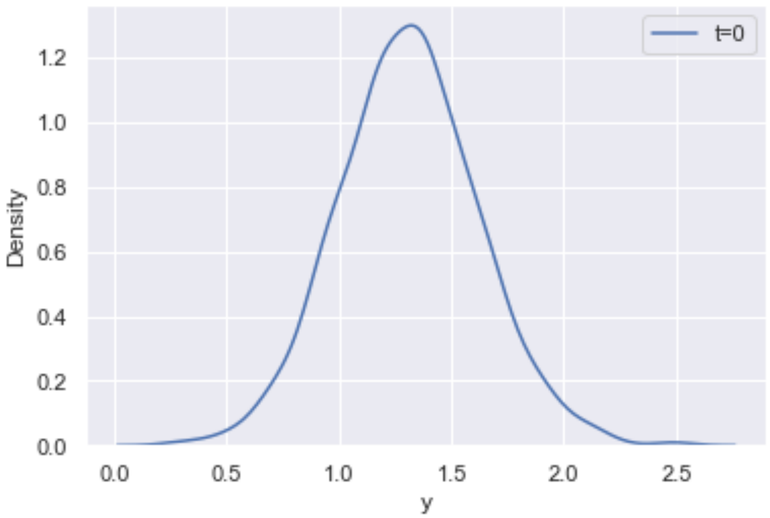
Hierarchical exp. growth model

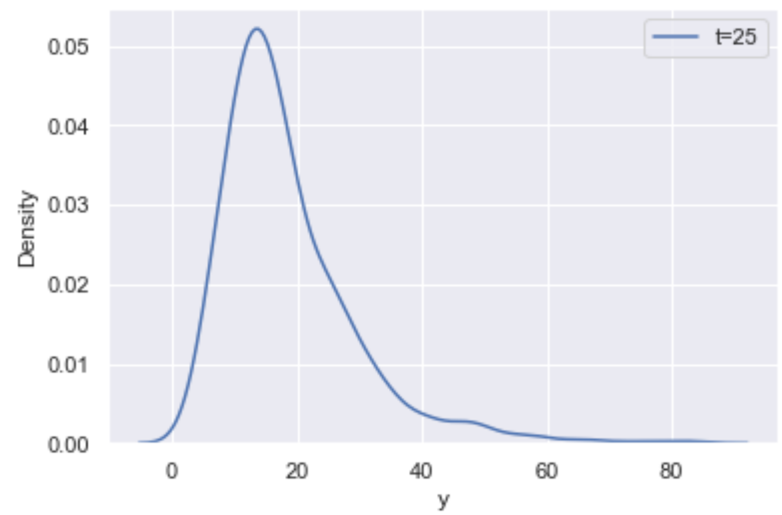
RECALL POSTERIORS FROM SLIDE 22:

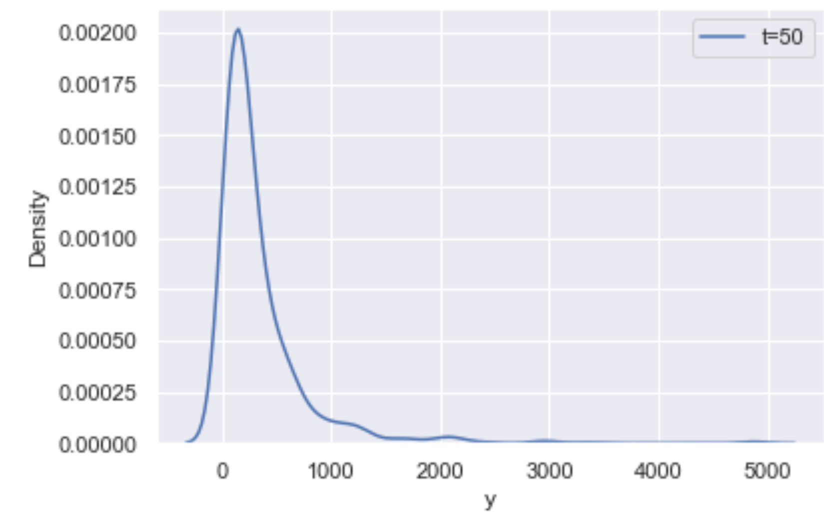


TIME SLICES OF DATA-GENERATING MODEL:

GAUSSIAN → **LOG-NORMAL (?)**







t

→ **GAUSSIAN POPULATION KDE APPROXIMATION IS NOT VERY ACCURATE.**

BUT: INFERENCE RESULTS ARE NEVERTHELESS GOOD. WHY?

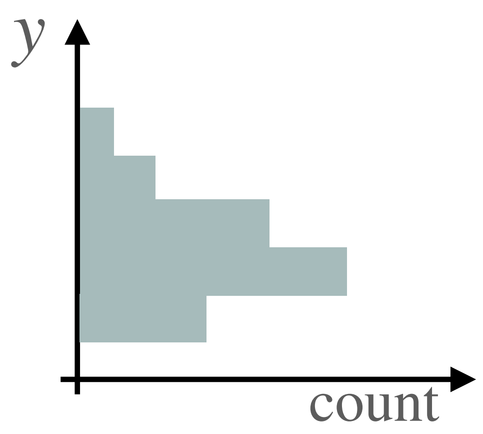
SHORT ANSWER Q1:

GAUSSIAN POPULATION KDE IS USED AS A 'FILTER' FOR BOTH THE OBSERVED DATA AND THE SAMPLES FOR A PROPOSED SET OF POPULATION PARAMETERS. THE POPULATION PARAMETERS ARE ACCEPTED RELATIVE TO THE LIKELIHOOD OF THE SAMPLES AND THE CONSISTENCY OF THE SAMPLES AND THE DATA WITH RESPECT TO THE SAME FILTER. SO, SUBJECT TO THE APPROPRIATENESS OF THE FILTER, THE INFERENCE RESULTS WILL BE GOOD.

ANSWER Q1: ABC REJECTION SAMPLING

LET'S THINK ABOUT ABC REJECTION SAMPLING:

1. **DATA IS SUMMARISED BY SUMMARY STATISTICS, E.G. MEAN AND VARIANCE (FILTERS).**



$$\hat{\mu}_{\text{obs}}, \hat{\sigma}^2_{\text{obs}}$$

2. **PROPOSE PARAMETERS, SAMPLE FROM CANDIDATE MODEL AND COMPUTE SUMMARY STATISTICS.**

$$\theta \quad \longrightarrow \quad y \sim p(\,\cdot\,|\,\theta) \quad \longrightarrow \quad \hat{\mu}, \hat{\sigma}^2$$

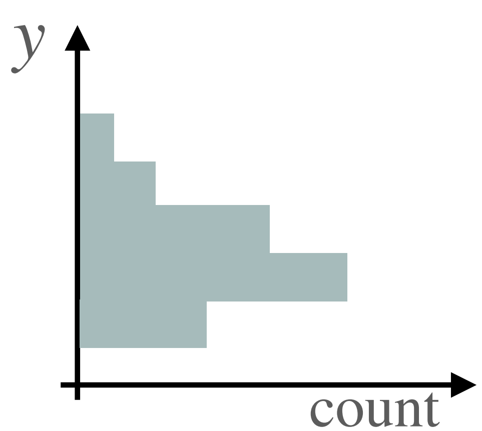
3. **ACCEPT IF SUMMARY STATISTICS ARE CLOSE TO EACH OTHER, ELSE REJECT.**

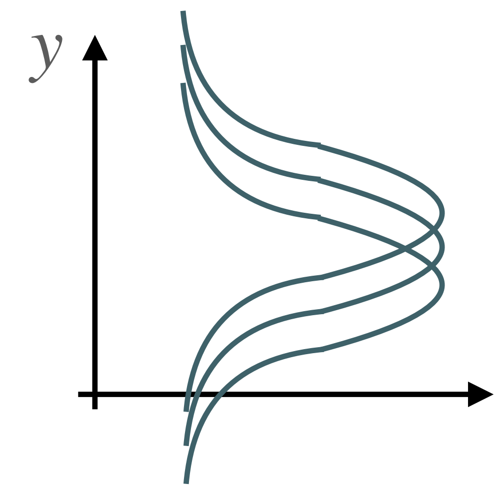
$$|\,\hat{\mu}_{\text{obs}} - \hat{\mu}\,| < \epsilon \quad \text{and} \quad |\,\hat{\sigma}^2_{\text{obs}} - \hat{\sigma}^2\,| < \epsilon$$

ANSWER Q1: GAUSSIAN KERNEL FILTER

POPULATION FILTER AS 'CONTINUOUS ABC':

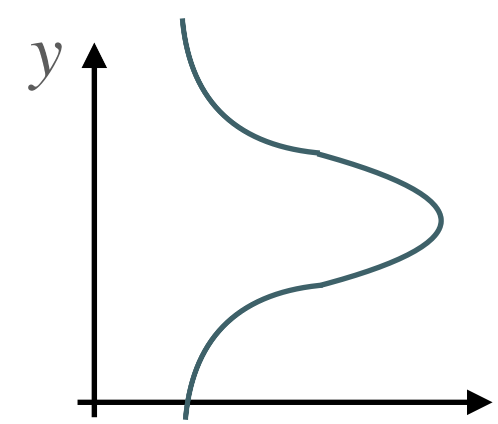
1. DATA IS SUMMARISED BY GAUSSIAN KERNELS WHICH ARE PROPOSED IN 2.:





2. PROPOSE PARAMETERS, SAMPLE FROM CANDIDATE MODEL AND ESTIMATE GAUSSIAN KERNEL:

θ $\qquad$ $y \sim p(\,\cdot\,|\,\theta)$



3. ACCEPT RELATIVE TO KERNEL LIKELIHOOD.

$$p_{\text{KDE}}(\theta, \{\epsilon_{ij}\}, \{\psi_i\}\,|\,\mathscr{D}) \propto p_{\text{KDE}}(\mathscr{D}, \{\epsilon_{ij}\}, \{\psi_i\}\,|\,\theta)\, p(\theta)$$

ANSWER Q1

1. GAUSSIAN POPULATION FILTER INFERENCE USES GAUSSIAN KERNELS AS FILTERS TO QUANTIFY THE CONSISTENCY OF THE OBSERVED DATA WITH THE PROPOSED SET OF POPULATION PARAMETERS.

   → INFERRED POSTERIORS ARE CORRECT, BUT MAY BE WIDER THAN NECESSARY SUBJECT TO THE INFORMATION LOSS INCURRED BY FILTERING.

2. CAVEATS IN PRACTICE: ONLY SAMPLES CONTRIBUTE TO FILTER CONSTRUCTION WHICH INTRODUCES AN ASYMMETRY.

   IF TRUE POPULATION DISTRIBUTION IS DIFFERENT FROM KERNEL:

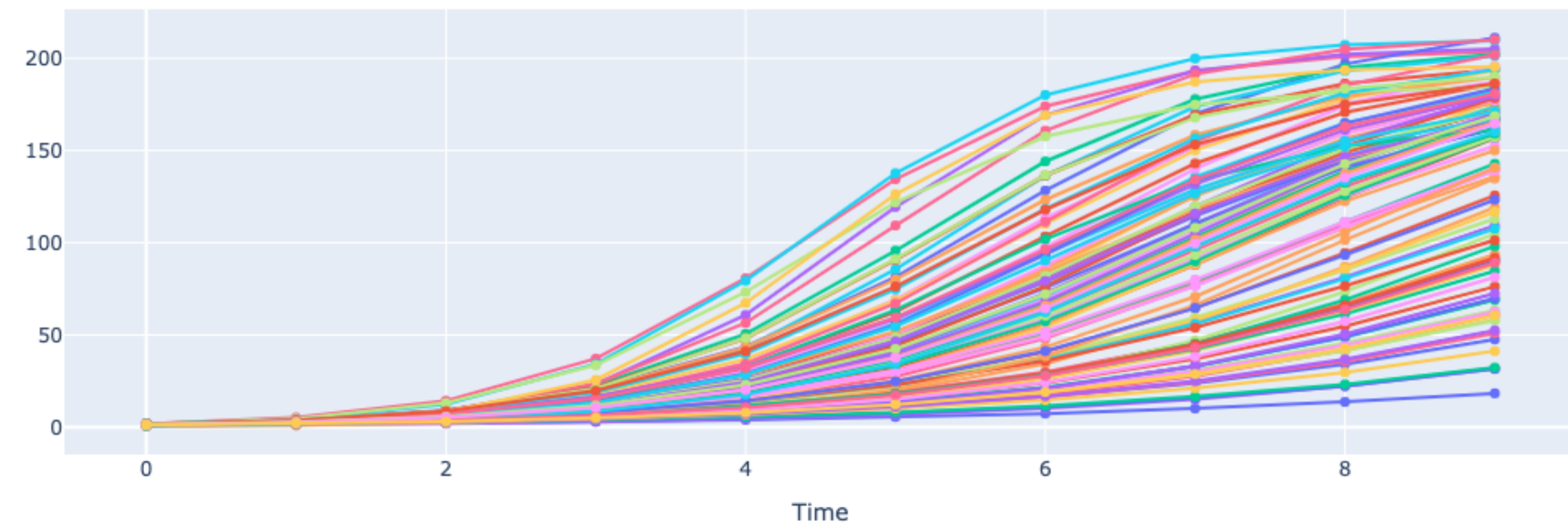
   i. FEW OBSERVATIONS, FEW SAMPLES: NO PROBLEM.
   ii. FEW OBSERVATIONS, MANY SAMPLES: PARAMETER UNCERTAINTY MAY BE UNDERESTIMATED.
   iii. MANY OBSERVATIONS, FEW SAMPLES: PARAMETER VARIANCE MAY BE OVERESTIMATED.
   iv. MANY OBSERVATIONS, MANY SAMPLES: NO PROBLEM.

   → IF POSSIBLE, RUN WITH AS MANY SAMPLES AS OBSERVATIONS, BUT NEVER MORE SAMPLES THAN OBSERVATIONS IF TRUE POPULATION DISTRIBUTION IS DIFFERENT FROM KERNEL.

3. POPULATION FILTER APPROACH IN PRACTICE: EXAMPLE 2

HIERARCHICAL LOGISTIC GROWTH MODEL:

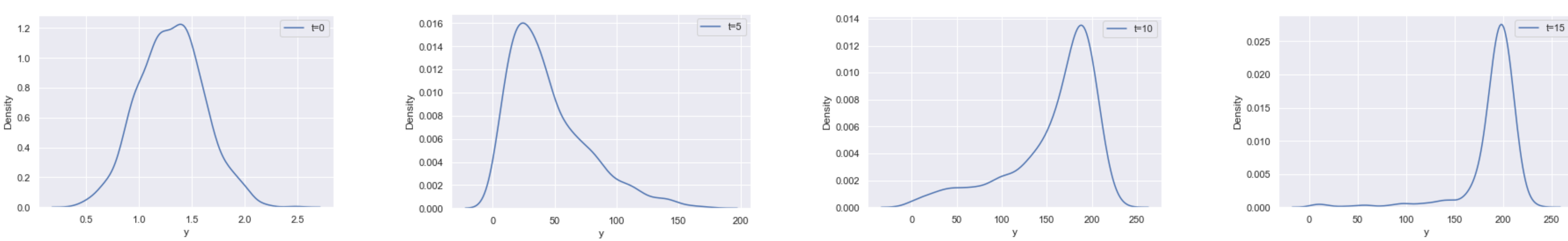
$$y = \frac{c\,\bar{y}_0}{\bar{y}_0 + (c - \bar{y}_0)\mathrm{e}^{-\lambda t}} + \epsilon$$



$$\bar{y}_0 \sim \mathcal{N}\left(\mu_{\bar{y}_0}, \sigma^2_{\bar{y}_0}\right)$$
$$\lambda \sim \mathcal{N}\left(\mu_{\lambda}, \sigma^2_{\lambda}\right)$$
$$c \sim \mathcal{N}\left(\mu_{c}, \sigma^2_{c}\right)$$
$$\epsilon \sim \mathcal{N}(0,\sigma^2)$$

TIME SLICES OF DATA-GENERATING MODEL:

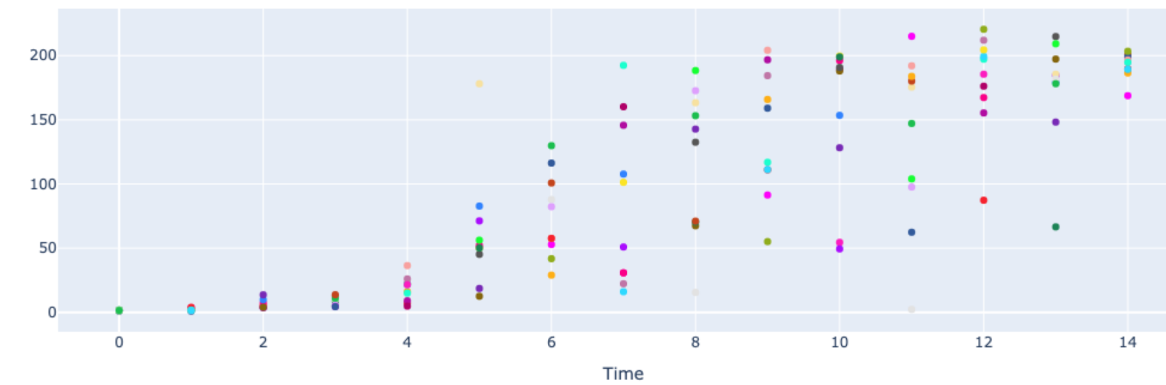
GAUSSIAN → LOG-NORMAL (?) → (?)



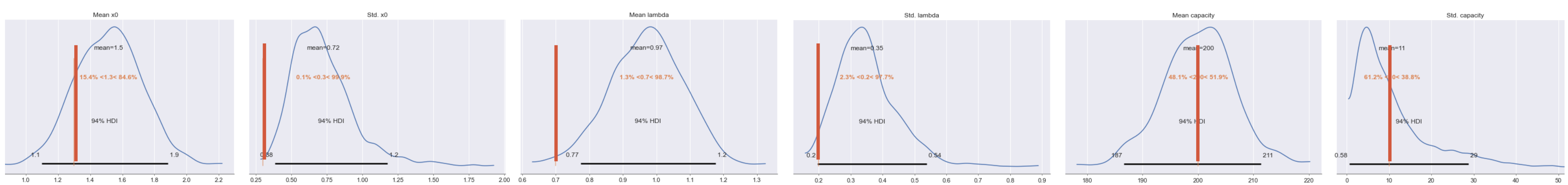
t

HIERARCHICAL LOGISTIC GROWTH: FEW OBSERVATIONS

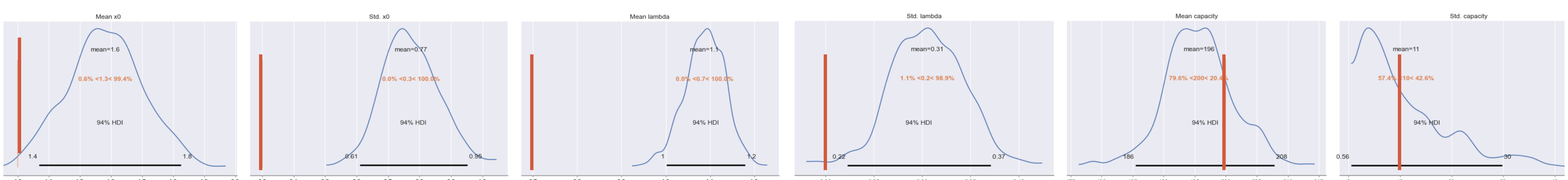
SNAPSHOT DATA: 10 INDIVIDUALS PER TIME POINT



10 SAMPLES:



100 SAMPLES:

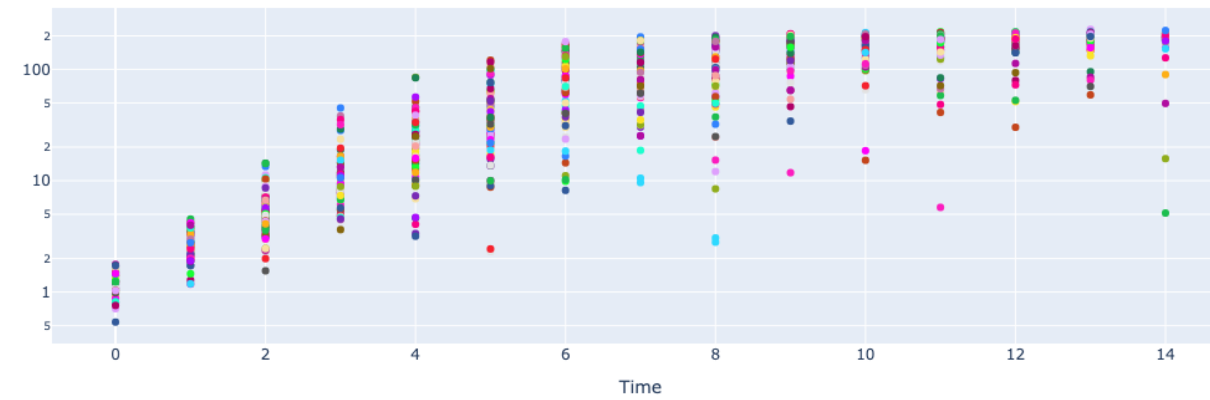


NUMBER OF SAMPLES

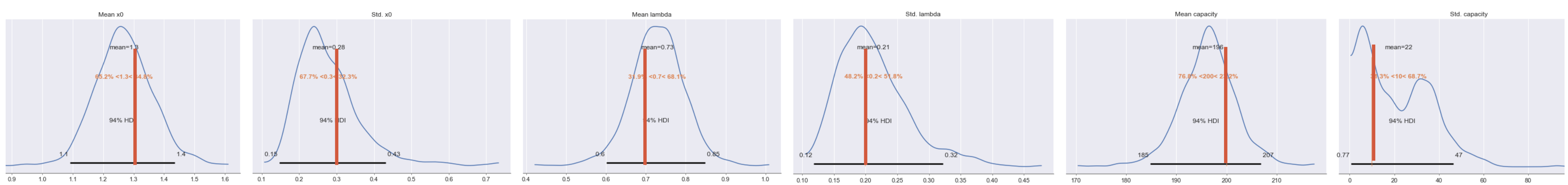
HEAVY TAILS OF TRUE DISTRIBUTION AND TOO LARGE SAMPLES-OBSERVATION RATIO LEAD TO UNDERESTIMATION OF UNCERTIANTY.

HIERARCHICAL LOGISTIC GROWTH: MORE OBSERVATIONS

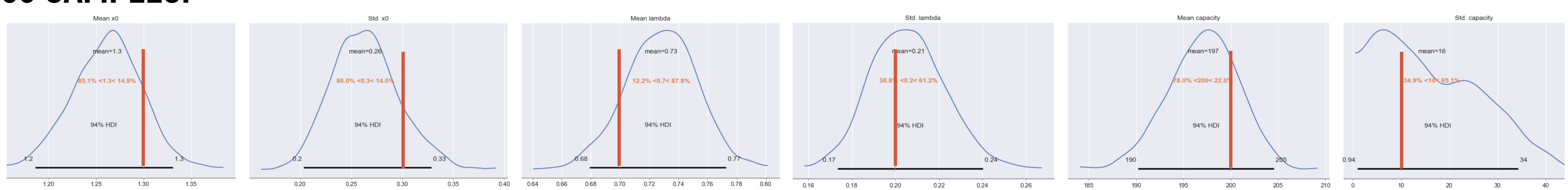
SNAPSHOT DATA: 100 INDIVIDUALS PER TIME POINT



10 SAMPLES:



100 SAMPLES:



NUMBER OF SAMPLES

HEAVY TAILS OF TRUE DISTRIBUTION AND TOO SMALL SAMPLES-OBSERVATION RATIO LEAD TO OVERESTIMATION OF UNCERTAINTY / SPURIOUS PEAKS IN POSTERIORS.