On Extension of Information Geometry of ‘Parameter Estimation’ to ‘State Estimation’

Rudolf Kulhavý

Honeywell Technology Center Europe &
Institute of Information Theory and Automation

General Regression

- Example [K.J. Åström]

$$y_k = (\mu + v_k)\, y_{k-1} + e_k$$

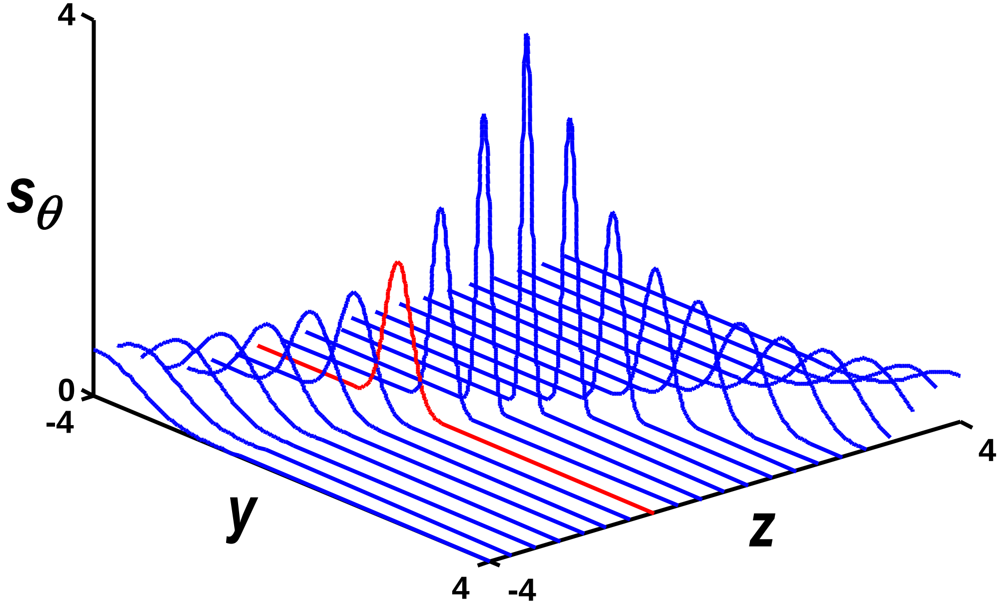
random fluctuation of AR(1) coefficient $= z_k$ *regressor*

- Assumptions
  - μ is constant
  - v_k is $N(0,\sigma_v^2)$ distributed
  - e_k is $N(0,\sigma_e^2)$ distributed

unknown parameters

$$\theta = (\mu, \sigma_e, \sigma_v)$$

Model Density

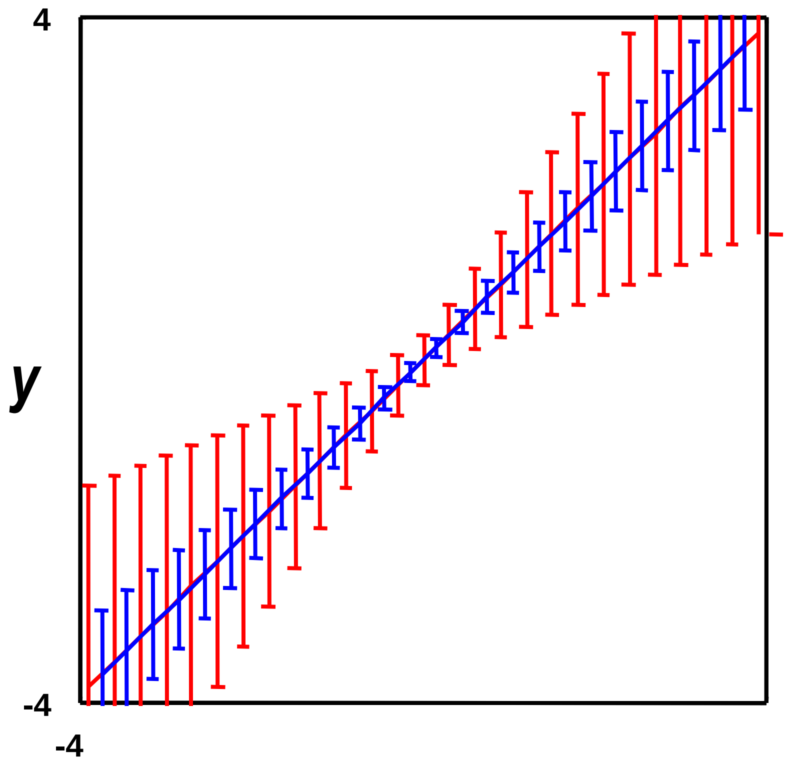


z-dependent variance

$$\sigma^2(z) = \sigma_e^2 + z^2\,\sigma_v^2$$

$$s_\theta(y \mid z) = \frac{1}{\sqrt{2\pi\sigma^2(z)}}\,\exp\left(-\frac{1}{2\sigma^2(z)}(y-\mu z)^2\right)$$

Model Density

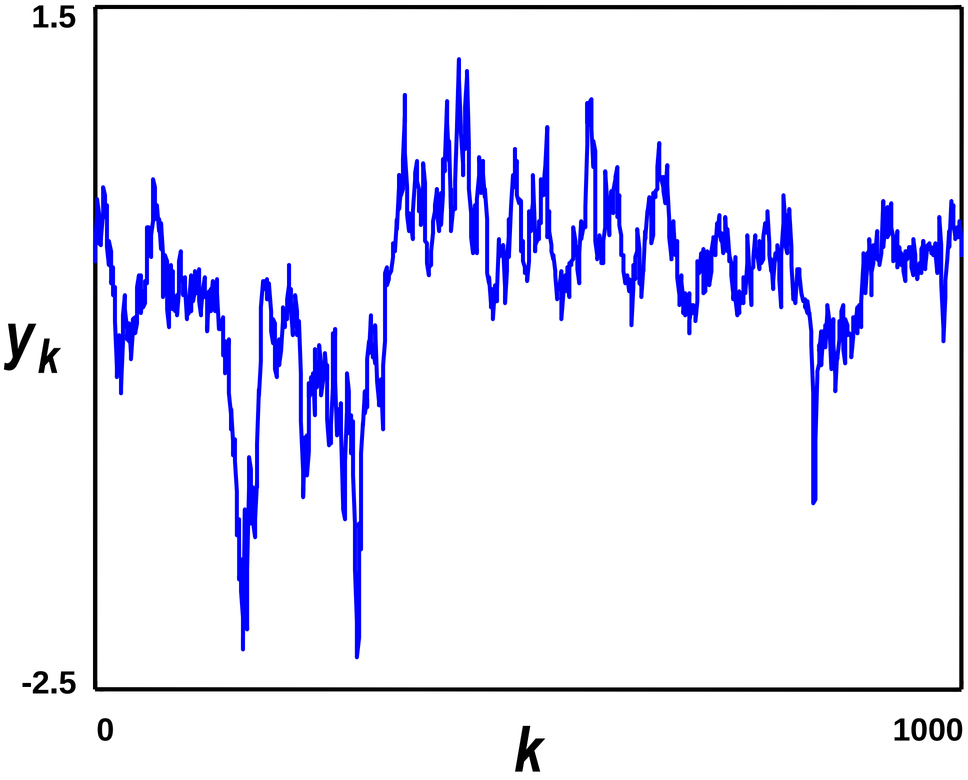


$$\mu z \pm \sqrt{\sigma_e^2 + z^2 \sigma_v^2}$$

$$\mu z \pm 3\sqrt{\sigma_e^2 + z^2 \sigma_v^2}$$

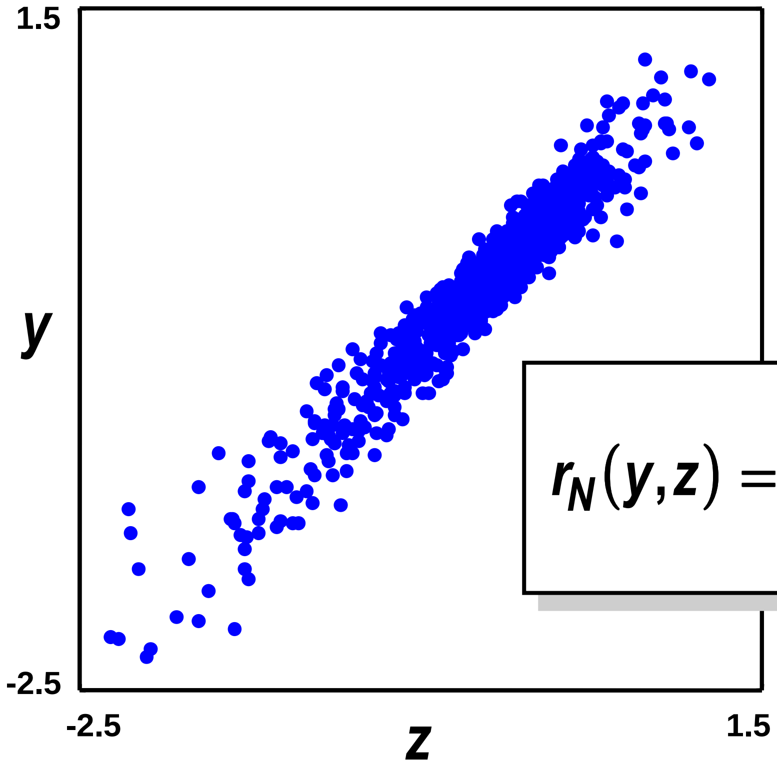
$$s_\theta(y \mid z) = \frac{1}{\sqrt{2\pi\sigma^2(z)}} \exp\left(-\frac{1}{2\sigma^2(z)} (y - \mu z)^2 \right)$$

Sample of Data



$$\mu = 0.98$$
$$\sigma_e = 0.1$$
$$\sigma_v = 0.2$$

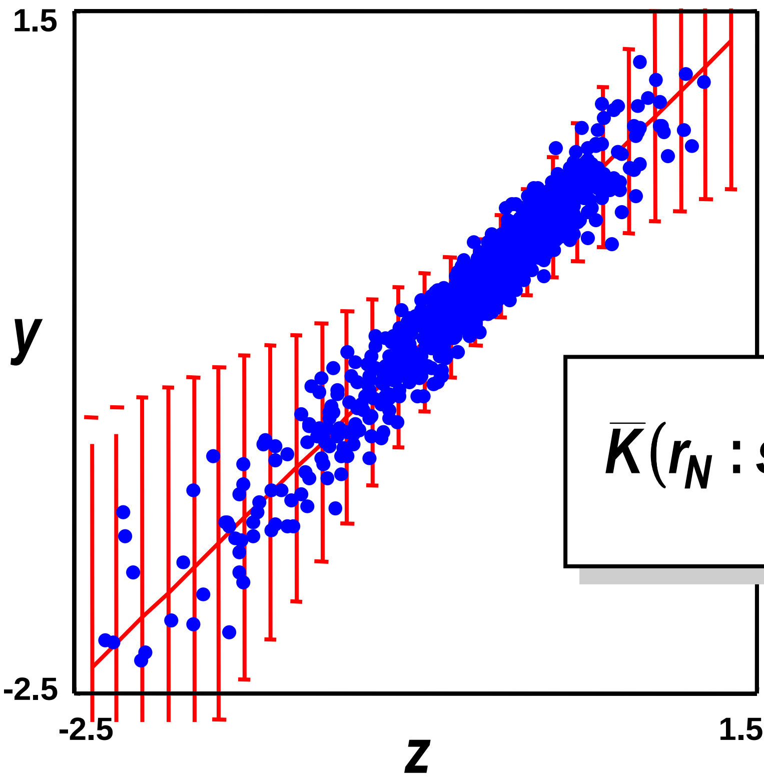
Empirical Density



Mixture of
Dirac functions

$$r_N(y,z) = \frac{1}{N}\sum_{k=m+1}^{N+m} \delta(y - y_k, z - z_k)$$

Probability Matching



Conditional inaccuracy

$$\overline{K}(r_N : s_\theta) = \iint r_N(y,z) \, \log \frac{1}{s_\theta(y \mid z)} \, dy \, dz$$

Inaccuracy

$$\bar{K}(r_N : s_\theta) = -\frac{1}{N} \log \prod_{k=m+1}^{N+m} s_\theta(y_k \mid z_k)$$

Likelihood

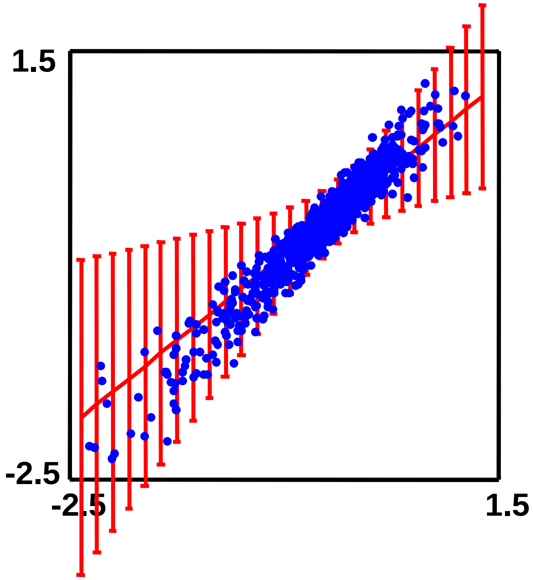
$$l_N(\theta) = c \exp\left(-N \bar{K}(r_N : s_\theta)\right)$$

Posterior

$$p_N(\theta) = c\, p_0(\theta) \exp\left(-N \bar{K}(r_N : s_\theta)\right)$$

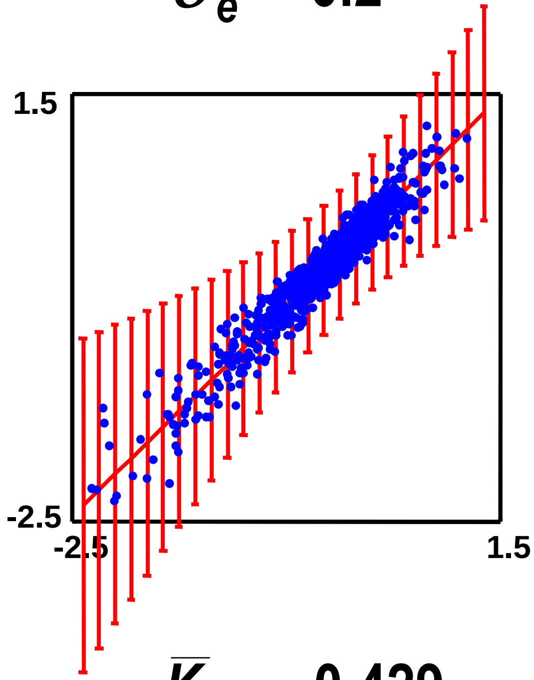
Estimation via Inaccuracy

$\mu = 0.8$



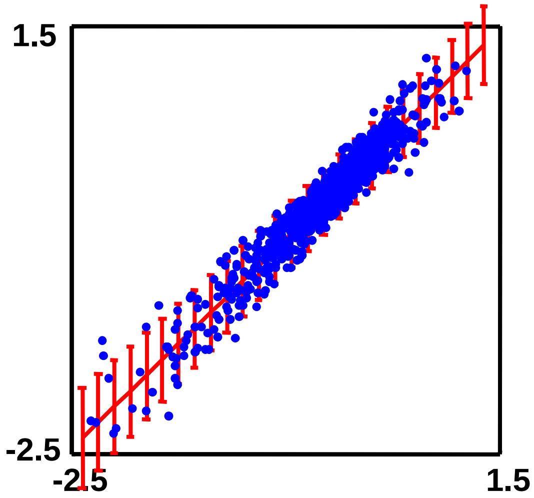
$\bar{K} = -0.552$

$\sigma_e = 0.2$



$\bar{K} = -0.429$

$\sigma_v = 0.05$



$\bar{K} = -0.517$

Ideal Compression of Data

- If $\{\log s_\theta(y \mid z)\}$ belongs to an affine space

$$\log s_\theta(y \mid z) = \log s_0(y \mid z) + \lambda^T(\theta)\, h(y,z) - \psi(\lambda(\theta)),$$

fixed $\neq f(z)$!

- then the inaccuracy

$$\bar{K}(r_N : s_\theta) = \bar{K}(r_N : s_0) - \lambda^T(\theta)\, \bar{h}_N + \psi(\lambda(\theta))$$

- depends on data only through the statistic

$$\bar{h}_N = \frac{1}{N} \sum_{k=m+1}^{N+m} h(y_k, z_k)\,.$$

➔ *The statistic carries all essential information.*

Construction of *h*-Statistic

Parameter grid
$\theta_1, \ldots, \theta_{n+1}$

- **Logarithm of density ratio**

$$h_i(y,z) = \log \frac{s_{\theta_i}(y \mid z)}{s_{\theta_{i+1}}(y \mid z)}, \quad i = 1, \ldots, n$$

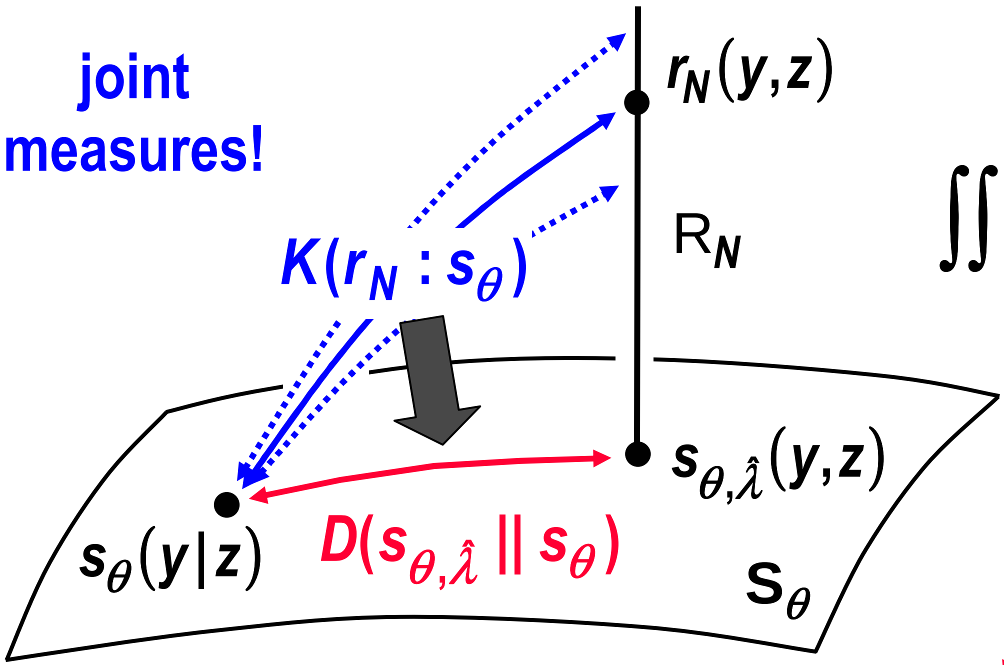
- **Inaccuracy difference**

$$\bar{h}_{i,N} = -\bar{K}(r_N : s_{\theta_i}) + \bar{K}(r_N : s_{\theta_{i+1}}), \quad i = 1, \ldots, n$$

- **Normalized logarithm of likelihood ratio**

$$\bar{h}_{i,N} = \frac{1}{N} \log \frac{l_N(\theta_i)}{l_N(\theta_{i+1})}, \quad i = 1, \ldots, n$$

Coping with Ambiguity



$$\iint r(y,z)\, h(y,z)\, dy\, dz = \bar{h}_N$$

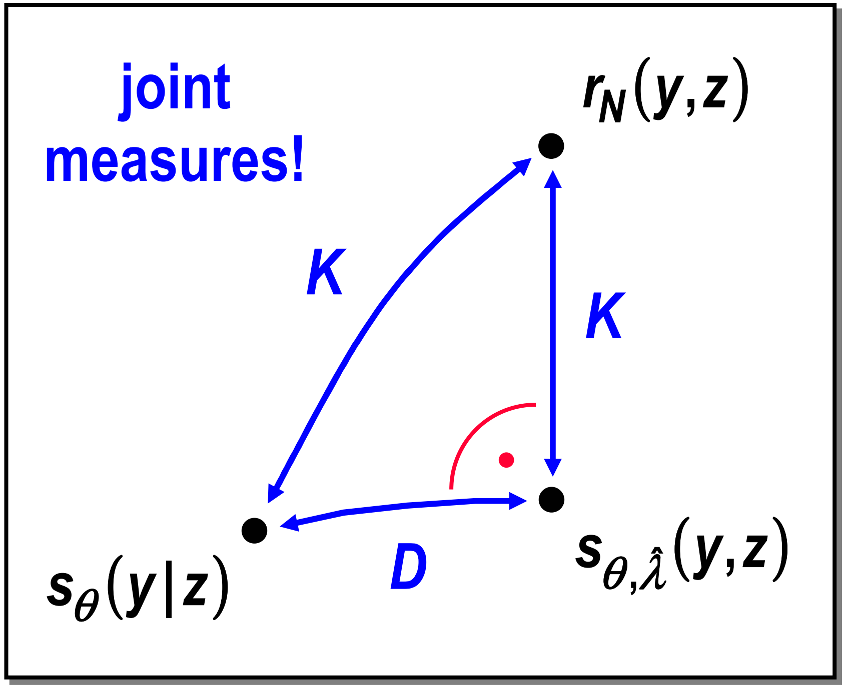
$$s_{\theta,\lambda}(y,z) = s_\theta(y|z)\exp(\lambda' h(y,z) - \psi(\theta,\lambda))$$

Pythagorean Relationship

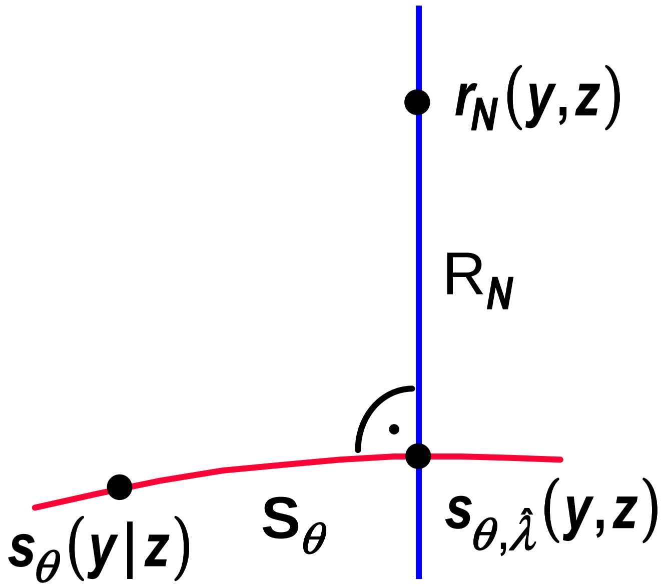
$$K(r_N : s_\theta) = K(r_N : s_{\theta,\hat{\lambda}}) + D(s_{\theta,\hat{\lambda}} \| s_\theta)$$

$$\min_{\lambda \in R^n} K(r_N : s_{\theta,\lambda})$$

$$\min_{r \in \mathrm{R}_N} D(r \| s_\theta)$$



MRE Approximation



$$D(\mathrm{R}_N \,\|\, s_\theta) = \min_{r \in \mathrm{R}_N} D(r \,\|\, s_\theta)$$

1. choose *h*-statistic so that

$$K(r_N : s_{\theta,\hat{\lambda}}) \;\approx\; \text{const.}$$

2. approximate inaccuracy

$$K(r_N : s_\theta) \;\approx\; D(\mathrm{R}_N \,\|\, s_\theta) + \text{const.}$$

3. approximate likelihood

$$\hat{l}_N(\theta) = c \exp\left(-N\, D(\mathrm{R}_N \,\|\, s_\theta)\right)$$

Lack of Normalization

Conditional!

$$D(r \,\|\, s) = \iint r(y,z) \log \frac{r(y,z)}{s(y\,|\,z)}\, dy\, dz$$

$$= \int r(z) \int r(y\,|\,z) \log \frac{r(y\,|\,z)}{s(y\,|\,z)}\, dy\, dz \;-\; \int r(z) \log \frac{1}{r(z)}\, dz$$

Conditional relative entropy

Marginal entropy

MRE Direct Computation

- Convex optimization problem

$$D(\mathrm{R}_N \,\|\, s_\theta) = \min_{r \in \mathrm{R}_N} D(r \,\|\, s_\theta)$$

$$= \max_{\lambda \in R^n} \left(\lambda' \overline{h}_N - \psi(\theta, \lambda) \right)$$

Statistic

Logarithm of normalizing divisor

- Entails multivariate integration

$$\psi(\theta, \lambda) = \log \iint s_\theta(y|z) \exp\left(\lambda' h(y,z)\right) \, dy \, dz$$

Extension to Filtering

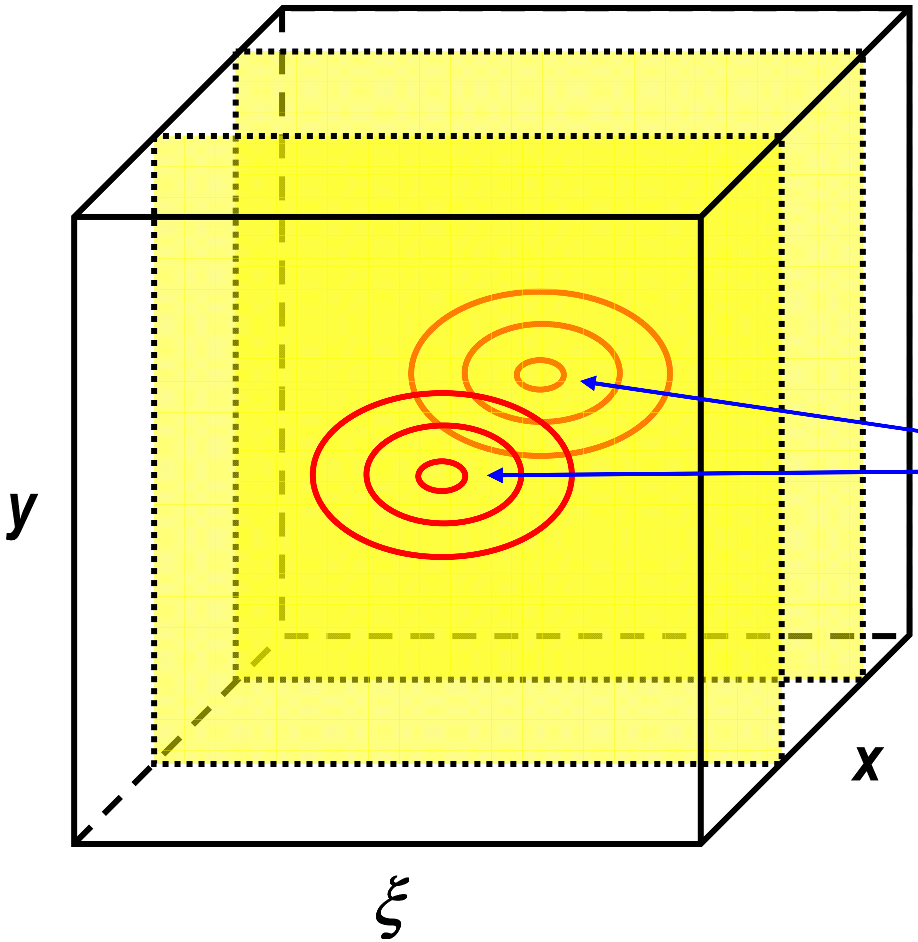
- State transition density

$q(x_{k+1} \mid x_k)$ if controlled: $q(x_{k+1} \mid x_k, u_k)$

- System output density

$s(y_k \mid x_k)$ if controlled: $s(y_k \mid x_k, u_k)$

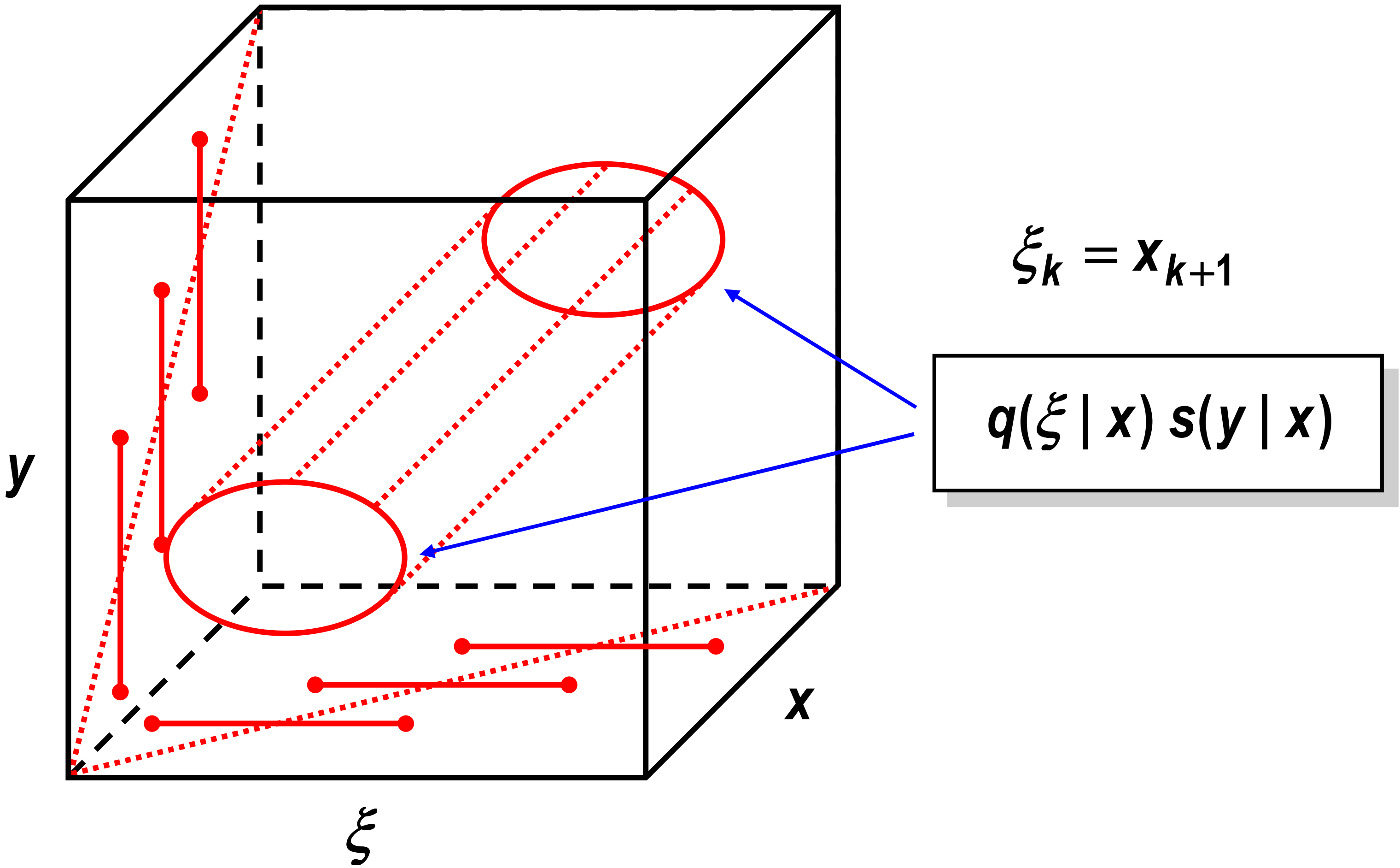
Model Density



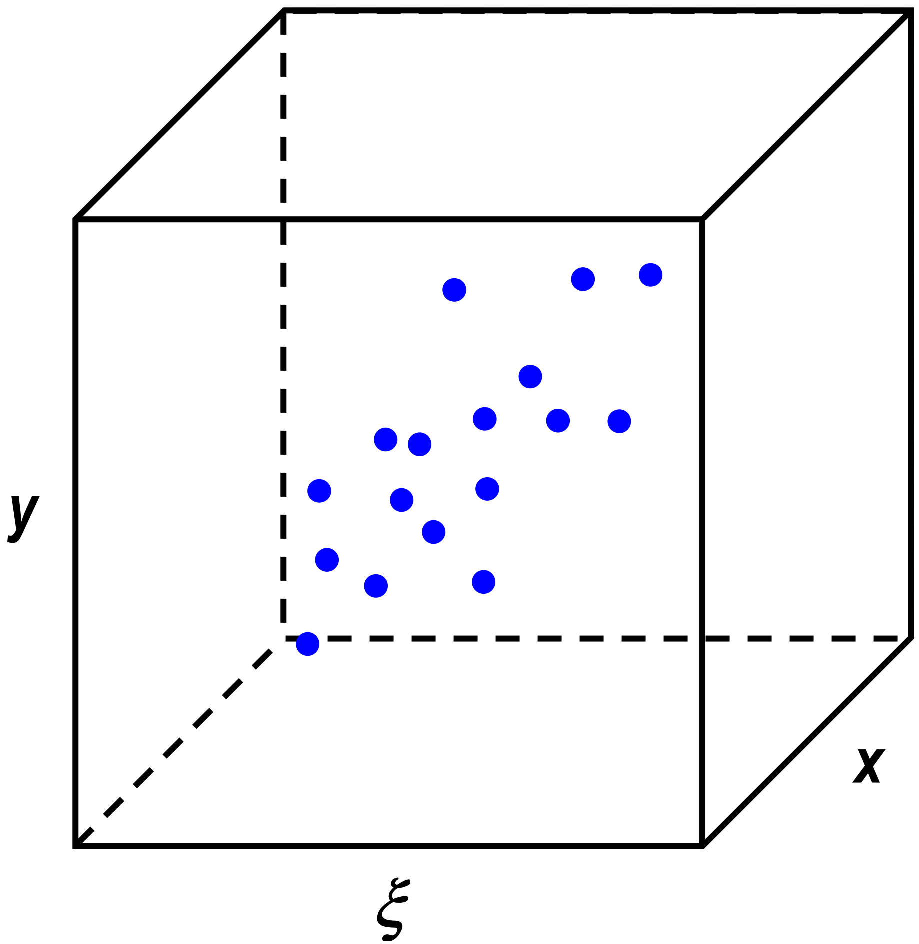
$\xi_k = x_{k+1}$

$q(\xi \mid x)\, s(y \mid x)$

Model Density



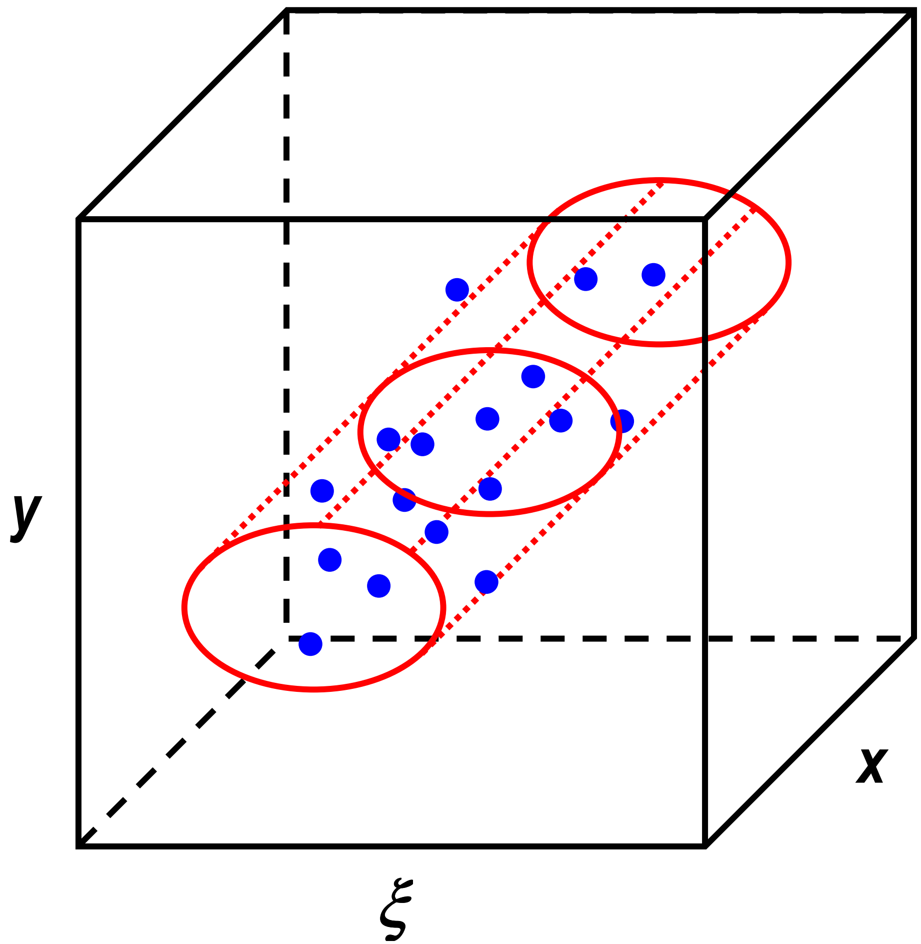
Empirical Density



Mixture of Dirac functions

$$r_N(\xi,x,y)=\frac{1}{N}\sum_{k=1}^{N}\delta(\xi-\xi_k,x-x_k,y-y_k)$$

Probability Matching



Conditional inaccuracy

$$\bar{K}(r_N : qs) = \iiint r_N(\xi, x, y) \ \log \frac{1}{q(\xi \mid x)\, s(y \mid x)} \, d\xi \, dx \, dy$$

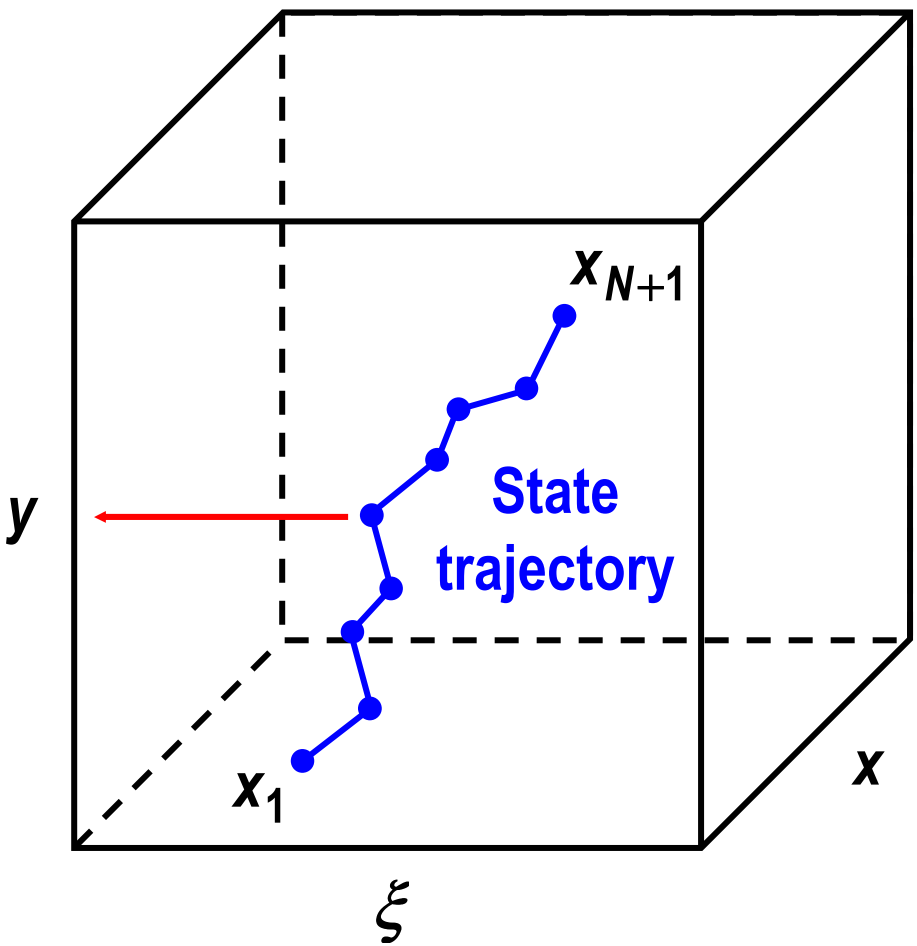
Inaccuracy

$$\bar{K}(r_N : qs) = -\frac{1}{N}\log\prod_{k=1}^{N} q(x_{k+1} \mid x_k)\, s(y_k \mid x_k)$$

Posterior

$$p_N(x_1, \ldots, x_{N+1}) = c\, p_0(x_1)\exp\left(-N\,\bar{K}(r_N : qs)\right)$$

Estimation of x_{N+1}



- Initial state fixed
- Terminal state fixed
- Measurements known

- Uncertainty $r_N \in \mathbf{R}_N$

$$\overline{K}(r_N : qs) \rightarrow D(\mathbf{R}_N \,\|\, qs)$$