A Geometric Approach to Statistical Estimation

Rudolf Kulhavý

Institute of Information Theory and Automation

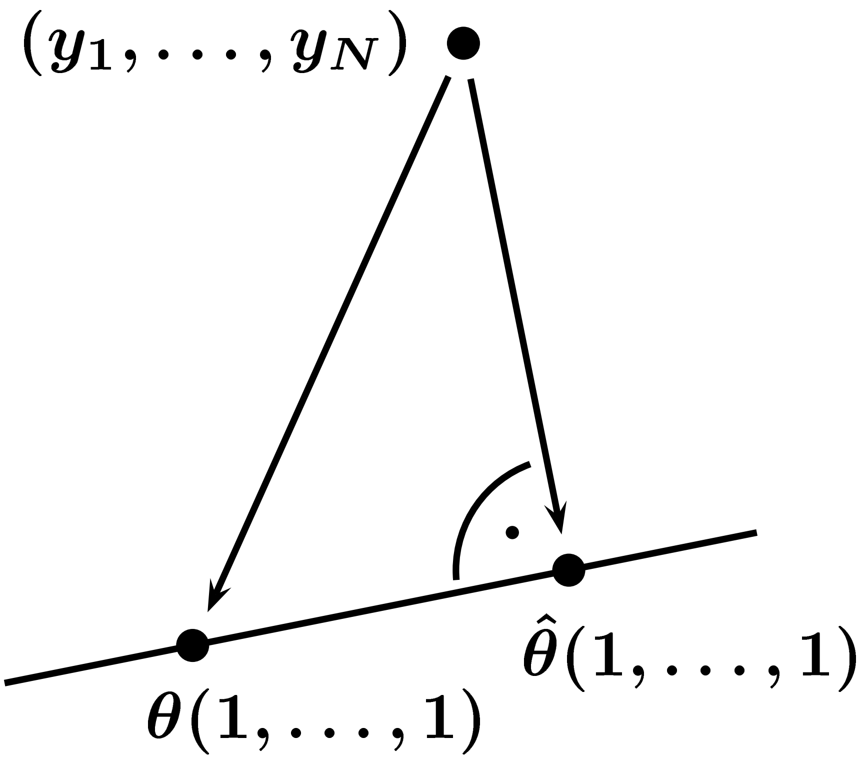
Academy of Sciences of the Czech Republic

Plan of the Talk

- **Estimation with *independent* data**
  - **Euclidean geometry of data**
  - **Geometry of probability distributions**
- **Estimation with *dependent* data**
  - **Euclidean geometry of data**
  - **Geometry of probability distributions**
- **Applications to nonstandard cases**

Euclidean Geometry of Data

Example: $Y_k = \theta + E_k$



min. distance projection

$$\min_{\theta} \|\underline{y} - \theta\,\underline{1}\|^2$$

orthogonal projection

$$\left(\underline{y} - \hat{\theta}\,\underline{1}\right)^T \underline{1} = 0$$

Pythagorean relation

$$\|\underline{y} - \theta\,\underline{1}\|^2 = \|\underline{y} - \hat{\theta}\,\underline{1}\|^2 + \|\theta\,\underline{1} - \hat{\theta}\,\underline{1}\|^2$$

From Data to Distrib's of Data

DATA: empirical density

$$r_N(y) = \tfrac{1}{N} \textstyle\sum_{k=1}^{N} \delta(y - y_k)$$

MODEL: exponential sampling density

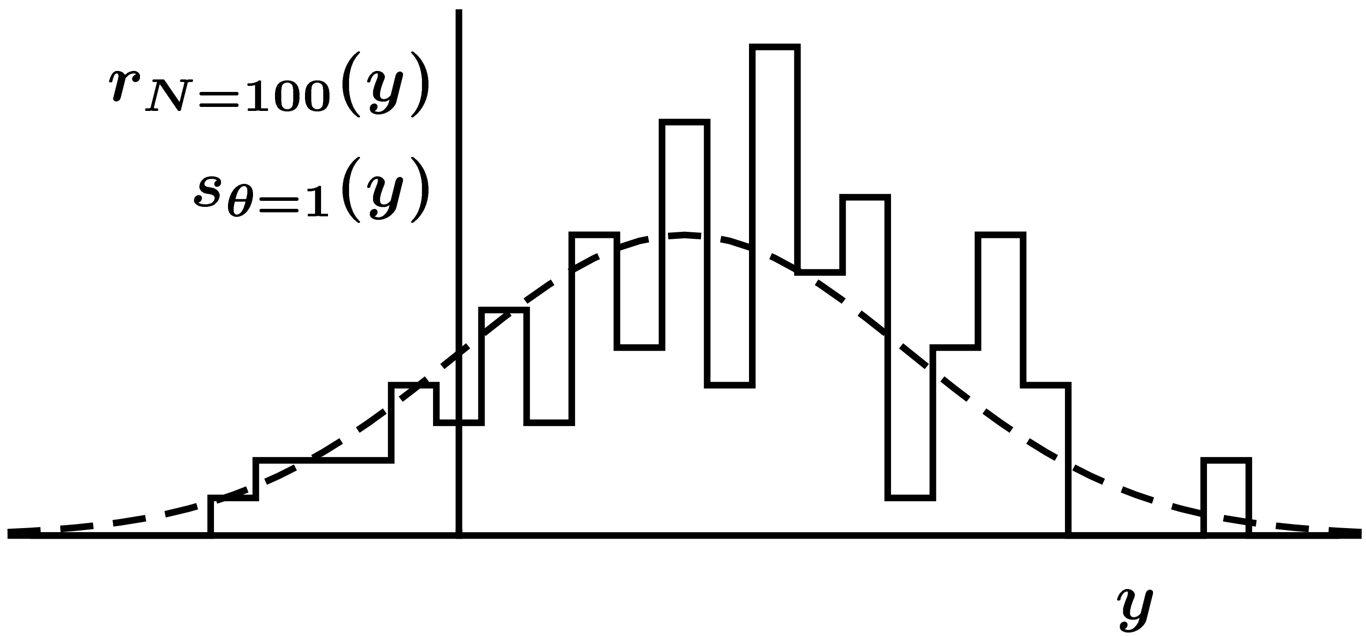
$$s_\theta(y) = s_0(y)\ \exp\big(\theta^T h(y) - \psi(\theta)\big)$$

LIKELIHOOD: inaccuracy [Kerridge, 1961]

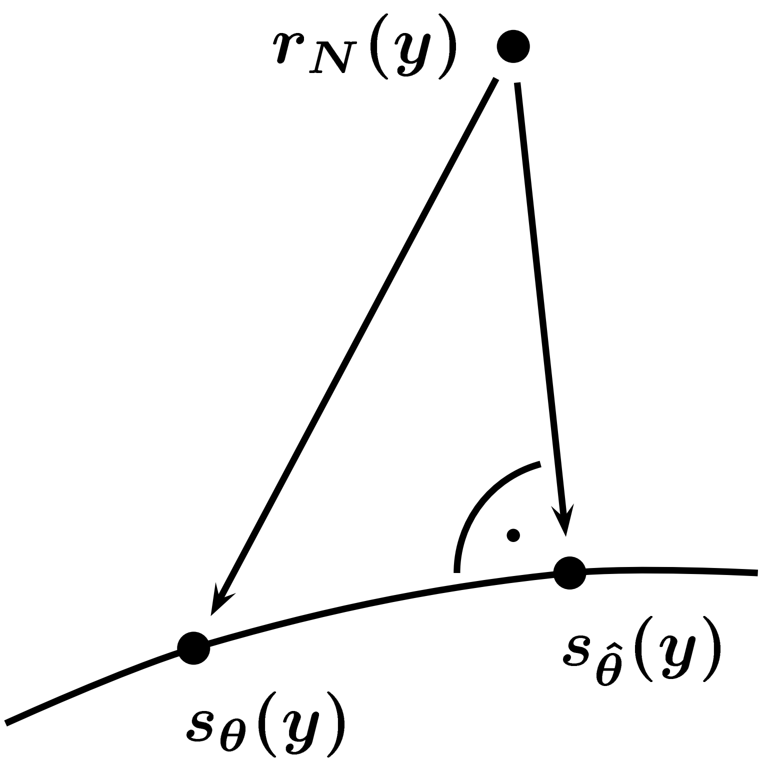
$$\underbrace{\prod_{k=1}^{N} s_\theta(y_k)}_{l_N(\theta)} = \exp\Big(-N \underbrace{\int r_N(y)\ \log \frac{1}{s_\theta(y)}\, \mathrm{d}y}_{K(r_N : s_\theta)}\Big)$$

Example: $Y_k \sim N(\theta, 1)$

$$s_\theta(y) = \underbrace{s_0(y)}_{\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}y^2\right)} \exp\left(\theta^T \underbrace{h(y)}_{y} - \underbrace{\psi(\theta)}_{\frac{1}{2}\theta^2}\right)$$



Geometry of Probability Distrib's



min. inaccuracy projection

$$\min_\theta K(r_N : s_\theta)$$

max. likelihood estimate

$\Rightarrow$ **orthogonal projection**

$$\int \big(r_N(y) - s_{\hat{\theta}}(y)\big) \cdot h(y)\,\mathrm{d}y = 0$$

Pythagorean Theorem

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

Kullback-Leibler distance [Kullback & Leibler, 1951]

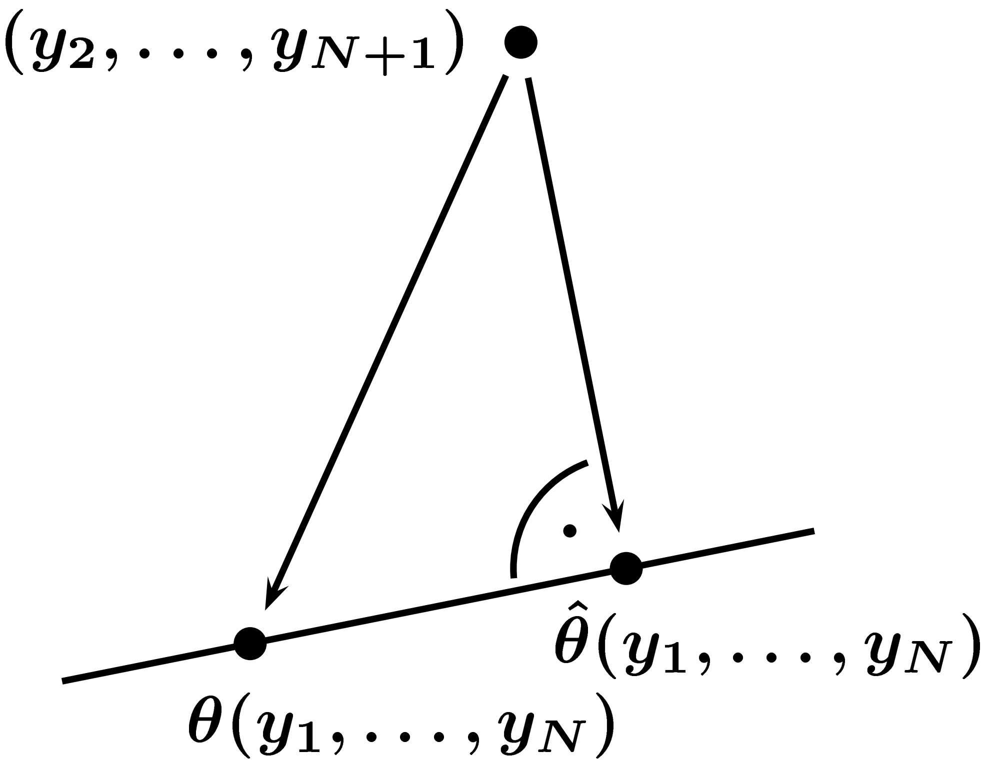
$$D(s_{\hat\theta} \,\|\, s_\theta) = \int s_{\hat\theta}(y)\, \log \frac{s_{\hat\theta}(y)}{s_\theta(y)}\, \mathrm{d}y$$

inaccuracy $\leftrightarrow$ likelihood

$$l_N(\theta) = l_N(\hat\theta)\, \exp\big(-N\, D(s_{\hat\theta} \,\|\, s_\theta)\big)$$

Euclidean Geometry of Data

Example: $Y_k = \theta Z_k + E_k, \;\; Z_k = Y_{k-1}$



min. distance projection

$$\min_{\theta} \|\underline{y} - \theta\, \underline{z}\|^2$$

orthogonal projection

$$\left(\underline{y} - \hat{\theta}\, \underline{z}\right)^T \underline{z} = 0$$

Pythagorean relation

$$\|\underline{y} - \theta\, \underline{z}\|^2 = \|\underline{y} - \hat{\theta}\, \underline{z}\|^2 + \|\theta\, \underline{z} - \hat{\theta}\, \underline{z}\|^2$$

From Data to Distrib's of Data

DATA: empirical *joint* density

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

MODEL: exponential *conditional* sampling density

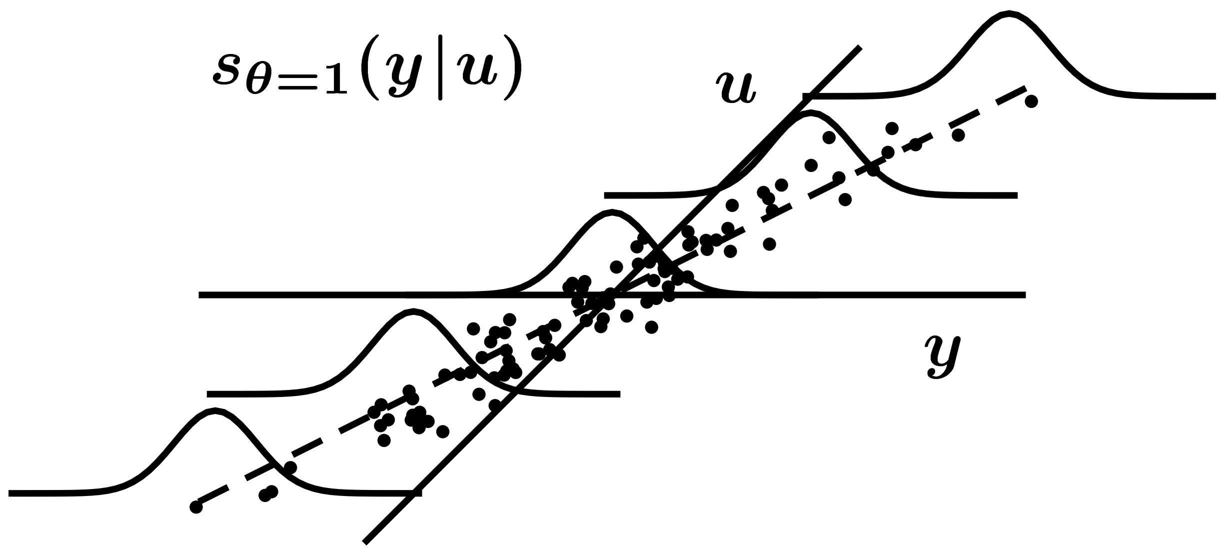
$$s_\theta(y|z) = s_0(y|z)\ \exp\big(\theta^T h(y,z) - \psi(\theta,z)\big)$$

LIKELIHOOD: *conditional* inaccuracy

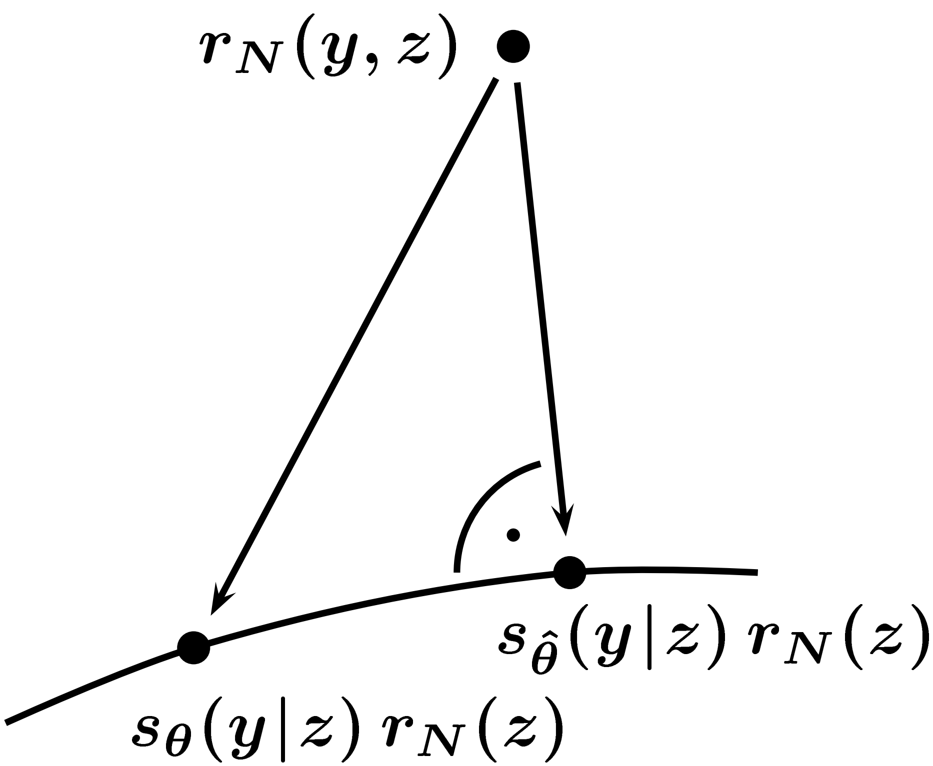
$$\underbrace{\sum_{k=1}^{N} s_\theta(y_k|z_k)}_{c\,l_N(\theta)} = \exp\Big(-N \underbrace{\int r_N(y,z)\ \log \frac{1}{s_\theta(y|z)}\,\mathrm{d}y}_{\bar{K}(r_N : s_\theta)}\Big)$$

Example: $Y_k \sim N(\theta\, U_k, 1)$

$$s_\theta(y|u) = \underbrace{s_0(y|u)}_{\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}y^2\right)} \exp\Big(\theta^T \underbrace{h(y,u)}_{yu} - \underbrace{\psi(\theta,u)}_{\frac{1}{2}\theta^2u^2}\Big)$$



Geometry of Probability Distrib's



min. inaccuracy projection

$$\min_\theta \bar{K}(r_N : s_\theta)$$

max. likelihood estimate
$\Rightarrow$ orthogonal projection

$$\int \big(r_N(y,z) - s_{\hat{\theta}}(y|z)\, r_N(z)\big) \cdot h(y,z)\, \mathrm{d}y = 0$$

Pythagorean Theorem

$$\bar{K}(r_N : s_\theta) = \bar{K}(r_N : s_{\hat{\theta}}) + \bar{D}(s_{\hat{\theta}} \,\|\, s_\theta \,|\, \tilde{r}_N)$$

conditional **Kullback-Leibler distance**

$$\bar{D}(s_{\hat{\theta}} \,\|\, s_\theta \,|\, \tilde{r}_N) = \int \tilde{r}_N(z) \int s_{\hat{\theta}}(y|z) \, \log \frac{s_{\hat{\theta}}(y|z)}{s_\theta(y|z)} \, \mathrm{d}y \; \mathrm{d}z$$

conditional **inaccuracy** $\leftrightarrow$ ***conditional*** **likelihood**

$$l_N(\theta) = l_N(\hat{\theta}) \, \exp\big(-N \, \bar{D}(s_{\hat{\theta}} \,\|\, s_\theta \,|\, \tilde{r}_N)\big)$$

Data Statistics

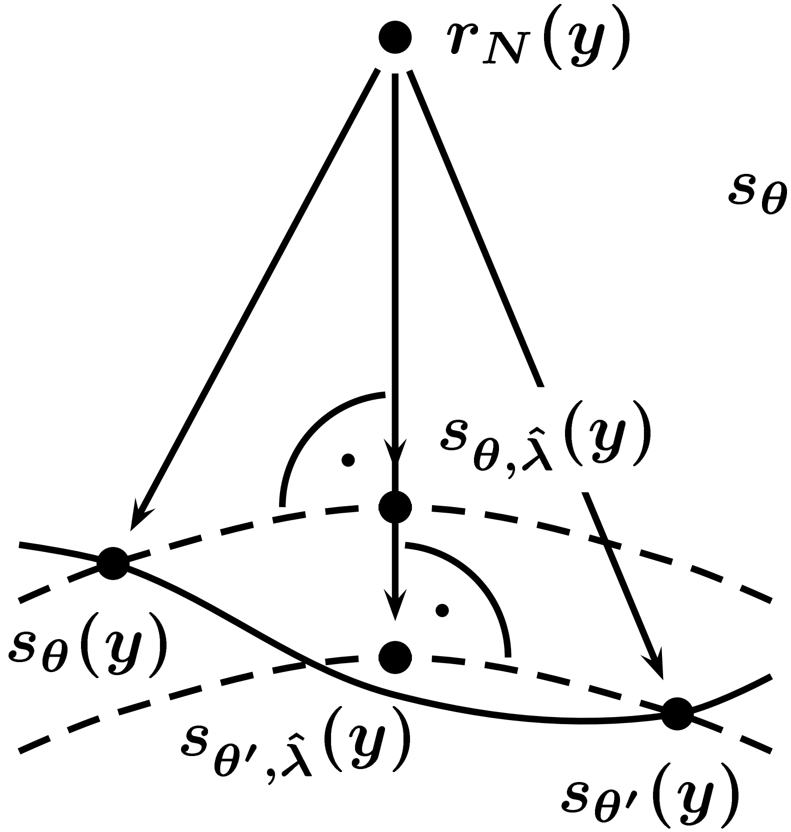
independent data

$$\underbrace{\int s_{\hat{\theta}}(y)\, h(y)\, \mathrm{d}y}_{\hat{h}_{\hat{\theta}}} = \underbrace{\int r_N(y)\, h(y)\, \mathrm{d}y}_{\bar{h}_N = \frac{1}{N}\sum_{k=1}^{N} h(y_k)}$$

dependent data

$$\int \tilde{r}_N(z) \underbrace{\int s_{\hat{\theta}}(y|z)\, h(y,z)\, \mathrm{d}y}_{\hat{h}_{\hat{\theta}}(z)} \mathrm{d}z = \underbrace{\int r_N(y,z)\, h(y,z)\, \mathrm{d}y\, \mathrm{d}z}_{\bar{h}_N = \frac{1}{N}\sum_{k=1}^{N} h(y_k, z_k)}$$

Application to Approximate Estimation



approximating expon. family

$$s_{\theta,\lambda}(y) = s_\theta(y)\ \exp\big(\lambda^T h(y) - \psi(\lambda)\big)$$

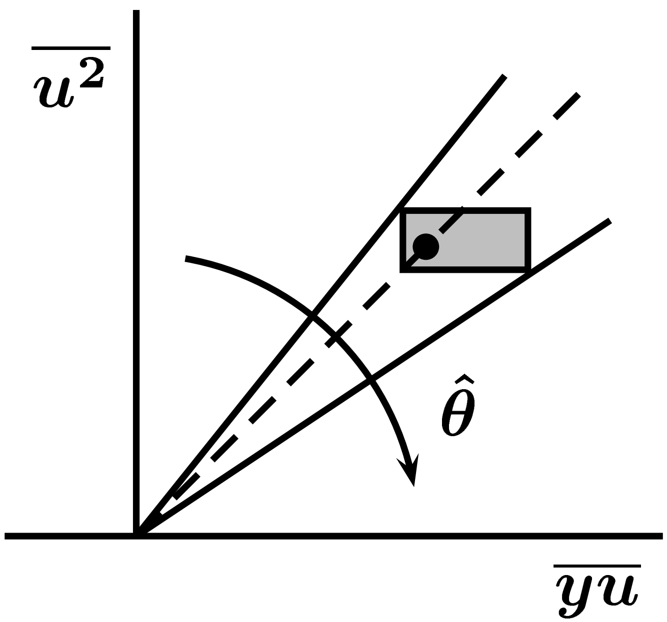
Pythagorean relation

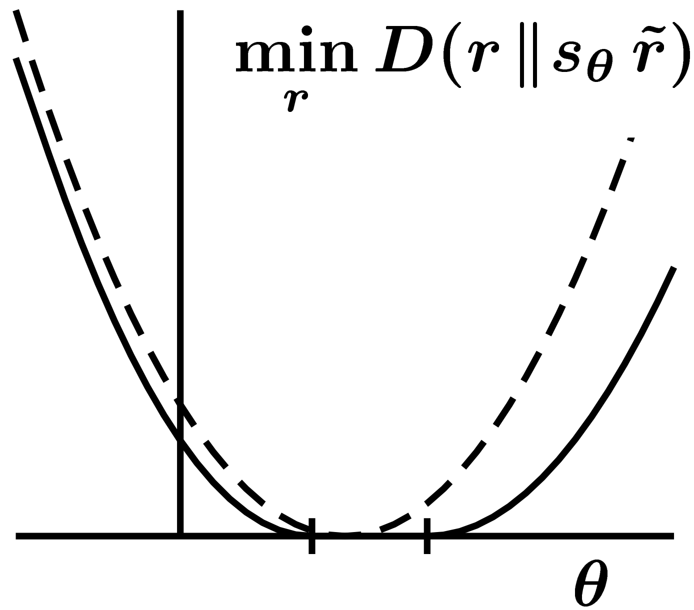
$$K(r_N : s_\theta) = K(r_N : s_{\theta,\hat{\lambda}}) + \underbrace{D(s_{\theta,\hat{\lambda}} \,\|\, s_\theta)}_{\min\limits_{E_r(h)=\bar{h}_N} D(r \,\|\, s_\theta)}$$

Example: $Y_k \sim N(\theta\, U_k, 1)$

unknown but bounded observations (statistics)

$$\hat{\theta} = \frac{\overline{yu}}{\overline{u^2}} \qquad \min_{\overline{yu},\,\overline{u^2}} \frac{1}{2}\Big(\theta - \frac{\overline{yu}}{\overline{u^2}}\Big)^2 \overline{u^2}$$





Why Geometry?

K.L. distance = information for discrimination

$$\mathcal{R}_N \subset \mathcal{R}'_N \quad \Rightarrow \quad D(\mathcal{R}_N \,\|\, s_\theta) \geq D(\mathcal{R}'_N \,\|\, s_\theta)$$