Geometric correction

A guided tour

Christos Dimitrakakis

Intelligent Systems Laboratory Amsterdam, University of Amsterdam, The Netherlands

July 17, 2009

Geometric correction

Definition

Estimating object (parameters) under (geometric) constraints

Objects

- N Objects: $\overline{x} \triangleq \{\overline{u}_\alpha\}_{\alpha=1}^{N}$, $\overline{u}_\alpha \in \mathcal{U}_\alpha \subset \mathbb{R}^\infty$.
- $\overline{x} \in \mathcal{X} \triangleq \times_{i=1}^{N} \mathbb{R}^{m_i}$
- Constraint $F : \mathcal{X} \longrightarrow \mathbb{R}^n$, with $F(\overline{x}) = 0$.

Observations

- Observations $u_\alpha = \overline{u}_\alpha + \Delta u_\alpha$, $u_\alpha \in \mathcal{U}_\alpha \subset \mathbb{R}^m$.
- Noise: $\Delta u_\alpha \in \mathcal{T}_{\overline{u}_\alpha}(U_\alpha)$, $\Delta u_\alpha \sim \mathcal{N}(0, \overline{V}(y_\alpha))$.

Geometric correction

Definition

Estimating object (parameters) under (geometric) constraints

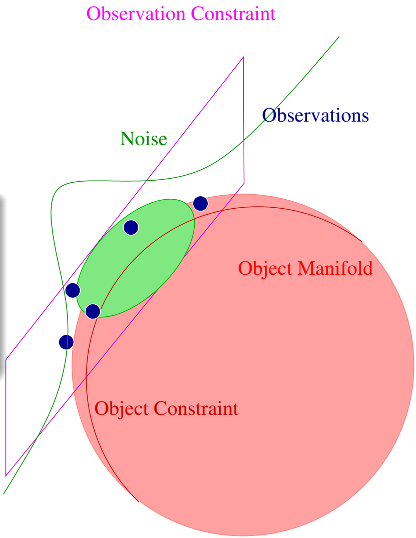
Objects

- N Objects: $\overline{x} \triangleq \{\overline{u}_\alpha\}_{\alpha=1}^{N}$, $\overline{u}_\alpha \in \mathcal{U}_\alpha \subset \mathbb{R}^\infty$.
- $\overline{x} \in \mathcal{X} \triangleq \times_{i=1}^{N} \mathbb{R}^{m_i}$
- Constraint $F : \mathcal{X} \to \mathbb{R}^n$, with $F(\overline{x}) = 0$.

Observations

- Observations $u = \overline{u} + \Delta u$, $u \in \mathcal{U} \subset \mathbb{R}^m$.
- Noise: $\Delta u \in \mathcal{T}_{\overline{u}}(\mathcal{U})$, $\Delta u \sim \mathcal{N}(0, \overline{V}(u))$.

The problem



Definition

- Given
  - Observations u
  - Object constraints $F(\overline{u}) = 0$
  - Noise constraints $\Delta u \in \mathcal{T}_{\overline{u}}$
- Estimate: $\hat{u}$ s.t. $F(\hat{u}) = 0$.

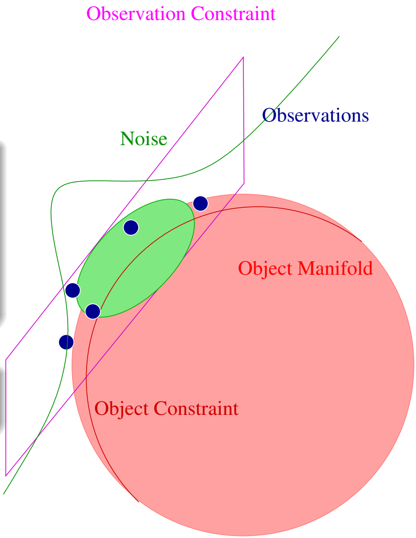
The problem

Definition

- Given
  - Observations u
  - Object constraints $F(\overline{u}) = 0$?
  - Noise constraints $\Delta u \in \mathcal{T}_{\overline{u}}$?
- Estimate: $\hat{u}$ s.t. $F(\hat{u}) = 0$.

A prayer

Let $\hat{u} \approx \overline{u}$.



Constrained optimisation

Constrained minimisation

For $g : \mathcal{X} \to \mathbb{R}$, $F : \mathcal{X} \to \mathbb{R}^n$, the minimum x^* satisfies:

$$g(x^*) \leq g(x), \qquad \forall x : F(x) = 0,$$

with $F(x^*) = 0$.

- Cost function: $g(\cdot)$.
- Constraints: $F(\cdot)$.

Example (Statistical parameter estimation)

Estimate parameters $x \in \mathcal{X}$ given:

- Observations u
- Constraints $F : \mathcal{X} \to \mathbb{R}^n$
- Model set $\Gamma = \{ p(\cdot|x) : x \in \mathcal{X} \}$

$$g(x) = -\ln p(u|x), \qquad F(x) = 0 \tag{1}$$

Constrained minimisation approaches

Penalty method

Define an augmented cost function for $c > 0$:

$$h_c(x) \triangleq g(x) + c\|F(x)\|, \qquad x_c^* \triangleq \arg\min_x h_c(x), \tag{2}$$

$$\lim_{c \to \infty} x_c^* = x^*, \qquad \text{since } \forall \epsilon > 0 \exists c_\epsilon : \forall c > c_\epsilon, \|x_c^* - x^*\| < \epsilon. \tag{3}$$

Lagrangian method

For $\lambda \in \mathbb{R}^n$, $F : \mathcal{X} \to \mathbb{R}^n$.

$$L(x, \lambda) \triangleq g(x) + \lambda^T F(x), \qquad \exists \lambda^* \in \mathbb{R}^n : \nabla_x L(x^*, \lambda^*) = 0$$

Other methods

- Barrier method (for inequality constraints).
- Projection method: Use $P : \mathcal{Z} \to \mathcal{X}$, such that $F(P(z)) = 0$ for all $z \in \mathcal{Z}$.

Lagrangian formulation

Constrained minimisation

Minimise $g(x)$, with $g : \mathcal{X} \to \mathbb{R}$, subject to $F(x) = 0$, with $F : \mathcal{X} \to \mathbb{R}^n$.

Lagrangian

$$L(x, \lambda) \triangleq g(x) + \lambda^T F(x)$$
$$\exists \lambda^* : \nabla_x L(x^*, \lambda^*) = 0$$

Optimality conditions

$$\nabla_x L(x^*, \lambda^*) = 0, \qquad \nabla_\lambda L(x^*, \lambda^*) = 0, \qquad \text{necessary}$$
$$y^T \nabla^2_{xx} L(x^*, \lambda^*) y > 0, \qquad \forall y \neq 0, y \in \mathcal{T}_{x^*} \qquad \text{sufficient}$$

Lagrangian formulation

Constrained minimisation

Minimise $g(x)$, with $g : \mathcal{X} \to \mathbb{R}$, subject to $F(x) = 0$, with $F : \mathcal{X} \to \mathbb{R}^n$.

Lagrangian

$$L(x, \lambda) \triangleq g(x) + \lambda^T F(x)$$
$$\exists \lambda^* : \nabla_x L(x^*, \lambda^*) = 0$$

Optimality conditions

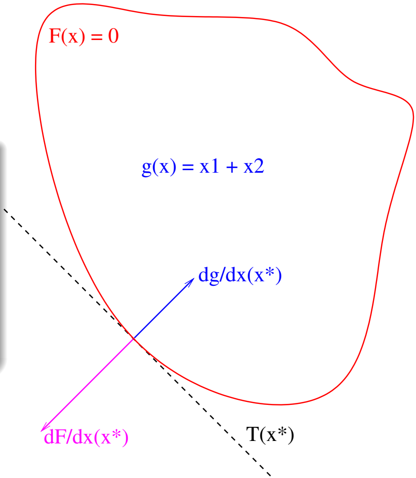
$$\nabla_x L(x^*, \lambda^*) = 0, \qquad \nabla_\lambda L(x^*, \lambda^*) = 0, \qquad \text{necessary}$$
$$y^T \nabla^2_{xx} L(x^*, \lambda^*) y > 0, \qquad \forall y \neq 0, y \in \mathcal{T}_{x^*} \qquad \text{sufficient}$$

Vector and matrix gradients

$x \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$, $F : \mathbb{R}^n \to \mathbb{R}^m$:

$$\nabla_x f(x^*) = \begin{pmatrix} \frac{\partial f(x^*)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x^*)}{\partial x_n} \end{pmatrix}, \qquad \nabla_x F(x^*) = [\nabla_x F_1(x^*) \cdots \nabla_x F_m(x^*)] \qquad (4)$$

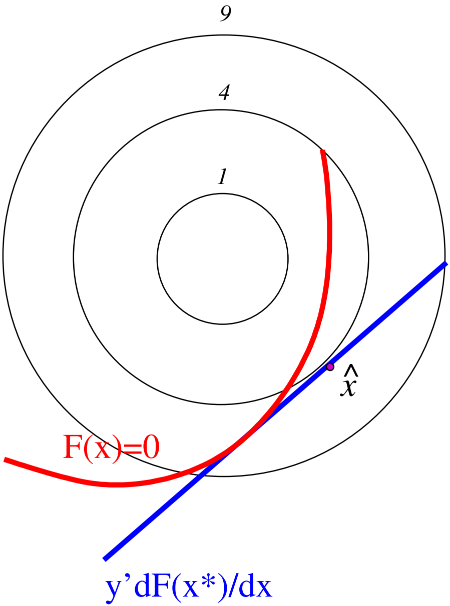
The Lagrangian



Optimality conditions

$$\nabla_x L(x^*, \lambda^*) = 0,$$
$$\nabla_\lambda L(x^*, \lambda^*) = 0,$$
$$y^T \nabla^2_{xx} L(x^*, \lambda^*) > 0, \forall y \neq 0, y \in \mathcal{T}_{x^*}$$
$$\mathcal{T}_{x^*} = \{y \in \mathbb{R}^m : \nabla_x F(x^*)^T y = 0\}$$

Linearisation algorithm



Linearising the constraints

$$F(x) = F(y) + (x - y)^T \nabla_x F(y) + \mathcal{O}\left(x^2\right)$$
$$\approx (x - y)^T \nabla_x F(y),$$

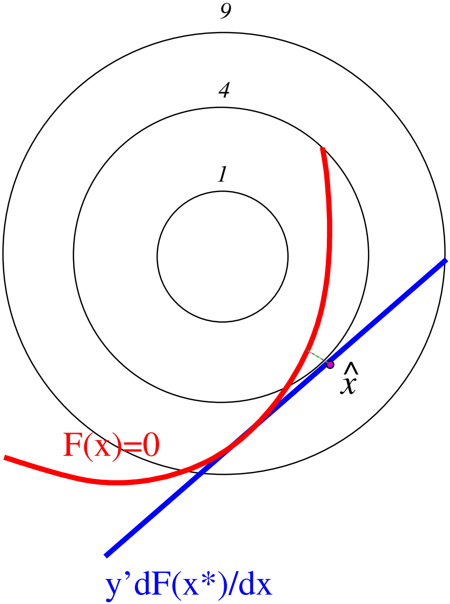
if $F(y) = 0$.

Example (Quadratic cost)

$$g(x) = x^T x,$$
$$F(x) \approx (x - y)^T \nabla_x F(y)$$

for all $y : F(y) = 0$.

Linearisation algorithm



Linearising the constraints

$$F(x) = F(y) + (x-y)^T \nabla_x F(y) + \mathcal{O}\left(x^2\right)$$
$$\approx (x-y)^T \nabla_x F(y),$$

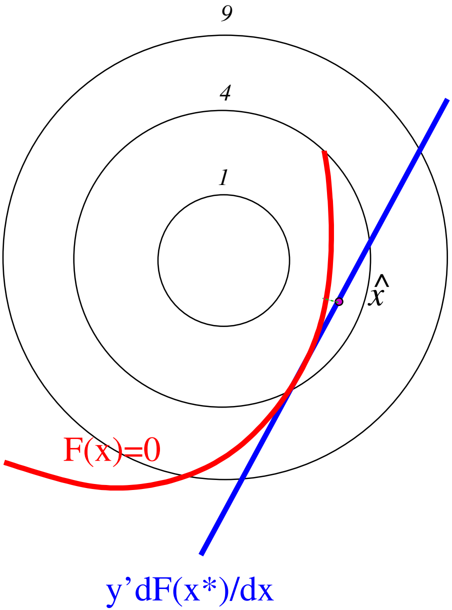
if $F(y) = 0$.

Example (Quadratic cost)

$$g(x) = x^T x,$$
$$F(x) \approx (x-y)^T \nabla_x F(y)$$

for all $y : F(y) = 0$.

Linearisation algorithm



Linearising the constraints

$$F(x) = F(y) + (x - y)^T \nabla_x F(y) + \mathcal{O}\left(x^2\right)$$
$$\approx (x - y)^T \nabla_x F(y),$$

if $F(y) = 0$.

Example (Quadratic cost)

$$g(x) = x^T x,$$
$$F(x) \approx (x - y)^T \nabla_x F(y)$$

for all $y : F(y) = 0$.

Optimisation for geometric estimation

Two sets of constraints

$$F(u) \approx \Delta u^T \nabla_u F(\bar{u})$$
$$M(\Delta u) = \Delta u^T v$$

Noise model

$$p(u|x) \propto \exp\left(-\frac{1}{2}(u - \bar{u})^T \Sigma^{-1} (u - \bar{u})\right), \qquad x = \mathcal{N}(\bar{u}, \Sigma). \tag{5}$$

Solution

- F is linear, g is quadratic, solve for $\lambda = WF$,

$$W = \nabla_u F^T V \nabla_u F.$$

- Noise constraints irrelevant.

Optimisation for geometric estimation

Two sets of constraints

$$F(u) \approx \Delta u^T \nabla_u F(\overline{u})$$
$$M(\Delta u) = \Delta u^T v$$

Noise model

$$p(u|x) \propto \exp\left(-\frac{1}{2}(u - \overline{u})^T \Sigma^{-1} (u - \overline{u})\right), \qquad x = \mathcal{N}(\overline{u}, \Sigma). \tag{5}$$

Problems

- $\Sigma = V[\overline{u}] \approx V[u]$
- Ill-defined problem: Constraints depend on $F(\overline{u})$

The covariance matrix

The noise and the constraints

- We need V to estimate λ

Estimating the covariance

- Approximate $\overline{V}$ (the actual covariance) with V (the local covariance).
- Problem: small $\|V - \hat{V}\|$ does not imply small $\|V^{-1} - \hat{V}^{-1}\|$.
- Kanatani's solution: Use linear algebra magic.

Estimating a good covariance matrix

Finding the Lagrange vector

F is linear, g is quadratic, solve for $\lambda = WF$,

$$W = \left(\nabla_u F^T V \nabla_u F\right)^{-1} \quad (6)$$

Estimating the covariance V

- Approximate $\overline{V}$ by V and $F(\overline{u})$ by $F(u)$.
- We know that the rank of $\overline{V}$ is r.

Estimating a good covariance matrix

Finding the Lagrange vector

F is linear, g is quadratic, solve for $\lambda = WF$,

$$W = \left(\nabla_u F^T V \nabla_u F \right)^{-1} \quad (6)$$

Some set-like notation

$W = Z^{-1}$, where $Z = (Z^{kl})$, $W = (W^{kl})$

$$Z = \left(\nabla_u F_k^T V \nabla_u F_l \right)$$

$$W = \left(\nabla_u F_k^T V \nabla_u F_l \right)^{-1}$$

Estimating the covariance V

- Approximate $\overline{V}$ by V and $F(\overline{u})$ by $F(u)$.
- We know that the rank of $\overline{V}$ is r.

Estimating a good covariance matrix

Finding the Lagrange vector

F is linear, g is quadratic, solve for $\lambda = WF$,

$$W = \left(\nabla_u F^T V \nabla_u F\right)^{-1} \tag{6}$$

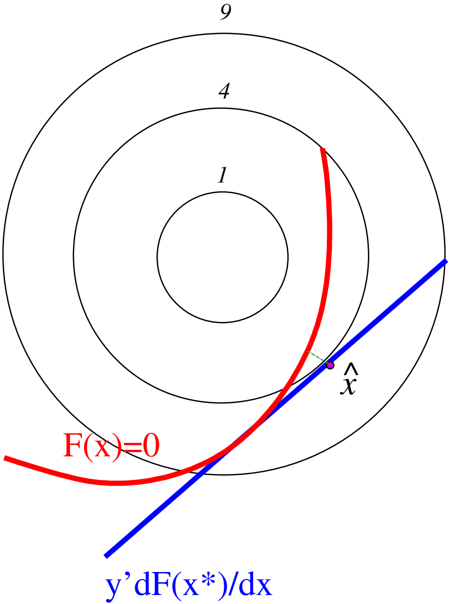
Estimating the covariance V

- Approximate $\overline{V}$ by V and $F(\overline{u})$ by $F(u)$.
- We know that the rank of $\overline{V}$ is r.

Rank-constrained generalized inverse

$$W_i = \left(\nabla_u F^T V[u] \nabla_u F\right)_r^{-1} \tag{7}$$

Iterated linearisation



Iterated linearised constrained optimisation

1: $\hat{u}_1 = u$.
2: **for** $t = 1, 2, \ldots$ **do**
3: $\Sigma_t = \mathcal{P}_{\hat{u}_t} V$
4: $\hat{F}_t = \Delta u^t \nabla_u F(\mathcal{P}_{\Sigma_t} \hat{u}_t)$
5: $g(u|\hat{u}, \Sigma_t) \triangleq \frac{1}{2}(u - \hat{u})^T \Sigma_t^{-1} (u - \hat{u})$.
6: $\hat{u}_{t+1} = \mathcal{P} \arg\min_{\hat{u}} g(u|\hat{u}, \Sigma_t)$.
7: **end for**

Projection

The projection $\mathcal{P}_{\Sigma_t}$ of $\hat{u}$ to $F(u) = 0$, where the linearization is performed, is done by minising the Mahanalobis distance.

Cost function changes at every step

$\Sigma_t \neq \Sigma_{t+1}$. Does it still converge? Convergence conditions unclear.

"A posteriori" covariance matrices

What is covariance here?

- "a priori" $m \times m$ covariance matrix V, assumed known (to some constant)
- For $\mathcal{T}$ a n-dimensional linear subspace of $\mathbb{R}^m$, $V_{\mathcal{T}} = \mathcal{P}_{\mathcal{T}} V$.
- $\mathcal{T}(u)$ is the tangent space to an n-dimensional manifold in $\mathbb{R}^m$, evaluated at u.
- $\bar{V} = V_{\mathcal{T}(\bar{u})}$, $V[u] = V_{\mathcal{T}_u}$.

What does "a posteriori" mean?

- Unrelated to conditional measures
- The "a priori" covariance matrix is merely the covariance evaluated at u.
- The "a posteriori" covariance matrix is the covariance evaluated at $\hat{u}$.

"Confidence regions" and noise

- Uncertainty about parameters must not be confused with observation noise.
- i.e. certainty that a coin is fair: $\theta = 0.5$ w.p. 1.
- Noisy measurements.

Finding the correct hypothesis

The setting

- Parameters/distribution $\theta \in \Theta$.
- Estimate $\widehat{\theta}_n \in \Theta$ from observations $z^n \triangleq \{z_1, \ldots, z_n\}$, $z_i \in \mathcal{Z}$.
- Obtain different estimate $\widehat{\theta}_n(H)$ under different hypotheses H, i.e. if each H corresponds to a different set of constraints on Θ. Which hypothesis to use?

The meaning of hypothesis testing

- Estimate how good the estimates (hypothesis) are
- Select the most suitable hypothesis, reporting error probability δ.
- Ultimately, a decision problem.

Frequentist principle

In repeated practical use of a statistical procedure, the long-run average actual error should not be greater than (and ideally should equal) the long-run average reported error.

Tail bound

Tail bound

Fix some $Z^n \subset \mathcal{Z}^n$. Then:

$$\mathbf{P}(z^n \notin Z^n | \theta) < f(\theta, Z^n),$$

f decreasing with $|Z^n|$.

Example (χ^2-test)

$$T(z) \triangleq \int_{R_\Sigma(z)}^{\infty} p_{\chi^2}(x) \, \mathrm{d}x \tag{8}$$

$$R_\Sigma(z) = \langle z, \Sigma^{-1} z \rangle \tag{9}$$

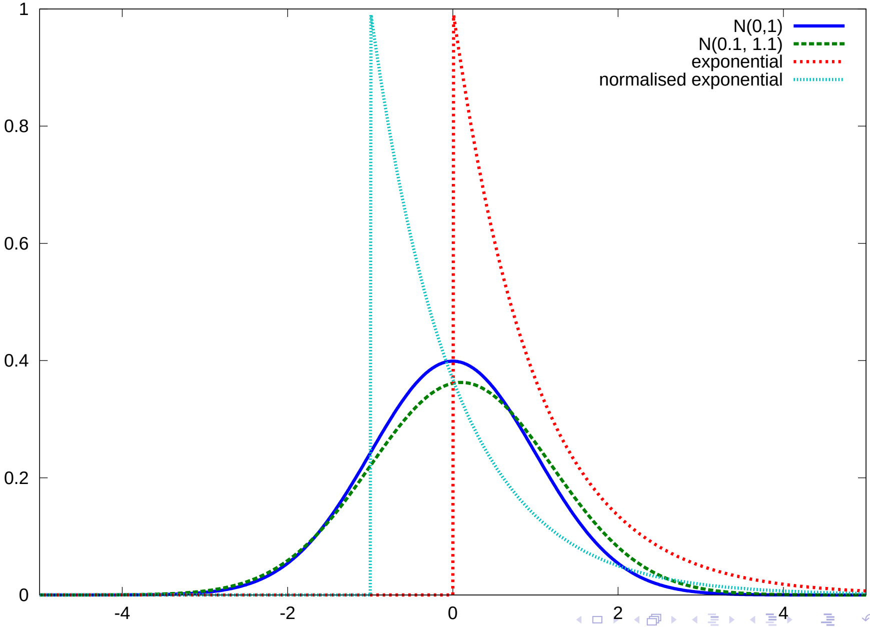
Has the property:

$$T(z) \sim \mathrm{Uniform}(0, 1), \qquad \text{if } z \sim \mathcal{N}(0, \Sigma). \tag{10}$$

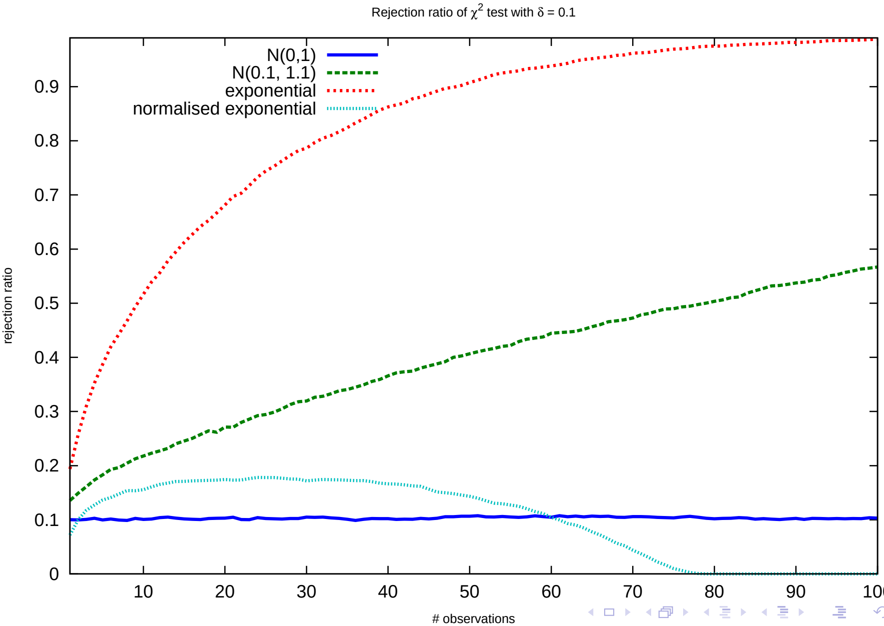
So:

$$\mathbf{P}(T(z) < \delta | z \sim \mathcal{N}(0, \Sigma)) < \delta, \qquad \forall \delta \in [0, 1]. \tag{11}$$

Testing for normality



The χ^2 test's performancer



Concentration inequality

Concentration inequality

Let D be a distance on Θ. Generally,

$$\mathbf{P}(D(\widehat{\theta}_n, \theta) > \epsilon | \theta) < \mathcal{O}\left(exp(-n\epsilon^2)\right), \qquad \forall \theta \in \Theta, \epsilon > 0. \tag{12}$$

Example (Hoeffding bound)

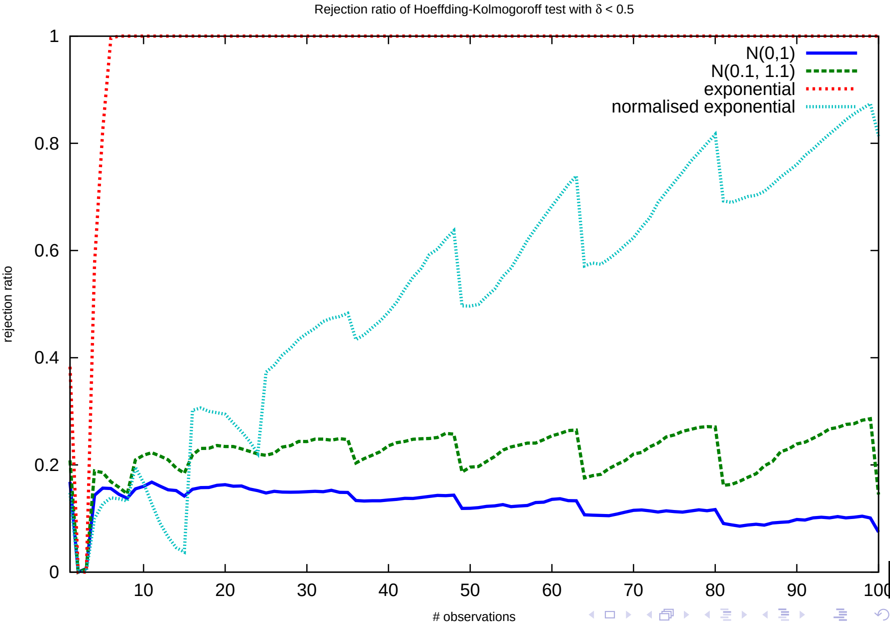
For $x \in [0,1]$, $\hat{x} \triangleq \frac{1}{n} \sum_{i=1}^{n} x_i$ and for any $\mathbf{P}$ and $\epsilon > 0$:

$$\mathbf{P}\left(\hat{x} \geq \mathbf{E}\, x + \epsilon\right) \leq \exp(-2n\epsilon^2) \Leftrightarrow \mathbf{P}\left(\hat{x} \geq \mathbf{E}\, x + \sqrt{\log(1/\delta)/2n}\right) \leq \delta. \tag{13}$$

Application to general measures

Let P_n be the empirical measure over $\sqrt{n}$ disjoint subsets S_i derived from z^n (i.e. a histogram with $\sqrt{n}$ bins). We can apply Hoeffding (or other concentration inequalities) to the distance between $P_n(z \in S_i)$ and $\mathbf{P}(z \in S_i)$, by setting $x^{(i)} = \mathbb{I}\{z \in S_i\}$.

The non-parametric Hoeffding-Kolmogoroff goodness-of-fit test



Bayesian hypothesis tests

Multiple hypotheses test

Given a set of hypotheses $H \triangleq \{ h_i : i = 1, \ldots, k \}$, with associated prior probabilities $\{ \pi(h_i) : i = 1, \ldots, k \}$, and data z, estimate

$$\pi(h_i|z) \triangleq \frac{\mathbf{P}(z|h_i)\pi(h_i)}{\sum_{j=1}^{k} \mathbf{P}(z|h_j)\pi(h_j)}. \tag{14}$$

ϵ-Null hypothesis test

Given a null hypothesis $h_0 = \mathbb{I}\{\theta \in \Theta_0\}$, with associated prior probability $\pi(h_0)$, construct $h_\epsilon \triangleq \mathbb{I}\{\theta \in \Theta_\epsilon\}$, where

$$\Theta_\epsilon = \{ \theta \in \Theta : \inf_{\theta' \in \Theta_0} D(\theta, \theta') < \epsilon \}$$

$$\pi(h_0|z) \leq \pi(h_\epsilon|z) \triangleq \frac{\mathbf{P}(z|h_\epsilon)\pi(h_\epsilon)}{\mathbf{P}(z|h_\epsilon)\pi(h_\epsilon) + \mathbf{P}(z|h_A)[1 - \pi(h_\epsilon)]}. \tag{15}$$

Coincidence

Assumptions and constraints

$$\bar{x}_1 = \bar{x}_2.$$

x_1, x_2 independent, $\mathbf{E}\, x_i = \bar{x}_i$.
Estimate $\hat{x}_i = x_i - \Delta x_i$.

Constrained cost minimisation

$$J(\hat{x}_i) \triangleq \sum_i g(x_i|\hat{x}_i, \Sigma), \qquad g(x_i|\hat{x}_i, \Sigma_i) \triangleq \frac{1}{2}(x_i - \hat{x}_i)^T \Sigma_i^{-1}(x_i - \hat{x}_i) \tag{16}$$

under constraints

$$\hat{x}_1 = \hat{x}_2, \qquad \Delta x_1, \Delta x_2 \text{ on the image plane.} \tag{17}$$

Coincidence

First order solution

$$\Delta x_1 = V[x_1]\mathbf{W}(x_1 - x_2) \tag{18}$$
$$\Delta x_2 = V[x_2]\mathbf{W}(x_2 - x_1) \tag{19}$$
$$\mathbf{W} \triangleq (V[x_1] + V[x_2])^{-1}. \tag{20}$$

Residual

"A posteriori" covariance matrix

$$V[\hat{x}] = V[x_1]\mathbf{W}V[x_2] = V[x_2](\mathbf{I} - \mathbf{W}V[x_2]) \tag{21}$$

Residual $\widehat{J} = \langle x_2 - x_1, \mathbf{W}x_2 - x1 \rangle$, with $\widehat{J} \sim \chi^2(2)$.

Hypothesis test

Perhaps better to test $\|x_2 - x_1\| < \epsilon$.

More examples??