Event Classification

31 May 2016

Shiro Ikeda
Institute of Statistical Mathematics

Personal Background.

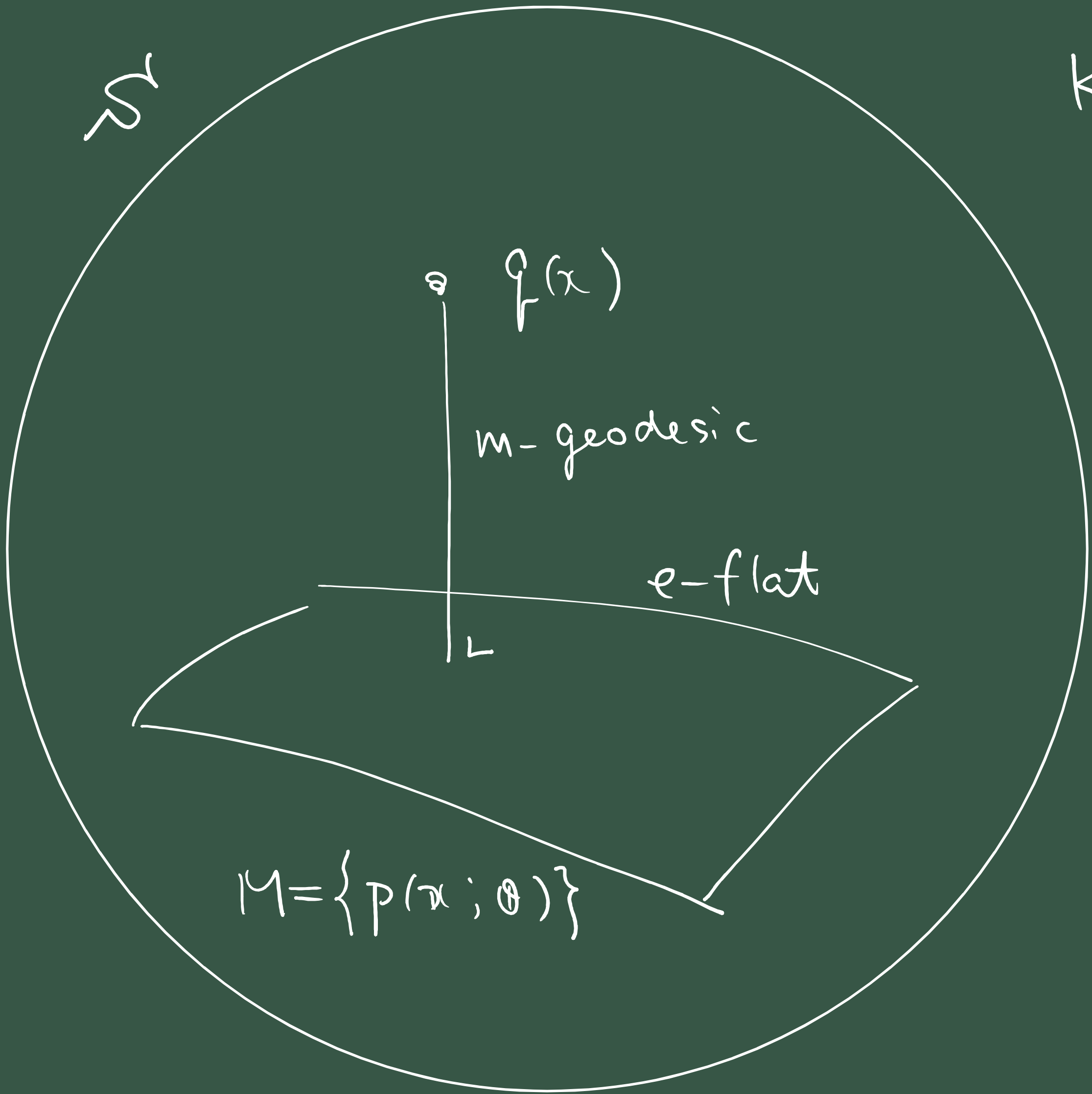
Theory

- Signal Processing
- Information Theory
- Information Geometry
- Statistics

Applications

- Speech Separation
- Fishery Data
- Astronomy

@ Information Geometry



$$KL(p_\theta, p_{\theta+\delta\theta}) = \frac{1}{2}\delta\theta^T I(\theta)\,\delta\theta$$

$I(\theta)$: Fisher matrix

$$\Gamma^{L-C}_{ij,k} = \frac{1}{2}\left(\Gamma^{e}_{ij,k} + \Gamma^{m}_{ij,k}\right)$$

- estimation
- test
- prediction
- approximate inference

Information Geometry anc ×

igaia.utia.cz

IGAIA IV

Information Geometry and its Applications IV

June 12-17, 2016, Liblice, Czech Republic

About the Conference

The fourth conference on Information Geometry and its Applications will be held on June 12-17, 2016, at Liblice Castle in Czech Republic. It is organized by the Institute of Information Theory and Automation of the Czech Academy of Sciences. The first three conferences took place in Pescara, Italy (**IGAIA I, 2002**), Tokio, Japan (**IGAIA II, 2005**), and Leipzig (**IGAIA III, 2010**).

The conference is financially supported by Max Planck Institute for Mathematics in the Sciences (Information Theory of Cognitive systems Group), Institute of Information Theory and Automation of Czech Academy of Sciences, and Università degli Studi di Roma Tor Vergata.

The aim of the conference is to highlight recent developments within the field of information geometry and to identify new directions of research. To this end, the conference will bring together leading experts in the field. It will consist of invited talks and poster sessions representing theoretical work as well as work within one of the many application fields where information geometry plays a role.

Confirmed invited speakers

Shun-ichi Amari, RIKEN Brain Science Institute, JP
Roman Belavkin, Middlesex University, UK
Michel Broniatowski, l'Université Pierre et Marie Curie, FR
Imre Csiszár, Hungarian Academy of Sciences, HU
Shinto Eguchi, The Institute of Statistical Mathematics, JP
Kenji Fukumizu, The Institute of Statistical Mathematics, JP
Davide Girolami, University of Oxford, UK
Peter Harremoës, Copenhagen Business College, DK
Frank Hansen, Tohoku University, JP
Shiro Ikeda, The Institute of Statistical Mathematics, JP
Jürgen Jost, Max Planck Institute for Mathematics in the Sciences, GE
Hông Vân Lê, Czech Academy of Sciences, CZ
Shunlong Luo, Chinese Academy of Sciences, Beijing, China
Hiroshi Matsuzoe, Nagoya Institute of Technology, JP
Peter Michor, Universität Wien, Austria

Shun-ichi Amari

This event honours the numerous scientific achievements of **Shun-ichi Amari** on the occasion of his 80th birthday.

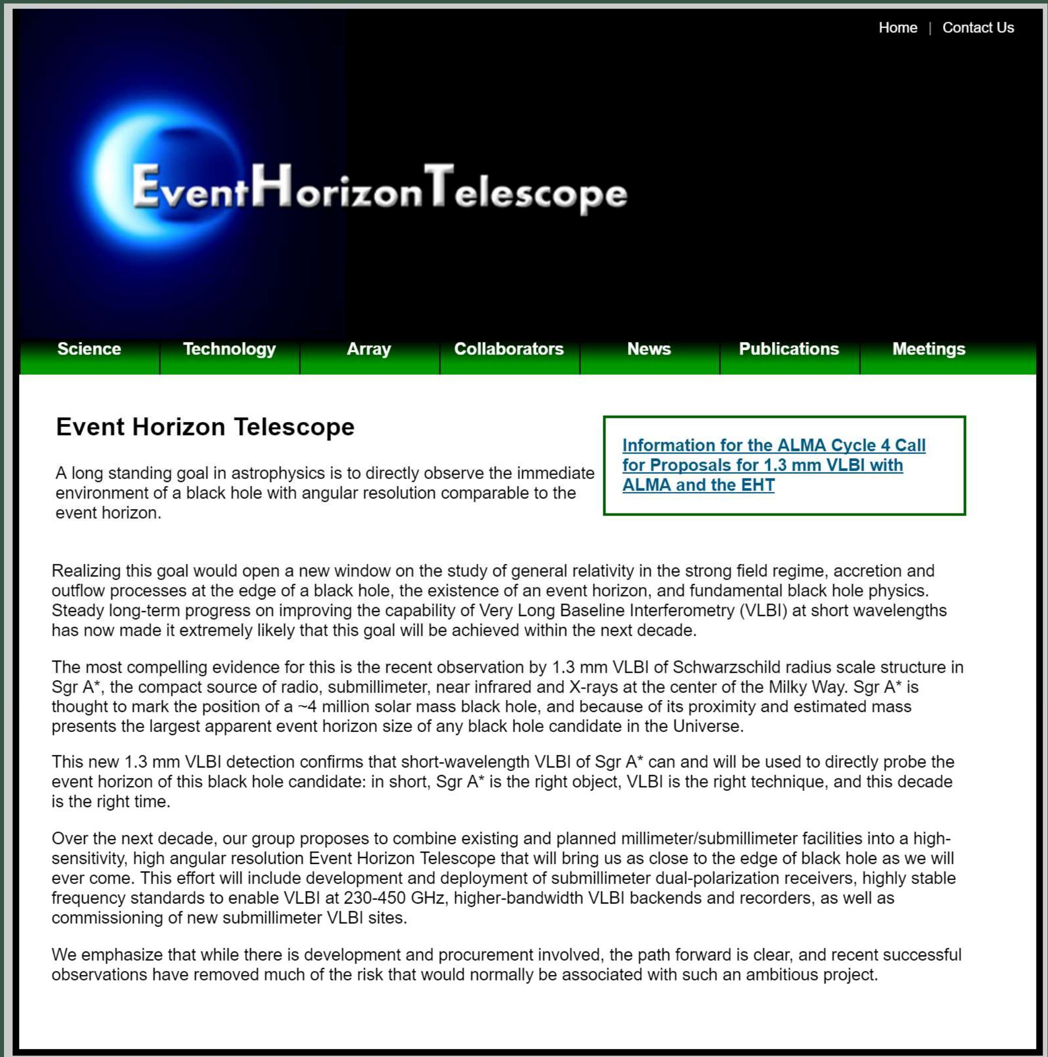
Poster session

The participants are invited to present a poster. The book of abstracts will be published prior to the conference. The poster boards are good for posters up to A0 size. There are no other poster specifications.

Dates and deadlines

April 14, 2016 - registration deadline
April 15, 2016 - notification of acceptance

@ EHT


Home | Contact Us

Event Horizon Telescope

Science | Technology | Array | Collaborators | News | Publications | Meetings

Event Horizon Telescope

A long standing goal in astrophysics is to directly observe the immediate environment of a black hole with angular resolution comparable to the event horizon.

Information for the ALMA Cycle 4 Call for Proposals for 1.3 mm VLBI with ALMA and the EHT

Realizing this goal would open a new window on the study of general relativity in the strong field regime, accretion and outflow processes at the edge of a black hole, the existence of an event horizon, and fundamental black hole physics. Steady long-term progress on improving the capability of Very Long Baseline Interferometry (VLBI) at short wavelengths has now made it extremely likely that this goal will be achieved within the next decade.

The most compelling evidence for this is the recent observation by 1.3 mm VLBI of Schwarzschild radius scale structure in Sgr A*, the compact source of radio, submillimeter, near infrared and X-rays at the center of the Milky Way. Sgr A* is thought to mark the position of a ~4 million solar mass black hole, and because of its proximity and estimated mass presents the largest apparent event horizon size of any black hole candidate in the Universe.

This new 1.3 mm VLBI detection confirms that short-wavelength VLBI of Sgr A* can and will be used to directly probe the event horizon of this black hole candidate: in short, Sgr A* is the right object, VLBI is the right technique, and this decade is the right time.

Over the next decade, our group proposes to combine existing and planned millimeter/submillimeter facilities into a high-sensitivity, high angular resolution Event Horizon Telescope that will bring us as close to the edge of black hole as we will ever come. This effort will include development and deployment of submillimeter dual-polarization receivers, highly stable frequency standards to enable VLBI at 230-450 GHz, higher-bandwidth VLBI backends and recorders, as well as commissioning of new submillimeter VLBI sites.

We emphasize that while there is development and procurement involved, the path forward is clear, and recent successful observations have removed much of the risk that would normally be associated with such an ambitious project.

http://www.eventhorizontelescope.org/

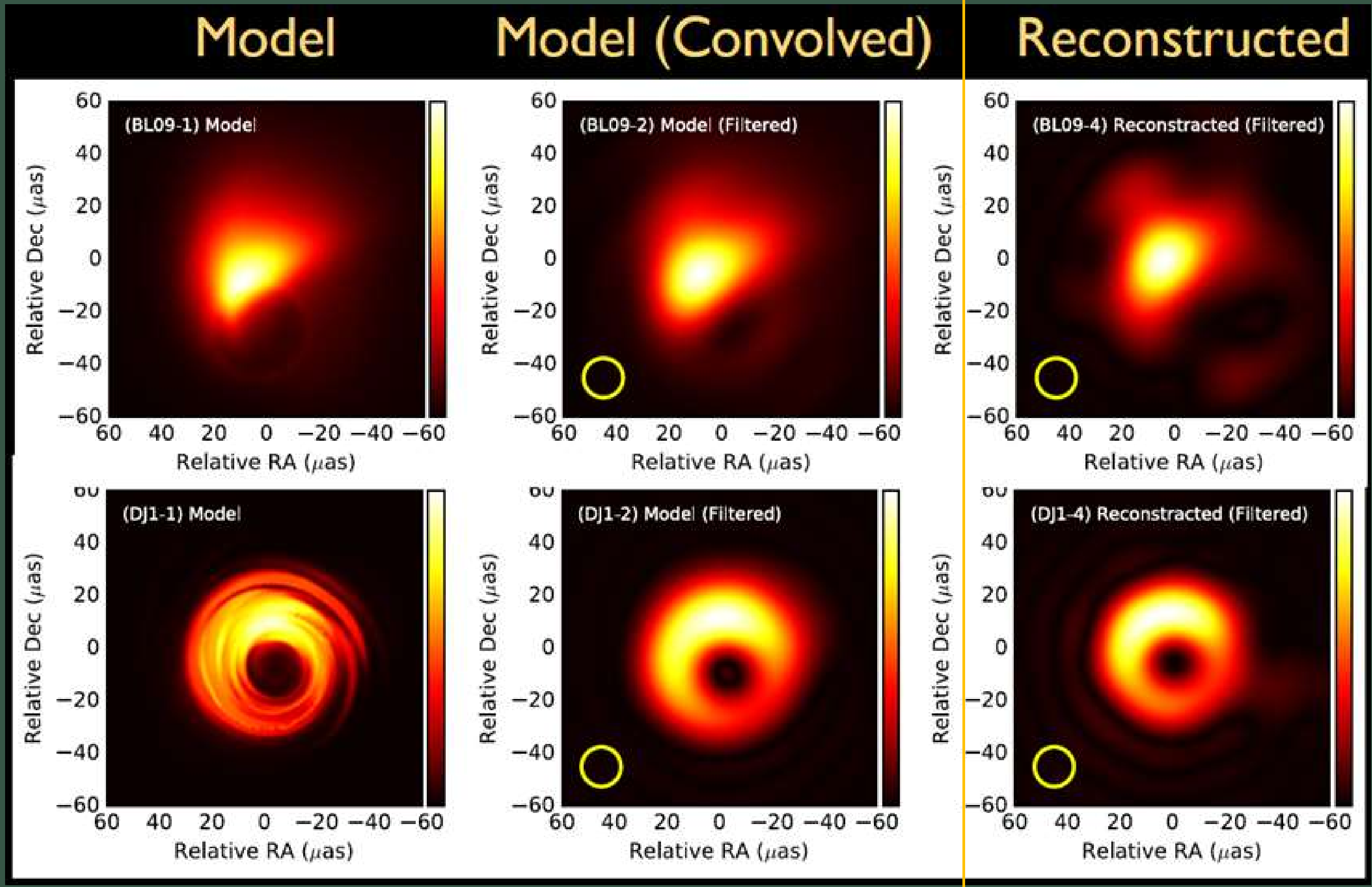
Image of Black Hole Shadow with Very Long Baseline Interferometer.

- @ Imaging is an ill-posed problem.

Collaboration:
Honma, Tazaki, Hada (NAOJ),
Akiyama(MIT),
Kuramochi (u-tokyo) ,...

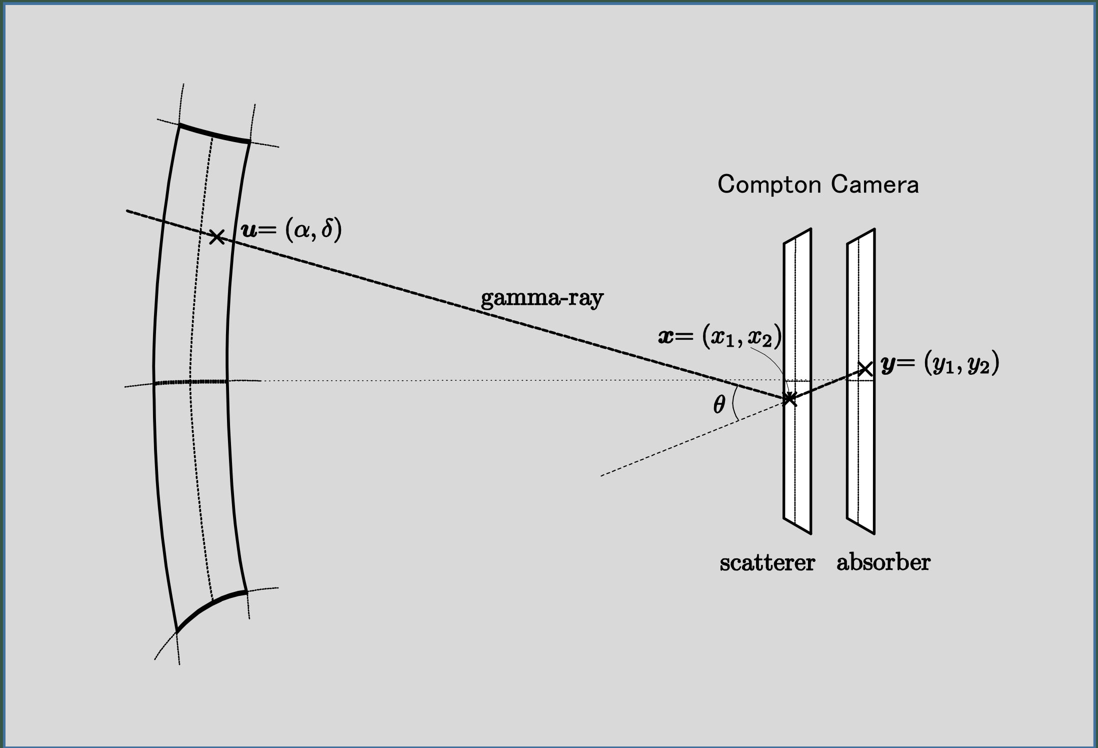
③ EHT imaging

Proposed Method



Ikeda, el.al, PASJ, 2016

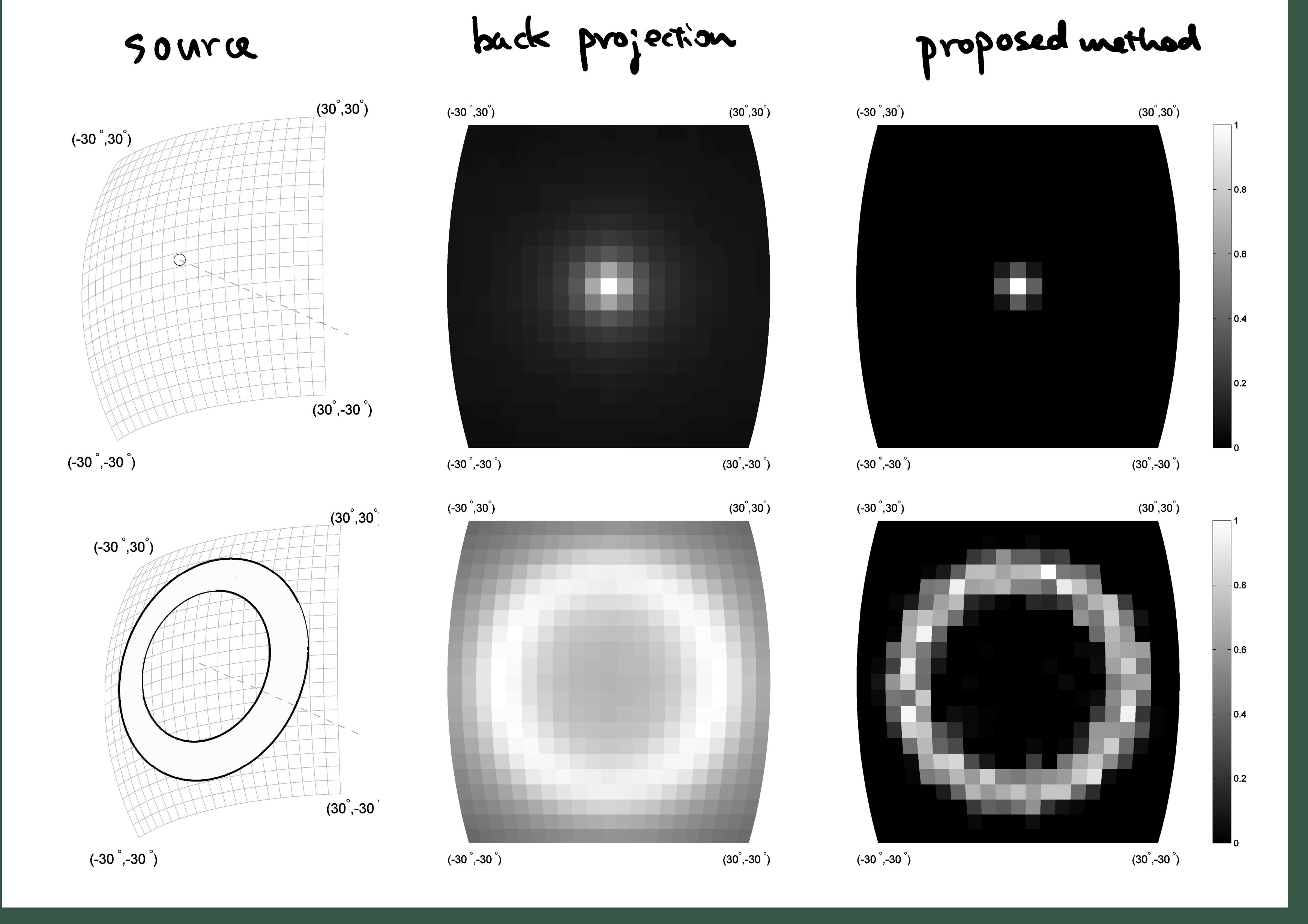
Compton Camera Imaging



Compton Scattering

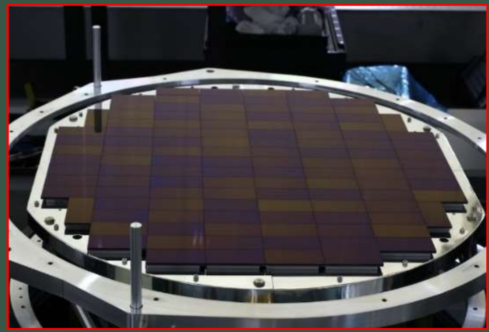
Ikeda, et.al, NIMA, 760, 2014

Imaging of Gamma ray sources



@ HSC

http://subarutelescope.org/Introduction/telescope.html



http://subarutelescope.org/Topics/2012/09/12/j_index.html

Collaboration:
Yoshida, Yasuda, Suzuki, Tanaka,...(IPMU),
Ueda,... (NTT), Morii(ISM)

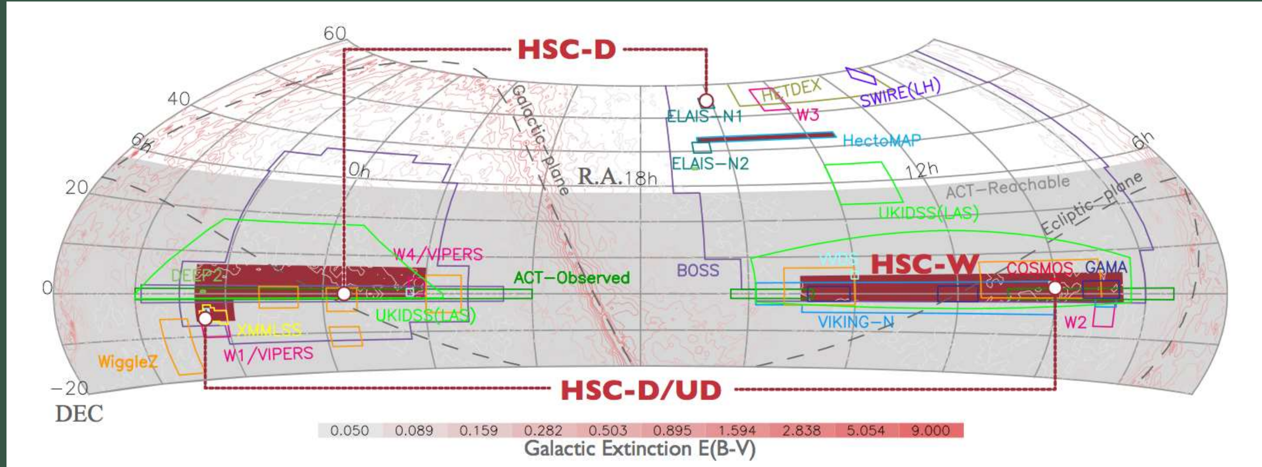


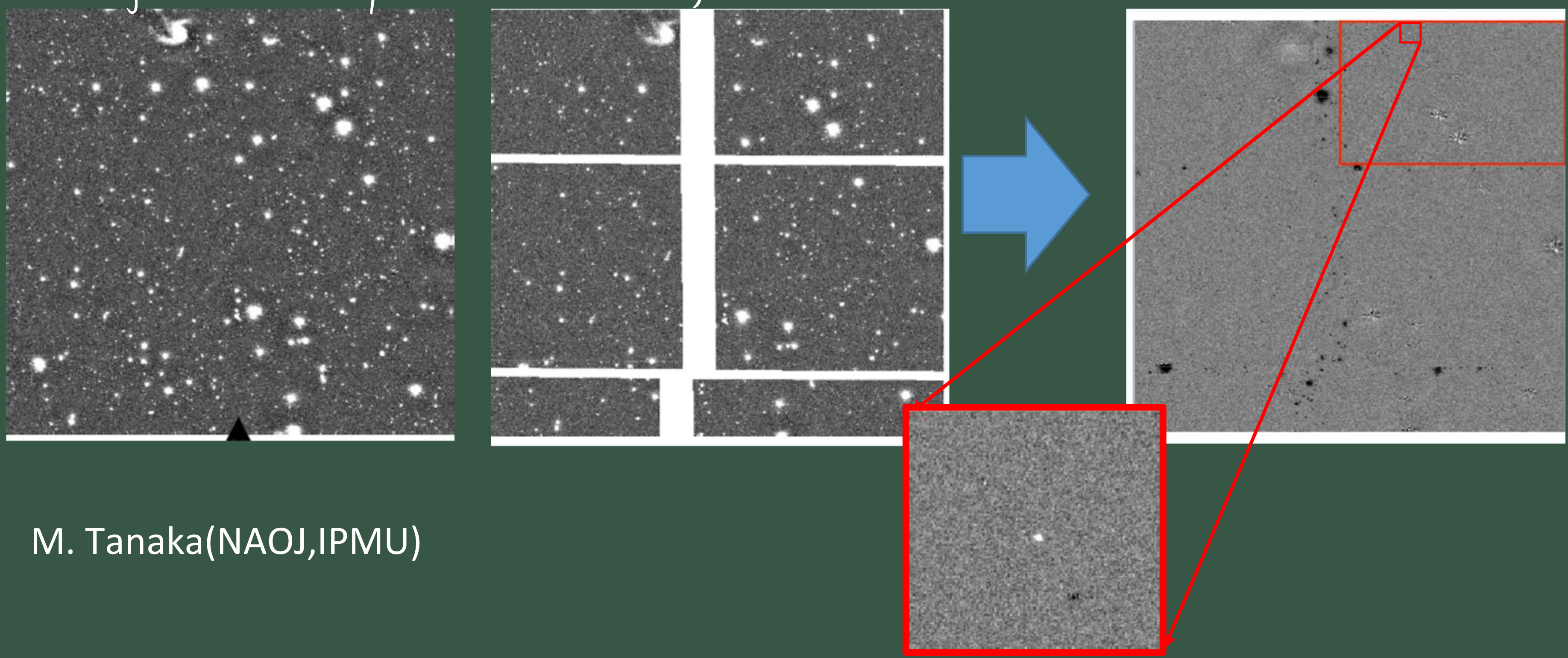
Figure 11: The location of the HSC-Wide, Deep (D) and Ultradeep (UD) fields on the sky in equatorial coordinates. A variety of external data sets and the Galactic dust extinction are also shown. The shaded region is the region accessible from the CMB polarization experiment, ACTPol, in Chile.

http://www.naoj.org/Projects/HSC/img/fields1.png

@ Finding Supernovea with HSC.

Transient Objects

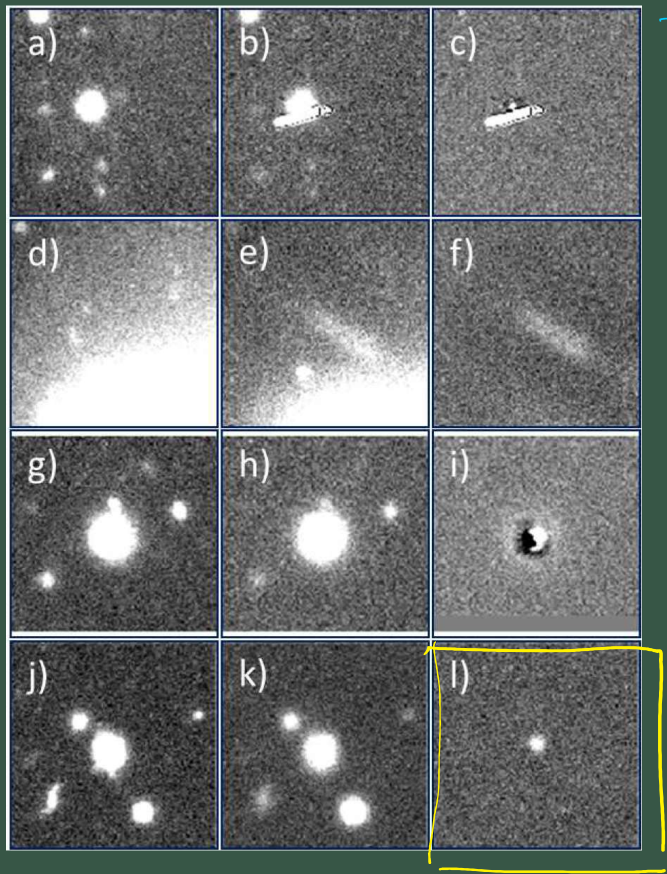
yesterday today Today − yesterday



candidate for supernovae

M. Tanaka(NAOJ,IPMU)

② Classification



transient

bogus

One night data.

\# of transient 61

\# of samples 16,665

<0.5 % are positive

$y = 1$ transient

$y = 0$ bogus

x : features.

$$p(y=1 \mid x) \leq 5.0 \times 10^{-3}$$

@ Tune a classifier which behaves like an astronomers.

Features

Feature variables	Description
Features related with a target object in the difference image	
mag_auto	Magnitude of a target object
magerr_auto	Error of magnitude of the target
a_image	Major radius of the elliptic shape of the target (pixel)
b_image	Minor radius of the elliptic shape of the target (pixel)
elongation	Elongation given by a_image/b_image
fwhm_image	Full width of half maximum (FWHM) of the target (pixel)
flux_auto	Flux of the target
fluxerr_auto	Error of flux of the target
flux_max	Maximal pixel value in the object profile
class_star	Quality flag between 0 and 1 (See Bertin & Arnouts 1996)
flags	Uncertainty of flux measurement (See Bertin & Arnouts 1996)
posi_nega	Flux increased: 1, decreased: 2
fwhm_ratio	FWHM normalized by the average over 4200×4200 pixel region
psffit_sigma_ratio	See text
psffit_peak_ratio	See text
frac_det	See text
density	Number of objects around the target within a square with 120 pixels centered at the target
Features related with reference stars in the template image	
flag_ref	1 if there is any reference star, or 0 if not.
a_ref	a_image value of the nearest reference star (NRS) in the template image (pixel)
b_ref	b_image value of NRS (pixel)
elongation_ref	elongation of NRS
fwhm_ref	fwhm_image of NRS (pixel)
fwhm_norm_ref	fwhm_ref / fwhm_image
mag_ref	mag_auto of NRS
magerr_ref	magerr_auto of NRS
class_star_ref	class_star of NRS
dist_ref	distance between the target object and NRS (pixel)

◎ Classification

If classifier says "bogus" for all,

it is correct with more than 98%.

Data is unbalanced.

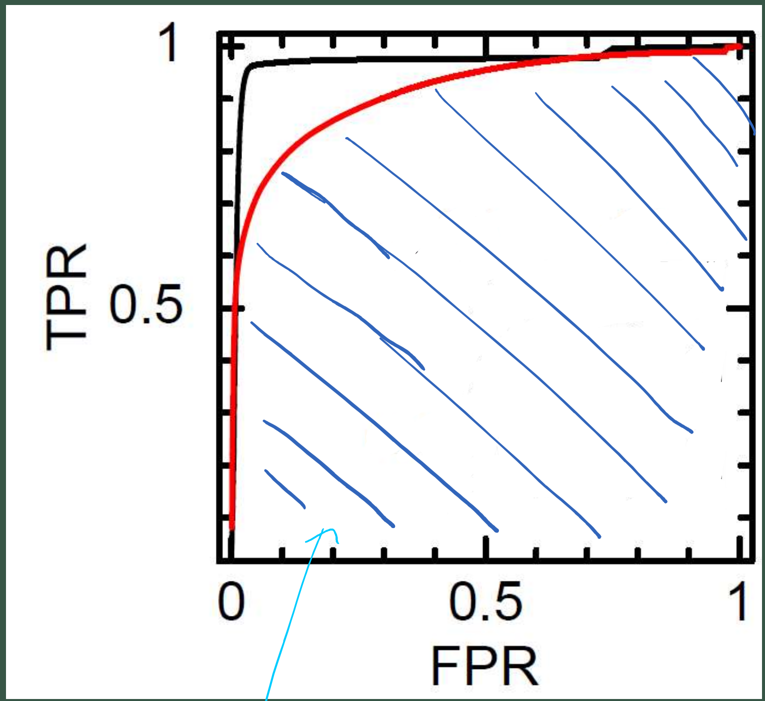
⇓

AUC-based learning

TPR : large ⟶ do not miss any supernova

FPR : low ⟶ # of candidate to check.

ROC and Area Under the Curve (AUC)



- TPR (true positive rate)

  classifier says $y=1$

  and truth is $y=1$

- FPR (false positive rate)

  classifier says $y=1$

  but truth is $y=0$

AUC. it is better if AUC is close to 1.

Learn the classifier to make AUC larger.

- Komori, O. (2011), Ann. Inst. Stat. Math, 63, 961, “A boosting method for maximization of the area under the ROC curve”

◎ Discriminative vs Informative Learning

Y.D. Rubinstein and T. Hastie, "Discriminative vs informative learning," KDD-97

x : observed variables $\quad x \in R^n$

y : class label $\quad y \in \{1, 2, \cdots, K\}$

Discriminative : $p(y \mid x)$

(LDA, Decision tree, Logistic regression $\cdots$)

Informative : $P(x \mid y)$

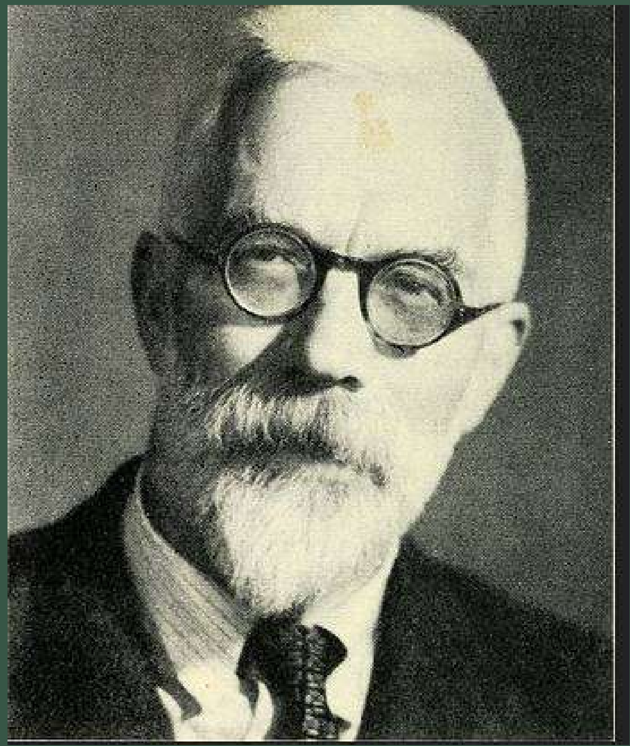
(Gaussian model, etc.)

@ which is better?

	Discriminative	Informative
Objective	$\sum_i \log p_\mu(x_i \mid y_i) \cdot p(y_i)$	$\sum_i \log p_\theta(y_i \mid x_i)$
Estimation	estimation is "easy" (if model is simple)	estimation is "hard"
Advantage	Efficient "if" model is correct	More flexible, robust

Y.D. Rubinstein and T. Hastie, "Discriminative vs informative learning," KDD-97

Bayesian or non-Bayesian



Asymtotic theory

Computional Cost.

@ Non-Bayesian Methods

- Point Estimation
- Hypothesis testing
- Design of Experiments

◎ Non-Bayesian Methods

- ◎ Point Estimation
- ◎ Hypothesis testing
- ◎ Design of Experiments
- ◎ Prediction ⋯ AIC.

@ Non-Bayesian Methods

- @ Point Estimation
- @ Hypothesis testing
- @ Design of Experiments
- @ Prediction ⋯ AIC, BIC, TIC, NIC, GIC

@ Notations

x_i : data, $i = 1, \cdots, N$

M_k : model

$\theta_k \in \Theta_k$: parameter

$\hat{\theta}_k$: MLE

$p(x_i ; \theta)$: likelihood

$l(x_i ; \theta)$: log-likelihood

@ AIC, BIC

$$AIC(k) = \underline{-\frac{1}{N}\sum_{i=1}^{N} l(x_i ; \hat{\theta}_k)} + \underline{\frac{1}{N}} \dim \Theta_k$$

(An/Akaike Information Criterion),

Akaike, H, 1974, IEEE trans. Auto. Cont.

$$BIC(k) = \underline{-\frac{1}{N}\sum_{i=1}^{N} l(x_i ; \hat{\theta}_k)} + \underline{\frac{\log N}{2N}} \dim \Theta_k$$

(Bayesian Information Criterion)

Schwarz, G., 1978, Annals of Statistics

@ AIC

$q(x)$: true dist.

$\hat{q}(x)$: empirical dist $\frac{1}{N}\sum_i \delta(x_i - x)$

$$\hat{\theta} = \arg\max_{\theta} E_{\hat{q}(x)}[\, l(x;\theta) \,] \qquad \text{(MLE)}$$

$$\theta^* = \arg\max_{\theta} E_{q(x)}[\, l(x;\theta) \,] \qquad \text{(true MLE)}$$

@ AIC

Th (White, H. 1982)

$$\sqrt{N}(\hat{\theta}-\theta^*) \sim \mathcal{N}(0, H_{\theta^*}^{-1} G_{\theta^*} H_{\theta^*}^{-1})$$

$$H_\theta = -E_{q(x)}[\partial_{\theta\theta} l(x;\theta)]$$

$$G_\theta = E_{q(x)}[\partial_\theta l(x;\theta)\, \partial_\theta l(x;\theta)]$$

$$(H_\theta = G_\theta \text{ if } q(x) = p(x;\theta))$$

@ AIC

IC : Information Criterion (generalization error)

$$IC = -E_{q(x)}\left[l(x;\hat{\theta}) \right]$$

test error is larger than true log-loss

$$\underline{-E_{q(x)}\left[l(x;\hat{\theta}) \right]} = \underline{-E_{q(x)}\left[l(x;\theta^{*}) \right]} + \frac{1}{2N} \operatorname{tr} H_{\theta^*}^{-1} G_{\theta^*}$$

training error is smaller than true log-loss

$$\underline{-E_{q(x)}\left[l(x;\theta^{*}) \right]} = \underline{-E_{\hat{q}(x)}\left[l(x;\hat{\theta}) \right]} + \frac{1}{2N} \operatorname{tr} H_{\theta^*}^{-1} G_{\theta^*}$$

◎ AIC

$$IC = -E_{q(x)}[l(x;\hat{\theta})]$$
$$= -E_{\hat{q}(x)}[l(x;\hat{\theta})] + \frac{1}{N}\mathrm{tr}\,\underline{H_{\theta^*}^{-1} G_{\theta^*}}$$

TIC (Takeuchi, 1976)

$$TIC = -E_{\hat{q}(x)}[l(x;\hat{\theta})] + \frac{1}{N}\mathrm{tr}\,\underline{H_{\hat{\theta}}^{-1} G_{\hat{\theta}}}$$

AIC

$$AIC = -E_{\hat{q}(x)}[l(x;\hat{\theta})] + \frac{1}{N}\dim\hat{\theta}$$

@ AIC : consistency

Assume $m_1 \subset m_2 \subset \cdots \subset m_K$ and $p(x; \theta_{k^*}^*) = q(x)$

m_{k^*} includes the true dist.

for model selected with AIC

$$\lim_{N \to \infty} p(\hat{k} > k^*) > 0$$

For BIC

$$\lim_{N \to \infty} p(\hat{k} > k^*) = 0$$

◎ consistency

$$-E_{\hat{q}(x)}\left[l(x;\hat{\theta}_k)\right] + \frac{c(N)}{N}\dim\hat{\theta}_k$$

($c(N) = \log\log N$ is the critical value)

AIC is an approximation of the generalization error.

@ BIC

$$x_1, \cdots x_N \sim p(x_1, \cdots, x_N) = \int p(x_1, \cdots x_N | \theta) \pi(\theta) d\theta.$$

$$\mathrm{BIC} = -\frac{1}{N} l(x_1, \cdots, x_N)$$

$$\simeq -E_{\hat{q}(x)}[l(x|\hat{\theta})] + \frac{\log N}{2N} \dim \hat{\theta}$$

@ from AIC to BIC

$$p(x_1, \cdots, x_N) = p(x_N | x_1, \cdots, x_{N-1})\, p(x_{N-1} | x_1, \cdots, x_{N-2}) \times$$

$$\cdots \times p(x_3 | x_1, x_2) \cdot p(x_2 | x_1) \cdot p(x_1)$$

$$\underline{-l(x_1, \cdots, x_N)} = -\sum_{i=1}^{N-1} l(x_{i+1} | x_1, \cdots, x_i) - l(x_1)$$

◎ from AIC to BIC

AIC

$$-\ell(x_{i+1} \mid x_1, \cdots, x_i) \simeq -E_{q(x)}[\ell(x \mid \theta^*)] + \frac{1}{2i} \dim \theta^*$$

$$-\frac{1}{N}\ell(x_1, \cdots, x_N) = -\frac{1}{N}\sum_{i=1}^{N-1} \ell(x_{i+1} \mid x_1, \cdots, x_i) - \ell(x_1)$$

$$\simeq -E_{q(x)}[\ell(x \mid \theta^*)] + \frac{1}{N}\sum_i \frac{1}{2i} \dim \theta^*$$

$$\simeq -E_{q(x)}[\ell(x \mid \theta^*)] + \frac{\log N}{2N} \dim \theta^*$$

@ from AIC to BIC

$$-\frac{1}{N}\ell(x_1,\cdots,x_N) \simeq -\underline{E_{q(x)}[\ell(x;\theta^*)]} + \frac{\log N}{2N}\dim\theta^*$$

$$\simeq -\underline{E_{\hat{q}(x)}[\ell(x;\hat{\theta})]} + \frac{1+\log N}{2N}\dim\hat{\theta}$$

$$\simeq -E_{\hat{q}(x)}[\ell(x;\hat{\theta})] + \frac{\log N}{2N}\dim\hat{\theta}$$

$$= \mathrm{BIC}$$

@ Conclusion

- ISM & IPMU .. (MOU in Jan/2016)
- Astronomy & particle physics.
- more people, meetings, discussions.