بسم الله الرحمن الرحیم

کارگاه مدلسازی داده:

Bayesian and Likelihood analysis

سیدمحمدصادق موحد

دانشکده فیزیک - دانشگاه شهید بهشتی

پژوهشکده فیزیک، پژوهشگاه دانشهای بنیادی (IPM)

گروه کیهان شناسی محاسباتی (GCC-SBU)

آزمایشگاه میان رشته ای ابن سینا

http://facultymembers.sbu.ac.ir/movahed/

کارگاه مدلسازی داده (شماره ۱)

۹۷/۲/۲۶

مهمترین منابع مورد استفاده

Institute for Research in Fundamental Sciences

دانشکده فیزیک

Books:

- Data analysis: A Bayesian Tutorial, by D.S. Sivia & J. Skilling, Oxford science Publication, 2010
- Data reduction and error analysis for the physical sciences, P. R. Bevington & D. K. Robinson, McGrawHill, 2003
- Error of Observations and their Treatment, J. Topping, 1972.
- Practical Physics, G. L. Squires, 1985.
- Data analysis in cosmology, "Lecture notes in physics 665"
- Mathematical statistics with applications, J.E. Freund's, Pearson education, 2005,
- All of Statistics, L. Wasserman, Springer, 2004
- Statistics, R.J. Barlow, Wiley, 2002.
- Introduction to statistics and Data analysis for Physicist, Gerhard Bohm, Günter Zech, 2010.
- Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data (Princeton Series in Modern Observational Astronomy), Željko Ivezic´, Andrew J. Connolly, Jacob T. VanderPlas, and Alexander Gray, PRINCETON UNIVERSITY PRESS, 2014
- Algebraic Topology of Random Fields and Complexes, Research thesis, Omer Bobrowski, 2012
- Random Fields and Geometry, Adler, R. J., Taylor, Jonathan E., Springer 2007.
- The Geometry of Random fields, Adler, 1981.
- GEOMETRY, TOPOLOGY and Physics, M. Nakahara, 2003
- Topological Complexity of Smooth Random Functions, Robert J. Adler, Jonathan E. Taylor, Springer 2009.

مهمترین منابع مورد استفاده

Papers and Lectures:

- Verde, Licia. "A practical guide to basic statistical techniques for data analysis in cosmology." arXiv preprint arXiv:0712.3028 (2007).
- Verde, Licia. "Statistical methods in cosmology." Lectures on Cosmology. Springer, Berlin, Heidelberg, 2010. 147-177.
- Liddle, Andrew R. "How many cosmological parameters." Monthly Notices of the Royal Astronomical Society 351.3 (2004): L49-L53.
- Heavens, Alan. "Statistical techniques in cosmology." arXiv preprint arXiv: 0906.0664 (2009).
- Efron, Bradley. "Bayesians, frequentists, and physicists." PHYSTAT2003: Statistical Problems in Particle Physics, Astrophysics, and Cosmology, SLAC, Stanford CA (2003): 17-24
- arXiv:physics/0406120
- **http://facultymembers.sbu.ac.ir/movahed/**

Institute for Research in Fundamental Sciences

مهمترین منابع مورد استفاده

دانشکده فیزیک

facultymembers.sbu.ac.ir/movahed/

SHAHID BEHESHTI UNIVERSITY

Department of Physics

MAIN MENU

- Home
- About Me
- Contact me
- Researches of Interest
- Publications
- Students
- Courses
- Talks & Presentations
- Useful Stuff
- Group Meetings
- My CV
- Collaborations
- Other Activities
- Photo
- Login
- پارسی (Persian)

Seyed Mohammad Sadegh Movahed Academic Homepage

News

NEW My program in the Autumn semester (1396-1397 (2017-2018)) (Download)

NEW Some proposals for Master Researches in my group (عناوین تعدادی پیشنهاده پایان نامه دوره کارشناسی ارشد) (Download)

Some proposals for Ph.D. Researches in my group (عناوین تعدادی پیشنهاده تز دوره دکتری) (Download)

Some proposed Books for the relation between Physics and Philosophy.

فهرست مطالب

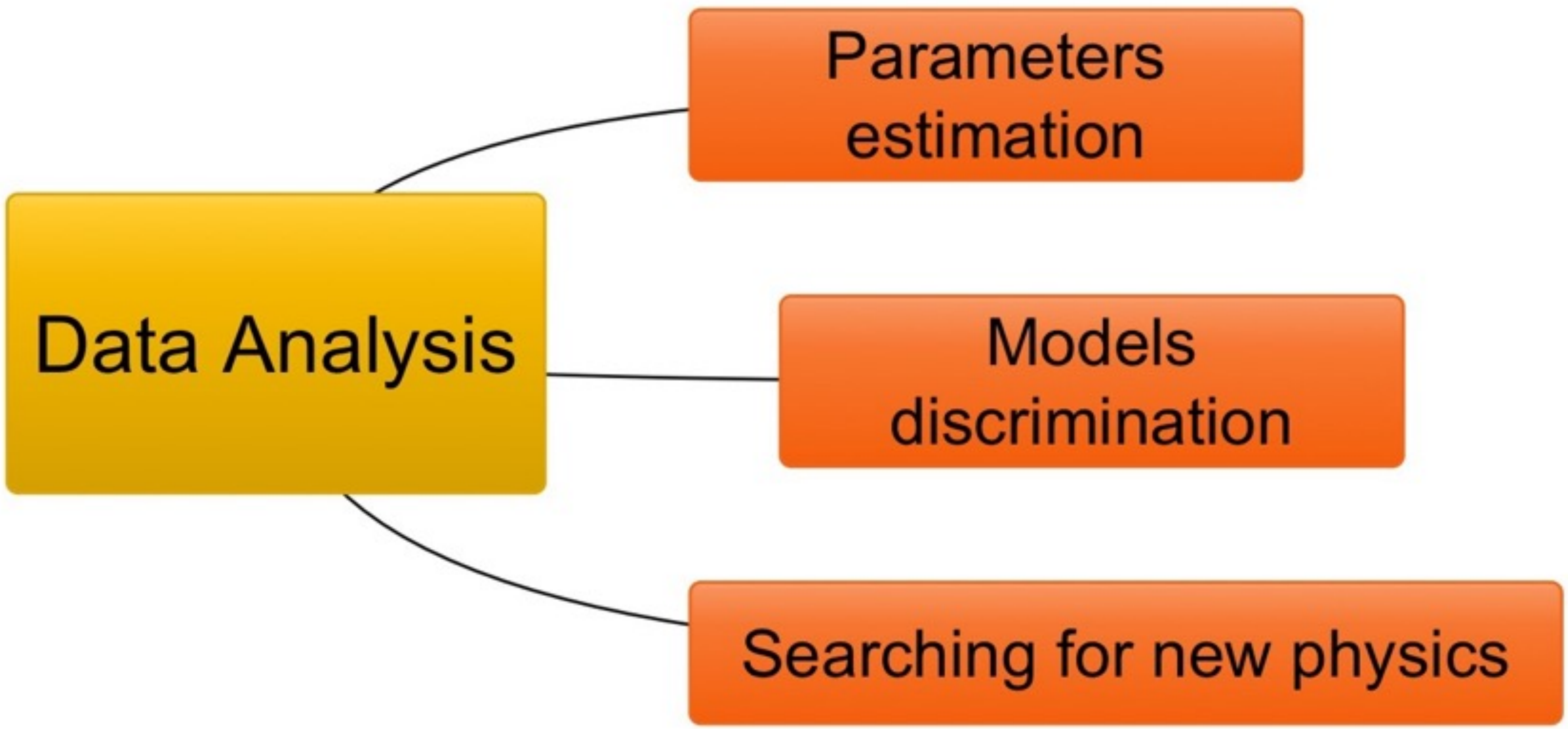
مقدمه

مدل کردن داده های حاصل از اندازه گیری
Data modeling

از داده ها چه چیزهایی یاد می گیریم؟

۱) منشاء و ماهیت
۲) چگونگی تحول
۳) احتمالاً پیش بینی
.
.

مدل کردن داده های حاصل از اندازه گیری
Data modeling



مراحل مدل کردن داده ها

۱) اندازه گیری (مستقیم یا شبیه سازی) (Measurment)

۲) برآورد خطا و ارزیابی انتشار خطا بر روی کمیت های ثانویه (Error estimation and error propagation)

۳) تدوین مدل با توجه به تابع مناسب

(model selection) (Merit function)

۴) انتخاب بینش تعیین مقادیر آزاد مدل

(Bayesian or Frequentist)

۵) استفاده از روشهای تعیین مقادیر کمیت های آزاد مدل و البته حوزه اعتبار آنها

(Parameter estimation and confidence interval)

۶) تعیین خوبی مدل Goodness of fit

تعیین مقدار برای کمیتهای آزاد

رهیافت Bayesian

۱) در این روش داده ها و نتایج بخشی از یک آنسامبل هستند

۲) کمیت های آزاد مدل مجهول هستند که ما تنها به صورت احتمالی می توانیم مقدار آنها را تعیین کنیم

۳) هر گونه اطلاعات اولیه در این رهیافت قابل استفاده هستند

۴) سازوکاری برای رهایی از کمیتهای اضافی (Nuisance) دارد

رهیافت Frequentist

۱) در این روش داده ها و نتایج ریزحالت یک پیکربندی به حساب می آیند و داده ها تکرار پذیرند

۲) کمیت های مدل مجهول ولی ثابت هستند

۳) هیچ گونه اطلاعی از مدل مورد استفاده نیست

۴) سازوکاری برای رهایی از کمیتهای اضافی (Nuisance) ندارد

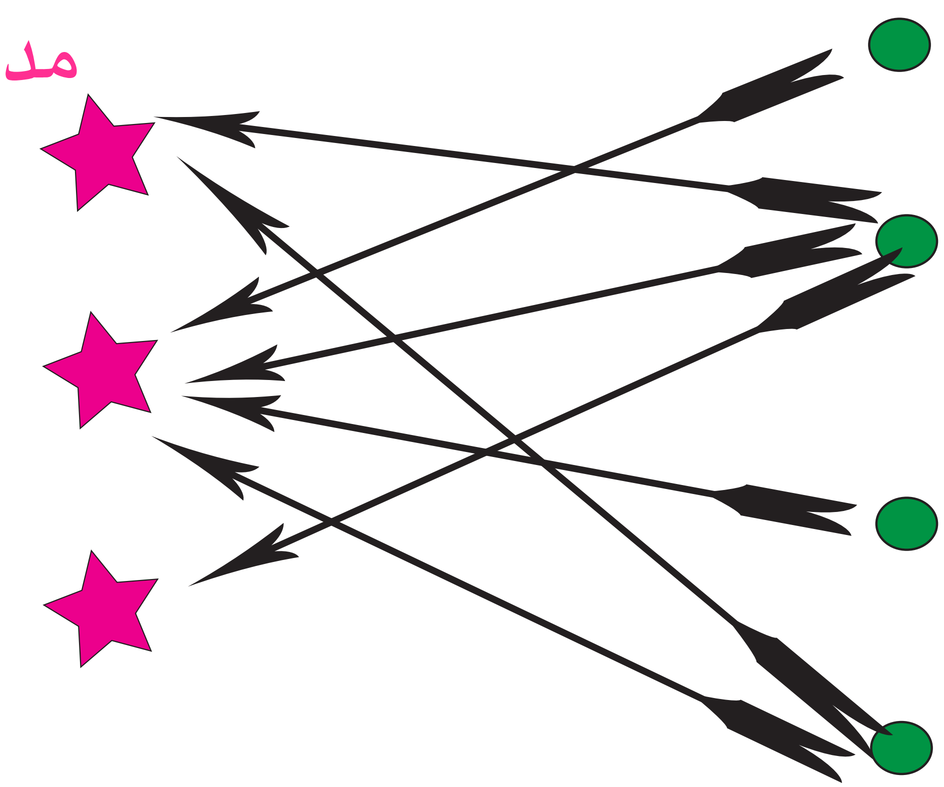
منطق استقرایی

Inductive logic: Bayesian

مشاهده رفتار تک تک اجزا منجر
به قانون کلی می شود. تمام
انسانهای مشاهده شده فانی
بودند در نتیجه انسان فانی است

مشاهدات

مقادیر ممکن
برای پارامترهای
مدل



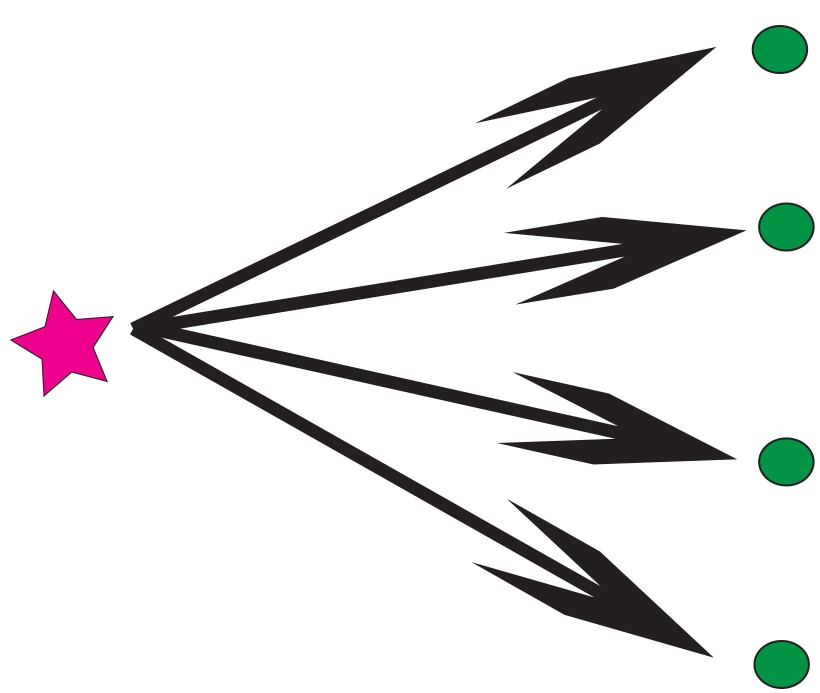
منطق قیاسی

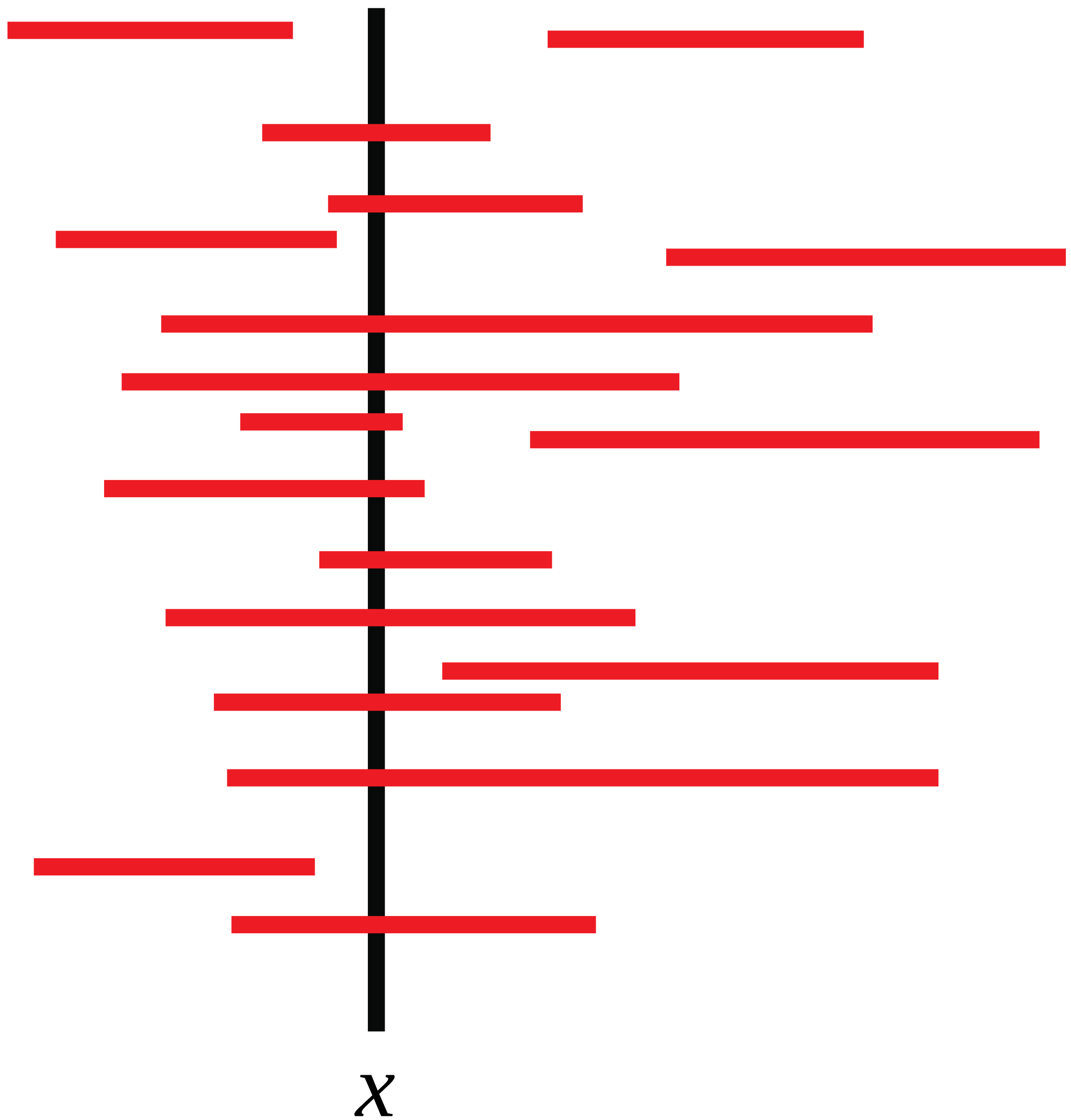
Deductive logic: Frequentist

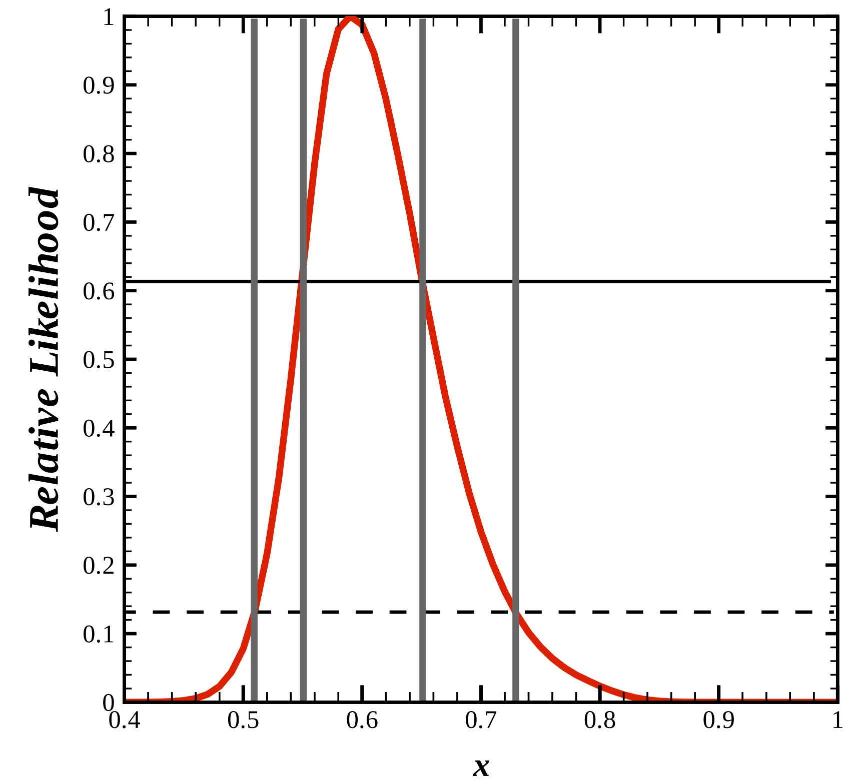
مشاهده مبتنی بر واقعیت های شناخته شده
که در کنار هم قرار می گیرند و نهایتا به
نتیجه می رسند. انسانها فانی هستند سقراط
انسان است لذا سقراط فانی است

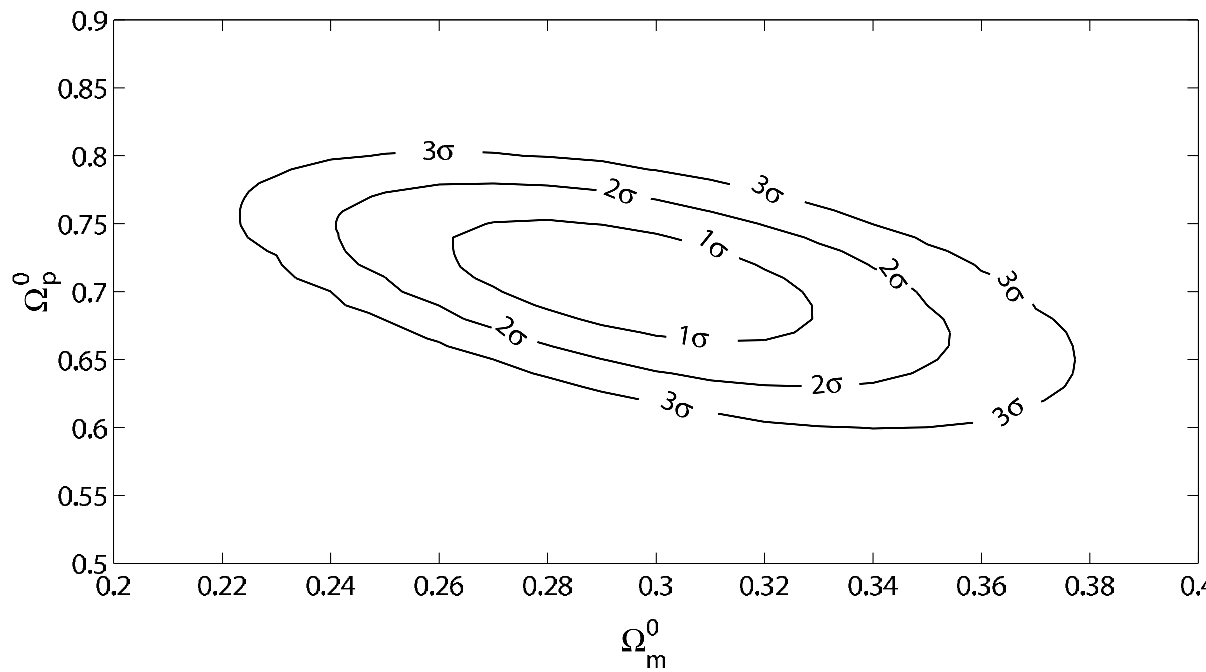
مقدار ثابت
برای مدل

مشاهدات









بر اساس نگرش frequentist
برای مثال ۹۰ درصد
مشاهدات به مقدار ثابت x
منجر شده است

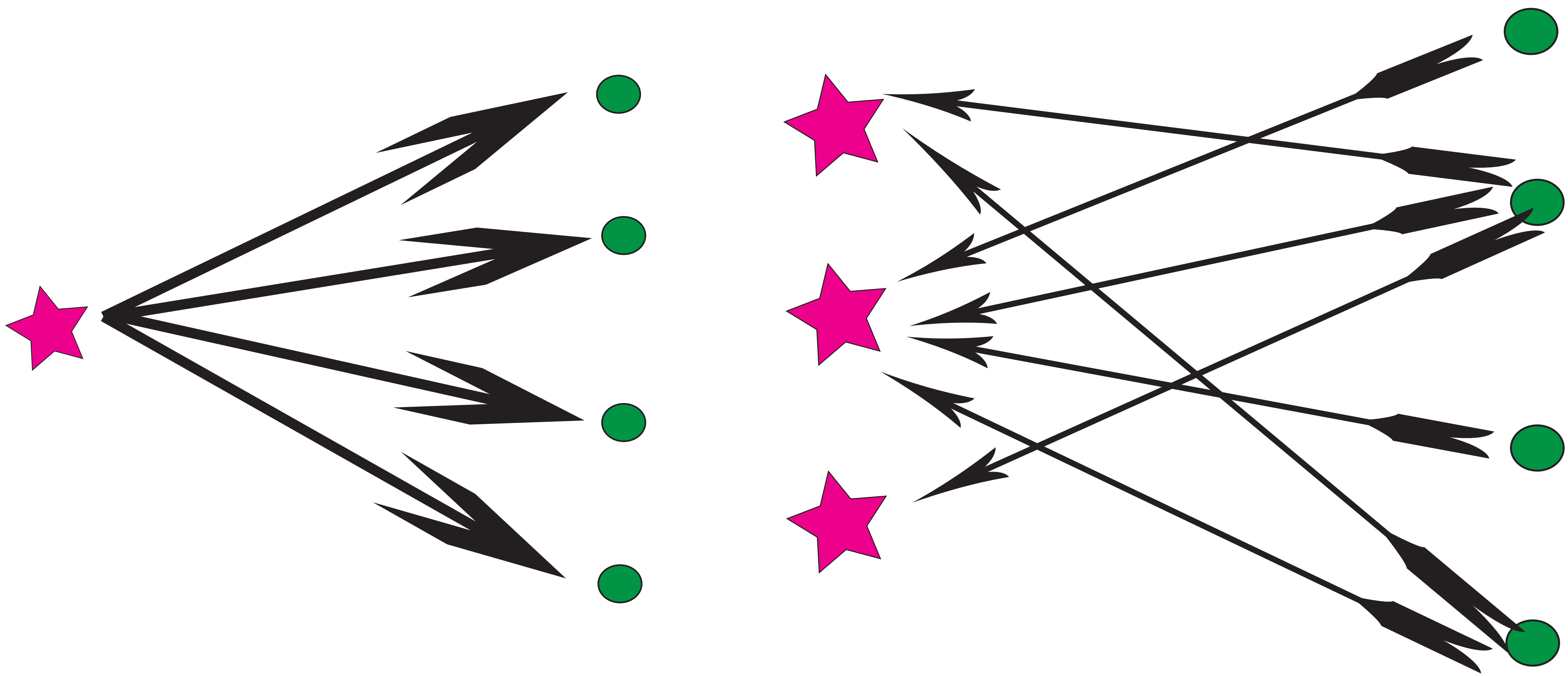
بر اساس نگرش Bayesian با
احتمال ۹۰ درصد مقدار دقیق
کمیت x دراین بازه قرار می گیرد

A history about Bayesian strategy

- By Thomas Bayes (1702-1761)
- Pierre Simon Laplace (1812)
- Fisher, Neyman, Wald,...
- Gelfand and Smith (1990)
- Now

هرچند که این ایده بسیار قدیمی است اما استفاده کاربردی از آن در دو دهه اخیر آغاز شده است چرا؟

با توجه به روشها می توان دریافت در روش Bayesian میزان و حجم محاسبات به مراتب بیشتر است بنابراین تا گسترش سخت افزار ها این ایده به تعویق افتاد



Bayesian approach

$D:\{x_i, y_i\} \quad i=1,\ldots,N$

مشاهدات: داده ها

$\Theta:\{\theta_j\} \qquad j=1,\ldots,M \qquad \rightarrow \qquad Y(x,\Theta)=\sum_{j=1}^{M}\theta_j f_j(x)$

برای مثال
یک مدل

$p(\Theta \mid D)$

احتمال یافتن مدل به شرطی که داده ها به دست آمده باشند. با چه احتمالی
مشاهدات به مدل منجر میشود Posterior

$p(D \mid \Theta)$

احتمال به دست آمدن مشاهدات به شرطی که مدل مشخص باشد. مدل مشخص با
چه احتمالی مشاهدات را به دست می دهد. Likelihood

$p(\Theta)$

هر گونه اطلاعات اولیه در مورد مقادیر کمیتهای مدل در این تابع توزیع وجود دارد.
Prior

$$p(\Theta \mid D) = \frac{p(\Theta, D)}{p(D)} = \frac{p(D, \Theta)}{p(D)} = \frac{p(D \mid \Theta) p(\Theta)}{p(D)}$$

$$p(D) = \int d\Theta \ p(D, \Theta) = \int d\Theta \ p(D \mid \Theta) p(\Theta)$$

$$p(\Theta \mid D) = \frac{p(D \mid \Theta) p(\Theta)}{\int d\Theta \ p(D \mid \Theta) p(\Theta)}$$

در حالتی که اطلاعات اولیه ای برای کمیت های مدل نداشته باشیم

Posterior=Likelihood

ویژگی های رهیافت بیزی

مثال: فرض کنید که برای یک دسته مشاهده داده شده، دو مدل داریم. کدامیک بامثال احتمال بیشتری انتخاب می شوند؟

ویژگی های رهیافت بیزی (I)

مثال: فرض کنید که برای یک دسته مشاهده داده شده، دو مدل داریم. کدامیک بامثال احتمال بیشتری انتخاب می شوند؟

$$D:\{x_i, y_i\}$$

$$\Theta_1 \quad \Theta_2$$

$$p(\Theta_1 \mid D) = \frac{p(D \mid \Theta_1) p(\Theta_1)}{\int d\Theta\; p(D \mid \Theta_1) p(\Theta_1)}$$

$$p(\Theta_2 \mid D) = \frac{p(D \mid \Theta_2) p(\Theta_2)}{\int d\Theta\; p(D \mid \Theta_2) p(\Theta_2)}$$

$$\frac{p(\Theta_1 \mid D)}{p(\Theta_2 \mid D)} = \frac{\dfrac{p(D \mid \Theta_1) p(\Theta_1)}{\int d\Theta\; p(D \mid \Theta_1) p(\Theta_1)}}{\dfrac{p(D \mid \Theta_2) p(\Theta_2)}{\int d\Theta\; p(D \mid \Theta_2) p(\Theta_2)}} = \frac{p(D \mid \Theta_1) p(\Theta_1)}{p(D \mid \Theta_2) p(\Theta_2)}$$

$$if \quad p(\Theta_1) = p(\Theta_2)$$

مدلی که درست نمایی بزرگتری داشته باشد با احتمال بیشتری توسط داده ها انتخاب می شود

ویژگی های رهیافت بیزی (II)

بهبود توزیع احتمال پارامترهای مدل

$$D_1 : \{x_i, y_i\}$$

$$\Theta : \{\theta_j\}$$

$$p(\Theta \mid D_1) = \frac{p(D_1 \mid \Theta) p(\Theta)}{\int d\Theta \; p(D_1 \mid \Theta) p(\Theta)}$$

$p(\Theta \mid D_1)$ به عنوان تابع اطلاعات اولیه در تحلیل جدید به حساب می آید

$$p(\Theta D_1 \mid D_2) = \frac{p(D_2 \mid \Theta D_1) p(\Theta \mid D_1)}{p(D_2 \mid D_1)}$$

$$p(D_2 \mid \Theta D_1) = \frac{p(D_2, D_1 \mid \Theta)}{p(D_1 \mid \Theta)}$$

$$p(\Theta D_1 \mid D_2) = \frac{p(D_2, D_1 \mid \Theta)}{p(D_1 \mid \Theta)} \times \frac{p(D_1 \mid \Theta) p(\Theta)}{p(D_1)} \times \frac{1}{p(D_2 \mid D_1)}$$

$$= \frac{p(D_2, D_1 \mid \Theta) p(\Theta)}{p(D_1, D_2)}$$

یعنی مشاهدات جدید را برای نتایج به دست آمده با مشاهدات
قدیمی به کار برده تا نهایتاً مقادیر کمیت ها به دست آیند

بیسکویت و شکلات

فرض کنید دو کاسه کاملاً شبیه به هم داریم. کاسه A دارای ۱۰ شکلات و ۳۰ بیسکویت است و کاسه B شامل ۲۰ عدد از هر کدام است.

سوال: با چه احتمالی بیسکویت از کاسه A انتخاب می شود؟

$$P(\theta_A \mid x) = \frac{P(x \mid \theta_A)P(\theta_A)}{p(x)} = \frac{P(x \mid \theta_A)P(\theta_A)}{p(x \mid \theta_A)p(\theta_A) + p(x \mid \theta_B)p(\theta_B)}$$

$$p(\theta_A) = p(\theta_B) = 1/2$$ چون دو کاسه (مدل) شبیه به هم هستند

$$p(x \mid \theta_A) = \frac{30}{40}$$

$$p(x \mid \theta_B) = \frac{20}{40}$$

$$P(\theta_A \mid x) = 0.6$$ احتمال مشاهده شدن بیسکویت از کاسه A بیشتر است

$$P(\theta_B \mid x) = 0.4$$

کاسه B زیباتر از A است پس مشاهده گر بیشتر گرایش به انتخاب کاسه B دارد پس:

$$P(\theta_A \mid x) = \frac{P(x \mid \theta_A)P(\theta_A)}{p(x)} = \frac{P(x \mid \theta_A)P(\theta_A)}{p(x \mid \theta_A)p(\theta_A) + p(x \mid \theta_B)p(\theta_B)}$$

$$p(\theta_A) = \frac{1}{4}$$

$$p(\theta_B) = \frac{3}{4}$$

$$p(x \mid \theta_A) = \frac{30}{40}$$

$$p(x \mid \theta_B) = \frac{20}{40}$$

$$P(\theta_A \mid x) = 0.34$$

$$P(\theta_B \mid x) = 0.66$$

احتمال مشاهده شدن بیسکویت از کاسه B بیشتر است

اکنون به دنبال این هستیم که
Posterior
را بیشینه کنیم

این رهیافت دقیقاً معادل یافتن
Best value for free parameters

بیشینه سازی تابع Posterior

$$D:\{x_i, y_i\} \quad i = 1,\dots,N$$

$$\Theta:\{\theta_j\} \qquad Y(x;\{\Theta\}) = \sum_{k=1}^{M} \theta_k f_k(x)$$

$$p(\Theta\,|\,D) = \frac{p(D\,|\,\Theta)\,p(\Theta)}{\int d\Theta\; p(D\,|\,\Theta)\,p(\Theta)}$$

$$if \quad p(\Theta) = cts$$

در صورتی که تابع prior وجود نداشته باشد در نتیجه تحلیل posterior به تحلیل Likelihood تبدیل می شود

$$p(\Theta\,|\,D) \sim p(D\,|\,\Theta)$$

$$p(D\,|\,\Theta) \equiv \prod_{i=1}^{N} P_i = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{\left[y_i - Y(x_i;\{\Theta\})\right]^2}{2\sigma_i^2}\right)$$

۱) با فرض اینکه اندازه گیری ها از یکدیگر مستقل باشند

۲) با توجه به قضیه حد مرکزی می توان در نظر گرفت که هر نقطه اندازه گیری حول مقدار بهینه اش به صورت گوسی توزیع شده است

$$\chi^2(\{\Theta\}) \equiv \sum_{i=1}^{N} \frac{\left[y_i - Y(x_i;\{\Theta\})\right]^2}{\sigma_i^2} = \sum_{i=1}^{N} \frac{\left[y_i - \sum_{k=1}^{M} \theta_k f_k(x_i)\right]^2}{\sigma_i^2}$$

$$p(D\,|\,\Theta) \sim e^{-\frac{\chi^2(\{\Theta\})}{2}}$$

بیشینه شدن Likelihood معادل با کمینه شدن chi^2 است

$$\chi^2(\{\Theta\}) \equiv \sum_{i=1}^{N} \frac{\left[y_i - \sum_{k=1}^{M} \theta_k f_k(x_i) \right]^2}{\sigma_i^2}$$

$$b_i \equiv \frac{y_i}{\sigma_i}$$

$$A_{ik} \equiv \frac{f_k(x_i)}{\sigma_i}$$

$$Min\left\| b - A\Theta \right\| \rightarrow \Theta_{Best}$$

Normal Equations
Singular Value Decomposition(SVD)

کمینه کردن chi^2

$$\chi^2(\{\Theta\}) = \sum_{i=1}^{N} \frac{\left[y_i - \sum_{k=1}^{M} \theta_k f_k(x_i) \right]^2}{\sigma_i^2} \quad \rightarrow \quad \chi^2(\{\Theta\} = \{\Theta\}_{Best}) = \chi^2_{\min}$$

$$\left.\frac{\partial \chi^2(\{\Theta\})}{\partial \theta_j}\right|_{\{\Theta\}=\{\Theta_{Best}\}} = 0$$

$$= \sum_{i=1}^{N} \left[y_i - \sum_{k=1}^{M} \theta_k f_k(x_i) \right] \times \frac{f_j(x_i)}{\sigma_i^2}$$

$$\sum_{i=1}^{N} \frac{y_i f_j(x_i)}{\sigma_i^2} = \sum_{i=1}^{N} \sum_{k=1}^{M} \frac{\theta_k f_k(x_i) f_j(x_i)}{\sigma_i^2}$$

$$\beta_j \equiv \sum_{i=1}^{N} \frac{y_i f_j(x_i)}{\sigma_i^2}$$

$$\alpha_{jk} \equiv \sum_{i=1}^{N} \frac{f_j(x_i) f_k(x_i)}{\sigma_i^2}$$

$$\beta_j = \sum_{k=1}^{M} \alpha_{jk} \theta_k$$

$$[\Theta] = [\alpha]^{-1} [\beta]$$

اکنون سوال این است که حوزه اعتبار کمیت Θ چقدر است؟

$$\beta_j \equiv \sum_{i=1}^{N} \frac{y_i f_j(x_i)}{\sigma_i^2}$$

$$\alpha_{jk} \equiv \sum_{i=1}^{N} \frac{f_j(x_i) f_k(x_i)}{\sigma_i^2}$$

$$[\Theta] = [\alpha]^{-1}[\beta]$$

$$\sigma^2(\theta_i) = \sum_{j=1}^{N} \sigma_j^2 \left(\frac{\partial \theta_i}{\partial y_j}\right)^2 + \sum_{j=1}^{N}\sum_{l \neq j}^{N} \left(\frac{\partial \theta_i}{\partial y_j}\right)\left(\frac{\partial \theta_i}{\partial y_l}\right) Cov\left(y_j y_l\right) \quad \rightarrow \quad \left(\frac{\partial \theta_i}{\partial y_j}\right) = \sum_{k=1}^{M} [\alpha_{ik}]^{-1} \frac{\partial \beta_k}{\partial y_j}$$

$$\left(\frac{\partial \theta_i}{\partial y_j}\right)^2 = \sum_{l=1}^{M}\sum_{k=1}^{M} [\alpha_{ik}]^{-1} [\alpha_{il}]^{-1} \frac{\partial \beta_k}{\partial y_j}\frac{\partial \beta_l}{\partial y_j}$$

$$\sigma^2(\theta_i) = \sum_{l=1}^{M}\sum_{k=1}^{M} [\alpha_{ik}]^{-1} [\alpha_{il}]^{-1} \sum_{j=1}^{N} \sigma_j^2 \frac{\partial \beta_k}{\partial y_j}\frac{\partial \beta_l}{\partial y_j} \quad \rightarrow \quad \sigma^2(\theta_i) = \sum_{l=1}^{M}\sum_{k=1}^{M} [\alpha_{ik}]^{-1} [\alpha_{il}]^{-1} [\alpha_{kl}]$$

$$\rightarrow \quad \sum_{k=1}^{M} [\alpha_{il}]^{-1} [\alpha_{kl}] = \delta_{ik}$$

$$\sigma^2(\theta_i) = [\alpha_{ii}]^{-1}$$

polynomial function

$$D : \{x_i, y_i\} \rightarrow \quad i = 1, \ldots, N$$

$$\Theta : \{\theta_k\} \quad \rightarrow \quad k = 1, \ldots, M \quad \rightarrow \quad Y(x;\Theta) = \theta_1 + \theta_2 x^1 + \ldots + \theta_M x^{M-1}$$

$$P(D \mid \Theta) = \prod_{i=1}^{N} P(D_i \mid \Theta) = \prod_{i=1}^{N} \left(\frac{1}{\sigma_i \sqrt{2\pi}} \right) \exp\left(-\frac{1}{2} \sum_{i=1}^{N} \frac{\left[y_i - \sum_{k=1}^{M} \theta_k x_i^{k-1} \right]^2}{\sigma_i^2} \right)$$

$$for \quad M = 3 \quad \chi^2(\theta_1, \theta_3, \theta_3) = \sum_{i=1}^{N} \frac{\left[y_i - \sum_{k=1}^{3} \theta_k x_i^{k-1} \right]^2}{\sigma_i^2}$$

$$0 = \frac{\partial \chi^2(\theta_1, \theta_2, \theta_3)}{\partial \theta_1} = -2 \sum_{i=1}^{N} \frac{\left[y_i - \sum_{k=1}^{3} \theta_k x_i^{k-1} \right]}{\sigma_i^2}$$

$$0 = \frac{\partial \chi^2(\theta_1, \theta_2, \theta_3)}{\partial \theta_2} = -2 \sum_{i=1}^{N} x_i \frac{\left[y_i - \sum_{k=1}^{3} \theta_k x_i^{k-1} \right]}{\sigma_i^2}$$

$$0 = \frac{\partial \chi^2(\theta_1, \theta_2, \theta_3)}{\partial \theta_3} = -2 \sum_{i=1}^{N} x_i^2 \frac{\left[y_i - \sum_{k=1}^{3} \theta_k x_i^{k-1} \right]}{\sigma_i^2}$$

$$\theta_1 = \frac{1}{\Delta}\begin{vmatrix} \sum_{i=1}^{N}\frac{y_i}{\sigma_i^2} & \sum_{i=1}^{N}\frac{x_i}{\sigma_i^2} & \sum_{i=1}^{N}\frac{x_i^2}{\sigma_i^2} \\ \sum_{i=1}^{N} y_i\frac{x_i}{\sigma_i^2} & \sum_{i=1}^{N}\frac{x_i^2}{\sigma_i^2} & \sum_{i=1}^{N}\frac{x_i^3}{\sigma_i^2} \\ \sum_{i=1}^{N} y_i\frac{x_i^2}{\sigma_i^2} & \sum_{i=1}^{N}\frac{x_i^3}{\sigma_i^2} & \sum_{i=1}^{N}\frac{x_i^4}{\sigma_i^2} \end{vmatrix}$$

$$\theta_2 = \frac{1}{\Delta}\begin{vmatrix} \sum_{i=1}^{N}\frac{1}{\sigma_i^2} & \sum_{i=1}^{N}\frac{y_i}{\sigma_i^2} & \sum_{i=1}^{N}\frac{x_i^2}{\sigma_i^2} \\ \sum_{i=1}^{N}\frac{x_i}{\sigma_i^2} & \sum_{i=1}^{N} y_i\frac{x_i}{\sigma_i^2} & \sum_{i=1}^{N}\frac{x_i^3}{\sigma_i^2} \\ \sum_{i=1}^{N}\frac{x_i^2}{\sigma_i^2} & \sum_{i=1}^{N} y_i\frac{x_i^2}{\sigma_i^2} & \sum_{i=1}^{N}\frac{x_i^4}{\sigma_i^2} \end{vmatrix}$$

$$\theta_3 = \frac{1}{\Delta}\begin{vmatrix} \sum_{i=1}^{N}\frac{1}{\sigma_i^2} & \sum_{i=1}^{N}\frac{x_i}{\sigma_i^2} & \sum_{i=1}^{N}\frac{y_i}{\sigma_i^2} \\ \sum_{i=1}^{N}\frac{x_i}{\sigma_i^2} & \sum_{i=1}^{N}\frac{x_i^2}{\sigma_i^2} & \sum_{i=1}^{N} y_i\frac{x_i}{\sigma_i^2} \\ \sum_{i=1}^{N}\frac{x_i^2}{\sigma_i^2} & \sum_{i=1}^{N}\frac{x_i^3}{\sigma_i^2} & \sum_{i=1}^{N} y_i\frac{x_i^2}{\sigma_i^2} \end{vmatrix}$$

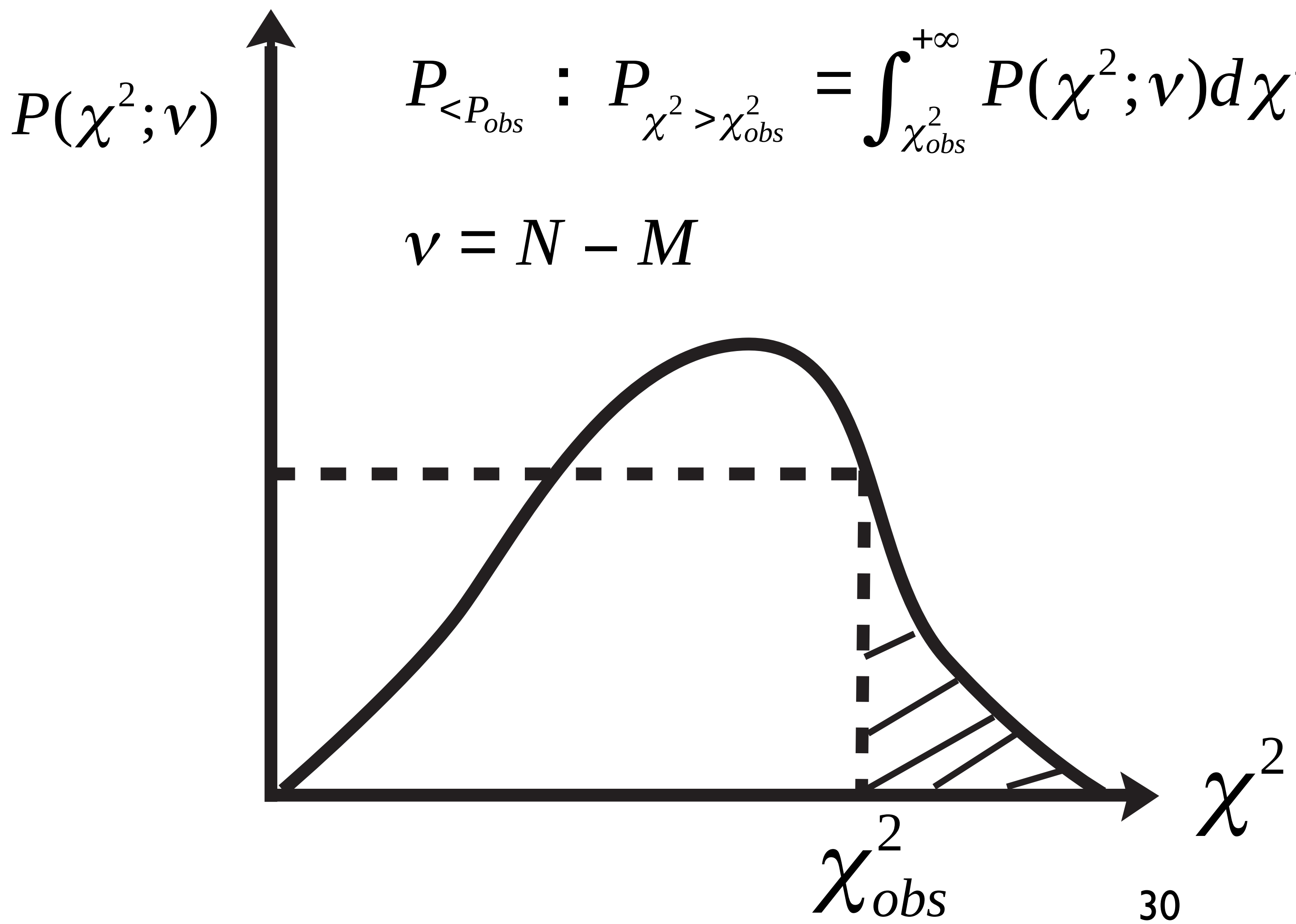
$$\Delta = \begin{vmatrix} \sum_{i=1}^{N}\frac{1}{\sigma_i^2} & \sum_{i=1}^{N}\frac{x_i}{\sigma_i^2} & \sum_{i=1}^{N}\frac{x_i^2}{\sigma_i^2} \\ \sum_{i=1}^{N}\frac{x_i}{\sigma_i^2} & \sum_{i=1}^{N}\frac{x_i^2}{\sigma_i^2} & \sum_{i=1}^{N}\frac{x_i^3}{\sigma_i^2} \\ \sum_{i=1}^{N}\frac{x_i^2}{\sigma_i^2} & \sum_{i=1}^{N}\frac{x_i^3}{\sigma_i^2} & \sum_{i=1}^{N}\frac{x_i^4}{\sigma_i^2} \end{vmatrix}$$

Goodness of fit
کیفیت برازش

یک سوال مهم: برای مقدار مشخص به دست آمده برای chi^2 احتمال اینکه مشاهده ای داشته باشیم که به تطابق بدتر منجر شود چقدر است؟

$$P_{<P_{obs}} : P_{\chi^2>\chi^2_{obs}} = \int_{\chi^2_{obs}}^{+\infty} P(\chi^2;\nu)\,d\chi^2 = \frac{1}{2}Q\left(-\frac{\nu-2}{2},\frac{\chi^2}{2}\right)$$ تابع بتای ناکامل

$$\nu = N - M$$



۱) در تحلیل ها یک مقدار برای احتمال بدتر بودن در نظر گرفته می شود . بعد اگر داشته باشیم: $\chi^2_{obs} \approx \chi^2_{\alpha}$

در آن صورت ادعا می شود که با احتمال $\alpha \times 100$ اندازه گیری هایی خواهیم داشت که با این مدل به chi^2 ای بیشتر منجر می شود. به بیانی دیگر این مدل با احتمال $(1-\alpha)\times 100$ مشاهده و اندازه گیری ها را تو صیف نمی کند.

۲) اگر داشته باشیم $\chi^2_{obs} > \chi^2_{\alpha}$ در آن صورت مدل به خوبی مشاهدات را توصیف کرده است. یعنی تفاوت مدل و مشاهدات معنا دار نیست.

۳) اگر داشته باشیم $\chi^2_{obs} \ll \chi^2_{\alpha}$ در آن صورت یا خطاهای مشاهدات خیلی بزرگتر از مقدار واقعی اش محاسبه شده اند یا که داده سازی شده است

برای رهایی از حالت ۳ بهتر است رهیافت دیگری را مد نظر بگیریم

وقتی بخواهیم که یک مدلی را تا احتمال ۹۵ درصد تایید کنیم یعنی اینکه با احتمال ۹۵ درصد با همین مدل به chi^2 بزرگتر برسیم

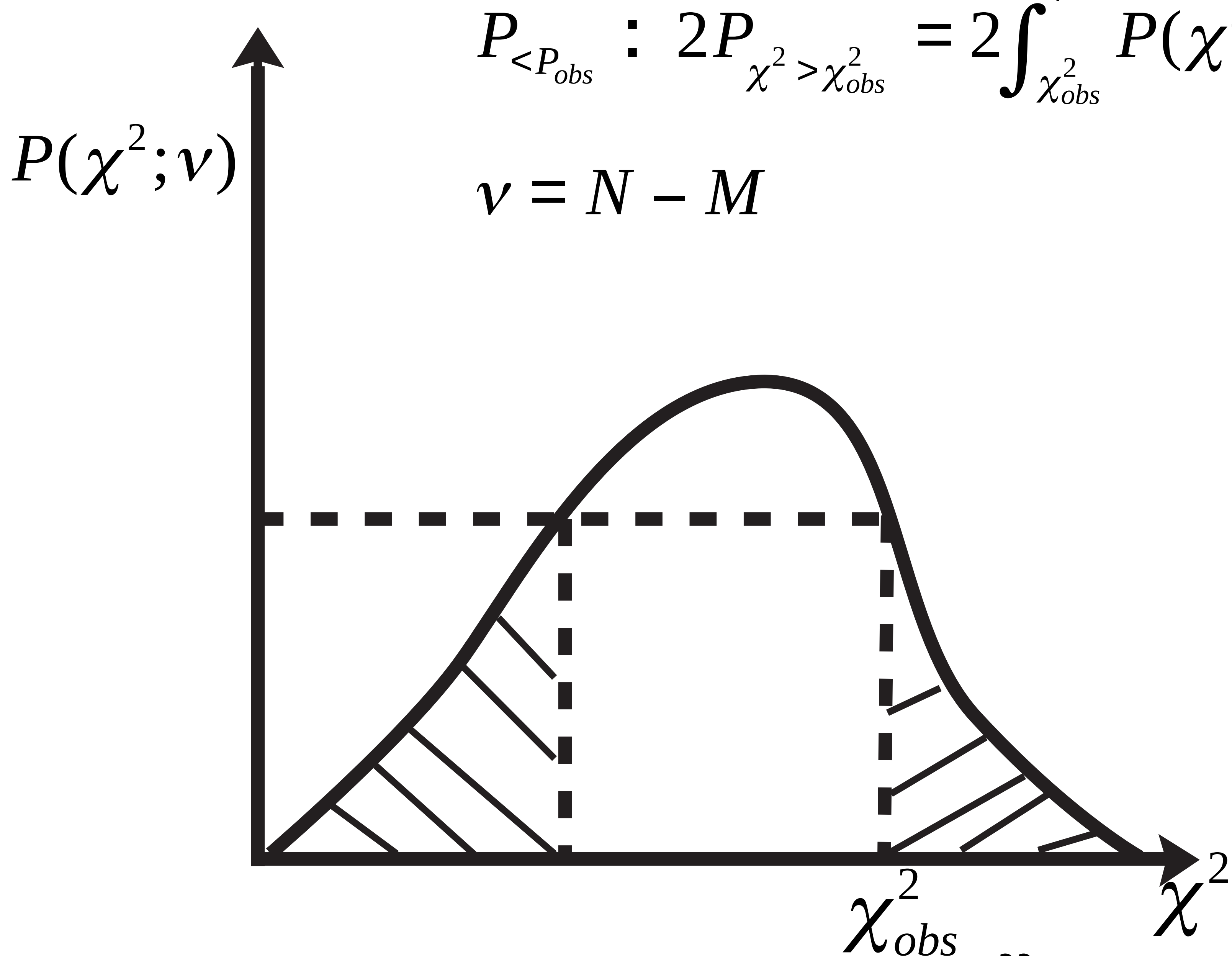
وقتی بخواهیم که یک مدلی را تا احتمال ۹۵ رد کنیم یعنی اینکه با احتمال ۵ درصد باید با همین مدل به chi^2 بزرگتر برسیم

در نظر می گیریم که وضعیت بهینه موقعی است که درصورتی اندازه گیری را دوباره تکرار کنیم به مقدار chi^2 محتمل برسیم بنابراین علاقه مندیم که احتمال برابر با ۵۰ درصد باشد بنابراین نه احتمال کم نه احتمال زیاد (معادل داده سازی یا غلط بودن واریانسها)

$$P_{<P_{obs}} : 2P_{\chi^2>\chi^2_{obs}} = 2\int_{\chi^2_{obs}}^{+\infty} P(\chi^2;\nu)d\chi^2 = Q\left(-\frac{\nu-2}{2},\frac{\chi^2}{2}\right)$$

$$\nu = N - M$$

تابع بتای ناکامل



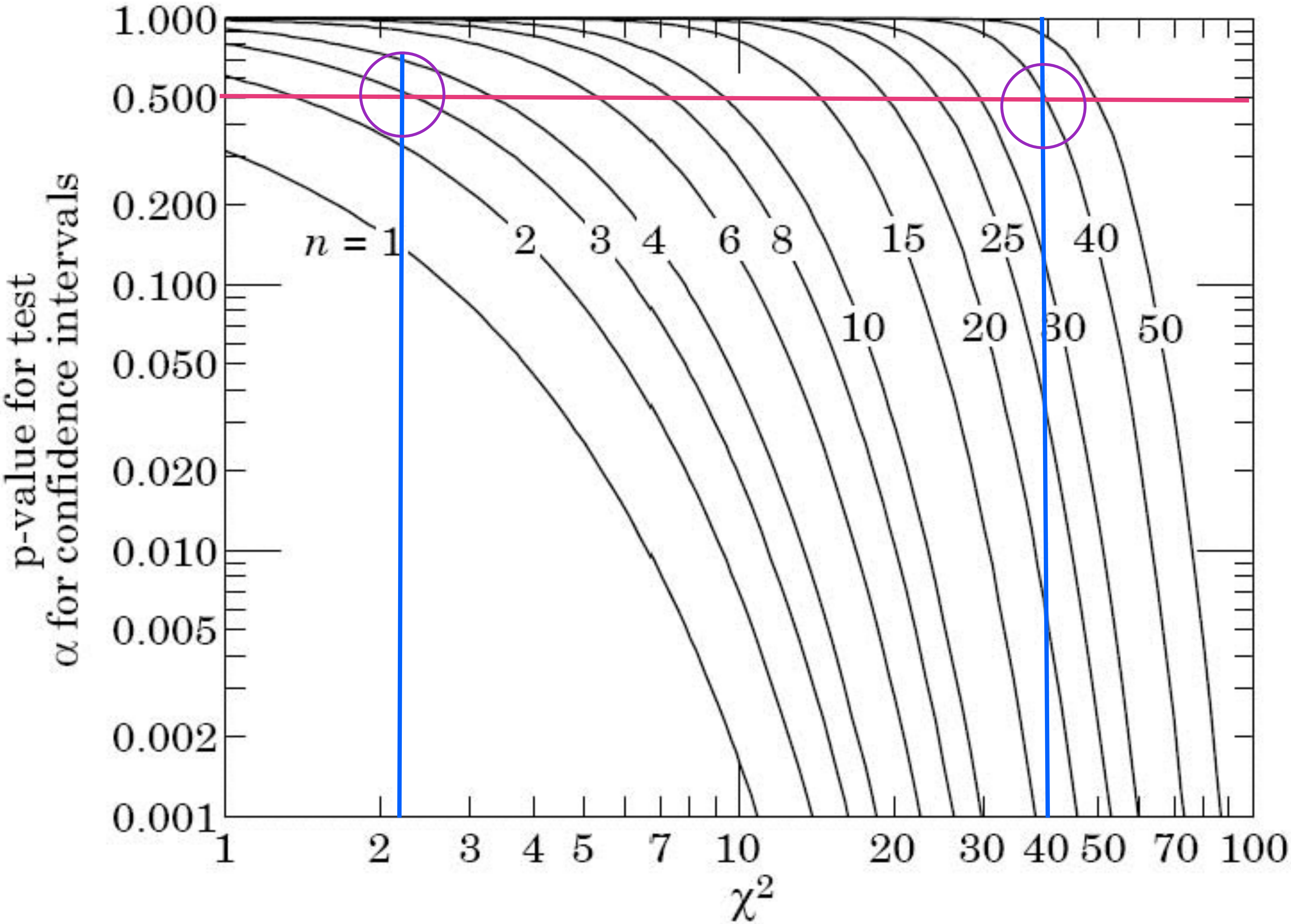
یک قاعده سرانگشتی برای بررسی میزان تطابق مدل با اندازه گیری

$$P(\chi^2;\nu) = \frac{(\chi^2)^{-\frac{(\nu-2)}{2}} \exp(-\chi^2/2)}{2^{\nu/2}\Gamma(\nu/2)}$$

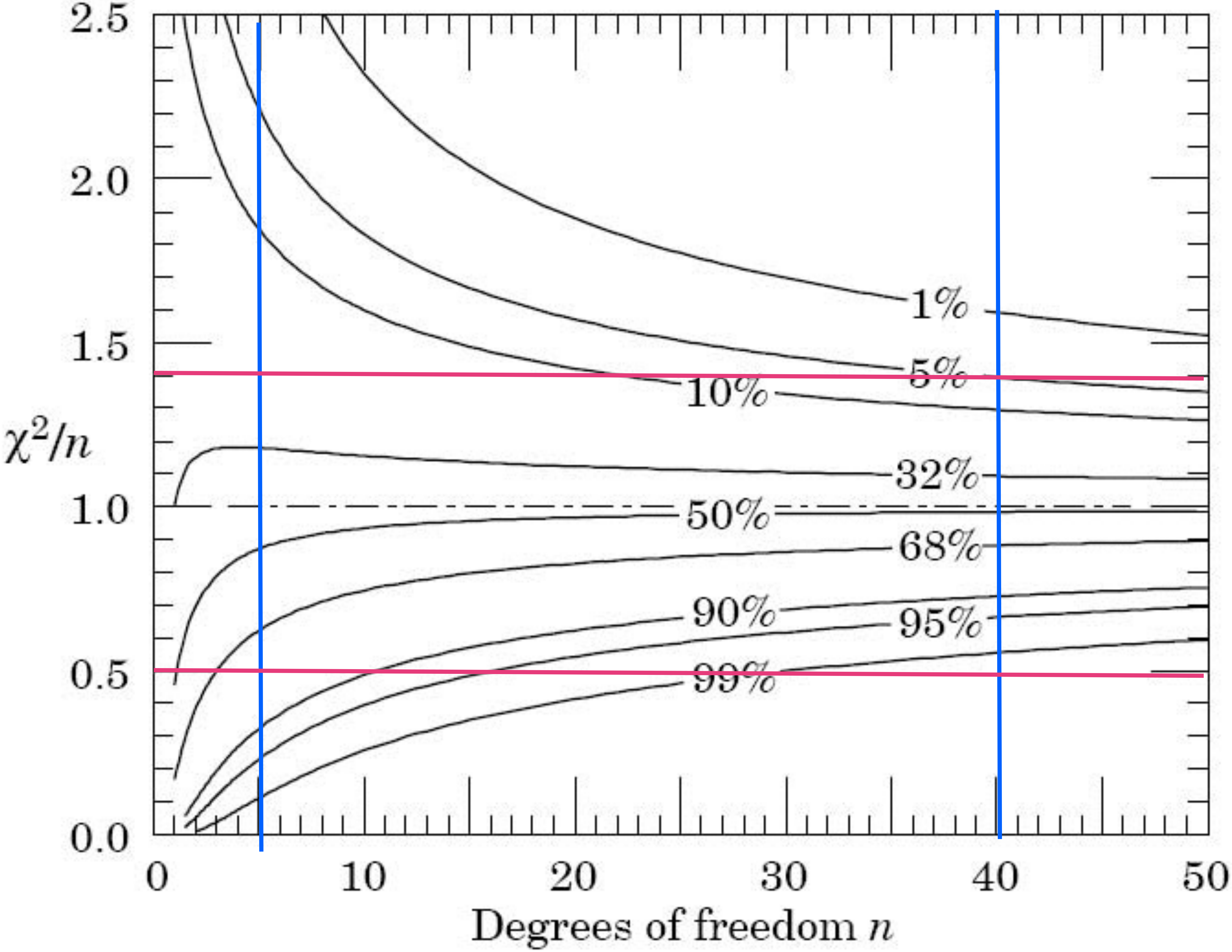
$$for \quad \nu >> 1$$

$$P(\chi^2;\nu) \to P_{Gaussian} : \exp\left(\frac{[\chi^2-\nu]^2}{2(2\nu)}\right)$$

$$\chi^2 = \nu \pm \sqrt{2\nu} \quad \to \quad \chi^2_\nu \equiv \frac{\chi^2}{\nu} = 1 \pm \sqrt{\frac{2}{\nu}}$$



W.-M. Yao et al., Journal of Physics G 33, 1 (2006)



W.-M. Yao et al., Journal of Physics G 33, 1 (2006)

محدویت توضیحات قبلی وقتی خواهد بود که تابع توزیع هر اندازه گیری گوسی نباشد بنابراین با تابع توزیع chi^2 مواجه نیستیم. نه اینکه نمی توانیم از chi^2 استفاده کنیم بلکه بحث در خصوص تابع توزیع است بنابراین باید ابتدا با کمک تبدیل تابع توزیع شکل تابع توزیع مانده را بدست آوریم و بعد مانند قبل به محاسبه بپردازیم

تراز تطابق
(Confidence interval)

۱) دیدیم که اگر مشاهدات مستقل و حول مقدار بهینه کمیت های مدل به صورت گوسی توزیع شده باشند نهایتاً به کمیت chi^2 می رسیم. با کمینه کردن این کمیت تابع Likelihood بیشینه می شود.

۲) بیشینه شدن تابع درست نمایی معادل یافتن بهترین مقادیر برای پارامترهای آزاد مدل خواهد بود.

$$P(D|\Theta) \sim \exp\left(-\frac{\chi^2}{2}\right)$$

۳) حال سوال این است که در منطق بیزی ناحیه ای که با احتمال مشخص مثلاً ۶۸ درصد می توان بهترین مقادیر را برای پارامترهای آزاد به دست آورد؟

۴) اکنون کمیت جدیدی تعریف می کنیم به نام Relative Likelihood:

$$\Gamma(D\,|\,\Theta) \equiv \frac{P(D\,|\,\Theta)}{P_{Max}(D\,|\,\Theta_*)} = \exp\left(-\frac{\left[\chi^2(\Theta)-\chi^2_{\min}\right]}{2}\right)$$

۵) از سویی دیگر از آنجا که به ازای مقدار بهینه کمیت های آزاد در مدل تابع Likelihood بیشینه می شود میتوان این تابع را بسط داد:

$$\log(P(D\,|\,\Theta)) = \log(P(D\,|\,\Theta=\Theta_{Best})) + \sum_{k=1}^{M}(\theta_k-\theta_k^{Best})\frac{\partial \log(P(D\,|\,\Theta))}{\partial\theta_k}$$
$$+\frac{1}{2}\sum_{k,l}^{M}\frac{\partial^2 \log(P(D\,|\,\Theta))}{\partial\theta_k\partial\theta_l}(\theta_k-\theta_k^{Best})(\theta_l-\theta_l^{Best}) + \mathrm{O}(\delta\Theta^3) + \ldots$$

۶) جمله دوم سمت راست معادله بالا صفر است با صرف نظر کردن از جملات مرتبه بالاتر داریم:

$$\Gamma(D\,|\,\Theta) = \exp\left(-\frac{\left[\chi^2(\Theta) - \chi^2_{\min}\right]}{2}\right) = \sqrt{\frac{\det[F]}{(2\pi)^M}}\exp\left(-\frac{\delta\Theta^T \bullet [F] \bullet \delta\Theta}{2}\right)$$

$$\Theta : \{\theta_k\} \qquad k = 1, \ldots, M$$

$$\delta\Theta \equiv \Theta - \Theta_{Best}$$

$$\chi^2_{\min} = \chi^2(\Theta = \Theta_{Best})$$

$$\Gamma(D\,|\,\Theta) : \ \exp\left(-\frac{\delta\Theta^T \bullet [F] \bullet \delta\Theta}{2}\right)$$

تعیین تراز تطابق با کمک تغییرات chi^2 به صورت عددی

دیدیم که می توان نوشت:

$$\Gamma(D\,|\,\Theta)=\exp\left(-\frac{\left[\chi^2(\Theta)-\chi^2_{\min}\right]}{2}\right)=\sqrt{\frac{\det[F]}{(2\pi)^M}}\exp\left(-\frac{\delta\Theta^T\bullet\left[F\right]\bullet\delta\Theta}{2}\right)$$

$$\Gamma(D\,|\,\theta_1)=\int d\theta_2\ldots d\theta_M\,\Gamma(D\,|\,\Theta) \sim \exp\left(-\frac{(\theta-\theta_{Best})^2}{2\sigma_\theta^2}\right)$$

$$\Delta\chi^2=\chi^2(\Theta)-\chi^2_{\min}=\frac{(\theta-\theta_{Best})^2}{\sigma_\theta^2}$$

$$n(\theta-\theta_{Best})\le n\sigma_\theta \qquad\rightarrow\qquad \Delta\chi^2\le n^2$$

$$\Gamma(D\,|\,\theta_1)=\exp\left(-\frac{n^2}{2}\right)\rightarrow\begin{cases}n=1\rightarrow\Gamma(D\,|\,\theta_1)=0.606\\ n=2\rightarrow\Gamma(D\,|\,\theta_1)=0.135\end{cases}$$

بنابراین اگر chi^2 مربوط به Marginalized Posterior از مقدار کمینه اش به اندازه n تغییر کند ناحیه احتمال $n\sigma$ برای کمیت مورد نظر به دست می آید

تراز تطابق در فضای M بُعدی

برای اینکار سوال میکنیم که تابع توزیع مربوط به chi^2 چگونه است؟ باز هم از تبدیل بین تابع توزیع ها استفاده می کنیم

$$\Gamma(D\,|\,\Theta) = \exp\left(-\frac{\left[\chi^2(\Theta)-\chi^2_{\min}\right]}{2}\right) = \sqrt{\frac{\det[F]}{(2\pi)^M}}\exp\left(-\frac{\delta\Theta^T\bullet\left[F\right]\bullet\delta\Theta}{2}\right)$$

$$:\ \exp\left(-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{\left[y_i-\sum_{k=1}^{M}f(\{\theta_i\},x_i)\right]^2}{\sigma_i^2}\right)\right)$$

$$D:\{x_i,y_i\}\quad\rightarrow i=1,\ldots,N$$

$$\Theta:\{\theta_M\}\qquad k=1,\ldots,M$$

$$if\quad i=1\quad\rightarrow\quad\Delta y=\left[y-\sum_{k=1}^{M}f(\{\theta_i\},x)\right]\quad\rightarrow\quad\chi^2=\Delta y^2$$

$$P(\Delta y)d\Delta y=P(\chi^2)d\chi^2\quad\rightarrow\quad P(\chi^2)=\frac{(\chi^2)^{-1/2}\exp(-\chi^2/2)}{\sqrt{2\pi}}$$

$$68.3\%=\int_{\sigma^-}^{\sigma^+}P(\theta)d\theta=\int_{\chi^2_{\min}}^{\chi^2}P(\chi^2)d\chi^2\rightarrow\Delta\chi^2=1$$

مانند قبل است

$$\Gamma(D\,|\,\Theta)=\exp\left(-\frac{\left[\chi^2(\Theta)-\chi^2_{\min}\right]}{2}\right)=\sqrt{\frac{\det[F]}{(2\pi)^M}}\exp\left(-\frac{\delta\Theta^T\bullet[F]\bullet\delta\Theta}{2}\right)$$

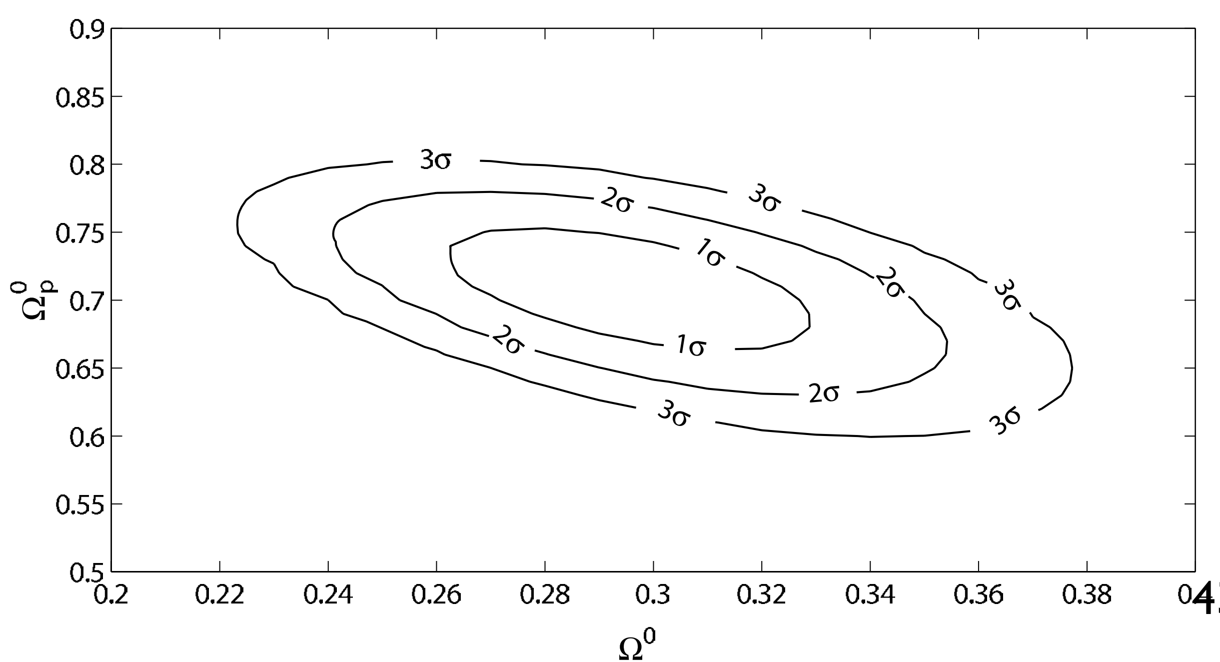
$$:\ \exp\left(-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{\left[y_i-\sum_{k=1}^{M}f(\{\theta_i\},x_i)\right]^2}{\sigma_i^2}\right)\right)$$

$$D:\{x_i,y_i\}\quad\rightarrow i=1,\ldots,N$$

$$\Theta:\{\theta_M\}\qquad k=1,\ldots,M$$

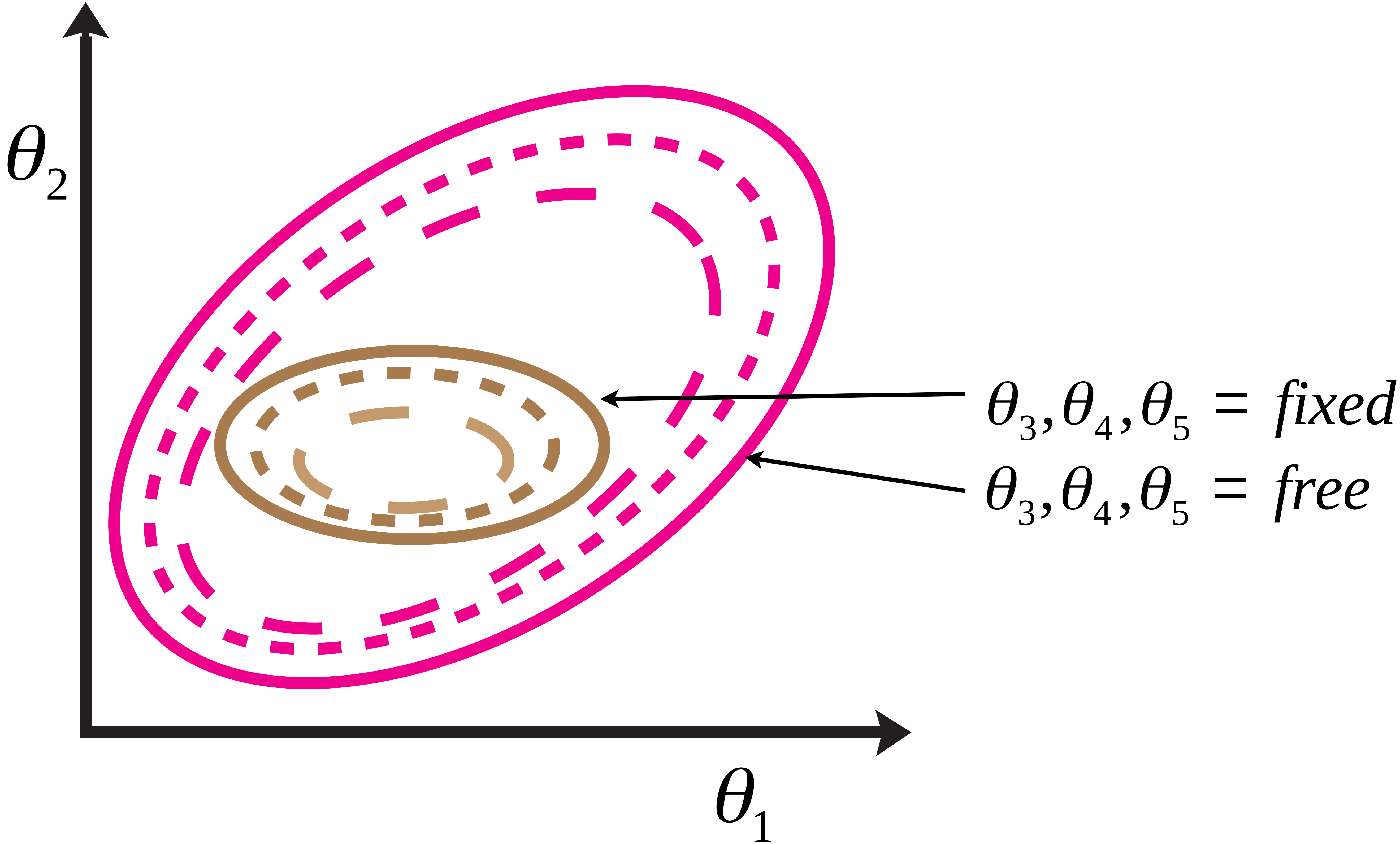
$$P(\Delta y)d\Delta y=P(\chi^2)d\chi^2\quad\rightarrow\quad P(\chi^2;\nu)=\frac{(\chi^2)^{-\frac{(\nu-2)}{2}}\exp(-\chi^2/2)}{2^{\nu/2}\Gamma(\nu/2)}$$

$$68.3\%=\int_{\sigma_1^-}^{\sigma_1^+}\int_{\sigma_2^-}^{\sigma_2^+}P(\theta_1,\theta_2)d\theta_1d\theta_2=\int_{\chi^2_{\min}}^{\chi^2}P(\chi^2;2)d\chi^2\rightarrow\Delta\chi^2=2.3$$



به لحاظ محاسباتی باید پس از یافتن مقدار بهینه کمیت ها اجازه دهیم که کمیت ها از مقدار بهینه اشان تغییر کنند به طوری که chi^2 از مقدار کمینه اش به اندازه مشخص شده جابجا شود در این وضعیت بیضی گون تراز تطابق در ۲ بُعد را به دست می آوریم

M	۱	۲	۳	۴	۵	۶
$p\%$	$\Delta\chi^{۲}$	$\Delta\chi^{۲}$	$\Delta\chi^{۲}$	$\Delta\chi^{۲}$	$\Delta\chi^{۲}$	$\Delta\chi^{۲}$
۶۸.۳%	۱.۰۰	۲.۳۰	۳.۵۳	۴.۷۲	۵.۸۹	۷.۰۴
۹۰%	۲.۷۱	۴.۶۱	۶.۲۵	۷.۷۸	۹.۲۴	۱۰.۶
۹۵.۴%	۴.۰۰	۶.۱۷	۸.۰۲	۹.۷۰	۱۱.۳۰	۱۲.۸۰
۹۹%	۶.۶۳	۹.۲۱	۱۱.۳۰	۱۳.۳۰	۱۵.۱۰	۱۶.۸۰
۹۹.۷۳%	۹.۰۰	۱۱.۸۰	۱۴.۲۰	۱۶.۳۰	۱۸.۲۰	۲۰.۱۰
۹۹.۹۹%	۱۵.۱۰	۱۸.۴۰	۲۱.۱۰	۲۳.۵۰	۲۵.۷۰	۲۷.۸۰



θ_2
$\theta_3, \theta_4, \theta_5$ = *fixed*
$\theta_3, \theta_4, \theta_5$ = *free*
θ_1

Chi^2 versus Median analysis

The χ^2 analysis assumes that:

(i) The experimental results are statistically independent;

(ii) There are no systematic errors;

(iii) The statistical errors follow a Gaussian distribution;

(iv) The standard deviation of the statistical errors is equal to the observational uncertainty.

The median analysis considers only following assumptions

(i) The experimental results are statistically independent;

(ii) There are no systematic errors;

the median statistic is not very sensitive to the presence of a few "outlier" SNIa objects.

In [9], it has been shown how the presence of one or very few "ill" data points could severely distort a χ^2 analysis, while the median results remained approximately the same.

In median statistics we search for those theoretical values for which, the half part of observed data points are located above and the rest is located below or equal theoretical prediction.

A. Barreira, P. P. Avelino, arXiv:1107.3971

J. R. Gott, III, M. S. Vogeley, S. Podariu, and B. Ratra, Astrophys. J. 549, 1 (2001), astro-ph/0006103

یافتن مقادیر کمیت ها به صورت عددی

1) Grid search Methods

- Deterministic approach
- Markov chain Monte Carlo MCMC
- Hamiltonian Monte Carlo HMC

2) Gradient Method

3) Expansion Method

.

.

.

در رهیافت تعینی با حرکت در فضای پارامتری و مقایسه تا جایی جلو می رویم تا تابع Posterior بیشینه شود.

برای فضای پارامتری در بُعد های کم توصیه می شود

MCMC

هدف یافتن دسته مقادیری که posterior را بیشینه می کنند

$1)\ \Theta_0 : \{\theta_0\}$

$2)\ g(\Theta_0 \to \Theta)$

$$3)\ \eta = \min\left[1, \frac{P(\Theta \mid D)g(\Theta_0 \to \Theta)}{P(\Theta_0 \mid D)g(\Theta \to \Theta_0)}\right]$$

$4)\ u : Uniform[0,1]$

$5)\ If \quad u \leq \eta \to \Theta_1 = \Theta$

$else$

$$\Theta_1 = \Theta_0$$

$6)\ Iterate$

معمولاً گوسی انتخاب می شود. هر چه که واریانس آن بزرگ باشد آهنگ پذیرش کمتر می شود

۱) ابتدا از یک سری مقادیر دلخواه برای کمیت های مدل شروع می کنیم

۲) با توجه به ماتریس پیشنهادی جهشی به مقادیر جدید می کنیم

۳) برای تصمیم به اینکه این جهش مورد قبول باشد یا نه آزمون مربوطه بررسی میشود

۴) این فرایند تکرار می شود تا جواب همگرا شود

Import Data $\{D\}$

Select $\{\theta\}_{old}$

Compute $\chi^2_{old}(\{\theta\}_{old})$, $\xi_{old} = e^{-\chi^2_{old}/2}$

loop on MCMC

Select $\{\theta\}_{New}$ by using $\{\theta\}_{old}$

Compute $\chi^2_{New}(\{\theta\}_{New})$

$\Delta\chi^2 = \chi^2_{New}(\{\theta\}_{New}) - \chi^2_{old}(\{\theta\}_{old})$

Check acceptance Rate $R \equiv \min\{1, e^{-\Delta\chi^2/2}\}$

$\xi =$ Call Random number

if $\xi \leq \exp(-\Delta\chi^2/2)$ then

$\{\theta\}_{old} = \{\theta\}_{New}$

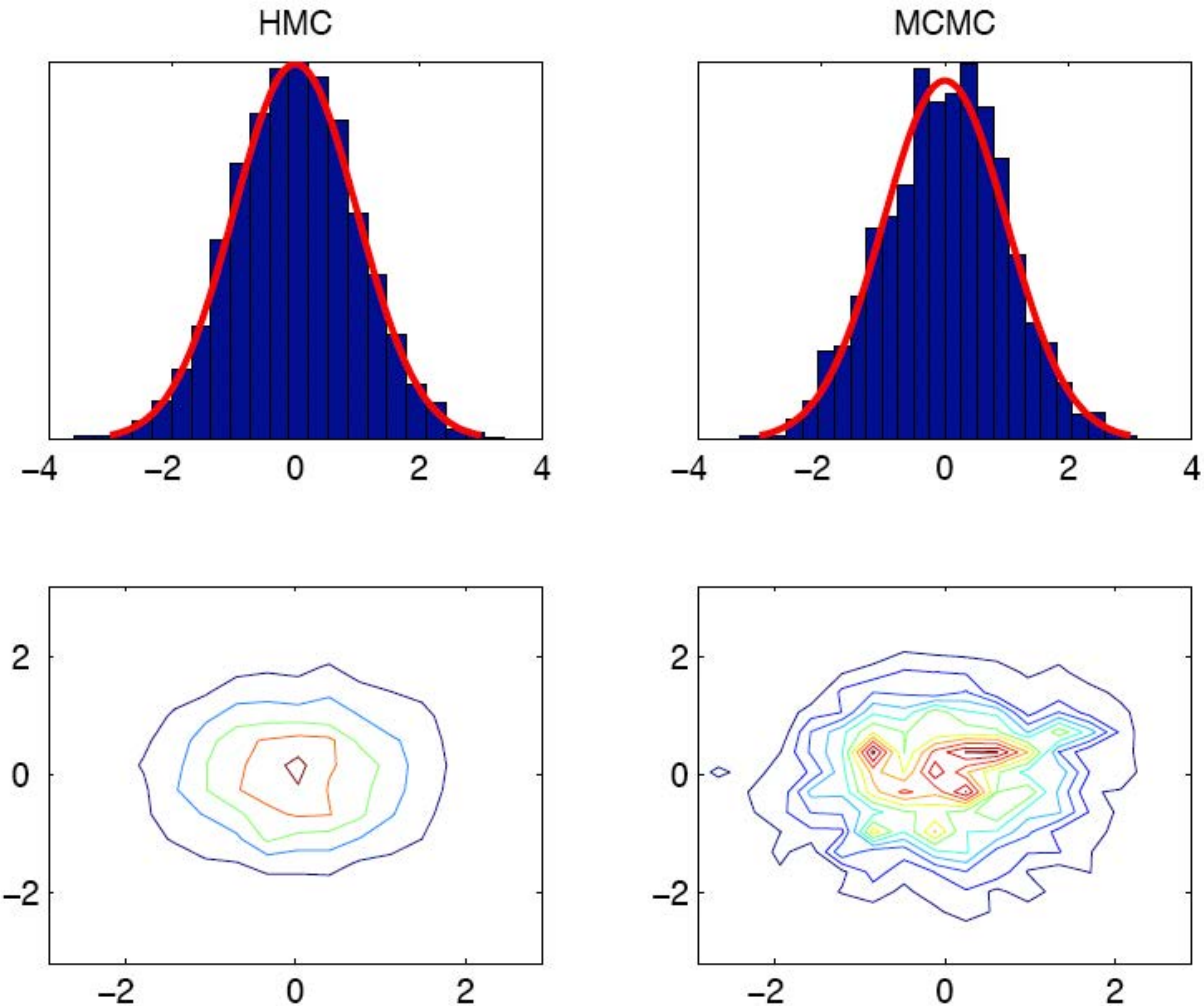
$\chi^2_{old} = \chi^2_{New}$

End if

Write $\{\theta\}_{old}$, χ^2_{old}

End loop

در واقع در اینجا داریم داده هایی تولید می کنیم که دارای تابع توزیع Chi^2 هستند



A. Hajian, PRD (75),083525 (2007)

Hamiltonian MC (HMC)

در این روش با توجه به تعریف یک هامیلتونی مسیر حرکت بر روی انرژی ثابت انجام می گیرد. این امر باعث می شود آهنگ پذیرش بالا رفته و مدت زمان محاسبات به طور قابل توجهی کاهش یابد.

برای این منظور یک انرژی پتانسیل و یک انرژی جنبشی تعریف می کنیم.

A. Hajian, PRD (75),083525 (2007)

Hamiltonian MC (HMC)

$1)\left[\Theta_0 : \{\theta_0\} \quad ; \quad u_0 : N(u,0,1)\right],$

$\mathrm{H}(\Theta,u) = U(\Theta) + K(u) \qquad U(\Theta) = -\ln P(\Theta \mid D), \quad K(u) = \frac{u^T u}{2}$

$\Gamma(\Theta \mid D; u) : \ \exp\left(-\mathrm{H}(\Theta,u)\right) = P(\Theta \mid D) \times N(u,0,1)$

2) *Make a leapfrog move* $\{\Theta_{(j-1)}; u_{(j-1)}\} \to \{\Theta_{(j)}; u_{(j)}\}$

3) $\{\Theta^*; u^*\} = \{\Theta(s); u(s)\}$

4) $\eta = \min\left[1, e^{\left(\mathrm{H}(\Theta,u) - \mathrm{H}(\Theta^*,u^*)\right)}\right]$

5) $\alpha : \ Uniform[0,1]$

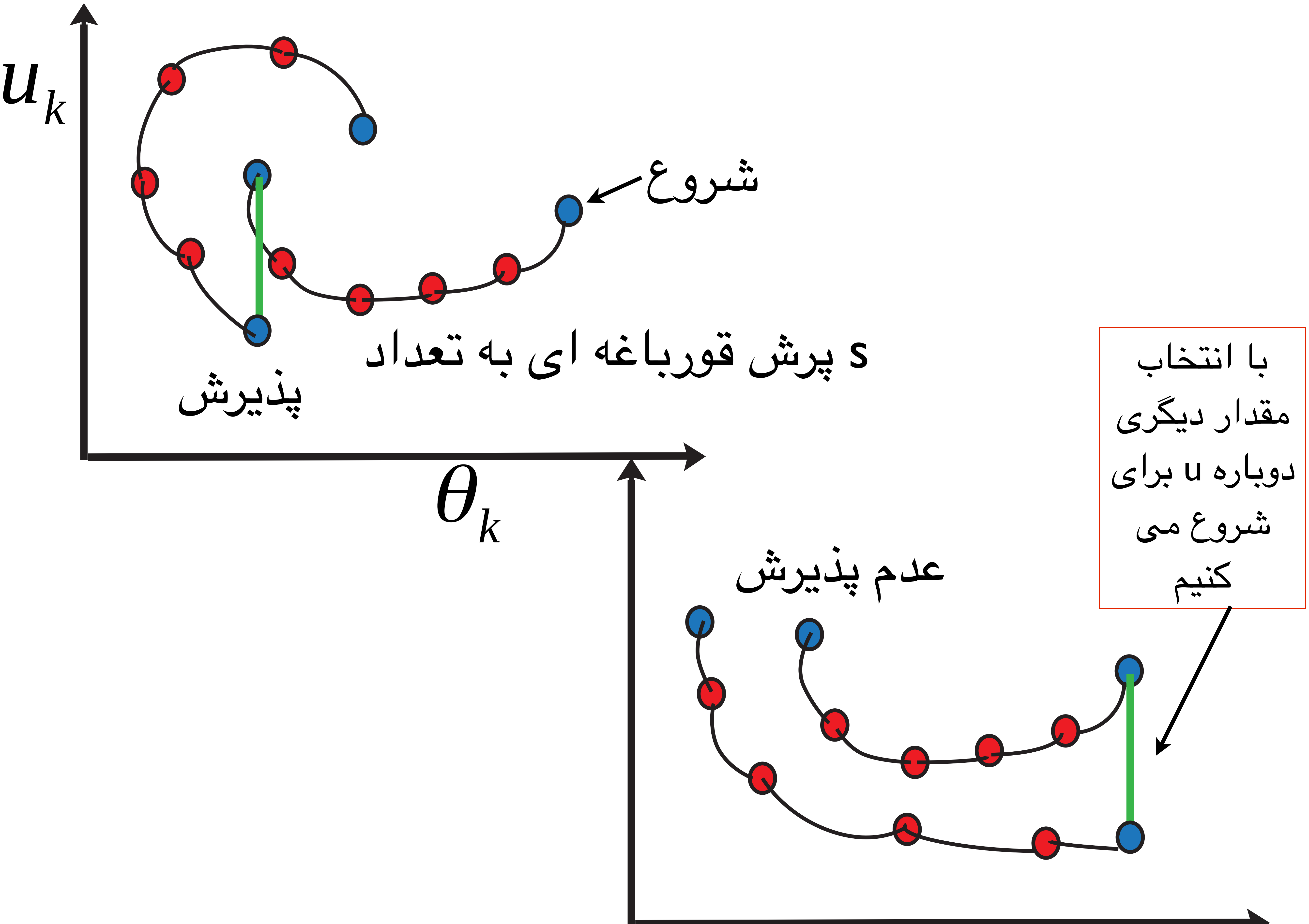
6) *If* $\alpha \le \eta \to \begin{cases} \Theta_1 = \Theta^* \\ u = u^* \end{cases}$

else

$\Theta_1 = \Theta_0$

7) *Iterate*

A. Hajian, PRD (75),083525 (2007)



u_k
شروع
پرش قورباغه ای به تعداد s
پذیرش
θ_k
با انتخاب مقدار دیگری دوباره u برای شروع می کنیم
عدم پذیرش

Import Data $\{D\}$

Select $\{\theta\}_{old}$, $\{u\}_{old} : N(u,0,1)$

$\chi^2_{old}(\{\theta\}_{old})$

$$\mathcal{H}_{old} = \frac{\chi^2_{old}}{2} + \frac{u^T u}{2}$$

$$\mathcal{S}_{old} = \exp(-\mathcal{H}_{old}) = \mathcal{L}(\{D\} \mid \{\theta\}_{old})\, N(u,0,1)$$

loop on MC

loop on HMC

Leapfrog or Verlet method
$$\begin{cases} \{\theta\}_{old} \longrightarrow \{\theta\}^{HMC}_{old} \\ \{u\}_{old} \longrightarrow \{u\}^{HMC}_{old} \end{cases}$$

End loop

$\{\theta\}_{New} = \{\theta\}^{HMC}_{old}$, $\{u\}_{New} = \{u\}^{HMC}_{old}$

$$\mathcal{H}_{New} = \frac{\chi^2_{New}}{2} + \frac{u^T u}{2}$$

$$\Delta\mathcal{H} = \mathcal{H}_{New} - \mathcal{H}_{old}$$

$$\mathcal{S} = \exp(-\Delta\mathcal{H})$$

Acceptance Rate $R = \min\{1, \mathcal{S}\}$

$\{\theta\}_{old} = \{\theta_{New}\}$, $\chi^2_{old} = \chi^2_{New}$

$\{u\}_{old} = N(u,0,1)$

Write $\{\theta\}_{old}$, $\exp(-\chi^2_{New}/2)$

End loop MC

HMC loop

loop on HMC j

Verlet Algorithm

$$\{\theta\}_{old}^{j+1} = \{\theta\}_{old}^{j} + \{u\}_{old}^{j}\epsilon + \frac{1}{2}\epsilon^2\left(-\frac{\partial \mathcal{H}}{\partial\{\theta\}_{old}^{j}}\right)$$

$$\{u\}_{old}^{j+1} = \{u_{old}\}^{j} + \frac{1}{2}\left[-\frac{\partial \mathcal{H}}{\partial\{\theta\}_{old}^{j+1}} - \frac{\partial \mathcal{H}}{\partial\{\theta\}_{old}^{j}}\right]\epsilon$$

End Loop

یا الگوریتم پرش قورباغه ای به صورت زیر

loop on HMC

$$\{u\}_{old}^{j+\frac{1}{2}} = \{u\}_{old}^{j} - \frac{\epsilon}{2}\frac{\partial \mathcal{H}}{\partial\{\theta\}_{old}^{j}}$$

$$\{\theta\}_{old}^{j+1} = \{\theta\}_{old}^{j} + \epsilon\{u\}_{old}^{j+1/2}$$

$$\{u\}_{old}^{j+1} = \{u\}_{old}^{j} - \epsilon/2\,\frac{\partial \mathcal{H}}{\partial\{\theta\}_{old}^{j+1}}$$

End Loop

روش حرکت مبتنی بر گرادیان chi^2

$$\nabla\chi^2 = \sum_{k=1}^{M} \frac{\partial\chi^2}{\partial\theta_k}\hat{\theta}_k$$

$$\left(\nabla\chi^2\right)_k = \frac{\partial\chi^2}{\partial\theta_k} : \frac{\chi^2(\theta_k + \Delta\theta_k) - \chi^2(\theta_k)}{\Delta\theta_k}$$

$$b_k \equiv \frac{\theta_k}{\Delta\theta_k}$$

بدون بُعد کردن

$$\frac{\partial\chi^2}{\partial b_k} = \frac{\partial\chi^2}{\partial\theta_k}\Delta\theta_k$$

$$\gamma_i = \frac{\dfrac{\partial\chi^2}{\partial b_k}}{\sqrt{\sum_{k=1}^{M}\left(\dfrac{\partial\chi^2}{\partial b_k}\right)}}$$

مولفه بردار یکه گرادیان

$$\theta_k(i+1) = \theta_k(i) - \gamma_i\Delta\theta_k$$

علامت منفی تضمین می کند که در جهتی حرکت کنیم که
chi^2 به سمت کمینه پیش رود

جمع بندی

۱) رهیافت بیزی و تکراری
۲) تحلیل Likelihood
۳) کمینه کردن chi^2
۴) کیفیت برازش
۵) تراز تطابق و ماتریس Fisher
۶) یافتن مقادیر بهینه کمیت ها به صورت عددی
۷) رهیافت MCMC
۸) رهیافت HMC
۹) رهیافت گرادیان

Institute for Research in Fundamental Sciences

COMPLEXLAB

Computational Cosmology Group Of Shahid Beheshti University

دانشکده فیزیک

از توجه شما سپاسگزارم