粒子物理与核物理实验中的数据分析

杨振伟
清华大学

第五讲：统计检验

本讲要点

- 假设，检验统计量，显著水平，功效
- 两种假设下的统计检验
- 纽曼-皮尔森引理
- 如何构造一个检验统计量
- Fisher甄别函数与神经网络
- 检验拟合优度，*P*-值定义与应用
- 信号观测的显著程度
- 皮尔逊的 χ^2 检验

概率与统计

统计的含义可以通过比较概率理论来理解

概率	统计(参量测定与假设检验)
从理论到数据	从数据到理论
通过计算某些可观测量(例如，平均值，分布等)来给出预期的实验分布。 例如：若宇称守衡，对一特定衰变分布有什么影响？	进行所谓的假设检验，比较理论预期的参量值或分布。从观察的实验数据中给出所研究参数的观测值和误差，并且在某一置信水平上检验理论的正确与否。 例如：观测到一特定衰变分布，是否可以断定宇称守衡？

统计分析的目标

假设检验

检验数据是否与某一特定理论相符(注意，该理论可包含一些自由参数）。

相符的程度由显著水平来表示。

参数拟合

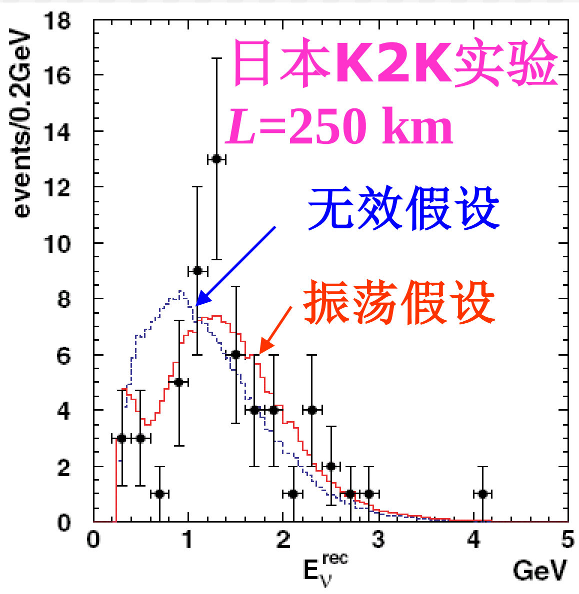
利用数据确定自由参数的大小。

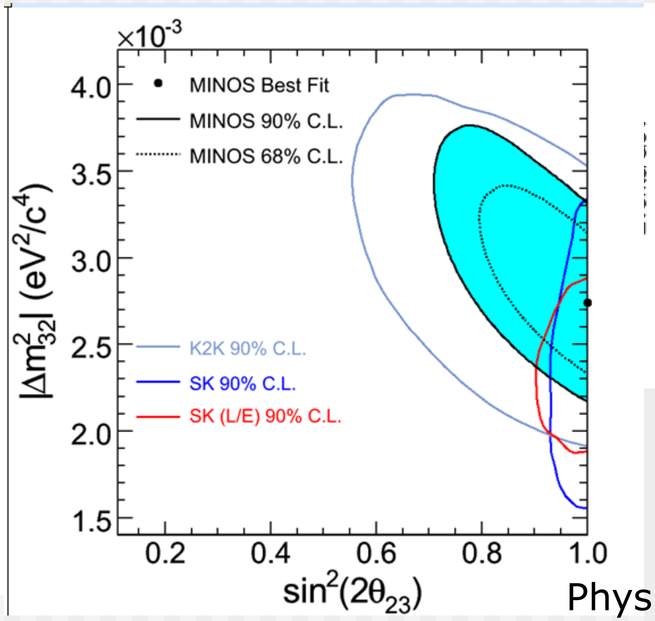
参数的准确程度由对应的不确定度大小来表示。

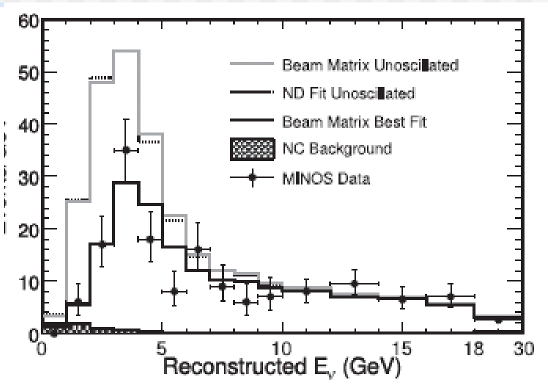
中微子振荡假设检验

利用加速器把中微子射往远处的探测器，观察有多少中微子发生了形态上的改变，即所谓的加速器中微子振荡实验

$$P(\nu_\mu \to \nu_\mu) = 1 - \sin^2(2\theta)\sin^2(1.27\Delta m^2 \frac{L}{E_{\nu_\mu}})$$







美国**MINOS**实验
L=700 km

Phys.Rev.D74,072003(2006)

Phys.Rev.Lett.97,191801(2006)

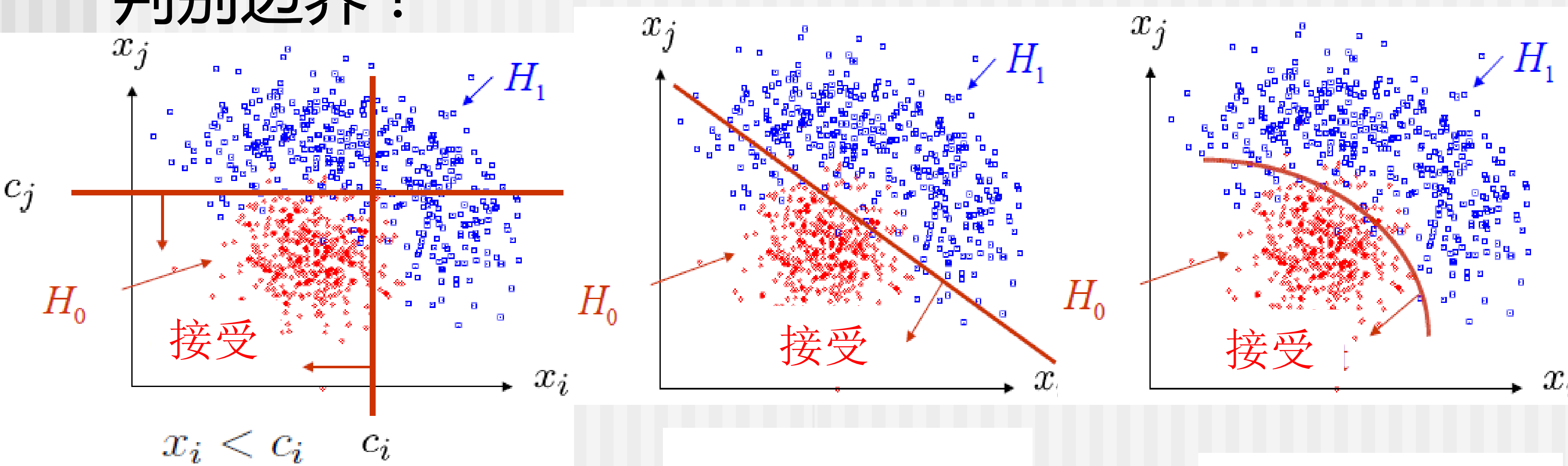
振荡假设符合概率：37%
无效假设符合概率：0.07%

美国实验证实了日本实验而且实验精度更高。

事例选择

假设实验数据中含有两种不同的事例，对应于假设H_0和假设H_1，我们希望选择出 H_0 类的事例。

每个事例对应于 $\vec{x}$ 空间中的一个点。应当如何选择判别边界？



$x_i < c_i$

$x_j < c_j$

$x_i + x_j < c$

$f(x_i, x_j) < c?$

1）如何给出“最优化”的选择?

2）多维空间有何困难?

假设检验

假如测量结果为 $\vec{x}=(x_1,x_2,\ldots,x_n)$, 例如: 正负电子对撞后所产生的事例中, 对于每个事例, 有下列测量量

x_1 = 产生的带电粒子数; x_2 = 粒子的平均横动量; x_3 = 产生的"喷注"数目; …

这里 $\vec{x}$ 服从在 $n-$维空间的某些与产生事例类型有关的联合概率密度函数, 例如: 正负电子对撞, 原子核与原子核碰撞, 等等。那么这些联合的概率密度函数 $f(\vec{x})$ 取决于采取何种假设。

$f(\vec{x}\,|\,H_0), f(\vec{x}\,|\,H_1)$, 等等

简单假设: $f(\vec{x})$ 无未定参数

复杂假设: $f(\vec{x};\alpha)$ 含未定参数 α

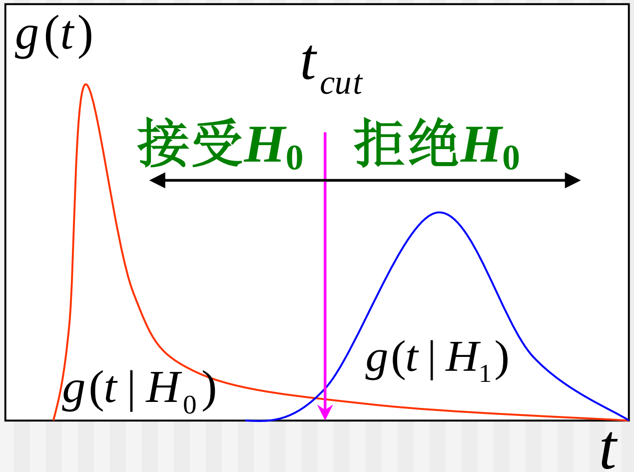
通常情况下很难处理多维的 $\vec{x}$ 问题, 因此, 常常构造低维的统计检验, 在不失去甄别各种假设能力的条件下, 使得 $t(\vec{x})$成为精简后的数据样本。

那么此时的统计量 t 具有概率密度函数 $g(t\,|\,H_0), g(t\,|\,H_1),\ldots$

拒绝域、第一与第二类错误

考虑统计检验量t 服从 $g(t|H_0), g(t|H_1),...$

定义拒绝域, 使得H_0假设为真时, t 不大可能发生



例如，在上述情况下, $t \geq t_{cut}$

如果观测量 t_{obs} 在拒绝域时,拒绝H_0, 否则接受H_0。

假若H_0为真，但被拒绝，这个可能性构成第一类错误

$$\alpha = \int_{t_{cut}}^{\infty} g(t|H_0)dt$$ (显著水平)

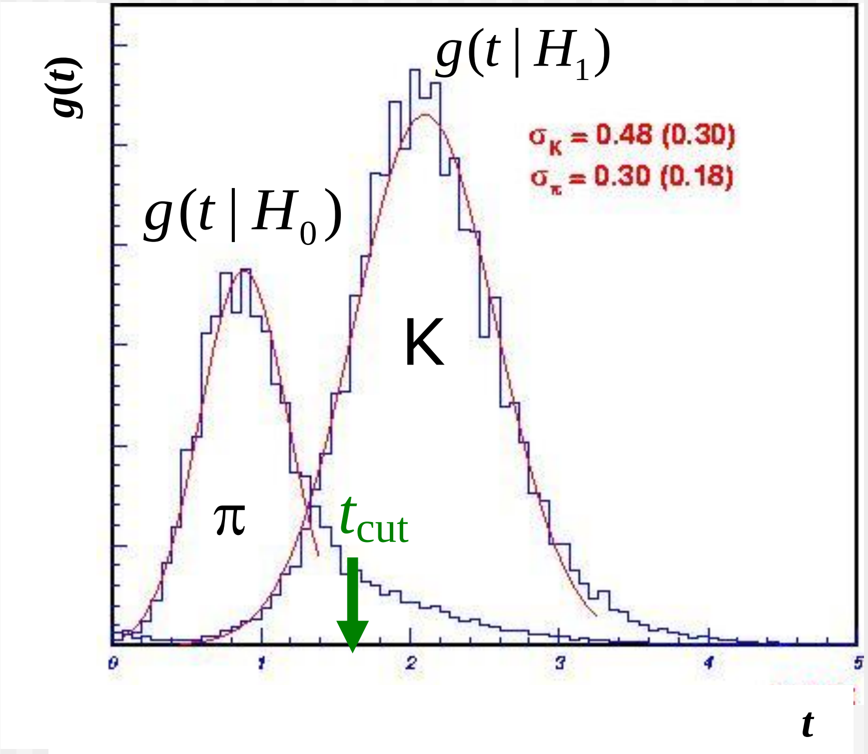
假若接受H_0真，但实际却是H_1为真，这个可能性构成第二类错误

$$\beta = \int_{-\infty}^{t_{cut}} g(t|H_1)dt$$ (1-β =功效)

例子:选择不同粒子

一束包含K/π粒子的束流穿过2厘米厚的闪烁体，根据电离能损的大小可以用来进行粒子鉴别。构造能量沉积测量量 t，并假设只有两种可能

$H_0 = \pi$（信号）
$H_1 = K$（本底）



通过要求 $t<t_{cut}$ 来选择 π 粒子，选择效率为

$$\epsilon_\pi = \int_{-\infty}^{t_{\rm cut}} g(t|\pi)\mathrm{d}t = 1-\alpha$$

$$\epsilon_K = \int_{-\infty}^{t_{\rm cut}} g(t|K)\mathrm{d}t = \beta$$

松选择：效率很高，但 K 本底高；
严选择：信号样本纯，但效率低。

π 的份额 a_π 可从 t 分布估计 $f(t;a_\pi) = a_\pi g(t|\pi) + (1-a_\pi) g(t|K)$

粒子鉴别的概率问题

对于一个具有测量值 t 的粒子，如何估计是 K 还是 π 的概率？

贝叶斯定理

$$h(K|t) = \frac{a_K g(t|K)}{a_K g(t|K) + a_\pi g(t|\pi)}$$

$$h(\pi|t) = \frac{a_\pi g(t|\pi)}{a_K g(t|K) + a_\pi g(t|\pi)}$$

对于贝叶斯论者：上式为粒子是 K 或 π 的可信程度

对于频率论者：给定 t 条件下，粒子是 K 或 π 的比率

两种解释均有道理

通常情况下，需要给出选择样本的纯度

$$p_\pi = \frac{N_\pi(t<t_{cut})}{N_{all}(t<t_{cut})} = \frac{\int_{-\infty}^{t_{cut}} a_\pi g(t|\pi)dt}{\int_{-\infty}^{t_{cut}} [a_\pi g(t|\pi) + (1-a_\pi)g(t|K)]dt} = \frac{\int_{-\infty}^{t_{cut}} h(\pi|t) f(t)dt}{\int_{-\infty}^{t_{cut}} f(t)dt}$$

$= \pi$ 粒子在区间 $(-\infty,\ t_{cut}]$ 的概率

注意: **$h(\pi|t)$** 有时会被解释为检验统计量。

纽曼- 皮尔森引理与拒绝域

考虑一个多维检验统计量 $t=(t_1,\ldots,t_m)$，有信号假设 H_0 与本底假设 H_1。

问题：如何选择一个最佳的拒绝域或者 **cut**?

纽曼-皮尔森引理：在给定效率条件下，要得到最高纯度的信号样本，或者在给定的显著水平下得到最高的功效，可以选择下列接受域来实现

$$\frac{g(\vec{t}\,|H_0)}{g(\vec{t}\,|\,H_1)} > c = \text{用以决定效率的常数}$$

对于不含未定参量的最优化一维检验统计量，

$$r = \frac{g(\vec{t}\,|\,H_0)}{g(\vec{t}\,|\,H_1)}$$

➡ 简单假设 H_0 与 H_1 的似然之比

实际应用中，$\boldsymbol{r}$ 最好是单调函数。

如何构造一个检验统计量

在只考虑两种可能性的情况下，对于每个事例，测量

$$\vec{x} = (x_1, \ldots, x_n)$$

根据纽曼-皮尔森引理，为了选择事例，可选择检验统计量

$$t(\vec{x}) = \frac{f(\vec{x} \mid H_0)}{f(\vec{x} \mid H_1)}$$

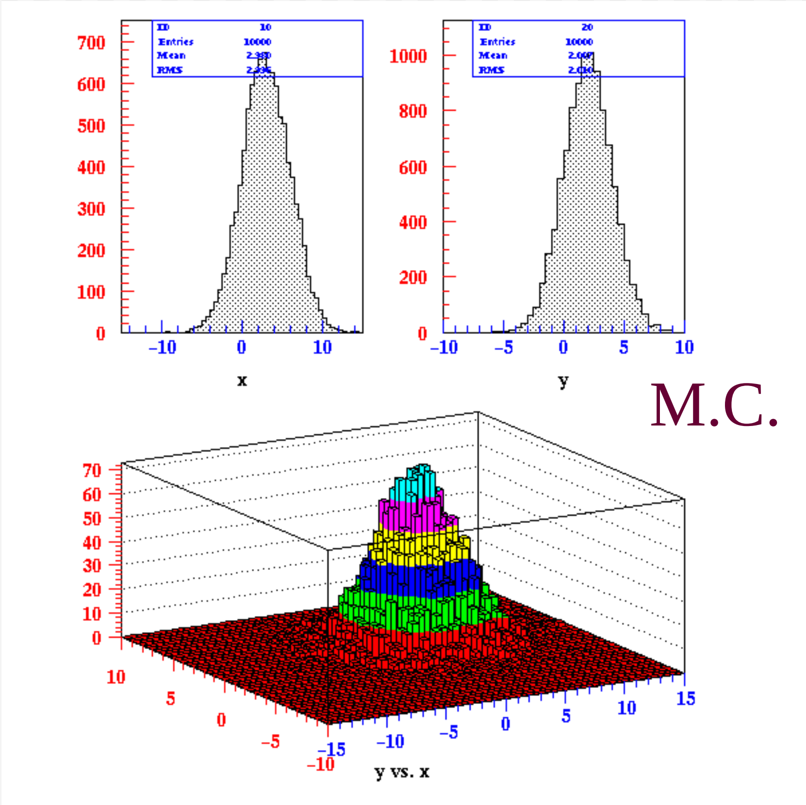
问题：如何知道这两个不同假设下的概率密度函数？

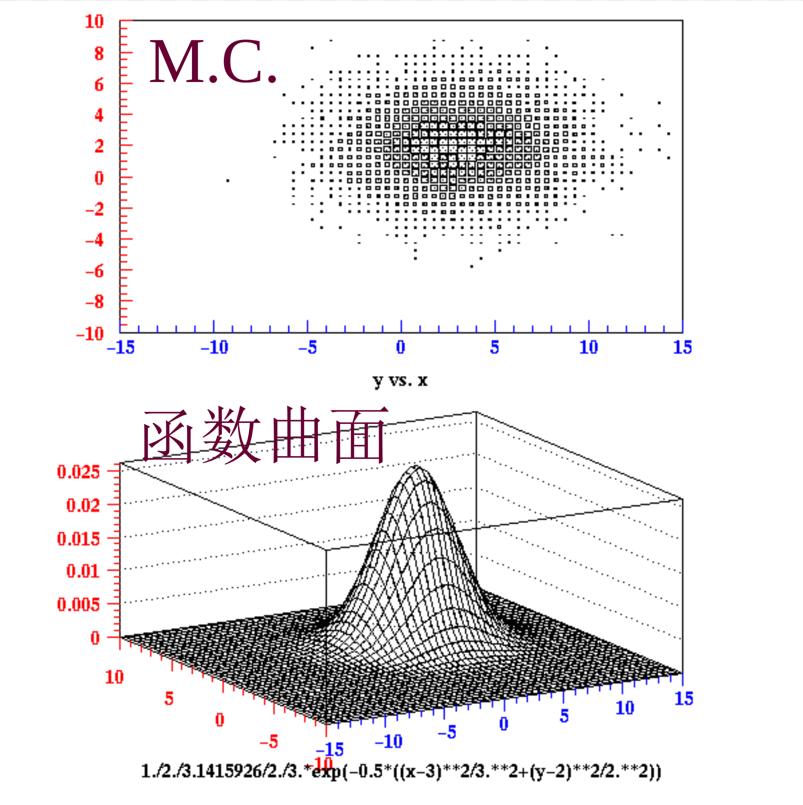
实际应用中，可以利用蒙特卡罗方法模拟物理过程与探测器响应，通过产生大量的样本，可以近似地得到上述概率密度函数的表达方式。

分别产生信号与本底事例, 并经过探测器模拟 → 对每个事例，得到测量量 $\vec{x}$，并填入$n-$维直方图。如果M为每个分量的区间数，则总单元数为M^n。 → $f(\vec{x} \mid H_0)$ $f(\vec{x} \mid H_1)$

但是如果 **n** 太大时，实际运用会很困难。

例子:蒙特卡罗近似求二维p.d.f.





分格子 → 统计每个格子的频数 → 近似的二维函数

如两者不相关 → 两个一维边缘分布 → $f(x,y)=f(x)f(y)$

线性检验统计量

当维数 >2时，用蒙特卡罗法找出多维概率密度函数依然较复杂。假设每一维研究均需要分M个区间，对于n-维问题，需要M^n个格子方能将密度度函数近似确定下来。为了简化处理此类问题，可以采用拟设的方法给出包含少量参数的检验统计量形式，通过确定参数(例如采用蒙特卡罗方法)，最大限度地区分 H_0与 H_1。

拟设：$t(\vec{x}) = \sum_{i=1}^{n} a_i x_i = \vec{a}^T \vec{x}$　　(即把测量量做线性叠加)

给定一个 $\vec{a}$，可以得到相应的概率密度函数 $g(t|H_0), g(t|H_1)$

通过选择 $\vec{a}$ 最大地区分 $g(t|H_0)$与$g(t|H_1)$ 的目的。

不同甄别的定义会导致在确定系数中有不同的规则，因此

必须定义所谓的区分量或甄别量。

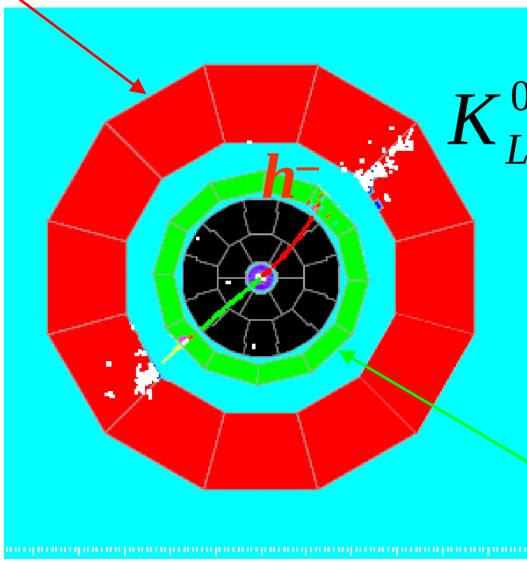
例子：对长寿命 *K* 介子的鉴别

- Two discriminating variables:

$$\delta_E = \frac{E-(E_{\pi^0})-P_h}{0.9\sqrt{P_h}} \quad (E:\ E_{HCAL}+E_{ECAL}),$$

$$\delta_\phi = sign \times |\phi_{barycenter} - \phi_{trackimpact}|.$$

强子量能器

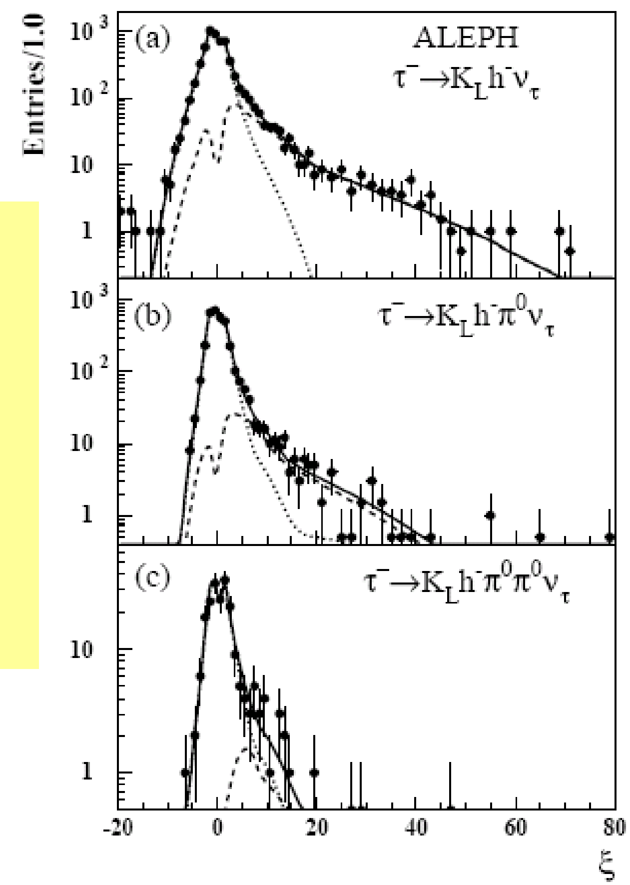


利用K_L^0粒子不受磁场影响而且较少发生电磁簇射的特点把它和带电强子区分开来。

$\Longrightarrow$ An event with K_L^0 corresponds to

$\delta_E > 0$ and $\delta_\phi < 0$.

电磁量能器



$$\xi = -\frac{\sqrt{2}}{2}(\delta_E - \Delta + \delta\phi)(\text{for } \delta_E > 0,\ \delta\phi > 0),$$

Δ 为常数，其余为实验观测量

$$\xi = \frac{\sqrt{2}}{2}(\delta_E - \Delta - \delta\phi)(\text{for } \delta_E > 0,\ \delta\phi \le 0),$$

Eur.Phys.J.C10，1（1999）

把一个**2-**维甄别问题简化为一维甄别问题。

对不同假设下的均值与方差要求

对已有的测量量，我们可以计算对应的期待值与协方差

$$(\mu_k)_i = \int x_i f(\vec{x}\,|\,H_k)d\vec{x} \qquad k = 0,1\ (假设)$$

$$(V_k)_{ij} = \int (x-\mu_k)_i (x-\mu_k)_j f(\vec{x}\,|\,H_k)d\vec{x} \qquad i,j = 1,\ldots,n\ (\vec{x}分量)$$

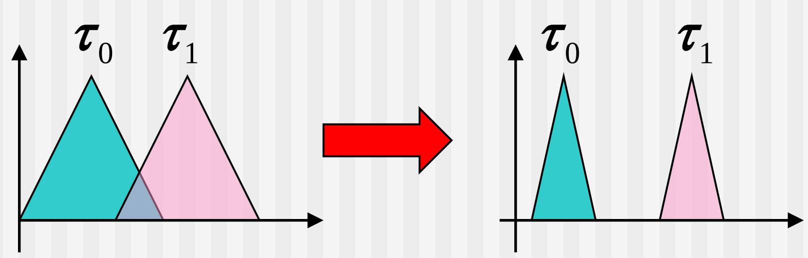
类似地，我们还可以导出计算 $t(\vec{x})$ 平均值与方差的公式

$$\tau_k = \int t(\vec{x})\, f(\vec{x}\,|\,H_k)d\vec{x} = \vec{a}^T \vec{\mu}_k$$

$$\Sigma_k^2 = \int (t(\vec{x}) - \tau_k)^2 f(\vec{x}\,|\,H_k)d\vec{x} = \vec{a}^T V_k \vec{a}$$

要求大的 $|\tau_0 - \tau_1|$ 与小的 Σ_0^2 , Σ_1^2

使得 pdfs 分布集中在均值附近。



Fisher 甄别函数的定义

Fisher 定义了一个甄别法

$$J(\vec{a})=\frac{(\tau_0-\tau_1)^2}{\sum_0^2+\sum_1^2}$$

$$=\sum_{i,j=1}^{n}a_ia_j(\mu_0-\mu_1)_i(\mu_0-\mu_1)_j=\sum_{i,j=1}^{n}a_ia_jB_{ij}=\vec{a}^TB\vec{a}$$

$$=\sum_{i,j=1}^{n}a_ia_j(V_0+V_1)_{ij}=\vec{a}^TW\vec{a}$$

则

$$J(\vec{a})=\frac{\vec{a}^TB\vec{a}}{\vec{a}^TW\vec{a}}$$

令 $\frac{\partial J}{\partial a_i}=0$ $\Rightarrow$ $\vec{a}\propto W^{-1}(\vec{\mu}_0-\vec{\mu}_1)$ (证明见习题)

因此定义了可求极值的**Fisher** 线性甄别函数 ***J***。

求Fisher 甄别函数的最大值

若将 $t(\vec{x})$ 写成

$$t(\vec{x}) = a_0 + \sum_{i=1}^{n} a_i x_i$$

用任意标度和偏置 a_0 去固定 τ_0，τ_1

求 $J(\vec{a}) = \dfrac{(\tau_0 - \tau_1)^2}{\Sigma_0^2 + \Sigma_1^2}$ 的最大值, 意味着要将下式最小化

$$\Sigma_0^2 + \Sigma_1^2 = E_0[(t-\tau_0)^2] + E_1[(t-\tau_1)^2]$$

与假设对应的期待值

求 Fisher 函数 $J(\vec{a})$ 的最大值就是以后介绍的最小二乘法原理中的一种。

高斯分布下Fisher 甄别量特点

假设 $f(\vec{x}\,|\,H_k)$是多变量高斯分布，具有平均值

$\vec{\mu}_0$ 为假设H_0的均值 $\vec{\mu}_1$ 为假设H_1的均值

而且，两者的协方差矩阵为 $V_0 = V_1 \equiv V$

含偏置的 **Fisher** 甄别量为 $t(\vec{x}) = a_0 + (\vec{\mu}_0 - \vec{\mu}_1)^T V^{-1} \vec{x}$

利用前面所述的似然比对给定效率条件下的最大纯度

$$r = \frac{f(\vec{x}\,|\,H_0)}{f(\vec{x}\,|\,H_1)} = \exp\left[-\frac{1}{2}(\vec{x}-\vec{\mu}_0)^T V^{-1}(\vec{x}-\vec{\mu}_0) + \frac{1}{2}(\vec{x}-\vec{\mu}_1)^T V^{-1}(\vec{x}-\vec{\mu}_1)\right]$$

$$\propto e^t$$

→ $t \propto \log(r)$ + 常数 (单调变化) →

Fisher 甄别量与似然比等效。

如果不是多变量高斯分布，上式不成立。

验后概率与逻辑函数

具有相同协方差矩阵的多变量 $\vec{x}$ 还可给出验后概率的简单表达式，例如

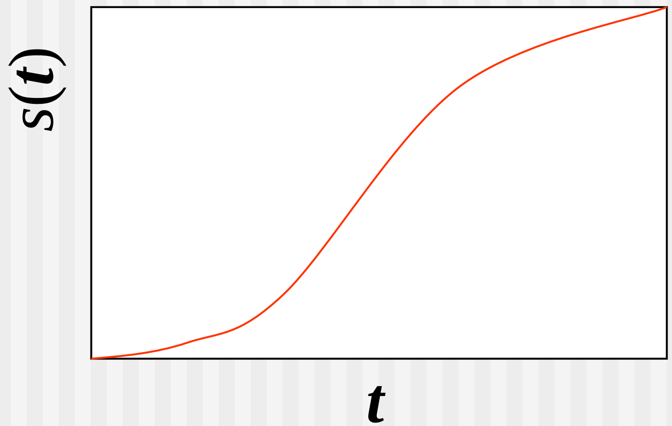
$$P(H_0|\vec{x}) = \frac{f(\vec{x}|H_0)P(H_0)}{f(\vec{x}|H_0)P(H_0)+f(\vec{x}|H_1)P(H_1)} = \frac{1}{1+\frac{P(H_1)}{P(H_0)r}}$$

贝叶斯定理

验前概率

选择恰当的偏置$\boldsymbol{a_0}$，利用高斯分布下**Fisher**甄别量的特点，上式可写为

$$P(H_0|\vec{x}) = \frac{1}{1+e^{-t}} \equiv s(t)$$



也就是所谓的“逻辑σ”函数

非线性检验统计，神经网络(一)

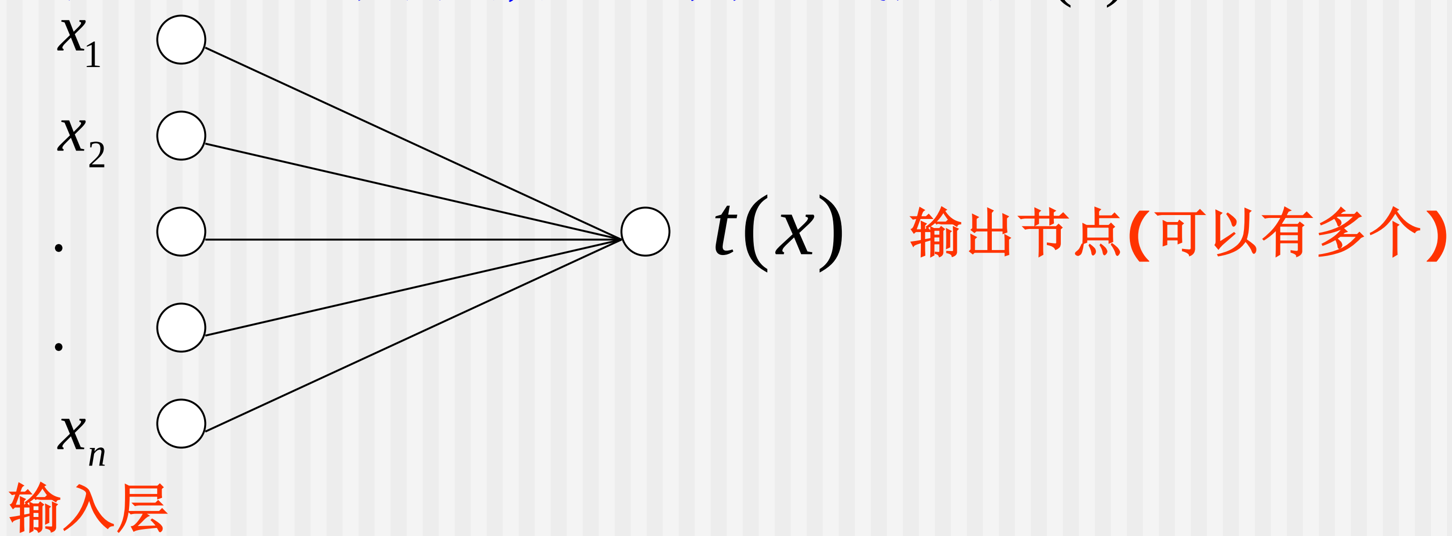
如果不同假设下得到实验观测量的概率密度函数 $f(\vec{x}\,|\,H_0)$ 与 $f(\vec{x}\,|\,H_1)$ 不是高斯或无共同的协方差矩阵，**Fisher**甄别方法不再适用。此时可以采用更为一般的所谓神经网络方法

假设统计检验量

$$t(\vec{x}) = s(a_0 + \sum_{i=1}^{n} a_i x_i)$$

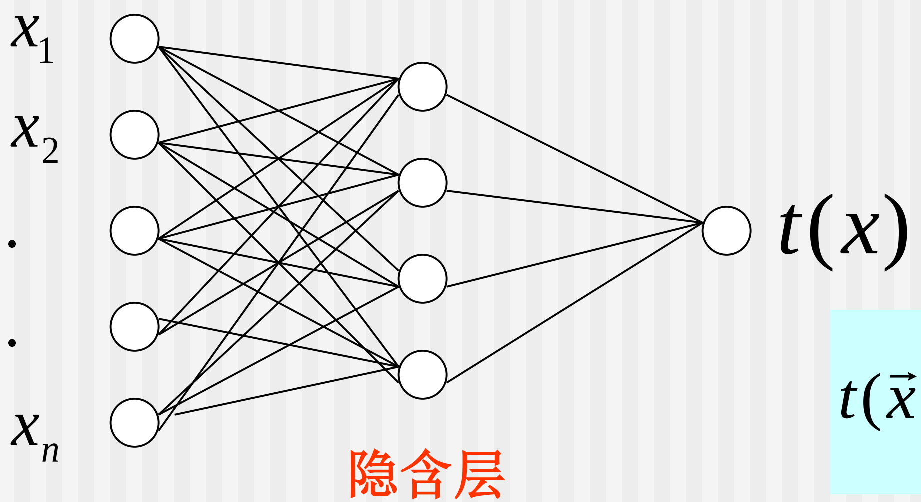
$$\boxed{\begin{aligned} s(u) &= (1+e^{-u})^{-1} \\ &= \text{激活函数} \end{aligned}}$$

是单层的感知器。s是单调的, 因此等效于线性的 $t(\vec{x})$



多层感知器的神经网络

推广到多层感知器



输出定义为

$$t(\vec{x}) = s[a_0 + \sum_{i=1}^{n} a_i h_i(\vec{x})]$$

上一层节点函数可写为

$$h_i(\vec{x}) = s(w_{i0} + \sum_{j=1}^{n} w_{ij} x_j)$$

越多节点 → 神经网络越接近优化的 $t(x)$

但需要定更多的参数!

a_i, w_{ij}为权重或者联结强度。

神经网络中的误差函数最小化

参数取值通常根据误差函数的最小化结果来决定

$$\varepsilon = E_0[(t-t^{(0)})^2] + E_1[(t-t^{(1)})^2]$$

这里 $t^{(0)}$，$t^{(1)}$ 为目标值，例如选 **0** 和 **1** 的逻辑 σ 函数值

实际应用中，通常以蒙特卡罗的训练样本平均值来取代期待值。**(调整参数值=神经网络的学习过程)**

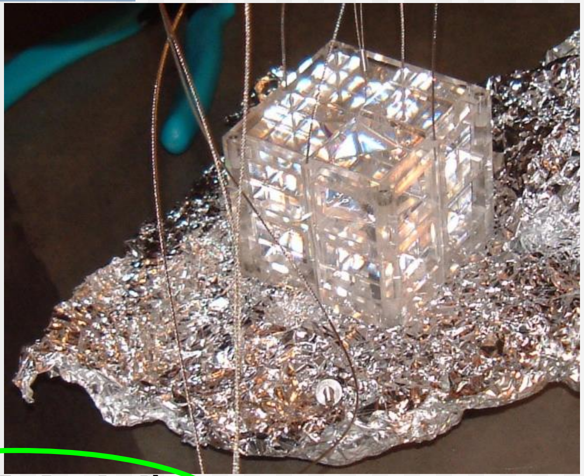
在核物理与粒子物理研究中，是通过定义信号与本底两个样本，从样本中给出每个事例的相关测量量**(**例如，动量，飞行时间**…)**，然后直接调用欧洲粒子物理实验室**(CERN)**提供的物理分析软件包**ROOT(**基于**C++)PAW(**基于**Fortran)**，得到训练后的参数与输出量，并将它们用于待分析的事例来决定其是本底还是信号。具体应用参见下列网站

PAW 用户：http://paw.web.cern.ch/paw/mlpfit/pawmlp.html

ROOT用户：http://root.cern.ch/root/html/examples/mlpHiggs.C.html

例子：用神经网络甄别中子信号

为了在一个**5**万吨水的超级神冈实验探测器上探测中子信号，进行了如下实验



~ 200 μs

Delayed

Prompt

$$\bar{\nu}_e + p \to e^+ + n$$

$$n + p \to \ldots\ldots \to n + p \to d + \gamma(2.2\ \mathrm{MeV})$$

[Am/Be]

Prompt

$$\alpha + {}^{9}\mathrm{Be} \to {}^{12}\mathrm{C}^* + n$$

Delayed

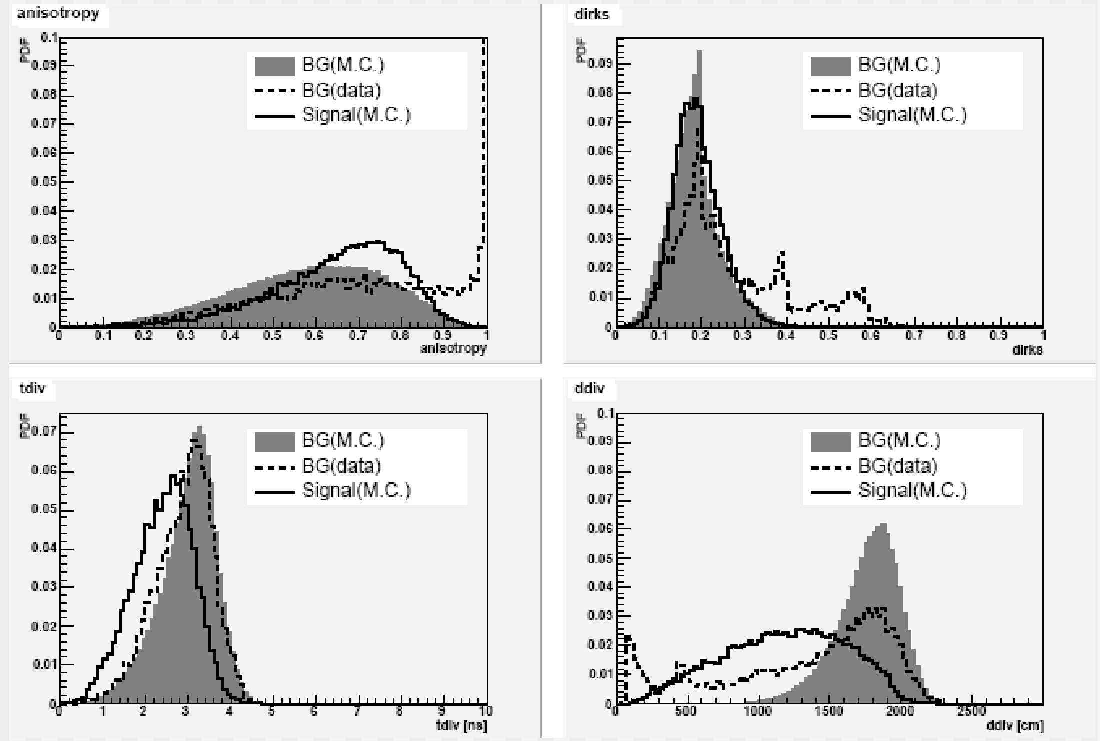
$$ {}^{12}\mathrm{C}^* \to {}^{12}\mathrm{C} + \gamma(4.4\ \mathrm{MeV})$$

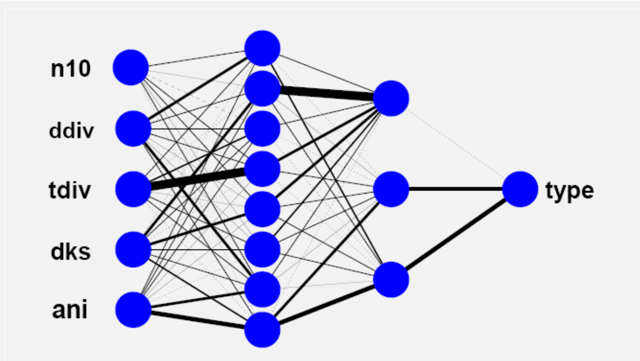
$$n + p \to \ldots\ldots \to n + p \to d + \gamma(2.2\ \mathrm{MeV})$$

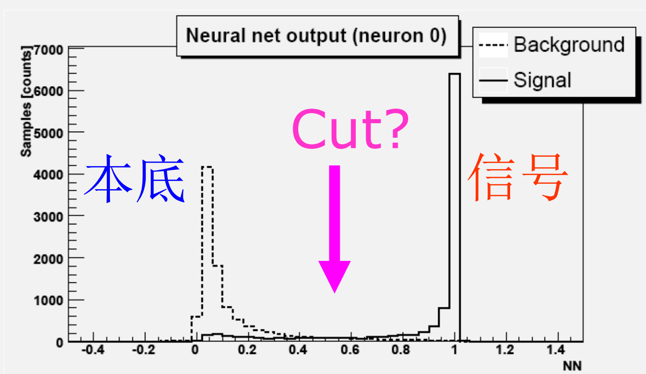
希望在纯水中观测到中子被水俘获的现象。已知数据中有大量包括诸如光电倍增管噪音的本底。

神经网络输入量与甄别量

神经网络
输入变量



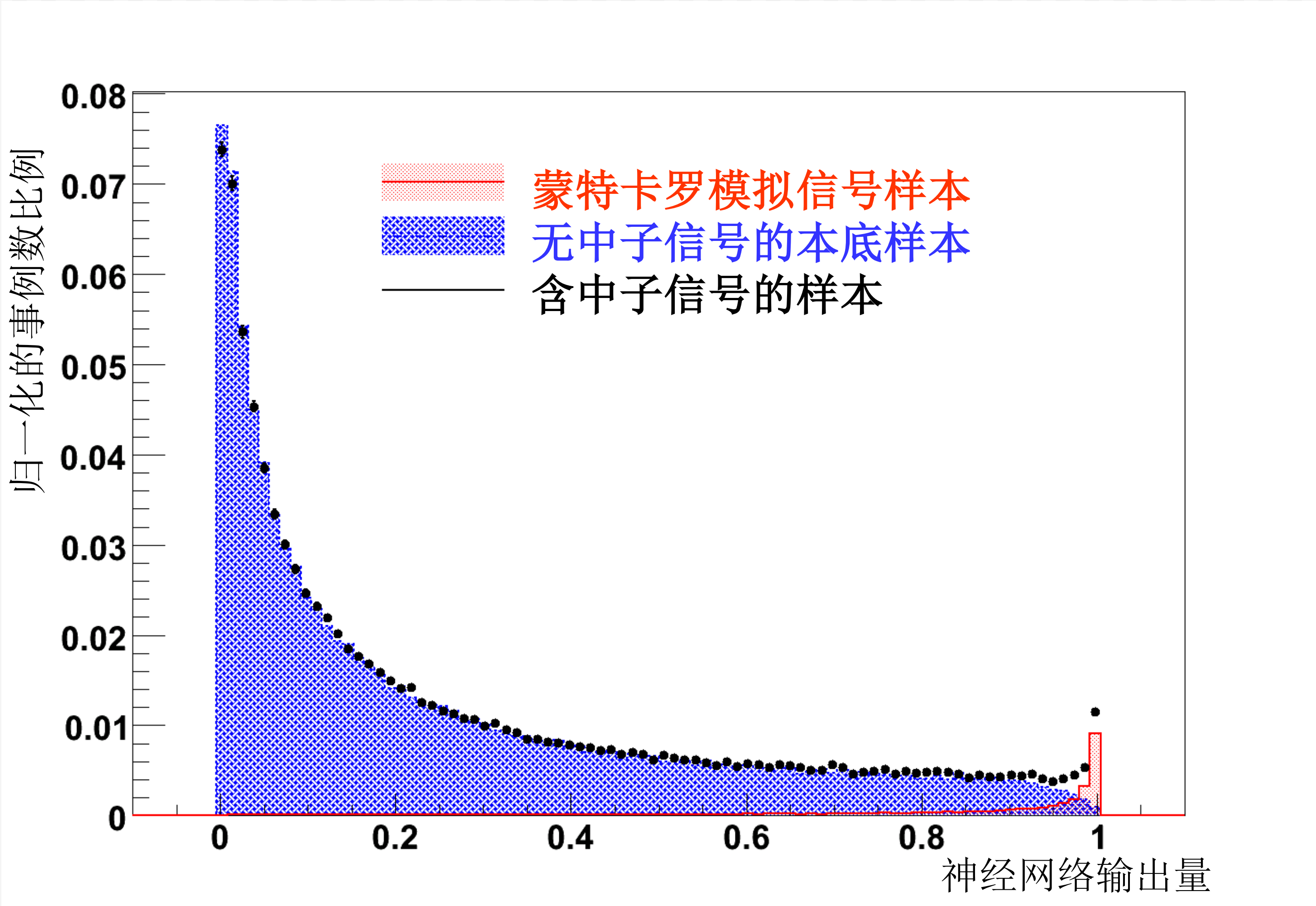




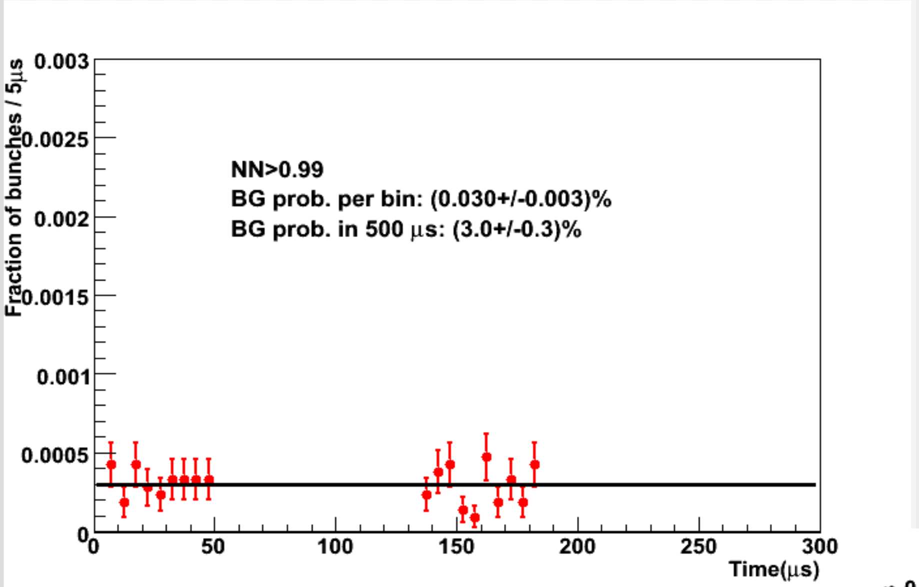
神经网络
输出量

选择拒绝域使得信噪比与效率最大。

从神经网络输出量中甄别中子

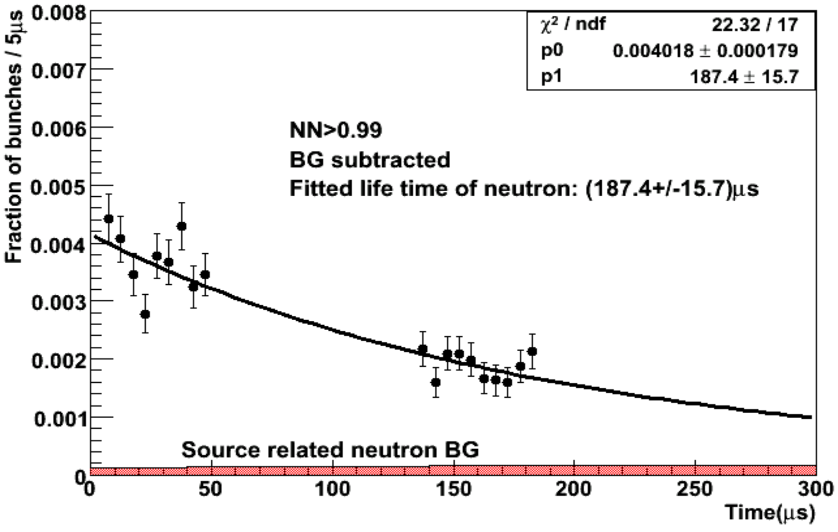


神经网络甄别后的中子信号



在本底无中子信号数据中看不到体现中子被水中的氢原子俘获的寿命特征

在含中子源信号的数据中看到了体现中子被水中的氢原子俘获的寿命特征



关于神经网络的输入变量问题

问题：是否输入量越多越好？

较少的输入量 → 较少的可调参数

→ 在有限的样本中，参数可以得到很好的确定

如果输入量之间中有很强的相关情形，应只保留一个。

如果输入量对甄别无太大影响，应弃之。

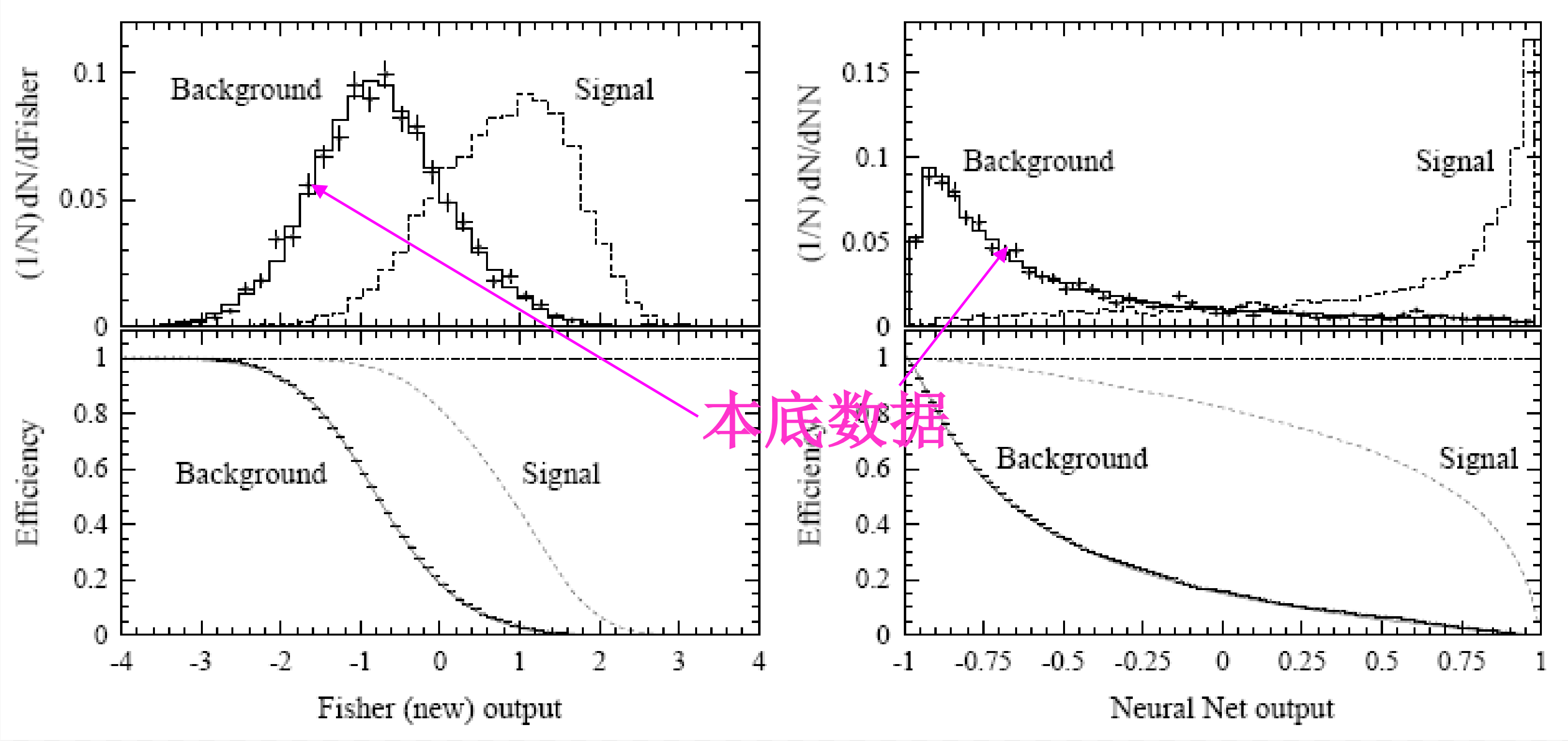
神经网络利用了较高阶矩的联合概率密度函数 $f(\vec{x}\,|\,H)$, 它们也许在训练的样本中找不到较好的模型来描述

→ 最好简化 $t(\vec{x})$, 只要它还能恰当地描述样本。

避免输入量和要研究的信号特征量相关联。

Fisher方法与神经网络

- **Fisher** 方法只适用于用线性方法构造统计量。
- 神经网络在应用上更具有普遍性和更大的甄别能力。



有研究表明，同等本底大小的情况下，神经网络有时能使效率增加15%。参见 arXiv:hep-ex/0107075

极端情况下的拟合优度检验

前面讲了统计检验的甄别问题，但在实际情况中还要处理极端情况下无效假设的拟合优度检验问题。

任意投掷一枚硬币，结果为正面与反面的概率都是**0.5**。

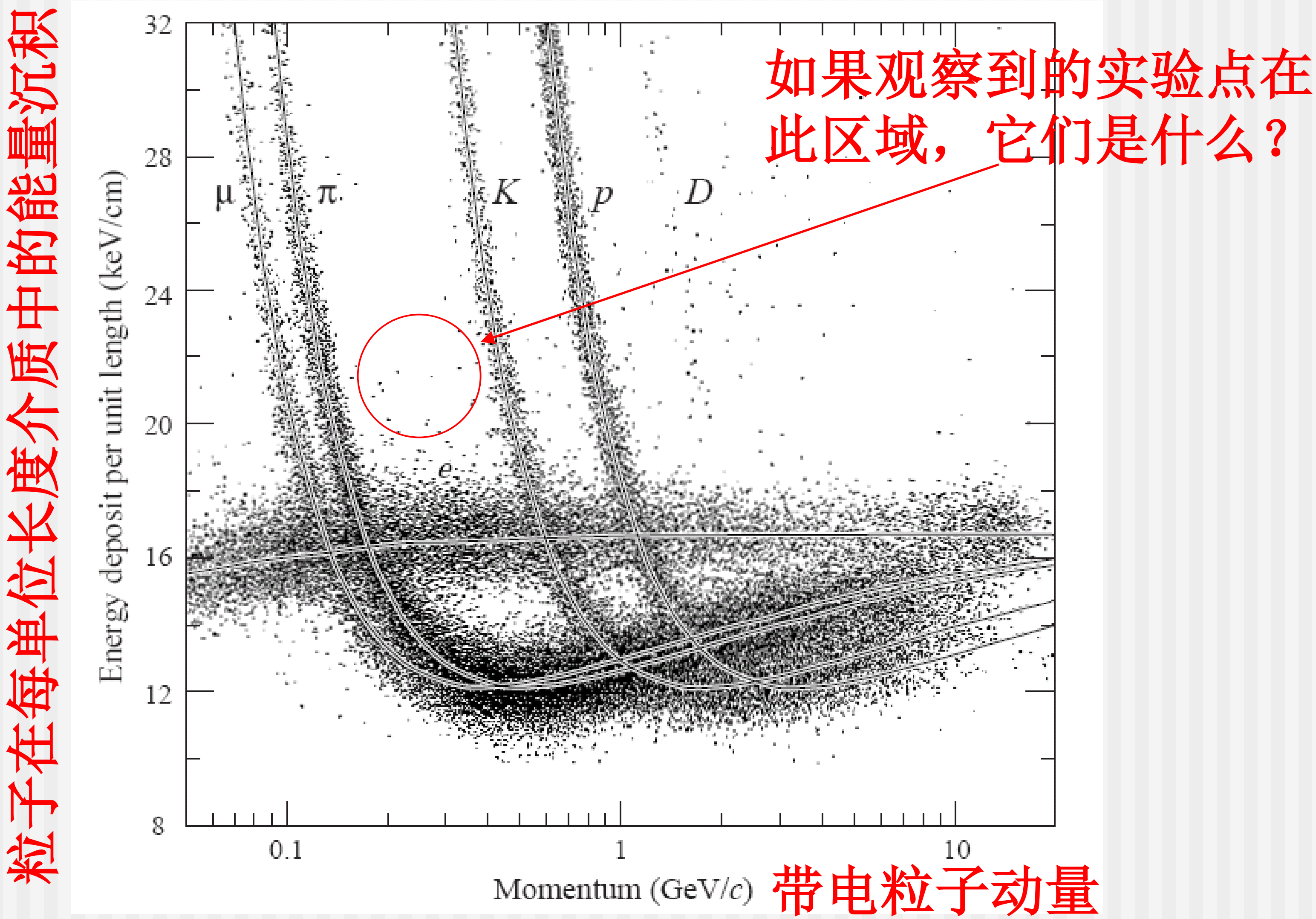
如果有人声称对此进行了检验。投了**20**次，得到了**17**次正面的结果。那么能否断定得到正面的概率应该是

$$p_h = 0.85 \pm 0.08$$

也就是说与预期值 **0.5** 有 **4** 个标准偏差呢？

问题：理论上允许这样的极端情况出现吗？ 或者说与这样一种极端情况相等或更高的概率有多大？

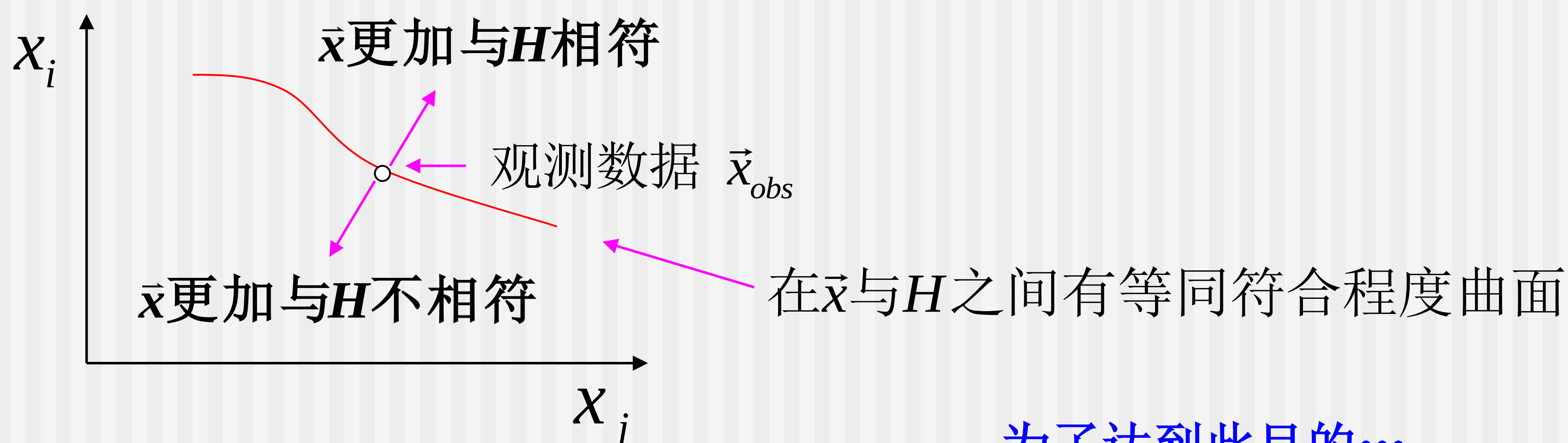
例子：粒子鉴别中常遇到的问题



检验拟合优度

如果假设H对数据中的一部分矢量$\vec{x}=(x_1, x_2,\ldots,x_n)$给出了预言$f(\vec{x}\,|\,H)$。我们在 $\vec{x}$-空间观察到一个点:$\vec{x}_{obs}$。从数据来看，对假设H的正确与否会得出什么样的结论呢?

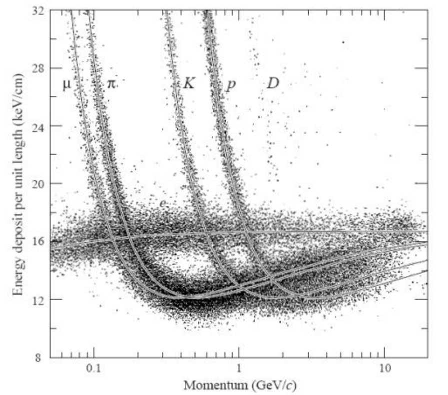
需要决定$\vec{x}$-空间中哪一部分比观测点$\vec{x}_{obs}$更能代表与假设H的不相符。



为了达到此目的…

检验统计量与拟合优度

通常需要构造统计检验量$t(\vec{x})$，它的大小可以反映出在$\vec{x}$与H之间符合的程度。例如



$$t = \left(\frac{(dE/dx)_K^{Theo} - (dE/dx)_K^{Obs}}{\sigma} \right)^2$$

小的 t → 数据与 H 更符合

大的 t → 数据与 H 更不符合

由于概率密度函数 $f(\vec{x}\,|\,H)$已知，因此在H假设条件件下检验统计量t 的概率密度函数 $g(t\,|\,H)$是完全可以确定。

P-值定义

将拟合优度用*P*-值表示（也称为观察的显著水平或置信水平）

P=观察到实验数据 $\vec{x}$ 或 $t(\vec{x})$像 $\vec{x}_{obs}$ 或 $t(\vec{x}_{obs})$一样，与假设H具有相同或较小符合程度的概率。

注意: 这不是 *H* 为真的概率。

在经典统计学上，我们从不涉及 ***P*(*H*)** 。

而在贝叶斯统计理论中，则把 ***H*** 当成了随机变量，并利用贝叶斯定理得到

$$P(H\,|\,t) = \frac{P(t\,|\,H)\pi(H)}{\int P(t\,|\,H)\pi(H)dH}$$

π(*H*):*H* 的先验概率

对所有可能性进行归一化积分

P-值与假设检验

根据 *P*-值的定义，对 *H* 假设拟合优度的检验可以通过计算*P*-值的大小来完成。但是应注意以下两点：

- 在*P*-值定义中不涉及别的假设。
- *P*-值是一个随机变量。前面的显著水平在检验时已经被指定为常数。

如果 *H* 为真，则对于连续的 $\vec{x}$，*P* 在[0,1]范围内均匀分布.

如果 *H* 非真，则 *P* 的概率密度函数通常很接近零。

例子:拟合优度检验

投 N 次硬币，观察到 n_h 次头朝上的概率服从二项式分布：

$$f(n_h; p_h, N) = \frac{N!}{n_h!(N-n_h)!} p_h^{n_h} (1-p_h)^{N-n_h}$$

假设H：硬币是公平的(朝上的 p_h = 朝下的 p_t = 0.5)

取拟合优度检验统计量 $t = |n_h - N/2|$

投 N=20 次硬币，观察到17次头朝上，则 $t_{obs} = |17 - 20/2| = 7$

在 t-空间中，具有相同或较少符合的区域为

$$t = (n_h - N/2) \geq 7$$

$$P - 值 = P(n_h = 0,1,2,3,17,18,19,20) = \sum_{i \leq 8} f_i \approx 0.0026$$

拟合优度检验中的问题

问题:当 *P*-值等于0.0026,是否意味着 *H* 假设是错的?

P-值并不回答此问题。它只是给出与观察到的结果一样,与 *H* 假设不符或者高于 *H* 假设($p_h=p_t=0.5$)的概率。

P-值="偶然"得到如此奇怪结果的概率

一种实用的检验方法是在同样的假设下,产生同样数目的事例足够多次。检查如此奇怪的结果发生的概率是否与*P*-值相当。

观测到一个信号的显著程度

假设观测 n 个事例，包含了

n_b=已知过程(或本底)的事例数　　n_s=新过程(或信号)的事例数

如果n_b，n_s服从泊松分布，均值为ν_b，ν_s，它们之和$n=n_b+n_s$也是服从泊松分布，均值为$\nu=\nu_b+\nu_s$:

$$P(n;\nu_s,\nu_b)=\frac{(\nu_s+\nu_b)^n}{n!}e^{-(\nu_s+\nu_b)}$$

如果 ν_b=0.5，而且观测到 n_{obs}=5 → 可否就此声称该迹象为新的发现？

假设 H：ν_s=0，即只有本底过程出现。

→ 也就是所谓的“无效假设”

观测到一个信号的显著程度(续)

对应的***P***−值

$$
\begin{aligned}
P-\text{值} &= P(n \geq n_{obs}) \\
&= \sum_{n=n_{obs}}^{\infty} P(n; \nu_s = 0, \nu_b) \\
&= 1 - \sum_{n=0}^{n_{obs}-1} \frac{\nu_b^n}{n!} e^{-\nu_b} \\
&\approx 1.7 \times 10^{-4} \\
&(\neq P(\nu_s = 0)!)
\end{aligned}
$$

给出了得到这种极端结果的概率：虽然很小但不为零！

潜在的问题之一

一个误导读者但又常常被使用的结果表示...

对 ν_s 估计时得到：$n_{obs}=5$

估计n 的标准偏差为:$\sqrt{n}=2.2$ } 信号

ν_s的估计值：$n_{obs}-\nu_b=4.5\pm2.2$ 即与零有两倍的标准偏差

实际想要的是：均值 ν_b=0.5的泊松变量给出观测量大于 5 概率是多少？

概率为1.7×10^{-4}

但上面的结果表示隐含了均值为4.5，σ = 2.2的高斯变量给出零或更少的概率：

$$\int_{-\infty}^{0}\frac{1}{\sqrt{2\pi}\times2.2}\exp\left(\frac{-(x-4.5)^2}{2\times2.2^2}\right)dx=0.021$$

如果ν_s >>1，没有问题，即 n 服从高斯分布。

潜在的问题之二

实际问题中会涉及系统误差，例如 ν_b=0.8，则概率变为

$$P\text{-值} = P(n \geq 5; \nu_b = 0.8, \nu_s = 0)$$

$$= \sum_{n=n_{obs}}^{\infty} P(n;\ \nu_b = 0.8, \nu_s = 0)$$

$$= 1 - \sum_{n=0}^{n_{obs}-1} \frac{\nu_b^n}{n!} e^{-\nu_b}$$

$$= 1.4 \times 10^{-3}$$

虽然本底只增大了0.3，但却比 ν_b=0.5 时小了一个量级。

建议给出与 ν_b 合理变化相对应的*P*-值范围

皮尔逊的 χ^2 检验

在观测的数据 $\vec{n}=(n_1,\ldots,n_N)$与预言的期待值 $\vec{\nu}=(\nu_1,\ldots,\nu_N)$ 之间进行比较的检验统计量

$$\chi^2=\sum_{i=1}^{N}\frac{(n_i-\nu_i)^2}{\nu_i}$$

如果 n_i 是相互独立而且服从均值为ν_i泊松分布，所有ν_i 并不太小(>5)，那么χ^2 将服从 N 个自由度的最小二乘概率密度函数分布。所观察的χ^2可给出P-值

$$P-\text{值}=\int_{\chi^2}^{\infty} f(z;N)dz$$

这里，$f(z;N)$ 自由度为N 的最小二乘概率密度函数。

皮尔逊的 χ^2 检验(续)

自由度为 N 的最小二乘概率密度函数的期待值为 $E(z)=N$

→ 通常以 χ^2/N 来体现符合的程度

最好分别给出 χ^2，N，例如

$$\chi^2=15,\quad N=10\rightarrow P-\text{值}=0.13$$

$$\chi^2=150, N=100\rightarrow P-\text{值}=9.0\times10^{-4}$$

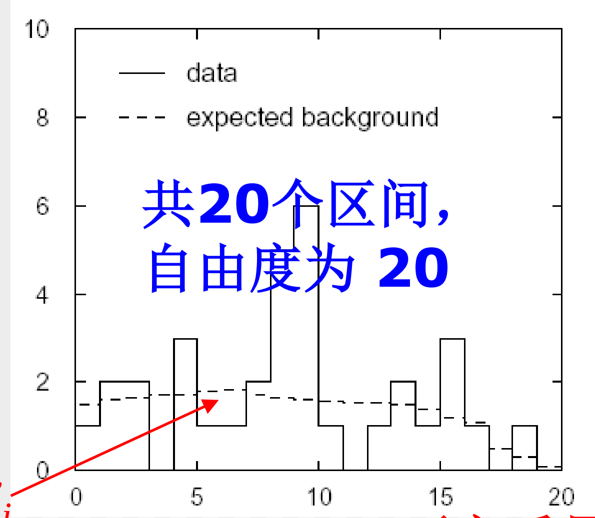
如果$n_{tot}=\sum_{i=1}^{N}n_i$固定，$n_i$服从二项式分布，$p_i=\nu_i/n_{tot}$,则

$$\chi^2=\sum_{i=1}^{N}\frac{(n_i-p_in_{tot})^2}{p_in_{tot}}$$

← 服从N-1自由度的χ^2分布($p_i\, n_{tot}>>1$)

例子: χ^2 检验

频数



ν_i

不变质量

$$\chi^2 = \sum_{i=1}^{N} \frac{(n_i - \nu_i)^2}{\nu_i} = 29.8$$

1）由于许多区间只有很少或根本没有计数，它将不服从 χ^2 的概率密度函数分布。

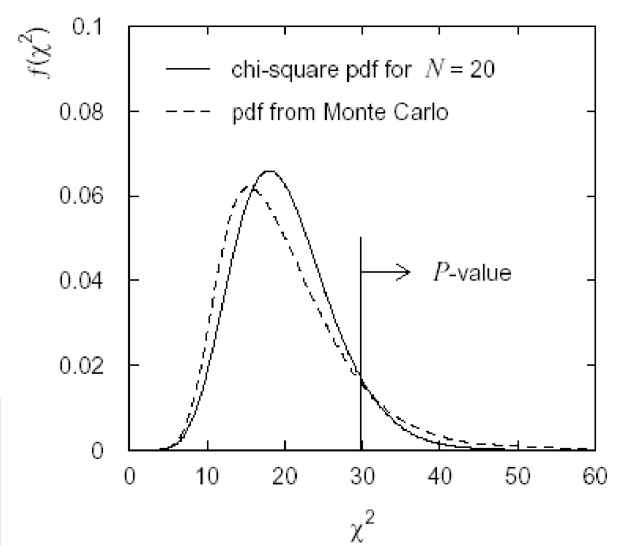
2）皮尔森 χ^2 仍可以作为一个检验统计量。

为计算*P*-值，先用蒙特卡罗方法得到 $f(\chi^2)$

产生 n_i 均值为 ν_i 的泊松分布，$i=1,\ldots,N$

计算 χ^2，填入直方图

重复足够多次



MC pdf：*P*-值=0.11

χ^2 pdf: *P*-值=0.073

对于统计检验的评论

在实际问题中，我们常常遇到对低统计量的情况下，需要判断所观察到的现象是否为真正的物理信号。利用***P*-**值的大小可以表示结果是否为已知过程的极端情形。由于每个人的信心不同，会造成同一个***P*-**值，结论却完全不一样的现象。

在统计误差范围内无新迹象。

结果虽然在统计误差范围，但有可能是新物理的信号。

发现了新物理的信号，误差为…

历史上类似故事的发生很多：J/Ψ粒子的发现，*W*粒子的发现，顶夸克的发现…

小结

- **统计检验：**

  检验在何种程度上，数据与假设相符。

- **检验统计量：**

  将矢量 $\vec{x}$ 简化为一个或几个分量的矢量 $t(\vec{x})$

- **检验的要点：**

  关键区，显著水平，功效，纯度，效率。

- **纽曼-皮尔森引理：**

  在给定效率条件下，给出纯度最大区。

- **构造检验统计量：**

  最好是似然比，但通常需太多待定参数。

- **统计分析中两种方法：**

  Fisher 甄别函数(线性的)；神经网络(非线性的)。

小结（续）

- **检验拟合优度，*P*-值定义与应用**

  P-值为得到数据像已观测的结果一样与假设不符或更不符合的概率。

- **信号观测的显著程度**

  很复杂，许多具有 10^{-4} 效应的结果最终证明是统计涨落的受害者。

- **皮尔逊 χ^2 检验**

  广泛用于检验统计量。对于小样本数据，它将不服从 χ^2 的概率密度函数分布。但仍可用蒙特卡罗得到概率密度函数分布。