

講義ノート4

- 統計の基礎
 - なぜ、統計がひつようか？
 - 確率密度関数

- 近年の高エネ分野における統計
 - 発見するとは？棄却するとは？
 - 正しさを追及して複雑化した検定。

統計の基礎

Why do we do experiments?

-- by Luis Lyons

- We set out to determine the numerical value of some physical quantity.

– ある物理量を測定する

例) 光速度を測定する。

- We are testing whether a particular theory is consistent with our data.

– 特定の理論/仮説(Hypothesis)を検証/棄却する

例) Hypothesis : 今月初めから光速度が3%増した。

光速度の測定値をこのHypothesisと比較

Why estimate uncertainty?

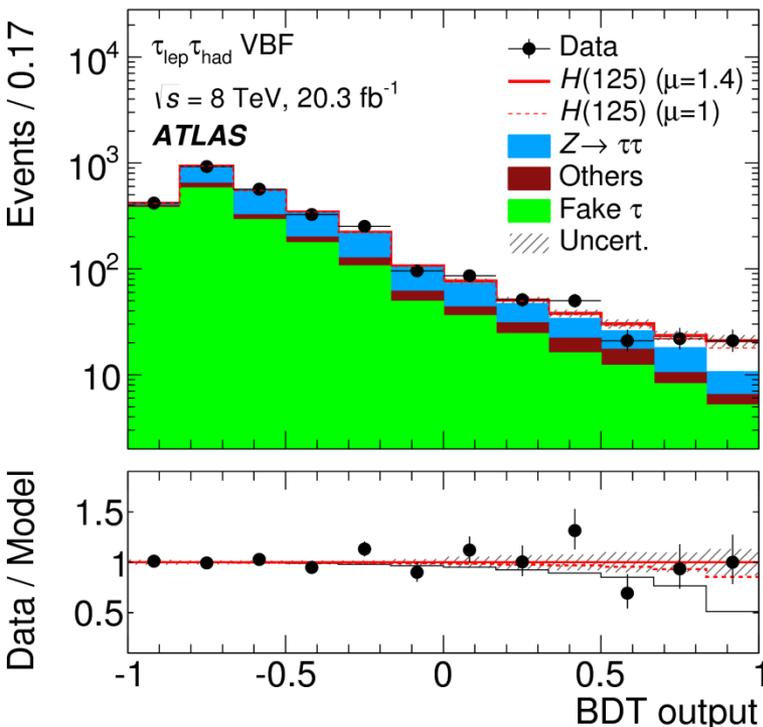
- 例) Hypothesis : 今月初めから光速度が3%増した。を検証する。
- ある実験で、光速度の測定値が $3.09 \times 10^8 \text{ m/s}$ だった。
 - 精度良い測定: $2.99792 \times 10^8 \text{ m/s}$ (本当はこれは実測値なので誤差がある) : Null Hypothesis
 - 3%増したのなら、 $3.08754 \times 10^8 \text{ m/s}$ になってるはず。: Test Hypothesis
- 可能性として、
 - $3.09 \pm 2 \times 10^8$ ($3 \pm 2 \times 10^8$)
 - そもそもこのHypothesisを検証できる精度ではない。
 - $3.09 \pm 0.15 \times 10^8$
 - 測定値は精度良い測定とも、Hypothesisとも無矛盾。
 - $3.09 \pm 0.01 \times 10^8$
 - よい精度でこのHypothesisを検証できた。
- このように、Uncertaintyなしの測定値は何の意味もない！

統計誤差と系統誤差

- **統計誤差(Statistical uncertainty)**
 - 統計のふらつきに由来する誤差
 - 繰り返しある物理量の測定を行ったと仮定したときのばらつき。通常**確率密度関数**で表し(ポアソン分布、ガウス分布、二項分布...)その**標準偏差**を統計誤差とする。
- **系統誤差(Systematics uncertainty)**
 - 統計誤差以外の誤差
 - 通常、測定器の誤差や理論的な不定性などが含まれる
例) 鉄球の直径をノギスで測った時、精度は0.05mm
 - あまり決まった方法がないことが多く、「えいや！」とつけることが多い(物理屋の良心)。ただ、統計誤差と並べて評価されることが多いため、大体の確率の目安は考えるべき。

(余談1) データ数のerror

- 高エネルギーの結果でよく見られるdataとmodelを比較するヒストグラムのデータ点にエラーバーが書かれているが、厳密には間違い。
- 1回の観測で得られたデータ数にはuncertaintyはなく、modelの方にuncertaintyがあるのが正しい。テーブル参照。**
- 但し、ATLASの結果は、慣例にならってPoissonのsigmaをエラーバーとして書くことが許されている。



Process/Category	VBF		
	All bins	Second to last bin	Last bin
BDT output bin			
Fake background	1680 ± 50	8.2 ± 0.9	5.2 ± 0.7
$Z \rightarrow \tau\tau$	877 ± 29	7.6 ± 0.9	4.2 ± 0.7
Top	82 ± 15	0.3 ± 0.4	0.5 ± 0.4
$Z \rightarrow \ell\ell (\ell \rightarrow \tau_{had})$	54 ± 26	1.0 ± 0.7	0.30 ± 0.28
Diboson	63 ± 11	1.0 ± 0.4	0.48 ± 0.20
ggF: $H \rightarrow \tau\tau (m_H = 125 \text{ GeV})$	16 ± 6	1.0 ± 0.4	1.2 ± 0.6
VBF: $H \rightarrow \tau\tau$	31 ± 8	4.5 ± 1.1	9.1 ± 2.2
WH: $H \rightarrow \tau\tau$	0.6 ± 0.4	< 0.1	< 0.1
ZH: $H \rightarrow \tau\tau$	0.16 ± 0.07	< 0.1	< 0.1
Total background	2760 ± 40	18.1 ± 2.3	10.7 ± 2.7
Total signal	48 ± 12	5.5 ± 1.3	10.3 ± 2.5
Data	2830	22	21

確率密度関数の例

中心極限定理

ある真の値がある場合、無作為の標本は
サンプルサイズを増やすと真の平均に近づく。

平均(mean)・分散(variance)・中央値(median)

余談2

- (真の値のある)母集団 x_i に対して

$$\text{平均: } \mu = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n}$$

$$\text{分散 (平均}\bar{x}\text{): } \sigma^2 = \frac{1}{m} \sum_{i=1}^m (\bar{x} - x_i)^2$$

$$\text{中央値: } Q_{\frac{1}{2}}(x) = \begin{cases} x'_{\frac{n+1}{2}}, & \text{if } n \text{ is odd.} \\ \frac{1}{2}(x'_{\frac{n}{2}} + x'_{\frac{n}{2}+1}), & \text{if } n \text{ is even.} \end{cases}$$

小さい方から x_1, x_2, \dots

- 連続確率分布関数 $f(x)$

$$\text{平均: } E[X] = \int_{-\infty}^{\infty} x f(x) dx (= \mu)$$

$$\text{分散 (平均}\mu\text{): } V[X] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

$$\text{標準偏差: } \sigma = \sqrt{V[X]}$$

$$\text{中央値: } \int_{-\infty}^m dF(x) \geq \frac{1}{2} \text{ and } \int_m^{\infty} dF(x) \geq \frac{1}{2}$$

標準偏差とRMS?

$$x_{\text{rms}} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_N^2}{N}}$$

$\mu = 0$ の時、 $\sigma = x_{\text{rms}}$

```
//
Double_t TH1::GetRMS(Int_t axis) const
{
    // For axis = 1, 2 or 3 returns the Sigma value of the histogram along
    // X, Y or Z axis
    // For axis = 11, 12 or 13 returns the error of RMS estimation along
    // X, Y or Z axis for Normal distribution
    //
    // Note that the mean value/sigma is computed using the bins in the currently
    // defined range (see TAxis::SetRange). By default the range includes
    // all bins from 1 to nbins included, excluding underflows and overflows.
    // To force the underflows and overflows in the computation, one must
    // call the static function TH1::StatOverflows(kTRUE) before filling
    // the histogram.
    // Note that this function returns the Standard Deviation (Sigma)
    // of the distribution (not RMS).
    // The Sigma estimate is computed as Sqrt((1/N)*(Sum(x_i-x_mean)^2))
    // The name "RMS" was introduced many years ago (Hbook/PAW times).
    // We kept the name for continuity.

    if (axis<1 || (axis>3 && axis<11) || axis>13) return 0;

    Double_t x, rms2, stats[kNstat];
    for (Int_t i=4; i<kNstat; i++) stats[i] = 0;
    GetStats(stats);
    if (stats[0] == 0) return 0;
    Int_t ax[3] = {2,4,7};
    Int_t axm = ax[axis%10 - 1];
    x = stats[axm]/stats[0];
    rms2 = TMath::Abs(stats[axm+1]/stats[0] -x*x);
    if (axis<10)
        return TMath::Sqrt(rms2);
    else {
        // The right formula for RMS error depends on 4th momentum (see Kendall-Stuart Vol 1 pag 243)
        // formula valid for only gaussian distribution ( 4-th momentum = 3 * sigma^4 )
        Double_t neff = GetEffectiveEntries();
        return ( neff > 0 ? TMath::Sqrt(rms2/(2*neff)) : 0. );
    }
}
```

二項分布

例1) 全住民の5%がある感染症に罹患しており、その中から無作為に500人を抽出する。ただし住民は500人よりずっと多いとする。このとき、抽出された集団の中に罹患者が30人以上いる確率はどれくらいか。抽出された集団の中に含まれる罹患者数を確率変数 X で表すとき、 X は $n = 500$ 、 $p = 0.05$ の二項分布に近似的に従う。

例2) 20GeVのelectron がEF_e24というtriggerを鳴らす確率が30%だったとする。1000event解析したとき、このときトリガーされた事象数 X は、 $n=1000$ $p=0.3$ の二項分布に従う。

$$P[X = k] = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for } k = 0, 1, 2, \dots, n \quad \text{where} \quad \binom{n}{k} = {}_n C_k = \frac{n!}{k!(n-k)!}$$

- 平均(期待値)

$$E[X] = np$$

- 分散

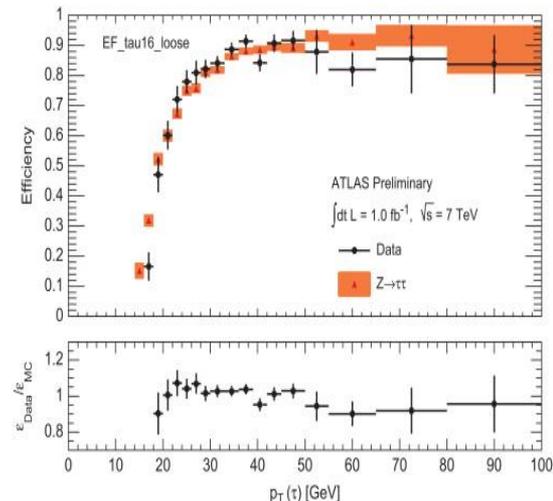
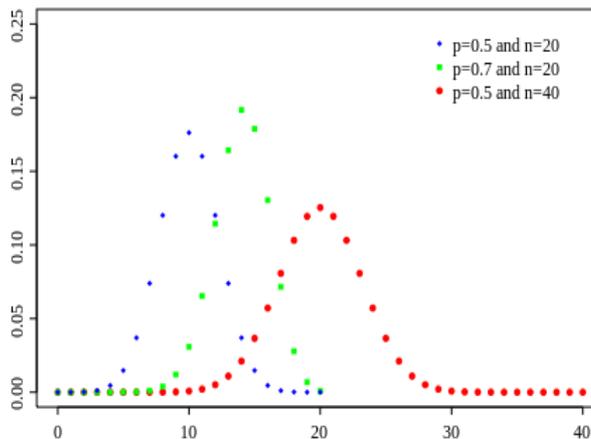
$$\text{Var}(X) = np(1-p)$$

- 正規分布近似

np および $np(1-p)$ が十分大きいとき (>5)

合意確率は、平均 p 標準偏差 $\sqrt{p(1-p)/n}$ の正規分布に近似できる。

- 二項分布は、 $n \rightarrow \infty$ かつ $p \rightarrow 0$ の極限でポアソン分布に近似できる。 ($p = \lambda/n$)



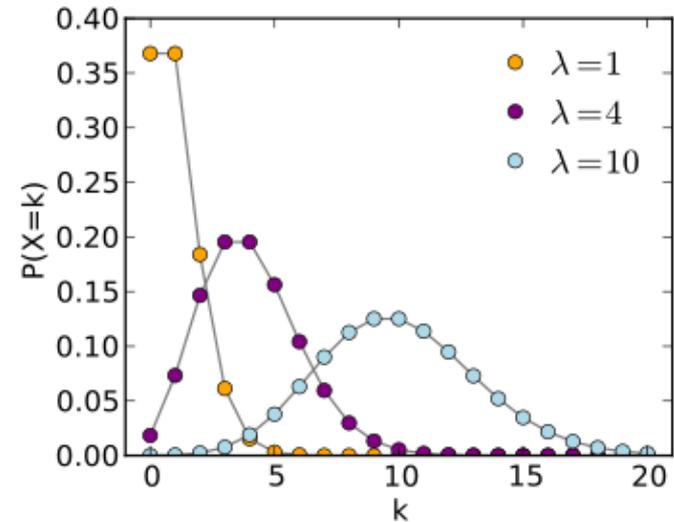
Poisson 分布

- (真の頻度がある)稀な事象が起こる回数
- 例1)、ある村で一日に起こる交通事故の数
- 例2)、LHCで一日につくるヒッグス粒子の数
- ポアソン分布

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

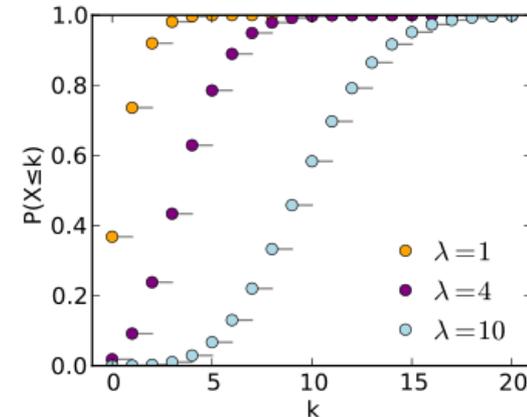
- 平均 $E[X] = \lambda$

- 分散 $V[X] = \lambda$ つまり標準偏差: $\sigma = \sqrt{\lambda}$



累積分布関数

cumulative distribution function



Gauss 分布

- (真の値があるとき)十分統計がある場合
分布はGauss分布に従う

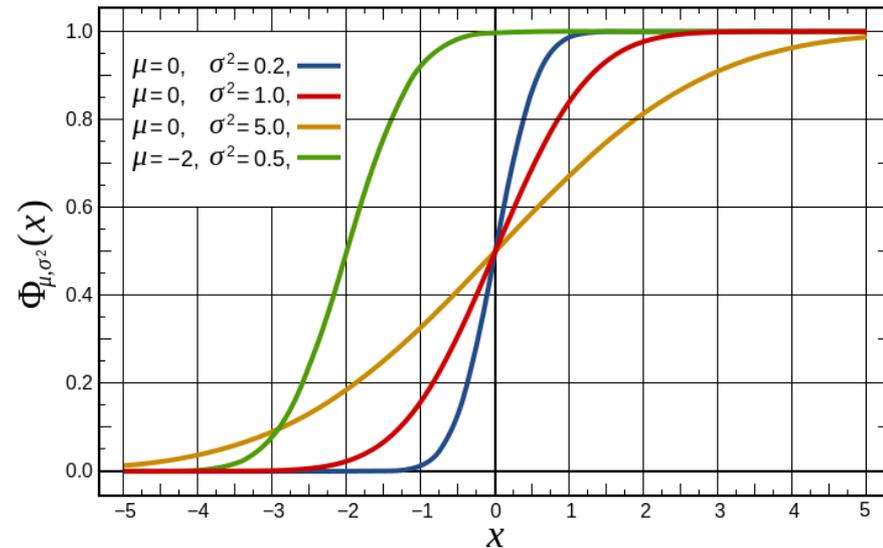
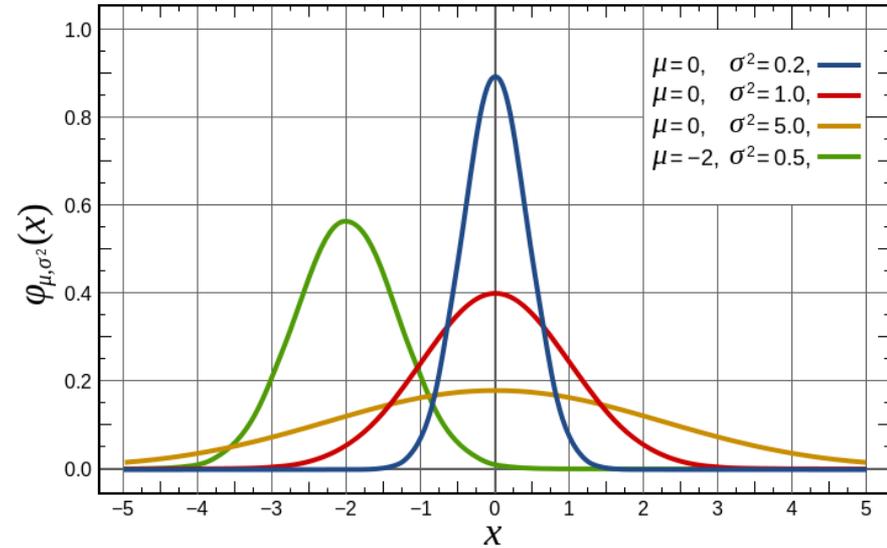
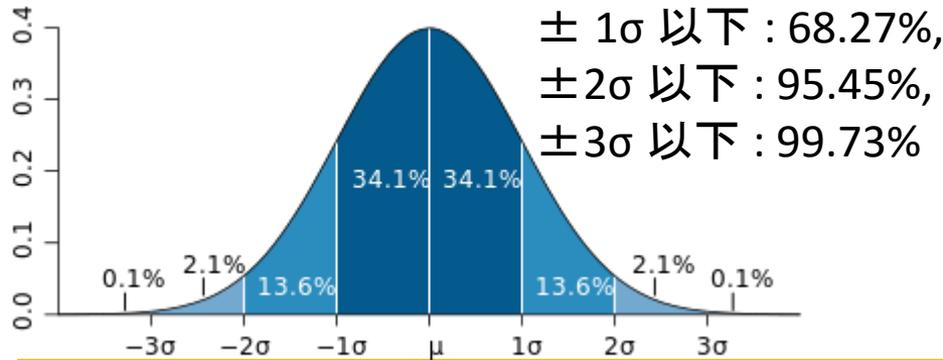
例) 半導体検出器をMIP通過した粒子が作る
電子-ホールペアの数 (300 μ mで2万対)
ガウス分布(正規分布)

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

平均 $E[X] = \mu$

分散 $V[X] = \sigma^2$

発見をクレームするときの 5σ とかいう数字は、
正規分布を仮定した標準偏差で定義されている



応用 (Hypothesis Test)

Statistical Test

- Maximum likelihood(最尤法)

- ある観測量 x_i が、(仮説の)確率密度関数 $f(x_i; \theta)$ (ここで、 $\theta = (\theta_1, \theta_2, \theta_3 \dots \theta_n)$ はLikelihood estimator) に従うとき、N個の観測値 $x_1, x_2, x_3 \dots x_N$ がどのくらいもっともらしく仮説を説明するか?を表すLikelihoodは、

$$L(\theta) = \prod_{i=1}^N f(x_i; \theta)$$

- Least squares(最小二乗法)

- ある観測量 y_i が、ガウス分布(平均 $F(y_i; \theta)$ 分散 σ_i^2)に従うとすると、もっともらしさを表す χ^2 は、(小さい方がよい一致を示す。)

$$\chi^2(\theta) \left(= -2 \ln L(\theta) + \text{constant} \right) = \sum_{i=1}^N \frac{(y_i - F(x_i; \theta))^2}{\sigma_i^2}$$

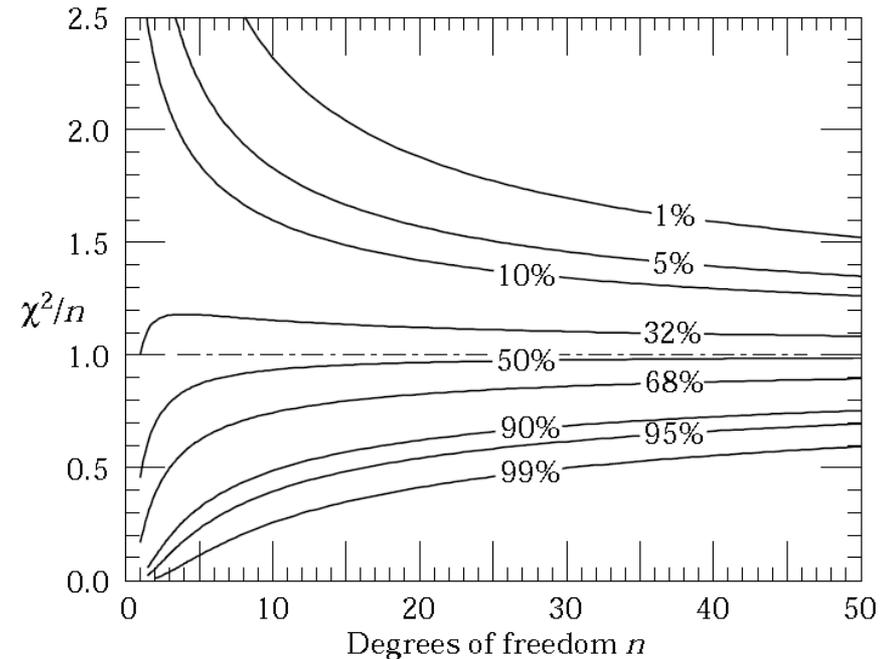
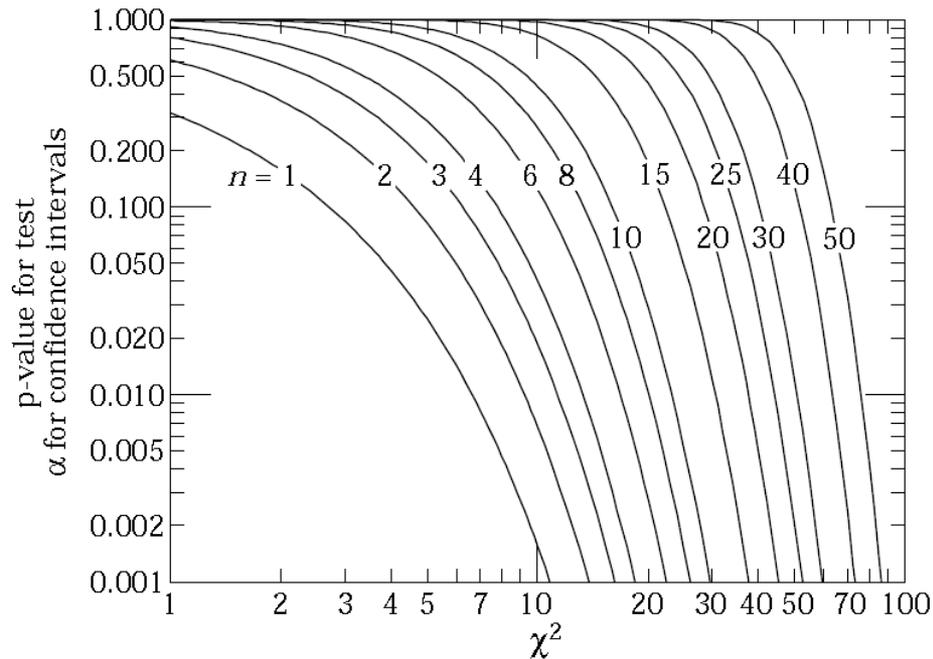
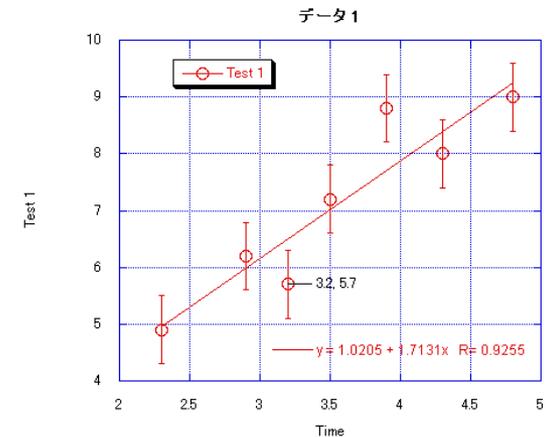
この値は基本的に自由度(θ の数)による。

χ^2 分布とフィットの整合性

- ある分布をフィットするとき以下のよう
に定義される χ^2 を最小にする。

$$\chi^2(\theta) = \sum_{i=1}^N \frac{(y_i - F(x_i; \theta))^2}{\sigma_i^2}$$

- フィットの整合性



Hypothesis Test

- 詳しくは後程。さっきの例
 - 例) Hypothesis : 今月初めから光速度が3%増した。を検証する。
 - 精度良い測定: 2.99792×10^8 m/s (本当はこれは実測値なので誤差がある) : Null Hypothesis
 - 3%増したのなら、 3.08754×10^8 m/sになってるはず。 : Test Hypothesis
- このNull Hypothesis とTest Hypothesis の確率密度関数の比($\lambda(x)$)を定義、(Test statistics)

$$\lambda(x) = \frac{f(x|H_0)}{f(x|H_1)}$$

Confidence levelと発見/棄却

- 実験の目的は最初に話した通り、
 - ある物理量を測定する
 - 特定の理論/仮説(Hypothesis)を検証/棄却する
- 検証や棄却は通常
 - **棄却**:存在する確率が5%以下の確率 →95% CL
 - **兆候**:存在する確率が 3σ (ガウス分布の確率を仮定)
 - **発見**:存在する確率が 5σ (ガウス分布の確率を仮定)
- どうやって出すか、は後程。

近年の高エネ分野における統計

発見するということ

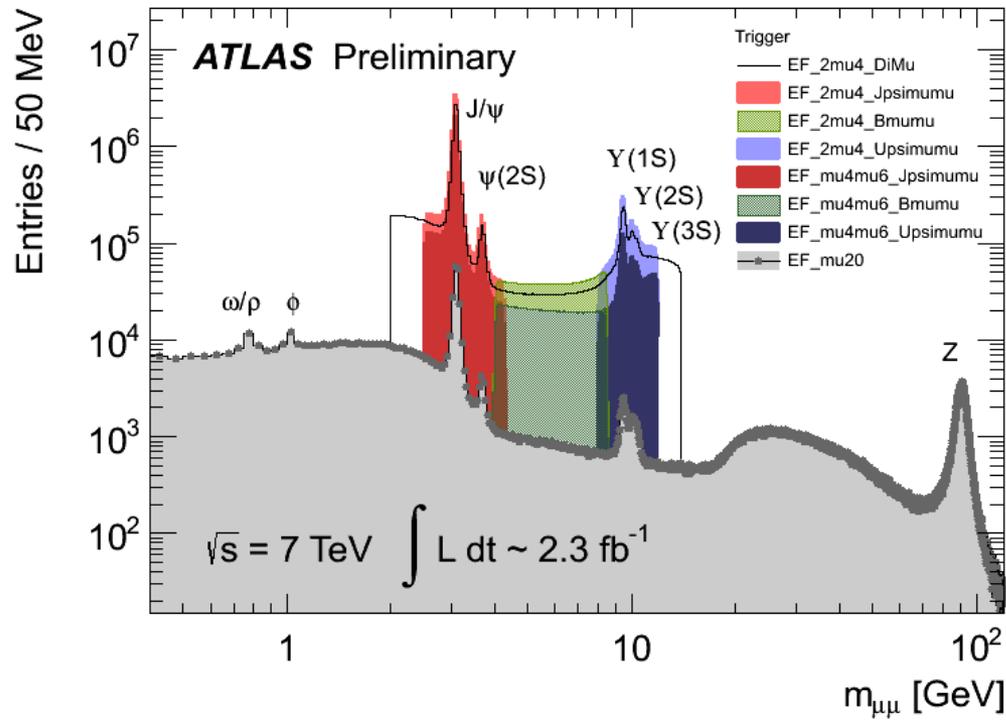
- 何を持って発見とするか。

Wikipedia

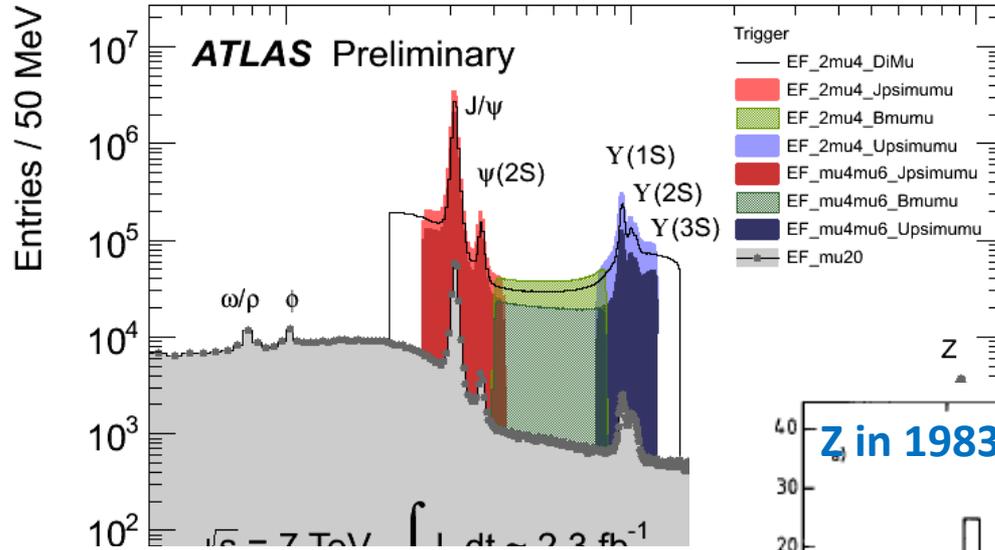
何をもって発見とするかについては、その対象によってやや異なる。事物であれば、問題は比較的簡単で、**それまで誰も見たことのなかったものを見つけることが発見である**。たとえば冥王星の発見などはこれに当たる。

- 素粒子の分野における発見

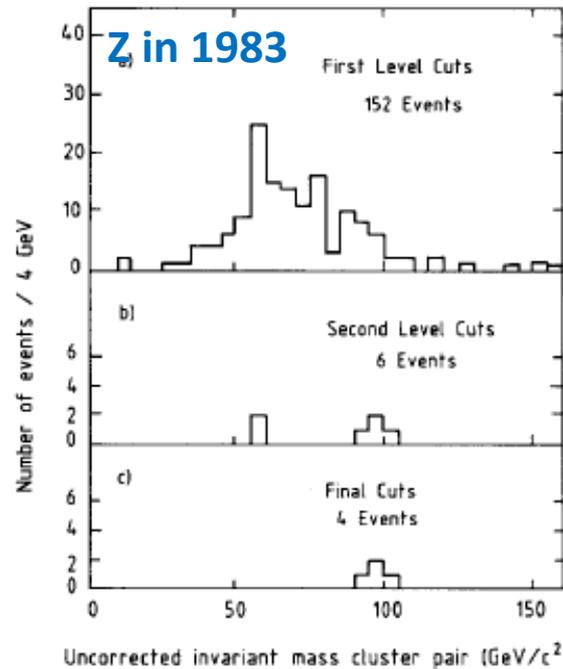
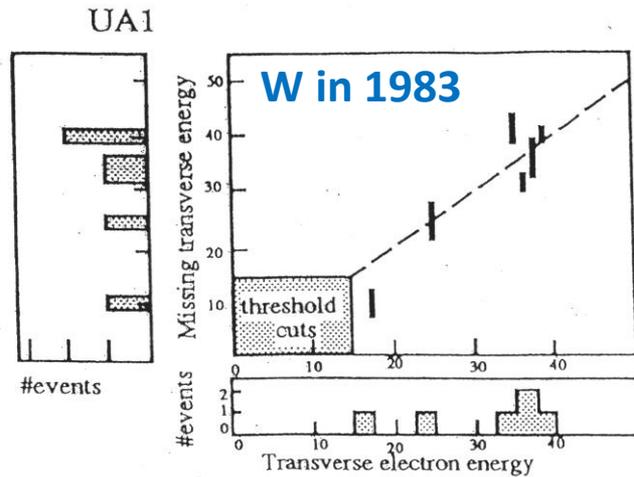
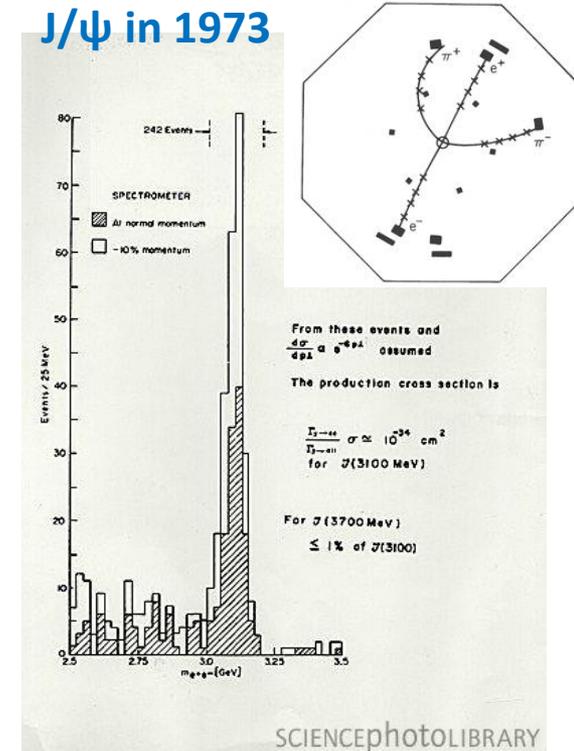
たとえば...



30-40年前の話

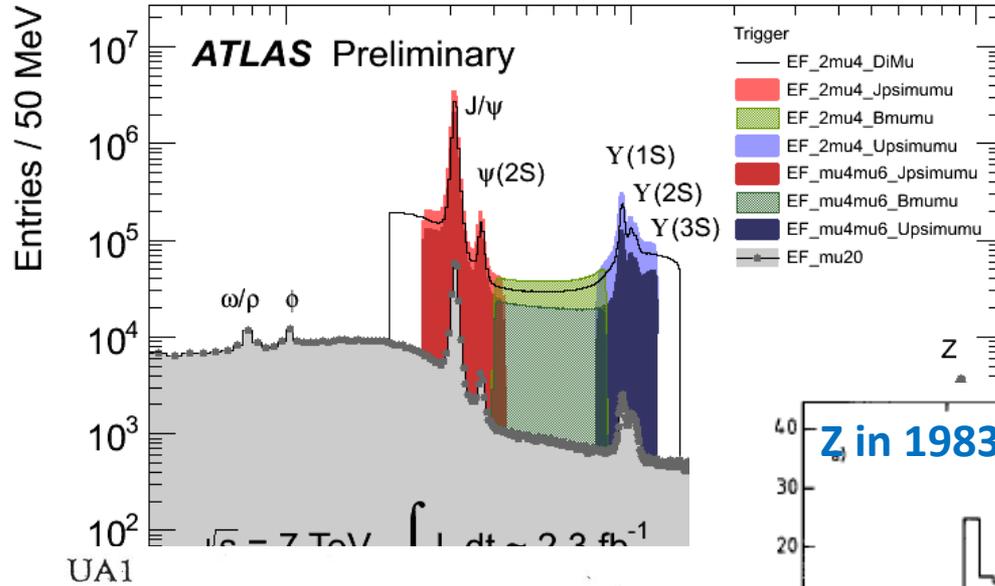


J/ψ in 1973

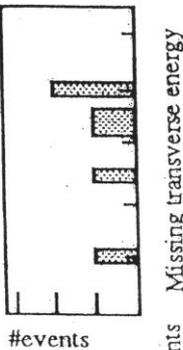
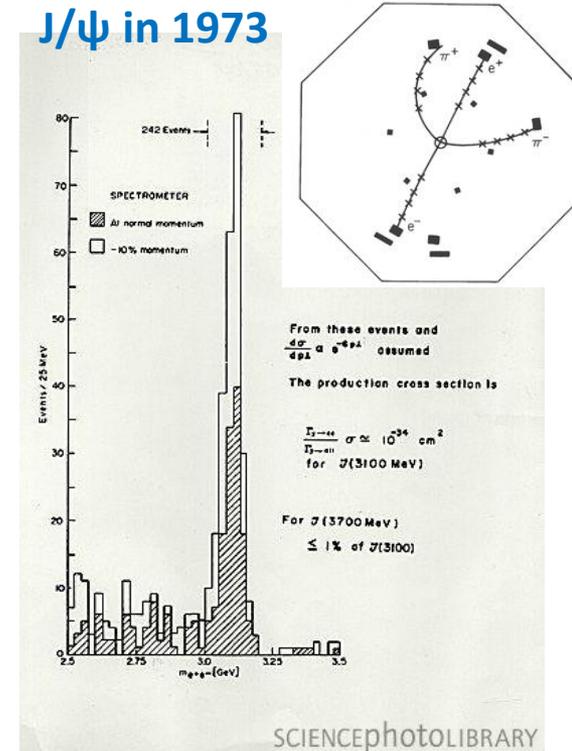


+1 di-muon event

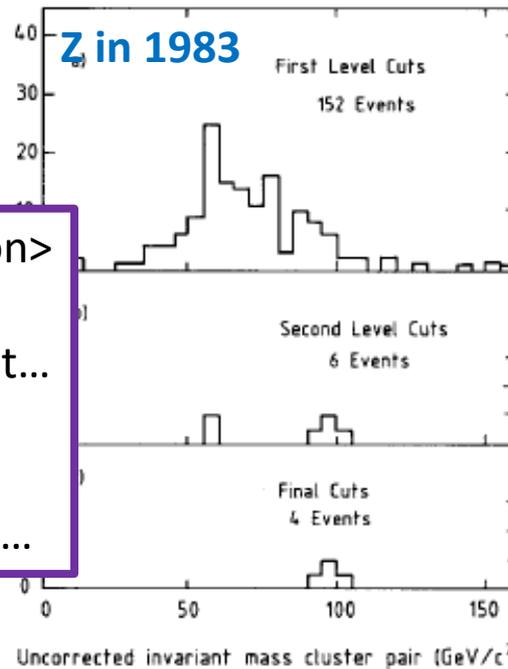
30-40年前の話



J/ψ in 1973

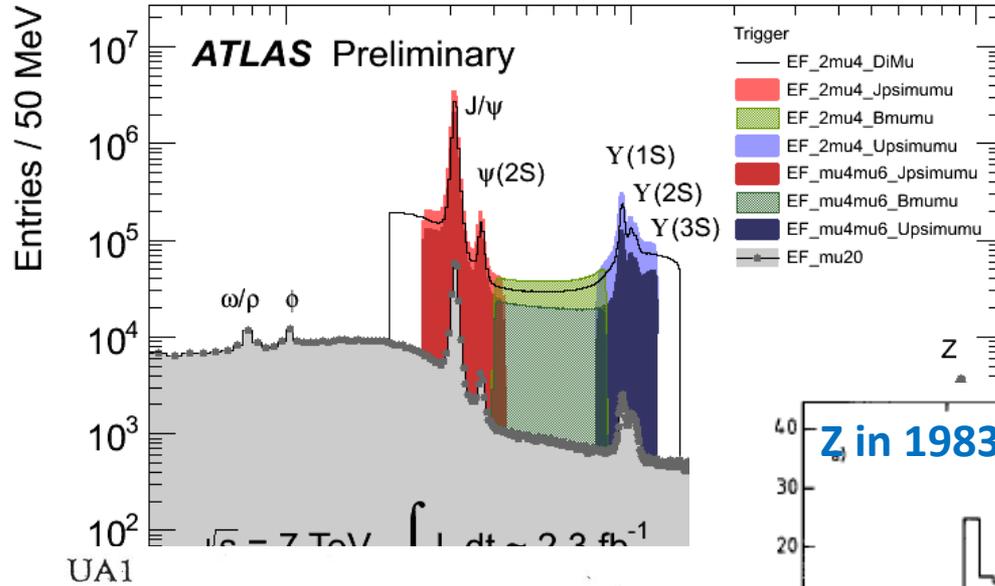


<background of Z observation>
 QCD Di-jet? → Negligible...
 Heavy Flavor Jet → 10⁻⁴event...
 Drell-Yan → Negligible...
 WW → Negligible...
 Onium decay? → Too small ...

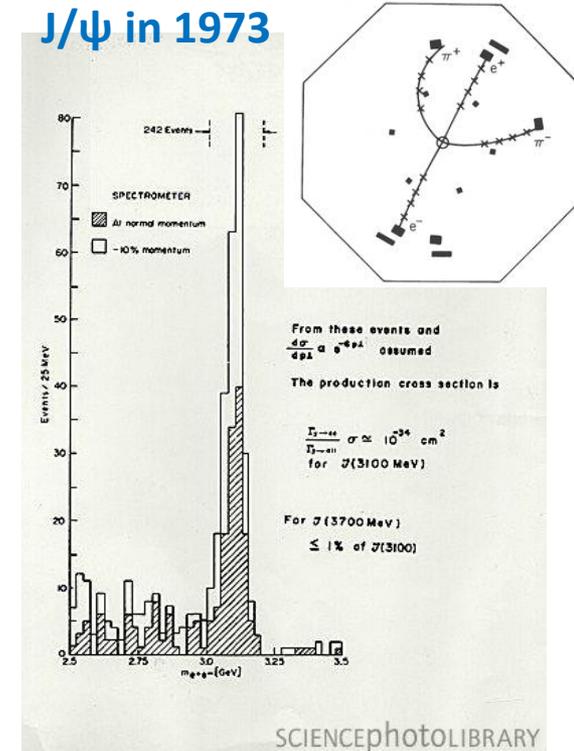


+1 di-muon event

30-40年前の話

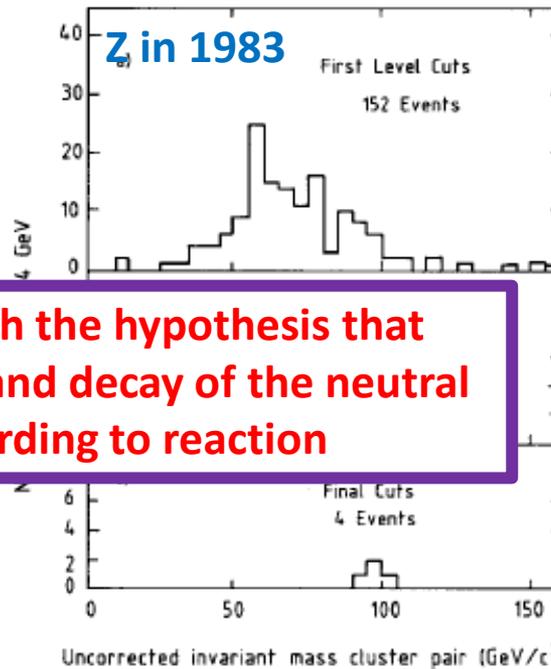


J/ψ in 1973



W in 1983

observations are in agreement with the hypothesis that events are due to the production and decay of the neutral intermediate vector boson Z^0 according to reaction



+1 di-muon event

発見するということ

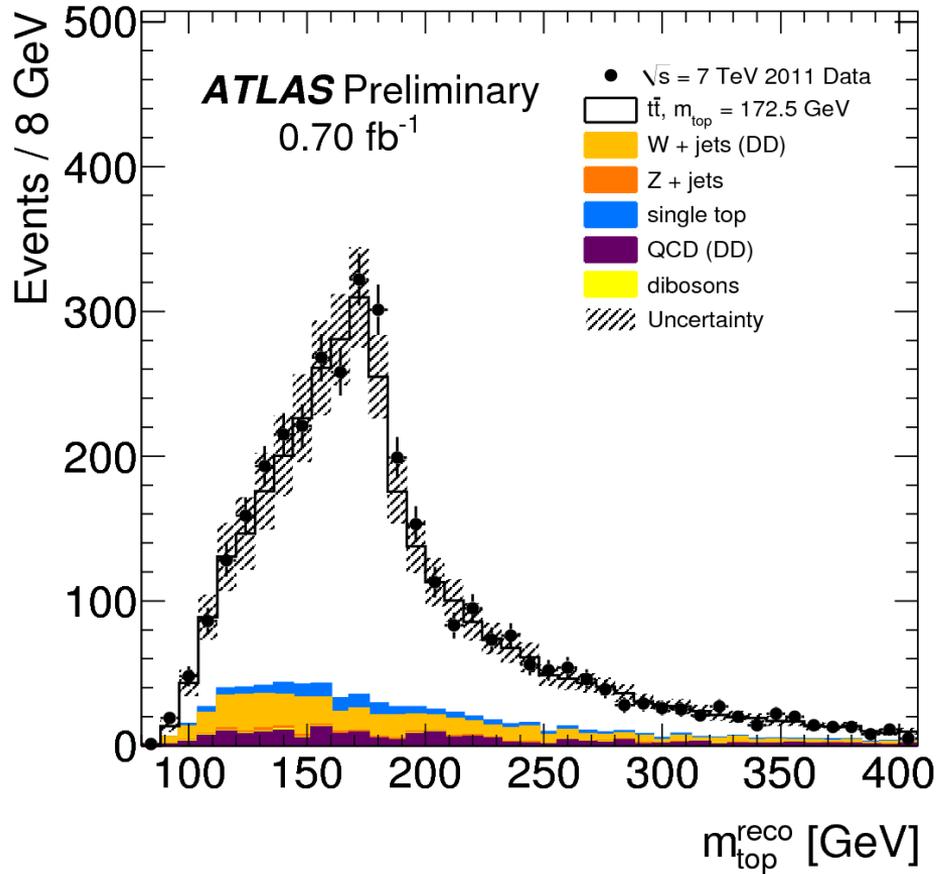
- 何を持って発見とするか。

Wikipedia

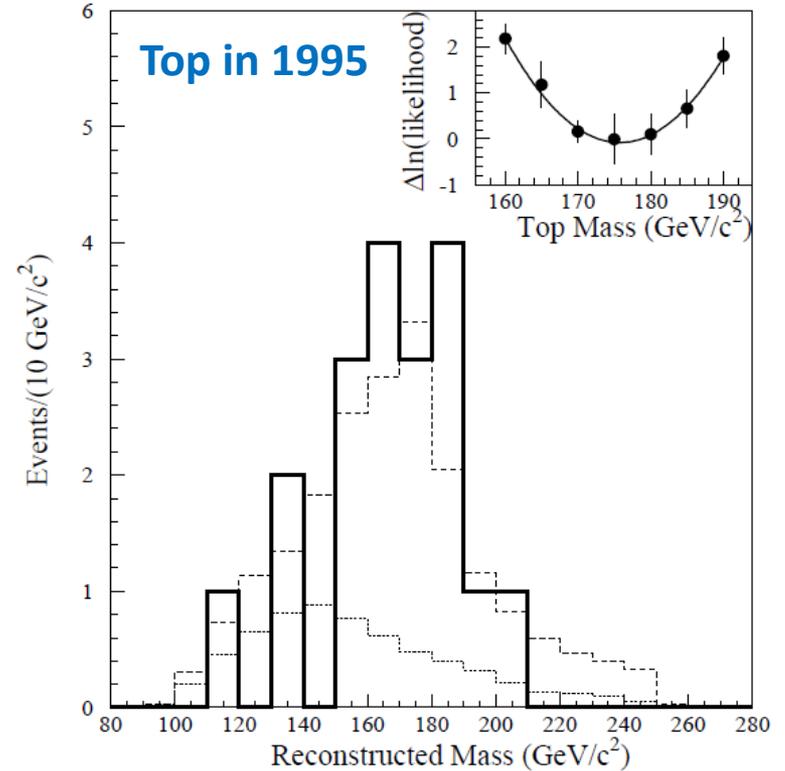
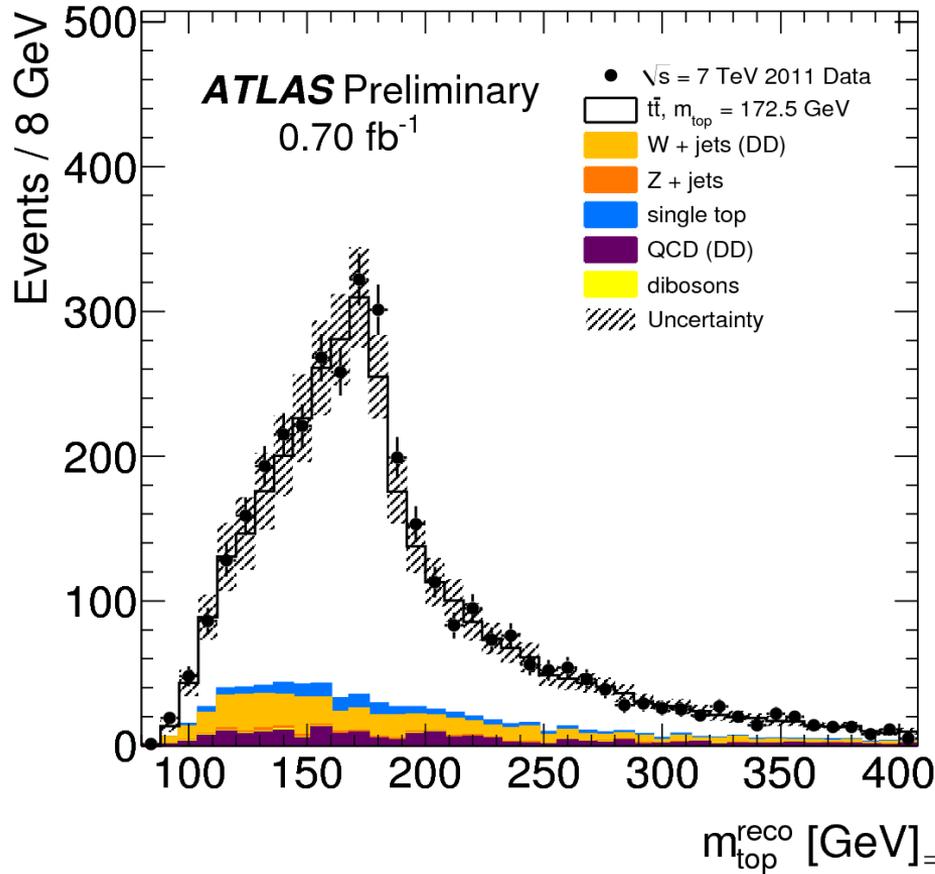
何をもって発見とするかについては、その対象によってやや異なる。事物であれば、問題は比較的簡単で、それまで誰も見たことのなかったものを見つけることが発見である。たとえば冥王星の発見などはこれに当たる。

- 素粒子の分野における発見
 - (30-40年前) 観測事象を十分議論した後の、誰の目にも明らかな過剰事象？

もう少し最近の話



もう少し最近の話



Background probability : 4.8σ
(+ Mass distribution : 5.0σ)

Channel:	SVX	SLT	Dilepton
observed	27 tags	23 tags	6 events
expected background	6.7 ± 2.1	15.4 ± 2.0	1.3 ± 0.3
background probability	2×10^{-5}	6×10^{-2}	3×10^{-3}

発見するということ

- 何を持って発見とするか。

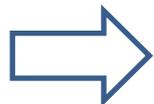
Wikipedia

何をもって発見とするかについては、その対象によってやや異なる。事物であれば、問題は比較的簡単で、それまで誰も見たことのなかったものを見つけることが発見である。たとえば冥王星の発見などはこれに当たる。

- 素粒子の分野における発見

- (30-40年前) 観測事象を十分議論した後の、誰の目にも明らかな過剰事象？
- 観測事象がバックグラウンドで説明できる確率が極めて小さいこと。(正規分布を仮定した 3σ , 5σ)

* 3σ ... 0.2699796%, 5σ ... 0.0000573303%



この“確率”をどうやって定義するかが問題

“確率”の定義

必要なもの

どの観測量を検定するか(test statistics) : N

例、観測事象数、Likelihood ratio....

検定量の確率分布(backgroundのみ) : $P(N|b)$

例、ポアソン分布(統計誤差) x ガウス分布(系統誤差)

実際の観測量: N_{obs}

例、データの観測事象数

確率(p-value)

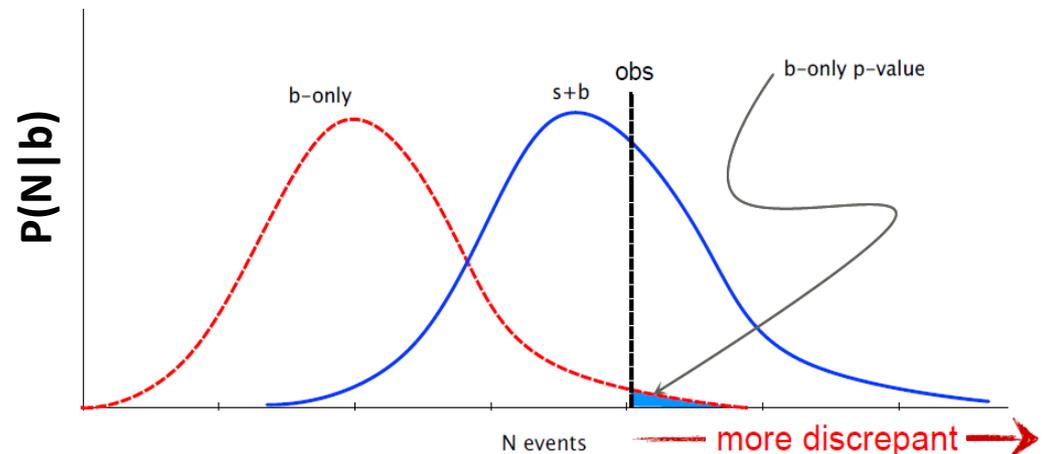
$$\mathbf{p\text{-value} = P(N > N_{obs} | b)}$$

逆に、 $1-p$ を、 CL_b と呼ぶ(発見した確率?)

$$\mathbf{CL_b = 1 - p = P(N < N_{obs} | b)}$$

予想感度(expected sensitivity)

$$\mathbf{p_{exp} = P(N > N_{s+b}^{med} | b)}$$



ある理論に対して制限を与える(limit)

Confidence Level (CL) : どのくらいの信頼度で制限を与えるか

理論が予想する信号のモデルが必要。

信号+背景事象(s+b)に対応する検定量の確率分布 : $P(N|s+b)$

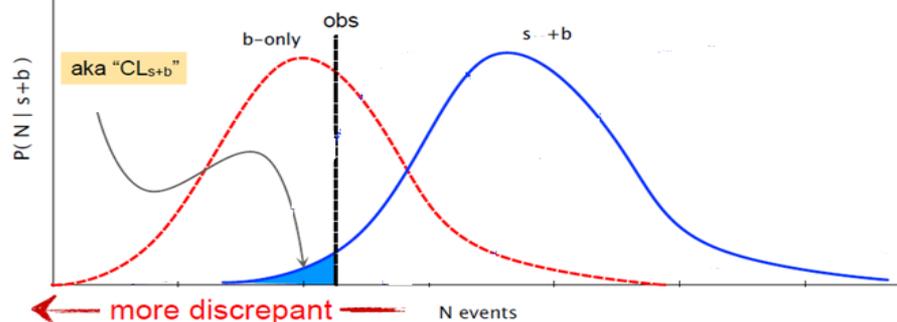
観測量がs+bで説明できる確率(CL_{s+b})

例、 N_{s+b} の確率分布

$$CL_{s+b} = P(N < N_{obs} | s+b)$$

b-onlyに近い: 小さい
s+bに近い: 大きい

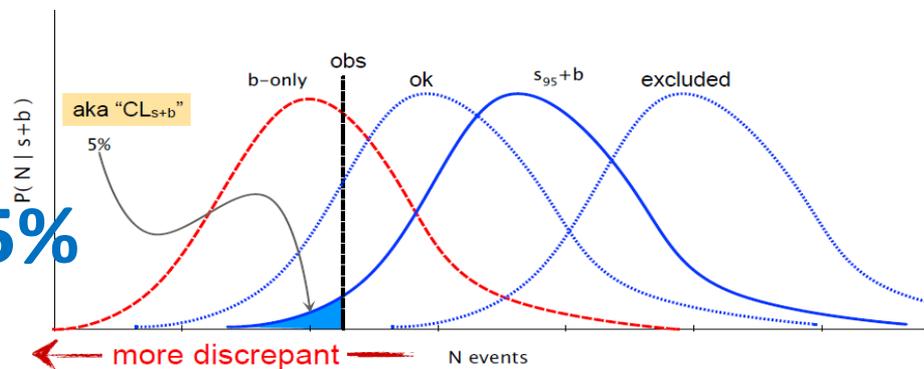
$CL_{s+b} < 5\%$: 95% CL で棄却



信号の強さ : μ 例、理論予想の何倍

$$CL_{s+b}(\mu) = P(N < N_{obs} | \mu s+b) = 5\%$$

となる μ を95% CLで棄却したという。



CL_{s+b} の弊害とPCL, CL_s

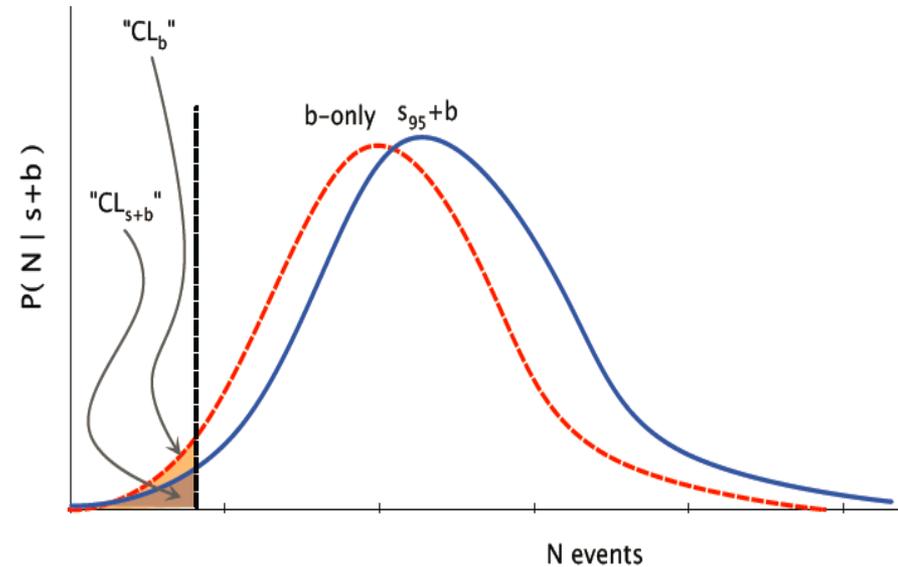
たとえば、感度がなくても...

極端な話...

$CL_b = P(N < N_{obs} | b) < 5\%$ の時、
(5%の確率で起こる)

$CL_{s+b}(0) = CL_b$ なので、

$CL_{s+b}(0) < 5\%$: $\mu=0$ を棄却!



CL_{s+b} の弊害とPCL, CL_s

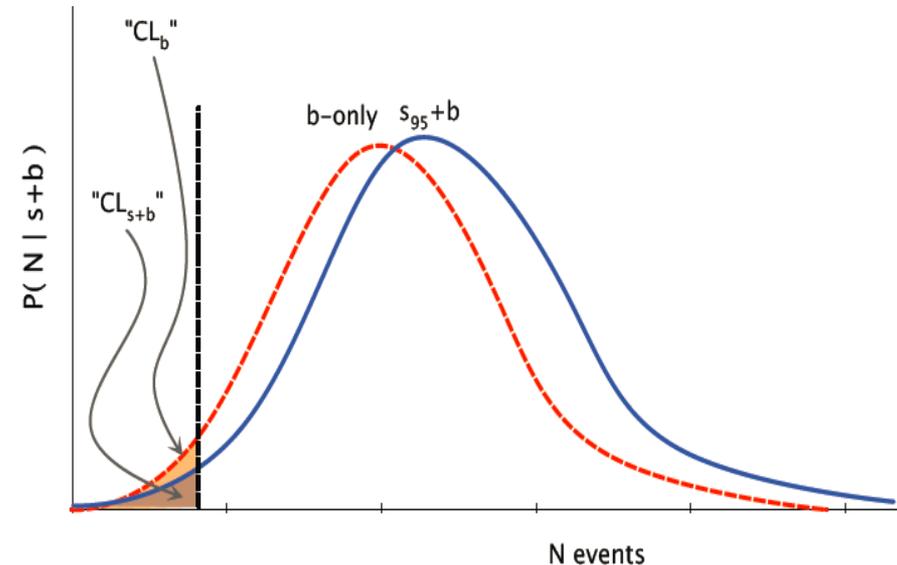
たとえ、感度がなくても...

極端な話...

$CL_b = P(N < N_{obs} | b) < 5\%$ の時、
(5%の確率で起こる)

$CL_{s+b}(0) = CL_b$ なので、

$CL_{s+b}(0) < 5\%$: $\mu=0$ を棄却!



“Can do this without running the experiment” by T. Junk

余談：水切りでHiggs棄却？

- 川で石を投げて水切りをやる。石とHiggs粒子の効果でHiggs粒子がある世界とない世界で多少の差が出るかもしれない。
- 実際はHiggs粒子の効果は非常に小さくて、ある場合とない場合の確率分布にほとんど差はない。
- たまたま水切りがうまく行って、10回はねたとする。

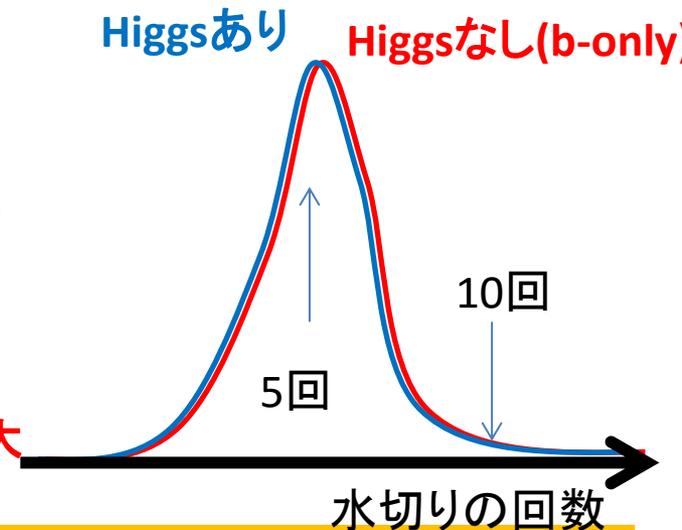
- こういうことが起こる確率はHiggsありのとき3%だったら、

$$CL_{s+b}(1) = P(N > N_{obs} | 1x s + b) = 3\%$$

$CL_{s+b}(1) < 5\%$ なのでHiggs粒子の存在を棄却した。

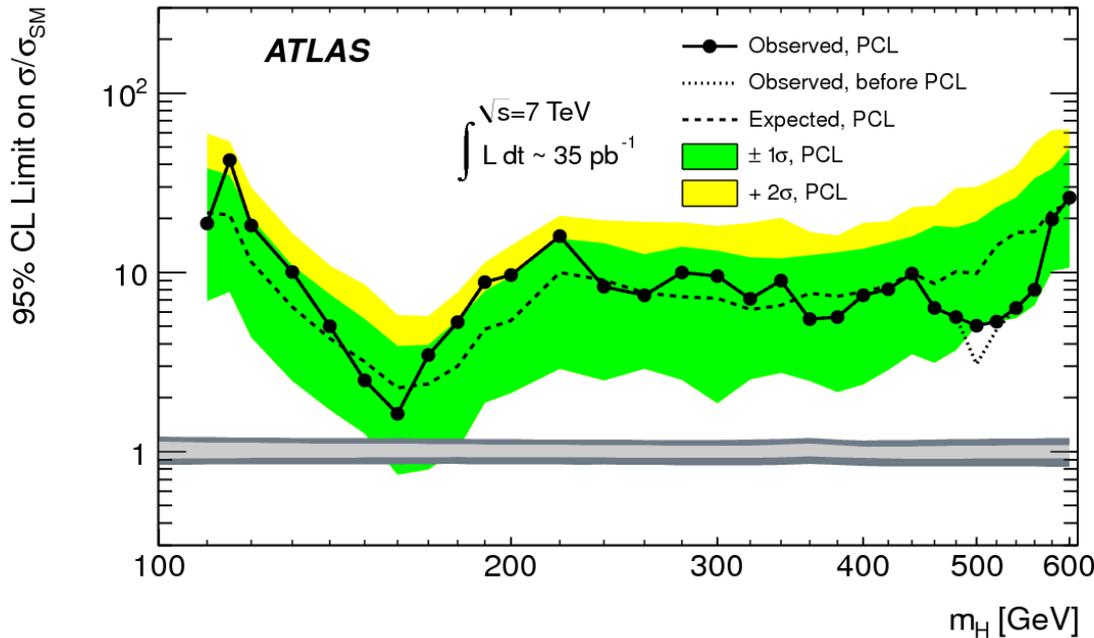
- 95% CL といっているのだから必ずしも間違いではないのかもしれないが... 問題は:
 - そもそも感度の無い実験で結論を出すのがいけない。
 - とはいえ、何をもちて感度があるというかは自明ではない。

- ある程度 CL_b (バックグラウンドで説明できる確率)が大きいことを要求すればよい？

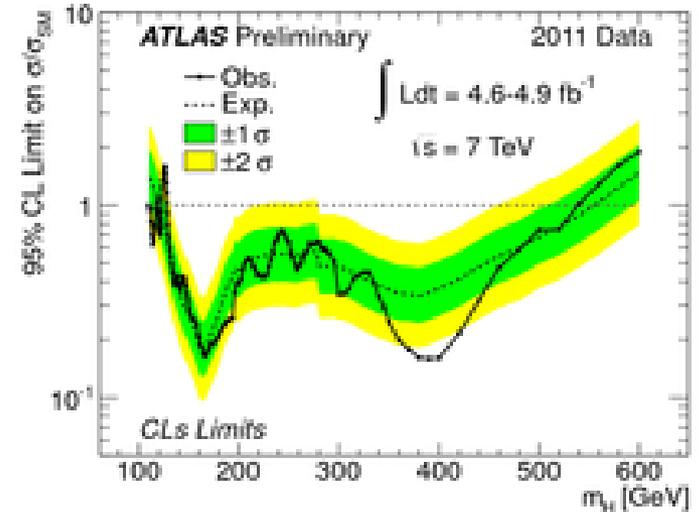


当初のATLASの結果

2011年 Moriond (2010年に取得した35fb⁻¹)



(c.f. ちょっと後の結果)



どこにも書いてませんが、これは CL_{s+b} の確率を用いて求めたlimit

なぜ -2σ のバンドがないのか？

PCLとは？

CL_{s+b}の弊害とPCL, CL_s

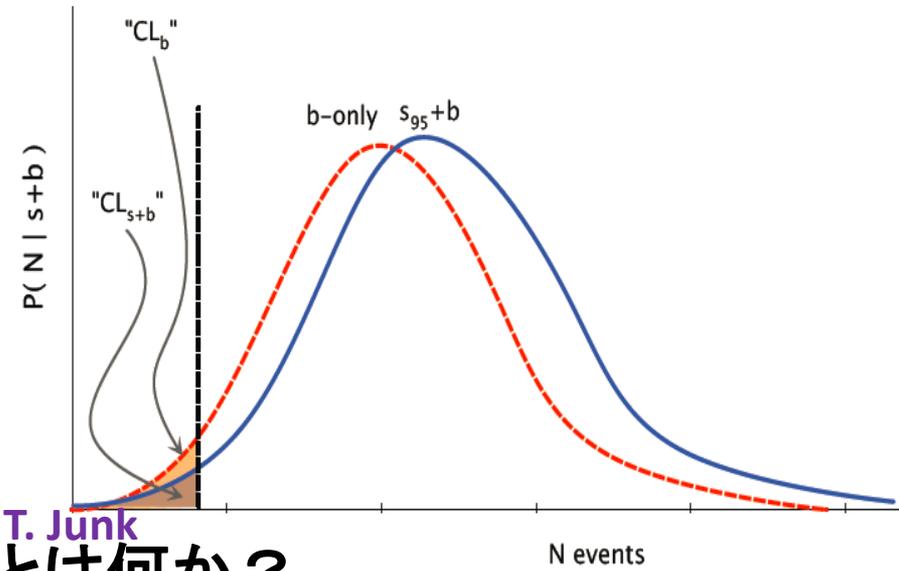
たとえば、感度がなくても...

極端な話...

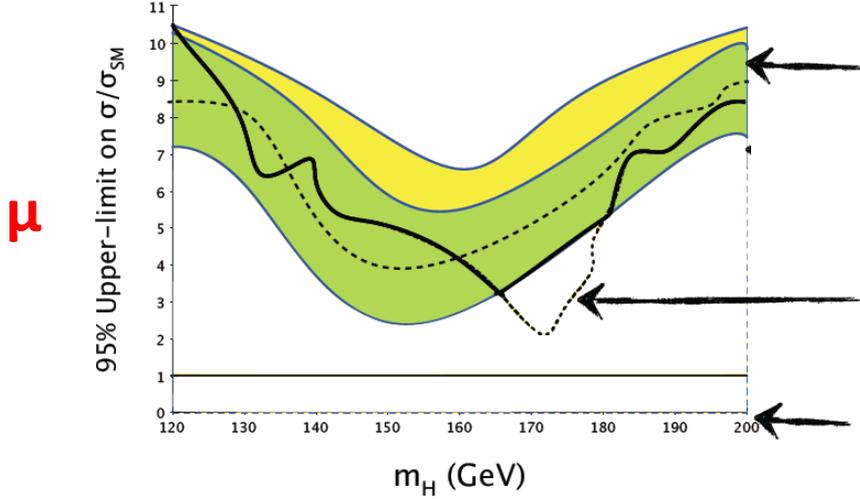
CL_b = P(N < N_{obs} | b) < 5%の時、
(5%の確率で起こる)

CL_{s+b}(0) = CL_bなので、

CL_{s+b}(0) < 5% : μ=0を棄却!



“Can do this without running the experiment” by T. Junk
なぜ -2σのバンドがないのか？ PCLとは何か？



点線CL_{s+b}で求めた予測limit (expected limit)
緑や黄色のバンドは、実験を繰り返した時に
上下に動く範囲(緑 1σ、黄2σ)
実線は実際の観測したlimit (observed limit)

CL_{s+b}の弊害とPCL, CL_s

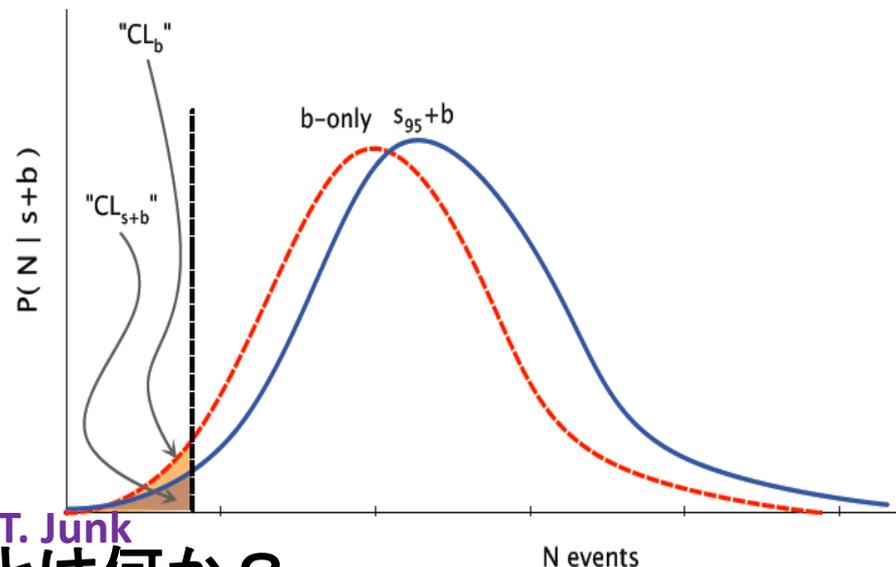
たとえば、感度がなくても...

極端な話...

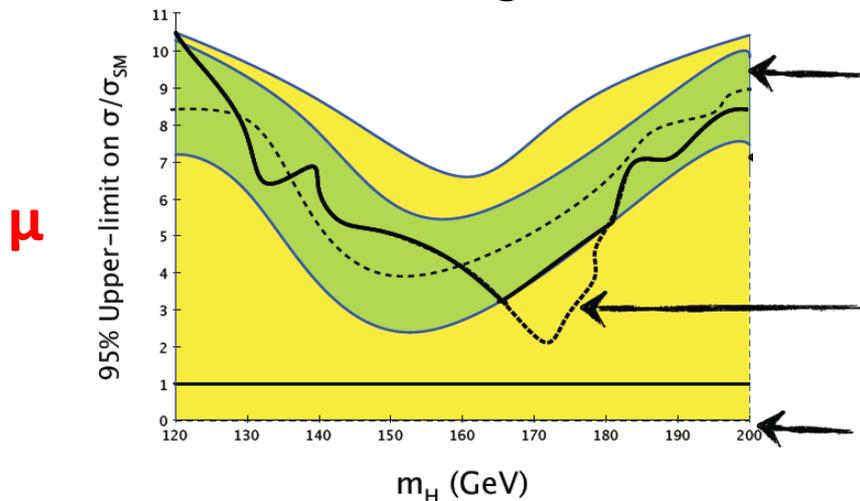
CL_b = P(N < N_{obs} | b) < 5%の時、
(5%の確率で起こる)

CL_{s+b}(0) = CL_bなので、

CL_{s+b}(0) < 5% : **μ=0を棄却!**



“Can do this without running the experiment” by T. Junk
なぜ -2σのバンドがないのか？ PCLとは何か？



点線CL_{s+b}で求めた予測limit (expected limit)
緑や黄色のバンドは、実験を繰り返した時に
上下に動く範囲(緑 1σ、黄2σ)
実線は実際の観測したlimit (observed limit)

μ=0 : -2σは、μ=0を棄却

CL_{s+b}の弊害とPCL, CL_s

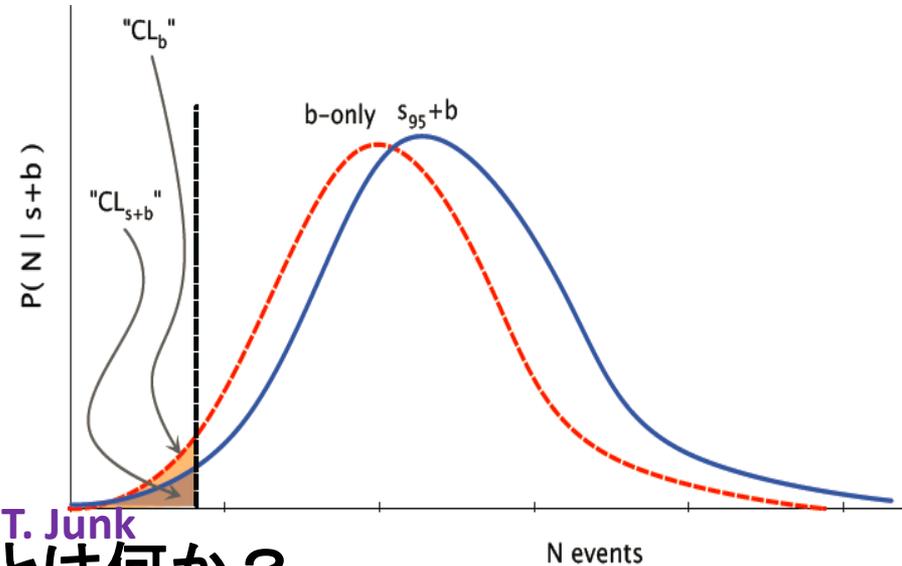
たとえば、感度がなくても...

極端な話...

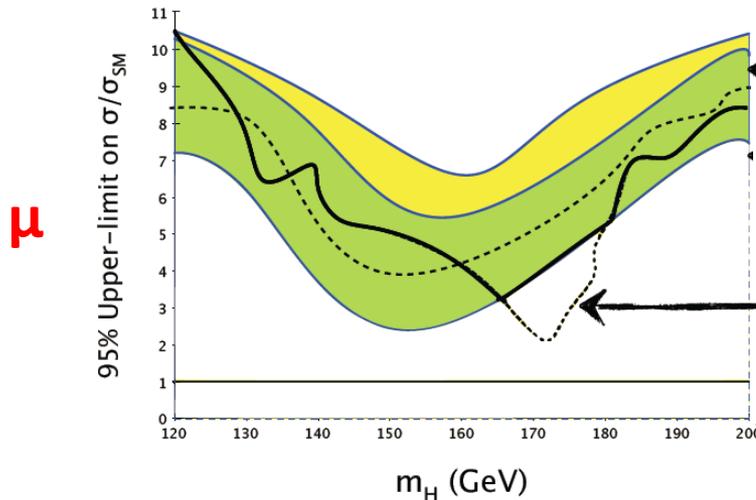
CL_b = P(N < N_{obs} | b) < 5%の時、
(5%の確率で起こる)

CL_{s+b}(0) = CL_bなので、

CL_{s+b}(0) < 5% : μ=0を棄却!



“Can do this without running the experiment” by T. Junk
なぜ -2σのバンドがないのか？ PCLとは何か？



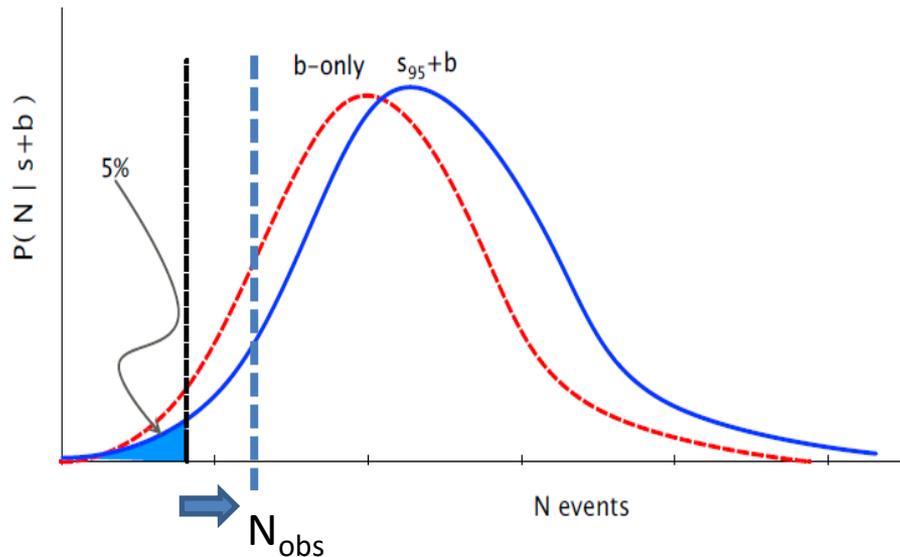
点線 CL_{s+b} で求めた予測limit (expected limit)
緑や黄色のバンドは、実験を繰り返した時に
上下に動く範囲(緑 1σ、黄 2σ)
実線は実際の観測したlimit (observed limit)

“too lucky”を防ぐために-1σで止める
CL_b < 16%の時は棄却しない

-2σの黄色のバンドは書かない!!

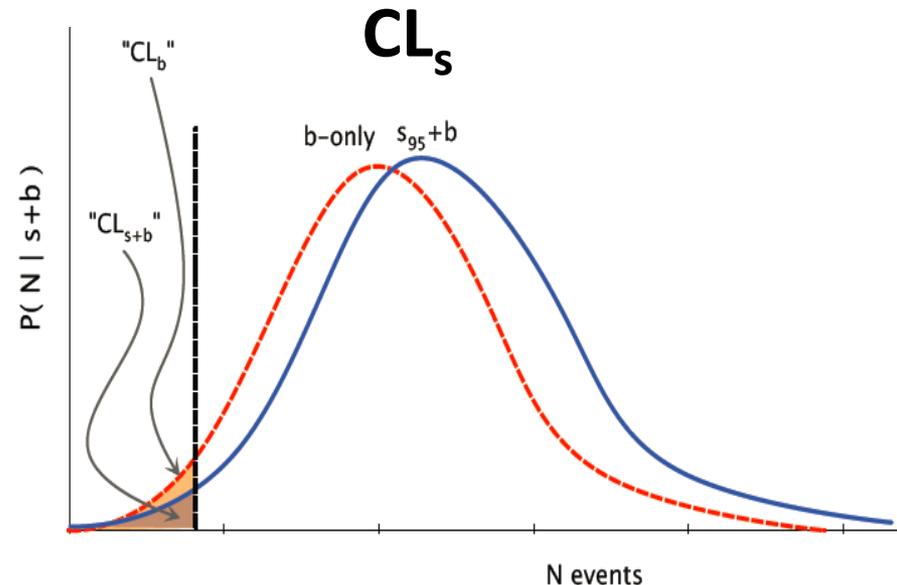
CL_{s+b}の弊害とPCL, CL_s

- PCLともう一つの解決策
Power Constraint Limit (PCL)



b-only の -1σ より観測量が低かったとき、
observed limit はexpected limit の -1σ で止める。

つまり、 N_{obs} が $N_{exp}^{-1\sigma}$ だったとする。



$$CL_s = CL_{s+b} / CL_b$$

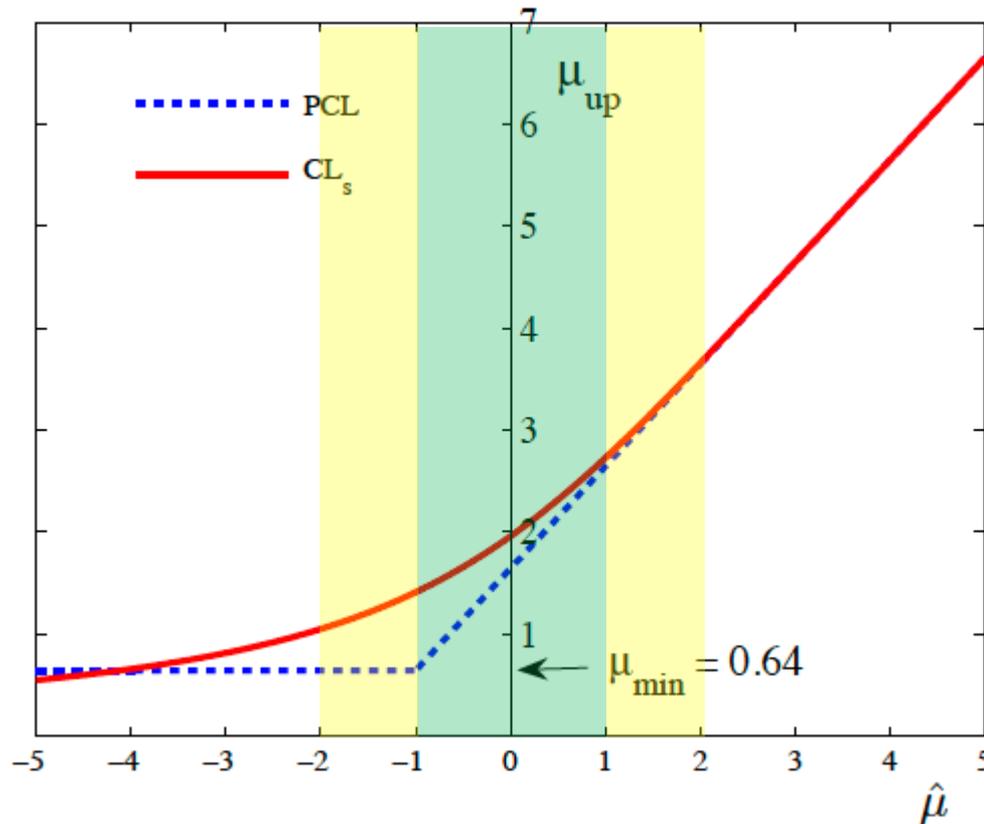
もはや確率ではないが...

CLs < 5% を棄却とするとCLs+bのような
あやまちはない。

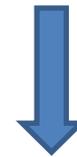
Note : b-onlyの中心が観測されたとき、
CLs = 2 * CLs+b : 95% CL は97.5% CLに相当

CL_{s+b}の弊害とPCL,CL_s

CL_{s+b}&PCL とCL_sの比較



確率としての正しさ
v.s
Conservative



CL_sが今のATLASの
recommendation

+1 σ より下に観測した場合CL_{s+b}&PCLの方が楽観的

再び“確率”の定義

必要なもの

どの観測量を検定するか(test statistics) : N

例、観測事象数、Likelihood ratio.....

- Test statistics をもう少し賢い量にしてみる。
 - 観測量(observable)は、
 - N_s, N_b, μ , 系統誤差(Nuisance Parameter)

2000年ごろ LEP の解析で発展 → Tevatron → LHC

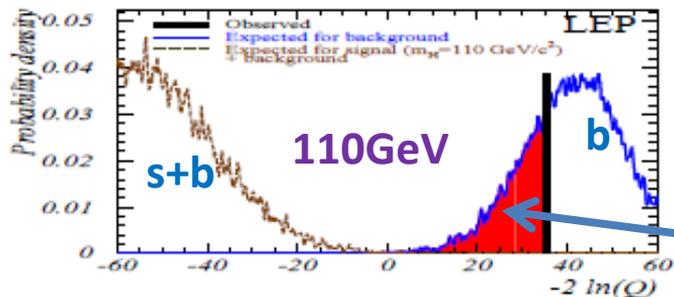
単純な観測事象数

→ LEP : Likelihood ratio

→ Tevatron : ratio of profile likelihood

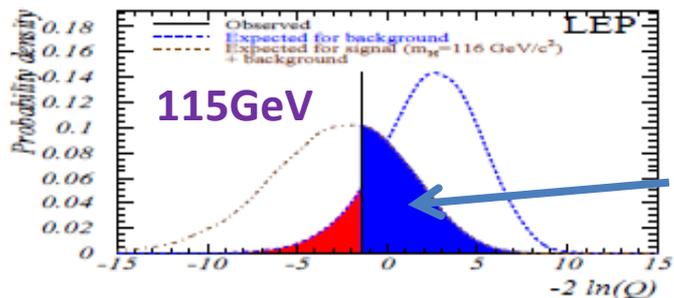
→ LHC : profile likelihood ratio

Test statistics : LEP Higgs

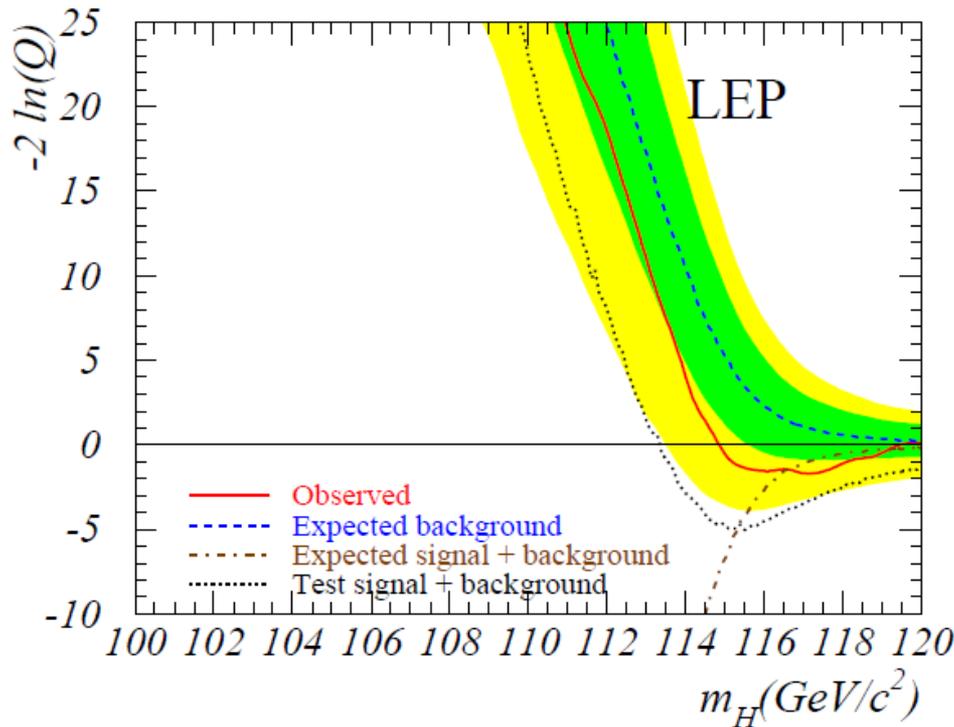
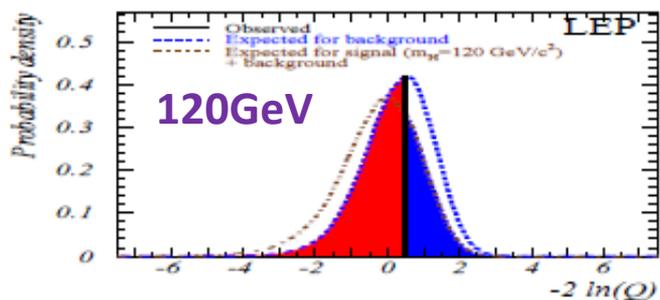


$$Q = \frac{\mathcal{L}(\text{data}|s + b)}{\mathcal{L}(\text{data}|b)} = \frac{e^{-(S_{\text{tot}}+B_{\text{tot}})}}{e^{-B_{\text{tot}}}} \prod_{i=1}^{N_{\text{bins}}} \frac{(s_i + b_i)^{N_i}}{b_i^{N_i}}$$

p-value (1-CL_b)

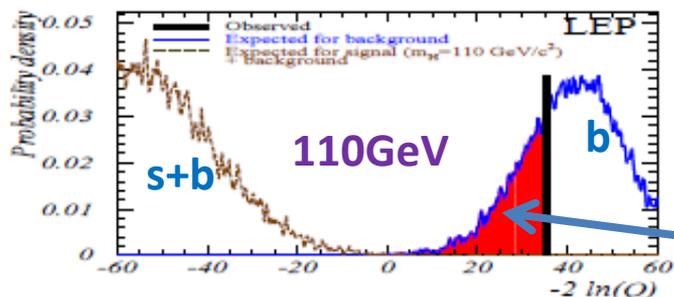


CL_{s+b}



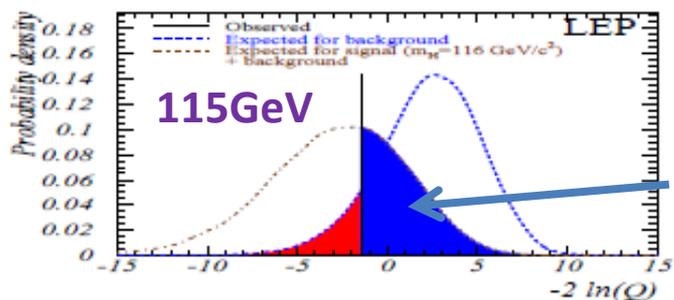
s+b like -2 ln(Q) b like

Test statistics : LEP Higgs

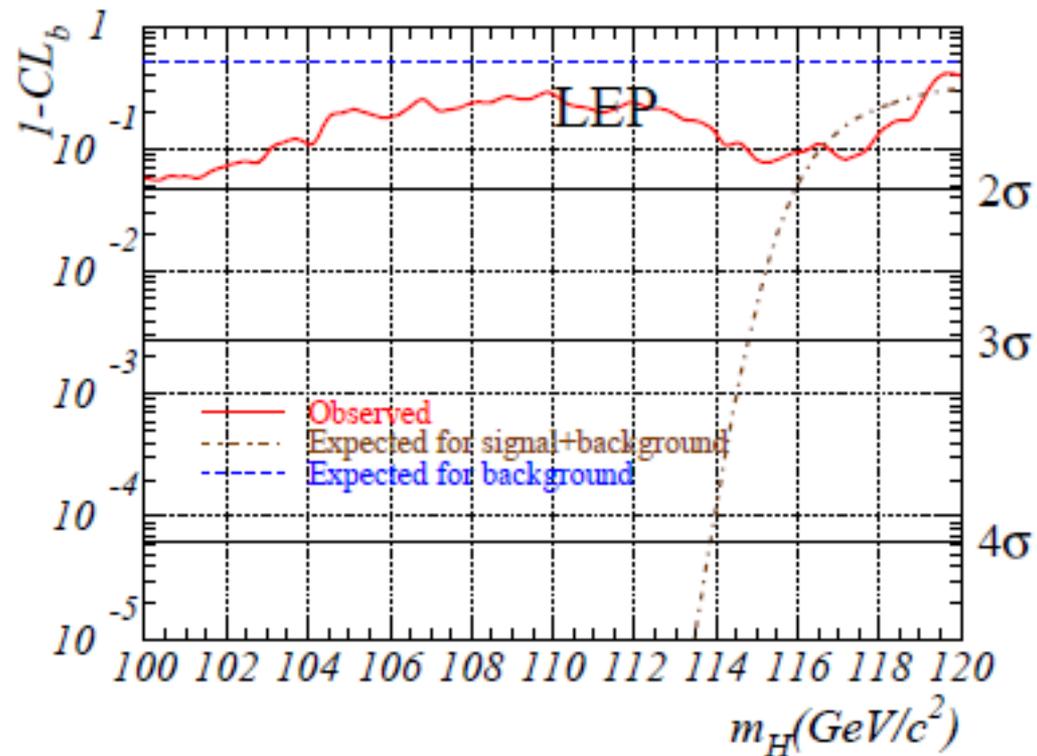
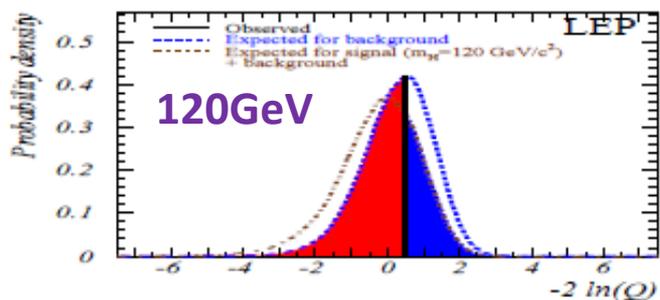


$$Q = \frac{\mathcal{L}(\text{data}|s + b)}{\mathcal{L}(\text{data}|b)} = \frac{e^{-(S_{tot} + B_{tot})}}{e^{-B_{tot}}} \prod_{i=1}^{N_{bins}} \frac{(s_i + b_i)^{N_i}}{b_i^{N_i}}$$

p-value (1-CL_b)



CL_{s+b}

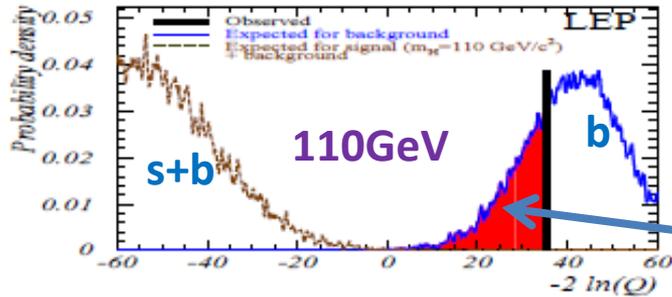


s+b like

-2 ln(Q)

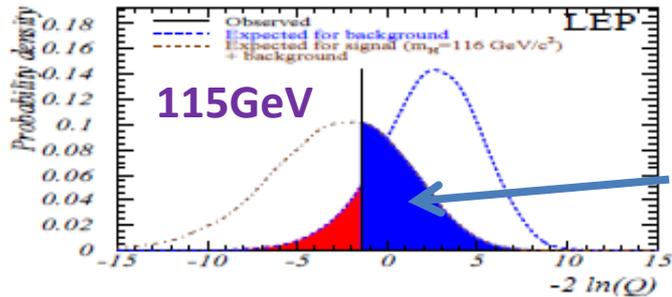
b like

Test statistics : LEP Higgs

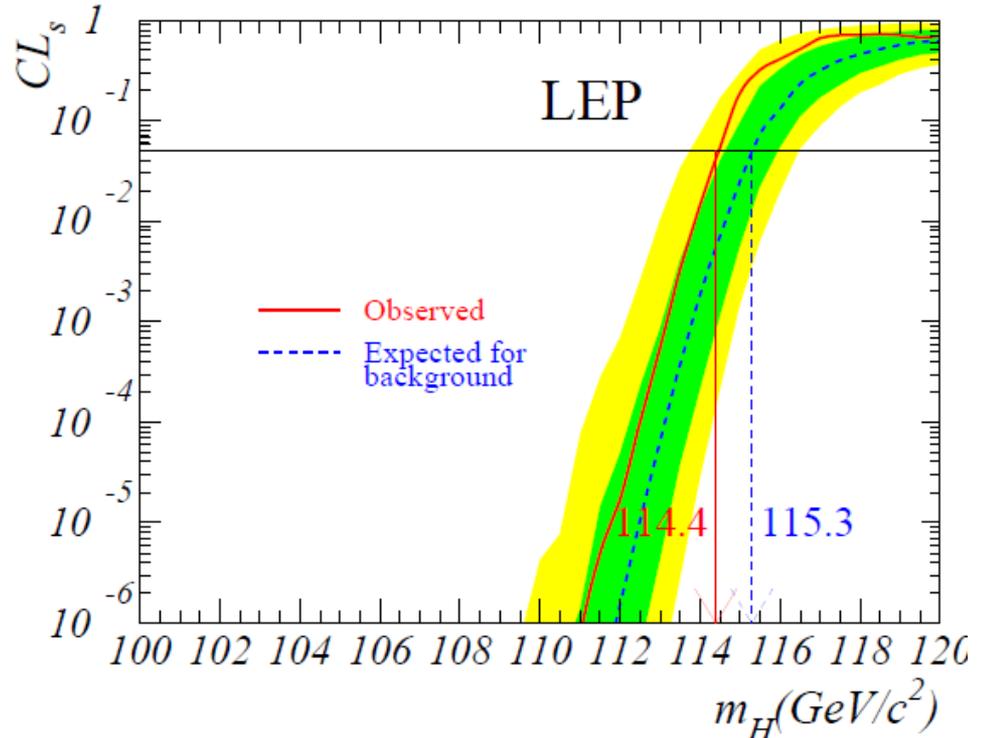
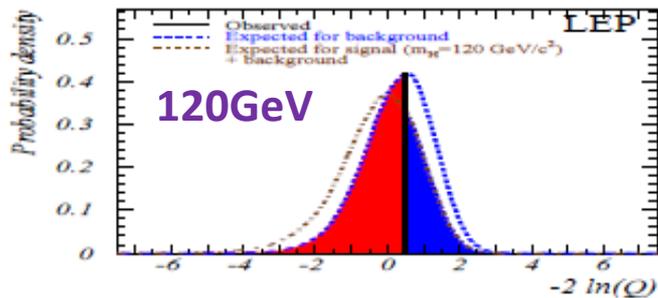


$$Q = \frac{\mathcal{L}(\text{data}|s + b)}{\mathcal{L}(\text{data}|b)} = \frac{e^{-(S_{tot} + B_{tot})}}{e^{-B_{tot}}} \prod_{i=1}^{N_{bins}} \frac{(s_i + b_i)^{N_i}}{b_i^{N_i}}$$

p-value (1-CL_b)



CL_{s+b}



s+b like -2 ln(Q) b like

Test statistics : Tevatron SingleTop

複雑なTest Statistics を使って発見をクレームした例 : Single-Top @ Tevatron

CDF Run II Preliminary, L = 3.2 fb⁻¹

- Single Top
- W+HF
- t \bar{t}
- QCD+Mistag
- Other
- Data

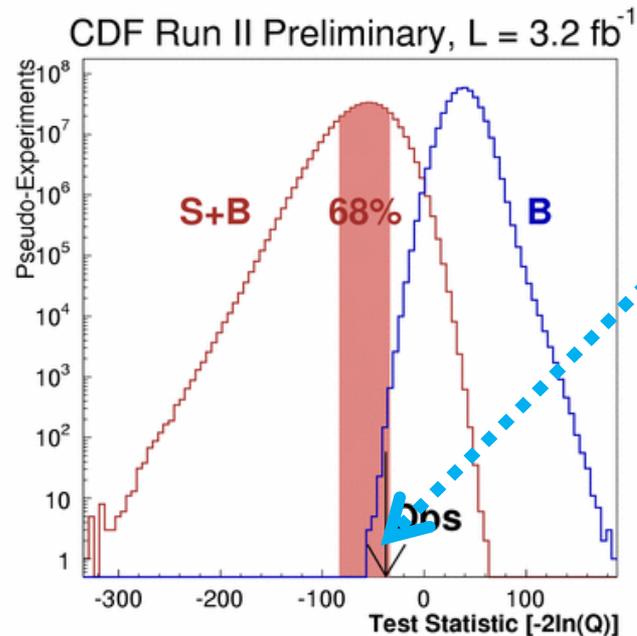
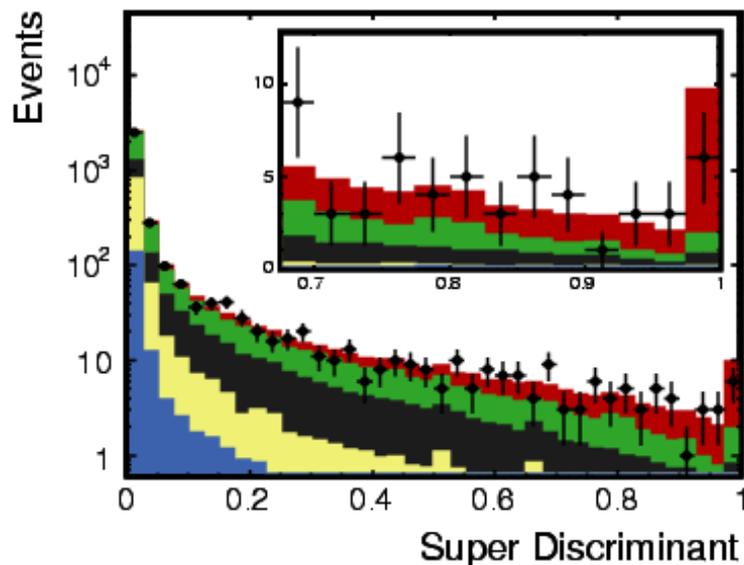
$$-2 \ln Q = -2 \ln \frac{p(\text{data} | H_1, \hat{\theta})}{p(\text{data} | H_0, \hat{\theta})}$$

H0 : b-only (null hypothesis)

H1 : s+b (test hypothesis)

θ : Nuisance parameter

このように系統誤差をNuisance parameterとして
多変数解析をすることを **Profiling** という。→ profile likelihood



p-value
 $3.1 \times 10^{-7} = 5\sigma$

Expected : $> 5.9\sigma$
(300M pe \approx 0event)

Test statistics : Tevatron SingleTop

複雑なTest Statistics を使って発見をクレームした例 : Single-Top @ Tevatron

CDF Run II Preliminary, L = 3.2 fb⁻¹

- Single Top
- W+HF
- t \bar{t}
- QCD+Mistag
- Other
- Data

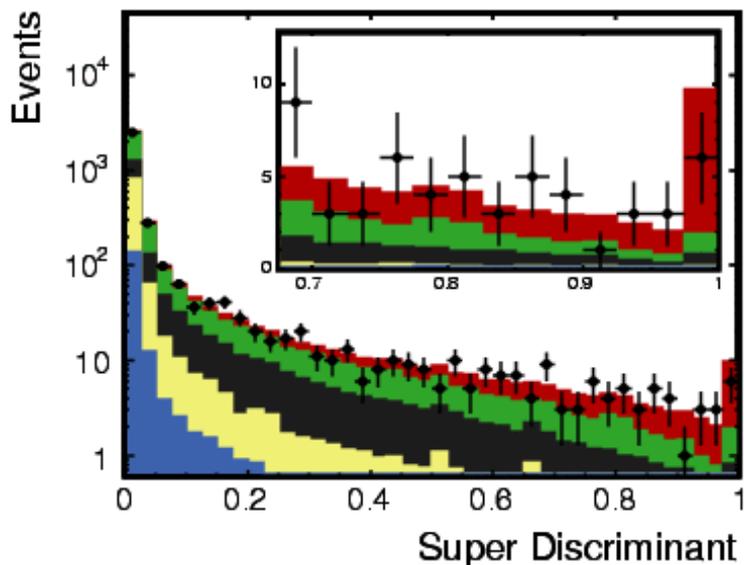
$$-2 \ln Q = -2 \ln \frac{p(\text{data} | H_1, \hat{\theta})}{p(\text{data} | H_0, \hat{\theta})}$$

H0 : b-only (null hypothesis)

H1 : s+b (test hypothesis)

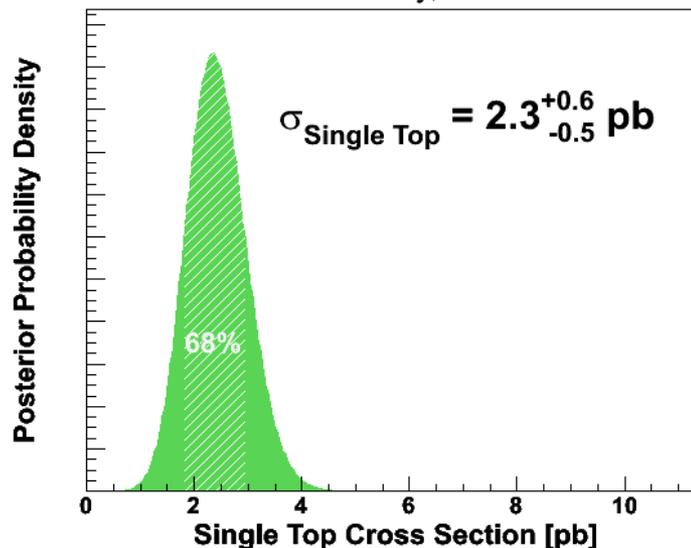
θ : Nuisance parameter

このように系統誤差をNuisance parameterとして
多変数解析をすることを**Profiling**という。→ profile likelihood



Profiling した結果で断面積をはかった。

CDF Run II Preliminary, L = 3.2 fb⁻¹



Test Statistics : LHC

LEP

$$Q = \frac{\mathcal{L}(\text{data}|s+b)}{\mathcal{L}(\text{data}|b)} = \frac{e^{-(S_{tot}+B_{tot})}}{e^{-B_{tot}}} \prod_{i=1}^{N_{bins}} \frac{(s_i + b_i)^{N_i}}{b_i^{N_i}}$$

Tevatron

$$-2 \ln Q = -2 \ln \frac{p(\text{data}|H_1, \hat{\theta})}{p(\text{data}|H_0, \hat{\theta})}$$

H0 : b-only(null hypothesis)
H1 : s+b (test hypothesis)
 θ : Nuisance parameter

LHC

Recommendation Document から

- Construct the likelihood function $\mathcal{L}(\mu, \theta)$ where μ is the signal strength and θ represent the nuisance parameters. An example of a likelihood function is given in section 5.
- Construct the test statistic \tilde{q}_μ based on the $\tilde{\lambda}(\mu)$ likelihood ratio:

$$\tilde{\lambda}(\mu) = \begin{cases} \frac{L(\mu, \hat{\theta}(\mu))}{L(\hat{\mu}, \hat{\theta})} & \hat{\mu} \geq 0, \\ \frac{L(\mu, \hat{\theta}(\mu))}{L(0, \hat{\theta}(0))} & \hat{\mu} < 0 \end{cases} \quad (1)$$

Here $\hat{\theta}(0)$ and $\hat{\theta}(\mu)$ refer to the conditional ML estimators of θ given a strength parameter of 0 or μ , respectively.

The test statistic \tilde{q}_μ is given by

$$\tilde{q}_\mu = \begin{cases} -2 \ln \tilde{\lambda}(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases}$$

Test Statistics : LHC

μ も θ もfreeにしてフィットしたときのbest fit

unconditional

μ を固定したときの θ のbest fit

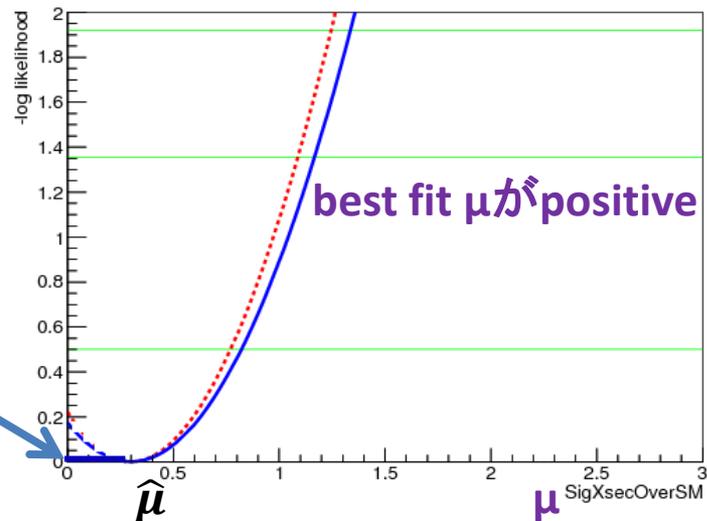
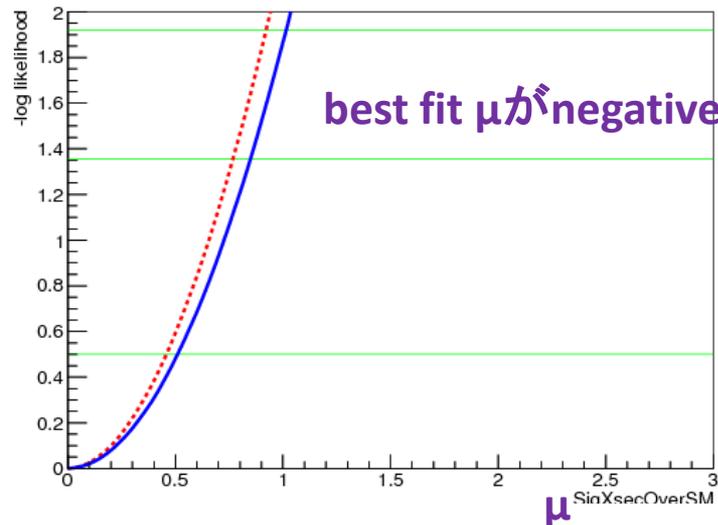
conditional

$$\tilde{\lambda}(\mu) = \begin{cases} \frac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(\hat{\mu}, \hat{\theta})} & \hat{\mu} \geq 0, \\ \frac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(0, \hat{\theta}(0))} & \hat{\mu} < 0 \end{cases}$$

$\hat{\mu} = 0$ とする。

best fit がnegative

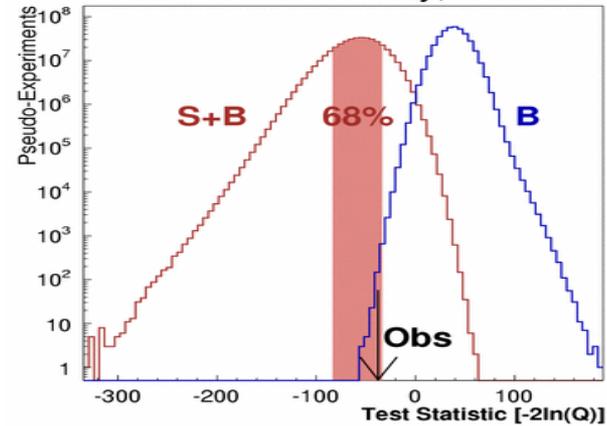
$$\tilde{q}_\mu = \begin{cases} -2 \ln \tilde{\lambda}(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases}$$



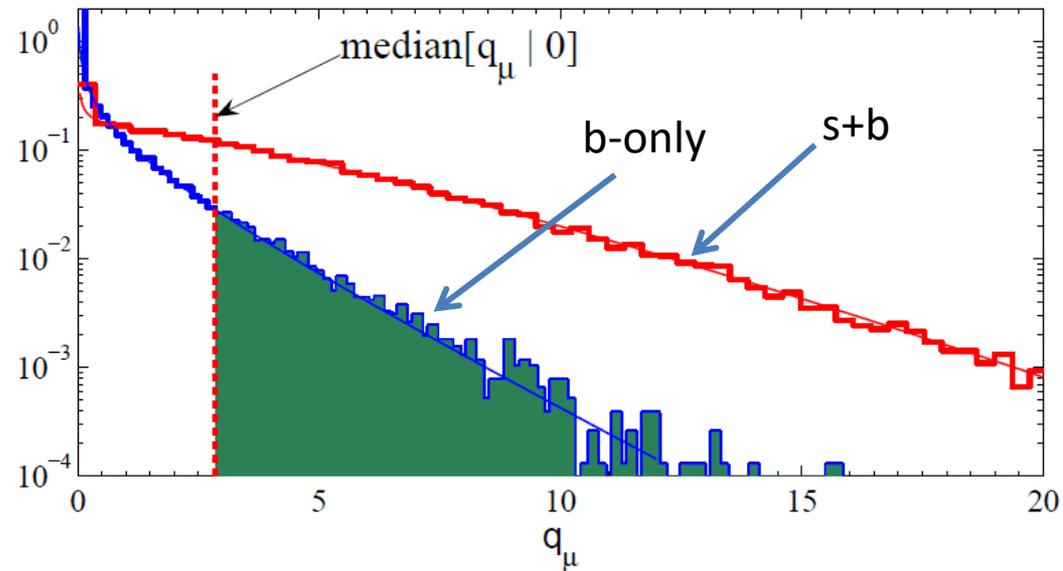
Test statistics の形

Tevatron

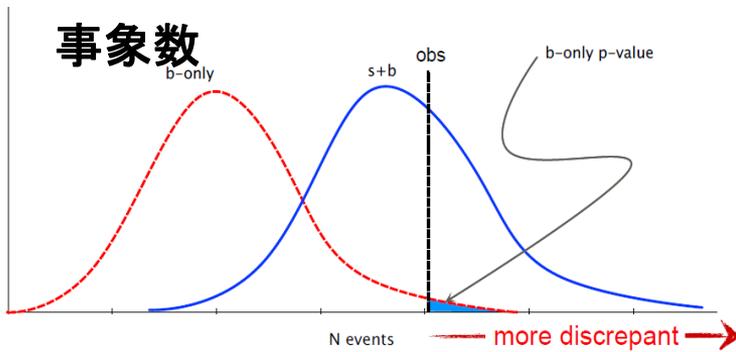
CDF Run II Preliminary, L = 3.2 fb⁻¹



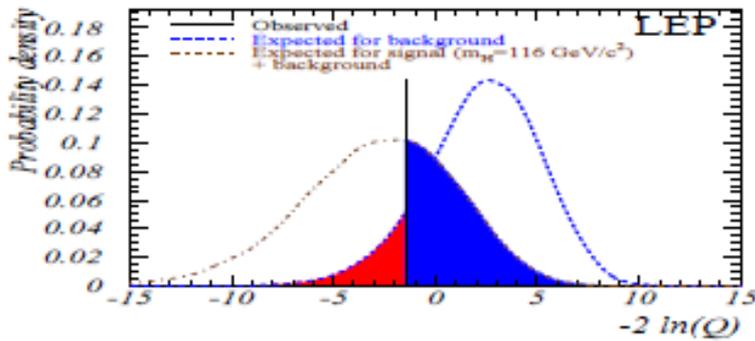
LHC



事象数



LEP



Profile Likelihood(例)

そもそも、 $\frac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(\hat{\mu}, \hat{\theta})}$ のLはどのように定義するのかという話

ここは、完全に解析による部分。
例えば、 $H \rightarrow \gamma\gamma$ と $H \rightarrow \tau\tau$ では全然違います。
例) Histofactory

$e \in \text{events}$

$b \in \text{bins}$

$c \in \text{channels}$

$s \in \text{samples}$

$p \in \text{parameters}$

Histofactor のちょっと変なところ
Luminosity だけは別扱い...

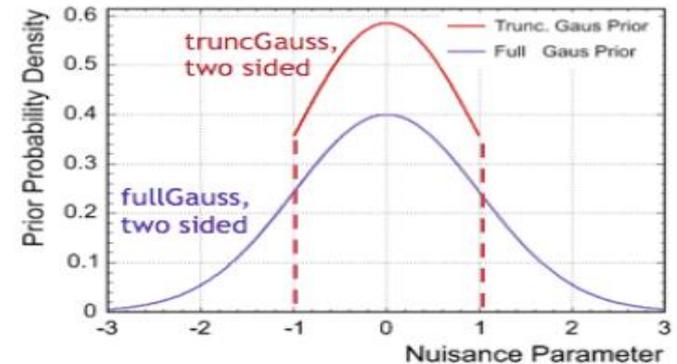
$$\mathcal{P}(n_{cb}, a_p | \phi_p, \alpha_p, \gamma_b) = \prod_{c \in \text{channels}} \prod_{b \in \text{bins}} \text{Pois}(n_{cb} | \nu_{cb}) \cdot G(L_0 | \lambda, \Delta_L) \cdot \prod_{p \in \mathbb{S} + \Gamma} P_p(a_p | \alpha_p)$$

ν_{cb} event 予想されているとき、
 n_{cb} event 観測する確率(統計:ポアソン)

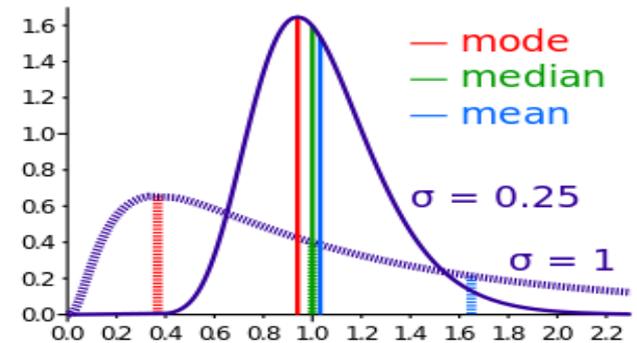
Nuisance parameter : α_p が、
別のところで計った大きさ a_p
でconstrainされている。

Constraint of Nuisance parameters

- 経験的にGauss分布?
→ 大きなerrorの問題
e.g. 60% uncertainty の -2σ はnegative!
- Tevatron Singletop :
 - Truncated Gaussian
- LHC Higgs :
 - Log normal constraint



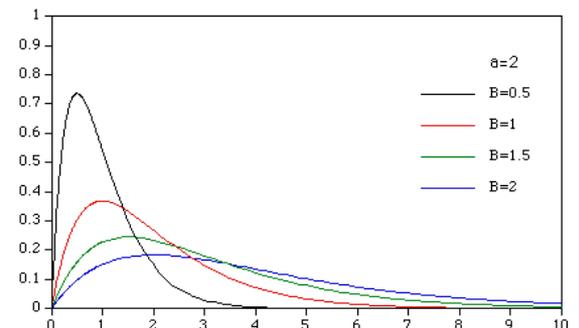
$$P_{LN}(n_p|\beta_p, \kappa_p) = \frac{1}{\sqrt{2\pi \ln \kappa}} \frac{1}{n_p} \exp \left[-\frac{\ln(n_p/\beta_p)^2}{2(\ln \kappa)^2} \right]$$



- 統計誤差?
– Gamma constraint

$$P_{\Gamma}(\beta_p|A = \tau_p, B = n_p - 1) = A(A\beta_p)^B e^{-A\beta_p} / \Gamma(B)$$

Poisson constraintと同じだが連続分布



Asymptotic とは？

- Test statistics が χ^2 分布に従う時、limit は test statistics から解析的に解ける。
- Test statistics:

$$q_\mu = -2 \ln \frac{L(\mu, \hat{\theta})}{L(\hat{\mu}, \hat{\theta})}; \quad \hat{\mu} < \mu$$

$$q_\mu = 0; \quad \hat{\mu} > \mu$$

を観測値から ∞ まで積分する。

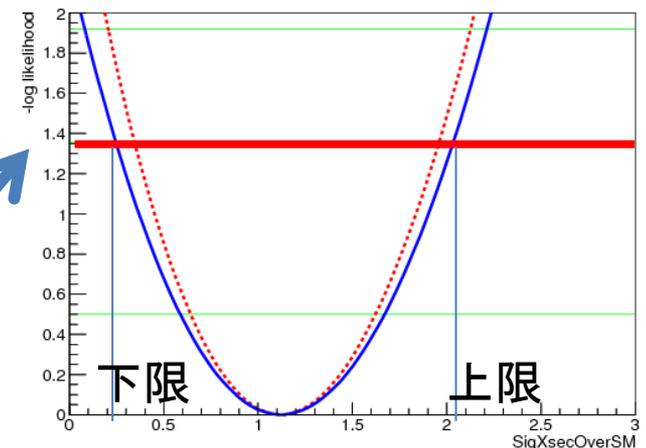
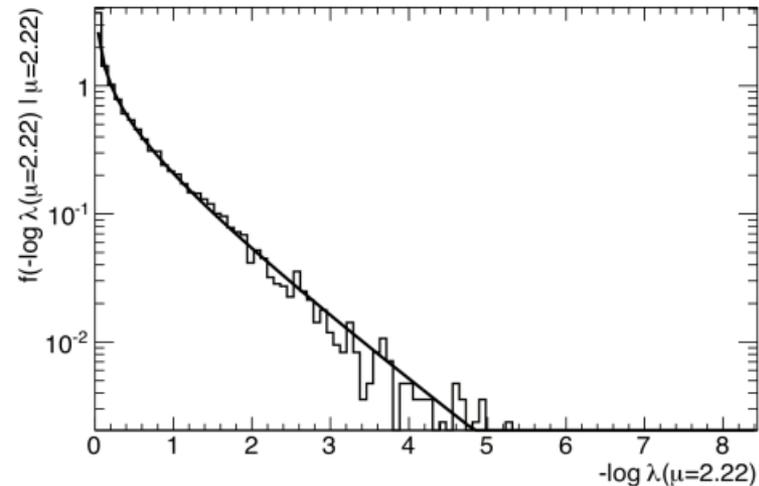
$$p_\mu = \int_{\hat{q}_{\mu, \text{obs}}}^{\infty} f(\tilde{q}_\mu | \mu, \hat{\theta}(\mu, \text{obs})) d\tilde{q}_\mu$$

これが CL に相当する確率。

$$p_{\mu_{95}} = 1 - \Phi(\sqrt{q_\mu}) = 1 - \Phi(1.64) = 0.05$$

つまり、test statistics の値が、 1.64^2 のときが 95%CL upper limit。

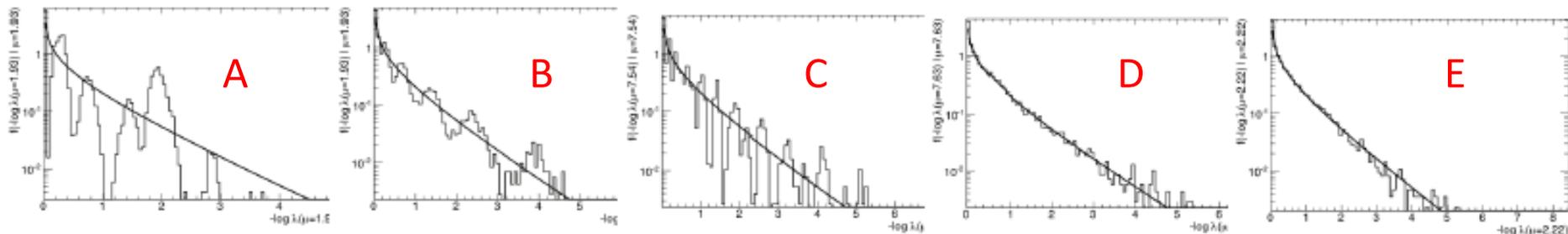
($q_\mu = -\log \lambda$ なので、 $-\log \lambda$ の plot の場合は $1.64^2/2 = 1.34$)



Asymptoticが使えないのは？

- では、どのようなときにAsymptotic近似が使えるのか、使えないのか？
- 基本的には、Toyを振ったlimitとAsymptotic limitが近いことを確認する必要がある。
- 単純な例でAsymptoticが破たんする例：小統計の場合

	s	b	Nobs	Approximate mu_95	Determine Regime to be
A	1 +/- 5%	0.1 +/- 5%	0	1.93	low count (clearly)
B	1 +/- 5%	5 +/- 5%	0	1.93	low count (to be safe)
C	1 +/- 1%	10 +/- 1%	10	7.54	intermediate
D	1 +/- 5%	10 +/- 5%	10	7.63	asymptotic
E	30 +/- 5%	200 +/- 5%	234	2.22	asymptotic (clearly)



Frequentist v.s. Bayesian

この論争は物理学だけでなく幅広い分野で起こっていて、いわば信仰みたいなもの。どちらにも長所短所があり、一概にどちらが良いとは言えない。

<真の値と観測量の取り扱い。>

- Frequentist: 真の値は固定であると考えて、確率の記述は観測量の再現性(頻度分布 “frequency”)となる。→つまり今まで話してきたことは、こっち。
- Bayesian: 観測量は固定であると考えて、確率は真の値を記述する(ある事前信念 “prior belief”が必要)。通常、この記述は、事後確率分布として表現される。(つまり、データの見解?みたいなもの。)

$$L'(data | r) = \int L(data | r, \theta) \pi(\theta) d\theta$$

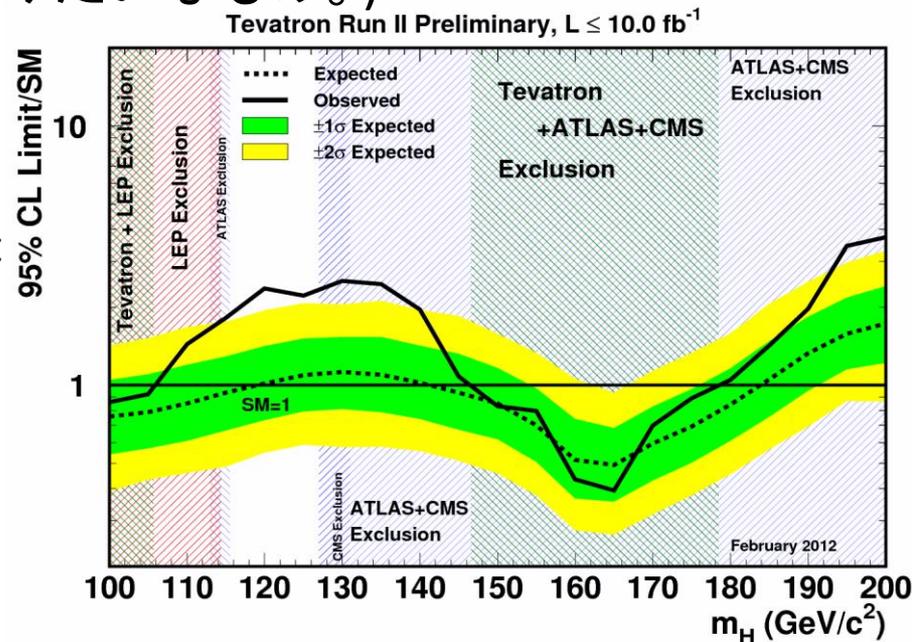
真の値(測定量) → ↑

系統誤差のPrior belief: e.g. Gauss constraint

$$0.95 = \int_0^{r_{lim}} L'(data | r) \pi(r) dr$$

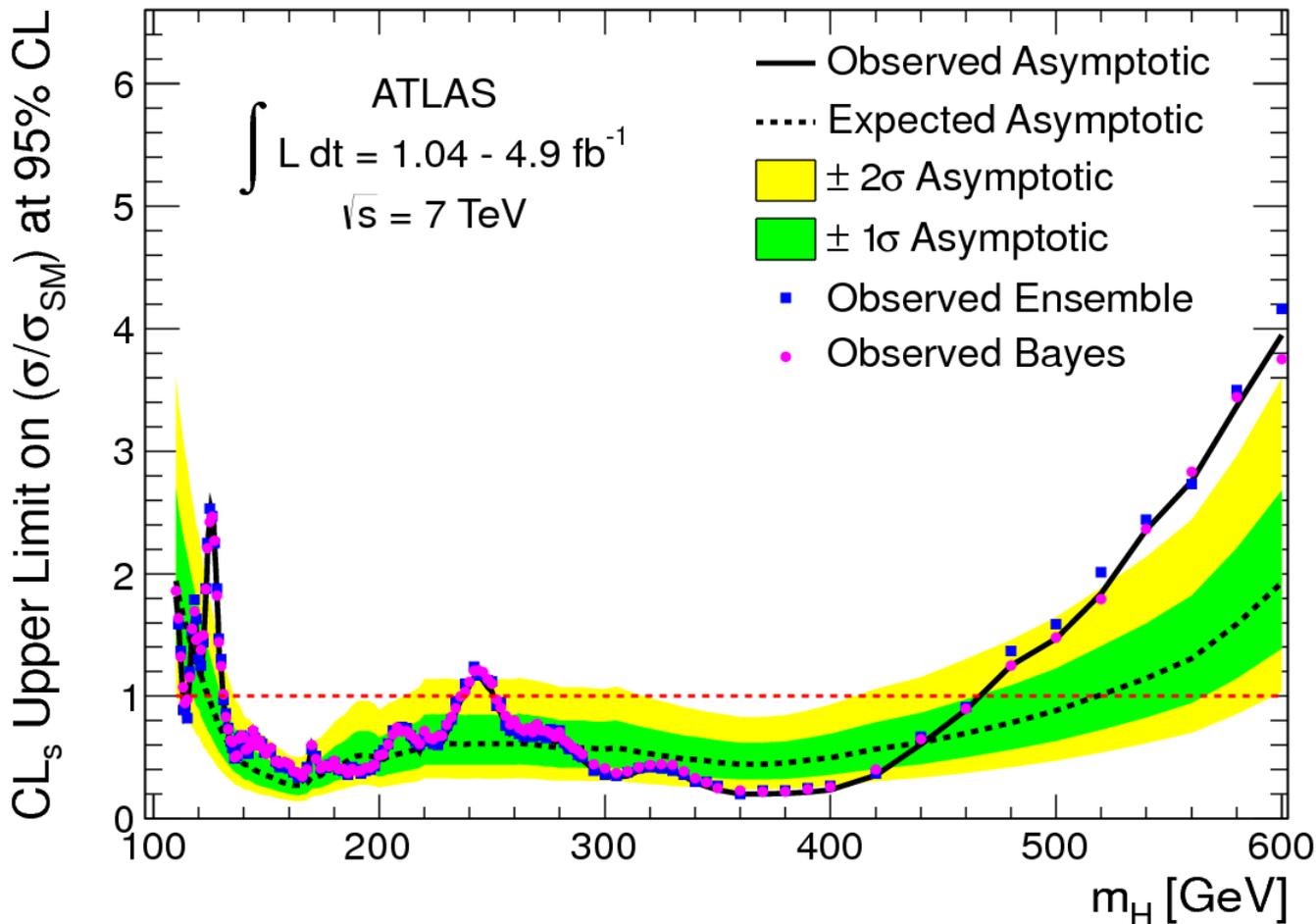
↑

真の値のPrior belief: constant (flat prior)



Tevatron のHiggs limitは、Bayesian.
ATLAS でも、一部使われている。

結果は違うのか？

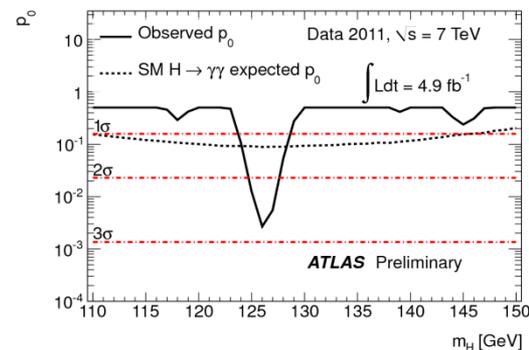
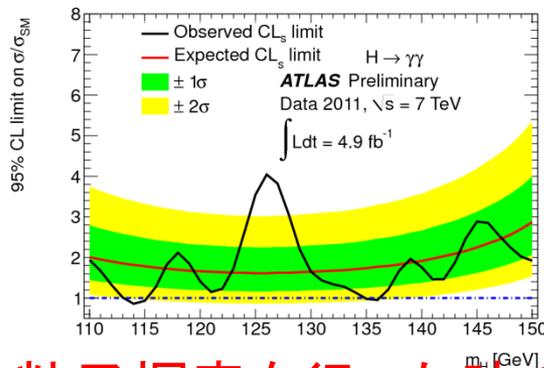


- 少なくともあるモデルを棄却する上ではFrequentistもBayesianもほとんど変わらない。ただし、発見・測定するときには要確認。

Look Elsewhere effectとは？

<http://arxiv.org/abs/1005.1891>

- $H \rightarrow \gamma\gamma$ @ 4.9fb^{-1}
 - Local p-value : 2.8σ
 - Global p-value : 1.5σ
- Local ? Global ? とは何か？



Local p-value : ある特定の質量の粒子探索を行った時の確率

Global p-value : 想定範囲でどこか一つの質量にあるものを探した時の確率

- 例1 100本に一本だけあたりがある宝くじを引く。
 - 一回だけ引くと当たる確率 : $1/100$ (local)
 - 10回引いて一度でも当たる確率: $1/10$ (global)
- 例2 バックグラウンドが 10^{-7} 事象予想されているとき1事象観測したら 5σ !
 - 10^7 回繰り返し実験をやったら1回くらい1事象観測してもそれは発見とは言えない。
(global p-value は、local p-value: 5σ に比べて圧倒的にちいさい。)

ここで、独立な実験を何回行ったか？が大事。

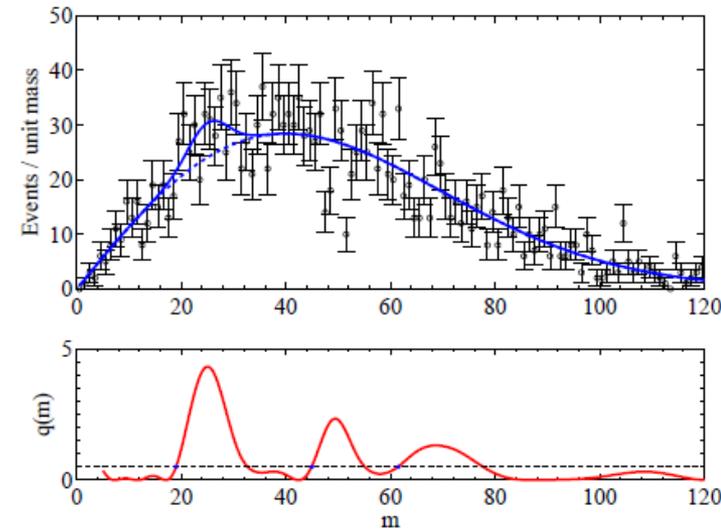
ヒッグスの質量領域でいうと、どのくらい質量点が離れたら独立と言えるのか？

とか、想定範囲内に何回相関の少ない実験が含まれているのか？

→信号の質量分布の幅が狭いほうがこの回数が多いはず。 $H \rightarrow \gamma\gamma$ や $H \rightarrow ZZ \rightarrow 4l$

Look Elsewhere effect

- 独立な実験の回数の決定
 - 質量の幅？
→バックグラウンドは相関あり。
これだけではだめ？
 - 実際のTest statisticsを使う。
 - あらゆる情報を含む。
(信号の幅、バックグラウンド)
 - ATLASでは、significance Z_0 とup cross する回数Nを用いる。
 - Z_0 は小さいほうが良い。→ ATLAS では、 $Z_0=0$ を用いている。



$$P_{global} = P_{local} + \underbrace{N e^{-(Z^2 - Z_0^2)/2}}_{\text{Look Elsewhere effect}}$$

Z: local significance N: 回数

この分だけ p-value が大きくなる。Sensitivityが下がる

backup

Documentations/useful information

- Statistics group mailing list :
hn-atlas-physics-statistics@cern.ch , atlas-phys-stat-root@cern.ch
- ATLASの論文で参照することになっているdocument
 - Asymptotic formulae for likelihoodbased tests of new physics
 - <http://arxiv.org/abs/1007.1727>
 - Presentation of search results: the CLs technique
 - A.L. Read, J. Phys. G 28 (2002) 2693.
- ATLAS 実験の統計Recommendation
 - <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/FrequentistLimitRecommendationImplementation>
- Tool/Tutorials
 - Histofactory
 - <https://twiki.cern.ch/twiki/pub/RooStats/WebHome/HistFactoryLikelihood.pdf>
 - <https://indico.cern.ch/conferenceDisplay.py?confId=118480>
 - <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/ExoticsStatisticsTutorialWorkspaces>
 - RooStat Tools
 - <https://svnweb.cern.ch/trac/atlasgrp/browser/Physics/StatForum/RooStatsTools>
- その他
 - PDG Statistics section.
 - CLs for ATLAS speaker : <https://twiki.cern.ch/twiki/pub/AtlasProtected/StatisticsTools/CLsInfo.pdf>
 - Bayesian limit :
<https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/BayesianLimitRecommendationImplementation>