Statistics in LHC Phenomenology

Tilman Plehn

MPI für Physik & University of Edinburgh

Bonn, 2/2007

Outline

Maximum signal significance

Neyman–Pearson lemma

Example: Higgs to muons

Supersymmetric parameter space

Markov chains

Higgs searches — life is tough

Life at LHC

– WBF $H \to \tau\tau$ in Standard Model [or MSSM]

– cut analysis promising enough

⇒ experimentalists at work [for example Atlas–Freiburg–Bonn]

– neural net better [non-trivially bounded signal regions]

– even better with LEP–type event weighting [not just counting experiment]

– Higgs discovery channel?

⇒ could we guess such an outcome? [or the opposite]

▸ Significance for 30 fb^{-1}:

Higgs Mass	Cut Analysis(Pois.)	Cut on NN	NN Sig. w/cut	NN Sig. w/LR
115	2.95	0.89	3.71	4.68
120	3.09	0.93	3.97	4.88
125	3.06	0.92	3.93	4.75
130	2.72	0.94	3.70	4.49
135	2.56	0.96	3.36	4.02
140	1.86	0.97	2.85	3.38

▸ Improvement of ~30% from Neural Nets

▸ Improvement of ~60% with Likelihood Ratio

[B. Quayle, ATLAS Higgs meeting, 2003]

Neyman–Pearson lemma

Answer: Neyman–Pearson lemma

- correct hypothesis m_1: Higgs signal
  wrong hypothesis m_2: SM background
- lemma: likelihood ratio $p(d|m_1)/p(d|m_2)$ most powerful estimator

  [lowest probability to mistake right for fluctuation of wrong (type-II error)]
- likelihood: for phase–space event $p(d|m) \sim |\mathcal{M}|^2$ [from Monte Carlo]
- estimator: plot density with estimator on x axis, cut signal–rich region

Application: optimal observables [invite Markus Diehl...]

- looking for best way to measure LEP physics
- use Neyman–Pearson theorem to construct correlated observables

Similar: matrix element method [CDF; DZero]

- event likelihood from data and Monte–Carlo [jet–parton identification]
- express likelihood of top events as function of m_t
- maximize probability $p(d|\mathrm{SM}, m_t)$ to measure m_t

$\Rightarrow$ likelihood hard to extract from data [single–top]

Neyman–Pearson lemma

Answer: Neyman–Pearson lemma

– correct hypothesis m_1: Higgs signal
  wrong hypothesis m_2: SM background

– lemma: likelihood ratio $p(d|m_1)/p(d|m_2)$ most powerful estimator

  [lowest probability to mistake right for fluctuation of wrong (type-II error)]

Optimal significance at parton level [Cranmer, TP]

– example: log-likelihood for n-event Poisson statistics [independent channels]

$$\text{Pois}(n|b) = \frac{e^{-b}\; b^n}{n!} \qquad \text{Pois}(n|s+b) = \frac{e^{-(s+b)}\; (s+b)^n}{n!}$$

$$q = \log \frac{\text{Pois}(n|s+b)}{\text{Pois}(n|b)} = -s + n\; \log\left(1 + \frac{s}{b}\right) \longrightarrow -\sum_j s_j + \sum_j n_j\; \log\left(1 + \frac{s_j}{b_j}\right)$$

– independent events with non–trivial distributions

$$q = \log\; \frac{\text{Pois}(n|s+b)\; \prod_{j=1}^n f_{s+b}^{(j)}}{\text{Pois}(n|b)\; \prod_{j=1}^n f_b^{(j)}} = -s + \sum_{j=1}^n \log\left(1 + \frac{s f_s^{(j)}}{b f_b^{(j)}}\right)$$

– continuous integration over phase space: $s\, f_s \to |\mathcal{M}_s|^2$

$$q(\vec{r}) = -\sigma_s \mathcal{L} + \log\left(1 + \frac{|\mathcal{M}_s(\vec{r})|^2}{|\mathcal{M}_b(\vec{r})|^2}\right)$$

Neyman–Pearson lemma

Answer: Neyman–Pearson lemma

– correct hypothesis m_1: Higgs signal
  wrong hypothesis m_2: SM background

– lemma: likelihood ratio $p(d|m_1)/p(d|m_2)$ most powerful estimator

[lowest probability to mistake right for fluctuation of wrong (type-II error)]

Optimal significance at parton level [Cranmer, TP]

– from likelihood map $q(\vec{r})$ to probability distribution pdf

– invert into single–event pdf

$$\rho_{1,b}(q_0) = \int d\vec{r}\, \frac{d\sigma_b(\vec{r})}{\sigma_{\text{tot},b}}\, \delta\left(q(\vec{r}) - q_0\right)$$

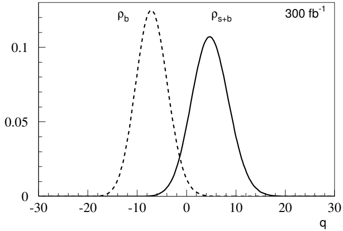
– Fourier–transform and compute n–event pdf: $\overline{\rho_{n,b}} = (\overline{\rho_{1,b}})^n$

– combine $n = 1, \dots$ into pdf

$$\rho_b(q) = \sum_n \text{Pois}(n|b) \times \rho_{n,b}(q)$$

$\Rightarrow$ integrate to $CL_b(q) = \int_q^\infty dq'\, \rho_b(q')$

[5σ is $CL_b = 2.85\, 10^{-7}$]



Sub–optimal: detector effects

Best of all worlds

– irreducible & unsmeared: identical signal and background phase space

$$\sigma_{tot} = \int dPS\, M_{PS}\, d\sigma_{PS} = \int d\vec{r}\, M(\vec{r})\, d\sigma(\vec{r})$$

– random numbers $\vec{r}$ basis for phase space configurations

⇒ don't be ridiculous $\Delta m_{\mu\mu}^{\text{width}} \ll \Delta m_{\mu\mu}^{\text{meas}}$

More realistic

– smear observable/random number transfer function W [Gaussian]

$$\sigma_{tot} = \int d\vec{r}_\perp dr_m^* \int_{-\infty}^{\infty} dr_m\, M(\vec{r})\, d\sigma(\vec{r})\, W(r_m, r_m^*)$$

– modified phase–space vector $\vec{r} = \{\vec{r}_\perp, r_m\}$ [without back door]

– likelihood map over smeared $\vec{r}$

⇒ same procedure as before

– complete smearing: replace phase space by set of distributions

– lose strict maximum significance claim

⇒ step–by–step into Whizard [Cranmer, TP, Reuter]

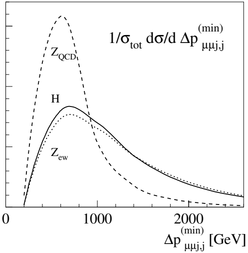
Example: Higgs to muons

Weak–boson–fusion Higgs with $H \to \mu\mu$

– number of signal events small [$\sigma \cdot BR \sim 0.25 fb$]

– no distribution with golden cut

$\Rightarrow$ perfect for multivariate analysis



Awful old results [TP, Rainwater]

$\sqrt{S}$	M_H	σ_H [fb]	σ_Z^{QCD} [fb]	σ_Z^{ew} [fb]	S/B	significance	$\triangle\sigma/\sigma$	$\mathcal{L}_{5\sigma}$ [fb^{-1}]
14	115	0.25	3.57	0.40	1/9.1	1.7 σ	60%	2600
14	120	0.22	2.60	0.33	1/7.5	1.8 σ	60%	2300
14	130	0.17	1.61	0.24	1/6.5	1.7 σ	65%	2700
14	140	0.10	1.11	0.19	1/7.5	1.2 σ	85%	4900
200	115	2.57	39.6	5.3	1/10.1	5.3 σ	20%	270
200	120	2.36	29.2	4.0	1/8.0	5.7 σ	20%	230
200	130	1.80	18.7	2.7	1/6.9	5.3 σ	20%	260
200	140	1.14	13.4	2.0	1/7.9	4.0 σ	27%	500

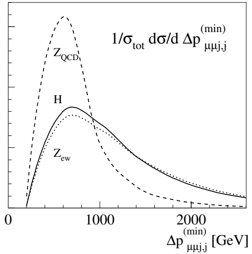
Example: Higgs to muons

Weak–boson–fusion Higgs with $H \to \mu\mu$

– number of signal events small [$\sigma \cdot BR \sim 0.25 fb$]

– no distribution with golden cut

$\Rightarrow$ perfect for multivariate analysis



Statistical promise

– mostly irreducible backgrounds

– smearing only relevant for $m_{\mu\mu}$ [mimic by Γ'_H?]

– compute likelihood map from matrix elements

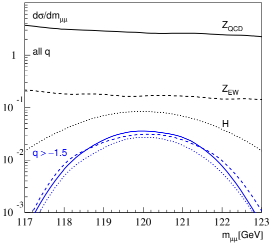
$\to$ upper limit (target?) on parton–level significance

$\to$ WBF $H \to \mu\mu$: 3.5 sigma in 300 fb^{-1}

[$\sim 4.4\sigma$ with mini-jet veto]

– physics: confirm Yukawa coupling

$\Rightarrow$ maybe, Jörn wants to have a look?



Supersymmetric parameter space

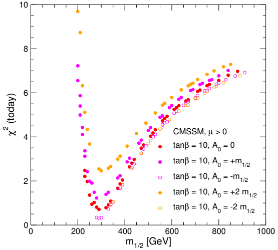
New physics at the LHC

- complex models, including dark matter, flavor physics, low-energy physics,...
- honest parameters: weak-scale Lagrangean
- measurements: masses or edges
  branching fractions
  cross sections
- errors: general correlation, statistics & systematics & theory
- problem in grid: huge phase space, local minimum?
  problem in fit: domain walls, global minimum? [also Fittino: Peter's talk]

First go at problem

- ask a friend how SUSY is broken ⇒ mSUGRA
- fit $m_0, m_{1/2}$
- no problem, include indirect constraints
- best fit pre-LHC [Ellis, Weinemeyer, Olive, Heiglein]

⇒ simple fit [no theory bias, except they know it is mSUGRA]



Supersymmetric parameter space

New physics at the LHC

– complex models, including dark matter, flavor physics, low-energy physics,...

– honest parameters: weak-scale Lagrangean

– measurements: masses or edges
branching fractions
cross sections

– errors: general correlation, statistics & systematics & theory

– problem in grid: huge phase space, local minimum?
problem in fit: domain walls, global minimum? [also Fittino: Peter's talk]

First go at problem

– ask a friend how SUSY is broken ⇒ mSUGRA

– fit $m_0, m_{1/2}, A_0, \tan\beta, y_t, ...$

⇒ best fit to LHC/ILC measurements

	SPS1a	ΔLHC masses	ΔLHC edges	ΔILC	ΔLHC+ILC
m_0	100	3.9	1.2	0.09	0.08
$m_{1/2}$	250	1.7	1.0	0.13	0.11
$\tan\beta$	10	1.1	0.9	0.12	0.12
A_0	-100	33	20	4.8	4.3

Supersymmetric parameter space

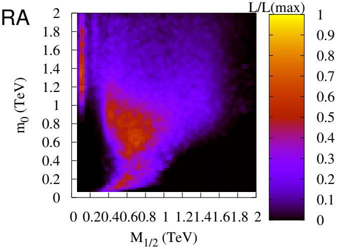
New physics at the LHC

- complex models, including dark matter, flavor physics, low-energy physics,...
- honest parameters: weak-scale Lagrangean
- measurements: masses or edges
  branching fractions
  cross sections
- errors: general correlation, statistics & systematics & theory
- problem in grid: huge phase space, local minimum?
  problem in fit: domain walls, global minimum? [also Fittino: Peter's talk]

First go at problem

- ask a friend how SUSY is broken ⇒ mSUGRA
- fit $m_0, m_{1/2}, A_0, \tan\beta, \text{sign}(\mu), y_t, \ldots$
- no problem, include indirect constraints
- probability map today [Allanach, Lester, Weber]

⇒ more complicated for MSSM@LHC



Supersymmetric parameter space

New physics at the LHC

– complex models, including dark matter, flavor physics, low-energy physics,...

– honest parameters: weak-scale Lagrangean

– measurements: masses or edges
branching fractions
cross sections

– errors: general correlation, statistics & systematics & theory

– problem in grid: huge phase space, local minimum?
problem in fit: domain walls, global minimum? [also Fittino: Peter's talk]

MSSM instead of mSUGRA

– technically painful
(1) grid for closed subset
(2) fit of other parameters
(3) complete fit

$\Rightarrow$ secondary minima???

	LHC	ILC	LHC+ILC	SPS1a
$\tan\beta$	10.22$\pm$9.1	10.26$\pm$0.3	10.06$\pm$0.2	10
M_1	102.45$\pm$5.3	102.32$\pm$0.1	102.23$\pm$0.1	102.2
M_3	578.67$\pm$15	fix 500	588.05$\pm$11	589.4
$M_{\tilde{\tau}_L}$	fix 500	197.68$\pm$1.2	199.25$\pm$1.1	197.8
$M_{\tilde{\tau}_R}$	129.03$\pm$6.9	135.66$\pm$0.3	133.35$\pm$0.6	135.5
$M_{\tilde{\mu}_L}$	198.7$\pm$5.1	198.7$\pm$0.5	198.7$\pm$0.5	198.7
$M_{\tilde{q}3_L}$	498.3$\pm$110	497.6$\pm$4.4	521.9$\pm$39	501.3
$M_{\tilde{t}_R}$	fix 500	420$\pm$2.1	411.73$\pm$12	420.2
$M_{\tilde{b}_R}$	522.26$\pm$113	fix 500	504.35$\pm$61	525.6
A_τ	fix 0	-202.4$\pm$89.5	352.1$\pm$171	-253.5
A_t	-507.8$\pm$91	-501.95$\pm$2.7	-505.24$\pm$3.3	-504.9
A_b	-784.7$\pm$35603	fix 0	-977$\pm$12467	-799.4

Markov chains

Probability maps of new physics

– Bayes' theorem: $p(m|d) = p(d|m)\ p(m)/p(d)$ [$p(d)$ through normalization]

– likelihood: data given a model $p(d|m) \sim |\mathcal{M}|^2$

– theorist's prejudice: model $p(m)$

$\Rightarrow$ given measurements: (1) compute map $p(m|d)$ of parameter space
(2) rank local maxima

Weighted Markov chains [Rauch, TP]

– classical: produce representative set of spin states
compute average energy based on this reduced sample

$\Rightarrow$ map (chain) based on probability of a state
expensive energy function on sample

– BSM physics: produce map $p(m|d)$ of parameter points
evaluate same probability from (binned) density

$\Rightarrow$ weighted Markov chains [like weighted Monte Carlo]

– already for mSUGRA: MCMC resolution not sufficient

$\Rightarrow$ use additional probability maximization to rank maxima

Markov chains

Probability maps of new physics

– Bayes' theorem: $p(m|d) = p(d|m)\ p(m)/p(d)$ [$p(d)$ through normalization]

– likelihood: data given a model $p(d|m) \sim |\mathcal{M}|^2$

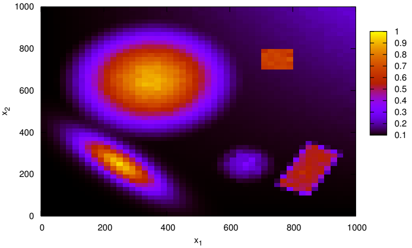
– theorist's prejudice: model $p(m)$

$\Rightarrow$ given measurements: (1) compute map $p(m|d)$ of parameter space
(2) rank local maxima

Toy model

– test function $V(\vec{x})$ in 5 dimensions [general high–dimensional extraction tool]

– Sfitter output #1: probability map
Sfitter output #2: list of local maxima [best fit]



V=74.9	(655	253	347	348	349)
V=59.9	(850	224	650	649	654)
V=58.2	(849	225	587	650	650)
V=25.1	(750	749	450	450	450)
V=16.0	(245	253	552	542	544)
V=12.1	(350	650	650	650	650)
...					

Markov chains

Probability maps of new physics

– Bayes' theorem: $p(m|d) = p(d|m)\ p(m)/p(d)$ [$p(d)$ through normalization]

– likelihood: data given a model $p(d|m) \sim |\mathcal{M}|^2$

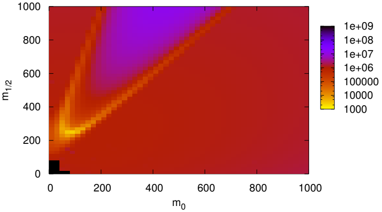
– theorist's prejudice: model $p(m)$

$\Rightarrow$ given measurements: (1) compute map $p(m|d)$ of parameter space
(2) rank local maxima

mSUGRA with LHC measurements [Lafaye, TP, Rauch, D.Zerwas]

– SPS1a kinematic edges with free m_b, m_t

– Sfitter output #1: probability map
Sfitter output #2: list of local maxima [best fit]



χ^2	m_0	$m_{1/2}$	$\tan\beta$	A_0	μ	m_t
0.3e-04	100.0	250.0	10.0	-99.9	+	171.4
27.42	99.7	251.6	11.7	848.9	+	181.6
54.12	107.2	243.4	13.3	-97.4	-	171.1
70.99	108.5	246.9	13.9	26.4	-	173.6
88.53	107.7	245.9	12.9	802.7	-	182.7
...						

Markov chains

Probability maps of new physics

– Bayes' theorem: $p(m|d) = p(d|m)\ p(m)/p(d)$ [$p(d)$ through normalization]

– likelihood: data given a model $p(d|m) \sim |\mathcal{M}|^2$

– theorist's prejudice: model $p(m)$

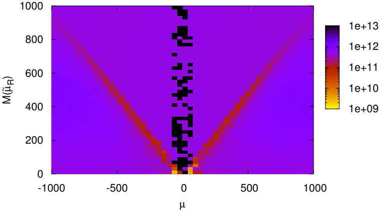
$\Rightarrow$ given measurements: (1) compute map $p(m|d)$ of parameter space
(2) rank local maxima

MSSM with LHC measurements

– complete weak–scale MSSM

– Sfitter output #1: probability map
Sfitter output #2: list of local maxima soon

$\Rightarrow$ ready to include bias [interpretation determined by quality of data]



To take home...

Statistics a powerful tool

- likelihood methods useful for phenomenology
- maximum significance from event generator
- topic for discussions between theorists and experimentalists

- complex new physics models at LHC
- secondary minima guaranteed, theory bias unavoidable
- another topic for discussions between theorists and experimentalists