Likelihood Functions for Supersymmetric Observables in Frequentist Analyses of the CMSSM and NUHM1 arXiv:0907.5568v1

Penn HEP Journal Club

Ryan D. Reece

University of Pennsylvania

September 29, 2009

SUSY

- The Minimal Supersymmetric Standard Model (MSSM) is the supersymmetric extension to the SM with the simplest Higgs sector, but otherwise general.
- The MSSM has two complex Higgs doublets: H_u and H_d that couple to up and down type fermions respectively.
- After electroweak symmetry breaking, the Higgs fields have two vacuum expectation values: v_u, v_d, $\tan\beta \equiv v_u/v_d$.
- Symmetry breaking still introduces 3 massless Goldstone bosons.
- The other 5 degrees of freedom give 5 Higgs bosons: $h, H, A, H^{\pm}$.
- 137 additional parameters

CMSSM and NUMH1

CMMSM

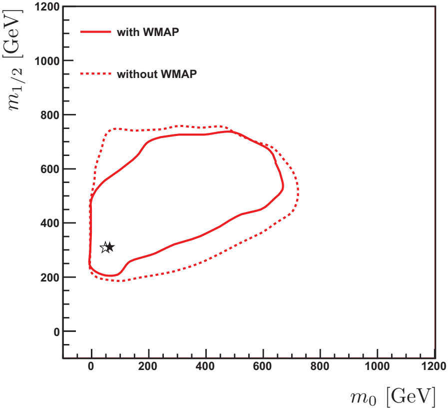
- The Constrained MSSM (CMSSM) is a subset of possible MSSMs in which "some universality relations are imposed on the soft SUSY-breaking parameters," limiting the number of parameters to a few. [arXiv:0808.4128v1]
- m_0 and $m_{1/2}$ are parameters that set the scale for the masses of the supersymmetric scalars and gauginos.

NUMH1

- Is a generalization of the CMSSM, dropping a universality relation in the Higgs sector.

χ^2 Fits and Confidence Contours

$$\chi^2 = \sum_i \frac{(C_i - P_i)^2}{\sigma(C_i)^2 + \sigma(P_i)^2} + \chi^2(\text{search constraints}) + \sum_i \frac{(f_{\mathrm{SM}i}^{\mathrm{obs}} - f_{\mathrm{SM}i}^{\mathrm{fit}})^2}{\sigma(f_{\mathrm{SM}i})^2}$$



[arXiv:0808.4128v1]

We'll discuss how one goes from the statistic on the left to the plot on the right.

Primer on the Maximum Likelihood Method

Consider a Gaussian distributed measurement:

$$f_1(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

If we repeat the measurement, the joint PDF is just a product:

$$f(\vec{x}|\mu,\sigma) = \prod_i f_1(x_i|\mu,\sigma)$$

The **likelihood function** is the same function as the PDF, only thought of as a function of the parameters, given the data. The experiment is over.

$$L(\mu,\sigma|\vec{x}) = f(\vec{x}|\mu,\sigma)$$

The **likelihood principle** states that the best estimate of the true parameters are the values which maximize the likelihood.

It is often more convenient to consider the log likelihood, which has the same maximum.

$$\ln L = \ln \prod f_1 = \sum \ln f_1$$
$$= \sum_i \left(-\frac{1}{2} \ln(2\pi\sigma^2) - \frac{(x_i - \mu)^2}{2\sigma^2} \right)$$

Maximize:

$$\frac{\partial \ln L}{\partial \mu} = \sum_i \frac{x_i - \hat{\mu}}{\sigma^2} = 0$$

$$\Rightarrow \qquad \sum_i (x_i - \hat{\mu}) = 0, \qquad \Rightarrow \qquad \hat{\mu} = \frac{1}{N} \sum_i x_i = \bar{x}$$

Which agrees with our intuition that the best estimate of the mean of a Gaussian is the sample mean.

Note that in the case of a Gaussian PDF, maximizing likelihood is equivalent to minimizing χ^2.

$$\ln L = \sum_i \left(-\frac{1}{2} \ln(2\pi\sigma^2) - \frac{(x-\mu)^2}{2\sigma^2} \right)$$

is maximized when

$$\chi^2 = \sum_i \frac{(x-\mu)^2}{\sigma^2}$$

is minimized.

This was a simple example of what statisticians call **point estimation**. Now we would like to quantify our error on this estimate.

One would think that if the likelihood function varies rather slowly near the peak, then there is a wide range of values of the parameters that are consistent with the data, and thus the estimate should have a large error.

To see the behavior of the likelihood function near the peak, consider the Taylor expansion of a general $\ln L$ of some parameter θ, near its maximum likelihood estimate $\hat{\theta}$:

$$\ln L(\theta) = \ln L(\hat{\theta}) + \cancelto{0}{\left.\frac{\partial \ln L}{\partial \theta}\right|_{\hat{\theta}}} (\theta - \hat{\theta}) + \frac{1}{2!} \underbrace{\left.\frac{\partial^2 \ln L}{\partial \theta^2}\right|_{\hat{\theta}}}_{-1/s^2} (\theta - \hat{\theta})^2 + \cdots$$

Dropping the remaining terms would imply that

$$L(\theta) = L(\hat{\theta}) \, \exp\left(\frac{-(\theta - \hat{\theta})^2}{2s^2} \right)$$

Note that the $\ln L(\theta)$ is parabolic.

So if this were a good approximation, we would expect that the variance of $\hat{\theta}$ would be given by

$$V[\hat{\theta}] \equiv \sigma_{\hat{\theta}}^2 = s^2 = -\left(\left. \frac{\partial^2 \ln L}{\partial \theta^2} \right|_{\hat{\theta}} \right)^{-1}$$

It turns out that there is more truth to this than you would think, given by an important theorem in statistics, the **Cramér-Rao Inequality**:

$$V[\hat{\theta}] \geq \left(1 + \frac{\partial b}{\partial \theta} \right)^2 / E\left[- \left. \frac{\partial^2 \ln L}{\partial \theta^2} \right|_{\hat{\theta}} \right]$$

An estimator's **efficiency** is defined to measure to what extent this inequality is equivalent:

$$\varepsilon[\hat{\theta}] \equiv \frac{1/E\left[- \left. \frac{\partial^2 \ln L}{\partial \theta^2} \right|_{\hat{\theta}} \right]}{V[\hat{\theta}]}$$

It can be shown that in the large sample limit:

Maximum likelihood estimators are unbiased and 100% efficient.

Therefore, in principle, one can calculate the variance of an ML estimator with

$$V[\hat{\theta}] = -\left(E\left[\left.\frac{\partial^2 \ln L}{\partial \theta^2}\right|_{\hat{\theta}}\right]\right)^{-1}$$

Calculating the expectation value would involve an analytic integration over the PDFs of all our possible measurements, or a Monte Carlo simulation of it. In practice, one usually uses the observed maximum likelihood estimate as the expectation.

$$\boxed{V[\hat{\theta}] = -\left(\left.\frac{\partial^2 \ln L}{\partial \theta^2}\right|_{\hat{\theta}}\right)^{-1}}$$

Let's go back to our simple example of a Gaussian likelihood to test this method of calculating the ML estimator's variance.

$$V[\hat{\mu}] = -\left(\left. \frac{\partial^2 \ln L}{\partial \mu^2} \right|_{\hat{\mu}} \right)^{-1}$$

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \mu^2} &= \frac{\partial^2}{\partial \mu^2} \sum_i \left(-\frac{1}{2} \ln(2\pi\sigma^2) - \frac{(x_i - \mu)^2}{2\sigma^2} \right) \\ &= \frac{\partial}{\partial \mu} \sum_i \frac{x_i - \hat{\mu}}{\sigma^2} = \sum_i \frac{-1}{\sigma^2} = \frac{-N}{\sigma^2} \end{aligned}$$

$$\Rightarrow \qquad V[\hat{\mu}] = \frac{\sigma^2}{N} \qquad \Rightarrow \qquad \sigma_{\hat{\mu}} = \frac{\sigma}{\sqrt{N}}$$

Which many of you will recognize as the proper error on the sample mean. If you are unfamiliar with it, we can actually derive it analytically in this case.

$$V[\bar{x}] = E[\bar{x}^2] - \cancel{E[\bar{x}]^2} \nearrow \mu$$

$$= E\left[\left(\frac{1}{N}\sum_i x_i\right)\left(\frac{1}{N}\sum_j x_j\right)\right] - \mu^2$$

$$= \frac{1}{N^2}\, E\left[\sum_{i\neq j} x_i\, x_j + \sum_i x_i^2\right] - \mu^2$$

$$= \frac{1}{N^2}\left(\sum_{i\neq j} \cancel{E[x]^2} \nearrow \mu^2 + \sum_i E[x^2]\right) - \mu^2$$

To find $E[x^2]$, consider

$$V[x] = \sigma^2 = E[x^2] - E[x]^2 = E[x^2] - \mu^2$$

$$\Rightarrow \qquad E[x^2] = \sigma^2 + \mu^2$$

$$\Rightarrow \qquad V[\bar{x}] = \frac{1}{N^2}\left(\sum_{i\neq j}\mu^2 + \sum_i(\sigma^2+\mu^2)\right) - \mu^2$$

$$= \frac{1}{N^2}\left((N^2-N)\mu^2 + N(\sigma^2+\mu^2)\right) - \mu^2$$

$$= \frac{\sigma^2}{N}$$

Which verifies the result we got from calculating derivatives of the likelihood function.

$$V[\hat{\theta}] = -\left(\left.\frac{\partial^2 \ln L}{\partial\theta^2}\right|_{\hat{\theta}}\right)^{-1}$$

In practice, one usually doesn't calculate this analytically, but instead:

- calculates the derivatives numerically, or
- uses the $\Delta \ln L$ or $\Delta\chi^2$ method, described now

Back to our Taylor expansion of $\ln L$:

$$\ln L(\theta) = \ln L(\hat{\theta}) + \frac{1}{2!} \underbrace{\left.\frac{\partial^2 \ln L}{\partial \theta^2}\right|_{\hat{\theta}}}_{-1/\sigma^2_{\hat{\theta}}} (\theta - \hat{\theta})^2 + \cdots$$

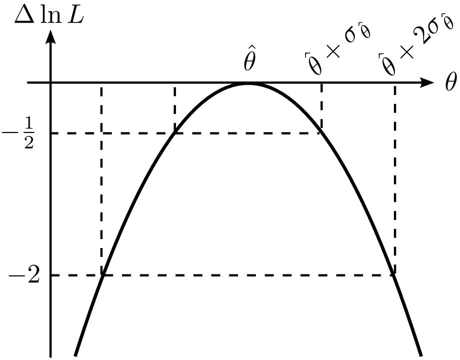
Let $\quad \Delta \ln L(\theta) \equiv \ln L(\theta) - \ln L(\hat{\theta})$

$$\Delta \ln L(\theta) \simeq -\frac{(\theta - \hat{\theta})^2}{2\sigma^2_{\hat{\theta}}}$$

$$\theta \rightarrow \hat{\theta} \pm n\sigma_{\hat{\theta}}$$

$$\Delta \ln L(\hat{\theta} \pm n\sigma_{\hat{\theta}}) = -\frac{(\pm n\sigma_{\hat{\theta}})^2}{2\sigma^2_{\hat{\theta}}}$$

$$\boxed{\Delta \ln L(\hat{\theta} \pm n\sigma_{\hat{\theta}}) = -\frac{n^2}{2}}$$



$$\Delta \ln L(\hat{\theta} \pm n\sigma_{\hat{\theta}}) = -\frac{n^2}{2}$$

This is the most common definition of the 68% and 95% **confidence intervals**:

68%/ 1 $\sigma_{\hat{\theta}}$: $\Delta \ln L = -\frac{1}{2}$
95%/ 2 $\sigma_{\hat{\theta}}$: $\Delta \ln L = -2$

Recall that in the case that the PDF is Gaussian, the $\ln L$ is just the χ^2 statistic.

$$\ln L = -\frac{\chi^2}{2}, \qquad \chi^2 = \sum \frac{(x-\theta)^2}{\sigma^2}$$

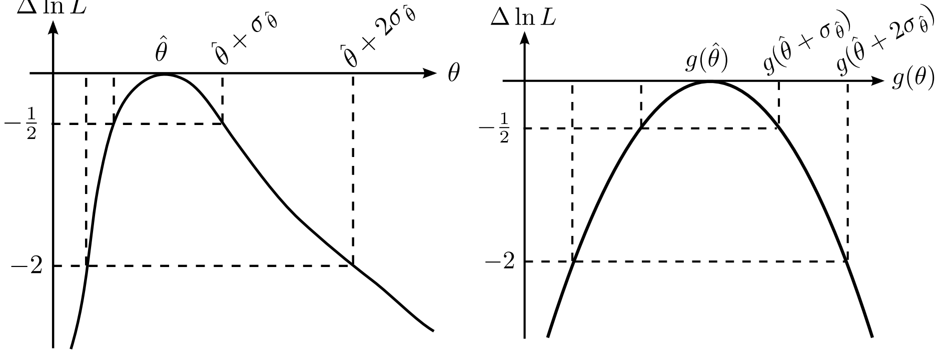
$$\Delta \ln L(\hat{\theta} \pm n\sigma_{\hat{\theta}}) = \ln L(\hat{\theta} \pm n\sigma_{\hat{\theta}}) - \ln L_{\max} = -\frac{n^2}{2}$$

$$\Rightarrow \qquad -\frac{1}{2}\left(\chi^2(\hat{\theta} \pm n\sigma_{\hat{\theta}}) - \chi^2_{\min}\right) = -\frac{n^2}{2}$$

$$\boxed{\Delta\chi^2(\hat{\theta} \pm n\sigma_{\hat{\theta}}) = n^2}$$

68%/ 1 $\sigma_{\hat{\theta}}$: $\Delta\chi^2 = 1$
95%/ 2 $\sigma_{\hat{\theta}}$: $\Delta\chi^2 = 4$

What if $L(\theta)$ is not Gaussian, *i.e.* $\ln L(\theta)$ is not parabolic?



Likelihood functions have an invariance property, such that if $g(x)$ is a monotonic function, then the maximum likelihood estimate of $g(\theta)$ is $g(\hat{\theta})$. In principle, one can find a change of variables function $g(\theta)$, for which the $\ln L(g(\theta))$ is parabolic as a function of $g(\theta)$. Therefore, using the invariance of the likelihood function, one can make inferences about a parameter of a non-Gaussian likelihood function *without actually finding such a transformation* [James p. 234].

[arXiv:0907.5568v1]

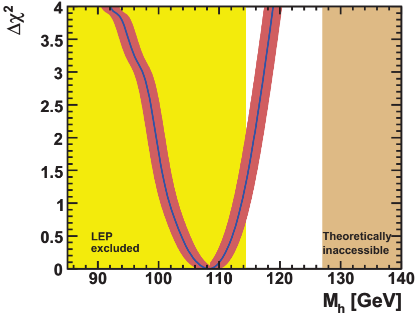


Figure 24. *The χ^2 functions for M_h in the CMSSM (left) and the NUHM1 (right), including the theoretical uncertainties (red bands). Also shown is the mass range excluded for a SM-like Higgs boson (yellow shading), and the ranges theoretically inaccessible in the supersymmetric models studied.*

Now you know: 68%/ 1 $\sigma_{\hat{\theta}}$: $\Delta\chi^2 = 1$, 95%/ 2 $\sigma_{\hat{\theta}}$: $\Delta\chi^2 = 4$

Multidimensional case:

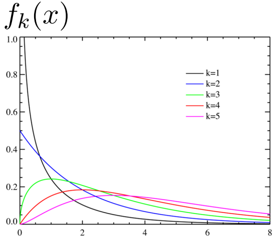
Assume $\hat{\vec{\theta}}$ is Gaussian distributed.

$$g(\hat{\vec{\theta}}|\vec{\theta}) = \frac{1}{(2\pi)^{n/2}|\mathbf{V}|^{1/2}} \exp\left(-\frac{1}{2}Q(\hat{\vec{\theta}},\vec{\theta})\right)$$

$$Q(\hat{\vec{\theta}},\vec{\theta}) \equiv (\hat{\vec{\theta}}-\vec{\theta})^{\mathrm{T}}\mathbf{V}^{-1}(\hat{\vec{\theta}}-\vec{\theta})$$

As in the one dimensional case, Q is distributed as a χ^2_n, with n degrees of freedom. The probability that we get a value of Q less than some constant Q_c is

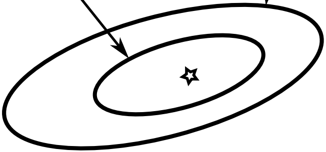
$$P(Q \leq Q_c) = \int_0^{Q_c} dx\, f_n(x) = F_n(Q_c) = c$$



where $f_n(x)$ is the PDF of χ^2_n, and F_n is its cumulative distribution.

$$\Rightarrow \qquad Q_c = F_n^{-1}(c)$$

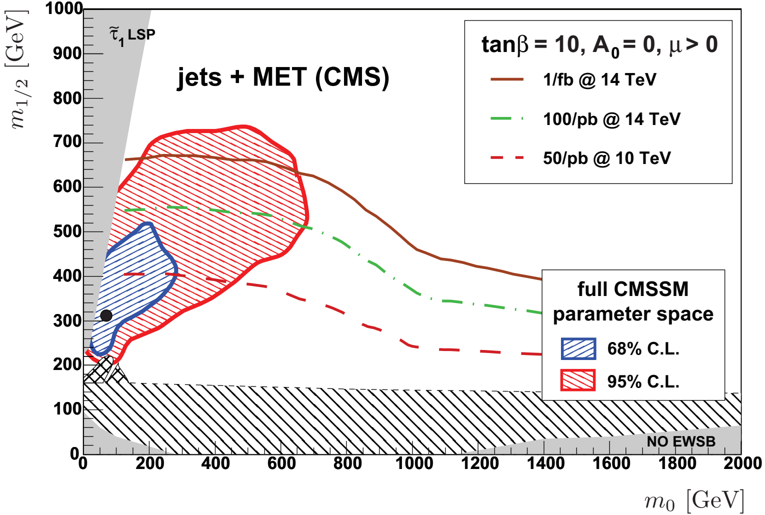
c	Q_c n=1	n=2	n=3
0.683	1.00	2.30	3.53
0.95	3.84	5.99	7.82
0.99	6.63	9.21	11.3



A note about confidence contours/intervals

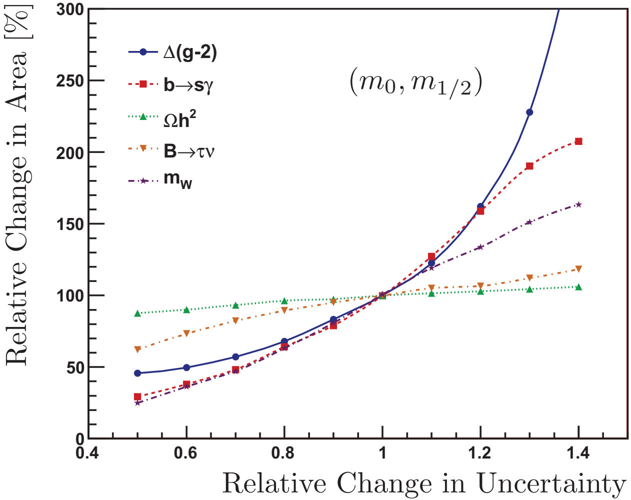
A 95% confidence contour doesn't mean that the true value of the parameter is in the contour with 95% probabilty. That would be a Bayesian probabilty (a probability of belief). It means that if the model is correct, we have properly estimated our errors, and if we were to repeat the experiment over and over again, each time creating a new contour, then the contour would cover the true parameter in 95% of the experiments (a frequentist probability).

CMSSM Contours and Exclusion



[arXiv:0808.4128v1]

Parameter Error Sensitivity



[arXiv:0808.4128v1]

$(g-2)_\mu$ and SUSY

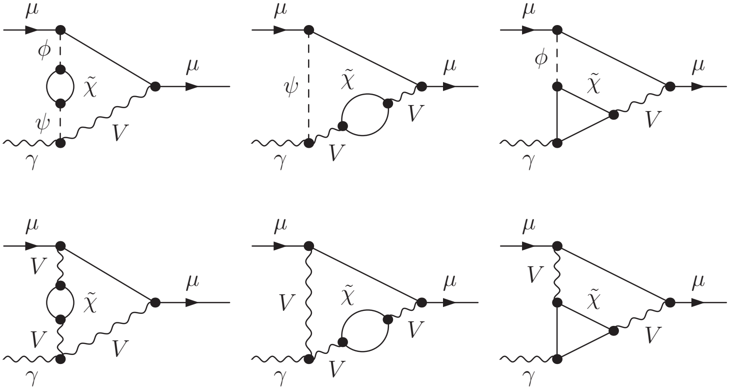


Figure 5: Generic two-loop diagrams to a_μ with a closed chargino/neutralino loop. ϕ, ψ denote the scalar particles h, H, A^0, $H^\pm$ and $G^{0,\pm}$; V denotes the vector bosons γ, Z, $W^\pm$; $\tilde{\chi}$ stands for the chargino/neutralino mass eigenstates $\tilde{\chi}^\pm_{1,2}$, $\tilde{\chi}^0_{1,2,3,4}$.

$a_\mu^{\mathrm{exp}} - a_\mu^{\mathrm{theo}} = (24.5 \pm 9.0) \times 10^{-10} : 2.7\sigma$ [hep-ph/0405255v1]

References

- Buchmueller, O., Cavanaugh, R., Roeck, D., Ellis, R., Fl, H., Heinemeyer, S., et al. Predictions for Supersymmetric Particle Masses using Indirect Experimental and Cosmological Constraints. (2008). [arXiv:0808.4128v1]
- Buchmueller, O., Cavanaugh, R., Roeck, A. D., Ellis, J. R., Heinemeyer, S., Isidori, G., et al. Likelihood Functions for Supersymmetric Observables in Frequentist Analyses of the CMSSM and NUHM1. (2009). [arXiv:0907.5568v1]
- Heinemeyer, S., Stckinger, D., & Weiglein, G. Electroweak and supersymmetric two-loop corrections to $(g-2)_\mu$. (2004). [hep-ph/0405255v1]
- Cowan, G. Statistical Data Analysis. (1998). Oxford University Press.
- James, F. Statistical Methods in Experimental Physics, 2nd Ed. (2006). World Scientific.