Asymptotic Formulae

Aaron Armbruster

University of Michigan

On behalf of ATLAS

February 13, 2013

Motivation

- Frequentist approach to constructing confidence intervals involves large numbers of pseudo-experiments to build test statistic distributions
- Several likelihood fits per toy becomes computationally expensive
  - Millions of CPU hours for large models like Higgs
- Asymptotics aims to analytically describe test statistic distributions under some approximations
  - Millions of CPU hours$\rightarrow$ several CPU minutes

Definitions

- Parameter of Interest μ, usually thought of as signal strength parameter
  - $\mu = 0 \Leftrightarrow H_0$
  - $\mu = 1 \Leftrightarrow H_1$
- Profile likelihood ratio $\lambda(\mu) = \frac{L(\mu,\hat{\hat{\theta}}_\mu)}{L(\hat{\mu},\hat{\theta})}$
  - Numerator maximizes L conditionally over θ for a particular μ
  - Denominator maximizes L unconditionally over all parameters μ, θ
- Alternative likelihood ratio for $\mu \geq 0$

$$\tilde{\lambda}(\mu) = \begin{cases} \frac{L(\mu,\hat{\hat{\theta}}_\mu)}{L(\hat{\mu},\hat{\theta})}, & \hat{\mu} \geq 0 \\ \frac{L(\mu,\hat{\hat{\theta}}_\mu)}{L(0,\hat{\hat{\theta}}_0)}, & \hat{\mu} < 0 \end{cases} \quad (1)$$

Common Test Statistic Variations

- Two sided "Feldman-Cousins" confidence interval: $t_\mu = -2\ln\lambda(\mu)$
- Upper limit: q_μ

$$q_\mu = \begin{cases} -2\ln\lambda(\mu), & \hat{\mu} \leq \mu \\ 0, & \hat{\mu} > \mu \end{cases} \tag{2}$$

- Upper limit with $\mu \geq 0$: $\tilde{q}_\mu$

$$\tilde{q}_\mu = \begin{cases} -2\ln\tilde{\lambda}(\mu), & \hat{\mu} \leq \mu \\ 0, & \hat{\mu} > \mu \end{cases} \tag{3}$$

- Discovery test statistic: q_0

$$q_0 = \begin{cases} -2\ln\lambda(0), & \hat{\mu} > 0 \\ 0, & \hat{\mu} < 0 \end{cases} \tag{4}$$

Signed ("uncapped") Test Statistics

- Upper limit with $\mu \geq 0$: $\tilde{r}_\mu$

$$\tilde{r}_\mu = \begin{cases} -2\ln\tilde{\lambda}(\mu), & \hat{\mu} \leq \mu \\ +2\ln\tilde{\lambda}(\mu), & \hat{\mu} > \mu \end{cases} \tag{5}$$

- Discovery test statistic: r_0

$$r_0 = \begin{cases} -2\ln\lambda(0), & \hat{\mu} > 0 \\ +2\ln\lambda(0), & \hat{\mu} < 0 \end{cases} \tag{6}$$

Definitions

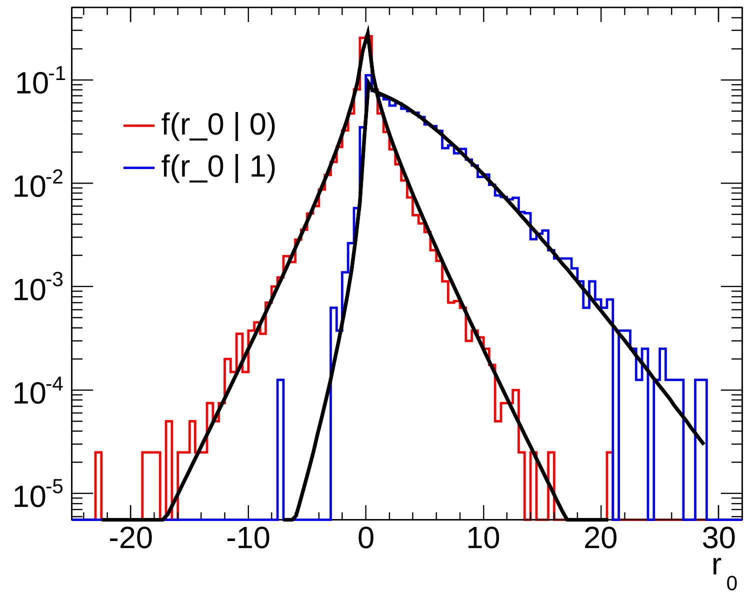
- $f(Q_\mu|\mu')$: Distribution of some test statistic Q_μ given the μ' hypothesis
- Asimov Dataset $A_{\mu'}$: Median dataset under μ' hypo, such that $Q_{\mu,A_{\mu'}} \sim \mathrm{med}\{Q_\mu|\mu'\}$
  - E.g., $N_{A_{\mu'}} = \mu' S + B$ for single bin
- Must make a choice of nuisance parameters $\hat{\theta}$ at which $A_{\mu'}$ is constructed
  - Toy procedure distributes global observables $\tilde{\theta}$ around $\hat{\theta}(\mu')$ through auxiliary constraint $\mathcal{A}(\tilde{\theta}|\hat{\theta})$
  - To be consistent, $\hat{\theta}(\mu')$ is used
  - Global observables $\tilde{\theta} = \hat{\theta}(\mu')$ must then be used when dealing with $A_{\mu'}$ to recover $\hat{\mu} = \mu'$
- Continuous likelihoods can be finely binned when constructing $A_{\mu'}$ with negligible bias

Approximations

- Wald: $t_\mu = -2\ln\lambda(\mu) = \frac{(\mu-\hat{\mu})^2}{\sigma^2} + \mathcal{O}(1/\sqrt{N})$
  - $-2\ln\lambda(\mu)$ is approximately parabolic for large sample size N
- Wilks: If there is a one-to-one correspondence between $\hat{\mu}$ and t_μ, and $\hat{\mu}$ is gaussian distributed, then t_μ is distributed as a non-central χ^2 with one degree of freedom
  - $f(t_\mu;\Lambda) = \frac{1}{2\sqrt{2\pi t_\mu}}[\exp(-\frac{1}{2}(\sqrt{t_\mu} + \frac{\mu-\mu'}{\sigma})^2) + \exp(-\frac{1}{2}(\sqrt{t_\mu} - \frac{\mu-\mu'}{\sigma})^2)]$
  - $\frac{(\mu-\mu')^2}{\sigma^2} \approx t_{\mu,A_{\mu'}}$ can be estimated by evaluating t_μ on $A_{\mu'}$
  - Similar formula for other test statistic variations
- Cumulative functions are straight forward

$$F(t_\mu|\mu') = \Phi(\sqrt{t_\mu} + \frac{\mu-\mu'}{\sigma}) + \Phi(\sqrt{t_\mu} - \frac{\mu-\mu'}{\sigma}) - 1 \quad (7)$$

Quantifying an Excess



- p-value $p_0 = \int_{r_{0,obs}}^{\infty} f(r_0|0)\mathrm{d}r_0$ used for discovery

$$p_0 = \begin{cases} 1 - \Phi(\sqrt{r_0}), & r_\mu \geq 0 \\ 1 - \Phi(-\sqrt{-r_0}), & r_\mu < 0 \end{cases} \quad (8)$$

- Expected p-value is $p_{0,\mathrm{exp}} = 1 - \Phi(\sqrt{r_{0,A_1}})$

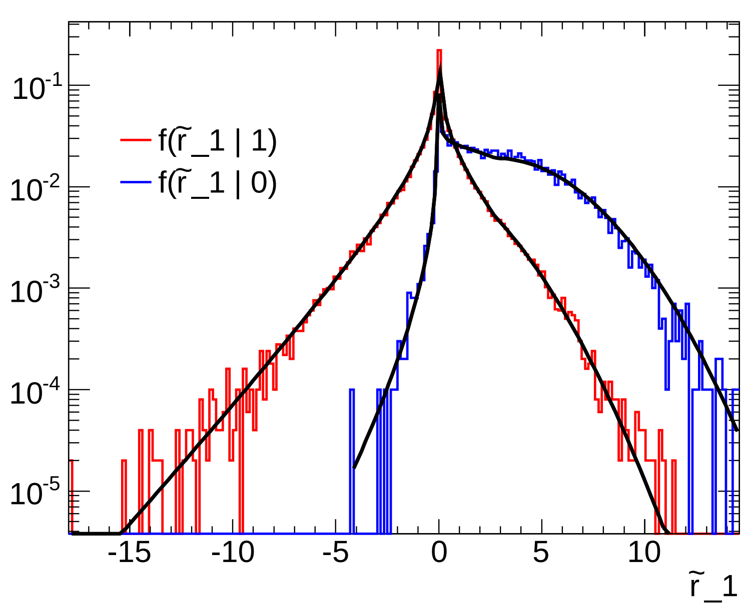
Upper Limits

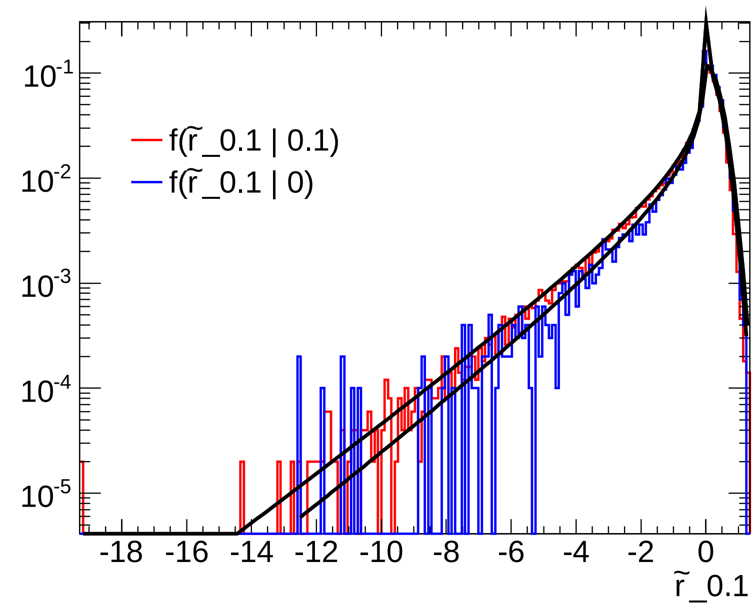
- Two p-values p_μ and p_b used for upper limits, with CLs = $\frac{p_\mu}{1-p_b}$

$$p_\mu = \int_{\tilde{r}_{\mu,obs}}^{\infty} f(\tilde{r}_\mu|\mu)\mathrm{d}\tilde{\mathrm{r}}_\mu = \begin{cases} 1-\Phi(-\sqrt{-\tilde{r}_\mu}), & \tilde{r}_\mu < 0 \\ 1-\Phi(\sqrt{\tilde{r}_\mu}), & 0 \leq \tilde{r}_\mu \leq \mu^2/\sigma^2 \\ 1-\Phi(\frac{\tilde{r}_\mu+\mu^2/\sigma^2}{2\mu/\sigma}), & \mu^2/\sigma^2 < \tilde{r}_\mu \end{cases}$$

$$p_b = \int_{-\infty}^{\tilde{r}_{\mu,obs}} f(\tilde{r}_\mu|0)\mathrm{d}\tilde{\mathrm{r}}_\mu = \begin{cases} 1-\Phi(-\sqrt{-\tilde{r}_\mu}-\frac{\mu}{\sigma}), & \tilde{r}_\mu < 0 \\ 1-\Phi(\sqrt{\tilde{r}_\mu}-\frac{\mu}{\sigma}), & 0 \leq \tilde{r}_\mu \leq \mu^2/\sigma^2 \\ 1-\Phi(\frac{\tilde{r}_\mu-\mu^2/\sigma^2}{2\mu/\sigma}), & \mu^2/\sigma^2 < \tilde{r}_\mu \end{cases} \quad (9)$$

Upper Limits





- 95% Upper limit μ_{up} found by scanning CLs to find $CLs(\mu_{up}) = 0.05$
- Toys distributions, and therefore p-values and μ_{up}, are very well modeled

Limit Bands

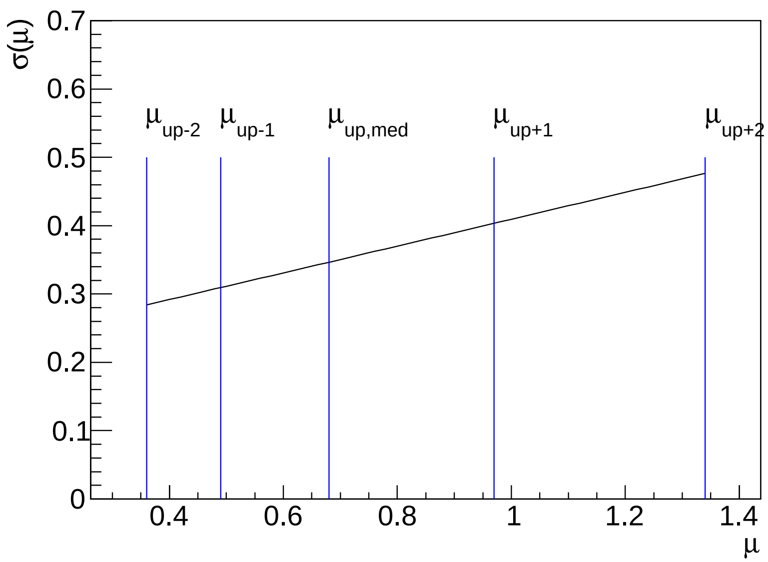
- Expected bands rely on describing distribution of upper limits $f(\mu_{up}|0)$ under b-only hypothesis
  - Relies on one-to-one mapping function $M : f(\hat{\mu}|0) \rightarrow f(\mu_{up}|0)$
- At N'th quantile, a b-only experiment with have a characteristic $\hat{\mu} = N\sigma$
- Test statistic will be $q_{\mu_{up+N}} = (\frac{\mu_{up+N} - N\sigma}{\sigma})^2$

$$\begin{aligned}
\mathrm{CLs}(\mu_{\mathrm{up+N}}) \quad \equiv \quad \alpha \quad & \approx \quad (1 - \Phi(\tfrac{\mu_{up+n} - N\sigma}{\sigma}))/(\Phi(\tfrac{\mu_{up+N}}{\sigma} - \tfrac{\mu_{up+N} - N\sigma}{\sigma})) \\
& = \quad (1 - \Phi(\tfrac{\mu_{up+n} - N\sigma}{\sigma}))/(\Phi(N)) \\
\Rightarrow \quad \mu_{up+N} \quad & = \quad \sigma\{\Phi^{-1}[1 - \alpha\Phi(N)] + N\}
\end{aligned} \tag{10}$$

- Derived using q_{μ} to simplify algebra, but $\tilde{r}_{\mu}$ gives equivalent answer

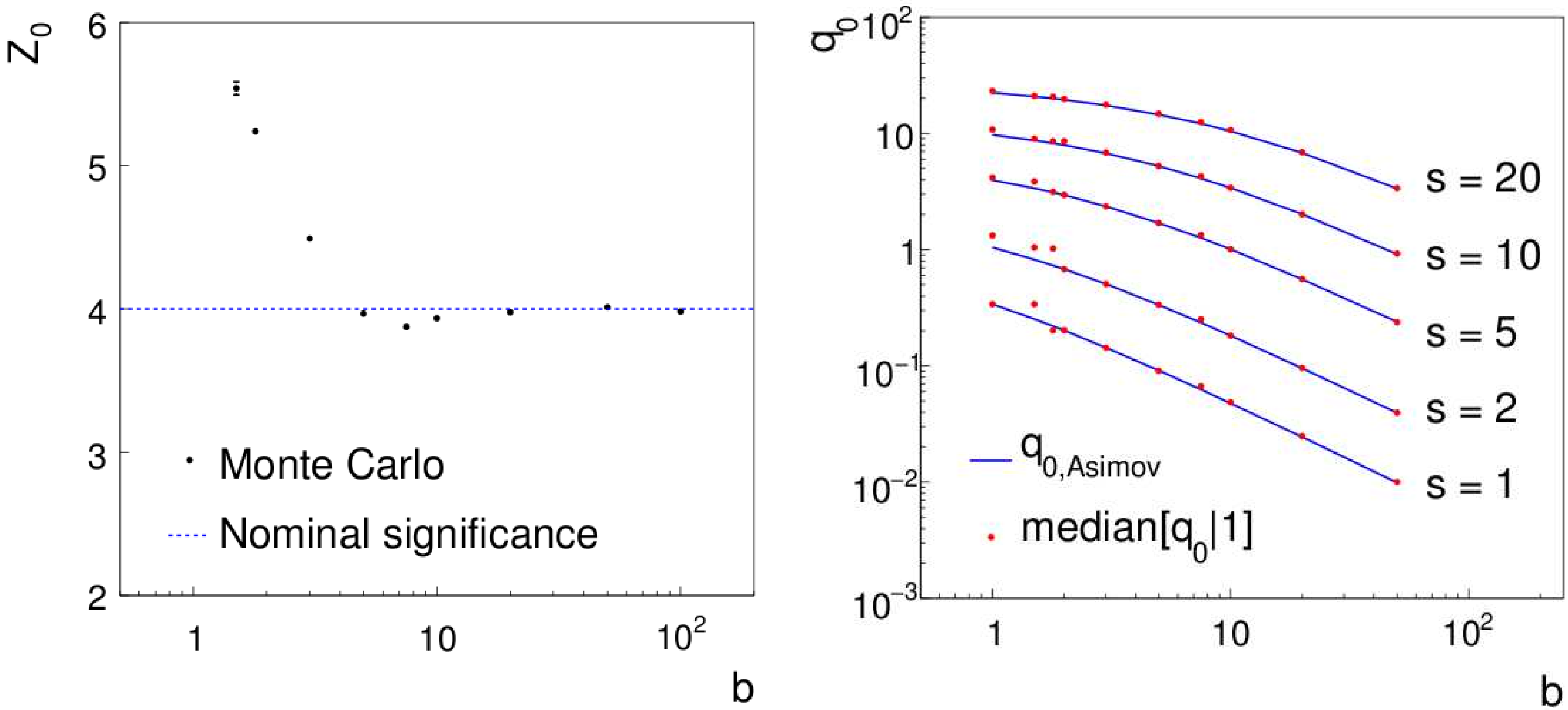
Limit Bands

- Denominator of CLs in derivation is $\Phi(\frac{\mu_{up+N}}{\sigma} - \frac{\mu_{up+N} - N\sigma}{\sigma})$
  - If σ is μ-independent, this reduces exactly to $\Phi(N)$
  - Otherwise, the eq is transcendental in μ_{up+N} and approximation fails
- In general, σ tends to have a slightly linear μ dependence



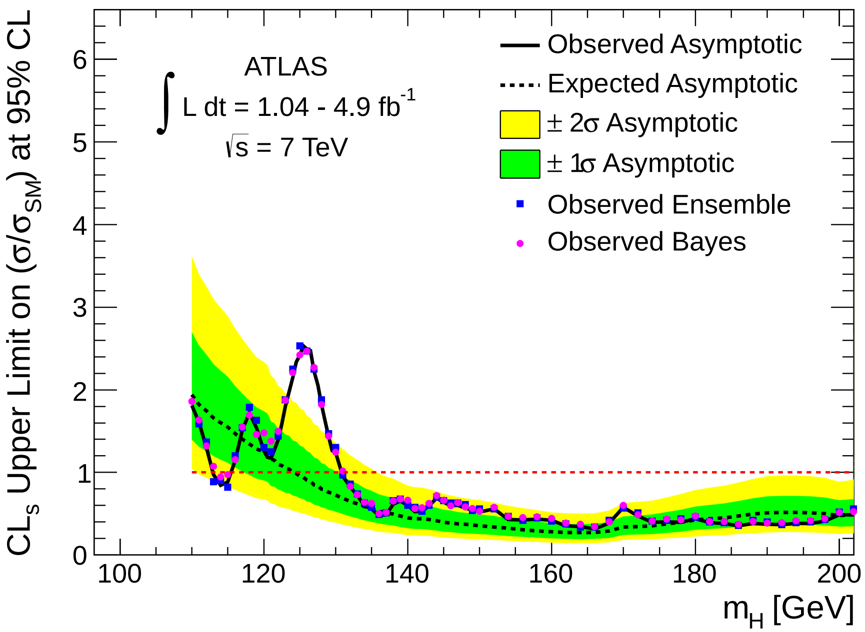
- More in depth discussion in "Discovery Experience" slides

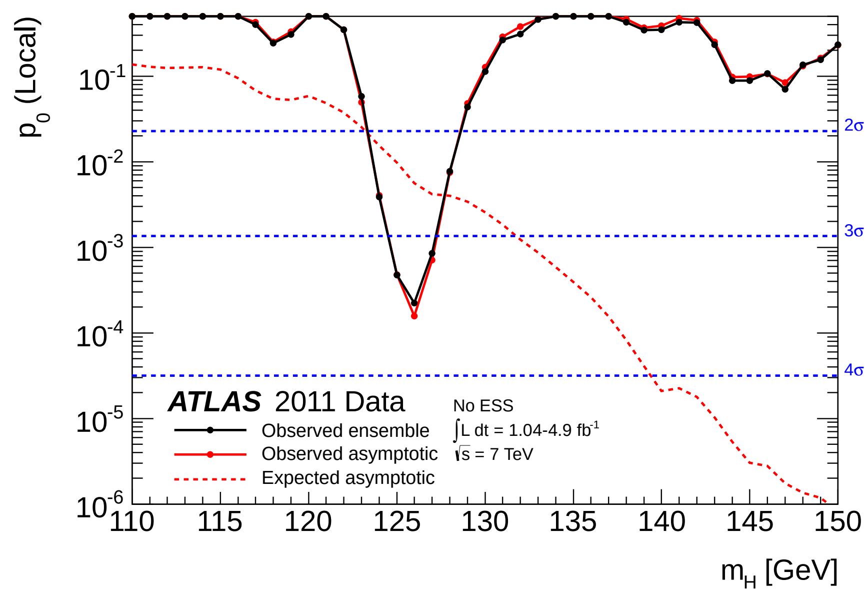
Validity vs Sample Size



- Tests vs toy MC show that asymptotic formulae hold well for $N_{exp} > 10$
- Asimov property holds even better, for $N_{exp} > 3$

Asymptotics in Higgs Search





- In Higgs model, asymptotics agrees well with toys
- Remember that toys have some statistical uncertainty associated with them, esp for small p-values

Summary

- Analytical formula can be used to approximate distribution of test statistic and p-values
  - Holds well in large N limit ($N > \mathcal{O}(10)$)
- Used within ATLAS for discovery, with some corrections for LEE
- Observed and median expected are straight forward. Bands require some second order corrections
- For more information
  - http://arxiv.org/abs/1007.1727 - Asimov paper
  - http://inspirehep.net/record/1193351 - Asymptotic distributions with upper and lower boundaries on POI