Introduction to HEP statistical data analysis

Shihan Zhao

School of Physics, SYSU

2026-07-20

Sino-Thai International Summer School on HEP and Astrophysics in 2026

See the full lecture notes for details and code examples.

Outline

Why Statistics in Particle Physics?

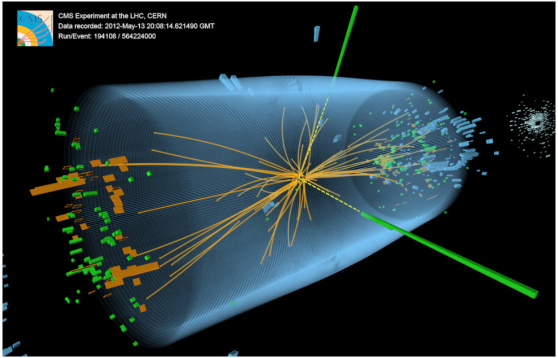
- Quantum behavior is intrinsically **random**
- Statistical methods are **irreplaceable** in particle physics

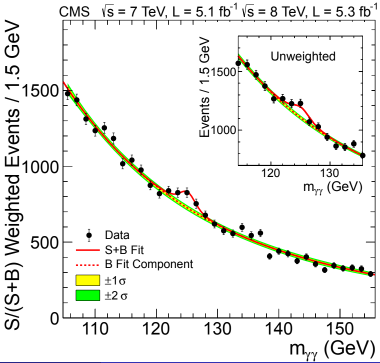
Why Statistics in Particle Physics?

- Quantum behavior is intrinsically **random**
- Statistical methods are **irreplaceable** in particle physics
- Two experiment types:
  - **Search / test**: hypothesis testing (e.g. Higgs discovery)
  - **Measurement**: parameter estimation (e.g. muon lifetime)

Why Statistics in Particle Physics?

- Quantum behavior is intrinsically **random**
- Statistical methods are **irreplaceable** in particle physics
- Two experiment types:
  - **Search / test**: hypothesis testing (e.g. Higgs discovery)
  - **Measurement**: parameter estimation (e.g. muon lifetime)





This Lecture's Goal

- Understand **fundamental statistical applications** in particle physics

This Lecture's Goal

- Understand **fundamental statistical applications** in particle physics
- Master two core methods:
  - **Maximum Likelihood Estimation** — parameter estimation
  - **Likelihood Ratio Test** — hypothesis testing

This Lecture's Goal

- Understand **fundamental statistical applications** in particle physics
- Master two core methods:
  - **Maximum Likelihood Estimation** — parameter estimation
  - **Likelihood Ratio Test** — hypothesis testing
- Learn how statistics bridges **raw data** and **physics conclusions**
- Appreciate the **rigor** required in experimental data analysis

A simple hands-on homework will follow after the lecture — data files provided.

Particle Decay & Breit–Wigner Distribution

Particle decay:

- Decay time exponential: $f(t) = \frac{1}{\tau} e^{-t/\tau}$
- Lifetime $\tau = \langle t \rangle$ is the expectation

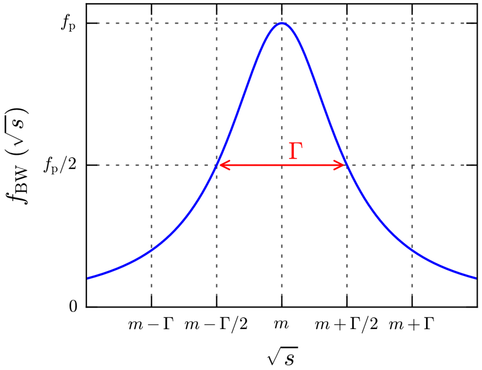
Particle Decay & Breit–Wigner Distribution

Particle decay:

- Decay time exponential: $f(t) = \frac{1}{\tau} e^{-t/\tau}$
- Lifetime $\tau = \langle t \rangle$ is the expectation

Breit–Wigner (mass broadening):

- Unstable particle mass not fixed
- $f(\sqrt{s}) = \dfrac{\Gamma}{2\pi} \dfrac{1}{(\sqrt{s} - m)^2 + \Gamma^2/4}$
- Γ = full width at half maximum
- $\Gamma = \hbar/\tau$ (width $\propto 1/$lifetime)



Counting Experiments & Poisson Distribution

- **Counting**: maybe the most important method in particle physics

Counting Experiments & Poisson Distribution

- **Counting**: maybe the most important method in particle physics
- Poisson distribution:

$$P(k=N) = \frac{n^N}{N!} e^{-n}$$

- Conditions: fixed expectation, independence between events

Counting Experiments & Poisson Distribution

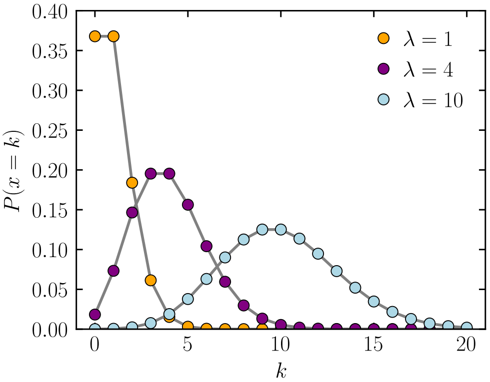
- **Counting**: maybe the most important method in particle physics
- Poisson distribution:

$$P(k = N) = \frac{n^N}{N!} e^{-n}$$

- Conditions: fixed expectation, independence between events
- **Asymptotic property** (large n, e.g. $n \gtrsim 20$):

$$\mathrm{Pois}(n) \to \mathcal{N}(n, \sqrt{n})$$

- **Uncertainty of a count N is $\sqrt{N}$!**



Maximum Likelihood Estimation

- **Core idea**: choose θ that maximizes the probability of observing the data

Maximum Likelihood Estimation

- **Core idea**: choose θ that maximizes the probability of observing the data
- **Likelihood function** (product of individual PDFs):

$$L(\boldsymbol{\theta}; \mathbf{x}) = \prod_{k=1}^{N} f(x_k; \boldsymbol{\theta})$$

- **Log-likelihood function** (sum is easier to work with):

$$\boxed{\log L(\boldsymbol{\theta}; \mathbf{x}) = \sum_{k=1}^{N} \log f(x_k; \boldsymbol{\theta})}$$

Maximum Likelihood Estimation

- **Core idea**: choose $\boldsymbol{\theta}$ that maximizes the probability of observing the data
- **Likelihood function** (product of individual PDFs):

$$L(\boldsymbol{\theta}; \mathbf{x}) = \prod_{k=1}^{N} f(x_k; \boldsymbol{\theta})$$

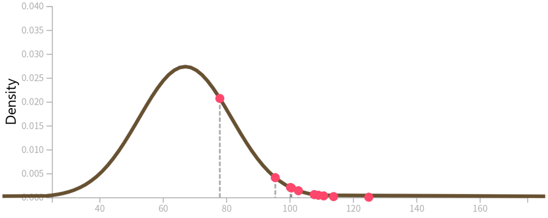
- **Log-likelihood function** (sum is easier to work with):

$$\boxed{\log L(\boldsymbol{\theta}; \mathbf{x}) = \sum_{k=1}^{N} \log f(x_k; \boldsymbol{\theta})}$$

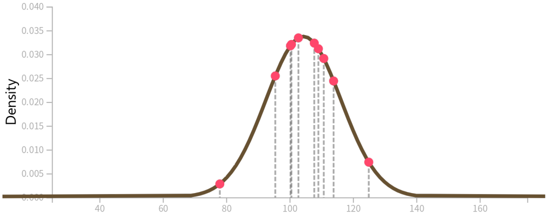
- **Maximum likelihood estimator**:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \log L(\boldsymbol{\theta}; \mathbf{x})$$

Maximum Likelihood Estimation



Small likelihood — poor fit



Maximum likelihood — best fit

- Try this at `https://rpsychologist.com/likelihood/`

Extended Maximum Likelihood Estimation

- In many analyses, the **total event count** n is also a random variable
- We need to fit both the shape parameters $\boldsymbol{\theta}$ *and* n_{sig}, n_{bkg}

Extended Maximum Likelihood Estimation

- In many analyses, the **total event count** n is also a random variable
- We need to fit both the shape parameters $\boldsymbol{\theta}$ *and* n_{sig}, n_{bkg}
- **Extended likelihood function** (product of Poisson count $\times$ shape):

$$L(n, \boldsymbol{\theta}; N, \mathbf{x}) = \frac{n^N}{N!} e^{-n} \prod_{k=1}^{N} f(x_k; \boldsymbol{\theta})$$

- **Extended log-likelihood** (drop constant $N!$):

$$\boxed{\log L(n, \boldsymbol{\theta}; N, \mathbf{x}) = -n + N \log n + \sum_{k=1}^{N} \log f(x_k; \boldsymbol{\theta})}$$

Extended Maximum Likelihood Estimation

- In many analyses, the **total event count** n is also a random variable
- We need to fit both the shape parameters $\boldsymbol{\theta}$ *and* n_{sig}, n_{bkg}
- **Extended likelihood function** (product of Poisson count $\times$ shape):

$$L(n, \boldsymbol{\theta}; N, \mathbf{x}) = \frac{n^N}{N!} e^{-n} \prod_{k=1}^{N} f(x_k; \boldsymbol{\theta})$$

- **Extended log-likelihood** (drop constant $N!$):

$$\boxed{\log L(n, \boldsymbol{\theta}; N, \mathbf{x}) = -n + N \log n + \sum_{k=1}^{N} \log f(x_k; \boldsymbol{\theta})}$$

- **Signal+background composite model**:

$$\rho(x) = n_{\text{sig}} f_{\text{sig}}(x) + n_{\text{bkg}} f_{\text{bkg}}(x)$$

- In RooFit: `RooAddPdf` automatically uses extended MLE

From Binned Likelihood to χ^2 Fit

Binned likelihood:

- Histogram with M bins: $\boldsymbol{n} = (n_1, \ldots, n_M)$
- Expected counts: $\boldsymbol{\mu}(\boldsymbol{\theta}) = (\mu_1, \ldots, \mu_M)$
- Each bin: $n_i \sim \mathrm{Pois}(\mu_i)$, independent

From Binned Likelihood to χ^2 Fit

Binned likelihood:

- Histogram with M bins: $\boldsymbol{n} = (n_1, \ldots, n_M)$
- Expected counts: $\boldsymbol{\mu}(\boldsymbol{\theta}) = (\mu_1, \ldots, \mu_M)$
- Each bin: $n_i \sim \mathrm{Pois}(\mu_i)$, independent

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{M} \big(n_i \log \mu_i(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta}) \big)$$

From Binned Likelihood to χ^2 Fit

Binned likelihood:

- Histogram with M bins: $\boldsymbol{n} = (n_1, \ldots, n_M)$
- Expected counts: $\boldsymbol{\mu}(\boldsymbol{\theta}) = (\mu_1, \ldots, \mu_M)$
- Each bin: $n_i \sim \mathrm{Pois}(\mu_i)$, independent

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{M} \big(n_i \log \mu_i(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta}) \big)$$

Large-n_i limit ($n_i \gtrsim 20$):

- Poisson $\to$ Gaussian: $n_i \sim \mathcal{N}(\mu_i, \sigma_i)$, $\sigma_i^2 \approx \mu_i \approx n_i$

From Binned Likelihood to χ^2 Fit

Binned likelihood:

- Histogram with M bins: $\boldsymbol{n} = (n_1, \ldots, n_M)$
- Expected counts: $\boldsymbol{\mu}(\boldsymbol{\theta}) = (\mu_1, \ldots, \mu_M)$
- Each bin: $n_i \sim \mathrm{Pois}(\mu_i)$, independent

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{M} \big(n_i \log \mu_i(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta})\big)$$

Large-n_i limit $(n_i \gtrsim 20)$:

- Poisson $\rightarrow$ Gaussian: $n_i \sim \mathcal{N}(\mu_i, \sigma_i)$, $\sigma_i^2 \approx \mu_i \approx n_i$

$$\boxed{\log L(\boldsymbol{\theta}) \;\approx\; -\frac{1}{2} \sum_{i=1}^{M} \frac{(n_i - \mu_i(\boldsymbol{\theta}))^2}{\sigma_i^2}}$$

From Binned Likelihood to χ^2 Fit

Binned likelihood:

- Histogram with M bins: $\boldsymbol{n} = (n_1, \ldots, n_M)$
- Expected counts: $\boldsymbol{\mu}(\boldsymbol{\theta}) = (\mu_1, \ldots, \mu_M)$
- Each bin: $n_i \sim \mathrm{Pois}(\mu_i)$, independent

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{M} \left(n_i \log \mu_i(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta})\right)$$

Large-n_i limit $(n_i \gtrsim 20)$:

- Poisson $\to$ Gaussian: $n_i \sim \mathcal{N}(\mu_i, \sigma_i)$, $\sigma_i^2 \approx \mu_i \approx n_i$

$$\boxed{\log L(\boldsymbol{\theta}) \approx -\frac{1}{2} \sum_{i=1}^{M} \frac{(n_i - \mu_i(\boldsymbol{\theta}))^2}{\sigma_i^2}}$$

Therefore:

$$\boxed{\log L \approx -\frac{1}{2}\chi^2}$$

Maximize $\log L \iff$ **minimize** χ^2

This is why least-squares fitting works for histograms with sufficient statistics.

From Binned Likelihood to χ^2 Fit

Binned likelihood:

- Histogram with M bins: $\boldsymbol{n} = (n_1, \ldots, n_M)$
- Expected counts: $\boldsymbol{\mu}(\boldsymbol{\theta}) = (\mu_1, \ldots, \mu_M)$
- Each bin: $n_i \sim \mathrm{Pois}(\mu_i)$, independent

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{M} \left(n_i \log \mu_i(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta}) \right)$$

Large-n_i limit $(n_i \gtrsim 20)$:

- Poisson $\to$ Gaussian: $n_i \sim \mathcal{N}(\mu_i, \sigma_i)$, $\sigma_i^2 \approx \mu_i \approx n_i$

$$\boxed{\log L(\boldsymbol{\theta}) \;\approx\; -\frac{1}{2} \sum_{i=1}^{M} \frac{(n_i - \mu_i(\boldsymbol{\theta}))^2}{\sigma_i^2}}$$

Therefore:

$$\boxed{\log L \approx -\frac{1}{2}\chi^2}$$

Maximize $\log L \iff$ **minimize** χ^2

This is why least-squares fitting works for histograms with sufficient statistics.

Common χ^2 variants:

- **Pearson** χ^2: $\displaystyle\sum_i \frac{(n_i - \mu_i)^2}{\mu_i}$
- **Neyman** χ^2: $\displaystyle\sum_i \frac{(n_i - \mu_i)^2}{n_i}$

Information matrix (HESSE):

- $\hat{V} = \hat{\mathcal{I}}^{-1}$, $\sigma_i = \sqrt{V_{ii}}$
- Symmetric CI: $\hat{\theta}_i \pm \sigma_i$
- Fast — assumes Gaussian asymptotics

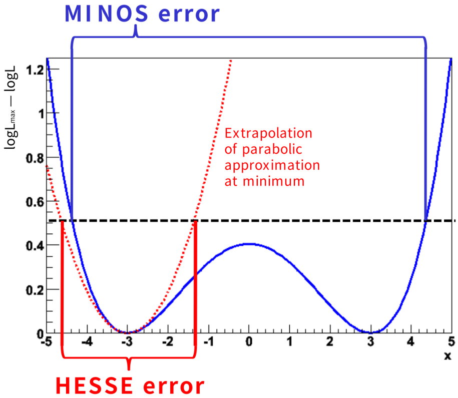
Uncertainty Estimation

Information matrix (HESSE):

- $\hat{V} = \hat{\mathcal{I}}^{-1}$, $\sigma_i = \sqrt{V_{ii}}$
- Symmetric CI: $\hat{\theta}_i \pm \sigma_i$
- Fast — assumes Gaussian asymptotics

Profile method (MINOS):

- $\lambda(\theta_i) = -2\left(\max_{\theta_i=\text{fixed}} \log L - \max \log L\right)$
- Asymmetric: $\hat{\theta}_i {}^{+\sigma_i^+}_{-\sigma_i^-}$
- More accurate for small N, computationally expensive



Reporting Measurement Results

- **Symmetric CI**: $\hat{\theta}_i \pm \hat{\sigma}_{\theta_i}$ (HESSE, 68.3% CL)
- **Asymmetric CI**: $\hat{\theta}_i{}^{+\hat{\sigma}^{+}_{\theta_i}}_{-\hat{\sigma}^{-}_{\theta_i}}$ (MINOS, 68.3% CL)

Reporting Measurement Results

- **Symmetric CI**: $\hat{\theta}_i \pm \hat{\sigma}_{\theta_i}$ (HESSE, 68.3% CL)
- **Asymmetric CI**: $\hat{\theta}_i{}^{+\hat{\sigma}^+_{\theta_i}}_{-\hat{\sigma}^-_{\theta_i}}$ (MINOS, 68.3% CL)
- **Significant figures**: uncertainty with 1–2 significant figures, value rounded to match
  - $\tau_\mu = 2197.81 \pm 0.66$ ns
  - J/ψ mass $= 3096.900 \pm 0.006$ MeV
- **Scientific notation**: power of ten / percent sign *outside*
  - $(g_\mu - 2)/2 = (11659205.9 \pm 2.2) \times 10^{-10}$
  - $D^+ \to K^-\pi^+\pi^+$ B.F. $= (9.38 \pm 0.16)\%$

Reporting Measurement Results

- **Symmetric CI**: $\hat{\theta}_i \pm \hat{\sigma}_{\theta_i}$ (HESSE, 68.3% CL)
- **Asymmetric CI**: $\hat{\theta}_i{}^{+\hat{\sigma}^+_{\theta_i}}_{-\hat{\sigma}^-_{\theta_i}}$ (MINOS, 68.3% CL)
- **Significant figures**: uncertainty with 1–2 significant figures, value rounded to match
  - $\tau_\mu = 2197.81 \pm 0.66$ ns
  - J/ψ mass $= 3096.900 \pm 0.006$ MeV
- **Scientific notation**: power of ten / percent sign *outside*
  - $(g_\mu - 2)/2 = (11659205.9 \pm 2.2) \times 10^{-10}$
  - $D^+ \to K^-\pi^+\pi^+$ B.F. $= (9.38 \pm 0.16)\%$
- **Stat. + sys. uncertainties**: report separately
  - $m_H = 125.3 \pm 0.4\,(\text{stat.}) \pm 0.5\,(\text{sys.})$ GeV
  - Compact: $125.3(4)(5)$ GeV
- **Notice**: significant-figure rounding only at the very end

ROOT & RooFit at a Glance

- **ROOT**: CERN's open-source data analysis framework for HEP
- **RooFit**: statistical library for PDF modeling, fitting, uncertainty estimation

ROOT & RooFit at a Glance

- **ROOT**: CERN's open-source data analysis framework for HEP
- **RooFit**: statistical library for PDF modeling, fitting, uncertainty estimation
- Core workflow:

$$\texttt{RooRealVar} \to \texttt{RooDataSet} \to \texttt{RooAbsPdf} \to \texttt{fitTo()} \to \text{plot}$$

ROOT & RooFit at a Glance

- **ROOT**: CERN's open-source data analysis framework for HEP
- **RooFit**: statistical library for PDF modeling, fitting, uncertainty estimation
- Core workflow:

  `RooRealVar` → `RooDataSet` → `RooAbsPdf` → `fitTo()` → plot

- Uncertainty estimation: **HESSE** (default, information matrix) vs **MINOS** (profile method)
- See: `https://root.cern/primer/` and `https://root.cern/manual/roofit/`

Hypothesis Testing Framework

Question

Two counting measurements:
$\hat{r}_1 = 0.3441 \pm 0.0098\ \mathsf{s}^{-1}$, $\hat{r}_2 = 0.314 \pm 0.013\ \mathsf{s}^{-1}$.
Can we conclude the count rates differ?

Hypothesis Testing Framework

Question

Two counting measurements:
$\hat{r}_1 = 0.3441 \pm 0.0098\ \mathrm{s}^{-1}$, $\hat{r}_2 = 0.314 \pm 0.013\ \mathrm{s}^{-1}$.
Can we conclude the count rates differ?

- **Core idea**: low-probability events do not occur (proof by contradiction)
- H_0 (null hypothesis): what we try to falsify
  - e.g. $r_1 = r_2$ or $n_{\mathrm{sig}} = 0$
- H_1 (alternative hypothesis): what we try to prove
  - e.g. $r_1 \neq r_2$ or $n_{\mathrm{sig}} > 0$

Hypothesis Testing Framework

Question

Two counting measurements:
$\hat{r}_1 = 0.3441 \pm 0.0098\ \mathrm{s}^{-1}$, $\hat{r}_2 = 0.314 \pm 0.013\ \mathrm{s}^{-1}$.
Can we conclude the count rates differ?

- **Core idea**: low-probability events do not occur (proof by contradiction)
- H_0 (null hypothesis): what we try to falsify
  - e.g. $r_1 = r_2$ or $n_{\mathrm{sig}} = 0$
- H_1 (alternative hypothesis): what we try to prove
  - e.g. $r_1 \neq r_2$ or $n_{\mathrm{sig}} > 0$
- **Test statistic** λ: measures extremeness of data under H_0
- **Acceptance region** vs **rejection region** $\rightarrow$ significance level α

Likelihood Ratio Test: Step by Step

1. **Formulate hypotheses**
   - $H_0 : \boldsymbol{\theta} \in \omega$ (constrained), $\quad H_1 : \boldsymbol{\theta} \in \Omega \setminus \omega$
   - e.g. $H_0 : r_1 = r_2, \;\; H_1 : r_1 \neq r_2$
2. **Fit both models & extract log-likelihood maxima**

$$\log L_0 = \max_{\boldsymbol{\theta} \in \omega} \log L(\boldsymbol{\theta}), \qquad \log L_1 = \max_{\boldsymbol{\theta} \in \Omega} \log L(\boldsymbol{\theta})$$

Likelihood Ratio Test: Step by Step

1. **Formulate hypotheses**
   - $H_0 : \boldsymbol{\theta} \in \omega$ (constrained), $\quad H_1 : \boldsymbol{\theta} \in \Omega \setminus \omega$
   - e.g. $H_0 : r_1 = r_2, \quad H_1 : r_1 \neq r_2$
2. **Fit both models & extract log-likelihood maxima**

$$\log L_0 = \max_{\boldsymbol{\theta} \in \omega} \log L(\boldsymbol{\theta}), \qquad \log L_1 = \max_{\boldsymbol{\theta} \in \Omega} \log L(\boldsymbol{\theta})$$

3. **Compute test statistic**

$$\boxed{\lambda = -2\big(\log L_0 - \log L_1\big)}$$

Likelihood Ratio Test: Step by Step

4. **Apply Wilks' theorem**
   - If N large and H_0 true: $\lambda \sim \chi^2(\mathrm{df}), \;\; \mathrm{df} = \dim \Omega - \dim \omega$

Likelihood Ratio Test: Step by Step

4. **Apply Wilks' theorem**
   - If N large and H_0 true: $\lambda \sim \chi^2(\mathrm{df})$, $\mathrm{df} = \dim\Omega - \dim\omega$
   - Count-rate test: Ω has 2 parameters (r_1, r_2), ω has 1 (r) $\Rightarrow$ $\mathrm{df} = 1$

Likelihood Ratio Test: Step by Step

4. **Apply Wilks' theorem**
   - If N large and H_0 true: $\lambda \sim \chi^2(\mathrm{df}), \;\; \mathrm{df} = \dim\Omega - \dim\omega$
   - Count-rate test: Ω has 2 parameters (r_1, r_2), ω has 1 $(r) \;\; \Rightarrow \;\; \mathrm{df} = 1$
   - Signal+background test (homework): Ω has 4 params $(n_{\mathrm{sig}}, n_{\mathrm{bkg}}, a_1, a_2)$, ω has 3 $(n_{\mathrm{bkg}}, a_1, a_2) \;\; \Rightarrow \;\; \mathrm{df} = 1$

Likelihood Ratio Test: Step by Step

4. **Apply Wilks' theorem**
   - If N large and H_0 true: $\lambda \sim \chi^2(\mathrm{df}),\ \ \mathrm{df} = \dim\Omega - \dim\omega$
   - Count-rate test: Ω has 2 parameters (r_1, r_2), ω has 1 (r) $\Rightarrow\ \mathrm{df} = 1$
   - Signal+background test (homework): Ω has 4 params $(n_{\mathrm{sig}}, n_{\mathrm{bkg}}, a_1, a_2)$, ω has 3 $(n_{\mathrm{bkg}}, a_1, a_2)$ $\Rightarrow\ \mathrm{df} = 1$
5. **Compute p-value**

$$p = P(\lambda \geq \hat{\lambda} \mid H_0) = 1 - F_{\chi^2(\mathrm{df})}(\hat{\lambda})$$

   - **Probability of observing data at least as extreme as seen**, *assuming H_0 is true*
   - **Small $p \Rightarrow H_0$ is unlikely; large $p \Rightarrow$ data consistent with H_0**

Likelihood Ratio Test: Step by Step

4. **Apply Wilks' theorem**
   - If N large and H_0 true: $\lambda \sim \chi^2(\mathrm{df})$, $\mathrm{df} = \dim\Omega - \dim\omega$
   - Count-rate test: Ω has 2 parameters (r_1, r_2), ω has 1 (r) $\Rightarrow$ $\mathrm{df} = 1$
   - Signal+background test (homework): Ω has 4 params $(n_{\mathrm{sig}}, n_{\mathrm{bkg}}, a_1, a_2)$, ω has 3 $(n_{\mathrm{bkg}}, a_1, a_2)$ $\Rightarrow$ $\mathrm{df} = 1$
5. **Compute p-value**

$$p = P(\lambda \geq \hat{\lambda} \mid H_0) = 1 - F_{\chi^2(\mathrm{df})}(\hat{\lambda})$$

   - **Probability of observing data at least as extreme as seen**, *assuming H_0 is true*
   - **Small $p \Rightarrow H_0$ is unlikely; large $p \Rightarrow$ data consistent with H_0**
6. **Compute significance** (Z-score)

$$Z = F^{-1}_{\mathcal{N}(0,1)}(1 - p)$$

   - Expresses p-value in units of Gaussian standard deviations (σ)

Likelihood Ratio Test: Step by Step

4. **Apply Wilks' theorem**
   - If N large and H_0 true: $\lambda \sim \chi^2(\mathrm{df})$, $\mathrm{df} = \dim\Omega - \dim\omega$
   - Count-rate test: Ω has 2 parameters (r_1, r_2), ω has 1 (r) $\Rightarrow$ $\mathrm{df} = 1$
   - Signal+background test (homework): Ω has 4 params $(n_{\mathrm{sig}}, n_{\mathrm{bkg}}, a_1, a_2)$, ω has 3 $(n_{\mathrm{bkg}}, a_1, a_2)$ $\Rightarrow$ $\mathrm{df} = 1$
5. **Compute p-value**

$$p = P(\lambda \geq \hat{\lambda} \mid H_0) = 1 - F_{\chi^2(\mathrm{df})}(\hat{\lambda})$$

   - **Probability of observing data at least as extreme as seen**, *assuming H_0 is true*
   - **Small $p \Rightarrow H_0$ is unlikely; large $p \Rightarrow$ data consistent with H_0**
6. **Compute significance** (Z-score)

$$Z = F^{-1}_{\mathcal{N}(0,1)}(1 - p)$$

   - Expresses p-value in units of Gaussian standard deviations (σ)
7. **Draw conclusion**: compare Z to threshold

Reporting Test Results

Z	p-**value**	**Typical conclusion**
$Z < 3$	$p > 0.0013$	No evidence for H_1
$3 \leq Z < 5$	$2.8 \times 10^{-7} < p \leq 0.0013$	Hints / evidence for H_1
$Z \geq 5$	$p \leq 2.8 \times 10^{-7}$	Discovery

Reporting Test Results

Z	p-**value**	**Typical conclusion**
$Z < 3$	$p > 0.0013$	No evidence for H_1
$3 \leq Z < 5$	$2.8 \times 10^{-7} < p \leq 0.0013$	Hints / evidence for H_1
$Z \geq 5$	$p \leq 2.8 \times 10^{-7}$	Discovery

Article-title conventions:

- $Z < 3$: "**Search for** ..." (e.g. MEG II: $\mu^+ \to e^+\gamma$)
- $3 \leq Z < 5$: "**Evidence for** ..."
- $Z \geq 5$: "**Observation of** ..." (e.g. CMS: Higgs boson, 5.0σ)

Reporting Test Results

Z	p-**value**	**Typical conclusion**
$Z < 3$	$p > 0.0013$	No evidence for H_1
$3 \leq Z < 5$	$2.8\times10^{-7} < p \leq 0.0013$	Hints / evidence for H_1
$Z \geq 5$	$p \leq 2.8 \times 10^{-7}$	Discovery

Article-title conventions:

- $Z < 3$: "**Search for** ..." (e.g. MEG II: $\mu^+ \to e^+\gamma$)
- $3 \leq Z < 5$: "**Evidence for** ..."
- $Z \geq 5$: "**Observation of** ..." (e.g. CMS: Higgs boson, 5.0σ)

Note

Sub-3σ results are still publishable — provide **upper limits** instead.

Likelihood Ratio Test: Counting Rate Comparison

Data:

- $T_1 = 3612$ s, $N_1 = 1243$, $\hat{r}_1 = 0.3441 \pm 0.0098$ s^{-1}
- $T_2 = 1803$ s, $N_2 = 567$, $\hat{r}_2 = 0.314 \pm 0.013$ s^{-1}

Data:

- $T_1 = 3612$ s, $N_1 = 1243$, $\hat{r}_1 = 0.3441 \pm 0.0098$ s^{-1}
- $T_2 = 1803$ s, $N_2 = 567$, $\hat{r}_2 = 0.314 \pm 0.013$ s^{-1}

Hypotheses:

$$H_0 : r_1 = r_2 \qquad H_1 : r_1 \neq r_2$$

Wilks' theorem: under H_0, $\lambda \sim \chi^2(1)$

Test statistic (analytic derivation):

$$\boxed{\lambda = 2\left(N_1 \log \frac{N_1}{\hat{r}T_1} + N_2 \log \frac{N_2}{\hat{r}T_2}\right)}$$

where $\hat{r} = \dfrac{N_1 + N_2}{T_1 + T_2}$ (pooled rate under H_0)

Likelihood Ratio Test: Counting Rate Comparison

Data:

- $T_1 = 3612$ s, $N_1 = 1243$, $\hat{r}_1 = 0.3441 \pm 0.0098$ s^{-1}
- $T_2 = 1803$ s, $N_2 = 567$, $\hat{r}_2 = 0.314 \pm 0.013$ s^{-1}

Hypotheses:

$$H_0: r_1 = r_2 \qquad H_1: r_1 \neq r_2$$

Test statistic (analytic derivation):

$$\boxed{\lambda = 2\left(N_1 \log \frac{N_1}{\hat{r}T_1} + N_2 \log \frac{N_2}{\hat{r}T_2}\right)}$$

where $\hat{r} = \dfrac{N_1 + N_2}{T_1 + T_2}$ (pooled rate under H_0)

Wilks' theorem: under H_0, $\lambda \sim \chi^2(1)$

Result:

- $\lambda = 3.20$
- $p = \texttt{TMath::Prob}(3.20, 1) = 0.074$
- $Z = F^{-1}_{\mathcal{N}(0,1)}(1 - p) = 1.4$

Conclusion:

$$Z < 3 \;\Rightarrow\; \text{no significant difference}$$

Homework: SiPM Single-Photoelectron Spectrum — Model

Physical principle:

- SiPM detects single photons; photo-electron (PE) count $k \sim \mathrm{Pois}(\mu_{\mathrm{PE}})$
- Each PE contributes $\approx$ same ADC amplitude; n PEs produce n times the amplitude

Homework: SiPM Single-Photoelectron Spectrum — Model

Physical principle:

- SiPM detects single photons; photo-electron (PE) count $k \sim \mathrm{Pois}(\mu_{\mathrm{PE}})$
- Each PE contributes $\approx$ same ADC amplitude; n PEs produce n times the amplitude

Multi-PE peak model $(n = 1, \ldots, 11)$:

- n-PE peak **mean**: $\mu_n = n \cdot s_{\mathrm{mean}} + b$
- n-PE peak **sigma**: $\sigma_n = \sqrt{n \cdot s_\sigma^2 + b_\sigma^2}$
- **Poisson weight** (zero-PE pedestal excluded):

$$w_n = \frac{N_{\mathrm{sig}} \cdot \mathrm{Pois}(n \mid \mu_{\mathrm{PE}})}{1 - e^{-\mu_{\mathrm{PE}}}}$$

Homework: SiPM Single-Photoelectron Spectrum — Model

Physical principle:

- SiPM detects single photons; photo-electron (PE) count $k \sim \mathrm{Pois}(\mu_{\mathrm{PE}})$
- Each PE contributes $\approx$ same ADC amplitude; n PEs produce n times the amplitude

Multi-PE peak model $(n = 1, \ldots, 11)$:

- n-PE peak **mean**: $\mu_n = n \cdot s_{\mathrm{mean}} + b$
- n-PE peak **sigma**: $\sigma_n = \sqrt{n \cdot s_\sigma^2 + b_\sigma^2}$
- **Poisson weight** (zero-PE pedestal excluded):

$$w_n = \frac{N_{\mathrm{sig}} \cdot \mathrm{Pois}(n \mid \mu_{\mathrm{PE}})}{1 - e^{-\mu_{\mathrm{PE}}}}$$

Composite model:

$$\boxed{\rho(\mathrm{ADC}) = \sum_{n=1}^{11} w_n \mathcal{N}\left(\mathrm{ADC} \mid n s_{\mathrm{mean}} + b,\ \sqrt{n s_\sigma^2 + b_\sigma^2}\right) + N_{\mathrm{bkg}} \mathcal{N}(\mathrm{ADC} \mid b, \sigma_{\mathrm{bkg}})}$$

Signal: Poisson-weighted multi-Gaussian. Background: Gaussian centered at baseline.

Homework: SiPM Single-Photoelectron Spectrum — Task

Parameters to extract:

- μ_{PE} — average number of PEs
- s_{mean} — single-PE amplitude (mV)
- s_σ — single-PE amplitude fluctuation
- b, b_σ — electronics baseline & noise

Homework: SiPM Single-Photoelectron Spectrum — Task

Parameters to extract:

- μ_{PE} — average number of PEs
- s_{mean} — single-PE amplitude (mV)
- s_σ — single-PE amplitude fluctuation
- $b,\ b_\sigma$ — electronics baseline & noise

Task:

1. Fit `sipm_adc_data.root` with the model on the previous slide
2. Plot the fit result and pull distribution
3. Report fitted parameters with their uncertainties

Tips:

- The ADC range must cover the main part of the spectrum
- Reasonable initial values & ranges are crucial for convergence
- Include both single-PE fluctuation *and* electronics noise

Note

This is a **measurement-type** exercise — contrast with the search-type new-physics task.

Homework: Searching for New Physics

Scenario

Future muon collider at $\sqrt{s} = 10$ TeV. Search for a new particle:

- Mass ≈ 4.2 TeV, Width ≈ 100 GeV
- Look for a mass peak on the invariant mass spectrum of the decay products
- Signal shape: **Breit–Wigner** (the particle's intrinsic mass distribution)
- Background shape: **quadratic polynomial**

Note

This analysis is **greatly simplified** for teaching purposes. In a real experiment, one would scan over a mass range and use the *local significance* to determine the most likely mass value, rather than fixing the mass beforehand.

Signal — Breit–Wigner:

$$\boxed{f_{\text{sig}}(m) = \frac{\Gamma}{2\pi}\,\frac{1}{(m-m_0)^2+\Gamma^2/4}}$$

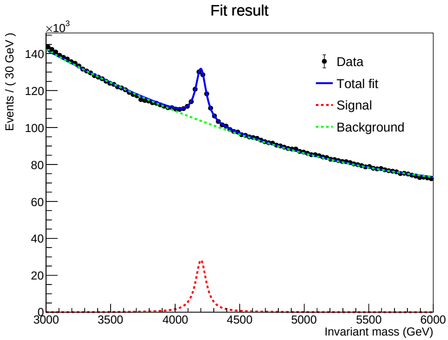
Background — quadratic polynomial:

$$\boxed{f_{\text{bkg}}(m) = A\left(a_2 m^2 + a_1 m + 1\right)}$$

Two competing composite models:

$$H_0:\ \rho_0(m) = n_{\text{bkg}}\, f_{\text{bkg}}(m; a_1, a_2)$$

$$H_1:\ \rho_1(m) = n_{\text{sig}}\, f_{\text{sig}}(m) + n_{\text{bkg}}\, f_{\text{bkg}}(m; a_1, a_2)$$



Signal (B–W) and background (polynomial) distributions

Homework: Method

Per-dataset procedure:

1. Fit ρ_0 and ρ_1 via extended MLE
2. Extract $\log L_0$ and $\log L_1$

Homework: Method

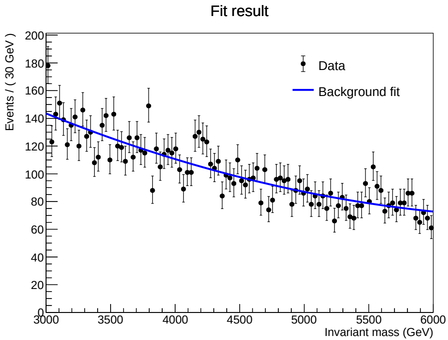
Per-dataset procedure:

1. Fit ρ_0 and ρ_1 via extended MLE
2. Extract $\log L_0$ and $\log L_1$
3. $\lambda = -2(\log L_0 - \log L_1)$
4. $p = \texttt{TMath::Prob}(\lambda, 1)$
5. $Z = \texttt{TMath::NormQuantile}(1 - p)$

Per-dataset procedure:

1. Fit ρ_0 and ρ_1 via extended MLE
2. Extract $\log L_0$ and $\log L_1$
3. $\lambda = -2(\log L_0 - \log L_1)$
4. $p = \texttt{TMath::Prob}(\lambda, 1)$
5. $Z = \texttt{TMath::NormQuantile}(1 - p)$
6. Draw conclusion using 5σ criterion
7. If significant, report n_{sig} with uncertainty

Example output:



H_0 fit result (background only)

See the full lecture notes for details and code examples.

Homework: Reporting

Task

6 datasets: `mucol_data0.root` through `mucol_data5.root`. Analyze **each** dataset as shown in the previous slides.

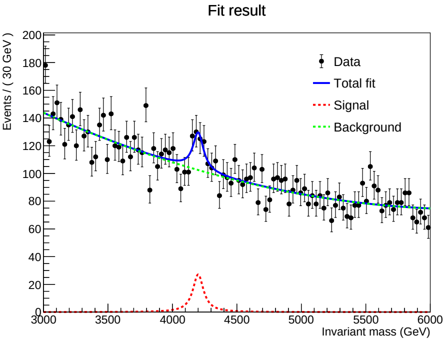
Task

6 datasets: `mucol_data0.root` through `mucol_data5.root`. Analyze **each** dataset as shown in the previous slides.

Submission

Work in **groups of 4–5**. Submit **one report per group** to `zhaoshh7@mail2.sysu.edu.cn` by the end of this week. The report should include fit result plots for each dataset, a completed summary table, final conclusions, and your thoughts.

Example output:



H_1 fit result (signal + background)

Thanks and good luck!