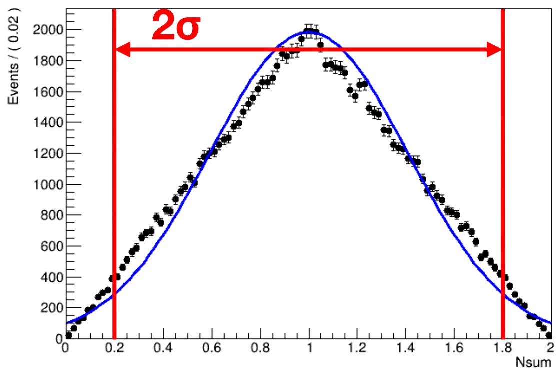
(Solutions - day 1)

Wouter Verkerke, NIKHEF

Solution – Exercise 1

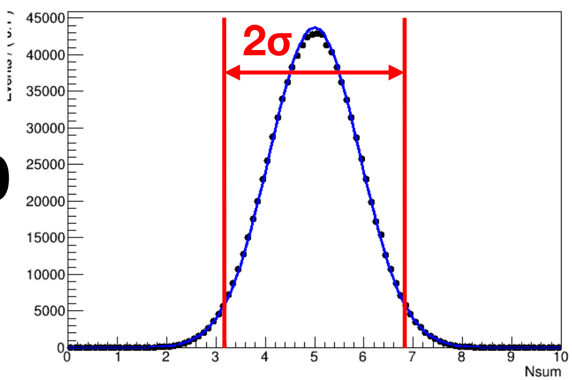
N=2

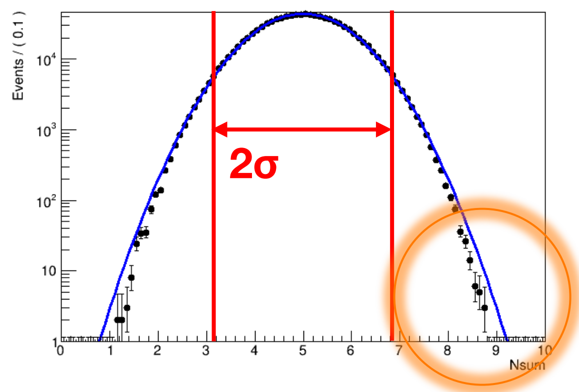


$$V(N=1) = E[x^2] - E[x]^2$$
$$= \int_0^1 x^2 dx - \left(\int_0^1 x\right)^2$$
$$= \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$$

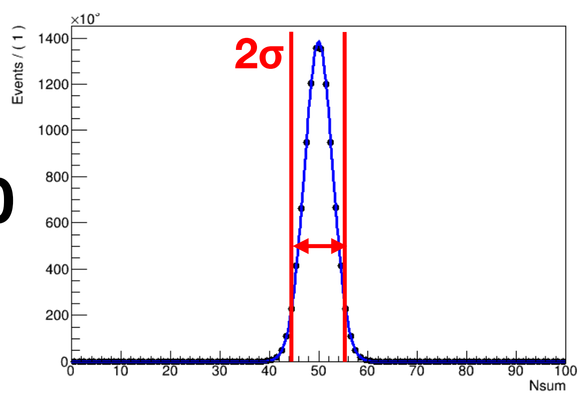
$$\sigma(N) = \frac{\sigma(1)}{\sqrt{N}} N$$
$$= \sqrt{\frac{N}{12}}$$

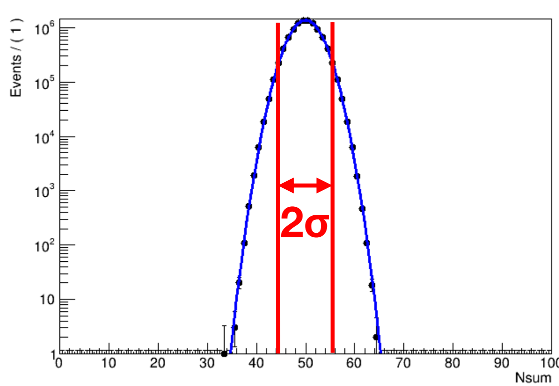
N=10





N=100

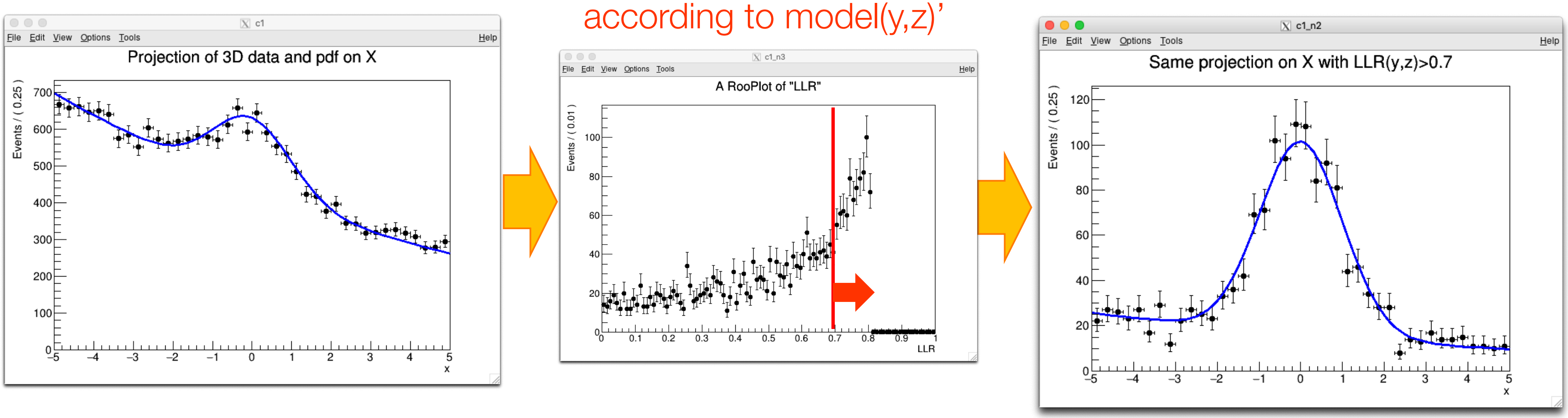




Solution – Exercise 3

- Analytical application of Neyman-Pearson lemma

'per-event signal purity according to model(y,z)'




```
RooFitResult: minimized FCN value: 135474, estimated distance to minimum: 2.25574e-06
              covariance matrix quality: Full, accurate covariance matrix
              Status : MINIMIZE=0 HESSE=0

    Floating Parameter    FinalValue +/-  Error
  --------------------  --------------------------
                  fsig    9.7816e-02 +/-  2.78e-03
```

→ 35.2σ

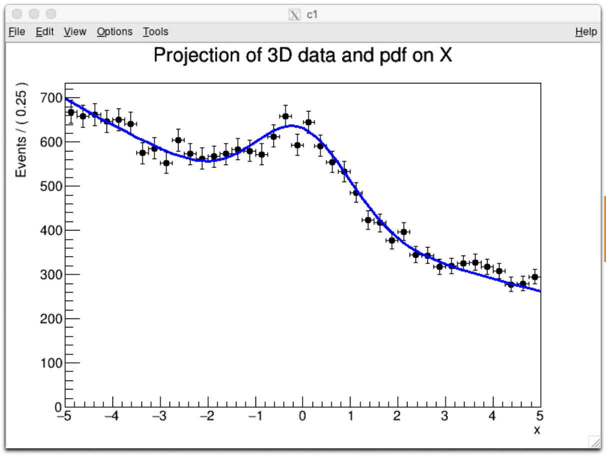
```
RooFitResult: minimized FCN value: 3096.86, estimated distance to minimum: 2.14652e-05
              covariance matrix quality: Full, accurate covariance matrix
              Status : MINIMIZE=0 HESSE=0

    Floating Parameter    FinalValue +/-  Error
  --------------------  --------------------------
                  fsig    5.6477e-01 +/-  1.92e-02
```

→ 29.4σ

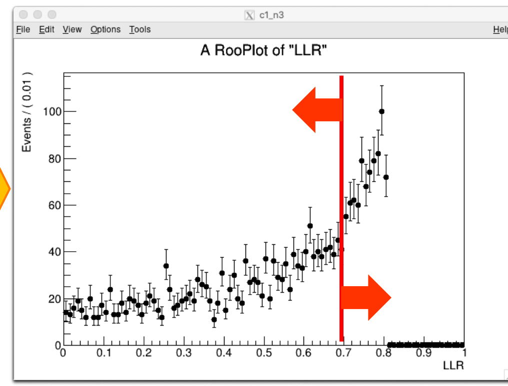
→ Note design choice to cut *only* on model f(y,z) with NP-lemma

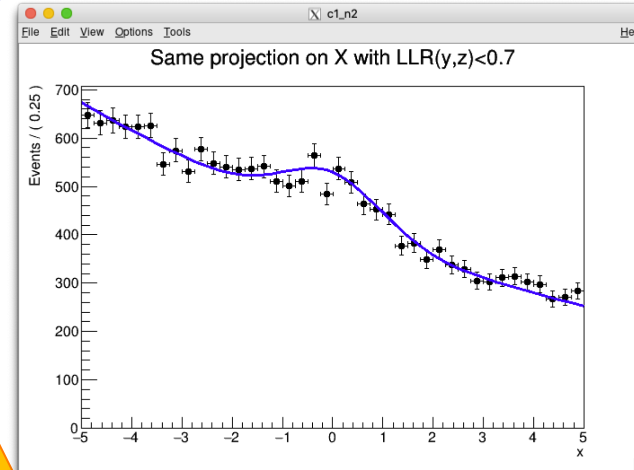
→ Can use f(x) to *measure* remaining background

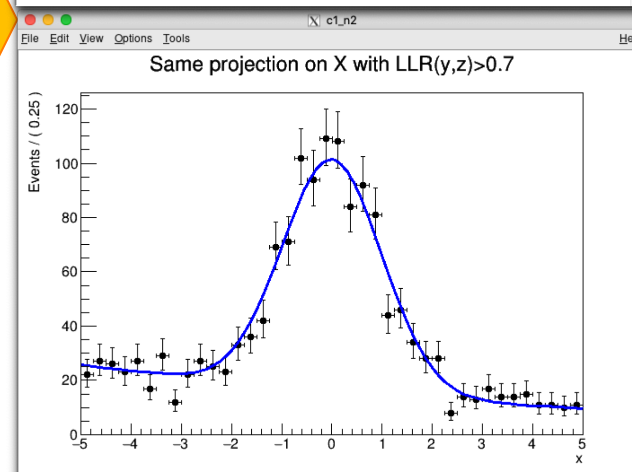
Solution – Exercise 3



'per-event signal purity according to model(y,z)'








```
RooFitResult: minimized FCN value: 135474, estimated distance to minimum: 2.25574e-06
              covariance matrix quality: Full, accurate covariance matrix
              Status : MINIMIZE=0 HESSE=0

  Floating Parameter    FinalValue +/-  Error
  --------------------  --------------------------
                fsig    9.7816e-02 +/-  2.78e-03
```

→ 35.2σ

```
RooFitResult: minimized FCN value: 3096.86, estimated distance to minimum: 2.14652e-05
              covariance matrix quality: Full, accurate covariance matrix
              Status : MINIMIZE=0 HESSE=0

  Floating Parameter    FinalValue +/-  Error
  --------------------  --------------------------
                fsig    5.6477e-01 +/-  1.92e-02
```

→ 29.4σ

```
RooFitResult: minimized FCN value: 41908.6, estimated distance to minimum: 5.35272e-05
              covariance matrix quality: Full, accurate covariance matrix
              Status : MINIMIZE=0 HESSE=0

  Floating Parameter    FinalValue +/-  Error
  --------------------  --------------------------
                fsig    4.8966e-02 +/-  5.49e-03
```

→ 8.9σ

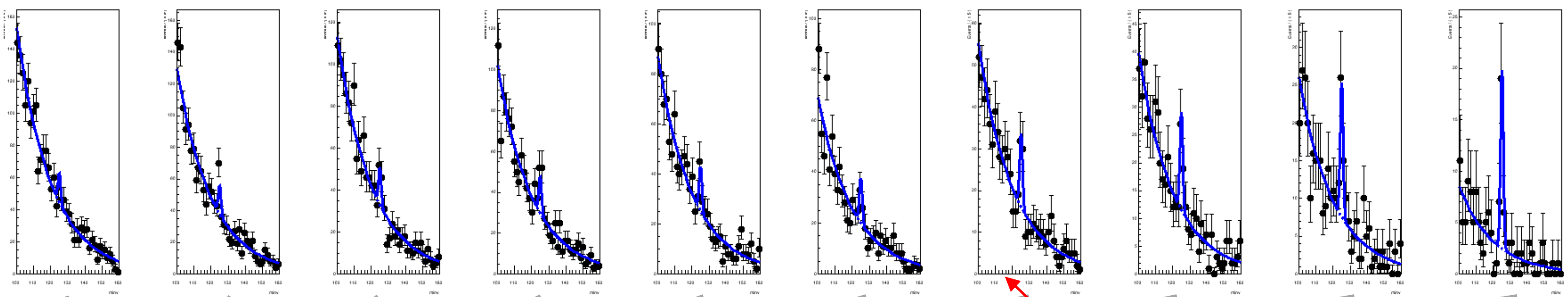
} 30.7σ

'not using prediction of event ratio over LLR cut'

NB: Feature!

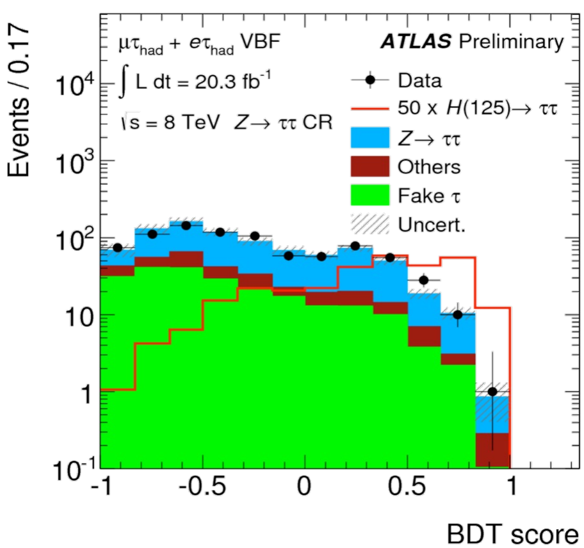
(Preview to day 3) → Can also split in >>2 regions

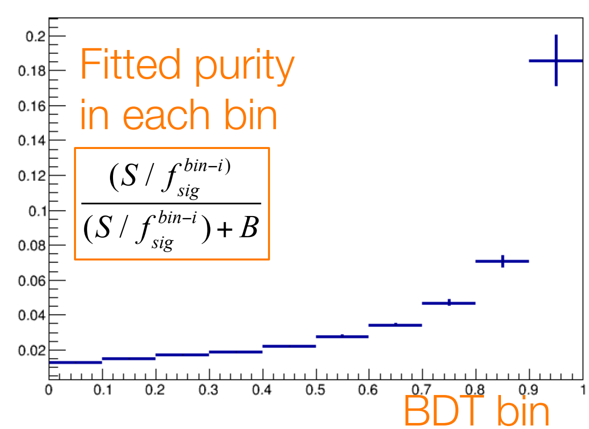
- In that case only exploit variation in purity, not distribution of LLR/BDT
  → *Analysis insensitive to mismodelling of LLR/BDT distribution!*



$$f_{bin-i}(m\,|\,S,B) = \frac{S}{f_{sig}^{bin-i}} f_S(m) + B_{bin-i} f_B(s)$$

Scale factor that ensures that every bin interprets S as the total signal yield





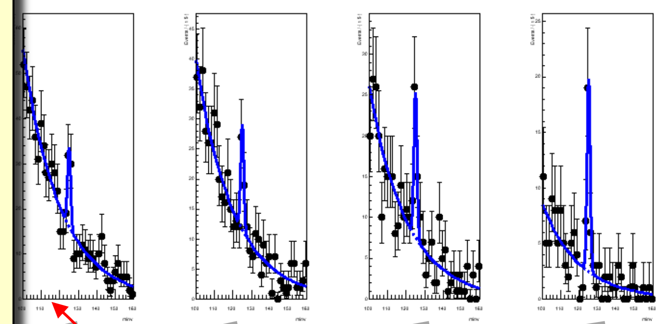
(Preview to day 3) → Can also split in >>2 regions

- In that case only exploit variation in purity, not distribution of LLR/BDT
  → *Analysis insensitive to mismodelling of LLR/BDT distribution!*

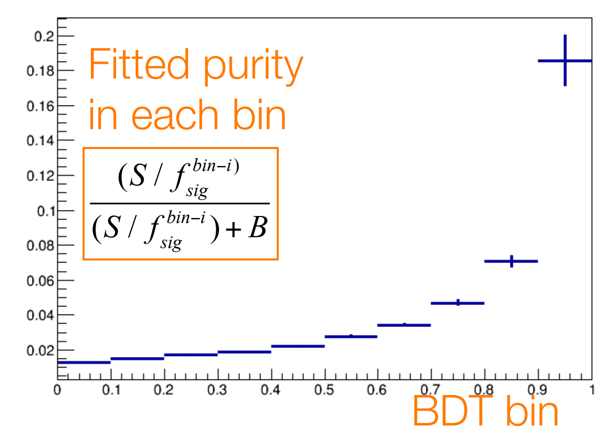
Joint PDF for this model

$$f(m,n_{BDT}|S,\vec{B}) = lookup(n_{BDT}) \begin{cases} f_{bin-0}(m|S,B_0) = \frac{S}{f_{sig}^{bin-0}} f_S(m) + B_{bin-0} f_B(s) \\ f_{bin-1}(m|S,B_1) = \frac{S}{f_{sig}^{bin-1}} f_S(m) + B_{bin-1} f_B(s) \\ f_{bin-2}(m|S,B_2) = \frac{S}{f_{sig}^{bin-2}} f_S(m) + B_{bin-2} f_B(s) \\ f_{bin-3}(m|S,B_3) = \frac{S}{f_{sig}^{bin-3}} f_S(m) + B_{bin-3} f_B(s) \\ \vdots \\ f_{bin-N}(m|S,B_N) = \frac{S}{f_{sig}^{bin-N}} f_S(m) + B_{bin-N} f_B(s) \end{cases}$$

```
// Construct template model
w.factory("SUM::fit_template(prod(Nsig[30,0,100],frac[1])*sig1,
                              Nbkg[1000,0,10000]*bkg1)") ;

// Construct joint model from template clones
w.factory("SIMCLONE::fitmodel(fit_template,
                    $SplitParam({Nbkg,frac},bdtBin))") ;
```



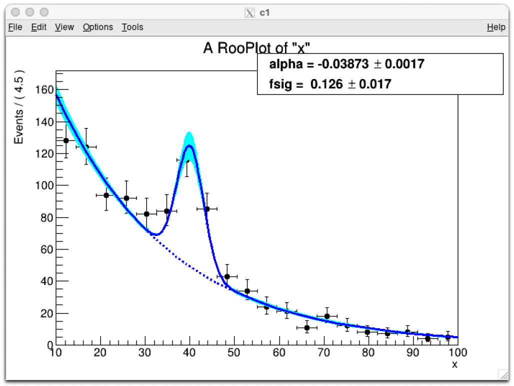
$$n|S,B) = \frac{S}{f_{sig}^{bin-i}} f_S(m) + B_{bin-i} f_B(s)$$

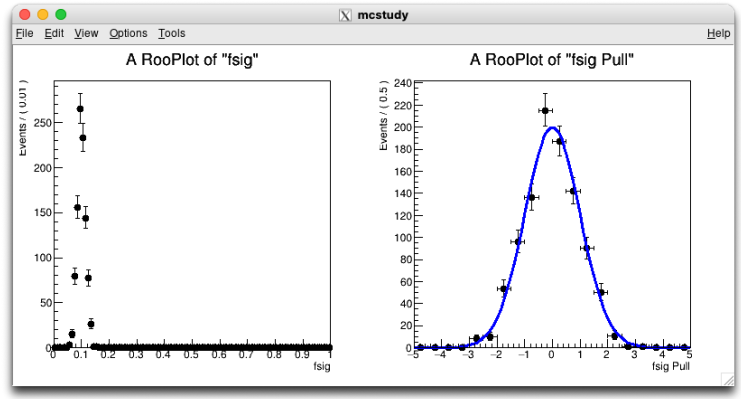


BDT score

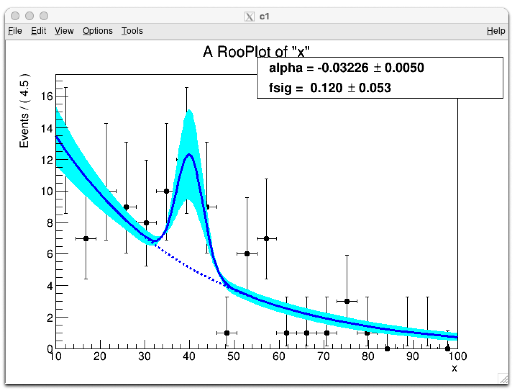
Solution – Exercise 5

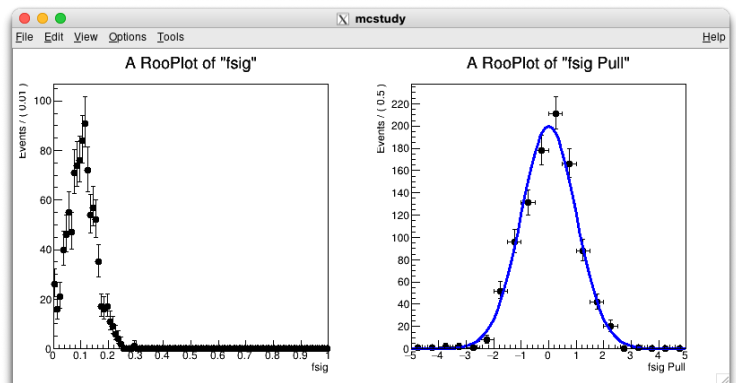
N=1000



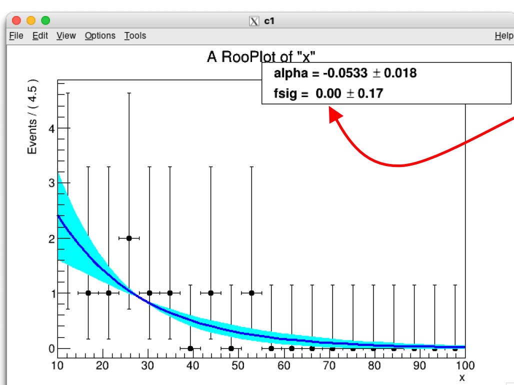


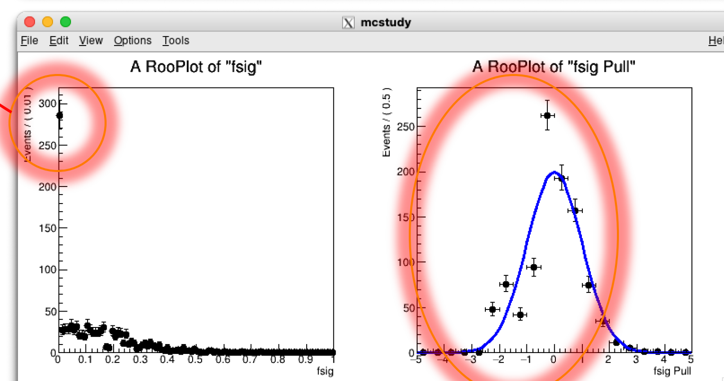
N=100





N=10



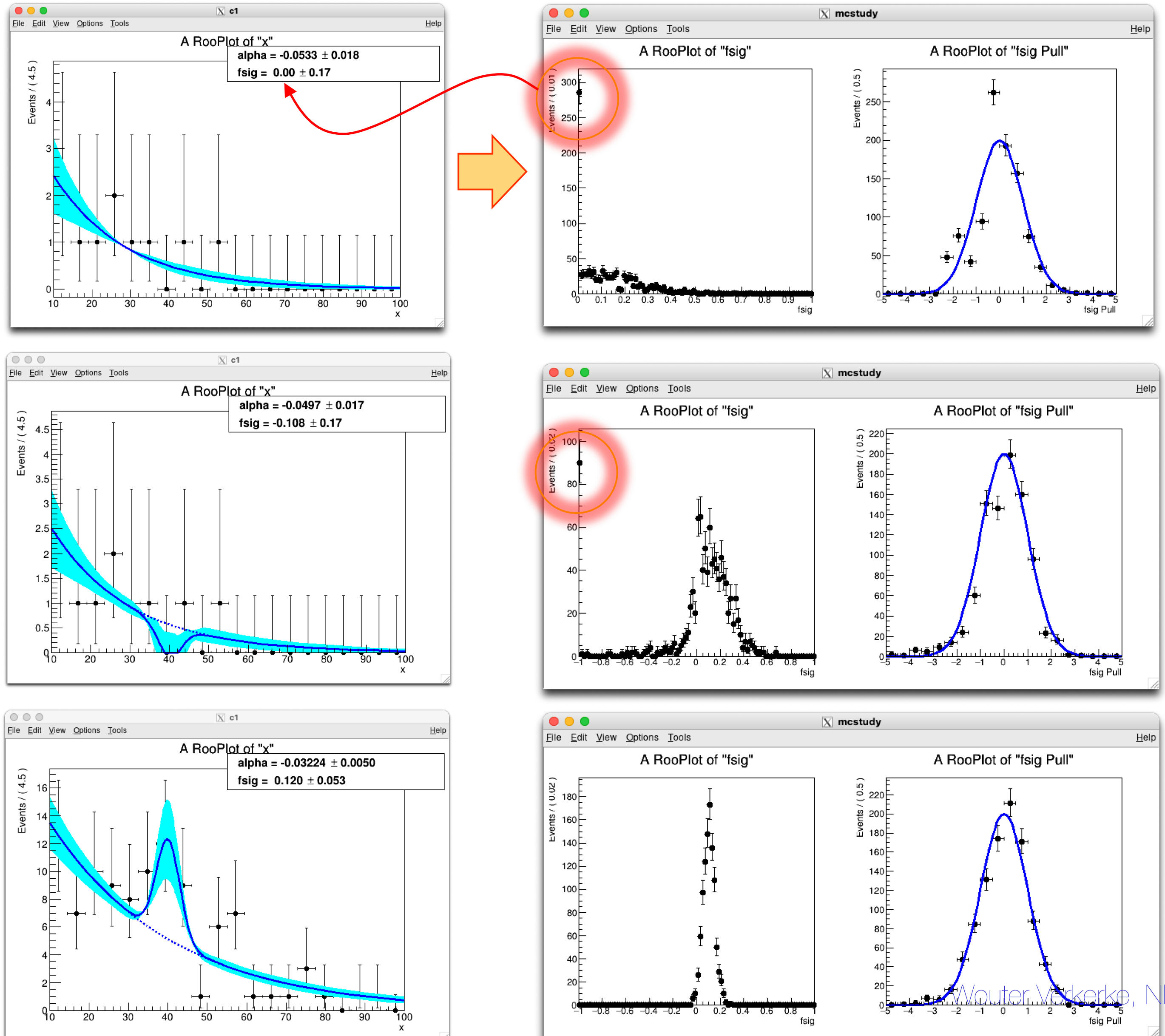


Solution – Exercise 5

(**N=10**)

fsig[-1,1]
N=10

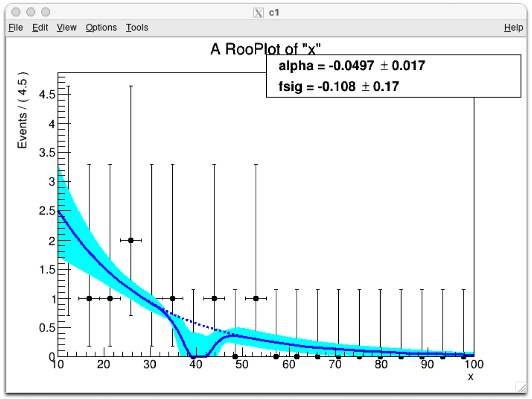
fsig[-1,1]
N=100

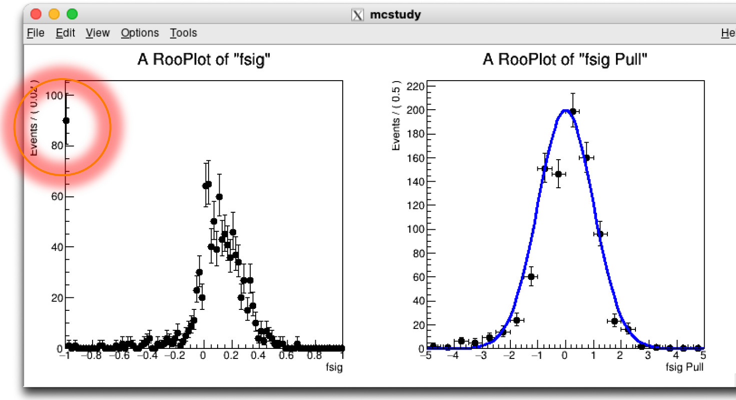


Solution – Exercise 5

fsig[-1,1]

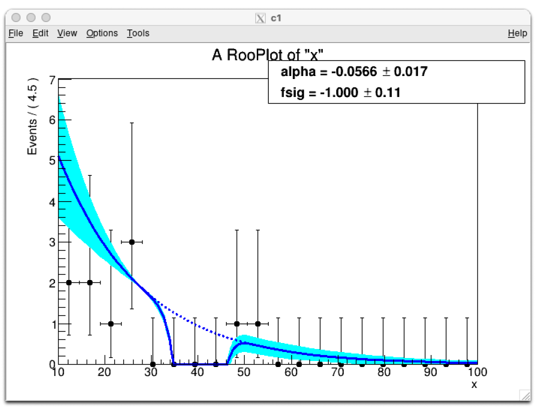
N=10

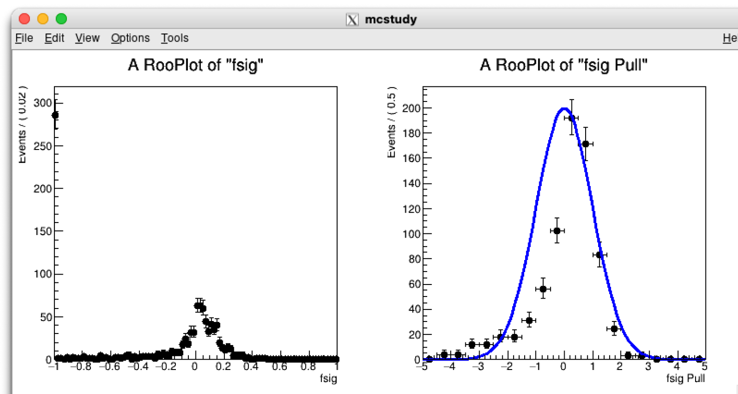




fsig=0

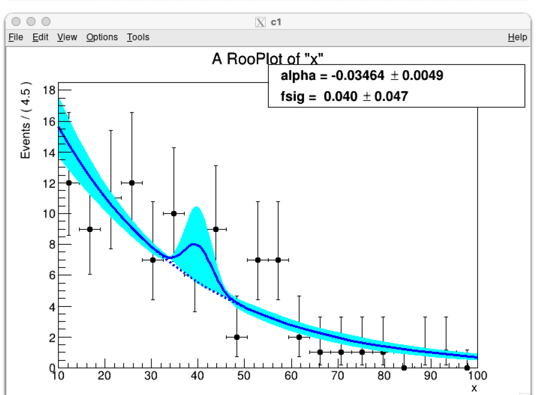
N=10

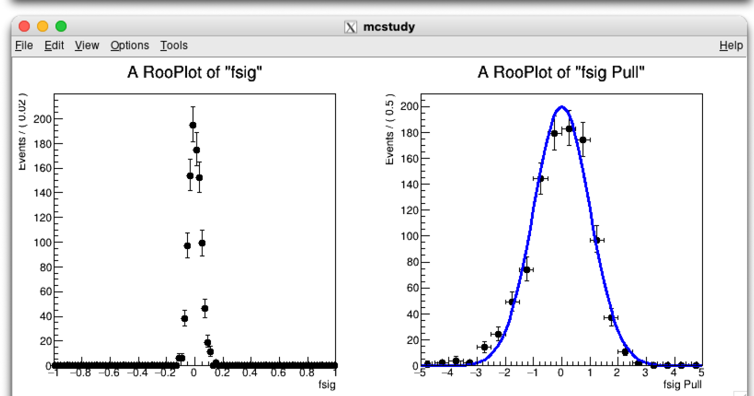




fsig=0

N=100





Solution – Exercise 6

- Normalized Exponential p.d.f

$$f(x) = \tau e^{-\tau x}$$

- Negative log-Likelihood (1 event)

$$-\log\big(f(x)\big) = -\log(\tau) - \log(e^{-\tau x})$$
$$= -\log(\tau) + \tau x$$

- Negative log-Likelihood (N events)

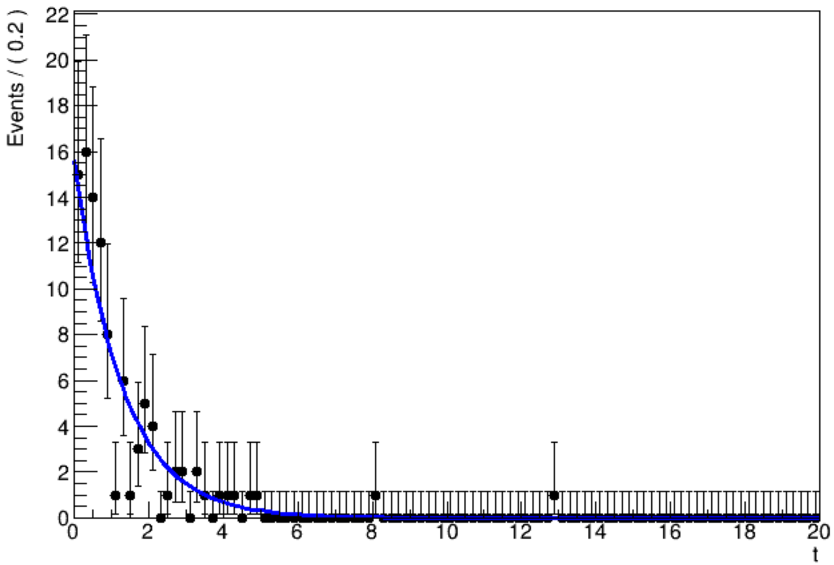
$$-\log(L) = \sum_{i=1}^{N} -\log(\tau) + \tau x_i$$
$$= -N\log(\tau) + \tau \sum_{i=1}^{n} x_i$$

- ML estimator for N events $0 = \frac{d\log L}{d\tau}$

$$0 = -N\frac{1}{\tau} + \sum_{i=1}^{n} x_i$$
$$\frac{N}{\tau} = \sum_{i=1}^{n} x_i \Rightarrow \frac{1}{\tau} = \frac{\sum_{i=1}^{n} x_i}{N} = \langle x \rangle$$

Solution – Exercise 6

- Analytical vs numeric calculation




```
 ..........
COVARIANCE MATRIX CALCULATED SUCCESSFULLY
FCN=124.735 FROM HESSE     STATUS=OK              5 CALLS          23 TOTAL
                    EDM=7.97722e-09    STRATEGY= 1      ERROR MATRIX ACCURATE
 EXT PARAMETER                                INTERNAL      INTERNAL
 NO.   NAME      VALUE            ERROR       STEP SIZE       VALUE
  1  tau          1.28062e+00   1.28048e-01   4.17606e-05  -1.07870e+00
                               ERR DEF= 0.5
EXTERNAL ERROR MATRIX.    NDIM=  25    NPAR=  1    ERR DEF=0.5
 1.640e-02
```

$$\frac{1}{\tau} = \frac{\sum_{i=1}^{n} x_i}{N} = \langle x \rangle$$

$\tau = 1.28063$

Agreement MINUIT/analytical within ~10^{-5} relative precision
Default MINUIT numerical precision (EDM 10^{-3}) ~4% of error → $5\text{x}10^{-3}$