Statistical Data Analysis
Discussion notes – week 7

- Problem sheet 4
- Maximum Likelihood examples

Problem sheet 4

Exercise 1(a) [2 marks] Let $x_N = \sum_{i=1}^{N} r_i$, where the r_i are independent and uniformly distributed between 0 and 1. Find the mean μ_N and standard deviation σ_N of x_N as a function of N.

1(a) [2 marks] The variable x_N is defined as

$$x_N = \sum_{i=1}^{N} r_i \,,$$

where the r_i are independent and uniformly distributed on [0,1]. From the lectures (or rederive) we have the mean and variance $E[r_i] = 1/2$ and $V[r_i] = 1/12$. The expectation value of x_N is therefore

$$E[x_N] \equiv \mu_N = E\left[\sum_{i=1}^{N} r_i\right] = \sum_{i=1}^{N} E[r_i] = \sum_{i=1}^{N} \frac{1}{2} = \frac{N}{2} \,.$$

The variance of x_N is

$$V[x_N] = V\left[\sum_{i=1}^{N} r_i\right] = \sum_{i=1}^{N} V[r_i] = \sum_{i=1}^{N} \frac{1}{12} = \frac{N}{12}$$

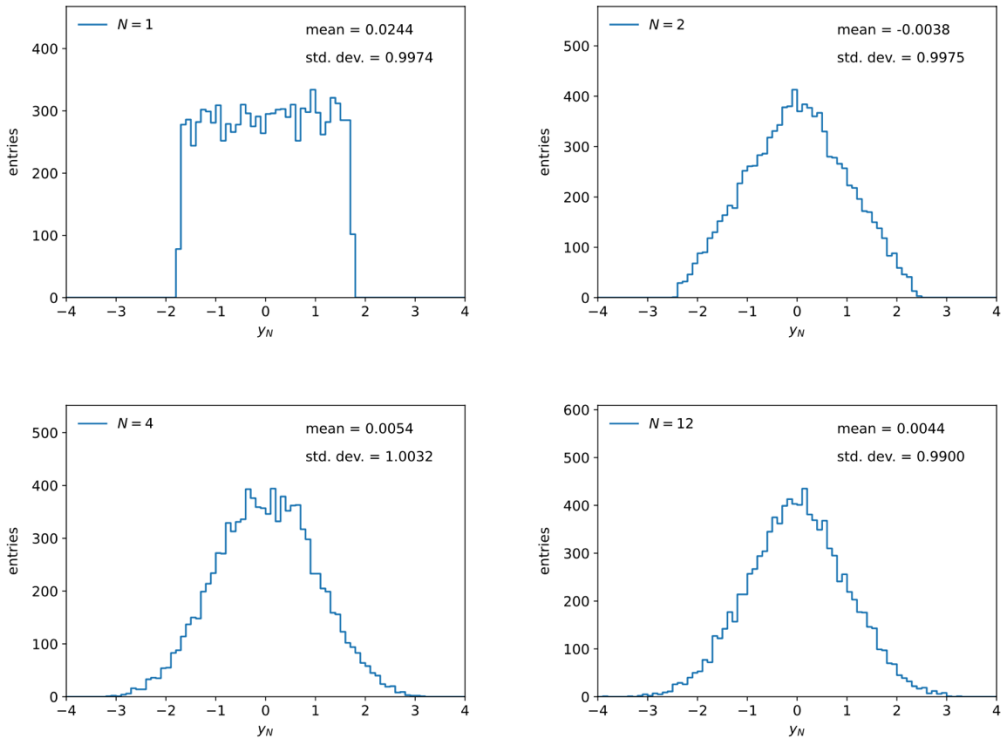
and so its standard deviation is

$$\sigma[x_N] = \sqrt{\frac{N}{12}} \,.$$

1(b) [5 marks] Using the results from (a), construct the standardized variable

$$y_N = \frac{x_N - \mu_N}{\sigma_N} = \sqrt{\frac{12}{N}} \left(\sum_{i=1}^{N} r_i - \frac{N}{2} \right) .$$

Using the **`simpleMC`** program (either C++ or Python) from problem sheet 3 as a starting point, write a computer program to make histograms of 10000 values of y_N as defined above for $N = 1, 2, 4, 12$. Make sure to set the limits of the histogram such that the entire distribution is plotted.



Histograms of y_N *for* N = 1, 2, 4, 12. The means and standard deviations are indicated on the plots. As shown, they are close to $\mu = 0$ and $\sigma = 1$, as must emerge by construction for the standardized variable y_N.

Exercise 2 Consider the pdf $f(x) = 4x^3$, $0 \leq x \leq 1$.

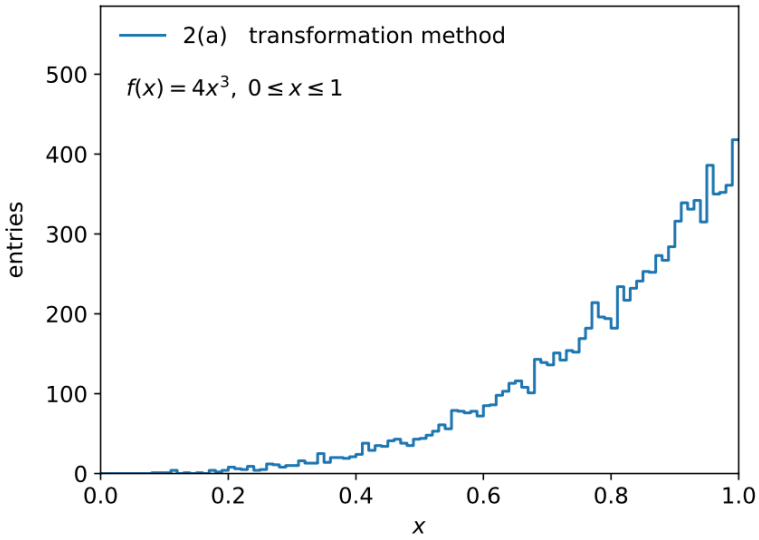
2(a) [4 marks] Use the transformation method to find the function $x(r)$ to generate random numbers according $f(x)$. Implement the method in a short computer program and make a histogram with 10000 values.

2(a) [4 marks] Find the transformation $x(r)$ to produce $x \sim f(x)$. First find the cumulative distribution

$$F(x) = \int_0^x 4x'^3\, dx' = x^4 \,.$$

Set $F(x) = r$ and solving for x gives

$$x(r) = r^{1/4} \,.$$



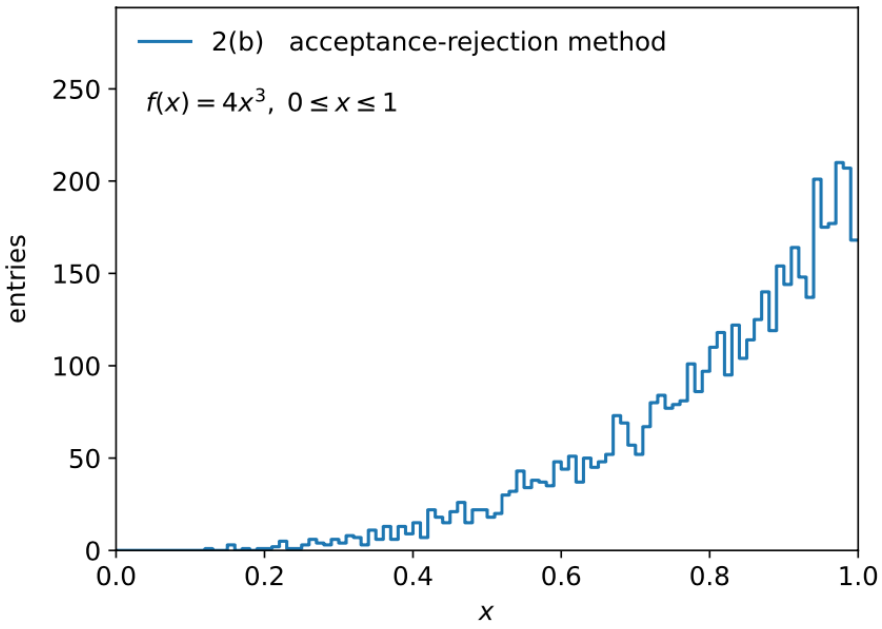

```
numVal = 10000
nBins = 100
rData = np.random.uniform(0., 1., numVal)
xMin=0.
xMax=1.
xData = pow(rData, 0.25)
xHist, bin_edges = np.histogram(xData, bins=nBins, range=(xMin, xMax))
plotHist(xHist, bin_edges, r'$x$', r'entries', r'2(a)   transformation method')
plt.figtext(0.18, 0.74, r'$f(x) = 4x^{3}, \; 0 \leq x \leq 1$')
```

2(b) [**4 marks**] Write a program to generate random numbers according to $f(x)$ using the acceptance-rejection technique. Plot a histogram of the results.

```
numVal = 20000
nBins = 100

def f(x):
    return 4. * x**3

xMin = 0.
xMax = 1.
fMax = 4.
r1 = np.random.uniform(0., 1., numVal)
x = xMin + r1*(xMax - xMin)
r2 = np.random.uniform(0., 1., numVal)
u = fMax*r2
xData = x[u<f(x)]
nBins = 100
xHist, bin_edges = np.histogram(xData, bins=nBins, range=(xMin, xMax))
plotHist(xHist, bin_edges, r'$x$', r'entries', r'2(b)   acceptance-rejection
method')
plt.figtext(0.18, 0.74, r'$f(x) = 4x^{3}, \; 0 \leq x \leq 1$')
```



Exercise 3 [5 marks] Suppose $\vec{x} = (x_1, \ldots, x_n)$ follows an n-dimensional Gaussian distribution $f(\vec{x}; \vec{\mu}, V)$ with $\vec{\mu} = (\mu_1, \ldots, \mu_n)$ and covariance matrix $V_{ij} = \mathrm{cov}[x_i, x_j]$. (In the formulas below regard $\vec{x}$ and $\vec{\mu}$ to be column vectors.) Suppose we have two hypotheses for the vector of means, $\vec{\mu}_0$ and $\vec{\mu}_1$, where for both one uses the same covariance matrix V, and consider the test statistic

$$t(\vec{x}) = \ln \frac{f(\vec{x}|\vec{\mu}_1)}{f(\vec{x}|\vec{\mu}_0)} \,.$$

Show that this $t(\vec{x})$ can be written in the form

$$t(\vec{x}) = w_0 + \sum_{i=1}^{n} w_i x_i \,,$$

or equivalently $t(\vec{x}) = w_0 + \vec{w}^T \vec{x}$, where $\vec{w}$ is a column vector of coefficients w_i, $i = 1, \ldots, n$.

i.e. $f(\vec{x}|\vec{\mu}_k) = \dfrac{1}{(2\pi)^{n/2}|V|^{1/2}} \exp\left[-\frac{1}{2}(\vec{x}-\vec{\mu}_k)^T V^{-1}(\vec{x}-\vec{\mu}_k)\right]$

Test statistic $t(x) = \ln \dfrac{f(\vec{x}|\vec{\mu}_1)}{f(\vec{x}|\vec{\mu}_0)}$

Ex. 3 (cont.)

$$\ln \frac{f(\vec{x}|\vec{\mu}_1)}{f(\vec{x}|\vec{\mu}_0)} = -\frac{1}{2}\left[(\vec{x}-\vec{\mu}_1)^T V^{-1}(\vec{x}-\vec{\mu}_1) - (\vec{x}-\vec{\mu}_0)^T V^{-1}(\vec{x}-\vec{\mu}_0)\right]$$

$$= -\frac{1}{2}\left[\vec{x}^T V^{-1}\vec{x} - \vec{\mu}_1^T V^{-1}\vec{x} - \vec{x}^T V^{-1}\vec{\mu}_1 + \vec{\mu}_1^T V^{-1}\vec{\mu}_1 - \vec{x}^T V^{-1}\vec{x} + \vec{\mu}_0^T V^{-1}\vec{x} + \vec{x}^T V^{-1}\vec{\mu}_0 - \vec{\mu}_0^T V^{-1}\vec{\mu}_0\right]$$

$\vec{x}^T V^{-1}\vec{\mu}_0 = \vec{\mu}_0^T V^{-1}\vec{x}$ since scalar $(\,) = (\,)^T$ and $V^{-1} = (V^{-1})^T$

Ex. 3 (cont.)

$$= -\frac{1}{2}\left[\vec{\mu}_1 V^{-1} \vec{\mu}_1 - \vec{\mu}_0^T V^{-1} \vec{\mu}_0\right] + (\vec{\mu}_1 - \vec{\mu}_0)^T V^{-1} \vec{x}$$

$$= w_0 + \vec{w}^T \vec{x}$$

$$\vec{w} = \left[(\vec{\mu}_1 - \vec{\mu}_0)^T V^{-1}\right]^T = (V^{-1})^T (\vec{\mu}_1 - \vec{\mu}_0)$$

$= V^{-1}$

If $W = V + V$, $W^{-1} = \frac{1}{2} V^{-1}$

$$\Rightarrow \vec{w} = 2W^{-1}(\vec{\mu}_1 - \vec{\mu}_0)$$

Example of Maximum Likelihood

Consider $f(x) = (1+\theta)\, x^{\theta}$, $0 \le x \le 1$

with i.i.d. sample $x_1, \ldots, x_n$

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} (1+\theta)\, x_i^{\theta}$$

$$\Rightarrow \ln L(\theta) = \sum_{i=1}^{n} \left[\ln(1+\theta) + \theta \ln x_i \right]$$

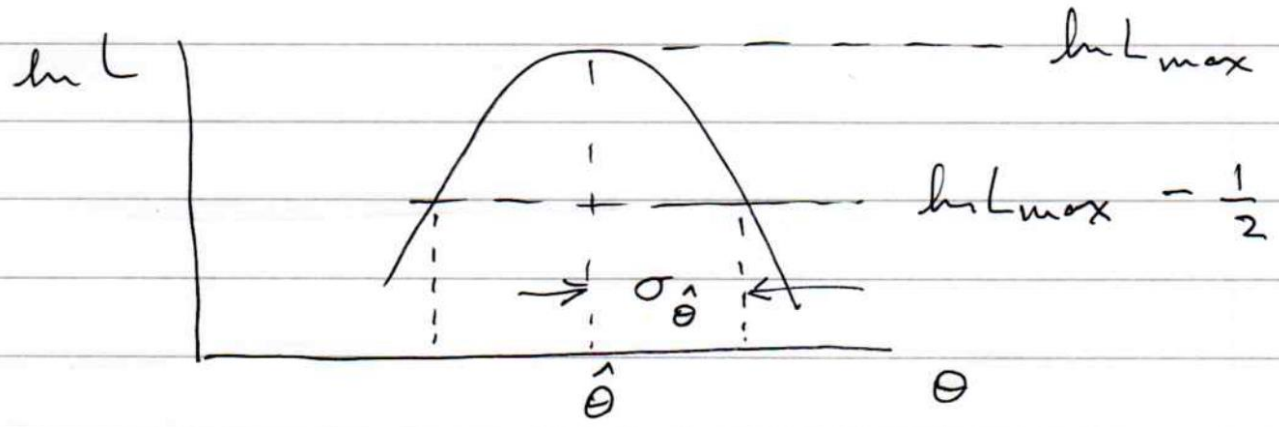
To find MLE,

$$\frac{\partial \ln L}{\partial \theta} = \frac{n}{1+\theta} + \sum_{i=1}^{n} \ln x_i \overset{\text{set}}{=} 0$$

$$\Rightarrow \hat{\theta} = -1 - \frac{n}{\sum_{i=1}^{n} \ln x_i}$$

MLE example (cont.)

For variance (graphical method)



$\sigma_{\hat{\theta}}$ found by moving θ away from $\hat{\theta}$ until $\ln L$ decreases by 1/2 from $\ln L_{max}$

Variance from asymptotic properties

Assume large n, $\overbrace{\quad}^{n} \rightarrow \approx 0$

$$V[\hat{\theta}] \approx MVB = -\frac{\left(1+\frac{\partial b}{\partial \theta}\right)^2}{E\left[\frac{\partial^2 \ln L}{\partial \theta^2}\right]}$$

$$\frac{\partial^2 \ln L}{\partial \theta^2} = -\frac{n}{(1+\theta)^2} \quad \text{or indep. of data } x_i$$

$$-E\left[\frac{\partial^2 \ln L}{\partial \theta^2}\right] = \frac{n}{(1+\theta)^2}$$

$$\Rightarrow V[\hat{\theta}] \approx \frac{(1+\theta)^2}{n} \Rightarrow \sigma_{\hat{\theta}} \approx \frac{1+\theta}{\sqrt{n}}$$

Now suppose problem is parametrized

using $\lambda = \exp\left[\frac{-1}{1+\theta}\right]$ ⟵ find MLE of λ.

Use $\hat{\lambda} = \lambda(\hat{\theta})$

$$\Rightarrow \hat{\lambda} = \exp\left[-\frac{1}{1+\left(-1-\frac{n}{\sum_{i=1}^{n}\ln x_i}\right)}\right]$$

$$= \exp\left[\frac{1}{n}\sum_{i=1}^{n}\ln x_i\right] = \exp\left[\sum_i \ln x_i^{1/n}\right]$$

$$= \prod_{i=1}^{n} x_i^{1/n}$$

Or the hard way: $\theta = -1 - \frac{1}{\ln \lambda}$

$$f(x;\lambda) = -\frac{1}{\ln \lambda} x^{\left(-1-\frac{1}{\ln \lambda}\right)}$$

$$L(\lambda) = \prod_{i=1}^{n} f(x_i;\lambda)$$

$\rightarrow \ln \lambda \longrightarrow \frac{\partial \ln L}{\partial \lambda} = 0 \rightarrow$ same $\hat{\lambda}$ as above.

MLE for number of taxis

The number plate of taxis in every canton in Switzerland ends with a number N from 1 to N_{tot}, where N_{tot} is the total number of taxis.

Model the probability for observing plate number N with

$$P(N|N_{\text{tot}}) = \frac{1}{N_{\text{tot}}}, \quad 1 \leq N \leq N_{\text{tot}}$$

MLE for N_{tot}

Suppose you observe one taxi at random with plate number N.

The likelihood function is $L(N_{\text{tot}}) = \dfrac{1}{N_{\text{tot}}}\,, \quad N_{\text{tot}} \geq N$

which is maximized for $\widehat{N}_{\text{tot}} = N$

The expectation value and bias of the MLE are

$$E[\widehat{N}_{\text{tot}}] = E[N] = \sum_{N=1}^{N_{\text{tot}}} \frac{N}{N_{\text{tot}}} = \frac{N_{\text{tot}} + 1}{2} \qquad b = \frac{1 - N_{\text{tot}}}{2}$$

For better estimators, see similar problem with tanks in WW2:
https://en.wikipedia.org/wiki/German_tank_problem

E.g. the minimum-variance unbiased estimator is: $\widehat{N}_{\text{tot}} = 2N - 1$

Cheap estimator for mass of W boson

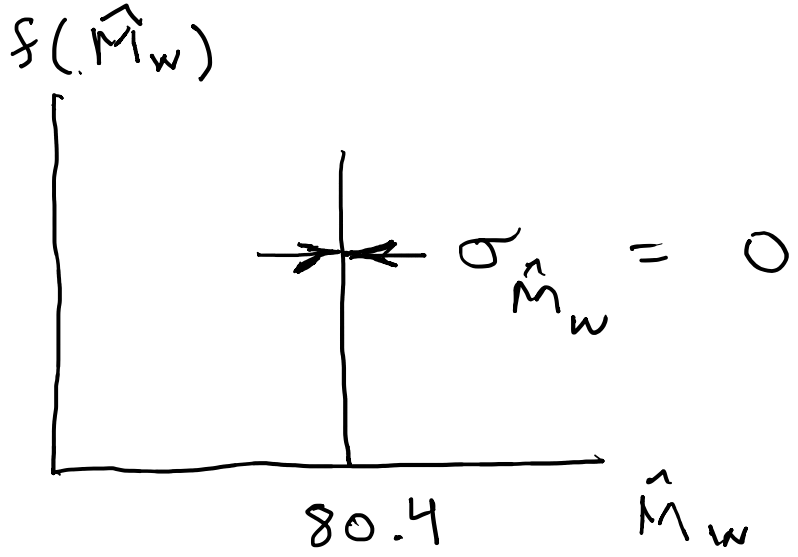
The Particle Physics community has spent huge sums trying to estimate the mass of the W boson with the smallest possible statistical and systematic uncertainty.

Here is an estimator with zero statistical uncertainty. And it's free!

$$\widehat{M_W} = 80.4\,\mathrm{GeV}$$

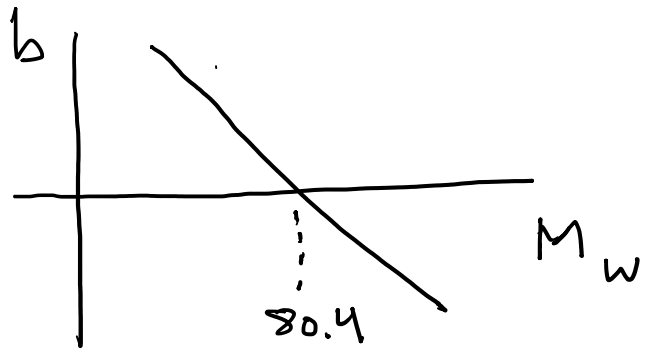
Here is its sampling distribution:

Does this violate the information inequality?



Cheap estimator for mass of W boson (2)

This estimator's bias is $b = E[\widehat{M_W}] - M_W = 80.4\,\text{GeV} - M_W$



Note best estimate of M_W is (in 2020) 80.379±0.012 GeV, so the numerical value of the bias may be fairly small.

But we have $\dfrac{\partial b}{\partial M_W} = -1$ and so

$$\text{MVB} = -\left(1 + \frac{\partial b}{\partial M_W}\right)^2 \Bigg/ E\left[\frac{\partial^2 \ln L}{\partial M_W^2}\right] = 0$$

So the information inequality is still satisfied.

Extended ML example

Consider two types of events (e.g., signal and background) each of which predict a given pdf for the variable x: $f_s(x)$ and $f_b(x)$.

We observe a mixture of the two event types, signal fraction = θ, expected total number = ν, observed total number = n.

Let $\mu_s = \theta\nu$, $\mu_b = (1-\theta)\nu$, goal is to estimate θ_s, θ_b.

$$f(x;\mu_s,\mu_b) = \frac{\mu_s}{\mu_s+\mu_b} f_s(x) + \frac{\mu_b}{\mu_s+\mu_b} f_b(x)$$

$$P(n;\mu_s,\mu_b) = \frac{(\mu_s+\mu_b)^n}{n!} e^{-(\mu_s+\mu_b)}$$

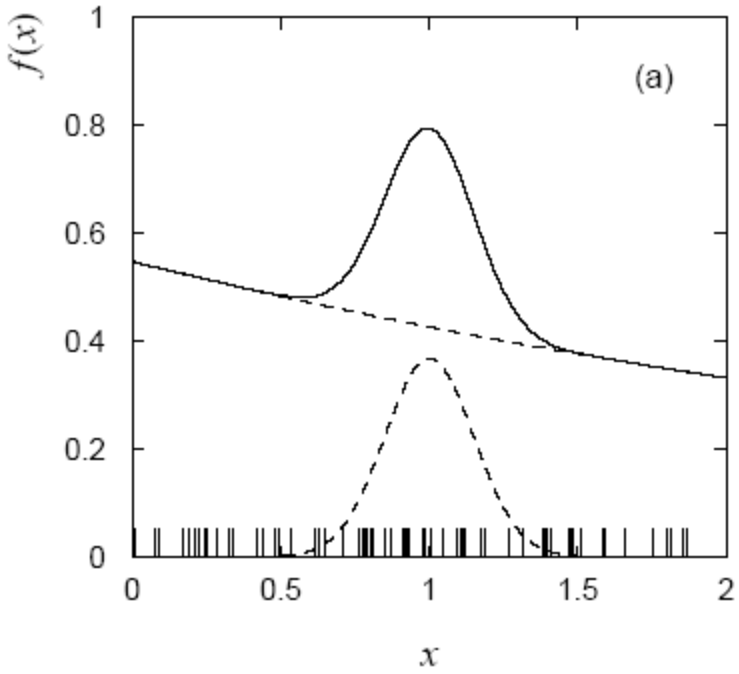
$$\rightarrow \quad \ln L(\mu_s,\mu_b) = -(\mu_s+\mu_b) + \sum_{i=1}^{n} \ln\left[(\mu_s+\mu_b) f(x_i;\mu_s,\mu_b)\right]$$

Extended ML example (2)

Monte Carlo example with combination of exponential and Gaussian:

$$\mu_s = 6$$

$$\mu_b = 60$$



Maximize log-likelihood in terms of μ_s and μ_b:

$$\hat{\mu}_s = 8.7 \pm 5.5$$

$$\hat{\mu}_b = 54.3 \pm 8.8$$

Here errors reflect total Poisson fluctuation as well as that in proportion of signal/background.