Post Lab: Fitting Data to an Assumed Distribution

1.

In[984]:=
```
y[x_] := RandomVariate[PoissonDistribution[x], 1000];
```

2.

In[1279]:=
```
distributionMean = 5;
Data = y[distributionMean];
DataCounts = BinCounts[Data, {0, 40, 1}];
bins = Table[i, {i, 0, Length[DataCounts] - 1}];
TwoDList = Multicolumn[Join[bins, DataCounts], 2] // First
```

Out[1283]=
```
{{0, 8}, {1, 26}, {2, 89}, {3, 123}, {4, 158}, {5, 175}, {6, 169}, {7, 127},
 {8, 60}, {9, 40}, {10, 15}, {11, 8}, {12, 2}, {13, 0}, {14, 0}, {15, 0},
 {16, 0}, {17, 0}, {18, 0}, {19, 0}, {20, 0}, {21, 0}, {22, 0}, {23, 0},
 {24, 0}, {25, 0}, {26, 0}, {27, 0}, {28, 0}, {29, 0}, {30, 0}, {31, 0},
 {32, 0}, {33, 0}, {34, 0}, {35, 0}, {36, 0}, {37, 0}, {38, 0}, {39, 0}}
```

3.

In[1284]:=
```
TotalPoints = Sum[DataCounts[[i]], {i, 1, Length[DataCounts]}]
NormedDataCounts = DataCounts / TotalPoints;
NormedTwoDList = Multicolumn[Join[bins, NormedDataCounts], 2] // First;
```

Out[1284]= 1000

4.

In[1287]:=
```
DataMean =
 Sum[(bins[[i]]) * DataCounts[[i]], {i, 1, Length[DataCounts]}] / TotalPoints // N
```

Out[1287]= 5.085

5.

In[1288]:= $g = \frac{1}{\sqrt{2\,\sigma^2\,\pi}}\, e^{-\frac{(x-\text{DataMean})^2}{2\,\sigma^2}};$

In[1289]:=
```
f = FindFit[NormedDataCounts, g, {σ}, x];
sig = f[[1, 2]]
(*mn=f[[2,2]]*)
```

Out[1290]= 2.41704

6.

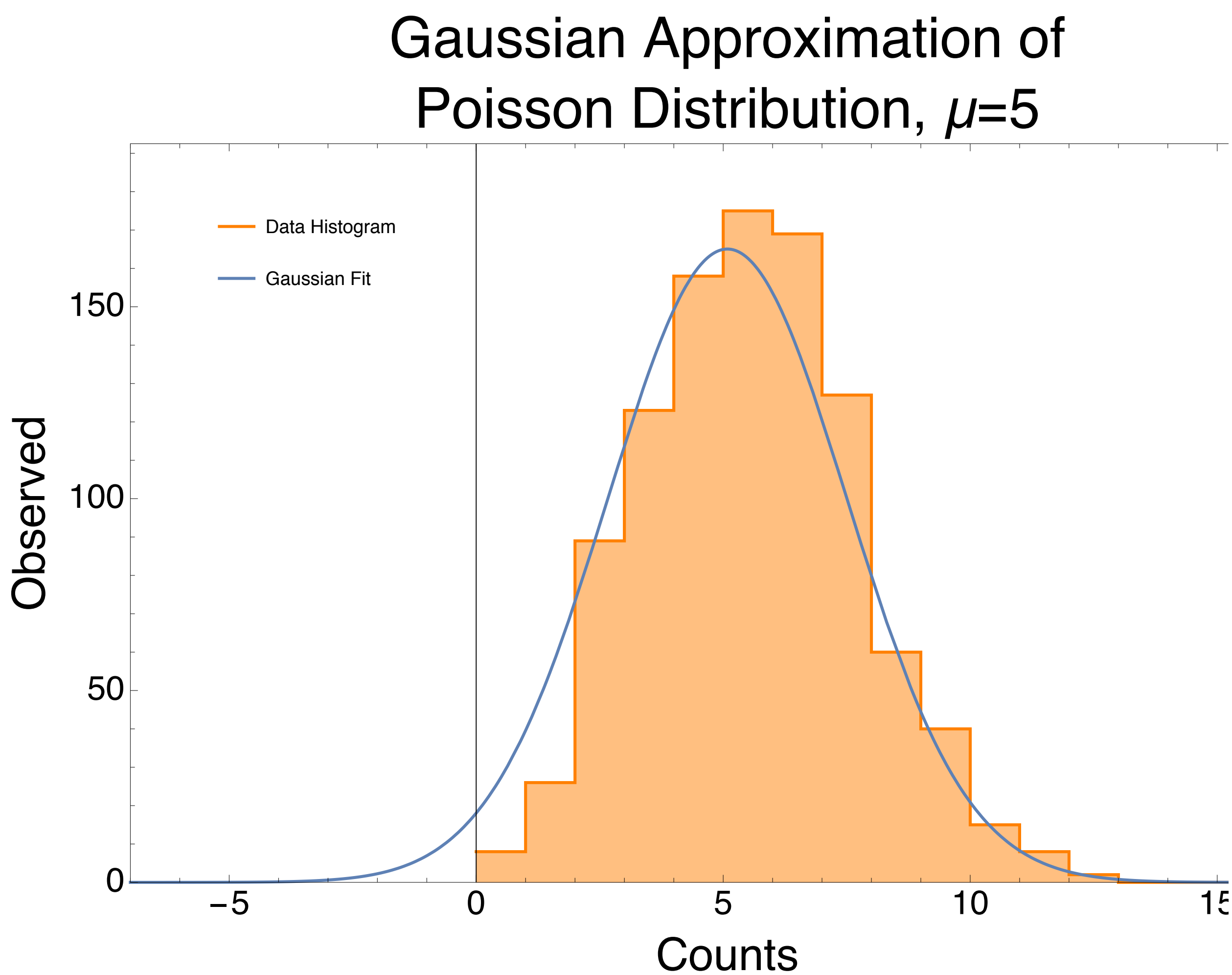
In[1291]:=
```
DataMean
```

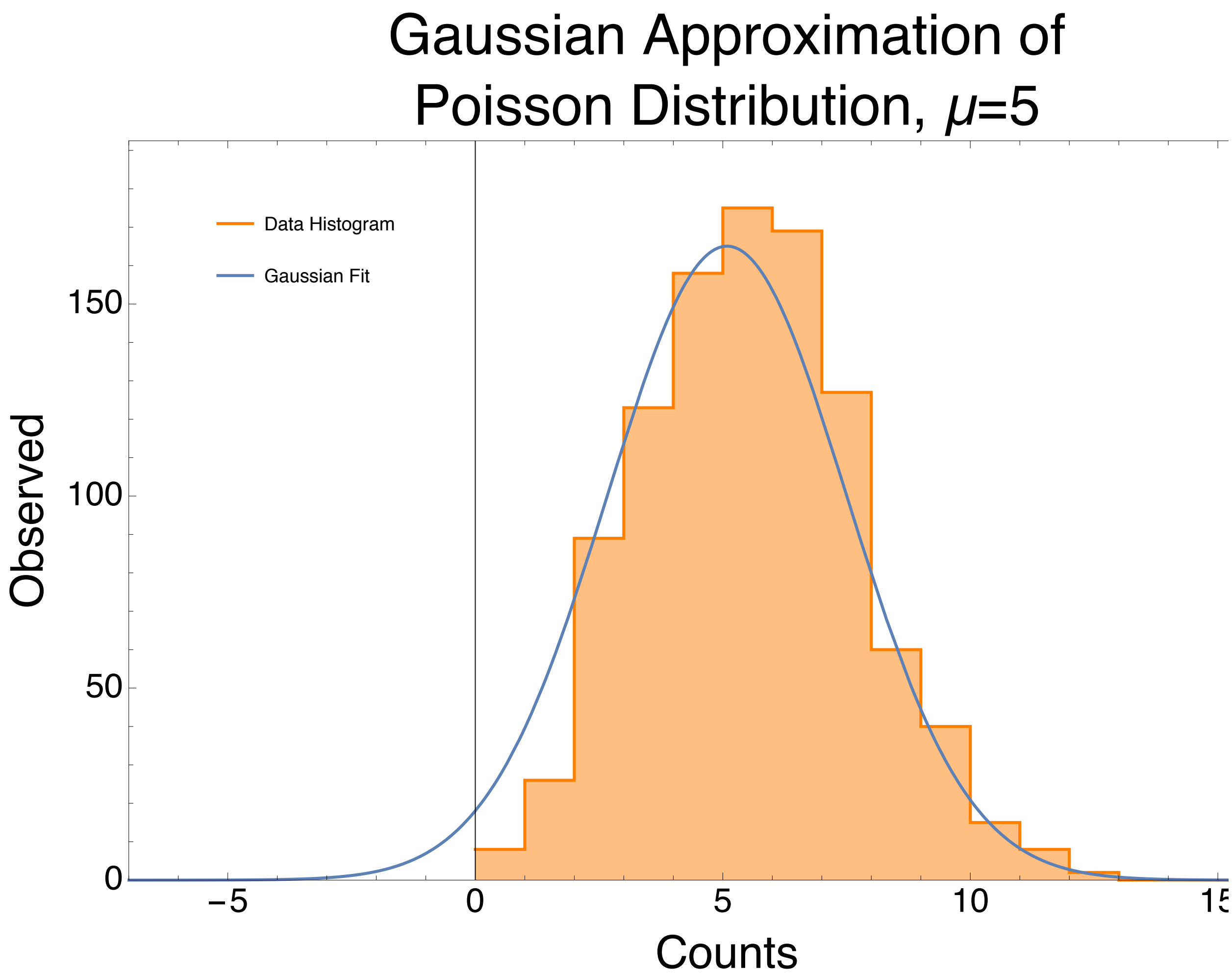
Out[1291]= 5.085

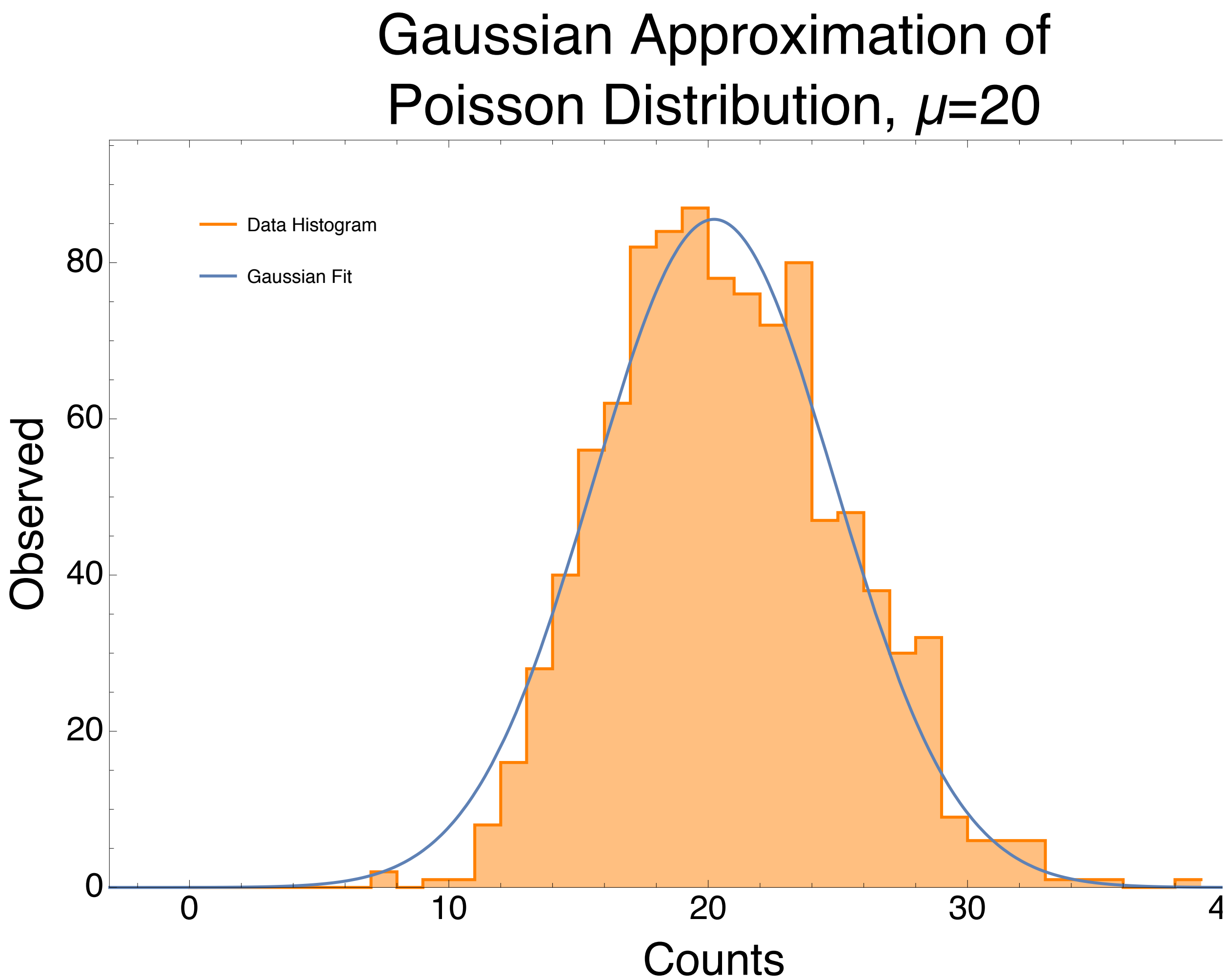
In[1292]:= $\text{myGauss}[x\_] := \frac{1}{\sqrt{2 * (\text{sig})^2 * \pi}}\, e^{-\frac{(x-\text{DataMean})^2}{2\,(\text{sig})^2}}$

In[1293]:=
```
(*Abscissa labels correspond to the value of the bin. The plotted
 Gaussian curve values correspond to the center of each bin*)
histTwo = Histogram[Data];
histThree = ListPlot[TwoDList,
   PlotRange → {{DataMean - 5 * sig, DataMean + 5 * sig}, {0, Max[TwoDList] * 1.1}},
   Joined → True, InterpolationOrder → 0, PlotRange → Automatic, Filling → Axis,
   FillingStyle → {Lighter[Orange, 0.5]}, PlotStyle → {Orange}, AxesOrigin → {0, 0},
   FrameLabel → {Style["Counts", Black, 24], Style["Observed", Black, 24]},
   Frame → True, PlotLabel →
    Style[StringJoin["Gaussian Approximation of\nPoisson Distribution, μ=",
      ToString[distributionMean]], Black, 30],
   FrameTicksStyle → Directive[Black, 18], ImageSize → 700,
   PlotLegends → Placed[{"Data Histogram"}, {0.15, 0.85}], PlotMarkers → ""];
curve = Plot[Length[Data] * myGauss[x], {x, DataMean - 5 * sig, DataMean + 5 * sig},
   AxesOrigin → {0, 0}, PlotLegends → Placed[{"Gaussian Fit"}, {0.15, 0.85}]];
Show[histThree, curve]
```

Out[1296]=




Gaussian Approximation of
Poisson Distribution, μ=5
Data Histogram
Gaussian Fit
Observed
150
100
50
0
−5
0
5
10
15
Counts



In the μ=5 case, the Gaussian fit seems to be shifted to the right of the histogram, probably due to the fact that observed counts will always be ≥0 while the Gaussian extends to infinity in each direction (even including only $\pm 5\sigma$ means that the graph will extend in the approximate range [-6.18,16.2]). In the μ=20 case this has less impact, and the Gaussian fits the data better. This is to be expected since as $N \longrightarrow \infty$, the Poisson distribution approaches the Gaussian distribution.