Exploring Distributions

- Use `disttool` to display a distribution and manipulate its parameters.
- Switch between CDF and PDF.
- Click and drag a slider arrow to change a parameter.
- Try a Gamma distribution.
- Click and drag in the figure window to measure the value of the distribution (dashed red line will slide as you move the mouse).

Random Distributions

- Use `randtool` to explore random distributions.
- Select Poisson distribution.
- Click the Resample button a few times.
- Increase the sample size to 1000 samples.
- Try resampling now.
- Click the Export button to export samples to the workspace.
  ```matlab
  hist(poissrv)
  plot(sort(poissrv), 'o')
  ```

Quick Line Fitting

```matlab
clear all
x = 0 : 0.1 : 5;
y = 3*x + rand(1, length(x));
scatter(x, y)
scatter(x, y, 20, 'r', 'filled')
lsline
```

Polynomial Fitting

```matlab
clear all
x = -10 : 0.25 : 10;
y = x.^3/10 – x.^2 + 5*sin(x);
polytool(x,y)
    Try using a Degree of 2, then 3.
    Click the Export button.
beta
clf, plot(x, y, 'LineWidth', 2)
hold on, plot(x, polyval(beta, x), 'ro')
```

Anonymous Functions

```matlab
f = @(x) 1 ./ (1+exp(-x.^2))
whos f
f(5)
f()
f(-1:4)
```
Plot Tools: Data Statistics

cf
x = randn(1000,1);
y = 1 - f(x)
plot(x, y, 'ro')

Undock the figure if it is docked.
Select Tools > Data Statistics, then check boxes for x mean and x std. dev.

Neurophysiology Exercise (cont.)

• Notice the two variables in your workspace.
• Make a diameter-vs-velocity scatter plot.
• Fit a line to this data using the Basic Fitting tool.
• What is the predicted conduction velocity of an axon 22 microns in diameter?
• What diameter value would give a conduction velocity of 6 meters/second?

Plot Tools: Basic Fitting

Select Tools > Basic Fitting
Check “cubic”
Check “Show equations”
Click the ➔ button
Click the next ➔ button
In the “Find y=f(x)” panel, enter -2 : 0.5 : 1.5
Click the Evaluate button, then check “Plot evaluated results”

Fitting A Gaussian

• Load a dataset of gasoline prices:
  clear all, load gas.mat
• Type the value of price2
• Let’s look at the distribution of values:
  hist(price2)
• Calculate some statistics:
  n = length(price2)
  mu = mean(price2)
  sigma = std(price2) * sqrt((n-1)/n)

Neurophysiology Exercise

• How does axon diameter in microns relate to conduction velocity in meters/second?

!wget www.cs.cmu.edu/~dst/Tutorials/Matlab/hursh.csv

type hursh.csv
clear all
Select: Import Data from the toolbar
Select the file hursh.csv
Select Import as: Column Vectors
Click the Import Selection button

Plot the Gaussian

x = min(price2) : 0.25 : max(price2)
y = normpdf(x, mu, sigma);
scaled_y = y * 4/max(y);
hold on
plot(x, scaled_y, 'r')

Could also do: histfit(price2, 10)
What Is the Likelihood?

- We estimated the mu and sigma parameters based on a small sample size (20 points).
- The true distribution may differ from our estimate.
- If we change mu and/or sigma slightly, how well does the new distribution fit our dataset?

Calculating Likelihood

function z = gauslike(mu, sigma, points)
    n = length(points);
    z = ones(size(mu));
    for i = 1 : n
        z = z .* normpdf(points(i), mu, sigma);
    end
end

Plot the Likelihood Surface

mus = 116 : 0.5 : 121;
sigmas = 2.5 : 0.1 : 5.5;
[x,y] = meshgrid(mus, sigmas);
z = gauslike(x, y, price2);
clf, surf(x, y, z*10)
xlabel mu
ylabel sigma
zlabel Likelihood
rotate3d on

Contour Plot of Likelihood

figure
[c,h] = contour(z*1e24);
clabel(c,h)

Interactive Contour Plot

fsurfht('gauslike', [116 122], [2.5 5.5], price2)
- Click and drag to move the crosshairs.
- Type the mean 118.5 into the X Value box.
- Type the sigma value 3.6401 into the Y Value box. Note the Z Value is 2.8386e-24
- Compute std(price2) and type that value into the Y-value box: the Z Value decreases.
- The peak is located at the sample mean, but not at the sample's standard deviation.
- Moral: the sigma value giving the greatest likelihood may not be the sample's sigma.

Nonlinear Regression

- Matlab can “tweak” parameters to fit an arbitrary model to a data set.
- First step: choose a model and determine its set of parameters. Example: a constant term plus an exponential function plus a noise term:
  \[ y_i = a_1 + a_2 \exp(-a_3 x_i) + \epsilon_i \]
- Write a Matlab function to evaluate the model given a parameter vector a and data x:
  \[ mdl = @(a,x) \ a(1) + a(2)\exp(-a(3)\times) \]
**Generate Some Test Data**

```
true_a = [1; 3; 2]
x = exprnd(2.5, 100, 1);  % 100x1 exp. distrib.
noise = normrnd(0, 0.1, 100, 1);
y = mdl(true_a, x) + noise;
scatter(x,y)
```

**Fitting The Model to the Data**

- Need a starting point for the parameter vector.
- Doesn't have to be accurate; just guess.
  ```
a_guess = [2; 2; 2]
```
- Now use nlinfit to estimate the parameters:
  ```
a_hat = nlinfit(x, y, mdl, a_guess)
```
- Pretty close to true_a!

**Examining the Fit**

```
xrange = min(x) : 0.01 : max(x);
clf, hold on
scatter(x, y)
plot(xrange, mdl(a_hat, xrange), 'r')
```

**Box Plots**

```
load carsmall
MPG
Origin
boxplot(MPG,Origin)
doc boxplot
```

**Anova Example**

```
doc anova1

Scroll down to Example 2. Are steel beams as strong as special alloy beams?

Cut and paste the sample code into your Matlab command window. Then do:

anova1(strength, alloy)
```

**To Learn More**

- Browse the Statistics Toolbox documentation:
  ```
doc stats
```
- Take Rob Kass' course on statistics for computational neuroscience.