Which Distribution for My Data?

Bob Cochran

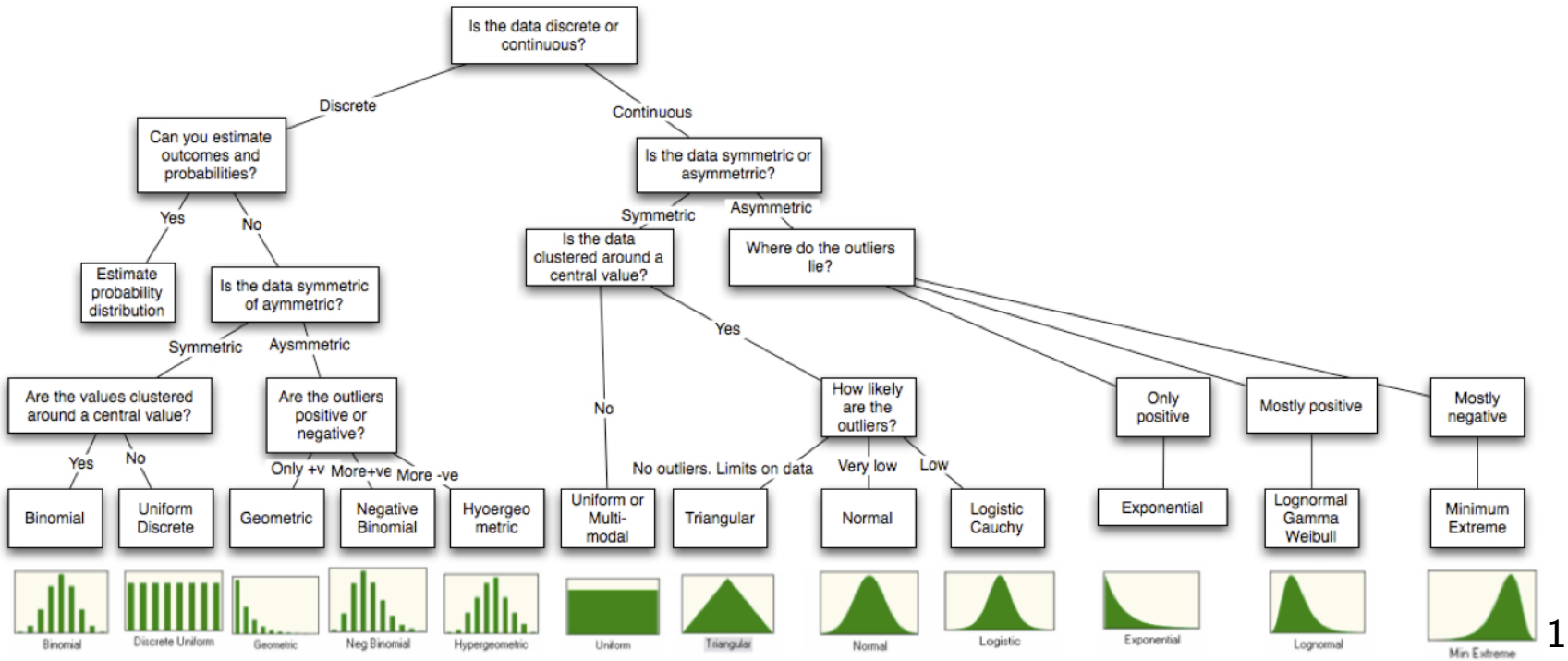
June 2, 2017

Which Distribution to Describe Uncertainty

- Your an engineer with data for a model parameter and need to decide which Dakota pdf to use
- Make use of the distribution fitting tools in R
  - See the MASS package
  - See the fitdistrplus package

A Starting Point

Figure 6A.15: Distributional Choices



[1]

[1] Probabilistic approaches to risk by Aswath Damodaran

Sample Convection Coefficient

#	h	#	h
1	8.43	11	9.58
2	10.7	12	11.6
3	10.5	13	10.9
4	10.3	14	10.7
5	9.46	15	9.65
6	8.82	16	9.50
7	10.1	17	9.32
8	11.6	18	9.75
9	10.8	19	10.4
10	10.6	20	9.90

Read the Data in From a File

```
getwd()          # get the current directory
setwd("path")    # set the current directory
list.files()     # list the files in the directory
dir()            # list the files in the directory

hdat <- read.table("hdata.dat", header=TRUE)  # header
hdat   # list what was read in
```

Process for `fitdistrplus` R Package

- load the package: `library(fitdistrplus)`
- Plot your data using: `plotdist(data, histo=TRUE, demp=TRUE)`
  - histogram on a density scale
  - empirical cumulative distribution function (CDF)
- `descdist(data, boot=1000)` provides classical descriptive statistics
  - Cullen and Frey plot of skewness and kurtosis for a variety of distributions
- `fit <- fitdist(data, "dist")`
  - dist is one of: `norm`, `lnorm`, `unif` and so-on ...
  - see R documentation on distributions

Plot the Raw h Data

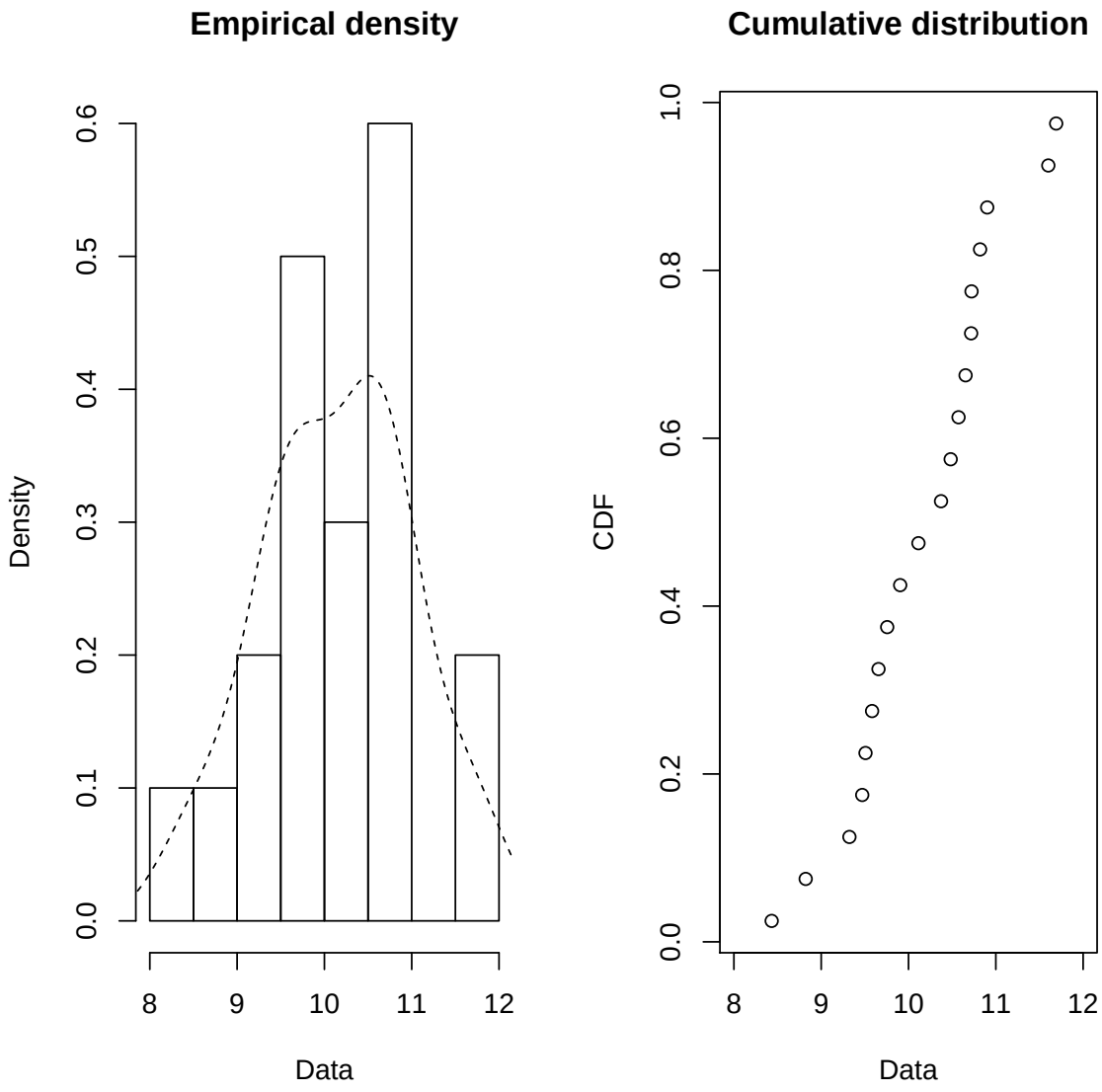


Figure 1: Distribution of h Data

Cullen and Frey Plot

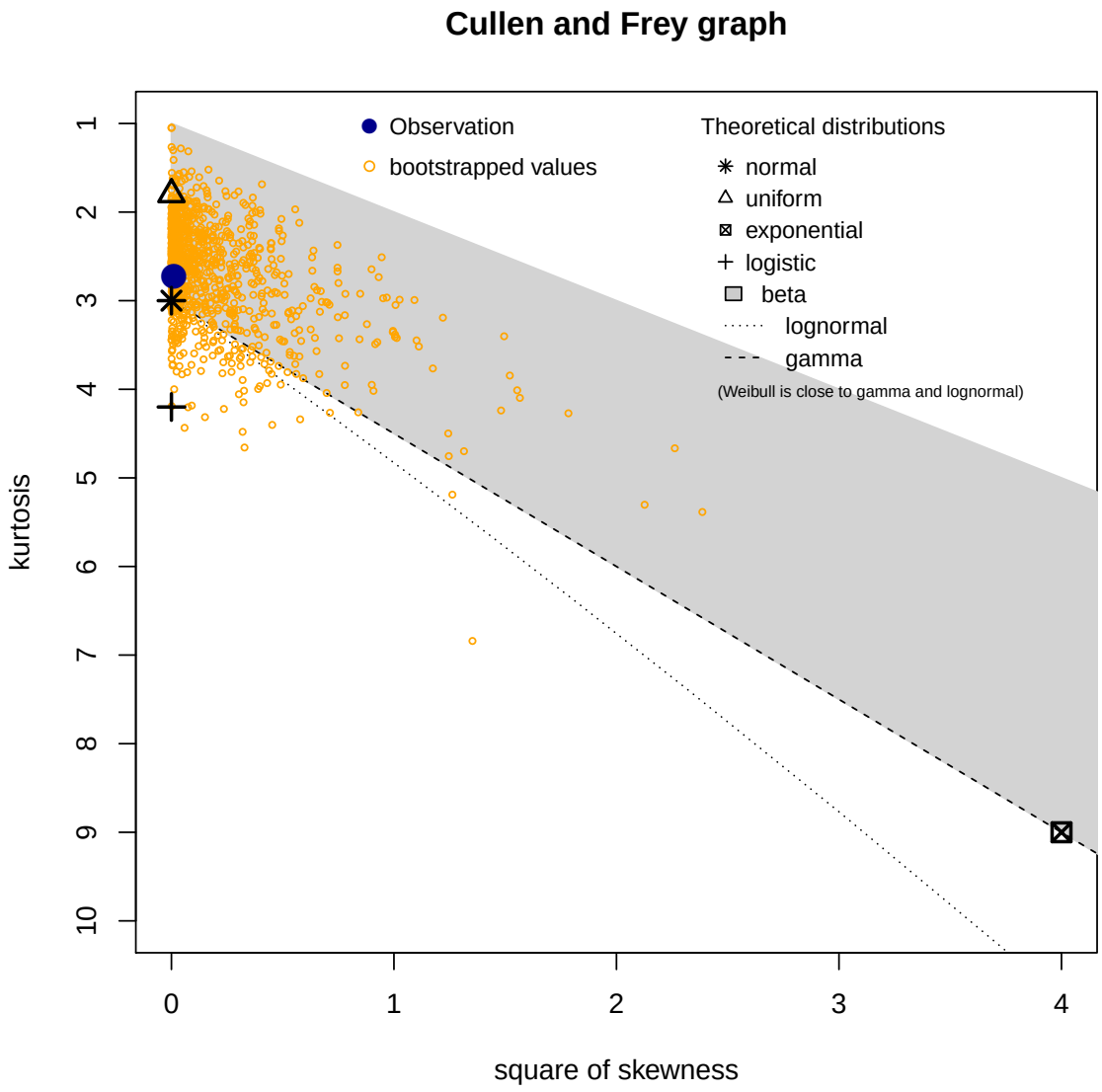


Figure 2: Cullen and Frey of h Data

Evaluate the Fit

- `summary(fit)` provides five quantitative measures:
  1. The parameter estimates
  2. The estimated standard errors
  3. The loglikelihood
  4. Akaic and Bayesian information criteria (AIC) and (BIC)
  5. The correlation matrix between parameter estimates
- `plot(fit)` provides four classical goodness-of-fit plots:
  1. A density plot
  2. A CDF plot
  3. A Q-Q plot
  4. A P-P plot

Goodness-of-fit Plots

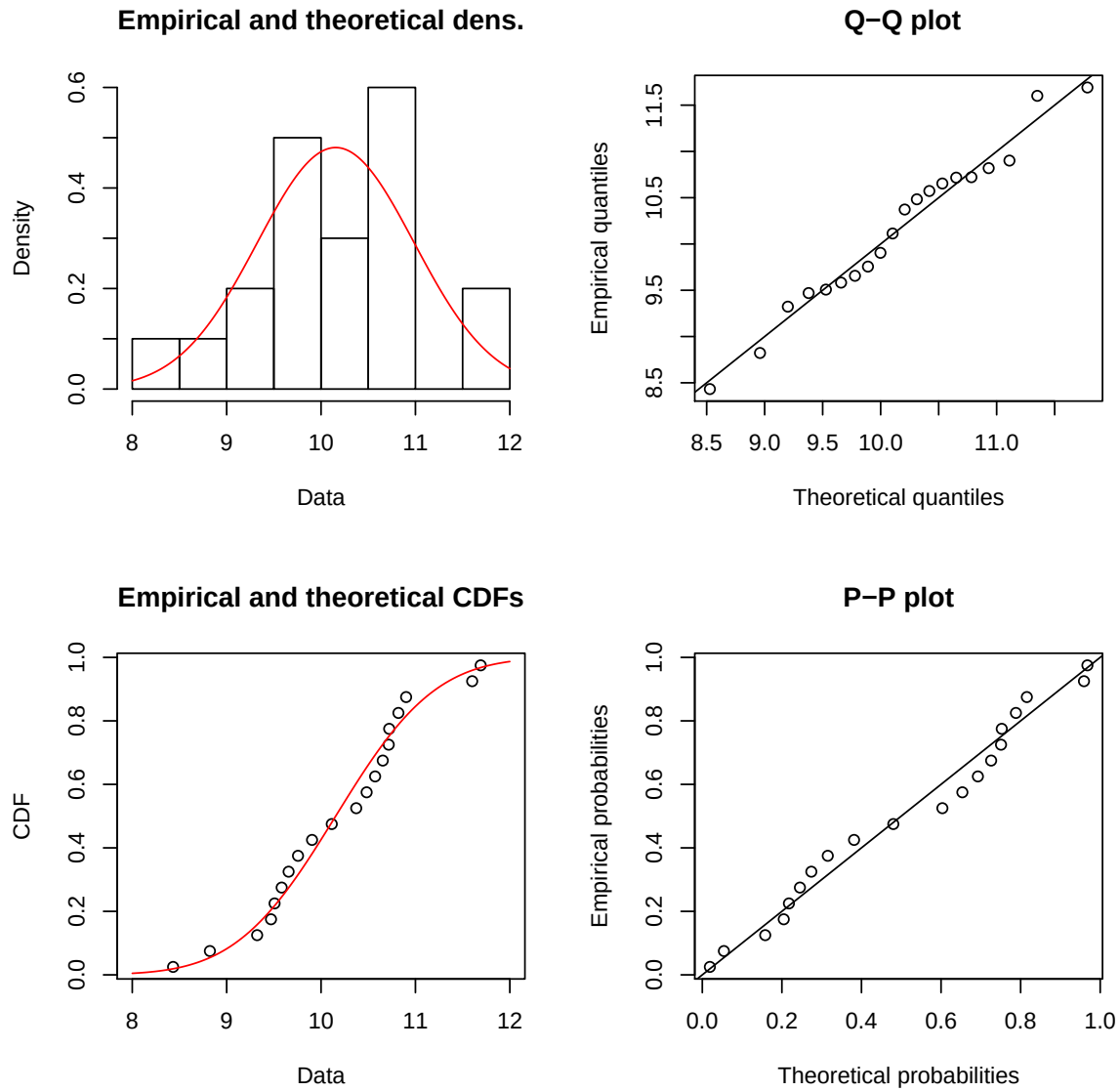


Figure 3: h Goodness-of-fit

Generating an Example Data Set

```
dataset <- rnorm(n = 20, mean = 10.0, sd = 2.0)
write.table(dataset, "dataset.txt")
```

Saving a Plot to PDF or PNG

```
pdf("filename.pdf")
plot( ... )
dev.off()

png("filename.png")
plot( ... )
dev.off()
```