Three Unrelated Topics:

Comparing Ranks through Simulation

Hierarchical Models

Quantile–Quantile (Q–Q) Plots

Comparing Ranks through Simulation

Want to make posterior inference on the relative ordering of the quality of care of the VA hospitals in 1992.

Observed data give one ranking, but no indication of uncertainty

- see the 6th ranked hospital
- many are very similar

Simulation with Transformations

- (Jointly) Ranks are a function of the proportions
  - Observed ranks are a function of observed proportions
  - Posterior for ranks is a function of the posterior for proportions
- A simulation from the posterior for proportions produces a simulation from the posterior for ranks
- Evaluate posterior for ranks via Monte Carlo

- Simulate a vector of proportions from the posterior

$$f_i|Y \sim Beta\,(y_i + 1, n_i - y_i + 1)$$

  $f_i|Y$'s conditionally independent

- Generate k samples from the posterior for proportions
- Transform each sample into a sample of posterior ranks

Conjugate Priors

- General principle: choose a prior which represents your beliefs
- In practice, use a convenient prior which approximately represents your beliefs
- Choose a family which is computationally convenient, but is also flexible
- A conjugate prior is one that leads to a posterior of the same form (it is likelihood-dependent)
- Examples:
  - Beta–binomial
  - Normal–normal
  - Gamma–gamma
  - Gamma–beta
  - Gamma–poisson

Why Hierarchical Models?

- Typical prior for binomial proportion parameter is a Beta
  - Beta is conjugate
  - Uniform is a special case
- What if the beta family is not flexible enough?
- What if our data are overdispersed?
  - Recall VA data are overdispersed

A Simple Hierarchical Model

- Typical prior is $Beta(a, b)$
- How do we choose a and b?
- Instead of choosing values for hyperparameters, treat them as unknown and also put priors on them
- Example:

$$
\begin{aligned}
y_i|\theta &\sim Bin(n_i, \theta) \\
\theta|\alpha, \beta &\sim Beta(\alpha, \beta) \\
\alpha|a, b &\sim \Gamma(a, b) \\
\beta|a, b &\sim \Gamma(a, b)
\end{aligned}
$$

- Advantages:
  - Hierarchical prior gives more flexibility to model
  - Posterior is less sensitive to choice of hyperparameters
- Disadvantages:
  - Model is harder to fit, as parameters are not "jointly conjugate"
    * Can marginalize
    * Can fit with MCMC
  - Parameters can be more difficult to interpret

A Different Hierarchical Model

- Account for overdispersion by fitting a separate θ_i for each VA hospital:

$$\begin{aligned} y_i|\theta_i &\sim Bin(n_i, \theta_i) \\ \theta_i|a, b &\sim Beta(a, b) \end{aligned}$$

- This is a random effects model
- Fit with MCMC
- By borrowing strength, can fit many parameters
- θ_i's are shrunk towards prior
- Could also put prior on hyperparameters

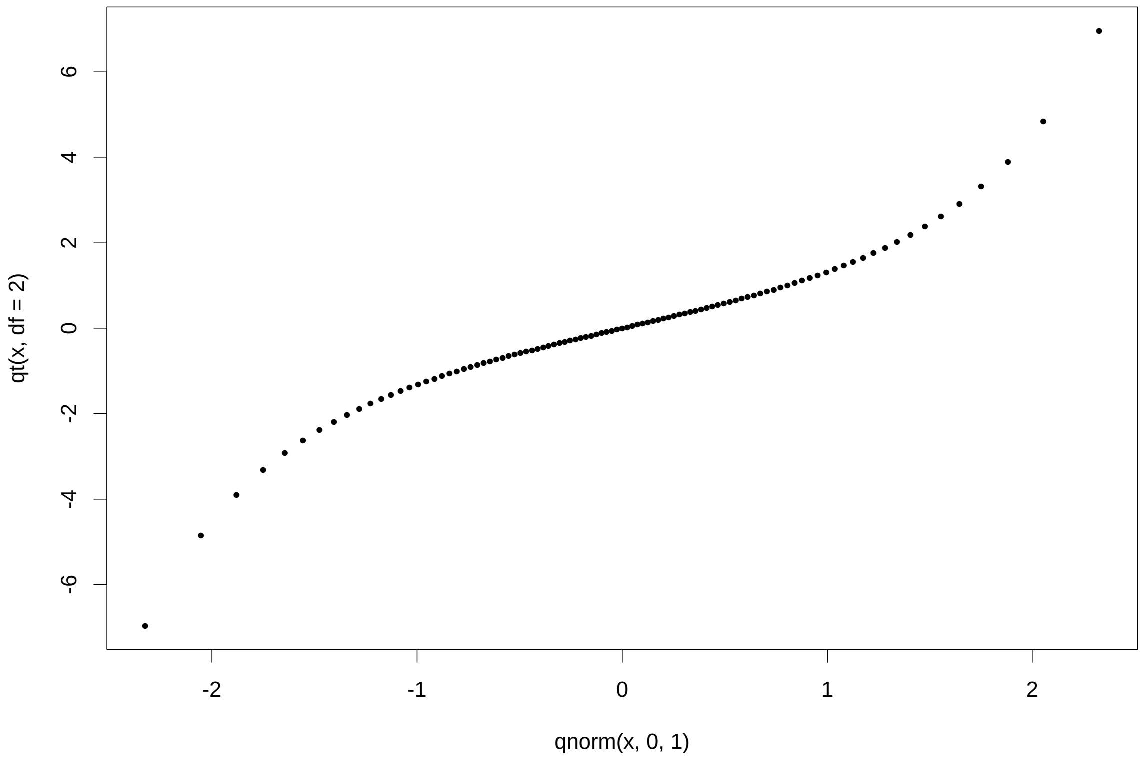
Quantile–Quantile (Q–Q) Plots

How do you compare two probability distributions?

Q–Q plots compare the quantiles of two distributions, either empirical or theoretical

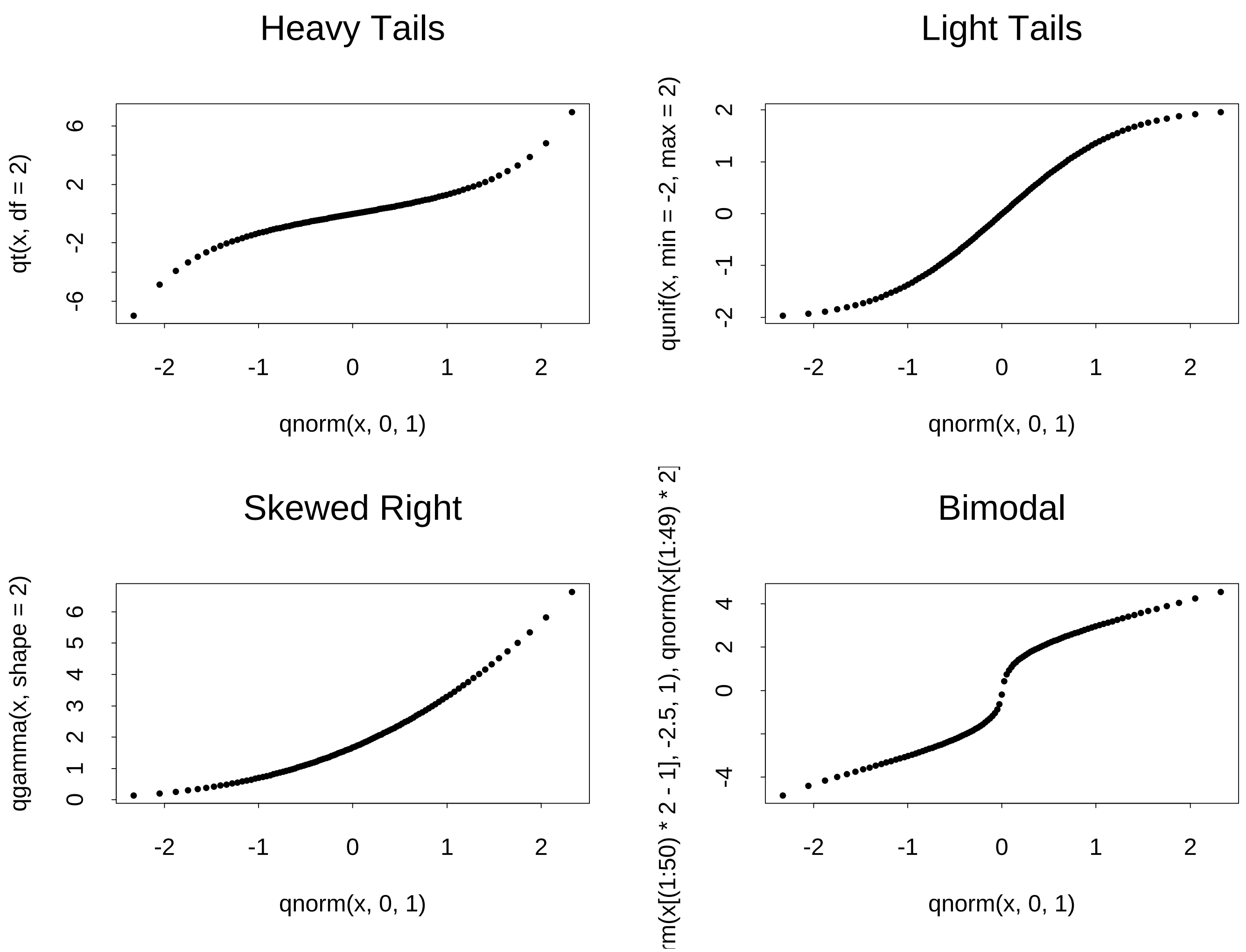
Simply plot quantiles of one distribution against the quantiles of the other

Example – normal versus t_2:



- If the two distributions are identical, all points will lie on the line $y = x$
- If the two distributions are from the same family but have a different central location, the line will have slope 1 but a non-zero intercept
- If the two distributions are from the same family but have a different variance, the line will pass through the origin but not have slope one
- If the two distributions are different, the line will not be straight – how it differs tells you how the distributions differ

Deviations from Normality



Q–Q Plots for Empirical Data

Plot

- quantiles of data, $x_{(1)}, \ldots, x_{(n)}$

against

- theoretical quantiles, $F^{-1}\left(\frac{i-\frac{1}{2}}{n}\right)$, $i = 1, \ldots, n$

Typically we are comparing to normality, although Q–Q plots are valid for all distributions