Class #11 Monday 21 February 2011

What did we cover last time?

- Discrete distributions
  - Binomial (N and p)
  - Geometric
  - Negative Binomial
  - Poissons
- Statistical Expectation
- Continuous distributions
  - Gaussian
    - Central Limit Theorem
    - Conditional Distribution
  - Gamma Distribution

3.4 Continuous Distributions cont...

❋(D) Beta Distribution

❋ Distributions from 0-1

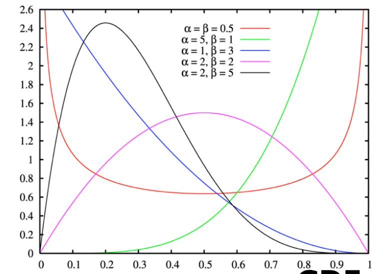
❋ Examples: Relative Humidity, cloud amount, probability, & sea ice

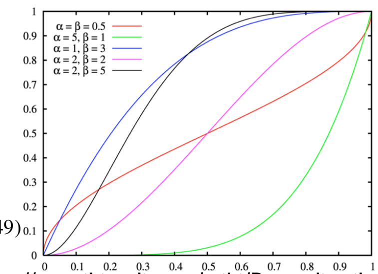
❋ p and q, 2 parameters, p=q symmetric,

❋ Use method of moments (get mean & var from sample).

❋ Can shift the limits [0,1] to [x,y].

$$f(x) = \left[\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}\right] x^{p-1}(1-x)^{q-1} \quad 0 \le x \le 1,\ p,q > 0 \quad (4.49)$$





http://en.wikipedia.org/wiki/Beta_distribution

3.4 Continuous Distributions cont...

❋(E) Extreme Value Distribution

❋ Describes behavior of largest of m values.

❋ Example: Table 4.7, 1951-1970, max precip (m=365)

❋ Extreme Types Theorem - the largest of m independent values , from a fixed distribution (General Extreme Value, GEV) will follow a known distribution as m increases.

❋ 3 parameters, location (ζ),scale (β) & shape (κ).

$$f(x) = \frac{1}{\beta}\left[1+\frac{\kappa(x-\zeta)}{\beta}\right]^{1-1/\kappa} \exp\left\{-\left[1+\frac{\kappa(x-\zeta)}{\beta}\right]^{1-1/\kappa}\right\},\ 1+\kappa(x-\zeta)/\beta > 0 \quad (4.54)$$

❋ Get parameters using Maximum Likelihood or L-moments. (choice depends on sample size)

❋ 3 special distribution cases of GEV

3.4 Continuous Distributions cont...

❋(E) Extreme Value Distribution cont...

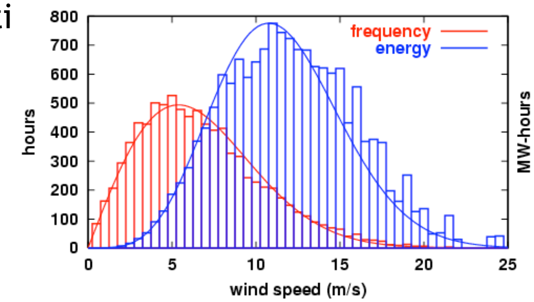
❋ 3 special distribution cases of GEV

- ❋ 1) Gumbel (Fisher-Tipett Type I)
  - ❋limiting case for κ==>0
  - ❋ use for extreme data drawn from distributions with well behaved tails (exponential)
  - ❋ Can estimate parameters using Method of moments
- ❋ 2) Freshet (Fisher-Tipett Type II)
  - ❋ κ>0
  - ❋ Have 'heavy tails', PDF decreases slowly for large values of x. The tail does not fall off as fast as for a Gaussian distribution.

Aside: Levy distribution, a stable distribution with heavy tails. (Stable means converges to self CLT).

3.4 Continuous Distributions cont...

- (E) Extreme Value Distribution cont...
- 3 special distribution cases of GEV cont...
  - 3) Weibull (Fisher-Tipett Type III)
    - κ<0
    - Fit with Maximum Likelihood or L-Moments
    - The Weibull model (alpha =2) closely mirrors the actual distribution of hourly wind speeds at many locati



http://en.wikipedia.org/wiki/Wind_power#Distribution_of_wind_speed

3.4 Continuous Distributions cont...

- (E) Extreme Value Distribution cont...
- Strong motivation for studying stats of extremes
- GEV requirements not always met (i.e. data collection changes over time, different causes of extreme rain) so check to make sure best to use.
- How to choose extremes?
  - Value in one year a max for another
  - Alternate: Pick n max out of total years (peak-over-threshold (POT) method). Serial correlation - take care of this.
- Extreme Probabilities as 'Average Return Periods', The CDF (F(x)) and avg sampling frequency (ω), BOARD, Go through Ex 4.10

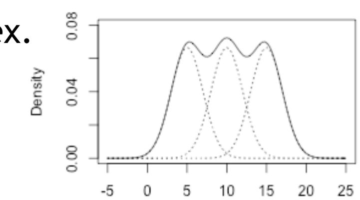
$$R(x) = \frac{1}{\omega[1 - F(x)]} \quad (4.62)$$

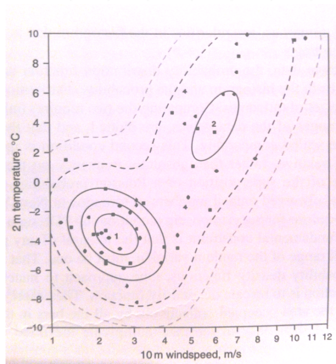
3.4 Continuous Distributions cont...

- (F) Mixture Distribution
- Use for Bimodal data, ENSO ex.
- Weighted average of a distribution, can combine any number, usually same type

$$f(x) = xf_1(x) + (1 - w)f_2(x) \quad (4.63)$$

- mean is weighted average
- variance more complex
- Fit with Maximum likelihood
- Gaussian most common but use two exponentials for nonzero daily precipitation.
- Ex: Two bivariate Gaussians, costs since many parameters to estimate.

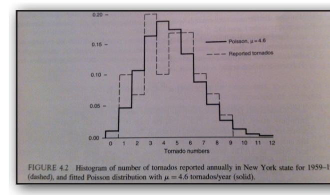




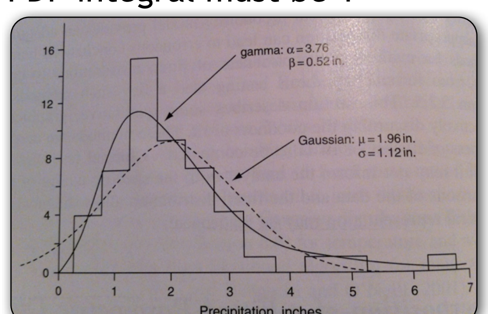
3.5 Qualitative Methods of Goodness of Fit

There exist quantitate methods for choice of distribution!

- (A) Superposition
- Simple, plot the sample values histogram & overlay the distribution & just make sure you have scaled properly.
- PDF integral must be 1



Tornados



Ithaca January rainfall
Gamma preferable

3.5 Qualitative Methods of Goodness of Fit

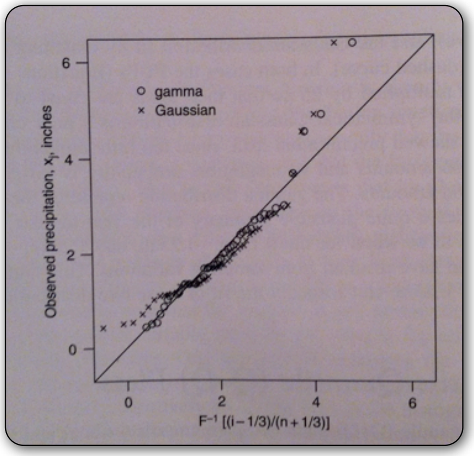
❋(B) Quantile-quantile plots

❋ Scatter plot, data value and it's estimate from quantile function of the fitted function (scaled).

❋ Perfect fit then all points fall on diagonal.

❋ large values underestimated

❋ Gamma better than Gaussian



3.6 Parameter fit using Maximum Likelihood

❋(A) Likelihood Function

❋Seeks values of parameters that maximize the Likelihood function.

❋ Bayesian: Most probable values of parameters given the data.

❋Likelihood function for Gaussian with n obs. is:

$$\Lambda(\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \prod_{i=1}^{n} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad (4.67)$$

❋ Looks like PDF for a Gaussian, so confusing!

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], -\infty < x < \infty \quad (4.23)$$

3.6 Parameter fit using Maximum Likelihood

❋(A) Likelihood Function cont...

❋ BOARD - how to solve...

❋(B) Newton-Raphson Method

❋ Gaussian MLEs were easy to calculate, usually calculate iteratively. Calculate roots.

❋ BOARD - example and how to solve...

❋ Example 4.12 Gamma Function

$$f(x) = \frac{(x/\beta)^{\alpha-1} \exp(-a/\beta)}{\beta\Gamma(\alpha)}, \quad x,\alpha,\beta > 0 (4.38)$$

❋(C) Expectation-Maximization (EM) Method

❋ Use for more than 3 parameters, more of an idea than a formulaic process. Book points to references for details on process.

3.7 Statistical Simulation

❋ Statistical Simulation: Generate 'seemingly' random numbers based on given PDF, EX: SAT forcing for an ocean model.

❋ Random number generator: really pseudo-random

❋(A) Uniform Random number generation

❋ Generate random uncorrelated samples between 0 and 1.

❋ How does this work? BOARD,

3.7 Statistical Simulation

❋(B) Nonuniform Random number generation

❋