Class #9 Monday 14 February 2011

What did we cover last time?

- Power Transformations
- Standardization of anomalies
- Paired data analysis
  - Scatter plots
  - Pearson's correlation coefficient
  - Spearman rank correlation
  - Kendall Tau test $\tau = \frac{N_C - N_D}{n(n-1)/2}$
  - Serial correlation, lag correlation
  - Autocorrelation (many variants)
  - correlation matrix

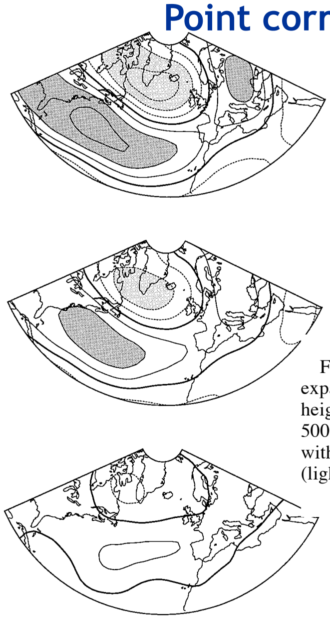
2.2.4 EDA for higher dimension data (Ref: Wilkes 3.6)

Transform data to:

- Reveal data features
- Adjust the distribution of data
- Variance stabilizing (reduce dependence of one variable on another)

1. Power Transformations
2. Standardization

Point correlation Plots 2-D



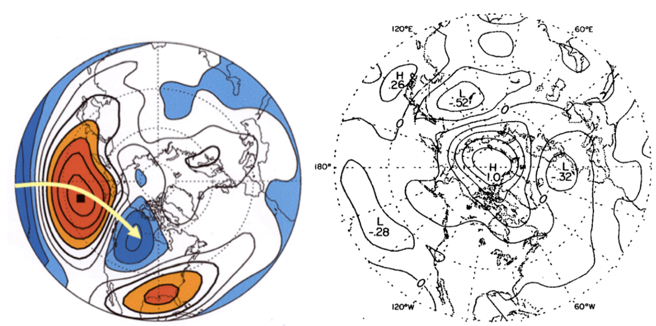
- ❋Useful to try to figure out sequence of events.
- ❋ Ocean lead atmosphere in tropics while atmosphere leads in midlatitudes.

FIG. 3. The 500-mb height correlations with the lag +2 SST SVD expansion coefficient, based on intraseasonal data: (top) 500-mb heights leading SST by 2 weeks, (middle) simultaneous, and (bottom) 500-mb heights lagging SST by 2 weeks. Contour interval = 0.2, with negative contours dashed and the zero contour darkened. Dark (light) shading indicates correlations >0.4 (<−0.4).

Deser and Timlin, 1998

One point correlation map

- ❋ Bjerknes, pressure one point correlation map
- ❋ Correlate variable at one point to same variable at all other points. Very useful.



Picture 2. Spatial distribution of correlation of the 500 Mb geopotential height anomaly time series (Seasonal JFM) at all points on the Northern hemisphere with the time series at a specified "base point" - North Pacific. Red colors positive correlation, blue colors negative correlation. Yellow arrow indicate meridional orientation of spatial structure existing in the correlation pattern. Picture courtesy of Prashant Sardeshmukh, CDC/OAR

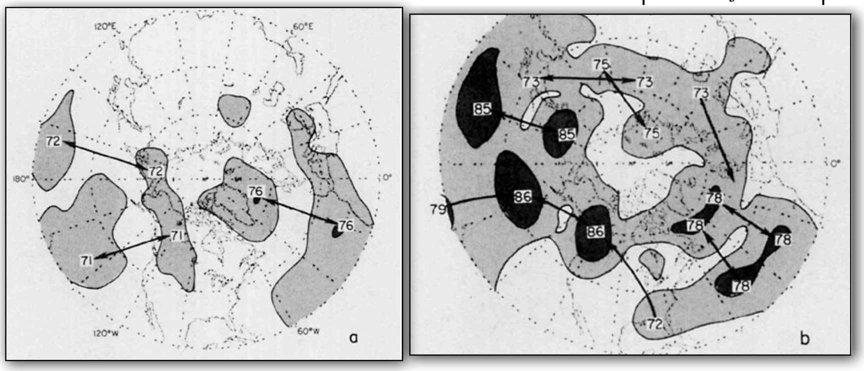
http://www.nws.noaa.gov/om/csd/pds/PCU2/statistics/Stats/part2/Corr_Space.htm

Teleconnectivity Map - One point Summary

5

- ✷ Wallace and Gutlzer classical paper
- ✷ Correlation matrix

$$T_i = \left| \min_j r_{i,j} \right|$$



(Fig 7 Wallace and Gutzler, 1981)

3.0 Parametric Probability Distributions (Ref: Wilkes Chapter 4)

6

- ✷ **3.1 Background (Ref: Wilkes 4.1)**
- ✷ **Idealize real data to theoretical distribution.**
- ✷ **Parametric idealizations are 'abstractions'**
- ✷ **Why useful?**
  - ✷ **Compactness (**reduce number of points so easier to work with**, Smoothing/Interpolation (**shows what is possible by filling in gaps**), & Extrapolation (**forecast probabilities outside your range of data**)**
- ✷ **Define parametric distribution (abstract mathematical form or characteristic shape, ex: Gaussian and bell shape)**
- ✷ **Parameters μ and σ,**

3.1 Background - Definitions

7

- ✷ Parameters vs. Statistics
  - ✷ Use Greek letters for parameters and Roman for statistics! Use μ and σ, m and s! Statistics are used to get parameters, Gaussian is a simple case.
- ✷ Discrete vs. Continuous Distribution
  - ✷ Two types of parametric distributions
  - ✷ Discrete - data that can take on only particular values (ex: 0 or 1, snow or no snow, one of three colors)
  - ✷ Continuous - data can take on any value in a range. Can be measured discretely but can be treated as continuous. (ex: T and p)

3.2 Discrete Distributions (Ref: Wilkes 4.2)

8

- ✷ (A) Binomial Distribution
  - ✷ One of two outcomes is possible at any time.
  - ✷ Calculate probability of N+1 trials if
    - a) Probability does not change from trial to trial.
    - b) Outcomes are mutually independent.

    Sore points: Diurnal cycle, Precip. can be highly correlated day-to-day.
- ✷Example: Coin Flip, p=0.5 Heads or tails (Par. N, p)
- ✷ X number of events and N number of trials, N=3 fair coins tossed, Want X number of heads (0,1,2,3 possible), get probability 1/8,3/8,3/8,1/8 (Board)

$$PR\{X = x\} = \binom{N}{x} p^x (1-p)^{N-x}, x = 0,1,2,3,...N \ (4.1), \quad \binom{N}{x} = \frac{N!}{x!(N-x)!} \ (4.2)$$

- (A) Binomial Distribution continued
  - Example 4.1 Lake Cayuga Freezing
    - Need p and N,
      - Freeze any next winter or any single winter then N=1, Freezing at least once in the next decade then N=10
      - Table 4.1 to get p, 10 freezes in 220 year p=0.045
      - once in 10 years, use Eq 4.1, get 0.30
  - Example 4.2 - Probability that it freezes at least once in 10 years. Sum of probabilities Pr{x=1}+Pr{x=2}+...Pr{x=10} = 0.37. Be clever and calculate 1-Pr{x=0}, since there are 11 possibilities in the next 10 years.
  - Can fit situations be be binomial! ex:Pr {T <32F}

3.2 Discrete Distributions (Ref: Wilkes 4.2)

- (B) Geometric Distribution
  - Like Binomial specifies the probabilities of a particular number of success in a given number of trials whereas Geometric specifies probabilities of the number of trials required to observe the next success. X- number of trials,

$$PR\{X = x\} = p(1-p)^{x-1}, x = 0,1,2,\ldots (4.5)$$

- Multiplicative law of probability, probability of success times x-1 consecutive failures.
- Aka ‘waiting distribution’, Has been applied to ‘dry spells’ or ‘wet spells’, waiting for wet or dry event, respectively.

3.2 Discrete Distributions (Ref: Wilkes 4.2)

- (C) Negative Binomial Distribution
  - Related to geometric distribution, parm p & k
  - Γ function, evaluate numerically

$$PR\{X = x\} = \frac{\Gamma(k+x)}{x!\Gamma(k)} p^k (1-p)^x, x = 0,1,2,\ldots (4.6), \Gamma(k+1) = k\Gamma(k)\,(4.8)$$

  - X is the number of failures before observing the kth success, so x+k is the waiting time for the kth success.
  - Ex 4.3

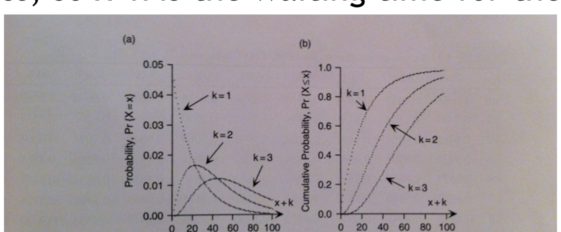


FIGURE 4.1 Probability distribution functions (a), and cumulative probability distribution functions (b), for the waiting time $x+k$ years for Cayuga Lake to freeze k times, using the Negative Binomial distribution, Equation 4.6.

- (C) Negative Binomial Distribution cont...
  - Find parameter values
  - Use first moment (mean) and second moment (variance), Since for the negative binomial distribution:

$$\mu = k(1-p)/p$$

$$\sigma^2 = k(1-p)/p^2$$

so

$$p = \frac{\bar{x}}{s^2} \quad 4.10a$$

$$k = \frac{\bar{x}^2}{s^2 - \bar{x}} \quad 4.10b$$

❋ (D) Poisson Distribution

❋ Describes a number of discrete events occurring in a sequence.Ex: The occurrence of Atlantic Hurricanes during a particular season.

❋ Param μ, average occurrence rate, intensity

$$\Pr\{X = x\} = \frac{\mu^x e^{-\mu}}{x!}, x = 0,1,2,\ldots(4.11)$$

❋ Use methods of moments to estimate μ, (mean occurrences per unit time).

❋ x is number of events, grows small if this number is big, makes sense.

❋ (D) Poisson Distribution

❋ Example 4.4, During 30 years there were 138 tornados, so μ is 138/30=4.6 tornados per year.

TABLE 4.3 Numbers of tornados reported annually in New York state, 1959–1988.

1959	3	1969	7	1979	3
1960	4	1970	4	1980	4
1961	5	1971	5	1981	3
1962	1	1972	6	1982	3
1963	3	1973	6	1983	8
1964	1	1974	6	1984	6
1965	5	1975	3	1985	7
1966	1	1976	7	1986	9
1967	2	1977	5	1987	6
1968	2	1978	8	1988	5

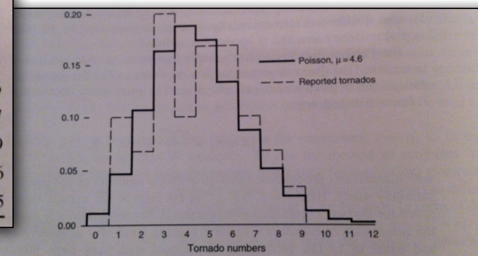


FIGURE 4.2 Histogram of number of tornados reported annually in New York state for 1959–1988 (dashed), and fitted Poisson distribution with $\mu = 4.6$ tornados/year (solid).

3.3 Statistical Expectations (Ref: Wilkes 4.3)

❋ Expected value of a random variable is the probability-weighted average of that variable.

❋

$$E[c] = c \quad (4.14a)$$

$$E[c\, g_1(x)] = c\, E[g_1(x)] \quad (4.14b)$$

$$E\left[\sum_{j=1}^{J} g_j(x)\right] = \sum_{j=1}^{J} E[g_j(x)], \quad (4.14c)$$

TABLE 4.4 Expected values (means) and variances for the four discrete probability distribution functions described in Section 4.2, in terms of their distribution parameters.

Distribution	Probability Distribution Function	$\mu = E[X]$	$\sigma^2 = Var[X]$
Binomial	Equation 4.1	Np	$Np(1-p)$
Geometric	Equation 4.5	$1/p$	$(1-p)/p^2$
Negative Binomial	Equation 4.6	$k(1-p)/p$	$k(1-p)/p^2$
Poisson	Equation 4.11	μ	μ

3.3 Statistical Expectations (Ref: Wilkes 4.3)

❋

TABLE 4.5 Binomial probabilities for $N = 3$ and $p = 0.5$, and the construction of the expectations $E[X]$ and $E[X^2]$ as probability-weighted averages.

X	$\Pr(X = x)$	$x \cdot \Pr(X = x)$	$x^2 \cdot \Pr(X = x)$
0	0.125	0.000	0.000
1	0.375	0.375	0.375
2	0.375	0.750	1.500
3	0.125	0.375	1.125
		$E[X] = 1.500$	$E[X^2] = 3.000$