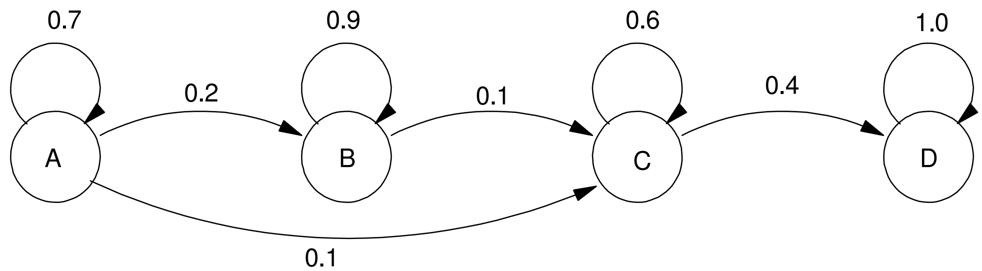
# ECE 341 - Test #3

Markov Chains and Data Analysis

Open-Book, Open Notes. Calculators, Matlab, Tarot cards, StatTrek allowed.

Giving or receiving help from others or from Chegg **not** allowed

**1) Markov Chains:** Four people are playing ball. Each second, a person either passes the ball or keeps it with probability p as shown below.




a) Express the probability that each person has the ball at time k+1 as:

$$X(k+1) = AX(k)$$

$$\begin{bmatrix} A(k+1) \\ B(k+1) \\ C(k+1) \\ D(k+1) \end{bmatrix} = \begin{bmatrix} 0.7 & 0 & 0 & 0 \\ 0.2 & 0.9 & 0 & 0 \\ 0.1 & 0.1 & 0.6 & 0 \\ 0 & 0 & 0.4 & 1 \end{bmatrix} \begin{bmatrix} A(k) \\ B(k) \\ C(k) \\ D(k) \end{bmatrix}$$

b) Assume the z-transform for person D having the ball is

$$Y(z) = \left( \frac{0.04z(z-0.7)}{(z-1)(z-0.9)(z-0.7)(z-0.6)} \right)$$

Find y(k) using z-Trasforms

Take the partial fractione expansion

$$Y = \left( \left( \frac{1}{z-1} \right) + \left( \frac{-1.3333}{z-0.9} \right) + \left( \frac{0}{z-0.7} \right) + \left( \frac{0.3333}{z-0.6} \right) \right) z$$

Take the inverse-z transform

$$y(k) = 1 - 1.3333(0.9)^k + 0(0.7)^k + 0.3333(0.6)^k \qquad k > 0$$

**2) t-Test (One data set).** A Monte-Carlo simulation was run for 8-card poker. Each simulation deals 100,000 hands of 8-cards. The number of times a hand contains 2-pair is recorded:

# hands = { 37625 37802 37611 37431 }

a) Determine the mean and standard deviation for this data

- **mean = 37,617**
- **std = 151.57**

b) (individual) If I run this experiment one more time, what number will I get with a confidence level of 90%? (5% tails)

5% tails with 3 degrees of freedom results in t = 2.355

$$\bar{x} - 2.355s < 2pair < \bar{x} + 2.355s$$

$$37,260 < 2pair < 37,974 \qquad p = 0.9$$

c) (population) From this data, what is the 90% confidence interval for the actual probability of getting 2-pair when dealt 8 cards?

Since you're dealing with the population, divide the variance by n

$$\bar{x} - 2.355\left(\frac{s}{\sqrt{4}}\right) < \mu < \bar{x} + 2.355\left(\frac{s}{\sqrt{4}}\right)$$

$$37,439 < \mu < 37,766 \qquad p = 0.9$$

**3) t-Test (Two data sets):** The global average temperature over two decades are as follows (source: NOAA):

| Time-Span | mean (milli-degrees F) | standard deviation (milli-degrees F) | # years |
|---|---|---|---|
| A: 1880 - 1889 | -176.58 | 80.98 | 10 |
| B: 1890 - 1899 | -243.58 | 92.92 | 10 |

a) (Individual) What is the probability that any given year in A is warmer than any given year in B?

Create a variable W = A - B

$$\bar{x}_w = \bar{x}_a - \bar{x}_b \qquad s_w = \sqrt{s_a^2 + s_b^2}$$

$$\bar{x}_w = 67.00 \qquad s_w = 123.255$$

The t-score for W (single tail test at W=0)

$$t = \left(\frac{\bar{x}_w - 0}{s_w}\right) = 0.5436$$

Degrees of freedom is about 9 (minimum dof for A and B)

From StatTrek, a t-score with 9 degrees of freedom is 70%

- **There is a 70.00% chance that a random year in A is warmer than a random year in B**
- **There is a 30.00% chance that a random year in A is colder than a random year in B**

b) (population) What is the probability that the temperature is rising? (mean of B is more than the mean of A)?

Here, you divide the variance by the sample size

$$s_w = \sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}} = 39.070$$

The t-score is now

$$t = \left(\frac{\bar{x}_w - 0}{s_w}\right) = 1.7190$$

This correspnds to a probability of 94.01%

- **It is 94.01% likely that decade A is warmer than decade B**
- **There is a 5.99% chance that decade A is colder than decade B**
-

4) Chi-Squared Test: The following Matlab code generated 100 random values for a 7 sided die.

```
Result = zeros(7,1);
for i=1:100
    die = ceil( 7*(rand ^ 0.8));
    Result(die) = Result(die) + 1;
    end
Result
```

a) Generate the frequency of rolling each number 1..7 with 100 rolls of the die

```
>> Result =    7    13    13    12    16    20    19
```

b) Determine if X has a uniform distribution (fair die) using a Chi-squared test.

| die roll | p | np | N | chi-squared |
|---|---|---|---|---|
| 1 | 1/7 | 14.28 | 7 | 3.71 |
| 2 | 1/7 | 14.28 | 13 | 0.11 |
| 3 | 1/7 | 14.28 | 13 | 0.11 |
| 4 | 1/7 | 14.28 | 12 | 0.36 |
| 5 | 1/7 | 14.28 | 16 | 0.21 |
| 6 | 1/7 | 14.28 | 20 | 2.29 |
| 7 | 1/7 | 14.28 | 19 | 1.56 |
| | | | **Total** | **8.36** |

From StatTrek, a chi-squared score of 8.36 with 6 degrees of freedom corresponds to a probability of 70.0%

**From the data, there is a 70% chance this is not a fair die**

5) ANOVA (Three data sets): The global average temperature over three decades are presented below. Determine the probability that the data sets have a different mean (temperatures are changing) using an F-test.

| Time-Span | mean (milli-degrees F) | standard deviation (milli-degrees F) | # years |
|---|---|---|---|
| A: 1880 - 1889 | -176.58 | 80.98 | 10 |
| B: 1890 - 1899 | -243.58 | 92.92 | 10 |
| C: 1900 - 1910 | -305.83 | 131.23 | 10 |

In matlab

```
Xa = -176.58;  %mean(A);
Va = 80.98^2;  %var(A);
Xb = -243.58;  %mean(B);
Vb = 92.92^2;  %var(B);
Xc = -305.83;  %mean(C);
Vc = 131.23^2;  %var(C);
Na = length(A);
Nb = length(B);
Nc = length(C);
k = 3;
N = Na + Nb + Nc
G = (Na*Xa + Nb*Xb + Nc*Xc) / N
MSSb = (Na*(Xa-G)^2 + Nb*(Xb-G)^2 + Nc*(Xc-G)^2) / (k-1)
MSSw = ((Na-1)*Va + (Nb-1)*Vb + (Nc-1)*Vc) / (N-k)
F = MSSb / MSSw


N =        30
G =     -241.9967
MSSb =     4.1783e+004
MSSw =     1.0804e+004
F =        3.8672
```

From StatTrek, an F-score with

- 2 degrees of freedom in the numerator
- 27 dof in the denominator
- F-score of 3.6872

results in p = 97%

**There is a 97% chance that these populations have different means**

- Back in 1890-1910, it looked like the eaeth was cooling off.
- There's some thought that, if left alone, we would be heading towards another ice age
- Meaning that global warming is so great, it's changing the natural cycles of the Earch