Another example of a priori error.

- Rainy days in Seattle in December.
- Calculate mean and std. dev. for each day
- Compare this to the monthly mean.
- If you find one day that is statistically sig. different, you must consider the probability of <u>every</u> day being significant - even though the data are independent.

Probability that none of the 31 days will pass 99% significance

$$(0.99)^{31} = 73\%.$$

Can also use composite analysis to compare variances in the F-test:

<u>Example</u>: Does QBO impact variance of T in the NH polar vortex.

40 yrs. of Jan. mean data

16 winters ⇒ east

16 winters ⇒ west.

$n_1 = n_2 = 16$

$s_1 = 12K$

$s_2 = 7.5K$

$$F = \frac{s_1^2}{s_2^2} = \frac{(12K)^2}{(7.5K)^2} = 2.56$$

Critical value for F-test = 3.52

$\therefore$ Conclude it is not significant.

df $\xrightarrow{N_2}$

F-table.

$N_1 \downarrow$

Sig. values.

Non-normal distributions

So far we've talked mainly about data that fit a normal distribution, or, if not, how to use a non-parametric test to evaluate.

if data are <u>not</u> normally distributed, then there are other theoretical distributions which can be used to fit data.

Table 4.6 Wilks p.87

Distribution	When to use
Gaussian (bell curve)	Data are approximately normally distributed.
Gamma Exponential Chi-square Pearson III	Data are skewed to one side.
Beta	Data restricted to bounded segments.
Extreme value	Focuses on the largest values in a given sample.

Focus on the gamma distribution and how it has been applied in characterizing precipitation via the standarized precipitation index. (refer to Don Edward's masters thesis notes)

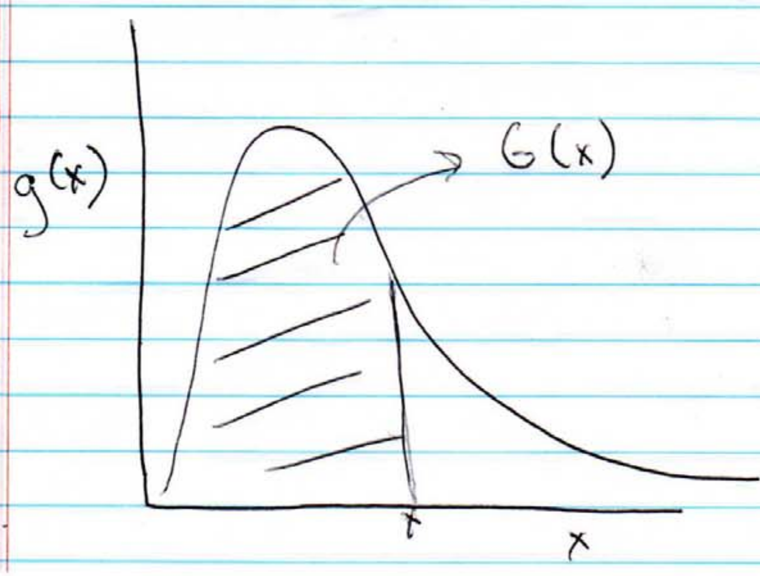
Gamma distribution pdf:

$$g(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}$$

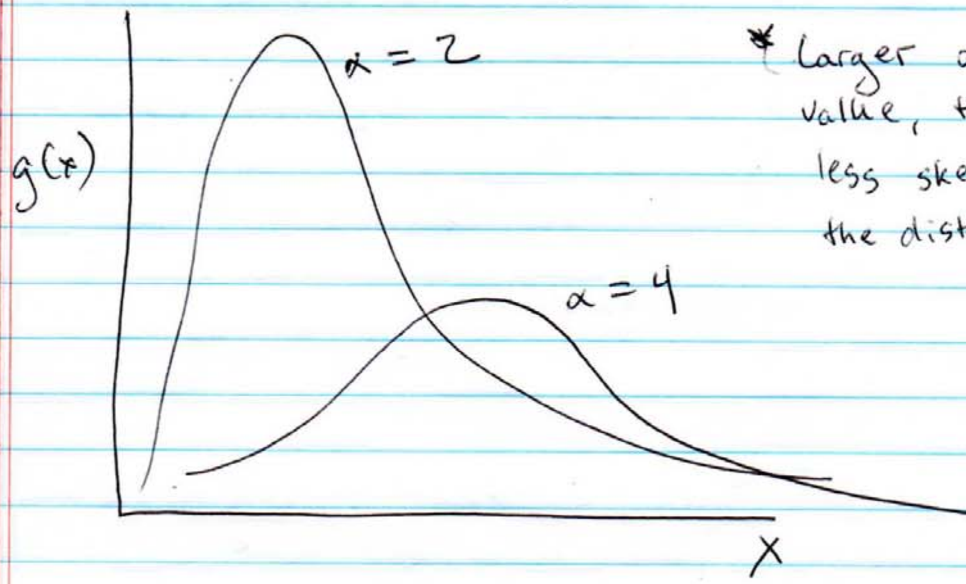
$\alpha > 0$ Shape parameter (α)
$\beta > 0$ β is a scale parameter
$x > 0$ given variable (precip.)

$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y}\, dy$ $\Gamma(\alpha)$ is gamma function.

Gamma distribution looks like:



With an increasing shape parameter α, the gamma distribution becomes more flat



* Larger α value, the less skewed the distribution.

Special limiting cases of gamma distribution

$\alpha = 1 \rightarrow$ Exponential distribution

$$g(x) = \frac{1}{\beta} \exp\left(\frac{-x}{\beta}\right) \quad (4.46)$$

Simply an exponential decay

$\beta = 2 \rightarrow$ Chi-square distribution

$$g(x) = \frac{x^{\alpha-1} \exp\left(\frac{-x}{2}\right)}{2^{\alpha}\,\Gamma(\alpha)}$$

Used to evaluate variance of normally dist. data.

SPI : Standardized precipitation index

Idea

- Precipitation typically does not fit a normal distribution, especially for dry climates
- Instead of gaussian distribution, fit a gamma probability density function to a given frequency distribution of precip. totals for a station (for a good many years of obs)
- Totals can be based on whatever time scale is desired (typically from 3 to 48 months)
- Cumulative probability of each precip. value transformed to a standard normal z score (ranges from typically -3 to 3)

Power of the method

- Allows for computation of a normalized index which indicates generally whether an area is dry or wet.
- Can be computed for different timescales, so can characterize short or long term wetness or dryness.
  → Different timescales will apply to different stakeholders !

if shown as a spatial plot, can compare areas with very different climatological characteristics in rainfall.

Example

Subtropical desert

- Tucson, AZ: Gets approximately 6 inches of rain during the monsoon season, so a total rainfall of 8" for the summer would be "wet"

Moist mid-latitude

- Chicago, IL: Gets approximately 15-20" of rain during the corresponding period (and its more normally distributed), so a total rainfall of 8" for the summer would be "dry"

Show plots of SPI from the western regional climate center..

Steps in computing SPI

1) Compute α and β parameters

$$\alpha = \frac{1}{4A}\left(1+\sqrt{1+\frac{4A}{3}}\right) \rightarrow \text{Thom estimator approx.}$$

$$\beta = \frac{\bar{x}}{\alpha}$$

$$A = \ln(\bar{x}) - \frac{\sum_{i=1}^{N} \ln(x_i)}{n}$$

Difference between nat. logs of ~~geomet~~ arithematic & geometric means.

x = precipitation observations (for a given period)

$\bar{x}$ = mean of precipitation observations

To be robust, need to use a relatively long base period to compute α and β, typically 30 years worth of data.

2) Compute the cumulative probability for each precipitation value

$$G(x) = \int_0^x g(x)\,dx = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} \int_0^x x^{\alpha-1} e^{-x/\beta}\,dx$$

Cumulative probability is:

$$H(x) = q + (1-q)\,G(x)$$

Where q is the probability of a zero (for which the gamma dist. is undefined)

3) Transform cumulative probability to standard normal variable z with a mean of zero and variance of 1.
This defines the SPI value.
(e.g. p. 22 of Edward's notes)

The specific equations for the transformations are:

p.24 Edwards

$$Z = SPI = -\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right)$$

$$0 < H(x) \leq 0.5$$

$$Z = SPI = +\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right)$$

$$0.5 < H(x) < 1.0.$$

The specific constants are given in Edwards masters thesis, uses an approximation developed by Abromowitz and Stegun (1965)

$$t = \sqrt{\ln\left(\frac{1}{(H(x))^2}\right)} \quad \text{for } 0 < H(x) < 0.5$$

$$t = \sqrt{\ln\left(\frac{1}{(1.0 - H(x))^2}\right)} \quad \text{for } 0.5 < H(x) < 1.0$$

4) Read SPI value and make physical inferences.

Conceptually, what this is doing...

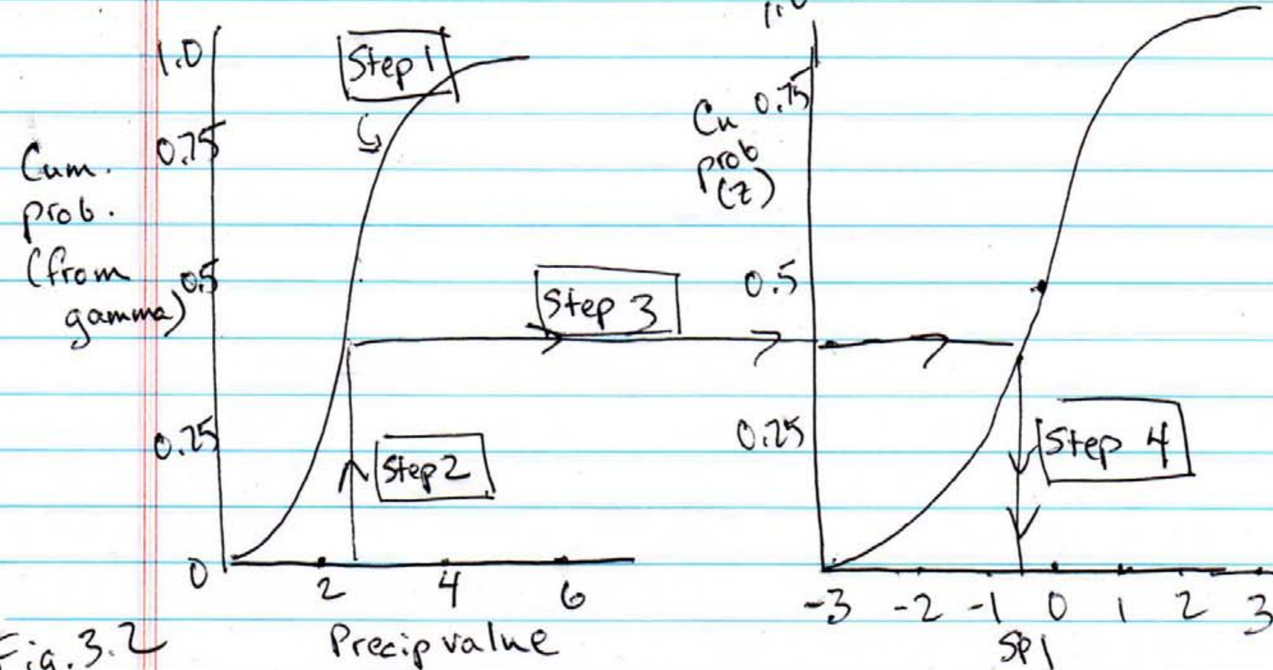


Fig. 3.2

Transforming the cumulative probability from a gamma distribution to a std. normal distribution.

Western Regional Climate Center Classification:

+3	Exceptionally wet
2-3	extremely wet
1.25-2	Very wet
0.75-1.24	Moderately wet
-0.74 to 0.74	Near normal
-1.25 to -0.75	Mod. dry
-2 to -1.25	Ext. dry
-3	Very dry

WRCC Bases their SPI statistics on the last 30 years of climate data.

Though this technique has mainly been used for precipitation, the statistical methodology behind it can be applied to any geophysical variable which may have positively skewed distributions.

Should be able to program this and compute SPI – homework assignment.