Statistical Forecasting

Wilks, 2011: “Statistical Methods in the Atmospheric Sciences”, Sections 6.2 and 6.3

Strengths of Statistical Forecasting

- Statistical methods are useful and necessary components of the forecasting enterprise
  - Simple, low computational cost
  - Help identify sources of predictability
  - Can serve as a useful benchmark for dynamical prediction
- Statistical methods are routinely used to post-process and enhance the results of dynamical forecasts

Wilks 2011

Linear vs. Nonlinear Statistical Models

- Linear Models
  1. Simple linear regression
  2. Multiple linear regression
  3. Analog forecasting
- Nonlinear Models
  4. Logistic Regression
  5. Poisson Regression

1. Simple Linear Regression

Assuming that the relationship between the predictor x and the predictand y can be approximated by a linear equation:

$$\hat{y} = a + bx$$

a and b are often determined using the linear least-squares method.
After we have determined a and b, we can estimate y based on x using the equation above.

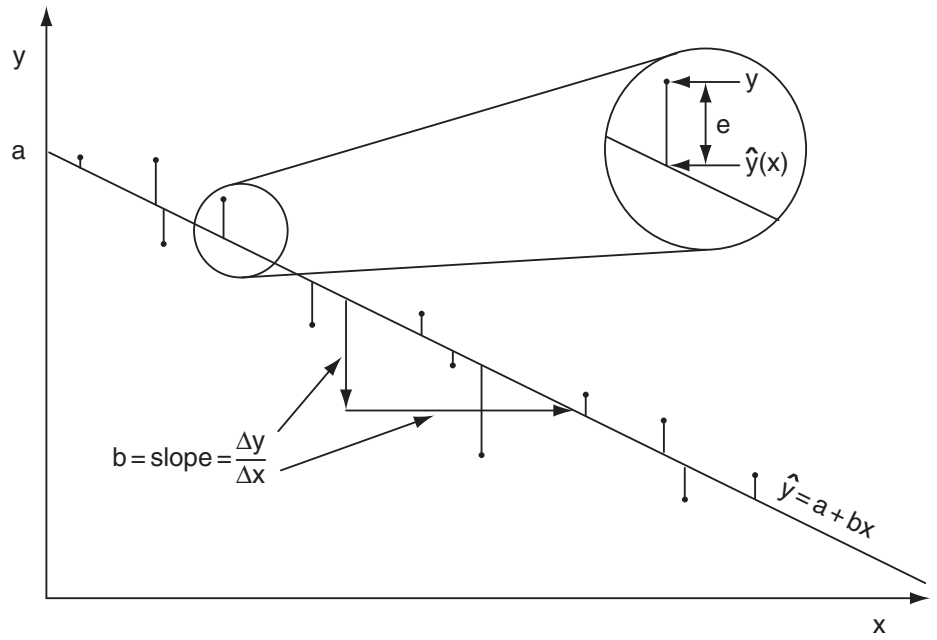

FIGURE 6.1 Schematic illustration of simple linear regression. The regression line, $\hat{y} = a + bx$, is chosen as the one minimizing some measure of the vertical differences (the residuals) between the points and the line. In least-squares regression that measure is the sum of the squared vertical distances. Inset shows the residual, e, as the difference between the data point and the regression line.

Wilks 2011

2. Multiple Linear Regression (MLR)

- There is still one single predictand, y, but there are multiple predictors (the number of predictors is denoted by K). The relationship between the predictand and the predictors can be approximated by a linear equation:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_K x_K.$$

- To develop the MLR model, we need to determine K+1 parameters, the intercept b_0 and the slopes for the individual predictors (b_1, .. b_k)
- Some questions you may have:
  - How do we select predictors?
  - How many predictors shall we select?
  - How do we assess the prediction skill?

Wilks 2011

3. Analog Forecasting

- The analog method identifies the best matches to the current climate conditions and then provides a forecast based on how the climate conditions evolved in those analog events.
  - Different studies may use different criteria to select “best matches”
  - “Best matches” are identified based on one or more meteorological variables
  - The selected “matches” can be used to construct a deterministic or probabilistic forecast.
- See a non-operational web site https://www.cpc.ncep.noaa.gov/products/people/wd51hd/

Nonlinear Models

- Nonlinear regression can be appropriate when the nature of the physical problem at hand dictates a nonlinear relationship between the predictand and predictors.
- In nonlinear regression, the fitting procedure is usually iterative and based on maximum likelihood methods.

Wilks 2011

Logistic Regression

- The predictand is a binary variable, taking on the values 0 and 1.
- The values of the predictand (0 and 1) can be viewed as probabilities of the dichotomous event not occurring or occurring, respectively, or above or below a cutoff threshold.
  - Although one can force a linear regression onto such a problem, the resulting forecasts may not lie on the unit interval.
- Logistic regressions are fit to binary predictands, according to the nonlinear equation

$$p_i = \frac{\exp(b_0 + b_1 x_1 + \cdots + b_K x_K)}{1 + \exp(b_0 + b_1 x_1 + \cdots + b_K x_K)} = \frac{1}{1 + \exp(-b_0 - b_1 x_1 - \cdots - b_K x_K)},$$

or

$$\ln\left(\frac{p_i}{1 - p_i}\right) = b_0 + b_1 x_1 + \cdots + b_K x_K.$$

Pi → 0 as $b_0 + b_1 x_1 + \ldots \longrightarrow -\infty$
Pi → 1 as $b_0 + b_1 x_1 + \ldots \longrightarrow +\infty$

Wilks 2011

Poisson Regression

- Poisson regression applies to the cases where the predictand is a nonnegative integer, such as the number of tornadoes.
- The Poisson mean of the predictand is specified as a nonlinear function of the predictors:

$$\mu_i = \exp[b_0 + b_1 x_1 + \cdots + b_K x_K],$$

or

$$\ln(\mu_i) = b_0 + b_1 x_1 + \cdots + b_K x_K.$$

- The equations ensure that the predicted value is nonnegative.

Wilks 2011

References

- Wilks, 2011: “Statistical Methods in the Atmospheric Sciences”, Sections 6.2 and 6.3