1 Nonparametric Regression

1.1 Introduction

- The *linear regression model*

$$Y_i \;=\; \beta_0 + \beta_1 X_i + \epsilon_i$$

  constrains the nonrandom part ($\beta_0 + \beta_1 X$) to take a particular **functional form** (linear).

  A regression model for which the nonrandom part is constrained to take a particular **functional form** is called a ***parametric regression*** model.

  All of the models we've seen so far are **parametric** models.

- ***Misspecification*** of the functional form in a parametric model (e.g. using linear when quadratic would've been more appropriate) can **bias** data analysis results and inflate prediction errors.

- ***Nonparametric regression*** refers to fitting a model of the form

$$Y_i \;=\; f(X_i) \;+\; \epsilon_i$$

  for *some* function $\boldsymbol{f(x)}$ whose **functional form** is left **unspecified** and allowed to be determined entirely by the data.

  We assume the errors have mean zero, i.e.

$$\mathrm{E}(\epsilon_i) \;=\; 0\,,$$

  so that $f(x)$ is the ***mean response***, i.e. for any given value of X,

$$\mathrm{E}(Y) \;=\; f(X).$$

  We might (or might not) assume a particular distribution (e.g. normal) for the errors.

- Because we don't specify a particular functional form for $f(x)$, there's **no risk** of **misspecification**.

- We'll suppose we have data:

Observation	Predictor Variable X	Response Variable Y
1	X_1	Y_1
2	X_2	Y_2
$\vdots$	$\vdots$	$\vdots$
n	X_n	Y_n

Our goal will be to obtain an ***estimate*** $\hat{f}(x)$ of the true (unknown) **mean response** function $f(x)$ from the data.

- All methods for obtaining the estimate $\hat{f}(x)$ involve a ***tuning parameter*** whose value controls how "wiggly" $\hat{f}(x)$ is allowed to be (akin to the *model complexity*).

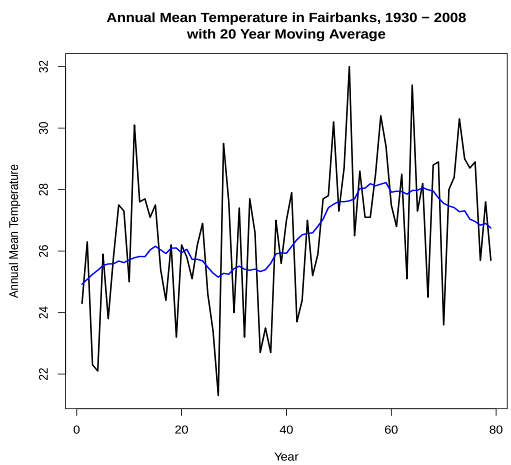
The **tuning parameter** value is often selected via **cross-validation**.

1.2 Moving Averages

- The simplest **nonparametric regression** method, for equally spaced X_i values such as a *time-series*, is the ***moving average***, which can ***simple*** (*unweighted*) or ***weighted***, ***one-sided*** or ***two-sided***.

1.2.1 Simple Moving Averages

- The blue curve below is a *two-sided moving average* for Fairbanks, Alaska annual mean temperatures for the years 1930 - 2008.



- The value of a ***two-sided moving average*** $\boldsymbol{\hat{f}(X_h)}$ at **time point** $\boldsymbol{X_h}$ is the average of the observations Y_i within a specified period of time $\boldsymbol{q}$ **before** *and* **after** time $\boldsymbol{X_h}$,

$$\hat{f}(X_h) \;=\; \frac{1}{2q+1}\sum_{i=h-q}^{h+q} Y_i\,. \tag{1}$$

- The moving average ***window size***, $\mathbf{2q+1}$, is the number of Y_i observations averaged in each $\hat{f}(X_h)$ value.

  As the ***window*** moves from left to right (in a scatterplot), the set of Y_i values contained within it changes and thus their average $\hat{f}(X_h)$ changes.

  In graphs, sequential values of $\hat{f}(X_h)$ are connected with lines to provide continuity.

- The **window size**, as determined by $\boldsymbol{q}$, is a ***tuning parameter*** that controls the degree of "wiggle" in the moving average.

  A **smaller window** results in a **more "wiggly"** moving average that **fits** the data **better** but **risks overfitting**. See Fig. 1.

  A **larger window** results in a **"smoother"** moving average but with potentially greater **"bias"** (analogous to model *misspecification*). See Fig. 1.

  The term ***bias-variance trade-off*** is sometimes used to describe the effect of altering $\boldsymbol{q}$ – the **small-$\boldsymbol{q}$** moving average has **high variance** ("wiggliness") but **low "bias"**, whereas the **large-$\boldsymbol{q}$** one has **low variance** but **high "bias"**.

  The optimal value of $\boldsymbol{q}$ can be selected via **cross-validation**.

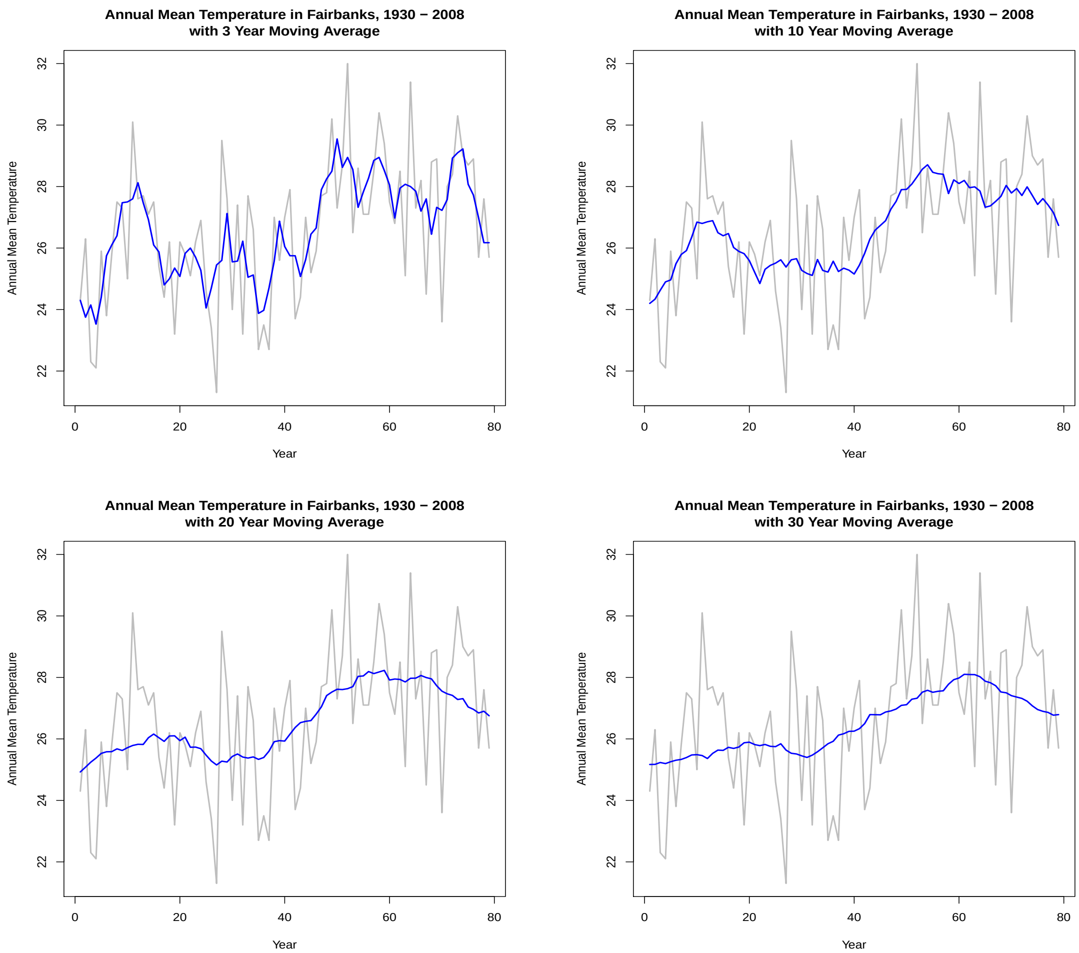


Figure 1: Moving averages for annual mean temperatures in Fairbanks, Alaska over the years 1930 - 2008, with window half-widths $q = 3$ (upper left), $q = 10$ (upper right), $q = 20$ (lower left), and $q = 30$ (lower right).

- The definition (1) of the two-sided moving average $\hat{f}(X_h)$ isn't valid for time points X_h within q time points of either end of the time series.

  One simple remedy is to extend the time series q periods of time at both ends by repeating the first observation Y_1 q times at the start and the last observation Y_n q times at the end.

- A ***one-sided moving average*** is similar to a two-sided moving average, except that only the q observations that **precede** Y_h in time are average, i.e.

$$\hat{f}(X_h) \;=\; \frac{1}{q+1} \sum_{i=h-q}^{h} Y_i \,.$$

1.2.2 Weighted Moving Averages

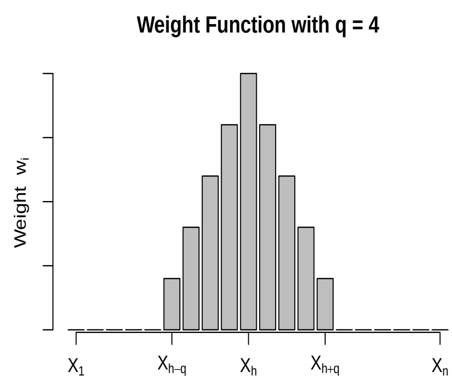
- The value of a (two-sided) ***weighted moving average*** $\boldsymbol{\hat{f}(X_h)}$ at **time point** $\boldsymbol{X_h}$ is

$$\hat{f}(X_h) \;=\; \frac{1}{2q+1} \sum_{i=h-q}^{h+q} w_i Y_i \,,$$

  where the ***weights*** $\boldsymbol{w_i}$ sum to **1**, and decrease with the distance of X_i away from X_h.

  This means that observations **farther** from $\boldsymbol{X_h}$ have **smaller weights**.

- The **weight** $\boldsymbol{w_i}$ associated with each observation Y_i is determined via a ***weight function***, for example the one shown below.



  Note that only the **closest** $\boldsymbol{q}$ observations to $\boldsymbol{X_h}$ have nonzero weights.