Climate Time Series Analysis:

Recent Climate Changes

Manfred Mudelsee

Climate Risk Analysis, Germany
Alfred Wegener Institute, Germany

Grenoble, 30 January 2012

Prologue

Climate: knowledge and experience,
space–time scales

Time series analysis: statistical concept

Computer: implementation of concept

Mathematical formulas: no sweat!

Chapters

1. Introduction

1.1

1.2 Noise and statistical distribution

1.3 Persistence

1.4

1.5 Aim and structure of this course

1. Introduction

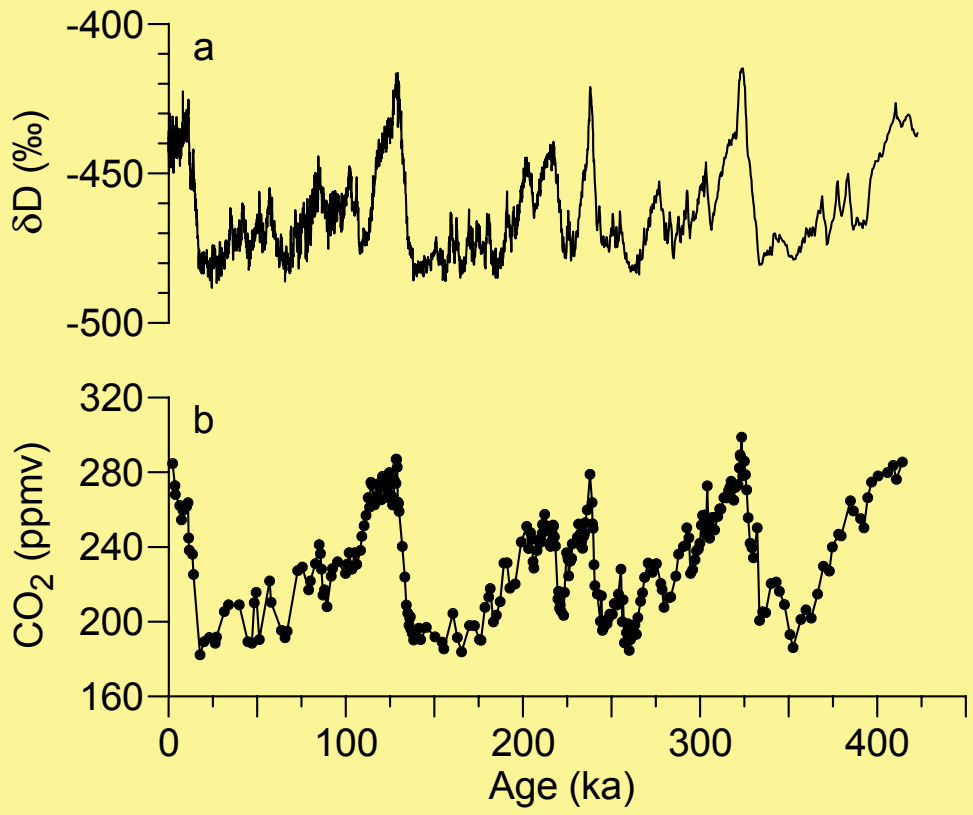


Figure 1.3. Ice core data: deuterium and CO_2 records from the Vostok station

$$\{t(i), x(i)\}_{i=1}^{n}$$

Time series sample

t, time value

x, climate-variable value

n, data size (sample size)

i, index

Climate time series

1. Introduction

$$X(T) = X_{\text{trend}}(T) + X_{\text{out}}(T) + S(T) \cdot X_{\text{noise}}(T), \qquad (1.1)$$

Climate equation, continuous time

T, time

X, climate variable observed

X_{trend}, trend component

X_{out}, outlier component

X_{noise}, noise component

S, variability

1. Introduction

$$X(i) = X_{\text{trend}}(i) + X_{\text{out}}(i) + S(i) \cdot X_{\text{noise}}(i), \qquad (1.2)$$

Climate equation, discrete time

$T(i)$, time points

$X(i) \equiv X(T(i))$, etc.

1. Introduction

Stochastic process $X(i)$ "Process level"

Time series $\{t(i), x(i)\}_{i=1}^{n}$ "Sample level"

Time series analysis uses sample to learn about process:
trend parameters, probability of extremes, cycles, etc.

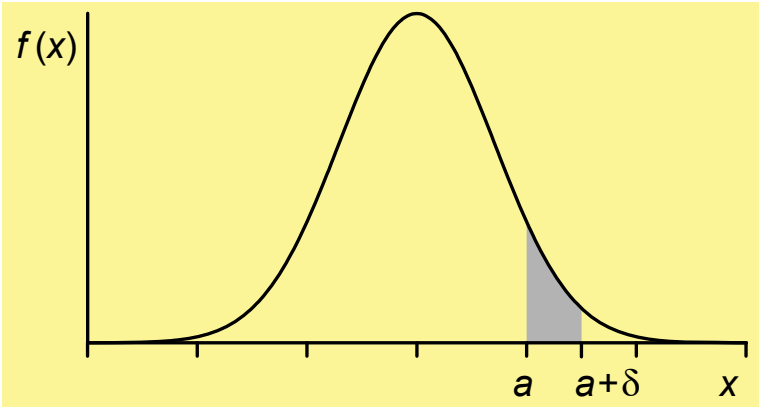
Univariate: $T(i)$, $X(i)$
Bivariate: $T(i)$, $X(i)$, $Y(i)$

1.2 Noise and statistical distribution

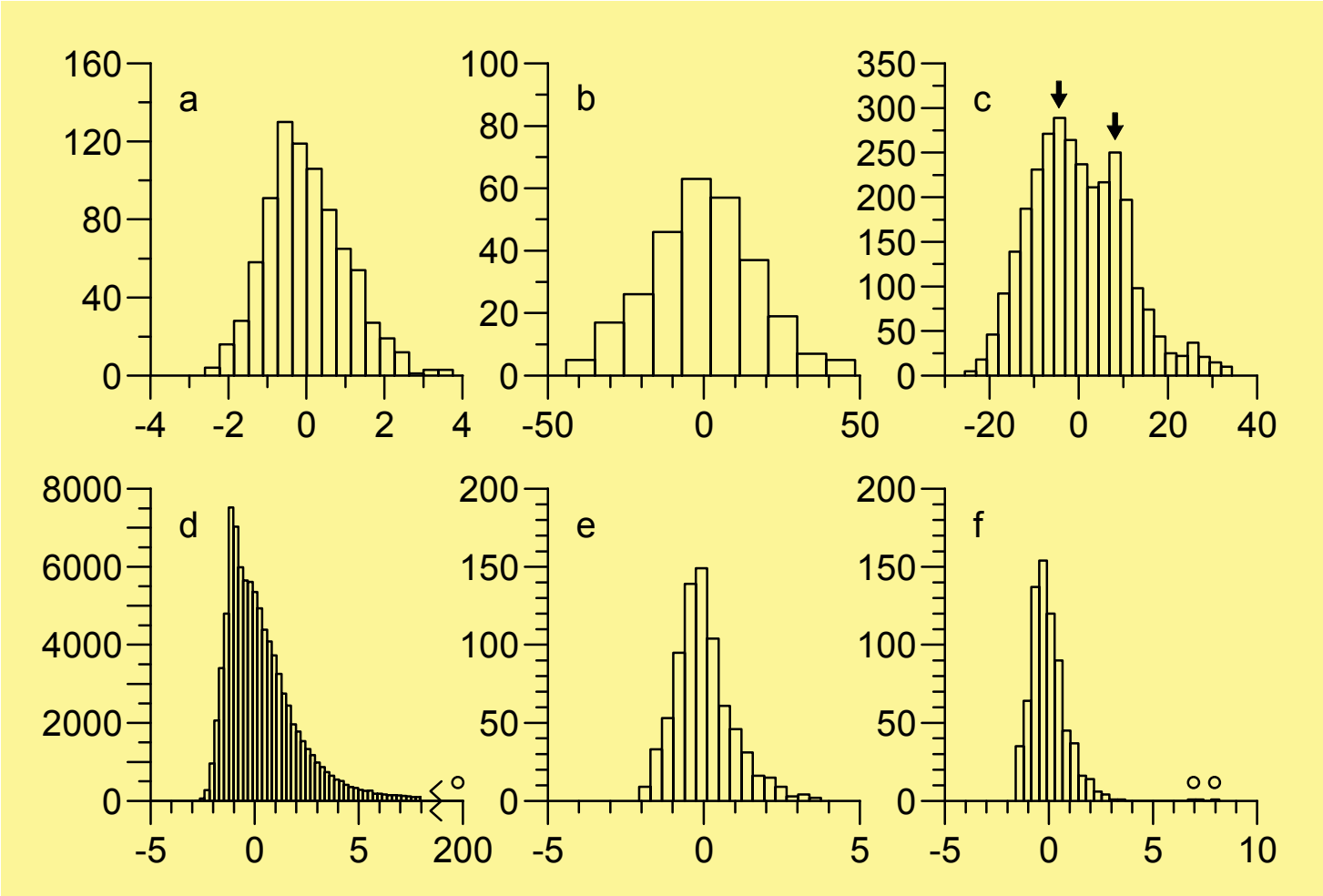
$$\text{prob}\left(a \leq X_{\text{noise}}(T) \leq a + \delta\right) \qquad = \int_{a}^{a+\delta} f(x)dx, \qquad (1.3)$$

Probability density function (PDF), *f*(*x*)

Gaussian or normal PDF



1.2 Noise and statistical distribution



Histogram: inference of PDF

Residuals, "detrended version" of the $x(i)$

Book, Fig. 1.11 (**b,** Vostok CO_2; **c,** Vostok δD)

1.3 Persistence

$$E\left[X_{\text{noise}}(T_1) \cdot X_{\text{noise}}(T_2)\right] \text{ for } \quad T_1 \neq T_2$$

Autocovariance

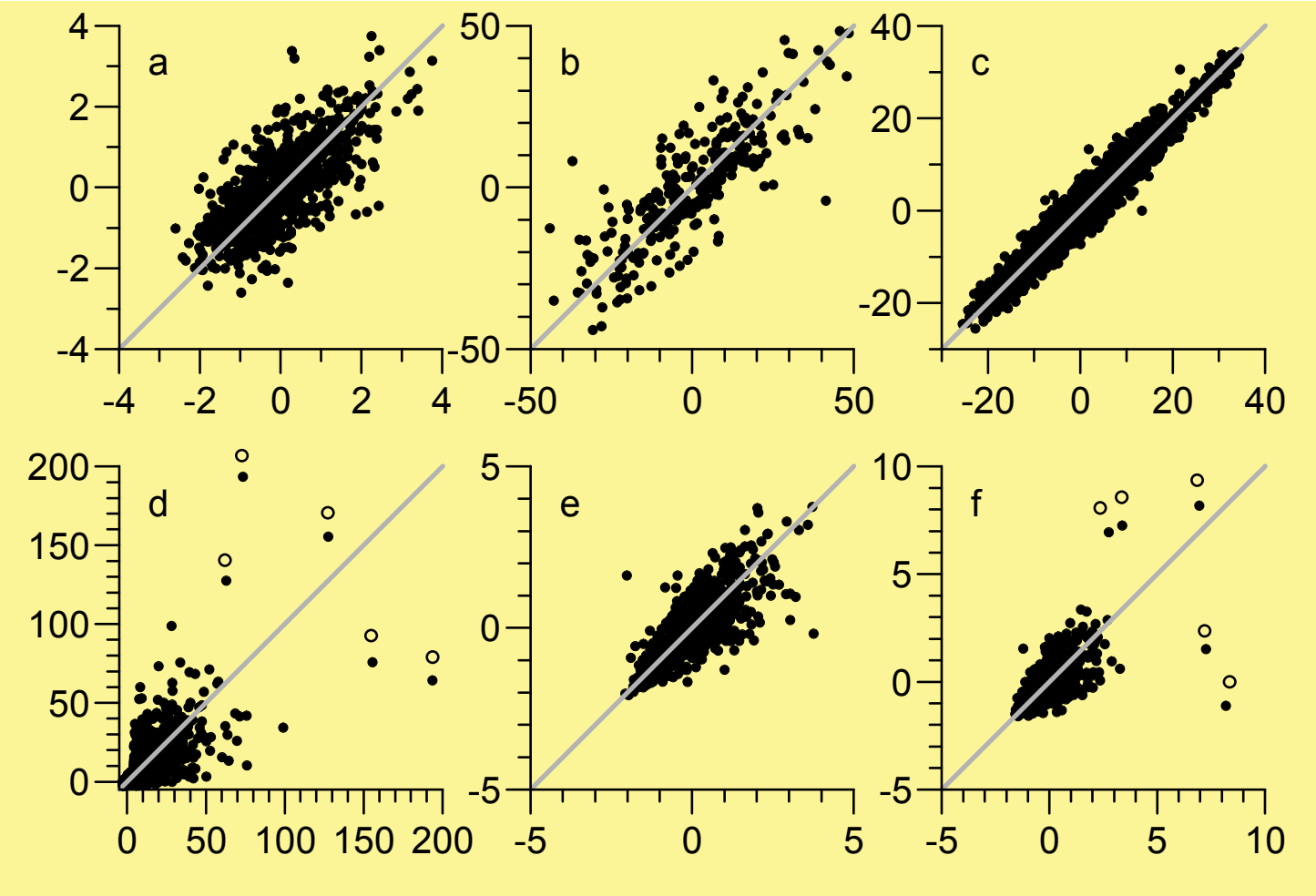
E, expectation operator (averaging)

$E[X_{\text{noise}}(T)] = 0$ (zero mean)

Positive autocovariance: persistence or “memory”

Book, p. 11–13

1.3 Persistence



Lag-1 scatterplot $x(i-1)$ vs $x(i)$: inference about persistence

Residuals, "detrended version" of the $x(i)$

Book, Fig. 1.12 (**b**, Vostok CO_2; **c**, Vostok δD)

1.3 Persistence

$$X(i) = X_{\text{trend}}(i) + X_{\text{out}}(i) + S(i) \cdot X_{\text{noise}}(i), \tag{1.2}$$

$\{t(i), x(i)\}_{i=1}^{n}$

Residuals:
use sample to determine trend, outliers and variability,
then
subtract trend, remove outliers and divide by variability
(details: Chapters 3 and 4)

1.5 Aim and structure of this course

$$X(i) = X_{\text{trend}}(i) + X_{\text{out}}(i) + S(i) \cdot X_{\text{noise}}(i), \qquad (1.2)$$

Climate equation, stochastic process $X(i)$

Time series $\{t(i), x(i)\}_{i=1}^{n}$

Time series analysis uses sample to learn about process: trend parameters, probability of extremes, cycles, etc.

Univariate: $T(i)$, $X(i)$

Bivariate: $T(i)$, $X(i)$, $Y(i)$

2. Persistence Models

2.1 First-order autoregressive model

2.2

2.3

2.4 Other models

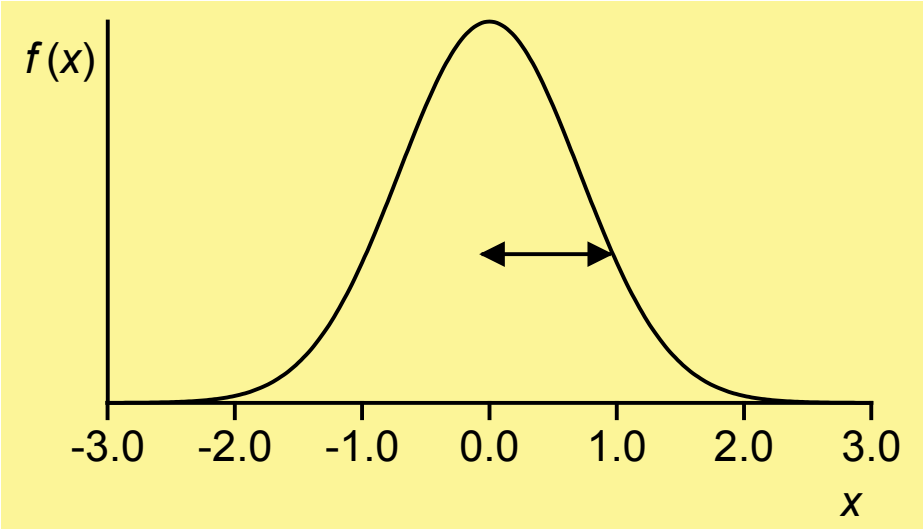
2.5

2.1 First-order autoregressive model

$$
\begin{aligned}
X_{\text{noise}}(1) &= \mathcal{E}_{\mathrm{N}(0,\,1)}(1), \\
X_{\text{noise}}(i) &= a \cdot X_{\text{noise}}(i-1) + \mathcal{E}_{\mathrm{N}(0,\,1-a^2)}(i), \qquad i = 2, \ldots, n.
\end{aligned}
\tag{2.1}
$$

AR(1) process, even spacing

$-1 < a < 1$



Innovation term $\mathcal{E}_{\mathrm{N}(\mu,\,\sigma^2)}(\cdot)$

Mean = μ = 0

Variance = σ^2 = 1

Gaussian shape

$f(x) = (2\pi)^{-1/2}\,\sigma^{-1}\exp[-(x-\mu)^2/(2\sigma^2)]$

$$f(x) = (2\pi)^{-1/2}\exp\left(-x^2/2\right)$$

2.1 First-order autoregressive model

$$
\begin{aligned}
X_{\text{noise}}(1) &= \mathcal{E}_{\mathrm{N}(0,\,1)}(1), \\
X_{\text{noise}}(i) &= a \cdot X_{\text{noise}}(i-1) + \mathcal{E}_{\mathrm{N}(0,\,1-a^2)}(i), \qquad i = 2, \ldots, n.
\end{aligned}
\tag{2.1}
$$

AR(1) process, even spacing

Mean: $E[X_{\text{noise}}(i)] = 0$ for all i

Variance: $VAR[X_{\text{noise}}(i)] = 1$ for all i

Strict stationarity

2.1 First-order autoregressive model

$$
\begin{aligned}
X_{\text{noise}}(1) &= \mathcal{E}_{\mathrm{N}(0,\,1)}(1), \\
X_{\text{noise}}(i) &= a \cdot X_{\text{noise}}(i-1) + \mathcal{E}_{\mathrm{N}(0,\,1-a^2)}(i), \qquad i = 2, \ldots, n.
\end{aligned}
\tag{2.1}
$$

AR(1) process, even spacing

$$
\begin{aligned}
\rho(h) &= \frac{E\Big[\big\{X_{\text{noise}}(i+h) - E\big[X_{\text{noise}}(i+h)\big]\big\} \cdot \big\{X_{\text{noise}}(i) - E\big[X_{\text{noise}}(i)\big]\big\}\Big]}{\Big\{VAR\big[X_{\text{noise}}(i+h)\big] \cdot VAR\big[X_{\text{noise}}(i)\big]\Big\}^{1/2}} \\
&= E\Big[X_{\text{noise}}(i+h) \cdot X_{\text{noise}}(i)\Big],
\end{aligned}
$$

Autocorrelation (lag *h*) $\rho(h) = a^{|h|}, \qquad h = 0, \pm 1, \pm 2, \ldots.$

2.1 First-order autoregressive model

$$
\begin{aligned}
X_{\text{noise}}(1) &= \mathcal{E}_{\mathrm{N}(0,\,1)}(1), \\
X_{\text{noise}}(i) &= a \cdot X_{\text{noise}}(i-1) + \mathcal{E}_{\mathrm{N}(0,\,1-a^2)}(i), \qquad i = 2, \ldots, n.
\end{aligned}
\tag{2.1}
$$

AR(1) process, even spacing

a > 0: persistence or "memory"

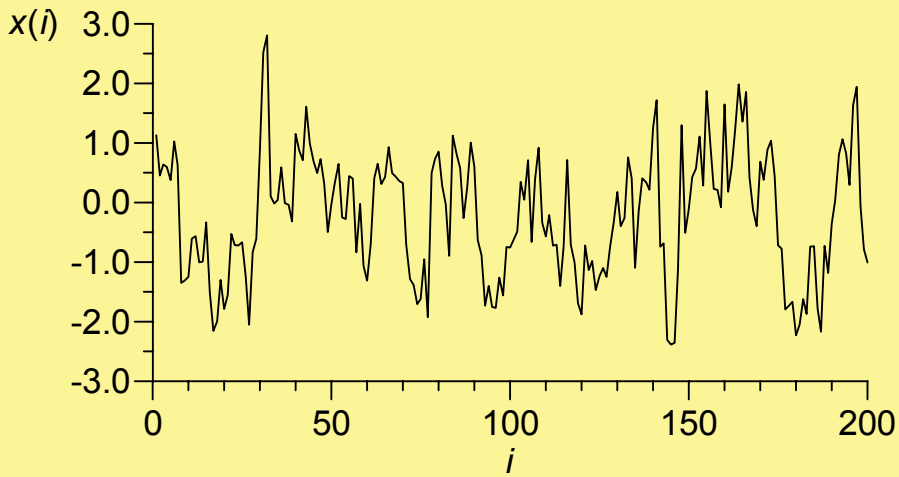


Figure 2.1. Realization of an AR(1) process (Eq. 2.1); $n = 200$ and $a = 0.7$.

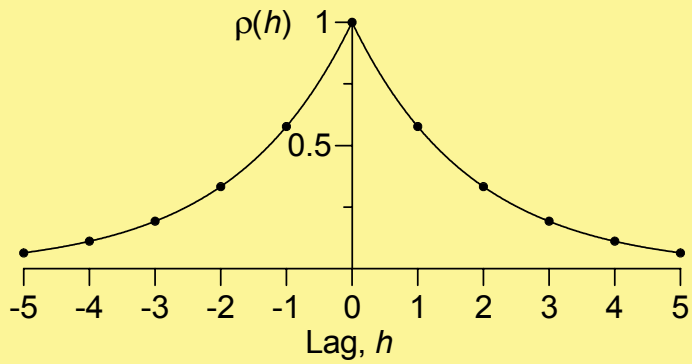


Figure 2.2. Autocorrelation function of the AR(1) process, $a > 0$. In the case of even

2.1 First-order autoregressive model

$$\begin{aligned} X_{\text{noise}}(1) &= \mathcal{E}_{\mathrm{N}(0,\,1)}(1), \\ X_{\text{noise}}(i) &= a \cdot X_{\text{noise}}(i-1) + \mathcal{E}_{\mathrm{N}(0,\,1-a^2)}(i), \qquad i = 2, \ldots, n. \end{aligned} \tag{2.1}$$

AR(1) process, even spacing

$$\widehat{a} = \sum_{i=2}^{n} x_{\text{noise}}(i) \cdot x_{\text{noise}}(i-1) \Bigg/ \sum_{i=2}^{n} x_{\text{noise}}(i)^2. \tag{2.4}$$

Estimation

Residuals $x_{\text{noise}}(i)$, “detrended version” of the $x(i)$

2.1 First-order autoregressive model

$$\begin{aligned} X_{\text{noise}}(1) &= \mathcal{E}_{\mathrm{N}(0,\,1)}(1), \\ X_{\text{noise}}(i) &= a \cdot X_{\text{noise}}(i-1) + \mathcal{E}_{\mathrm{N}(0,\,1-a^2)}(i), \qquad i = 2, \ldots, n. \end{aligned} \tag{2.1}$$

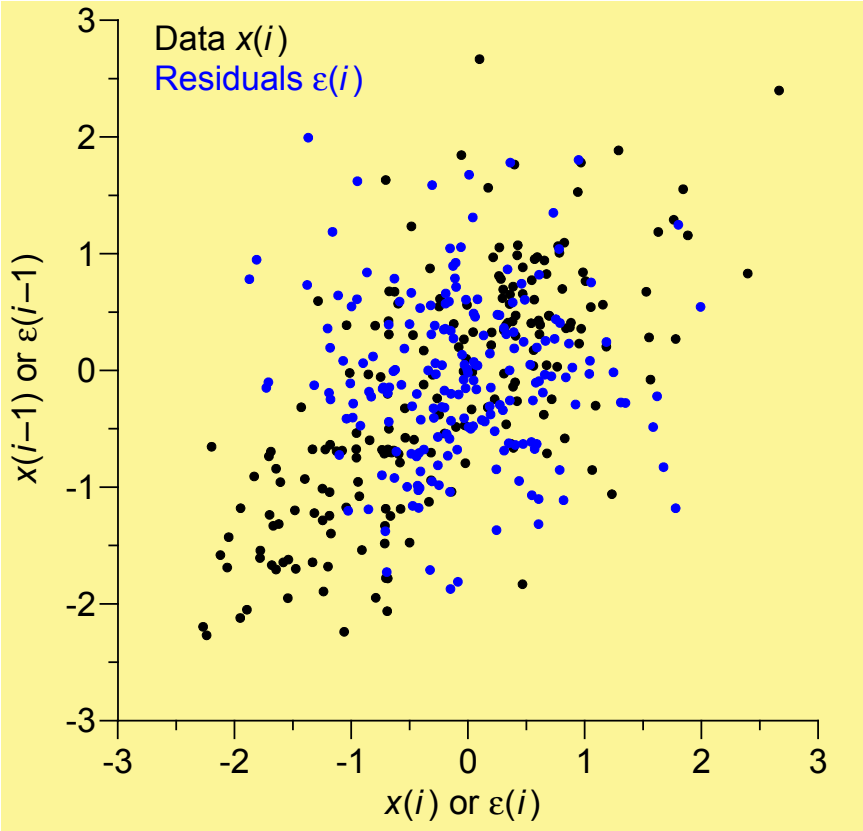
AR(1) process, even spacing

$$\epsilon(i) = x_{\text{noise}}(i) - \widehat{a} \cdot x_{\text{noise}}(i-1), \qquad i = 2, \ldots, n. \tag{2.5}$$

White-noise residuals (estimation residuals)

Graphical test of model suitability

2.1 First-order autoregressive model

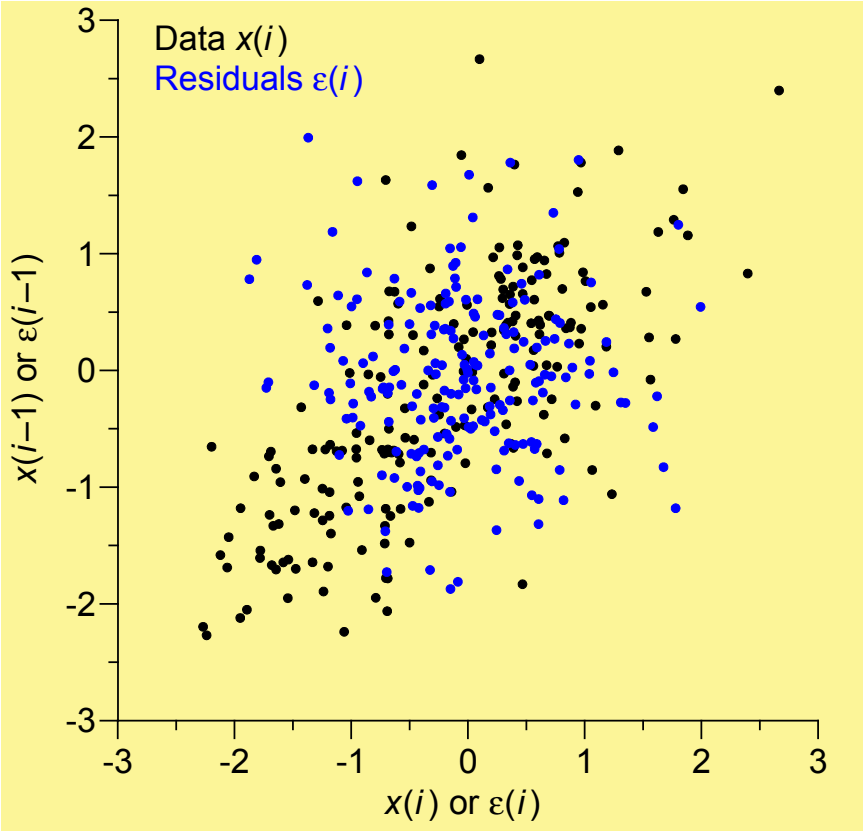


Artificial time series (n = 200, a = 0.7)

$\hat{a}$ = 0.717

Data from Book, Fig. 2.1

2.1 First-order autoregressive model



Artificial time series (n = 200, a = 0.7)

$\hat{a}$ = 0.717

Theoretically:

$$E(\widehat{a}) \simeq (1 - 2/n)\, a.$$ = 0.693

Negative bias, underestimation

$$VAR(\widehat{a}) \simeq \left(1 - a^2\right)/n$$ = 0.003

Estimation variance

$VAR[\hat{a}]^{1/2} \approx 0.055$

Data from Book, Fig. 2.1; Book, Eqs. (2.39), (2.42)

2.1 First-order autoregressive model

$$
\begin{aligned}
X_{\text{noise}}(1) &= \mathcal{E}_{\mathrm{N}(0,\,1)}(1), \\
X_{\text{noise}}(i) &= a \cdot X_{\text{noise}}(i-1) + \mathcal{E}_{\mathrm{N}(0,\,1-a^2)}(i), \qquad i = 2, \ldots, n.
\end{aligned}
\tag{2.1}
$$

AR(1) process, even spacing

a > 0: persistence,
fewer independent data points,
effective data size *n′* < *n*,
larger estimation variance

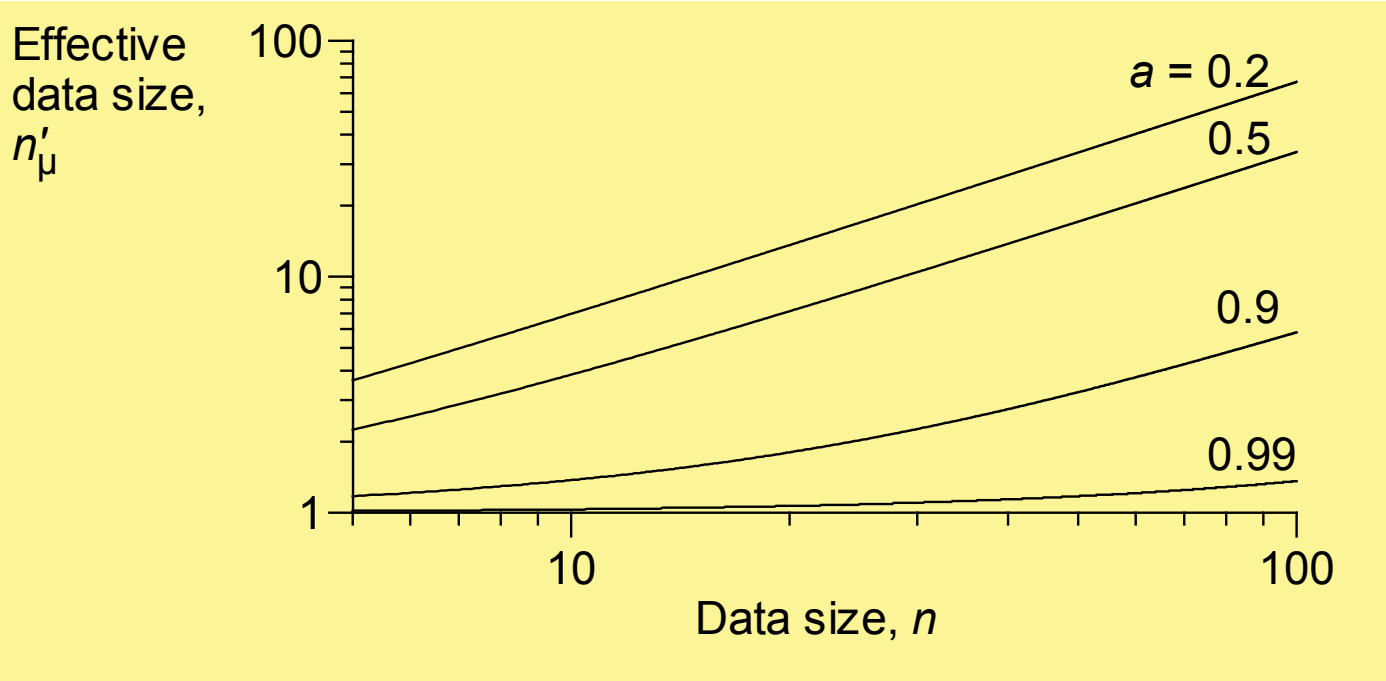
2.1 First-order autoregressive model

$\bar{X} = \sum_{i=1}^{n} X(i)/n,$

Mean estimation

$$n'_\mu = n \left[1 + 2 \sum_{i=1}^{n-1} (1 - i/n)\, \rho(i) \right]^{-1}$$

Effective data size

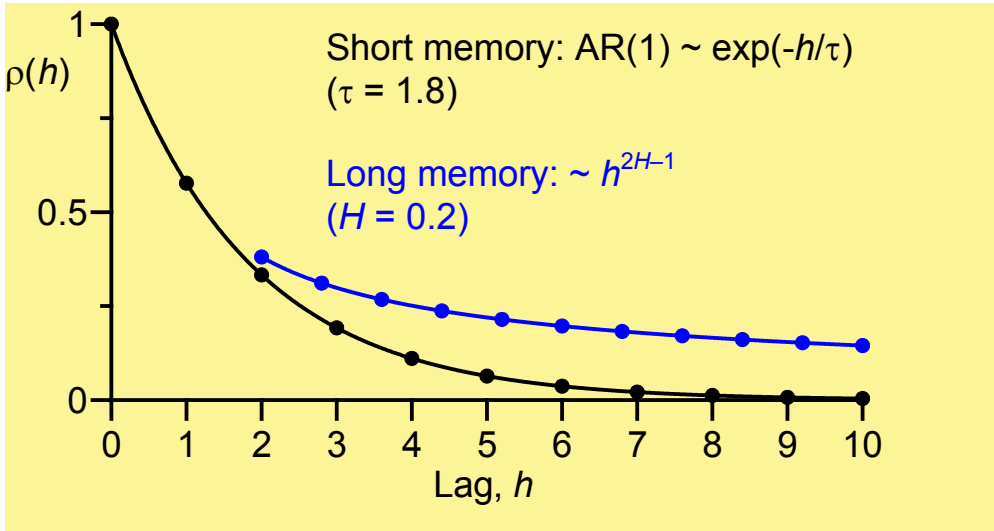


Book, Eqs. (2.6), (2.7), Fig. 2.10a (modified)

2.4 Other models

Long memory processes

- Fractional Gaussian noise
- ARFIMA(p, δ, q)



Long memory

- Hyperbolically decreasing acf
- (for $h \to \infty$)

Mandelbrot (1983), Beran (1994, 1997, 1998), Doukhan et al. (2003), Robinson (2003); Book, Sections 2.4.1, 2.5.2, 2.5.3

2.4 Other models

$$
\begin{aligned}
X_{\text{noise}}(1) &= \mathcal{E}_{\mathrm{N}(0,\,1)}(1), \\
X_{\text{noise}}(i) &= a \cdot X_{\text{noise}}(i-1) + \mathcal{E}_{\mathrm{N}(0,\,1-a^2)}(i), \qquad i = 2, \ldots, n.
\end{aligned}
\tag{2.1}
$$

Gaussian AR(1) process, even spacing

Gaussian innovation term $\mathcal{E}_{\mathrm{N}(\mu,\,\sigma^2)}(\cdot)$

Non-Gaussian processes

Non-Gaussian innovation terms

3. Bootstrap Confidence Intervals

3.1 Error bars and confidence intervals

3.2 Bootstrap principle

3.3 Bootstrap resampling

3.4 Bootstrap confidence intervals

3.5

3.6 Bootstrap hypothesis tests

3.1 Error bars and confidence intervals

$$X(i) = X_{\text{trend}}(i) + X_{\text{out}}(i) + S(i) \cdot X_{\text{noise}}(i), \qquad (1.2)$$

Climate equation, stochastic process *X*(*i*)

Time series $\{t(i), x(i)\}_{i=1}^{n}$

Time series analysis uses sample to learn about process: trend parameters, probability of extremes, cycles, etc.

3.1 Error bars and confidence intervals

Estimator

A recipe how to calculate parameter(s) of interest

Estimation

We apply recipe to data and get result.

Estimate

Numerical value of result

Hat notation

Denotes estimator/estimate.

Estimation of AR(1) parameter *a*

$$\begin{aligned} X_{\text{noise}}(1) &= \mathcal{E}_{\mathrm{N}(0,\,1)}(1), \\ X_{\text{noise}}(i) &= a \cdot X_{\text{noise}}(i-1) + \mathcal{E}_{\mathrm{N}(0,\,1-a^2)}(i), \qquad i = 2, \ldots, n. \end{aligned} \tag{2.1}$$

$$\widehat{a} = \sum_{i=2}^{n} x_{\text{noise}}(i) \cdot x_{\text{noise}}(i-1) \Bigg/ \sum_{i=2}^{n} x_{\text{noise}}(i)^2. \tag{2.4}$$

3.1 Error bars and confidence intervals

True parameter	Sample	Estimator, estimate
	Time series	
θ	$\{t(i), x(i)\}_{i=1}^{n}$	$\hat{\theta}$
("Theta")		("Theta hat")
	$n < \infty$,	
	Observation errors (measurement, proxy, calibration, climate model, etc.)	$\hat{\theta} \neq \theta$

3.1 Error bars and confidence intervals

Estimate does not equal true (but unknown) value,
it deviates.

However,
we can say how large this deviation typically is:
error bars, confidence intervals, etc.

We need statistical language,
we make a statistical inference.

3.1 Error bars and confidence intervals—Philosophical excursion

Assumptions implicitly made sofar include:

(1) Truth exists (logic, realism)

(2) Time arrow

(3) Ergodicity

(4) Foundation of probability (axiomatic approach)

Recommendation:

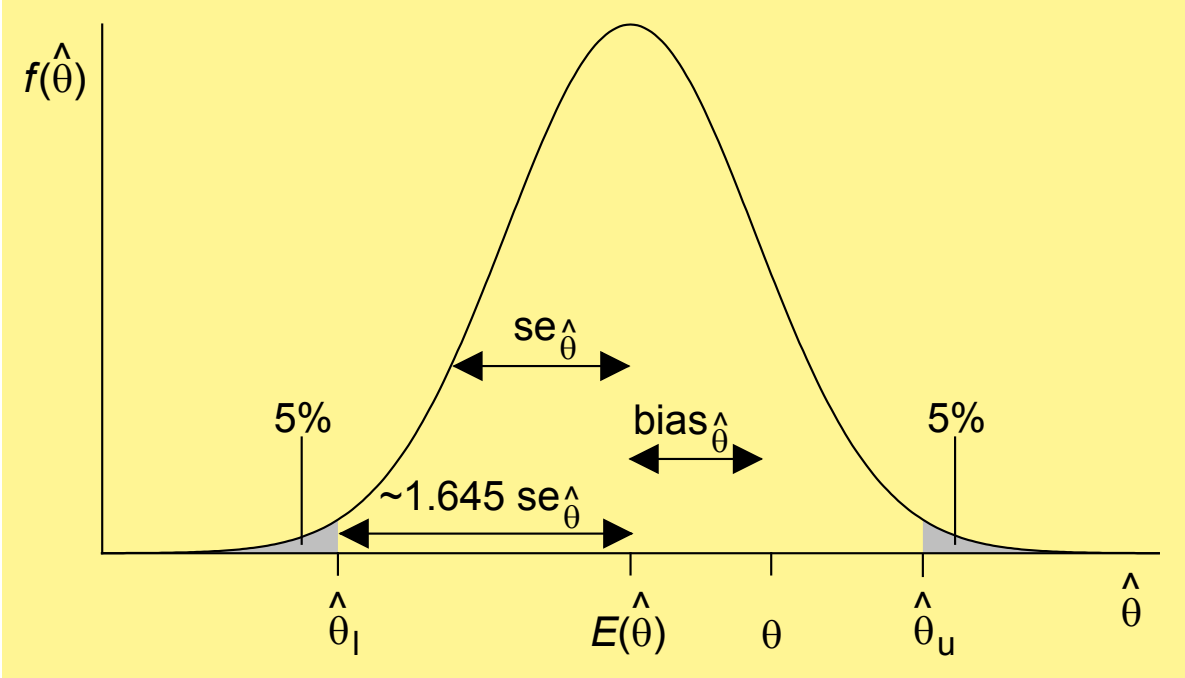
Be cautious with “blog science” or “post-normal” stuff.

Kant (1781), Popper (1935), Einstein (1949), Polanyi (1958), Kuhn (1970)

Feynman (1974) Cargo cult science. *Engineering and Science*, June 1974, p. 10–13

3.1 Error bars and confidence intervals

PDF of an estimator



$$\mathrm{se}_{\widehat{\theta}} = \left[VAR\left(\widehat{\theta}\right)\right]^{1/2}.$$

Standard error

$$\mathrm{bias}_{\widehat{\theta}} = E\left(\widehat{\theta}\right) - \theta.$$

Bias

$$\mathrm{RMSE}_{\widehat{\theta}} = \left\{E\left[\left(\widehat{\theta} - \theta\right)^2\right]\right\}^{1/2} = \left(\mathrm{se}_{\widehat{\theta}}{}^2 + \mathrm{bias}_{\widehat{\theta}}{}^2\right)^{1/2}.$$

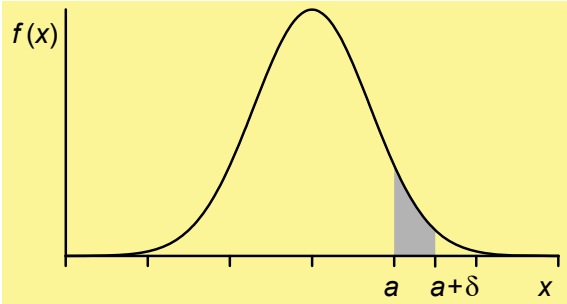
Root mean squared error

$$\mathrm{CI}_{\widehat{\theta},1-2\alpha} = \left[\widehat{\theta}_{\mathrm{l}}; \widehat{\theta}_{\mathrm{u}}\right],$$

Confidence interval (e.g., equi-tailed)

3.1 Error bars and confidence intervals

Mean estimation of Gaussian white noise



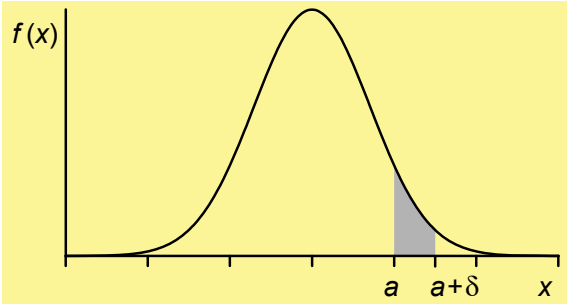
$$X(i) = \mathcal{E}_{\mathrm{N}(\mu,\,\sigma^2)}(i), \qquad i = 1, \ldots, n,$$

$$\widehat{\mu} = \bar{X} = \sum_{i=1}^{n} X(i)/n.$$

What is the expectation?

3.1 Error bars and confidence intervals

Mean estimation of Gaussian white noise



$$X(i) = \mathcal{E}_{\mathrm{N}(\mu,\,\sigma^2)}(i), \qquad i = 1, \ldots, n,$$

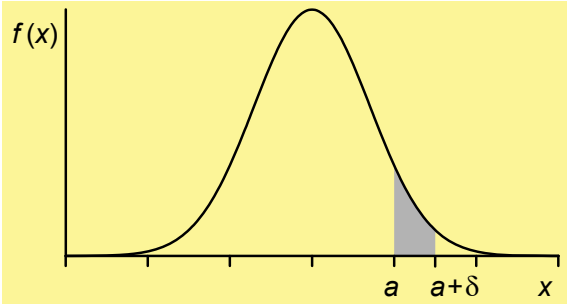
$$\widehat{\mu} = \bar{X} = \sum_{i=1}^{n} X(i)/n.$$

What is the expectation?

$(\mu + \mu + \mu + \ldots + \mu)/n = \mu \times n/n = \mu$ (zero bias)

3.1 Error bars and confidence intervals

Mean estimation of Gaussian white noise



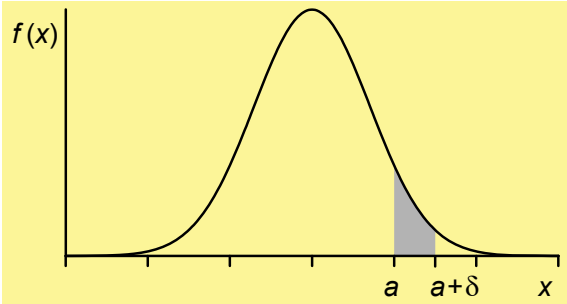
$$X(i) = \mathcal{E}_{\mathrm{N}(\mu,\,\sigma^2)}(i), \qquad i = 1, \ldots, n,$$

$$\widehat{\mu} = \bar{X} = \sum_{i=1}^{n} X(i)/n.$$

What is the variance?

3.1 Error bars and confidence intervals

Mean estimation of Gaussian white noise



$$X(i) = \mathcal{E}_{\mathrm{N}(\mu,\,\sigma^2)}(i), \qquad i = 1, \ldots, n,$$

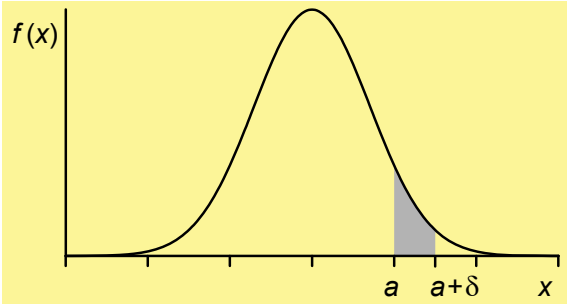
$$\widehat{\mu} = \bar{X} = \sum_{i=1}^{n} X(i)/n.$$

What is the variance?

$$(\sigma^2 + \sigma^2 + \sigma^2 + \ldots + \sigma^2)/n^2 = \sigma^2 \times n/n^2 = \sigma^2/n \qquad \text{(standard error } \sigma n^{-1/2}\text{)}$$

3.1 Error bars and confidence intervals

Mean estimation of Gaussian white noise



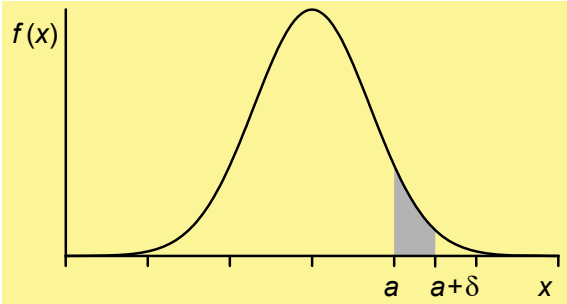
$$X(i) = \mathcal{E}_{\mathrm{N}(\mu,\,\sigma^2)}(i), \qquad i = 1, \ldots, n,$$

$$\widehat{\mu} = \bar{X} = \sum_{i=1}^{n} X(i)/n.$$

What is the distribution of the estimator?

3.1 Error bars and confidence intervals

Mean estimation of Gaussian white noise



$$X(i) = \mathcal{E}_{\mathrm{N}(\mu,\,\sigma^2)}(i), \qquad i = 1, \ldots, n,$$

$$\widehat{\mu} = \bar{X} = \sum_{i=1}^{n} X(i)/n.$$

What is the distribution of the estimator?

Student's *t*-distribution (we can calculate CIs)

Book, Eqs. (3.9), (3.10), (3.15), Section 3.1.1; Johnson et al. (1995)

3.1 Error bars and confidence intervals

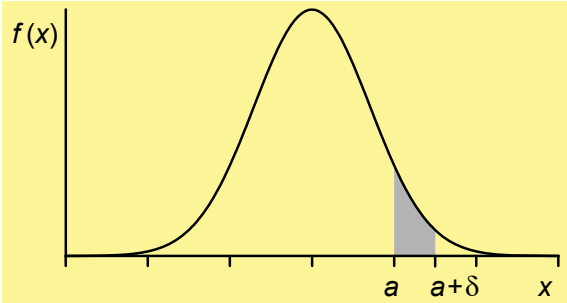
Table 3.1. Monte Carlo experiment, mean estimation of a Gaussian purely random process. $n_{\text{sim}} = 4{,}750{,}000$ random samples of $\{X(i)\}_{i=1}^{n}$ were generated after Eq. (3.9) with $\mu = 1.0$, $\sigma = 2.0$ and various n values. An exact confidence interval $\text{CI}_{\bar{x},1-2\alpha}$ was constructed for each simulation after Eq. (3.18) with $\alpha = 0.025$. Average CI length, empirical $\text{RMSE}_{\bar{X}}$ and empirical coverage were determined subsequently. The entries are rounded.

n	$\text{RMSE}_{\bar{x}}^{\text{a}}$	$Nominal^{\text{b}}$	$\langle\ CI\ length\ \rangle^{\text{c}}$	$Nominal^{\text{d}}$	$\gamma_{\bar{x}}^{\text{e}}$	$Nominal$
10	0.6327	0.6325	2.7832	2.7832	0.9499	0.9500
20	0.4474	0.4472	1.8476	1.8476	0.9498	0.9500
50	0.2828	0.2828	1.1310	1.1310	0.9501	0.9500
100	0.2000	0.2000	0.7916	0.7917	0.9499	0.9500
200	0.1415	0.1414	0.5570	0.5571	0.9499	0.9500
500	0.0894	0.0894	0.3513	0.3513	0.9500	0.9500
1000	0.0633	0.0632	0.2482	0.2482	0.9499	0.9500

[a] Empirical $\text{RMSE}_{\bar{X}}$, given by $\left[\sum_{i=1}^{n_{\text{sim}}} (\bar{x} - \mu)^2 / n_{\text{sim}}\right]^{1/2}$.
[b] $\sigma \cdot n^{-1/2}$.
[c] Average value over n_{sim} simulations.
[d] $2 \cdot t_{n-1}(1-\alpha) \cdot \sigma \cdot c \cdot n^{-1/2}$, where c is given by Eq. (3.24).
[e] Empirical coverage, given by the number of simulations where $\text{CI}_{\bar{x},1-2\alpha}$ contains μ, divided by n_{sim}. Standard error of $\gamma_{\bar{x}}$ is (Efron and Tibshirani 1993) nominally $[2\alpha(1-2\alpha)/n_{\text{sim}}]^{1/2} = 0.0001$.

3.1 Error bars and confidence intervals

Mean estimation of Gaussian white noise



$$X(i) = \mathcal{E}_{\mathrm{N}(\mu,\,\sigma^2)}(i), \qquad i = 1, \ldots, n,$$

$$\widehat{\mu} = \bar{X} = \sum_{i=1}^{n} X(i)/n.$$

Simple data generating process,

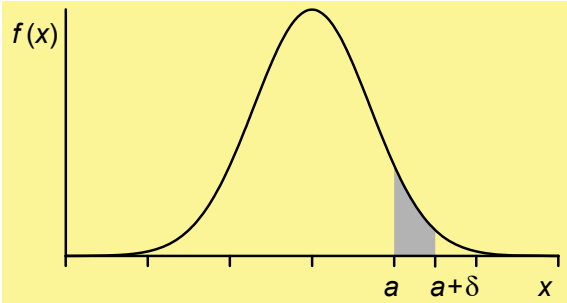
simple estimation problem:

- we can write down PDF of estimator (i.e., we have everything),
- CIs are exact.

3.1 Error bars and confidence intervals

Standard deviation estimation

of Gaussian white noise



$$X(i) = \mathcal{E}_{\mathrm{N}(\mu,\,\sigma^2)}(i), \qquad i = 1, \ldots, n,$$

$$\widehat{\sigma} = S_{n-1} = \left\{ \sum_{i=1}^{n} \left[X(i) - \bar{X} \right]^2 / (n-1) \right\}^{1/2}$$

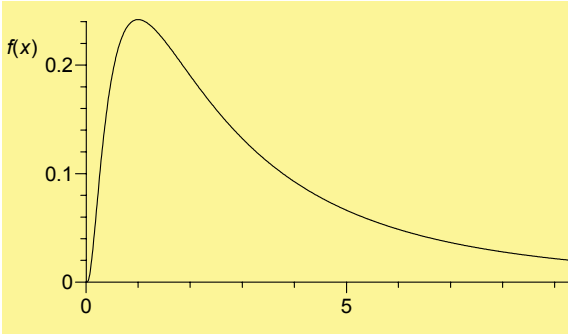
Simple data generating process,

simple estimation problem:

- we can write down PDF of estimator (i.e., we have everything),
- CIs are exact.

3.1 Error bars and confidence intervals

Mean estimation of lognormal white noise



$$X(i) = \exp\left[\mathcal{E}_{\mathrm{N}(\mu,\,\sigma^2)}(i)\right],\ \ i = 1, \ldots, n,$$

$$\widehat{\mu} = \bar{X} = \sum_{i=1}^{n} X(i)/n.$$

Distributional assumption violated:

data generating process lognormal, not Gaussian.

What happens to CI (from Student's *t*-distribution)?

3.1 Error bars and confidence intervals

Table 3.3. Monte Carlo experiment, mean and median estimation of a lognormal purely random process. $n_{\text{sim}} = 4{,}750{,}000$ random samples of $\{X(i)\}_{i=1}^{n}$ were generated after $X(i) = \exp\left[\mathcal{E}_{\mathrm{N}(\mu,\,\sigma^2)}(i)\right]$, $i = 1, \ldots, n$, with $\mu = 1.0$, $\sigma = 1.0$ and various n values. The density function is skewed (Fig. 3.2). Analysed as estimators of the centre of location of the distribution were the sample mean (Eq. 3.16) and the sample median, $\widehat{m}$ (see background material, Section 3.8). $\mathrm{CI}_{\bar{x},1-2\alpha}$ was constructed after Eq. (3.18) with $\alpha = 0.025$.

n	$\gamma_{\bar{x}}^{\mathrm{a}}$	*Nominal*	C^{b}
10	0.8392	0.9500	−0.1108
20	0.8670	0.9500	−0.0830
50	0.8991	0.9500	−0.0509
100	0.9170	0.9500	−0.0330
200	0.9296	0.9500	−0.0204
500	0.9399	0.9500	−0.0101
1000	0.9442	0.9500	−0.0058

3.1 Error bars and confidence intervals

Real world

- Violated assumptions (e.g., distributional shape)
- CIs not exact, but only approximate
- CIs have coverage error, C
- C decreases with n
  - $C \sim 1/n^{1/2}$ first-order accurate CI
  - $C \sim 1/n$ second-order accurate CI

3.1 Error bars and confidence intervals

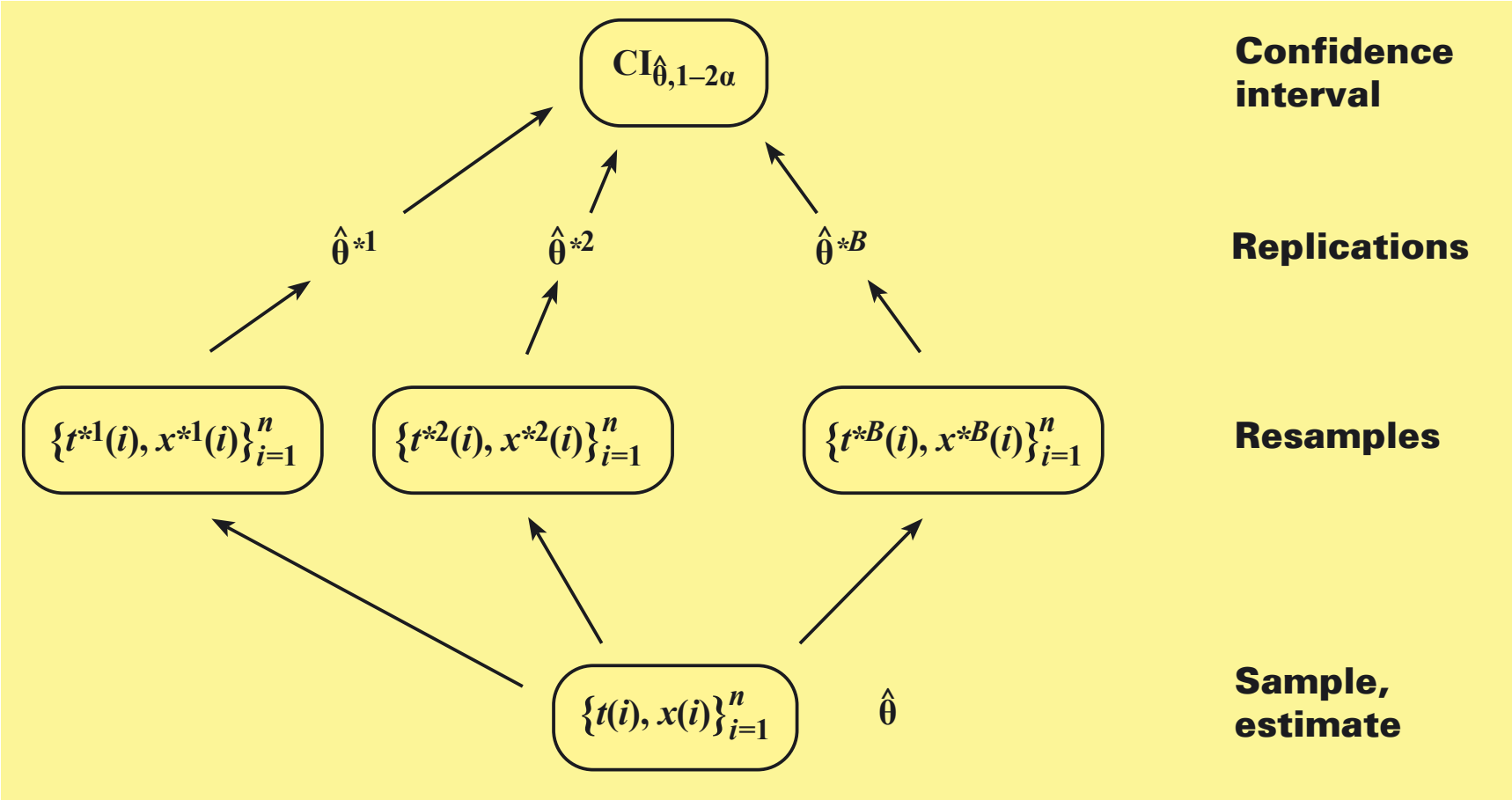
Real world

- Complex processes (distributional shape, persistence, spacing)
- Complex estimation problems
- PDF of estimator cannot be derived analytically
- Approximate CIs

Bootstrap

- Real world
- Method to construct (approximate) CIs

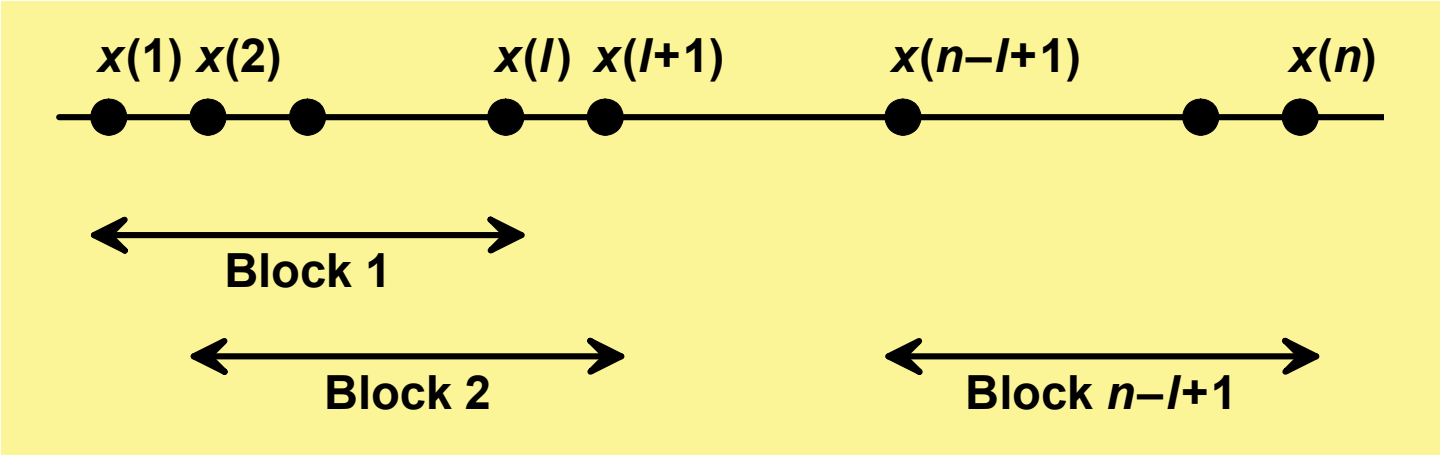
3.2 Bootstrap principle



3.2 Bootstrap principle

- Bootstrap
  - Resampling
    - Moving block bootstrap (MBB)
    - Autoregressive bootstrap (ARB)
    - Other
  - Confidence interval construction
    - Normal
    - Student's *t*
    - Percentile
    - BCa

3.3 Bootstrap resampling



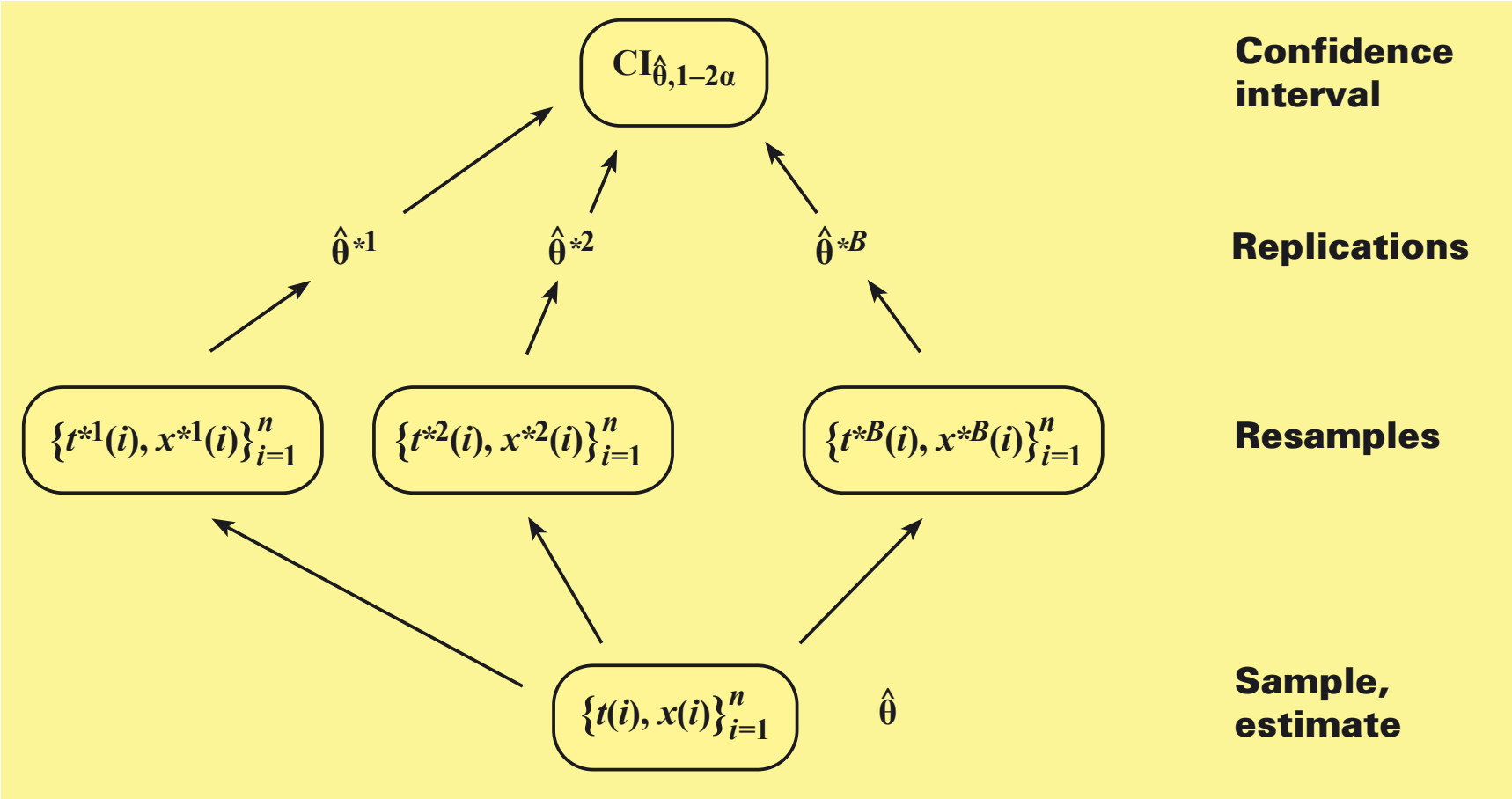
Moving block bootstrap (MBB)

3.3 Bootstrap resampling

Step 1	Data	$\{t(i), x(i)\}_{i=1}^{n}$
Step 2	Resampled times	$\{t^*(i)\}_{i=1}^{n} = \{t(i)\}_{i=1}^{n}$
	unchanged	
Step 3	Residuals (Eq. 3.29)	$r(i) = [x(i) - \widehat{x}_{\text{trend}}(i) - \widehat{x}_{\text{out}}(i)] \Big/ \widehat{S}(i)$
Step 4	Apply MBB	
	(Algorithm 3.1)	
	to residuals	$\{r(i)\}_{i=1}^{n}$
Step 5	Resampled residuals	$\{r^*(i)\}_{i=1}^{n}$
Step 6	Use resampled residuals	
	to produce resamples	$x^*(i) = \widehat{x}_{\text{trend}}(i) + \widehat{x}_{\text{out}}(i) + \widehat{S}(i) \cdot r^*(i)$

Algorithm 3.3. MBB for realistic climate processes, which comprise trend, outlier and variability components.

3.4 Bootstrap confidence intervals



3.4 Bootstrap confidence intervals

$$\widehat{\mathrm{se}}_{\widehat{\theta}^*} = \left\{ \sum_{b=1}^{B} \left[\widehat{\theta}^{*b} - \left\langle \widehat{\theta}^{*b} \right\rangle \right]^2 \Big/ (B-1) \right\}^{1/2}, \qquad (3.30)$$

Bootstrap standard error

3.4 Bootstrap confidence intervals

$$\mathrm{CI}_{\widehat{\theta},1-2\alpha} = \left[\widehat{\theta} + z(\alpha) \cdot \widehat{\mathrm{se}}_{\widehat{\theta}*}; \widehat{\theta} - z(\alpha) \cdot \widehat{\mathrm{se}}_{\widehat{\theta}*}\right], \tag{3.31}$$

Normal confidence interval

$z(\alpha)$ Percentage point of normal distribution

Example

$z(1 - 0.025) \approx 1.959964$

"±2-sigma interval is 95% CI."

3.4 Bootstrap confidence intervals

$$\mathrm{CI}_{\widehat{\theta},1-2\alpha} = \left[\widehat{\theta} + t_\nu(\alpha) \cdot \widehat{\mathrm{se}}_{\widehat{\theta}*}; \widehat{\theta} - t_\nu(\alpha) \cdot \widehat{\mathrm{se}}_{\widehat{\theta}*}\right], \qquad (3.32)$$

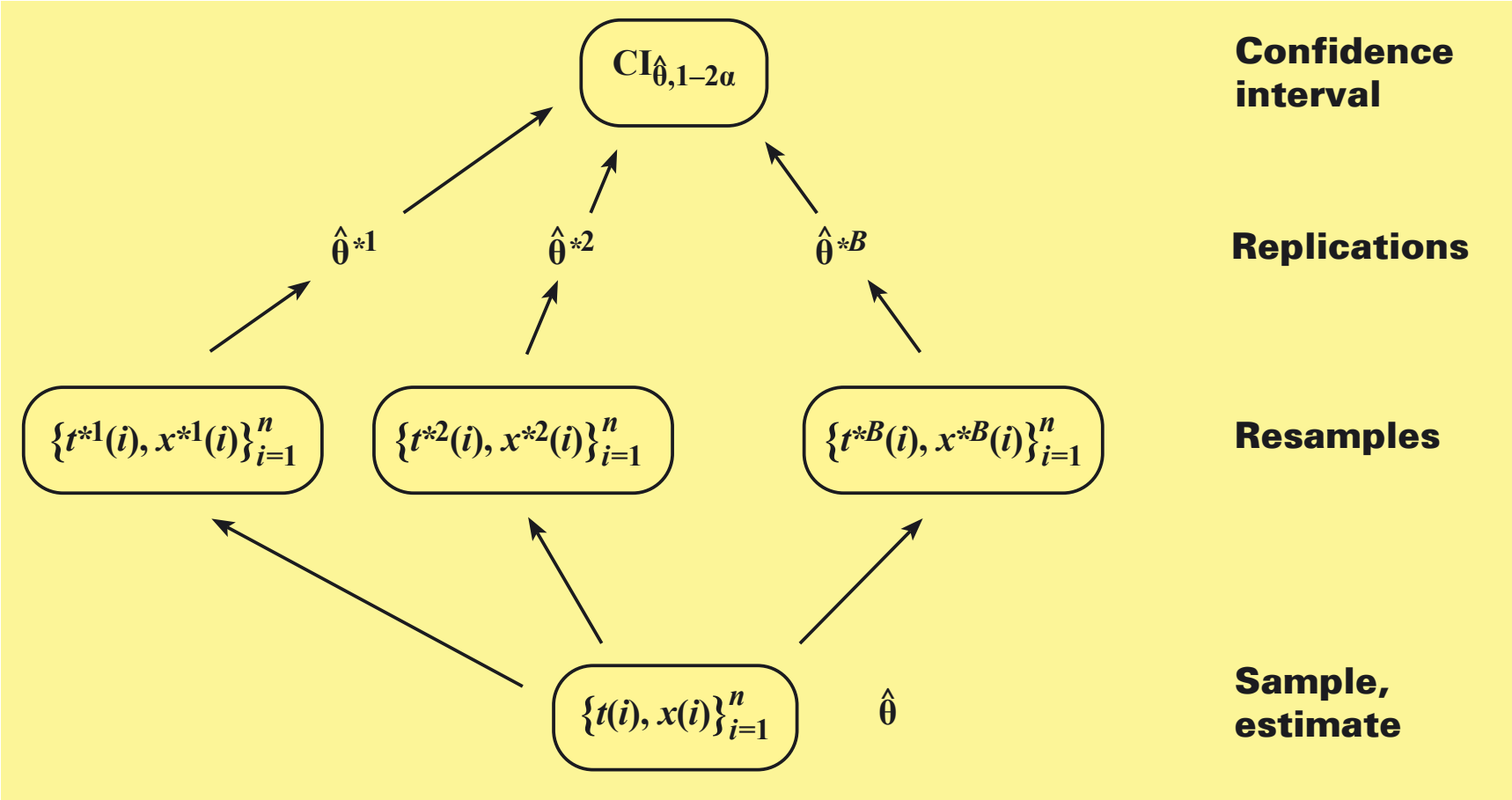
Student's *t* confidence interval

$t_v(\alpha)$ Percentage point of Student's *t*-distribution
with v degrees of freedom ($v = n$ – number of parameters)

Takes into account that not only standard error,
but also fit value estimated.

Negligible difference to normal CI if v above ~ 30

3.4 Bootstrap confidence intervals

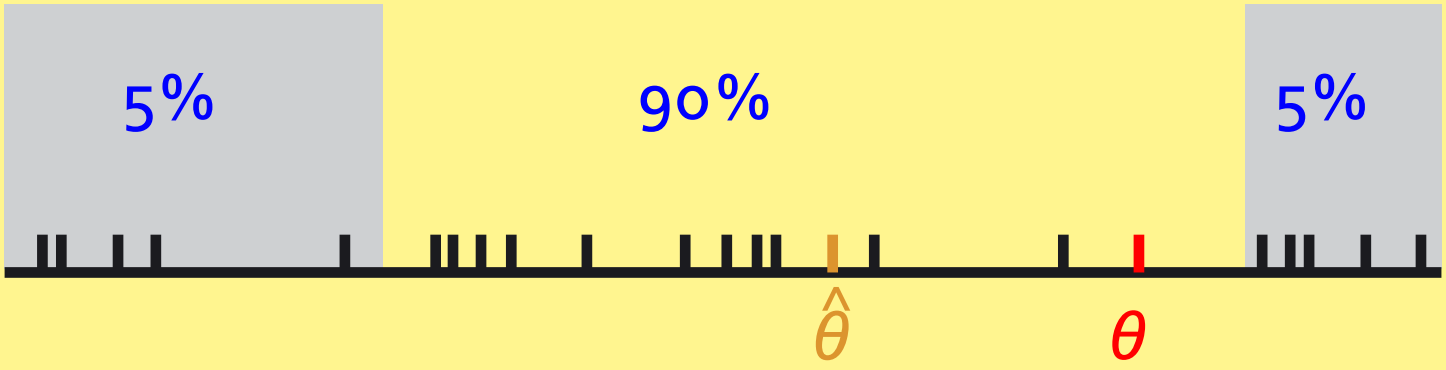


3.4 Bootstrap confidence intervals

Construction of equi-tailed percentile CIs

2 Percentile CI

1 $\hat{\theta}$, $\{\hat{\theta}^{*b}\}_{b=1}^{B}$



3.4 Bootstrap confidence intervals

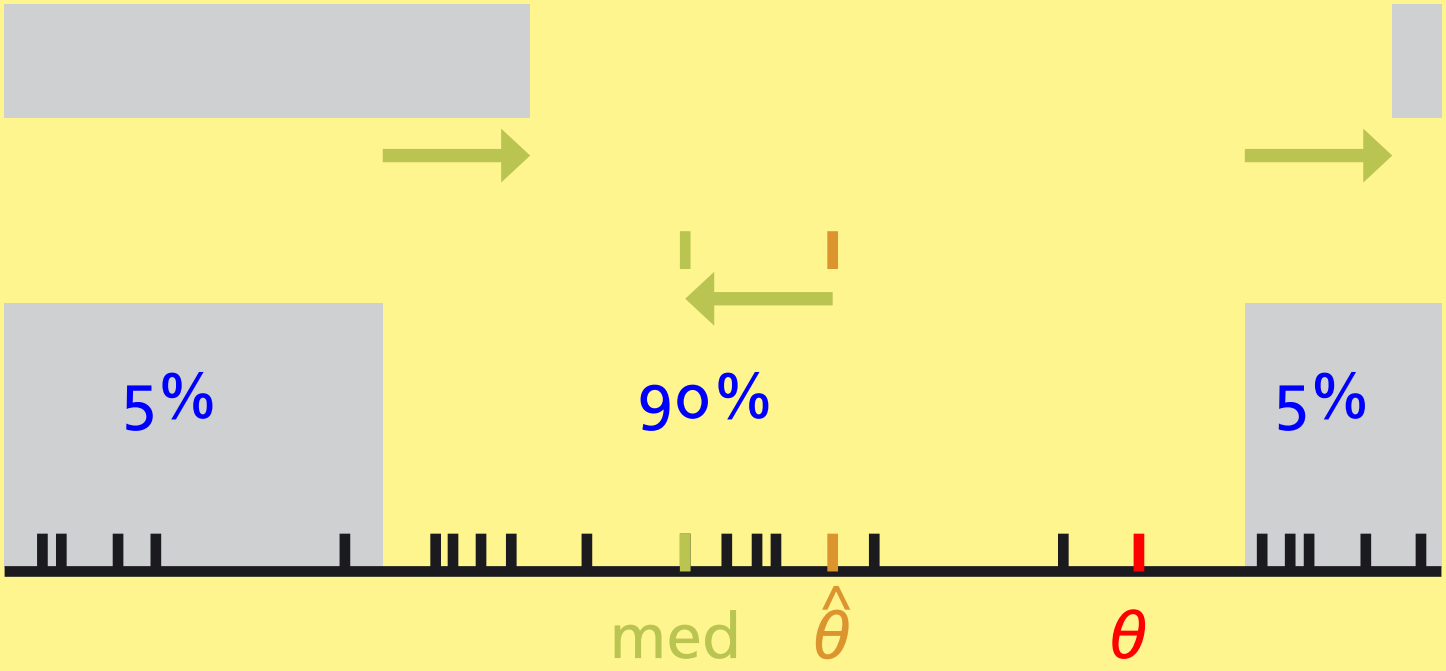
Construction of equi-tailed percentile CIs

4 Bias correction

3 $\hat{\theta}$, $\text{med}\{\hat{\theta}^{*b}\}_{b=1}$

2 Percentile CI

1 $\hat{\theta}$, $\{\hat{\theta}^{*b}\}_{b=1}^{B}$



3.6 Bootstrap hypothesis tests

Statistical inference

- Estimating parameter θ
- Testing hypothesis (true/false)

3.6 Bootstrap hypothesis tests

H_o	Null hypothesis
H_1	Alternative hypothesis (H_o and H_1 mutually exclusive)
U, u	Test statistic
	Our choice, a guide to help us decide (H_o or H_1)
$F_o(u)$	Distribution function of U "under H_o"

$$\begin{aligned} P &= \text{prob}\,(U \geq u \mid H_0) \\ &= 1 - F_0(u). \end{aligned}$$

$$P = \text{prob}\,(|U| \geq |u| \mid H_0)\,.$$

P	P-value (one-sided)	P-value (two-sided)

If P is small, then we reject H_o and accept H_1.

3.6 Bootstrap hypothesis tests

H_o Null hypothesis ("Gaussian white noise process, mean $\mu = 0$")

H_1 Alternative hypothesis (H_o and H_1 mutually exclusive) ("$\mu \neq 0$")

U, u Test statistic (sample mean, $U = \Sigma X(i)/n$, $u = \Sigma x(i)/n$)

Our choice, a guide to help us decide (H_o or H_1)

$F_o(u)$ Distribution function of U "under H_o" (Student's t-distribution)

$$\begin{aligned} P &= \text{prob}\,(U \geq u \mid H_0) \\ &= 1 - F_0(u). \end{aligned}$$

$$P = \text{prob}\,(|U| \geq |u| \mid H_0)\,.$$

P P-value (one-sided)

P-value (two-sided)

If P is small, then we reject H_o and accept H_1.

3.6 Bootstrap hypothesis tests

Real world

- Complex processes (distributional shape, persistence, spacing)
- Complex constructions of test statistic u (e.g., fingerprint test)
- $F_o(u)$ cannot be derived analytically

Bootstrap

- Real world
- Method to approximate $F_o(u)$

Book, Section 3.6; Hasselmann (1993), Hegerl et al. (1996), Lehmann and Romano (2005)

3.6 Bootstrap hypothesis tests

Give *P*-values (do not just state whether e.g. $P < 0.01$).

A tested hypothesis should be realistic, not a "straw man"!

A confidence interval contains more quantitative information than a hypothesis test result.

4. Regression I

4.1 Linear regression

4.2 Nonlinear regression

4.3

4. Regression I

$$X(i) = X_{\text{trend}}(i) + X_{\text{out}}(i) + S(i) \cdot X_{\text{noise}}(i), \qquad (1.2)$$

Climate equation, stochastic process $X(i)$

Time series $\{t(i), x(i)\}_{i=1}^{n}$

Time series analysis uses sample to learn about process:
trend parameters, probability of extremes, cycles, etc.

Ignore $X_{\text{out}}(i)$,
focus on $X_{\text{trend}}(i)$ and also $S(i)$.

Brückner (1890), Hann (1901), Köppen (1923)

4. Regression I

$X_{\text{trend}}(i)$

Parametric trend model

- Linear
- Nonlinear

Nonparametric trend model

4.1 Linear regression

$$X(i) = \beta_0 + \beta_1 T(i) + S(i) \cdot X_{\text{noise}}(i). \tag{4.3}$$

Linear trend model

$$SSQW(\beta_0, \beta_1) = \sum_{i=1}^{n} \left[x(i) - \beta_0 - \beta_1 t(i)\right]^2 / S(i)^2 \ , \tag{4.4}$$

Estimation

Weighted least squares (WLS)

4.1 Linear regression

$$\widehat{\beta}_0 = \left[\sum_{i=1}^{n} x(i)/S(i)^2 - \widehat{\beta}_1 \sum_{i=1}^{n} t(i)/S(i)^2\right] \Big/ W, \tag{4.5}$$

$$\widehat{\beta}_1 = \left\{\left[\sum_{i=1}^{n} t(i)/S(i)^2\right]\left[\sum_{i=1}^{n} x(i)/S(i)^2\right] \Big/ W - \sum_{i=1}^{n} t(i)\, x(i)/S(i)^2\right\}$$

$$\times \left\{\left[\sum_{i=1}^{n} t(i)/S(i)^2\right]^2 \Big/ W - \sum_{i=1}^{n} t(i)^2/S(i)^2\right\}^{-1}, \tag{4.6}$$

$$W = \sum_{i=1}^{n} 1/S(i)^2. \tag{4.7}$$

4.1 Linear regression

$$X(i) = \beta_0 + \beta_1 T(i) + S(i) \cdot X_{\text{noise}}(i). \qquad (4.3)$$

Linear trend model

$$SSQW(\beta_0, \beta_1) = \sum_{i=1}^{n} \left[x(i) - \beta_0 - \beta_1 t(i)\right]^2 / S(i)^2 \ , \qquad (4.4)$$

Estimation

- Ordinary least squares (OLS)
  - $S(i) = S$ = const. (known or unknown)
  - no weighting, even easier to calculate than WLS

4.1 Linear regression

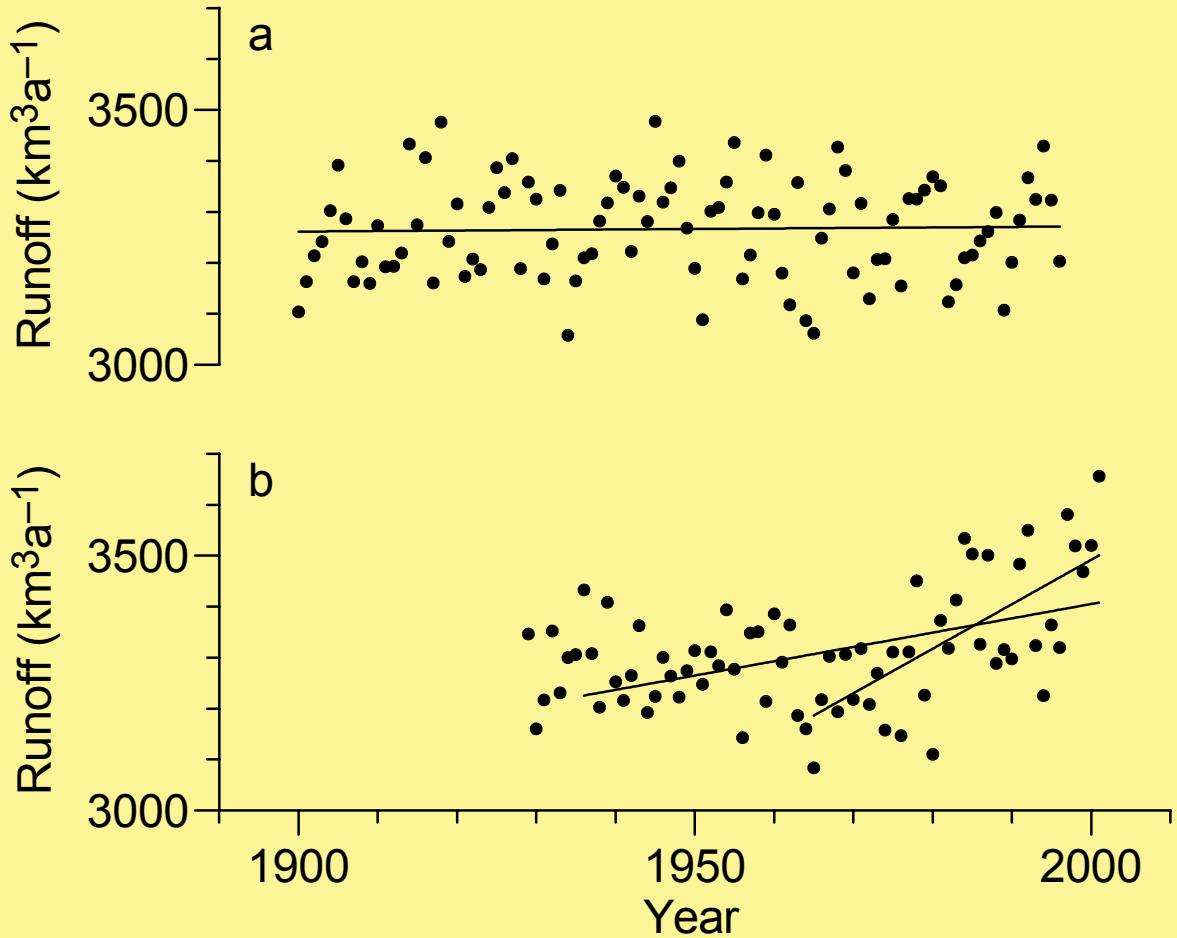


Figure 4.1. Linear regression models fitted to modelled Arctic river runoff (Fig. 1.9). **a** Natural forcing only; **b** combined anthropogenic and natural forcing. Following

Example (OLS)

a $\hat{\beta}_0$ = 3068 km³a⁻¹, $\hat{\beta}_1$ = 0.102 km³a⁻²

4.1 Linear regression

$$X(i) = \beta_0 + \beta_1 T(i) + S(i) \cdot X_{\text{noise}}(i). \tag{4.3}$$

Linear trend model

$$SSQG(\boldsymbol{\beta}) = (\mathbf{x} - \mathbf{T}\boldsymbol{\beta})' \, \mathbf{V}^{-1} \, (\mathbf{x} - \mathbf{T}\boldsymbol{\beta}) \, . \tag{4.9}$$

Estimation

- Generalized least squares (GLS)
  - ***V***, covariance matrix (contains persistence and variability)

4.1 Linear regression

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \text{ (parameter vector)}, \tag{4.10}$$

$$\mathbf{x} = \begin{bmatrix} x(1) \\ \vdots \\ x(n) \end{bmatrix} \text{ (data vector)}, \tag{4.11}$$

$$\mathbf{T} = \begin{bmatrix} 1 & t(1) \\ \vdots & \vdots \\ 1 & t(n) \end{bmatrix} \text{ (time matrix)} \tag{4.12}$$

4.1 Linear regression

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{T}'\mathbf{V}^{-1}\mathbf{T}\right)^{-1}\mathbf{T}'\mathbf{V}^{-1}\mathbf{x}. \tag{4.13}$$

$$V(i_1, i_2) = S(i_1) \cdot S(i_2) \cdot E\left[X_{\text{noise}}(i_1) \cdot X_{\text{noise}}(i_2)\right], \tag{4.14}$$

Covariance matrix elements

$$\widehat{V}(i_1, i_2) = \widehat{S}(i_1) \cdot \widehat{S}(i_2) \cdot \exp\left[-|t(i_1) - t(i_2)|/\widehat{\tau}'\right], \tag{4.15}$$

Covariance matrix elements

- estimated
- AR(1) persistence
- bias-corrected persistence time

4.1 Linear regression

$$r(i) = \left[x(i) - \widehat{\beta}_0 - \widehat{\beta}_1 t(i)\right] \Big/ \widehat{S}(i), \qquad (4.16)$$

Regression residuals, weighted

$$e(i) = \left[x(i) - \widehat{\beta}_0 - \widehat{\beta}_1 t(i)\right]$$

Regression residuals, unweighted

4.1 Linear regression

Assume

1. Noise has Gaussian shape
2. Covariance matrix known (or correctly estimated)
3. Linear regression model suitable

Then

$$\mathrm{CI}_{\widehat{\beta}_j,1-2\alpha} = \left[\widehat{\beta}_j + t_{n-2}(\alpha)\cdot \mathrm{se}_{\widehat{\beta}_j}; \widehat{\beta}_j + t_{n-2}(1-\alpha)\cdot \mathrm{se}_{\widehat{\beta}_j}\right], \qquad (4.20)$$

Classical CI (j = 0, intercept; j = 1, slope)

$$\mathrm{se}_{\widehat{\beta}_j} = [C(j,j)]^{1/2}, \qquad \mathbf{C} = \left(\mathbf{T}'\mathbf{V}^{-1}\mathbf{T}\right)^{-1}.$$

OLS

special case of GLS

zero off-diagonal and equal ***V*** diagonal matrix elements

$$\mathrm{se}_{\widehat{\beta}_0} = MS_E^{1/2}\left\{1/n + \left(\sum_{i=1}^{n} t(i)/n\right)^2 \Big/ \left[\sum_{i=1}^{n} t(i)^2 - \left(\sum_{i=1}^{n} t(i)\right)^2 \Big/ n\right]\right\}^{1/2}, \tag{4.24}$$

$$\mathrm{se}_{\widehat{\beta}_1} = MS_E^{1/2}\left[\sum_{i=1}^{n} t(i)^2 - \left(\sum_{i=1}^{n} t(i)\right)^2 \Big/ n\right]^{-1/2}. \tag{4.25}$$

$$\widehat{S}(i) = \widehat{S} = \left\{\sum_{i=1}^{n}\left[x(i) - \widehat{\beta}_0 - \widehat{\beta}_1 t(i)\right]^2 \Big/ (n-2)\right\}^{1/2} = MS_E^{1/2}. \tag{4.8}$$

MS_E, residual mean square

4.1 Linear regression

Conclusion 1

If we know

variability $S(i)$

persistence

τ (uneven spacing)

a (even spacing)

(that means, covariance matrix $\boldsymbol{V}$)

$\Rightarrow$ Simple calculations, straightforward estimation

Conclusion 2

$$\text{OLS} \in \text{WLS} \in \text{GLS}$$

4.1 Linear regression

Real world

unknown variability $S(i)$

unknown persistence

τ (uneven spacing)

a (even spacing)

that means, unknown covariance matrix $\boldsymbol{V}$

⇒ Iteratively (Prais–Winsten procedure)

⇒ Quick-and-dirty: OLS with effective data size

4.1 Linear regression

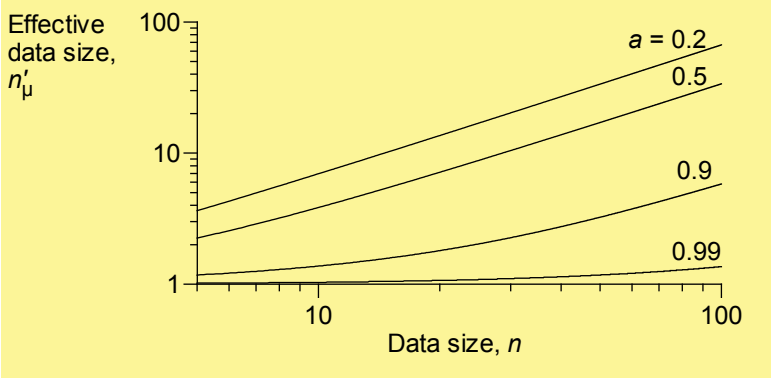
Via effective data size, $n'_\mu < n$

$$\widehat{S}(i) = \widehat{S} = \left\{ \sum_{i=1}^{n} \left[x(i) - \widehat{\beta}_0 - \widehat{\beta}_1 t(i) \right]^2 \Big/ (n-2) \right\}^{1/2} = MS_E^{1/2}. \quad (4.8)$$

Insert in Eq. (4.8) n'_μ for n

⇒ Larger residual mean square

⇒ Wider classical OLS CI



Caveats

No re-estimation of persistence parameter

n'_μ applies to mean estimation, not regression

4.1 Linear regression

Table 4.1. Monte Carlo experiment, linear OLS regression with AR(1) noise of normal shape, even spacing: CI coverage performance. $n_{\text{sim}} = 47{,}500$ random samples were generated from $X(i) = 2 + 2T(i) + X_{\text{noise}}(i)$, where $T(i) = i, i = 1, \ldots, n$ and the noise is a Gaussian AR(1) process (Eq. 2.1) with $a = 1/\mathrm{e} \approx 0.37$. Two CI types for the estimated slope were constructed, classical and bootstrap. Construction of classical CIs either ignored persistence and calculated via n (Eqs. 4.8, 4.20 and 4.25) or used n'_μ (Section 4.1.4.3). The bootstrap CIs used ARB (Algorithm 3.4) or MBB (Algorithm 3.1) resampling and the BCa method (Section 3.4.4) with $B = 1999$ and $\alpha = 0.025$.

n	$\gamma^{\mathrm{a}}_{\widehat{\beta}_1}$		*Nominal*
	CI type		
	Classical		
	Via n	*Via* n'_μ	
10	0.851	0.900	0.950
20	0.832	0.915	0.950
50	0.819	0.932	0.950
100	0.817	0.941	0.950
200	0.819	0.945	0.950
500	0.818	0.947	0.950
1000	0.816	0.950	0.950

4.1 Linear regression

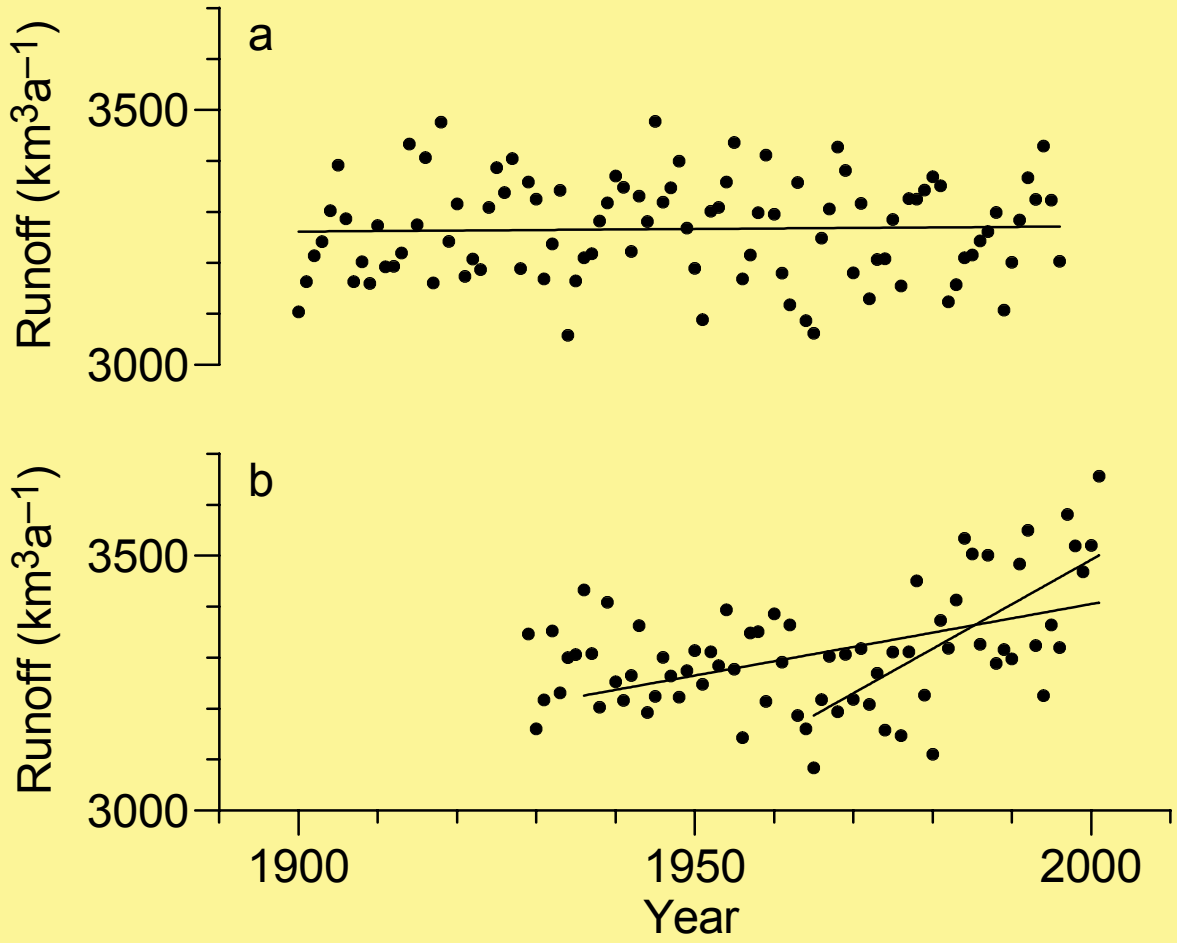


Figure 4.1. Linear regression models fitted to modelled Arctic river runoff (Fig. 1.9). **a** Natural forcing only; **b** combined anthropogenic and natural forcing. Following

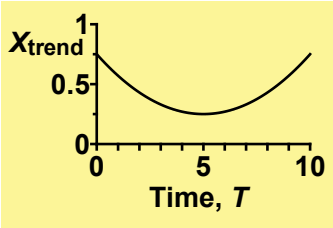
Example (OLS)

a $\hat{\beta}_0 = 3068 \pm 694$ km³a⁻¹, $\hat{\beta}_1 = 0.102 \pm 0.356$ km³a⁻² (via *n;* Wu et al. 2005)

a $\hat{\beta}_0 = 3068 \pm 859$ km³a⁻¹, $\hat{\beta}_1 = 0.102 \pm 0.441$ km³a⁻² (via n'_μ)

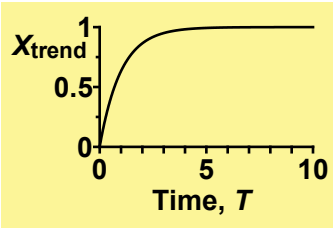
4.2 Nonlinear regression

$$X(i) = \beta_0 + \beta_1 T(i) + \beta_2 T(i)^2 + S(i) \cdot X_{\text{noise}}(i), \qquad (4.37)$$



Parabolic trend model: not really nonlinear (i.e, in parameters)

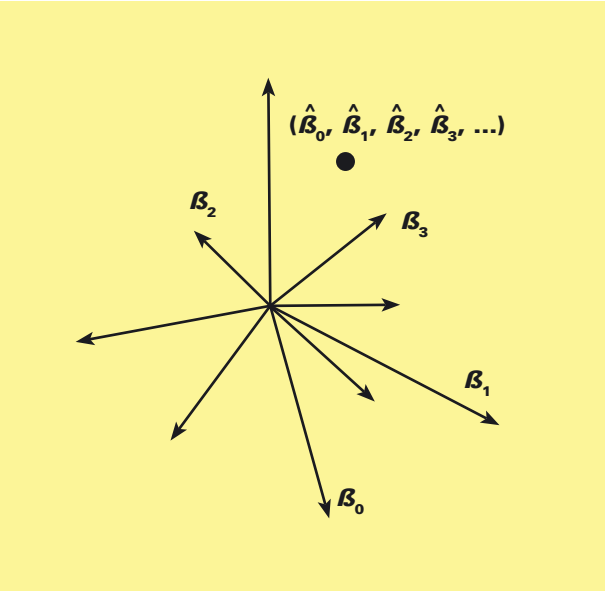
$$X(i) = \beta_0 \left\{1 - \exp\left[-\beta_1 T(i)\right]\right\} + S(i) \cdot X_{\text{noise}}(i). \qquad (4.38)$$



Saturation trend function: nonlinear

4.2 Nonlinear regression

High-dimensional parameter space ($\boldsymbol{\beta}$)



Estimation

Search for minimum of

$SSQG(\boldsymbol{\beta})$

Often no exact solution,

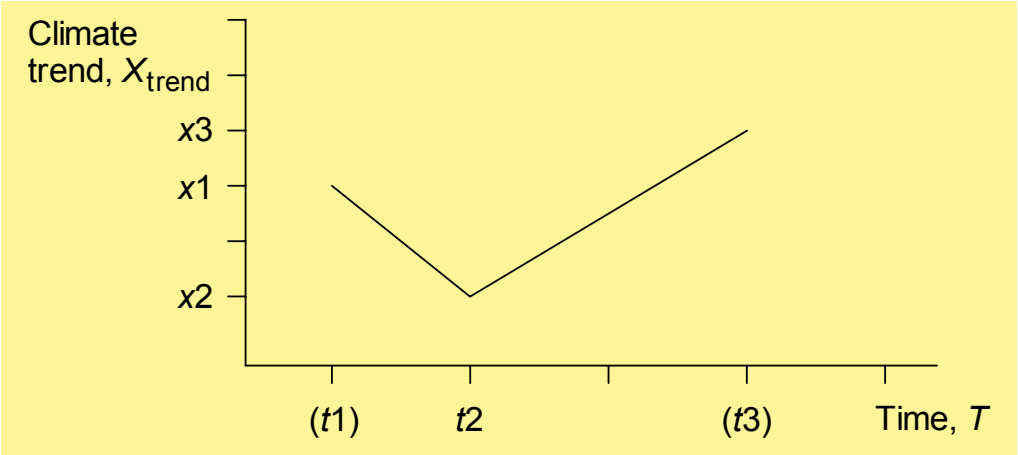
but numerical techniques

(1) Guess starting values.

(2) Go step (of defined size) into negative gradient direction.

(3) If changes small: exit, otherwise go back to (1).

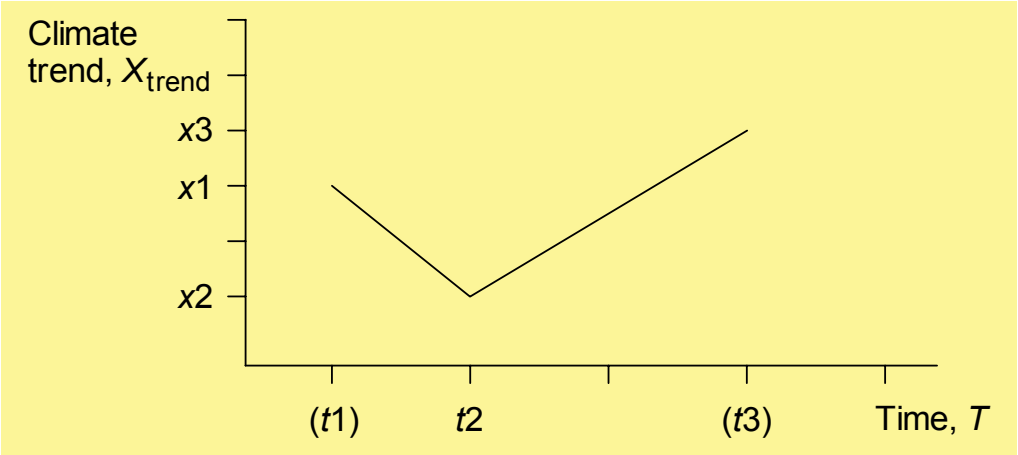
4.2 Nonlinear regression



Trend-change model: break

$$X_{\text{break}}(T) = \begin{cases} x1 + (T - t1)(x2 - x1)/(t2 - t1) & \text{for } T \leq t2, \\ x2 + (T - t2)(x3 - x2)/(t3 - t2) & \text{for } T > t2, \end{cases} \tag{4.47}$$

4.2 Nonlinear regression



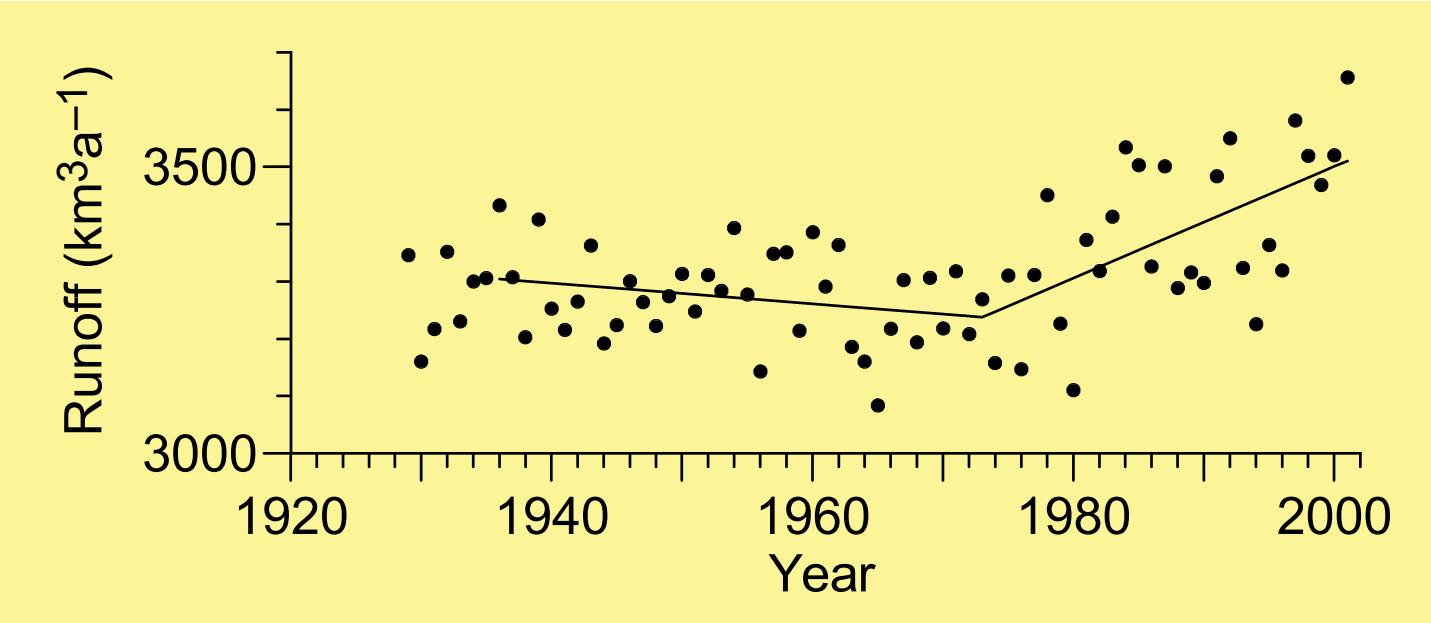
Trend-change model: break

$$SSQW(x1, t2, x2, x3) = \sum_{i=1}^{n} \left[x(i) - x_{\text{break}}(i)\right]^2 / S(i)^2 \,, \qquad (4.48)$$

Estimation (WLS combined with brute-force search for *t*2)

Book, Fig. 4.11

4.2 Nonlinear regression



Break point in Arctic river runoff

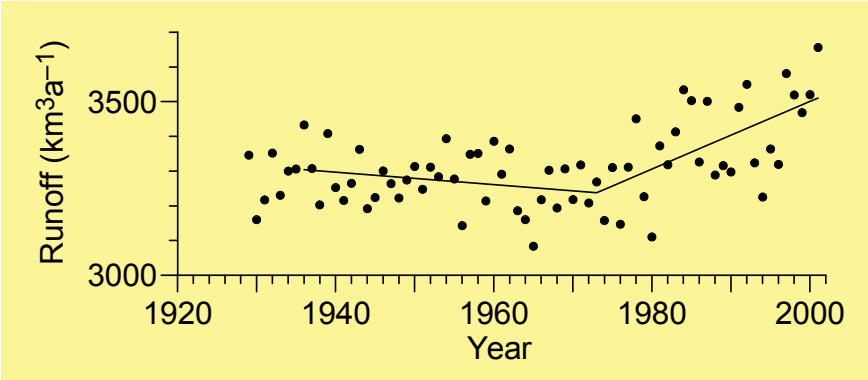
$\hat{t}_2 = 1973$

$\hat{\beta}_1 = -1.8\ \mathrm{km^3a^{-2}}$

$\hat{\beta}_2 = 9.7\ \mathrm{km^3a^{-2}}$

4.2 Nonlinear regression

Error bars?



Classical formulas do not exist, gradient not defined

⇒ Bootstrap resampling

Break point in Arctic river runoff

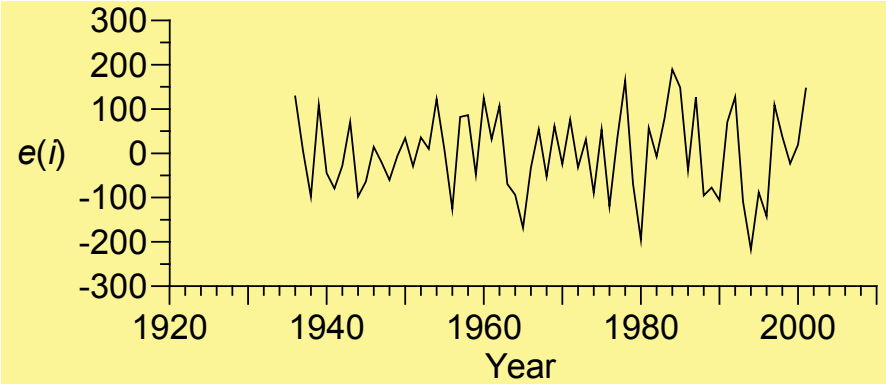
$\hat{t}_2 = 1973$ $\hat{\beta}_1 = -1.8\ \mathrm{km^3a^{-2}}$

$\hat{\beta}_2 = 9.7\ \mathrm{km^3a^{-2}}$

4.2 Nonlinear regression

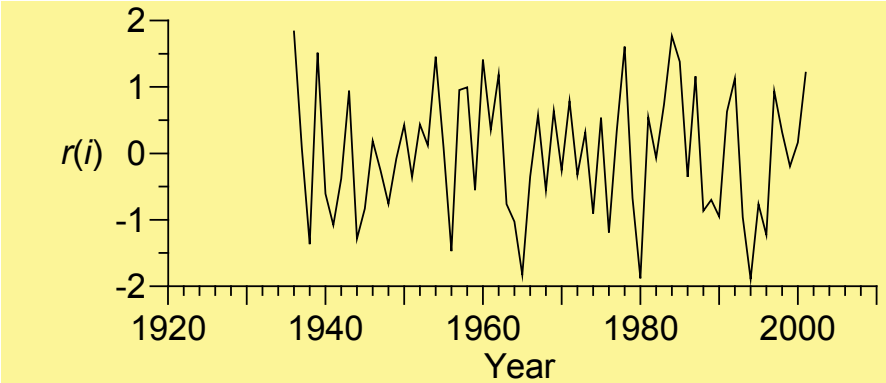
$$e(i) = x(i) - \widehat{x}_{\text{break}}(i), i = 1, \ldots, n;$$

Unweighted residuals



$$r(i) = e(i) \Big/ \widehat{S}(i)\ , \qquad i = 1, \ldots, n.$$

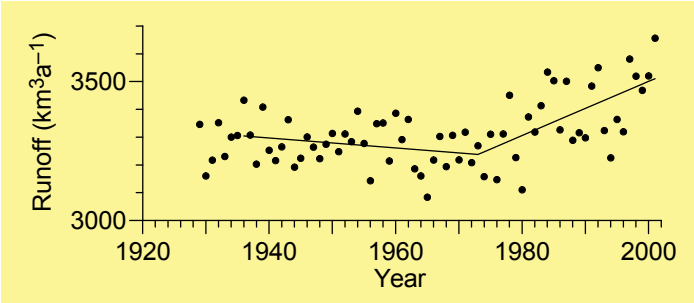
Weighted residuals

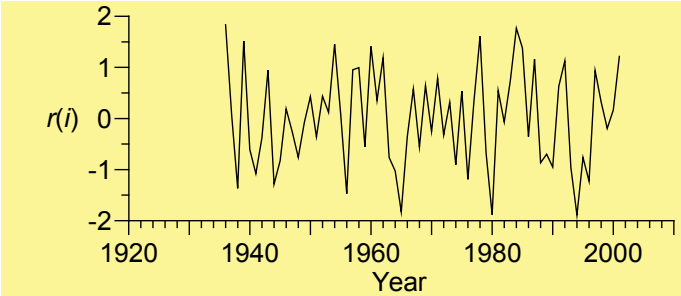


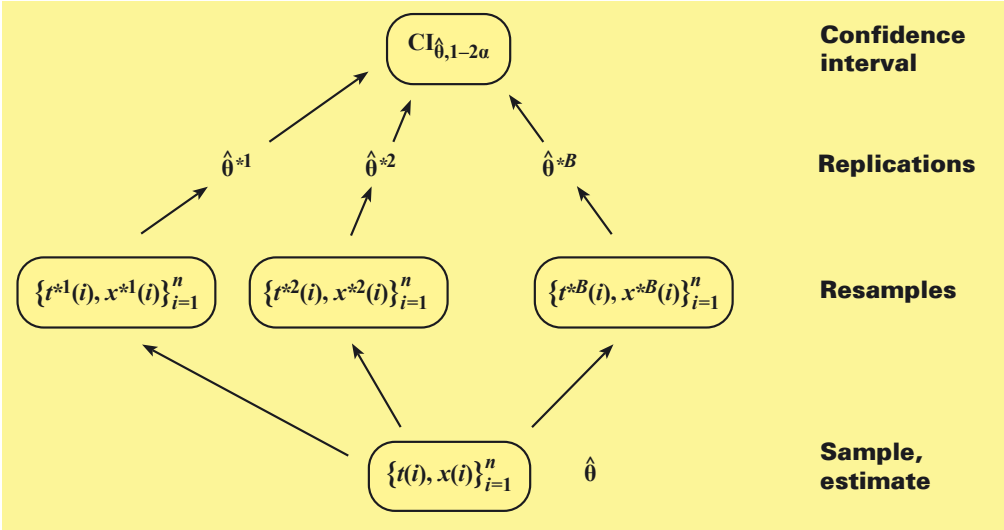
4.2 Nonlinear regression

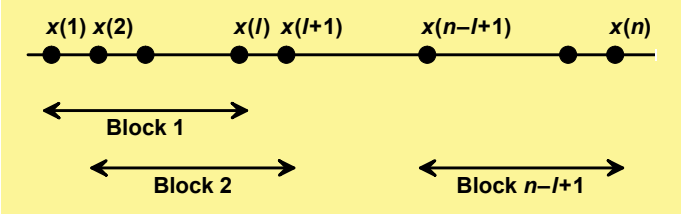
Bootstrap error bars

- Take fit curve
- Add blocks of residuals
- Re-estimate: replication
- Repeat many times
- Calculate standard deviation (replications)

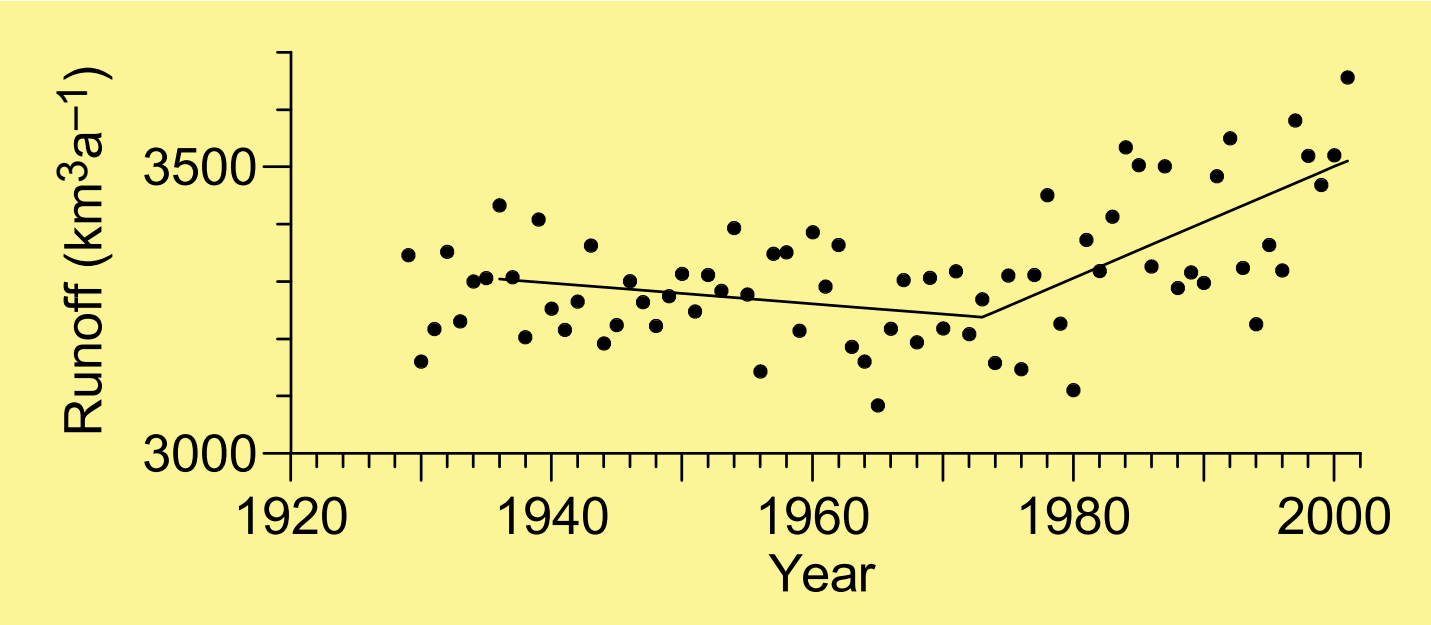








4.2 Nonlinear regression



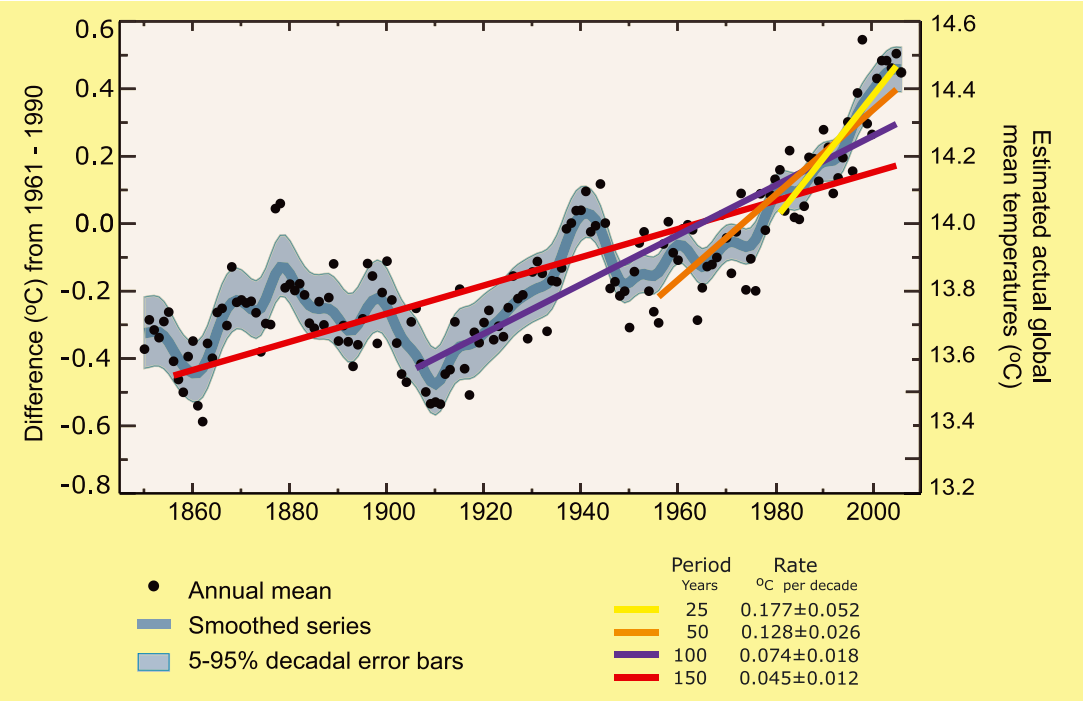
Break point in Arctic river runoff

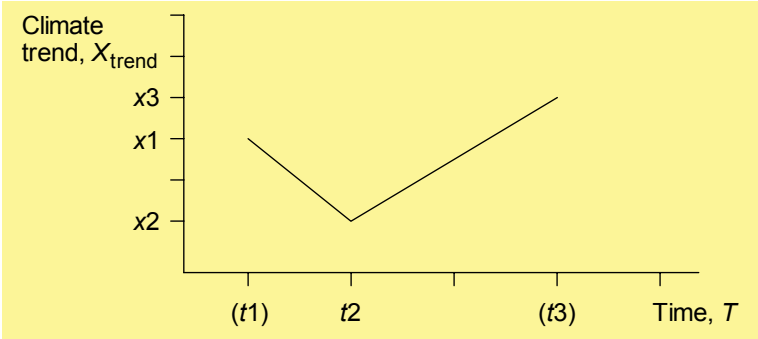
$\hat{t}2 = 1973 \pm 6$ $\quad \hat{\beta}_1 = -1.8 \pm 1.6\ \mathrm{km^3a^{-2}}$

$\hat{\beta}_2 = 9.7 \pm 3.6\ \mathrm{km^3a^{-2}}$

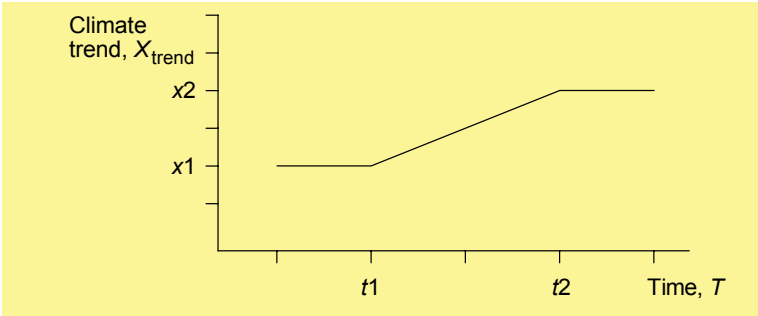
Book, Fig. 4.12; bootstrap error bars: MBB, $B = 400$

4.2 Nonlinear regression—Homework





Break regression model



Ramp regression model

Solomon et al. 2007 (p. 37)

www.manfredmudelsee.com/soft/BREAKFIT
www.manfredmudelsee.com/soft/RAMPFIT

Climate Time Series Analysis: Conclusions

(1) $\hat{\theta} \neq \theta$

(2) Computational techniques (bootstrap, Monte Carlo) help to obtain realistic error bars.

(3) Estimation problems (linear, break, ramp, etc.):
Be creative!

Climate Time Series Analysis: Conclusions

(1) $\hat{\theta} \neq \theta$

(2) Computational techniques (bootstrap, Monte Carlo) help to obtain realistic error bars.

(3) Estimation problems (linear, break, ramp, etc.): Be creative!

Thanks!

Postdoc job: www.climate-risk-analysis.com

2.5 Climate theory

Climate, $X(T)$
Weather, $Y(T)$

$$\begin{aligned} \frac{dX(T)}{dT} &= F(X(T), Y(T)), \qquad \text{timescale } \tau_X, \\ \frac{dY(T)}{dT} &= G(X(T), Y(T)), \qquad \text{timescale } \tau_Y, \end{aligned}$$

Weather components
"leave a trace" in climate.

$$\begin{aligned} \frac{dX(T)}{dT} &\simeq F(X(0), Y(T)), \\ &= W(T), \end{aligned}$$

$$X(T+1) = X(T) + \mathcal{E}_{\mathrm{N}(0,\,\sigma^2)}(T).$$

Climate does not run away:
negative feedback (~ X)

$$X(T+1) = a \cdot X(T) + \mathcal{E}_{\mathrm{N}(0,\,\sigma^2)}(T),$$

$W(T)$, Wiener process ("continuous-time noise increment")
Hasselmann (1976); Book, Eqs. (2.25), (2.26), (2.28), (2.29), (2.30), Section 2.5.1

2.5 Climate theory

Support of simple AR(1) climate noise model

1. Weather components (Hasselmann 1976)
2. Climate equation $$X(i) = X_{\text{trend}}(i) + X_{\text{out}}(i) + S(i) \cdot X_{\text{noise}}(i),$$
3. We have also trend, outliers/extremes and variability to describe climate.

Book, Eq. (1.2)

$$l_{\text{opt}} = NINT\left\{\left[6^{1/2}\cdot\widehat{\bar{a}}\Big/\left(1-\widehat{\bar{a}}^2\right)\right]^{2/3}\cdot n^{1/3}\right\}, \qquad (3.28)$$

Block length selector (MBB) after Carlstein (1986)

Other block length selectors via

- persistence time
- autocorrelation function

Step 1 Make an initial guess of the variability, $\widehat{S}(i)$. Make an initial guess of the persistence time, $\widehat{\tau}$, for uneven spacing or the autocorrelation parameter, $\widehat{a}$, for even spacing. (For notational convenience, we omit writing a superscript.) Often the simple choices of constant variability and absent persistence are sufficient.

Step 2 Update the $\mathbf{V}$ matrix after Eq. (4.15).

Step 3 Perform GLS (Eq. 4.13) to obtain updates of regression parameter estimates. Update also $SSQG$ (Eq. 4.9).

Step 4 If regression parameters, and possibly also $SSQG$, have not changed strongly at the preceding step, take this solution and go to Step 8.

Step 5 Calculate the unweighted regression residuals $e(i)$. Update the variability estimate $\widehat{S}(i)$ by using the $e(i)$.

Step 6 Calculate the weighted regression residuals, $r(i)$, after Eq. (4.16). Update the persistence estimate $\widehat{\tau}$ or $\widehat{a}$ by using the $r(i)$. Take bias correction (Section 2.6) into account.

Step 7 Go to Step 2.

Classical CI

Prais–Winsten procedure

Step 7 Go to Step 2.

Step 8 After regression parameters, and possibly also $SSQG$, approached the solution at Step 4, calculate the residuals $r(i)$ and the standard errors (Eq. 4.21). The residuals can be used for graphical analysis (Montgomery and Peck 1992). One looks whether they fail to reveal more structure than a Gaussian AR(1) process, that means, whether the linear regression model with AR(1) noise is a suitable description of the data. The standard errors can be used for classical CI construction (Eq. 4.20).

Classical CI

Prais–Winsten procedure