

7. Begriffe der Schätztheorie

Inhalt

7.1 Unterdrückung von Störgrößen

- Störgrößenkompensation
- Mittelwertbildung
- Schätzung

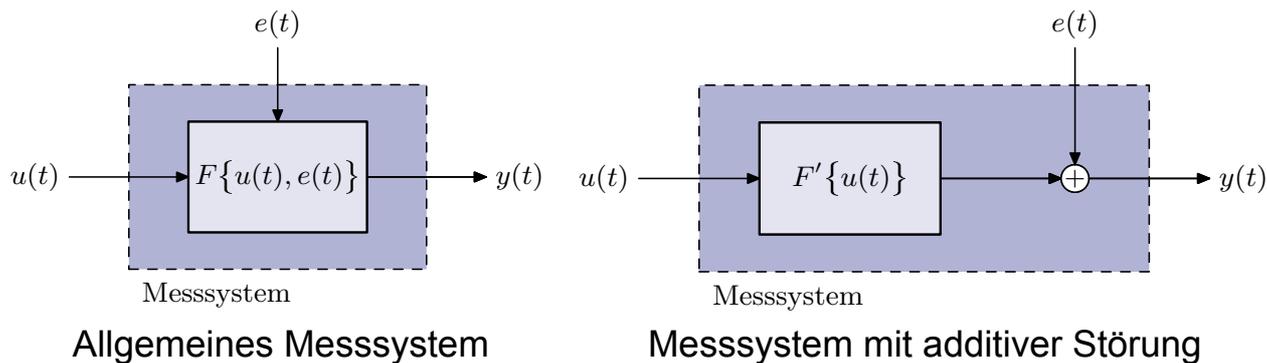
7.2 Modellbildung

- Fehlermodell
- Lineares Signalmodell durch Zerlegung in Basisfunktionen
- AR-Signalmodell
- Markov-Prozess

7.3 Beurteilungskriterien von Schätzfiltern

- Erwartungstreue
- Konsistenz
- Effizienz

Messen als Schätzproblem



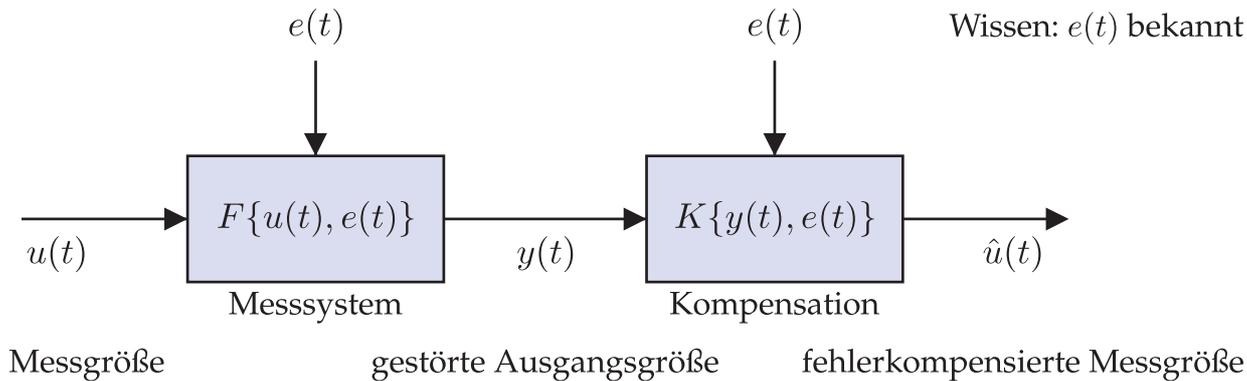
- **Ziel der Signalverarbeitung:** wahre Größe $u(t)$ aus der Messgröße $y(t)$ mit Hilfe von Schätzalgorithmen möglichst gut rekonstruieren
- **Lösungsansatz:** gesamtes verfügbares Wissen in Form eines **Modells** formulieren:
 - Wissen über erwarteten Signalverlauf $u(t)$
 - Wissen über Verfälschungen durch das Messsystem (Struktur von F)
 - Wissen über externe Fehlereinflüsse $e(t)$

7.1 Unterdrückung von Störgrößen

- Es gibt verschiedene Ansätze zur Unterdrückung von Störungen: **Kompensation**, **Filterung** und **Schätzung**
- Im Folgenden werden die Unterschiede zwischen ihnen erläutert

7.1.1 Störgrößenkompensation

- Wenn eine Störgröße $e(t)$ systematisch (deterministisch) ist, kann sie geschätzt und **kompensiert** werden



7.1.1 Störgrößen-Kompensation

Beispiel: Additiv überlagerte Sinusschwingung

- Gestörtes Signal:

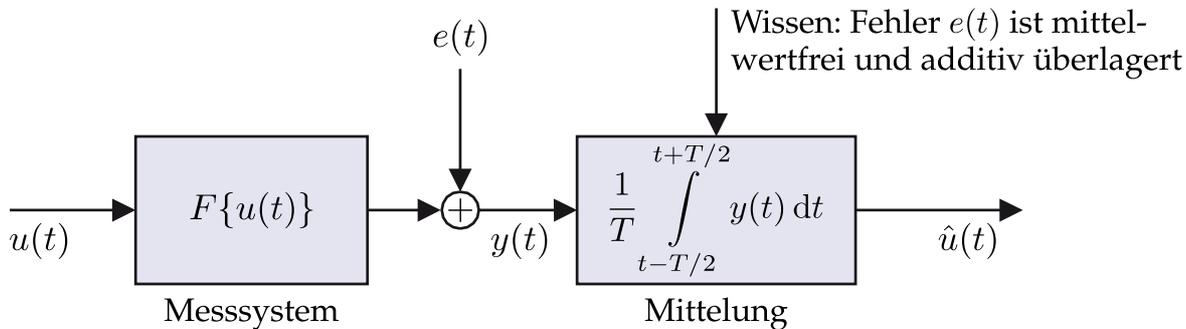
$$y(t) = u(t) + \underbrace{z_0 \cdot \sin(2\pi ft + \varphi)}_{e(t)}$$

- Die Störgröße $e(t)$ sei systematisch, d. h. durch längere Beobachtung können die Parameter z_0 , f und φ eindeutig bestimmt werden
- Kompensation durch Überlagerung einer um π phasenverschobenen Sinusschwingung gleicher Amplitude und Frequenz:

$$\hat{u}(t) = y(t) - \hat{e}(t) = u(t) + e(t) - \hat{e}(t) \approx u(t)$$

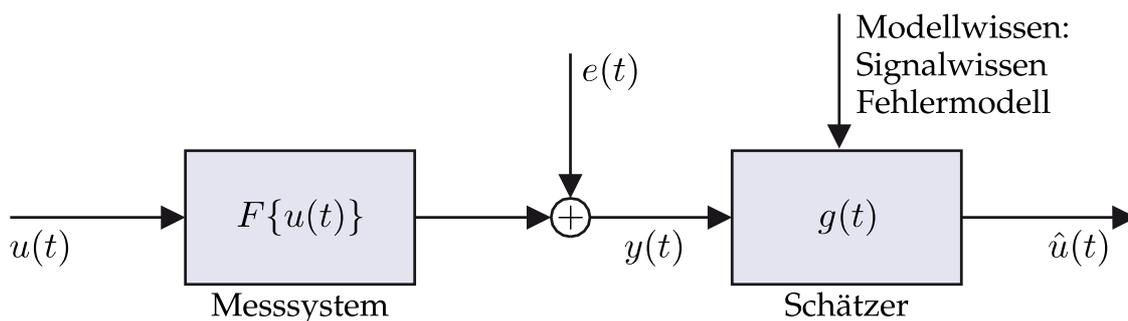
- Praktische Anwendung: Geräuschkompensation im Flugzeug in den Kopfhörern der Piloten

- Additiv überlagerte, mittelwertfreie Störgrößen können durch eine **Mittelwertbildung** unterdrückt werden
- Weniger Wissen über Störung als bei Störgrößenkompensation erforderlich, aber dafür stärkere Veränderung des Nutzsignals (TP-Filterung)
→ Mittelwertbildung vertretbar, wenn $u(t)$ im Wesentlichen TP-Signal ist



7.1.3 Schätzung

- Weniger Vorwissen über Störung verfügbar
- Schätzung des Signals basierend auf Signal- und Fehlermodell (Kap. 8)
- Zum Entwurf eines Schätzfilters mit der Impulsantwort $g(t)$ benötigt man eine ausreichend genaue Beschreibung über:
 - die Anzeigegröße $y(t)$ ohne Störung
 - die Eigenschaften des Fehlers $e(t)$
- Diese Beschreibung heißt **Modell**; das Schätzergebnis hängt entscheidend von dessen Qualität ab



7.2 Modellbildung

7.2 Modellbildung

- **Modell:** Analytische (ggf. approximative) Beschreibung des Signalverlaufs oder des dynamischen Systemverhaltens

7.2.1 Fehlermodell

- Annahme: additiver Fehler: $y(t) = x(t) + e(t)$
- Ohne Vorwissen: häufige Annahme der Störung als mittelwertfreies, weißes Rauschen

$$E\{e(t)\} = 0$$

$$r_{ee}(\tau) = E\{e(t + \tau) e(t)\} = \sigma_e^2 \cdot \delta(\tau)$$

- **Fehlermodell** beim Übergang auf zeitdiskrete Signale der Länge N :

$$\mathbf{e}(n) = [e(n) \ e(n-1) \ \dots \ e(n-N+1)]^T$$

$$\mathbf{R}_{ee} = E\{\mathbf{e}(n) \mathbf{e}^T(n)\} \stackrel{\text{Mittelwertfreiheit}}{=} \mathbf{C}_{ee} \stackrel{\text{Unkorreliertheit}}{=} \sigma_e^2 \cdot \mathbf{I}$$

- Signal $y(t)$ wird durch eine Linearkombination verschiedener Basisfunktionen beschrieben:

$$y(t) = \sum_{k=0}^{K-1} b_k \varphi_k(t) + e(t)$$

Dabei ist K gleich der Anzahl der Parameter bzw. Basisfunktionen

- Stehen N Messwerte zur Verfügung, so lassen sich N Gleichungen für das Signal zu unterschiedlichen Abtastzeitpunkten aufstellen:

$$y(t) = b_0 \cdot \varphi_0(t) + \dots + b_{K-1} \cdot \varphi_{K-1}(t) + e(t)$$

$$y(t - t_A) = b_0 \cdot \varphi_0(t - t_A) + \dots + b_{K-1} \cdot \varphi_{K-1}(t - t_A) + e(t - t_A)$$

$$\vdots$$

$$y(t - (N - 1)t_A) = b_0 \cdot \varphi_0(t - (N - 1)t_A) + \dots + b_{K-1} \cdot \varphi_{K-1}(t - (N - 1)t_A) + e(t - (N - 1)t_A)$$

- Messvektor:

$$\mathbf{y}(t) = [y(t) \ y(t - t_A) \ \dots \ y(t - (N - 1)t_A)]^T$$

- Fehlervektor:

$$\mathbf{e}(t) = [e(t) \ e(t - t_A) \ \dots \ e(t - (N - 1)t_A)]^T$$

- Beobachtungsmatrix:

$$\Phi(t) = \begin{bmatrix} \varphi_0(t) & \dots & \varphi_{K-1}(t) \\ \varphi_0(t - t_A) & \dots & \varphi_{K-1}(t - t_A) \\ \vdots & \ddots & \vdots \\ \varphi_0(t - (N - 1)t_A) & \dots & \varphi_{K-1}(t - (N - 1)t_A) \end{bmatrix}$$

- Signalmodell in Vektorschreibweise:

$$\mathbf{y}(t) = \Phi(t) \cdot \mathbf{b} + \mathbf{e}(t)$$

- Die Zeilenvektoren und Spaltenvektoren der Beobachtungsmatrix sind die sog. **Regressionsvektoren** bzw. **Basisvektoren**

Beispiel: Lineares Signalmodell

- Messwerte im Beobachtungszeitraum werden als Potenzen der Zeit angenommen:

$$y(t) = b_0 + b_1 \cdot t + b_2 \cdot t^2 + e(t)$$

- Die Basisfunktionen sind die k -ten Potenzen der Zeit t :

$$\varphi_k(t) = t^k$$

- Dann beschreibt der Parameter b_0 den stationären Anteil und die Parameter b_1 sowie b_2 die lineare bzw. quadratische Änderung des Signals mit der Zeit

Zeitdiskretisierung

- Messvektor:

$$\mathbf{y}(n) = [y(n) \ y(n-1) \ \dots \ y(n-N+1)]^T$$

- Fehlervektor:

$$\mathbf{e}(n) = [e(n) \ e(n-1) \ \dots \ e(n-N+1)]^T$$

- Beobachtungsmatrix:

$$\Phi(n) = \begin{bmatrix} \varphi_0(n) & \dots & \varphi_{K-1}(n) \\ \vdots & \ddots & \vdots \\ \varphi_0(n-N+1) & \dots & \varphi_{K-1}(n-N+1) \end{bmatrix}$$

- Basisvektoren:

$$\varphi_k(n) = [\varphi_k(n) \ \dots \ \varphi_k(n-N+1)]^T$$

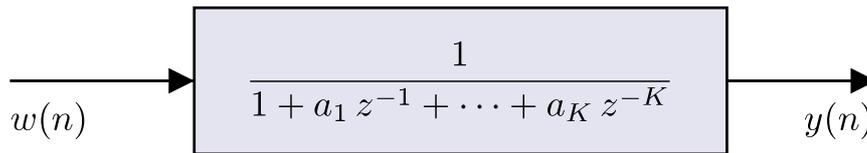
- Zeitdiskretes Signalmodell: $\mathbf{y}(n) = \Phi(n) \cdot \mathbf{b} + \mathbf{e}(n)$

$$= \sum_{k=0}^{K-1} b_k \varphi_k(n) + \mathbf{e}(n)$$

- **Stochastisches Signalmodell** für schwach stationäre Prozesse $y(n)$
- Signal $y(n)$ ist der Ausgang eines **autoregressiven** Systems $G(z)$, das mit weißem Rauschen $w(n)$ angeregt wird:

$$y(n) = w(n) - \sum_{k=1}^K a_k y(n-k)$$

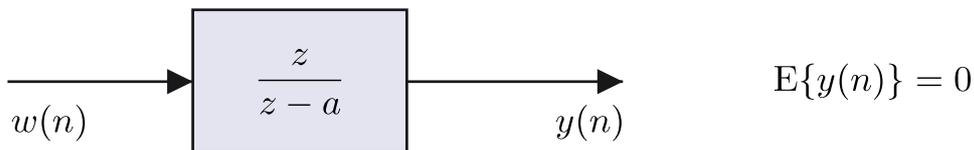
- Das AR-System hat eine gebrochen rationale Übertragungsfunktion $G(z)$, die nur ein Nennerpolynom $A(z)$ enthält:



- Die Koeffizienten a_k des AR-Signalmodells sind so zu bestimmen, dass der Systemausgang $y(n)$ einen gegebenen Signalverlauf („Mustersignal“) möglichst gut annähert
- **Vorteil:** Ansatz führt zu linearem Schätzer (siehe Abschn. 8.5)

7.2.4 Markov-Prozess

- **Markov-Prozess:** stochastischer Prozess $y(n)$, der am Ausgang eines LTI-Systems 1. Ordnung resultiert, das mit weißem Rauschen $w(n)$ angeregt wird $\rightarrow y(n)$ ist ein **farbiger** (korrelierter) Rauschprozess



- Zustand $y(n)$ hängt nur vom Vorzustand $y(n-1)$ und vom Eingang $w(n)$ ab:

$$y(n) = a \cdot y(n-1) + w(n) \quad \Leftrightarrow \quad Y(z) = a \cdot z^{-1} Y(z) + W(z)$$

$$\Rightarrow \quad Y(z) = \frac{z}{z-a} W(z) = G(z) W(z)$$

- Durch rekursives Einsetzen von $y(n)$ erhält man mit $0 < a < 1$ (Stabilität!):

$$y(n) = \sum_{i=0}^{\infty} a^i w(n-i) = g(n) * w(n) \quad \text{mit} \quad g(n) = a^n \cdot \sigma(n)$$

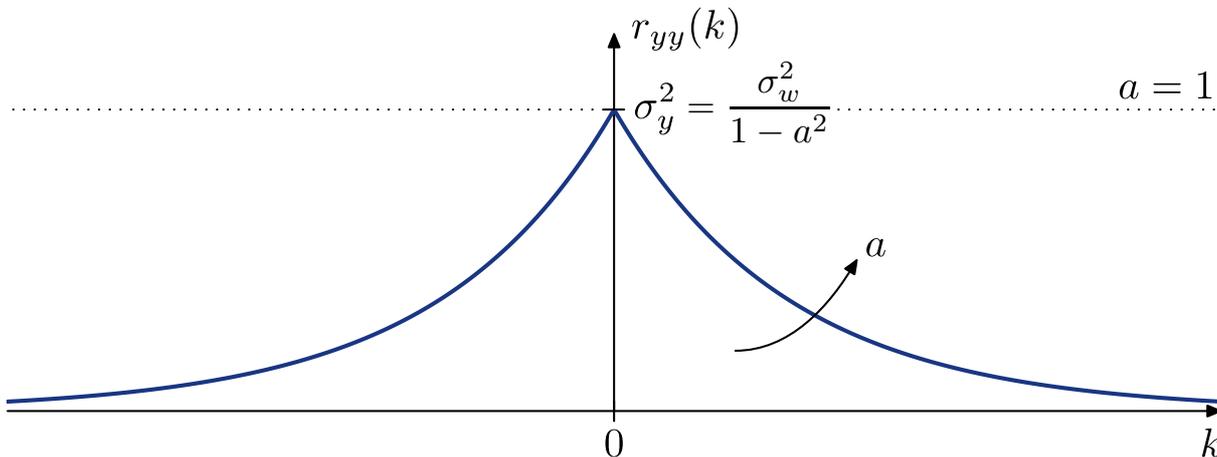
- Leistungsdichte und AKF des weißen Eingangsprozesses $w(n)$:

$$S_{ww}(z) = \sigma_w^2 \quad \bullet \text{---} \circ \quad r_{ww}(k) = E\{w(n+k)w(n)\} = \sigma_w^2 \cdot \delta(k)$$

- Leistungsdichte und AKF des Ausgangsprozesses $y(n)$:

$$S_{yy}(z) = S_{ww}(z) |G(z)|^2 = \sigma_w^2 S_{gg}^E(z) \quad \bullet \text{---} \circ \quad r_{yy}(k) = \sigma_w^2 r_{gg}^E(k) = \frac{\sigma_w^2}{1-a^2} a^{|k|}$$

- Einstellung der Charakteristik des farbigen Rauschens über a



- Für mittelwertfreie Prozesse $\mathbf{y}(n) = [y(n) \dots y(n-N+1)]^T$ gilt:

$$\begin{aligned} \mathbf{R}_{yy} &= \mathbf{C}_{yy} \\ &= \begin{bmatrix} r_{yy}(0) & r_{yy}(1) & \dots & r_{yy}(N-1) \\ r_{yy}(-1) & r_{yy}(0) & \dots & r_{yy}(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ r_{yy}(1-N) & r_{yy}(2-N) & \dots & r_{yy}(0) \end{bmatrix} = \frac{\sigma_w^2}{1-a^2} \begin{bmatrix} 1 & a & \dots & a^{N-1} \\ a & 1 & \dots & a^{N-2} \\ \vdots & \ddots & \ddots & \vdots \\ a^{N-1} & \dots & \dots & 1 \end{bmatrix} \end{aligned}$$

- Die **inverse Autokovarianzmatrix** kann geschlossen berechnet werden:

$$\mathbf{C}_{yy}^{-1} = \frac{1}{\sigma_w^2} \begin{bmatrix} 1 & -a & & & \mathbf{0} \\ -a & 1+a^2 & \ddots & & \\ & \ddots & \ddots & -a & \\ & & -a & 1+a^2 & -a \\ \mathbf{0} & & & -a & 1 \end{bmatrix}$$

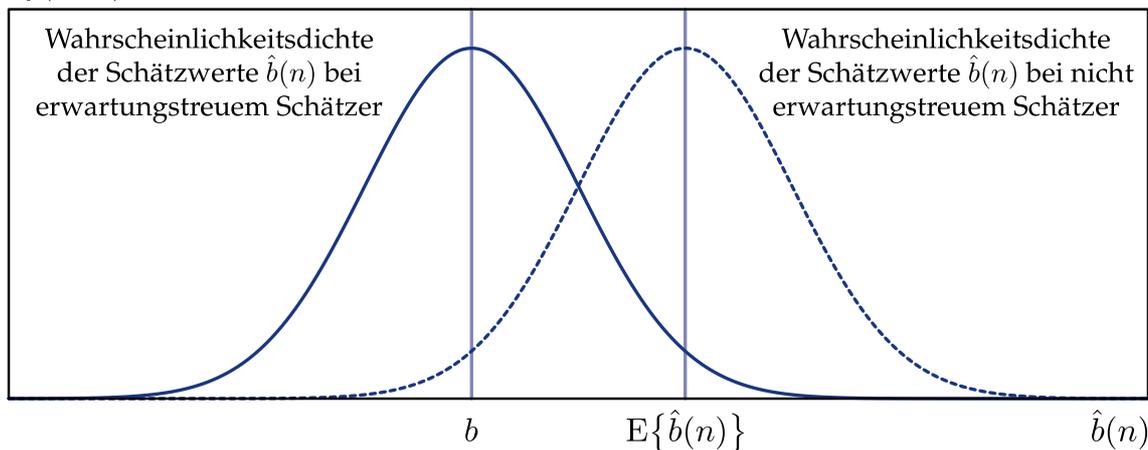
Dies ist ein Grund für die häufige Verwendung von Markov-Prozessen zur Modellierung farbiger Rauschprozesse (vgl. Abschn. 6.1.3)

7.3 Beurteilungskriterien von Schätzfiltern

7.3.1 Erwartungstreue

- **Aufgabenstellung:** Schätzung eines Parametervektors (wahrer Wert: \mathbf{b})
- Schätzwert des Parametervektors $\hat{\mathbf{b}}(n)$ wird als Zufallsvektor interpretiert, der bei jeder neuen Schätzung zu einem anderen Ergebnis führt

■ **Erwartungstreue:** $E\{\hat{\mathbf{b}}(n)\} = \mathbf{b}$

 $f_{\hat{\mathbf{b}}}(b(n))$


Erwartungstreue bei einer skalaren Schätzgröße b

- Der **systematische Schätzfehler** (engl.: *bias*) berechnet sich wie folgt:

$$\tilde{\mathbf{b}} = \mathbb{E}\{\hat{\mathbf{b}}(n)\} - \mathbf{b}$$

- Hängt der *Bias* vom wahren Wert \mathbf{b} ab, so kann dieser **nicht** korrigiert werden

Asymptotische Erwartungstreue

- Für **asymptotisch** erwartungstreue Schätzer gilt (N : Stichprobenumfang)

$$\lim_{N \rightarrow \infty} \mathbb{E}\{\hat{\mathbf{b}}(n)\} = \mathbf{b}$$

- Jeder erwartungstreue Schätzer ist auch asymptotisch erwartungstreu, aber nicht umgekehrt
- Erwartungstreue ist zwar eine wünschenswerte Eigenschaft; allerdings kann es nicht erwartungstreue Schätzer geben, die bessere Konvergenzeigenschaften besitzen als jeder erwartungstreue Schätzer:

Bei solchen Schätzern ist der mittlere quadratische Schätzfehler bei gegebenem Stichprobenumfang N oftmals kleiner

Konsistenz

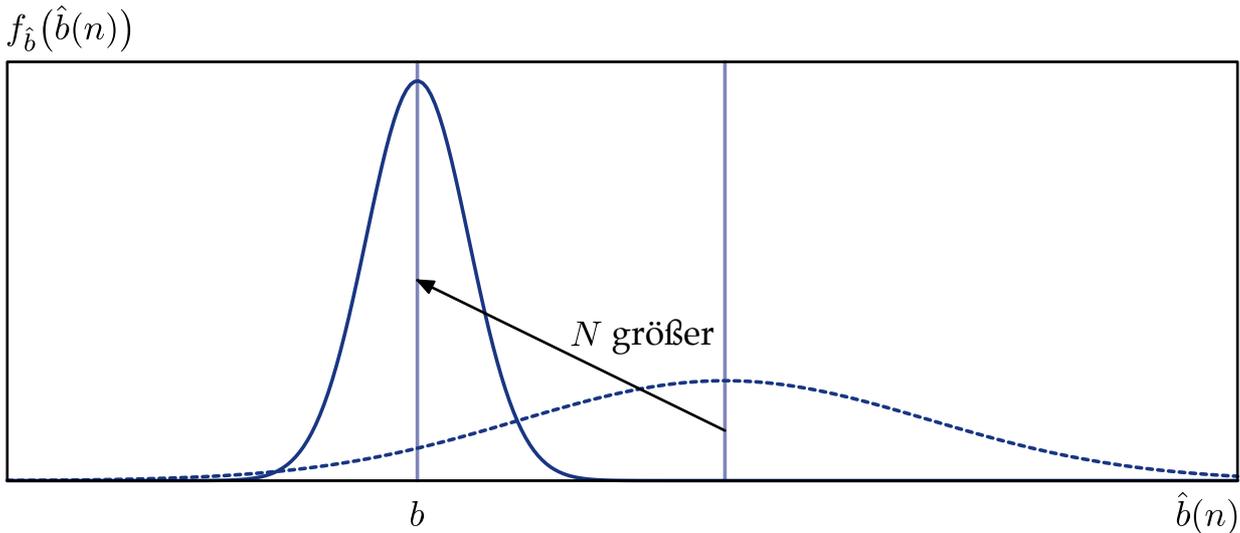
- Bei einem **konsistenten** Schätzer strebt die Schätzfehlerkovarianz mit wachsendem Stichprobenumfang $N \rightarrow \infty$ gegen null:

$$\lim_{N \rightarrow \infty} \mathbf{C}_{\tilde{\mathbf{b}}\tilde{\mathbf{b}}}(n) = \mathbf{0} \quad \text{mit} \quad \tilde{\mathbf{b}} = \hat{\mathbf{b}}(n) - \mathbf{b}$$

- Für $N \rightarrow \infty$ nähert sich der Schätzwert $\hat{\mathbf{b}}(n)$ einer einzelnen Schätzung dem wahren Wert \mathbf{b} an:

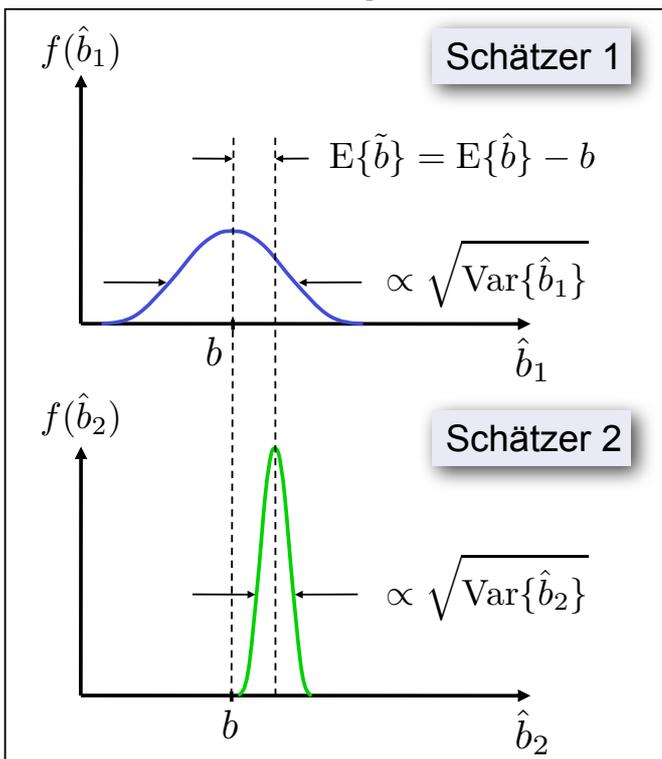
$$\lim_{N \rightarrow \infty} \mathbb{E}\{\hat{\mathbf{b}}(n)\} = \mathbf{b}$$

- Daraus folgt, dass bei Konsistenz eine Erhöhung des Stichprobenumfangs N die Qualität der Schätzung verbessert
- Jeder konsistente Schätzer ist zumindest **asymptotisch erwartungstreu**
- Es gibt Schätzer, die erwartungstreu, aber nicht konsistent sind
- Ist eine größere Anzahl N an Messwerten verfügbar, so ist ein konsistenter Schätzer vorzuziehen, sofern ein solcher Schätzer existiert



Konsistenz bei einem asymptotisch erwartungstreuen Schätzer:
 Eine Erhöhung der Anzahl der Messwerte N
 für eine Schätzung verbessert das Ergebnis der Schätzung

Diskussion: Erwartungstreue, Varianz, mittlerer quadratischer Fehler



b sei wahrer Wert

Schätzer \hat{b}_1 : erwartungstreu

Schätzer \hat{b}_2 : nicht erwartungstreu, aber:

$$\sqrt{\text{Var}\{\hat{b}_2\}} < \sqrt{\text{Var}\{\hat{b}_1\}}$$

$\text{Var}\{.\}$: **stochastischer Fehler**

Bias $E\{\tilde{b}\}$: **systematischer Fehler**

$$E\{(\hat{b} - b)^2\} = E^2\{\tilde{b}\} + \text{Var}\{\hat{b}\} : \text{mittlerer quadratischer Fehler}$$

Welcher Schätzer ist besser?

Effizienz (Wirksamkeit)

- Die Schätzfehlerkovarianz ist bei einem **effizienten** Schätzer für ein vorgegebenes Prozessmodell und einen vorgegebenen Stichprobenumfang N minimal, d. h. es gibt kein anderes Schätzfilter mit einer kleineren Schätzfehlerkovarianz unter diesen Bedingungen:

$$C_{\tilde{\mathbf{b}}\tilde{\mathbf{b}}}(n) = E\left\{(\hat{\mathbf{b}}(n) - \mathbf{b})(\hat{\mathbf{b}}(n) - \mathbf{b})^T\right\} \xrightarrow{!} \min$$

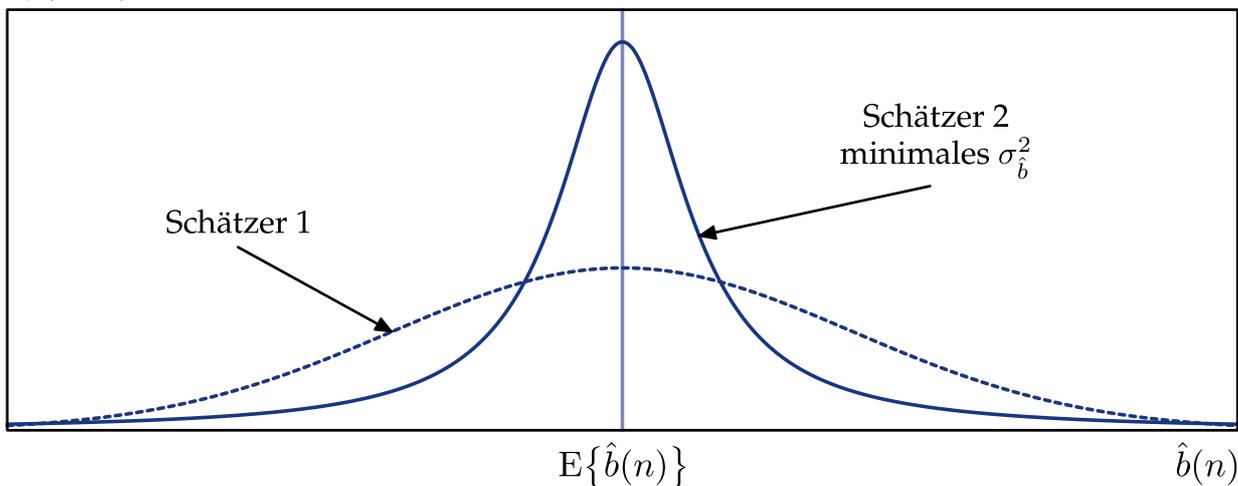
Dabei ist $\tilde{\mathbf{b}} = \hat{\mathbf{b}}(n) - \mathbf{b}$ der Schätzfehler.

Anmerkungen

- Bei **erwartungstreuen** Schätzern gibt es eine theoretische Untergrenze für die Schätzfehlerkovarianz (**Cramér-Rao-Grenze**) – Schätzer, die diese Grenze erreichen, sind **effizient**
- Zu einem gegebenen Problem existiert jedoch nicht zwangsläufig ein Schätzer, der diese Grenze erreicht

7.3.3 Effizienz

$f_{\hat{\mathbf{b}}}(\hat{\mathbf{b}}(n))$



Ein effizienter Schätzer hat eine minimale Schätzfehlerkovarianz