Vortrag im Rahmen des Astrophysikalischen Seminars:

Bayessche Datenanalyse in der Astronomie

Milena Oehlers

06. Juli 2021

Bayessche Datenanalyse in der Astronomie

- Einführung & Motivation am linearen Beispiel
- Grundlagen der Bayesschen Theorie
- Wahl des Priors
- Markov Chain Monte Carlo (MCMC)
    - Mathematische Grundlagen
    - Numerische Methoden: (Adaptive) Metropolis Hastings
    - Konvergenzbestimmung
- Anwendung: Lineares Beispiel; Binäres System aus Stern und Exoplanet
- Zusammenfassung

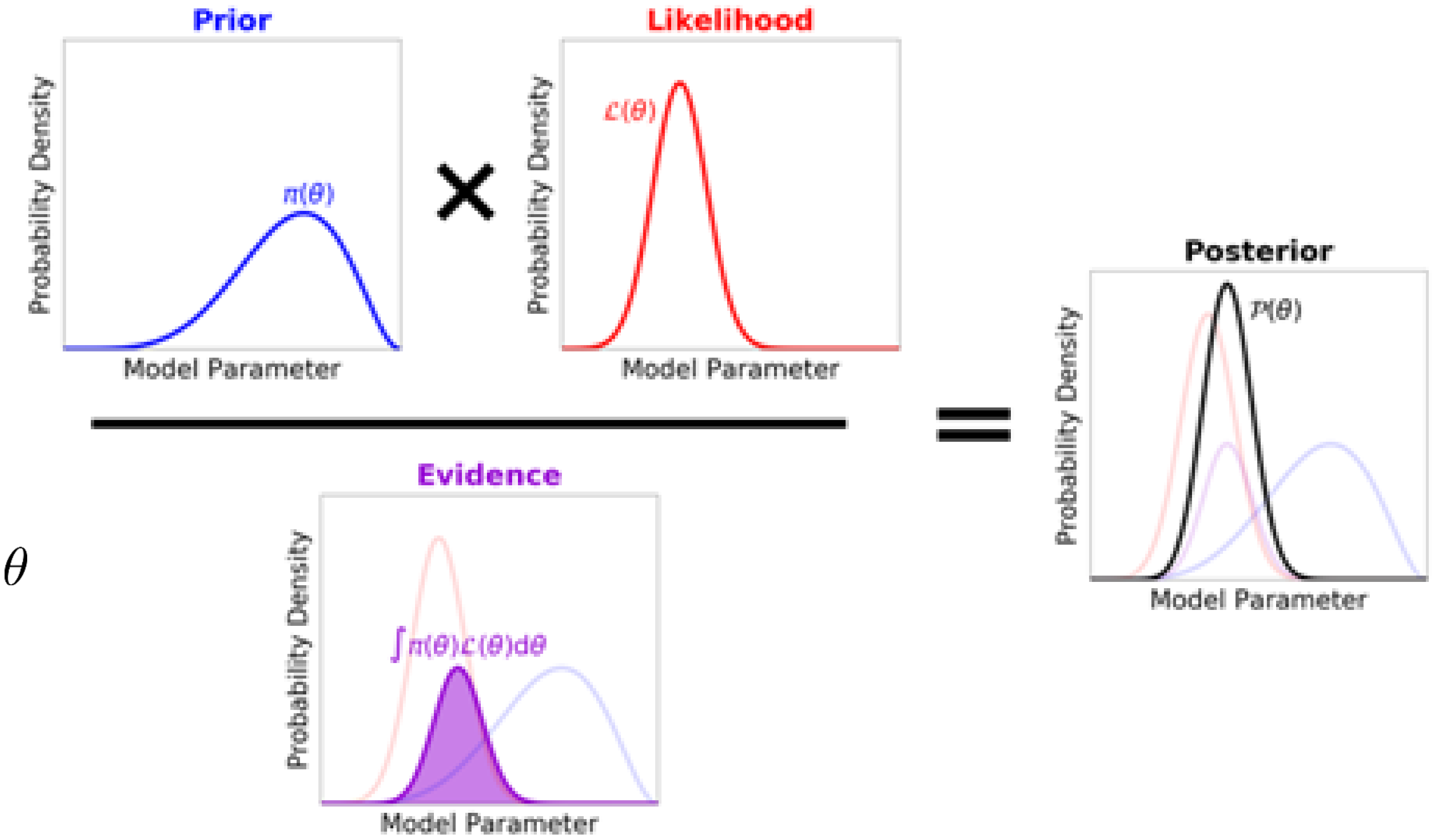
Idee der Bayesschen Theorie

Bekannt:

- Daten
- Modell
- Zusatzinfo

Gesucht:

Modellparameter θ



Motivation am linearen Beispiel

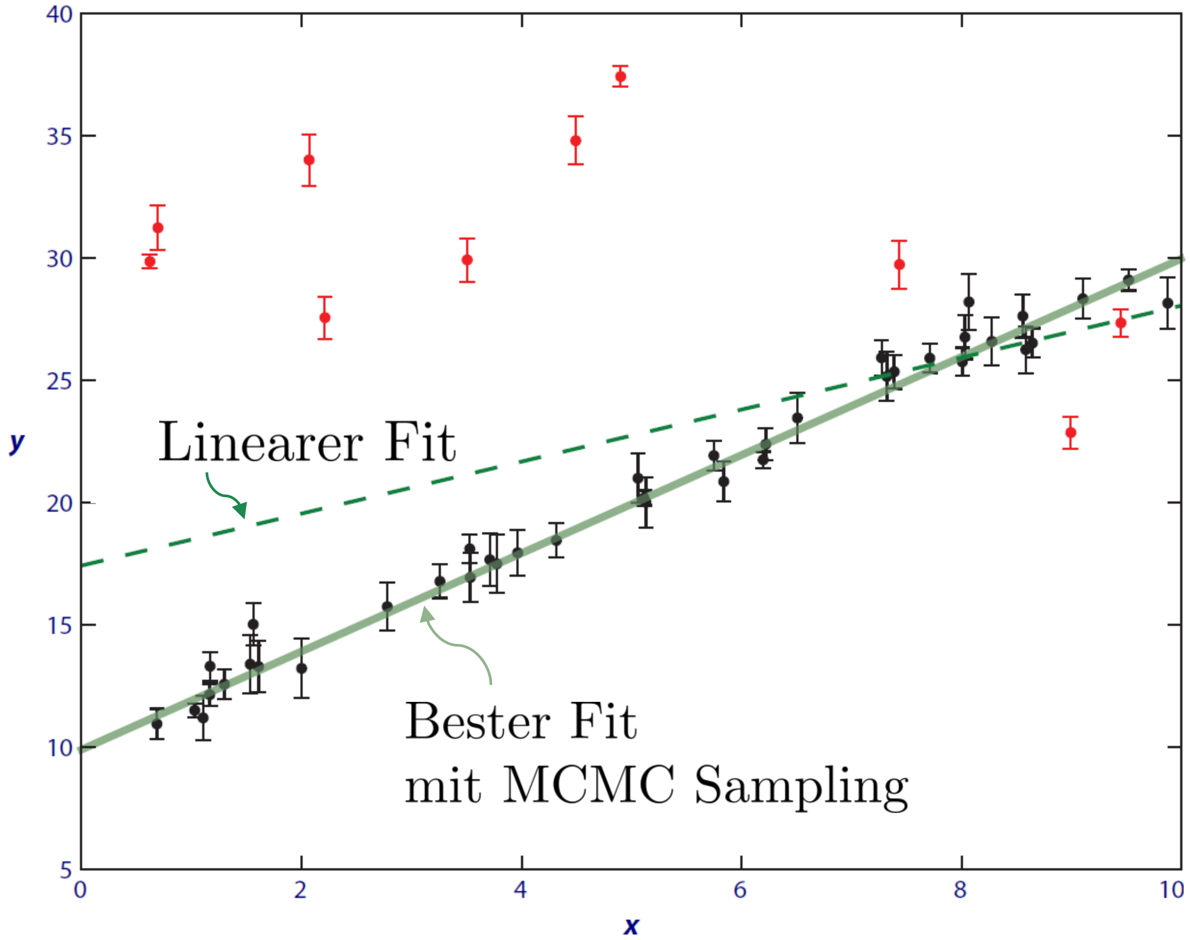
Bekannt:

- Datenpunkte $z_j = (x, y)_j$
- Messunsicherheit σ_z

Modellannahmen:

- Real: $y_j = m \cdot x_j + c$
- Ausreißer: $y_j \sim \mathcal{N}(\mu_a, \sigma_a)$

Ziel: Parameter m und c
ohne Ausreißer schätzen



Beispiel und Abbildung adaptiert aus [1]

Motivation am linearen Beispiel (adaptiert aus [1])

Modellannahmen $\Rightarrow$ Wsk., Messpunkte $Z = \{z_1, ..., z_N\}$ zu erhalten
für gegebene Modellparameter $\theta = (m, c, \mu_a, \sigma_a, P_a)$:

$$p(Z|\theta, \sigma_z) = \prod_{j=1}^{N} [p(z_j|\mu_a, \sigma_a, \sigma_z)P_a + p(z_j|m, c, \sigma_z)(1 - P_a)]$$

mit $$p(z_j|m, c, \sigma_z) = \frac{1}{\sqrt{2\pi}\sigma_z} \exp\left[-\frac{(y_j - mx_j - c)^2}{2\sigma_z^2}\right]$$

und $$p(z_j|\mu_a, \sigma_a, \sigma_z) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y_j - \mu_a)^2}{2\sigma^2}\right] \quad \text{mit} \quad \sigma^2 = \sigma_z^2 + \sigma_a^2$$

<u>aber gesucht:</u> $p(\theta|Z, \sigma_z)$, denn $\mathbb{E}(\theta) = \int \theta \cdot p(\theta|Z, \sigma_z) d\theta$
und $\sigma_\theta = \int (\theta - \mathbb{E}(\theta))^2 \cdot p(\theta|Z, \sigma_z) d\theta$

Grundlagen der Bayesschen Theorie

	θ_1	θ_2	...	Gesamt
Z_1	$p(Z_1, \theta_1)$	$p(Z_1, \theta_2)$	...	$p(Z_1)$
Z_2	$p(Z_2, \theta_1)$	$p(Z_2, \theta_2)$	...	$p(Z_2)$
...	...	...	...	...
Gesamt	$p(\theta_1)$	$p(\theta_2)$	...	1.0

bekannte Information
hier: Messunsicherheit σ_z

Produktregel: $p(Z, \theta | I) = p(\theta | Z, I) p(Z | I) = p(Z | \theta, I) p(\theta | I)$

$\Rightarrow$ Satz von Bayes: $p(\theta | Z, I) \propto p(Z | \theta, I) \cdot p(\theta | I)$

Posterior $\propto$ Likelihood $\cdot$ Prior

gesucht Modellannahmen zu wählen

Wahl des Priors $p(\theta|I)$ (adaptiert aus [1])

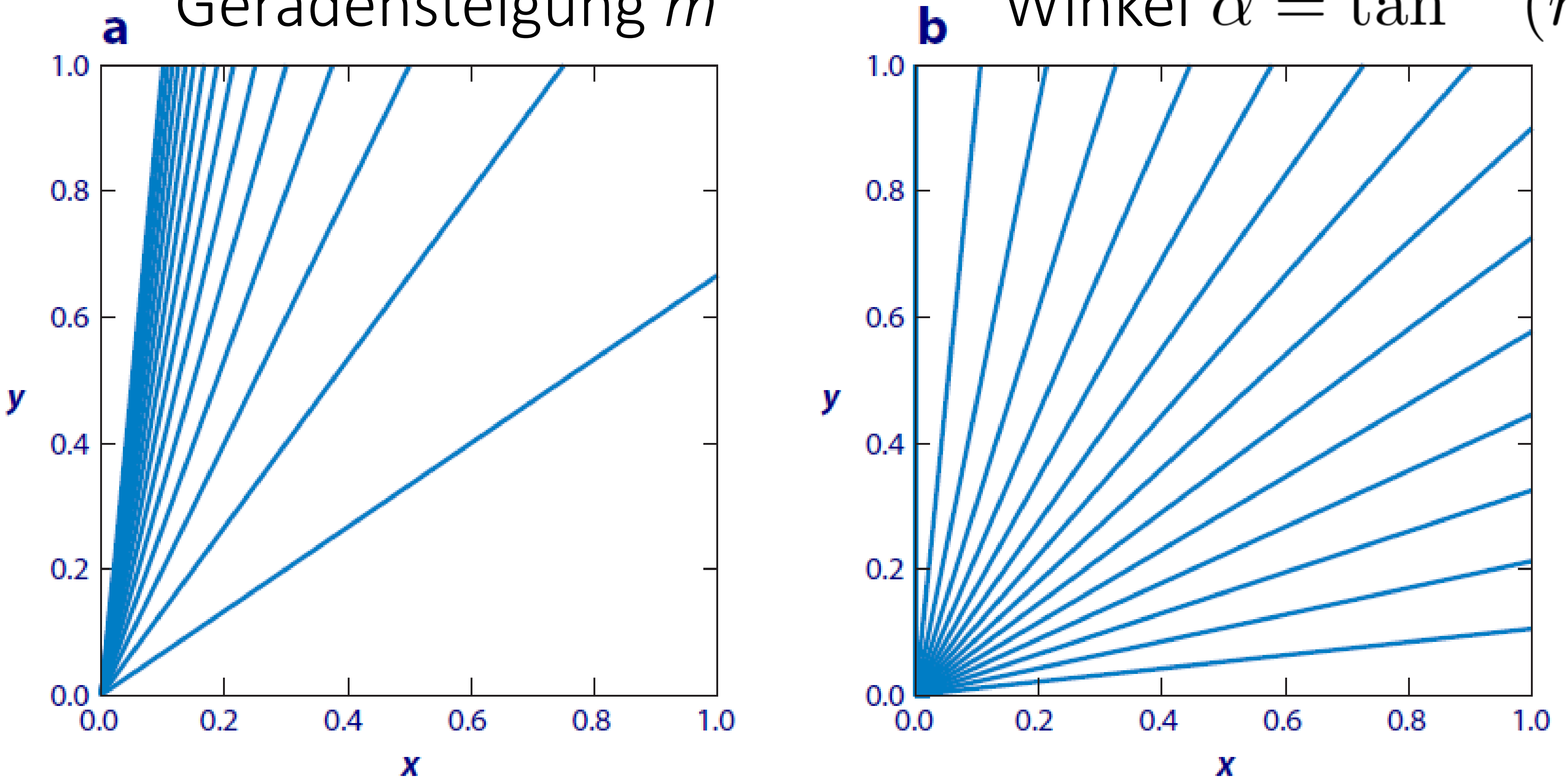
- *Informativer Prior:* kann durch Info aus vergangenen Analysen sehr restriktiv gewählt werden, wichtig bei nicht aussagekräftigen Daten
- *Ignorance Prior:* bei aussagekräftigen Daten reicht glatter Prior in Bereichen hoher Parameter-Likelihood für gute Ergebnisse, z.B. durch
  - Gleichverteilung in möglichem Parameterraum
  - Invarianz der Likelihood $p(x'|\theta')dx' = p(x|\theta)dx$ bezüglich Transformations-gruppen $(\theta', x') = h(\theta, x)$ wie Translation, Skalierung, Rotation…
  - Maximierung der Entropie $S = -\sum_i p_i \log p_i$ von Priors, die Einschränkungen erfüllen

Motivation am linearen Beispiel (adaptiert aus [1])

Nutze Rotationsinvarianz, um Verzerrungen des Priors zu vermeiden:

Gleichverteilt in Geradensteigung m

Gleichverteilt in Winkel $\alpha = \tan^{-1}(m)$



Markov Chain Monte Carlo (MCMC) [1]

<u>Ziel:</u> $\mathbb{E}(\theta) = \int \boxed{\theta} \cdot p(\theta|Z,\sigma_z)d\theta$ und $\sigma_\theta = \int \boxed{(\theta - \mathbb{E}(\theta))^2} \cdot p(\theta|Z,\sigma_z)d\theta$

Integrale meist nicht analytisch lösbar. MCMC zieht stattdessen iterativ Stichproben θ_i mithilfe einer Markovkette, deren Gleichgewichtsverteilung $p^*(\theta) = p(\theta|Z,\sigma_z)$ ist

Für Markovkette mit Übergangswsk. $K(\theta_i,\theta_{i+1}) = p(\theta_{i+1}|\theta_i)$
$\exists$ stationäre Wsk.verteilung $p^* \Leftarrow \boxed{p^*(\theta_i)K(\theta_i,\theta_{i+1}) = p^*(\theta_{i+1})K(\theta_{i+1},\theta_i)}$

Detailed Balance Bedingung

$$\left.\begin{array}{r}\text{stationär}\\ \text{\& irreduzibel}\end{array}\right\} \Rightarrow \left.\begin{array}{r}\left\{\begin{array}{c}p^* \text{ ist eindeutig}\\ \text{\& positiv rekurrent}\end{array}\right\}\\ \\ \text{aperiodisch}\end{array}\right] \Rightarrow \lim_{i\to\infty} ||pK^i - p^*|| = 0$$

Monte Carlo Integration: $\mathbb{E}_{p^*}[g(\theta)] = \int \boxed{g(\theta)} p^*(\theta)d\theta = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n g(\theta_i)$

MCMC-Methoden: Metropolis-Hastings (MH) [1]

Setze Übergangswahrscheinlichkeit $K(\theta_i, \theta_{i+1}) = q(\theta_{i+1}|\theta_i)\alpha(\theta_i, \theta_{i+1})$ mit Akzeptanzwahrscheinlichkeit

$$\alpha(\theta_i, \theta_{i+1}) = \min\,[1, a_i] \quad \text{mit} \quad a_i = \frac{p^*(\theta_{i+1})q(\theta_i|\theta_{i+1})}{p^*(\theta_i)q(\theta_{i+1}|\theta_i)}$$

um *Detailed Balance* zu erfüllen.

Setze Vorschlagsdichte $q(\theta_{i+1}|\theta_i)$

- *Symmetrisch* $= q(\theta_i|\theta_{i+1}) \Rightarrow a_i = p^*(\theta_{i+1})/p^*(\theta_i)$
- *Random Walk* $= q(\theta_{i+1} - \theta_i)$, z.B. $\mathcal{N}(\theta_{i+1}; \theta_i, \Sigma)$
- …

$$\boxed{\Sigma^* \overset{[2]}{=} (2.38^2/D)\Sigma_{p^*}}$$

nicht Bekannt!

MCMC-Methoden: Adaptive (RW) Metropolis

<u>*Gewählt:*</u> Vorschlagsdichte $q(\theta_{i+1}|\theta_i) = \mathcal{N}(\theta_{i+1}; \theta_i, s_i\Sigma_i)$
Nach jedem Schritt i werden zusätzlich die Vorschlagsdichte-parameter $\rho_i = (s_i, \mu_i, \Sigma_i)$ aktualisiert, sodass $\rho_i \overset{i\to\infty}{\longrightarrow} \rho^*$, für welche die Markovkette am schnellsten konvergiert:
<u>*Setze:*</u>

$$\log s_{i+1} - \log s_i = \gamma_{i+1}(\alpha_i - \alpha^*) \text{ mit } \alpha^* = \alpha(D) \text{ bekannt}$$
$$\text{und } \Sigma_{i+1} - \Sigma_i = \gamma_{i+1}[(\theta_{i+1} - \mu_i)(\theta_{i+1} - \mu_i)^T - \Sigma_i]$$
$$\text{mit } \mu_{i+1} - \mu_i = \gamma_{i+1}(\theta_{i+1} - \mu_i)$$

<u>*Wobei:*</u> $\sum_{i=1}^{\infty} \gamma_i = \infty$ und $\sum_{i=1}^{\infty} \gamma_i^2 < \infty$

z.B.: $\gamma_i = \gamma / i^\beta$
mit $\beta \in]0, 1[$

MCMC Konvergenzbestimmung [1]

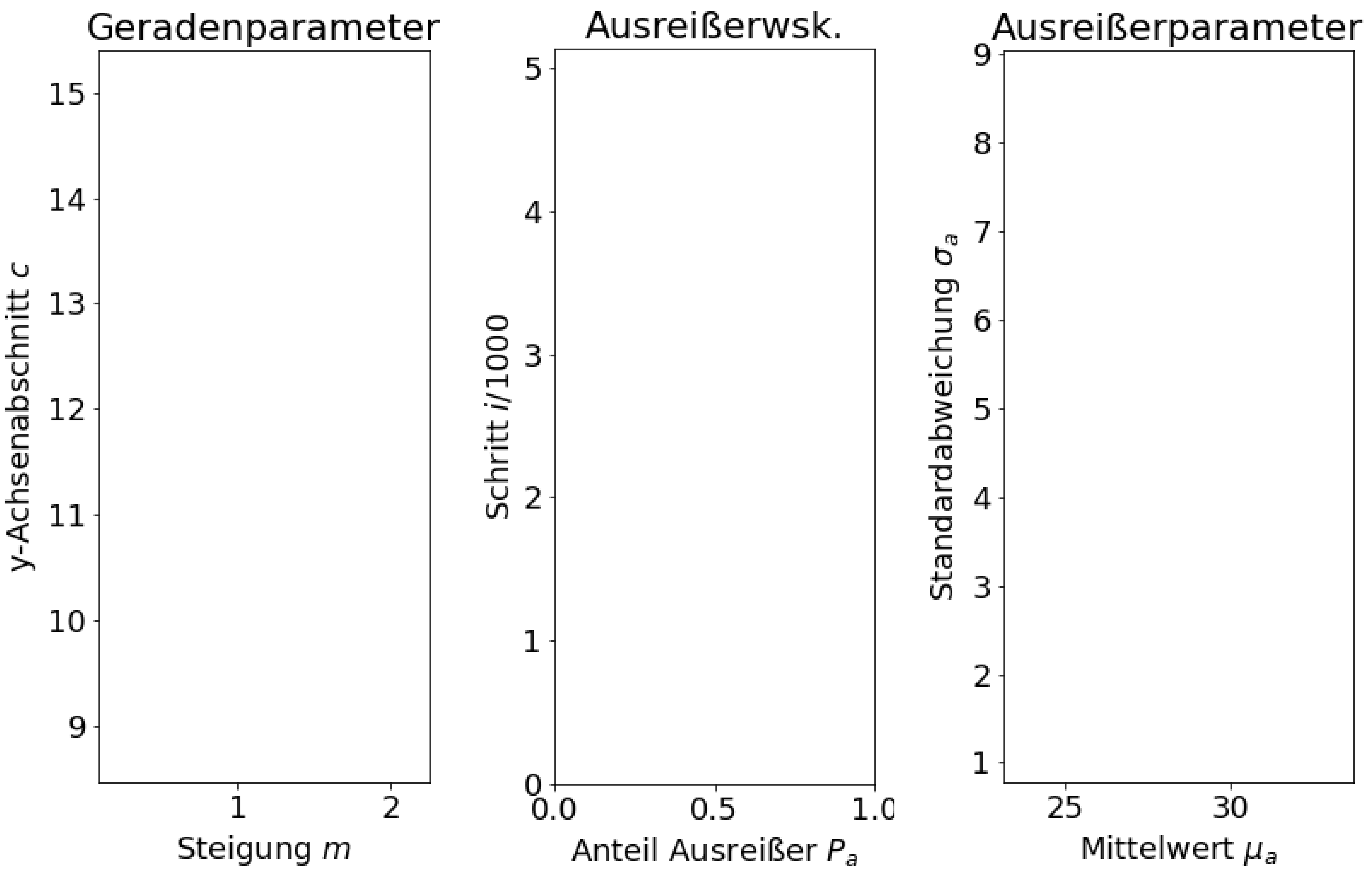
<u>Zwischenkettenvarianz:</u> Berechne K Markovketten $k = 1, \ldots, K$ geichzeitig, welche $p^*(\theta)$ aus den Ziehungen der letzten S Schritte ableiten.

- Jede Kette hat eine innere Varianz $I_k = \frac{1}{S-1} \sum_{s=1}^{S} (\theta_{s,k} - \overline{\theta}_k)^2$,
  $\Rightarrow$ durchschnittliche Innerkettenvarianz $I = \frac{1}{K} \sum_{k=1}^{K} I_k$
- Zwischenkettenvarianz $Z = \frac{1}{1-K} \sum_{k=1}^{K} (\overline{\theta}_k - \overline{\theta})^2$
- gepoolte Varianz $V = \frac{K+1}{K} Z + \frac{S-1}{S} I$

Wenn $Z \to 0$ geht Konvergenzratio $R = \sqrt{V/I} \to 1$

Motivation am linearen Beispiel

Adaptive RW Metropolis mit Konvergenzkriterium Zwischenkettenvarianz



Kette $k \in [1, ..., 5]$,

Akzeptiere die letzten 1000 Modellparameterwerte $\theta = (m, c, \mu_a, \sigma_a, P_a)$

$[\theta_{i-1000}, \theta_i]_k$

$KI_{0.95}[q_k(\theta_{i+1}|\theta_i)]$,

$\overline{\theta}_k$ und $KI_{0.95}(\theta_k)$

$\overline{\theta}$ und $KI_{0.95}(\overline{\theta}_k)$

KI: Konfidenzintervall

Anwendungsbeispiel: [1]

Binäres System aus Stern und Exoplanet

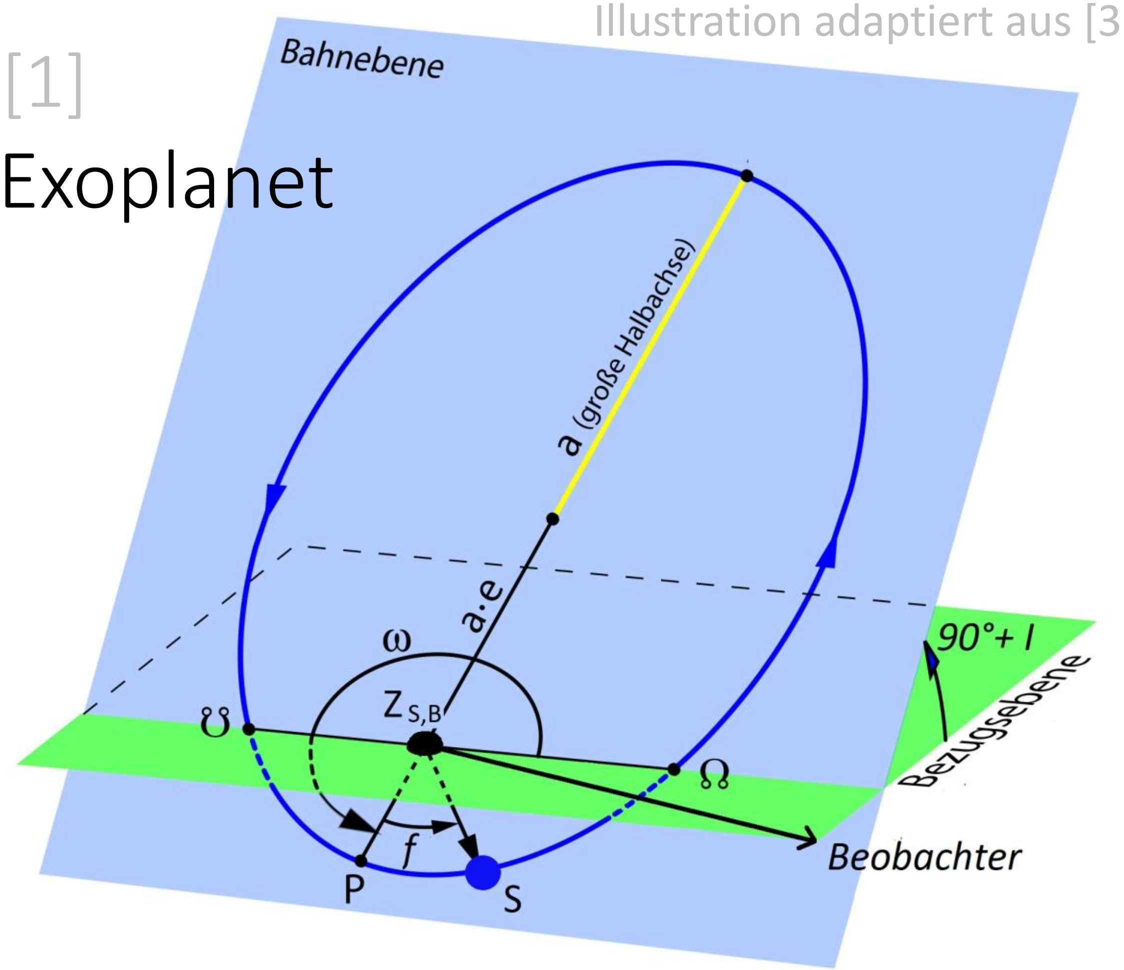
$$v(t) = \kappa[\cos(f+\omega) + e\cos(\omega)] + \overline{v_Z}$$

$$\text{mit} \quad \kappa = \frac{(2\pi G)^{1/3} m \sin(I)}{T^{1/3}(M+m)^{2/3}\sqrt{1-e^2}}$$

$$\tan(f/2) = \sqrt{\frac{1+e}{1-e}}\tan(u/2)$$

$$u - e\sin(u) = \frac{2\pi}{T}(t-\tau)$$

Illustration adaptiert aus [3]



$v(t)$	Radialgeschw. des Sterns S	ω	Winkel von Perizentrum P zu aufsteigendem Knoten Ω
M,m	Masse des Sterns/ Exoplaneten	τ	Durchgangszeit durch P
I, e	Inklination/ Exzentrizität	T	Umlaufzeit
f	wahre Anomalie		

Anwendungsbeispiel: [1]
Binäres System aus Stern und Exoplanet

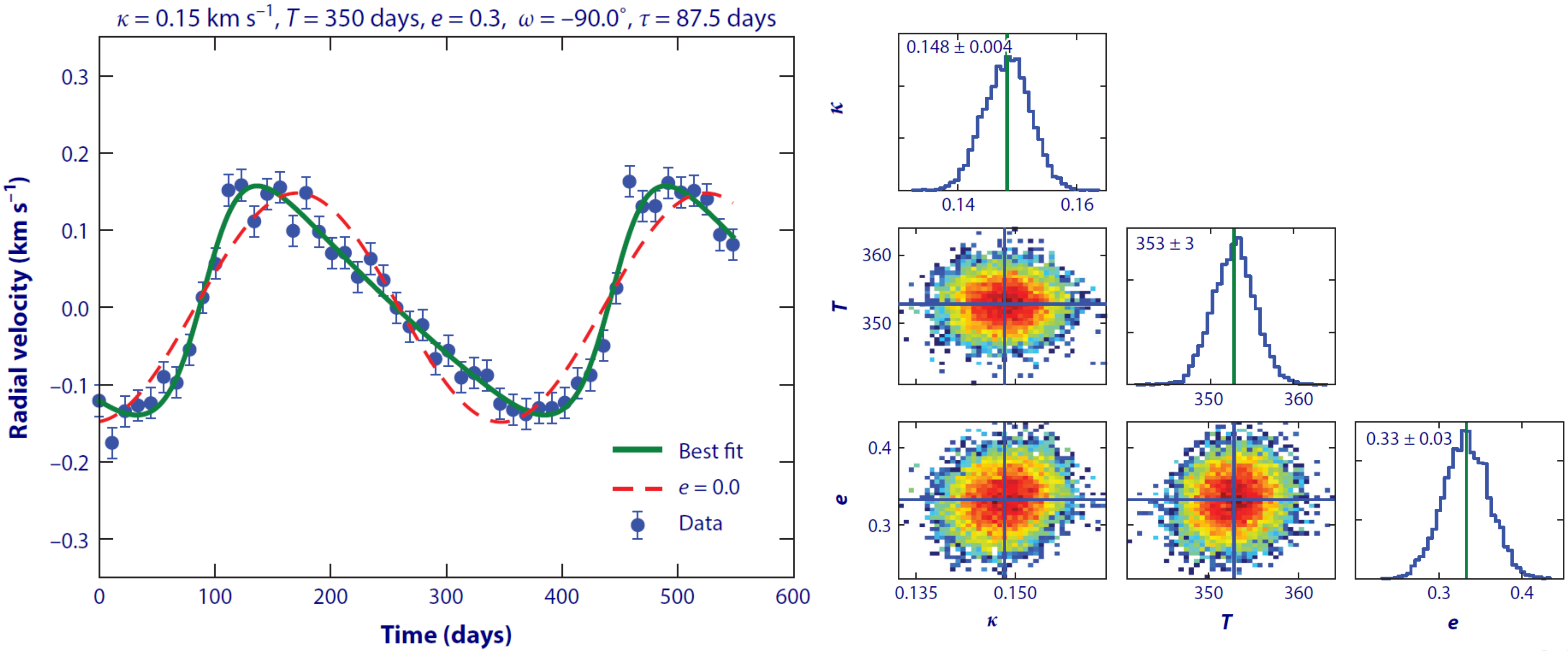
Gesucht: Modellparameter $\theta = (\overline{v_Z}, \kappa, T, e, \tau, \omega, S)$ mit Spektralvarianz S

Bekannt:

- Datenpunkte $D = [d_i] = [(t, v)_i]$ mit $i = 1, ..., M$
- Messunsicherheit σ_v

Aus Modell folgt *Posterior:* $p(\theta|D) \propto p(\theta) \prod_{i=1}^{M} \mathcal{N}(v_i | v(t_i;\theta), \sigma_v^2 + S^2)$

Anwendungsbeispiel: [1]
Binäres System aus Stern und Exoplanet



Illustrationen aus [1]

Zusammenfassung

- Motivation am linearen Beispiel: Wie ohne Ausreißer Datenpunkte fitten?
- Grundlagen: Vierfeldertafel, Produktregel, Satz von Bayes
- Wahl des Priors: Ignorance Prior mit geringstem Bias durch Likelihood-Invarianz
- Markov Chain Monte Carlo (MCMC): Methode zur Anwendung Bayesscher Datenanalyse
    - Mathematische Grundlagen: Markov Chain, Detailed Balance, Monte Carlo
    - Numerische Methoden: Adaptive Metropolis Hastings zur gleichzeitig iterativen Anpassung der Kovarianz der Vorschlagsdichte für schnelle Konvergenz der Markovkette
    - Konvergenzbestimmung mithilfe von Zwischenkettenvarianz $Z \to 0$ mehrerer gleichzeitig berechneter Markovketten, während durchschnittliche Innerkettenvarianz $I \to I^{\infty} > 0$
- Anwendung auf das linearen Beispiel: Illustration der fünf Markovketten, ihrer Vorschlagsdichten $q_k(\theta_{i+1}|\theta_i)$ und Innerkettenvarianzen I_k, sowie der Zwischenkettenvarianz Z
- Anwendung auf das binäre System aus Stern und Exoplanet für Fit der Radialgeschwindigkeit

Referenzen

[1] Sharma, Sanjib. "Markov chain monte carlo methods for bayesian data analysis in astronomy." *Annual Review of Astronomy and Astrophysics* 55 (2017): 213-259.

[2] Gelman, Andrew, Gareth O. Roberts, and Walter R. Gilks. "Efficient Metropolis jumping rules." *Bayesian statistics* 5.599-608 (1996): 42.

[3] https://de.wikipedia.org/wiki/Bahnneigung

Backup-Folien

- Wahl des Priors (extended)
- MCMC-Methoden: Metropolis-Hastings (extended)
- MCMC-Methoden: Gibbs Sampling
- MCMC-Methoden: Adaptive (RW) Metropolis-Hastings (extended)
- Konvergenzbestimmung (extended)

Wahl des Priors $p(\theta|I)$ (adaptiert aus [1])

- Suche in entarteten Parameterräumen durch Prior-Wahl einschränken
- *Informativer Prior:* kann durch Info aus vergangenen Analysen sehr restriktiv gewählt werden, wichtig bei nicht aussagekräftigen Daten
- *Ignorance Prior:* bei aussagekräftigen Daten reicht glatter Prior in Bereichen hoher Parameter-Likelihood für gute Ergebnisse, z.B. durch
  - Gleichverteilung in möglichem Parameterraum
  - Invarianz der Likelihood $p(x'|\theta')dx' = p(x|\theta)dx$ bezüglich Transformationsgruppen $(\theta', x') = h(\theta, x)$ wie Translation, Skalierung, Rotation…
  - Maximierung der Entropie $S = -\sum_i p_i \log p_i$ von Priors, die Einschränkungen erfüllen

MCMC-Methoden: Metropolis-Hastings (MH) [1]

Setze $K(\theta_i, \theta_{i+1}) = q(\theta_{i+1}|\theta_i)\alpha(\theta_i, \theta_{i+1})$ mit Akzeptanzwahrscheinlichkeit $\alpha(\theta_i, \theta_{i+1}) = \min\,[1, a_i]$ und $a_i = \frac{p^*(\theta_{i+1})q(\theta_i|\theta_{i+1})}{p^*(\theta_i)q(\theta_{i+1}|\theta_i)}$ um *Detailed Balance* zu erfüllen.

$$\boxed{\Sigma^* \overset{[2]}{=} (2.38^2/D)\Sigma_{p^*}}$$

Setze Vorschlagsdichte $q(\theta_{i+1}|\theta_i)$

- *Symmetrisch* $= q(\theta_i|\theta_{i+1}) \Rightarrow a_i = p^*(\theta_{i+1})/p^*(\theta_i)$
- *Random Walk* $= q(\theta_{i+1} - \theta_i)$, z.B. $U(\theta_i - \sigma, \theta_i + \sigma)$ oder $\mathcal{N}(\theta_i, \Sigma)$
- *Unabhängig* $= q(\theta_{i+1}) \mathrel{\hat{=}} [\,p^*(\theta_{i+1})$ *mit längeren Ausläufern*$\,]$
- *Langevin* $\sim \mathcal{N}(\theta_i + \frac{\sigma^2}{2}\nabla \log p^*(\theta_i), \sigma^2)$ falls $\nabla \log p^*(\theta_i)$ bekannt

MCMC-Methoden: Gibbs Sampling [1]

Bei jedem Schritt i, ziehe iterativ jeden Modellparameter $\theta_i^j = (m, c, \mu_a, \sigma_a, P_a)_i$ einzeln:

$$K(\theta_i, \theta_{i+1}) = \kappa_{1\to d}(\theta_{i+1}|\theta_i) = \prod_{j=1}^{d} p^*(\theta_{i+1}^j|\theta_{i+1}^1, \dots, \theta_{i+1}^{j-1}, \theta_i^{j+1}, \dots, \theta_i^d)$$

$$\Rightarrow \quad p^*(\theta_i)\kappa_{\boxed{1\to d}}(\theta_{i+1}|\theta_i) = p^*(\theta_{i+1})\kappa_{\boxed{d\to 1}}(\theta_i|\theta_{i+1})$$ ~~Detailed Balance~~

$$\Leftrightarrow \quad \boxed{\int p^*(\theta_i)\kappa(\theta_{i+1}|\theta_i)d\theta_i = p^*(\theta_{i+1})}$$ Def. von Stationarität

Zufällige Parameterreihenfolge beim Ziehen $\Rightarrow$ *Detailed Balance*

Falls Ziehen aus bedingter Dichte schwierig: *Metropolis Within Gibbs* für einzelne Parameter oder Parametersets

MCMC-Methoden: Adaptive (RW) Metropolis

Gewählt: Vorschlagsdichte $q(\theta_{i+1}|\theta_i) = \mathcal{N}(\theta_i, s_i\Sigma_i)$
Nach jedem Schritt i werden zusätzlich die Vorschlagsparameter
$\rho_i = (s_i, \mu_i, \Sigma_i)$ aktualisiert, sodass $\rho_i \xrightarrow{i\to\infty} \rho^*$, für welche
die Markovkette am schnellsten konvergiert:

Gesucht: $\langle N(\rho^*)\rangle = \hat{M}(\rho^*) = m^* = f(\theta^*) \xleftarrow{i\to\infty} f(\theta_i)$

z.B.: $\gamma_i = \gamma/i^\beta$
mit $\beta \in]0,1[$

Setze $\rho_{i+1} - \rho_i = \gamma_i[f(\theta_{i+1}) - N(\rho_i)]$ mit $\sum_{i=1}^{\infty} \gamma_i = \infty$ und $\sum_{i=1}^{\infty} \gamma_i^2 < \infty$

$\Rightarrow \quad \log s_{i+1} - \log s_i = \gamma_{i+1}(\alpha_i - \alpha^*)$ mit $\alpha^* = \alpha(D)$ bekannt
und $\quad \Sigma_{i+1} - \Sigma_i = \gamma_{i+1}[(\theta_{i+1} - \mu_i)(\theta_{i+1} - \mu_i)^T - \Sigma_i]$
mit $\quad \mu_{i+1} - \mu_i = \gamma_{i+1}(\theta_{i+1} - \mu_i)$

Konvergenzbestimmung [1]

- *Effective Sample Size (ESS):* Für N unabhängige Stichproben gilt $\sigma_\theta^{MC} = \sigma_\theta/\sqrt{N}$. Markovkette hat jedoch Autokorrelation
$$\rho_{\theta\theta}(t) = \frac{\mathbb{E}[(\theta_i - \overline{\theta})(\theta_{i+t} - \overline{\theta})]}{\mathbb{E}[(\theta_i - \overline{\theta})^2]} \neq 0$$
sodass $N \to N^{eff} = N/(2\tau_{int,\theta})$ mit $\tau_{int,\theta} = \frac{1}{2}\sum_{t=-\infty}^{\infty} \rho_{\theta\theta}(t)$

- Bei m gleichzeitg berechneten Ketten, welche $p^*(\theta)$ aus den letzten n Ziehungen ableiten, ergibt sich die gepoolte Varianz $V = \frac{m+1}{m}Z + \frac{n-1}{n}I$ mit Zwischen- und durchschnittlicher Innerkettenvarianz Z und I, sodass $R = \sqrt{V/I} \to 1$ Konvergenz anzeigt