week12a: Kapitel 7: Effizienz von Schätzern, Teil 3
Die Cramer-Rao Abschätzung für das Beispiel 3

Erinnern wir uns an das Beispiel 3, das war der zeitdiskrete Ornstein-Uhlenbeck Prozess oder kurz OU-Prozess:

Beispiel 3: Wir diskretisieren das Zeitintervall $[0,T]$ mit einem Δt gemäss

$$[0,T] \;\approx\; \bigl\{0, \Delta t, 2\Delta t, \cdots, k\Delta t, \cdots, N\Delta t\bigr\} \;=:\; \{t_0, t_1, t_2, \cdots, t_k, \cdots, t_n = T\}$$

mit $t_k = k\Delta t$ und $t_N = N\Delta t = T$ so dass $N = T/\Delta t$. Ein zeitdiskreter Ornstein-Uhlenbeck oder kurz OU-Prozess $\{x_{t_k}\}_{k=1}^N$ ist dann gegeben durch die Rekursion

$$x_{t_k} \;=\; x_{t_{k-1}} \;+\; \kappa(\mu - x_{t_{k-1}})\Delta t \;+\; \sigma\sqrt{\Delta t}\,\phi_k \tag{1}$$

mit standard-normalverteilten Zufallszahlen ϕ_k und einem deterministisch vorgegebenen Startwert $x_{t_0} = x_0$, etwa $x_0 = \mu$. Das ist dann äquivalent zu der folgenden stochastischen Rekursionsvorschrift

$$x_{t_k} \;=\; \alpha\, x_{t_{k-1}} \;+\; \beta \;+\; \eta\,\phi_k \tag{2}$$

wenn wir die Abkürzungen

$$\begin{aligned} \alpha &:= 1 - \kappa\Delta t \\ \beta &:= \kappa\mu\Delta t \\ \eta &:= \sigma\sqrt{\Delta t} \end{aligned} \tag{3}$$

benutzen. Die Likelihood-Funktion war gegeben durch

$$L(\alpha,\beta,\eta) \;=\; \frac{1}{(2\pi\eta^2)^{n/2}}\;\exp\Bigl\{-\frac{1}{2\eta^2}\sum_{k=1}^{n}\bigl(\,x_{t_k} - [\alpha x_{t_{k-1}} + \beta]\,\bigr)^2\Bigr\}\;\prod_{k=1}^{n} dx_{t_k}$$

so dass das $p_\theta(x)$ in diesem Fall also durch (mit $x_k \equiv x_{t_k}$)

$$p_\theta(x) \;=\; p_{\alpha,\beta,\eta}(x_1,\cdots,x_n) \;=\; \prod_{k=1}^{n}\left\{\,\frac{1}{\sqrt{2\pi\eta^2}}\,e^{-\frac{(\,x_k - [\alpha x_{k-1}+\beta]\,)^2}{2\eta^2}}\,\right\} \tag{4}$$

gegeben ist.

Die Maximum-Likelihood-Schätzer waren gegeben durch

$$\begin{aligned}
\hat{\alpha}_{\rm ML} &= \frac{n\sum_{k=1}^n x_{t_{k-1}}x_{t_k} - \sum_{k=1}^n x_{t_{k-1}}\sum_{k=1}^n x_{t_k}}{n\sum_{k=1}^n x_{t_{k-1}}^2 - \bigl(\sum_{k=1}^n x_{t_{k-1}}\bigr)^2} \\
\hat{\beta}_{\rm ML} &= \frac{\sum_{k=1}^n x_{t_{k-1}}^2\sum_{k=1}^n x_{t_k} - \sum_{k=1}^n x_{t_{k-1}}x_{t_k}\sum_{k=1}^n x_{t_{k-1}}}{n\sum_{k=1}^n x_{t_{k-1}}^2 - \bigl(\sum_{k=1}^n x_{t_{k-1}}\bigr)^2}
\end{aligned}$$

und

$$\widehat{\eta^2}_{\rm ML} = \frac{1}{n}\sum_{k=1}^n \bigl(x_{t_k} - [\hat{\alpha}x_{t_{k-1}} + \hat{\beta}]\bigr)^2$$

mit denen dann gemäss

$$\begin{aligned}
\kappa &= \tfrac{1-\alpha}{\Delta t} \\
\mu &= \tfrac{\beta}{1-\alpha} \\
\sigma^2 &= \tfrac{\eta^2}{\Delta t}
\end{aligned} \tag{5}$$

die ursprünglichen Modell-Parameter (κ,μ,σ) geschätzt werden können.

In dem week9a.pdf hatten wir uns die empirischen, die simulierten Verteilungen von $\hat{\kappa}_{\rm ML}$, $\hat{\mu}_{\rm ML}$ und $\widehat{\sigma^2}_{\rm ML}$ angeschaut, das waren die Histogramme, und insbesondere festgestellt, dass die Streuungen der Verteilungen für $\hat{\kappa}_{\rm ML}$ und $\hat{\mu}_{\rm ML}$ für $n\to\infty$ nicht kleiner wurden, sondern das schien so mehr oder weniger gegen eine Grenzverteilung zu konvergieren. Die Schätzer sind nicht konsistent. Dieses Verhalten kann man nun mit Hilfe der Cramer-Rao Abschätzung etwas besser verstehen, und das wollen wir uns jetzt hier anschauen:

Die Cramer-Rao Abschätzung war gegeben durch

$$\mathsf{V}[\,\hat{\theta}_k\,] \;\geq\; \bigl[\, I^{-1}(\theta)\,\bigr]_{k,k} \tag{6}$$

mit der Fisher-Informationsmatrix

$$I(\theta) = \Bigl(-\mathsf{E}\Bigl[\, \tfrac{\partial^2 \log p_\theta}{\partial\theta_k\partial\theta_\ell}\,\Bigr]\,\Bigr)_{k,\ell=1,\ldots,m} \;\in\; \mathbb{R}^{m\times m} \tag{7}$$

In unserem Fall ist $m=3$ und

$$\theta = (\theta_1,\theta_2,\theta_3) = (\alpha,\beta,\eta^2) =: (\alpha,\beta,\nu)$$

Wir wollen das $I(\theta)$ berechnen. Anstelle von x_{t_k} schreiben wir für den Moment einfach nur x_k. Dann haben wir

$$\log p_\theta = -\frac{n}{2}\log(2\pi\nu) - \frac{1}{2\nu}\sum_{k=1}^n (x_k - \alpha x_{k-1} - \beta\,)^2$$

mit den folgenden Ableitungen:

$$
\begin{aligned}
\frac{\partial \log p_\theta}{\partial \alpha} &= +\frac{1}{\nu}\sum_{k=1}^{n}(\,x_k - \alpha x_{k-1} - \beta\,)x_{k-1} \\
\frac{\partial \log p_\theta}{\partial \beta} &= +\frac{1}{\nu}\sum_{k=1}^{n}(\,x_k - \alpha x_{k-1} - \beta\,) \\
\frac{\partial \log p_\theta}{\partial \nu} &= -\frac{n}{2\nu} + \frac{1}{2\nu^2}\sum_{k=1}^{n}(\,x_k - \alpha x_{k-1} - \beta\,)^2
\end{aligned}
$$

und

$$
\begin{aligned}
\frac{\partial^2 \log p_\theta}{\partial \alpha^2} &= -\frac{1}{\nu}\sum_{k=1}^{n} x_{k-1}^2 \\
\frac{\partial^2 \log p_\theta}{\partial \beta^2} &= -\frac{1}{\nu}\sum_{k=1}^{n} 1 = -\frac{n}{\nu} \\
\frac{\partial^2 \log p_\theta}{\partial \nu^2} &= +\frac{n}{2\nu^2} - \frac{1}{\nu^3}\sum_{k=1}^{n}(\,x_k - \alpha x_{k-1} - \beta\,)^2
\end{aligned}
$$

und den off-diagonalen Termen

$$
\begin{aligned}
\frac{\partial^2 \log p_\theta}{\partial \beta \partial \alpha} &= -\frac{1}{\nu}\sum_{k=1}^{n} x_{k-1} \\
\frac{\partial^2 \log p_\theta}{\partial \nu \partial \alpha} &= -\frac{1}{\nu^2}\sum_{k=1}^{n}(\,x_k - \alpha x_{k-1} - \beta\,)x_{k-1} \\
\frac{\partial^2 \log p_\theta}{\partial \nu \partial \beta} &= -\frac{1}{\nu^2}\sum_{k=1}^{n}(\,x_k - \alpha x_{k-1} - \beta\,)
\end{aligned}
$$

Wir müssen die Erwartungswerte von den zweiten Ableitungen berechnen. Aus der stochastischen Rekursionsgleichung des OU-Prozesses (2) bekommen wir

$$
x_{t_k} - \alpha\, x_{t_{k-1}} - \beta = \eta\,\phi_k \tag{8}
$$

mit standard-normalverteilten ϕ_k's. Also

$$
\mathsf{E}[\,x_k - \alpha x_{k-1} - \beta\,] = \mathsf{E}\big[\,\eta\,\phi_k\,\big] = 0
$$

und wegen $x_{k-1} = x_{k-1}(\phi_{k-1}, \phi_{k-2}, ..., \phi_1)$, also in dem x_{k-1} kommt kein ϕ_k drin vor, haben wir auch

$$
\begin{aligned}
\mathsf{E}\big[\,(\,x_k - \alpha x_{k-1} - \beta\,)x_{k-1}\big] &= \mathsf{E}\big[\,\eta\,\phi_k\,x_{k-1}\big] \\
&= \eta\,\mathsf{E}[\,\phi_k\,]\,\mathsf{E}\big[\,x_{k-1}\,\big] \\
&= \eta \times 0 \times \mathsf{E}\big[\,x_{k-1}\,\big] = 0
\end{aligned}
$$

Weiterhin,

$$\mathsf{E}\big[\,(\,x_k - \alpha x_{k-1} - \beta\,)^2\,\big] \;\;=\;\; \eta^2\,\mathsf{E}[\,\phi_k^2\,] \;\;=\;\; \eta^2 \;\;=\;\; \nu$$

Damit bekommen wir also

$$\begin{aligned}
\mathsf{E}\left[\,\frac{\partial^2 \log p_\theta}{\partial\alpha^2}\,\right] &= -\,\frac{1}{\nu}\sum_{k=1}^{n}\mathsf{E}[\,x_{k-1}^2\,] \\
\mathsf{E}\left[\,\frac{\partial^2 \log p_\theta}{\partial\beta^2}\,\right] &= -\,\frac{n}{\nu} \\
\mathsf{E}\left[\,\frac{\partial^2 \log p_\theta}{\partial\nu^2}\,\right] &= +\,\frac{n}{2\nu^2} \;-\; \frac{1}{\nu^3}\sum_{k=1}^{n}\nu \;\;=\;\; -\,\frac{n}{2\nu^2}
\end{aligned}$$

und die off-diagonalen Terme

$$\begin{aligned}
\mathsf{E}\left[\,\frac{\partial^2 \log p_\theta}{\partial\beta\partial\alpha}\,\right] &= -\,\frac{1}{\nu}\sum_{k=1}^{n}\mathsf{E}[\,x_{k-1}\,] \\
\mathsf{E}\left[\,\frac{\partial^2 \log p_\theta}{\partial\nu\partial\alpha}\,\right] &= 0 \\
\mathsf{E}\left[\,\frac{\partial^2 \log p_\theta}{\partial\nu\partial\beta}\,\right] &= 0
\end{aligned}$$

Also,

$$\begin{aligned}
I(\theta) &= \begin{pmatrix} \frac{1}{\nu}\sum_{k=1}^{n}\mathsf{E}[\,x_{k-1}^2\,] & \frac{1}{\nu}\sum_{k=1}^{n}\mathsf{E}[\,x_{k-1}\,] & 0 \\ \frac{1}{\nu}\sum_{k=1}^{n}\mathsf{E}[\,x_{k-1}\,] & \frac{n}{\nu} & 0 \\ 0 & 0 & \frac{n}{2\nu^2} \end{pmatrix} \\
&= \frac{n}{\nu}\begin{pmatrix} \frac{1}{n}\sum_{k=1}^{n}\mathsf{E}[\,x_{k-1}^2\,] & \frac{1}{n}\sum_{k=1}^{n}\mathsf{E}[\,x_{k-1}\,] & 0 \\ \frac{1}{n}\sum_{k=1}^{n}\mathsf{E}[\,x_{k-1}\,] & 1 & 0 \\ 0 & 0 & \frac{1}{2\nu} \end{pmatrix} \\
&=: \frac{n}{\nu}\begin{pmatrix} \langle x^2\rangle & \langle x\rangle & 0 \\ \langle x\rangle & 1 & 0 \\ 0 & 0 & \frac{1}{2\nu} \end{pmatrix}
\end{aligned} \tag{9}$$

mit den Grössen (es war $T = n\Delta t$)

$$\langle x^2\rangle \;\;\equiv\;\; \langle x^2\rangle_T \;\;:=\;\; \frac{1}{n}\sum_{k=1}^{n}\mathsf{E}[\,x_{k-1}^2\,] \;\;=\;\; \frac{1}{n\Delta t}\sum_{k=1}^{n}\mathsf{E}[\,x_{t_{k-1}}^2\,]\,\Delta t \;\;\overset{\Delta t\to 0}{\approx}\;\; \frac{1}{T}\int_0^T \mathsf{E}[\,x_t^2\,]\,dt \tag{10}$$

und analog

$$\langle x\rangle \;\;\equiv\;\; \langle x\rangle_T \;\;:=\;\; \frac{1}{n}\sum_{k=1}^{n}\mathsf{E}[\,x_{k-1}\,] \;\;=\;\; \frac{1}{n\Delta t}\sum_{k=1}^{n}\mathsf{E}[\,x_{t_{k-1}}\,]\,\Delta t \;\;\overset{\Delta t\to 0}{\approx}\;\; \frac{1}{T}\int_0^T \mathsf{E}[\,x_t\,]\,dt \tag{11}$$

Das Inverse von $I(\theta)$ ist jetzt kein Problem mehr,

$$I^{-1}(\theta) \;=\; \frac{\nu}{n}\begin{pmatrix} \frac{1}{\langle x^2\rangle - \langle x\rangle^2} & -\frac{\langle x\rangle}{\langle x^2\rangle - \langle x\rangle^2} & 0 \\ -\frac{\langle x\rangle}{\langle x^2\rangle - \langle x\rangle^2} & \frac{\langle x^2\rangle}{\langle x^2\rangle - \langle x\rangle^2} & 0 \\ 0 & 0 & 2\nu \end{pmatrix} \tag{12}$$

und die Cramer-Rao Abschätzung liefert dann die folgenden unteren Schranken für alle erwartungstreuen Schätzer $\hat{\alpha}, \hat{\beta}$ und $\hat{\nu}$:

$$\mathsf{V}[\hat{\alpha}] \;\geq\; \frac{\nu}{n}\,\frac{1}{\langle x^2\rangle - \langle x\rangle^2} \tag{13}$$

$$\mathsf{V}[\hat{\beta}] \;\geq\; \frac{\nu}{n}\,\frac{\langle x^2\rangle}{\langle x^2\rangle - \langle x\rangle^2} \tag{14}$$

$$\mathsf{V}[\hat{\nu}] \;\geq\; \frac{2\nu^2}{n} \tag{15}$$

Die Gössen $\langle x^2\rangle$ und $\langle x\rangle^2$ berechnen wir in dem folgenden Lemma:

Lemma 7.3.1: Die $\{x_{t_k}\}_{k=1}^n$ seien gegeben durch einen zeitdiskreten OU-Prozess (1) oder (2) mit $T = n\Delta t$ und die Grössen $\langle x\rangle_T$ und $\langle x^2\rangle_T$ seien definiert durch (10) und (11) weiter oben. Der Startwert x_0 des OU-Prozesses sei gegeben durch den Mean Reversion Level μ, also $x_{t_0} := \mu$. Dann gilt im Limes $\Delta t \to 0$:

$$\langle x\rangle_T \;=\; \mu \tag{16}$$

$$\langle x^2\rangle_T \;\overset{\Delta t\to 0}{\approx}\; \mu^2 \;+\; \frac{\sigma^2}{2\kappa}\,\frac{2\kappa T - 1 + e^{-2\kappa T}}{2\kappa T} \tag{17}$$

Insbesondere,

$$\langle x^2\rangle_T \;-\; \langle x\rangle_T^2 \;\overset{\Delta t\to 0}{\approx}\; \frac{\sigma^2}{2\kappa}\,\frac{2\kappa T - 1 + e^{-2\kappa T}}{2\kappa T} \tag{18}$$

Beweis: ..machen wir am Donnerstag. $\blacksquare$

Jetzt schauen wir uns die Schätzer für die ‘ursprünglichen’ Modellparameter (κ, μ, σ) an. Der Zusammenhang zu den ‘abgeleiteten’ Modellparametern $(\alpha, \beta, \nu = \eta^2)$ war gegeben durch (5), das waren die Gleichungen

$$\kappa \;=\; \tfrac{1-\alpha}{\Delta t}$$

$$\mu \;=\; \tfrac{\beta}{1-\alpha}$$

$$\sigma^2 \;=\; \tfrac{\nu}{\Delta t}$$

‘Ursprünglich’ meint hier, dass diese Parameter unabhängig von der Zeitdiskretisierung Δt sind und die abgeleiteten Modellparameter hängen dann also von dem Δt ab. Wenn wir also das κ mit

$$\hat{\kappa} \;=\; \frac{1}{\Delta t}\,(1 \,-\, \hat{\alpha}) \tag{19}$$

mit einem erwartungstreuen $\hat\alpha$ schätzen, dann bekommen wir notwendigerweise

$$
\begin{aligned}
\mathsf{V}[\,\hat\kappa\,] \quad &= \quad \mathsf{V}\left[\,\frac{1}{\Delta t}\,(1\,-\,\hat\alpha)\,\right] \\
&= \quad \frac{1}{(\Delta t)^2}\,\times\,\mathsf{V}[\,1\,-\,\hat\alpha\,] \\
&= \quad \frac{1}{(\Delta t)^2}\,\times\,\mathsf{V}[\,\hat\alpha\,] \\
&\overset{\text{Cramer}-\text{Rao (13)}}{\geq} \quad \frac{1}{(\Delta t)^2}\,\times\,\frac{\nu}{n}\,\frac{1}{\langle x^2\rangle-\langle x\rangle^2} \\
&= \quad \frac{1}{(\Delta t)^2}\,\times\,\frac{\sigma^2\Delta t}{n}\,\times\,\frac{1}{\langle x^2\rangle-\langle x\rangle^2} \\
&= \quad \frac{\sigma^2}{T}\,\times\,\frac{1}{\langle x^2\rangle-\langle x\rangle^2} \\
&\overset{\Delta t\to 0}{=} \quad \frac{\sigma^2}{T}\,\times\,\frac{1}{\frac{\sigma^2}{2\kappa}\,\frac{2\kappa T-1+e^{-2\kappa T}}{2\kappa T}} \\
&= \quad \frac{4\kappa^2}{2\kappa T-1+e^{-2\kappa T}}
\end{aligned}
\tag{20}
$$

und dieser Ausdruck geht also nicht nach 0 auch wenn das $\Delta t\to 0$ geht und das $n = T/\Delta t\to\infty$ geht. Ein erwartungstreuer Schätzer für das κ kann also erst im Limes $T\to\infty$ konsistent werden, erst im Limes grosser T kann seine Varianz beliebig klein gemacht werden. Für festes T ist die Varianz immer grösser oder gleich einem festem positiven Wert, gegeben durch die rechte Seite von (20).

Schauen wir uns die Schätzer für das σ^2 an, die sind gegeben durch

$$\widehat{\sigma^2} \;=\; \frac{1}{\Delta t}\,\hat\nu \tag{21}$$

und die Cramer-Rao Abschätzung liefert in diesem Fall

$$
\begin{aligned}
\mathsf{V}[\,\widehat{\sigma^2}\,] \quad &= \quad \mathsf{V}\left[\,\frac{1}{\Delta t}\,\hat\nu\,\right] \\
&= \quad \frac{1}{(\Delta t)^2}\,\times\,\mathsf{V}[\hat\nu] \\
&\overset{\text{Cramer}-\text{Rao (15)}}{\geq} \quad \frac{1}{(\Delta t)^2}\,\times\,\frac{2\nu^2}{n} \\
&= \quad \frac{1}{(\Delta t)^2}\,\times\,\frac{2\,(\sigma^2\Delta t)^2}{n} \\
&= \quad \frac{2\sigma^4}{n}
\end{aligned}
\tag{22}
$$

und in diesem Fall geht die untere Schranke, die rechte Seite von (22), nach 0 wenn n nach unendlich oder Δt nach 0 geht. Das deckt sich also genau mit dem Verhalten, was wir im week9a.pdf beobachtet haben.