Definition 6.2. Ein *statistisches Modell* ist ein Tripel $(\mathcal{M} =)\big(\mathcal{X}, \mathscr{F}, (P_\vartheta)_{\vartheta\in\Theta}\big)$, wo $\mathcal{X} \neq \varnothing$ Menge („Beobachtungs- oder Stichprobenraum"), $\mathscr{F} \subset 2^{\mathcal{X}}$ eine σ-Algebra, Θ eine Menge (mit $|\Theta| > 1$) und für jedes $\vartheta \in \Theta$ ist P_ϑ ein W'maß auf $(\mathcal{X}, \mathscr{F})$.

Das Modell $\mathcal{M}$ heißt *parametrisch*, wenn $\Theta \subset \mathbb{R}^d$ für ein $d \in \mathbb{N}$, speziell *einparametrisch*, wenn $d = 1$.

$\mathcal{M}$ heißt *diskret*, wenn $\mathcal{X}$ abzählbar ist, $\mathcal{M}$ heißt *stetig*, wenn $\mathcal{X} \subset \mathbb{R}^n$ und jedes P_ϑ eine Dichte $\rho_\vartheta : \mathcal{X} \to [0, \infty]$ besitzt.

Ein diskretes oder stetiges Modell heißt ein *Standardmodell*.

$\big(\mathcal{X}^n, \mathscr{F}^{\otimes n}, (P_\vartheta^{\otimes n})_{\vartheta\in\Theta}\big)$ heißt das n-fache Produktmodell von $\mathcal{M}$ (für Produktmaße vgl. Def. 2.21).

Definition 6.3. $\big(\mathcal{X}, \mathscr{F}, (P_\vartheta)_{\vartheta\in\Theta}\big)$ statistisches Modell, $(S, \mathscr{A})$ messbarer Raum.

1. Eine Zufallsvariable X (definiert auf $(\mathcal{X}, \mathscr{F})$ mit Werten in S, d.h. $X : \mathcal{X} \to S$ ist $\mathscr{F}$-$\mathscr{A}$-messbar) heißt eine *Statistik* (manchmal auch: „Stichprobe").

2. Sei $\tau : \Theta \to \mathbb{R}$ eine reelle Kenngröße (oder „Parametermerkmal"), eine Statistik $T : \mathcal{X} \to \mathbb{R}$ heißt ein *Schätzer* (genauer: „Punktschätzer") für τ.

3. Ein Schätzer T für τ heißt *erwartungstreu* (oder „unverzerrt"), wenn gilt

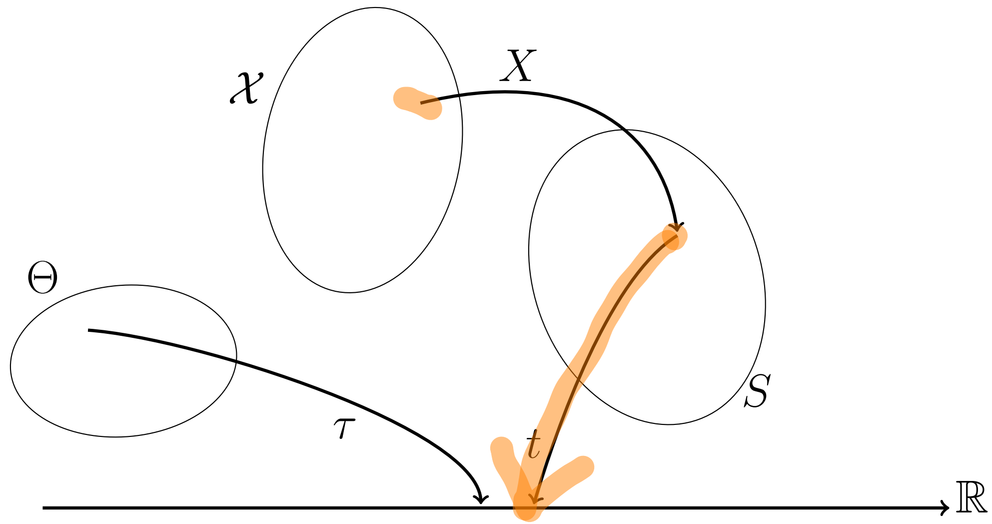
$$\forall\, \vartheta \in \Theta : \quad \mathbb{E}_\vartheta[T] = \tau(\vartheta).$$

   $b_\vartheta(T) := \mathbb{E}_\vartheta[T] - \tau(\vartheta)$ heißt die *Verzerrung* (englisch: bias) von T.

Die typische Konstruktion / Situation eines Schätzers ist $T = t(X)$ für eine Funktion $t : S \to \mathbb{R}$.

Man schreibt / benennt einen Schätzer für τ oft $\widehat{\tau}$.

$(\mathcal{X}, \mathscr{F}, (P_\vartheta)_{\vartheta\in\Theta})$ statistisches Modell, $X: \mathcal{X} \to S$ eine Statistik, $\tau: \Theta \to \mathbb{R}$ eine reelle Kenngröße (oder „Parametermerkmal“), $T: \mathcal{X} \to \mathbb{R}$ (hier definiert via $T = t \circ X$) ein *Schätzer* für τ



Schematische Darstellung eines Schätzers $T = t(X)$ für τ

Das Startbeispiel 6.1 in der Formalisierung von Definition 6.2 und 6.3 ausgedrückt:

- Statistisches Modell: $\mathcal{X} = \{0,1\}^n$, $\mathscr{F} = 2^{\mathcal{X}}$, $\Theta = [0,1]$, $P_\vartheta = \mathrm{Ber}_\vartheta^{\otimes n}$
- Parametermerkmal: $\tau: \Theta \to \mathbb{R}$, $\tau(\vartheta) = \vartheta$
- Statistik: $S = \mathcal{X}$, $X = \mathrm{Id}_{\mathcal{X}}$
- Schätzer: $T: \mathcal{X} \to \mathbb{R}$, $T\big((x_1, \dots, x_n)\big) = \dfrac{x_1 + \dots + x_n}{n}$

Dies ist ein Standardmodell im Sinne von Def. 6.2, es gilt $\mathbb{E}_\vartheta[T] = \vartheta = \tau(\vartheta)$ für alle $\vartheta \in [0,1]$, d.h. T ist hier ein erwartungstreuer Schätzer für τ.

Beispiel 6.4 (Erwartungstreue Schätzer für Mittelwert und Varianz im Produktmodell)**.** Für $\vartheta \in \Theta$ sei Q_ϑ ein W'maß auf $\mathbb{R}$ mit endlichem Mittelwert

$$m(\vartheta) := \int_{\mathbb{R}} x \, Q_\vartheta(dx)$$

und endlicher Varianz

$$v(\vartheta) := \int_{\mathbb{R}} \big(x - m(\vartheta)\big)^2 \, Q_\vartheta(dx).$$

Unter P_ϑ seien $X_1, \dots, X_n$ u.i.v., $X_i \sim Q_\vartheta$.

(In der Formalisierung von Definition 6.2 und 6.3 könnten wir wählen: $\mathcal{M} = \big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (P_\vartheta)_{\vartheta \in \Theta}\big)$ mit $P_\vartheta = Q_\vartheta^{\otimes n}$ für $\vartheta \in \Theta$, als Statistik betrachten wir $X = (X_1, \dots, X_n)$ mit $X_i : \mathbb{R}^n \to \mathbb{R}$ die Projektion auf die i-te Koordinate.

Bemerke: dies ist u.U. kein parametrisches Modell, man könnte z.B.

$$\Theta := \left\{Q : Q \text{ ist W'maß auf } \mathbb{R} \text{ mit } \int_{\mathbb{R}} x^2 \, Q(dx) < \infty\right\}$$

wählen.)

Dann ist

$$\overline{X} := \frac{1}{n} \sum_{i=1}^n X_i \quad \text{ein erwartungstreuer Schätzer für } m(\vartheta),$$

$$S^2 := \frac{1}{n-1} \sum_{i=1}^n \big(X_i - \overline{X}\big)^2 \quad \text{ein erwartungstreuer Schätzer für } v(\vartheta).$$

(In diesem Kontext heißt $\overline{X}$ auch der empirische Mittelwert oder Stichprobenmittelwert, S^2 die korrigierte Stichprobenvarianz, vgl. auch die Diskussion in Kapitel 6.1 über deskriptive Statistik.)

Für $\vartheta \in \Theta$ gilt nämlich

$$\mathbb{E}_\vartheta[\overline{X}] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_\vartheta[X_i] = \frac{1}{n}\cdot n\, m(\vartheta) = m(\vartheta),$$

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

$$\begin{aligned}\mathbb{E}_\vartheta\Big[\sum_{i=1}^{n}\big(X_i-\overline{X}\big)^2\Big] &= n\mathbb{E}_\vartheta\big[(X_i-\overline{X})^2\big] = n\mathrm{Var}_\vartheta\big[X_i-\overline{X}\big] \quad (+\,0^2)\\ &= n\mathrm{Var}_\vartheta\Big[\frac{n-1}{n}X_1 - \frac{1}{n}\sum_{i=2}^{n}X_i\Big] = n\Big(\big(\tfrac{n-1}{n}\big)^2\mathrm{Var}_\vartheta[X_1] + \tfrac{n-1}{n^2}\mathrm{Var}_\vartheta[X_1]\Big)\\ &= (n-1)\mathrm{Var}_\vartheta[X_1],\end{aligned}$$

u.a.

also

$$\mathbb{E}_\vartheta\big[S^2\big] = \frac{1}{n-1}\mathbb{E}_\vartheta\Big[\sum_{i=1}^{n}\big(X_i-\overline{X}\big)^2\Big] = v(\vartheta).$$

Beobachtung und Definition 6.5. Betrachten wir in der Situation von Beispiel 6.4 die Stichprobengröße n als variabel (formal: wir gehen zum unendlichen Produktmodell $\mathcal{M} = (\mathbb{R}^\infty, \mathcal{B}^{\otimes\infty}, (Q_\vartheta^{\otimes\infty})_{\vartheta\in\Theta})$ über (vgl. Beobachtung und Bericht 2.23), mit $X_i : \mathbb{R}^\infty \to \mathbb{R}$ Projektion auf i-te Koordinate).

Dann gilt für jedes $\vartheta \in \Theta$

$$\overline{X}_n := \frac{1}{n}\sum_{i=1}^n X_i \underset{n\to\infty}{\longrightarrow} m(\vartheta) \quad \text{stochastisch bzgl. } P_\vartheta \quad \text{und}$$

$$S_n^2 := \frac{1}{n-1}\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2 \underset{n\to\infty}{\longrightarrow} v(\vartheta) \quad \text{stochastisch bzgl. } P_\vartheta.$$

stoch. unter P_ϑ

Man sagt: Diese (Folgen von) Schätzer(n) sind *konsistent*.

Für $\overline{X}_n$ folgt dies direkt aus dem Gesetz der großen Zahlen (siehe Korollar 4.2), weiterhin ist

$$\frac{n-1}{n}S_n^2 = \frac{1}{n}\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2 = \left(\frac{1}{n}\sum_{i=1}^n X_i^2\right) - 2\left(\frac{1}{n}\sum_{i=1}^n X_i\overline{X}_n\right) + \left(\overline{X}_n\right)^2 = \underbrace{\left(\frac{1}{n}\sum_{i=1}^n X_i^2\right)}_{\xrightarrow[n\to\infty]{P} E_\vartheta[X_1^2]} - \left(\overline{X}_n\right)^2$$

$$\xrightarrow[n\to\infty]{P_\vartheta} \int_{\mathbb{R}} x^2\, Q_\vartheta(dx) - \left(m(\vartheta)\right)^2 = v(\vartheta)$$

gemäß dem Gesetz der großen Zahlen (siehe Bericht 4.8) zusammen mit Lemma 4.5 und obigem, wegen $\frac{n-1}{n} \to 1$ folgt mit Lemma 4.5 die Behauptung.

Bericht. Tatsächlich gilt sogar

$$\overline{X}_n \underset{n\to\infty}{\longrightarrow} m(\vartheta) \quad P_\vartheta\text{-f.s.} \qquad \text{und} \qquad S_n^2 \underset{n\to\infty}{\longrightarrow} v(\vartheta) \quad P_\vartheta\text{-f.s.}$$

(diese Schätzer sind auch „stark konsistent").

Für $\overline{X}_n$ folgt dies unter den Voraussetzungen von Beispiel 6.4 aus der Version des starken Gesetzes der großen Zahlen, die wir in Satz 4.6 bewiesen haben, für S_n^2 folgt dies aus Bericht 4.8.

6.2.1 Maximum-Likelihood-Schätzer

Definition 6.6. Sei $\mathscr{M} = (\mathcal{X}, \mathscr{F}, (P_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Standardmodell mit

$$\text{Gewichten } \rho_\vartheta(\cdot) \quad \text{bzw.} \quad \text{Dichte } \rho_\vartheta(\cdot) \qquad \text{für } \vartheta \in \Theta.$$

Die Funktion

$$\begin{array}{rcl} \rho: \ \mathcal{X} \times \Theta & \to & [0, \infty) \\ \rotatebox{90}{$\in$} & & \\ (x, \vartheta) & \mapsto & \rho(x, \vartheta) := \rho_\vartheta(x) \end{array}$$

heißt *Likelihood-Funktion* (manchmal auch „Plausibilitäts-Funktion“), für $x \in \mathcal{X}$ heißt

$$L_x : \Theta \to [0, \infty), \qquad L_x(\vartheta) = \rho(x, \vartheta)$$

die Likelihood-Funktion zum Beobachtungswert x.

Ein Schätzer $T : \mathcal{X} \to \Theta$ heißt (ein) Maximum-Likelihood-Schätzer, wenn

$$\Big(L_x(T(x)) = \Big)\, \rho(x, T(x)) = \max_{\vartheta\in\Theta} \rho(x, \cancel{\theta}\,\vartheta) \quad \forall x \in \mathcal{X}$$

(auch kurz ML-Schätzer genannt, engl. MLE = maximum likelihood estimator).

Beispiel 6.7. 1. („Rückfangmethode“, engl. “capture-recapture”) Ein Teich enthalte ϑ Fische (einer gewissen Art, $\vartheta \in \mathbb{N}$ ist der unbekannte Parameter), fange und markiere m, setze wieder aus. Wenn sich die markierten Fische gut verteilt haben, fange erneut n Fische.

Nehmen wir an, wir beobachten unter den erneut gefangenen x markierte Fische.

$\frac{x}{n} \approx \frac{m}{\vartheta}$

$\Updownarrow$

$\vartheta \approx \frac{n}{x} \cdot m$

Formalisierung als statistisches Modell:
$\mathcal{X} = \{0, 1, \ldots, n\}$, $\Theta = \{(m \vee n), (m \vee n) + 1, (m \vee n) + 2, \ldots\}$, $P_\vartheta = \mathrm{Hyp}_{m,\vartheta-m,n}$
Die Likelihood-Funktion ist

$$\rho(x, \vartheta) = \frac{\binom{m}{x}\binom{\vartheta-m}{n-x}}{\binom{\vartheta}{n}},$$

der ML-Schätzer ist

$$\widehat{\vartheta}_{\mathrm{ML}} = T(x) = \left\lfloor \frac{n}{x} \cdot m \right\rfloor,$$

denn

$$\begin{aligned}
\frac{\rho(x,\vartheta)}{\rho(x,\vartheta-1)} &= \frac{\binom{\vartheta-m}{n-x}\binom{\vartheta-1}{n}}{\binom{\vartheta}{n}\binom{\vartheta-1-m}{n-x}} \\
&= \frac{(\vartheta-m)(\vartheta-n)}{\vartheta(\vartheta-m-n+x)} = 1 - \frac{\vartheta x - mn}{\vartheta(\vartheta-m-n+x)} \begin{cases} > 1, & \vartheta < \frac{mn}{x}, \\ = 1, & \vartheta = \frac{mn}{x}, \\ < 1, & \vartheta > \frac{mn}{x} \end{cases}
\end{aligned}$$

(beachte: stets ist $\vartheta - m \geq n - x$, es gibt im Teich mindestens so viele unmarkierte Fische wie in der Rückfang-Stichprobe).

2. (Erfolgsw'keit im Binomialmodell)

$$\rho(x,\vartheta) = L_x(\vartheta) = \binom{n}{x}\vartheta^x(1-\vartheta)^{n-x}, \quad \big(P_\vartheta = \mathrm{Bin}_{n,\vartheta} \text{ für } \theta \in \Theta = [0,1],\ x \in \mathcal{X} = \{0,1,\dots,n\}\big),$$

$$\begin{aligned}\frac{d}{d\vartheta}\log L_x(\vartheta) &= \frac{d}{d\vartheta}\Big(\log\binom{n}{x} + x\log\vartheta + (n-x)\log(1-\vartheta)\Big)\\ &= \frac{x}{\vartheta} - \frac{n-x}{1-\vartheta} = 0 \iff \vartheta = \frac{x}{n},\end{aligned}$$

d.h. hier ist $\widehat{\vartheta}_{\mathrm{ML}} = \dfrac{x}{n}$.

(Es ist $\frac{d}{d\vartheta}\log L_x(\vartheta) > 0$ für $\vartheta < x/n$ und $\frac{d}{d\vartheta}\log L_x(\vartheta) < 0$ für $\vartheta > x/n$, d.h. es handelt sich tatsächlich um ein Maximum; Inspektion zeigt, dass auch in den Randfällen $x = 0$ und $x = n$ $\widehat{\vartheta}_{\mathrm{ML}} = \frac{x}{n}$ gilt.)

3. (Normales Modell mit bekannter Varianz) n Beobachtungen seien u.i.v. $\sim \mathcal{N}_{\vartheta,\sigma^2}$, $\sigma^2 > 0$ sei bekannt.
Mit $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ ist

$(\Theta = \mathbb{R},\ P_\vartheta = \mathcal{N}_{\vartheta,\sigma^2}^{\otimes n})$

$$\rho(x,\vartheta) = L_x(\vartheta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x_i-\vartheta)^2}{2\sigma^2} \Big)$$
$$= (2\pi\sigma^2)^{-n/2} \exp\Big(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i-\vartheta)^2 \Big)$$

d.h.

$$L_x(\vartheta) \stackrel{!}{=} \max \iff \sum_{i=1}^{n} (x_i-\vartheta)^2 \stackrel{!}{=} \min .$$

Mit $m(x) := \frac{1}{n} \sum_{i=1}^{n} x_i$ ist

$$\sum_{i=1}^{n} (x_i-\vartheta)^2 = \sum_{i=1}^{n} \big(x_i - m(x)\big)^2 + \big(m(x)-\vartheta\big)^2,$$

d.h. es ist $\widehat{\vartheta}_{\mathrm{ML}} = m(x)$, das empirische Mittel der Beobachtungen.

4. (Normales Modell, unbekannter Erwartungswert und unbekannte Varianz) n Beobachtungen seien u.i.v. $\sim \mathcal{N}_{\mu,v}$ mit unbekannten $\mu \in \mathbb{R}$ und $v \in (0, \infty)$.
(Formalisierung: $\mathcal{X} = \mathbb{R}^n$, $\Theta = \{\vartheta = (\mu, v) : \mu \in \mathbb{R}, v > 0\}$, $P_{(\mu,v)} = \mathcal{N}_{\mu,v}^{\otimes n}$)
Wie in 3. ist

$$\log L_x\big((\mu, v)\big) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log v - \frac{1}{2v}\sum_{i=1}^{n}(x_i - \mu)^2,$$

nach obigem ist $\widehat{\mu}_{\mathrm{ML}} = \frac{1}{n}\sum_{i=1}^{n} x_i$ Maximierer bezüglich μ (für jeden Wert von v), weiter ist

$$\frac{\partial}{\partial v}\log L_x\big((\mu, v)\big)\Big|_{\mu=\widehat{\mu}_{\mathrm{ML}}} = -\frac{n}{2v} + \frac{1}{2v^2}\sum_{i=1}^{n}(x_i - \widehat{\mu}_{\mathrm{ML}})^2,$$

also $\frac{\partial}{\partial v}\log L_x\big((\widehat{\mu}_{\mathrm{ML}}, v)\big) = 0 \iff v = \widehat{v}_{\mathrm{ML}} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \widehat{\mu}_{\mathrm{ML}})^2.$

(Und man prüft: $\log L_x\big((\widehat{\mu}_{\mathrm{ML}}, v)\big)$ ist wachsend für $v < \widehat{v}_{\mathrm{ML}}$, fallend für $v > \widehat{v}_{\mathrm{ML}}$.)

Beachte: Der ML-Schätzer für die unbekannte Varianz ist hier die (unkorrigierte) Stichprobenvarianz, also ist er nicht erwartungstreu (vgl. Beob. und Def. 6.5).

5. n Beobachtungen seien u.i.v. uniform auf $[0, \vartheta]$ (mit einem unbekannten $\vartheta \in (0, \infty)$).

Es ist

$$\widehat{\vartheta}_{\mathrm{ML}} = \max\{x_1, x_2, \ldots, x_n\},$$

denn

$$L_{(x_1,\ldots,x_n)}(\vartheta) = \begin{cases} \dfrac{1}{\vartheta^n}, & \text{falls } \vartheta \geq x_1, x_2, \ldots, x_n, \\ 0, & \text{sonst.} \end{cases}$$

Bericht 6.8 (Cramér-Rao-Schranke und „beste“ Schätzer)**.** Sei $(\mathcal{X}, \mathscr{F}, (P_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Standardmodell, $\rho(x,\vartheta)$ die Likelihoodfunktion.

Ein erwartungstreuer Schätzer T für ein reelles Parametermerkmal $\tau(\vartheta)$ heißt *varianzminimierend* (auch „gleichmäßig bester Schätzer“, engl. UMVU (= uniformly minimum variance unbiased) estimator), falls für jeden anderen erwartungstreuen Schätzer $\widetilde{T}$ für τ gilt

$$\mathrm{Var}_\vartheta[T] \le \mathrm{Var}_\vartheta[\widetilde{T}] \quad \text{für alle } \vartheta \in \Theta.$$

(In diesem Sinne ist T optimal und beantwortet – so existent – auf diese Weise die Frage „Wie gut kann man $\tau(\vartheta)$ anhand der Beobachtungen überhaupt schätzen?“)

Ein einparametriges Standardmodell (d.h. $\Theta \subset \mathbb{R}$) heißt regulär, falls gilt:

(i) $\Theta \subset \mathbb{R}$ ist ein offenes Intervall.

(ii) Likelihood-Funktion $\rho(x,\vartheta)$ ist strikt positiv auf $\mathcal{X} \times \Theta$ und für jedes x ist $\vartheta \mapsto \rho(x,\vartheta)$ stetig diff'bar.

(iii) $U_\vartheta(x) := \frac{d}{d\vartheta} \log \rho(x,\vartheta)$ erfüllt $I_\vartheta := \mathrm{Var}_\vartheta[U_\vartheta] \in (0,\infty)$
(U_ϑ heißt die „Scorefunktion“ und I_ϑ heißt die *Fisher-Information*) und es gilt

$$\int_{\mathcal{X}} \frac{d}{d\vartheta}\rho(x,\vartheta)\,dx = \frac{d}{d\vartheta}\int_{\mathcal{X}} \rho(x,\vartheta)\,dx \quad (=0).$$

Weiter heißt ein Schätzer T regulär, wenn für jedes $\vartheta \in \Theta$ gilt

$$\frac{d}{d\vartheta}\int_{\mathcal{X}} T(x)\rho(x,\vartheta)\,dx = \int_{\mathcal{X}} T(x)\frac{d}{d\vartheta}\rho(x,\vartheta)\,dx.$$

(Wenn $\mathcal{X}$ diskret ist, so ist jeweils das Integral $\int_{\mathcal{X}} \ldots dx$ durch die Summe $\sum_{x\in\mathcal{X}} \ldots$ zu ersetzen.)

$U_\vartheta(x) := \frac{d}{d\vartheta} \log \rho(x, \vartheta)$ die Scorefunktion, $I_\vartheta := \mathrm{Var}_\vartheta[U_\vartheta]$ die Fisher-Information.

Sei $\tau : \Theta \to \mathbb{R}$ ein stetig differenzierbares Parametermerkmal, T ein regulärer, erwartungstreuer Schätzer für τ in einem regulären Standardmodell. Dann gilt die Cramér-Rao-Schranke[1]:

$$\mathrm{Var}_\vartheta[T] \geq \frac{(\tau'(\vartheta))^2}{I(\vartheta)} \quad \forall\, \vartheta \in \Theta,$$

wobei Gleichheit genau dann gilt, wenn

$$T(x) - \tau(\vartheta) = \frac{\tau'(\vartheta) U_\vartheta(x)}{I(\vartheta)}.$$

[1]nach Harald Cramér, 1983–1985 und Calyampudi Radhakrishna Rao, *1920