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: $\mathcal{X} = \mathbb{R}^n$, $\Theta = \{\vartheta = (\mu, v) : \mu \in \mathbb{R}, v > 0\}$, $P_{(\mu,v)} = \mathcal{N}_{\mu,v}^{\otimes n}$)

$\left(\frac{1}{\sqrt{2\pi v}}\right)^n$

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
$$\widehat{\vartheta}_{\mathrm{ML}} = \max\{x_1, x_2, \dots, x_n\},$$
denn
$$L_{(x_1,\dots,x_n)}(\vartheta) = \begin{cases} \dfrac{1}{\vartheta^n}, & \text{falls } \vartheta \geq x_1, x_2, \dots, x_n, \\ 0, & \text{sonst.} \end{cases}$$

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

$$\left(\mathbb{E}_\vartheta[U_\vartheta] =\right) \int_{\mathcal{X}} \frac{d}{d\vartheta}\rho(x,\vartheta)\,dx = \frac{d}{d\vartheta} \underbrace{\int_{\mathcal{X}} \rho(x,\vartheta)\,dx}_{=1} \quad (= 0).$$

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]:

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

$I(\vartheta) = I_\vartheta$

wobei Gleichheit genau dann gilt, wenn

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

(insbes. $\frac{\tau'(\vartheta) U_\vartheta(x)}{I(\vartheta)} + \tau(\vartheta)$ darf nicht von ϑ abhängen)

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

Beispiel(-klasse). Exponentielle Familien (bzgl. der Statistik T)

Sei

$$\rho(x, \vartheta) = h(x) \cdot \exp\big(a(\vartheta) \cdot T(x) - b(\vartheta)\big)$$

für gewisse Funktionen $a, b : \Theta \to \mathbb{R}$ und $h : \mathcal{X} \to \mathbb{R}$

Dann ist

$$U_\vartheta(x) = a'(\vartheta)T(x) - b'(\vartheta),$$

$(\Rightarrow$
$0 = a'(\vartheta)\mathbb{E}_\vartheta[T] - b'(\vartheta))$

also

$$\mathbb{E}_\vartheta[T] = \frac{b'(\vartheta)}{a'(\vartheta)} =: \tau(\vartheta),$$

man kann zeigen, dass

$$\begin{aligned} I(\vartheta) &= \mathrm{Var}_\vartheta[U_\vartheta] \\ &= a'(\vartheta) \cdot \tau'(\vartheta), \end{aligned}$$

d.h. es gilt

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

T ist demnach in dieser Situation ein varianzminimierender erwartungstreuer Schätzer für τ.

Beispiel-Instanzen.

1. Binomialverteilungen: $P_\vartheta = \mathrm{Bin}_{n,\vartheta}$, $\vartheta \in [0,1]$

$$\rho(x,\vartheta) = \binom{n}{x}\vartheta^x(1-\vartheta)^{n-x} = \underbrace{\binom{n}{x}}_{=h(x)}\exp\Big(\underbrace{\frac{x}{n}}_{T(x)}\underbrace{n\log(\frac{\vartheta}{1-\vartheta})}_{=a(\vartheta)} + \underbrace{n\log(1-\vartheta)}_{=-b(\vartheta)}\Big), \quad x \in \mathbb{N}_0,$$

$T(x) = \frac{x}{n}$ ist varianzminimierender erwartungstreuer Schätzer für $\tau(\vartheta) = b'(\vartheta)/a'(\vartheta) = \vartheta$.

2. Poissonverteilungen: $P_\vartheta = \mathrm{Poi}_\vartheta$, $\vartheta \in (0,\infty)$

$b(\vartheta) = \vartheta$

$$\rho(x,\vartheta) = e^{-\vartheta}\frac{\vartheta^x}{x!} = \underbrace{\frac{1}{x!}}_{=h(x)}\, e^{\overbrace{x}^{T(x)}\overbrace{\log\vartheta}^{=a(\vartheta)} - \overbrace{\vartheta}^{b(\vartheta)}}$$

$a(\vartheta) = \log(\vartheta)$

Es ist $\tau(\vartheta) = \frac{1}{1/\vartheta} = \vartheta$, $T(x) = x$ ist varianzminimierender erwartungstreuer Schätzer für ϑ, seine Varianz ist $\frac{(\tau'(\vartheta))^2}{a'(\vartheta)\tau'(\vartheta)} = \frac{1^2}{\frac{1}{\vartheta}\cdot 1} = \vartheta$.

3. Normalverteilungen bei bekannter Varianz: $P_\vartheta = \mathcal{N}_{\vartheta,\sigma^2}$ mit festem $\sigma^2 > 0$

$x^2 - 2\vartheta x + \vartheta^2$

$$\rho(x,\vartheta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(- \frac{1}{2\sigma^2}(x-\vartheta)^2 \Big)$$

$$= \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/(2\sigma^2)}}_{=h(x)} \cdot \exp\Big(- \underbrace{x}_{=T(x)} \cdot \underbrace{\frac{\vartheta}{\sigma^2}}_{=a(\vartheta)} - \underbrace{\frac{\vartheta^2}{2\sigma^2}}_{=b(\vartheta)} \Big),$$

$\left(a(\vartheta) = -\frac{\vartheta}{\sigma^2} \right)$

also: $T(x) = x$ ist varianzminimierender erwartungstreuer Schätzer für $\vartheta = \tau(\vartheta) = \frac{b'(\vartheta)}{a'(\vartheta)}$, seine Varianz ist $\sigma^2 = \frac{1}{I(\vartheta)} = 1^2/(a'(\vartheta)\tau'(\vartheta))$.

Bemerkung. Das n-fache Produktmodell $\mathscr{M}^{\otimes n}$ eines regulären Modells ist wiederum regulär, seine Fisher-Information erfüllt $I^{(n)}(\vartheta) = n \cdot I(\vartheta)$.

Ist $\mathscr{M}$ exponentielles Modells bzgl. der Statistik T, so ist $\mathscr{M}^{\otimes n}$ ebenfalls ein exponentielles Modell, und die zugrundeliegende Statistik ist

$$T_n(x_1,\ldots,x_n) = \frac{1}{n}\sum_{i=1}^n T(x_i),$$

denn

$$\rho^{\otimes n}((x_1,\ldots,x_n),\vartheta) = \prod_{i=1}^n \rho(x_i,\vartheta) = \prod_{i=1}^n h(x_i) \exp\left(na(\vartheta) \cdot \frac{1}{n} \cdot \sum_{i=1}^n (T(x_i)) - nb(\vartheta) \right).$$

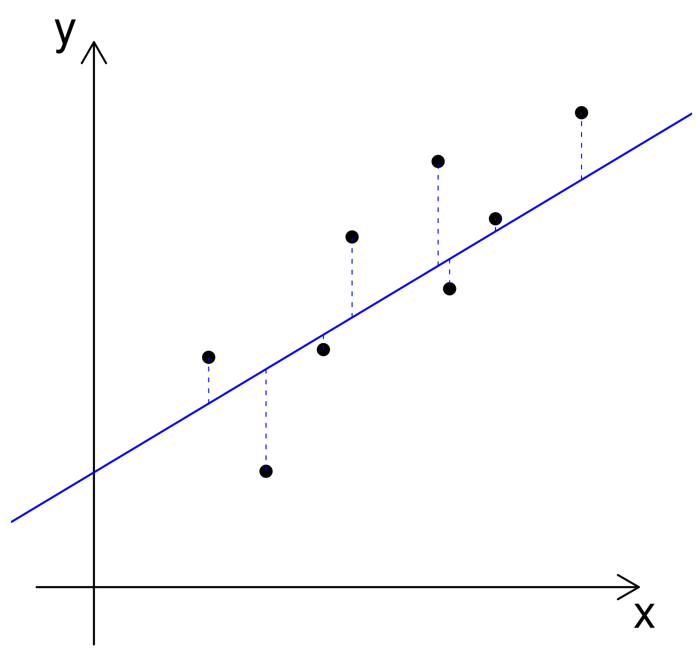
6.2.2 Eine Anmerkung zu linearer Regression und kleinste-Quadrate-Schätzung

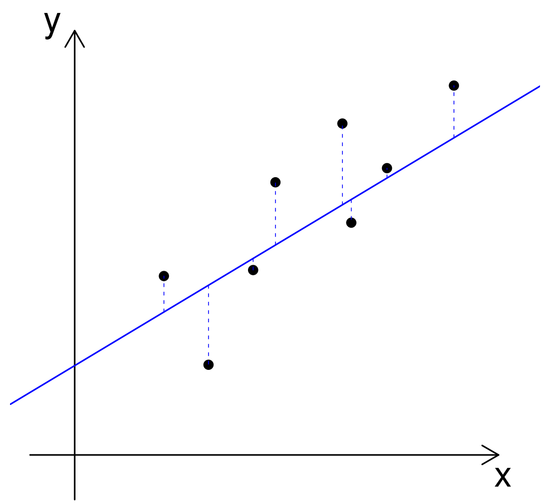
Beobachtung 6.9 (Lineare Regression als kleinste-Quadrate-Schätzer)**.** Nehmen wir an, die Beobachtungen bestehen aus n Messwertpaaren (x_i, y_i), $i = 1, \dots, n$ (Werte in $\mathbb{R}^2$) und wir vermuten aus theoretischen Gründen einen (affin-)linearen Zusammenhang, d.h. bei „perfekter" Messung gälte $y_i = \beta_0 + \beta_1 x_1$ für gewisse (uns unbekannte) Zahlen β_0 und β_1.

$y_i = \beta_0 + \beta_1 x_i$

(Ein „Lehrbuchbeispiel": y_i ist die Länge einer Stahlfeder bei Zugbelastung mit Gewicht x_i innerhalb des Gültigkeitsbereich des Hooke'schen Gesetzes.)

Aufgrund beispielsweise von Messungenauigkeiten (oder womöglich auch weil der lineare Zusammenhang in Wirklichkeit nur approximativ gilt) werden die realen Datenpunkte typischerweise nicht auf einer Geraden liegen.





Formulierung als statistisches Modell: $x_1, \dots, x_n$ seien feste (bekannte) Werte (x ist die „erklärende Variable"), für $\vartheta = (\beta_0, \beta_1) \in \Theta = \mathbb{R}^2$ sei unter P_ϑ

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \; i = 1, \dots, n \quad \text{mit } \varepsilon_i \text{ u.i.v. mit } \mathbb{E}[\varepsilon_i] = 0, \; \mathrm{Var}[\varepsilon_i] = \sigma^2$$

und wir fassen die beobachteten y_i-Werte als Realisierungen der Y_i auf (y ist die „abhängige Variable" oder „Zielgröße").

Ein naheliegender Ansatz, $\vartheta = (\beta_0, \beta_1)$ zu schätzen, ist der *kleinste-Quadrate-Schätzer*: Finde $\widehat{\beta}_0, \widehat{\beta}_1$ so, dass

$$\sum_{i=1}^n \left(y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_i)\right)^2 = \min_{(\beta_0,\beta_1)\in\mathbb{R}^2} \sum_{i=1}^n \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2$$

Gesucht:

$$\widehat{\beta}_0, \widehat{\beta}_1 \quad \text{so, dass} \quad \sum_{i=1}^n \left(y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_i)\right)^2 = \min_{(\beta_0,\beta_1)\in\mathbb{R}^2} \sum_{i=1}^n \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2$$

Die Lösung kennen wir schon (vgl. Beob. 3.21, die wir hier gewissermaßen nur „statistisch aussprechen“): Mit

$$\overline{x} := \frac{1}{n}\sum_{i=1}^n x_i, \quad \overline{y} := \frac{1}{n}\sum_{i=1}^n y_i, \quad \sigma_x^2 := \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^2, \quad \sigma_y^2 := \frac{1}{n}\sum_{i=1}^n (y_i - \overline{y})^2$$

$$\mathrm{cov}_{x,y} := \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})$$

ist

$$\widehat{\beta}_1 = \frac{\mathrm{cov}_{x,y}}{\sigma_x^2}, \quad \widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x}. \tag{6.1}$$

(($\bar{x}, \bar{y}$) liegt auf der Regressionsgerade)

(Betrachte nämlich eine ZV $(\widetilde{X}, \widetilde{Y})$ mit Werten in $\mathbb{R}^2$, deren Verteilung die empirische Verteilung der Datenpunkte ist, d.h. $\frac{1}{n}\sum_{i=1}^n \delta_{(x_i,y_i)}$, so ist $\mathbb{E}[\widetilde{X}] = \overline{x}$, $\mathbb{E}[\widetilde{Y}] = \overline{y}$, $\mathrm{Var}[\widetilde{X}] = \sigma_x^2$, $\mathrm{Var}[\widetilde{Y}] = \sigma_y^2$, $\mathrm{Cov}[\widetilde{X}, \widetilde{Y}] = \mathrm{cov}_{x,y}$ und die Behauptung folgt wörtlich aus Beob. 3.21, man kann natürlich auch die Rechnung dort nochmals hier wiederholen).

Übrigens: Wenn man zusätzlich annimmt, dass die ε_i u.i.v. $\sim \mathcal{N}_{0,\sigma^2}$ sind, so ist der kleinste-Quadrate-Schätzer hier auch zugleich der Maximum-Likelihood-Schätzer (mit einer Rechnung analog zu Bsp. 6.7, 3.).

6.3 Konfidenzintervalle (und Konfidenzbereiche)

Definition 6.10. Sei $\mathscr{M} = (\mathcal{X}, \mathscr{F}, (P_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell, $\tau(\vartheta)$ reelles Parametermerkmal, L, R Statistiken mit $L \le R$, $\alpha \in (0,1)$.

Das (zufällige) Intervall $I := [L, R]$ heißt ein *Konfidenzintervall* (manchmal auch „Vertrauensintervall") für τ zum (Sicherheits-)Niveau $1-\alpha$ (bzw. Irrtumsniveau α), wenn gilt

$L \leq \tau(\vartheta) \leq R$

$$\forall\, \vartheta \in \Theta : \quad P_\vartheta\big(\tau(\vartheta) \in I\big) \ge 1-\alpha.$$

Beachte: I ist zufällig, nicht aber ϑ (zumindest in unserer (sogenannten frequentistischen) Interpretation).

Allgemeiner heißt eine in Abhängigkeit von den Beobachtungen $x \in \mathcal{X}$ konstruierte Menge $C(x) \subset \Theta$ heißt ein *Konfidenzbereich* für τ zum (Sicherheits-)Niveau $1-\alpha$, wenn gilt

$$\forall\, \vartheta \in \Theta : \quad P_\vartheta\big(\{x \in \mathcal{X} : C(x) \ni \tau(\vartheta)\}\big) \ge 1-\alpha.$$

Offenbar möchte man i.A. I so kurz wie möglich wählen (soweit verträglich mit dem geforderten Niveau).

(das ist anders in der sog. Bayes-Statistik, die wir in dieser Vorlesung nicht behandeln)

Beispiel 6.11 (Konfidenzintervall für den Mittelwert im normalen Modell bei bekannter Varianz). Unter P_ϑ seien $X_1, X_2, \dots, X_n$ u.i.v. $\sim \mathcal{N}_{\vartheta,\sigma^2}$ mit $\vartheta \in \Theta := \mathbb{R}$, $\sigma^2 > 0$ sei bekannt (und fest).

Sei $q := \Phi^{-1}(1 - \frac{\alpha}{2})$ das $(1 - \alpha/2)$-Quantil der Standardnormalverteilung, $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

$$I := \left[\overline{X} - q \cdot \frac{\sigma}{\sqrt{n}}, \overline{X} + q \cdot \frac{\sigma}{\sqrt{n}}\right]$$

ist ein Konfidenzintervall für ϑ zum (Sicherheits-)Niveau $1 - \alpha$.

Denn unter P_ϑ ist $\overline{X} \sim \mathcal{N}_{\vartheta, \sigma^2/n}$

$$P_\vartheta\left(\overline{X} - q \cdot \frac{\sigma}{\sqrt{n}} \le \vartheta \le \overline{X} + q \cdot \frac{\sigma}{\sqrt{n}}\right)$$

$$= P_\vartheta\left(-q \le \underbrace{\frac{\overline{X} - \vartheta}{\sigma/\sqrt{n}}}_{\sim \mathcal{N}_{0,1} \text{ unter } P_\vartheta} \le q\right) = P(-q \le \overset{\sim \mathcal{N}_{0,1}}{Z} \le q)$$

$$= \underbrace{P(Z \le q)}_{= 1 - \frac{\alpha}{2}} - \underbrace{P(Z \le -q)}_{= \alpha/2} = 1 - \alpha$$

$\lrcorner$