

(3.96) Definition (Doppel-exponentialverteilung):

Die Doppel-exponentialverteilung (auch: Laplace-Verteilung) mit Skalensparameter $\lambda > 0$ ist eine stetige Wahrscheinlichkeitsverteilung auf $\mathbb{R}$ mit Lebesgue-dichte f_λ , gegeben durch

$$f_\lambda(t) = \frac{\lambda}{2} \exp(-\lambda |t|), \quad t \in \mathbb{R}.$$

(3.97) Satz:

Wählen wir unter den Gegebenheiten von Definition 3.92 die a priori -Dichte als $f_{\tilde{\beta}}(\beta) = \prod_{j=1}^p f_{\lambda_j}(\beta_j)$, nehmen wir also a priori stochastisch unabhängige, identisch Laplace(λ) - verteilte Regressionskoeffizienten an, so ist der resultierende MAP - Schätzer für β ein L_1 - Norm penalisierter Maximum - Likelihood - Schätzer.

Beweis:

Für die a posteriori - Dichte von $\tilde{\beta}$ gilt

$$f_{\tilde{\beta}|Y=y}(\beta) \propto \ell(\beta; y) \cdot \prod_{j=1}^p \exp(-\lambda_j |\beta_j|).$$

Damit ist

$$\hat{\beta}_{\text{post.}} = \arg \max_{\beta \in \mathbb{R}^p} f_{\tilde{\beta}|Y=y}(\beta)$$

$$= \arg \max_{\beta \in \mathbb{R}^p} \{ \ln(\ell(\beta; y)) - \lambda \sum_{j=1}^p |\beta_j| \}$$

$$= \sigma_y \max_{\beta \in \mathbb{R}^p} \{ \ln(\ell(\beta|y)) - \lambda \cdot \|\beta\|_1 \}. \quad \square$$

(3.38) Bemerkungen:

(a) Es gilt äquivalent

$$\hat{\beta}_{\text{post.}} = \sigma_y \min_{\beta \in \mathbb{R}^p} \{ -\ln(\ell(\beta|y)) + \lambda \|\beta\|_1 \}.$$

Im Falle normalverteilter Responserwartungen ist $-\ln(\ell(\beta|y))$ isotone Transformation der Fehlerquadratsumme.

(b) Im klassischen ANCOVA-Modell mit unbekannter Fehlervarianz σ^2 bietet es sich wieder an, die bedingte a priori - Verteilung $\mathcal{L}(\beta | \sigma^2 = \sigma^2)$ als $[\text{Laplace}(\lambda/\sigma)]^p$ zu wählen, da dann für jede a priori - inverse Gamma-Verteilung von σ^2 sichergestellt ist, dass die a posteriori - Dichte unimodal ist (siehe Park und Casella, 2008).

(c) Der Skalierungs- bzw. Regularisierungsparameter λ kann entweder (z.B. durch Kreuzvalidierung oder marginale Likelihoodmaximierung) explizit gewählt werden oder es wird eine zusätzliche Hierarchiestufe in die Bayesianische Modellierung eingeführt, indem eine Hyper-a priori - Verteilung für $\tilde{\lambda}$ gewählt wird. Park und Casella (2008) empfehlen

eine Gammaverteilung für $\tilde{z}^2$.

(d) Unter den Gegebenheiten von (b), also klassischer ANCOVA, entspricht $\hat{\beta}_{\text{post}}$ aus Satz 3.97 dem lasso (least absolute shrinkage and selection operator) - Schätzer aus Tibshirani (1996).

(e) L_1 -Regularisierung führt im Gegensatz zur L_2 -Regularisierung oft implizit zu einer Variablenselektion. Dazu schreiben wir die Zielkriterien der beiden Verfahren um:

$$\hat{\beta}_{\text{lasso}} = \arg \min_{\mu_1} \{ -\ln(\ell(\beta, y)) \}, \quad \mu_1 := \{ \beta \in \mathbb{R}^p : \|\beta\|_1 \leq c_1 \},$$

$$\hat{\beta}_{\text{ridge}} = \arg \min_{\mu_2} \{ -\ln(\ell(\beta, y)) \}, \quad \mu_2 := \{ \beta \in \mathbb{R}^p : \|\beta\|_2^2 \leq c_2 \},$$

wobei $c_1 \equiv c_1(n)$ und $c_2 \equiv c_2(n)$ Transformationen des jeweiligen Regularisierungsparameters sind.

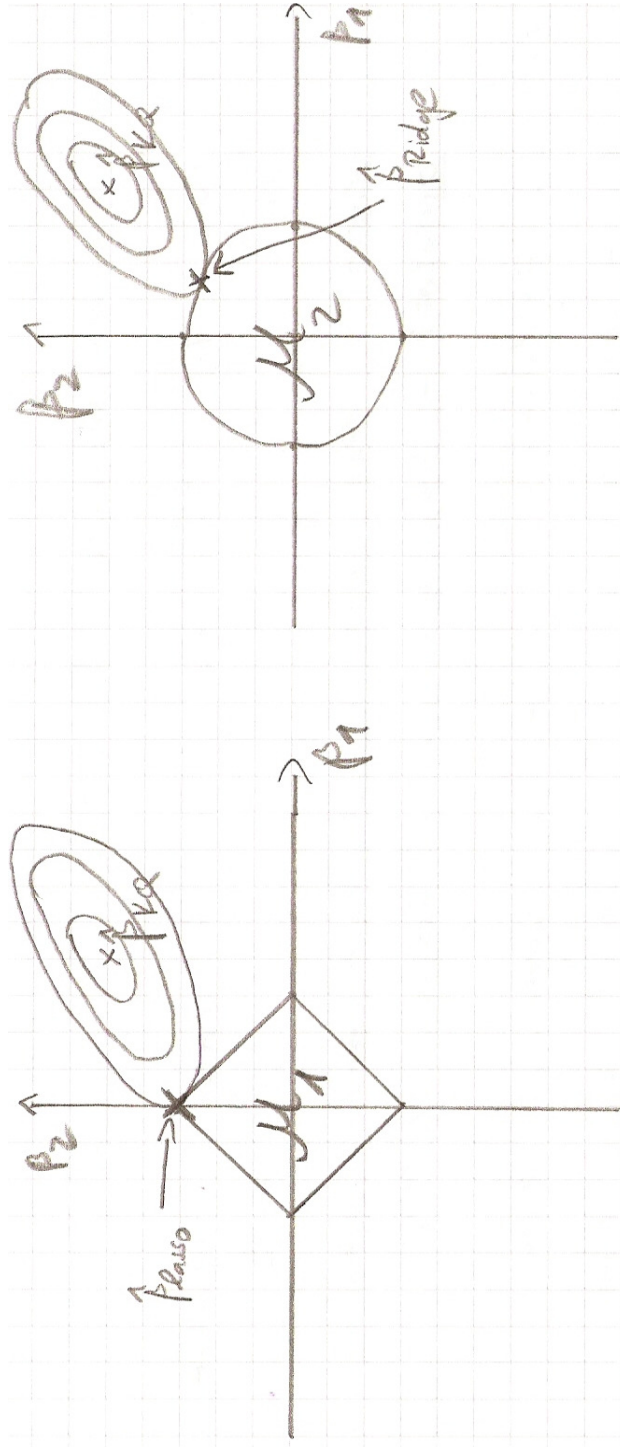
Im ANCOVA-Fall ist

$-\ln(\ell(\beta, y))$ äquivalent zur Fehlerquadratsumme

$$\sum_{i=1}^n (y_i - \sum_{j=1}^p \beta_j x_{ij})^2, \quad \text{deren Konturlinien für } p=2$$

bzüglich $\beta \in \mathbb{R}^2$ Ellipsen sind.

Wir erhalten die folgenden beiden Schaubilder adaptiert nach Tibshirani (1996).



(f) Es existieren eine Reihe verallgemeinerter Regularisierungstechniken, so zum Beispiel Bridge-Regression (L_q -Norm Regularisierung mit $q \geq 0$) oder "elastic net" (Strafterm ist gewichtetes Mittel aus L_1 - und L_2 -Norm Regularisierungstermen). Die Herleitung von (asymptotischen) Verteilungsaussagen für pensidierte ML- bzw. KQ-Schätzer ist Gegenstand aktueller Forschung.

Modellauswahl durch multiples Testen

(Bauer, Pötcher und Hackl (1988) Statistik 1, 39-44)

Modell: $(\Omega, F, (P_{\eta})_{\eta \in \Delta})$ mit $\dim(\Theta) = p > 1$
 $\eta \in \Delta$ $\nwarrow$ $p \in \mathbb{N}$

Annahme: $\exists I_0 \subseteq \{1, \dots, p\} : \beta_j = 0 \quad \forall j \in I_0$
 η : Störparameter, ggfs. unendlich-dimensional

Beispiel: ANCOVA-Modell 3.14:

Variaten: $E_{\beta} [Y_i | \bar{X}_i = \bar{x}_i] = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$

Es mögen Kovariaten $\{x_j : j \in I_0\}$ existieren, die keinen Einfluss auf

die Response haben, also $\beta_j = 0, j \in I_0$.

Variaten: $\text{Var}(Y_i) = \sigma^2$ Störparameter, nicht selbst Ziel der statistischen Inferenz

O.B.d.A.: $I_0 = \{r+1, \dots, p\}, 0 \leq r \leq p$ und $I_1 := \{1, \dots, p\} \setminus I_0 = \{1, \dots, r\}$.

Algorithmus (Multipler Test):

(1) Sei $\hat{\eta}_{jn}$, $1 \leq j \leq p$, Punktschätzer für η_j ,
basierend auf einer Stichprobe $Y = (Y_1, \dots, Y_n)^T$
mit $Y \sim P_{\eta, \eta}$ und zugehöriger Realisierung
 $y = (y_1, \dots, y_n)^T$.

Seien $\hat{\sigma}_{jn}$ (σ_{jn}) $1 \leq j \leq p$, positive reelle Werte
und $\hat{\sigma}_{jn}$ (σ_{jn}) Schätzer (für σ_{jn}).

(2) Schätze $\hat{I}_0 = \{1 \leq j \leq p : |\hat{\eta}_{jn}| \hat{\sigma}_{jn}^{-1} \leq c_j(n)\}$
und $\hat{I}_1 = \{1, \dots, p\} \setminus \hat{I}_0$, wobei

$\forall 1 \leq j \leq p$: $c_j(n)$ ein positiver Schwellenwert mit der Eigenschaft $c_j(n) \xrightarrow[n \rightarrow \infty]{} \infty$ ist.

Sollte $\hat{\sigma}_{jn} = 0$ für ein $1 \leq j \leq p$ sein, $\hat{\sigma}_{jn} = 0$ falls $\hat{\sigma}_{jn} = 0$ so setzen wir $|\hat{\sigma}_{jn}|^{\frac{1}{p-1}} = 0$, falls $\hat{\sigma}_{jn} = 0$ ist und ∞ sonst.

Satz (Schwache Konsistenz):

Falls $\forall 1 \leq j \leq p$ gilt:

$$(a) \quad \mathbb{E}_{\hat{\sigma}_{jn}} [(\hat{\sigma}_{jn} - \sigma_j)^2] \leq \sigma_{jn}^{-2} \quad \text{ist beschränkt}$$

(durch M_j).

$$(b) \quad \hat{\sigma}_{jn}^{\frac{1}{p-1}} \xrightarrow[n \rightarrow \infty]{} 1 \quad \text{in Wahrscheinlichkeit,}$$

$$(c) \quad \lim_{n \rightarrow \infty} \hat{\sigma}_{jn} c_j(n) \rightarrow 0,$$

$$\text{so ist} \quad \lim_{n \rightarrow \infty} \mathbb{P}_{\hat{\sigma}_{jn}} (\hat{\sigma}_{jn} \neq \sigma_j) = 0.$$

Beweis:

Typ I - Fehler:

Sei j mit $j \in I_0$ beliebig aber fest.

Dann ist $\mathbb{P}_{\hat{\sigma}_{jn}} (|\hat{\sigma}_{jn}| \hat{\sigma}_{jn}^{-1} [\hat{c}_j(n)]^{-1} > \delta)$ für alle $\delta > 0$:

$$\leq \delta^{-2} \mathbb{E}_{\hat{\sigma}_{jn}} [\hat{\sigma}_{jn}^2] \hat{\sigma}_{jn}^{-2} [\hat{c}_j(n)]^{-2}$$

nach Tschebyscheff'scher Ungleichung.

Wegen Annahme (a) und $j \in I_0$ ist

$$\mathbb{E} \left[\hat{\beta}_{jn}^2 \right] \hat{\sigma}_{jn}^{-2} \leq M_j \quad \text{und} \quad \text{wfr}$$

erhalten $S^{-2} M_j [c_j(n)]^{-2} \xrightarrow{(n \rightarrow \infty)} 0$, da

$c_j(n)$ divergent ist.

Insgesamt bedeutet dies, dass $|\hat{\beta}_{jn}| \hat{\sigma}_{jn}^{-1} [c_j(n)]^{-1} \xrightarrow{(n \rightarrow \infty)} 0$ in Wahrscheinlichkeit.

Wegen Annahme (b) gilt ebenso

$$|\hat{\beta}_{jn}| \hat{\sigma}_{jn}^{-1} [c_j(n)]^{-1} \xrightarrow{(n \rightarrow \infty)} 0 \quad \text{in Wahrscheinlichkeit.}$$

Damit strebt insbesondere

$$P_{\beta, j, n} (|\hat{\beta}_{jn}| \hat{\sigma}_{jn}^{-1} [c_j(n)]^{-1} > 1) \xrightarrow{(n \rightarrow \infty)} 0, \quad \text{also}$$

$$\lim_{n \rightarrow \infty} P_{\beta, j, n} (j \neq I_0) = 0.$$

Typ II-Fehler:

Sei $\mathcal{L}$ mit $\mathcal{L} \in I_1$ beliebig aber fest und

$0 < \varepsilon < 1$ eine fest gewählte Zahl.

Dann ist

$$P_{\beta, j, n} (|\hat{\beta}_{jn}| \hat{\sigma}_{jn}^{-1} \leq c_\varepsilon(n))$$

$$\leq P_{\beta, j, n} (|\hat{\beta}_{jn}| \hat{\sigma}_{jn}^{-1} \hat{\sigma}_{en}^{-1} \leq c_\varepsilon(n)),$$

$$|\hat{\sigma}_{en} \hat{\sigma}_{en}^{-1} - 1| < \varepsilon) +$$

$P_{\lambda, n} (|\hat{\sigma}_n - 1| \geq \varepsilon)$ wegen des
Satzes von der totalen Unabhinlichkeit und
Monotonie von $P_{\lambda, n}$. Also ist

$$P_{\lambda, n} (|\hat{\sigma}_n|^{1-\varepsilon} \leq c_\varepsilon(n)) \\ \leq P_{\lambda, n} (|\hat{\sigma}_n| \hat{\sigma}_n^{-1} (1+\varepsilon)^{-1} \leq c_\varepsilon(n)) \\ + P_{\lambda, n} (|\hat{\sigma}_n| \hat{\sigma}_n^{-1} - 1 \geq \varepsilon).$$

Der zweite Summand strebt für $n \rightarrow \infty$
gegen Null wegen Annahme (b).

Bleibt, den ersten Summanden zu diskutieren:

$$P_{\lambda, n} (|\hat{\sigma}_n| \hat{\sigma}_n^{-1} (1+\varepsilon)^{-1} \leq c_\varepsilon(n)) \\ = P_{\lambda, n} (\{|\hat{\sigma}_n| - |\hat{\sigma}_n|\} \hat{\sigma}_n^{-1} \geq |\hat{\sigma}_n| \hat{\sigma}_n^{-1} - c_\varepsilon(n) (1+\varepsilon)^{-1}) \\ \leq P_{\lambda, n} (|\hat{\sigma}_n - \hat{\sigma}_n| \hat{\sigma}_n^{-1} \geq |\hat{\sigma}_n| \hat{\sigma}_n^{-1} - c_\varepsilon(n) (1+\varepsilon)^{-1}) \\ \leq E_{\lambda, n} [(\hat{\sigma}_n - \hat{\sigma}_n)^2] \cdot \{|\hat{\sigma}_n| - \hat{\sigma}_n c_\varepsilon(n) (1+\varepsilon)^{-1}\}^{-2}.$$

Annahmen (a) und (c) zusammengekommen
ergeben schließlich, dass

$$\lim_{n \rightarrow \infty} P_{\lambda, n} (|\hat{\sigma}_n| \hat{\sigma}_n^{-1} \leq c_\varepsilon(n)) = 0 \\ = \lim_{n \rightarrow \infty} P_{\lambda, n} (e \in I_0^1). \quad \square$$

Korollar:

Unter den Voraussetzungen des multivariaten zentralen Grenzwertsatzes 3.28 für den KQ-Schätzer im klassischen ANCOVA-Modell liefert die Variablenselektionsprozedur durch

multipler Testen eine schwach konsistente Modellauswahl, falls $\forall 1 \leq j \leq p: \hat{\sigma}_{jn}^2 := \text{Var}(\hat{\beta}_j(n))$

gewählt sowie $\hat{\sigma}_{jn}^{12}$ gemäß Korollar 3.26

über die Residuenquadratsumme konstruiert wird und die Schwellenwerte die Bedingung

$$\forall 1 \leq j \neq l: \lim_{n \rightarrow \infty} c_j(n) n^{-1/2} = 0 \quad \text{erfüllen.}$$