Projekt LVS-IR-Taubenstein

Projektpartner: Sascha Filimon, Roman Ossner

Gruppenbetreuer: Dr. André Klima

Projektgruppe: Alexander Fogus, Lea Vanheyden, Zorana Spasojević

2. Mai 2020

Ludwig Maximilians Universität

Inhaltsverzeichnis

Hintergrund

- Konfliktsituation zwischen Mensch und Natur bzw. Tierreich im Alpengebiet
- Kooperation des Departments für Geographie an der LMU, Lawinencamp Bayern, Gebietsbetreuer Mangfallgebirge, Alpenregion Tegernsee/Schliersee und dem Deutschen Alpenverein München (DAV)
- Spezielle Untersuchungen am Spitzingsee
  ↳ Territorium von Wildtieren
  ↳ Beliebte Gegend für Sportler
- Untersuchung über die Mitnahme von LVS-Geräten anhand von Checkpoints und manueller Datenerhebung
  ↳ Ziel: Das Verhalten der Besucher analysieren

Datengrundlage

- Untersuchungsgegenstand: Wintersportler (vorrangig Skitourengänger & Schneeschuhgeher)
- Untersuchungszeitraum der Checkpoints Wintersaison 18/19
  ↳ Genauer Zeitraum: 25.12.2018 - 13.04.2019
- Checkpoints an zwei Routen (Nord- und Südseite) erfassen:
  ↳ Messungen insgesamt: 37216
  ↳ Personen mit LVS-Gerät: 8305
  ↳ Personen ohne LVS-Gerät: 28911
- Untersuchungszeitraum der manuellen Erhebung am 27.02 und 28.02.2019
  ↳ Genaue Daten aus Alex Deskriptiv

Datengrundlage

- Grafik zu Checkpoints Wintersaison 18/19
- Graf zu Studentischer Zählung 19/20

Lineares Regressionsmodell

- Erwartungswert einer Zufallsvariable wird durch Linearkombination von Kovariablen beschrieben
- Problem: Oft unzureichend, da auch nicht lineare Einflüsse auf Zielvariable wirken können
- Tabelle unterteilt in lineare und nicht lineare Kovariablen?

Univariate Glättung

Ermöglicht flexible Modellierung genau einer metrischen Kovariable auf Zielvariable

Daten

(y_i, z_i) zu der metrischen Zielvariable y und einer metrischen Kovariable z

Modell

$y_i = f(z_i) + \epsilon_i, \text{ wobei } \epsilon_i \text{ i.i.d. } N(0, \sigma^2)$

Daraus folgt:

$E(y_i) = f(z_i) \text{ und } Var(y_i) = \sigma^2, \; i = 1, ..., n$

$\Rightarrow$ Erwartungswert der Zielvariable wird durch Funktion f modelliert

Penalisierte Splines basierend auf B-Splines

- Funktion f durch Spline modellieren
- Konstruktion des Penalisierten Splines durch B-Spline Basisfunktion und Strafterm

Penalisierte Splines basierend auf B-Splines

B-Spline Basisfunktion

$$f(z) = \sum_{j=1}^{d} \gamma_j B_j(z), \;\; wobei \; d = m + l - 1$$

$$B_j^l(z) = \frac{z - k_j}{k_{j+l} - k_j} B_j^{l-1}(z) + \frac{k_{j+l+1} - z}{k_{j+l+1} - k_j + 1} B_{j+1}^{l-1}(z).$$

- Basisfunktion aus $(l+1)$ Polynomstücke des gewünschten Grades (l)

  ↳ werden an den Knoten $(l - 1 - mal)$ stetig zusammengesetzt
- Basisfunktion $(B_j(z))$ mit dem Kleinste-Quadrate-Schätzer $\hat{\gamma}_j$ skalieren
- Skalierte Basisfunktionen addieren

Penalisierte Splines basierend auf B-Splines

Strafterm

$$\lambda \int (f''(z))^2 dz$$

- Ziel: *Rauen* Schätzfunktionen entgegenwirken, d.h. zu starke Anpassung der Daten verhindern
- Bestrafung durch Glättungsparameter λ
- Für $\lambda \to \infty$, annähernd lineare Schätzfunktion

Zyklische P-Splines

Additives Modell

- Stellt ein nicht-parametrisches Regressionsmodell dar
- Vorteil: Neben linearen Effekten können auch nicht-lineare Einflüsse von Kovariablen auf Zielvariable modelliert werden

Standardmodell der additiven Regression (rausnehmen):

$$y_i = \underbrace{f_1(z_{i1}) + ... + f_q(z_{iq})}_{\text{nicht-parametrische Effekte}} + \underbrace{\beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik}}_{\text{parametrische Effekte}} + \epsilon_i$$

Generalisiertes additives Modell

- Keine normalverteilte Zielvariable
- Zielvariable nimmt typischerweise Verteilung aus Exponentialfamilie an

  ↳ Entweder normal-, binomial-, poisson- oder gammaverteilt

Generalisiertes additives Modell

Zielgröße

Beobachtete Zielgröße ist Binomial-verteilt $y_i|x_i \sim B(1, \pi_i)$:

$$y_i = \begin{cases} 1 & \text{,wenn Person mit LVS-Gerät identifiziert wird} \\ 0 & \text{,wenn Person ohne LVS-Gerät identifiziert wird.} \end{cases}$$

Erwartungswert

Für den Erwartungswert der Zielvariable gilt:

$E(y_i) = \pi_i = \frac{exp(\eta_i)}{1+exp(\eta_i)} = h(\eta_i)$

Logit-Link

$\log(\pi_i) = \log(\frac{\pi_i}{1-\pi_i})$

Generalisiertes additives Modell

Additiver Prädiktor:

Der Erwartungswert μ der Zielvariable y wird mit dem additiven Prädiktor:

$\eta_i = f_{1,2}(Uhrzeit_i, Datum_i)^* + f_3(Lawinenwarnstufe_i) + f_4(Wochentag_i) + f_5(Temperatur\_Residuen_i)^{**} + f_6(Sonneneinstrahlung\_Resiuden_i)^{**} + f_7(Schneehoehe\_Residuen_i)^{**} + \beta_0 + \beta_1(Ferientag_i)$

Bemerkung*: Modell kann durch Interaktionseffekt ergänzt werden
↳ Nicht-lineare Interaktion zwischen Datum und Uhrzeit

Bemerkung**: Smooth-Funktion für Residuen der Kovariablen Temperatur, Sonneneinstrahlung und Schnehöhe, da ansonsten
↳ *Concurvity*, stellt nicht-lineare Form der Kollinearität dar

Soll Tabelle mit Parametrischen und nicht Parametrischen Variablen eingefügt werden?

Gruppierte Daten (Überschriften für Spalten d Matrix hinzuf.)

- Bisher: Betrachtung von Individualdaten
- Weitere Möglichkeit: Gruppierte Daten
  ↳ Nach identische Zeilen der Kovariablen-Datenmatrix gruppieren

$$\begin{matrix} 2018-12-25 \\ \vdots \\ 2019-01-15 \end{matrix} \begin{bmatrix} 39 \\ \vdots \\ NA \end{bmatrix} \begin{bmatrix} 40 \\ \vdots \\ NA \end{bmatrix} \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \vdots & \vdots \\ x_{G1} & \cdots & x_{110k\cdot} \end{bmatrix}$$

- G=Anzahl verschiedener Kovariablenvektoren
  $\Rightarrow$ G deutlich kleiner als n=Stichprobenumfang
- Problem: Überdispersion

Überdispersion (Formeln Varianz u. Dispersionspara.?)

- Varianz ist ein wichtiges Instrument für Charakterisierung einer Verteilung
- Gilt als Streuungsmaß und beschreibt Konzentration der Verteilung um den Erwartungswert
- Abschätzung der Varianz für gruppierte Daten möglich
- Problem: In Praxis ist *empirische* Varianz oft höher als durch Binomial-Modell geschätzt
  $\Rightarrow$ Unterschätzung der Varianz wird als Überdispersion bezeichnet

Gamm