Poisson Regression

Verallgemeinerte Lineare Modelle (GLMs)

28.11.2011

Poisson Regression

Aus der Einführungsvorlesung

Poisson-Verteilung ist in der Regel gut geeignet, um Anzahlen zu modellieren.

Frage

Können wir die Poisson-Verteilung auch in einem Regressions- oder ANOVA-Kontext verwenden?

Wir nehmen also an, dass wir **Zähldaten (count data)** Y_i haben. Zu jeder Beobachtung Y_i haben wir auch **erklärende Variablen** $\underline{x}_i$.

Beispiel Schiffs-Havarien

Y	Anzahl Schaden-Ereignisse einer **Flotte** von Schiffen	
X	Anzahl Betriebs-Monate	M
	Schiffs-Typ	T: A, B, C, D, E
	Baujahr-Periode	C: 60, 65, 70, 75
	Beobachtungs-Periode	P: 0 (1960–74), 1 (1975–79)

	T	C	P	M	Y
1	A	60	0	127	0
2	A	60	1	63	0
3	A	65	0	1095	3
		...			
33	E	70	1	2161	12
34	E	75	1	542	1

Können wir dies in ein Regressionsmodell packen?

Wenn wir annehmen, dass $Y_i \sim Pois(\lambda_i)$, dann gilt

$$E[Y_i] = \lambda_i, \quad Var(Y_i) = \lambda_i,$$

wobei $\lambda_i > 0$.

Regression heisst: Wir **modellieren den Erwartungswert** in Abhängigkeit von den erklärenden Variablen.

Hier heisst das: λ_i hängt von den erklärenden Variablen ab.

Zusätzlich ist damit auch gerade die Varianz festgelegt, da die beiden Grössen identisch sind.

Wir haben aber die Restriktion $\lambda_i > 0$!

Mit geeigneter **Link-Funktion** können wir dieses Problem eliminieren (siehe logistische Regression).

Nehme Logarithmus als Link-Funktion und modelliere dann **linear**:

$$\log(\lambda_i) = \underline{x}_i^T \underline{\beta} \in \mathbb{R}.$$

Das ist äquivalent dazu, den Erwartungswert direkt als

$$\lambda_i = e^{\underline{x}_i^T \underline{\beta}} > 0$$

zu modellieren (vielleicht erscheint dies intuitiver).

Jetzt ist man eigentlich schon fast am Ziel.

Wir können wie früher die Likelihood aufschreiben als Funktion von $\underline{\beta}$.

Mit dem **Maximum-Likelihood Prinzip** kann man die Parameter schätzen.

Wieder wird diese mit **iterativen Verfahren** maximiert $\rightsquigarrow \widehat{\underline{\beta}}$.

Tests / Vertrauensintervalle etc. wie bei der logistischen Regression:

- **Wald Tests** aus dem iterativen Vefahren ("Z-Tests").
- **Likelihood-Ratio Tests** via Devianzen.
  Diese sind in der Regel vorzuziehen, da mehr Power.

Poisson Regression: Interpretation der Parameter

Schauen wir das Modell noch etwas genauer an. Es gilt

$$\begin{aligned} E[Y_i \,|\, \underline{x}_i] &= e^{\beta_0} \cdot e^{\beta_1 x_i^{(1)}} \cdots e^{\beta_m x_i^{(m)}} \\ &= e^{\beta_0} \cdot \exp\{\beta_1\}^{x_i^{(1)}} \cdots \exp\{\beta_m\}^{x_i^{(m)}} . \end{aligned}$$

D.h. ändert man $x^{(j)}$ um eine Einheit, bewirkt dies eine **Multiplikation** des Erwartungswertes mit dem **Faktor** $\exp\{\beta_j\}$.

Dies ist eigentlich auch sinnvoll so: Wir wollen ja in der Regel Aussagen im Stil von:

"In der einen Gruppe ist die Anzahl um einen gewissen **Faktor** höher (z.B. 30%)".

Ist β_j **positiv**, so bedeutet dies, dass der Erwartungswert **grösser** wird mit zunehmendem $x^{(j)}$. Umgekehrt entsprechend bei negativem β_j

Zurück zum Beispiel der Schiffs-Havarien

Es ist natürlich anzunehmen, dass die Anzahl Havarien **proportional** zur Anzahl Betriebsmonate ist.

In der Regression hätten wir einfach die **Rate** Y_i/M_i als "normalisierte" Zielgrösse verwendet.

Dies ist hier leider **nicht** möglich. Wieso? Weil Y_i/M_i **keiner** Poisson-Verteilung mehr folgt.

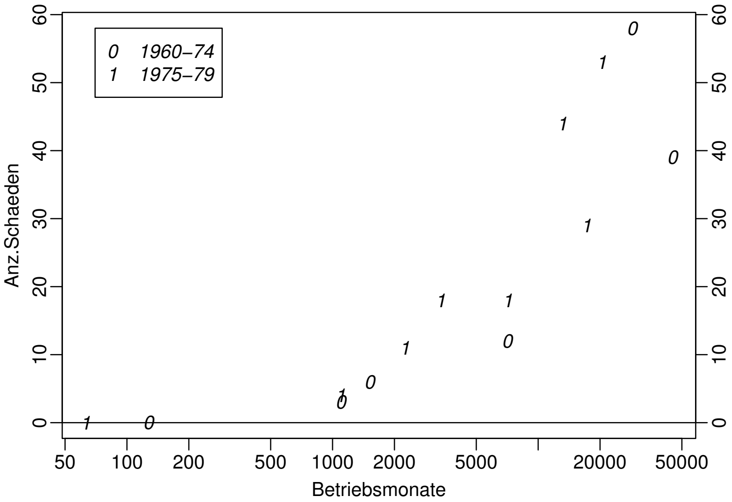
Unser gewünschtes Modell lautet eigentlich $\lambda_i = M_i \cdot e^{\underline{x}_i^T \underline{\beta}}$, bzw.

$$\log(\lambda_i) = \log(M_i) + \underline{x}_i^T \underline{\beta}.$$

Für M_i gibt es also **keinen** Parameter zu schätzen!

In **R** kann man dies der Funktion `glm` mit dem Argument `offset` mitteilen.

Illustration der Daten



Angepasstes Modell

```
glm(formula = incidents ~ type + period + year, family = poisson,
    data = d.ships, offset = log(service))

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -6.40590    0.21744 -29.460  < 2e-16 ***
typeB       -0.54334    0.17759  -3.060  0.00222 **
typeC       -0.68740    0.32904  -2.089  0.03670 *
typeD       -0.07596    0.29058  -0.261  0.79377
typeE        0.32558    0.23588   1.380  0.16750
period75     0.38447    0.11827   3.251  0.00115 **
year65       0.69714    0.14964   4.659 3.18e-06 ***
year70       0.81843    0.16977   4.821 1.43e-06 ***
year75       0.45343    0.23317   1.945  0.05182 .

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 146.328  on 33  degrees of freedom
Residual deviance:  38.695  on 25  degrees of freedom
AIC: 154.56

Number of Fisher Scoring iterations: 5
```

Interpretation Modell Output

Referenzlevel beim Schiffstyp ist Typ A. Wollen wir z.B. Typ B mit C vergleichen, so haben wir

$$\exp\{-0.543 - (-0.687)\} = 1.15.$$

D.h. Typ B hat **15% mehr Unfälle** pro Monat als Typ C **(multiplikativer Effekt!)**.

Vergleichen wir z.B. Typ A mit Typ C, so erhalten wir $\exp\{-0.687\} = 0.5$.

Typ C hat also **halb so viele** Unfälle wie Typ A.

Ein **approx. 95%-Vertrauensintervall** für diesen Faktor ist

$$[\exp\{-1.33\}, \exp\{-0.04\}] = [0.26,\ 0.96],$$

da

$$-0.68740 \pm 1.96 \cdot 0.32904 = [-1.33,\ -0.04].$$

Verallgemeinerte Lineare Modelle

Das Vorgehen scheint also sehr ähnlich zu sein wie bei der logistischen Regression!

Wir transformieren jeweils den Erwartungswert mit der **Link-Funktion** in den Raum des **linearen Prädiktors**, wo wir keine Einschränkungen mehr haben:

$$g\left(E[Y_i \,|\, \underline{x}_i]\right) = \eta_i = \underline{x}_i^T \underline{\beta} \in \mathbb{R}.$$

Die Verteilungen (und die Link-Funktionen) waren jedoch unterschiedlich:

- Lineare Regression: Normalverteilung
- Logistische Regression: Bernoulli- bzw. Binomialverteilung
- Poisson Regression: Poissonverteilung

Gute Nachricht: Wir können dies ganz allgemein aufschreiben.
Schlechte Nachricht: Das wird theoretisch.

Exponentialfamilie

Die **Exponentialfamilie** hat W'keitsfunktion bzw. Dichte

$$f\,(y;\theta,\phi,\omega) = \exp\left\{\frac{y\theta - b(\theta)}{\phi}\,\omega + c(y;\phi,\omega)\right\},$$

wobei

θ	**Kanonischer Parameter**
ϕ	**Dispersionsparameter** (Störparameter)
ω	Gewicht bei gruppierten Daten
b	Funktion die Verteilung festlegt
c	Normierung auf W'keit 1

Exponentialfamilie: Erwartungswert und Varianz

Es gilt:

$$\mu = E[Y] = b'(\theta), \quad \mathrm{Var}(Y) = b''(\theta) \cdot \frac{\phi}{\omega} = V(\mu) \cdot \frac{\phi}{\omega}$$

mit geeigneter Funktion V.

Wir können $\theta = (b')^{-1}(\mu)$ schreiben und erhalten somit

$$V(\mu) = b''\left((b')^{-1}(\mu)\right).$$

Zu merken ist:

“Die Varianzfunktion V ist eine Funktion des Erwartungswertes μ”.

Beispiel: Poissonverteilung

Wir haben

$$P(Y=y)=\frac{1}{y!}\lambda^y e^{-\lambda}, \quad \log(P(Y=y))=-\log(y!)+y\log(\lambda)-\lambda$$

Somit erhalten wir

$$\begin{aligned}
\theta &= \log(\lambda)\\
b(\theta) &= \exp(\theta)=\lambda\\
\phi &= 1\\
\omega &= 1\\
c(y;\phi,\omega) &= -\log(y!).
\end{aligned}$$

Beispiel: Normalverteilung

$$\begin{aligned}\log\left(f(y;\mu,\sigma^2)\right) &= -\log\left(\sqrt{2\pi}\sigma\right) - \frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2 \\ &= \frac{y\mu - \frac{1}{2}\mu^2}{\sigma^2} - \frac{y^2}{2\sigma^2} - \log\left(\sqrt{2\pi}\sigma\right)\end{aligned}$$

Hier haben wir

$$\begin{aligned}\theta &= \mu \\ \phi &= \sigma^2 \\ b(\theta) &= \theta^2/2 \\ c(y;\phi) &= -\frac{1}{2}\left(\frac{y^2}{\phi} + \log(2\pi\phi)\right).\end{aligned}$$

Weitere Beispiele

- Binomialverteilung
- Exponentialverteilung
- Gamma-Verteilung
- Weibull-Verteilung

Die Exponentialfamilie ist also eine **grosse Klasse** von Verteilungen!

Link-Funktion

Durch Verwendung der Link-Funktion sollen unmögliche Werte vermieden werden. Wir haben schon gesehen:

$$
\begin{array}{ll}
g(\mu) = \mu & \text{wenn } E[Y] \text{ beliebig} \\
g(\mu) = \log(\mu) & \text{wenn } E[Y] > 0 \\
g(\mu) = \text{logit}(\mu) = \log\left(\frac{\mu}{1-\mu}\right) & \text{wenn } 0 < E(Y) < 1.
\end{array}
$$

Kanonische Link-Funktion

Die Link-Funktion heisst **kanonische Link-Funktion**, falls

$$\eta = g(\mu) = \theta = (b')^{-1}(\mu).$$

Oder anders gesagt: Wenn der lineare Prädiktor gerade dem kanonischen Parameter entspricht.

Die erklärenden Variablen wirken dann also linear auf den kanonischen Parameter.

$\rightsquigarrow$ Wähle $g = (b')^{-1}$!

Normalverteilung	$g(\mu) = \mu$
Poissonverteilung	$g(\mu) = \log(\mu)$
Binomialverteilung	$g(\mu) = \text{logit}(\mu)$

Vorteile: Existenz und Eindeutigkeit, einfachere Schätzgleichungen

Schätzungen und Tests

Eigentlich haben wir schon alles gesehen:

- Schätzung mit **Maximum-Likelihood Prinzip**.
- **Iterative Verfahren** zur numerischen Maximierung (Iteratively Reweighted Least Squares).
- Tests mit **Wald-Tests** bzw. eher mit **Likelihood-Ratio Tests (Devianzen)**.

Übergrosse Streuung (Overdispersion)

Zurück zur **Poisson-Regression**.

Bei der Poissonverteilung (wie z.B. auch bei der Binomialverteilung) ist die **Varianz durch das Modell festgelegt**.

Oft hat man aber in den Daten eine **grössere Streuung** als durch das Modell erwartet würde.

Mögliche Gründe:

- Fehlende erklärende Variablen im Modell. Falls vorhanden $\rightsquigarrow$ benutzen!
- Korrelierte Beobachtungen (z.B. in Gruppen)
- Fehlende zufällige Effekte (λ enthält random effect). Verwende z.B. Negativbinomial-Verteilung.
- ...

Die einfachste Art, eine grössere Streuung zuzulassen, besteht darin, den Dispersionsparameter nicht mehr auf 1 festzulegen.

Zum Beispiel würden wir bei Poisson Regression das Modell erweitern zu

$$V(\mu) = \phi\mu$$

mit $\phi \neq 1$.

Man spricht dann von sogenannten **Quasi-Likelihood Modellen**.

Wie können wir die Overdispersion entdecken?

Verwende z.B. **Pearson-Chiquadrat-Statistik**

$$T = \sum_{i=1}^{n} \frac{\omega_i (y_i - \widehat{\mu}_i)^2}{\phi V(\widehat{\mu}_i)}.$$

Wenn ϕ nicht aus den Daten geschätzt werden muss, ist diese χ^2- verteilt mit $n - p$ Freiheitsgraden.

Alternativ kann man auch die **Residuendevianz** anschauen (gross bedeutet in der Regel, dass etwas nicht stimmt).

Man kann aber auch einfach den Dispersionsparameter schätzen und schauen, was für Auswirkungen dies hat.

Wie können wir die Overdispersion schätzen?

$$\widehat{\phi} = \frac{1}{n-p} \sum_{i=1}^{n} \frac{\omega_i (y_i - \widehat{\mu}_i)^2}{V(\mu_i)}.$$

D.h. zuerst werden die "gewöhnlichen" Parameter geschätzt, und **am Schluss** wird noch ϕ angepasst.

Wie macht man das in R?

Verwende `family = quasipoisson` in `glm`. Bei `regr` wird dies in der Regel standardmässig verwendet.

Für technisch Interessierte:

Die Funktion `glm` kümmert sich **nicht** um den Dispersionsparameter, dieser wird erst **am Schluss** in `summary.glm` berechnet (siehe Formel oben!).

Overdispersion: Auswirkungen

Was hat das alles für Konsequenzen?

Auf die **Schätzung** der Parameter hat das **keinen Einfluss** (siehe vorher).

Der Overdispersion-Parameter verändert aber die Varianz um den Faktor $\widehat{\phi}$.

Daher ändern sich die Standardfehler um den Faktor $\sqrt{\widehat{\phi}}$.

Und somit werden die Vertrauensintervalle um diesen Faktor **breiter**!

Beispiel Schiffs-Havarien

```
glm(formula = incidents ~ type + period + year, family = quasipoisson,
    data = d.ships, offset = log(service))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -6.40590    0.28276 -22.655  < 2e-16 ***
typeB       -0.54334    0.23094  -2.353  0.02681 *
typeC       -0.68740    0.42789  -1.607  0.12072
typeD       -0.07596    0.37787  -0.201  0.84230
typeE        0.32558    0.30674   1.061  0.29864
period75     0.38447    0.15380   2.500  0.01935 *
year65       0.69714    0.19459   3.583  0.00143 **
year70       0.81843    0.22077   3.707  0.00105 **
year75       0.45343    0.30321   1.495  0.14733

(Dispersion parameter for quasipoisson family taken to be 1.691028)

    Null deviance: 146.328  on 33  degrees of freedom
Residual deviance:  38.695  on 25  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 5
```

Residuenanalyse

Siehe logistische Regression. Es ex. mehrere mögliche Definitionen.

- **Response Residuals, Raw Residuals**

$$R_i = Y_i - \widehat{\mu}_i, \ \ \widehat{\mu}_i = g^{-1}(\underline{x}_i^T \underline{\widehat{\beta}})$$

- **Pearson Residuals**

$$R_i^{(P)} = R_i / \sqrt{V(\widehat{\mu}_i)/\omega_i} \quad \text{(standardisiert)}$$

- **Working Residuals, Link Residuals**

  Berechnung via: iterativ gewichtete Kleinste Quadrate
  $\rightsquigarrow$ lineare Näherung $\rightsquigarrow$ Residuen: “working residuals”.

- **Devianz-Residuen**

$$R_i^{(D)} = \text{sign}(Y_i - \widehat{\mu}_i)\sqrt{d_i},$$

  wobei d_i der entsprechende Summand der Residuendevianz ist.

Merkpunkte

- Verallgemeinerte lineare Modelle umfassen als Verteilungen der Zielgrösse:
  - Normalverteilung
  - Bernoulli- und Binomial-Verteilung,
  - Poisson-Verteilung,
  - Exponential- und Gamma-Verteilung.
- Theorie und Algorithmus für alle diese Modelle:
  - Iteratively Reweighted Least Squares
  - Maximum Likelihood-Schätzung
  - Begriff der Devianzen
- Residuen sind weniger nützlich, da "künstliche" Strukturen auftreten. Man braucht sie trotzdem!