Bayesiansk statistik, 732g43, 7.5 hp

Moment 2 - Linjär regressionsanalys

Bertil Wegmann

STIMA, IDA, Linköpings universitet

Översikt moment 2: linjär regressionsanalys

- Bayesiansk linjär regression utan förklaringsvariabler
- Bayesiansk enkel linjär regression
- Bayesiansk multipel linjär regression
- Kod_Moment2.R (kan laddas ned på kurshemsidan)

Normal data, okänd varians

- Modell:

$$Y_1, ..., Y_n | \mu, \sigma^2 \overset{iid}{\sim} N(\mu, \sigma^2)$$

$$\mu = \beta_0$$

där både medelvärdet μ och variansen σ^2 är okända.

- Modellen kan skrivas som en linjär regressionsmodell utan förklaringsvariabler:

$$y_i = \beta_0 + \epsilon_i$$

$$\epsilon_i \overset{iid}{\sim} N\left(0, \sigma^2\right)$$

- Prior:

$$p(\mu, \sigma)$$

- Om μ och σ antas oberoende apriori (boken):

$$p\left(\mu, \sigma\right) = p\left(\mu\right) p\left(\sigma\right).$$

Normal data, okänd varians

- Priorn för μ och σ^2 kan också specificeras enligt:

$$p\left(\mu, \sigma^2\right) = p\left(\mu|\sigma^2\right) p\left(\sigma^2\right)$$

- Standard icke-informativ prior är en uniform prior för $(\mu, \ln \sigma)$:

$$p\left(\mu, \sigma^2\right) \propto \frac{1}{\sigma^2}.$$

- Fördelen med standard icke-informativ prior:
  - betingade posteriorn $p\left(\mu|\sigma^2\right)$ och den marginella posteriorn $p\left(\sigma^2\right)$ följer kända fördelningar.
  - ger acceptabla resultat om man har mycket data.
- Nackdelar:
  - om man har lite data bör man specificera en rimligare prior (se oberoende priors i boken), eftersom priorn blir mer viktig vid lite data.

Normal data, okänd varians, linjär regressionsmodell

- Modell för linjär regression:

$$Y_1, ..., Y_n | \mu, \sigma^2, \boldsymbol{x} \overset{iid}{\sim} N(\mu, \sigma^2)$$

$$\mu = \boldsymbol{\beta} \boldsymbol{x}',$$

  där variansen σ^2 är okänd och vektor med förklaringsvariabler $\boldsymbol{x} = (1\, x_1\, \ldots\, x_k)$ och parametrar $\boldsymbol{\beta} = (\beta_0\, \beta_1\, \ldots\, \beta_k)$ i stället för endast $\mu = \beta_0$ i föregående modell.

- Om $\boldsymbol{\beta}$ och σ antas oberoende apriori (boken):

$$p(\boldsymbol{\beta}, \sigma) = \prod_{j=0}^{k} p(\beta_j)\, p(\sigma).$$

- Standard icke-informativ prior är här en uniform prior för parametrarna $(\boldsymbol{\beta}, \ln \sigma)$:

$$p(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{x}) \propto \frac{1}{\sigma^2}.$$

Normal data, okänd varians, linjär regressionsmodell

- Fördelen med den standard icke-informativa priorn $p(\boldsymbol{\beta}, \ln \sigma)$:
  - betingade posteriorn $p(\boldsymbol{\beta}|\sigma^2)$ och den marginella posteriorn $p(\sigma^2)$ följer även här kända fördelningar.
  - ger acceptabla resultat om man har mycket data jämfört med antalet förklaringsvariabler i $\boldsymbol{x}$.
- Nackdelar:
  - om man har lite data eller många förklaringsvariabler, så bör man specificera en rimligare prior (se oberoende priors i boken).

Normal data, okänd varians - exempel

- Ett slumpmässigt urval av 32 bilar har dragits från *1974 Motor Trend US magazine*, se datamaterialet *mtcars* från R:s dataexempel.
- Slutgiltigt mål: multipel linjär regressionsanalys med den beroende variabeln $y = \textit{miles/(US) gallon}$ för en bils bränsleförbrukning.
- 1 *miles/gallon* motsvarar ungefär 0,43 kilometer per liter. Transformera om y till kilometer per liter.
- Modell utan förklaringsvariabler:

$$Y_1, ..., Y_n | \mu, \sigma^2 \overset{iid}{\sim} N(\mu, \sigma^2)$$

$$\mu = \beta_0$$

där både medelvärdet μ och variansen σ^2 är okända.

- Oberoende priors för μ och σ (använd t.ex. webbverktyget för att elicitera priorn):

$$\mu \sim N\left(10, 10^2\right)$$

$$\sigma \sim \textit{Uniform}\left(0, 20\right)$$

Normal data, okänd varians - visualisera priors

- Plotta priors för μ och σ för att plotta dina antaganden apriori.
- Plotta data y för att se vad dina priors ger för apriori information om antal kilometer per liter.
  1. Dra många värden på μ och σ från priorfördelningarna.
  2. Dra betingade värden $y|\mu, \sigma^2$ givet värdena i punkt 1.
- Verkar dina priors rimliga? Om inte, ändra priors för μ och σ tills du blir nöjd.
- Kontakta experter, t.ex. bilhandlare, bilverkstäder, bilföreningar, etc., om du behöver hjälp med att elicitera en rimlig prior.

Normal data, okänd varians - kvadratisk approximation

- Kvadratisk approximation med funktionen **map** i R-paketet **rethinking**, se R-koden **Kod_Moment2.R**.
- Problem med kvadratisk approximation för σ, eftersom standardavvikelse- eller variansparametrar har en tendens att vara skeva åt höger.
- Lösning: kvadratisk approximation blir i bland bättre för $\ln \sigma$.
- Marginell posteriorfördelning för σ bestäms genom att antilogaritmera posteriorfördelningen för $\ln \sigma$:

$$\sigma_i = exp\left[(\ln \sigma)_i\right],$$

där i är den i:te samplade dragningen från respektive posteriorfördelning.

- Jämför posteriorfördelningarna för σ med respektive kvadratisk approximation för σ och $\ln \sigma$.
- Om $\ln \sigma | y_1, \ldots, y_n \sim N(\mu_n, \sigma_n)$, så följer posteriorfördelningen för σ en *log-normal* fördelning med parametrar μ_n och σ_n.

Normal data, okänd varians - posterior för nya obs.

- Posteriorn för $\hat{y}$ i linjär regression utan förklaringsvariabler är posteriorn för medelvärdet μ.
- Posterior prediktiv fördelning för nya observationer $\tilde{y}$ givet data $y_1, \ldots, y_n$, $p(\tilde{y}|y)$.
- Modellutvärdering med replikerade data (*in-sample fit*):
  - plotta $p(\tilde{y}|y)$ genom att dra värden från posteriorfördelningen $(\mu, \sigma)\,|y_1, \ldots, y_n$:

1. Dra många värden på (μ, σ) från posteriorfördelningen.
2. Dra nya observationer $\tilde{y}$ från $\tilde{y}|\mu, \sigma^2 \sim N(\mu, \sigma^2)$ givet (μ, σ) i punkt 1.

Normal data, okänd varians - posteriorn för μ, σ

- Även om μ och σ antas vara oberoende apriori, så tillåts dom vara beroende aposteriori.
- Funktion för kovariansen mellan μ och σ från posteriorn: *vcov* ().
- Korrelationsmatris: *cov2cor* ().
- Dra posteriorvärden för μ och σ direkt från multivariat normalfördelning (kvadratisk approximation):

$$mvrnorm\,(n = Nsamples,\ mu = coef\ (),\ Sigma = vcov\ ())$$

Normal data, okänd varians - uniform prior $p(\mu, \ln \sigma)$

- Modell:

$$Y_1, ..., Y_n | \mu, \sigma^2 \overset{iid}{\sim} N(\mu, \sigma^2)$$

- Prior:

$$p(\mu, \sigma^2) \propto (\sigma^2)^{-1}$$

$$p(\mu | \sigma^2) \propto c$$

$$p(\sigma^2) \propto (\sigma^2)^{-1}$$

- Betingad posterior för $\mu | \sigma^2$:

$$p(\mu | \sigma^2, y_1, ..., y_n) \quad \propto \quad p(y_1, ..., y_n | \mu, \sigma^2)\, p(\mu | \sigma^2)$$

$$\mu | \sigma^2, y_1, ..., y_n \sim N(\bar{y}, \sigma^2 / n)$$

Normal data, okänd varians - uniform prior

- Marginell posterior för σ^2:

$$p(\sigma^2|y_1,...,y_n) = \int p(\mu,\sigma^2|\,y_1,...,y_n)\,d\mu$$

$$= \int p(y_1,...,y_n|\mu,\sigma^2)\,p\left(\mu,\sigma^2\right)\,d\mu$$

$$\Longrightarrow \sigma^2|y_1,...,y_n \sim Inv-\chi^2(n-1,s^2),$$

  där s^2 är urvalsvariansen för data $y_1,\ldots,y_n$.

- $Inv-\chi^2\left(n-1,s^2\right)$ är en skalad invers χ^2-fördelning:

$$\frac{(n-1)\,s^2}{\sigma^2}\,|\,y_1,\ldots,y_n \sim \chi^2\,(n-1)\,.$$

Normal data, okänd varians - sampla från posteriorn

- Dra värden från posteriorfördelningen för σ^2:
  1. Dra ett värde $f\left(\sigma^2\right) = \frac{(n-1)s^2}{\sigma^2}$ från $\chi^2\left(n-1\right)$. (använd funktion rchisq i R)
  2. Beräkna det dragna värdet för σ^2, givet värdet $f\left(\sigma^2\right)$ i punkt 1, enligt:

$$\sigma^2 = \frac{\left(n-1\right)s^2}{f\left(\sigma^2\right)}$$

  3. Upprepa denna procedur många gånger för att få många dragna värden från posteriorfördelningen för σ^2.

- Dra värden från den betingade posteriorfördelningen $\mu|\sigma^2$:

$$\mu_i|\sigma_i^2, y_1, ..., y_n \sim N(\bar{y}, \sigma_i^2/n),$$

där i är den i:te samplade dragningen från respektive posteriorfördelning för σ^2 och $\mu|\sigma^2$.

Enkel linjär regression

- Enkel linjär regression:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

$$\epsilon_i \overset{iid}{\sim} N\left(0, \sigma^2\right)$$

- Modellen kan skrivas som:

$$Y_1, ..., Y_n | \mu_i, \sigma^2 \overset{iid}{\sim} N(\mu_i, \sigma^2)$$

$$\mu_i = \beta_0 + \beta_1 x_i,$$

där parametrarna β_0, β_1 och variansen σ^2 är okända.

- Prior:

$$p(\beta_0, \beta_1, \sigma)$$

- Om $(\beta_0, \beta_1, \sigma)$ antas oberoende apriori (boken):

$$p\left(\beta_0, \beta_1, \sigma\right) = p\left(\beta_0\right) p\left(\beta_1\right) p\left(\sigma\right).$$

Enkel linjär regression - exempel

- Enkel linjär regressionsanalys med den beroende variabeln $y = \textit{miles/(US) gallon}$ för en bils bränsleförbrukning och förklaringsvariabeln $x = \textit{vikt}$ i ton för en bil.
- 1 *miles/gallon* motsvarar ungefär 0,43 kilometer per liter. Transformera om y till kilometer per liter.
- 1 *pound* motsvarar ungefär 0,45 kilo. Transformera om x till ton.
- Modell:

$$Y_1, ..., Y_n | \mu_i, \sigma^2 \overset{iid}{\sim} N(\mu_i, \sigma^2)$$
$$\mu_i = \beta_0 + \beta_1 x_i$$

- Oberoende priors för $(\beta_0, \beta_1, \sigma)$:

$$\beta_0 \sim N\left(10, 10^2\right)$$
$$\beta_1 \sim N\left(-5, 5^2\right)$$
$$\ln \sigma \sim N\left(\ln 10, (\ln 2)^2\right)$$

Enkel linjär regression - posteriorresultat

- Kvadratisk approximation med funktionen **map** i R-paketet **rethinking**, se R-koden **Kod_Moment2.R**.
- Sammanfattning av posteriorn sker oftast genom att presentera tabeller och plottar över posteriorresultatet.
- Plottar av posteriorn ger oftast mer information om posteriorn än vad tabeller ger. All osäkerhet om olika kvantiteter i modellen kan plottas men inte återges i tabeller.
- Man kan ge mer viktning till tabeller när man blir mer van vid att tolka posteriorresultatet.
- Typisk tabell inkluderar posterior medelvärdet, standardavvikelsen och kredibilitetsintervall (t.ex. 90.9 % och 95.2 %).
- Oddset för positiv eller negativ lutning kan beräknas för lutningsparametern β_1.

Enkel linjär regression - posteriorresultat

- Hög korrelation mellan interceptet β_0 och lutningen β_1 ($r_{\beta_0,\beta_1} = -0.957$) kan medföra svårigheter att skatta modellen i mer komplicerade modeller med fler förklaringsvariabler $\Longrightarrow$ Centrera eller standardisera förklaringsvariablerna.
- Om inte möjliga värden skiljer sig avsevärt mellan förklaringsvariablerna, så räcker det med centrering $x_c = x - \bar{x}$. Annars är det bättre med standardisering $x_s = \frac{x-\bar{x}}{\sigma_X}$.
- Skattade modeller med standardiserade förklaringsvariabler kan vara mer svårtolkade, men man kan konvertera tillbaka till estimationsresultat på originalskala för förklaringsvariablerna.
- Fördelen med standardiserade förklaringsvariabler är att man kan jämföra magnituderna på lutningsparametrarna för förklaringsvariablerna, eftersom förklaringsvariablernas värden är standardiserade till samma skala. Viktigt om mycket blir signifikant pga mycket data.

Enkel linjär regression - posteriorresultat

- Plotta den skattade regressionslinjen från map (maximum a posteriori) skattningarna.
- Alla regressionslinjer från alla posteriordragningar för β_0 och β_1 kan plottas enligt:

$$\mu_{ij} = \beta_{0j} + \beta_{1j} x_i,$$

där i gäller för observation i och j är den j:te samplade dragningen från posteriorfördelningen för (β_0, β_1).

- Posteriorfördelningen för förväntad bensinförbrukning för en bil med vikt 1.5 ton (motsvarande vikt för x_0):

$$\mu_{1500,j} = \beta_{0j} + \beta_{1j} \cdot x_0,$$

där j är den j:te samplade dragningen från posteriorfördelningen för (β_0, β_1).

Enkel linjär regression - posteriorintervall för y och $\hat{y}$

- Posteriorfördelningen för förväntad bensinförbrukning för alla möjliga bilar som väger mellan 0.8 och 2.4 ton.
- Kredibilitetsintervall för $\hat{y}_i$ = kredibilitetsintervall för μ_i som funktion av olika bilvikter x i en figur.
- Prediktionsintervall för y_i som funktion av olika bilvikter x i en figur.

Bayesiansk förklaringsgrad

- Beräkna den klassiska förklaringsgraden för varje samplat värde s från posteriorn:

$$R_s^2 = \frac{SSR_s}{SST} = \frac{\sum_{i=1}^{n} \left(\hat{y}_{is} - \bar{y}\right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2}$$

- Gelman et al (2017) argumenterar för att en alternativ förklaringsgrad är bättre, eftersom förklaringsgraden ovan kan leda till $R^2 > 1$ vid lite data och informativa priorfördelningar, se följande dokument:

  http://www.stat.columbia.edu/∼gelman/research/unpublished/bayes_R2.pdf

- Alternativ Bayesiansk förklaringsgrad:

$$\textit{Bayesian } R_s^2 = \frac{SSR_s}{SSR_s + SSE_s} = \frac{\sum_{i=1}^{n} \left(\hat{y}_{is} - \bar{y}\right)^2}{\sum_{i=1}^{n} \left(\hat{y}_{is} - \bar{y}\right)^2 + \sum_{i=1}^{n} \left(y_i - \hat{y}_{is}\right)^2}$$

Bayesiansk residualanalys

- Ej standard med residualanalys i Bayesiansk statistik. Val av modell är subjektivt och det är mer vanligt att utvärdera konkurrerande modeller mot varandra utifrån två huvuddrag:
  1. hur bra är modellen på att replikera data (*in-sample fit*)
  2. hur bra prediktionsförmåga har modellen (*out-of sample fit*).
- Om residualanalys används, så kan den utvärderas på "vanligt" sätt. Obs! För varje observation i data har man en posteriorfördelning över residualen för denna observation: $r_{is} = y_i - \hat{y}_{is}$ för varje samplat värde s.
- Plotta residualerna mot μ för att undersöka om det är konstant variation σ kring μ. Undersök även här hur bra det linjära antagandet verkar vara för hur μ är länkat till förklaringsvariabler.
- Undersök om residualerna är normalfördelade med histogram.
- Undersök om residualerna verkar vara beroende av varandra över observationsordning.

Multipel linjär regression - utvärdera posteriorn för (β, σ)

- I stort sett alla posteriorutvärderingar kan göras för den multipla linjära regressionsmodellen som för den enkla linjära regressionsmodellen, dvs

  1. Tabeller över posteriorresultat för respektive parameters marginella posteriorfördelning, t.ex. medelvärde, standardavvikelse och kredibilitetsintervall.
  2. Visualisering av marginella posteriorfördelningen för varje lutningsparameter. Inte viktigt vid kvadratisk approximation eller i fall där posteriorn är lik en multivariat normalfördelning. Fokusera på vissa marginella posteriorfördelningar som visar något avvikande.
  3. Oddset för positiv eller negativ lutning kan beräknas för respektive lutningsparametern β_j till en förklaringsvariabel j.
  4. Visualisering av bivariata posteriorfördelningar kan vara intressant för att undersöka hur effekten från två förklaringsvariabler samvarierar. Contour plots är vanligt.
  5. Parametrarnas korrelationsmatris kan redovisas.

Multipel linjär regression - utvärdera posteriorn för $\tilde{y}$ och $\hat{y}$

- I stort sett alla posteriorutvärderingar kan göras för den multipla linjära regressionsmodellen som för den enkla linjära regressionsmodellen, dvs

  1. Posteriorfördelningen för μ kan redovisas för specifika värden på vektorn med förklaringsvariabler x.
  2. Prediktionsintervall och den prediktiva fördelningen för y kan redovisas för specifika värden på vektorn med förklaringsvariabler x.
  3. Replikering av data kan jämföras med faktiska data för modellutvärdering. (*in-sample fit*)
  4. Prediktioner för nya värden kan redovisas från den prediktiva fördelningen $p(\tilde{y}|y)$. Den prediktiva förmågan kan utvärderas mellan modeller utifrån olika prediktionsmått. (*out-of sample fit*)
  5. Och mycket annat...

Multipel linjär regression - exempel

- Multipel linjär regressionsanalys med
  - beroende variabel $y = \textit{miles}/(\textit{US})\,\textit{gallon}$ för en bils bränsleförbrukning
  - $x_1 =$ manuell växellåda (=1)
  - $x_2 =$ vikt i ton
  - $x_3 =$ antal hästkrafter
  - $x_4 =$ tid i sek på en kvarts mile
  - $x_5 =$ antal framåtväxlar
- 1 *miles/gallon* motsvarar ungefär 0,43 kilometer per liter. Transformera om y till kilometer per liter.
- 1 *pound* motsvarar ungefär 0,45 kilo. Transformera om x_1 till ton.
- Standardisera alla förklaringsvariabler förutom dummyvariabeln x_1.

Multipel linjär regression - exempel

- Modell:

$$Y_1, ..., Y_n | \mu_i, \sigma^2 \overset{iid}{\sim} N(\mu_i, \sigma^2)$$

$$\mu_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_1 x_{3i} + \beta_2 x_{4i} + \beta_1 x_{5i} = \boldsymbol{\beta} \boldsymbol{x}_i',$$

där variansen σ^2 är okänd och med vektorn av förklaringsvariabler som $\boldsymbol{x_i} = (1\, x_{1i}\, \ldots\, x_{ki})$ för observation i samt vektorn med parametrar $\boldsymbol{\beta} = (\beta_0\, \beta_1\, \ldots\, \beta_k)$.

- Om $\boldsymbol{\beta}$ och σ antas oberoende apriori (boken):

$$p(\boldsymbol{\beta}, \sigma) = \prod_{j=0}^{k} p(\beta_j)\, p(\sigma).$$

Multipel linjär regression - uniform prior $p(\boldsymbol{\beta}, \sigma^2)$

- Modell:

$$Y_1, ..., Y_n | \mu, \sigma^2 \overset{iid}{\sim} N(\mu, \sigma^2)$$
$$\mu = \boldsymbol{\beta} \boldsymbol{x}'$$

- Prior:

$$p(\boldsymbol{\beta}, \sigma^2) \propto (\sigma^2)^{-1}$$
$$p(\boldsymbol{\beta} | \sigma^2) \propto c$$
$$p(\sigma^2) \propto (\sigma^2)^{-1}$$

- Betingad posterior:

$$\boldsymbol{\beta} | \sigma^2, y, \boldsymbol{X} \sim N\left(\hat{\boldsymbol{\beta}}, V_\beta \sigma^2\right),$$

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}' y$$
$$V_\beta = (\boldsymbol{X}'\boldsymbol{X})^{-1},$$

där $\boldsymbol{X}$ är en $n \times k$ matris med förklaringsvariabler.

Multipel linjär regression - uniform prior $p(\boldsymbol{\beta}, \sigma^2)$

- Marginell posterior för $\sigma^2|\boldsymbol{x}, y$:

$$\sigma^2|y \sim Inv - \chi^2(n-k, s^2),$$

där $s^2 = \frac{(y-\boldsymbol{x}\boldsymbol{\beta}')'(y-\boldsymbol{x}\boldsymbol{\beta}')}{n-k}$

- $Inv - \chi^2\left(n-k, s^2\right)$ är en skalad invers χ^2−fördelning:

$$\frac{(n-k)\,s^2}{\sigma^2}\,|\,y \sim \chi^2\,(n-k)\,.$$

Multipel linjär regression - uniform prior: posterior samples

- Dra värden från posteriorfördelningen för σ^2:
  1. Dra ett värde $f\left(\sigma^2\right) = \frac{(n-k)s^2}{\sigma^2}$ från $\chi^2\left(n-k\right)$. (använd funktion rchisq i R)
  2. Beräkna det dragna värdet för σ^2, givet värdet $f\left(\sigma^2\right)$ i punkt 1, enligt:

$$\sigma^2 = \frac{\left(n-k\right)s^2}{f\left(\sigma^2\right)}$$

  3. Upprepa denna procedur många gånger för att få många dragna värden från posteriorfördelningen för σ^2.

- Dra värden från den betingade posteriorfördelningen $\boldsymbol{\beta}|\sigma^2, y, \boldsymbol{X}$:

$$\boldsymbol{\beta}_i|\sigma_i^2, y, \boldsymbol{X} \sim N\left(\hat{\boldsymbol{\beta}}, V_\beta\sigma_i^2\right),$$

där i är den i:te samplade dragningen från respektive posteriorfördelning för $\sigma^2|\boldsymbol{X}, y$ och $\boldsymbol{\beta}|\sigma^2, y, \boldsymbol{X}$.