4.3 Bayes Schätzer

Wir nehmen an, dass wir unsere Unsicherheit über unbekannte Parameter mit einer Wahrscheinlichkeitsdichte beschreiben können — die apriori Dichte

$$p(\underline{\theta})$$

Die apriori Dichte widerspiegelt unser Wissen/Glauben/ Vorurteile über unbekannte Parameter. Nach emnid war die Unterstützung der CSU in der Sonntagsfrage vom 20.4.2012 46%. Was sollte man als apriori Verteilung dieses Parameters für eine Umfrage morgen nehmen? Könnte es sein, dass Mitarbeiter von Herrn Seehofer und Herrn Pronold andere apriori Verteilungen vorschlagen würden?

Die Formel von Bayes wird angewendet, um die Resultate einer Stichprobe (neue Informationen) auszuwerten und eine aposteriori Dichte zu finden.

$$p(\underline{\theta}|\underline{x}) = \frac{f(x_1,\ldots,x_n|\underline{\theta})p(\underline{\theta})}{\int f(x_1,\ldots,x_n|\underline{\theta})p(\underline{\theta})\mathrm{d}\underline{\theta}}$$

$$= \frac{f(x_1,\ldots,x_n|\underline{\theta})p(\underline{\theta})}{f(\underline{x})}$$

$$= \frac{L_\theta(\underline{\theta}, \underline{x}) p(\underline{\theta})}{\int L_\theta(\underline{\theta}, \underline{x}) p(\underline{\theta}) \mathrm{d}\underline{\theta}}$$

Apriori Glauben $+$ Information $\rightarrow$ aposteriori Glauben

und für eine neue Stichprobe $\underline{y}$ gilt die Vorhersage

$$f(\underline{y}|\underline{x}) = \int f(\underline{y}|\underline{\theta}) p(\underline{\theta}|\underline{x}) \mathrm{d}\underline{\theta}$$

Falls neue Daten $\underline{y}$ beobachtet werden, würden wir die alte aposteriori Verteilung als apriori nehmen und eine aktualisierte aposteriori Verteilung berechnen.

$$p(\underline{\theta}|\underline{y}) \propto f(\underline{y}|\underline{\theta}) p(\underline{\theta}|\underline{x})$$

4.3.1 Bayes Beispiel mit der Binomial Verteilung

$$X \sim B(N, p)$$

Unter der Annahme der apriori Dichte

$$p \sim G(0, 1)$$

ist die aposteriori Dichte

$$f(p|x) = \frac{\binom{N}{x} p^x (1-p)^{N-x} * 1}{\int_0^1 \binom{N}{x} p^x (1-p)^{N-x} * 1 \mathrm{d}p}$$

Da

$$\int_0^1 p^x (1-p)^{N-x} * 1 \mathrm{d}p = \frac{\Gamma(x+1)\Gamma(N-x+1)}{\Gamma(N+2)}$$

haben wir

$$f(p|x) = (N+1)\binom{N}{x} p^x (1-p)^{N-x} \qquad 0 \leq p \leq 1$$

$$= \frac{\Gamma(N+2)}{\Gamma(x+1)\Gamma(N-x+1)} p^{(x+1)-1}(1-p)^{(N-x+1)-1}$$

eine Betaverteilung mit Parametern

$$a = x + 1 \quad b = N - x + 1$$

$$E[p|x] = \frac{a}{a+b} = \frac{x+1}{N+2}$$

Wenn wir diesen Wert als unseren Schätzer für p nehmen, haben wir

$$\hat{p} = \frac{x+1}{N+2} = \frac{2}{N+2}\frac{1}{2} + \frac{N}{N+2}\frac{x}{N}$$

eine gewichtete Mischung des apriori Erwartungswert $\frac{1}{2}$ und des ML Schätzers $\frac{x}{N}$.

In diesem Fall hatten wir
apriori $+$ Information $\rightarrow$ aposteriori

$$G(0,1) \quad + \quad B(N,p) \quad \rightarrow \quad Beta$$

aber die Gleichverteilung ist eine Betaverteilung mit $a = 1 \quad b = 1$. Im allgemeinen gilt

$$Beta(a,b) \quad + \quad B(N,p) \quad \rightarrow \quad Beta(a',b')$$

$$a' = a + x \qquad b' = b + N - x$$

$$\mu_{apriori} = \frac{a}{a+b}$$

$$\mu_{aposteriori} = \frac{a'}{a'+b'} = \frac{a+x}{a+b+N}$$

$$= \left(\frac{a+b}{a+b+N}\right)\left(\frac{a}{a+b}\right) + \left(\frac{N}{a+b+N}\right)\left(\frac{x}{N}\right)$$

eine gewichtete Summe von $\mu_{apriori}$ und dem ML Schätzer.

Die Betaverteilung ist eine konjugierte apriori Familie für die Binomialverteilung.

Statt den aposteriori Erwartungswert als Schätzer zu nehmen, könnten wir auch den Modalschätzer wählen, d.h. das Maximum aus der aposteriori Dichte. Um einen Intervallschätzer zu berechnen, benutzen wir die aposteriori Dichte und nehmen, aus verständlichen Gründen, das kürzeste 95% Intervall (was nicht immer leicht zu bestimmen sein mag):

$$P_{aposteriori}(c < \theta < d) = 0.95$$

$$\min_{c<\theta<d} f(\theta) \geq \max_{\theta \leq c, \theta \geq d} f(\theta)$$

4.3.2 Poisson Verteilung

$$Y \sim P(\theta)$$

$$p(y|\theta) \propto \theta^{\sum y_i} e^{-n\theta}$$

Eine konjugierte apriori Verteilung wäre dann

$$p(\theta) \propto \theta^{\alpha-1} e^{-\beta\theta}$$

eine $\Gamma(\alpha, \beta)$ Verteilung

$$\Rightarrow \theta|y \sim \Gamma(\alpha + n\bar{y}, \beta + n)$$

Für $n = 1$ haben wir ein interessantes Resultat:

$$p(y) = \frac{p(y|\theta)p(\theta)}{p(\theta|y)}$$

$$= \frac{e^{-\theta}\frac{\theta^y}{y!}\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}}{\frac{(\beta+1)^{\alpha+y}}{\Gamma(\alpha+y)}\theta^{\alpha+y-1}e^{-(\beta+1)\theta}}$$

$$= \frac{\Gamma(\alpha + y)}{\Gamma(\alpha)y!}\left(\frac{\beta}{(1+\beta)}\right)^\alpha \left(\frac{1}{1+\beta}\right)^y$$

$\Rightarrow$ Die Negativbinomial Verteilung kann man als eine Poisson Verteilung betrachten, wo der Parameter eine Γ apriori Dichte hat. (α ganzzahlig)

Verallgemeinerte Poissonverteilung

$$Y_i \sim P(x_i\theta) \qquad x_i \text{ bekannt}$$

z.B. Krankheitsraten in verschiedenen Gebieten $\{i\}$.

$$p(y|\theta) \propto \theta^{\sum y_i} e^{-\theta \sum x_i}$$

Nehmen wir als apriori Dichte, $p(\theta)$, $\Gamma(\alpha, \beta)$

$$\Rightarrow \theta | y \sim \Gamma(\alpha + \sum y_i, \beta + \sum x_i)$$

4.3.3 Bayes und die Normalverteilung

$$X \sim N(\mu, \sigma^2) \quad \sigma^2 \text{ bekannt}$$

Sei meine apriori Dichte für μ

$$\mu \sim N(\mu_0, \sigma_0^2) \quad \mu_0, \sigma_0^2 \text{ bekannt}$$

$$f(\mu|x_1, \ldots, x_n) =$$

$$\frac{(\frac{1}{\sigma\sqrt{2\pi}})^n e^{-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2} \frac{1}{\sigma_0\sqrt{2\pi}} e^{-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2}}{\int_{-\infty}^{\infty} (\text{Zähler}) \mathrm{d}\mu}$$

$$\propto e^{-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2 - \frac{1}{2\sigma_0^2}(\mu-\mu_0)^2}$$

$\Rightarrow$ die aposteriori Dichte von μ ist eine Normaldichte.

$$\mu_{aposteriori} \sim N(\mu_1, \sigma_1^2)$$

$$\sigma_1^2 = \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}}$$

$$\mu_1 = \sigma_1^2 * \left(\frac{\sum x_i}{\sigma^2} + \frac{\mu_0}{\sigma_0^2} \right)$$

Die Normalverteilung ist eine konjugierte Familie für sich selbst (σ^2 bekannt).

Für den Sonderfall $\sigma_0^2 = \sigma^2$ haben wir

$$\sigma_1^2 = \frac{\sigma^2}{n+1} \text{ und } \mu_1 = \frac{n\bar{x} + \mu_0}{n+1}$$

Der apriori Erwartungswert bekommt ein Gewicht von 1 und der Stichprobenmittelwert ein Gewicht von n.

Im allgemeinen Fall mit $\sigma_0^2 >> \frac{\sigma^2}{n}$ (d.h. große apriori Unsicherheit oder eine große Stichprobe):

$$\sigma_1^2 = \frac{\sigma^2}{n} \text{ und } \mu_1 = \bar{x}$$

$$\mu \sim N(\bar{x}, \frac{\sigma^2}{n})$$

$$P_\mu(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}) = 0.95$$

ein echtes 95% Wahrscheinlichkeitsintervall, kein KI.

Für $\sigma_0^2 << \frac{\sigma^2}{n}$ gilt

$$\mu \sim N(\mu_0, \sigma_0^2)$$

Beim festen Glauben, ändert neue Information nichts.

4.3.4 Schwierigkeiten

Was passiert, wenn wir auch σ^2 nicht kennen? Wie sollten wir eine gemeinsame apriori Dichte für μ und σ^2 wählen?

4.3.5 Nichtkonjugierte Verteilungen

z.B.

$$X \sim P(\lambda) \qquad \lambda \sim G(5, 10)$$

$$f(\lambda|x) = \frac{e^{-\lambda}\frac{\lambda^x}{x!}\frac{1}{5}}{\int\limits_5^{10} e^{-\lambda}\frac{\lambda^x}{x!}\frac{1}{5}\mathrm{d}\lambda} \qquad 5 \leq \lambda \leq 10$$

In diesem Fall geht es nur mit einer direkten Auswertung. In komplexeren Fällen ist die Auswertung nicht möglich. Wenn analytische und numerische Methoden nicht klappen, simuliert man mit klugen rechnerintensiven Methoden:

MCMC Monte Carlo Markov Chain.