Seminar in Statistik - FS 2008

Nonparametric Bayes

Handout verfasst von

Ivo Francioni und Philippe Muller

Zürich, 17. März 2008

1 Einleitung

Bis jetzt haben wir in der Bayes'schen Statistik immer mit einem endlichdimensionalen Parameterraum Θ gearbeitet. Nun aber betrachten wir nichtparametrische Probleme, d.h.

$$\Theta = \{\text{Wahrscheinlichkeitsmasse auf dem messbaren Raum } (H, \mathcal{A})\} =: \mathcal{P}.$$

Θ ist also unendlichdimensional.

Wir wollen nun in einem ersten Teil eine a priori Verteilung auf dem ∞-dimensionalen Raum $(\mathcal{P}, \mathcal{C})$ definieren, die sich besonders gut eignet um die a posteriori Verteilung für gegebene Daten zu bestimmen. Hierbei ist $\mathcal{C}$ die kleinste σ-Algebra, die von den Mengen der Form $\{P : P(A) < r\}$ erzeugt wird, wobei $A \in \mathcal{B}$ und $r \in [0, 1]$.

In einem zweiten Teil werden wir die Nützlichkeit dieser a priori Verteilung anhand von einigen Beispielen demonstrieren.

2 Der Dirichlet-Prozess, hilfreich bei nichtparametrischen Problemen

2.1 Random probability measures (RPM)

Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Durch eine $\mathbb{R}^k$-wertige ZV auf Ω wird ein eindeutiges Wahrscheinlichkeitsmass auf $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ festgelegt. Den gleichen Zusammenhang gibt es auch hier, in unserer Situation, zwischen einem Wahrscheinlichkeitsmass auf $(\mathcal{P}, \mathcal{C})$ und einer Mass-wertigen ZV'n. Diese masswertigen ZV'n nennen wir „random probability measures".
Unser Ziel ist es ein passendes RPM zu finden:

$$\begin{aligned} P:\ &\Omega \to \mathcal{P},\ \omega \mapsto P_\omega \\ \text{mit } P.(A):\ &\Omega \to \mathbb{R},\ \omega \mapsto P_\omega(A) \quad \mathcal{F}\text{-}\mathcal{B}(\mathbb{R})\text{-messbar } \forall A \in \mathcal{A}. \end{aligned}$$

2.2 Der Dirichlet-Prozess

Das für unsere Zwecke am besten geeignete RPM ist der Dirichlet-Prozess.

Definition 2.2.1 *Sei $(H, \mathcal{A})$ ein messbarer Raum und α ein endliches Mass auf $(H, \mathcal{A})$, das ungleich Null ist. Dann ist ein Dirichlet-Prozess P auf $(H, \mathcal{A})$ mit Parameter α ein RPM auf $(H, \mathcal{A})$ s.d. für alle Partitionen $H = \bigcup_{i=1}^k A_i$*

($A_i \in \mathcal{A}$) von H gilt: Der Zufallsvektor $\omega \mapsto (P_\omega(A_1), P_\omega(A_2), \ldots, P_\omega(A_k))$ hat Dirichlet-Verteilung mit Parameter $(\alpha(A_1), \alpha(A_2), \ldots, \alpha(A_k))$.

Definition 2.2.2 *Sei:*

- $(\alpha_1, \alpha_2, \ldots, \alpha_k)$ *ein Vektor mit* $\alpha_j \geq 0\ \forall j$ *und* $\sum \alpha_j > 0$.
- z_{α_j}, $j = 1, 2, \ldots, k$ *unabhängige* $\mathcal{G}(\alpha_j, 1)$*-verteilte ZV'n.*
- $z = \sum z_{\alpha_j}$ *und* $y_j = \frac{z_{\alpha_j}}{z}$ *für* $j = 1, 2, \ldots, k$.

Dann ist die k-dimensionale Dirichlet-Verteilung mit Parametern $\alpha_1, \alpha_2, \ldots, \alpha_k$ (Notation: $\mathcal{D}(\alpha_1, \alpha_2, \ldots, \alpha_k)$) definiert als die gemeinsame Verteilung des Zufallsvektors $(y_1, y_2, \ldots, y_k)$, der Werte im k-dimensionalen Simplex Δ_k annimmt.

Bemerkung 2.2.3 *Falls $X = (X_1, X_2, \ldots, X_k) \sim \mathcal{D}(\alpha_1, \ldots, \alpha_k)$, dann sind die Randverteilungen gegeben durch: $X_i \sim Beta(\alpha_i, \sum_{j=1}^k \alpha_j - \alpha_i)$.*

Der Dirichlet-Prozess gibt uns ein Wahrscheinlickeitsmass auf $(\mathcal{P}, \mathcal{C})$. Wir werden nun zwei mögliche Konstruktionen dieses Prozesses kurz erläutern. Sei dafür α ein endliches Mass auf $(H, \mathcal{A})$, das ungleich Null ist.

Erste Konstruktion: Axiomatisch (nach Thomas Ferguson)
Lege für jedes $n \in \mathbb{N}$ und jede Partition $A_1, A_2, \ldots, A_n$ von H die gemeinsame Verteilung $(P(A_1), P(A_2), \ldots, P(A_n)) \sim \mathcal{D}(\alpha(A_1), \alpha(A_2), \ldots, \alpha(A_n))$ fest. Mit dem Konsistenzsatz von Kolmogorov kann man so ein eindeutiges Wahrscheinlichkeitsmass auf $[0, 1]^{\mathcal{A}}$ bestimmen.

Zweite Konstruktion: „Stick-breaking Construction“ (nach Jayaram Sethuraman)
$\forall A \in \mathcal{A}$, setze $P(A) = \sum_{i=1}^\infty p_i \delta_{Y_i}(A)$, mit

$$Y_i : \Omega \to H \text{ iid } \sim \frac{\alpha(\cdot)}{\alpha(H)} \quad \text{und}$$

$$p_i = U_i \cdot \prod_{j=1}^{i-1} (1 - U_j),\ U_j \text{ iid } \sim Beta(1, \alpha(H)).$$

Man kann zeigen, dass das so konstruierte RPM ein Dirichlet-Prozess ist.

Jetzt kommen wir zum wichtigsten Resultat dieses Vortrags:

Satz 2.2.4 *Es sei $P \sim \mathcal{D}(\alpha)$ ein Dirichlet-Prozess, und $X_1, \ldots, X_n$ eine Folge von Zufallsvariablen, die gegeben P iid P verteilt sind. Dann ist die a posteriori Verteilung von P für gegebene Daten $(X_1, X_2, \ldots, X_n)$ wieder ein Dirichlet-Prozess, und zwar $\mathcal{D}(\alpha + \sum \delta_{X_i})$.*

Dieser Satz ist der Grund, warum Dirichlet-Prozesse so wichtig für die nichtparametrische Bayes-Statistik sind. Näheres dazu werden wir im nächsten Teil, wo wir ein paar Beispiele betrachten, erläutern.

3 Beispiele

In diesem Abschnitt werden wir zwei Beispiele von Bayes-Schätzern anschauen. Im Folgenden sei

$$\Theta = \{\text{Menge aller Wahrscheinlichkeitsverteilungen auf } (\mathbb{R}, \mathcal{B})\},$$

wobei $\mathcal{B}$ die Borel-σ-Algebra ist.

Als nützlich wird sich die Eigenschaft erweisen, dass mit einem Dirichlet-prior $\mathcal{D}(\alpha)$ die a posteriori Verteilung $\mathcal{D}(\alpha + \sum_i \delta_{X_i})$ ist. Dies erlaubt uns den Bayes-Schätzer für das „no sample“ Problem (also für a posteriori Verteilung = a priori Verteilung $\mathcal{D}(\alpha)$) zu berechnen und dann den Bayesschätzer durch ersetzen von α durch $\alpha + \sum_i \delta_{X_i}$ zu erhalten.

3.1 Verteilungsschätzer

Im ersten Beispiel werden wir die Verteilung selbst schätzen.
Es sei:

$$F(t) = P((-\infty, t]), \ t \in \mathbb{R}$$

für $P \in \Theta$. Wir wollen den Bayesschätzer für die a posteriori Verteilung $\mathcal{D}(\alpha)$ (no sample) bestimmen unter dem Verlust

$$L(F, \hat{F}) = \int\limits_{\mathbb{R}} (F(t) - \hat{F}(t))^2 dW(t),$$

wobei W ein endliches Mass auf $(\mathbb{R}, \mathcal{B})$ ist.
Der Bayesschätzer minimiert den erwarteten Verlust unter der a posteriori

Verteilung und daher ist

$$F_{Bayes} = \arg\min_{\hat{F}} E[\int (F(t) - \hat{F}(t))^2 dW(t)] \tag{1}$$

$$\overset{Tonelli}{=} \int \arg\min_{\hat{F}} \{E[(F(t) - \hat{F}(t))^2]\} dW(t). \tag{2}$$

Für festes t gilt also:

$$\begin{aligned} F_{Bayes}(t) &= \arg\min_{\hat{F}} E[(F(t) - \hat{F}(t))^2] \\ &= E[F(t)] \\ &= E[P((-\infty, t])]. \end{aligned}$$

Wir wissen das die Randverteilung eines k-dimensional Dirichlet-verteilten Zufallsvektors Beta-verteilt ist. In diesem Fall ist

$$P(-\infty, t]) \sim \text{Beta}(\alpha((-\infty, t]), \alpha((t, \infty)))$$

Also ist im no sample Fall der Bayesschätzer

$$\begin{aligned} F_{Bayes}(t) &= \frac{\alpha((-\infty, t])}{\alpha(\mathbb{R})} \\ &= F_0(t). \end{aligned}$$

Durch ersetzen der a priori Verteilung durch die a posteriori Verteilung $\mathcal{D}(\alpha + \sum_i \delta_{X_i})$ erhalten wir für die Realisierungen $X_1, \ldots, X_n$ den Bayesschätzer:

$$\begin{aligned} \hat{F}_n(t|X_1, \ldots, X_n) &= \frac{\alpha((-\infty, t]) + \frac{1}{n} \sum \delta_{X_i}((-\infty, t])}{\alpha(\mathbb{R}) + n} \\ &= p_n F_0(t) + (1 - p_n) F_n(t|X_1, \ldots, X_n), \end{aligned}$$

wobei $p_n = \frac{\alpha(\mathbb{R})}{\alpha(\mathbb{R})+n}$ und $F_n(t|X_1, \ldots, X_n)$ die empirische Verteilungsfunktion ist. p_n ist die Wahrscheinlichkeit, dass die a priori Dichte stimmt und man sieht, dass diese durch $\alpha(\mathbb{R})$ bestimmt wird.

3.2 Mittelwert

Ein weiteres Beispiel wäre ein Schätzer für der Mittelwert $\mu = \int x P(dx)$ für $P \in \Theta$. Für den Verlust

$$L(P, \hat{\mu}) = (\mu - \hat{\mu})^2$$

erhält man den Bayesschätzer

$$\hat{\mu}_n(X_1, \ldots, X_n) = p_n\mu_0 + (1 - p_n)\bar{X}_n,$$

wobei p_n wie oben, $\bar{X}_n$ der Mittelwert der Beobachtungen und μ_0 der a priori Erwartungswert ist, d.h.

$$\mu_0 = \int_{\mathbb{R}} x \frac{d\alpha(x)}{\alpha(\mathbb{R})}.$$

Literatur

[1] Thomas S. Ferguson. A Bayesian Analysis of some Nonparametric Problems. *The Annals of Statistics*, 1(2):209-230, 1973.

[2] Mark J. Schervish. *Theory of Statistics.* Springer-Verlag, N. Y., 1995.

[3] J. Sethuraman. A Constructive Definition of Dirichlet Priors. *Statistica Sinica*, 4:639-650, 1994.