Herleitung des Posteriors für einen Dirichlet-Prozess-Prior

Marco Frei

Diese Notizen halten sich notationell an den Vortrag von Ivo Francioni und Philippe Muller. Der Beweis orientiert sich an Schervish [1995, Kapitel 1.6, Theorem 1.94].

Es sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $(H, \mathcal{A})$ ein messbarer Raum und α ein endliches von Null verschiedenes Mass auf $(H, \mathcal{A})$. Es bezeichne fernerhin $\mathcal{P}$ die Menge aller Wahrscheinlichkeitsmasse auf $(H, \mathcal{A})$ und $\mathcal{C}$ sei die σ-Algebra auf $\mathcal{P}$, die von Mengen der Form $\{Q \in \mathcal{P} : Q(A_i) \leq t_i, i = 1, \ldots, n\}$ mit $H = \biguplus A_i, A_i \in \mathcal{A}$ und $t_i \in [0, 1]$ erzeugt wird. Schliesslich sei $\mathbf{P} : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathcal{P}, \mathcal{C})$ ein Dirichletprozess mit Parameter α (notiert als $\mathbf{P} \sim \mathcal{D}(\alpha)$) und $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (H, \mathcal{A})$ eine Zufallsvariable mit[1] $X|\mathbf{P} = Q \sim Q$. Dann gilt:

Theorem. $\mathbf{P}|X = x$ ist ein Dirichletprozess mit Parameter $\alpha + \delta_x$.

Beweis: Wir bezeichnen die Verteilung von $\mathbf{P}$ mit $\mu_{\mathbf{P}}$, die bedingte Verteilung $\mathbf{P}|X = x$ mit $\mu_{\mathbf{P}|X}(\cdot|x)$ und die Randverteilung von X mit μ_X. Es gilt $\forall A \in \mathcal{A}$:

$$\mu_X(A) = \mathbb{P}[X \in A] = E[\mathbb{P}[X \in A|\mathbf{P}]] = E[\mathbf{P}[A]] = \frac{\alpha(A)}{\alpha(H)}.$$

Die gemeinsame Verteilung von $(\mathbf{P}, X)$ lässt sich wie folgt faktorisieren:

$$\forall C \in \mathcal{C}, A \in \mathcal{A} : \quad \mathbb{P}[\mathbf{P} \in C, X \in A] = \int_A \mu_{\mathbf{P}|X}(C|x)\mu_X(dx).$$

Die Idee ist, diese Faktorisierung explizit durchzuführen; dann kann man $\mu_{\mathbf{P}|X}(\cdot|x)$ ablesen. Es reicht die Erzeuger von $\mathcal{C}$ zu betrachten, d.h. oBdA sei $C = \{Q \in \mathcal{P} : Q(A_1) \leq t_1, \ldots, Q(A_n) \leq t_n\}$. Zunächst verfeinern wir die Mengen $A_1, \ldots, A_n, A, A^c$ zu einer gemeinsamen ($2n$-elementigen) Partition von H. Zu diesem Zwecke definieren wir

$$A_i^0 = A_i \cap A, \quad A_i^1 = A_i \cap A^c.$$

Dann gilt offenbar

$$H = \biguplus_{i,j} A_i^j, \quad A = \biguplus_i A_i^0$$

und

$$\mathbf{P}(A) = \sum_i \mathbf{P}(A_i^0).$$

Wir definieren fernerhin

$$C_A = \{(z_1, z_2, \ldots, z_{2n}) : z_i + z_{n+i} \leq t_i, i = 1, \ldots, n\}.$$

Mit dieser Notation gilt sodann

$$Q \in C \Leftrightarrow \big(Q(A_1^0), \ldots, Q(A_n^1)\big) \in C_A.$$

Wir kürzen ausserdem ab:

$$\beta_1 = \alpha(A_1^0), \ldots, \beta_n = \alpha(A_n^0), \beta_{n+1} = \alpha(A_1^1), \ldots, \beta_{2n} = \alpha(A_n^1).$$

[1] Im folgenden verwenden wir für eine (generische) Realisierung von $\mathbf{P}$ jeweils den Buchstaben Q, d.h. für $Q \in \mathcal{P}$ bezeichnet $\{\mathbf{P} = Q\}$ das Ereignis $\{\omega \in \Omega : \mathbf{P}_\omega = Q\}$

Wir nehmen an, dass alle $\beta_i > 0$ sind; die folgenden Überlegungen bleiben jedoch mit kleinen Anpassungen auch richtig, falls gewisse β_i verschwinden. Wir beginnen mit der eigentlichen Faktorisierung:

$$\begin{aligned}
\mathbb{P}[\mathbf{P} \in C, X \in A] = E[\mathbb{P}[\mathbf{P} \in C, X \in A|\mathbf{P}]] = E[I_{\{\mathbf{P}\in C\}}\mathbf{P}(A)] \\
&= \int_C Q(A)\mu_{\mathbf{P}}(dQ) = \int_C Q\left(\biguplus_{j=1}^{n} A_j^0\right)\mu_{\mathbf{P}}(dQ) = \int_C \sum_{j=1}^{n} Q(A_j^0)\mu_{\mathbf{P}}(dQ) \\
&\overset{(i)}{=} \int_{C_A} \left(\sum_{j=1}^{n} z_j\right) \frac{\Gamma(\alpha(H))}{\prod_{i=1}^{2n}\Gamma(\beta_i)} \prod_{i=1}^{2n-1} z_i^{\beta_i - 1}\left(1 - \sum_{i=1}^{2n-1} z_i\right)^{\beta_{2n}-1} dz_1 dz_2 \dots dz_{2n-1} \qquad (1) \\
&\overset{(ii)}{=} \sum_{j=1}^{n} \frac{\beta_i}{\alpha(H)} \int_{C_A} \frac{\Gamma(\alpha(H)+1)}{\prod_{i=1}^{2n}\Gamma(\beta_i^j)} \prod_{i=1}^{2n-1} z_i^{\beta_i^j - 1}\left(1 - \sum_{i=1}^{2n-1} z_i\right)^{\beta_{2n}^j-1} dz_1 dz_2 \dots dz_{2n-1},
\end{aligned}$$

wobei $\beta_j^j = \beta_j + 1$ und $\beta_i^j = \beta_i$ für $i \neq j$. Gleichheit (i) benutzt, dass die endlichdimensionalen Verteilungen von $\mathbf{P}$ Dirichlet sind; bei (ii) erweitern wir mit $\alpha(H)$ und beherzigen für die Gammafunktion die funktionale Identität

$$\Gamma(z+1) = z \cdot \Gamma(z).$$

Wir bezeichnen im folgenden die Verteilung eines $\mathcal{D}(\alpha + \delta_x)$ Prozesses $\mathbf{P}^*$ mit $\nu(\cdot|x)$. Für $x \in A$ ist die gemeinsame Verteilung der $\mathbf{P}^*(A_i^j)$ eine Dirchlet-Verteilung mit Parametern $(\beta_1^j, \dots, \beta_{2n}^j)$, wobei j so gewählt ist, dass $x \in A_j^0$. Wir haben also nach Definition

$$x \in A_j^0 \Rightarrow \nu(C|x) = \int_{C_A} \frac{\Gamma(\alpha(H)+1)}{\prod_{i=1}^{2n}\Gamma(\beta_i^j)} \prod_{i=1}^{2n-1} z_i^{\beta_i^j - 1}\left(1 - \sum_{i=1}^{2n-1} z_i\right)^{\beta_{2n}^j-1} \prod_{i=1}^{2n-1} dz_i,$$

respektive für allgemeines $x \in H$ in Indikatornotation

$$\nu(C|x) = \sum_{j=1}^{n} I_{A_j^0}(x) \int_{C_A} \frac{\Gamma(\alpha(H)+1)}{\prod_{i=1}^{2n}\Gamma(\beta_i^j)} \prod_{i=1}^{2n-1} z_i^{\beta_i^j - 1}\left(1 - \sum_{i=1}^{2n-1} z_i\right)^{\beta_{2n}^j-1} \prod_{i=1}^{2n-1} dz_i.$$

Wir integrieren nun $\nu(C|x)$ über die Menge A bezüglich der Randverteilung μ_X und erhalten:

$$\begin{aligned}
\int_A \nu(C|x)\mu_X(dx) = \frac{1}{\alpha(H)} \int_A \nu(C|x)\alpha(dx) \\
&= \sum_{j=1}^{n} \frac{\beta_i}{\alpha(H)} \int_{C_A} \frac{\Gamma(\alpha(H)+1)}{\prod_{i=1}^{2n}\Gamma(\beta_i^j)} \prod_{i=1}^{2n-1} z_i^{\beta_i^j - 1}\left(1 - \sum_{i=1}^{2n-1} z_i\right)^{\beta_{2n}^j-1} \prod_{i=1}^{2n-1} dz_i. \qquad (2)
\end{aligned}$$

Ein Vergleich von (1) und (2) zeigt, dass

$$\mathbb{P}[\mathbf{P} \in C, X \in A] = \int_A \nu(C|x)\mu_X(dx),$$

und damit folgt

$$\mu_{\mathbf{P}|X}(\cdot|x) = \nu(\cdot|x) = \mathcal{D}(\alpha + \delta_x),$$

was zu zeigen war. □

Literatur

M. J. Schervish. *Theory of Statistics*. Springer-Verlag, N. Y., 1995.