Integration im TrueSkill Verfahren

Johannes Loevenich

7. Juli 2014

1 Problemstellung

1.1 Gaussian Density Filtering

Angenommen es sei ein Parameter Θ mit $P(\Theta) = N(\Theta; \mu; \Sigma)$ und eine Likelyhood Wahrscheinlichkeit $P(x|\Theta)$ gegeben. Bezeichne die Likelyhood Wahrscheinlichkeit die Funktion $t_x(\Theta)$, die nur vom Parameter Θ abhängt. Dann ist die Wahrscheinlichkeit $P(\Theta|x)$ nicht zwingend länger gaußverteilt,

$$P(\Theta|x) = \frac{t_x(\Theta)P(\Theta)}{\int t_x(\Theta')P(\Theta')d\Theta'} \tag{1.1}$$

Vom ADF wissen wir, dass wir diese Wahrscheinlichkeit mithilfe der Gaußverteilung $N(\Theta, \mu_x^{'}, \Sigma_x^{'})$ so approximieren können, dass die KL-Divergenz minimiert wird. Im Allgemeinen ergeben sich

$$\mu_x = \mu + \Sigma_x,\ \Sigma_x^{'} = \Sigma - \Sigma(g_x g_x^T - 2G_x)\Sigma, \tag{1.2}$$

wobei der Vektor g_x und die Matrix G_x durch

$$g_x := \frac{\partial log(Z_x(\mu^{'}, \Sigma^{'}))}{\partial \mu^{'}},\ G_x := \frac{\partial log(Z_x(\mu^{'}, \Sigma^{'}))}{\partial \Sigma^{'}} \tag{1.3}$$

gegeben sind.

1.2 Multidimensional korrigiert und abgeschnittene Gaußverteilung

Wir nennen x korrigiert und abgeschnitten gaußverteilt, wenn $\mathbf{x} \sim \mathcal{R}(\mathbf{x}; \mu, \Sigma^2, \alpha, \beta)$ und dies bedeutet, dass die Dichte von $\mathbf{x}$ durch gegeben ist. Es gibt keine effizienten analytischen Ausdrücke für die Normalisierungskonstante und jegliche Momente dieser Verteilung. Ist die Dimension von x jedoch nicht zu groß, so lassen sich die Normalisierungskonstante und Momente mithilfe des Genz Algorithmus approximieren.

1.3 Transformationstechniken und Genz Algorithmus

Im vorherigen Abschnitt haben wir gesehen, dass Integrale der Form

$$F(\alpha, \beta) = (2\pi)^{-\frac{n}{2}}|\Sigma|^{-\frac{1}{2}} \int_{\alpha_1}^{\beta_1} ... \int_{\alpha_n}^{\beta_n} exp(-\frac{1}{2}(\mathbf{x} - \mu)^T \mathbf{\Sigma}^{-1}(\mathbf{x} - \mu))g(\mathbf{x})dx \tag{1.4}$$

gelöst werden müssen. Auch wenn die in vorgestellten Integrationsmethoden in manchen Fällen zu einer Lösung dieses Integrals führen würden, wird die Konvergenz dieser Verfahren durch die hier vorgestellten Verfahren deutlich verbessert. Das Hauptaugenmerk soll dabei auf der Transformation auf den Einheitswürfel $(0, 1)^d$ liegen, um das so erhaltene Integral dann mithilfe von verschiedenen Integrationsmethoden leicht lösen zu können.

In einem ersten Schritt kann der Einfluss des Erwartungswertes μ durch die geschickte Substitution $\mathbf{y} = \mathbf{x} - \mu$ eliminiert werden,

$$F(\alpha, \beta) = (2\pi)^{-\frac{n}{2}}|\Sigma|^{-\frac{1}{2}} \int_{\alpha_1-\mu_1}^{\beta_1-\mu_1} ... \int_{\alpha_n-\mu_n}^{\beta_n-\mu_n} exp(-\frac{1}{2}\mathbf{y}^T \mathbf{\Sigma}^{-1}\mathbf{y})g(\mathbf{y} + \mu)dx \tag{1.5}$$

Sei nun $\mathbf{\Sigma} = \mathbf{LL^T}$ die Cholesky-Zerlegung der Kovarianzmatrix $\mathbf{\Sigma}$ und $\mathbf{L}$ dabei untere Dreiecksmatrix, dann schreibe $\mathbf{y} = \mathbf{Lz}$. Da $\mathbf{y}$ untere Dreicksmatrix ist, kann $\mathbf{z}$ iterativ durch

$$z_i = \frac{y_i - \sum_{j=1}^{i-1} L_{i,j}z_j}{L_{i,i}} \tag{1.6}$$

bestimmt werden. Durch die positiv Definitheit von Σ gilt stets $L_{i,i} > 0$ und z_i ist strikt monoton steigende Funktion in Abhängigkeit von y_i. Außerdem gilt $\mathbf{y^T\Sigma^{-1}y}$ $= \mathbf{z^TL^T(LL^T)^{-1}Lz} = \mathbf{z^Tz}$ und

$$d\mathbf{y} = |\mathbf{L}|d\mathbf{z} = |\mathbf{\Sigma}^{\frac{1}{2}}|d\mathbf{z}. \tag{1.7}$$

Mit diesen zwei Eigenschaften lässt sich (1.5) umformen zu

$$F(\alpha, \beta) = (2\pi)^{-\frac{n}{2}}|\Sigma|^{-\frac{1}{2}} \int_{\alpha'_1}^{\beta'_1} \mathcal{N}(z_1) \int_{\alpha'_n(z_1)}^{\beta'_n(z_1)} \mathcal{N}(z_2) ... \int_{\alpha'_n(z_1,...,z_n)}^{\beta'_n(z_1,...,z_n)} \mathcal{N}(z_n)g(\mathbf{y} + \mu)dx \tag{1.8}$$

, wobei Funktionen α' und β' durch

$$
\begin{aligned}
\alpha'_i(z_1, ..., z_{i-1}) &= \frac{\alpha_i - \mu_i - \sum_{j=1}^{i-1} L_{i,j} z_j}{L_{i,i}} \\
\beta'_i(z_1, ..., z_{i-1}) &= \frac{\beta_i - \mu_i - \sum_{j=1}^{i-1} L_{i,j} z_j}{L_{i,i}}
\end{aligned} \tag{1.9}
$$

Als nächster Schritt wird eine Transformation der einzelnen Koordination mit der inversen Normalverteilung $\Phi^{-1}(z_i) = v_i$ durchgeführt. Damit ergibt sich mit

$$
\begin{aligned}
\alpha''_i(v_1, ..., v_{i-1}) &= \Phi(\frac{\alpha_i - \mu_i - \sum_{j=1}^{i-1} L_{i,j} \Phi^{-1}(v_j)}{L_{i,i}}) \\
\beta''_i(v_1, ..., v_{i-1}) &= \Phi(\frac{\beta_i - \mu_i - \sum_{j=1}^{i-1} L_{i,j} \Phi^{-1}(v_j)}{L_{i,i}})
\end{aligned} \tag{1.10}
$$

für $F(\alpha, \beta)$ die Darstellung

$$
F(\alpha, \beta) = \int_{\alpha''_1}^{\beta''_1} \int_{\alpha'_2(v_1)}^{\beta'_2(v_1)} ... \int_{\alpha'_n(v_1,...,v_n)}^{\beta'_n(v_1,...,v_n)} g(\mathbf{L}[\Phi^{-1}(v_1), ..., \Phi^{-1}(v_n)] + \mu) d\mathbf{v} \tag{1.11}
$$

Zum Schluss führt eine Anwendung des Transformationssatz mit der linearen Transformation $v_i = \alpha''_i + w_i(\beta''_i - \alpha''_i)$ zu

$$
F(\alpha, \beta) = (\alpha''_1 \beta''_1) \int_0^1 ...(\alpha''_n \beta''_n) \int_0^1 g(\mathbf{L}[\Phi^{-1}(v_1), ..., \Phi^{-1}(v_n)] + \mu) d\mathbf{w} \tag{1.12}
$$

, wobei w_i uniform in $[0, 1]$ verteilt ist.

Wir wollen bemerken, dass F als die Erwartung der Funktion $g(\mathbf{y}(\mathbf{w}))$ interpretiert werden kann. Es ist klar, dass diese Erwartung invariant unter Permutation der Indizes von $\mathbf{w}$ ist. Diese Darstellung erlaubt die direkte Anwendung von mehrdimensionalen Integrationsmethoden. Auch ist an Gleichung (1.12) zu erkennen, dass sich das Integrationsproblem um eine Dimension reduziert hat, da die rechte Seite der Gleichung nicht von w_n abhängt und diese Variable daher herausintegriert werden kann.