

Column ～ 指数型分布族 ～

本章で取り上げる尤度関数、共役事前分布には一般形が存在する。まず、尤度関数については一般的に次のような形のものを想定すると都合がいい。この形で表される分布の族のことを、「指数型分布族」と呼ぶ。

$$p(\vec{x}|\vec{\eta}) = h(\vec{x}) \exp(\vec{\eta}^T \vec{t}(\vec{x}) - a(\vec{\eta}))$$

η は自然パラメータ、 $\vec{t}(\vec{x}_n)$ は十分統計量、 $h(\vec{x})$ は基底測度、 $a(\vec{\eta})$ は対数分配関数と呼ばれている。対数分配関数 $a(\vec{\eta})$ は $p(\vec{x}|\vec{\eta})$ を積分して 1 になるように保証してくれるもので、

$$\int h(\vec{x}) \exp(\vec{\eta}^T \vec{t}(\vec{x}) - a(\vec{\eta})) dx = 1$$

$$a(\vec{\eta}) = \ln \int h(\vec{x}) \exp(\vec{\eta}^T \vec{t}(\vec{x})) dx$$

この指数型分布族に対して都合のいい共役事前分布は次のようなものが知られている。

$$p_{\lambda}(\vec{\eta}) = h(\vec{\eta}) \exp(\vec{\eta}^T \vec{\lambda}_1 - a(\vec{\eta})\lambda_2 - a_c(\vec{\lambda}))$$

この共役事前分布を用いて事後分布を計算すると、

$$\begin{aligned} p(\vec{\eta}|X) &\propto p_{\lambda}(\vec{\eta}) \prod_{n=1}^N p(\vec{x}_n|\vec{\eta}) \\ &= h(\vec{\eta}) \exp(\vec{\eta}^T \vec{\lambda}_1 - a(\vec{\eta})\lambda_2 - a_c(\vec{\lambda})) \prod_{n=1}^N h(\vec{x}) \exp(\vec{\eta}^T \vec{t}(\vec{x}) - a(\vec{\eta})) \\ &= h(\vec{\eta}) \exp(\vec{\eta}^T \vec{\lambda}_1 - a(\vec{\eta})\lambda_2 - a_c(\vec{\lambda})) \left\{ \prod_{n=1}^N h(\vec{x}) \right\} \exp\left(\vec{\eta}^T \sum_{n=1}^N (\vec{t}(\vec{x}) - Na(\vec{\eta}))\right) \\ &\propto h(\vec{\eta}) \exp\left(\vec{\eta}^T \left(\vec{\lambda}_1 + \sum_{n=1}^N \vec{t}(\vec{x}_n)\right) - a(\vec{\eta})(\lambda_2 + N)\right) \end{aligned}$$

このように事後分布も事前分布と同じ形になる。事後分布のパラメータは、

$$\vec{\lambda}_1 = \vec{\lambda}_1 + \sum_{n=1}^N \vec{t}(\vec{x}_n), \quad \lambda_2 = \lambda_2 + N$$

予測分布についても同じように指数分布族で表すことができ、

$$\begin{aligned}
p(\vec{x}_*|X) &= \int p(\vec{x}_*|\vec{\eta})p(\vec{\eta}|X)d\vec{\eta} \\
&= \int d\vec{\eta} \{h(\vec{x}) \exp(\vec{\eta}^T \vec{t}(\vec{x}) - a(\vec{\eta}))h_c(\vec{\eta}) \exp(\vec{\eta}^T \vec{\lambda}_1 - a(\vec{\eta})\lambda_2 - a_c(\vec{\lambda}))\} \\
&= \dots\dots\dots = h(\vec{x}_*) \frac{\exp(a_c(\vec{\lambda}_1 + \vec{t}(\vec{x}_*), \lambda_2 + 1))}{\exp(a_c(\vec{\lambda}_1, \lambda_2))}
\end{aligned}$$

となり、一般的には指数型分布族にはならないことがわかる。ベルヌーイ分布、ガウス分布など、多くの分布が指数型分布族として表せることが知られており、この形で表すことができれば、あとは式 () や () に代入するだけで事後分布、予測分布を計算することができる。

また、計算の都合上、 $g(\vec{\eta}) = \exp(-a(\vec{\eta}))$ と置き、() の両辺に関して $\vec{\eta}$ について勾配を取ると、

$$\begin{aligned}
\nabla_{\vec{\eta}} \int h(\vec{x}) \exp(\vec{\eta}^T \vec{t}(\vec{x}) - a(\vec{\eta})) dx &= 0 \\
\nabla_{\vec{\eta}} g(\vec{\eta}) \int h(\vec{x}) \exp(\vec{\eta}^T \vec{t}(\vec{x})) dx \\
+ \int d\vec{x} \{h(\vec{x}) \exp(\vec{\eta}^T \vec{t}(\vec{x}))\vec{t}(\vec{x})\} &= 0 \\
-\frac{1}{g(\vec{\eta})} \nabla_{\vec{\eta}} = g(\vec{\eta}) \int d\vec{x} \{h(\vec{x}) \exp(\vec{\eta}^T \vec{t}(\vec{x}))\vec{t}(\vec{x})\} &= E[\vec{t}(\vec{x})]
\end{aligned}$$

よって次の結果が得られる。

$$-\nabla_{\vec{\eta}} \ln g(\vec{\eta}) = \nabla_{\vec{\eta}} a(\vec{\eta}) = E[\vec{t}(\vec{x})]$$

このことから、対数分配関数 $a(\vec{\eta})$ の $\vec{\eta}$ に関する勾配は十分統計量 $\vec{t}(\vec{x})$ の期待値になる。

また 2 階の偏微分は十分統計量の共分散になることもわかる。

$$\begin{aligned}
\frac{\partial^2 a(\vec{\eta})}{\partial \eta_i \partial \eta_j} &= E[\vec{t}_i(\vec{x})\vec{t}_j(\vec{x})] - E[\vec{t}_i(\vec{x})]E[\vec{t}_j(\vec{x})] \\
&= Cov[\vec{t}_i(\vec{x}), \vec{t}_j(\vec{x})]
\end{aligned}$$