ことわり；本レジュメの最後には Appendix の頁を付けている。

Abbreviation の約束

・r.v.；random value

・p.d.f.；probability density function

・c.d.f.；cumulative distribution function

1 確率的識別モデル

1.1 プロビット回帰

ここでは、2 クラス分類問題へのアプローチ手法の 1 つであるプロビット回帰について、最尤法によるパラメタ決定に至るまでの準備、もっと言うと尤度関数を書きくだせるまでの準備を行う。最終的に尤度関数は「自然科学の統計学 p.238 の（8.6)」のように確率関数を乗じた形で表せることがわかっており、このサブセクションでは、その 1 つ 1 つの確率関数を得る段階までを議論する[*1]。

まず、特徴量が与えられたときにクラス C_1 に分類される確率を求めたい。今、一般化線形モデルの議論をしていることもあり、特徴量 $\boldsymbol{\phi}$ を露わに書くことはしない。代わりに、活性化関数の引数である $a = \boldsymbol{w}^T\boldsymbol{\phi}$ を入力データとして意味を持たせ、これを露わに書くことにする。活性化関数は次のように定義する。

$$p(t = 1|a) = f(a) \tag{1}$$

式（4.112）のように、入力データ a が閾値 θ を超えたときクラス C_1 に分類することにすると、クラス C_1 への分類確率は次のように書ける。なお、閾値の r.v. を Θ で表す。

$$\begin{aligned} p(t = 1|a) &= P(\Theta \leq a) \\ &= F(a) \ (F \text{ は何某かの p.d.f. に対する c.d.f.}) \\ &= \int_{-\infty}^{a} p(\theta)d\theta \ (p \text{ は何某かの p.d.f.}) \end{aligned} \tag{2}$$

θ は r.v. Θ を確率関数で写像した先の実数値であることに注意されたい。式 1 により、この c.d.f. 表示が活性化関数になるということも言える。

今 p.d.f. が標準正規分布で与えられるとすると、上の c.d.f. は

$$\Phi(a) = \int_{-\infty}^{a} N(\theta|0, 1)d\theta \tag{3}$$

と書けて、これはプロビット関数の逆関数として知られている。この関数はシグモイド関数とよく似た形をしており、最尤法を適用する際[*2]に、どちらの関数を尤度関数に組み込むべきかがよく議論される[*3]。

[*1] しかし、この本では尤度方程式の議論をしないのに、ここで寸止めするのは違和感を感じました。むしろ、この後の 4.5.2 において、プロビット関数の逆関数がシグモイド関数に似ているという事実を使うために、その準備としてプロビット関数の紹介をしたかったのではないかと思っています。

[*2] 繰り返しますが、このレジュメでは尤度方程式などの議論までは踏み込みません。

[*3] 両者の違いについては「自然科学の統計学」レジュメ Chap.8 頁 3 の下半分あたりの記述を参照されたい（「probit か logit か？」の項)。

1.2　正準連結関数

ある条件を満たす連結関数[*4]を選ぶことで、誤差関数の勾配が「(誤差)×(特徴量ベクトル)」の形で得られることを示す。この形は今まで導いてきた形（例. ロジスティック回帰 ; (4.91)）を一般化してきたものになっていることも確かめる。

まず、誤差関数の勾配について考える。準備として、以下のように目的変数 t についての条件付き確率分布を指数型分布族の形式で与えておく（4.84 を利用)。

$$p(t|\eta, s) = \frac{1}{s}h(\frac{t}{s})g(\eta)\exp\frac{\eta t}{s} \tag{4}$$

次に、t の条件付き期待値を求める。この期待値の計算は既に p.113 の（2.226）で学んでいるが、もう一度計算方法のエッセンスだけ以下に整理しておく。

・式 4 の右辺にパラメタ η に関して nabla 演算子をかます。

・gradient 計算を行う（積の微分公式を用いることになる)。

・(gradient) = 0 としてやると、目的変数 t についての条件付き期待値 $E(t|\eta)$ が式中に表れる。

・$E(t|\eta)$ について等式変形してやれば（4.119）が得られる。後の議論のため、この期待値を y としておく。

(4.119）を見れば、y と η には関係があることが分かる。互いに陽に表すことは難しそうなので、ひとまず $\eta = \psi(y)$ と書いておく。この関係を後の式変形で用いる（式 6)。

さらなる準備として、上で定めた y が一般化線形モデルの形（4.120）になっていると約束しておく。この関係も後の式変形で用いる（式 6)。

この後、対数尤度関数の勾配を求めることによって、誤算関数の勾配を求める手続きを踏む。そのために目的変数 t についての対数尤度関数を準備しておく必要がある。目的変数のデータが N 個存在すると考えると t の対数尤度関数は

$$\begin{aligned}\ln p(\boldsymbol{t}|\eta, s) &= \Pi_{n=1}^{N} \ln[\frac{1}{s}h(\frac{t_n}{s})g(\eta_n)\exp(\frac{\eta_n t_n}{s})] \\ &= \sum \ln\frac{1}{s} + \sum \ln\frac{t_n}{s} + \sum \ln g(\eta_n) + \sum \ln\exp(\frac{\eta_n t_n}{s}) \\ &= \mathrm{const.} + \sum(\ln g(\eta_n) + \frac{\eta_n t_n}{s})\end{aligned} \tag{5}$$

の形で表現できる。ここで、スケールパラメタ s は全ての観測データで共通であると仮定している。このと

[*4] 活性化関数の逆関数のこと。事前に定義もなく突然出て来たことに驚きを隠せません。

き、式 5 に対してパラメタ $\boldsymbol{w}$ に関する勾配を取る。

$$
\begin{aligned}
\nabla_{\boldsymbol{w}} \ln p(\boldsymbol{t}|\eta, s) &= \nabla \sum (\ln g(\eta_n) + \frac{\eta_n t_n}{s}) \\
&= \sum (\frac{\partial}{\partial \boldsymbol{w}} \ln g(\eta_n) + \frac{\partial}{\partial \boldsymbol{w}} \frac{\eta_n t_n}{s}) \leftarrow \text{ナブラをパラメタによる微分で書き換えた。} \\
&= \sum (\frac{\partial}{\partial \eta_n} \frac{\partial \eta_n}{\partial \boldsymbol{w}} \ln g(\eta_n) + \frac{\partial}{\partial \eta_n} \frac{\partial \eta_n}{\partial \boldsymbol{w}} \frac{\eta_n t_n}{s}) \leftarrow \text{Chain rule} \\
&= \sum (\frac{\partial}{\partial \eta_n} \ln g(\eta_n) + \frac{\partial}{\partial \eta_n} \frac{\eta_n t_n}{s}) \frac{\partial \eta_n}{\partial \boldsymbol{w}} \\
&= \sum (\frac{\partial}{\partial \eta_n} \ln g(\eta_n) + \frac{t_n}{s}) \frac{\partial \eta_n}{\partial y_n} \frac{\partial y_n}{\partial \boldsymbol{w}} \leftarrow \text{Chain rule} \\
&= \sum (\frac{\partial}{\partial \eta_n} \ln g(\eta_n) + \frac{t_n}{s}) \frac{\partial \eta_n}{\partial y_n} \frac{\partial y_n}{\partial a_n} \frac{\partial a_n}{\partial \boldsymbol{w}} \leftarrow \text{Chain rule} \\
&= \sum (-\frac{y_n}{s} + \frac{t_n}{s}) \frac{\partial \eta_n}{\partial y_n} \frac{\partial y_n}{\partial a_n} \boldsymbol{\phi}_n \quad \because \text{(4.119)、(4.105)} \\
&= \sum \frac{1}{s} (t_n - y_n) \psi'(y_n) f'(a_n) \boldsymbol{\phi}_n \ (\because \eta = \psi(y), y_n = f(a_n))
\end{aligned} \tag{6}
$$

最後の等式のところで、先で言及した y と η の関係性や y が入力変数の線形結合による非線形関数で書けることを用いている。

ここで連結関数、すなわち活性化関数の逆関数を $f^{-1}(y) = \psi(y)$ のように取ると

$$
\begin{aligned}
f^{-1}(y) &= \psi(y) \\
f(\psi(y)) &= y \\
f'(\psi(y))\psi'(y) &= 1 \\
f'(a)\psi'(y) &= 1 \ (\because a = f^{-1}(y) = \psi(y))
\end{aligned} \tag{7}
$$

式 7 を式 6 に挿入すると、結局誤差関数の勾配は

$$
\sum \frac{1}{s}(t_n - y_n)\boldsymbol{\phi}_n \tag{8}
$$

という形になる。確かに、誤算関数の勾配が「(誤差)×(特徴量ベクトル)」の形で得られていることがわかる。尺度パラメタについて、$s = \beta^{-1}$ とすればガウス分布の誤差関数となり、$s = 1$ とすればロジスティックモデルの誤差関数となる。

2　ラプラス近似

事後分布がガウス分布でなくなるときに、パラメタ $\boldsymbol{w}$ 上で解析的に積分計算を実行するため、何らかの近似が必要となる。ここでは、その近似において広く用いられている手法であるラプラス近似の適用方法について学ぶ。ラプラス近似においては、真の分布の 2 階微分を求めることがポイントとなる[*5]。また、真の分布 $p(\boldsymbol{z})$ と近似分布 $q(\boldsymbol{z})$ の関係性について、視覚的なイメージを掴む。なお、ラプラス近似が対象とするのは確率密度分布が連続の場合であることは断っておく[*6]。

まず、簡単な場合、変数が 1 次の場合について、ラプラス近似のフレームワークを整理する。この近似手法は以下の手続きを踏むことになる。

[*5] 厳密には定数倍のズレがあります。

[*6] 関数が微分できることを前提とした近似手法であるため、連続であることは間違いなく保証されている必要があります。

・真の分布 $p(z)$ を次の形で仮定する。

$$p(z) = \frac{1}{Z} f(z) \tag{9}$$

・$p(z)$ のモードを見つける。つまり以下の関係式を満たす z_0 を見つける[*7]。

$$\frac{df(z)}{dz}|_{z=z_0} = 0 \tag{10}$$

・$\ln f(z)$ について $z = z_0$ 周りでテーラー展開する。z について 2 次の項まで残す。
・上記手続きにより $f(z) \propto$（ガウス分布）となるので、全積分が 1 になるように規格化すればよい。
・規格化することで求められた確率密度分布を $q(z)$ と表す。これが求めるべき近似分布となる。
上記手続きに従えば、近似分布を計算することができる。

ここで、真の分布と近似分布の関係性を視覚的に確かめる。図 4.14 左を見ると、確かに真の分布 p(z)（灰色）のモードが近似分布（点線）のモードと一致していることがわかる。

多変量の場合も同じフレームワークを辿ることで近似分布の表式（4.134）を導くことができる。先の例（1 変数の場合）と異なる点としては、精度が精度行列に変更されること、その行列の determinant を考える必要が出てくること、すなわち多変量解析で出てくる行列やベクトルを取り扱う必要があるというところである。

ラプラス近似においては、式 9 で定義した Z を計算しようとするステップが存在しない。これは次のサブセクションで取り組む「モデルの比較」によるアプローチと異なる点である。

なお、本手法の強み・弱みについては読み合わせとする。

2.1 モデルの比較と BIC

先のセクションでは、真の分布 $p(\boldsymbol{z})$ 全体を正規化ガウス分布で近似するという手法を学んだ。このサブセクションでは、真の分布 $p(\boldsymbol{z})$ の正規化係数 Z について（4.133）を用いて近似的な表式を得るというアプローチを取る。この近似された Z を用いることで、ラプラス近似のもとでの対数エビデンスモデルを求めることができる。この対数エビデンスモデルから、ベイズ情報量規準（BIC）が得られ、この BIC について 1 章で学んだ AIC との比較を行う。

まず、正規化係数 Z を計算する。式（4.133）を Z の定義 $Z = \int f(z)dz$ に代入してやればよい。多変量ガウス積分を実行することで Z は容易に計算できて、（4.135）が得られる。

この Z がなぜモデルエビデンスの議論と結びつくのかはよくわからん。

3 ベイズロジスティック回帰

ロジスティック回帰についてベイズ的取り扱いを学ぶ。ただし、厳密に解析的な計算を実行することは難しいため、事後確率分布と予測分布の両方に対して近似的な評価方法を得ることにする。

[*7] $p(z)$ のモードを中心とするガウス分布を探すことがラプラス近似の肝です。

3.1 ラプラス近似

ここでは事後確率分布の近似式を得る。近似手法としてはラプラス近似を用いることにする。今回は事後分布について 2 階微分の評価が必要になる[*8]。

本セクションで学ぶラプラス近似のフレームワークをセクション 4.4 で整理したフレームワークと比較を行う。そのために、4.4 で整理したフレームワークの要点をもう一度以下に載せる。

・真の分布のモードを見つける。

・$f(z)$ について 2 階微分をかます。

・$f(z) \propto$ ガウス分布 となるので、全積分が 1 になるように規格化する。

・規格化によって求めた確率密度分布を $q(z)$ とする。これが求めるべき近似分布となる。

次に、本セクションで学ぶフレームワークを以下で整理する。先のフレームワークと異なるところ（追加・変更点）については青字で表現する。

・真の分布が事後分布であることに注意する（追加)。

・準備として事前分布と尤度関数を定義し、ベイズの定理を用いて事後分布の表式を得る（追加)。

・真の分布のモードを見つける。

・事後確率（変更）[*9]について 2 階微分をかます。

・事後確率（変更）$\propto$ ガウス分布 となるので、全積分が 1 になるように規格化する。

・規格化によって求めた確率密度分布を $q(z)$ とする。これが求めるべき近似分布となる。

詳細な計算は省略するが、モードを求めた時点で近似分布（ガウス分布）の期待値が得られ、そして、事後確率の 2 階微分を求めた時点で近似分布の共分散行列が得られる。したがって式（4.144）が得られることになる。

次のサブセクションでは予測を行う。予測分布を求めるには、上で求めた事後確率分布（4.144）を用いてやればよい。

3.2 予測分布

ここでは、先のサブセクションで得た事後分布をもとに予測分布を求める。今回はクラス分類問題を考えていく。求めるべき予測分布は、新しい特徴量（データ）が与えられた際のクラス C_1 に対する予測分布である。

そのためには、まず予測分布の計算式（3.57）[*10]を用いて、パラメタ $\boldsymbol{w}$ について周辺化積分を実行してやればよい。なお、積分の実行にあたってはデルタ関数の取り扱いがポイントとなる[*11]。

パラメタについて積分をしてやると、クラス C_1 の予測分布について、ひとまず（4.151）まで辿り着ける。しかしながら、この形でもまだ解析的な計算は実行できない。そのためプロビット関数を用いて、さらなる近似を行い予測分布を解析的な形（4.155）まで持っていく。

ひとまず予測分布の周辺化積分の式を素直に書き下し、そのうえで真の分布を近似分布で置き換えておく。

$$p(C_1|\phi, \boldsymbol{t}) = \int p(C_1|\phi, \boldsymbol{w})p(\boldsymbol{w}|\boldsymbol{t})d\boldsymbol{w} \simeq \int \sigma(\boldsymbol{w}^T\boldsymbol{\phi})q(\boldsymbol{w}|\boldsymbol{t})d\boldsymbol{w} \tag{11}$$

[*8] セクション 4.4 ではベイズ的な考え方は持ち出して議論することはありませんでした。もっとも、密度関数の 2 階微分を考えるという点では同じですが。

[*9] 厳密に言えば事後確率の対数ですが。

[*10] p.155 参照。

[*11] 突然ディラックのデルタ関数を持ち出されても、初見の人には理解不能なのではないかと思います。付録に解説がついているわけでも無いですし。ずいぶんと不親切な説明だと思いました。

第二辺から第三辺への変形において、クラス C_1 の事後確率がシグモイド関数で表現できることを用いている[*12]。

ここで、以下のように $\sigma(\boldsymbol{w}^T\boldsymbol{\phi})$ をデルタ関数を用いた表現に書き換えることにする。

$$\sigma(\boldsymbol{w}^T\boldsymbol{\phi}) = \int \delta(a - \boldsymbol{w}^T\boldsymbol{\phi})\sigma(a)da \tag{12}$$

式 12 は $a = \boldsymbol{w}^T\boldsymbol{\phi}$ となるときのみ、積分値が $\sigma(\boldsymbol{w}^T\boldsymbol{\phi})$ として取り出されるということを意味している。これを式 11 に代入すると

$$\begin{aligned} p(C_1|\phi, \boldsymbol{t}) &\simeq \int_{\boldsymbol{w}}\int_a \delta(a - \boldsymbol{w}^T\phi)\sigma(a)daq(\boldsymbol{w})d\boldsymbol{w} \\ &= \int_a \sigma(a)\int_{\boldsymbol{w}} \delta(a - \boldsymbol{w}^T\phi)q(\boldsymbol{w})d\boldsymbol{w}da \\ &= \int_a \sigma(a)p(a)da \end{aligned} \tag{13}$$

ここで $p(a)$ は次のように定義している。

$$p(a) = \int \delta(a - \boldsymbol{w}^T\phi)q(\boldsymbol{w})d\boldsymbol{w} \tag{14}$$

式 13 の被積分関数について σ の部分はシグモイド関数であることはわかるが、一方で p(a) の形はわかっていない。そこで p(a) の形を考えてみたい。

p.85 の式（2.83）の結果（同時分布がガウス分布であるとき、その周辺分布もガウス分布となる）を使うと、$q(\boldsymbol{w})$ がガウス分布であることから、その周辺分布 $p(a)$ もガウス分布となると言える[*13]。したがって、期待値と分散を求めれば分布がユニークに定まることになる。

まず $p(a)$ の期待値 μ_a から計算する。

$$\begin{aligned} \mu_a &= \int_a p(a)ada \\ &= \int_a\int_{\boldsymbol{w}} \delta(a - \boldsymbol{w}^T\phi)q(\boldsymbol{w})ada \\ &= \int_{\boldsymbol{w}}\int_a \delta(a - \boldsymbol{w}^T\phi)adaq(\boldsymbol{w})d\boldsymbol{w} \\ &= \int_{\boldsymbol{w}} \boldsymbol{w}^T\phi q(\boldsymbol{w})d\boldsymbol{w} \ (\because a = \boldsymbol{w}^T\phi \text{のみ取り出し}) \\ &= \int q(\boldsymbol{w})\boldsymbol{w}^T d\boldsymbol{w}\boldsymbol{\phi} \\ &= \boldsymbol{w}_{MAP}\boldsymbol{\phi} \ (\because q(\boldsymbol{w}) \text{ は } \boldsymbol{w}_{MAP}\text{（Gauss 分布の期待値を与える）でモードをとる}) \end{aligned} \tag{15}$$

[*12] p.157 の式（4.57）を参照。

[*13] 非常に怪しい。なぜなら実際のところ、同時分布はガウス分布にデルタ関数が掛かった形になっており、純粋なガウス分布ではないからです。しかしながら、ひとまず正しいと仮定して議論を進めています。

次に $p(a)$ の分散 σ_a^2 を計算する。

$$
\begin{aligned}
\sigma_a^2 &= \int p(a)(a^2-\mu_a^2)da \\
&= \int_a \int_{\boldsymbol{w}} \delta(a-\boldsymbol{w}^T\phi)q(\boldsymbol{w})d\boldsymbol{w}(a^2-\mu_a^2)da \\
&= \int_{\boldsymbol{w}} q(\boldsymbol{w}) \int_a \delta(a-\boldsymbol{w}^T\phi)(a^2-\mu_a^2)dad\boldsymbol{w} \\
&= \int_{\boldsymbol{w}} q(\boldsymbol{w})((\boldsymbol{w}^T\phi)^2-\mu_a^2)d\boldsymbol{w} \ (\because a=\boldsymbol{w}^T\phi\text{のみ取り出し、}\int \delta(a-\boldsymbol{w}^T\phi)da=1) \\
&= \int q(\boldsymbol{w})((\boldsymbol{w}^T\phi)^2-(\boldsymbol{w}_{MAP}^T\phi)^2)d\boldsymbol{w} \ (\because (4.149)\text{ を代入}^{*14}\text{。}) \\
&= \int q(\boldsymbol{w})[(\boldsymbol{w}^T\boldsymbol{\phi})^T(\boldsymbol{w}^T\boldsymbol{\phi})-(\boldsymbol{w}_{MAP}^T\boldsymbol{\phi})^T(\boldsymbol{w}_{MAP}^T\boldsymbol{\phi})]d\boldsymbol{w} \\
&= \int q(\boldsymbol{w})(\boldsymbol{\phi}^T\boldsymbol{w}\boldsymbol{w}^T\boldsymbol{\phi}-\boldsymbol{\phi}^T\boldsymbol{w}_{MAP}\boldsymbol{w}_{MAP}^T\boldsymbol{\phi})d\boldsymbol{w} \\
&= \boldsymbol{\phi}^T\int q(\boldsymbol{w})(\boldsymbol{w}\boldsymbol{w}^T-\boldsymbol{w}_{MAP}\boldsymbol{w}_{MAP}^T)d\boldsymbol{w}\boldsymbol{\phi} \\
&= \boldsymbol{\phi}^T var(\boldsymbol{w})\boldsymbol{\phi} \\
&= \boldsymbol{\phi}^T\boldsymbol{S}_N\boldsymbol{\phi} \ (\because \boldsymbol{S}_N\text{は事後分布の共分散行列})
\end{aligned} \tag{16}
$$

以上により、$p(a)$ の期待値、分散について露わな形で表現することができたので、$p(a)$ はユニークに定まったと言える。ここで改めて、クラス C_1 の予測分布 $p(C_1|\boldsymbol{t})$ を書き下す。

$$
p(C_1|\boldsymbol{t}) = \int \sigma(a)p(a)da = \int \sigma(a)N(a|\mu_a,\sigma_a^2)da \tag{17}
$$

しかしながら、残念なことに被積分関数の形は定まったが、この積分そのものを実行することはできない。ここで、我々がとるアプローチは、ロジスティックシグモイド関数と「プロビット関数の逆関数」の類似性を用いるというものである。すなわち、式 17 に現れるシグモイド関数をそれによく似た関数である「プロビット関数の逆関数」で置き換えてしまおうという発想である*15。ただし、単純に置き換えるのではなく、スケールの調整は必要になることは注意されたい。

スケールの調整を行い*16、式 17 にてシグモイド→プロビット関数の逆関数への置き換えをしてやる。また、簡便な表記のため、サブスクリプトの a を省略する。なお、プロビット関数の逆関数は p.210 の式 (4.114) で定義されている。

$$
\begin{aligned}
p(C_1|\boldsymbol{t}) &= \int \Phi(\lambda a)N(a|\mu_a,\sigma_a^2)da \\
&=> \text{Go to App.1} \\
&= \Phi(\frac{\mu}{(\lambda^{-2}+\sigma^2)^{1/2}})
\end{aligned} \tag{18}
$$

*14 p.219 の式（4.150）を見ると $\boldsymbol{m}_N^T$ で表現しているのですが、この表現を使う動機が私にはわかりませんでした。ここまでの一連の議論を考えれば $\boldsymbol{w}_{MAP}$ を使う方が自然なので、このレジュメではそのようにさせてもらっています。確かに、p.151 の（3.49）とラプラス近似での議論を踏まえると、$\boldsymbol{m}^T$ と $\boldsymbol{w}_{MAP}$ が一致するだろうというのはわかるのですが...

*15 両者がどれほど類似しているかは p.195 の図 4.9 を参照して確かめましょう。

*16 申し訳ないですが、スケールパラメタ λ を見出す計算は省略します。

ここからはほとんど代入操作のみのため、ポイントだけかいつまんで説明する。まず、式 18 内のプロビット関数の逆関数をシグモイド関数に戻してやると式（4.153）が得られる。そして、再び期待値や分散のサブスクリプト a を復活させると式（4.155）を導くことができて、これがクラス C_1 の予測分布を近似的に表したものとなっている。

(4.152) 導出のフレームワーク

・ $a = \mu + \sigma z$ で変数変換（標準化による計算のシンプル化）

・(4.152) の各辺を別個に微分し、両者が一致することを示す。

・導関数について原始関数も一致することを示し、導出完了。 ←ダメでした。

$$(\text{LHS of } 4.152) = \int_a \int_{-\infty}^{\lambda a} N(\theta|0,1)\,d\theta\, N(a|\mu,\sigma^2)\,da \quad (\because (4.114))$$

$a = \mu + \sigma z$ とすると

$$\frac{da}{dz} = \sigma$$

よって

$$= \int_a \int_{-\infty}^{\lambda(\mu+\sigma z)} N(\theta|0,1)\,d\theta\, \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{z^2}{2}\right)\sigma\,dz$$

$$= \int_a \int_{-\infty}^{\lambda(\mu+\sigma z)} N(\theta|0,1)\,d\theta\, \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)dz \quad \cdots ①$$

これを μ で微分する。

$$\frac{d}{d\mu}① = \int_a \frac{d}{d\mu}\int_{-\infty}^{\lambda(\mu+\sigma z)} N(\theta|0,1)\,d\theta\, \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)dz$$

$$= \int_a \lambda N(\lambda(\mu+\sigma z)|0,1)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)dz$$

（∵微積分学の基本定理）

$$= \frac{\lambda}{\sqrt{2\pi}}\int \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(\lambda(\mu+\sigma z))^2}{2}\right\}\exp\left(-\frac{z^2}{2}\right)dz$$

$$= \frac{\lambda}{2\pi}\int \exp\left(-\frac{\lambda^2(\mu^2+2\mu\sigma z+\sigma^2 z^2)+z^2}{2}\right)dz$$

$$= \frac{\lambda}{2\pi} \int \exp\left(-\frac{1+\lambda^2\sigma^2}{2} \left(z + \frac{\lambda^2 \mu \sigma}{1+\lambda^2\sigma^2} \right)^2 \right) \exp\left(\frac{-\lambda^2\mu^2}{2(1+\lambda^2\sigma^2)} \right) dz$$

$$= \frac{\lambda}{2\pi} \exp\left(\frac{-\lambda^2\mu^2}{2(1+\lambda^2\sigma^2)} \right) \sqrt{\frac{\pi}{\frac{1+\lambda^2\sigma^2}{2}}}$$

$$= \frac{\lambda}{\sqrt{2\pi}} \sqrt{\frac{1}{1+\lambda^2\sigma^2}} \exp\left(\frac{-\lambda^2\mu^2}{2(1+\lambda^2\sigma^2)} \right)$$

$$= \frac{\lambda}{\sqrt{2\pi}} \frac{1}{\lambda} \sqrt{\frac{1}{\lambda^{-2}+\sigma^2}} \exp\left(-\frac{1}{2} \left(\frac{\mu^2}{\lambda^{-2}+\sigma^2} \right) \right)$$

$$= \frac{1}{\sqrt{2\pi}} \frac{1}{(\lambda^{-2}+\sigma^2)^{1/2}} \exp\left(-\frac{1}{2} \left(\frac{\mu}{(\lambda^{-2}+\sigma^2)^{1/2}} \right)^2 \right) \quad \cdots \text{①}$$

次に (4.152) の右辺を μ で微分する。この導関数が ① と等しくなればよい。

$$\frac{d}{d\mu}(\text{RHS of } 4.152) = \frac{d}{d\mu} \Phi\left(\frac{\mu}{(\lambda^{-2}+\sigma^2)^{1/2}} \right)$$

$$= \frac{d}{d\mu} \int_{-\infty}^{\frac{\mu}{(\lambda^{-2}+\sigma^2)^{1/2}}} N(\theta \mid 0, 1)\, d\theta \quad (\because (4.114))$$

$$= \frac{1}{(\lambda^{-2}+\sigma^2)^{1/2}} N\left(\frac{\mu}{(\lambda^{-2}+\sigma^2)^{1/2}} \,\middle|\, 0, 1 \right)$$

($\because$ 基本定理)

$$= \frac{1}{(\lambda^{-2}+\sigma^2)^{1/2}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{\mu}{(\lambda^{-2}+\sigma^2)^{1/2}} \right)^2 \right)$$

$$= \text{①}$$

よって、(4.152) の各辺の導関数が一致することは示された。

原始関数の一致性はわからなかった…

> Back to 頁 8