5.2.3.2 の再パラメータ化勾配について、確認してみる。

文献 [60] の 2.4 を見ると、

$$q_\phi(z|x)\prod_i dz_i = p(\epsilon)\prod d\epsilon_i \tag{1}$$

となっている。

これを前提とすると、以下のようになる。(5.38) より、

$$\begin{aligned}I(\xi) &= \nabla_\xi \int f(w)q(w;\xi)dw = \nabla_\xi \int f(w)q(\epsilon)d\epsilon = \nabla_\xi \int f(g(\xi;\epsilon))q(\epsilon)d\epsilon = \int \nabla_\xi (f(g(\xi;\epsilon))q(\epsilon))d\epsilon \\ &= \int q(\epsilon)\nabla_\xi f(g(\xi;\epsilon))d\epsilon = \int q(\epsilon)f'(g(\xi;\epsilon))\nabla_\xi g(\xi;\epsilon)d\epsilon = \mathbb{E}_{q(\epsilon)}[f'(g(\xi;\epsilon))\nabla_\xi g(\xi;\epsilon)]\end{aligned} \tag{2}$$

となり、(5.42) が求まる。

(1) について、詳細を見てみる。

一般的な重積分の変数変換を考えると以下のようになる。

$$\begin{aligned}\int\int\cdots\int f(x_1,x_2,\cdots,x_n)dx_1dx_2\cdots dx_n &= \int\int\cdots\int f(x_1,x_2,\cdots,x_n)\prod_i dx_i \\ &= \int\int\cdots\int g(u_1,u_2,\cdots,u_n)\left|\frac{\partial(x_1,x_2,\cdots,x_n)}{\partial(u_1,u_2,\cdots,u_n)}\right|\prod_i du_i \\ &= \int\int\cdots\int g(u_1,u_2,\cdots,u_n)\left|\frac{\partial(x_1,x_2,\cdots,x_n)}{\partial(u_1,u_2,\cdots,u_n)}\right| du_1du_2\cdots du_n\end{aligned} \tag{3}$$

ただし、

$$g(u_1,u_2,\cdots,u_n) \equiv f(x_1(u_1,u_2,\cdots,u_n),x_2(u_1,u_2,\cdots,u_n),\cdots,x_n(u_1,u_2,\cdots,u_n)) \tag{4}$$

これについては例えば、https://www.math.is.tohoku.ac.jp/ junya/lecture/calculus/change-of-variables.pdf や https://www2.math.kyushu-u.ac.jp/ hara/lectures/06/zoku15-060511.pdf に記載がある。

また、(3.11) を、上記のヤコビ行列の行列式の記載方法を参考に記載すると確率密度関数の変数変換は以下のようになる。

$$p_y(y) = p_x(x)\left|\frac{\partial(x_1,x_2,\cdots,x_n)}{\partial(u_1,u_2,\cdots,u_n)}\right| \tag{5}$$

さて、これらを踏まえて、再度、$I(\xi)$ を検討する。

$$I(\xi) = \nabla_\xi \int f(w)q(w;\xi)dw = \nabla_\xi \int f(g(\xi;\epsilon))q(g(\xi;\epsilon);\xi)\left|\frac{\partial(w_1,w_2,\cdots,w_n)}{\partial(\xi_1,\xi_2,\cdots,\xi_n)}\right|d\epsilon = \nabla_\xi \int f(g(\xi;\epsilon))q(\epsilon)d\epsilon \tag{6}$$

最初の等号で重積分の変数変換、2 番目の等号で、確率密度関数の変数変換を行っている。

よって、(1) のように考えられることがわかる。

さて、(1) を踏まえて、(5.45) を考えてみる。

$$\begin{aligned}\frac{\partial}{\partial\mu}\int f(w)q(w;\xi)dw &= \frac{\partial}{\partial\mu}\int f(g(\xi;\epsilon))q(\epsilon)d\epsilon = \int q(\epsilon)\frac{\partial}{\partial\mu}f(g(\xi;\epsilon))d\epsilon = \int q(\epsilon)f'(g(\xi;\epsilon))\frac{\partial}{\partial\mu}g(\xi;\epsilon)d\epsilon \\ &= \int q(\epsilon)f'(g(\xi;\epsilon))d\epsilon = \int f'(g(\xi;\epsilon))q(\epsilon)d\epsilon = \int f'(w)q(w;\xi)dw\end{aligned} \tag{7}$$

つまり、

$$I(\mu) = \frac{\partial}{\partial \mu} \int f(w) q(w;\xi) dw = \int f'(w) q(w;\xi) dw = \mathbb{E}_{q(w;\xi)}[f'(w)] \tag{8}$$

となり、(5.47) が求まる。

なお、(5.44) より、

$$w \equiv g(\xi;\epsilon) = \mu + \sigma\epsilon \tag{9}$$

となり、

$$\frac{\partial w}{\partial \mu} = \frac{\partial}{\partial \mu} g(\xi;\epsilon) = 1 \tag{10}$$

同様に (5.46) を考える。

まず、

$$\frac{\partial w}{\partial \sigma} = \frac{\partial}{\partial \sigma} g(\xi;\epsilon) = \epsilon \tag{11}$$

それを踏まえると、

$$\begin{aligned}\frac{\partial}{\partial \sigma} \int f(w) q(w;\xi) dw = \frac{\partial}{\partial \sigma} \int f(g(\xi;\epsilon)) q(\epsilon) d\epsilon = \int q(\epsilon) \frac{\partial}{\partial \sigma} f(g(\xi;\epsilon)) d\epsilon = \int q(\epsilon) f'(g(\xi;\epsilon)) \frac{\partial}{\partial \sigma} g(\xi;\epsilon) d\epsilon \\ = \int q(\epsilon) f'(g(\xi;\epsilon)) \epsilon d\epsilon = \int f'(g(\xi;\epsilon)) \epsilon q(\epsilon) d\epsilon = \int f'(w) \frac{(w-\mu)}{\sigma} q(w;\xi) dw\end{aligned} \tag{12}$$

つまり、

$$I(\sigma) = \frac{\partial}{\partial \sigma} \int f(w) q(w;\xi) dw = \int f'(w) \frac{(w-\mu)}{\sigma} q(w;\xi) dw = \mathbb{E}_{q(w;\xi)}[f'(w) \frac{(w-\mu)}{\sigma}] \tag{13}$$

となり、(5.48) が求まる。

(5.42) は、式を求めているが (5.45)-(5.48) とはあまり関係していない気がする。