第11章　正規分布の確率集中不等式

11.1　正規分布の部分積分公式と補間公式

$X \sim \mathsf{N}(0, \sigma^2)$ $(0 < \sigma < \infty)$ とする．すると X の p.d.f.

$$\mathsf{p}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \qquad (x \in \mathbb{R})$$

をもち

$$x\mathsf{p}(x) = -\sigma^2 \overset{\bullet}{\mathsf{p}}(x), \qquad \overset{\bullet}{\mathsf{p}}(x) = \frac{\mathrm{d}\mathsf{p}}{\mathrm{d}x}(x)$$

なる関係式が成立することがわかる．このことをから, 以下の部分積分の公式が得られる．

定理 11.1. $h : \mathbb{R} \longrightarrow \mathbb{R}$ は連続微分可能な関数で, $\mathsf{E}[|Xh(X)|] < \infty$ と $\mathsf{E}\big[|\overset{\bullet}{h}(X)|\big] < \infty$ をみたすとする．このとき

$$\mathsf{E}[Xh(X)] = \mathsf{E}[X^2]\mathsf{E}\big[\overset{\bullet}{h}(X)\big] \tag{11.1}$$

が成立する．

注意 11.2 ([24] から借用)**.** $X \sim \mathsf{N}(0, 1)$ とし, $h : \mathbb{R} \longrightarrow \mathbb{R}$ は 1 回連続微分[1] とする．．このとき, $\mathsf{E}\big[|\overset{\bullet}{h}(X)|\big] < \infty$ ならば, $\mathsf{E}[|Xh(X)|] < \infty$ かつ $\mathsf{E}\big[|h(X)|\big] < \infty$ となることがわかる．実際, Fubini の定理を用いると

$$\begin{aligned}
\int_0^\infty |xh(x)| e^{-x^2/2}\,\mathrm{d}x &= \int_0^\infty \left| x \int_0^x \overset{\bullet}{h}(y)\,\mathrm{d}y \right| e^{-x^2/2}\,\mathrm{d}x \\
&\leq \int_0^\infty x \left\{ \int_0^x |\overset{\bullet}{h}(y)|\,\mathrm{d}y \right\} e^{-x^2/2}\,\mathrm{d}x \\
&= \int_0^\infty |\overset{\bullet}{h}(y)| \underbrace{\left\{ \int_y^\infty x e^{-x^2/2}\,\mathrm{d}x \right\}}_{=e^{-y^2/2}} \mathrm{d}y \qquad (\because \text{Fubini}) \\
&= \int_0^\infty |\overset{\bullet}{h}(y)| e^{-y^2/2}\,\mathrm{d}y
\end{aligned}$$

[1] 実は, 絶対連続を仮定するだけでよい．

より

$$\mathsf{E}\big[|Xh(X)|\big] \le \mathsf{E}\big[|\overset{\bullet}{h}(X)|\big] < \infty$$

がわかる．さらに

$$\begin{aligned}|h(x)| &= \big|h(x)\mathbb{1}_{[0,1]}(x)\big| + \big|h(x)\mathbb{1}_{(1,\infty)}(x)\big| \\ &\le \sup_{|x|\le 1}|h(x)| + \big|xh(x)\mathbb{1}_{(1,\infty)}(x)\big| \\ &\le \sup_{|x|\le 1}|h(x)| + \big|xh(x)\big| =: C + +\big|xh(x)\big|\end{aligned}$$

となる．ただし, C はある定数である．よって

$$\mathsf{E}\big[|h(X)|\big] \le C + \mathsf{E}\big[|Xh(X)|\big] < \infty$$

がわかる．

また, $\mathsf{E}\big[|\overset{\bullet}{h}(X)|\big] < \infty$ もわかるので

$$\lim_{|x|\to\infty} h(x)e^{-x^2/2}\,\mathrm{d}x = 0$$

が成立することを証明できる．詳しくは, [32, pp.284-285] を参照のこと．
□

定理 11.1 は多変量正規分布に対する拡張はできる．

$$\boldsymbol{X} = (X_1,\, X_2,\, \ldots,\, X_d)^\top \sim \mathsf{N}_d(\mathbf{0}_d,\, \Sigma)$$

とする．ただし, $\mathbf{0}_d = (\underbrace{0,\, 0, \ldots,\, 0}_{d\text{ 個}})^\top$ で, Σ は $d\times d$ の正定値対称行列である．すなわち, $\Sigma = \Sigma^\top$ で, 任意の $\boldsymbol{a} \ne \mathbf{0}_d$ に対して, $\boldsymbol{a}^\top\Sigma\boldsymbol{a} > 0$ が成立する．さらに, $h_j : \mathbb{R}^n \longrightarrow \mathbb{R}\,(j = 1,\, 2,\, \ldots,\, d)$ は連続微分可能な関数とし, $\boldsymbol{h} = (h_1,\, h_2,\, \ldots,\, h_d)^\top$ と書くことにする．

定理 11.3. $\boldsymbol{h}$ はゆるい成長条件をみたすと仮定する．このとき

$$\mathsf{E}\big[X_1\boldsymbol{h}(\boldsymbol{X})\big] = \sum_{j=1}^{d} \mathsf{E}[X_1X_j]\mathsf{E}\left[\frac{\partial \boldsymbol{h}}{\partial x_j}(\boldsymbol{X})\right] \tag{11.2}$$

が成立する．ただし

$$\begin{aligned}\mathsf{E}\big[X_1\boldsymbol{h}(\boldsymbol{X})\big] &= \big(\mathsf{E}\big[Xh_1(\boldsymbol{X})\big],\, \mathsf{E}\big[Xh_2(\boldsymbol{X})\big],\, \ldots,\, \mathsf{E}\big[Xh_d(\boldsymbol{X})\big]\big)^\top, \\ \mathsf{E}\left[\frac{\partial \boldsymbol{h}}{\partial x_j}(\boldsymbol{X})\right] &= \left(\mathsf{E}\left[\frac{\partial h_1}{\partial x_j}(\boldsymbol{X})\right],\, \mathsf{E}\left[\frac{\partial h_2}{\partial x_j}(\boldsymbol{X})\right],\, \ldots,\, \mathsf{E}\left[\frac{\partial h_3}{\partial x_j}(\boldsymbol{X})\right]\right)^\top\end{aligned}$$

である．等式 (11.2) を多変量正規分布の部分積分公式という．

Proof. $\sigma_1^2 = \mathsf{E}[X_1^2]$ とする．このとき, 確率ベクトル $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_d)^\top$ の各成分を

$$Y_j = X_j - \lambda_j X_1, \qquad \lambda_j = \frac{\mathsf{E}[X_j X_1]}{\sigma_1^2},$$

で定まる．すると $\boldsymbol{Y}$ と X_1 は独立となる．実際, $j = 1, 2, \ldots, d$ に対して

$$\begin{aligned}\mathsf{Cov}[Y_j, X_1] &= \mathsf{Cov}[X_j, X_1] - \lambda_j \mathsf{Cov}[X_1, X_1] \\ &= \mathsf{E}[X_j X_1] - \frac{\mathsf{E}[X_j X_1]}{\sigma_1^2} \times \underbrace{\mathsf{E}[X_1^2]}_{=\sigma_1^2} = 0\end{aligned}$$

からわかる．よって, 正規分布の性質から $\boldsymbol{Y}$ と X_1 は独立である．

$\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_d)^\top$ とおくと

$$\boldsymbol{X} = \boldsymbol{Y} + X_1 \boldsymbol{\lambda}$$

と書ける．$\mathsf{E}_1[\,\cdot\,] := \mathsf{E}[\,\cdot\,|\boldsymbol{Y}]$ と表記する．(11.1) から

$$\begin{aligned}\mathsf{E}_1[X_1 h_j(\boldsymbol{X})] &= \mathsf{E}_1[X_1 h_j(\boldsymbol{Y} + X_1\boldsymbol{\lambda})] &\quad (11.3)\\ &= \sigma_1^2 \mathsf{E}\left[\frac{\partial h_j}{\partial t}(\boldsymbol{Y} + t\boldsymbol{\lambda})\bigg|_{t=X_1}\right] \qquad (j = 1, 2, \ldots, d) &\quad (11.4)\end{aligned}$$

となる．ただし, すべての $\boldsymbol{Y}$ に対して

$$\lim_{t\to\pm\infty} h(\boldsymbol{Y} + t\boldsymbol{\lambda}) = 0$$

をみたし, (11.4) の両辺の期待値は有限の値を持つと仮定する．よって

$$\mathsf{E}_1[X_1 \boldsymbol{h}(\boldsymbol{X})] = \sigma_1^2 \mathsf{E}\left[\frac{\partial \boldsymbol{h}}{\partial t}(\boldsymbol{Y} + t\boldsymbol{\lambda})\bigg|_{t=X_1}\right] \tag{11.5}$$

である．さらに

$$X_1 \boldsymbol{h}(\boldsymbol{X}) \qquad と \qquad \frac{\partial \boldsymbol{h}}{\partial t}(\boldsymbol{Y} + t\boldsymbol{\lambda})\bigg|_{t=X_1}$$

は絶対可積分としたとき, (11.5) の両辺の期待値を $\boldsymbol{Y}$ に関して取り, Fubini の定理を用いると

$$\mathsf{E}[X_1 \boldsymbol{h}(\boldsymbol{X})] = \sigma_1^2 \mathsf{E}\left[\frac{\partial \boldsymbol{h}}{\partial t}(\boldsymbol{Y} + t\boldsymbol{\lambda})\bigg|_{t=X_1}\right] \tag{11.6}$$

を得る．ここで, $k = 1, 2, \ldots, d$ に対して

$$\frac{\partial h_k}{\partial t}(\boldsymbol{Y} + t\boldsymbol{\lambda})\bigg|_{t=X_1} = \sum_{j=1}^{d} \lambda_j \frac{\partial h_k}{\partial x_j}(\boldsymbol{Y} + X_1\boldsymbol{\lambda}) = \sum_{j=1}^{d} \lambda_j \frac{\partial h_k}{\partial x_j}(\boldsymbol{X})$$

となることに注意すると

$$\mathsf{E}[X_1\boldsymbol{h}(\boldsymbol{X})] = \sum_{j=1}^{d} \mathsf{E}[X_1X_j]\mathsf{E}\left[\frac{\partial \boldsymbol{h}}{\partial x_j}(\boldsymbol{X})\right]$$

を得る. □

注意 11.4. 成長条件についてのコメントを加筆すること. □

次に, 中心化された多変量正規分布に従う独立な 2 つの確率ベクトルの線型補間を考えるために, 確率ベクトルを

$$\boldsymbol{X} = (X_1, X_2, \ldots, X_d)^\top, \qquad \boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_d)^\top, \qquad \boldsymbol{X} \text{ と } \boldsymbol{Y} \text{ は独立}$$

と書こう. さらに

$$\sigma_{ij} := \mathsf{E}[X_iX_j], \qquad \psi_{ij} := \mathsf{E}[Y_iY_j] \qquad (i, j = 1, 2, \ldots, d)$$

とする. このとき, $0 \leq t \leq 1$ に対して

$$\boldsymbol{Z}(t) := \sqrt{t}\boldsymbol{X} + \sqrt{1-t}\boldsymbol{Y} \tag{11.7}$$

と定める. $\boldsymbol{\Sigma} := (\sigma_{ij})$, $\boldsymbol{\Psi} = (\psi_{ij})$ とおくと

$$\boldsymbol{Z}(t) \sim \mathsf{N}_d\big(\boldsymbol{0}_d, t\boldsymbol{\Sigma} + (1-t)\boldsymbol{\Psi}\big) \qquad (0 \leq t \leq 1)$$

となる. $\boldsymbol{Z}(t)$ を $\boldsymbol{X}$ と $\boldsymbol{Y}$ の線型補間という.

$j = 1, 2, \ldots, d$ に対して, $h_j : \mathbb{R}^d \longrightarrow \mathbb{R}$ を 2 回連続微分可能な関数とし, $\boldsymbol{h} = (h_1, h_2, \ldots, h_d)^\top$ とおく. $\boldsymbol{h}$ の期待値を

$$\boldsymbol{g}(t) := \mathsf{E}[\boldsymbol{h}(\boldsymbol{Z}(t))] \qquad (0 \leq t \leq 1)$$

と書くことにする. 定め方から

$$\boldsymbol{g}(1) = \mathsf{E}[\boldsymbol{h}(\boldsymbol{Y})], \qquad \boldsymbol{g}(0) = \mathsf{E}[\boldsymbol{h}(\boldsymbol{X})]$$

となる. $\boldsymbol{h}$ の各成分 $h_k\,(k = 1, 2, \ldots, d)$ の 1 次導関数は指数成長条件をみたすとき, 微分記号と積分記号の交換ができることに注意する. $\boldsymbol{g} = (g_1, g_2, \ldots, g_d)^\top$ と書いたとき, $k = 1, 2, \ldots, d$ に対して

$$\begin{aligned}\dot{g}_k(t) &:= \frac{\mathrm{d}}{\mathrm{d}t}g_k(t) = \frac{\mathrm{d}}{\mathrm{d}t}\mathsf{E}\big[h_k(\boldsymbol{Z}(t))\big] = \mathsf{E}\left[\frac{\mathrm{d}}{\mathrm{d}t}h_k(\boldsymbol{Z}(t))\right] \\ &= \mathsf{E}\left[\sum_{j=1}^{d}\frac{\mathrm{d}h_k}{\mathrm{d}x_j}(\boldsymbol{Z}(t))\frac{\mathrm{d}}{\mathrm{d}t}Z_j(t)\right] =: \mathsf{E}\left[\sum_{j=1}^{d}\frac{\mathrm{d}h_k}{\mathrm{d}x_j}(\boldsymbol{Z}(t))\,\dot{Z}_j(t)\right]\end{aligned}$$

となる. ただし

$$\overset{\bullet}{Z}_j(t) = \frac{\mathrm{d}}{\mathrm{d}t} Z_j(t) = \frac{1}{2\sqrt{t}} X_j - \frac{1}{2\sqrt{1-t}} Y_j$$

である. $\overset{\bullet}{Z}_j(t)$ と $Z_i(t)\,(i = 1, 2, \ldots, d)$ の共分散は

$$\begin{aligned}
\mathsf{E}\big[\overset{\bullet}{Z}_j(t) Z_i(t)\big] &= \mathsf{E}\left[\left(\frac{1}{2\sqrt{t}} X_j - \frac{1}{2\sqrt{1-t}} Y_j\right)(\sqrt{t} X_i + \sqrt{1-t} Y_i)\right] \\
&= \mathsf{E}\left[\frac{1}{2} X_i X_j - \frac{\sqrt{t}}{2\sqrt{1-t}} X_j Y_j + \frac{\sqrt{1-t}}{2\sqrt{t}} X_j Y_i - \frac{1}{2} Y_i Y_j\right] \\
&= \frac{1}{2}\Big\{\mathsf{E}[X_i X_j] - \mathsf{E}[Y_i Y_j]\Big\} \\
&= \frac{1}{2}(\sigma_{ij} - \psi_{ij})
\end{aligned}$$

となる. $\forall 1 \le i, j \le d$ に対して, ある $c_1 > 0, c_2 > 0$ が存在して

$$\left|\frac{\partial \boldsymbol{h}}{\partial x_i}(\boldsymbol{x})\right| \le c_1 e^{c_2 |\boldsymbol{x}|_{2,d}} \tag{11.8}$$

と

$$\left|\frac{\partial \boldsymbol{h}^2}{\partial x_i \partial x_j}(\boldsymbol{x})\right| \le c_1 e^{c_2 |\boldsymbol{x}|_{2,d}} \qquad \text{または} \qquad \sigma_{i,j} = \psi_{i,j} \tag{11.9}$$

のいずれかが成立すると仮定すると, 正規分布の部分積分公式 (定理 11.3) を用いると $k = 1, 2, \ldots, d$ に対して

$$\begin{aligned}
\mathsf{E}\left[\frac{\partial h_k}{\partial x_j}(\boldsymbol{Z}(t))\, \overset{\bullet}{Z}_j(t)\right] &= \sum_{i=1}^{d} \mathsf{E}\big[Z_i(t)\, \overset{\bullet}{Z}_j(t)\big] \mathsf{E}\left[\frac{\partial^2 h_k}{\partial x_i \partial x_j}(\boldsymbol{Z}(t))\right] \\
&= \frac{1}{2}\sum_{i=1}^{d} (\sigma_{ij} - \psi_{ij}) \mathsf{E}\left[\frac{\partial^2 h_k}{\partial x_i \partial x_j}(\boldsymbol{Z}(t))\right]
\end{aligned}$$

を得る. 1 から d まで j をたし合わせると

$$\overset{\bullet}{g}_k(t) = \frac{\mathrm{d}}{\mathrm{d}t} \mathsf{E}\big[h_k(\boldsymbol{Z}(t))\big] = \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \big(\sigma_{ij} - \psi_{ij}\big) \mathsf{E}\left[\frac{\partial^2 h_k}{\partial x_i \partial x_j}(\boldsymbol{Z}(t))\right] \tag{11.10}$$

を得る. これをベクトルで表現すると

$$\overset{\bullet}{\boldsymbol{g}}(t) = \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \big(\sigma_{ij} - \psi_{ij}\big) \mathsf{E}\left[\frac{\partial^2 \boldsymbol{h}}{\partial x_i \partial x_j}(\boldsymbol{Z}(t))\right]$$

となる.

関数 $h\colon \mathbb{R} \longrightarrow \mathbb{R}$ は 2 回連続微分可能で, ある種の成長条件をみたすとする. $Z \sim \mathsf{N}(0,\,1)$ のとき

$$g(t,\,x) := \mathsf{E}\big[h(x+\sqrt{t}Z\big]$$

とおくと

$$\begin{aligned}
\frac{\partial g}{\partial t} &= \mathsf{E}\left[\frac{\partial}{\partial t}h(x+\sqrt{t}X)\right] \\
&= \mathsf{E}\left[\frac{Z}{2\sqrt{t}}\,\dot{h}\,(x+\sqrt{t}Z)\right] \\
&= \frac{1}{2\sqrt{t}}\int_{-\infty}^{\infty} z\,\dot{h}\,(x+\sqrt{t}z)\frac{1}{\sqrt{2\pi}}e^{-z^2/2}\,\mathrm{d}z \\
&= \frac{1}{2\sqrt{t}}\left[-\,\dot{h}\,(x+\sqrt{t}z)\frac{1}{\sqrt{2\pi}}e^{-z^2/2}\right]_{-\infty}^{\infty} \\
&\qquad + \frac{1}{2\sqrt{t}}\int_{-\infty}^{\infty}\frac{\partial}{\partial z}\,\dot{h}\,(x+\sqrt{t}z)\frac{1}{\sqrt{2\pi}}e^{-z^2/2}\,\mathrm{d}z \\
&= \frac{1}{2}\mathsf{E}\big[\ddot{h}\,(x+\sqrt{t}Z)\big] \qquad (11.11)
\end{aligned}$$

となる. 一方

$$\begin{aligned}
\frac{\partial g}{\partial x^2} &= \frac{\partial^2}{\partial x^2}\mathsf{E}\big[h(x+\sqrt{t}Z)\big] = \mathsf{E}\left[\frac{\partial^2}{\partial x^2}h(x+\sqrt{t}Z)\right] \\
&= \mathsf{E}\left[\ddot{h}\,(x+\sqrt{t}Z)\right] \qquad (11.12)
\end{aligned}$$

となる. (11.11) と (11.12) を合わせると

$$\frac{\partial g}{\partial t}(t,\,x) = \frac{1}{2}\frac{\partial g}{\partial x^2}(t,\,x) \qquad (11.13)$$

を得る. 境界条件 $g(0,\,x) = h(x)$ を与えると熱方程式 (11.13) を解くことができる.

11.2 ガウス確率集中不等式

Lipschitz 関数 $h : \mathbb{R}^d \longrightarrow \mathbb{R}$ を考える. すなわち, ある $L > 0$ に対して

$$\big|h(\boldsymbol{x}) - h(\boldsymbol{y})\big| \leq L\big|\boldsymbol{x}-\boldsymbol{y}\big|_{2,\,d} \qquad (\forall \boldsymbol{x},\,\boldsymbol{y} \in \mathbb{R}^d) \qquad (11.14)$$

をみたす. このような L のうちの最小のものを関数 h の Lipschitz 半ノルム $\|h\|_{\mathrm{Lip}}$ と呼ぶことにする.

定理 11.5. $Z_1, Z_2, \ldots, Z_d \overset{\text{i.i.d.}}{\sim} \mathsf{N}(0, 1)$ とし, $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_d)^\top$ とおく. h を $\mathbb{R}^d$ の Lipschitz 関数とする. このとき, 任意の $t \geq 0$ に対して

$$\mathsf{Pr}\Bigg(h(\boldsymbol{Z}) - \mathsf{E}[h(\boldsymbol{Z})] \geq t\Bigg) \leq \exp\Bigg(-\frac{t^2}{4\|h\|_{\mathrm{Lip}}^2}\Bigg)$$

が成り立つ.

Proof. **I h は微分可能で勾配が有限の場合の証明**. h は微分可能で, その勾配が $L > 0$ なる上限を持つと仮定する. すなわち

$$\big|\nabla h\big|_{2,\,d} \leq L, \qquad \nabla = \left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \ldots, \frac{\partial}{\partial x_d}\right)^\top, \qquad \boldsymbol{x} = (x_1, x_2, \ldots, x_d)^\top$$

である. $\lambda \geq 0$ を取り, $\boldsymbol{Z}_1$, $\boldsymbol{Z}_2$ を $\boldsymbol{Z}$ の独立複製とし, Gauss 補間

$$g(t) = \mathsf{E}\Bigg[\exp\Bigg\{\lambda\Bigg(h(\sqrt{t}\boldsymbol{Z}_1 + \sqrt{1-t}\boldsymbol{Z}) - h(\sqrt{t}\boldsymbol{Z}_2 + \sqrt{1-t}\boldsymbol{Z})\Bigg)\Bigg\}\Bigg] \tag{11.15}$$

を考える.
① $g(1) \leq e^{\lambda^2 L^2}$ **の証明**. 上の $g(t)$ を (11.7) の形に直す. 関数 $G : \mathbb{R}^{2d} \longrightarrow \mathbb{R}$ を

$$\begin{aligned}&G(x_1, x_2, \ldots, x_d, x_{d+1}, \ldots, x_{2d})\\ &\quad = \exp\Bigg\{\lambda\Bigg(h(x_1, x_2, \ldots, x_n) - h(x_{d+1}, x_{d+2}, \ldots, x_{2d})\Bigg)\Bigg\}\end{aligned}$$

と定め

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{Z}_1 \\ \boldsymbol{Z}_2 \end{pmatrix}, \qquad \boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Z} \\ \boldsymbol{Z} \end{pmatrix}$$

とおく. このとき, (11.7) の形の Gauss 補間は

$$g(t) = \mathsf{E}\big[G\big(\sqrt{t}\boldsymbol{X} + \sqrt{1-t}\boldsymbol{Y}\big)\big]$$

と書き直すことができる. さらに

$$\boldsymbol{\Sigma} := \mathsf{Cov}[\boldsymbol{X}] = \begin{bmatrix} \mathbf{I}_d & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_d \end{bmatrix}, \qquad \boldsymbol{\Psi} := \mathsf{Cov}[\boldsymbol{Y}] = \begin{bmatrix} \mathbf{I}_d & \mathbf{I}_d \\ \mathbf{I}_d & \mathbf{I}_d \end{bmatrix}$$

となることに注意する. したがって

$$\boldsymbol{\Sigma} - \boldsymbol{\Psi} = \begin{bmatrix} \mathbf{0} & -\mathbf{I}_d \\ -\mathbf{I}_d & \mathbf{0} \end{bmatrix} = (\sigma_{i,j} - \psi_{i,j}) \tag{11.16}$$

となる．また, $h_j = \dfrac{\partial h}{\partial x_j}\,(j = 1,\, 2,\, \ldots,)$ と置いたとき, 簡単な計算[2] から, $j = 1,\, 2,\, \ldots,\, d$ に対して

$$\frac{\partial G^2}{\partial x_j \partial x_{j+d}} = -\lambda^2 G h_j(x_1,\, x_2,\, \ldots,\, x_d) h_j(x_{d+1},\, x_{d+2},\, \ldots,\, x_{2d})$$

を得る．仮定から, $\bigl|\nabla h\bigr|_{2,\,d} \leq L$ なので, $h_j\,(j = 1,\, 2, \ldots,\, d)$ はすべて有界となる．したがって, $c_1,\, c_2 \in \mathbb{R}$ が存在して

$$\bigl|h(\boldsymbol{x})\bigr| \leq c_1 + c_2 |\boldsymbol{x}|_{2\,,d}$$

と評価[3] できる．よって, G に対して, 成長条件 (11.9) が満たされるので, (11.10) と (11.16) から

$$\begin{aligned}
\overset{\bullet}{g}\,(t) &= \frac{1}{2}\sum_{i,\,j=1}^{d} (\sigma_{i,j} - \psi_{i,j})\mathsf{E}\Bigl[\frac{\partial^2 G}{\partial x_i \partial x_j}\Bigr] \\
&= \lambda^2 \mathsf{E}\Bigl[G(\sqrt{t}\boldsymbol{X} + \sqrt{1-t}\boldsymbol{Y}) \sum_{j=1}^{d} h_j(\sqrt{t}\boldsymbol{Z}_1 + \sqrt{1-t}\boldsymbol{Z}) h_j(\sqrt{t}\boldsymbol{Z}_1 + \sqrt{1-t}\boldsymbol{Z})\Bigr]
\end{aligned}$$

---

[2]実際

$$\begin{aligned}
\frac{\partial G^2}{\partial x_j \partial x_{j+d}} &= \frac{\partial}{\partial x_{j+d}}\biggl\{\frac{\partial}{\partial x_j}\exp\Bigl(\lambda\bigl(h(x_1,\, x_2,\, \ldots,\, x_n) - h(x_{d+1},\, x_{d+2},\, \ldots,\, x_{2d})\bigr)\Bigr)\biggr\} \\
&= \frac{\partial}{\partial x_{j+d}}\biggl\{\lambda h_j(x_1,\, x_2,\, \ldots,\, x_n)\exp\Bigl(\lambda\bigl(h(x_1,\, x_2,\, \ldots,\, x_n) - h(x_{d+1},\, x_{d+2},\, \ldots,\, x_{2d})\bigr)\Bigr)\biggr\} \\
&= -\lambda h_j(x_{d+1},\, x_{d+2},\, \ldots,\, x_{2d})\lambda h_j(x_1,\, x_2,\, \ldots,\, x_n) \\
&\qquad \times \exp\Bigl(\lambda\bigl(h(x_1,\, x_2,\, \ldots,\, x_n) - h(x_{d+1},\, x_{d+2},\, \ldots,\, x_{2d})\bigr)\Bigr)\biggr\} \\
&= -\lambda^2 h_j(x_{d+1},\, x_{d+2},\, \ldots,\, x_{2d})\lambda h_j(x_1,\, x_2,\, \ldots,\, x_n)
\end{aligned}$$

からわかる．

[3]実際, $|h_j| \leq c_3$ とすると

$$\begin{aligned}
\bigl|h(\boldsymbol{y}) - h(\boldsymbol{x})\bigr| &= \Bigl|\int_0^1 \frac{d}{dt} h\bigl(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x})\bigr)\,\mathrm{d}t\Bigr| = \Bigl|\int_0^1 \sum_{j=1}^{d} (y_j - x_j) h_j\,\mathrm{d}t\Bigr| \\
&\leq \int_0^1 \sum_{j=1}^{d} |(y_j - x_j)| \times |h_j|\,\mathrm{d}t \leq c_3 \sum_{j=1}^{d} |y_j - x_j| \leq \sqrt{d} c_3 \sqrt{\sum_{j=1}^{d} (y_j - x_j)^2} \\
&= \sqrt{d} c_3 |\boldsymbol{y} - \boldsymbol{x}|_{2\,,d}
\end{aligned}$$

となる．ここで, $\boldsymbol{y} = \boldsymbol{0}$ とおくと

$$|h(\boldsymbol{x})| - \underbrace{|h(\boldsymbol{0})|}_{=:c_1} \leq \underbrace{\sqrt{d} c_3}_{=:c_2} |\boldsymbol{x}|_{2\,,d}$$

よりわかる．

がわかる. Cauchy-Schwarz の不等式から

$$\begin{aligned}&\sum_{j=1}^{d} h_j(\sqrt{t}\boldsymbol{Z}_1+\sqrt{1-t}\boldsymbol{Z})h_j(\sqrt{t}\boldsymbol{Z}_1+\sqrt{1-t}\boldsymbol{Z})\\&\qquad\le \bigl|\nabla h(\sqrt{t}\boldsymbol{Z}_1+\sqrt{1-t}\boldsymbol{Z})\bigr|_{2,\,d}\times\bigl|\nabla h(\sqrt{t}\boldsymbol{Z}_2+\sqrt{1-t}\boldsymbol{Z})\bigr|_{2,\,d}\\&\qquad\le L^2\end{aligned}$$

となる. よって

$$\overset{\bullet}{g}(t)\le\lambda^2L^2\mathsf{E}\bigl[G(\sqrt{t}\boldsymbol{X}+\sqrt{1-t}\boldsymbol{Y})\bigr]=\lambda^2L^2g(t)$$

を得る. これより

$$\frac{d}{dt}\Bigl\{g(t)e^{-\lambda^2L^2t}\Bigr\}=e^{-\lambda^2L^2t}\Bigl(\overset{\bullet}{g}(t)-\lambda^2L^2g(t)\Bigr)\le 0$$

となるので, 関数 $t\mapsto g(t)e^{-\lambda^2L^2t}$ は非増加である. よって, $g(1)\le e^{\lambda^2L^2}g(0)$ となる. さらに, (11.15) から $g(0)=1$ がわかるので

$$g(1)=\mathsf{E}\bigl[\exp\bigl\{\lambda\bigl(h(\boldsymbol{Z}_1)-h(\boldsymbol{Z}_2)\bigr)\bigr\}\bigr]\le e^{\lambda^2L^2} \tag{11.17}$$

がわかる.

<u>② $\mathsf{Pr}(h(\boldsymbol{Z})-\mathsf{E}[h(\boldsymbol{Z})]\ge t)\le e^{-\lambda^2/L^2}$ の証明.</u> 関数 $x\mapsto\exp(-\lambda x)$ は凸なので, Jensen の不等式 ($\boldsymbol{Z}_2$ の期待値に関するもの) から

$$\exp\Bigl\{\lambda\Bigl(h(\boldsymbol{Z}_1)-\mathsf{E}_2[h(\boldsymbol{Z}_2)]\Bigr)\Bigr\}\le\mathsf{E}_2\Bigl[\exp\Bigl\{\lambda\Bigl(h(\boldsymbol{Z}_1)-h(\boldsymbol{Z}_2)\Bigr)\Bigr\}\Bigr]$$

を得る. ただし, $\mathsf{E}_j[\,\cdot\,]$ は $\boldsymbol{Z}_j\ (j=1,2)$ に関する期待値を表す. さらに, 上記の不等式を $\boldsymbol{Z}_1$ に関して積分をし, $\boldsymbol{Z}_1$, $\boldsymbol{Z}_2$ は $\boldsymbol{Z}$ と同じ分布を持つことと (11.17) から

$$\begin{aligned}\mathsf{E}\Bigl[\exp\Bigl\{\lambda\Bigl(h(\boldsymbol{Z})-\mathsf{E}\bigl[h(\boldsymbol{Z})\bigr]\Bigr)\Bigr\}\Bigr]&=\mathsf{E}_1\Bigl[\exp\Bigl\{\lambda\Bigl(h(\boldsymbol{Z}_1)-\mathsf{E}_2\bigl[h(\boldsymbol{Z}_2)\bigr]\Bigr)\Bigr\}\Bigr]\\&\le\mathsf{E}_1\Bigl[\mathsf{E}_2\Bigl[\exp\Bigl\{\lambda\Bigl(h(\boldsymbol{Z}_1)-\bigl[h(\boldsymbol{Z}_2)\bigr]\Bigr)\Bigr\}\Bigr]\Bigr]\\&=\mathsf{E}\Bigl[\exp\Bigl\{\lambda\Bigl(h(\boldsymbol{Z}_1)-\bigl[h(\boldsymbol{Z}_2)\bigr]\Bigr)\Bigr\}\Bigr]\\&\le e^{\lambda^2L^2}\end{aligned}$$

を得る. さらに, $\lambda > 0$, $t > 0$ とし, Markov の不等式を用いると

$$\begin{aligned}\mathsf{Pr}\Big(h(\boldsymbol{Z}) - \mathsf{E}[h(\boldsymbol{Z})] \geq t\Big) &= \mathsf{Pr}\Big(\exp\Big(\lambda\big(h(\boldsymbol{Z}) - \mathsf{E}[h(\boldsymbol{Z})]\big)\Big) \geq e^{\lambda t}\Big)\\ &\leq e^{-\lambda t}\mathsf{E}\Big[\exp\Big(\lambda\big(h(\boldsymbol{Z}) - \mathsf{E}[h(\boldsymbol{Z})]\big)\Big)\Big]\\ &\leq \exp\Big(-\lambda t + \lambda^2 L^2\Big)\\ &= \exp\Big(L^2\big(\lambda - t/(2L^2)\big)^2 - \frac{t^2}{4L^2}\Big)\end{aligned}$$

を得る. $\lambda > 0$ は任意だったので, $\lambda = t/(2L^2)$ とおくと

$$\mathsf{Pr}\Big(h(\boldsymbol{Z}) - \mathsf{E}[h(\boldsymbol{Z})] \geq t\Big) \leq \exp\Big(-\frac{t^2}{4L^2}\Big)$$

を得る.

II h は Lipschitz の場合の証明. 標準的な smoothing technique を用いる. すなわち, $\epsilon > 0$ に対して

$$h_\epsilon(\boldsymbol{x}) := \mathsf{E}\big[h(\boldsymbol{x} + \epsilon\widetilde{\boldsymbol{Z}}\big] \tag{11.18}$$

と定める. ただし, $\widetilde{\boldsymbol{Z}} \sim \mathsf{N}_d(\boldsymbol{0}_d, \mathrm{I}_d)$ である. すると関数 h_ϵ も Lipschitz とこと [4] がわかる. しかし, 微分可能 [5] である. 関数 h_ϵ に I の結果を適用すると, $\forall t > 0$ に対して

$$\mathsf{Pr}\big(h_\epsilon(\boldsymbol{Z}) - \mathsf{E}[h_\epsilon(\boldsymbol{Z}) \geq t\big) \leq \exp\Big(-\frac{t^2}{4\|h_\epsilon\|_{\mathrm{Lip}}}\Big) \tag{11.19}$$

[4] 実際

$$\big|h_\epsilon(\boldsymbol{x}) - h_\epsilon(\boldsymbol{y})\big| \leq \mathsf{E}\big[\big|h(\boldsymbol{x} + \epsilon\boldsymbol{Z}) - h(\boldsymbol{y} + \epsilon\boldsymbol{Z})\big|\big] \leq L\|\boldsymbol{x} - \boldsymbol{y}|_{2,\,d}$$

よりわかる.

[5] なぜならば

$$\begin{aligned}\mathsf{E}\big[h(\boldsymbol{x} + \epsilon\boldsymbol{Z})\big] &= \frac{1}{(\sqrt{2\pi})^d}\int_{\mathbb{R}^d} h(\boldsymbol{x} + \epsilon\boldsymbol{y})e^{-|\boldsymbol{y}|_{2,\,d}^2/2}\,\mathrm{d}\boldsymbol{y}\\ &= \frac{1}{(\epsilon\sqrt{2\pi})^d}\int_{\mathbb{R}^d} h(\boldsymbol{t})e^{-|\boldsymbol{t}-\boldsymbol{x}|_{2,\,d}^2/(2\epsilon^2)}\,\mathrm{d}\boldsymbol{t}\\ &\quad\big(\because \boldsymbol{y} = (\boldsymbol{t} - \boldsymbol{x})/\epsilon \text{ と変換 }\big)\end{aligned}$$

とかけることからわかる.

となる. 十分小さな $\epsilon > 0$ に対して

$$
\begin{aligned}
\bigl|h(\boldsymbol{x}) - h_\epsilon(\boldsymbol{x})\bigr| &= \left|\mathsf{E}\left[h(\boldsymbol{x}) - h(\boldsymbol{x} + \epsilon\widetilde{\boldsymbol{Z}})\right]\right| \\
&\le \mathsf{E}\left[\bigl|h(\boldsymbol{x}) - h(\boldsymbol{x} + \epsilon\widetilde{\boldsymbol{Z}})\bigr|\right] \\
&\le \mathsf{E}\left[L\epsilon\bigl|\widetilde{\boldsymbol{Z}}\bigr|_{2,\,d}\right] \qquad \bigl(\because (11.14)\bigr) \\
&\le L\epsilon\bigl|\widetilde{\boldsymbol{Z}}\bigr|_{2,\,d} \overset{\epsilon\to 0}{\longrightarrow} 0
\end{aligned}
$$

から, h_ϵ は h を一様に近似することがわかる. よって

$$
\begin{aligned}
&\mathsf{Pr}\Bigl(h(\boldsymbol{Z}) - \mathsf{E}[h(\boldsymbol{Z})] \ge t\Bigr) \\
&\quad = \mathsf{Pr}\Bigl(h_\epsilon(\boldsymbol{Z}) - \mathsf{E}[h_\epsilon(\boldsymbol{Z})] \ge t - \bigl\{h(\boldsymbol{Z})) - h_\epsilon(\boldsymbol{Z})) - \mathsf{E}[h(\boldsymbol{Z})) - h_\epsilon(\boldsymbol{Z}))]\bigr\}\Bigr) \\
&\quad \le \mathsf{Pr}\Bigl(h_\epsilon(\boldsymbol{Z}) - \mathsf{E}[h_\epsilon(\boldsymbol{Z})] \ge t - 2L\epsilon\mathsf{E}[|\widetilde{Z}|_{2,\,d}\Bigr) \\
&\quad \le \exp\left(-\frac{\bigl(t - 2L\epsilon\mathsf{E}[|\widetilde{Z}|_{2,\,d}\bigr)^2}{4\|h_\epsilon\|_{\mathrm{Lip}}}\right) \\
&\quad \overset{\epsilon\to 0}{\longrightarrow} \exp\left(-\frac{t^2}{4\|h\|_{\mathrm{Lip}}}\right)
\end{aligned}
$$

がわかる. よって, 定理は証明された. □

以下では, Gauss 確率集中不等式の適用例を述べる.

例 11.6. □

例 11.7. □

例 11.8. □

例 11.9. □

11.3 Gauss 分布比較

定理 11.10 (Slepian の不等式)**.** $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)^\top$, $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_d)^\top$ は $\mathbb{R}^d$ 上の中心化された正規分布に従う確率ベクトル[6] で

1. $\mathsf{E}[X_j^2] = \mathsf{E}[Y_j]^2$ $(j = 1, 2, \ldots, d)$,

[6] $\boldsymbol{X}$ と $\boldsymbol{Y}$ は独立でなくともよい.

2. $\mathsf{E}[X_j X_k] \le \mathsf{E}[Y_j Y_k]\,(j,\,k = 1,\,2,\ldots,\,d)$

をみたすとする．このとき，$\forall t_j\,(j = 1,\,2,\ldots,\,d)$ に対して

$$\mathsf{Pr}\Bigl(\bigcap_{j=1}^{d}\bigl\{X_j \le t_j\bigr\}\Bigr) \le \mathsf{Pr}\Bigl(\bigcap_{j=1}^{d}\bigl\{Y_j \le t_j\bigr\}\Bigr) \tag{11.20}$$

が成り立つ．さらに

$$\mathsf{E}\Bigl[\max_{j=1,\,2,\,\ldots,\,d} X_j\Bigr] \ge \mathsf{E}\Bigl[\max_{j=1,\,2,\,\ldots,\,d} Y_j\Bigr] \tag{11.21}$$

が成立する．

Proof. **I** (11.20) の証明．まず

$$\mathbb{1}\Bigl(\bigcap_{j=1}^{d}\bigl\{x_j \le t_j\bigr\}\Bigr) = \prod_{j=1}^{d}\mathbb{1}\bigl\{x_j \le t_j\bigr\}$$

に注意する．滑らかな非負値非増加関数 $\varphi_j(\,\cdot\,)\,(j = 1,\,2,\,\ldots,\,d)$ で $\mathbb{1}_{(-\infty,\,t_j]}(\,\cdot\,)$ を近似する．ここで

$$\varphi(\boldsymbol{x}) = \prod_{j=1}^{d}\varphi_j(x_j), \qquad \boldsymbol{x} = (x_1,\,x_2,\,\ldots,\,d)$$

とおき，Gauss 補間

$$g(t) := \mathsf{E}\bigl[\varphi\bigl(\sqrt{t}\boldsymbol{X} + \sqrt{1-t}\boldsymbol{Y}\bigr)\bigr]\mathsf{E}\bigl[\varphi\bigl(\sqrt{t}\boldsymbol{Z}(t)\bigr)\bigr]$$

を考える．すると (11.10) から

$$\dot{g}\,(t) = \frac{1}{2}\sum_{i,\,j=1}^{d}\underbrace{\bigl(\mathsf{E}[X_iX_j] - \mathsf{E}[Y_iY_j]\bigr)}_{\le 0}\mathsf{E}\Bigl[\frac{\partial^2\varphi}{\partial x_i\partial x_j}(\boldsymbol{Z}(t))\Bigr] \le 0$$

となる．実際，$\dot{\varphi}_i\,(x_i) \le 0$ かつ $\dot{\varphi}_j\,(x_j) \le 0\,(i \ne j)$ なので

$$\frac{\partial^2}{\partial x_i\partial x_j}\prod_{\ell=1}^{d}\varphi_\ell(x_\ell) = \underbrace{\dot{\varphi}_i\,(x_i)\,\dot{\varphi}_j\,(x_j)}_{\ge 0}\prod_{\ell\ne i,\,j}\underbrace{\varphi_\ell(x_\ell)}_{\ge 0} \ge 0$$

となることからわかる．よって，g は非増加となるので

$$g(1) = \mathsf{E}[\varphi(\boldsymbol{X})] \le g(0) = \mathsf{E}[\varphi(\boldsymbol{Y})]$$

がわかる. φ_j は指示関数 $\mathbb{1}_{(-\infty,\, t_j]}$ に収束するので

$$\begin{aligned}\mathsf{Pr}\biggl(\bigcap_{j=1}^{d}\{X_j \le t_j\}\biggr) &= \mathsf{E}\biggl[\bigcap_{j=1}^{d}\{X_j \le t_j\}\biggr] \le \mathsf{E}\biggl[\bigcap_{j=1}^{d}\{Y_j \le t_j\}\biggr] \\ &= \mathsf{Pr}\biggl(\bigcap_{j=1}^{d}\{Y_j \le t_j\}\biggr)\end{aligned}$$

がわかる.

<u>**II** (11.23) の証明</u>. (11.20) から (11.23) を証明する. (11.20) において $t_j = t\,(j = 1,\, 2,\, \ldots,\, d)$ とおくと

$$\begin{aligned}\mathsf{Pr}\Bigl(\max_{j=1,\,2,\,\ldots,\,d} X_j \le t\Bigr) &= \mathsf{Pr}\biggl(\bigcap_{j=1}^{d}\{X_j \le t\}\biggr) \le \mathsf{Pr}\biggl(\bigcap_{j=1}^{d}\{Y_j \le t\}\biggr) \\ &= \mathsf{Pr}\Bigl(\max_{j=1,\,2,\,\ldots,\,d} Y_j \le t\Bigr) \qquad (11.22)\end{aligned}$$

となる. ここで

$$\begin{aligned}&X := \max_{j=1,\,2,\,\ldots,\,d} X_j, \quad X^+ := \max\{X,\, 0\}, \quad X^- := \max\{-X,\, 0\}, \\ &Y := \max_{j=1,\,2,\,\ldots,\,d} Y_j, \quad Y^+ := \max\{Y,\, 0\}, \quad Y^- := \max\{-Y,\, 0\}\end{aligned}$$

とおく. すると $X^+,\, X^-,\, Y^+,\, Y^-$ は非負値確率変数で

$$X = X^+ - X^-, \quad Y = Y^+ - Y^-$$

となる. すると $t \ge 0$ に対して, $\{X^+ \ge 0\} \cap \{X^- \ge 0\} = \varnothing$ となること注意して, (11.22) を用いると

$$\mathsf{Pr}\bigl(X^+ - X^- > t\bigr) \ge \mathsf{Pr}\bigl(Y^+ - Y^- > t\bigr) \;\Rightarrow\; \mathsf{Pr}\bigl(X^+ > t\bigr) \ge \mathsf{Pr}\bigl(Y^+ > t\bigr)$$

となる. さらに

$$\begin{aligned}&\mathsf{Pr}\bigl(X^+ - X^- > -t\bigr) \ge \mathsf{Pr}\bigl(Y^+ - Y^- > -t\bigr) \\ &\quad\Rightarrow\; \mathsf{Pr}\bigl(X^- - X^+ < t\bigr) \ge \mathsf{Pr}\bigl(Y^- - Y^+ < t\bigr) \\ &\quad\Rightarrow\; \mathsf{Pr}\bigl(X^- - X^+ \ge t\bigr) \le \mathsf{Pr}\bigl(Y^- - Y^+ \ge t\bigr) \\ &\quad\Rightarrow\; \mathsf{Pr}\bigl(X^- \ge t\bigr) \le \mathsf{Pr}\bigl(Y^- \ge t\bigr)\end{aligned}$$

がわかる. よって, 命題 1.32 から

$$
\begin{aligned}
\mathsf{E}[X] &= \mathsf{E}[X^+] - \mathsf{E}[X^-] \\
&= \int_0^t \mathsf{Pr}(X^+ > t)\,\mathrm{d}t - \int_0^t \mathsf{Pr}(X^- > t)\,\mathrm{d}t \\
&= \int_0^t \mathsf{Pr}(X^+ > t)\,\mathrm{d}t - \int_0^t \mathsf{Pr}(X^- \geq t)\,\mathrm{d}t \\
&\geq \int_0^t \mathsf{Pr}(Y^+ > t)\,\mathrm{d}t - \int_0^t \mathsf{Pr}(Y^- \geq t)\,\mathrm{d}t \\
&\geq \int_0^t \mathsf{Pr}(Y^+ > t)\,\mathrm{d}t - \int_0^t \mathsf{Pr}(Y^- > t)\,\mathrm{d}t \\
&= \mathsf{E}[Y^+] - \mathsf{E}[Y^-] = \mathsf{E}[Y]
\end{aligned}
$$

となる. よって, 定理は証明された. □

次に, (11.23) をより制限のすくない条件のもとで証明しよう. $i, j \in \{1, 2, \ldots, d\}$, $i \neq j$ に対して

$$
\begin{aligned}
\mathsf{E}\big[(X_i - X_j)^2\big] &= \mathsf{E}[X_i^2] + \mathsf{E}[X_j^2] - \mathsf{E}[X_i X_j], \\
\mathsf{E}\big[(Y_i - Y_j)^2\big] &= \mathsf{E}[Y_i^2] + \mathsf{E}[Y_j^2] - \mathsf{E}[Y_i Y_j]
\end{aligned}
$$

となるので, 定理 11.10 の最初の条件のもと

$$
\begin{aligned}
&\mathsf{E}[X_i X_j] \leq \mathsf{E}[Y_i Y_j]\ (\forall 1 \leq i, j \leq d) \\
&\qquad\qquad \Leftrightarrow\ \mathsf{E}\big[(X_i - X_j)^2\big] \geq \mathsf{E}\big[(Y_i - Y_j)^2\big]\ (\forall 1 \leq i, j \leq d)
\end{aligned}
$$

となる. 2 番目の条件を置き換えることにより, 自動的に 1 番目の条件が成立することを示す.

定理 11.11. $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)^\top$, $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_d)^\top$ は $\mathbb{R}^d$ 上の中心化された正規分布に従う確率ベクトルで

$$
\mathsf{E}\big[(X_i - X_j)^2\big] \geq \mathsf{E}\big[(Y_i - Y_j)^2\ (\forall 1 \leq i, j \leq d)
$$

をみたすとする. このとき

$$
\mathsf{E}\left[\max_{j=1,2,\ldots,d} X_j\right] \geq \mathsf{E}\left[\max_{j=1,2,\ldots,d} Y_j\right] \tag{11.23}
$$

が成立する.

Proof. ① $\displaystyle\max_{j=1,2,\ldots,d} x_j$ **の平滑化**. 補間法を用いて証明する. しかし, 最大値関数 $\overline{\boldsymbol{x} = (x_1, x_2, \ldots, x_d) \mapsto}\ \displaystyle\max_{j=1,2,\ldots,d} x_j$ に対する特別な平滑化関数

$$
F(\boldsymbol{x}) = \frac{1}{\beta} \log \sum_{j=1}^d e^{\beta x_j}
$$

を用いる．ただし, $\beta > 0$ である．関数 F を最大値の平滑化関数とみる理由は以下である．

$$\max_{j=1,2\ldots,d} x_j \le \frac{1}{\beta}\log\left(\sum_{j=1}^{d} e^{\beta x_j}\right) \le \max_{j=1,2\ldots,d} x_j + \frac{\log n}{\beta} \tag{11.24}$$

が成り立つこと[7]からわかる．すると

$$\mathsf{E}[F(\boldsymbol{X})] \stackrel{\beta\to\infty}{\longrightarrow} \mathsf{E}\left[\max_{j=1,2,\ldots,d} X_j\right]; \qquad \mathsf{E}[F(\boldsymbol{Y})] \stackrel{\beta\to\infty}{\longrightarrow} \mathsf{E}\left[\max_{j=1,2,\ldots,d} Y_j\right]$$

から

$$\mathsf{E}[F(\boldsymbol{X})] \ge \mathsf{E}[F(\boldsymbol{Y})] \Rightarrow \mathsf{E}\left[\max_{j=1,2,\ldots,d} X_j\right] \ge \mathsf{E}\left[\max_{j=1,2,\ldots,d} Y_j\right]$$

がわかる．

② $\mathsf{E}[F(\boldsymbol{X})] \ge \mathsf{E}[F(\boldsymbol{Y})]$ の証明. Gauss 補間 (11.10) を用いるために, F の導関数を求める．まず, $j = 1, 2, \ldots, d$ に対して

$$F_j(\boldsymbol{x}) := \frac{\partial F}{\partial x_j} = \frac{e^{\beta x_j}}{\sum_{\ell=1}^{n} e^{\beta x_\ell}}$$

となることに注意する．さらに

$$F_{jj}(\boldsymbol{x}) := \frac{\partial^2 F}{\partial x_j^2} = \beta\Big(F_j(\boldsymbol{x}) - \{F_j(\boldsymbol{x})\}^2\Big), \tag{11.25}$$

$$F_{ij}(\boldsymbol{x}) := \frac{\partial^2 F}{\partial x_i \partial x_j^2} = -\beta F_i(\boldsymbol{x}) F_j(\boldsymbol{x}) \qquad (i \ne j) \tag{11.26}$$

となる．よって, Gauss 補間 (11.10) は以下のようになる．$\boldsymbol{Z}(t) = \sqrt{t}\boldsymbol{X} +$

[7]実際

$$\begin{aligned}\max_{j=1,2\ldots,d} x_j &= \frac{1}{\beta}\sum_{j=1}^{d} \beta x_j = \frac{1}{\beta}\log\exp\left(\sum_{j=1}^{d}\beta x_j\right) \le \frac{1}{\beta}\log\left(\sum_{j=1}^{d} e^{\beta x_j}\right)\\ &\le \frac{1}{\beta}\log\left(n\exp\{\beta\max_{j=1,2\ldots,d} x_j\}\right) = \frac{1}{\beta}\left(\log n + \beta\max_{j=1,2\ldots,d} x_j\right)\\ &= \max_{j=1,2\ldots,d} x_j + \frac{\log n}{\beta}\end{aligned}$$

からわかる．

$\sqrt{1-t}\boldsymbol{Y}$ とおくと

$$
\begin{aligned}
\overset{\bullet}{g}(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\mathsf{E}\big[F(\boldsymbol{Z}(t)\big] \\
&= \frac{1}{2}\sum_{i,j=1}^{d}\big\{\mathsf{E}[X_iX_j]-\mathsf{E}[Y_iY_j]\big\}\mathsf{E}\left[\frac{\partial^2 F}{\partial x_i\partial x_j}(\boldsymbol{Z}(t))\right] \\
&= \frac{\beta}{2}\sum_{j=1}^{d}\big\{\mathsf{E}[X_j^2]-\mathsf{E}[Y_j^2]\big\}\mathsf{E}\big[F_j(\boldsymbol{Z}(t))-F_j^2(\boldsymbol{Z}(t))\big] \\
&\quad -\frac{\beta}{2}\sum_{i\neq j}\big\{\mathsf{E}[X_iX_j]-\mathsf{E}[Y_iY_j]\big\}\mathsf{E}\Big[F_i(\boldsymbol{Z}(t))F_j(\boldsymbol{Z}(t))\Big] \qquad (11.27)
\end{aligned}
$$

を得る．F をうまく選んで

$$
\sum_{j=1}^{d}F_j(\boldsymbol{x})=\sum_{j=1}^{d}\frac{e^{\beta x_j}}{\sum_{k=1}^{d}e^{\beta x_k}}=1
$$

とする．すると

$$
F_j(\boldsymbol{x})-F_j^2(\boldsymbol{x})=F_j(\boldsymbol{x})\big\{1-F_j(\boldsymbol{x})\big\}=F_j(\boldsymbol{x})\sum_{i\neq j}F_i(\boldsymbol{x})=\sum_{j\neq i}F_i(\boldsymbol{x})F_j(\boldsymbol{x})
$$

がわかる．これから

$$
\begin{aligned}
&\sum_{j=1}^{d}\big\{\mathsf{E}[X_j^2]-\mathsf{E}[Y_j^2]\big\}\mathsf{E}\big[F_j(\boldsymbol{Z}(t))-F_j^2(\boldsymbol{Z}(t))\big] \\
&\qquad = \sum_{j=1}^{d}\big\{\mathsf{E}[X_j^2]-\mathsf{E}[Y_j^2]\big\}\sum_{i\neq j}\mathsf{E}\big[F_i(\boldsymbol{Z}(t))F_j(\boldsymbol{Z}(t))\big] \\
&\qquad = \sum_{i=1}^{d}\sum_{j\neq i}^{d}\big\{\mathsf{E}[X_j^2]-\mathsf{E}[Y_j^2]\big\}\mathsf{E}\big[F_i(\boldsymbol{Z}(t))F_j(\boldsymbol{Z}(t))\big]
\end{aligned}
$$

となる．上式の右辺の i と j を入れ替えると

$$
\begin{aligned}
&\sum_{j=1}^{d}\big\{\mathsf{E}[X_j^2]-\mathsf{E}[Y_j^2]\big\}\mathsf{E}\big[F_j(\boldsymbol{Z}(t))-F_j^2(\boldsymbol{Z}(t))\big] \\
&\qquad = \sum_{i=1}^{d}\sum_{j\neq i}^{d}\big\{\mathsf{E}[X_i^2]-\mathsf{E}[Y_i^2]\big\}\mathsf{E}\big[F_i(\boldsymbol{Z}(t))F_j(\boldsymbol{Z}(t))\big]
\end{aligned}
$$

となる．上の 2 つの式の辺々を加えると

$$
\begin{aligned}
&\sum_{j=1}^{d}\big\{\mathsf{E}[X_j^2]-\mathsf{E}[Y_j^2]\big\}\mathsf{E}\big[F_j(\boldsymbol{Z}(t))-F_j^2(\boldsymbol{Z}(t))\big] \\
&\qquad = \frac{1}{2}\sum_{i=1}^{d}\sum_{j\neq i}\Big(\mathsf{E}[X_i^2]+\mathsf{E}[X_j^2]-\mathsf{E}[Y_i^2]-\mathsf{E}[Y_j^2]\Big)\mathsf{E}\big[F_i(\boldsymbol{Z}(t))F_j(\boldsymbol{Z}(t))\big]
\end{aligned}
$$

を得る. この式を (11.27) の最右辺の 1 項目に入れると

$$
\begin{aligned}
\overset{\bullet}{g}(t) &= \frac{\beta}{4}\sum_{i=1}^{d}\sum_{j\neq i}\Big(\mathsf{E}[X_i^2]+\mathsf{E}[X_j^2]-\mathsf{E}[Y_i^2]-\mathsf{E}[Y_j^2]\Big)\mathsf{E}\big[F_i(\boldsymbol{Z}(t))F_j(\boldsymbol{Z}(t))\big] \\
&\quad -\frac{\beta}{2}\sum_{i=1}^{d}\sum_{j\neq i}\big\{\mathsf{E}[X_iX_j]-\mathsf{E}[Y_iY_j]\big\}\mathsf{E}\Big[F_i(\boldsymbol{Z}(t))F_j(\boldsymbol{Z}(t))\Big] \\
&= \frac{\beta}{4}\sum_{i=1}^{d}\sum_{j\neq i}\Big(\mathsf{E}\big[(X_i-X_j)^2\big]-\mathsf{E}\big[(Y_i-Y_j)^2\big]\Big)\mathsf{E}\Big[F_i(\boldsymbol{Z}(t))F_j(\boldsymbol{Z}(t))\Big] \qquad (11.28) \\
&\geq 0
\end{aligned}
$$

を得る. これから

$$
\mathsf{E}[F(\boldsymbol{X})]=\mathsf{E}[F(\boldsymbol{Z}(1)]=g(1)\geq g(0)=\mathsf{E}[F(\boldsymbol{Z}(0)]=\mathsf{E}[F(\boldsymbol{Y})]
$$

がわかる. □

例 11.12. □

系 11.13. $\boldsymbol{X}=(X_1,\,X_2,\,\ldots,\,X_d)^\top$, $\boldsymbol{Y}=(Y_1,\,Y_2,\,\ldots,\,Y_d)^\top$ は $\mathbb{R}^d$ 上の中心化された正規分布に従う確率ベクトルで, ある $\epsilon>0$ に対して

$$
\Big|\mathsf{E}\big[(X_i-X_j)^2\big]-\mathsf{E}\big[(Y_i-Y_j)^2\big]\Big|\leq\frac{\epsilon}{d} \qquad (\forall i,\,j=1,\,2,\,\ldots,\,d) \quad (11.29)
$$

をみたすとき

$$
\left|\mathsf{E}\left[\max_{j=1,\,2,\,\ldots,\,d}X_j\right]-\mathsf{E}\left[\max_{j=1,\,2,\,\ldots,\,d}Y_j\right]\right|\leq\sqrt{\epsilon\log d}
$$

が成り立つ.

Proof. $\beta>0$ に対して

$$
F(\boldsymbol{x})=\frac{1}{\beta}\log\left(\sum_{j=1}^{d}e^{\beta x_j}\right)
$$

おく. すると, $i\neq j$ に対して

$$
\begin{aligned}
&\frac{\partial F}{\partial x_j}(\boldsymbol{x})=:F_j(\boldsymbol{x})=\frac{e^{\beta x_i}}{\sum_{k=1}^{d}e^{\beta x_k}}, \\
&\qquad \frac{\partial^2 F}{\partial x_i\partial x_j}(\boldsymbol{x})=\frac{1}{\beta}\frac{\partial}{\partial x_j}\left\{\frac{\beta e^{\beta x_i}}{\sum_{k=1}^{d}e^{\beta x_k}}\right\}=-\frac{1}{\beta}\Bigg\{\underbrace{\frac{\beta e^{\beta x_i}}{\sum_{k=1}^{d}e^{\beta x_k}}}_{\beta F_i(\boldsymbol{x})}\Bigg\}\left\{\frac{\beta e^{\beta x_j}}{\sum_{k=1}^{d}e^{\beta x_k}}\right\} \\
&\qquad\qquad =-\beta F_i(\boldsymbol{x})F_j(\boldsymbol{x})=:F_{i,j}(\boldsymbol{x})
\end{aligned}
$$

となる. $\sum_{j=1}^d F_j(\boldsymbol{X}) = 1,\ 0 \le F_i(\boldsymbol{x}) \le 1\,(j=1,\ 2,\ldots,\ d)$ から

$$\left|\frac{\partial^2 F}{\partial x_i \partial x_j}(\boldsymbol{x})\right| \le \beta, \quad F_{i,i}(\boldsymbol{x}) = -\sum_{j\neq i} F_{i,j}(\boldsymbol{x})$$

がわかる. さらに

$$\begin{aligned}\overset{\bullet}{g}\,(t) &= \mathsf{E}\left[\frac{\mathrm{d}}{\mathrm{d}t}F(\boldsymbol{Z}(t))\right]\\ &= \mathsf{E}\left[\sum_{j=1}^d F_j(\boldsymbol{Z}(t))\left(\frac{1}{2\sqrt{t}}X_j - \frac{1}{2\sqrt{1-t}}Y_j\right)\right] \end{aligned}\tag{11.30}$$

となる. 定理 11.3 から

$$\begin{aligned}\mathsf{E}\Big[F_j(\boldsymbol{Z}(t))X_j\Big] &= \sqrt{t}\sum_{i=1}^d \mathsf{E}[X_jX_i]\mathsf{E}\left[\frac{\partial F^2}{\partial x_i \partial x_j}(\boldsymbol{Z}(t))\right],\\ \mathsf{E}\Big[F_j(\boldsymbol{Z}(t))Y_j\Big] &= \sqrt{1-t}\sum_{i=1}^d \mathsf{E}[Y_jY_i]\mathsf{E}\left[\frac{\partial F^2}{\partial x_i \partial x_j}(\boldsymbol{Z}(t))\right]\end{aligned}$$

となる. 上の 2 式を (11.30) に入れると

$$\begin{aligned}\overset{\bullet}{g}\,(t) &= \frac{1}{2}\sum_{j=1}^d\sum_{i=1}^d \big\{\mathsf{E}[X_iX_j] - \mathsf{E}[Y_iY_j]\big\}\mathsf{E}\left[\frac{\partial F^2}{\partial x_i \partial x_j}(\boldsymbol{Z}(t))\right]\\ &= \frac{1}{2}\sum_{j=1}^d\Bigg\{\big\{\mathsf{E}[X_j^2] - \mathsf{E}[Y_j^2]\big\}\mathsf{E}\left[\frac{\partial F^2}{\partial x_j \partial x_j}(\boldsymbol{Z}(t))\right]\\ &\qquad + \sum_{i\neq j}\big\{\mathsf{E}[X_iX_j] - \mathsf{E}[Y_iY_j]\big\}\mathsf{E}\left[\frac{\partial F^2}{\partial x_i \partial x_j}(\boldsymbol{Z}(t))\right]\Bigg\}\end{aligned}\tag{11.31}$$

となる. さらに, (11.26) と $\sum_{k=1}^d F_k(\boldsymbol{Z}(t)) = 1$ から

$$\begin{aligned}&\sum_{j=1}^d \big\{\mathsf{E}[X_j^2] - \mathsf{E}[Y_j^2]\big\}\mathsf{E}\left[\frac{\partial F^2}{\partial x_j \partial x_j}(\boldsymbol{Z}(t))\right]\\ &\quad = \beta\sum_{j=1}^d \big\{\mathsf{E}[X_j^2] - \mathsf{E}[Y_j^2]\big\}\mathsf{E}\Big[F_i(\boldsymbol{Z}(t))\underbrace{\big\{1 - F_i(\boldsymbol{Z}(t))\big\}}_{=-\sum_{j\neq i}F_j(\boldsymbol{Z}(t))}\Big]\\ &\quad = -\beta\sum_{j=1}^d\sum_{j\neq i} \big\{\mathsf{E}[X_j^2] - \mathsf{E}[Y_j^2]\big\}\mathsf{E}\big[F_i(\boldsymbol{Z}(t))F_j(\boldsymbol{Z}(t))\big]\end{aligned}$$

となる. 上の式の右辺で i と j を入れ替えると

$$\sum_{j=1}^{d}\bigl\{\mathsf{E}[X_j^2]-\mathsf{E}[Y_j^2]\bigr\}\mathsf{E}\left[\frac{\partial F^2}{\partial x_j\partial x_j}(\boldsymbol{Z}(t))\right]$$
$$=-\beta\sum_{i=1}^{d}\sum_{i\neq j}\bigl\{\mathsf{E}[X_i^2]-\mathsf{E}[Y_i^2]\bigr\}\mathsf{E}\bigl[F_i(\boldsymbol{Z}(t))F_j(\boldsymbol{Z}(t))\bigr]$$

となる. これらのふたつの式を合わせると

$$\sum_{j=1}^{d}\bigl\{\mathsf{E}[X_j^2]-\mathsf{E}[Y_j^2]\bigr\}\mathsf{E}\left[\frac{\partial F^2}{\partial x_j\partial x_j}(\boldsymbol{Z}(t))\right]$$
$$=-\frac{\beta}{2}\sum_{j=1}^{d}\sum_{i\neq j}\bigl\{\mathsf{E}[X_i^2]+\mathsf{E}[X_j^2]-\mathsf{E}[Y_i^2]+\mathsf{E}[Y_j^2]\bigr\}\mathsf{E}\bigl[F_i(\boldsymbol{Z}(t))F_j(\boldsymbol{Z}(t))\bigr]$$

を得る. この式を (11.31) に代入すると

$$\overset{\bullet}{g}(t)=-\frac{\beta}{4}\sum_{j=1}^{d}\sum_{i\neq j}\Bigl\{\mathsf{E}[X_i^2]+\mathsf{E}[X_j^2]-\mathsf{E}[Y_i^2]-\mathsf{E}[Y_j^2]-2\mathsf{E}[X_iX_j]+2\mathsf{E}[Y_iY_j]\Bigr\}$$
$$\times\,\mathsf{E}\bigl[F_i(\boldsymbol{Z}(t))F_j(\boldsymbol{Z}(t))\bigr]$$
$$=-\frac{\beta}{4}\sum_{j=1}^{d}\sum_{i\neq j}\Bigl\{\mathsf{E}[(X_i-X_j)^2]-\mathsf{E}[(Y_i-Y_j)^2]\Bigr\}\mathsf{E}\bigl[F_i(\boldsymbol{Z}(t))F_j(\boldsymbol{Z}(t))\bigr]$$

を得る. したがって, (11.29) から

$$\bigl|\overset{\bullet}{g}(t)\bigr|\le\frac{\epsilon\beta}{4d}\Bigl|\sum_{j=1}^{d}\sum_{i\neq j}\mathsf{E}\bigl[F_i(\boldsymbol{Z}(t))F_j(\boldsymbol{Z}(t))\bigr]\Bigr|$$
$$=\frac{\epsilon\beta}{4d}\Bigl|\sum_{j=1}^{d}\mathsf{E}\bigl[\underbrace{\sum_{i\neq j}F_i(\boldsymbol{Z}(t))}_{=1-F_j(\boldsymbol{Z}(t))}F_j(\boldsymbol{Z}(t))\bigr]\Bigr|$$
$$\le\frac{\epsilon\beta}{4d}\sum_{j=1}^{d}\mathsf{E}[|F_j(\boldsymbol{Z}(t))|]$$
$$\le\frac{\epsilon\beta}{4}$$

を得る. 以上から

$$\Bigl|\mathsf{E}[F(\boldsymbol{X})]-\mathsf{E}[F(\boldsymbol{Y})]\Bigr|=\bigl|g(1)-g(0)\bigr|=\Bigl|\int_0^1\overset{\bullet}{g}(t)\,\mathrm{d}t\Bigr|$$
$$\le\mathsf{E}\bigl[\bigl|\overset{\bullet}{g}(t)\bigr|\bigr]\le\frac{\beta\epsilon}{4}\tag{11.32}$$

となる. (11.27) から (11.28) から

$$\begin{aligned} F(\boldsymbol{X}) &\leq \max_{j=1,2,\ldots,d} X_j + \frac{\log d}{\beta}, \\ F(\boldsymbol{Y}) &\geq \max_{j=1,2,\ldots,d} Y_j \end{aligned}$$

となるので

$$F(\boldsymbol{X}) - \max_{j=1,2,\ldots,d} X_j - \Big\{F(\boldsymbol{Y}) - \max_{j=1,2,\ldots,d} Y_j\Big\} \leq \frac{\log d}{\beta}$$

となる. 同様に

$$F(\boldsymbol{Y}) - \max_{j=1,2,\ldots,d} Y_j - \Big\{F(\boldsymbol{X}) - \max_{j=1,2,\ldots,d} X_j\Big\} \leq \frac{\log d}{\beta}$$

となるので

$$\left|F(\boldsymbol{X}) - \max_{j=1,2,\ldots,d} X_j - \Big\{F(\boldsymbol{Y}) - \max_{j=1,2,\ldots,d} Y_j\Big\}\right| \leq \frac{\log d}{\beta} \tag{11.33}$$

となる. したがって, (11.32) と (11.33) から

$$\begin{aligned} &\left|\mathsf{E}\left[\max_{j=1,2,\ldots,d} X_j\right] - \mathsf{E}\left[\max_{j=1,2,\ldots,d} X_j\right]\right| \\ &\quad \leq \Big|\mathsf{E}[F(\boldsymbol{Z})] - \mathsf{E}[F(\boldsymbol{Y})\Big| \\ &\qquad + \mathsf{E}\left[\left|F(\boldsymbol{X}) - \max_{j=1,2,\ldots,d} X_j - \Big\{F(\boldsymbol{Y}) - \max_{j=1,2,\ldots,d} Y_j\Big\}\right|\right] \\ &\quad \leq \frac{\beta\epsilon}{4} + \frac{\log d}{\beta} \end{aligned}$$

を得る. ここで, $\beta = 2\sqrt{\dfrac{\log d}{\epsilon}}$ とおくと

$$\left|\mathsf{E}\left[\max_{j=1,2,\ldots,d} X_j\right] - \mathsf{E}\left[\max_{j=1,2,\ldots,d} X_j\right]\right| \leq \sqrt{\epsilon \log d}$$

を得る. よって, 系は証明された. □

定理 11.14 (Sudakov minoration)**.** $X_1, X_2, \ldots, X_d$ を中心かされた正規分布に従う確率変数列とする. このとき

$$\mathsf{E}\left[\max_{j=1,2,\ldots,d} X_j\right] \geq \frac{1}{2} \min_{i \neq j} \sqrt{\mathsf{E}\big[(X_i - X_j)^2\big] \log d}$$

が成立する.

Proof. □

注意 11.15. Sudakov minoration の不等式の驚くべき結果は, Gauss 過程の最大値に対する下限を与えることである. 実際, $\mathcal{T}$ で添え字づけれた中心化された Gauss 過程 $\{X_t : t \in \mathcal{T}\}$ を考える. $\mathcal{T}$ 上の疑似距離 $d(t, s)^2 = \mathsf{E}[(X_t - X_s)^2]$ で定める. このとき, $0 < \forall\epsilon \le D := \sup_{s, t \in \mathcal{T}} d(s, t)$ に対して

$$\mathsf{E}\left[\sup_{t \in \mathcal{T}} X_t\right] \ge \frac{\epsilon}{2}\sqrt{H(\epsilon, \mathcal{T}, d)}$$

をもつことである.