Frontiers of Statistics: Contraction theory for posterior distributions **Spring 2019**

Lecture 7: March 19

Lecturer: Anirban Bhattacharya & Debdeep Pati *Scribes: Se Yoon Lee & Satwik Acharyya*

Note: *LaTeX template courtesy of UC Berkeley EECS dept & CMU's convex optimization course taught by Ryan Tibshirani.*

7.1 Upper bound of probability of shifted small ball: Anderson's theorem

In 1955, T.W Anderson established the *elegant moving-set inequality*, which probably the most widely sited result in multivariate statistical analysis. It has variant versions. Here, we start from general version which does not involve probability measure, and eventually narrows down to multivariate Gaussian measure version. We take this top-down approach because the general version of Anderson's theorem is much widely applicable than the Gaussian version.

Theorem 7.1. *(General version) Let f be a non-negative, symmetric unimodal, and integrable function on $\mathbb{R}^n$, i.e., $f(\boldsymbol{x}) \geq 0, f(-\boldsymbol{x}) = f(\boldsymbol{x})$, and $\int_{\mathbb{R}^n} f(\boldsymbol{x})d\boldsymbol{x} < \infty$ for all $\boldsymbol{x} \in \mathbb{R}^n$. Let K be a symmetric convex subset of $\mathbb{R}^n$, i.e., K is convex and $K = -K$. Define a function*

$$h(t, \boldsymbol{y}) = \int_{K+t\boldsymbol{y}} f(\boldsymbol{x})d\boldsymbol{x} \quad : \mathbb{R} \times \mathbb{R}^n \to [0, \infty),$$

where $\boldsymbol{y} \in \mathbb{R}^n$ is independently chosen from $\boldsymbol{x}$. Fix $\boldsymbol{y}$. Then:

(a) $h(t, \boldsymbol{y})$ is a symmetric unimodal function w.r.t. t.

(b) $h(t, \boldsymbol{y})$ achieves its maximum value at $t = 0$.

Proof. We prove some mild version. Assume that f is univariate, unimodal, and continuous on $\mathbb{R}$. Without loss of generality, consider an interval $K = [-a, a]$, where $0 < a < \infty$ and positive direction $y = 1$. Then

$$h(t, 1) = \int_{-a+t}^{a+t} f(x)dx.$$

For notational simplicity, let $g(t) = h(t, 1)$. First, we prove that g is an even function.

$$g(-t) = \int_{-a-t}^{a-t} f(x)dx = \int_{a+t}^{-(a-t)} f(-z) - dz = \int_{-a+t}^{a+t} f(z)dz = g(t).$$

Because f is continuous, by the fundamental theorem of calculus, we know that g is differentiable on $\mathbb{R}$, and by unimodality assumption on f (think) we have

$$g'(t) = f(t + a) - f(t - a) \leq 0, \text{for } t > 0.$$

Therefore, g is monotonically decreasing on $(0, \infty)$. By symmetric property (even function) of g, g is monotonically increasing on $(-\infty, 0)$. Because g is differentiable, trivially, g attains its maximum at $t = 0$. □

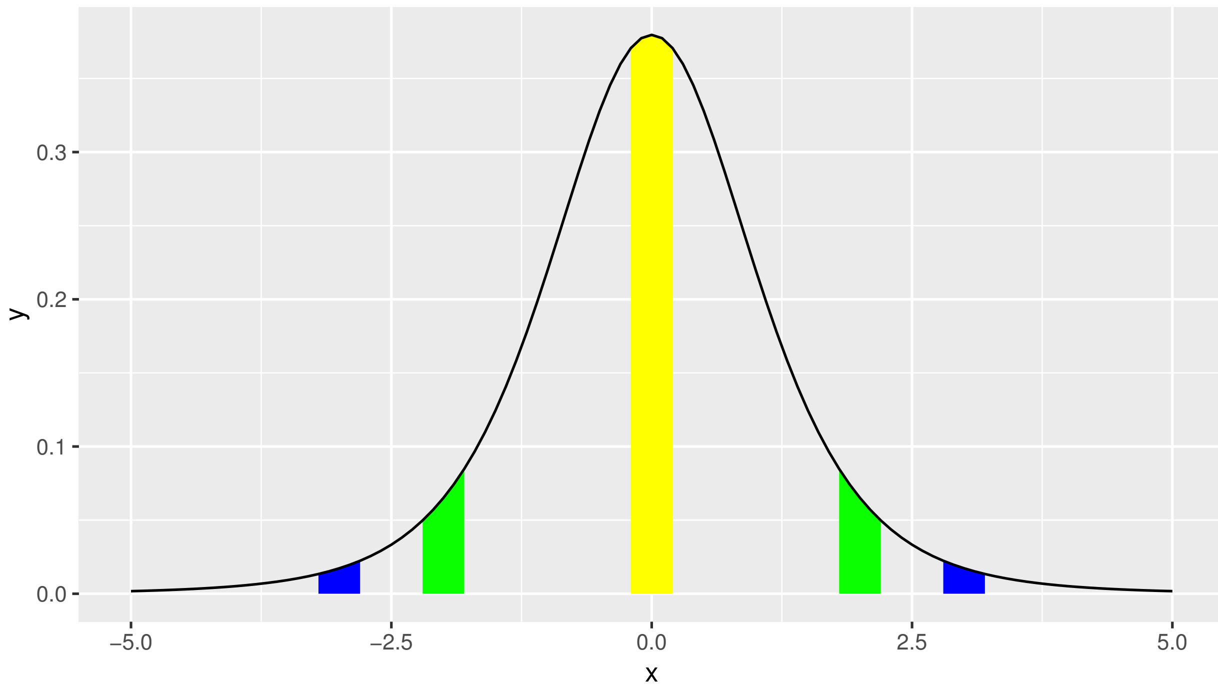

Figure 7.1: Visualization of Anderson's theorem in student t-distribution with df $= 5$, i.e., $f(x) = t_{df=5}(x)$. We selected the convex set $K = [-0.2, 0.2]$ and direction $y = 1$, therefore, $g(t) = h(t, 1) = \int_{-0.2+t}^{0.2+t} t_{df=5}(x)dx$ is a consideration from Anderson's theorem. Three choice of $t = 0, 2$, and 3, are shown in yellow, green, and blue colors, respectively. By Anderson's theorem, $g(t)$ is an even function and attains its maximum at $t = 0$, which corresponds to the yellow region.

The proof for multivariate version is not trivial, and hence omitted. By Theorem 7.1, for any $\mathbf{y} \in \mathbb{R}$, we have $h(t, \mathbf{y}) \le h(0, \mathbf{y}) = \int_K f_{\mathbf{x}} d\mathbf{x}$ so $\mathbf{y}$ disappear in the rhs, which is useful in probability theory to obtain an upper bound in certain situation. The Figure 7.1 shows illustration of Theorem 7.1 using t-density with degree of freedom 5. The beauty of the theorem is its wide applicability because of its fairly lenient assumption on f.

Corollary 7.2. *(Probability measure version) Given a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, suppose that* $X : \Omega \to \mathbb{R}^n$ *is an* $\mathbb{R}^n$*-valued random variable with probability density function* $f : \mathbb{R}^n \to [0, \infty)$*. Assume that* f *is symmetric unimodal and integrable on* $\mathbb{R}^n$*. Let* $K \in \mathbb{R}^n$ *be any origin-symmetric convex body. Let* $Y : \Omega \to \mathbb{R}^n$ *be an random variable independent of* X*. Then*

$$\mathbb{P}[X \in K] \ge \mathbb{P}[X - tY \in K], \quad \textit{for all } t \in \mathbb{R}.$$

Proof. Without loss of generality, assume that Y is continuous random variable with density $h(\mathbf{y})$. Define a function

$$h(t, Y) = \int_{K+tY} f(\mathbf{x})d\mathbf{x} = \int \mathcal{I}_{[\mathbf{x} \in K + t\mathbf{y}]} f(\mathbf{x})d\mathbf{x}.$$

Start with

$$\mathbb{P}[X - tY \in K] = \mathbb{E}[\mathcal{I}_{[X-tY\in K]}] = \mathbb{E}[\mathbb{E}[\mathcal{I}_{[X-tY\in K]}|Y]] = \mathbb{E}[\mathbb{E}[\mathcal{I}_{[X\in K+tY]}|Y]] = \int \left(\int \mathcal{I}_{[\mathbf{x}\in K+t\mathbf{y}]} f(\mathbf{x})d\mathbf{x} \right) h(\mathbf{y})d\mathbf{y}.$$

By Anderson's theorem 7.1, we have

$$\begin{aligned}\mathbb{P}[X - tY \in K] &= \int \left(\int \mathcal{I}_{[\mathbf{x} \in K + t\mathbf{y}]} f(\mathbf{x}) d\mathbf{x} \right) h(\mathbf{y}) d\mathbf{y} \le \int \left(\int \mathcal{I}_{[\mathbf{x} \in K]} f(\mathbf{x}) d\mathbf{x} \right) h(\mathbf{y}) d\mathbf{y} \\ &= \left(\int \mathcal{I}_{[\mathbf{x} \in K]} f(\mathbf{x}) d\mathbf{x} \right) \cdot \int h(\mathbf{y}) d\mathbf{y} = \int \mathcal{I}_{[\mathbf{x} \in K]} f(\mathbf{x}) d\mathbf{x} = \mathbb{P}[X \in K].\end{aligned}$$

□

Many kind of distribution follows the assumptions stated in the Theorem 7.2. Those include Laplace distribution, Gaussian distribution, student t-distribution with some degree freedome, and double generalized Pareto distribution with some shape parameter. However, Cauchy distribution may not applicable because it is not integrable. Note that in the Corollary 7.2, Y is any random quantity which is independent of X. The following version consider Y is the degenerated random quantity, saying $Y = \mathbf{y}$, with of Gaussian probability measure.

Corollary 7.3. *(Gaussian measure version) Let $K \subset \mathbb{R}^n$ be a symmetric convex subset and $X \sim \mathcal{N}_n(\boldsymbol{0}, \Sigma)$ for some covariance matrix Σ. Then for any $\boldsymbol{y} \in \mathbb{R}^n$, we have*

$$\mathbb{P}[X \pm \boldsymbol{y} \in K] \le \mathbb{P}[X \in K].$$

Proof. Use Corollary 7.2 with $X \sim \mathcal{N}_n(\mathbf{0}, \Sigma)$, $Y = \mathbf{y} \in \mathbb{R}^n$ (degenerated random variable), and $t = \pm 1$. □

Denote a δ-Euclidean ball $B_\delta(\mathbf{x}_0) = \{\mathbf{x} \in \mathbb{R}^n | \|\mathbf{x} - \mathbf{x}_0\|_2 < \delta\}$. Suppose $K = B_\delta(\mathbf{0})$ and $\mathbf{y} \in \mathbb{R}^n$. Then $K + \mathbf{y} = B_\delta(\mathbf{y})$. By Corollary 7.3, if $X \sim \mathcal{N}_n(\mathbf{0}, \Sigma)$, then

$$\mathbb{P}[X - \mathbf{y} \in K] = \mathbb{P}[X \in K + \mathbf{y} = B_\delta(\mathbf{y})] = \mathbb{P}[\|X - \mathbf{y}\|_2 < \delta] \le \mathbb{P}[\|X\|_2 < \delta], \quad \text{for any } \delta > 0. \tag{7.1}$$

Graphical illustration is shown in Figure 7.2. As Gaussian distribution is unimodal with symmetric, it is very clearly that the inequality (7.1) should hold for any $\delta > 0$ in intuitive sense. The inequality implies that the probability $\mathbb{P}[\|X - \mathbf{y}\|_2 < \delta]$ is upper bounded by the measure of center region $\mathbb{P}[\|X\|_2 < \delta]$ for any $\delta > 0$ and any $\mathbf{y} \in \mathbb{R}^n$. As $\mathbf{y} \in \mathbb{R}^n$ changes, the the green ball in Figure 7.2 is moving. For this reason, Anderson's result is also often described as a *moving set inequality.*

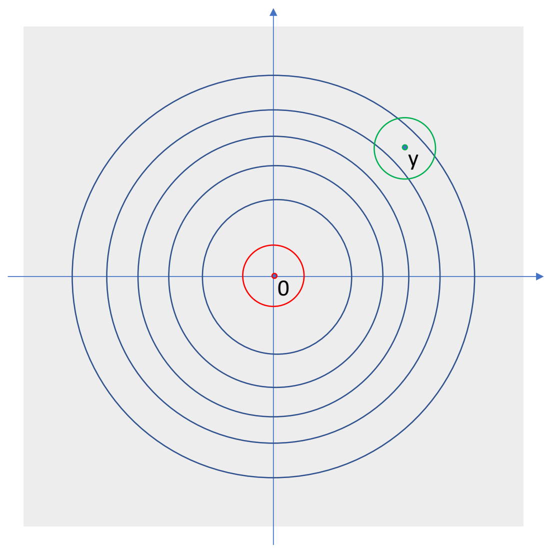

Figure 7.2: Gaussian distribution example for the inequality: $\mathbb{P}[\|X-\mathbf{y}\|_2 < \delta] \le \mathbb{P}[\|X\|_2 < \delta]$, for any $\delta > 0$. The panels shows level curves are of density from $X \sim \mathcal{N}_n(0, \Sigma)$, when $n = 2$.

7.2 Lower bound of probability of shifted small ball

Anderson's inequality provides an upper bound of $\mathbb{P}[\|X-\mathbf{y}\|_2 < \delta]$. Now, we seek to find lower bound of $\mathbb{P}[\|X-\mathbf{y}\|_2 < \delta]$ when $X \sim \mathcal{N}_n(\mathbf{0}, \Sigma)$.

Theorem 7.4. *Let $X \sim \mathcal{N}_n(\boldsymbol{0}, \Sigma)$ and $\boldsymbol{y} \in \mathbb{R}^n$. Then*

$$\mathbb{P}[\|X-\boldsymbol{y}\|_2 < \delta] \ge \exp\left(-\frac{\boldsymbol{y}^\top \Sigma^{-1} \boldsymbol{y}}{2}\right) \cdot \mathbb{P}[\|X\|_2 < \delta], \quad \textit{for any } \delta > 0. \tag{7.2}$$

Proof. Let $K = B_\delta(\mathbf{y}) = \{\mathbf{x} \in \mathbb{R}^n | \|\mathbf{x}-\mathbf{y}\|_2 < \delta\}$. Start with

$$\mathbb{P}[\|X-\mathbf{y}\|_2 < \delta] = \int_K \mathcal{N}_n(\mathbf{x}|\mathbf{0}, \Sigma) d\mathbf{x} = \int_{\|\mathbf{x}-\mathbf{y}\|_2 < \delta} \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{\mathbf{x}^\top \Sigma^{-1} \mathbf{x}}{2}\right) d\mathbf{x}.$$

Let $\mathbf{z} = \mathbf{x} - \mathbf{y}$. Then

$$\begin{aligned}
\mathbb{P}[\|X-\mathbf{y}\|_2 < \delta] &= \int_{\|\mathbf{z}\|_2 < \delta} \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{(\mathbf{z}+\mathbf{y})^\top \Sigma^{-1} (\mathbf{z}+\mathbf{y})}{2}\right) d\mathbf{z} \\
&= \exp\left(-\frac{\mathbf{y}^\top \Sigma^{-1} \mathbf{y}}{2}\right) \cdot \int_{\|\mathbf{z}\|_2 < \delta} \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left\{-\frac{1}{2}(\mathbf{z}^\top \Sigma^{-1} \mathbf{z} + 2\mathbf{z}^\top \Sigma^{-1} \mathbf{y})\right\} d\mathbf{z} \\
&= \exp\left(-\frac{\mathbf{y}^\top \Sigma^{-1} \mathbf{y}}{2}\right) \cdot \int_{\|\mathbf{z}\|_2 < \delta} \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}\mathbf{z}^\top \Sigma^{-1} \mathbf{z}\right) \cdot \exp\left(-\mathbf{z}^\top \Sigma^{-1} \mathbf{y}\right) d\mathbf{z}.
\end{aligned}$$

Use some trick:

$$
\begin{aligned}
I &= \int_{\|\mathbf{z}\|_2<\delta} \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}\mathbf{z}^\top\Sigma^{-1}\mathbf{z}\right)\cdot\exp\left(-\mathbf{z}^\top\Sigma^{-1}\mathbf{y}\right)d\mathbf{z} \\
&= \frac{1}{2}\cdot\int_{\|\mathbf{z}\|_2<\delta} \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}\mathbf{z}^\top\Sigma^{-1}\mathbf{z}\right)\cdot\exp\left(-\mathbf{z}^\top\Sigma^{-1}\mathbf{y}\right)d\mathbf{z} \\
&\quad+\frac{1}{2}\cdot\int_{\|\mathbf{z}\|_2<\delta} \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}\mathbf{z}^\top\Sigma^{-1}\mathbf{z}\right)\cdot\exp\left(-\mathbf{z}^\top\Sigma^{-1}\mathbf{y}\right)d\mathbf{z}.
\end{aligned}
$$

For the second integral, let $\mathbf{t}=-\mathbf{z}$:

$$
\begin{aligned}
&\int_{\|\mathbf{z}\|_2<\delta} \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}\mathbf{z}^\top\Sigma^{-1}\mathbf{z}\right)\cdot\exp\left(-\mathbf{z}^\top\Sigma^{-1}\mathbf{y}\right)d\mathbf{z} \\
&= \int_{\|-\mathbf{t}\|_2<\delta} \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}(-\mathbf{t})^\top\Sigma^{-1}(-\mathbf{t})\right)\cdot\exp\left(-(-\mathbf{t})^\top\Sigma^{-1}\mathbf{y}\right)-d\mathbf{t} \\
&= -\int_{\|\mathbf{t}\|_2<\delta} \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}\mathbf{t}^\top\Sigma^{-1}\mathbf{t}\right)\cdot\exp\left(\mathbf{t}^\top\Sigma^{-1}\mathbf{y}\right)d\mathbf{t} \\
&= \int_{\|\mathbf{t}\|_2<\delta} \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}\mathbf{t}^\top\Sigma^{-1}\mathbf{t}\right)\cdot\exp\left(\mathbf{t}^\top\Sigma^{-1}\mathbf{y}\right)d\mathbf{t},
\end{aligned}
$$

where in the last we used a fact that $K=\{\mathbf{t}\in\mathbb{R}^n|\|\mathbf{t}\|<\delta\}$ satisfies the symmetric convexity, i.e., $-K=-\{\mathbf{t}\in\mathbb{R}^n|\|\mathbf{t}\|<\delta\}=\{-\mathbf{t}\in\mathbb{R}^n|\|\mathbf{t}\|<\delta\}=\{\mathbf{t}\in\mathbb{R}^n|\|\mathbf{t}\|<\delta\}=K$.
Going back to I, we have

$$
\begin{aligned}
I &= \frac{1}{2}\cdot\int_{\|\mathbf{z}\|_2<\delta} \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}\mathbf{z}^\top\Sigma^{-1}\mathbf{z}\right)\cdot\exp\left(-\mathbf{z}^\top\Sigma^{-1}\mathbf{y}\right)d\mathbf{z} \\
&\quad+\frac{1}{2}\cdot\int_{\|\mathbf{z}\|_2<\delta} \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}\mathbf{z}^\top\Sigma^{-1}\mathbf{z}\right)\cdot\exp\left(\mathbf{z}^\top\Sigma^{-1}\mathbf{y}\right)d\mathbf{z} \\
&= \frac{1}{2}\cdot\int_{\|\mathbf{z}\|_2<\delta} \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}\mathbf{z}^\top\Sigma^{-1}\mathbf{z}\right)\cdot\left\{\exp\left(-\mathbf{z}^\top\Sigma^{-1}\mathbf{y}\right)+\exp\left(\mathbf{z}^\top\Sigma^{-1}\mathbf{y}\right)\right\}d\mathbf{z} \\
&\geq \frac{1}{2}\cdot\int_{\|\mathbf{z}\|_2<\delta} \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}\mathbf{z}^\top\Sigma^{-1}\mathbf{z}\right)\cdot 2d\mathbf{z} = \int_{\|\mathbf{z}\|_2<\delta} \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}\mathbf{z}^\top\Sigma^{-1}\mathbf{z}\right)d\mathbf{z} \\
&= \mathbb{P}[\|X\|_2<\delta],
\end{aligned}
$$

where we used $a+1/a\geq 2$. □

Actually, using Cameron-Martin formula and Holder inequality, then Theorem 7.4 can be extended to Gaussian process version, extending to non-parametric statistics.
The following Corollary is summary of Theorems we have worked on so far. In conclusion, by (7.1) which is derived by Anderson's theorem and Theorem 7.4which is derived by using inequality $a+1/a\geq 2$, we have

Corollary 7.5. *Let $X\sim\mathcal{N}_n(\boldsymbol{0},\Sigma)$ and $\boldsymbol{y}\in\mathbb{R}^n$. Then*

$$
\exp\left(-\frac{1}{2}\cdot\boldsymbol{y}^\top\Sigma^{-1}\boldsymbol{y}\right)\cdot\mathbb{P}[\|X\|_2<\delta]\leq\mathbb{P}[\|X-\boldsymbol{y}\|_2<\delta]\leq\mathbb{P}[\|X\|_2<\delta],\quad \textit{for any } \delta>0.
$$

7.3 Toy example of Corollary 7.5 in Bayesian statistics

Suppose $\theta \sim \mathcal{N}_1(0, \sigma^2)$ and $\theta_0 \in \mathbb{R}$ is fixed (non-stochastic). We have [2]

$$\exp\left(-\frac{\theta_0^2}{2\sigma^2}\right) \cdot \mathbb{P}[|\theta| < \delta] \leq \mathbb{P}[|\theta - \theta_0| < \delta] \leq \mathbb{P}[|\theta| < \delta], \quad \text{for any } \delta > 0. \tag{7.3}$$

Figure 7.3 displays the story of inequality (7.3). Two values are crucially important in probability measure of shifted small ball under Gaussianity: σ and θ_0. Having smaller σ or larger θ_0 lead to the lower bound of (7.3) to be small, and eventually, prone to produce numerically zero. In other words, the smaller σ, the more the Gaussian measure concentrated about 0.

$$\sigma_1 \leq \sigma_2 \Longrightarrow \mathbb{P}_{\sigma_1}[K] \geq \mathbb{P}_{\sigma_2}[K], \quad \text{for any symmetric convex subset } K \subset \mathbb{R}.$$

This argument can be extended to multivariate Gaussian distribution[3]: consider $\boldsymbol{\theta}_1 \sim \mathcal{N}_p(\mathbf{0}, \Sigma_1)$ and $\boldsymbol{\theta}_2 \sim \mathcal{N}_p(\mathbf{0}, \Sigma_2)$, then

$$\Sigma_1 \leq \Sigma_2 \Longrightarrow \mathbb{P}_{\Sigma_1}[K] \geq \mathbb{P}_{\Sigma_2}[K], \quad \text{for any symmetric convex subset } K \subset \mathbb{R},$$

where $A \leq B$ means $B - A$ is non-negative definite.

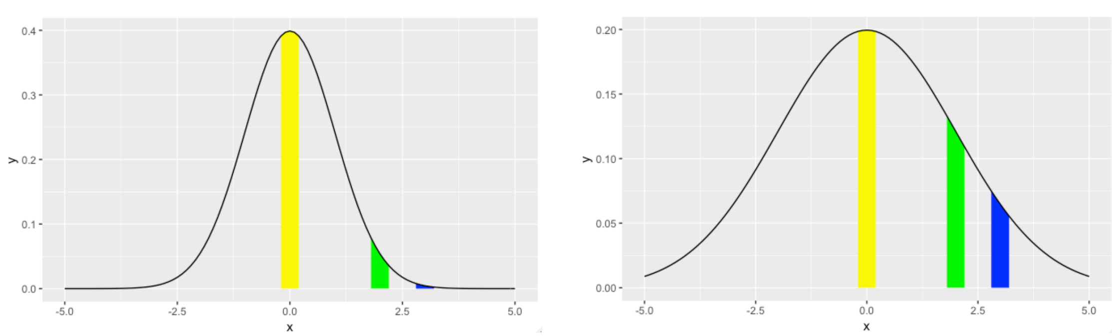

Figure 7.3: Gaussian distribution with $\mathcal{N}_1(0,1)$(left) and $\mathcal{N}_1(0,2)$ (right). $\delta = 0.2$

Examples : [Gaussian Regression]

Consider the data as $(y_i, X_i)_{i=1}^n$ where $X_i \in [0,1]$. We can write the non-parametric regression as following.

$$y_i = f(X_i) + \epsilon_i, \qquad \epsilon_i \sim N(0,1)$$

$$f(t) = \sum_{j=1}^{k} \theta_j B_j(t)$$

where B_j are some fixed basis. Let $\theta^{(k)} = (\theta_1, \ldots, \theta_k)|k \sim N(0, \tau^2 I_k)$. Prior on k is given as π_k and assume that $\pi(k=j) \geq e^{-j \log j}$ for large j.

Let f_0 be the true function. Under restrictive assumption, we can say that there exits a k_0 and $\theta_{01}, \ldots, \theta_{0k_0}$ such that $f_0 = \sum_i^{k_0} \theta_{0j} B_j$. We can also consider more relaxed assumption such as $f_0 \in C^\alpha[0,1]$ where $\alpha > 0$. $C^\alpha[0,1]$ is space of α-Holder smooth function on $[0,1]$. Functions of α-Holder are $\lfloor \alpha \rfloor$-times continuously differentiable and $\lfloor \alpha \rfloor$-th derivative satisfies the following condition.

$$\mid f^{(\lfloor \alpha \rfloor)}(x) - f^{(\lfloor \alpha \rfloor)}(x') \mid \leq C \mid x - x' \mid^{\alpha - \lfloor \alpha \rfloor} \quad \forall x, x' \in [0,1]$$

Comment We can set a hierarchical prior in this non-parametric regression set up but verifying the prior mass condition may be difficult. Under $C^\alpha[0,1]$ space, we can obtain a minimax rate : $\epsilon_n \asymp n^{-\alpha/(2\alpha+1)}$.

Notations: $p_{0i} \equiv N(f_0(X_i), 1)$ and $p_i \equiv N(f(X_i), 1)$.

$$\frac{1}{n}\sum_{i=1}^{n} D(p_{0i} \mid\mid p_i) = \frac{1}{n}\sum_{i=1}^{n}[f(X_i) - f_0(X_i)]^2 = \mid\mid f - f_0 \mid\mid_{2,n}^2$$

Suppose the true density is given by $p_0^{(n)} = \prod_{i=1}^n p_{0,i}$. Now the prior mass condition can verified with following steps.

$$\left[\frac{1}{n}\sum_{i=1}^{n} D(p_{0i} \mid\mid p_i) < \epsilon_n^2, \frac{1}{n}\sum_{i=1}^{n}\{V(p_{0i} \mid\mid p_i) - D^2(p_{0i} \mid\mid p_i)\} < \epsilon_n^2\right]$$

$$= \{f : \mid\mid f - f_0 \mid\mid_{2,n} \leq \epsilon_n^2\} \supset \{f : \mid\mid f - f_0 \mid\mid_\infty \leq \epsilon_n^2\}$$

where $\mid\mid g \mid\mid_\infty = \sup_{x \in [0,1]} \mid g(x) \mid$.
$\{f : \mid\mid f - f_0 \mid\mid_\infty \leq \epsilon_n^2\} = \{$ f and f_0 are closed in sup norm and subset of $\mid\mid f - f_0 \mid\mid_{2,n}\}$.

Fact : Suppose $f_0 \in C^\alpha[0,1]$ with $k_n \asymp n^{1/2\alpha+1} \implies \exists\, \theta_0 = (\theta_{01}, \ldots, \theta_{0k_n})$ such that $f_{0,n}(t) = \sum_{j=1}^{k_n} \theta_{0j} B_j(t)$ where $\mid\mid f_0 - f_{0,n} \mid\mid_\infty \leq \epsilon/2$.

Comment: There is a connection between k_n and ϵ_n. If the true function is not smooth then we need large value of k_n basically we need more basis to express the function. Less no of basis is required in case of a smooth function. Minimax rates are associated with number of basis of functions.

Remark 1: $\mid\mid f - f_0 \mid\mid_\infty \leq \mid\mid f - f_{0,n} \mid\mid_\infty + \mid\mid f_{0,n} - f_0 \mid\mid_\infty \leq \epsilon_n/2 + \epsilon_n/2 = \epsilon_n$

Remark 2: $\mid f - f_{0,n} \mid = \mid \sum_{j=1}^{k_n} \theta_j B_j(t) - \sum_{j=1}^{k_n} \theta_{0j} B_j(t) \mid \leq \sum_{j=1}^{k_n} \mid \theta_j - \theta_{0j} \mid B_j(t) \mid$

$\leq M \sum_{j=1}^{k_n} \mid \theta_j - \theta_{0j} \mid = \mid\mid \theta_j - \theta_{0j} \mid\mid_1$

Verification of prior mass condition in Gaussian regression :

$$\Pi(\| f - f_0 \|_\infty < \epsilon_n) \geq \Pi(\| f - f_0 \|_\infty < \epsilon_n) \geq \Pi(k = k_n)\Pi(\| \theta - \theta_0 \|_1 < \frac{\epsilon_n}{2M} \mid k)$$

$$\geq e^{-k_n log k_n} \prod_{j=1}^{k_n} \Pi(| \theta_j - \theta_{0j} |) \geq e^{-k_n log k_n} \left(\frac{\epsilon_n}{2Mk_n}\right)^{k_n} \asymp e^{-ck_n(log n)^t} \asymp e^{-ck_n\epsilon_n^2}.$$

References

[1] PERLMAN and MICHAEL D, "Anderson's Theorem on the Integral over a symmetric convex set, and its applications in probability and statistics," *Technical report, university of Washington*, 1988, pp. 1–6.

[2] LI, WENBO V and SHAO, Q-M, "Gaussian processes: inequalities, small ball probabilities and applications," *Handbook of Statistics, Elsevier*, 2001, pp. 533–597.

[3] EATON, MORRIS L and PERLMAN, MICHAEL D, "Concentration inequalities for multivariate distributions: I. Multivariate normal distributions," *Statistics & probability letters, Elsevier*, 1991, pp. 487–504.