Lecture-14: Tensorization of experiments

1 Tensor product of experiments

Tensor product is a way to define a high-dimensional model from low-dimensional models.

Definition 1.1. Consider parameter space Θ_i, input space $\mathcal{X}_i$, label space $\mathcal{Y}_i$, prediction space $\mathcal{Y}'_i$, estimand $T_i : \Theta_i \to \mathcal{Y}_i$, estimator $\hat{T}_i : \mathcal{X}_i \times [0,1] \to \mathcal{Y}_i$, and the corresponding loss function $L_i : \mathcal{Y}_i \times \mathcal{Y}'_i \to \mathbb{R}$ for all $i \in [d]$. We respectively define the tensor product of parameter, input, label, and prediction spaces, as

$$\Theta \triangleq \prod_{i\in[d]} \Theta_i, \qquad \mathcal{X} \triangleq \prod_{i\in[d]} \mathcal{X}_i, \qquad \mathcal{Y} \triangleq \prod_{i\in[d]} \mathcal{Y}_i, \qquad \mathcal{Y}' \triangleq \prod_{i\in[d]} \mathcal{Y}'_i.$$

The corresponding tensor product of estimand $T : \Theta \to \mathcal{Y}$, estimator $\hat{T} : \mathcal{X} \times [0,1]^d \to \mathcal{Y}'$, and loss function $L : \mathcal{Y} \times \mathcal{Y}' \to \mathbb{R}$ are defined respectively for all $\theta \in \Theta$, $(X,U) \in \mathcal{X} \times [0,1]$, and $(y,y') \in \mathcal{Y} \times \mathcal{Y}'$, as

$$T(\theta) \triangleq (T_i(\theta_i) : i \in [d]), \qquad \hat{T}(X,U) \triangleq (\hat{T}_i(X,U) : i \in [d]), \qquad L(y,y') \triangleq \sum_{i=1}^{d} L_i(y_i,y'_i).$$

Given statistical experiments $\mathcal{P}_i \triangleq \left\{P_{\theta_i} : \theta_i \in \Theta_i\right\}$ and for each $i \in [d]$, their tensor product refers to the following statistical experiment

$$\mathcal{P} \triangleq \left\{ P_\theta \triangleq \prod_{i=1}^{d} P_{\theta_i} : \theta \in \Theta \right\}.$$

Remark 1. The observation X consists of independent and not identically distributed $X_i \sim P_{\theta_i}$ and the loss function takes a *separable* form. This should be contrasted with the multiple-observation model, in which m *i.i.d.* observations drawn from the same distribution P_θ are given.

Theorem 1.2 (Minimax risk of tensor product). *The following equality holds true for minimax risk of the tensorized experiment*

$$\sum_{i=1}^{d} R_B^*(\mathcal{P}_i) \leqslant R^*(\mathcal{P}) \leqslant \sum_{i=1}^{d} R^*(\mathcal{P}_i).$$

Consequently, if minimax theorem holds for each experiment, i.e., $R^*(\mathcal{P}_i) = R_B^*(\mathcal{P}_i)$, then it also holds for the tensorized experiment, i.e. $R^*(\mathcal{P}) = \sum_{i=1}^{d} R^*(\mathcal{P}_i)$.

Proof. The upper bound follows by taking a sub-class of estimators where $\hat{T}_i(X,U) \triangleq \hat{T}_i(X_i,U_i)$. We can rewrite the minimax risk for the tensorized experiment as

$$\begin{aligned} R^*(\mathcal{P}) &= \inf_{\hat{T}} \sup_{\theta\in\Theta} \mathbb{E}[L(T(\theta),\hat{T}(X,U)) \mid \theta] \leqslant \inf_{\hat{T}} \sup_{\theta\in\Theta} \sum_{i=1}^{d} \mathbb{E}[L_i(T_i(\theta_i),\hat{T}_i(X,U)) \mid \theta] \\ &\leqslant \sum_{i=1}^{d} \inf_{\hat{T}_i(X_i,U_i)} \sup_{\theta_i\in\Theta_i} \mathbb{E}[L_i(T_i(\theta_i),\hat{T}_i(X_i,U_i)) \mid \theta] = \sum_{i=1}^{d} R^*(\mathcal{P}_i). \end{aligned}$$

For the lower bound, we take a product prior $\pi \triangleq \prod_{i=1}^{d} \pi_i$ under which $\theta : \Omega \to \Theta$ is an independent vector, and consequently $X : \Omega \to \mathcal{X}$ is an independent vector. Let $\hat{T}_i(X,U) \triangleq \hat{T}_i(X,U_i)$ where $U : \Omega \to [0,1]^d$ is an *i.i.d.* uniform vector independent of X and θ. Defining $\tilde{U}_i \triangleq ((X_j : j \in [d] \setminus \{i\}),U_i)$, we observe that $\tilde{U}_i$ is independent of X_i from independence of X and independence of U. Further, we can write $\hat{T}_i(X,U_i) = \hat{T}_i(X_i,\tilde{U}_i)$ and we observe that

$$R_{\pi_i}(\hat{T}_i(X_i,\tilde{U}_i)) = \mathbb{E}L_i(T_i(\theta_i),\hat{T}_i(X_i,\tilde{U}_i)) \geqslant \inf_{\hat{T}_i} R_{\pi_i}(\hat{T}_i(X_i,\tilde{U}_i)) = R^*_{\pi_i}.$$

From the fact that sup is greater than mean, and the linearity of expectation, we get

$$\sup_{\theta\in\Theta}\mathbb{E}[L(T(\theta),\hat{T}(X,U))\mid\theta]\geqslant\mathbb{E}[L(T(\theta),\hat{T}(X,U))]=\sum_{i=1}^{d}\mathbb{E}L_i(T_i(\theta_i),\hat{T}_i(X,U))=\sum_{i=1}^{d}R_{\pi_i}(\hat{T}_i(X_i,\tilde{U}_i)).$$

Since the choices of prior π_i and estimator $\hat{T}$ were arbitrary, the lower bound follows. □

Example 1.3 (Unstructured GLM). Consider statistical decision theory simple setting with $\mathcal{Y}=\mathcal{Y}'=\Theta\triangleq\mathbb{R}^d$. An unstructured GLM statistical model $\mathcal{P}\triangleq(\mathcal{N}(\theta,\sigma^2 I_d):\theta\in\Theta)$ with quadratic loss $L:(\theta,\hat{\theta})\mapsto\left\|\theta-\hat{\theta}\right\|_2^2=\sum_{i=1}^d(\theta_i-\hat{\theta}_i)^2$ is simply the d-fold tensor product of the one-dimensional GLM. Since minimax theorem holds for the GLM, Theorem 1.2 shows the minimax risks sum up to $d\sigma^2$.

Remark 2. In general, it is possible that the minimax risk of the tensorized experiment is strictly less than the sum of individual minimax risks. This may appear surprising since X_i only carries information about θ_i and it makes sense intuitively to estimate θ_i based solely on X_i. However, this is not always true.

Example 1.4 (Minimax risk of tensorized experiment strictly less than the sum of individual minimax risks). Consider statistical decision theory simple setting with label space $\mathcal{Y}=\mathcal{Y}'=\Theta\triangleq\mathbb{N}$, observation $X\triangleq\theta Z$ where $Z:\Omega\to\{0,1\}$ is an independent Bernoulli random variable with $\mathbb{E}Z=\frac{1}{2}$, and the loss function $L:(\theta,\hat{\theta})\mapsto\mathbb{1}_{\{\hat{\theta}<\theta\}}$. If $Z=0$, then all information about θ is erased. Therefore for any estimator $\hat{\theta}\triangleq\hat{T}(X,U)$, the risk is lower bounded by

$$R_\theta(\hat{\theta})=P(\{\hat{\theta}<\theta\}\mid\theta)\geqslant P(\{\hat{\theta}<\theta,Z=0\}\mid\theta)=\frac{1}{2}P(\{\hat{\theta}<\theta\}\mid\{Z=0\},\theta).$$

Taking supremum on both sides, we obtain $\sup_\theta R_\theta(\hat{\theta})\geqslant\frac{1}{2}$. It follows that minimax risk $R^*\geqslant\frac{1}{2}$. For an estimator $\hat{T}(X,U)\triangleq X$, we obtain risk $R_\theta(\hat{\theta})=\mathbb{E}[\mathbb{1}_{\{\theta Z<\theta\}}]=\mathbb{E}\mathbb{1}_{\{Z=0\}}=\frac{1}{2}$. It follows that minimax risk $R^*=\frac{1}{2}$. Recall that $R^*_\pi=0$ in this case for any prior $\pi\in\mathcal{M}(\Theta)$.

Next consider the tensor product of two copies of this experiment with $\Theta\triangleq\mathbb{N}^2$ observation $X\triangleq\theta\circ Z$ where *i.i.d.* random vector $Z:\Omega\to\{0,1\}^2$ with $\mathbb{E}Z_1=\frac{1}{2}$, and the loss function $L(\theta,\hat{\theta})=\mathbb{1}_{\{\hat{\theta}_1<\theta_1\}}+\mathbb{1}_{\{\hat{\theta}_2<\theta_2\}}$. Consider the following estimator

$$\hat{\theta}_1=\hat{\theta}_2\triangleq X_1\vee X_2+\mathbb{1}_{\{X_1=X_2=0\}}=(\theta_1\vee\theta_2)Z_1Z_2+\theta_1Z_1\bar{Z}_2+\theta_2\bar{Z}_1Z_2+\bar{Z}_1\bar{Z}_2.$$

Since $\theta_1,\theta_2\in\mathbb{N}$, we can write the indicators

$$\mathbb{1}_{\{\hat{\theta}_1<\theta_1\}}=\mathbb{1}_{\{\theta_2<\theta_1\}}\bar{Z}_1Z_2+\mathbb{1}_{\{1<\theta_1\}}\bar{Z}_1\bar{Z}_2,\qquad\mathbb{1}_{\{\hat{\theta}_2<\theta_2\}}=\mathbb{1}_{\{\theta_1<\theta_2\}}Z_1\bar{Z}_2+\mathbb{1}_{\{1<\theta_2\}}\bar{Z}_1\bar{Z}_2.$$

Since Z is *i.i.d.* Bernoulli random vector with $\mathbb{E}Z_1=\frac{1}{2}$, we get $\mathbb{E}\bar{Z}_1Z_2=\mathbb{E}Z_1\bar{Z}_2=\mathbb{E}\bar{Z}_1\bar{Z}_2=\frac{1}{4}$. Therefore, for any $\theta_1,\theta_2\in\mathbb{N}$, averaging over Z_1,Z_2, we can find the mean loss

$$\mathbb{E}L(\theta,\hat{\theta})=\mathbb{E}\mathbb{1}_{\{\hat{\theta}_1<\theta_1\}}+\mathbb{E}\mathbb{1}_{\{\hat{\theta}_2<\theta_2\}}\leqslant\frac{1}{4}(\mathbb{1}_{\{\theta_1<\theta_2\}}+\mathbb{1}_{\{\theta_2<\theta_1\}}+2)\leqslant\frac{3}{4}.$$

2 Minimax risk of GLM with non-quadratic loss

Lemma 2.1. *Let $Z\sim\mathcal{N}(0,1)$. Then for $1\leqslant q<\infty$, we have $\min_{y\in\mathbb{R}}\mathbb{E}\,|y+Z|^q=\mathbb{E}\,|Z|^q$.*

Proof. We fix $a\in\mathbb{R}_+,y\in\mathbb{R}$, and denote the distribution of random variable Z by $F_Z:\mathbb{R}\to[0,1]$. From the symmetry of F_Z around 0 and unimodality at 0, we observe that

$$P\{|y+Z|\leqslant a\}=F_Z(a-y)-F_Z(-a-y)\leqslant F_Z(a)-F_Z(-a)=P\{|Z|\leqslant a\}.$$

The equality is achieved at $y = 0$, and the result follows from the following observation

$$\mathbb{E}|y+Z|^q = \int_{x\in\mathbb{R}_+} P\{|y+Z|^q > x\}\,dx \geqslant \int_{x\in\mathbb{R}_+} P\{|Z|^q > x\}\,dx = \mathbb{E}|Z|^q.$$

□

Theorem 2.2. *Consider the statistical decision theory simple setting under unstructured Gaussian location model on $\Theta = \mathcal{X} = \mathbb{R}^d$, and* i.i.d. *observation sample $X : \Omega \to \mathcal{X}^m$ with common distribution $\mathcal{N}(\theta, \sigma^2 I_d)$. Denoting $Z \sim \mathcal{N}(0, I_d)$, for $1 \leqslant q < \infty$, we have*

$$\inf_{\hat{\theta}} \sup_{\theta\in\Theta} \mathbb{E}[\|\theta - \hat{\theta}\|_q^q \mid \theta] = \frac{1}{m^{q/2}} \mathbb{E}\|Z\|_q^q.$$

Proof. Consider the statistical decision theory simple setting under unstructured Gaussian location model on $\Theta = \mathcal{X} = \mathbb{R}^d$. Denoting estimate $\hat{\theta} \triangleq \hat{T}(X, U)$, we observe that the loss function $L : \mathcal{Y} \to \mathcal{Y}' \to \mathbb{R}_+$ defined for all $\theta, \hat{\theta} \in \Theta$ as $\|\theta - \hat{\theta}\|_q^q = \sum_{i=1}^d |\theta_i - \hat{\theta}_i|^q$ is separable. We further note that $\mathcal{N}(\theta, I_d)$ is a product distribution. Thus the experiment is a d-fold tensor product of the one-dimensional version. From Theorem 1.2, the minimax risk for this experiment lies between the aggregate minimax and the aggregate worst case Bayes risk for $d = 1$.

For $d = 1$, the upper bound is achieved by the sample mean $\bar{X} \triangleq \frac{1}{m}\sum_{i=1}^m X_i$, which is distributed according to $\mathcal{N}(\theta, \frac{1}{m})$ and is a sufficient statistic for θ. For the lower bound, we consider a Gaussian prior $\pi = \mathcal{N}(0, s)$ for which the posterior distribution is also Gaussian $P_{\theta|X} = \mathcal{N}(\frac{sm}{1+sm}, \frac{s}{1+sm})$. From Lemma 2.1, it follows that the Bayes estimator is simply the conditional mean, and hence the Bayes risk is

$$R^*_\pi = \mathbb{E}|\theta - \mathbb{E}[\theta \mid X]|s^q = \Big(\frac{s}{1+sm}\Big)^{\frac{q}{2}} \mathbb{E}|Z|^q.$$

Taking limit as $s \to \infty$ proves the matching lower bound. □

3 Log-concavity, Anderson's lemma, and exact minimax risk in GLM

Computing the exact minimax risk is frequently difficult especially in high dimensions. Nevertheless, for the special case of unconstrained GLM, the minimax risk is known exactly in arbitrary dimensions for a large collection of loss functions. We have previously seen in Theorem 2.2 that this is possible for loss functions of the form $L : (\theta, \hat{\theta}) \mapsto \|\theta - \hat{\theta}\|_q^q$. Examining the proof of this result, we note that the major limitation is that it only applies to separable loss functions, so that tensorization allows us to reduce the problem to one dimension. This does not apply to (and actually fails) for non-separable loss, since Theorem 1.2, if applicable, dictates the risk to grow linearly with the dimension, which is not always the case. We next discuss a more general result that goes beyond separable losses.

Definition 3.1. A function $\rho : \mathbb{R}^d \to \mathbb{R}_+$ is called *bowl-shaped* if its sublevel set $K_c \triangleq \{x \in \mathbb{R}^d : \rho(x) \leqslant c\}$ is convex and even symmetric i.e. $K_c = -K_c$ for all $c \in \mathbb{R}$.

Theorem 3.2. *Consider the statistical decision theory simple setting for unstructured GLM with $\Theta = \mathcal{X} = \mathcal{Y} = \mathcal{Y}' \triangleq \mathbb{R}^d$,* i.i.d. *observation sample $X : \Omega \to \mathcal{X}^m$ with the common distribution $\mathcal{N}(0, I_d)$, and the loss function be $L(\theta, \hat{\theta}) = \rho(\theta - \hat{\theta})$ where $\rho : \mathbb{R}^d \to \mathbb{R}_+$ is bowl-shapm:ed and lower-semicontinuous. Let $Z \sim \mathcal{N}(0, I_d)$, then the minimax risk is given by*

$$R^* \triangleq \inf_{\hat{\theta}} \sup_{\theta\in\mathbb{R}^d} \mathbb{E}[\rho(\theta - \hat{\theta}) \mid \theta] = \mathbb{E}\rho\Big(\frac{Z}{\sqrt{m}}\Big).$$

Furthermore, the upper bound is attained by $\bar{X} = \frac{1}{m}\sum_{i=1}^m X_i$.

Proof. Let $Z \sim \mathcal{N}(0, I_d)$. To show the upper bound, we take the estimator $\hat{\theta} \triangleq \hat{T}(X) \triangleq \bar{X}$. For this estimator, the distribution of $(\theta - \hat{\theta})$ and $\frac{Z}{\sqrt{m}}$ are identical, and for all $\theta \in \Theta$

$$\mathbb{E}[\rho(\theta - \hat{\theta}) \mid \theta] = \mathbb{E}\rho\Big(\frac{Z}{\sqrt{m}}\Big).$$

We lower bound the minimax risk R^* by the Bayes risk R^*_π with the prior $\pi = \mathcal{N}(0, sI_d)$. We take the estimate $\hat{\theta}^* \triangleq \mathbb{E}[\theta \mid X]$. Under the Gaussian prior π and estimate $\hat{\theta}^* = \mathbb{E}[\theta \mid X]$, we observe that

$\theta - \hat{\theta}^* \sim \mathcal{N}(0, \frac{s}{1+sm})$ which is identical to the distribution of $\sqrt{\frac{s}{1+sm}}Z$. From Anderson's Lemma A.7, we obtain that for bowl shaped functions ρ,

$$\mathbb{E}\rho\Big(\frac{Z}{\sqrt{m}}\Big) = \mathbb{E}_\pi \rho(\theta - \hat{\theta}^*) = \inf_{\hat{\theta}} \mathbb{E}_\pi \rho(\theta - \hat{\theta}^* + \hat{\theta}^* - \hat{\theta}) = R_\pi^*.$$

Since ρ is lower semicontinuous, sending $s \to \infty$ and applying Fatou's lemma, we obtain

$$R^* \geqslant \lim_{s\to\infty} R_\pi^*(s) = \lim_{s\to\infty} \mathbb{E}\rho\Big(\sqrt{\frac{s}{1+sm}}Z\Big) \geqslant \mathbb{E}\rho\Big(\frac{Z}{\sqrt{m}}\Big).$$

□

Corollary 3.3. *Consider a map $\rho : \mathbb{R}^d \to \mathbb{R}_+$ defined as $\rho(x) \triangleq \|x\|^q$ for some $q > 0$ and arbitrary norm $\|\cdot\|$ on $\mathbb{R}^d$. Then $R^* = \frac{\mathbb{E}\|Z\|^q}{m^{\frac{q}{2}}}$.*

Proof. It follows from Theorem 3.2 applied to bowl shaped loss function $\|\theta - \hat{\theta}\|^q$. □

Example 3.4 (Applications of Corollary 3.3). Consider the statistical decision theory simple setting where $\Theta = \mathcal{Y} = \mathcal{Y}'$, and the loss function $L : \theta \times \hat{\theta} \mapsto L(\theta, \hat{\theta}) \triangleq \rho(\theta - \hat{\theta})$ is defined in terms of bowl-shaped loss functions $\rho : \mathbb{R}^d \times \mathbb{R}_+$ for all $x \in \mathbb{R}^d$.

- For $\Theta \subseteq \mathbb{R}^d$ and $\rho(x) \triangleq \|x\|_2^2$, the minimax risk is $R^* \asymp \frac{1}{m}\mathbb{E}\,\|Z\|^2 = \frac{d}{m}$.
- For $\Theta \subseteq \mathbb{R}^d$ and $\rho(x) \triangleq \|x\|_\infty$, we have $\mathbb{E}\,\|Z\|_\infty \asymp \sqrt{\ln d}$ and the minimax risk is $R^* = \sqrt{\frac{d}{m}}$.
- For $\Theta \subseteq \mathbb{R}^{d\times d}$ and $\rho(\theta) = \|\theta\|_{\text{op}}$ denote the operator norm that is the maximum singular value. In this case, $\mathbb{E}\,\|Z\|_{\text{op}} \asymp \sqrt{d}$ and so minimax risk is $R^* = \sqrt{\frac{d}{m}}$.
- For $\Theta \subseteq \mathbb{R}^{d\times d}$ and $\rho(\theta) = \|\theta\|_F$, the minimax risk $R^* \asymp \frac{d}{\sqrt{m}}$.

Remark 3. We can also phrase the result of Corollary 28.8 in terms of the sample complexity n?(?) as defined in (28.16). For example, for q = 2 we have n?(?) = ??E[?Z?2]/???. The above examples show that the scaling of n?(?) with dimension depends on the loss function and the ?rule of thumb? that the sampling complexity is proportional to the number of parameters need not always hold. Finally, for the sake of high-probability (as opposed to average) risk bound, consider ?(? ? ??) = 1?? ? ??? ¿ ?, which is lower semicontinuous and bowl-shaped. Then the exact expression R? = P ???Z? ? ??n??. This result is stronger since the sample mean is optimal simultaneously for all ?, so that integrating over ? recovers (28.20).

A Log-concavity

Definition A.1. A measure $\mu \in \mathcal{M}(\mathbb{R}^d)$ is said to be log-concave if for any $A, B \in \mathcal{B}(\mathbb{R}^d)$ and $\lambda \in [0,1]$, we have

$$\mu(\lambda A + (1-\lambda)B) \geqslant \mu(A)^\lambda \mu(B)^{1-\lambda}.$$

Theorem A.2. *Let $\mathcal{X} \triangleq \mathbb{R}^d$ and $\mu \in \mathcal{M}(\mathcal{X})$ that has a density $f \triangleq \frac{d\mu}{d\text{vol}} \in \mathbb{R}_+^{\mathcal{X}}$ with respect to Lebesgue measure* vol $\in \mathcal{M}(\mathcal{X})$. *Then, μ is log-concave iff f is log-concave.*

Example A.3 (Lebesgue measure). Let $\mu = \text{vol}$ be the Lebesgue measure on $\mathbb{R}^d$,which satisfies Theorem A.2 for $f \equiv 1$. It follows that for any $\lambda \in [0,1]$,

$$\text{vol}(\lambda A + (1-\lambda)B) \triangleq \text{vol}(A)^\lambda \text{vol}(B)^{1-\lambda}.$$

Example A.4 (Guassian measure). Let $\mu \triangleq \mathcal{N}(0,\Sigma)$ with positive semidefinite covariance matrix $\Sigma \succ 0$. Then, it has a log-concave density f, since

$$\ln f(x) = -\frac{d}{2}\ln 2\pi - \frac{1}{2}\ln\det(\Sigma) - \frac{1}{2}x^T\Sigma^{-1}x$$

is concave in x.

Theorem A.5 (Brunn-Minkowski). *Let $d \in \mathbb{N}$ and $\mathcal{X} \triangleq \mathbb{R}^d$. Then for any $A, B \in \mathcal{B}(\mathcal{X}), \lambda \in [0,1]$, we get*

$$\mathrm{vol}(A+B)^{\frac{1}{d}} \geqslant \mathrm{vol}(A)^{\frac{1}{d}} + \mathrm{vol}(B)^{\frac{1}{d}}.$$

Proof. Let $A, B \in \mathcal{B}(\mathcal{X})$, and define two unit volume sets

$$A' \triangleq \mathrm{vol}(A)^{-\frac{1}{d}}A, \qquad B' \triangleq \mathrm{vol}(B)^{-\frac{1}{d}}B.$$

Taking $\lambda \triangleq \frac{\mathrm{vol}(A)^{\frac{1}{d}}}{\mathrm{vol}(A)^{\frac{1}{d}}+\mathrm{vol}(B)^{\frac{1}{d}}}$ and $A', B' \in \mathcal{B}(\mathcal{X})$ in Example A.3, we obtain

$$\frac{\mathrm{vol}(A+B)}{(\mathrm{vol}(A)^{\frac{1}{d}} + \mathrm{vol}(B)^{\frac{1}{d}})^d} = \mathrm{vol}(\lambda A' + (1-\lambda)B') \geqslant \mathrm{vol}(A')^{\lambda}\mathrm{vol}(B')^{1-\lambda} = 1.$$

□

Lemma A.6. *Let $K \subseteq \mathbb{R}^d$ be an even symmetric convex set and $X \sim \mathcal{N}(0,\Sigma)$. Then*

$$\max_{y\in\mathbb{R}^d} P\{X + y \in K\} = P\{X \in K\}.$$

Proof. From Example A.4 and Theorem A.2, it follows that distribution of X is log-concave. Let $y \in \mathbb{R}$, then we observe that $\frac{1}{2}(K+y) + \frac{1}{2}(K-y) = K$ from the convexity of K. From the log-concavity of distribution of X for $\lambda = \frac{1}{2}$ and measurable sets $A \triangleq K+y, B \triangleq K-y$, we obtain

$$P\{X \in K\} = P\left\{X \in \frac{1}{2}(K+y) + \frac{1}{2}(K-y)\right\} \geqslant \sqrt{P\{X \in K+y\}P\{X \in K-y\}}.$$

The equality is obtained for $y = 0$. From even symmetry of K, we have $K = -K$, and hence $\{X \in K-y\} = \{X \in -K-y\}$. Since X has an even symmetric distribution, we obtain $P\{X \in -K-y\} = P\{X \in K+y\}$. It follows that $P\{X \in K\} \geqslant P\{X \in K+y\}$, with the equality at $y = 0$, and the result follows. □

Lemma A.7 (Anderson). *Let $X \sim \mathcal{N}(0,\Sigma)$ for some positive definite $\Sigma \succ 0$ and $\rho : \mathbb{R}^d \to \mathbb{R}_+$ a bowl-shaped loss function. Then,* $\min_{y\in\mathbb{R}^d} \mathbb{E}\rho(y+X) = \mathbb{E}\rho(X)$.

Proof. Denote the sublevel set $K_c \triangleq \left\{x \in \mathbb{R}^d : \rho(x) \leqslant c\right\}$ for each $c \in \mathbb{R}$. Since ρ is bowl- shaped, K_c is convex and even symmetric. For $y \in \mathbb{R}^d$, we write the mean

$$\mathbb{E}\rho(y+X) = \int_{x\in\mathbb{R}_+} P\{\rho(y+X) > x\}\,dx = \int_{x\in\mathbb{R}_+} P\{X + y \notin K_x\}\,dx.$$

From Lemma A.6, we have $\min_{y\in\mathbb{R}^d} P\{X+y \notin K_x\} = P\{X \notin K_x\}$, and it follows that

$$\min_{y\in\mathbb{R}^d} \mathbb{E}\rho(y+X) = \int_{x\in\mathbb{R}_+} P\{X \notin K_x\}\,dx = \int_{x\in\mathbb{R}_+} P\{\rho(X) > x\}\,dx = \mathbb{E}\rho(X).$$

□