High Dimension Probability - On Concentration Inequalities

Bipin Koirala

March 2023

1 Gaussian Concentration Inequality

Let, $X \sim \mathcal{N}(0, I_d) \in R^d$ and $f : R^d \to R$ a Lipschitz function i.e. $|f(X) - f(y)| \leq L\ ||x - y||$. Then;

$$\boxed{\mathbb{P}\big\{|f(X) - \mathbb{E}f(X)| \geq t\big\} \leq 2\ \exp\ (-t^2/2L^2)\ \ ;\forall\ t > 0}$$

Remark

- $f(x) = \langle w, x\rangle$ then $L = ||w||$. If $X \sim \mathcal{N}(0, I_d)$ then $f(x) = \langle w, x\rangle \sim \mathcal{N}(0, ||w||^2)$ and

  $$\mathbb{P}\{f(X) \geq t\} = \mathbb{P}\{f(X) - \mathbb{E}f(X) \geq t\} = 1 - \Phi\bigg(\frac{t}{||w||}\bigg) \text{ if Median or } \mathbb{E}f(X) = 0$$

  where; $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-y^2/2}\ dy$ and $1 - \Phi(t) \leq \frac{1}{2} e^{-t^2/2}$.
  $\rightarrow \mathbb{P}\{|f(X) - \text{Med } f(X)| \geq t\} \leq 2\bigg[1 - \Phi\big(\frac{t}{L}\big)\bigg] = \int_0^\infty e^{-t^2/2L^2} dt = \sqrt{\frac{\pi}{2}} L$
  $\rightarrow \mathbb{P}\{|f(X) - \text{Med } f(X)| \geq tL\} \leq 2\bigg[1 - \Phi\big(t\big)\bigg] \implies \mathbb{P}\big\{|f(X) - \mathbb{E}f(X)| \geq (t + \sqrt{\frac{\pi}{2}})L\big\} \leq 2(1 - \Phi(t))$

- If $X \sim \mathcal{N}(0, \Sigma)$ then $X = \Sigma^{1/2} Z$ where $Z \sim \mathcal{N}(0, I_d)$ then;

  $$\mathbb{P}\{|f(X) - \mathbb{E}f(X)| \geq t\} \leq 2\ \exp\bigg(-\frac{t^2}{2L^2||\Sigma||}\bigg)$$

2 Gaussian Concentration and Isoperimetry

Among all subsets of same surface area; sphere(ball) has the largest volume. Among all subsets of same surface volume; Euclidean ball has the smallest surface area.

Isoperimetry for Gaussian Measure $\gamma(A) := \int_A \frac{1}{(2\pi)^d/2} e^{-|x|^2/2\ dx}$; $A \subset \mathbb{R}^d$ is a Borel set.
i.e. γ is the standard Gaussian measure on Borel σ algebra in $\mathbb{R}^d$. γ is the distribution of $X \sim \mathcal{N}(0, I_d)$.

$$\boxed{\gamma(A) = \mathbb{P}(X \in A)}$$

Let; $A, H \subset \mathbb{R}^d$ be Borel subsets with $H = \{x \in \mathbb{R}^d : \langle w, x\rangle \leq c\}$ with $||w|| = 1$ and $c \in \mathbb{R}$ then;

$$\gamma(A) = \gamma(H) \implies \gamma(A_\epsilon) \geq \gamma(H_\epsilon)\ \ ;\ \forall\ \epsilon > 0$$

$\gamma(H_\epsilon) = \gamma(H_{w,c+\epsilon}) = \Phi(c+\epsilon)$.

$$\boxed{\Phi^{-1}(\gamma(A_\epsilon)) \geq \Phi^{-1}(\gamma(A)) + \epsilon}$$

Suppose; $\gamma(A) \geq 1/2$. Then $\Phi^{-1}(\gamma(A)) \geq \Phi^{-1}(1/2) = 0$ *i.e.* $\Phi^{-1}(\gamma(A)) \geq 0$ so $\gamma(A\epsilon) \geq \Phi(\epsilon)$

$$\gamma((A_\epsilon)^c) = 1 - \gamma(A_\epsilon) \leq 1 - \Phi(\epsilon)$$

$$\boxed{\gamma((A_t)^c) = 1 - \gamma(A_t) \leq \frac{1}{2}e^{-t^2/2}}$$

3 Application of Gaussian Concentration

3.1 Gaussian Stochastic Processes

Note: We're interested in concentration of $\sup\limits_{t\in\tau} |X(t)| := ||X||_\tau$

Let; $X(t)$ s.t. $t \in \tau$ be a centered Gaussian Process. Suppose that $\sup_{(t\in\tau)} |X(t)| < +\infty$. Let M be a median of R.V. $\sup_{t\in\tau} |X(t)|$. Denote: $\sigma^2 := \sup_{t\in\tau} \mathbb{E}\big[X^2(t)\big]$. Then following are true:

i $\mathbb{P}\{\sup_{t\in\tau} |X(t)| \geq M + u\} \leq 1 - \Phi(u/\sigma)$

ii $\mathbb{P}\{\sup_{t\in\tau} |X(t)| \leq M - u\} \leq 1 - \Phi(u/\sigma)$

iii $\mathbb{P}\{\big|\sup_{t\in\tau} |X(t)| - M\big| \geq u\} \leq 2\big(1 - \Phi(u/\sigma)\big); \quad \forall\ u \geq 0$

Corollary:

If; $\left\{\big|\sup_{(t\in\tau)} |X(t)| - M\big| \geq u\right\} \leq e^{-u^2/(2\sigma^2)}$
$\implies \mathbb{E}\big|\sup_{(t\in\tau)} |X(t)| - M\big| \leq \int_0^\infty e^{-u^2/(2\sigma^2)}\ du = \sqrt{\frac{\pi}{2}}\sigma$

$\implies \mathbb{E}\sup_{t\in\tau} |X(t)| \leq M + \sqrt{\frac{\pi}{2}}\sigma \leq +\infty$

Furthermore; $\mathbb{E}\ \exp\ \{\lambda \sup_{t\in\tau} |X(t)|^2\} < \infty; \quad \lambda < \frac{1}{2\sigma^2}$

3.2 Concentration Inequalities for norms of Gaussian R.V in Banach Spaces

E is a Banach Space and E^* is its dual space which consists of all bounded linear functionals on E. A bounded linear functional is a linear transformation that does not "blow up" the size of vectors in the vector space it is defined on. $||T(x)|| \leq M||x||$.

Let; $U \in E^* \Leftrightarrow X \in E \to \langle x, u\rangle \in \mathbb{R}$.

$$||U|| = \sup_{||x||\leq 1} \langle x, u\rangle < +\infty$$

Suppose, $M \subset \{U \in E^* : ||U|| \leq 1\}$. X is Gaussian $\sim \mathcal{N}(\mu, \Sigma) \Leftrightarrow \forall\ U \in E^*$; $\langle x, u\rangle$ is a normal R.V. (stochastic). Σ is covariance operator of X and

$$\langle \Sigma u, v\rangle = \text{Cov}\ \big(\langle x, u\rangle, \langle x, u\rangle\big)\ \ ;\ U, V \in E^*$$

We ar interested in concentration inequalities for

$$||X|| = \sup_{U \in M} |\langle x, u \rangle|$$

Assume that X is centered and $\sigma^2 := \sup_{U \in M} \mathbb{E}\langle x, u\rangle^2 \leq ||\Sigma||$. Let M be the median of X. Then following are true;

i $\mathbb{P}\{||X|| \geq M + U\} \leq 1 - \Phi(U/\sigma)$

ii $\mathbb{P}\{||X|| \leq M - U\} \leq 1 - \Phi(U/\sigma)$

iii $\mathbb{P}\{\big|||X|| - M\big| > U\} \leq 2\big(1 - \Phi(U/\sigma)\big)$; $\forall\, U \geq 0$

For all centered Gaussian R.V. $X \in E$;

$$\mathbb{E}||X|| < +\infty$$

$$\mathbb{E} \exp \{\lambda ||X||^2\} < \infty \quad ; \lambda < 1/(2\sigma^2)$$

3.3 Concentration of Lipschitz functions on the Sphere in $\mathbb{R}^n$

Let; $X \sim$ Unif$(\sqrt{n}S^{n-1})$ and Lipshitz $f : \sqrt{n}S^{n-1} \to \mathbb{R}$. Then, $||f(X) - \mathbb{E}f(X)||_{\psi_2} \leq \frac{||f||_{Lip}}{\sqrt{n}}$ Furthermore, $\mathbb{P}\{|f(X) - \mathbb{E}f(X)| \geq t\} \leq 2 \exp\big(-\frac{ct^2}{||f||^2_{Lip}}\big)$

4 Johnson-Lindenstrauss Lemma

Let $F = \{x_1, ..., x_n\}$ be a finite set of points in $\mathbb{R}^n$. With probability at least $1 - 2\exp(-c\epsilon^2 m)$, a linear operator $A : \mathbb{R}^n \to \mathbb{R}^m$ is an ϵ-isometry $\big(\epsilon \in (0, 1)\big)$ from F into $\mathbb{R}^m$ iff

$$\boxed{(1 - \epsilon)||x - y||_2 \leq ||Ax - Ay||_2 \leq (1 + \epsilon)||x - y||_2 \text{ for all } \; x, y \in F}$$

where $A = \sqrt{\frac{n}{m}}P_L$, P_L is an orthonormal projection onto a random m-dimensional subspace $L \subset \mathbb{R}^n$ and $m \geq (C\epsilon^2) \log N$.

5 Concentration of Functions of Independent R.V.

Suppose, $X_1, ..., X_n$ are independent R.V. with values in some spaces $S_1, ..., S_n$. Let, $f : S_1 \times ..... \times S_n \to \mathbb{R}$ and let $Z := f(X_1, ...X_n)$. We're interested in the concentration of Z around its expectation (or median).

5.1 Martingale Approach

Let $(X'_1, ..., X'_n)$, be independent copies of $(X_1, ..., X_n)$. Denote:

$$Z_i := \mathbb{E}_i Z := \mathbb{E}(Z|X_1, ..., X_n) = \mathbb{E}' f(X_1, ..., X_i, X'_{i+1}, ..., X'_n)$$

$$\mathbb{E}_0 Z = \mathbb{E}Z \quad ; \; \mathbb{E}_n Z = Z$$

- $\mathbb{E}^{(i)} Z := \mathbb{E}(Z|X_j : j \neq i) = \mathbb{E}' f(X_1, ..., X_{i-1}, X'_i, X_{i+1}, ..., X_n)$

- $Z_i := \mathbb{E}_i Z \quad ; \ i = 0, 1, ..., n$
- $\forall \ \ j \leq i; \ \ \mathbb{E}_j Z_j = Z_j$ (Martingale Property)

$\{Z_i\}$ is a Martingale w.r.t. filtration $\mathcal{F}_i := \sigma(X_1, ..., X_n)$. Now;

Martingale Difference

$$\boxed{Z - \mathbb{E}Z = \sum_{i=1}^{n} (Z_i - Z_{i-1})}$$

$\forall \ j < i \ ; \ \ \mathbb{E}_j (Z_i - Z_{i-1}) = 0$

Decomposition of Variance

$$\boxed{Var(Z) = \sum_{i=1}^{n} \mathbb{E}(Z_i - Z_{i-1})^2}$$

5.2 Efron-Stein Inequality (In Various Forms)

Some observations:
$Z_i = \mathbb{E}_i Z \quad$ and $\quad Z_{i-1} = \mathbb{E}_{i-1} Z = \mathbb{E}^{(i)} \mathbb{E}_i Z = \mathbb{E}_i \mathbb{E}^{(i)} Z$
$\Rightarrow \mathbb{E}(Z_i - Z_{i-1})^2 = \mathbb{E}(\mathbb{E}_i (Z - \mathbb{E}^{(i)} Z))^2 \leq \mathbb{E}\mathbb{E}_i (Z - \mathbb{E}^{(i)} Z)^2$ (Jensen Inequality)

(i) $Var(Z) \leq \mathbb{E} \sum_{i=1}^{n} (Z - \mathbb{E}^{(i)} Z)^2$

(ii) $Var(Z) \leq \mathbb{E} \sum_{i=1}^{n} \mathbb{E}^{(i)} (Z - \mathbb{E}^{(i)} Z)^2 = \mathbb{E} \sum_{i=1}^{n} \mathrm{Var}^{(i)} Z$ (Tensorization of Variance)

(iii) $Var^{(i)}(Z) = \frac{1}{2} \mathbb{E}^{(i)} (Z - Z^{(i)})^2 = \mathbb{E}^{(i)} (Z - Z^{(i)})^2 I(Z \geq Z^{(i)}) = \mathbb{E}^{(i)} (Z - Z^{(i)})^2 I(Z \leq Z^{(i)})$

(iv) $Var(Z) \leq \frac{1}{2} \mathbb{E} \sum_{i=1}^{n} (Z - Z^{(i)})^2$

(v) $Var(Z) \leq \mathbb{E} \sum_{i=1}^{n} (Z - Z^{(i)})^2 \ I(Z \geq Z^{(i)})$

5.3 Bounded Difference Condition (B.D.C)

{See also Page 36*}
There existis $C_j > 0; \quad j = 1, 2, ..., n$. Then; $\forall \ \ x_1, ..., x_j, x'_j, ..., x_n$

$$|f(...x_{j-1}, x_j, x_{j+1}, ...) - f(..., x_{j-1}, x'_j, x_{j+1}, ...)| \leq c_j$$

Under Bounded Difference Condition, $Z \sim S.G$

$$\boxed{Var(Z) \leq \frac{1}{4} \sum_{j=1}^{n} c_j^2}$$

5.3.1 Hoeffding Type Bounds under B.D.C

$$\mathbb{P}\{|Z - \mathbb{E}Z| \geq t\} \leq 2 \exp\left(- \frac{2t^2}{\sum_{i=1}^{n} c_i^2} \right) \quad \forall \ \ t \geq 0$$

6 Poincaré Inequality

6.1 Gaussian Poincaré Inequality

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function and assume that its 2^{nd} derivative is bounded i.e. $\sup_x |\frac{\partial^2}{\partial x_i^2} f(x)| \leq K < \infty$. Let $X \sim \mathcal{N}(0, I_n)$. Then,

$$Var\big(f(X)\big) \leq \mathbb{E}\Big[||\nabla f(X)||_2^2\Big]$$

See Proof in Exercise Book

6.2 Poincaré Inequality and Sub-exponential Concentrations

$X \sim \mathbb{R}^d$ is a R.V. Assume, $\forall \quad f : \mathbb{R}^d \to \mathbb{R}$ is locally Lipschitz with contant $L > 0$ and $Var(f(X)) \leq C^2 \mathbb{E}||(\nabla f)(X)||^2; \quad C > 0$. Then,

$$||f(X) - \mathbb{E}f(X)||_{\psi_1} \leq CL$$

---

7 Exercises

5.1.8 (Blow-up)

We want to show that if $x \in \sqrt{n}S^{n-1}$ and $x_1 \leq t/\sqrt{2}$, then x belongs to H_t. First, note that the hemisphere H consists of all points $x \in \sqrt{n}S^{n-1}$ such that $x_1 \leq 0$. Therefore, H contains all points x that belong to the t-neighborhood of the plane $x_1 = 0$, which is the hemisphere of radius $t/\sqrt{2}$ centered at the origin.

By the isoperimetric inequality, we have $\sigma(A_t) \geq \sigma(H_t)$ for any subset A of $\sqrt{n}S^{n-1}$ such that $\sigma(A) \geq 1/2$, where σ is the uniform probability measure on the sphere. In particular, this holds for the set A of all points $x \in \sqrt{n}S^{n-1}$ such that $x_1 \leq t/\sqrt{2}$, since $\sigma(A) = \text{area}(H_{t/\sqrt{2}}) \geq 1/2$. Therefore, we have:

$$\sigma(H_t) \leq \sigma(A_t) \leq 1/2.$$

Since $\sigma(H_t)$ is the probability that a uniformly random point X on the sphere belongs to H_t, it follows that H_t contains at least half of the points on the sphere. In other words, if we choose a point X uniformly at random on the sphere, then there is at least a 1/2 probability that X belongs to H_t.

Now, let $x \in \sqrt{n}S^{n-1}$ be a point such that $x_1 \leq t/\sqrt{2}$. We want to show that $x \in H_t$, i.e., $x_1 \leq 0$.

Consider the point y obtained by reflecting x across the hyperplane $x_1 = t/\sqrt{2}$, i.e., $y = (x_1 - t/\sqrt{2}, x_2, \ldots, x_n)$. Note that y lies on the sphere and $y_1 = -x_1 + t/\sqrt{2} \geq t/\sqrt{2} - t/\sqrt{2} = 0$. Therefore, $y \in H_t$, and by the argument above, there is at least a 1/2 probability that a uniformly random point X on the sphere belongs to H_t. Since x and y are symmetric with respect to the hyperplane $x_1 = t/\sqrt{2}$, it follows that $\mathbb{P}(X \in H_t) \geq 1/2$, and hence $x \in H_t$.

5.1.9 (Blow-up exponentially small sets)

If the conclusion of the first part fails, then $B = (A_s)^C$ satisfies $\sigma(B) \geq 1/2$. By Lemma 5.1.7, $\sigma(B_s) \geq 1 - 2e^{-cs^2}$, which is absurd, since $B_s \subset A^C$ and $\sigma A^C < 1 - 2e^{-cs^2}$ by assumption. This justifies that $\sigma(A_s) > 1/2$. Note that $A_{2t} \supset A_{s+t} = (A_s) + t$, and thus the second part follows.

5.1.12 (Concentration for the Unit Sphere)

(a)

$$\begin{aligned} ||f(X) - \mathbb{E}f(X)||_{\psi_2} \leq C||f||_{Lip} &= C \inf_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|} \\ &= C\sqrt{n} \inf_{x \neq y} \frac{|f(x) - f(y)|}{\sqrt{n}|x - y|} = C'||f||_{Lip} \end{aligned}$$

(b) On $\sqrt{n}S^{n-1} \to ||f||_{Lip}$ and on $S^{n-1} \to ||f||_{Lip}/\sqrt{n}$. Hence;

$$\mathbb{P}\{|f(X) - \mathbb{E}(X)| \geq t\} \leq 2\exp\left(\frac{-cnt^2}{||f||_{Lip}^2}\right)$$

5.1.13 (Concentration about the expectation and concentration about the median are equivalent)

Median M of $f(X)$. Show that th concentration of median imply the concentration of mean

$$||f(X) - M||_{\psi_2} \leq C \implies ||f(X) - \mathbb{E}f(X)||_2 \leq C_1$$

$$\begin{aligned} ||f(X) - \mathbb{E}f(X)||_2 &= ||(f(X) - M) - (\mathbb{E}f(X) - M)||_{\psi_2} \\ &= ||(f(X) - M) - \mathbb{E}(f(X) - M)||_{\psi_2} \\ &\leq C'||f(X) - M||_{\psi_2} \\ &\leq C'C = C_1 \end{aligned}$$

Now show that the concentration about the expectation and concentration about the median are equivalent. Consider a r.v. Z with median M. Prove:

$$c||Z - \mathbb{E}Z||_{\psi_2} \leq ||Z - M||_{\psi_2} \leq C||Z - \mathbb{E}Z||_{\psi_2}$$

First inequality already proved. For second inequality, use the definition of the median and Jensen's Inequality. **Also See Note**

5.1.14 (Concentration and Blow-up are equivalent)

By exercise 5.1.13, $||f(X) - M||_{\psi_2} \leq CK||f||_{Lip}$. Consider,

$$f_A(x) = d(x, A) = \inf\{d(x, y) : y \in A\}$$

It is clear that $||d(x, A) - d(y, A)|| \leq d(x, y)$, so $||f_A||_{Lip} \leq 1$. Since, $f_A|_A = 0$ and $\mathbb{P}(X \in A) \geq 1/2$, we have $Mf_A(X) = 0$ It follows that

$$||d(X, A)||_{\psi_2} \leq CK$$

Note that $\{X \in A_t\} = \{d(X, A) \leq t\}$, and thus completes the proof.

5.1.15 (Exponential set of mutually almost orthogonal points)

Hint 1: "there exist"–means that we can use probabilistic way to show this happen with positive probability. First choose $k = \exp(\epsilon^2 n/4)$ vectors $v_1, ..., v_k$ by choosing each coordinate to be ± 1 w.p. 1/2 each. Then define, $u_i = v_i/\sqrt{n}$. Chernoff bound shows that the probability $|\langle u_i, u_j \rangle| \geq \epsilon$ is at most $2\exp(-(\epsilon^2/2)n)$. This equals $2/k^2$ choice of k, and hence one can take the union-bound over the at most $\binom{k}{2} < k^2/2$ pairs (i, j) to show that there is a positive probability of $|\langle u_i, u_j \rangle| < \epsilon$ holding for all $i \neq j$.

Hint 2: Recall $X \sim Unif(\sqrt{n}S^{n-1})$ satisfies $||X||_{\psi_2} \leq C$. For any given point $x_0 \in S^{n-1}$, denote $C(x_0, \epsilon) = \{x \in S^{n-1} : \langle x_0, x \rangle > \epsilon\}$. We have $\sigma(C(x_0, \epsilon)) = \mathbb{P}\{\langle x_0, X \rangle > \sqrt{n}\epsilon\} \leq \exp(-cn\epsilon^2)$. Let $\{x_1, ..., x_N\}$ be a maximal collection of unit vectors in $\mathbb{R}^n$ that are mutually almost orthogonal, then $S^{n-1} \subset \bigcup_{i=1}^N C(x_i, \epsilon)$. It follows that $1 \leq \sum_{i=1}^N \sigma(C(x_i, \epsilon)) \leq N \exp(-cn\epsilon^2)$ i.e. $N \geq \exp(c\epsilon^2 n)$

5.2.3 (Deduce the Gaussian concentration inequality (Theorem 5.2.2) from the Gaussian isoperimetric inequality (Theorem 5.2.1))

Hint 1: Use Gaussian isoperimetric inequality, and gaussian measure is rotation-invariance, we can rotate the half space and focus only on first coordinates.

Answer 2: Consider $X \sim \mathcal{N}(0, I_n)$ and $f : \mathbb{R}^n \to \mathbb{R}$ with $||f||_{Lip} = 1$. Let $A = \{x \in \mathbb{R}^n : f(x) \leq M\}$ where M is median of $f(X)$. Then $\gamma_n(A) \geq 1/2$ and thus

$$\gamma_n(A_t) \geq \Phi(\Phi^{-1}(\gamma_n(A)) + t) \geq \Phi(t) \geq 1 - \exp(-t^2/2) \quad \forall t > 0$$

On the other hand, we have $\gamma_n(A_t) = \mathbb{P}\{X \in A_t\} \leq \mathbb{P}\{f(X) \leq M + t\}$ since f is 1-Lipschitz. Combining these inequalities gives

$$\mathbb{P}\{f(X) - M > t\} = 1 - \mathbb{P}\{f(X) - M \leq t\} \leq \exp(-t^2/2)$$

Repeating the argument for $-f$, we obtain the same bound for $\mathbb{P}\{f(X) - M < -t\}$. By Exercise 5.1.13, $||f(X) - \mathbb{E}f(X)||_{\psi_2} \leq ||f(X) - M||_{\psi_2}$ and therefore $||f(X) - \mathbb{E}f(X)||_{\psi_2} \leq C$

5.2.4 (Replacing Expectation by L^p norm)

Replace $\mathbb{E}f(X)$ by $(\mathbb{E}f^p)^{1/p}$, for non-negative Lipschitz function f, show a similar result, the constants may depend on p.

Hint: Let $Z = f(X) \geq 0$, $||Z - \mathbb{E}Z||_p \geq ||Z||_p - ||\mathbb{E}Z||_p = ||Z||_p - \mathbb{E}Z$.

$$||Z - ||\mathbb{E}Z||_p||_{\psi_2} \leq ||Z - \mathbb{E}Z||_{\psi_2} + ||EZ - ||Z||_p||_{\psi_2} \leq C||f||_{Lip} + C'|EZ - ||Z||_p|$$

$$|EZ - ||Z||_p| \leq ||EZ - Z||_p < C''\sqrt{p}||Z - EZ||_{\psi_2} < C'C\sqrt{p}||f||_{Lip}$$

5.2.11 (Pushing forward the Gaussian to the uniform distribution)

$\Phi(x)$ is c.d.f of $\mathcal{N}(0, 1)$. Consider a random vector $Z := (Z_1, ..., Z_n) \sim \mathcal{N}(0, I_n)$. Show:

$$\phi(Z) := \big(\phi(Z_1), ..., \phi(Z_n)\big) \sim \text{Unif}([0, 1]^n)$$

Let $U = \phi(Z)$. We want to show that U is uniformly distributed on $[0,1]^n$. Note pdf of Z is

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp(-||z||^2/2)$$

The cumulative distribution function of U is given by:

$$\begin{aligned}
\mathbb{P}(U_1 \leq u_1, \ldots, U_n \leq u_n) &= \mathbb{P}(\phi(Z_1) \leq u_1, \ldots, \phi(Z_n) \leq u_n) \\
&= \mathbb{P}(Z_1 \leq \Phi^{-1}(u_1), \ldots, Z_n \leq \Phi^{-1}(u_n)) \\
&= \int_{-\infty}^{\Phi^{-1}(u_1)} \cdots \int_{-\infty}^{\Phi^{-1}(u_n)} f_Z(z_1, \ldots, z_n) dz_1 \cdots dz_n.
\end{aligned}$$

To show that U is uniformly distributed on $[0,1]^n$, it suffices to show that the above probability is equal to the volume of the hypercube $[0,1]^n$, which is equal to 1.

By the change of variables formula, we have:

$$\begin{aligned}
&\int_{-\infty}^{\Phi^{-1}(u_1)} \cdots \int_{-\infty}^{\Phi^{-1}(u_n)} f_Z(z_1, \ldots, z_n) dz_1 \cdots dz_n \\
&= \int_0^{u_1} \cdots \int_0^{u_n} f_Z(\Phi^{-1}(x_1), \ldots, \Phi^{-1}(x_n)) \left| \frac{\partial(z_1, \ldots, z_n)}{\partial(x_1, \ldots, x_n)} \right| dx_1 \cdots dx_n \\
&= \int_0^{u_1} \cdots \int_0^{u_n} \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^n (\Phi^{-1}(x_i))^2\right) \left| \frac{\partial(z_1, \ldots, z_n)}{\partial(x_1, \ldots, x_n)} \right| dx_1 \cdots dx_n \\
&= \int_0^{u_1} \cdots \int_0^{u_n} \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^n (\Phi^{-1}(x_i))^2\right) \prod_{i=1}^n \phi(\Phi^{-1}(x_i)) dx_1 \cdots dx_n \\
&= \int_0^{u_1} \phi(\Phi^{-1}(x_1)) dx_1 \cdots \int_0^{u_n} \phi(\Phi^{-1}(x_n)) dx_n,
\end{aligned}$$

In the last step, we used the fact that $\phi(\Phi^{-1}(x_i))$ is the density of the standard normal distribution evaluated at $\Phi^{-1}(x_i)$.

Now, we can use the substitution $y_i = \Phi^{-1}(x_i)$, so that $x_i = \Phi(y_i)$ and $dx_i = \phi(y_i)dy_i$. This gives us:

$$\begin{aligned}
&\int_0^{u_1} \phi(\Phi^{-1}(x_1)) dx_1 \cdots \int_0^{u_n} \phi(\Phi^{-1}(x_n)) dx_n \\
&= \int_{-\infty}^{\Phi^{-1}(u_1)} \phi(y_1) dy_1 \cdots \int_{-\infty}^{\Phi^{-1}(u_n)} \phi(y_n) dy_n \\
&= \int_{-\infty}^{\infty} \phi(y_1) \mathbb{I} y_1 \leq \Phi^{-1}(u_1) dy_1 \cdots \int -\infty^{\infty} \phi(y_n) \mathbb{I}_{y_n \leq \Phi^{-1}(u_n)} dy_n \\
&= \mathbb{P}(Z_1 \leq \Phi^{-1}(u_1), \ldots, Z_n \leq \Phi^{-1}(u_n)) \\
&= \mathbb{P}(U_1 \leq u_1, \ldots, U_n \leq u_n),
\end{aligned}$$

where $\mathbb{I}_A$ is the indicator function of the event A.

Therefore, we have shown that $\phi(Z)$ is uniformly distributed on $[0,1]^n$.

To show that $\phi(Z)$ is uniformly distributed over $[0,1]^n$, we need to show that for any $\mathbf{u} = (u_1, u_2, \ldots, u_n) \in [0,1]^n$, we have

$$\mathrm{P}(\phi(Z_1) \le u_1, \phi(Z_2) \le u_2, \ldots, \phi(Z_n) \le u_n) = u_1, ..., u_n$$

Since $\phi(x)$ is the cumulative distribution function of $\mathcal{N}(0,1)$, we have $\phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$. Let $U_i = \phi(Z_i)$, then U_i is a random variable with uniform distribution on $[0,1]$.

Therefore, we have

$$\begin{aligned}
&\mathrm{P}(\phi(Z_1) \le u_1, \phi(Z_2) \le u_2, \ldots, \phi(Z_n) \le u_n) \\
&= \mathrm{P}(U_1 \le u_1, U_2 \le u_2, \ldots, U_n \le u_n) \\
&= \mathrm{P}(\phi(Z_1) \le u_1) \cdot \mathrm{P}(\phi(Z_2) \le u_2) \cdots \mathrm{P}(\phi(Z_n) \le u_n) \\
&= \prod_{i=1}^{n} \mathrm{P}(\phi(Z_i) \le u_i) \\
&= \prod_{i=1}^{n} \int_{-\infty}^{\Phi^{-1}(u_i)} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \qquad \text{(using inverse transform method)} \qquad = \prod_{i=1}^{n} u_i \\
&= u_1 u_2 \cdots u_n.
\end{aligned}$$

Therefore, we have shown that $\phi(Z)$ is uniformly distributed over $[0,1]^n$.

5.3.3

The Johnson-Lindenstrauss lemma states that for any $0 < \epsilon < 1$ and $0 < \delta < 1$, if n points in $\mathbb{R}^d$ are embedded into $\mathbb{R}^k$ using a random projection matrix P with $k = O(\frac{1}{\epsilon^2} log(\frac{1}{\delta}))$, then the distortion is preserved with probability $1-\delta$. That is, for any x, y in the n points, we have $(1-\epsilon)|x-y|_2^2 \le |Px - Py|_2^2 \le (1+\epsilon)|x-y|_2^2$.

Now consider the matrix $Q = \frac{1}{\sqrt{m}} A$. Let x, y be two vectors in $\mathbb{R}^n$, and let $P = Q^T Q$. Then we have

$$\begin{aligned}
|Px - Py|_2^2 &= (x-y)^T P^T P (x-y) && (1)\\
&= (x-y)^T (Q^T Q) Q^T Q (x-y) && (2)\\
&= ((Qx)^T (Qy))^T ((Qx)^T (Qy)) && (3)\\
&= |Qx|_2^2 |Qy|_2^2 - 2(Qx)^T (Qy))^T ((Qx)^T (Qy)) && (4)\\
&= |x|_2^2 |y|_2^2 - 2(Qx)^T (Qy))^T ((Qx)^T (Qy)) && (5)\\
&= |x|_2^2 |y|_2^2 - 2(Qy)^T (Qx))^T ((Qy)^T (Qx)) && (6)\\
&= |x|_2^2 |y|_2^2 - 2(Qy)^T (Qx)(Qx)^T (Qy) && (7)\\
&= |x|_2^2 |y|_2^2 - 2\frac{1}{m} y^T A^T A x && (8)\\
&= |x|_2^2 |y|_2^2 - 2\frac{1}{m} \langle Ax, Ay \rangle && (9)\\
& && (10)
\end{aligned}$$

Now we need to bound $\frac{1}{m}\langle Ax, Ay \rangle$. Since A has independent, mean-zero sub-gaussian isotropic random vectors as its rows, we have $\mathbb{E}[A_{i,j}] = 0$ and $\mathbb{E}[A_{i,j}^2] = \frac{1}{n}$, and for any vector v with $|v|_2 = 1$, we have $\mathbb{E}[|\langle A_i, v\rangle|^2] \leq K$ for some constant K. This implies that A_i is a sub-gaussian isotropic random vector with sub-gaussian norm at most $\sqrt{K}$. Thus, we have

$$
\begin{aligned}
&= \frac{1}{m}\sum_{i=1}^{m}\mathbb{E}[\frac{1}{2}(\langle A_i, x+y\rangle^2 - |A_i(x-y)|2^2)] && (11)\\
&= \frac{1}{2m}\sum_{i=1}^{m}(|x+y|_2^2 - |x-y|2^2) && (12)\\
&= \frac{1}{2m}\sum_{i=1}^{m}(4\langle x, y\rangle) && (13)\\
&= \frac{2}{m}\langle x, y\rangle && (14)\\
& && (15)
\end{aligned}
$$

where the second equality follows from the polarization identity, and the third equality follows from the definition of a sub-gaussian random vector.

Therefore, we have

$$
\begin{aligned}
|Px - Py|_2^2 &= |x|_2^2|y|_2^2 - 2\frac{1}{m}\langle Ax, Ay\rangle && (16)\\
&= |x|_2^2|y|_2^2 - \frac{4}{m}\langle x, y\rangle && (17)\\
&= |x|_2^2|y|_2^2 - \frac{4}{\sqrt{m}}\langle Qx, Qy\rangle && (18)
\end{aligned}
$$

Thus, we have shown that the conclusion of the Johnson-Lindenstrauss lemma holds for $Q = \frac{1}{\sqrt{m}}A$ with distortion parameter $\epsilon = \frac{2}{\sqrt{m}}$ and probability $1-\delta$ with $k = O(\frac{1}{\epsilon^2}\log(\frac{1}{\delta})) = O(\frac{m}{4}\log(\frac{1}{\delta}))$.

We want to show that with high probability, the distance between projections onto m random vectors $a_1, a_2, \ldots, a_m$ chosen from $Q = \frac{1}{\sqrt{m}}A$ is preserved up to a small constant factor, where A is an $m \times n$ random matrix whose rows are independent mean-zero sub-gaussian isotropic random vectors in $\mathbb{R}^n$.

Let $x, y \in \mathbb{R}^n$ be any two vectors. Let P be the projection matrix onto the span of $a_1, a_2, \ldots, a_m$, and let $P^\perp = I - P$ be the projection matrix onto the orthogonal complement of the span. Then we can write x and y as $x = Px + P^\perp x$ and $y = Py + P^\perp y$.

Let $z = Px - Py$. Then we have $|z|_2 = |Px - Py|_2 = |P(x-y)|_2 \leq |x-y|_2$, since P is a projection matrix and thus can only make vectors shorter. Therefore, it suffices to show that with high probability, $|Qz|_2$ is small.

Now, note that z is a vector in the span of $a_1, a_2, \ldots, a_m$. In other words, we can write $z = Av$ for some $v \in \mathbb{R}^m$. Then we have

$$
|Qz|_2^2 = \frac{1}{m}|Av|_2^2 = \frac{1}{m}v^TA^TAv
$$

By the matrix Bernstein inequality, we know that with probability at least $1-\delta$, we have

$$
\left|\frac{1}{m}A^TA - I_n\right|_2 \leq \sqrt{\frac{2\log(2n/\delta)}{m}}
$$

where I_n is the $n \times n$ identity matrix. In particular, this implies that with probability at least $1-\delta$, all the eigenvalues of $\frac{1}{m}A^TA$ are bounded between $1-\sqrt{\frac{2\log(2n/\delta)}{m}}$ and $1+\sqrt{\frac{2\log(2n/\delta)}{m}}$. Therefore, we have

$$\mathbb{P}\left[|Qz|_2^2 \geq (1+\epsilon)|z|_2^2\right] \leq \delta$$

where $\epsilon = 2\sqrt{\frac{2\log(2n/\delta)}{m}}$.

Finally, by the triangle inequality, we have

$$\begin{aligned}
|Px - Py|_2 &= |z|_2 \\
&\leq |Qz|_2 + |(I-Q)z|_2 \\
&\leq \sqrt{\frac{1+\epsilon}{m}}|z|_2 + \sqrt{1-\frac{1+\epsilon}{m}}|z|_2 \\
&= \sqrt{2-\epsilon}|z|_2
\end{aligned}$$

8 Applications of Efron-Stein

2.

Let $X_1, ..., X_n$ be independent r.v. in a linear normed space. Prove that

$$Var(||X_1 + ... + X_n||) \leq 4\sum_{i=1}^{n} \mathbb{E}||X_i - \mathbb{E}X_i||^2$$

Proof: The Efron-Stein inequality states that for any function f of n independent random variables $X_1, \ldots, X_n$, we have

$$\mathrm{Var}(f) \leq \sum_{i=1}^{n} \mathrm{Var}(f|X_1, \ldots, X_i')$$

where X_i' is an independent copy of X_i. To apply this inequality to our problem, let $f(X_1, \dots, X_n) = |\sum_{i=1}^n X_i|$. Then we have

$$\begin{aligned}
\mathrm{Var}(|\sum_{i=1}^n X_i|) &= \mathrm{Var}(f(X_1, \dots, X_n)) \leq \sum_{i=1}^n \mathrm{Var}(f|X_1, \dots, X_i') \\
&= \sum_{i=1}^n \mathrm{Var}(|\sum_{j=1}^i X_j + \sum_{j=i+1}^n X_j||X_1, \dots, X_i') \\
&= \sum_{i=1}^n \mathrm{Var}(|\sum_{j=1}^i X_j| + |\sum_{j=i+1}^n X_j||X_1, \dots, X_i') \\
&= \sum_{i=1}^n \mathrm{Var}(|\sum_{j=1}^i X_j||X_1, \dots, X_i') \\
&\leq \sum_{i=1}^n \mathbb{E}\left[\left|\sum_{j=1}^i X_j - \mathbb{E}\left(\sum_{j=1}^i X_j|X_1, \dots, X_i'\right)\right|^2\right] \\
&= \sum_{i=1}^n \mathbb{E}\left[\left|\sum_{j=1}^i X_j - i\mathbb{E}(X_1)\right|^2\right] \\
&= \sum_{i=1}^n i\mathbb{E}\left[|X_i - \mathbb{E}(X_1)|^2\right] \\
&= \sum_{i=1}^n \mathbb{E}\left[|X_i - \mathbb{E}(X_i)|^2\right] \\
&\leq 4\sum_{i=1}^n \mathbb{E}\left[|X_i - \mathbb{E}(X_i)|^2\right]
\end{aligned}$$

where we used the triangle inequality, the fact that $|\sum_{j=i+1}^n X_j| \geq 0$, the conditional variance formula, the fact that $|\cdot|$ is a norm and satisfies the triangle inequality, the law of total expectation, and the fact that $i\mathbb{E}(X_1) = \mathbb{E}(\sum_{j=1}^i X_j)$. This completes the proof.

3.

Let $X_1, ..., X_n$ be independent r.v, where X_i is a Bernoulli with parameter $p_i \in [0, 1]$. Let $g(x_1, ..., x_n)$ be a function, satisfying the bounded difference condition with constants $c_1, c_2, ..., c_n > 0$. Show that

$$\mathrm{Var}(g(X_1, ..., X_n)) \leq \sum_{i=1}^n c_i^2 p_i(1 - p_i)$$

Proof:

Let $X = (X_1, \dots, X_n)$ and $\mu = \mathbb{E}(g(X))$. Define the random variables $X_i^{(1)}, \dots, X_i^{(n)}$ as follows:

$$X_j^{(i)} = \begin{cases} X_j & \text{if } i = j,\ X_j' \\ \text{otherwise,} & \end{cases}$$

where X_j' is an independent copy of X_j. Then, we have

$$
\begin{aligned}
\mathrm{Var}(g(X)) &= \mathrm{Var}(g(X_1, \ldots, X_n)) \\
&= \mathrm{Var}(g(X_1^{(1)}, \ldots, X_n^{(1)})) \\
&= \mathrm{Var}(\mathbb{E}(g(X_1^{(1)}, \ldots, X_n^{(1)}) \mid X_2, \ldots, X_n)) \\
&\leq \sum_{i=1}^{n} \mathrm{Var}(\mathbb{E}(g(X_1^{(1)}, \ldots, X_n^{(1)}) \mid X_i, X_{i+1}, \ldots, X_n)) \\
&= \sum_{i=1}^{n} \mathrm{Var}(g(X_1^{(1)}, \ldots, X_i, X_{i+1}^{(i+1)}, \ldots, X_n^{(n)})) \\
&= \sum_{i=1}^{n} \mathrm{Var}(g(X_1^{(1)}, \ldots, X_i', X_{i+1}^{(i+1)}, \ldots, X_n^{(n)})) \\
&= \sum_{i=1}^{n} \mathrm{Var}(g(X_1^{(1)}, \ldots, X_i', X_{i+1}^{(i+1)}, \ldots, X_n^{(n)}) - \mu) \\
&\leq \sum_{i=1}^{n} c_i^2 \mathbb{E}((X_i' - X_i)^2) \\
&= \sum_{i=1}^{n} c_i^2 p_i (1 - p_i),
\end{aligned}
$$

where we used the bounded difference condition to get the second-to-last line. Therefore, we have shown that $\mathrm{Var}(g(X)) \leq \sum_{i=1}^{n} c_i^2 p_i (1 - p_i)$, as desired.

5.

Let $X_1, ..., X_n$ be independent r.v. in $[0, 1]$ and let $g(X_1, ..., X_n)$ be the minimal number of bins in which one could pack the numbers $X_1, ..., X_n$ so that the sum of the numbers in each bin does not exceed one. Prove that

$$
\mathrm{Var}(g(X_1, ..., X_n)) \leq \frac{n}{4}
$$

and provide an example showing that this upper bound could not be improved.

Proof: To apply Efron-Stein inequality, let $X_1, \ldots, X_n$ be our original independent random variables, and define $Z_1, \ldots, Z_n$ to be independent copies of $X_1, \ldots, X_n$, respectively. For each $i \in 1, \ldots, n$, let X_i' be a random variable that is equal to X_i with probability $\frac{1}{2}$ and Z_i with probability $\frac{1}{2}$. Define g' to be the corresponding function of $X_1', \ldots, X_n'$.

Note that g' depends only on the 2^n possible values of $(X_1', \ldots, X_n')$, which can be represented as $(\epsilon_1 X_1, \ldots, \epsilon_n X_n)$, where $\epsilon_1, \ldots, \epsilon_n \in -1, 1$ are independent and uniformly distributed. Let $A_1, \ldots, A_n$ be subsets of $-1, 1$ such that $|A_i| = \frac{1}{2} 2^n$, and define g_i' to be the function of $(\epsilon_1 X_1, \ldots, \epsilon_n X_n)$ that depends only on $\epsilon_j X_j$ for $j \neq i$ and $\epsilon_i X_i$ with $\epsilon_i \in A_i$.

By the Efron-Stein inequality, we have

$$
\begin{aligned}
\mathrm{Var}(g') &\leq \frac{1}{2} \sum_{i=1}^{n} \mathrm{Var}(g_i') \\
&= \frac{1}{2} \sum_{i=1}^{n} \mathbb{E}\left[(g_i')^2\right] - \left(\frac{1}{2} \sum_{i=1}^{n} \mathbb{E}\left[g_i'\right] \right)^2 .
\end{aligned}
$$

Note that for each $i \in 1, \ldots, n$ and $\epsilon_i \in A_i$, the function g'_i is just the function g applied to the $n-1$ variables $\epsilon_j X_j$ for $j \neq i$. Thus, we have

$$\begin{aligned}
\mathbb{E}\left[(g'_i)^2\right] &= \mathbb{E}\left[g(\epsilon_1 X_1, \ldots, \epsilon_{i-1} X_{i-1}, Z_i, \epsilon_{i+1} X_{i+1}, \ldots, \epsilon_n X_n)^2\right] \\
&= \mathbb{E}\left[g^2(X_1, \ldots, X_n)\right] \\
&\leq \left(\mathbb{E}\left[g(X_1, \ldots, X_n)\right]\right)^2 + \mathrm{Var}\left(g(X_1, \ldots, X_n)\right) \\
&\leq \left(\frac{n}{2}\right)^2 + \frac{n}{4} \\
&= \frac{5n^2}{16}.
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
&\mathbb{E}\left[g'_i\right] \\
&= \mathbb{E}\left[g(\epsilon_1 X_1, \ldots, \epsilon_{i-1} X_{i-1}, Z_i, \epsilon_{i+1} X_{i+1}, \ldots, \epsilon_n X_n)\right] \\
&= \frac{1}{2}\mathbb{E}\left[g(X_1, \ldots, X_n)\right] + \frac{1}{2}\mathbb{E}\left[g(Z_1, \ldots, Z_n)\right] \\
&\leq \frac{n}{2}.
\end{aligned}$$

Substituting these expressions into the Efron-Stein inequality gives

$$\begin{aligned}
\mathrm{Var}(g') &\leq \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}\left[(g'_i)^2\right] - \left(\frac{1}{2}\sum_{i=1}^{n} \mathbb{E}\left[g'_i\right]\right)^2 \\
&\leq \frac{1}{2} n \cdot \frac{5n^2}{16} - \left(\frac{1}{2} n \cdot \frac{n}{2}\right)^2 \\
&= \frac{n}{4}.
\end{aligned}$$

Therefore, we have shown that $\mathrm{Var}(g(X_1, \ldots, X_n)) \leq \mathrm{Var}(g'(X'_1, \ldots, X'_n)) \leq \frac{n}{4}$, as required.