Learning Theory:
Concentration Inequalities

Daniel Khashabi

Fall 2013
Last Update: October 19, 2016

1 Introduction

To measure up how the modeling using samples is close to the original (unknown) system, researchers have developed various. One of the ways to quantify this is to use *probabilistic* approach for modeling the *concentration* (closeness) of the model to the original system. Such analysis need using various probabilistic inequalities that could bound the results.

2 Markov's inequality; the most basic bound

Theorem 2.1. *If X is a non-negative random variable, for any $t > 0$ we have*

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}X}{t}$$

Proof.

$$\begin{aligned}
\mathbb{E}X &= \int_{X \in \mathcal{X}} P_X(dx) \\
&= \int_{X \in \mathcal{X}, X < t} X P_X(dx) + \int_{X \in \mathcal{X}, X \geq t} X P_X(dx) \\
&\geq \int_{X \in \mathcal{X}, X \geq t} X P_X(dx) \\
&\geq \int_{X \in \mathcal{X}, X \geq t} t P_X(dx) = t \int_{X \in \mathcal{X}, X \geq t} P_X(dx) = t\mathbb{P}(X \geq t)
\end{aligned}$$

■

3 Chebyshev's inequality

Theorem 3.1.

$$\mathbb{P}\left(|X - \mathbb{E}X| \geq t\right) \leq \frac{VarX}{t^2}$$

Proof. There is a simple proof using Markov's inequality. Based Markov's inequality we know,

$$\mathbb{P}(X \geq t) = \mathbb{P}(\phi(X) \geq \phi(t)) \leq \frac{\mathbb{E}\phi(X)}{\phi(t)}$$

Now assume $\phi(x) = x^2 (x \geq 0)$, which gives the desired result. In other words, using the Markov's inequality

$$\begin{aligned}\mathbb{P}\left(|X - \mathbb{E}X| \geq t\right) &= \mathbb{P}\left(|X - \mathbb{E}X|^2 \geq t^2\right) \\ &\leq \frac{\mathbb{E}\left|X - \mathbb{E}X\right|^2}{t^2} = \frac{\mathbb{E}X^2 - (\mathbb{E}X)^2}{t^2} = \frac{\mathrm{Var}X}{t^2}\end{aligned}$$

∎

Remark 3.2. In general the right side can be any moment (not just the second moment), i.e. $\phi(x) = x^q (x \geq 0)$, for any positive q, using which we have,

$$\mathbb{P}\left(|X - \mathbb{E}X| \geq t\right) \leq \frac{\mathbb{E}\left|X - \mathbb{E}X\right|^q}{t^q}$$

Since the goal is to find tighter upper-bounds one can minimize the right-side with respect to q.

4 Chernoff's Trick

For any $s > 0$,

$$\mathbb{P}(X > t) = \mathbb{P}(e^{sX} > e^{st}) \leq e^{-st}\mathbb{E}\left[e^{sX}\right]$$

To make the bound as tight as possible,

$$\mathbb{P}(X > t) \leq \inf_{s>0} e^{-st}\mathbb{E}\left[e^{sX}\right]$$

5 Hoeffding's inequality; a special case

Lemma 5.1. *If $X \in \mathcal{X}$ is a random variable with $\mathbb{E}X = 0$, and $\exists a, b \in \mathcal{X}$, s.t. $\mathbb{P}(a \leq X \leq b) = 1$, then for any $s > 0$*

$$\mathbb{E}\left[e^{sX}\right] \leq e^{\frac{1}{8}s^2(b-a)^2}$$

Proof. Assume any point $x \in [a, b]$, which can be represented as $x = \beta.b + (1-\beta).a, 0 \leq \beta \leq 1$. Let us assume a function $\phi(x) = e^{sx}$ for any $s > 0$ which is a convex function on $\mathcal{X} = \mathbb{R}$. Then we have,

$$e^{sx} \leq \beta e^{sb} + (1-\beta)e^{sa}$$

By replacing β with $\frac{x-a}{b-a}$, and since $\mathbb{E}X = 0$, we have,

$$\mathbb{E}\left[e^{sx}\right] \leq \frac{b}{b-a}e^{sa} - \frac{a}{b-a}e^{sb}$$

Now set $p = \frac{a}{a-b}$, and thus $1 - p = \frac{b}{b-a}$. We continue the previous expression:

$$\begin{aligned}\mathbb{E}\left[e^{sx}\right] &\leq \frac{b}{b-a}e^{sa} - \frac{a}{b-a}e^{sb} \\ &\leq (1-p)e^{sa} + pe^{sb} = e^{sa}\left(1 - p + pe^{s(b-a)}\right)\end{aligned}$$

Define $u \triangleq s(b-a)$ and define $\Phi(u) \triangleq -pu + \log(1 - p + pe^u)$. Then we have

$$\mathbb{E}e^{sX} \leq e^{\Phi(u)}$$

which holds for any value of u, for $p \in [0, 1]$. One show that the $\Phi(u)$ is upper-bounded by $\frac{1}{8}u^2 = \frac{1}{8}s^2(b-a)^2$, which proves the desired result. ■

Theorem 5.2 (Hoeffding's inequality)**.** *Let $X_1, \ldots, X_n$ be independent random variables, and for any X_i, $\exists a_i, b_i \in \mathcal{X} = \mathbb{R}$, s.t. $\mathbb{P}(a_i \leq X \leq b_i) = 1$. Let $S_n \triangleq \frac{1}{n}\sum_{i=1}^n X_i$ then for any $t > 0$,*

$$\mathbb{P}\left(|S_n - \mathbb{E}S_n| \geq t\right) \leq 2\exp\left(-\frac{2n^2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right) \tag{1}$$

In the special case when $[a_i, b_i] = [0, 1]$ we will have:

$$\mathbb{P}\left(|S_n - \mathbb{E}S_n| \geq t\right) \leq 2\exp\left(-2nt^2\right) \tag{2}$$

Proof. We can divide the inequality into two parts,

$$\mathbb{P}\left(S_n - \mathbb{E}S_n \geq t\right) \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right) \tag{3}$$

$$\mathbb{P}\left(S_n - \mathbb{E}S_n \leq -t\right) \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right) \tag{4}$$

To use the Lemma 5.1, we can replace each variable $X_i \to X_i - \mathbb{E}X_i$, so that $\mathbb{E}X_i = 0$; then we have, $\mathbb{E}\left[e^{sX_i}\right] \leq e^{s^2(b_i - a_i)^2/8}$. Using Chernoff's trick,

$$\mathbb{P}\left(S_n \geq t\right) = \mathbb{P}\left(e^{S_n} \geq e^t\right) \leq e^{-st}\mathbb{E}\left[e^{sS_n}\right]$$

And since X_i's are independent,

$$\mathbb{E}\left[e^{sS_n}\right] = \mathbb{E}\left[e^{s(X_1 + \ldots + X_n)}\right] = \mathbb{E}\left[\prod_{i=1}^n e^{sX_i}\right] = \prod_{i=1}^n \mathbb{E}\left[e^{sX_i}\right]$$

$$\Rightarrow \mathbb{P}\left(S_n \geq t\right) \leq e^{-st}\prod_{i=1}^n e^{s^2(b_i - a_i)^2/8} = \exp\left\{-st + \frac{s^2}{8}\sum_{i=1}^n (b_i - a_i)^2\right\}$$

To minimize the right hand-side with respect to s we can choose it to be $s = \frac{4t}{\sum_{i=1}^n (b_i - a_i)^2}$. This proves the Equation 3. The proof for the Equation 4 is similar. ■

Remark 5.3. When dealing with one-sided inequalities, the upper-bound probability will get half:

$$\mathbb{P}\left(S_n - \mathbb{E}S_n \geq t\right) = \mathbb{P}\left(S_n - \mathbb{E}S_n \leq -t\right) \leq \exp\left(-\frac{2n^2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right) \tag{5}$$

Another way of saying the previous inequality is the following corollary.

Corollary 5.4. *With probability at least $1-\delta$, the following inequality holds:*

$$|S_n - \mathbb{E}S_n| \leq t$$

for $t = \sqrt{\frac{\ln(2/\delta)}{2n^2}\sum_{i=1}^{n}(b_i - a_i)^2}$. *In the special case* $[a_i, b_i] = [0, 1]$, $t = \sqrt{\frac{\ln(2/\delta)}{2n}}$.

6 McDiarmid's inequality: bounded differences

This is also known as *Hoeffding-Azuma*, *Bounded Differences* or *Martingale concentration* inequality.

Definition 6.1 (A function with bounded differences)**.** *A function is said to have bounded differences, if changing only one variable, and keep everything the save, the absolute value of the difference is always bounded. In other words, if we have a function* $f : \mathcal{X}^n \to \mathbb{R}$, *then,*

$$\begin{aligned} &\exists c_i \in \mathbb{R},\ s.t.\ \forall x_1, \ldots, x_n, x_i' \in \mathbb{X}, \\ &\qquad \sup \left| f(x_1, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n) \right| \leq c_i \end{aligned}$$

Theorem 6.1 (McDiarmid's inequality)**.** *Let* $X = (X_1, \ldots, X_n) \in \mathcal{X}^n$, *and* $f : \mathcal{X}^n \to \mathbb{R}$ *has bounded differences. Then for any* $t > 0$,

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

TBW. ■

Example 6.2.

1. Given a real-valued random variable Z, such that,

   $$\log \mathbb{E}\left[e^{sZ}\right] \leq \frac{vs^2}{2(1-cs)}, \tag{6}$$

   for some $v, c \in \mathbb{R}^+$ and every $s \in [0, \frac{1}{c}]$. Prove that

   $$\mathbb{P}\left(Z \geq \sqrt{2vt} + ct\right) \leq e^{-t} \tag{7}$$

   for all $t > 0$.

   Hint: We know $\{X_i\}_{i=1}^n$ are i.i.d. and $X_i \sim \mathcal{N}(0,1)$. Then $U = \sum_{i=1}^n X_i^2$ with $U \sim \chi_n^2$ (Chi-Squared distribution with degree-of-freedom of n.).

2. Given the definitions in the previous part, we want to prove that,

   $$\mathbb{P}\left(U - n \geq 2\sqrt{nt} + 2t\right) \leq e^{-t}.$$

   Hint: First prove Lemma 6.3 !

Lemma 6.3. *The following holds for any* $s \in (0, 1/2)$

$$-s - \frac{1}{2}\log(1-2s) \leq \frac{s^2}{1-2s}, \quad 0 < s < 1/2 \tag{8}$$

Example 6.4. We consider we have a set of i.i.d. random variables $\{x_i\}_{i=1}^n$, and $x_i \sim \text{Binomial}(\theta)$. Clearly if $S \triangleq \sum_{i=1}^n x_i$,

$$S \sim \mathrm{B}(n, \theta) \Rightarrow \begin{cases} \mathbb{E}[S] = n\theta, \\ \mathbb{V}[S] = n\theta(1-\theta) \end{cases}$$

Now prove each part.

1. First prove the following holds

$$\mathbb{P}(S \geq n\alpha) \leq e^{-n \times d(\alpha || \theta)}, \ \forall \alpha \in [\theta, 1] \tag{9}$$

   where, $d(\alpha||\theta) = \alpha \log \frac{\alpha}{\theta} + (1-\alpha) \log \frac{(1-\alpha)}{(1-\theta)}$ is the KL-divergence between α and θ, two Bernoulli random variables.

2. Now prove that the bound in Equation 9 is indeed tighter than,

$$\mathbb{P}(S \geq n\alpha) \leq e^{-2n(\alpha-\theta)^2}, \ \forall \alpha \in [\theta, 1]$$

Example 6.5. We assume having i.i.d. random variables, $\{Y_i\}_{i=1}^k$. For a given real number $y \in \mathbb{R}$,

$$N = \left| \{1 \leq j \leq k : Y_j \geq y\} \right|. \tag{10}$$

1. Considering, $\max_{1 \leq j \leq k} \mathbb{P}(Y_j \geq y) \leq p$, we want to show that there is exists a pair $(\tilde{N}, \tilde{S})$ of jointly distributed integer-valued random variables such that
   - $\tilde{N}$ has the same distribution as N.
   - $\tilde{S}$ has the Binomial(k, p).
   - $\tilde{N} \leq \tilde{S}$.

2. Assuming that $T_{k,p}(t)$ is the probability that Binomial(k,p) is a random variable greater or equal to t, we want to show that

$$\mathbb{P}(N \geq t) \leq T_{k,p}(t) \tag{11}$$

7 Bennet's and Bernstein's Inequalities

Bennet's inequality offers some improvement over Hoeffding's when the variances of the summands are small compared to their almost-sure bounds on the absolute values:

Lemma 7.1 (Bennet's inequality)**.** *Let $Z_1, \ldots, Z_n$ be independent random variables with zero mean, and assume that $Z_i \leq 1$ with probability 1. Let*

$$\sigma^2 \geq \frac{1}{n} \sum_{i=1}^n \mathbb{E}\left[Z_i^2\right]$$

Then for any $\epsilon > 0$,

$$P\left(\sum_{i=1}^n Z_i > \epsilon\right) \leq e^{-m\sigma^2 h\left(\frac{\epsilon}{m\sigma^2}\right)}$$

where

$$h(a) = (1+a)\log(1+a) - a$$

Hoeffding's inequality only assumes the summands are bounded almost surely, while Bennett's inequality offers some improvement when the variances of the summands are small compared to their almost sure bounds. In both inequalities, unlike some other inequalities or limit theorems, there is no requirement that the component variables have identical or similar distributions (only requirement is the independence).

One can write an equivalent martingale version of this inequality:

$$\mathrm{P}\left(S_n - E_n > \epsilon\right) \leq \exp\left(-V_n h(\frac{t}{V_n})\right)$$

where $S_n = \sum_{i=1}^{n} X_i$, $E_n = \sum_{i=1}^{n} \mathbb{E}\left[X_i\right]$ and $V_n = \sum_{i=1}^{n} \mathbb{V}\left[X_i\right]$.

Using the inequality that $h(a) \leq a^2/(2 + 2a/3)$ it is possible to prove the following:

Lemma 7.2 ((Bernstein's inequality)**.** *Let $Z_1, \ldots, Z_n$ be i.i.d. random variables with zero mean, and assume that $|Z_i| \leq M$ with probability 1. For any $t > 0$,*

$$P\left(\sum_{i=1}^{n} Z_i > t\right) \leq \exp\left(-\frac{t^2/2}{\sum \mathbb{E}\left[Z_i^2\right] + Mt/3}\right)$$

This is a generalization of Hoeffding's since it can handle not only independent variables but also weakly-dependent variables.

The martingale form of this inequality is as following:

$$\mathrm{P}\left(S_n - E_n > \epsilon\right) \leq 2\exp\left(-\frac{t^2}{V_n + t/3}\right)$$

8 Bibliographical notes

?

9 Basic inequalities

Lemma 9.1 (The exponential inequality)**.** *For all* $x \in \mathbb{R}$,

$$1 + x \leq e^x.$$

Proof. Both sides are equal at $x = 0$. The derivative of the left side is a constant 1, and the derivative on the right is e^x. For all $x \geq 0$, $e^x \geq 1$, so the right side grows faster than the left. For $x \leq 0$, $e^x \leq 1$, so as x goes down from 0 to $-\infty$, the left side decreases faster than the right. ■

Lemma 9.2 (the logarithmic inequality)**.** *For all* $x > -1$,

$$\ln(1 + x) \leq x.$$

Lemma 9.3 (Bernoulli inequalities)**.**

- *If* $x, y \geq 0$, *then*

  $$(1 - x)(1 - y) \geq 1 - x - y.$$

  If in addition $y \leq 1$, *then*

  $$(1 + x)(1 + y) \geq 1 + x + y.$$

- *If* $x_1, x_2, \ldots, x_k \geq 0$, *then*

  $$(1 + x_1)(1 + x_2) \cdots (1 + x_k) \geq 1 + x_1 + \cdots + x_k.$$

  If in addition $x_2, \ldots, x_k \leq 1$, *then*

  $$(1 - x_1)(1 - x_2) \cdots (1 - x_k) \geq 1 - x_1 - x_2 - \cdots - x_k.$$

Proof.

(a) Since $xy \geq 0$, we have

$$(1 + x)(1 + y) = 1 + x + y + xy \geq 1 + x + y$$

and

$$(1 - x)(1 - y) = 1 - x - y + xy \geq 1 - x - y.$$

(b) The base case for $k = 2$ is given by (a). For $k > 2$, since $x_1 + \cdots + x_{k-1} > 0$, we have by induction and base k,

$$\begin{aligned}(1 + x_1) \cdots (1 + x_k) &\geq (1 + x_1 + \cdots + x_{k-1})(1 + x_k) \\ &\geq 1 + x_1 + \cdots + x_k\end{aligned}$$

and

$$\begin{aligned}(1 - x_1) \cdots (1 - x_k) &\geq (1 - x_1 - \cdots - x_{k-1})(1 - x_k) \\ &\geq 1 - x_1 - \cdots - x_k.\end{aligned}$$

■

Lemma 9.4. *For* $0 \le x \le 1/2$,

$$\ln(1+x) \ge x + x^2$$

and

$$\ln(1-x) \ge -x - x^2$$

Lemma 9.5. *For* $x \in [-1/2, 1/2]$,

$$\ln(1-x) \ge x + x^2.$$

Proof. For $x \le 0$,

$$\ln(1-x) = \ln(1+(-x)) \ge -x + x^2 \ge -x - x^2.$$

For $x \in\ge 0$,

$$\ln(1-x) \ge -x - x^2$$

$\blacksquare$

Proof. For $x \le 0$,

$$\ln(1+x) = \ln(1-(-x)) \ge -(-x) - (-x)^2 = x - x^2 \ge x/2.$$

$\blacksquare$

10 Proofs

10.1 Proof of lemma 6.3

Proof. Let's label each side of the inequality,

$$\begin{cases} A(s) = -s - \frac{1}{2}\log(1-2s) \\ B(s) = \frac{s^2}{1-2s} \end{cases}$$

We prove this inequality in two steps,

- $A(s) = B(s),$ for $s = 0$:
  This is easy to show that $A(0) = B(0) = 0$
- $\frac{dA(s)}{ds} \leq \frac{dB(s)}{ds}, \forall s \in (0, 1/2)$:

$$\frac{dA(s)}{ds} = -1 - \frac{1}{2}\frac{-2}{1-2s} = \frac{2s}{1-2s}$$
$$\frac{dB(s)}{ds} = \frac{2s(1-s)}{(1-2s)^2}$$
$$\frac{dB(s)}{ds} - \frac{dA(s)}{ds} = \frac{2s(1-s)}{(1-2s)^2} - \frac{2s}{1-2s} = \frac{2s^2}{(1-2s)^2} \geq 0$$

This gives us the desired result. ■

11 Answers to examples

Here answers to some of the questions are included. The answers are mostly by the authors, and might be buggy. Therefore, read cautiously!

11.1 Answer to example 6.2

11.1.1 First part :

Proof. Based on the Markov inequality we know that we have,

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}.$$

We apply the Chernoff trick to find better (smaller=tighter) upper bound for our desired upper-bound:

$$\mathbb{P}(X \geq a) = \mathbb{P}\left(e^{sX} \geq e^{sa}\right) \leq \frac{\mathbb{E}\left[e^{sX}\right]}{e^{sa}} = e^{-sa}\mathbb{E}\left[e^{sX}\right], \quad \forall s > 0. \tag{12}$$

$$\mathbb{P}(X \geq a) \leq \inf_{s\in\mathbb{R}^+} e^{-sa}\mathbb{E}\left[e^{sX}\right]. \tag{13}$$

From the Equation 7, and combining it with the Equation 13 we have,

$$\begin{aligned}\mathbb{P}(X \geq a) \leq \inf_{s\in\mathbb{R}^+} e^{-sa}\mathbb{E}\left[e^{sX}\right] &\leq \inf_{s\in\mathbb{R}^+}\left\{e^{-sa}\exp\left[\frac{vs^2}{2(1-cs)}\right]\right\} \\ &= \inf_{s\in\mathbb{R}^+}\left\{\exp\left[-sa + \frac{vs^2}{2(1-cs)}\right]\right\}.\end{aligned}$$

The minimizer of the right term could be found:

$$\frac{d}{ds}\left[-sa + \frac{vs^2}{2(1-cs)}\right] = -a + \frac{2vs(1-cs) + cvs^2}{2(1-cs)^2} = 0$$

$$\Rightarrow \mathbb{P}(X \geq a) \leq e^{-sa}\mathbb{E}\left[e^{sX}\right] \leq \left\{e^{-sa}\exp\left[\frac{vs^2}{2(1-cs)}\right]\right\}.$$

Since the minimization could be cumbersome, we do reverse engineering using the modified objective,

$$\begin{aligned}\mathbb{P}\left(Z \geq \sqrt{2vt} + ct\right) &= \mathbb{P}\left(e^{sZ} \geq e^{s\left(\sqrt{2vt}+ct\right)}\right) \\ &\leq \inf_{s\in\mathbb{R}^+} \frac{\mathbb{E}e^{sZ}}{e^{s\left(\sqrt{2vt}+ct\right)}} \\ &\leq \inf_{s\in\mathbb{R}^+} \exp\left\{\frac{vs^2}{2(1-cs)} - s\left(\sqrt{2vt}+ct\right)\right\}\end{aligned}$$

Now we just need to show that,

$$\inf_{s\in\mathbb{R}^+} \exp\left\{\frac{vs^2}{2(1-cs)} - s\left(\sqrt{2vt}+ct\right)\right\} \leq e^{-t}$$

We can make it looser and show that,

$$\exists s \in \mathbb{R}^+ \quad s.t. \quad \exp\left\{\frac{vs^2}{2(1-cs)} - s\left(\sqrt{2vt}+ct\right)\right\} = e^{-t}$$
$$\Rightarrow \exists s \in \mathbb{R}^+ \quad s.t. \quad \exp\left\{\frac{vs^2}{2(1-cs)} - s\left(\sqrt{2vt}+ct\right)+1\right\} = 1$$
$$\Rightarrow \exists s \in \mathbb{R}^+ \quad s.t. \quad \frac{vs^2}{2(1-cs)} - s\left(\sqrt{2vt}+ct\right)+1 = 0$$

The expression above could be turned into a nice closed form:

$$\begin{aligned}\frac{vs^2}{2(1-cs)} - s\left(\sqrt{2vt}+ct\right)+1 &= \frac{vs^2 - 2s\left(\sqrt{2vt}+ct\right)(1-cs)+2(1-cs)}{2(1-cs)}\\ &= \frac{vs^2 - 2s\sqrt{2vt}(1-cs)+2(1-cs)^2}{2(1-cs)}\\ &= \frac{\left(s\sqrt{v}-(1-cs)\sqrt{2t}\right)^2}{2(1-cs)} = 0\\ \Rightarrow s &= \frac{\sqrt{2t}}{c\sqrt{2t}-\sqrt{v}}\end{aligned}$$

This holds only when $s \in [0, \frac{1}{c}]$, since we used Equation 6. Since $v, t \in \mathbb{R}^+$, for any $t \geq \frac{v}{2c^2}$, s belongs to $[0, \frac{1}{c}]$ and this bound holds. ■

11.1.2 Second part :

Proof. We first prove that the inequality in Equation 6 holds, when we choose $Z = U - n = \sum_{i=1}^n X_i^2 - n$,

$$\begin{aligned}\mathbb{E}\left[e^{sZ}\right] &= \int_{\mathbf{x}\in\mathcal{X}} e^{s\left(\sum_{i=1}^n x_i^2 - n\right)}(2\pi)^{-k/2}\left|\mathbf{\Sigma}\right|^{-1/2} e^{-\frac{1}{2}(\mathbf{x}-\mu)^\top \mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)} d\mathbf{x}\\ &= e^{-sn}\prod_{i=1}^n \int_{x_j} e^{sx_i^2}\frac{1}{\sqrt{2\pi}}e^{-\frac{x_i^2}{2}}dx_i\\ &= e^{-sn}\prod_{i=1}^n \int_{x_j} \frac{1}{\sqrt{2\pi}}e^{-x_i^2\left(\frac{1}{2}-s\right)}dx_i\\ &= e^{-sn}\prod_{i=1}^n \sqrt{\frac{1}{1-2s}}\\ &= e^{-sn}(1-2s)^{-n/2}\end{aligned}$$

Taking logarithm of both sides, and using the result of Lemma 6.3,

$$\log \mathbb{E}\left[e^{sZ}\right] = n\left(-s - \frac{1}{2}(1-2s)\right) \leq n \times \frac{s^2}{1-2s}, \quad 0 < s < 1/2$$

Which is of the form given in the equation 6, when $v = n$ and $c = 2$. Given this, we can use the bound we found in part (a), with replacing the values $v = n$ and $c = 2$, which gives us the desired result. ■

11.2 Answer to example 6.4

11.2.1 First part :

Proof. Based on the Markov inequality we know that,

$$\mathbb{P}(S \geq n\alpha) \leq \inf_{\lambda \in \mathbb{R}^+} \left[e^{-\lambda n\alpha} \mathbb{E}_{S\sim \mathrm{B}(n,\theta)} e^{S\lambda} \right] \tag{14}$$

It is easy to show that the Moment Generating Function (MGF) for a binomial distribution is as following,

$$M_X(t) = \mathbb{E}_{X\sim \mathrm{Binomial}(n,\theta)} e^{tX} = (1-\theta+\theta e^t)^n$$

Using MGF formula, we can simplify our bound in Equation 14.

$$\begin{aligned}\mathbb{P}(S \geq n\alpha) \leq \inf_{\lambda \in \mathbb{R}^+} \left[e^{-\lambda n\alpha} (1-\theta+\theta e^{\lambda})^n \right] &= \inf_{\lambda \in \mathbb{R}^+} \left[\left((1-\theta)e^{-\lambda\alpha} + \theta e^{\lambda(1-\alpha)} \right)^n \right] \\ &= \left(\inf_{\lambda \in \mathbb{R}^+} \left[(1-\theta)e^{-\lambda\alpha} + \theta e^{\lambda(1-\alpha)} \right] \right)^n\end{aligned}$$

We define $A(\lambda) = (1-\theta)e^{-\lambda\alpha} + \theta e^{\lambda(1-\alpha)}$, and find minimizer of $A(\lambda)$,

$$\frac{\partial A}{\partial \lambda} = -\alpha(1-\theta)e^{-\lambda\alpha} + (1-\alpha)\theta e^{\lambda(1-\alpha)} = 0 \Rightarrow \lambda = \ln \frac{\alpha(1-\theta)}{(1-\alpha)\theta}$$

$$\Rightarrow A(\lambda) = (1-\theta)e^{-\lambda\alpha} + \theta e^{\lambda(1-\alpha)} = (1-\theta)e^{-\lambda\alpha} \left(1 + \frac{\theta}{1-\theta} e^{\lambda} \right)$$

$$\begin{aligned}\Rightarrow A(\lambda)\big|_{\lambda = \ln \frac{\alpha(1-\theta)}{(1-\alpha)\theta}} &= (1-\theta)e^{-\lambda\alpha} \left(1 + \frac{\theta}{1-\theta} e^{\lambda} \right) \Bigg|_{\lambda = \ln \frac{\alpha(1-\theta)}{(1-\alpha)\theta}} \\ &= \left(\frac{\theta}{\alpha} \right)^{\alpha} \left(\frac{1-\theta}{1-\alpha} \right)^{1-\alpha} \\ &= -d(\alpha||\theta)\end{aligned}$$

$$\Rightarrow \mathbb{P}(S \geq n\alpha) \leq e^{-nd(\alpha||\theta)}$$

■

11.2.2 Second part :

Proof. We want to show,

$$e^{-n\times d(\alpha||\theta)} \leq e^{-2n(\alpha-\theta)^2},\ \forall \alpha \in [\theta, 1]$$

Or we want to show,

$$d(\alpha||\theta) \geq 2(\alpha-\theta)^2,\ \forall \alpha \in [\theta, 1]$$

First note that for $\alpha = \theta$ both bounds are the same, as they are both zero. Then if we show that the derivative of $d(\alpha||\theta)$ is always greater than $2(\alpha-\theta)^2$, this would imply that $d(\alpha||\theta) \geq 2(\alpha-\theta)^2$, for all $\alpha \in [\theta, 1]$. Equivalently, we show that $\frac{\partial \lambda}{\partial \alpha} \geq 0$ for $\theta \geq \alpha \geq 1$, where

$$\lambda = d(\alpha||\theta) - 2(\alpha-\theta)^2$$

$$\frac{\partial \lambda}{\partial \alpha} = \ln \frac{\alpha}{\theta} + 1 + \ln \frac{1-\alpha}{1-\theta} - 1$$

To show that $\frac{\partial \lambda}{\partial \alpha} \geq 0$, we do the same trick; since $\frac{\partial \lambda}{\partial \alpha}\big|_{\alpha=\theta} = 0$, we just need to show that, $\frac{\partial^2 \lambda}{\partial \alpha^2} \geq 0$. Since $\frac{\partial^2 \lambda}{\partial \alpha^2} = \frac{1}{\alpha(1-\alpha)} - 4$, and using the Arithmetic-Geometric inequality*, we have

$$\alpha(1-\alpha) \geq 4$$

Then $\frac{\partial^2 \lambda}{\partial \alpha^2} \geq 0$ which finishes our proof. ■

11.3 Answer to example 6.5

11.3.1 First part :

Proof. In practice the original random variables $\{Y_i\}_{i=1}^k$, could be distributed with any arbitrary distribution. But we can easily replace them with any arbitrary distribution, and come up with the same distribution for Equation 10, as long as it satisfies certain conditions. To make everything simple, we use uniform distribution, and we define the set of uniform i.i.d. random variables $\{U_i\}_{i=1}^k$, such that for a fixed y,

$$\mathbb{P}(U_j \geq y) = \mathbb{P}(Y_j \geq y), \quad \forall j \in \{1, \ldots, k\} \tag{15}$$

Now we define a modified version of the Equation 10 for the U_j random variables,

$$\tilde{N} = |\{1 \leq j \leq k : U_j \geq y\}| \,. \tag{16}$$

Referring to the conditions in the Equation 15, it is easy to see that $\tilde{N}$ and N have the same distributions, or in other words,

$$\mathbb{P}(\tilde{N} \geq t) = \mathbb{P}(N \geq t),\ \forall t \in \mathbb{N} \cup \{0\}$$

In a similar way, we define the set of i.i.d. random variables $\{V_i\}_{i=1}^k$, with the condition that

$$\mathbb{P}(V_j \geq y) = p, \quad \forall j \in \{1, \ldots, k\} \tag{17}$$

Similar to $\tilde{N}$ we define a modified version of the Equation 10 for the V_j random variables,

$$\tilde{S} = |\{1 \leq j \leq k : V_j \geq y\}| \,. \tag{18}$$

Based on the condition in Equation 17, since

$$\mathbb{P}(Y_j \geq y) = \mathbb{P}(U_j \geq y) \leq p = \mathbb{P}(V_j \geq y)$$

it is clear that,

$$\mathbb{P}(\tilde{S} \geq t) \geq \mathbb{P}(\tilde{N} \geq t),\ \forall t \in \mathbb{N} \cup \{0\} \tag{19}$$

Also it is easy to see that $\tilde{S} \sim \text{Binomial}(k, p)$. To show this, we define

$$\tilde{S}_j = \mathbf{1}\{V_j \geq y\} \,.$$

Each $\tilde{S}_j$ is distributed with the Bernoulli, with $\mathbb{P}(\tilde{S}_j = 1) = \mathbb{P}(V_j \geq y) = p$. Clearly $\tilde{S} = \sum_{j=1}^k \tilde{S}_j$, and since each $\tilde{S}_j \sim \text{Bernoulli}(p)$,

$$S \sim \text{Binomial}(k, p) \tag{20}$$

■

* $\frac{x+y}{2} \geq \sqrt{xy} \Rightarrow \frac{\alpha+(1-\alpha)}{2} \geq \sqrt{(1-\alpha)\alpha} \Rightarrow \alpha(1-\alpha) \geq 4$

11.3.2 Second part :

Proof. Referring back to Equation 19, we have,

$$\mathbb{P}(\tilde{S} \geq t) \geq \mathbb{P}(\tilde{N} \geq t), \ \forall t \in \mathbb{N} \cup \{0\}$$

And based on the result in Equation 20, it gives us the desired answer. ■