Advanced Topics in Artificial Intelligence: EE6180
Indian Institute of Technology Madras — October 6, 2020
Instructor: Abhishek Sinha

Solution to PSET 1

- **Do not** distribute the solutions outside the class.

---

1. **(Boosting Randomized Algorithms)** Let $Z_i = 1$ if the i^{th} run of the algorithm yields correct result and is -1 otherwise. The r.v.s are i.i.d. and $\mathbb{E}(Z_i) = 2\delta$. The probability of a wrong final decision may be bounded as follows:

$$\mathbb{P}(\frac{1}{N}\sum_{i=1}^{N} Z_i - 2\delta < -2\delta) \leq \exp(-2N\delta^2) < \epsilon.$$

if $N \geq \frac{1}{2}\delta^{-2}\ln(\epsilon^{-1})$.

2. **(Concentration of the L_2 norm of a random vector)** From the theorem we proved in the class, we have

$$\mathbb{P}(|||X||_2 - \sqrt{n}| \geq t) \leq 2\exp(-\frac{ct^2}{K^4}).$$

(1) We can write

$$\begin{aligned}\mathbb{E}\big(|||X||_2 - \sqrt{n}|\big) &= \int_0^\infty \mathbb{P}\big(|||X||_2 - \sqrt{n}| \geq t\big)dt \\ &\leq 2\int_0^\infty \exp(-\frac{ct^2}{K^4})dt \\ &\leq CK^2\end{aligned}$$

for some absolute constant C.
(2) From the variational characterization of the variance, we have:

$$\begin{aligned}\mathrm{Var}(||X||_2) &\leq \mathbb{E}\big(||X||_2 - \sqrt{n}\big)^2 \\ &= \int_0^\infty \mathbb{P}\big(|||X||_2 - \sqrt{n}| \geq \sqrt{t}\big)dt \\ &\leq 2\int_0^\infty \exp(-\frac{ct}{K^4})dt \\ &\leq CK^4.\end{aligned}$$

3. **(Concentration for Spin Glasses)**

(a) Since the *free energy* $F_d(\boldsymbol{\theta})$ can be expressed as the logarithm of summation of exponentials of linear functions, it is convex.

(b) By directly computing the derivative, we have

$$\frac{\partial F_d}{\partial \theta_{ij}} = \frac{1}{\sqrt{d}} \frac{\sum_{\boldsymbol{x}\in\{\pm 1\}^d} \exp(\frac{1}{\sqrt{d}}\sum_{i<j}\theta_{ij}x_ix_j)x_ix_j}{\sum_{\boldsymbol{x}\in\{\pm 1\}^d} \exp(\frac{1}{\sqrt{d}}\sum_{i<j}\theta_{ij}x_ix_j)}.$$

Hence, using triangle inequality, we have $|\frac{\partial F_d}{\partial \theta_{ij}}| \leq \frac{1}{\sqrt{d}}$. Since the dimension of $\boldsymbol{\theta}$ vector is $d(d-1)/2$, it follows that

$$||\nabla F_d(\boldsymbol{\theta}))||_2 \leq \sqrt{\frac{d(d-1)}{2d}} \leq \sqrt{d/2}.$$

Finally, for some ζ

$$|F_d(\theta) - F_d(\theta')| = |\nabla F_d(\zeta)\cdot(\theta-\theta')| \overset{\text{(Cauchy-Schwartz)}}{\leq} \sqrt{d/2}||\theta-\theta'||_2.$$

(c) As established in part (b), $F_d(\boldsymbol{\theta})$ is a Lipschitz function of Gaussian $\mathcal{N}(0,\sigma^2)$ random variables with Lipschitz constant $\sqrt{d/2}$ (hence, Lipschitz with constant $\sigma\sqrt{d/2}$ w.r.t. the standard Gaussian variables). Hence, using the Gaussian concentration inequality, we have

$$\mathbb{P}\big(F_d(\boldsymbol{\theta}) \geq \mathbb{E}F_d(\boldsymbol{\theta}) + \delta\big) \leq \exp\big(-\frac{\delta^2}{d\sigma^2}\big). \tag{1}$$

Using Jensen's inequality for the concave function $\log(\cdot)$, we have

$$\mathbb{E}F_d(\boldsymbol{\theta}) \leq \log\Bigg(\sum_{\boldsymbol{x}\in\{\pm1\}^d} \mathbb{E}\exp(\frac{1}{\sqrt{d}}\sum_{i<j}\Theta_{ij}x_ix_j)\Bigg). \tag{2}$$

Notice that $\frac{1}{\sqrt{d}}\sum_{i<j}\Theta_{ij}x_ix_j$ is a Gaussian random variable with mean zero and variance $= \frac{d^2-d}{2d}\sigma^2 \leq \frac{d\sigma^2}{2}$. Hence, from Equation (2), we have

$$\mathbb{E}F_d(\Theta) \leq \log\big(2^d\exp(d\sigma^2/4)\big) = d\log 2 + d\sigma^2/4.$$

Combining the above with Eqn. (1), we have

$$\mathbb{P}\big(F_d(\boldsymbol{\Theta}) \geq d\log 2 + d\sigma^2/4 + \delta\big) \leq \mathbb{P}\big(F_d(\boldsymbol{\Theta}) \geq \mathbb{E}F_d(\boldsymbol{\Theta}) + \delta\big) \leq \exp\big(-\frac{\delta^2}{d\sigma^2}\big)$$

Finally, using $t = \delta/d$, we have

$$\mathbb{P}\big(\frac{F_d(\boldsymbol{\Theta})}{d} \geq \log 2 + \sigma^2/4 + t\big) \leq \exp\big(-\frac{dt^2}{\sigma^2}\big).$$

4. **(Balls-in-Bins)** (a) Let S be a set of C bins, and let X_S be the random variable denoting the total number of balls in the bins in the set S. Clearly, $X_S \sim \text{Binom}(T, \frac{1}{2})$. We can equivalently express $X_S = \sum_{i=1}^{T} Y_i$, where the r.v. Y_i is the indicator variable for the event that the i^{th} ball lands in the set S. Clearly, the r.v.s $\{Y_i\}$ are i.i.d $\sim$ Bernoulli$(\frac{1}{2})$. Hence, for any $\lambda > 0$, we have

$$\mathbb{E}\bigg(\exp\big(\lambda(X_S - \frac{T}{2})\big)\bigg) \leq \exp(\lambda^2 T/8).$$

Thus, $X_S \sim \text{subG}(\sqrt{T}/2)$. Note that $M_C(T) = \max_{S:|S|=C} X_S$. Using Massart's lemma, we have

$$\mathbb{E}(M_C(T)) \leq \frac{T}{2} + \frac{\sqrt{T}}{2}\sqrt{2\ln\binom{2C}{C}} \leq \frac{T}{2} + \sqrt{CT}.$$

(b)

(Case: $C = 1$**)** Let the r.v. $Z \equiv \sum_{t=1}^{T} W_t$, denote the summation of T i.i.d. uniform Bernoulli random variables. Hence Z is Binomially distributed with the parameters $(T, 1/2)$. Using linearity of expectation, we can write Observe that, we can write

$$\max\big\{Z, T - Z\big\} \;\; = \;\; \frac{T}{2} + |Z - T/2|. \tag{3}$$

The mean absolute deviation for a symmetric binomial random variable may be computed in closed form by using De Moivre's formula ([1], Eqn. (1)) as follows:

$$\mathbb{E}\big|Z - \frac{T}{2}\big| = \frac{1}{2^T}\Big(\lfloor\frac{T}{2}\rfloor + 1\Big)\binom{T}{\lfloor\frac{T}{2}\rfloor + 1}. \tag{4}$$

Eqn. (4), in combination with a non-asymptotic form of Stirling's formula [2], yields the following *non-asymptotic* lower bound

$$\mathbb{E}\big|Z - \frac{T}{2}\big| \geq \sqrt{\frac{T}{2\pi}} - \frac{1}{2\sqrt{2\pi T}}, \quad \forall T \geq 1. \tag{5}$$

(Case: Arbitrary $C \geq 1$**)** We index the bins sequentially as $1, 2, \ldots, 2C$. Next, we logically combine every two consecutive bins $\{(2i-1, 2i)\}, 1 \leq i \leq C$, to obtain C *Super bins* (See Figure 1). Let us denote the (random) number of balls in the i^{th} super bin by $X_i, j = 1, 2, \ldots, C$. Conditioned on the r.v. X_i, the number of balls in the corresponding bins: $2i-1$ and $2i$ are jointly distributed as $(Z, X_i - Z)$, where Z is a binomial random variable with parameter $(X_i, \frac{1}{2})$. Let H_i denote the maximum

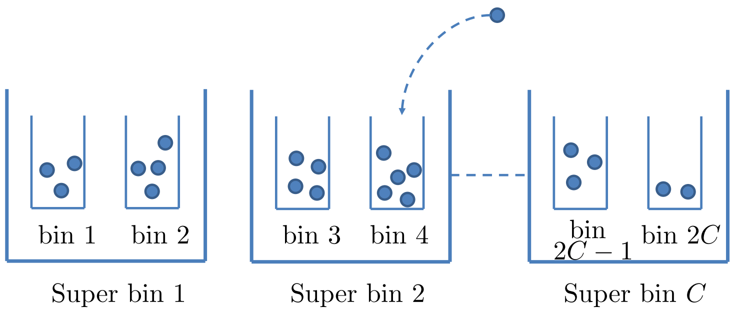

Figure 1: Illustrating the construction of Super bins

number of balls between the corresponding bins $2i-1$ and $2i$. Then, as shown above, when $X_i > 0$:

$$\mathbb{E}(H_i|X_i) \geq \frac{X_i}{2} + \sqrt{\frac{X_i}{2\pi}} - \frac{1}{2\sqrt{2\pi X_i}}, \quad \forall 1 \leq i \leq C. \tag{6}$$

Since $M_C \geq \sum_{i=1}^{C} H_i$, we have

$$\begin{aligned}
& \mathbb{E}(M_C) \\
\geq \;& \mathbb{E}\left(\sum_{i=1}^{C} H_i\right) = \sum_{i=1}^{C} \mathbb{E}(H_i) \overset{(a)}{=} C\mathbb{E}(H_1) \overset{(b)}{=} C\mathbb{E}\big(H_1 \mathbb{1}(X_1 > 0)\big) \\
\overset{(c)}{=} \;& C\mathbb{E}\mathbb{E}(H_1 \mathbb{1}(X_1 > 0)|X_1) \\
\overset{(d)}{\geq} \;& C\mathbb{E}\left(\frac{X_1 \mathbb{1}(X_1 > 0)}{2} + \sqrt{\frac{X_1}{2\pi}} \mathbb{1}(X_1 > 0) - \frac{\mathbb{1}(X_1 > 0)}{2\sqrt{2\pi X_1}}\right) \\
\overset{(e)}{=} \;& \frac{C}{2}\mathbb{E}(X_1) + \frac{C}{\sqrt{2\pi}}\mathbb{E}(\sqrt{X_1}) - \frac{C}{2\sqrt{2\pi}}\mathbb{E}\left(\frac{\mathbb{1}(X_1 > 0)}{\sqrt{X_1}}\right),
\end{aligned} \tag{7}$$

where the equality (a) follows from the fact that the random variables $\{H_i\}_{i=1}^{C}$ have identical distribution. For equation (b), we write

$$H_1 = H_1 \mathbb{1}(X_1 = 0) + H_1 \mathbb{1}(X_1 > 0).$$

Now, observe that if $X_1 = 0$ then $H_1 = 0$ a.s. Hence, almost surely, we have $H_1 = H_1 \mathbb{1}(X_1 > 0)$. The equation (c) follows from the tower property of conditional expectation, the inequality (d) follows from the bound (6), and the equality (e) follows from the facts that $X_1 = X_1 \mathbb{1}(X_1 > 0)$, $\sqrt{X_1} = \sqrt{X_1}\mathbb{1}(X_1 > 0)$. The Lemma now follows by using the bounds on moments of the binomial distribution as computed next.

Bounding the Expectations in Eqn. (7)**:** For bounding the middle term in Eqn. (7), consider the factorization:

$$\sqrt{x} - \left(1 + \frac{x-1}{2} - \frac{(x-1)^2}{2}\right) = \frac{\sqrt{x}}{2}(\sqrt{x}-1)^2(\sqrt{x}+2).$$

The RHS is non-negative for any $x \geq 0$. Thus, we have the following algebraic inequality:

$$\sqrt{x} \geq 1 + \frac{x-1}{2} - \frac{(x-1)^2}{2}, \quad \forall\ x \geq 0.$$

Replacing the variable x with the random variable $\frac{X_1}{\mathbb{E}(X_1)}$ point wise, we have almost surely,

$$\sqrt{\frac{X_1}{\mathbb{E}(X_1)}} \geq 1 + \frac{\frac{X_1}{\mathbb{E}(X_1)} - 1}{2} - \frac{(\frac{X_1}{\mathbb{E}(X_1)} - 1)^2}{2}.$$

Taking expectation of both sides, the above yields

$$\mathbb{E}(\sqrt{X_1}) \geq \sqrt{\mathbb{E}(X_1)}\left(1 - \frac{\mathsf{Var}(X_1)}{2(\mathbb{E}(X_1))^2}\right). \tag{8}$$

Finally, recall that $X_1 \sim \mathsf{Binom}(T, \frac{1}{C})$. Hence, $\mathbb{E}(X_1) = \frac{T}{C}$ and $\mathsf{Var}(X_1) = T\frac{1}{C}(1-\frac{1}{C}) \leq \frac{T}{C}$. Using this, Eqn. (8) yields the following lower bound

$$\mathbb{E}(\sqrt{X_1}) \geq \sqrt{\frac{T}{C}} - \frac{1}{2}\sqrt{\frac{C}{T}}. \tag{9}$$

For bounding the last term in Eqn. (7), we write

$$\frac{\mathbb{1}(X_1 > 0)}{\sqrt{X_1}} \leq 1\mathbb{1}\big(X_1 \leq \frac{T}{2C}\big) + \sqrt{\frac{2C}{T}}.$$

Recall that $X_1 \sim \mathsf{Binom}(T, \frac{1}{C})$. Taking expectation of both sides of the above inequality, we have

$$\mathbb{E}\left(\frac{\mathbb{1}(X_1 > 0)}{\sqrt{X_1}}\right) \leq \mathbb{P}(X_1 \leq \frac{1}{2}\mathbb{E}(X_1)) + \sqrt{\frac{2C}{T}}.$$

The probability term in the above expression may be bounded using Chebyshev's inequality as follows:

$$\begin{aligned}
\mathbb{P}\big(X_1 \leq \frac{1}{2}\mathbb{E}(X_1)\big) &= \mathbb{P}\big(X_1 - \mathbb{E}(X_1) \leq -\frac{1}{2}\mathbb{E}(X_1)\big) \\
&\leq \mathbb{P}\big(|X_1 - \mathbb{E}(X_1)| \geq \frac{1}{2}\mathbb{E}(X_1)\big) \\
&\leq \frac{4\mathsf{Var}(X_1)}{(\mathbb{E}(X_1))^2} = 4\frac{C-1}{T}
\end{aligned}$$

Taking the above bounds together, we obtain

$$\mathbb{E}\bigg(\frac{\mathbb{1}(X_1 > 0)}{\sqrt{X_1}}\bigg) \leq \sqrt{\frac{2C}{T}} + \frac{4C}{T}. \tag{10}$$

Finally, combining Eqns. (7), (9), and (10) together, we obtain

$$\mathbb{E}(M_C) \geq \frac{T}{2} + \sqrt{\frac{CT}{2\pi}} - \frac{(\sqrt{2}+1)C^{3/2}}{2\sqrt{2\pi T}} - \sqrt{\frac{2}{\pi}}\frac{C^2}{T}. \quad \blacksquare$$

See reference [3] for an application of the above problem in online caching.

5. **(Predicting Binary Sequences)** Consider any partition P from the set of all partitions $\mathcal{P}_n$ that partitions the n-length sequence $\boldsymbol{y}_1^n$ into k parts. Consider the function $\phi_n(P, \boldsymbol{y})$ defined as follows:

$$\phi_n(P, \boldsymbol{y}) = \frac{1}{n}\sum_{j=1}^{k} \min\big(\sum_{i \in P_j} y_i, \sum_{i \in P_j}(1-y_i)\big),$$

where P_j denotes the j^{th} partition. Finally, we define the function ϕ_n as

$$\phi_n(\boldsymbol{y}) = \inf_{P \in \mathcal{P}} \phi_n(P, \boldsymbol{y}) + R_n,$$

where R_n is chosen such that $\mathbb{E}\phi_n = \frac{1}{2}$, where the expectation is taken over i.i.d. uniform Bernoulli $\boldsymbol{y}$. Note that, $\phi_n(P, \boldsymbol{y})$ may be alternatively expressed as:

$$\phi_n(P, \boldsymbol{y}) = \frac{1}{2} - \frac{1}{n}\sum_{j=1}^{k} \big|\sum_{i \in P_j} y_i - \frac{|P_j|}{2}\big|.$$

Thus, we require

$$R_n = \frac{1}{n}\mathbb{E}\bigg(\sup_{P \in \mathcal{P}} \sum_{j=1}^{k} \big|\sum_{i \in P_j} y_i - \frac{|P_j|}{2}\big|\bigg).$$

Clearly, the function ϕ_n satisfies the stability condition. Define

$$g(P) = \sum_{j=1}^{k} \big|\sum_{i \in P_j} y_i - \frac{|P_j|}{2}\big|.$$

For any $\lambda > 0$, the MGF of $g(P)$ for any partition P is bounded as:

$$\mathbb{E}\exp(\lambda g(P)) \leq \prod_{j=1}^{k} 2 \prod_{i \in P_j} \mathbb{E}\big(\exp(\lambda(y_i - \frac{1}{2})\big) \leq 2^k \exp(n\lambda^2/4).$$

Using the standard soft-max trick, we have

$$\exp(\lambda \mathbb{E} \sup_P g(P)) \leq \mathbb{E}(\exp(\lambda \sup_P g(P)) \leq |\mathcal{P}| 2^k \exp(n\lambda^2/4).$$

Thus,

$$\mathbb{E} \sup_P g(P) \leq \frac{k \ln 2 + \ln |\mathcal{P}|}{\lambda} + \frac{n\lambda}{4}.$$

From basic combinatorics, we have

$$\log |\mathcal{P}| \leq \log \binom{n}{k} \leq k \log \frac{ne}{k}.$$

Choosing appropriate $\lambda > 0$, we conclude that

$$R_n \leq c\sqrt{\frac{k}{n} \log \frac{ne}{k}},$$

for some universal constant c.

References

[1] Daniel Berend and Aryeh Kontorovich. A sharp estimate of the binomial mean absolute deviation with applications. *Statistics & Probability Letters*, 83(4):1254–1259, 2013.

[2] Herbert Robbins. A remark on stirling's formula. *The American mathematical monthly*, 62(1):26–29, 1955.

[3] Rajarshi Bhattacharjee, Subhankar Banerjee, and Abhishek Sinha. Fundamental limits on the regret of online network-caching. In *Abstracts of the 2020 SIGMETRICS/Performance Joint International Conference on Measurement and Modeling of Computer Systems*, pages 15–16, 2020.