Assignment 4 - Solutions

00103335: Deep Learning and Reinforcement Learning

Fall 2026

Problem 1. Bennett's inequality

(a) Boundedness justifies integrating the exponential power series. For every integer $k \geq 2$,

$$|\mathbb{E}X_i^k| \leq \mathbb{E}|X_i|^k \leq b^{k-2}\mathbb{E}X_i^2 = b^{k-2}\sigma_i^2.$$

Since $\mathbb{E}X_i = 0$ and $\lambda \geq 0$,

$$\begin{aligned}\mathbb{E}e^{\lambda X_i} &= 1 + \sum_{k=2}^{\infty} \frac{\lambda^k \mathbb{E}X_i^k}{k!} \\ &\leq 1 + \frac{\sigma_i^2}{b^2} \sum_{k=2}^{\infty} \frac{(\lambda b)^k}{k!} = 1 + \frac{\sigma_i^2}{b^2}(e^{\lambda b} - 1 - \lambda b).\end{aligned}$$

Applying $\log(1+x) \leq x$ for $x \geq 0$ proves

$$\boxed{\log \mathbb{E}e^{\lambda X_i} \leq \frac{\sigma_i^2}{b^2}(e^{\lambda b} - 1 - \lambda b).}$$

(b) Independence and part (a) give

$$\log \mathbb{E} \exp\left(\lambda \sum_{i=1}^{n} X_i\right) \leq \frac{nv}{b^2}(e^{\lambda b} - 1 - \lambda b).$$

Consequently, for every $\lambda \geq 0$, exponential Markov gives

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq nt\right) \leq \exp\left[-n\lambda t + \frac{nv}{b^2}(e^{\lambda b} - 1 - \lambda b)\right].$$

The exponent is minimized at $\lambda = b^{-1}\log(1 + bt/v) > 0$. Writing $u = bt/v$, its value there is

$$-\frac{nv}{b^2}\{(1+u)\log(1+u) - u\} = -\frac{nv}{b^2}h(u).$$

This proves the upper-tail bound. Applying the same argument to $-X_1, \ldots, -X_n$ gives the lower-tail bound

$$\boxed{\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \leq -t\right) \leq \exp\left[-\frac{nv}{b^2}h\left(\frac{bt}{v}\right)\right].}$$

(c) Set

$$g(u) = h(u) - \frac{u^2}{2(1+u/3)} = h(u) - \frac{3u^2}{2(u+3)}.$$

Direct differentiation gives $g(0) = g'(0) = 0$ and

$$g''(u) = \frac{1}{1+u} - \frac{27}{(u+3)^3} = \frac{u^2(u+9)}{(1+u)(u+3)^3} \geq 0 \qquad (u \geq 0).$$

Thus $g'(u) \geq 0$ and then $g(u) \geq 0$. Substituting this inequality into part (b) yields

$$\frac{nv}{b^2} h\left(\frac{bt}{v}\right) \geq \frac{nt^2}{2(v + bt/3)}.$$

The union bound for the two tails therefore gives

$$\boxed{\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| \geq t\right) \leq 2\exp\left[-\frac{nt^2}{2(v + bt/3)}\right].}$$

Problem 2. Bernstein tails and expected error

(a) For $t \geq 0$, the inequality $\nu^2 + bt \leq 2\max\{\nu^2, bt\}$ implies

$$\exp\left[-\frac{t^2}{2(\nu^2 + bt)}\right] \leq e^{-t^2/(4\nu^2)} + e^{-t/(4b)}.$$

Combining the given tail bound with the bound by 1, and using $\min\{1, a + c\} \leq \min\{1, a\} + \min\{1, c\}$ for $a, c \geq 0$, gives

$$\mathbb{P}(Z \geq t) \leq \min\{1, Ce^{-t^2/(4\nu^2)}\} + \min\{1, Ce^{-t/(4b)}\}.$$

Let $a = 2\nu\sqrt{\log C}$. For $s \geq 0$, $(a + s)^2 \geq a^2 + s^2$ and $Ce^{-a^2/(4\nu^2)} = 1$. Hence

$$\begin{aligned}\int_0^\infty \min\{1, Ce^{-t^2/(4\nu^2)}\}\, dt &\leq a + C\int_a^\infty e^{-t^2/(4\nu^2)}\, dt \\ &\leq 2\nu\sqrt{\log C} + \int_0^\infty e^{-s^2/(4\nu^2)}\, ds = 2\nu\sqrt{\log C} + \nu\sqrt{\pi}.\end{aligned}$$

The second integral is explicit:

$$\int_0^\infty \min\{1, Ce^{-t/(4b)}\}\, dt = 4b\log C + C\int_{4b\log C}^\infty e^{-t/(4b)}\, dt = 4b(1 + \log C).$$

Tonelli's identity $\mathbb{E}Z = \int_0^\infty \mathbb{P}(Z > t)\, dt$ now gives

$$\begin{aligned}\mathbb{E}Z &\leq 2\nu\sqrt{\log C} + \nu\sqrt{\pi} + 4b(1 + \log C) \\ &\leq \boxed{2\nu(\sqrt{\pi} + \sqrt{\log C}) + 4b(1 + \log C)}.\end{aligned}$$

(b) The moment assumption bounds $|\mathbb{E}X_1^k|$, rather than $\mathbb{E}|X_1|^k$ for odd k, so we first justify the MGF expansion. For $0 \leq s < 1/b$, Tonelli's theorem for the nonnegative series of $\cosh(sX_1)$ gives

$$\begin{aligned}\mathbb{E}\cosh(sX_1) &= 1 + \sum_{m=1}^{\infty} \frac{s^{2m}\mathbb{E}X_1^{2m}}{(2m)!} \\ &\leq 1 + \frac{\sigma^2}{2}\sum_{m=1}^{\infty} s^{2m}b^{2m-2} = 1 + \frac{\sigma^2 s^2}{2(1 - b^2s^2)} < \infty.\end{aligned}$$

Since $e^{s|x|} \le 2\cosh(sx)$, this proves $\mathbb{E}e^{s|X_1|} < \infty$. Thus, for $|\lambda| < 1/b$, the exponential series is absolutely integrable, and exchanging expectation and summation is valid. Using the signed-moment bounds and $\mathbb{E}X_1 = 0$,

$$\begin{aligned}\mathbb{E}e^{\lambda X_1} &= 1 + \sum_{k=2}^{\infty} \frac{\lambda^k \mathbb{E}X_1^k}{k!} \\ &\le 1 + \frac{\sigma^2}{2} \sum_{k=2}^{\infty} |\lambda|^k b^{k-2} = 1 + \frac{\sigma^2\lambda^2}{2(1-b|\lambda|)}.\end{aligned}$$

It follows that

$$\log \mathbb{E}e^{\lambda X_1} \le \frac{\sigma^2\lambda^2}{2(1-b|\lambda|)} \qquad (|\lambda| < 1/b).$$

If $\sigma = 0$, then $X_i = 0$ almost surely and the conclusion is immediate. Suppose $\sigma > 0$. By independence, with $\nu_n^2 = \sigma^2/n$ and $b_n = b/n$,

$$\log \mathbb{E}e^{\lambda \overline{X}_n} \le \frac{\nu_n^2\lambda^2}{2(1-b_n|\lambda|)} \qquad (|\lambda| < 1/b_n).$$

For $t > 0$, apply exponential Markov with $\lambda = t/(\nu_n^2 + b_n t) < 1/b_n$. The resulting upper-tail bound is $\exp[-t^2/(2(\nu_n^2 + b_n t))]$. The same argument for $-\overline{X}_n$, followed by a union bound, yields

$$\mathbb{P}(|\overline{X}_n| \ge t) \le 2\exp\left[-\frac{t^2}{2(\nu_n^2 + b_n t)}\right] = 2\exp\left[-\frac{nt^2}{2(\sigma^2 + bt)}\right].$$

The bound also holds at $t = 0$. Applying part (a) with $Z = |\overline{X}_n|$, $C = 2$, $\nu = \sigma/\sqrt{n}$, and scale b/n proves

$$\boxed{\mathbb{E}|\overline{X}_n| \le \frac{2\sigma}{\sqrt{n}}(\sqrt{\pi} + \sqrt{\log 2}) + \frac{4b}{n}(1 + \log 2).}$$

Problem 3. Maxima of light-tailed variables

The exponential-moment assumptions imply integrability of the maxima: for any admissible $\lambda > 0$, $e^{\lambda|X_i|} \le e^{\lambda X_i} + e^{-\lambda X_i}$. None of the following arguments uses independence.

(a) For $\lambda > 0$, Jensen's inequality and the pointwise bound $e^{\lambda \max_i X_i} \le \sum_i e^{\lambda X_i}$ give

$$\exp\left(\lambda \mathbb{E} \max_{i \le n} X_i\right) \le \mathbb{E}e^{\lambda \max_{i \le n} X_i} \le \sum_{i=1}^{n} \mathbb{E}e^{\lambda X_i} \le n e^{\sigma^2\lambda^2/2}.$$

Taking logarithms and dividing by λ yields

$$\mathbb{E} \max_{i \le n} X_i \le \frac{\log n}{\lambda} + \frac{\sigma^2\lambda}{2}.$$

For $n > 1$ and $\sigma > 0$, choose $\lambda = \sqrt{2\log n}/\sigma$ to obtain

$$\boxed{\mathbb{E} \max_{i \le n} X_i \le \sigma\sqrt{2\log n}.}$$

When $n = 1$, both the desired bound and $\mathbb{E}X_1$ equal zero. When $\sigma = 0$, letting $\lambda \to \infty$ in the preceding bound also gives the conclusion.

(b) Apply part (a) to the $2n$ variables $X_1, \dots, X_n, -X_1, \dots, -X_n$. They have the same sub-Gaussian variance proxy, and their maximum is $\max_i |X_i|$. Therefore

$$\mathbb{E} \max_{i \le n} |X_i| \le \sigma \sqrt{2 \log(2n)}.$$

For $n \ge 2$, $\log(2n) \le 2 \log n$, so

$$\boxed{\mathbb{E} \max_{i \le n} |X_i| \le \sigma \sqrt{2 \log(2n)} \le 2\sigma \sqrt{\log n}.}$$

(c) The argument in part (a) now gives only

$$\mathbb{E} \max_{i \le n} X_i \le \frac{\log n}{\lambda} + \frac{\nu^2 \lambda}{2} \qquad (0 < \lambda < 1/b).$$

Write $a = \log n > 0$ and $s = \sqrt{2a}$. Choose

$$\lambda = \frac{s}{\nu + bs}.$$

Because $\nu > 0$, this choice satisfies the strict restriction $0 < \lambda < 1/b$. Moreover,

$$\begin{aligned} \frac{a}{\lambda} + \frac{\nu^2 \lambda}{2} &= ba + \frac{\nu s}{2} + \frac{\nu^2 s}{2(\nu + bs)} \\ &\le ba + \nu s. \end{aligned}$$

Thus

$$\boxed{\mathbb{E} \max_{i \le n} X_i \le \nu \sqrt{2 \log n} + b \log n.}$$

For the maximum of absolute values, use the $2n$ variables $X_1, \dots, X_n, -X_1, \dots, -X_n$ and repeat the same calculation with $a = \log(2n)$. The assumption for negative λ supplies the required MGF bound for $-X_i$. This proves

$$\boxed{\mathbb{E} \max_{i \le n} |X_i| \le \nu \sqrt{2 \log(2n)} + b \log(2n).}$$

Problem 4. Symmetric Pareto variables and truncation

(a) For $u \ge 1$, symmetry and integration give

$$\mathbb{P}(|X_i| > u) = \alpha \int_u^\infty x^{-\alpha - 1} \, dx = u^{-\alpha}.$$

Also $\mathbb{E}|X_i| = \alpha/(\alpha - 1) < \infty$, so symmetry implies $\mathbb{E}X_i = 0$. Finally,

$$\boxed{\sigma_\alpha^2 = \mathbb{E}X_i^2 = \alpha \int_1^\infty x^{1 - \alpha} \, dx = \frac{\alpha}{\alpha - 2}.}$$

(b) The truncation is symmetric and bounded, so $\mathbb{E}Y_i = 0$ and $|Y_i| \le u$ almost surely. In fact,

$$\mathbb{E}Y_i^2 = \alpha \int_1^u x^{1 - \alpha} \, dx = \frac{\alpha}{\alpha - 2}(1 - u^{2 - \alpha}) \le \sigma_\alpha^2.$$

The Y_i are independent because each is a function of X_i alone.

(c) On the event $\max_i |X_i| \le u$, the sums of the X_i and Y_i agree. Consequently,

$$\begin{aligned}\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| \ge t\right) &\le \mathbb{P}\left(\left|\sum_{i=1}^n Y_i\right| \ge t\right) + \mathbb{P}\left(\max_{i\le n} |X_i| > u\right) \\ &\le 2\exp\left[-\frac{t^2}{2(n\sigma_\alpha^2 + ut/3)}\right] + nu^{-\alpha}.\end{aligned}$$

The last line uses Bernstein's inequality from Problem 1(c), with bound u and total variance at most $n\sigma_\alpha^2$, and a union bound for truncation. At $u = 1$, the variables Y_i vanish almost surely, so the same estimate remains valid.

(d) Since $nu^{-\alpha}$ decreases in u, its smallest admissible threshold is

$$\boxed{u = \left(\frac{2n}{\delta}\right)^{1/\alpha}.}$$

This is larger than 1, and it gives $nu^{-\alpha} = \delta/2$. Put $L = \log(4/\delta)$. With $t = n\varepsilon$, requiring the exponential term to equal $\delta/2$ is equivalent to

$$\frac{n\varepsilon^2}{2(\sigma_\alpha^2 + u\varepsilon/3)} = L, \qquad n\varepsilon^2 - \frac{2Lu}{3}\varepsilon - 2L\sigma_\alpha^2 = 0.$$

The unique positive root is

$$\boxed{\varepsilon_{n,\delta} = \frac{uL}{3n} + \sqrt{\frac{2\sigma_\alpha^2 L}{n} + \frac{u^2L^2}{9n^2}}, \qquad u = \left(\frac{2n}{\delta}\right)^{1/\alpha}, \quad L = \log\frac{4}{\delta}.}$$

Both terms in the probability bound from part (c) now equal $\delta/2$, hence

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| \ge n\varepsilon_{n,\delta}\right) \le \delta.$$

For fixed δ, $u = (2/\delta)^{1/\alpha} n^{1/\alpha}$. Because $\alpha > 2$, we have $u/n = o(n^{-1/2})$ and $u^2/n^2 = o(n^{-1})$. Thus the leading order is

$$\boxed{\varepsilon_{n,\delta} = \sigma_\alpha\sqrt{\frac{2\log(4/\delta)}{n}}\,(1 + o(1)).}$$

Problem 5. Covering the Euclidean ball

Throughout, $\varepsilon > 0$. An ε-net covers the set by closed Euclidean balls of radius ε centered at its net points. Let V_d denote the positive volume of B_2^d.

(a) If $z_1, \dots, z_M \in B_2^d$ form an ε-net, then

$$B_2^d \subseteq \bigcup_{j=1}^M (z_j + \varepsilon B_2^d).$$

By subadditivity, translation invariance, and scaling of volume,

$$V_d \le \sum_{j=1}^{M} \mathrm{vol}(z_j + \varepsilon B_2^d) = M\varepsilon^d V_d.$$

Dividing proves $M \ge \varepsilon^{-d}$, and therefore

$$\boxed{N(B_2^d, \varepsilon) \ge \varepsilon^{-d}.}$$

This argument holds for every $\varepsilon > 0$; for $\varepsilon \ge 1$, the lower bound is at most 1.

(b) Construct points in B_2^d greedily, adding a new point whenever its distance from every previously chosen point is strictly greater than ε. Any M points chosen this way have pairwise disjoint open balls of radius $\varepsilon/2$ around them. All these balls lie in the ball of radius $1 + \varepsilon/2$ about the origin. Comparing their volumes gives

$$M(\varepsilon/2)^d V_d \le (1 + \varepsilon/2)^d V_d, \qquad M \le (1 + 2/\varepsilon)^d.$$

In particular, the greedy procedure must stop after finitely many steps: otherwise, it would produce a finite initial segment violating this bound. At termination, every point of B_2^d lies within distance ε of a chosen point. The chosen points therefore form a net, all its centers belong to B_2^d, and

$$\boxed{N(B_2^d, \varepsilon) \le (1 + 2/\varepsilon)^d.}$$

No restriction such as $\varepsilon \le 1$ was used.

(c) Run the same greedy construction with all chosen points required to lie on S^{d-1}. They still lie in B_2^d, so the open balls of radius $\varepsilon/2$ around them are disjoint and contained in the ball of radius $1 + \varepsilon/2$. The same volume calculation forces termination after at most $(1 + 2/\varepsilon)^d$ points. Termination means that these points cover S^{d-1} within distance ε, and the centers are on the sphere by construction. Thus, for every $\varepsilon > 0$,

$$\boxed{N(S^{d-1}, \varepsilon) \le (1 + 2/\varepsilon)^d.}$$

The volume comparison uses ambient balls, not the volume of the sphere itself. At $\varepsilon = 1/4$, the bound is

$$\boxed{N(S^{d-1}, 1/4) \le 9^d.}$$