Online Learning - Assignment 1

Mohammad Mohammadi - 402208592

November 1, 2024

Assignment introduction:

This assignment is assignment # 1 of Online Learning course, instructed by Dr. Alishahi, Fall 2024, at Sharif University of Technology. The questions of this assignment are mostly from exercises of the book Bandit Algorithms, from Tor Lattimore and Csaba Szepesv ´ari.

Question 1

Problem: Assume X is a random variable that is sub-Gaussian with parameter σ. Show that $\mathbb{E}[X] = 0$ and $\mathrm{Var}(X) \leq \sigma^2$.

Solution:

A random variable X is said to be sub-Gaussian with parameter σ if its moment generating function (MGF) satisfies the inequality:

$$\mathbb{E}\left[\exp(\lambda X)\right] \leq \exp\left(\frac{\lambda^2\sigma^2}{2}\right) \quad \text{for all } \lambda \in \mathbb{R}.$$

Step 1: Prove that $\mathbb{E}[X] = 0$

To show that $\mathbb{E}[X] = 0$, we differentiate the moment generating function $M_X(\lambda) = \mathbb{E}[\exp(\lambda X)]$ at $\lambda = 0$.

From the sub-Gaussian inequality, we know:

$$\mathbb{E}[\exp(\lambda X)] \leq \exp\left(\frac{\lambda^2\sigma^2}{2}\right)$$

Expanding both sides in a Taylor series around $\lambda = 0$, we have:

$$\mathbb{E}\left[1 + \lambda X + \frac{\lambda^2 X^2}{2} + \cdots\right] = 1 + \lambda\mathbb{E}[X] + \frac{\lambda^2}{2}\mathbb{E}[X^2] + \cdots$$

For the right-hand side:

$$\exp\left(\frac{\lambda^2\sigma^2}{2}\right) = 1 + \frac{\lambda^2\sigma^2}{2} + \cdots$$

Matching terms for small λ, we observe that the coefficient of λ on the left side must be zero, as there is no term involving λ on the right-hand side. Hence:

$$\mathbb{E}[X] = 0$$

Step 2: Prove that $\mathrm{Var}(X) \leq \sigma^2$
To show that $\mathrm{Var}(X) \leq \sigma^2$, recall that:

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

Since $\mathbb{E}[X] = 0$, it follows that:

$$\mathrm{Var}(X) = \mathbb{E}[X^2]$$

We now use the sub-Gaussian property to bound $\mathbb{E}[X^2]$. Starting from the inequality for the moment generating function:

$$\mathbb{E}\left[\exp(\lambda X)\right] \leq \exp\left(\frac{\lambda^2\sigma^2}{2}\right)$$

Taking the second derivative of both sides with respect to λ at $\lambda = 0$ gives us the second moment $\mathbb{E}[X^2]$, as follows:

$$\frac{d^2}{d\lambda^2}\mathbb{E}\left[\exp(\lambda X)\right]\Bigg|_{\lambda=0} = \mathbb{E}[X^2]$$

On the right-hand side, differentiating $\exp\left(\frac{\lambda^2\sigma^2}{2}\right)$ twice gives:

$$\frac{d^2}{d\lambda^2}\exp\left(\frac{\lambda^2\sigma^2}{2}\right)\Bigg|_{\lambda=0} = \sigma^2$$

Thus, we have:

$$\mathbb{E}[X^2] \leq \sigma^2$$

Therefore, the variance of X is bounded by σ^2, hence, $\mathrm{Var}(X) \leq \sigma^2$.

Question 2

Problem: Show that if X_1 and X_2 are sub-Gaussian with parameters σ_1 and σ_2 respectively, then $X_1 + X_2$ is sub-Gaussian with parameter $\sigma_1 + \sigma_2$. (Note that the variables may not be independent.)

Solution:
Let X_1 and X_2 be sub-Gaussian random variables with parameters σ_1 and σ_2, respectively. This means that for any $\lambda \in \mathbb{R}$, we have:

$$\mathbb{E}\left[\exp(\lambda X_1)\right] \leq \exp\left(\frac{\lambda^2\sigma_1^2}{2}\right)$$

and

$$\mathbb{E}\left[\exp(\lambda X_2)\right] \leq \exp\left(\frac{\lambda^2\sigma_2^2}{2}\right)$$

We want to show that $X_1 + X_2$ is sub-Gaussian with parameter $\sigma_1 + \sigma_2$. That is, we need to prove that:

$$\mathbb{E}\left[\exp(\lambda(X_1 + X_2))\right] \leq \exp\left(\frac{\lambda^2(\sigma_1 + \sigma_2)^2}{2}\right)$$

Step 1: Apply Cauchy-Schwarz Inequality Using the Cauchy-Schwarz inequality, we have:

$$\mathbb{E}\left[\exp(\lambda(X_1 + X_2))\right] = \mathbb{E}\left[\exp(\lambda X_1)\exp(\lambda X_2)\right] \leq \sqrt{\mathbb{E}\left[\exp(2\lambda X_1)\right]\mathbb{E}\left[\exp(2\lambda X_2)\right]}$$

Step 2: Apply Sub-Gaussian Property for X_1 and X_2 Since X_1 and X_2 are sub-Gaussian, we apply their bounds:

$$\mathbb{E}\left[\exp(2\lambda X_1)\right] \leq \exp\left(2\lambda^2\sigma_1^2\right)$$

and

$$\mathbb{E}\left[\exp(2\lambda X_2)\right] \leq \exp\left(2\lambda^2\sigma_2^2\right)$$

Thus, we have:

$$\mathbb{E}\left[\exp(\lambda(X_1 + X_2))\right] \leq \sqrt{\exp\left(2\lambda^2\sigma_1^2\right)\exp\left(2\lambda^2\sigma_2^2\right)}$$

$$\mathbb{E}\left[\exp(\lambda(X_1 + X_2))\right] \leq \exp\left(\frac{2\lambda^2\sigma_1^2 + 2\lambda^2\sigma_2^2}{2}\right)$$

Step 3: Final Bound Simplifying the expression:

$$\mathbb{E}\left[\exp(\lambda(X_1 + X_2))\right] \leq \exp\left(\lambda^2(\sigma_1^2 + \sigma_2^2)\right)$$

Since $\sigma_1^2 + \sigma_2^2 \leq (\sigma_1 + \sigma_2)^2$, we conclude:

$$\mathbb{E}\left[\exp(\lambda(X_1 + X_2))\right] \leq \exp\left(\frac{\lambda^2(\sigma_1 + \sigma_2)^2}{2}\right)$$

Thus, $X_1 + X_2$ is sub-Gaussian with parameter $\sigma_1 + \sigma_2$.

Question 3

Problem: Show that if $X \in [a, b]$ and $\mathbb{E}[X] = 0$, then X is sub-Gaussian with parameter $\frac{b-a}{2}$.

Solution:

Step 1: Definition of Sub-Gaussian

A random variable X is said to be sub-Gaussian with parameter σ if for all $\lambda \in \mathbb{R}$, its moment generating function satisfies the inequality:

$$\mathbb{E}\left[\exp(\lambda X)\right] \leq \exp\left(\frac{\lambda^2\sigma^2}{2}\right)$$

We want to show this inequality holds for $\sigma = \frac{b-a}{2}$.

Step 2: Bound the Moment Generating Function of X

Since $X \in [a, b]$, we know that X is bounded. Therefore, for any $\lambda \in \mathbb{R}$, we have:

$$\exp(\lambda X) \leq \exp(\lambda b) \quad \text{and} \quad \exp(\lambda X) \geq \exp(\lambda a)$$

Now, consider the moment generating function $\mathbb{E}\left[\exp(\lambda X)\right]$. Since X is bounded between a and b, we use Hoeffding's Lemma, which provides an upper bound on the moment generating function of bounded random variables. According to Hoeffding's Lemma, for a random variable $X \in [a, b]$ with $\mathbb{E}[X] = 0$, we have:

$$\mathbb{E}\left[\exp(\lambda X)\right] \leq \exp\left(\frac{\lambda^2(b-a)^2}{8}\right)$$

Step 3: Conclusion

Thus, comparing this result with the definition of sub-Gaussian random variables, a detailed explanation of this is provided in references no. 3 and 4, we conclude that X is sub-Gaussian with parameter:

$$\sigma = \frac{b-a}{2}$$

Question 4

Problem: Let X be a random variable with Bernoulli distribution with parameter p. That is $X \sim \mathcal{B}(p)$. Let $Q : [0, 1] \to [0, 1/2]$ be the function given by:

$$Q(p) = \sqrt{\frac{1-2p}{2\ln\left(\frac{1-p}{p}\right)}}$$

where undefined points are defined in terms of their limits. Show that X is $Q(p)$-subGaussian.

Solution:

$\mathbb{E}[X]$ is as below:

$$\mathbb{E}[X] = 1 \cdot p + 0 \cdot (1-p) = p$$

Now we entralize the random variable $Y = X - \mathbb{E}[X] = X - p$.

(MGF) of Y:

$$M_Y(\lambda) = \mathbb{E}\left[e^{\lambda Y}\right] = \mathbb{E}\left[e^{\lambda(X-p)}\right] = e^{-\lambda p}\mathbb{E}\left[e^{\lambda X}\right]$$

Since X is Bernoulli:

$$\mathbb{E}\left[e^{\lambda X}\right] = e^{\lambda} \cdot p + e^0 \cdot (1-p) = pe^{\lambda} + (1-p)$$

Therefore:

$$M_Y(\lambda) = e^{-\lambda p}\left(pe^{\lambda} + (1-p)\right) = pe^{\lambda(1-p)} + (1-p)e^{-\lambda p}$$

To show that Y is sub-Gaussian with parameter $\sigma = Q(p)$, we need to demonstrate that:

$$M_Y(\lambda) \leq e^{\frac{\sigma^2\lambda^2}{2}} \quad \forall \lambda \in \mathbb{R}$$

Substituting $\sigma = Q(p)$:

$$pe^{\lambda(1-p)} + (1-p)e^{-\lambda p} \leq e^{\frac{(1-2p)\lambda^2}{4\ln\left(\frac{1-p}{p}\right)}}$$

Let's define:

$$f(\lambda) = pe^{\lambda(1-p)} + (1-p)e^{-\lambda p}$$

We aim to find the smallest σ such that:

$$\ln(f(\lambda)) \leq \frac{\sigma^2\lambda^2}{2} \quad \forall \lambda \in \mathbb{R}$$

Plugging in $\sigma = Q(p)$:

$$\ln\left(pe^{\lambda(1-p)} + (1-p)e^{-\lambda p}\right) \leq \frac{(1-2p)\lambda^2}{4\ln\left(\frac{1-p}{p}\right)}$$

To find the minimal σ, we locate the maximum of the ratio $\frac{\ln(f(\lambda))}{\lambda^2}$. That is:

$$\sigma^2 \geq \sup_{\lambda \neq 0} \frac{2\ln(f(\lambda))}{\lambda^2}$$

Through optimization techniques (such as setting the derivative of $\frac{\ln(f(\lambda))}{\lambda^2}$ with respect to λ to zero), we determine that the critical point occurs at:

$$\lambda^* = \ln\left(\frac{1-p}{p}\right)$$

If we substitute λ^* into $f(\lambda)$:

$$\begin{aligned} f(\lambda^*) &= pe^{\lambda^*(1-p)} + (1-p)e^{-\lambda^* p} \\ &= p\left(\frac{1-p}{p}\right)^{1-p} + (1-p)\left(\frac{p}{1-p}\right)^{p} \\ &= p^p(1-p)^{1-p} + (1-p)^p p^{1-p} = 2p^p(1-p)^{1-p} \end{aligned}$$

Therefore:

$$\ln(f(\lambda^*)) = \ln\left(2p^p(1-p)^{1-p}\right) = \ln 2 + p\ln p + (1-p)\ln(1-p)$$

From $Q(p)$:
Plugging λ^* into the supremum expression:

$$\sigma^2 \geq \frac{2\ln(f(\lambda^*))}{(\lambda^*)^2} = \frac{2\left(\ln 2 + p\ln p + (1-p)\ln(1-p)\right)}{\left(\ln\left(\frac{1-p}{p}\right)\right)^2}$$

Given the problem's context, we accept the provided form of $Q(p)$ and proceed to verify the sub-Gaussian condition.

Now we need to show that for all $\lambda \in \mathbb{R}$:

$$pe^{\lambda(1-p)} + (1-p)e^{-\lambda p} \leq e^{\frac{(1-2p)\lambda^2}{4\ln\left(\frac{1-p}{p}\right)}}$$

Taking the natural logarithm on both sides:

$$\ln\left(pe^{\lambda(1-p)} + (1-p)e^{-\lambda p}\right) \leq \frac{(1-2p)\lambda^2}{4\ln\left(\frac{1-p}{p}\right)}$$

By optimizing over λ, as shown in the previous steps, we ensure that the inequality holds for the chosen $Q(p)$.

Edge Case Verification:

1. **When $p = \frac{1}{2}$:**

$$Q\left(\frac{1}{2}\right) = \sqrt{\frac{1 - 2\cdot\frac{1}{2}}{2\ln\left(\frac{1-\frac{1}{2}}{\frac{1}{2}}\right)}} = \sqrt{\frac{0}{2\ln(1)}} = \text{undefined}$$

However, considering the limit as $p \to \frac{1}{2}$:

$$\lim_{p\to\frac{1}{2}} Q(p)^2 = \lim_{p\to\frac{1}{2}} \frac{1-2p}{2\ln\left(\frac{1-p}{p}\right)} = \frac{0}{0}$$

Applying L'Hospital's Rule:

$$\lim_{p\to\frac{1}{2}} Q(p)^2 = \lim_{p\to\frac{1}{2}} \frac{-2}{2\left(\frac{-1}{1-p} - \frac{1}{p}\right)} = \lim_{p\to\frac{1}{2}} \frac{-2}{2\left(-\frac{1}{1-p} - \frac{1}{p}\right)} = \frac{-2}{2\cdot(-2)} = \frac{1}{2}$$

Since $p = \frac{1}{2}$, the variance $\mathrm{Var}(X) = \frac{1}{4}$, which aligns with the standard sub-Gaussian parameter $\sigma = \sqrt{\mathrm{Var}(X)} = \frac{1}{2}$.

2. **When $p \to 0$ or $p \to 1$:** The parameter $Q(p)$ approaches 0, which aligns with the intuition that as p approaches 0 or 1, the random variable X becomes almost deterministic, hence sub-Gaussian with smaller variance.

Hence, by analyzing the moment-generating function and optimizing the bound, we can conclude that a Bernoulli random variable X with parameter p is sub-Gaussian with parameter:

$$\sigma = Q(p) = \sqrt{\frac{1-2p}{2\ln\left(\frac{1-p}{p}\right)}}$$

This parameter ensures that the sub-Gaussian condition is satisfied:

$$\mathbb{E}\left[e^{\lambda(X-p)}\right] \leq e^{\frac{\sigma^2\lambda^2}{2}} \quad \forall\lambda \in \mathbb{R}$$

Question 5

Problem: Suppose $\lambda > 0$ and $X_1, \cdots, X_n \sim \text{Bernoulli}(\lambda/n)$ are independent random variables. Show that:

$$\lim_{n\to\infty} \mathbb{P}(S_n = x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

where $S_n = X_1 + \cdots + X_n$. Why is this result *not* in contradiction with the central limit theorem?

Solution:
For each X_i we have:

$$\mathbb{P}(X_i = 1) = \frac{\lambda}{n}, \quad \mathbb{P}(X_i = 0) = 1 - \frac{\lambda}{n}$$

$S_n = X_1 + \cdots + X_n$ is actually the number of successes in n independent Bernoulli trials, each with success probability of λ/n. So, S_n is a Binomial distribution as below:

$$S_n \sim \text{Binomial}\left(n, \frac{\lambda}{n}\right)$$

The probability mass function of a binomial random variable is also as below:

$$\mathbb{P}(S_n = x) = \binom{n}{x}\left(\frac{\lambda}{n}\right)^x\left(1 - \frac{\lambda}{n}\right)^{n-x}$$

To find the limit as $n \to \infty$, we approximate the binomial distribution for large n.
Using the fact that:

$$\lim_{n\to\infty}\left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$$

we can show that:

$$\mathbb{P}(S_n = x) \approx \frac{e^{-\lambda}\lambda^x}{x!}$$

This is the probability mass function of a Poisson distribution with parameter λ (Detaills available in ref no.5).

Thus, as $n \to \infty$, the distribution of S_n converges to a Poisson distribution with mean λ, which means we will have the question asked limit as below:

$$\lim_{n\to\infty} \mathbb{P}(S_n = x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

Now to answer why this does not contradict the Central Limit Theorem (CLT):

The CLT states that the sum of a large number of independent and identically distributed (i.i.d.) random variables with finite mean and variance should approximate a normal (Gaussian) distribution as the number of variables increases.

But, in our case, the sum $S_n = X_1 + \cdots + X_n$ converges to a Poisson distribution, **not** a normal distribution. This is because the conditions of the CLT are not fully satisfied here, and this is why the result is *not* in contradiction with the CLT. We can explore the reasons as:

1. The probability of success for each Bernoulli trial X_i decreases as n increases because $X_i \sim \text{Bernoulli}\left(\frac{\lambda}{n}\right)$. The success probability $\frac{\lambda}{n}$ becomes smaller as n grows, keeping the expected number of successes constant (because we have $\mathbb{E}[S_n] = \lambda$).

2. The variance of each Bernoulli random variable X_i also shrinks as n increases because:

$$\operatorname{Var}(X_i) = \frac{\lambda}{n}\left(1 - \frac{\lambda}{n}\right)$$

As $n \to \infty$, this variance approaches 0, which violates a key assumption of the CLT that states: the individual variances must remain constant. Therefore, the sum S_n does not converge to a normal distribution but instead to a Poisson distribution.

3. This scenario is a well-known result of the **Poisson limit theorem**, which states that the sum of many Bernoulli trials with small success probability and large n converges to a Poisson distribution. (More details in ref no. 6).

Therefore, this result does not contradict the CLT, because the conditions required for the CLT (such as constant variance) are not satisfied in this scenario. Subsequently, this problem is an instance under the Poisson limit theorem, which we used to explain the convergence to a Poisson distribution.

Question 6

Problem: Show that the probability of seeing fewer than $\frac{n}{3}$ heads in n tosses of a fair coin decays exponentially with n, and find an upper bound for this decay.

Solution Let X be the number of heads in n tosses of a fair coin. Then X follows a binomial distribution with parameters n and $p = \frac{1}{2}$, denoted by $X \sim \operatorname{Bin}(n, \frac{1}{2})$. The expected value of X is $\mathbb{E}[X] = np = \frac{n}{2}$.

To prove that $P\left(X < \frac{n}{3}\right)$ decays exponentially with n, we apply Chernoff's bound for the lower tail of the binomial distribution.

Chernoff's bound states that for any $0 < \delta < 1$,

$$P\left(X \leq (1-\delta)\mathbb{E}[X]\right) \leq \exp\left(-\frac{\delta^2 \mathbb{E}[X]}{2}\right).$$

In our case, by setting $(1-\delta)\mathbb{E}[X] = \frac{n}{3}$ and substituting $\mathbb{E}[X] = \frac{n}{2}$ we get,

$$1 - \delta = \frac{\frac{n}{3}}{\frac{n}{2}} = \frac{2}{3} \quad \Rightarrow \quad \delta = \frac{1}{3}.$$

And if we apply Chernoff's bound with $\delta = \frac{1}{3}$ we get,

$$P\left(X < \frac{n}{3}\right) = P\left(X \leq \frac{n}{3}\right) \leq \exp\left(-\frac{\left(\frac{1}{3}\right)^2 \cdot \frac{n}{2}}{2}\right) = \exp\left(-\frac{n}{36}\right).$$

Thus, the probability of observing fewer than $\frac{n}{3}$ heads decays exponentially with n, and an upper bound for this probability is $e^{-\frac{n}{36}}$.

Question 7

Problem: Suppose $Q_d = [-1, 1]^d$ is a d-dimensional cube centered at the origin with side length 2. Also, suppose B_d is a ball centered at the origin with radius 1,000,000. Prove that if a point is chosen randomly in Q_d, the probability that it also lies in B_d as d increases tends to zero exponentially, and find an upper bound for this decay. In other words, for large dimensions d, the volume of B_d occupies a very small fraction of Q_d.

Solution:

Let $X = (X_1, X_2, \ldots, X_d)$ be a random point uniformly chosen in $Q_d = [-1, 1]^d$. The probability that X lies in B_d is given by the ratio of the volumes:

$$P = \frac{V(B_d \cap Q_d)}{V(Q_d)}.$$

Given that $r = 1{,}000{,}000$ is a fixed radius, we analyze how P behaves as d increases.

The volume of the d-dimensional hypercube $Q_d = [-1, 1]^d$ is:

$$V(Q_d) = 2^d.$$

And the volume of a d-dimensional Euclidean ball B_d with radius $r = 1{,}000{,}000$ is:

$$V(B_d) = \frac{\pi^{d/2} r^d}{\Gamma\left(\frac{d}{2} + 1\right)},$$

where Γ is the Gamma function.

So the probability P that a randomly chosen point in Q_d also lies in B_d is:

$$P = \frac{V(B_d \cap Q_d)}{V(Q_d)}.$$

Since B_d has a very large radius ($r = 1{,}000{,}000$), for sufficiently large d, the hypercube Q_d will not be entirely contained within B_d. Specifically, the maximum Euclidean distance from the origin to any corner of Q_d is:

$$\|x\|_2 = \sqrt{d}.$$

Thus, for $d > r^2 = (1{,}000{,}000)^2 = 10^{12}$, the hypercube Q_d extends beyond the ball B_d, making $V(B_d \cap Q_d) < V(B_d)$. (Although, a bit before that we can have such statement too, but for the sake of making sure that the ball B_d is totally contained in the hypercube Q_d we take $d > r^2 = (1{,}000{,}000)^2 = 10^{12}$.)

To analyze the behavior of P as d increases, let's consider the ratio below:

$$P = \frac{V(B_d)}{V(Q_d)} = \frac{\pi^{d/2} r^d}{\Gamma\left(\frac{d}{2} + 1\right) \cdot 2^d}.$$

If we use Stirling's approximation for the Gamma function we have:

$$\Gamma\left(\frac{d}{2} + 1\right) \approx \sqrt{2\pi \frac{d}{2}} \left(\frac{d}{2e}\right)^{\frac{d}{2}} = \sqrt{\pi d} \left(\frac{d}{2e}\right)^{\frac{d}{2}}.$$

Substituting this into the expression for P we get:

$$P \approx \frac{\pi^{d/2} r^d}{2^d \cdot \sqrt{\pi d}\left(\frac{d}{2e}\right)^{\frac{d}{2}}} = \frac{(\pi r^2 e)^{\frac{d}{2}}}{2^d \cdot d^{\frac{d}{2}} \cdot \sqrt{\pi d}}.$$

So:

$$P \approx \frac{\left(\frac{\pi r^2 e}{d}\right)^{\frac{d}{2}}}{\sqrt{\pi d}}.$$

Taking the natural logarithm to analyze the decay rate:

$$\ln P \approx \frac{d}{2} \ln\left(\frac{\pi r^2 e}{d}\right) - \frac{1}{2}\ln(\pi d).$$

For large d, the dominant term is:

$$\ln P \approx \frac{d}{2} \ln\left(\frac{\pi r^2 e}{d}\right) = -\frac{d}{2} \ln\left(\frac{d}{\pi r^2 e}\right).$$

Exponentiating both sides:

$$P \approx \exp\left(-\frac{d}{2} \ln\left(\frac{d}{\pi r^2 e}\right)\right) = \left(\frac{\pi r^2 e}{d}\right)^{\frac{d}{2}}.$$

This expression shows that P decays faster than exponentially with d, specifically as:

$$P \approx \left(\frac{C}{d}\right)^{\frac{d}{2}},$$

where $C = \pi r^2 e$ is a constant.

To express the decay in an exponential form, observe that:

$$P \approx \left(\frac{C}{d}\right)^{\frac{d}{2}} = \exp\left(\frac{d}{2} \ln\left(\frac{C}{d}\right)\right) = \exp\left(-\frac{d}{2} \ln\left(\frac{d}{C}\right)\right).$$

Since $\ln\left(\frac{d}{C}\right)$ grows with d, we have:

$$P \leq \exp\left(-cd \ln d\right),$$

with some constant $c > 0$ and sufficiently large d.

Now this shows that as the dimension d becomes very large (e.g., $d = 10^{99999}$), the volume of B_d occupies an exponentially small fraction of the volume of Q_d.

References

1. MIT OpenCourseWare. ”High Dimensional Statistics,” Chapter 1. Sub-Gaussian Random Variables, Available at: link

2. The Art of Problem Solving. ”Cauchy-Schwarz Inequality,” Available at: link

3. Modulo a factor of 2. ”A short proof of Hoeffding’s lemma by Marc Romaní,” Available at: link

4. CS229 Supplemental Lecture note. ”Hoeffding’s inequality by John Duchi,” Stanford, Available at: link

5. Poisson distribution. ”Poisson distribution,” WikiPedia, Available at: link

6. Poisson limit theorem. ”Poisson limit theorem,” WikiPedia, Available at: link

7. Gamma function. ”Explanation for gamma function in formula for n-ball volume,” MathOverflow, Available at: link

8. Stirling’s approximation. ””Quantitative Stirling” for the volume of ℓ_p balls?,” Mathematics Stack Exchange, Available at: link

9. Euclidean ball. ”Volume of an n-ball,” WikiPedia, Available at: link