Chapter 2 Concentration of sums of independent random variables

Desh Raj

June 2, 2018

Exercise 2.1.4.

Solution Given, $g \sim \mathcal{N}(0,1)$. On differentiating, we get $g'(t) = -tg(t)$.

$$
\begin{aligned}
\mathbb{E}\, X^2 \mathbb{1}_{X>t} &= \int_t^\infty x^2 g(x)dx \\
&= \int_t^\infty x(xg(x))dx \\
&= x\int_t^\infty xg(x)dx + \int_t^\infty g(x)dx && \text{(Integration by parts)} \\
&= -xg(x)\Big]_t^\infty + \mathbb{P}(X>t) \\
&= tg(t) + \mathbb{P}(X>t) && \text{(Taking limit of the first term)}
\end{aligned}
$$

This is the desired result.

Exercise 2.2.3. Bounding the hyperbolic cosine

Solution We have,

$$
\begin{aligned}
\cosh(x) &= \frac{e^x + e^{-x}}{2} \\
&= \frac{1}{2}\left(\sum_{i=0}^\infty \frac{x^i}{i!} + \sum_{i=0}^\infty \frac{(-x)^i}{i!}\right) && \text{(Using Taylor series expansion)} \\
&= \sum_{i=0}^\infty \frac{x^{2i}}{(2i)!} \geq \sum_{i=0}^\infty \frac{x^{2i}}{2^i i!} \\
&\text{(Consider the denominator of the } i^{th} \text{ term. Each product term in } (2i)! \text{ is strictly greater than that in } 2^i i!.) \\
&= \exp\left(\frac{x^2}{2}\right). && \text{(Again using Taylor series expansion)}
\end{aligned}
$$

Exercise 2.2.7.

Solution Given, $X_1, \ldots, X_N$ are independent random variables such that $X_i \in [m_i, M_i]$. Let $S_N = X_1 + \ldots + X_N$. Then, for $t > 0$, we have

$$
\begin{aligned}
\mathbb{P}\{S_N - \mathbb{E}\, S_N \geq t\} &= \mathbb{P}\{e^{s(S_N - \mathbb{E}\, S_N)} \geq e^{st}\} && \text{(For some } s > 0\text{)} \\
&\leq \frac{\mathbb{E}\, e^{s(S_N - \mathbb{E}\, S_N)}}{e^{st}} && \text{(Using Markov's inequality)} \\
&= e^{-st} \prod_{i=1}^{N} \mathbb{E}\left[e^{s(X_i - \mathbb{E}\, X_i)}\right] \\
&\leq e^{-st} \prod_{i=1}^{N} \exp\left(\frac{s^2 (M_i - m_i)^2}{8}\right) && \text{(Using Hoeffding's lemma, since } \mathbb{E}(X_i - \mathbb{E}\, X_i) = 0\text{)} \\
&= \exp\left(-st + \frac{s^2}{8} \sum_{i=1}^{N} (M_i - m_i)^2\right).
\end{aligned}
$$

Let $g(s) = \exp\left(-st + \frac{s^2}{8}\sum_{i=1}^{N}(M_i - m_i)^2\right)$. $g(s)$ achieves its minimum at $s = \frac{4t}{\sum_{i=1}^{N}(M_i - m_i)^2}$. Substituting this value in the inequality, we get

$$
\mathbb{P}\{S_N - \mathbb{E}\, S_N \geq t\} \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{N}(M_i - m_i)^2}\right).
$$

Exercise 2.2.8.

Solution Let X_i denote the Bernoulli random variable indicating that the answer is wrong in the i^{th} turn. Then $\mathbb{E}\, X_i = \frac{1}{2} - \delta$, and $X_i \in [0, 1]$. We have to bound the probability that the final answer is wrong with probability ϵ, i.e., $\mathcal{P}(\text{majority of decisions are wrong}) \leq \epsilon$. From Hoeffding's inequality, we have

$$
\begin{aligned}
&\mathcal{P}\left(\sum_{i=1}^{N} (X_i - \mathbb{E}\, X_i) \geq t\right) \leq \exp\left(\frac{-2t^2}{\sum_{i=1}^{N}(M_i - m_i)^2}\right) \\
\Rightarrow &\mathcal{P}\left(\sum_{i=1}^{N} \left(X_i - \frac{1}{2} + \delta\right) \geq t\right) \leq \exp\left(\frac{-2t^2}{N}\right) && \text{(Since } M_i = 1 \text{ and } m_i = 0\text{)}
\end{aligned}
$$

To get a wrong answer finally, we require that at least half of all answers must be wrong, i.e.

$$
\begin{aligned}
&\sum_{i=1}^{N} X_i \geq \frac{N}{2} \\
\Rightarrow &\sum_{i=1}^{N} \left(X_i - \frac{1}{2} + \delta\right) \geq \delta N
\end{aligned}
$$

Plugging $t = \delta N$ in earlier inequality, we get

$$
\mathcal{P}\left(\sum_{i=1}^{N} \left(X_i - \frac{1}{2} + \delta\right) \geq \delta N\right) \leq \exp\left(-2N\delta^2\right)
$$

For this probbility to be bounded by ϵ, we require that

$$\begin{aligned}
&\exp(-2N\delta^2) \leq \epsilon \\
\Rightarrow &-2N\delta^2 \leq \log(\epsilon) && \text{(Taking log both sides)} \\
\Rightarrow &N \geq \frac{\log(\epsilon^{-1})}{2\delta^2}.
\end{aligned}$$

Exercise 2.2.9.

Solution Part 1

We want to bound the probability that the sample mean deviates from the true mean by a quantity greater than ϵ. Using Chebyshev's inequality, we have

$$\mathcal{P}\left(\|\frac{1}{N}\sum_{i=1}^{N} X_i - \mu\| \geq \epsilon\right) \leq \frac{\operatorname{var}\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right)}{\epsilon^2} \tag{1}$$

Since the samples are drawn i.i.d, we have

$$\begin{aligned}
&\operatorname{var}(X_1 + \ldots + X_N) = \operatorname{var}(X_1) + \ldots + \operatorname{var}(X_N) \\
\Rightarrow &\operatorname{var}(X_1 + \ldots + X_N) = N\sigma^2 && \text{(Since var}(X_i)=\sigma^2) \\
\Rightarrow &\frac{1}{N^2}\operatorname{var}(X_1 + \ldots + X_N) = \frac{\sigma^2}{N} && \text{(Dividing both sides by } \sigma^2)
\end{aligned}$$

$$\Rightarrow \operatorname{var}\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) = \frac{\sigma^2}{N}$$

Plugging this value in 1, we get

$$\mathcal{P}\left(\|\frac{1}{N}\sum_{i=1}^{N} X_i - \mu\| \geq \epsilon\right) \leq \frac{\sigma^2}{N\epsilon^2}$$

We want that an ϵ-accurate estimate should be computed with probability at least 3/4, i.e., the error probability should be at most 1/4. Using this bound, we get $\frac{\sigma^2}{N\epsilon^2} \leq \frac{1}{4}$, which implies $N = \mathcal{O}(\frac{\sigma^2}{\epsilon^2})$.

Part 2

Using Theorem 2.2.6 with $X_i \in [-\sigma, \sigma]$, we get

$$\begin{aligned}
\mathcal{P}\left(\frac{1}{N}\sum_{i=1}^{N} X_i - \mu \geq \epsilon\right) &= \mathcal{P}\left(\sum_{i=1}^{N}(X_i - \mu) \geq N\epsilon\right) && \text{(Multiplying both sides by } N) \\
&\leq \exp\left(-\frac{2N^2\epsilon^2}{\sum_{i=1}^{N}(2\sigma)^2}\right) && \text{(From Theorem 2.2.6)} \\
&= \exp\left(-\frac{2N^2\epsilon^2}{4N\sigma^2}\right) \\
&= \exp\left(-\frac{N\epsilon^2}{2\sigma^2}\right)
\end{aligned}$$

We want that an ϵ-accurate estimate should be computed with probability at least $1-\delta$, i.e., the error probability should be at most δ. This means that

$$\begin{aligned}
&\exp\left(-\frac{N\epsilon^2}{2\sigma^2}\right) \leq \delta \\
&\Rightarrow -\frac{N\epsilon^2}{2\sigma^2} \leq \log(\delta) && \text{(Taking log both sides)} \\
&\Rightarrow N = \mathcal{O}(\log(\delta^{-1})\frac{\sigma^2}{\epsilon^2}),
\end{aligned}$$

which is the desired result.

Exercise 2.2.10.

Solution Part 1

Let f_{X_i} denote the probability density function. Since X_i is non-negative, it's MGF is given as

$$M_{X_i}(-t) = \int_0^\infty e^{-tx} f_{X_i}(x)dx$$

Since $e^{-tx} \geq 0$, $e^{-tx} f_{X_i} \leq e^{-tx}\max\|f_{X_i}\| = e^{-tx}$ (since $\max\|f_{X_i}\| = 1$). Using this in above equation, we get

$$M_{X_i}(-t) \leq \int_0^\infty e^{-tx}dx = \frac{1}{t}$$

Part 2

We have

$$\begin{aligned}
\mathcal{P}\left(\sum_{i=1}^N X_i \leq \epsilon N\right) &= \mathcal{P}\left(\sum_{i=1}^N -X_i \geq -\epsilon N\right) \\
&= \mathcal{P}\left(\sum_{i=1}^N -\frac{X_i}{\epsilon} \geq -N\right) \\
&= \mathcal{P}\left(\exp(\sum_{i=1}^N -\frac{X_i}{\epsilon} \geq \exp(-N)\right) && \text{(Since exp is monotonously increasing)} \\
&\leq e^N\,\mathbb{E}\exp\left(\sum_{i=1}^N -\frac{X_i}{\epsilon}\right) && \text{(Using Markov inequality)} \\
&= e^n \prod_{i=1}^N \mathbb{E}\exp\left(-\frac{X_i}{\epsilon}\right) \\
&\leq e^N \prod_{i=1}^N \epsilon && \text{(Using result in part 1)} \\
&= (e\epsilon)^N.
\end{aligned}$$

Exercise 2.3.2.

Solution Since $t < \mu$, so the function $f(x) = \left(\frac{t}{\mu}\right)^x$ is monotonically decreasing. So we can write

$$\begin{aligned}
\mathcal{P}(S_N \leq t) &= \mathcal{P}\left(\left(\frac{t}{\mu}\right)^{S_N} \geq \left(\frac{t}{\mu}\right)^t\right) \\
&\leq \frac{\mathbb{E}\left(\frac{t}{\mu}\right)^{S_N}}{\left(\frac{t}{\mu}\right)^t} && \text{(By Markov's inequality)}
\end{aligned}$$

We have

$$\begin{aligned}
\mathbb{E}\,\alpha^{S_N} = \mathbb{E}\,\alpha^{\sum_{i=1}^N X_i} = \mathbb{E}\prod_{i=1}^N \alpha^{X_i} \\
&= \prod_{i=1}^N \left[p_i\alpha^1 + (1-p_i)\alpha^0\right] \\
&\qquad \text{(Since } X_i \text{ can only take values 1 and 0 with probability } p_i \text{ and } 1-p_i\text{)} \\
&= \prod_{i=1}^N (1 + (\alpha - 1)p_i) \\
&\leq \prod_{i=1}^N \exp((\alpha-1)p_i) && \text{(Using } 1 + x \leq e^x\text{)} \\
&= \exp(\sum_{i=1}^N ((\alpha-1)p_i)) \\
&= \exp[(\alpha-1)\mu]
\end{aligned}$$

Using this inequality in above, we get

$$\begin{aligned}
\mathcal{P}(S_N \leq t) &\leq \exp(t-\mu)\left(\frac{\mu}{t}\right)^t \\
&= e^{-\mu}\left(\frac{e\mu}{t}\right)^t,
\end{aligned}$$

which is the desired result.

Exercise 2.3.3.

Solution Since $X \sim \text{Pois}(\lambda)$, by Poisson limit theorem, we can say that X is the sum of N Bernoulli random variables $X_1, \ldots, X_N$, for some large N, where $\mathbb{E}\,X_i = \lambda$. As such, we can use the Chernoff concentration bound for the sum of Bernoulli random variables to bound X, hence giving the desired result.

Exercise 2.3.5.

Solution First, using $t = (1+\delta)\mu$ in Chernoff bounds for upper tails, we get

$$\mathcal{P}\{X - \mu \geq \delta\mu\} \leq e^{-\mu} \left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu}. \tag{2}$$

Now, using $t = (1-\delta)\mu$ in Chernoff bounds for lower tails (proved in previous problem), we get

$$\mathcal{P}\{X - \mu \leq -\delta\mu\} \leq e^{-\mu} \left(\frac{e}{1-\delta}\right)^{(1-\delta)\mu}. \tag{3}$$

Adding 2 and 3, we get

$$\mathcal{P}\{\|X - \mu\| \geq \delta\mu\} \leq e^{-\mu} \left[\left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu} + \left(\frac{e}{1-\delta}\right)^{(1-\delta)\mu}\right]$$

We can bound the terms inside the bracket on the RHS as follows.

$$\begin{aligned}
\left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu} &= \exp\left[(1+\delta)\mu \log\left(\frac{e}{1+\delta}\right)\right] \\
&= \exp\left[\mu(1+\delta)(1-\log(1+\delta))\right] \\
&\leq \exp\left[\mu(1+\delta)(1-\frac{\delta}{1+\delta/2})\right] && (\text{Since } \log(1+x) \geq \tfrac{x}{1+x/2}) \\
&= \exp\left[\mu(1+\delta)\frac{1-\frac{\delta}{2}}{1+\frac{\delta}{2}}\right] \\
&= \exp\left[\mu\left(1-\frac{\delta^2}{2+\delta}\right)\right] \\
&\leq \exp\left[\mu\left(1-\frac{\delta^2}{3}\right)\right] && (\text{Since } 2+\delta \leq 3)
\end{aligned}$$

and

$$\begin{aligned}
\left(\frac{e}{1-\delta}\right)^{(1-\delta)\mu} &= \exp\left[(1-\delta)\mu \log\left(\frac{e}{1-\delta}\right)\right] \\
&= \exp\left[\mu(1-\delta)(1-\log(1-\delta))\right] \\
&\leq \exp\left(\mu(1-\delta)(1-\frac{-\delta}{\sqrt{1-\delta}})\right] && (\text{Since } \log(1-x) \geq \tfrac{-x}{\sqrt{1-x}}) \\
&= \exp\left[\mu(1-\delta+\delta\sqrt{1-\delta})\right] \\
&\leq \exp\left[\mu\left(1-\frac{\delta^2}{3}\right)\right] && (\text{Since } \sqrt{1-\delta} \leq 1-\tfrac{\delta}{2})
\end{aligned}$$

Using these bounds, we get

$$\begin{aligned}
\mathcal{P}\{\|X - \mu\| \geq \delta\mu\} &\leq 2\exp\left[-\mu\left(1-\frac{\delta^2}{3}\right)\right] \\
&\leq 2e^{-c\mu\delta^2}.
\end{aligned}$$

Exercise 2.3.6.

Solution Since $X \sim \text{Pois}(\lambda)$, by Poisson limit theorem, we can say that X is the sum of N Bernoulli random variables $X_1, \ldots, X_N$, for some large N, where $\mathbb{E} X_i = \lambda$. Using the result in Exercise 2.3.5, we get

$$\mathcal{P}\{\|X - \lambda\| \geq \delta\lambda\} \leq 2\exp(-c\lambda\delta^2). \tag{4}$$

Now substituting $t = \delta\lambda$ in 4, we get the desired result.

Exercise 2.3.8.

Solution We will first show that the sum of independent Poisson distributions is a Poisson distribution. Let X and Y be Poisson distributions with means λ_1 and λ_2, respectively. Their MGFs are given as $M_X(s) = e^{-\lambda_1(1-s)}$ and $M_Y(s) = e^{-\lambda_2(1-s)}$. Then, we have

$$M_{X+Y}(s) = M_X(s)M_Y(s) = e^{-(\lambda_1+\lambda_2)(1-s)}.$$

Hence, $X + Y \sim \text{Pois}(\lambda_1 + \lambda_2)$.

It is given that $X \sim \text{Pois}(\lambda)$. Therefore, X can be considered as a sum of λ Poisson distributions with mean 1, using the above result. Now consider the distribution $Z = \frac{X-\lambda}{\sqrt{\lambda}}$. We can directly apply the Central Limit Theorem to Z to obtain the desired result, since $\mathbb{E} X = \lambda$ and $\text{var}(X) = \lambda$.

Exercise 2.4.2.

Solution Consider a fixed vertex i. Using Markov inequality, we can bound the probability that its degree d_i is larger than $\mathcal{O}(\log n)$, as

$$\mathcal{P}\{d_i \geq \mathcal{O}(\log n)\} \leq \frac{\mathbb{E}\, d_i}{\mathcal{O}(\log n)} = c$$

We can now use the union bound to bound the probability that such a vertex exists as follows

$$\mathcal{P}\{\exists i : d_i \geq \mathcal{O}(\log n)\} = \sum_{i=1}^{n} \mathcal{P}\{d_i \geq \mathcal{O}(\log n)\} \leq cn$$

We can choose a constant c such that this probability is very low, and therefore, the probability of its counterpart is very high.

Exercise 2.4.3.

Solution Similar to the previous exercise, we can show that if the expected degree is $d = \mathcal{O}(1)$, then with a high probability, all vertices of G have degrees $\mathcal{O}(1)$. Furthermore, we know from asymptotic notations that $\mathcal{O}(1) \leq \mathcal{O}\left(\frac{\log n}{\log \log n}\right)$. Hence, the result is satisfied.

Exercise 2.4.4.

Solution Since degrees are not independent, let us define d_i' as the out-degree of the i^{th} vertex of the corresponding directed graph. Since we know that degrees are always larger than out-degrees, if we can show that with high probability, there exists an i such that $d_i' = 10d$, then we are done. For this, we can assume that in the limit of a large n, the degree distribution (which is actually a Binomial) approximates a Poisson distribution. With this assumption, we can now use the Poisson tail bound for an exact value (2.9).

Note: I have not been able to derive the exact result for this problem.

Exercise 2.4.5.

Solution Since the degrees approximately follow the Poisson distribution for large n, and given that the expected degrees are $\mathcal{O}(1)$ we have

$$\mathcal{P}\{d_i = k\} = e^{-1}\frac{1^k}{k!} \geq \frac{e^{-1}}{k^k}, \tag{5}$$

where the last inequality is obtained by Stirling's approximation.

For the given k, we have

$$\begin{aligned}
\log k^k &= k \log k \\
&= \frac{\log n}{\log\log n}(\log\log n - \log\log\log n) \\
&\approx \log n && \text{(Since } \log\log n >> \log\log\log n\text{)} \\
\therefore k^k &\approx n
\end{aligned}$$

Using this approximation in 5 and taking union bound over all vertices, we see that the required probability is at least e^{-1}, which is sufficiently large.

Exercise 2.5.1.

Solution We have

$$\begin{aligned}
\mathbb{E}\, X^p &= \int_{-\infty}^{\infty} x^p \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{p-1}(x e^{-x^2/2}) dx \\
&= \frac{1}{\sqrt{2\pi}} \left[-x^{p-1} e^{-x^2/2}|_{-\infty}^{\infty} + (p-1)\int_{-\infty}^{\infty} x^{p-2} x e^{-x^2/2} dx \right] && \text{(Integration by parts)} \\
&= \frac{p-1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{p-2} x e^{-x^2/2} dx && \text{(Since the first term is 0)} \\
&= (p-1)\,\mathbb{E}\, X^{p-2} \\
&= (p-1)(p-3)\ldots(3)(1) && \text{(Since } \mathbb{E}\, X^0 = 1\text{)} \\
&= \frac{p!}{2.4\ldots p} \\
&= \frac{p!}{2^{p/2}(\frac{p}{2})!} \\
\therefore \|X\|_p = (\mathbb{E}\, X^p)^{1/p} &= \frac{1}{\sqrt{2}} \left(\frac{p!}{(p/2)!} \right)^{1/p},
\end{aligned}$$

and the equivalence to the desired result can be seen by using the definition $\Gamma(z) = (z-1)!$.

We can further write

$$
\begin{aligned}
\|X\|_p = (\mathbb{E}\, X^p)^{1/p} &= \frac{1}{\sqrt{2}} \left(\frac{p!}{(p/2)!} \right)^{1/p} \\
&= \frac{1}{\sqrt{2}} \left(p(p-1)\dots(\frac{p}{2}+1) \right)^{1/p} \\
&\leq \frac{1}{\sqrt{2}} \left(p^{p/2} \right)^{1/p} && \text{(Since each term in product is less than } p) \\
&= \mathcal{O}(\sqrt{p})
\end{aligned}
$$

as $p \to \infty$.

For the moment-generating function, we have

$$
\begin{aligned}
\mathbb{E} \exp(\lambda X) &= \int_{-\infty}^{\infty} e^{\lambda x} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2+\lambda x} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2-2\lambda x+\lambda^2)+\frac{1}{2}\lambda^2} dx && \text{(Adding and subtracting } \tfrac{1}{2}\lambda^2) \\
&= \frac{e^{\lambda^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-\lambda)^2}{2}} dx \\
&= e^{\lambda^2/2} \mathcal{N}(\lambda, 1) \\
&= e^{\lambda^2/2},
\end{aligned}
$$

which is the desired result.

Exercise 2.5.4.

Solution

$$
\begin{aligned}
\mathbb{E} \exp(\lambda^2 X^2) &= \mathbb{E} \exp(\lambda X)^2 \\
&= \mathbb{E}\, exp \sum_{i=1}^{\lambda X} \lambda X \\
&= \mathbb{E} \prod_{i=1}^{\lambda X} \exp(\lambda X) \\
&\leq \mathbb{E} \prod_{i=1}^{\lambda X} \exp(\lambda^2) && \text{(Using Property 5 with } K_5 = 1) \\
&= \lambda^3\, \mathbb{E}\, X \leq \exp(K^2 \lambda^2) && \text{(Using Property 4)}
\end{aligned}
$$

This inequality can only hold for all $\lambda \in \mathcal{R}$ if $\mathbb{E}\, X = 0$.

Exercise 2.5.5.

Solution **Part 1**

We can calculate the MGF of X^2 for $X \sim \mathcal{N}(0,1)$ as follows.

$$\begin{aligned}
\mathbb{E}\,\lambda X^2 &= \int_{-\infty}^{\infty} \lambda x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx \\
&= \frac{2\lambda}{\sqrt{2\pi}} \int_0^{\infty} x(xe^{-\frac{x^2}{2}})dx \\
&= \frac{2\lambda}{\sqrt{2\pi}} \left[-e^{-\frac{x^2}{2}}|_0^{\infty} + \int_0^{\infty} e^{-\frac{x^2}{2}}\,dx\right] && \text{(Integration by parts)} \\
&= \lambda\left(\frac{\sqrt{2}}{\pi} + 1\right)
\end{aligned}$$

This is a monotonously increasing function in λ, and therefore it is only finite in a bounded neighborhood of 0.

Part 2

We have

$$\begin{aligned}
&\mathbb{E}\exp(\lambda^2 X^2) \le \exp(K\lambda^2) \\
&\Rightarrow \mathbb{E}\left(1 + \frac{\lambda^2 X^2}{1!} + \frac{(\lambda^2 X^2)^2}{2!} + \dots\right) \le 1 + \frac{K\lambda^2}{1!} + \frac{(K\lambda^2)^2}{2!} + \dots && \text{(Using Taylor expansion)} \\
&\therefore \forall i \in 0, 1, \dots, \quad \lambda^{2i}\,\mathbb{E}\,X^{2i} \le \lambda^{2i}(\sqrt{K})^{2i} \\
& && \text{(Comparing individual terms since all terms are positive)} \\
&\therefore \mathbb{E}\,X^{2i} \le K^i \\
&\Rightarrow \mathbb{E}((X^2)^i) \le K
\end{aligned}$$

$\therefore \|X\|_\infty \le K$

Exercise 2.5.7.

Solution To prove that a function is a norm, it needs to satisfy the following 3 properties:

1. $p(\mathbf{u} + \mathbf{v}) \le p(\mathbf{u}) + p(\mathbf{v})$
2. $p(a\mathbf{v}) = |a|p(\mathbf{v})$
3. If $p(\mathbf{v}) = 0$, then $\mathbf{v} = 0$

We will now show each of these properties for the sub-gaussian norm $\|\cdot\|_{\psi_2}$.

1. Let $f(x) = e^{x^2}$. We have

$$
\begin{aligned}
f\left(\frac{|X+Y|}{a+b}\right) &\le f\left(\frac{|X|+|Y|}{a+b}\right) \\
&\le \frac{a}{a+b} f\left(\frac{|X|}{a}\right) + \frac{b}{a+b} f\left(\frac{|Y|}{b}\right) && \text{(Jensen's inequality for convex functions)} \\
\mathbb{E}\, f\left(\frac{|X+Y|}{a+b}\right) &\le \frac{a}{a+b}\,\mathbb{E}\, f\left(\frac{|X|}{a}\right) + \frac{b}{a+b}\,\mathbb{E}\, f\left(\frac{|Y|}{b}\right) && \text{(Taking } \mathbb{E} \text{ on both sides)} \\
\mathbb{E}\, f\left(\frac{|X+Y|}{\|X\|_{\psi_2} + \|Y\|_{\psi_2}}\right) &\le \frac{a}{a+b} 2 + \frac{b}{a+b} 2 = 2 && \text{(Taking } a = \|X\|_{\psi_2} \text{ and } b = \|Y\|_{\psi_2}\text{)}
\end{aligned}
$$

So $\|X\|_{\psi_2} + \|Y\|_{\psi_2}$ is in the set $t > 0 : \mathbb{E}\exp(X^2/t^2) \le 2$, and the proof is complete.

2. We have

$$
\begin{aligned}
\|aX\|_{\psi_2} &= \inf t > 0 : \mathbb{E}\exp(-(aX)^2/t^2) \le 2 \\
&= \inf au > 0 : \mathbb{E}\exp(-X^2/t^2) \le 2 && \text{(Taking } t = au\text{)} \\
&= a\|X\|_{\psi_2}
\end{aligned}
$$

3. We have

$$
\begin{aligned}
& \|X\|_{\psi_2} = 0 \\
& \Rightarrow \inf t > 0 : \mathbb{E}\exp(-X^2/t^2) \le 2 = 0 \\
& \Rightarrow \mathbb{E}\exp(-X^2/t^2) \le 2 \\
& \Rightarrow \exp(-\,\mathbb{E}\, X^2/t^2) \le 2 && \text{(By Jensen's inequality)} \\
& \Rightarrow \mathbb{E}\, X^2 \le -t^2 \log 2 \\
& \Rightarrow \mathbb{E}\, X^2 \le \lim_{t \to 0} -t^2 \log 2 && \text{(Taking } \lim_{t\to 0} \text{ both sides)} \\
& \Rightarrow \mathbb{E}\, X^2 \le 0 \\
\therefore\ & X = 0
\end{aligned}
$$

Exercise 2.5.9.

Solution (i) $X \sim \text{Pois}(\lambda)$. We have

$$
\begin{aligned}
\mathcal{P}\{X \ge t\} &> \mathcal{P}\{X = t\} \\
&= e^{-\lambda}\frac{\lambda^t}{t!}
\end{aligned}
$$

Exercise 2.5.10.

Solution

Exercise 2.5.11.

Solution