6th Practice Class
Concentration inequalities II

6.1 (a) Let us suppose that we have an unfair die. That is, it has a shape of a general parallelepiped but the sum of the opposite sides are still 7. Let X_i be the outcome of the ith roll (the rolls are independent), and let $S_n = X_1 + \cdots + X_n$. Using Bernstein's, give estimates for

$$\mathbb{P}\left(\left|S_n - n\frac{7}{2}\right| > n\right).$$

(b) Give an estimate by using Hoeffding's inequality!

Solution

(a) First, recall Bernstein's inequality. If $\mathbb{P}(|X - \mathbb{E}(X)| \leq K) = 1$ and $0 \leq \lambda \leq \frac{\mathbb{D}(S_n)}{K}$ then

$$\mathbb{P}(|S_n - \mathbb{E}(S_n)| \geq \lambda \mathbb{D}(S_n)) \leq 2\exp\left(\frac{-\lambda^2}{2\left(1+\lambda\frac{K}{\mathbb{D}(S_n)}\right)^2}\right).$$

Thus, we need to calculate the expectation and the standard deviation of S_n. Moreover, we need to determine K and choose and appropriate λ.

Notice by the symmetry

$$\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = 6) = p,$$
$$\mathbb{P}(X_i = 2) = \mathbb{P}(X_i = 5) = q,$$
$$\mathbb{P}(X_i = 3) = \mathbb{P}(X_i = 4) = r.$$

Moreover, we have

$$2p + 2q + 2r = 1$$
$$p + q + r = \frac{1}{2}.$$

The expectation

$$\mathbb{E}(X_i) = 1p + 2q + 3r + 4r + 5q + 6p = 7\underbrace{(p+q+r)}_{=1/2} = \frac{7}{2}.$$

The second moment

$$\begin{aligned}\mathbb{E}\left(X_i^2\right) &= 1^2p + 2^2q + 3^2r + 4^2r + 5^2q + 6^2p\\ &= 37p + 29q + 25r\\ &= 37p + 29q + 25(1/2 - p - q)\\ &= 12p + 4q + 25/2.\end{aligned}$$

Therefore, the variance

$$\mathbb{D}^2(X_i) = \mathbb{E}\left(X_i^2\right) - \mathbb{E}(X_i)^2 = 12p + 4p + \frac{25}{2} - \frac{49}{4} = 12p + 4q + \frac{1}{4}.$$

It implies the following bounds on $\mathbb{D}^2(X_i)$ (since $p, q \in (0, 1/2)$)

$$\frac{1}{4} < \mathbb{D}^2(X_i) < \frac{25}{4}$$

Therefore, for the expectation and variance of S_n we have the estimates

$$\mathbb{E}(S_n) = \frac{7}{2}n \qquad \text{and} \qquad \frac{1}{4}n < \mathbb{D}^2(S_n) < \frac{25}{4}n.$$

In particular, for the standard deviation of S_n we have the estimates

$$\frac{1}{2}\sqrt{n} < \mathbb{D}(S_n) < \frac{5}{2}\sqrt{n}.$$

Now it is clear that by choosing $K = \frac{5}{2}$ we have

$$\mathbb{P}(|X_i - \mathbb{E}(X_i)| \leq K) = \mathbb{P}(|X_i - 7/2| \leq 5/2) = 1$$

since $X_i \in \{1, 2, 3, 4, 5, 6\}$.
Now let $\lambda = \frac{1}{5}\sqrt{n} \geq 0$. Then

$$\frac{\mathbb{D}(S_n)}{K} > \frac{1}{2}\sqrt{n} \cdot \frac{2}{5} = \frac{\sqrt{n}}{5} = \lambda \geq 0.$$

Moreover, notice that

$$\lambda \mathbb{D}(S_n) < \frac{\sqrt{n}}{5} \cdot \frac{5}{2}\sqrt{n} = \frac{1}{2}n.$$

Thus,

$$\left|S_n - n\frac{7}{2}\right| > n \implies \left|S_n - n\frac{7}{2}\right| > \frac{1}{2}n \implies \left|S_n - n\frac{7}{2}\right| > \lambda \mathbb{D}(S_n).$$

This proves the bound

$$\mathbb{P}\left(\left|S_n - n\frac{7}{2}\right| > n\right) \leq \mathbb{P}\left(\left|S_n - n\frac{7}{2}\right| > \lambda \mathbb{D}(S_n)\right) = \mathbb{P}(|S_n - \mathbb{E}(S_n)| > \lambda \mathbb{D}(S_n)). \tag{1}$$

Finally, we can apply Bernstein's inequality to the right hand side of the inequality in (1).

$$\mathbb{P}(|S_n - \mathbb{E}(S_n)| > \lambda \mathbb{D}(S_n)) \leq 2\exp\left(\frac{-\lambda^2}{2\left(1 + \lambda\frac{K}{\mathbb{D}(S_n)}\right)^2}\right) = 2\exp\left(\frac{-\left(\frac{1}{5}\sqrt{n}\right)^2}{2\left(1 + \frac{1}{5}\sqrt{n}\frac{\frac{5}{2}}{\mathbb{D}(S_n)}\right)^2}\right)$$

We do not have an explicit formula for $\mathbb{D}(S_n)$ but replacing it in the upper bound with something bigger gives a slightly bigger estimate. Thus, using that $\mathbb{D}(S_n) < 5/2\sqrt{n}$

$$2\exp\left(\frac{-\left(\frac{1}{5}\sqrt{n}\right)^2}{2\left(1 + \frac{1}{5}\sqrt{n}\frac{\frac{5}{2}}{\mathbb{D}(S_n)}\right)^2}\right) \leq 2\exp\left(\frac{-\left(\frac{1}{5}\sqrt{n}\right)^2}{2\left(1 + \frac{1}{5}\sqrt{n}\frac{\frac{5}{2}}{\frac{5}{2}\sqrt{n}}\right)^2}\right) = 2\exp\left(-\frac{1}{72}n\right).$$

We conclude

$$\mathbb{P}\left(\left|S_n - n\frac{7}{2}\right| > n\right) \leq 2\exp\left(-\frac{1}{72}n\right).$$

(b) First, recall Hoeffding's inequality. If $\mathbb{P}(a_i \leq X_i \leq b_i) = 1$, then

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

Clearly, $\mathbb{P}(1 \leq X_i \leq 6) = 1$. Thus

$$\mathbb{P}\left(\left|S_n - n\frac{7}{2}\right| > n\right) = \mathbb{P}(|S_n - \mathbb{E}(S_n)| > n) \leq 2\exp\left(\frac{-2n^2}{\sum_{i=1}^n (6-1)^2}\right) = 2\exp\left(-\frac{2}{25}n\right).$$

6.3 Let X be a random variable. The function $\widehat{I}(\lambda) = \log(\mathbb{E}\left(e^{\lambda X}\right))$ is called logarithmic moment generating function.

(a) Find the logarithmic moment generating function of $aX + b$.
(b) Let X_1, X_2 be independent random variables, find the LMGF of $X_1 + X_2$.

Solution

(a) The LMGF of $aX + b$ is

$$\log\left(\mathbb{E}\left(e^{\lambda(aX+b)}\right)\right) = \log\left(e^{\lambda b}\mathbb{E}\left(e^{\lambda aX}\right)\right) = \log(e^{\lambda b}) + \log\left(\mathbb{E}\left(e^{\lambda aX}\right)\right) = \lambda b + \widehat{I}(\lambda a).$$

(b) The LMGF of $X_1 + X_2$ is

$$\begin{aligned}\log\left(\mathbb{E}\left(e^{\lambda(X_1+X_2)}\right)\right) = \log\left(\mathbb{E}\left(e^{\lambda X_1}e^{\lambda X_2}\right)\right) &\overset{\text{indep.}}{=} \log\left(\mathbb{E}\left(e^{\lambda X_1}\right)\mathbb{E}\left(e^{\lambda X_2}\right)\right) \\ &= \log\left(\mathbb{E}\left(e^{\lambda X_1}\right)\right) + \log\left(\mathbb{E}\left(e^{\lambda X_2}\right)\right) = \widehat{I}_1(\lambda) + \widehat{I}_2(\lambda).\end{aligned}$$

6.5 Let us define the *rate function* of the random variable X by the Legendre transform of the $I(x) = \sup_{\lambda\in\mathbb{R}}\{\lambda x - \widehat{I}(\lambda)\}$, where $\widehat{I}(\lambda) = \log(\mathbb{E}\left(e^{\lambda X}\right))$ is the logarithmic moment generating function of X. Suppose that $\widehat{I}(\lambda)$ exists for every $\lambda \in \mathbb{R}$

(a) Using the strict convexity of $\widehat{I}$ show that $I(x) = x \cdot (\widehat{I}')^{-1}(x) - \widehat{I}((\widehat{I}')^{-1}(x))$.
(b) Show that I is C^∞, $I(\mathbb{E}(X)) = 0$ and $I(x) > 0$ for every $x \neq \mathbb{E}(X)$.

Solution

(a) The formula was proven during the lecture.

(b) Since $\widehat{I}(\lambda)$ exists for every $\lambda \in \mathbb{R}$, it follows $\widehat{I} \in C^\infty$. As a consequence $\widehat{I}' \in C^\infty$. Since $\widehat{I}$ is strictly convex, $\widehat{I}'$ is strictly monotone increasing. From the inverse function theorem it follows that $(\widehat{I}')^{-1} \in C^\infty$. Thus, the right hand side of

$$I(x) = x \cdot (\widehat{I}')^{-1}(x) - \widehat{I}((\widehat{I}')^{-1}(x))$$

is in C^∞.
Notice

$$\widehat{I}(0) = \log(1) = 0.$$

Moreover,

$$\widehat{I}'(0) = \frac{\mathbb{E}\left(X \cdot e^{0\cdot X}\right)}{\mathbb{E}(e^{0\cdot X})} = \mathbb{E}(X).$$

Thus,

$$(\widehat{I}'(\mathbb{E}(X)))^{-1} = 0.$$

From the formula it follows

$$\begin{aligned}I(\mathbb{E}(X)) &= \mathbb{E}(X)(\widehat{I}(\mathbb{E}(X)))^{-1} - \widehat{I}((\widehat{I}(\mathbb{E}(X)))^{-1}) \\ &= \mathbb{E}(X) \cdot 0 - \widehat{I}(0) \\ &= 0.\end{aligned}$$