Lecture 11

Instructor: Quan Zhou

For more details about the materials covered in this note, see Chapters 2.2 and 2.3 of Vershynin [1].

11.1 Some concentration inequalities

In this section, we assume $X_1, X_2, \ldots$ are *i.i.d.* random variables with expectation μ. Let $\bar{X}_n$ denote the average of the first n random variables.

Lemma 11.1. *Let Z be a random variable with mean 0 and variance σ^2.*

(i) Hoeffding's lemma: If $Z \in [a,b]$, then $E[e^{\lambda Z}] \leq e^{\lambda^2 (b-a)^2/8}$ for $\lambda > 0$.

(ii) If $|Z| \leq K$, then for $0 < \lambda < 3/K$,

$$E\left[e^{\lambda Z}\right] \leq \exp\left(\frac{\lambda^2\sigma^2/2}{1-\lambda K/3}\right).$$

(iii) If $|Z| \leq K$, $E[e^{\lambda Z}] \leq \exp\left\{\frac{\sigma^2}{K^2}(e^{\lambda K} - 1 - \lambda K)\right\}$ for $\lambda > 0$.

Proof. To prove part (i), note that $e^{\lambda z}$ is a convex function in z and thus for any $z \in [a,b]$,

$$e^{\lambda z} \leq \frac{b-z}{b-a} e^{\lambda a} + \frac{z-a}{b-a} e^{\lambda b}.$$

Taking expectation on both sides and using $E[Z] = 0$, we get

$$E[e^{\lambda Z}] \leq \frac{b}{b-a} e^{\lambda a} - \frac{a}{b-a} e^{\lambda b} = e^{-\lambda\theta(b-a)} \left(1 - \theta + \theta e^{\lambda(b-a)}\right),$$

where $\theta = -a/(b-a)$. Let $u = \lambda(b-a)$ and consider

$$\psi(u) = -\theta u + \log(1 - \theta + \theta e^u), \quad u \geq 0.$$

Direct calculation yields that $\psi(0) = \psi'(0) = 0$ and $\psi''(u) \leq 1/4$ for every u. Hence, by Taylor theorem (and the mean-value form for the remainder),

$$\psi(u) = \frac{1}{2}\psi''(v)$$

for some $v \in [0, u]$. Therefore, $\psi(u) \le u^2/8$ for $u \ge 0$ and

$$E[e^{\lambda Z}] \le e^{\psi(u)} \le e^{\lambda^2 (b-a)^2/8}.$$

To prove part (ii), first we verify that for $|z| < 3$,

$$e^z \le 1 + z + \frac{z^2/2}{1 - |z|/3}.$$

This can be shown by Taylor expansion:

$$2\frac{e^z - 1 - z}{z^2} = \sum_{k=2}^{\infty} \frac{z^{k-2}}{k!/2} \le \sum_{k=2}^{\infty} \frac{|z|^{k-2}}{k!/2} \le \sum_{k=2}^{\infty} \frac{|z|^{k-2}}{3^{k-2}} = \frac{1}{1 - |z|/3}.$$

Then, taking expectation on both sides and using $e^x \ge 1 + x$, we find that

$$E[e^{\lambda Z}] \le E\left[\exp\left(\frac{\lambda^2 Z^2/2}{1 - \lambda|Z|/3}\right)\right] \le E\left[\exp\left(\frac{\lambda^2 Z^2/2}{1 - \lambda K/3}\right)\right],$$

provided that $\lambda < 3/K$ (so that $\lambda|Z| \le 3$.)

To prove part (iii), apply Taylor expansion to obtain that, for $z \in (-K, K)$,

$$\begin{aligned} e^{\lambda z} &= 1 + \lambda z + \sum_{n=2}^{\infty} \frac{\lambda^n z^n}{n!} \le 1 + \lambda z + \sum_{n=2}^{\infty} \frac{\lambda^n z^2 |z|^{n-2}}{n!} \\ &\le 1 + \lambda z + \sum_{n=2}^{\infty} \frac{\lambda^n z^2 K^{n-2}}{n!} = 1 + \lambda z + \frac{z^2}{K^2} \sum_{n=2}^{\infty} \frac{\lambda^n K^n}{n!} \\ &= 1 + \lambda z + \frac{z^2}{K^2}(e^{\lambda K} - 1 - \lambda K). \end{aligned}$$

Taking expectation on both sides, we get

$$E\left(e^{\lambda Z}\right) \le 1 + \frac{\sigma^2}{K^2}\left(e^{\lambda K} - 1 - \lambda K\right) \le \exp\left\{\frac{\sigma^2}{K^2}(e^{\lambda K} - 1 - \lambda K)\right\},$$

which completes the proof. □

Theorem 11.1 (Hoeffding's inequality)**.** *If $X_i \in [m, M]$ (i.e. bounded),*

$$\mathsf{P}(\bar{X}_n - \mu \ge t) \le \exp\left(-\frac{2nt^2}{(M-m)^2}\right), \qquad \forall\, t \ge 0.$$

Proof. We use the Chernoff bound. For any $\lambda > 0$,

$$\begin{aligned} \mathsf{P}(\bar{X}_n - \mu \geq t) &= \mathsf{P}\left\{ e^{\lambda(\bar{X}_n - \mu)} \geq e^{\lambda t} \right\} \\ &\leq e^{-\lambda t} E\left[e^{\lambda(\bar{X}_n - \mu)} \right] \\ &= e^{-\lambda t} \prod_{i=1}^{n} E\left[e^{\lambda(X_i - \mu)/n} \right], \end{aligned}$$

where the last equality follows from the independence between $X_1, \dots, X_n$. Now we apply Hoeffding's lemma to obtain

$$\mathsf{P}(\bar{X}_n - \mu \geq t) \leq e^{-\lambda t} \prod_{i=1}^{n} \exp\left(\frac{\lambda^2 (M-m)^2}{8n^2} \right) = \exp\left\{ \frac{\lambda^2 (M-m)^2}{8n} - \lambda t \right\}.$$

This inequality holds for any $\lambda > 0$. We should choose the best one (i.e. the one that minimizes the upper bound), which is given by

$$\lambda^* = \frac{4nt}{(M-m)^2},$$

for every $t \geq 0$. The result then follows. □

Theorem 11.2 (Bernstein's inequality)**.** *Suppose* $E[X_i] = \mu = 0$*,* $\mathrm{Var}(X_i) = \sigma^2$ *and* $|X_i| \leq K$*. Then,*

$$\mathsf{P}(\bar{X}_n \geq t) \leq \exp\left(-\frac{nt^2/2}{\sigma^2 + Kt/3} \right), \qquad \forall\, t \geq 0.$$

Proof. Applying Chernoff bound with Lemma 11.1 (ii), we obtain

$$\mathsf{P}(\bar{X}_n \geq t) \leq \exp\left\{ \frac{\lambda^2 \sigma^2 / 2n}{1 - \lambda K/(3n)} - \lambda t \right\}.$$

for $0 < \lambda < 3n/K$. Let's further assume $1 - \lambda K/3n \geq c$. Then,

$$\frac{\lambda^2 \sigma^2 / 2n}{1 - \lambda K/3} - \lambda t \leq \frac{\lambda^2 \sigma^2}{2nc} - \lambda t = f(\lambda).$$

$f(\lambda)$ is maximized at $\lambda^* = tcn/\sigma^2$, which gives $f(\lambda^*) = -\dfrac{t^2 cn}{2\sigma^2}$. Thus,

$$\mathsf{P}(\bar{X}_n \geq t) \leq \exp\left\{ -\frac{t^2 cn}{2\sigma^2} \right\}. \tag{1}$$

However, the choice of c is not arbitrary and must satisfy

$$1 - \frac{\lambda^* K}{3n} \geq c.$$

Some algebra yields that this is equivalent to $c \leq \{Kt/(3\sigma^2) + 1\}^{-1} = c^*$. Plugging c^* black into (1), we obtain the asserted inequality. $\square$

Theorem 11.3 (Bennett's inequality)**.** *Under the same assumption as Bernstein's inequality, we also have*

$$\mathsf{P}(\bar{X}_n \geq t) \leq \exp\left(-\frac{n\sigma^2}{K^2}\, h\left(\frac{Kt}{\sigma^2}\right)\right), \qquad \forall\, t \geq 0,$$

where $h(x) = (1+x)\log(1+x) - x$.

Proof. By Chernoff bound and Lemma 11.1 (iii),

$$\mathsf{P}(\bar{X}_n \geq t) \leq \exp\left\{\frac{n\sigma^2}{K^2}\left(e^{\lambda K/n} - 1 - \frac{\lambda K}{n}\right) - \lambda t\right\}.$$

By differentiating the exponent and setting it to zero, we get

$$\lambda^* = \frac{n}{K}\log\left(1 + \frac{tK}{\sigma^2}\right).$$

Some routine algebra then yields the asserted inequality. $\square$

Example 11.1. Consider a triangular array $\{Y_{n,k} : 1 \leq k \leq n,\ n \geq 1\}$ where for each n, $Y_{n,1}, \dots, Y_{n,n}$ are i.i.d. with $\mathsf{P}(Y_{n,k} = n) = 1/n$ and $\mathsf{P}(Y_{n,k} = 0) = 1 - 1/n$. Define $\bar{Y}_n = (Y_{n,1} + \cdots + Y_{n,n})/n$ for each n (i.e., the average of the n-th row). It is straightforward to compute that

$$E[Y_{n,1}] = 1, \quad \mathrm{Var}(Y_{n,1}) = n - 1, \quad E[\bar{Y}_n] = 1, \quad \mathrm{Var}(\bar{Y}_n) = \frac{n-1}{n}.$$

Hence, $\mathrm{Var}(\bar{Y}_n)$ is asymptotically equal to 1. Observe that

$$\mathsf{P}\left(\frac{|\bar{Y}_n - E[\bar{Y}_n]|}{\sqrt{\mathrm{Var}(\bar{Y}_n)}} \geq n - 1\right) \geq \mathsf{P}(\bar{Y}_n = n) = n^{-n} = e^{-n\log n},$$

which is a slower rate than e^{-cn^2} for any $c > 0$.

To apply the three concentration inequalities, we first center the random variables by letting $X_{n,k} = Y_{n,k} - 1$. Then, $|X_{n,k}| \leq K = n - 1$ and $\mathrm{Var}(X_{n,k}) = \sigma^2 = n - 1$. So we obtain that

$$\begin{aligned}
\text{Hoeffding:} \quad & \mathsf{P}(\bar{X}_n \geq t) \leq e^{-2t^2/n}, \\
\text{Bernstein:} \quad & \mathsf{P}(\bar{X}_n \geq t) \leq \exp\left(-\frac{n}{n-1} \frac{t^2/2}{1 + t/3} \right), \\
\text{Bennett:} \quad & \mathsf{P}(\bar{X}_n \geq t) \leq \exp\left(-\frac{n}{n-1} [(1+t)\log(1+t) - t] \right).
\end{aligned}$$

Now choose $t = n$, and one can check that only Bennett's inequality gives the right order.

References

[1] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.