Probability and Stochastic Processes
Lecture 18: Concentration Bounds and Limit Theorems

Prof. Washim Uddin Mondal
Department of Electrical Engineering
Indian Institute of Technology Kanpur

Markov's Inequality

Proposition 18.1

If X is a positive valued random variable defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$, then

$$\mathbf{P}(X \geq a) \leq \frac{\mathbf{E}[X]}{a}, \ \forall a \in (0, \infty)$$

Proof of Proposition 18.1 (Continuous Case)

$$\begin{aligned} \mathbf{E}[X] = \int_0^\infty x f_X(x) \mathrm{d}x &= \int_0^a x f_X(x) \mathrm{d}x + \int_a^\infty x f_X(x) \mathrm{d}x \\ &\geq \int_a^\infty x f_X(x) \mathrm{d}x \geq a \int_a^\infty f_X(x) \mathrm{d}x = a\mathbf{P}(X \geq a) \end{aligned}$$

Markov's Inequality

Example

Let X be a positive random variable defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ such that $\mathbf{E}[X] < \infty$. Show that $\mathbf{P}(X < \infty) = 1$.

Solution

- Define $A_n = \{X \geq n\}$. Observe that $\{X = \infty\}$ is identical to $\cap_{n \in \mathbb{N}} A_n$. Moreover, $A_{n+1} \subset A_n, \forall n \in \mathbb{N}$. Invoking the continuity of probability measures, we get

$$\mathbf{P}\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \lim_{n \to \infty} \mathbf{P}(A_n) \overset{(a)}{\leq} \lim_{n \to \infty} \frac{\mathbf{E}[X]}{n} = 0$$

- Relation (a) follows from Markov's inequality.

Chebyshev's Inequality

Proposition 18.2

If X is a random variable defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ such that $\mathbf{E}[X] = \mu < \infty$ and $\mathrm{Var}(X) = \sigma^2 < \infty$, then the following holds $\forall \epsilon > 0$.

$$\mathbf{P}\left(|X - \mu| \geq \epsilon\right) \leq \frac{\sigma^2}{\epsilon^2}$$

Proof of Proposition 18.2

$$\mathbf{P}\left(|X - \mu| \geq \epsilon\right) = \mathbf{P}\left((X - \mu)^2 \geq \epsilon^2\right) \overset{(a)}{\leq} \frac{\mathbf{E}\left[(X - \mu)^2\right]}{\epsilon^2} = \frac{\sigma^2}{\epsilon^2}$$

- Step (a) is due to Markov's Inequality.

Chebyshev's Inequality

Example

If X is a random variable defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ such that $\mathbf{E}[X] = \mu < \infty$, and $\mathrm{Var}(X) = 0$, then show that $\mathbf{P}(X = \mu) = 1$.

Solution

- Let $A_n = \{|X - \mu| < \frac{1}{n}\}$. Note that $\{X = \mu\}$ is the same event as $\cap_{n \in \mathbb{N}} A_n$. Moreover, $A_{n+1} \subset A_n, \forall n \in \mathbb{N}$. Invoking the continuity of probability measures, we get

$$\mathbf{P}\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \lim_{n \to \infty} \mathbf{P}(A_n) = 1 - \lim_{n \to \infty} \mathbf{P}\left(|X - \mu| \geq \frac{1}{n}\right) \overset{(a)}{\geq} 1$$

- Step (a) uses Chebyshev's inequality.

Chernoff's Inequality

Proposition 18.3

Let X be a random variable defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$. If $a \in \mathbb{R}$, then

$$\mathbf{P}\left(X \geq a\right) \leq \inf_{t>0} e^{-at} M_X(t) \quad \text{and} \quad \mathbf{P}(X \leq a) \leq \inf_{t<0} e^{-at} M_X(t)$$

Proof of Proposition 18.3

- If $t > 0$, then we can obtain the following using Markov's inequality.

$$\mathbf{P}(X \geq a) = \mathbf{P}(tX \geq ta) = \mathbf{P}(e^{tX} \geq e^{ta}) \leq \frac{\mathbf{E}(e^{tX})}{e^{ta}} = e^{-ta} M_X(t)$$

- The term $\mathbf{P}(X \leq a)$ can be bounded similarly.

Chernoff's Inequality

Example

Using Chernoff's inequality, derive an upper bound of the probability $\mathbf{P}(X \geq \mu + \epsilon)$, where $X \sim \mathcal{N}(\mu, \sigma^2)$, and $\epsilon > 0$.

Solution

- Note that $\inf_{t>0} e^{-at} M_X(t) = \exp\left(\inf_{t>0}\{-at + \log M_X(t)\}\right)$.
- Recall that $\log M_X(t) = \mu t + \frac{1}{2}\sigma^2 t^2$. Therefore,

$$\inf_{t>0}\{-(\mu+\epsilon)t + \log M_X(t)\} = \inf_{t>0}\left\{-\epsilon t + \frac{1}{2}\sigma^2 t^2\right\} = -\frac{\epsilon^2}{2\sigma^2}$$

- Hence, $\mathbf{P}(X \geq \mu + \epsilon) \leq \exp\left(-\epsilon^2/2\sigma^2\right)$.

Chernoff's Inequality

Example

Using Chernoff's inequality, derive an upper bound of the probability $\mathbf{P}(X \geq a)$, where $a > 0$ and $X \sim \text{Poisson}(\lambda)$ for some $\lambda > 0$.

Solution

- Recall that $\log M_X(t) = \lambda(e^t - 1)$. Moreover,

$$\inf_{t>0} \left\{-at + \lambda(e^t - 1)\right\} = -a \log\left(\frac{a}{\lambda}\right) + a - \lambda$$

- Thus, $\mathbf{P}(X \geq a) \leq (\lambda/a)^a \exp(a - \lambda)$.

Weak Law of Large Numbers

Proposition 18.4

Let $(X_1, X_2 \cdots)$ be a sequence of independent and identically distributed (IID) random variables, defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$, with a common mean $\mu < \infty$ and a common variance σ^2. If $\bar{X}_n = (X_1 + \cdots + X_n)/n$, then $\bar{X}_n \xrightarrow{P} \mu$.

Proof of Proposition 18.4 (Assuming $\sigma^2 < \infty$)

- Recall that $\mathbf{E}[\bar{X}_n] = \mu$ and $\mathrm{Var}(\bar{X}_n) = \sigma^2/n$. For any $\epsilon > 0$,

$$\mathbf{P}\left(|\bar{X}_n - \mu| > \epsilon\right) \leq \frac{\mathrm{Var}(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

- Hence, $\lim_{n\to\infty} \mathbf{P}(|\bar{X}_n - \mu| > \epsilon) = 0$.

Strong Law of Large Numbers

Proposition 18.5

Let $(X_1, X_2, \cdots)$ be a sequence of independent and identically distributed (IID) random variables, defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$, with a common mean $\mu < \infty$. If $\bar{X}_n = (X_1 + \cdots + X_n)/n$, then $\bar{X}_n \xrightarrow{a.s.} \mu$.

Proof of Proposition 18.5 (Special Case)

- Define $Y_i = X_i - \mu$, $\forall i \in \mathbb{N}$. We have $\mathbf{E}[Y_i] = 0$, $\forall i \in \mathbb{N}$.
- Define $\bar{Y}_n = S_n/n$, where $S_n = Y_1 + \cdots + Y_n$, $\forall n \in \mathbb{N}$.
- We have to prove that $\bar{Y}_n \xrightarrow{a.s.} 0$. Let $\mathbf{E}[Y_i^2] = \sigma^2 < \infty$ and $\mathbf{E}[Y_i^4] = \tau < \infty$, $\forall i \in \mathbb{N}$.

Strong Law of Large Numbers

Proof of Proposition 18.5 (Continued)

- Note that $\mathbf{E}[Y_i^2 Y_j Y_k] = 0$ and $\mathbf{E}[Y_i Y_j Y_k Y_l] = 0$ if $i \neq j \neq k \neq l$.
- Also, $\mathbf{E}[Y_i^3 Y_j] = 0$ and $\mathbf{E}[Y_i^2 Y_j^2] = \sigma^4$ if $i \neq j$. Hence,

$$\mathbf{E}[S_n^4] = \sum_{1 \leq i \leq n} \mathbf{E}[Y_i^4] + \binom{4}{2} \sum_{1 \leq i < j \leq n} \mathbf{E}[Y_i^2 Y_j^2] = n\tau + 3n(n-1)\sigma^4$$

- We have the following result.

$$\mathbf{E}\left[\sum_{n=1}^{\infty} \bar{Y}_n^4\right] = \mathbf{E}\left[\sum_{n=1}^{\infty} \frac{S_n^4}{n^4}\right] = \tau \sum_{n=1}^{\infty} \frac{1}{n^3} + 3\sigma^4 \left(\sum_{n=1}^{\infty} \frac{1}{n^2} - \sum_{n=1}^{\infty} \frac{1}{n^3}\right)$$

Strong Law of Large Numbers

Proof of Proposition 18.5 (Continued)

- The series $\sum_{n=1}^{\infty} n^{-p}$ converges for any $p > 1$.
- Hence, $\mathbf{E}\left[\sum_{n=1}^{\infty} \bar{Y}_n^4\right] < \infty$, which implies that $\mathbf{P}\left(\sum_{n=1}^{\infty} \bar{Y}_n^4 < \infty\right) = 1$.
- Moreover, $\sum_{n=1}^{\infty} \left[\bar{Y}_n(\omega)\right]^4 < \infty$ implies that $\lim_{n\to\infty} \bar{Y}_n(\omega) = 0$.
- Therefore, $\mathbf{P}(\lim_{n\to\infty} \bar{Y}_n = 0) \geq \mathbf{P}\left(\sum_{n=1}^{\infty} \bar{Y}_n^4 < \infty\right) = 1$.

Note

- Both laws hold even with infinite variances. However, the proofs are more involved.

Central Limit Theorem (CLT)

Theorem 18.6

Let $(X_1, X_2, \cdots)$ be a sequence of independent and identically distributed (IID) random variables, defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$, with a common mean $\mu < \infty$ and a common variance $\sigma^2 < \infty$. If $\bar{X}_n = (X_1 + \cdots + X_n)/n$, then

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} Z \sim \mathcal{N}(0, 1)$$

In other words, the following holds $\forall a \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbf{P}\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq a\right) = \Phi(a)$$

Central Limit Theorem (CLT)

Lemma 18.7

Let $(Z_1, Z_2, \cdots)$ be a sequence of random variables, with $(M_{Z_n})_{n\in\mathbb{N}}$ as their associated sequence of MGFs. If Z is a random variable with M_Z as its associated MGF such that $\lim_{n\to\infty} M_{Z_n}(t) = M_Z(t)$, $\forall t \in \mathbb{R}$, then $Z_n \xrightarrow{D} Z$.

Proof of Theorem 18.6

- Define $Y_n = (X_n - \mu)/\sigma$, $\forall n \in \mathbb{N}$. Clearly, Y_n's are IID random variables.
- Let $M_Y(\cdot)$ be the common MGF of Y_n's.
- Note that $Z_n \triangleq (\bar{X}_n - \mu)/(\sigma/\sqrt{n}) = (Y_1 + \cdots + Y_n)/\sqrt{n}$.

Central Limit Theorem (CLT)

Proof of Theorem 18.6 (Continued)

$$M_{Z_n}(t) = \mathbf{E}\left[\prod_{i=1}^{n} \exp\left(\frac{t}{\sqrt{n}} Y_i\right)\right] = \left[M_Y\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

- We have the following limiting value.

$$\lim_{n\to\infty} \log M_{Z_n}(t) = \lim_{n\to\infty} n \log M_Y\left(\frac{t}{\sqrt{n}}\right) = t^2 \lim_{y\to 0} \frac{\log M_Y(y)}{y^2}$$

- The last equality uses the substitution that $y = t/\sqrt{n}$.
- Since $\log M_Y(0) = 0$, we can apply L'Hospital's rule.

Central Limit Theorem (CLT)

Proof of Theorem 18.6 (Continued)

$$\lim_{n\to\infty} \log M_{Z_n}(t) = \frac{t^2}{2} \lim_{y\to 0} \frac{M_Y^{(1)}(y)}{M_Y(y)y} = \frac{t^2}{2} \lim_{y\to 0} \frac{M_Y^{(1)}(y)}{y}$$

- Since $M_Y^{(1)}(0) = \mathbf{E}[Y_n] = 0, \forall n \in \mathbb{N}$, we can use L'Hospital's rule again.

$$\lim_{n\to\infty} \log M_{Z_n}(t) = \frac{t^2}{2} \lim_{y\to 0} M_Y^{(2)}(y) \overset{(a)}{=} \frac{t^2}{2}$$

  where (a) uses $M_Y^{(2)}(0) = \mathbf{E}[Y_n^2] = \mathbf{E}[(X_n - \mu)^2]/\sigma^2 = 1, \forall n \in \mathbb{N}$.

- Therefore, $\lim_{n\to\infty} M_{Z_n}(t) = e^{\frac{t^2}{2}}, \forall t \in \mathbb{R}$.

Central Limit Theorem (CLT)

Example

An instructor has 50 exams that will be graded in sequence. The times required to grade the exams are independent, with a common distribution that has mean 20 minutes and standard deviation 4 minutes. Approximate the probability that the instructor will grade at least 25 of the exams in the first 450 minutes of work.

Solution

- Let X_i denote the time needed to grade the ith exam.
- Thus, $\mu = \mathbf{E}[X_i] = 20$ and $\sigma^2 = \mathrm{Var}(X_i) = 16, \forall i \in \{1, \cdots, 50\}$.
- Define $\bar{X}_n = (X_1 + \cdots + X_n)/n$, and $Z_n = (\bar{X}_n - \mu)/(\sigma/\sqrt{n}), \forall n \in \mathbb{N}$.

Central Limit Theorem (CLT)

Solution (Continued)

- The required probability is as follows.

$$\begin{aligned} &\mathbf{P}\left(X_1 + \cdots + X_{25} \le 450\right) \\ &= \mathbf{P}\left(\bar{X}_{25} \le 18\right) = \mathbf{P}\left(Z_{25} \le \frac{18-20}{4/\sqrt{25}}\right) \\ &\qquad = \mathbf{P}\left(Z_{25} \le -2.5\right) \overset{(a)}{\approx} \Phi(-2.5) = 1 - \Phi(2.5) = 0.006 \end{aligned}$$

- Step (a) invokes the CLT and approximates Z_{25} as a standard normal random variable.