Probability and Stochastic Processes Course Notes

Chia-Hao Chang, chchangkh@utexas.edu

Fall 2018

1 Taylor Series and the Limit of Exponential Function

Theorem 1 (Taylor Series). *Let f be an $n+1$ times differentiable function. Then the* **Truncated Taylor Series** *of f is*

$$f(z+h) = \sum_{k=1}^{m} \frac{h^k}{k!} f^{(k)}(z) + \frac{h^{n+1}}{(n+1)!} f^{(n+1)}(w)$$

for some $w \in [z, z+h]$. Further, if f is infinitely differentiable, we can write it as infinite series sum

$$f(z+h) = \sum_{n=1}^{\infty} \frac{h^k}{k!} f^{(k)}(z)$$

where $f^{(k)}(\cdot) = \frac{d^k f}{dx^k}$ is the k^{th} derivative of f.

Proposition 1 (Exponential Limit of a sequence). *Let (x_n) be a sequence on $\mathbb{R}$ convergening to x. i.e. $\lim_{n\to\infty} x_n = x$. Then the sequence (y_n)*

$$y_n = \left(1 + \frac{x_n}{n}\right)^n$$

converges to e^x. Namely,

$$\lim_{n\to\infty} \left(1 + \frac{x_n}{n}\right)^n = e^{\lim\limits_{n\to\infty} x_n} = e^x$$

Proposition 2 (Limit of a Function on a Sequence[1]). *Let f be a continuous function on $\mathbb{R}$ and (a_n) be a sequence on $\mathbb{R}$. If $\lim_{n\to\infty} x_n = x$ and if f is continuous, then*

$$\lim_{n\to\infty} f(x_n) = f(\lim_{n\to\infty} x_n) = f(x)$$

2 Law of Large Numbers

In this section, we introduce three versions of **Law of Large Numbers** (LLN).

Theorem 2 (Weak Law of large Numbers)**.** *Suppose* $\{X_1, X_2, \cdots, X_n, \cdots\}$ *is a sequence of* **i.i.d**. *random variables with* $\mathbb{E}[X_i] = \mu$, $\forall i$. *Let* S_n *be the (random) partial sum of* X_i. *i.e.*

$$S_n = \sum_{i=1}^{n} X_i$$

$\frac{S_n}{n}$ **converges to** μ **in probability** *i.e*

$$\frac{S_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n} \to \mu \quad \text{in probability}$$

Theorem 3 (Strong Law of Large Numbers)**.** *Suppose* $\{X_1, X_2, \cdots, X_n, \cdots\}$ *is a sequence of* **i.i.d**. *random variables with* $\mathbb{E}[X_i] = \mu$, $\forall i$. *Let* S_n *be the (random) partial sum of* X_i. *Then* $\frac{S_n}{n}$ **converges to** μ **almost surely** *i.e.*

$$\frac{S_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n} \to \mu \quad \text{almost surely}$$

Theorem 4 (LLN in Mean Square Sense)**.** *Suppose* $\{X_1, X_2, \cdots, X_n, \cdots\}$ *is a sequence of* **i.i.d**. *random variables with* $\mathbb{E}[X_i] = \mu$, *and* $\mathbb{E}[X_i^2] < c\ \forall i$ *for some* $c \in \mathbb{R}$. *Let* S_n *be the (random) partial sum of* X_i. *Then* $\frac{S_n}{n}$ **converges to** μ **in mean square**. *i.e.*

$$\frac{S_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n} \to \mu \quad \text{mean square}$$

To illustrate the spirit behind LLN, we can find that the **average** of the sum of i.i.d random variables "Collapse" exactly to a **fixed** point (or a constant random variable i.e. μ) indefinitely. That is, when the number of samples $\{X_1, \cdots, X_n\}$ are large enough, we are certain that their average contracts to a point in $\mathbb{R}$, namely their expected value $\mathbb{E}[X_i] = \mu$. Note when we say "average" what we mean is that we do a random experiment and get the ω from Ω. With such ω, we are able to define an (infinite) sequence on $\mathbb{R}$. i.e.

$$X_1(\omega), X_2(\omega), X_3(\omega), \cdots, X_n(\omega), \cdots$$

The Law of Large Numbers states that, as we take more and more samples $X_i(\omega)$ and average over them, we will find that $\frac{S_n}{n}$ converges to μ **indefinitely no natter what sample** ω **we use**.

3 Central Limit Theorem

Theorem 5 (Central Limit Theorem)**.** *Let* $\{X_1, X_2, \cdots, X_n, \cdots\}$ *be a sequence of i.i.d. random variables with* $\mathbb{E}[X_i] = \mu$ *and* $Var[X_i] = \sigma^2$, $\forall i$. $S_n = \sum_{i=1}^{n} X_i$ *is the partial sum of* X_i. *Then, we have,*

$$\frac{S_n - n\mu}{\sqrt{n}} \longrightarrow Z \quad \text{in distribution}$$

where $Z \sim \mathcal{N}(0, \sigma^2)$. *Or, equivalently*

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \longrightarrow \mathcal{N}(0, 1) \quad \text{in distribution}$$

One might view the CLT as a weaker version of the LLN in the sense that LLN demands the average of the sequence converges exactly to a point (in fact, μ). On the other hand, CLT requires only the average (in some sense) to be distributed like a zero-mean Gaussian random variable.

The problem why CLT does not collapse to the mean of the random variables lies in $\sqrt{n}$ in the denominator. The partial sum in CLT is divided by $\sqrt{n}$ while the other is divided by n in LLN. In fact, it can be shown that if the denominator is of n^α, only $\alpha = \frac{1}{2}$ gives a finite, non-zero variance. For $\alpha > 1/2$, the variance goes to 0 which is similar to LLN; for $\alpha < 1/2$, $Var[S_n] \to \infty$ as $n \to \infty$.

4 Chernoff Bound

Definition 1 (Convex Function)**.** Let $\varphi(\cdot) : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ (except $-\infty$). φ is said to be a **convex** function if $\forall a, b \in \mathbb{R}$, $a \le b$ and $\forall \lambda \in [0,1]$,

$$\varphi(\lambda a + (1-\lambda)b) \le \lambda\varphi(a) + (1-\lambda)\varphi(b)$$

That is, the **linear interpolation** is always greater than or equal to the function value.

Definition 2 (Concave function)**.** A function φ is said to be concave if $-\varphi$ is convex.

Theorem 6 (Jensen's Inequality)**.** *Let φ be a convex function and X be a random variable s.t. $\mathbb{E}[X] < \infty$. Then,*

$$\mathbb{E}[\varphi(X)] \ge \varphi(\mathbb{E}[X])$$

Corollary 1 (Jensen's Inequality for Concave functions)**.** *Let g be concave function and X be a random variable with finite mean, then*

$$\mathbb{E}[g(X)] \le g(\mathbb{E}[X])$$

Theorem 7 (Chernoff Bound for a Single Random Variable)**.**

$$\mathbb{P}(\{X > a\}) \le e^{-\max\limits_{\theta \ge 0}(\theta a - \log M_X(\theta))}$$

where $M_X(\theta) = \mathbb{E}[e^{\theta X}]$ is the MGF of X.

Theorem 8 (Chernoff bound for Sum of Random Variables)**.** *Given a sequence of i.i.d. random variables $\{X_1, X_2, \cdots, X_n, \cdots\}$. The tail probability $\mathbb{P}(S_n \ge a)$ of the partial sum $S_n = \sum_{i=1}^n X_i$ is bounded by*

$$\mathbb{P}(S_n \ge na) \le exp(-n \sup_{\theta \ge 0}\{\theta a - \log M_X(\theta)\})$$

i.e. it is bounded by the same quantity as a single random variable to the power of n.

Definition 3 (Rate Function)**.** The rate function $I(a)$ is

$$I(a) = \sup_{\theta \in \mathbb{R}}\{\theta a - \log M_X(\theta)\}$$

Theorem 9 (Cramer's Theorem)**.** *Let $\{X_i : i = 1, , 2, \cdots\}$ be a sequence of i.i.d. random variables and $I(\cdot)$ be the rate function associated with X_i. Then, for any $\varepsilon > 0$, we have*

(a)

$$\mathbb{P}(\frac{1}{n}\sum_{i=1}^{n} X_i > a) \le e^{-nI(a)}, \quad \forall n \ge 1$$

(b)

$$\exists\ N_\varepsilon \in \mathbb{N} \quad \text{s.t.} \quad \forall n \ge N_\varepsilon, \quad \mathbb{P}(\frac{1}{n}\sum_{i=1}^{n} X_i > a) \ge e^{-n(I(a)+\varepsilon)}$$

That is,

$$\lim_{n\to\infty} -\frac{1}{n} \log \left\{ \mathbb{P}(\frac{1}{n}\sum_{i=1}^{n} X_i > a) \right\} = I(a)$$

References

[1] W. Rudin: Principles of Mathematical Analysis, *McGraw-Hill*, 1976.