Section 1: Introduction to probability and convergence

Shruti Sawant and Animesh Kumar Email: `animesh@ee.iitb.ac.in`

1 The need for random variables in engineering systems

Random variables are used to model uncertainty in engineering systems. If all the parameters of an engineering system are not known, they are often modeled by one or many random variables. This is often done since keeping track of many parameters in a system is either impossible or expensive. We illustrate this idea with the following examples.

Example 1.1 (Expected time of arrival in transport)**.** *Consider the expected time of arrival (ETA) problem encountered in map services. A vehicle has to travel from a starting point A to a destination B. It is desirable to know the transit time or ETA. The transit time is a function of the velocity profile as a function of time of own vehicle. This profile is further dependent on (i) one's own driving pattern, (ii) sequence of traffic lights at various intersections through the transit, (iii) traffic at various intersections, (iv) traffic all along the road, and (v) multiple path options between A and B. For an oracle, which will know all these quantities, it is easy to predict the ETA. For a map service system, it is difficult to ascertain these quantities exactly to calculate the transit time. In such setup, a map service will utilize random variables to estimate the transit time for a given vehicle.*

Example 1.2 (RSRP in wireless communication)**.** *A mobile wireless network consists of a base station (BTS) and a receiver named user-equipment (UE). The UE receives a wireless signal from BTS for successful wireless reception/transmission. The reference signal received power (RSRP) for a wireless network measures the strength of signal received by a UE (in dB). The knowledge of RSRP at a UE ensures communication at the fastest possible rate. At any instant and at any point in space, the RSRP varies due to the paths of wireless signal propagation (multipath), geometry of various objects in the path (like trees, buildings, wires), shadowing due to obstacles, and changing position of the UE. It is difficult to exactly calculate RSRP since too many variables are involved. This uncertainty present in wireless channel can be modeled using random variables.*

These examples suggest that in engineering systems, determinism can require too much information. In such situations random variables can be used to model uncertainty or uncertain parameters.

A little history of random variables and probability is mentioned here. Forms of probability and statistics were developed by Arab mathematicians studying cryptology between the eighth and thirteenth century [2]. Exact mathematical descriptions for probability was developed much later. In the sixteenth century, Cardano stated without proof that the accuracies of empirical statistics tend to improve with the number of trials [1]. This conjecture was proved in a limited setup by Bernoulli (dubbed as the Golden theorem) for $\mathrm{Ber}(p)$ random variables in the eighteenth century. Many mathematicians since then–including Poisson, Chebychev, Borel, Markov, Kolmogorov and Khinchine–worked on the development of probability as well as Cardano's conjecture. Some of the universal laws that have emerged are summarized next. We begin by stating some of the contemporary convergence laws available in the literature.

2 Basic convergence laws in statistics

We first begin with the weak law of large numbers (WLLN), which relates empirical averages to distribution average (or expectation).

Fact 2.1 (Weak law of large numbers)**.** *Let $X_1, X_2, X_3, \ldots$ be independent and identically distributed random variables with expectation $\mathbb{E}(x)$.As all random variables are identically distributed, they have common $eE(X)$. Then by the WLLN,*

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}(X)\right| > \varepsilon\right) = 0. \tag{1}$$

In other words, the probability that empirical average of $X_1, X_2, \ldots$ deviates from $\mathbb{E}(X)$ by more than ε is negligible as n becomes large. Here the empirical average is function of random variables $X_1, \ldots, X_n$; therefore, it is random variable and it converges as $n \to \infty$.

The convergence can be understood for finite but large n as follows. For any $\delta > 0$, there exists an n_0 such that

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}(X)\right| > \varepsilon\right) \leq \delta \tag{2}$$

for $n > n_0$. Formally, this manner of convergence is called 'in probability' convergence.

To elucidate how the convergence of empirical average happens around $\mathbb{E}(X)$, the central limit theorem (CLT) was developed.

Fact 2.2 (Central limit theorem)**.** *Let $X_1, X_2, X_3, \ldots$ be independent and identically distributed random variables with expectation $\mathbb{E}(X)$ and variance σ_X^2. Then by the CLT*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{X_i - \mathbb{E}(X)\} \leq t\right) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{t} \exp\left(\frac{xX^2}{2\sigma_X^2}\right) dx. \tag{3}$$

The subtraction of $\mathbb{E}(X)$ from each random variable X_i is called as centering. This results in random variables $\{X_i - \mathbb{E}(X)\}, i = 1, 2, \ldots$ with zero mean. The limit on the left-hand side in the above equation converges to the probability of a Gaussian random variable with zero mean and variance σ_X^2 being smaller than t. This manner of convergence is called as convergence in distribution or weak convergence. In the limit, the left-hand side probability expression converges the corresponding value on the right hand side, for every $t \in \mathbb{R}$. Approximately, it can be argued that the distribution of $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{X_i - \mathbb{E}(X)\}$ is approximately Gaussian with zero mean and a finite variance. As a result, $\frac{1}{n}\sum_{i=1}^{n} X_i$ is expected to lie within $\mathbb{E}(X) \pm \frac{\sigma_X}{\sqrt{n}}$ with a significant probability.

However, the CLT does not give any law for a finite n as it is a limiting distribution. It is desirable to obtain a bound (as opposed to a convergence law) which is useful for every value of n. Such bounds are often called as 'finite sample size' bound. The Chernoff bound provides us with the same.

Fact 2.3 (Chernoff bound)**.** *Let $X_1, X_2, X_3, \ldots$ be independent and identically distributed random variables with a moment generating function $g_X(t)$. Then Chernoff bound states that,*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}(X)\right| > \varepsilon\right) \leq 2\exp(n\mu_X(\varepsilon)) \tag{4}$$

where $\mu_X(\varepsilon) < 0$ in a lot of cases. Both, WLLN and CLT involve limits unlike this bound. As a result, exact calculation of probability for finite n is possible for the Chernoff bound. Chernoff bound is useful only when $\mu_X(\varepsilon)$ is negative as probability is limited by one and when the moment generating function exists.

References

[1] Law of Large Numbers: *https://en.wikipedia.org/wiki/Law_of_large_numbers*

[2] Probability: *https://en.wikipedia.org/wiki/Probability*