Basic Bounds and Limit Theorems

Kevin Shi

May 2025

1 Introduction

In this section, I will go over some of the fundamental limit theorems that are used in probability, attempting to prove them when necessary. The textbook I used for this was *Introduction to Probability*, by Bertsekas and Tsitsiklis.

2 Bounds/Inequalities

2.1 Markov Inequality

The Markov inequality states that:

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}, \forall a > 0 \tag{1}$$

With a key condition being that the sample space of X must be strictly positive numbers. To prove this, consider the new random variable:

$$Y_a = 0 \text{ if } X < a; \, a \text{ otherwise.} \tag{2}$$

Then:

$$\mathbb{E}[Y_a] = aP(Y = a) = aP(X \geq a) \tag{3}$$

$$\tag{4}$$

However, we know that $\mathbb{E}[Y_a] \leq \mathbb{E}[X]$, as it is always that $Y_a \leq X$. Therefore, we derive that:

$$aP(X \geq a) \leq \mathbb{E}[X] \tag{5}$$

$$\rightarrow P(X \geq a) \leq \frac{\mathbb{E}[X]}{a} \tag{6}$$

In general however, this is a fairly loose bound. It is essentially saying that the smaller the expected value of X, the less probable the random variable takes large values.

2.2 Chebyshev's Inequality

A benefit of Chebyshev's inequality is that it doesn't require that X take strictly positive values. This can be seen in its formulation:

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2} \tag{7}$$

This can be proved using the Markov inequality:

$$P((X-\mu)^2 \geq \varepsilon^2) \leq \frac{\mathbb{E}[(X-\mu)^2]}{\varepsilon^2} = \frac{\text{Var}(X)}{\varepsilon^2} \tag{8}$$

Note the that the event $|X-\mu| \geq \varepsilon$ is the same as $(X-\mu)^2 \geq \varepsilon^2$, as both events require that $-\varepsilon \leq X-\mu \leq \varepsilon$. Therefore we can also say that $P((X-\mu)^2 \geq \varepsilon^2) = P(|X-\mu| \geq \varepsilon)$, completing the proof:

$$P(|X-\mu| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2} \tag{9}$$

It essentially states that if X has a small variance, then the probability it takes a value far from the mean is low. Although this gives tighter bounds than the Markov inequality, it is still a loose bound. However, this is sufficient to prove the weak law of large numbers, so if you wish, you can skip the next two bounds.

If we wish to extend similar notions to multivariate RVs, we need the notion of sub-Gaussian random variables.

3 Limit Theorems

3.1 Weak Law of Large Numbers

For a sequence of random variables: $X_1, \ldots, X_n$ sampled iid from a distribution with mean μ and variance σ^2, the weak law of large numbers states that for $\overline{X} = \frac{X_1 + \ldots + X_n}{n}$:

$$\lim_{n\to\infty} P(|\overline{X} - \mu| > \epsilon) \to 0 \tag{10}$$

The proof follows from using Chebyshev's inequality:

$$P(|\overline{X} - \mu| \geq \epsilon) \leq \frac{\text{Var}(\overline{X})}{\varepsilon^2} = \frac{\text{Var}(X_1 + \ldots + X_n)}{n^2\varepsilon^2} \tag{11}$$

$$\to P(|\overline{X} - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \tag{12}$$

Which means that as we take $n \to \infty$ for any $\varepsilon > 0$, the probability bound also goes to 0. In other words, we say that the RV $\overline{X}$ (the sample mean) converges in probability to the true mean.

3.2 Central Limit Theorem (CLT

Consider the scaled random variable Z_n:

$$Z_n = \frac{X_1 + \ldots + X_n - n\mu}{\sigma\sqrt{n}} \tag{13}$$

The central limit theorem states that for a standard normal distribution $\Phi(z)$, we have that the CDF of Z_n converges to $\Phi(z)$.

$$\lim_{n\to\infty} P(Z_n \leq z) = \Phi(z) \tag{14}$$

Which can be proven as using moment generating functions. However, I will save such analysis for a separate document where I will go over moment generating functions and more advanced bounds, as well as attempt a foray into sub-Gaussians.

4 Connection to Frequentist Statistics

The central limit theorem is just a theorem, and has uses in both frequentist and Bayesian inference. In these notes I would prefer to focus on applications of the two limit theorems introduced towards classical inferencce.

Probably the easiest example is that of point estimation, where we can use the weak law of large numbers to estimate certain parameters and statistics. For example, the sample mean is both and unbiased and consistent estimator for the true mean. As another example, the variance is the expected value of the RV: $(X - \mu)^2$. It follows from the weak law of large numbers that we can also use the sample mean of this new RV to form reasonable estimators for the true variance.

It is also important that for large amounts of data, the CDF of the sample mean approaches the normal, which allows for simpler modeling of data.

This is what I can think of off the top of my head, but I will certainly add more as my knowledge expands and I can think of more examples. I imagine that there are more sophisticated methods that also rely on these fundamental limit theorems.