10. Central Theorems

Chris Piech and Mehran Sahami

May 2017

In probability there are a set of general propositions that are not self-evident but can be proved by a chain of reasoning. In this chapter we introduce a set of theorems that are central towards probability theory. This chapter is lighter than some of the others and it is worth making sure you follow the ideas presented in lecture (especially with respect to the central limit theorem).

Inequalities

The following inequalities are useful when you know very little about your distribution, but you would still like to make probabilistic claims. They most often show up in proofs.

Markov's Inequality

If X is a *non-negative* random variable:

$$P(X \geq a) \leq \frac{E[X]}{a} \qquad \text{for all } a > 0$$

Chebyshev's Inequality

If X is a random variable with $E[X] = \mu$ and $\text{Var}(X) = \sigma^2$:

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2} \qquad \text{for all } k > 0$$

Law of Large Numbers

Consider IID random variables $X_1, X_2 \ldots$ such that $E[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$. Then for any $\varepsilon > 0$, the Weak Law of Large Numbers states:

$$P(|X - \mu| \geq \varepsilon) \xrightarrow[n \to \infty]{} 0$$

The Strong Law of Large Numbers states:

$$P\left(lim_{n \to \infty} \left(\frac{X_1 + X_2 + \cdots + X_n}{n} \right) = \mu \right) = 1$$

Central Limit Theorem

The central limit theorem proves that the averages of equally sized samples from *any* distribution themselves be normally distributed. Consider IID random variables $X_1, X_2 \ldots$ such that $E[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$. Let

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Mathematically, the central limit theorem states:

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \qquad \text{as } n \to \infty$$

It is often expressed in terms of the standard normal, Z:

$$Z = \frac{(\sum_{i=1}^{n} X_i) - n\mu}{\sigma\sqrt{n}} \qquad \text{as } n \to \infty$$

At this point you probably think that the central limit theorem is awesome. But it gets even better. With some algebraic manipulation we can show that if the sample mean of IID random variables is normal, it follows that the sum of equally weighted IID random variables must also be normal. Let's call the sum of IID random variables $\bar{Y}$:

$$\begin{aligned}
\bar{Y} &= \sum_{i=1}^{n} X_i = n \cdot \bar{X} && \text{If we define } \bar{Y} \text{ to be the sum of our variables} \\
&\sim N(n\mu, n^2\frac{\sigma^2}{n}) && \text{Since } \bar{X} \text{ is a normal and } n \text{ is a constant.} \\
&\sim N(n\mu, n\sigma^2) && \text{By simplifying.}
\end{aligned}$$

In summary, the central limit theorem explains that both the sample mean of IID variables is normal (regardless of what distribution the IID variables came from) and that the sum of equally weighted IID random variables is normal (again, regardless of the underlying distribution).

Example 1

Say you have a new algorithm and you want to test its running time. You have an idea of the variance of the algorithm's run time: $\sigma^2 = 4\text{sec}^2$ but you want to estimate the mean: $\mu = t\text{sec}$. You can run the algorithm repeatedly (IID trials). How many trials do you have to run so that your estimated runtime = $t \pm 0.5$ with 95% certainty? LetX_i be the run time of the i-th run (for $1 \leq i \leq n$).

$$0.95 = P(-0.5 \leq \frac{\sum_{i=1}^{n} X_i}{n} - t \leq 0.5)$$

By the central limit theorem, the standard normal Z must be equal to:

$$\begin{aligned}
Z &= \frac{(\sum_{i=1}^{n} X_i) - n\mu}{\sigma\sqrt{n}} \\
&= \frac{(\sum_{i=1}^{n} X_i) - nt}{2\sqrt{n}}
\end{aligned}$$

Now we rewrite our probability inequality so that the central term is Z:

$$\begin{aligned}
0.95 = P(-0.5 \leq \frac{\sum_{i=1}^{n} X_i}{n} - t \leq 0.5) &= P(\frac{-0.5\sqrt{n}}{2} \leq \frac{\sum_{i=1}^{n} X_i}{n} - t \leq \frac{0.5\sqrt{n}}{2}) \\
&= P(\frac{-0.5\sqrt{n}}{2} \leq \frac{\sqrt{n}}{2}\frac{\sum_{i=1}^{n} X_i}{n} - \frac{\sqrt{n}}{2}t \leq \frac{0.5\sqrt{n}}{2}) = P(\frac{-0.5\sqrt{n}}{2} \leq \frac{\sum_{i=1}^{n} X_i}{2\sqrt{n}} - \frac{\sqrt{n}}{\sqrt{n}}\frac{\sqrt{n}t}{2} \leq \frac{0.5\sqrt{n}}{2}) \\
&= P(\frac{-0.5\sqrt{n}}{2} \leq \frac{\sum_{i=1}^{n} X_i - nt}{2\sqrt{n}} \leq \frac{0.5\sqrt{n}}{2}) \\
&= P(\frac{-0.5\sqrt{n}}{2} \leq Z \leq \frac{0.5\sqrt{n}}{2})
\end{aligned}$$

And now we can find the value of n that makes this equation hold.

$$\begin{aligned}
0.95 &= \phi(\frac{\sqrt{n}}{4}) - \phi(-\frac{\sqrt{n}}{4}) = \phi(\frac{\sqrt{n}}{4}) - (1 - \phi(\frac{\sqrt{n}}{4})) \\
&= 2\phi(\frac{\sqrt{n}}{4}) - 1 \\
0.975 &= \phi(\frac{\sqrt{n}}{4}) \\
\phi^{-1}(0.975) &= \frac{\sqrt{n}}{4} \\
1.96 &= \frac{\sqrt{n}}{4} \\
n &= 61.4
\end{aligned}$$

Thus it takes 62 runs. If you are interested in how this extends to cases where the variance is unknown, look into variations of the students' t-test.

Example 2

You will roll a 6 sided dice 10 times. Let X be the total value of all 10 dice = $X_1 + X_2 + \cdots + X_10$. You win the game if $X \leq 25$ or $X \geq 45$. Use the central limit theorem to calculate the probability that you win.

Recall that $E[X_i] = 3.5$ and $\text{Var}(X_i) = \frac{35}{12}$.

$$\begin{aligned}
P(X \leq 25 \text{ or } X \geq 45) &= 1 - P(25.5 \leq X \leq 44.5) \\
&= 1 - P(\frac{25.5 - 10(3.5)}{\sqrt{35/12}\sqrt{10}} \leq \frac{X - 10(3.5)}{\sqrt{35/12}\sqrt{10}} \leq \frac{44.5 - 10(3.5)}{\sqrt{35/12}\sqrt{10}} \\
&\approx 1 - (2\phi(1.76) - 1) \approx 2(1 - 0.9608) = 0.0784
\end{aligned}$$