9. Central Limit Theorem

170B Probability Theory, Puck Rombach

Last updated: November 2, 2016

Bertsekas & Tsitsiklis: Section 5.4. Assumed knowledge: weak law of large numbers, moment generating functions.

9.1 Poisson Approximation to the Binomial

When we have a binomial distribution with \(n\) large compared to \(np\), then we can approximate this with a Poisson distribution. In the limit \(n \to \infty\) (if \(np\) remains constant) this approximation becomes exact.

**Theorem 9.1.** Let \(X\) be binomial with parameters \(n\) and \(p\). Such that \(np = \lambda\) is a constant. Then,

\[
\lim_{n \to \infty} p_X(x) = \frac{\lambda^x e^{-\lambda}}{x!}.
\]

**Proof.** We need the following facts:

\[
\lim_{n \to \infty} n - k = \lim_{n \to \infty} n,
\]

and

\[
\lim_{n \to \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda},
\]

which we showed in class using a Taylor approximation. We have

\[
\lim_{n \to \infty} p_X(x) = \lim_{n \to \infty} \binom{n}{x} p^x (1 - p)^{n-x}
= \lim_{n \to \infty} \frac{n(n-1)(n-2)\ldots(n-x+1) \lambda^x}{x!} \frac{1}{n^x} \left(1 - \frac{\lambda}{n}\right)^{n-x}
= \lim_{n \to \infty} \frac{n^x \lambda^x}{x! n^x} \left(1 - \frac{\lambda}{n}\right)^{n-x}
= \frac{x^x}{x!} \lim_{n \to \infty} \left(1 - \frac{\lambda}{n}\right)^n
= \frac{\lambda^x e^{-\lambda}}{x!},
\]

showing that \(p_X(x)\) approaches a Poisson distribution when \(x\) is small compared to \(n\) (close to \(\lambda\)). Of course, when \(x\) is not small compared to \(n\), we will get \(p_X(x)\) close to 0 anyway. The “interesting” region is around \(\lambda\).

□
9.2 Convergence in Distribution

We have seen convergence in distribution in the central limit theorem, where we say that samples start to “behave” like normal distributions, in the sense that their CDFs converge to the CDF of a normal distribution.

Let \( \{X_n\} \) be a sequence of random variables with CDFs \( F_{X_n}(x) \), we say that

- \( \{X_n\} \) converges in distribution to \( Y \), or
- \( d\lim_{n \to \infty} X_n = Y \), or
- \( X_n \xrightarrow{d} Y \),

if \( \lim_{n \to \infty} F_{X_n}(x) = F_Y(x) \), for all \( x \) where \( F_Y(x) \) is continuous.

Exercise 9.2. We have \( X_n \) be a sequence of random variables with CDFs.

\[
F_{X_n}(x) = \begin{cases} 
0, & \text{if } x < 0, \\ 
1 - (1 - x)^n, & \text{if } 0 \leq x \leq 1, \\ 
1, & \text{if } x > 1.
\end{cases}
\]

Prove that this sequence converges in distribution. Does it converge in probability? Figure 1 shows a plot for the first few CDFs in this sequence.

![Figure 1: \( F_{X_n}(x) \) for \( n = 1, \ldots, 20 \) (dark to light).](image)
9.3 Central Limit Theorem

Let \( X_1, X_2, \ldots, X_n \) be i.i.d. random variables, with \( X_i \sim X \), and \( \mathbb{E}(X) = \mu \), \( \text{var}(X) = \sigma^2 \). You can think of this as taking a random sample from some underlying distribution. We do not even need to know what the underlying distribution is. All that matters is that the random variables are i.i.d.

We may be interested in the following derived random variables:

\[
S_n = X_1 + X_2 + \ldots + X_n,
\]

\[
M_n = \frac{X_1 + X_2 + \ldots + X_n}{n},
\]

\[
Z_n = \frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma \sqrt{n}}.
\]

Then we have the following, very powerful theorem. Let \( N(\mu, \sigma^2) \) be the normal distribution with parameters \( \mu \) and \( \sigma^2 \).

**Theorem 9.3** (The Central Limit Theorem). For \( Z_n \) as defined above, we have

\[
\lim_{n \to \infty} \mathbb{P}(Z_n \leq z) = \Phi(z).
\]

In other words, \( Z_n \distr N(0, 1) \). This also implies that \( S_n \distr N(n\mu, n\sigma^2) \) and \( M_n \distr N(\mu, \sigma^2/n) \).

**Proof.** We assume that \( \mu = 0 \). (It is easy to shift a distribution by a constant.) We have

\[
M_{Z_n}(s) = \mathbb{E}(e^{sZ_n})
= \mathbb{E}\left(e^{\frac{s}{\sigma \sqrt{n}} \sum_{i=1}^{n} X_i}\right)
= \prod_{i=1}^{n} \mathbb{E}\left(e^{\frac{s}{\sigma \sqrt{n}} X_i}\right)
= \prod_{i=1}^{n} M_{X}\left(\frac{s}{\sigma \sqrt{n}}\right)
= \left(M_{X}\left(\frac{s}{\sigma \sqrt{n}}\right)\right)^n.
\]

When we take a Taylor expansion around \( s = 0 \) (or use our known facts about moment generating functions), we obtain

\[
M_X(s) = 1 + \mathbb{E}(X)s + \frac{\mathbb{E}(X^2)}{2} s^2 + \ldots \approx 1 + \frac{\sigma^2}{2} s^2,
\]

where the higher order terms are small compared to the first terms. This gives us

\[
M_X\left(\frac{s}{\sigma \sqrt{n}}\right) \approx 1 + \frac{s^2}{2n}.
\]

Now, we have

\[
\lim_{n \to \infty} M_{Z_n}(s) = \lim_{n \to \infty} \left(1 + \frac{s^2}{2n}\right)^n = e^{s^2/2}.
\]

By the inversion property of moment generating functions, we have now shown that \( Z_n \) converges in distribution to a standard normal. \( \square \)
9.4 De Moivre - Laplace Approximation to the Binomial

De Moivre-Laplace approximation is a refinement of the normal approximation to the binomial implied by the Central Limit Theorem. Let $S_n = X_1 + \ldots + X_n$, with $X_i \sim Ber(p)$ iid, be a binomial random variable with parameters $n$ and $p$. The CLT gives us

$$P(k \leq S_n \leq l) \approx \Phi\left(\frac{l - np}{\sqrt{np(1 - p)}}\right) - \Phi\left(\frac{k - np}{\sqrt{np(1 - p)}}\right).$$

However, we are dealing with an approximation of a discrete distribution by a continuous one. For example, $P(S_n = t)$ corresponds to the probability that the corresponding normal random variable falls in the interval $[t - \frac{1}{2}, t + \frac{1}{2}]$. Therefore, a better approximation of $P(k \leq S_n \leq l)$ is given by

$$P(k \leq S_n \leq l) \approx \Phi\left(\frac{l + \frac{1}{2} - np}{\sqrt{np(1 - p)}}\right) - \Phi\left(\frac{k - \frac{1}{2} - np}{\sqrt{np(1 - p)}}\right).$$