AID-521 Mathematics for Data Science

Module: Probability | Lecture: 9

LIMIT THEOREMS IN PROBABILITY THEORY

chebyshev’s theorem, law of large numbers, central limit theorem

Chebyshev's Theorem

X is a r.v. with mean μ and s.d. σ.

For any constant $K > 0$, we have

$$P(\ |X - \mu|\ <\ K \cdot \sigma\ )\ \ \geq\ \ 1\ -\ \frac{1}{K^2}$$

→ For any data set (regardless of the shape of the distribution), at least $[(1 - \frac{1}{K^2}) \times 100]$ % of observations will lie within $K (\geq 1)$ standard deviations of the mean.

Chebyshev's Theorem

Try yourself

A r.v. X has mean 24 and variance 9.

What is the lower bound for the probability that X assumes values between 16.5 to 31.5.

Chebyshev's Theorem

Try yourself – solution

Bounds from Chebyshev's theorem looks like this:

$$-K \cdot \sigma + \mu \quad < \quad X \quad < \quad K \cdot \sigma + \mu$$

Required bounds:

$$16.5 \quad < \quad X \quad < \quad 31.5$$

Solve the above to get $K = 2.5$, which can be simply substituted in the r.h.s. of Chebyshev's theorem to obtain 0.84 as the answer.

(Weak) Law of Large Numbers

$X_1, X_2, ..., X_n$ is a set of pairwise independent r.v.s with $E[X_i] = \mu$ and $Var(X_i) = \sigma^2$.

For any constant $\epsilon > 0$, as $n \to \infty$, we have

$$P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right| < \epsilon\right) \to 1.$$

$\rightarrow$ If the sample size n is large, the sample mean rarely deviates from the mean of the distribution of X.

(Weak) Law of Large Numbers

Try yourself

Verify the law of large numbers for i.i.d. Bernoulli r.v.s $X_1, X_2, \ldots, X_n$ with parameter p.

(Weak) Law of Large Numbers

Try yourself – solution

$E[X_i] = p$, $Var(X_i) = p(1-p)$

$$
\begin{aligned}
P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - p\right| < \epsilon\right) \\
&= P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - p\right| < \frac{\epsilon}{\sigma} \cdot \sigma\right) \\
&\geq\ 1 - \frac{1}{(\epsilon/\sigma)^2}, \text{ by Chebyshev's thm.} \\
&\geq\ 1 - \frac{1}{\epsilon^2} \cdot \frac{1}{n} p(1-p) \\
&\to 1 \text{ as } n \to \infty
\end{aligned}
$$

Central Limit Theorem (CLT)

$X_1, X_2, \ldots, X_n$ is a random sample from a population with mean μ and variance σ^2.

Then, we have

$$lim_{n\to\infty}\ P(\sqrt{n}\,\frac{\bar{X}-\mu}{\sigma} \leq z) \ \sim\ F^{cdf}_{\mathcal{N}(0,1)} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-t^2/2}dt.$$

- $\rightarrow$ A widely used result.
- $\rightarrow$ The z-transform of the sample mean of a r.v. X is asymptotically standard normal, irrespective of the shape of the distribution of X.

Central Limit Theorem (CLT)

Try yourself

The service times for customers at a store are independent r.v.s with mean 1.5 minutes and variance 1.0.

Find the probability that 100 customers can be served in less than 2 hours of total service time.

Central Limit Theorem (CLT)

Try yourself – solution

The r.v. is clearly the time needed for each customer. Let X_i denote the service time needed for ith customer.

The required probability then becomes

$$P\Big(\sum_{i=1}^{100} X_i \le 120\Big) \;=\; P(\bar{X} \le 1.20)$$

$$= P\left(\sqrt{100}\frac{\bar{X} - 1.50}{1} \;\le\; \sqrt{100}\frac{1.20 - 1.50}{1}\right) \overset{\text{CLT}}{\approx} P(Z \le -3) = 0.0013,$$

where $Z \sim \mathcal{N}(0, 1)$ is the standard normal r.v.