Central Limit Theorem
03/15/2025

Inferential statistics analyzes a sample randomly selected from the object population and provides inferences on the population's characteristics based on the sample results. The Central Limit Theorem establishes the foundation for such inferences.

Markov's Inequality

Let X be a nonnegative random variable and t a positive number, then

$$P(X \geq t) \leq \frac{E(X)}{t}.$$

Proof:

$$E(X) = \sum_{x} xp(x) = \sum_{x<t} xp(x) + \sum_{x \geq t} xp(x) \geq \sum_{x \geq t} xp(x) \geq \sum_{x \geq t} tp(x) = tP(X \geq t) \Box$$

Chebyshev's Inequality

Let X be a random variable with mean μ and variance σ^2. For any $t > 0$,

$$P(|X - \mu| > t) \leq \frac{\sigma^2}{t^2}$$

Proof: Let $Y = (X - \mu)^2$, therefore $E(Y) = \sigma^2$.
$P(|X - \mu| > t) = P((X - \mu)^2 > t^2) = P(Y > t^2) \leq E(Y)/t^2 = \sigma^2/t^2$ by following Markov's inequality$\Box$.

Law of Large Numbers

Let $X_1, X_2, \ldots, X_i, \ldots$ be a sequence of independent random variables with mean μ and variance σ^2. Further, let n be the number of elements in a subset of X and the mean of the subset as:

$$\overline{X_n} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Then for any $\epsilon > 0$:

$P(|\overline{X_n} - \mu| > \epsilon) \to 0 \ as \ n \to \infty$

Proof:

$$E(\overline{X_n}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{n\mu}{n} = \mu$$

$$Var(\overline{X_n})\frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Based on Chebyshev's inequality, we have the following:

$$P(|\overline{X_n} - \mu| > \epsilon) \leq \frac{Var(\overline{X_n})}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0, \ as \ n \to \infty \square$$

Moment-Generating Function

The moment-generating function of a random variable, X, is

$M(t) = E(e^{tX})$

It has the following properties:

$M'(0) = E(X)$

$M''(0) = E(X^2)$

For the standard normal distribution:

$M(t) = e^{t^2/2}, \ E(X) = 0 \ and \ Var(X) = 1$

If X and Y are independent random variables, with moment-generating functions $M_X(t)$ and $M_Y(t)$, and $Z = X + Y$, the $M_Z(t) = M_X(t)\dot{M}_Y(t)$.

Classic Central Limit Theorem

Let $X_1, X_2, ...$ be a sequence of independent random variables having mean 0 and variance σ^2. Let

$$S_n = \sum_{i=1}^{n} X_i$$

Then

$$\lim_{n \to \infty} P\left(\frac{S_n}{\sigma\sqrt{n}} \le z\right) = \Phi(z)$$

where $\Phi(z)$ is the cumulative standard normal distribution.

__Path to the Proof__: Let $Z_n = S_n/(\sigma\sqrt{n})$. Using Taylor expansion, it can be shown the moment generating function of Z_n:

$$M_{Z_n}(t) = \left(1 + \frac{t^2}{2n} + \epsilon_n\right)^n \to e^{t^2/2} \; as \; n \to \; \infty$$

Therefore, $M_{Z_n}(t)$ approaches the moment generating function of the standard normal distribution that leads to the same μ and σ^2.

__Central Limit Theorem with t Distribution__

If Z is a standard normal random variable, the distribution of $U = Z^2$ is called the chi-square distribution with 1 degree of freedom. Further, if $U_1, U_2, \ldots, U_n$ are independent chi-square random variables with 1 degree of freedom, the distribution of $U_1 + U_2 +, \ldots, +U_n$ is called the chi-square distribution with n degrees of freedom, denoted by χ_n^2.

If $Z \sim N(0,1)$ and $U \sim \chi_n^2$ and Z and U are independent, then the distribution of $Z/\sqrt{U/n}$ is called the t distribution with n degrees of freedom.

Let $X_1, X_2, \ldots, X_n$ be independent $\sim N(\mu, \sigma^2)$ random variables, and note

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \; and \; S^2 = \frac{1}{n-1}\sum_{i-1}^{n}(X - \overline{X})^2$$

Then

$$E(\overline{X}) = \mu \; and \; Var(\overline{X}) = \frac{\sigma^2}{n}$$

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

Proof:

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} = \frac{\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)}{\sqrt{S^2/\sigma^2}}$$

Recall:

$$S^2 = \frac{1}{n-1}\sum_{i-1}^{n}(X-\overline{X})^2$$

The right hand side of the first equality is in the form of $Z/\sqrt{U/(n-1)}$ as the t distribution with $n-1$ degree of freedom. $\square$

Concluding Remarks

The law of large numbers defines the sample mean of a sequence of random variables (the population set) and suggests that the sample mean becomes an adequate point estimate of the population mean as the sample size is sufficiently large. This provides the foundation for inferential statistical estimates based on the population mean. The classic central limit theorem expands the inferential capability further by specifying the “overlapping” probability of the sample mean over the population mean. This probability is conveniently the cumulative standard normal distribution of a “z-score”, again, when the sample size is sufficiently large and the standard deviation of the population happens to be known. Once a probability range is chosen (the confidence interval), the margin of error can be computed for the point estimate of the population mean. However, this magic power of inferential statistics is constructed based on the assumption that the standard deviation of the population is known which often is not the case. The version of central limit theorem based on the t distribution defines a term called standard error using the square root of the sample-mean variance. Applying the standard error of a sufficiently large number of samples, one can compute the margin of error of the estimated population mean without requiring the population standard deviation. The probability, however, is the cumulative t distribution with n-1 degrees of freedom, not the cumulative standard normal distribution any more.

At this juncture, the inferential capability of statistics of random variables is firmly established, not only for predicting the population mean but bracketing this prediction with margin of error at any given confidence interval.