Probabilistic proof [edit]

Markov's inequality states that for any real-valued random variable *Y* and any positive number *a*, we have Pr(|*Y*| > *a*) ≤ E(|*Y*|)/*a*. One way to prove Chebyshev's inequality is to apply Markov's inequality to the random variable $Y = (X - \mu)^2$ with $a = (k\sigma)^2$.

It can also be proved directly. For any event *A*, let I_A be the indicator random variable of *A*, i.e. I_A equals 1 if *A* occurs and 0 otherwise. Then

$$\begin{aligned}\Pr(|X-\mu| \geq k\sigma) &= \operatorname{E}\left(I_{|X-\mu|\geq k\sigma}\right) \\ &= \operatorname{E}\left(I_{\left(\frac{X-\mu}{k\sigma}\right)^2 \geq 1}\right) \\ &\leq \operatorname{E}\left(\left(\frac{X-\mu}{k\sigma}\right)^2\right) \\ &= \frac{1}{k^2}\frac{\operatorname{E}((X-\mu)^2)}{\sigma^2} \\ &= \frac{1}{k^2}.\end{aligned}$$

The Inequality of direct proof shows why the bounds are quite loose in typical cases:

1. If $0 \leq \left(\frac{X-\mu}{k\sigma}\right)^2 < 1$, instead of taking the indicating value 0 as given by the left side of the inequality, a positive value of $\left(\frac{X-\mu}{k\sigma}\right)^2$ is counted.
2. If $\left(\frac{X-\mu}{k\sigma}\right)^2 \geq 1$, instead of taking the indicating value 1 as given by the left side of the inequality, a value $\left(\frac{X-\mu}{k\sigma}\right)^2$ greater or equal to 1 is counted. In some cases it exceeds 1 by a very wide margin.

Proof of the weak law [edit]

Given X_1, X_2, ... an infinite sequence of i.i.d. random variables with finite expected value $E(X_1) = E(X_2) = ... = \mu < \infty$, we are interested in the convergence of the sample average

$$\overline{X}_n = \tfrac{1}{n}(X_1 + \cdots + X_n).$$

The weak law of large numbers states:

Theorem: $\overline{X}_n \xrightarrow{P} \mu \qquad \text{when } n \to \infty.$ **(law. 2)**

Proof using Chebyshev's inequality assuming finite variance [edit]

This proof uses the assumption of finite variance $\operatorname{Var}(X_i) = \sigma^2$ (for all i). The independence of the random variables implies no correlation between them, and we have that

$$\operatorname{Var}(\overline{X}_n) = \operatorname{Var}(\tfrac{1}{n}(X_1 + \cdots + X_n)) = \frac{1}{n^2}\operatorname{Var}(X_1 + \cdots + X_n) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

The common mean μ of the sequence is the mean of the sample average:

$$E(\overline{X}_n) = \mu.$$

Using Chebyshev's inequality on $\overline{X}_n$ results in

$$\operatorname{P}(\left|\overline{X}_n - \mu\right| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

This may be used to obtain the following:

$$\operatorname{P}(\left|\overline{X}_n - \mu\right| < \varepsilon) = 1 - \operatorname{P}(\left|\overline{X}_n - \mu\right| \geqslant \varepsilon) \geqslant 1 - \frac{\sigma^2}{n\varepsilon^2}.$$

As n approaches infinity, the expression approaches 1. And by definition of convergence in probability, we have obtained

$$\overline{X}_n \xrightarrow{P} \mu \qquad \text{when } n \to \infty.$$ **(law. 2)**

Classical CLT [edit]

Let $\{X_1, \ldots, X_n\}$ be a random sample of size n—that is, a sequence of independent and identically distributed (i.i.d.) random variables drawn from a distribution of expected value given by μ and finite variance given by σ^2. Suppose we are interested in the sample average

$$S_n := \frac{X_1 + \cdots + X_n}{n}$$

of these random variables. By the law of large numbers, the sample averages converge in probability and almost surely to the expected value μ as $n \to \infty$. The classical central limit theorem describes the size and the distributional form of the stochastic fluctuations around the deterministic number μ during this convergence. More precisely, it states that as n gets larger, the distribution of the difference between the sample average S_n and its limit μ, when multiplied by the factor $\sqrt{n}$ (that is $\sqrt{n}(S_n - \mu)$), approximates the normal distribution with mean 0 and variance σ^2. For large enough n, the distribution of S_n is close to the normal distribution with mean μ and variance σ^2/n. The usefulness of the theorem is that the distribution of $\sqrt{n}(S_n - \mu)$ approaches normality regardless of the shape of the distribution of the individual X_i. Formally, the theorem can be stated as follows:

> **Lindeberg–Lévy CLT.** Suppose $\{X_1, X_2, \ldots\}$ is a sequence of i.i.d. random variables with $\mathrm{E}[X_i] = \mu$ and $\mathrm{Var}[X_i] = \sigma^2 < \infty$. Then as n approaches infinity, the random variables $\sqrt{n}(S_n - \mu)$ converge in distribution to a normal $N(0,\sigma^2)$:[3]
>
> $$\sqrt{n}\left(S_n - \mu\right) \xrightarrow{d} N\left(0, \sigma^2\right).$$

In the case $\sigma > 0$, convergence in distribution means that the cumulative distribution functions of $\sqrt{n}(S_n - \mu)$ converge pointwise to the cdf of the $N(0, \sigma^2)$ distribution: for every real number z,

$$\lim_{n\to\infty} \Pr\left[\sqrt{n}(S_n - \mu) \le z\right] = \lim_{n\to\infty} \Pr\left[\frac{\sqrt{n}(S_n - \mu)}{\sigma} \le \frac{z}{\sigma}\right] = \Phi\left(\frac{z}{\sigma}\right),$$

where $\Phi(z)$ is the standard normal cdf evaluated at z. The convergence is uniform in z in the sense that

$$\lim_{n\to\infty} \sup_{z\in\mathbb{R}} \left|\Pr\left[\sqrt{n}(S_n - \mu) \le z\right] - \Phi\left(\frac{z}{\sigma}\right)\right| = 0,$$

where sup denotes the least upper bound (or supremum) of the set.[4]