Markov, Chebyshev, and the Weak Law of Large Numbers

The Law of Large Numbers is one of the fundamental theorems of statistics. One version of this theorem, The Weak Law of Large Numbers, can be proven in a fairly straightforward manner using Chebyshev's Theorem, which is, in turn, a special case of the Markov Inequality.

Markov's Inequality: If X is a random variable that takes only nonnegative values, then for any value $c > 0$, $P(X \geq c) \leq \dfrac{E(X)}{c}$.

The discrete and continuous forms of this theorem are quite similar, varying only in using sums $\left(E(X) = \sum_i x_i\, p(x_i)\right)$ or integrals $\left(E(X) = \int_{-\infty}^{\infty} x\, p(x)\right)$. Here, we give the continuous form.

If X is a continuous random variable with density function $p(x)$, then

$$E(X) = \int_0^{\infty} x\, p(x)\, dx .$$

We can break this interval into two pieces, those values of x less than or equal to c and those values of x greater than or equal to c. So

$$E(X) = \int_0^{c} x\, p(x)\, dx + \int_c^{\infty} x\, p(x)\, dx .$$

If we take only those values of x greater than or equal to c, we have

$$E(X) \geq \int_c^{\infty} x\, p(x)\, dx .$$

Since the values of x in this integrand have the property that $x \geq c$, if we replace x with c, we have

$$E(X) \geq \int_c^{\infty} x\, p(x)\, dx \geq \int_c^{\infty} c\, p(x)\, dx = c\int_c^{\infty} p(x)\, dx .$$

This last integrand is just the probability that X is greater than or equal to c, so

$$E(X) \geq a\ P(X \geq c) \text{ so } P(X \geq c) \leq \frac{E(X)}{c} .$$

This means that for any random variable with nonnegative values which has a mean of 10, the probability that $X > 15$ is less than or equal to $\frac{10}{15} = 0.67$, regardless of the variance of the random variable.

Chebyshev's Theorem: Let *X* be a random variable with mean m and standard deviation s. Let *k* be any positive number. Then $P(|X - m| > k) < \frac{s^2}{k^2}$.

Now that we have Markov's Inequality, we recognize that $(X - m)^2$ is a nonnegative random variable and apply the inequality to $E\left[(X - m)^2\right]$ with $c = k^2$. Instead, we will use a similar argument to that used for the Markov Inequality, but consider a discrete random variable.

By definition, we have

$$s^2 = \sum_{i=1}^{n} (x_i - m)^2 \, p(x_i).$$

If we delete some terms in the sum, the sum will get smaller. So, delete all terms for which $|x_i - m| \le k$. We will denote the terms remaining with x^*. Now sum only those values x^* x^* where $|x^* - m| > k$ to find that

$$s^2 \ge \sum_j \left(x_j^* - m\right)^2 p\left(x_j^*\right).$$

The subscript *j* is used to indicate that this summation may not be as extensive as the summation on *i*. Since we are using only those values x^* where $|x^* - m| > k$, if we replace each $\left(x_j^* - m\right)$ with *k*, we have

$$s^2 \ge \sum_j \left(x_j^* - m\right)^2 p\left(x_j^*\right) > \sum_j k^2 \, p\left(x_j^*\right) = k^2 \sum_j p\left(x_j^*\right) \text{ and } \frac{s^2}{k^2} > \sum_j p\left(x_j^*\right)$$

As before, the summation $\sum_j p\left(x_j^*\right)$ is just $P(|X - m| > k)$ and so $P(|X - m| > k) < \frac{s^2}{k^2}$. Thus, for example, in any population with mean 10 and standard deviation 2, at most one-fourth of the values will be more than 4 units from 10. Consequently, at least 75% of the values will be between 6 and 14.

The Weak Form of the Law of Large Numbers: Let a population be specified by a random variable X with mean $m_{\overline{X}}$ and standard deviation $s_{\overline{X}}$. Let $\overline{X}$ be the mean of a random sample of size *n* drawn with replacement from this population. Let *c* be any positive number. Then $\lim_{n \to \infty} P\left(m_X - c \le \overline{X} \le m_X + c\right) = 1$.

To prove this form of the Law of Large Numbers, invoke Chebyshev's Theorem on the random variable $\overline{X}$ with mean $m_{\overline{X}}$ and standard deviation $s_{\overline{X}}$. Thus,

$$P\left(\left|\overline{X} - m_{\overline{X}}\right| > c\right) < \frac{s_{\overline{X}}^2}{c^2}.$$

We need only rewrite the mean $m_{\overline{X}}$ and standard deviation $s_{\overline{X}}$ in terms of the population mean m_X and population standard deviation s_X. We know that $m_{\overline{X}} = m_X$ and $s_{\overline{X}}^2 = \frac{s_X^2}{n}$. So we have

$$P\left(\left|\overline{X} - m_X\right| > c\right) < \frac{s_X^2}{nc^2} \text{ and so } P\left(\left|\overline{X} - m_X\right| \le c\right) > 1 - \frac{s_X^2}{nc^2}.$$

Thus, $P\left(m_X - c \le \overline{X} \le m_X + c\right)$ is bound between $1 - \frac{s_X^2}{nc^2}$ and 1. As $n \to \infty$, the probability approaches 1. So $\lim_{n\to\infty} P\left(m_X - c \le \overline{X} \le m_X + c\right) = 1$.

References:

Goldberg, Samuel, *Probability: An Introduction*, Dover Publications, New York, New York, 1960.

Ross, Sheldon, *A First Course in Probability*, Macmillan Publishing Company, New York, New York, 1976.