Section 8.2 Markov and Chebyshev Inequalities and the Weak Law of Large Numbers

THEOREM (Markov's Inequality): Suppose that X is a random variable taking only non-negative values. Then, for any $a > 0$ we have

$$P\{X \geq a\} \leq \frac{E[X]}{a}$$

Proof: Consider the random variable

$$1_{\{X \geq a\}} = \begin{cases} 1 & \text{if } X \geq a \\ 0 & \text{else} \end{cases}$$

Note that because $X \geq 0$,

$$1_{\{X \geq a\}} \leq \frac{X}{a}$$

Therefore,

$$\frac{E[X]}{a} \geq E\left[1_{\{X \geq a\}}\right] = P\{1_{\{X \geq a\}} = 1\} = P\{X \geq a\}$$

Rearranging terms give s the result. ■

EXAMPLE: Let X be uniform$(0,1)$. Then for any $a > 0$ Markov's inequality gives

$$P\{X \geq a\} \leq \frac{1}{2a}$$

whereas

$$P\{X \geq a\} = \int_a^1 dx = 1 - a, \quad 0 < a < 1$$

It is easy to show that $\frac{1}{2a} > 1 - a$ for any $a > 0$. Moreover, $f(a) = \frac{1}{2a} + a - 1$ attains its minimum on $(0,1)$ at $a = \frac{1}{\sqrt{2}}$ with

$$f\left(\frac{1}{\sqrt{2}}\right) = \sqrt{2} - 1 = 0.4142...$$

EXAMPLE: A post office handles, on average, 10,000 letters a day. What can be said about the probability that it will handle at least 15,000 letters tomorrow.

EXAMPLE: A post office handles, on average, 10,000 letters a day. What can be said about the probability that it will handle at least 15,000 letters tomorrow.

Solution: Let X be the number of letters handled tomorrow. Then $E(X) = 10,000$. Further, by Markov's inequality we get

$$P(X \geq 15,000) \leq \frac{E(X)}{15,000} = \frac{10,000}{15,000} = \frac{2}{3}$$

THEOREM (Chebyshev's Inequality): Let X be a random variable with finite mean μ and variance σ^2, then for any $k > 0$ we have

$$P\{|X - \mu| \geq k\} \leq \frac{\sigma^2}{k^2}$$

THEOREM (Chebyshev's Inequality): Let X be a random variable with finite mean μ and variance σ^2, then for any $k > 0$ we have

$$P\{|X-\mu| \geq k\} \leq \frac{\sigma^2}{k^2}$$

Proof: Note that $(X-\mu)^2 \geq 0$, so we may use Markov's inequality. Apply with $a = k^2$ to find that

$$P\{(X-\mu)^2 \geq k^2\} \leq \frac{E[(X-\mu)^2]}{k^2} = \frac{\sigma^2}{k^2}$$

But $(X-\mu)^2 \geq k^2$ if and only if $|X-\mu| \geq k$, so

$$P\{|X-\mu| \geq k\} \leq \frac{\sigma^2}{k^2} \ \blacksquare$$

REMARK: Note that if $k = a\sigma$ in Chebyshev's inequality, then

$$P(|X-\mu| \geq a\sigma) \leq \frac{1}{a^2}$$

This gives that probability of being a deviates from the expected value drops like $1/a^2$. Of course, we also have

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

EXAMPLE: Let X be uniform$(0,1)$. Then for any $a > 0$ Chebyshev's Inequality gives

$$P\left(\left|X - \frac{1}{2}\right| < a\sigma\right) \geq 1 - \frac{1}{a^2}$$

whereas

$$P\left(\left|X - \frac{1}{2}\right| < a\sigma\right) = P\left(-a\sigma < X - \frac{1}{2} < a\sigma\right) = P\left(-a\sigma + \frac{1}{2} < X < a\sigma + \frac{1}{2}\right)$$

$$= \int_{-a\sigma+1/2}^{a\sigma+1/2} dx = 2a\sigma = \frac{2a}{\sqrt{12}} = \frac{a}{\sqrt{3}}, \quad 0 < a \leq \sqrt{3}$$

One can show that $\frac{a}{\sqrt{3}} > 1 - \frac{1}{a^2}$ for any $a > 0$. Moreover, $f(a) = \frac{a}{\sqrt{3}} + \frac{1}{a^2} - 1$ attains its minimum on $(0,\infty)$ at $a = \sqrt[6]{12} = 1.513...$ with

$$f\left(\sqrt[6]{12}\right) = 0.3103...$$

REMARK: Note that $P\left(\left|X - \frac{1}{2}\right| < a\sigma\right) = 1$ if $a \geq \sqrt{3}$ and $1 - \frac{1}{a^2} \to 1$ as $a \to \infty$.

EXAMPLE: A post-office handles 10,000 letters per day with a variance of 2,000 letters. What can be said about the probability that this post office handles between 8,000 and 12,000 letters tomorrow? What about the probability that more than 15,000 letters come in?

EXAMPLE: A post-office handles 10,000 letters per day with a variance of 2,000 letters. What can be said about the probability that this post office handles between 8,000 and 12,000 letters tomorrow? What about the probability that more than 15,000 letters come in?

Solution: We want

$$P(8,000 < X < 12,000) = P(-2,000 < X - 10,000 < 2,000)$$

$$= P(|X - 10,000| < 2,000) = 1 - P(|X - 10,000| \geq 2,000)$$

By Chebyshev's inequality we have

$$P(|X - 10,000| \geq 2,000) \leq \frac{\sigma^2}{(2,000)^2} = \frac{1}{2,000} = 0.0005$$

Thus,

$$P(8,000 < X < 12,000) \geq 1 - 0.0005 = 0.9995$$

Now to bound the probability that more than 15,000 letters come in. We have

$$P\{X \geq 15,000\} = P\{X - 10,000 \geq 5,000\} \leq P\{(X - 10,000 \geq 5,000) \cup (X - 10,000 \leq -5,000)\}$$

$$= P\{|X - 10,000| \geq 5,000\} \leq \frac{2,000}{5,000^2} = \frac{1}{12500} = 0.00008$$

Recall, that the old bound (no information about variance) was 2/3.

EXAMPLE: Let X be the outcome of a roll of a die. We know that

$$E[X] = 21/6 = 3.5 \quad \text{and} \quad \text{Var}(X) = 35/12$$

Thus, by Markov,

$$0.167 \approx 1/6 = P(X \geq 6) \leq \frac{E[X]}{6} = \frac{21/6}{6} = 21/36 \approx 0.583$$

Chebyshev gives

$$P\ (\ X \leq 2 \text{ or } X \geq 5) \ = 2/3 = P(|X - 3.5| \geq 1.5) \leq \frac{35/12}{1.5^2} \approx 1.296$$

THEOREM: If $\text{Var}(X) = 0$, then

$$P\{X = E[X]\} = 1$$

THEOREM: If $\mathrm{Var}(X) = 0$, then

$$P\{X = E[X]\} = 1$$

Proof. We will use Chebyshev's inequality. Let $E[X] = \mu$. For any $n \geq 1$ we have

$$P\left\{|X - \mu| \geq \frac{1}{n}\right\} = 0$$

Let $E_n = \left\{|X - \mu| \geq \frac{1}{n}\right\}$. Note that

$$E_1 \subseteq E_2 \subseteq \ldots \subseteq E_n \subseteq E_{n+1} \subseteq \ldots$$

so this is an increasing sequence of sets. Therefore,

$$0 = \lim_{n\to\infty} P\left\{|X - \mu| \geq \frac{1}{n}\right\} = P\left\{\lim_{n\to\infty}\left\{|X - \mu| \geq \frac{1}{n}\right\}\right\} = P\left\{\bigcup_{n=1}^{\infty} E_n\right\} = P\{X \neq \mu\}$$

Hence,

$$P\{X = \mu\} = 1 - P\{X \neq \mu\} = 1 \blacksquare$$

THEOREM (The Weak Law of Large Numbers): Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having the finite mean $E[X_i] = \mu$. Then for any $\epsilon > 0$ we have

$$P\left\{\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \geq \epsilon\right\} \to 0 \quad \text{as} \quad n \to \infty$$

THEOREM (The Weak Law of Large Numbers): Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having the finite mean $E[X_i] = \mu$. Then for any $\epsilon > 0$ we have

$$P\left\{\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \geq \epsilon\right\} \to 0 \quad \text{as} \quad n \to \infty$$

Proof: We will also assume that σ^2 is finite to make the proof easier. We know that

$$E\left[\frac{X_1 + \ldots + X_n}{n}\right] = \mu \quad \text{and} \quad \text{Var}\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{\sigma^2}{n}$$

Therefore, a direct application of Chebyshev's inequality shows that

$$P\left\{\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \geq \epsilon\right\} \leq \frac{\sigma^2}{n\epsilon^2} \to 0 \quad \text{as} \quad n \to \infty \blacksquare$$

Statistics and Sample Sizes

Suppose that we want to estimate the mean μ of a random variable using a finite number of independent samples from a population (IQ score, average income, average surface water temperature, etc.).

Major Question: How many samples do we need to be 95% sure our value is within .01 of the correct value?

Let μ be the mean and σ^2 be the variance and let

$$\overline{X} = \frac{X_1 + \ldots + X_n}{n}$$

be the sample mean. We know that

$$E[\overline{X}] = \mu, \qquad \text{Var}(\overline{X}) = \text{Var}\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{\sigma^2}{n}$$

By Chebyshev we have

$$P\{|\overline{X} - \mu| \geq \epsilon\} \leq \frac{\sigma^2}{\epsilon^2 n}$$

Now, we want that for some n

$$P\{|\overline{X} - \mu| < .01\} \geq .95 \quad \Longrightarrow \quad P\{|\overline{X} - \mu| \geq .01\} \leq .05$$

Thus, it is sufficient to guarantee that

$$\frac{\sigma^2}{(.01)^2 n} = .05 \quad \Longrightarrow \quad n \geq \frac{\sigma^2}{(.01)^2(.05)} = 200,000\sigma^2$$

To be 95% sure we are within .1 of the mean we need to satisfy

$$\frac{\sigma^2}{(.1)^2 n} = .05 \quad \Longrightarrow \quad n \geq \frac{\sigma^2}{(.1)^2(.05)} = 2,000\sigma^2$$

Section 8.3 The Central Limit Theorem

EXAMPLE: Consider rolling a die 100 times (each X_i is the output from one roll) and adding outcomes. We will get around 350, plus or minus some. We do this experiment 10,000 times and plot the number of times you get each outcome. The graph looks like a bell curve.

EXAMPLE: Go to a library and go to the stacks. Each row of books is divided into $n \gg 1$ pieces. Let X_i be the number of books on piece i. Then $\sum\limits_{i=1}^{n} X_i$ is the number of books on a given row. Do this for all rows of the same length. You will get a plot that looks like a bell curve.

Recall, if $X_1, X_2, \ldots$ are independent and identically distributed random variables with mean μ, then by the weak law of large numbers we have that

$$\frac{1}{n}[X_1 + \ldots + X_n] \to \mu$$

or

$$\frac{1}{n}[X_1 + \ldots + X_n - n\mu] \to 0$$

In a sense, this says that scaling by n was "too much." Why? Note that the random variable

$$X_1 + X_2 + \ldots + X_n - n\mu$$

has mean zero and variance

$$\mathrm{Var}(X_1 + \ldots + X_n) = n\mathrm{Var}(X)$$

Therefore,

$$\frac{X_1 + \ldots + X_n - n\mu}{\sigma\sqrt{n}}$$

has mean zero and variance one.

THEOREM (Central Limit Theorem): Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with mean μ and variance σ^2. Then the distribution of

$$\frac{X_1 + \ldots + X_n - n\mu}{\sigma\sqrt{n}}$$

tends to the standard normal as $n \to \infty$. That is, for $-\infty < a < \infty$ we have

$$P\left\{\frac{X_1 + \ldots + X_n - n\mu}{\sigma\sqrt{n}} \le a\right\} \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} e^{-x^2/2}dx \quad \text{as} \quad n \to \infty$$

To prove the theorem, we need the following:

LEMMA: Let $Z_1, Z_2, \ldots$ be a sequence of random variables with distribution functions F_{Z_n} and moment generating functions M_{Z_n}, $n \ge 1$; and let Z be a random variable with a distribution function F_Z and a moment generating function M_Z. If

$$M_{Z_n}(t) \to M_Z(t)$$

for all t, then $F_{Z_n}(t) \to F_Z(t)$ for all t at which $F_Z(t)$ is continuous.

REMARK: Note, that if Z is a standard normal, then

$$M_Z(t) = e^{t^2/2}$$

which is continuous everywhere, so we can ignore that condition in our case. Thus, we now see that if

$$M_{Z_n}(t) \to e^{t^2/2} \quad \text{as} \quad n \to \infty$$

for each t, then

$$F_{Z_n}(t) \to \Phi(t)$$

Proof: Assume that $\mu = 0$ and $\sigma^2 = 1$. Let $M(t)$ denote the moment generating function of the common distribution. The moment generating function of $X_i/\sqrt{n}$ is

$$E[e^{tX_i/\sqrt{n}}] = M\left(\frac{t}{\sqrt{n}}\right)$$

Therefore, the moment generating function of the sum $\sum_{i=1}^{n} \frac{X_i}{\sqrt{n}}$ is $\left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n$. We need

$$\left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n \to e^{t^2/2} \quad \Longleftrightarrow \quad \ln\left(\left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n\right) \to \frac{t^2}{2}$$

Let

$$L(t) = \ln M(t)$$

Note that

$$L(0) = \ln M(0) = 0, \qquad L'(0) = \frac{M'(0)}{M(0)} = \mu = 0$$

and

$$L''(0) = \frac{M''(0)M(0) - M'(0)^2}{[M(0)]^2} = E[X^2] - (E[X])^2 = 1$$

Putting pieces together, we get

$$\ln\left(\left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n\right) = n \ln M\left(\frac{t}{\sqrt{n}}\right) = nL\left(\frac{t}{\sqrt{n}}\right)$$

Finally, we obtain

$$\begin{aligned}
\lim_{n\to\infty} \frac{L(t/\sqrt{n})}{n^{-1}} &= \lim_{n\to\infty} \frac{-L'(t/\sqrt{n})n^{-3/2}t}{-2n^{-2}} \qquad \text{(by L'Hopital's rule)} \\
&= \lim_{n\to\infty} \frac{L'(t/\sqrt{n})t}{2n^{-1/2}} \\
&= \lim_{n\to\infty} \frac{-L''(t/\sqrt{n})t^2n^{-3/2}}{-2\cdot 2\cdot (1/2)\cdot n^{-3/2}} \qquad \text{(by L'Hopital's rule)} \\
&= \lim_{n\to\infty} \frac{L''(t/\sqrt{n})t^2}{2} \\
&= \frac{t^2}{2}
\end{aligned}$$

So, we are done if $\mu = 0$ and $\sigma^2 = 1$. For the general case, just consider

$$X_i^* = \frac{X_i - \mu}{\sigma}$$

which has mean zero and variance one. Then,

$$P\left\{\frac{X_1^* + \ldots + X_n^*}{\sqrt{n}} \le t\right\} \to \Phi(t) \quad \Longrightarrow \quad P\left\{\frac{X_1 + \ldots + X_n - n\mu}{\sigma\sqrt{n}} \le t\right\} \to \Phi(t)$$

which is the general result. ■

EXAMPLE: An astronomer is measuring the distance to a star. Because of different errors, each measurement will not be precisely correct, but merely an estimate. He will therefore make a series of measurements and use the average as his estimate of the distance. He believes his measurements are independent and identically distributed with mean d and variance 4. How many measurements does he need to make to be 95% sure his estimate is accurate to within $\pm.5$ light-years?

Solution: We let the observations be denoted by X_i. From the central limit theorem we know that

$$Z_n = \frac{\sum_{i=1}^n X_i - nd}{2\sqrt{n}}$$

has approximately a normal distribution. Therefore, for a fixed n we have

$$P\left\{-.5 \le \frac{1}{n}\sum_{i=1}^n X_i - d \le .5\right\} = P\left\{-.5\frac{\sqrt{n}}{2} \le Z_n \le .5\frac{\sqrt{n}}{2}\right\}$$

$$= \Phi\left(\frac{\sqrt{n}}{4}\right) - \Phi\left(-\frac{\sqrt{n}}{4}\right)$$

$$= 2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1$$

To be 95% certain we want to find n^* so that

$$2\Phi\left(\frac{\sqrt{n^*}}{4}\right) - 1 = .95 \quad \text{or} \quad \Phi\left(\frac{\sqrt{n^*}}{4}\right) = .975$$

Using the chart we see that

$$\frac{\sqrt{n^*}}{4} = 1.96 \quad \text{or} \quad n^* = (7.84)^2 = 61.47$$

Thus, he should make 62 observations.

It's tricky to know how good this approximation is. However, we could also use Chebyshev's inequality if we want a sharp bound. Since

$$E\left[\frac{1}{n}\sum_{i=1}^n X_i\right] = d \quad \text{and} \quad \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{4}{n}$$

Chebyshev's inequality yields

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^n X_i - d\right| > .5\right\} \le \frac{4}{n(.5)^2} = \frac{16}{n}$$

Therefore, to guarantee the above probability is less than or equal to .05, we need to solve

$$\frac{16}{n} = .05 \quad \Longrightarrow \quad n = \frac{16}{.05} = 320$$

That is, if he makes 320 observations, he can be 95% sure his value is within ±.5 light-years of the true value.

EXAMPLE: Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with mean μ and standard deviation σ. Set $S_n = X_1 + \ldots + X_n$. For a large n, what is the approximate probability that S_n is between $E[S_n] - k\sigma_{S_n}$ and $E[S_n] + k\sigma_{S_n}$ (the probability of being k deviates from the mean)?

Solution: We have

$$E[S_n] = E[X_1 + \ldots + X_n] = n\mu \quad \text{and} \quad \sigma_{S_n} = \sqrt{\text{Var}(S_n)} = \sqrt{\text{Var}(X_1 + \ldots + X_n)} = \sigma\sqrt{n}$$

Thus, letting $Z \sim N(0,\ 1)$, we get

$$\begin{aligned}
P\{E[S_n] - k\sigma_{S_n} \le S_n \le E[S_n] + k\sigma_{S_n}\} &= P\{n\mu - k\sigma\sqrt{n} \le S_n \le n\mu - k\sigma\sqrt{n}\} \\
&= P\{-k\sigma\sqrt{n} \le S_n - n\mu \le k\sigma\sqrt{n}\} \\
&= P\{-k \le \frac{S_n - n\mu}{\sigma\sqrt{n}} \le k\} \\
&\approx P(-k \le Z \le k) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-k}^{k} e^{-x^2/2} dx \\
&= \begin{cases} .6826 & k = 1 \\ .9545 & k = 2 \\ .9972 & k = 3 \\ .9999366 & k = 4 \end{cases}
\end{aligned}$$

Compare: Chebyshev's inequality gives

$$\begin{aligned}
P\{|S_n - n\mu| < k\sigma\sqrt{n}\} &= 1 - P\{|S_n - n\mu| \ge k\sigma\sqrt{n}\} \\
&\ge 1 - \frac{\sigma^2 n}{k^2\sigma^2 n} \\
&= 1 - \frac{1}{k^2} \\
&= \begin{cases} 0 & k = 1 \\ .75 & k = 2 \\ .8889 & k = 3 \\ .9375 & k = 4 \end{cases}
\end{aligned}$$

Section 8.4 The Strong Law of Large Numbers

THEOREM (The Strong Law of Large Numbers): Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having a finite mean $\mu = E[X_i]$ and finite fourth moment $E[X_i^4] = K$. Then, with probability one,

$$\frac{X_1 + \ldots + X_n}{n} \to \mu \quad \text{as} \quad n \to \infty$$

Compare: By the Weak Law we have

$$P\left\{\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \geq \epsilon\right\} \to 0 \quad \text{as} \quad n \to \infty$$

Imagine you are rolling a die and you are tracking

$$\overline{X}_n = \frac{X_1 + \ldots + X_n}{n}$$

The Weak Law says that for large n, *the probability* that $\overline{X}_n$ is more than ϵ away from μ goes to zero. However, we could still have

$$|\overline{X}_{100} - \mu| > \epsilon, \quad |\overline{X}_{1000} - \mu| > \epsilon, \quad |\overline{X}_{10000} - \mu| > \epsilon, \quad |\overline{X}_{100000} - \mu| > \epsilon, \text{ etc.,}$$

it's just that it happens very rarely.

The Strong Law says that *with probability one* for a given $\epsilon > 0$, there is an N such that for all $n > N$

$$|\overline{X}_n - \mu| \leq \epsilon$$

Proof: As we did in the proof of the central limit theorem, we begin by assuming that the mean of the random variables is zero. That is

$$E[X_i] = 0$$

Now we let

$$S_n = X_1 + \ldots + X_n$$

and consider

$$E[S_n^4] = E[(X_1 + \ldots + X_n)^4]$$

Expanding the quartic results in terms of the form

$$X_i^4, \quad X_i^3 X_j, \quad X_i^2 X_j^2, \quad X_i^2 X_j X_k, \quad X_i X_k X_j X_\ell$$

However, we know that they are independent and have a mean of zero. Therefore, lots of the terms are zero

$$E[X_i^3 X_j] = E[X_i^3]E[X_j] = 0$$

$$E[X_i^2 X_j X_k] = E[X_i^2]E[X_i]E[X_j] = 0$$

$$E[X_i X_k X_j X_\ell] = E[X_i]E[X_k]E[X_j]E[X_\ell] = 0$$

Therefore, the only terms that remain are those of the form

$$X_i^4, \quad X_i^2 X_j^2$$

Obviously, there are exactly n terms of the form X_i^4. As for the other, we note that for each pair of i and j, $\binom{4}{2} = 6$ ways to choose i two times from the four. Thus, each $X_i^2 X_j^2$ will appear 6 times. Also, there are $\binom{n}{2}$ ways to choose 2 objects from n, so the expansion is

$$E[S_n^4] = nE[X_i^4] + 6\binom{n}{2}E[X_i^2]E[X_j]^2 = nK + 3n(n-1)(E[X_i^2])^2 \tag{1}$$

We have

$$0 \leq \text{Var}(X_i^2) = E[X_i^4] - (E[X_i^2])^2$$

and so

$$(E[X_i^2])^2 \leq E[X_i^4] = K$$

Therefore, by (1) we get

$$E[S_n^4] \leq nK + 3n(n-1)K$$

or

$$E\left[\frac{S_n^4}{n^4}\right] \leq \frac{K}{n^3} + \frac{3K}{n^2}$$

Hence,

$$E\left[\sum_{n=1}^{\infty}\frac{S_n^4}{n^4}\right] = \sum_{i=1}^{\infty}E\left[\frac{S_n^4}{n^4}\right] \leq \sum_{n=1}^{\infty}\left(\frac{K}{n^3} + \frac{3K}{n^2}\right) < \infty$$

Could

$$\sum_{n=1}^{\infty}\frac{S_n^4}{n^4} = \infty$$

with probability $p > 0$? If so, then the expected value would be infinite, which we know it isn't. Therefore, with a probability of one we have that

$$\sum_{n=1}^{\infty}\frac{S_n^4}{n^4} < \infty$$

But then, with a probability of one,

$$\frac{S_n^4}{n^4} \to 0 \quad \text{as} \quad n \to \infty \qquad \text{or} \qquad \frac{S_n}{n} \to 0 \quad \text{as} \quad n \to \infty$$

which is precisely what we wanted to prove.

In the case when $\mu \neq 0$, we apply the preceding argument to $X_i - \mu$ to show that with a probability of one we have

$$\lim_{n\to\infty}\sum_{i=1}^{n}\frac{X_i - \mu}{n} = 0$$

or, equivalently,

$$\lim_{n\to\infty}\sum_{i=1}^{n}\frac{X_i}{n} = \mu$$

which proves the result. ■