en time averages and expected ıg modeled.

NCTIONS: We next show that y of an event as well as to the *indicator function* of A, to be a points in A and has the value 0 useful because they allow us to ıbles to events. For the situation being independently repeated n ıt of events in each other repeti- lative frequency of A over the n 's divided by n. This is best ex-

$$\frac{\sum_{i=1}^{n} I_{A_i}}{n} \qquad (28)$$

e indicator functions, and, from

$$\left| \geq \varepsilon \right) \leq \frac{\sigma^2}{n\varepsilon^2} \qquad (29)$$

obability theory for real world obability of an event A (theory) akes an analogy between inde- iments (real world). It then ap- ility and relative frequency.

y call "FOUL," because of the real world must "act like" the periments, and we cannot pre-

integrated circuits, the relative erent integrated circuits in the , at different times of the day, ıdependent and identically dis- ocesses and the theory.

ive frequency *within the theory* es a type of consistency, within

the theory, between relative frequency and probability. This is all that can be expected. Theorems establish rigorous results within a model of reality, but cannot prove things about the real world itself. These theorems, however, provide us with the framework to understand and interpret the behavior of the real world. They both allow us to improve our models, and to predict future behavior to the extent that the models reflect reality.

WEAK LAW WITH INFINITE VARIANCE: We now establish the law of large numbers without assuming a finite variance.

THEOREM 1: WEAK LAW OF LARGE NUMBERS: Let $S_n = X_1+...+X_n$ where $X_1, X_2, \ldots$ are IID random variables with a finite mean E[X]. Then for any $\varepsilon>0$,

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - E[X]\right| \geq \varepsilon\right) = 0 \qquad (30)$$

Proof*[6]: We use a truncation argument; such arguments are used frequently in dealing with random variables that have infinite variance. Let b be a real number (which we later take to be increasing with n), and for each variable X_i, define a new random variable $\tilde{X}_i$ (see figure 1.9) by

$$\begin{aligned} \tilde{X}_i &= X_i & &\text{if } |X_i - E[X]| \leq b \\ \tilde{X}_i &= E[X] + b & &\text{if } X_i - E[X] \geq b \\ \tilde{X}_i &= E[X] - b & &\text{if } X_i - E[X] \leq -b \end{aligned} \qquad (31)$$

The variables $\tilde{X}_i$ are IID and we let $E[\tilde{X}]$ be the mean of $\tilde{X}_i$. As shown in exercise 1.10, the variance of $\tilde{X}$ can be upper bounded by the second moment around any other value, so $\sigma_{\tilde{X}}^2 \leq E[(\tilde{X} - E[X])^2]$. This can be further upper bounded by

$$\sigma_{\tilde{X}}^2 \leq E\left[\left(\tilde{X}-E[X]\right)^2\right] = \int_{-\infty}^{\infty} (x-E[X])^2 dF_{\tilde{X}}(x) \leq b\int_{-\infty}^{\infty} \left|x - E[X]\right| dF_{\tilde{X}}(x)$$

The last inequality follows since $|x-E[X]| \leq b$ over the range of $\tilde{X}$. We next use the fact that $F_{\tilde{X}}(x) = F_X(x)$ for $E[X]-b < x < E[X]+b$ to upper bound the final integral.

$$\sigma_{\tilde{X}}^2 \leq b\int_{-\infty}^{\infty} \left|x-E[X]\right| dF_X(x) = b\,\alpha \quad \text{where } \alpha = \int_{-\infty}^{\infty} \left|x-E[X]\right| dF_X(x) \qquad (32)$$

The quantity α in (32) is the mean of $|X-E[X]|$ and must exist since we assume that X has a mean (see example 6). Now, letting $\tilde{S}_n = \tilde{X}_1+...+\tilde{X}_n$, (and using $\varepsilon/2$ in place of ε), (25) becomes

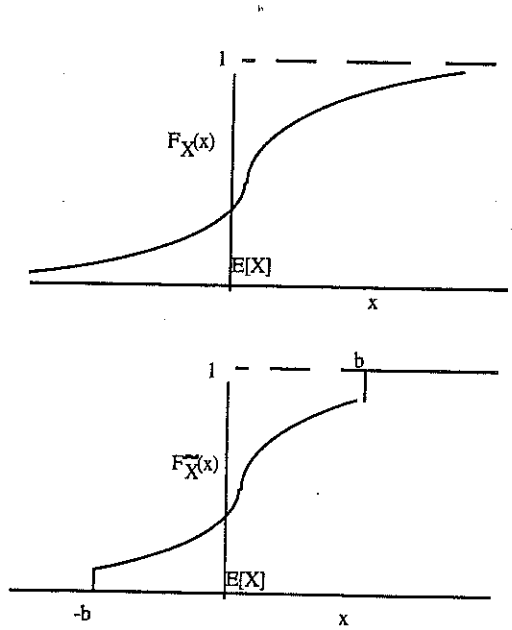

Figure 1.9. Truncated variable $\tilde{X}$.

$$P\left(\left|\frac{\tilde{S}_n}{n} - E[\tilde{X}]\right| \geq \frac{\varepsilon}{2}\right) \leq \frac{4\,\sigma_{\tilde{X}}^2}{n\,\varepsilon^2} \leq \frac{4b\alpha}{n\varepsilon^2}$$

As b increases, $E[\tilde{X}]$ approaches E[X]. Thus for sufficiently large b, $|E[\tilde{X}] - E[X]| < \varepsilon/2$ and

$$P\left(\left|\frac{\tilde{S}_n}{n} - E[X]\right| \geq \varepsilon\right) \leq \frac{4\,b\,\alpha}{n\,\varepsilon^2} \tag{33}$$

Now $\tilde{S}_n$ and S_n have the same value for sample points where $|X_i - E[X]| \leq b$ for all i, $1 \leq i \leq n$. Thus, using the union bound (which says that the probability of a union of events is less than or equal to the sum of the probabilities of the individual events),

$$P(\tilde{S}_n \neq S_n) \leq n\,P\left(\left|X - E[X]\right| > b\right) \tag{34}$$

The event $\{|(S_n/n) - E[X]| \geq \varepsilon\}$ can only occur if either $|(\tilde{S}_n/n) - E[X]| \geq \varepsilon$ or if $\tilde{S}_n \neq S_n$. Thus, combining (33) and (34), and letting $\delta = b/n$, we have

$$P\left(\left|\frac{S_n}{n} - E[X]\right| \geq \varepsilon\right) \leq \frac{4\delta\alpha}{\varepsilon^2} + \frac{1}{\delta}\left[\delta n\, P\left(\left|X\text{-}E[X]\right| > \delta n\right)\right] \tag{35}$$

Since (33) and (34) are valid for arbitrary n and sufficiently large $b>0$, (35) is valid for arbitrary $\delta>0$ and sufficiently large n. For any given $\varepsilon>0$, we now choose δ to make the first term on the right of (35) as small as desired. From (20), the final term in brackets in (35) can be made arbitrarily small by choosing n large enough, and thus the right hand side of (35) can be made arbitrarily small by choosing n large enough, thus completing the proof.

EXAMPLE 6: The Cauchy random variable Z has the probability density $f_Z(z) = 1/[\pi(1+z^2)]$. The mean of Z does not exist, and Z has the very peculiar property that $[Z_1+Z_2+...+Z_n]/n$ has the same density as Z for all n. Thus the law of large numbers does not hold for the Cauchy distribution, which is not surprising since the mean doesn't exist. Recall that the mean of a random variable exists only if

$$\int_{-\infty}^{0} x\,dF_X(x) > -\infty \quad \text{and} \quad \int_{0}^{\infty} x\,dF_X(x) < \infty, \quad \text{or equivalently,} \tag{36}$$

$$\int_{-\infty}^{\infty} |x|\,dF_X(x) < \infty$$

From symmetry, we note that, for the Cauchy distribution, the integral $\int_{z=-b}^{b} dF_Z(z)$ is zero for all b, and thus the integral exists in the Cauchy principal value sense, but not in the ordinary sense of (36). In this text, the existence of integrals always refers to existence in the ordinary rather than Cauchy principal value sense.

1.8 STRONG LAW OF LARGE NUMBERS

We next discuss the strong law of large numbers. We will not prove this result here, but will prove a slightly weaker form of it after discussing martingales in Chapter 7.

THEOREM 2: STRONG LAW OF LARGE NUMBERS (Version 1): Let $S_n = X_1+...+X_n$ where $X_1, X_2, ...$ are IID random variables with a finite mean $\bar{X}$. Then for any $\varepsilon>0$,

$$\lim_{n\to\infty} P\left(\sup_{m\geq n}\left|\frac{S_m}{m} - \bar{X}\right| > \varepsilon\right) = 0 \tag{37}$$

The notation *sup* above stands for supremum (see note 2). The supremum of a set $\{Y_i; i\geq 1\}$ of random variables is a random variable. For each sample point ω in the sample

up$_i$ $Y_i(\omega)$ is the supremum of
that maps each sample point

$> \varepsilon\}$ as the event that $|S_m/m$ -
n states that the probability of
em states that, for large n, it is
eeds ε. An alternative way to
each δ>0, there is an integer

δ (38)

increasing n) for S_n/n to devi-
that the *sequence* S_1, $S_2/2$, $S_3/3$,
y term in the sequence beyond
ng example illustrates the dif-
his example, the random vari-
ited. Since both the strong and
difference for the IID case.

ent random variables for which
(see figure 1.10). Then $E[S_n]$
preted as a sum of dependent
$_{i-1}$, but the argument really just
Markov inequality shows that
see that $\sup_{m\geq n} S_m/m$ takes on
$S_m/m=0$ for all m≥n. This is an
..19) can be seen to be zero for

$\frac{S_m}{m} > \varepsilon\Big) = 1$

n fact likely to be 0), but that if
as a value exceeding ε (and in

One sees from this example that the strong law is saying something about a sample outcome of an infinite sequence of random variables. If one views X_i as a time sequence, then the sample output from the sequence of sample averages S_n/n can be viewed as a sequence of more and more elaborate attempts to estimate the mean. There is clearly some advantage to being able to say that this sequence of attempts not only gets close to the mean with high probability but also stays close to the mean.

Despite the above rationalization, the difference between the strong and weak law almost appears to be mathematical nit picking. On the other hand, we shall discover, as we use these results, that the strong law is often much easier to use than the weak law. The useful form of the strong law, however, is the following theorem. The statement of this theorem is deceptively simple, and it will take some care to understand what the theorem is saying.

THEOREM 3: STRONG LAW OF LARGE NUMBERS (Version 2): Let $S_n = X_1+...+X_n$ where $X_1, X_2,...$ are IID random variables with a finite mean $\overline{X}$. Then with probability 1,

$$\lim_{n\to\infty} \frac{S_n}{n} = \overline{X} \tag{39}$$

For each sample point ω, $S_n(\omega)/n$ is a sequence of real numbers that might or might not have a limit. If this limit exists for all sample points, then $\lim_{n\to\infty} S_n/n$ is a random variable that maps each sample point ω into $\lim_{n\to\infty} S_n(\omega)/n$. Usually this limit does not exist for all sample points, but the theorem implicitly asserts that the limit does exist for all sample points except a set of probability 0. Thus $\lim_{n\to\infty} S_n/n$ is still regarded as a random variable. The theorem asserts not only that $\lim_{n\to\infty} S_n(\omega)/n$ exists for all sample points except a set of zero probability, but also asserts that the limit is equal to $\overline{X}$ for all sample points except a set of probability 0. A sequence of random variables S_n/n that converges in the sense of (39) is said to *converge with probability 1*.

EXAMPLE 8: Suppose the X_i are Bernoulli with equiprobable ones and zeros. Then $\overline{X}$ = 1/2. We can easily construct sequences for which the sample average is not 1/2; for example the sequence of all zeros, the sequence of all ones, sequences with 1/3 zeros and 2/3 ones, and so forth. The theorem says, however, that collectively those sequences have zero probability.

Proof of theorem 3: We assume theorem 2 (which we do not prove until Chapter 7) in order to prove theorem 3. Consider the event illustrated in figure 1.11 in which tighter and tighter bounds are placed on successive elements of the sequence $\{S_n/n;\ n\geq 1\}$. In particular, for some increasing set of positive integers $n_1, n_2, ...$, we consider the bound $|S_n/n - \overline{X}| \leq 2^{-k}$ for $n_k \leq n < n_{k+1}$. For any sample point ω, if $\{S_n(\omega);\ n\geq 1\}$ satisfies all these constraints, then $\lim_{n\to\infty} S_n(\omega)/n = \overline{X}$. The probability of the complementary set of sample points for which one of these bounds is *unsatisfied* is given by

$$P\left\{\cup_{k\ge1}\left[\cup_{n_k\le n<n_{k+1}}\left(\left|\frac{S_n}{n}-\bar{X}\right|>2^{-k}\right)\right]\right\} = P\left\{\cup_{k\ge1}\left[\cup_{n_k\le n}\left(\left|\frac{S_n}{n}-\bar{X}\right|>2^{-k}\right)\right]\right\} \qquad (40)$$

$$\le \sum_{k=1}^{\infty} P\left[\cup_{n\ge n_k}\left(\left|\frac{S_n}{n}-\bar{X}\right|>2^{-k}\right)\right] \qquad (41)$$

$$= \sum_{k=1}^{\infty} P\left(\sup_{n\ge n_k}\left|\frac{S_n}{n}-\bar{X}\right|>2^{-k}\right) \qquad (42)$$

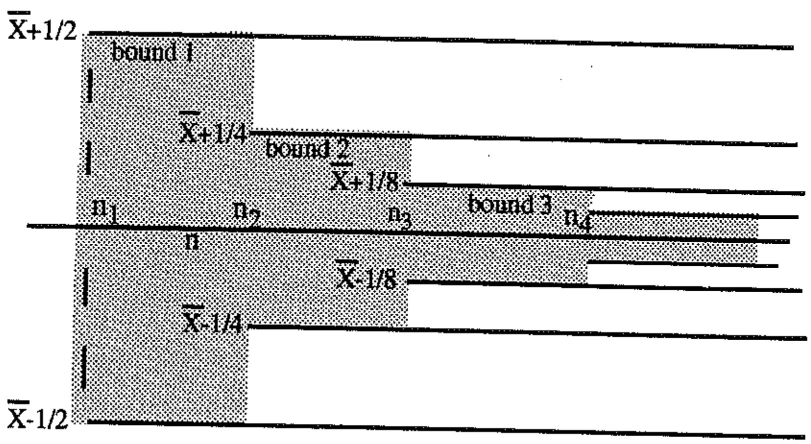

Figure 1.11. Illustration of the union of events in (40); the k[th] sub-event in (40) is the set of sample points for which S_n/n falls outside of the k[th] bound for some n, $n_k \le n < n_{k+1}$; i.e., for which $|S_n/n - \bar{X}| > 2^{-k}$ for some $n_k \le n < n_{k+1}$.

The first equality is most easily visualized in figure 1.11; if $|S_n/n - \bar{X}| > 2^{-k}$ for one value of k, then $|S_n/n - \bar{X}| > 2^{-k'}$ for all k'≥k. In going from (40) to (41), we have used the union bound; this says that the probability of a union of events is less than or equal to the sum of the probabilities of the individual events. Finally (42) follows because the supremum of a sequence exceeds 2^{-k} if and only if one of the elements exceeds 2^{-k}.

From (38), for any ε,δ>0, there is an n(ε,δ) such that $P(\sup_{n\ge n(\varepsilon,\delta)}|S_n/n - \bar{X}|>\varepsilon) \le \delta$. For given δ_0, We then choose n_k in the bound above as $n_k = n(2^{-k}, \delta_0 2^{-k})$, i.e., so that $P(\sup_{n\ge n_k} |S_m/m - \bar{X}| > 2^{-k}) \le \delta_0 2^{-k}$. Substituting this in (42), we have

$$\sum_{k=1}^{\infty} P\left(\sup_{n\ge n_k}\left|\frac{S_n}{n}-\bar{X}\right|>2^{-k}\right) \le \sum_{k=1}^{\infty} \delta_0 2^{-k} = \delta_0 \qquad (43)$$

It follows that the set of samp
ability at least $1-\delta_0$, and as we ha
points. Since this is true for any δ
$= \bar{X}$ must have probability 1, con

Note that as δ_0 is decreased,
the set of sample points that fall w
verges very slowly to $\bar{X}$ for a gi
$\{S_n(\omega)/n; n\ge1\}$ to stay within the

1.9 SUMMARY

This chapter has provided a brief r
the basic ingredients of sample sp
to random variables, and then to
understanding the underlying str

TABLE OF STANDARD RANI

The following table summarizes t
density or PMF is specified only

Name	*Density or PMF*
(Continuous rv, $f_X(x)$)	
Exponential	$\lambda\exp(-\lambda x)$; $x\ge0$
Erlang	$\frac{\lambda^n x^{n-1}\exp(-\lambda x)}{(n-1)!}$
Gaussian	$\frac{\exp(-x^2/2\sigma^2)}{\sqrt{2\pi}\,\sigma}$
Uniform	$\frac{1}{a}$; $0\le x\le a$
(Integer rv, $P_N(n)$)	
Bernoulli	$P_N(0)=1-p$; $P_N(1)$
Binomial	$\frac{k!}{n!(k-n)!}p^n(1-p)$
Geometric	$(1-p)p^n$; $n\ge0$
Poisson	$\frac{\lambda^n\exp(-\lambda)}{n!}$; $n\ge$