Topics in
Probability Theory and Stochastic Processes
Steven R. Dunbar

Large Deviations

Rating

Mathematicians Only: prolonged scenes of intense rigor.
Section Starter Question

From the proof of the Weak Law of Large Numbers, we know
\[
P_n \left[ \left| \frac{S_n}{n} - p \right| > \epsilon \right] < \frac{p(1-p)}{n\epsilon^2}
\]
or more roughly, the deviation of the sample mean of Bernoulli trials from \( p \) is \( O(1/n) \). Is it possible to do provide a more refined estimate of the probability of such a deviation? Why do you think so? What would be the possible statement of a better result?

---

Key Concepts

1. Large Deviations Theory is about the remote tail of a probability distribution. Roughly speaking, large deviation theory estimates the exponential decay of the probability of extreme events, as the number of observations grows arbitrarily large.

2. For \( 0 < \epsilon < 1 - p \) and \( n \geq 1 \),
\[
P_n \left[ \frac{S_n}{n} > p + \epsilon \right] \leq e^{-nh_+(\epsilon)}
\]

where \( h_+(\epsilon) = (p + \epsilon) \ln \left( \frac{p+\epsilon}{p} \right) + (1 - p - \epsilon) \ln \left( \frac{1-p-\epsilon}{1-p} \right) \)
Vocabulary

1. **Large Deviations Theory** is about the remote tail of a probability distribution. Roughly speaking, large deviation theory estimates the exponential decay of the probability of extreme events, as the number of observations grows arbitrarily large.

Mathematical Ideas

The Large Deviations Estimate

**Large Deviations Theory** is about the remote tail of a probability distribution. Roughly speaking, large deviation theory estimates the exponential decay of the probability of extreme events, as the number of observations grows arbitrarily large. Large deviations theory was perhaps discovered by Swedish mathematician Harald Cramér who applied it to model the insurance business. More general results were later obtained by H. Chernoff, among other people. An incomplete list of mathematicians who have made important advances would include S. S. R. Varadhan, D. Ruelle and O. E. Lanford.

Given parameter value $p$, define $h_+(\epsilon)$ for $0 < \epsilon < 1 - p$ as a function of $\epsilon$

$$h_+(\epsilon) = (p + \epsilon) \ln \left( \frac{p + \epsilon}{p} \right) + (1 - p - \epsilon) \ln \left( \frac{1 - p - \epsilon}{1 - p} \right).$$

See Figure 1.

The first two lemmas are Taylor expansions of $h_+(\epsilon)$ and $e^{-nh_+(\epsilon)}$ in $\epsilon$ to give a sense of the asymptotics of these functions.

**Lemma 1.**

$$h_+(\epsilon) = \frac{1}{2p(1 - p)} \epsilon^2 + \frac{2p - 1}{6p^2(1 - p)^2} \epsilon^3 + \frac{1 - 3p + 3p^2}{12p^3(1 - p)^3} \epsilon^4 + O(\epsilon^5)$$
See also Figure 1.

Lemma 2.

\[ e^{-nh_+(\epsilon)} = 1 - \frac{n}{2p(1-p)} \epsilon^2 + \frac{n(2p - 1)}{6p^2(1-p)^2} \epsilon^3 + \frac{2n - 6np + 6np^2 - 3n^2p + 3n^2p^2}{24p^3(1-p)^3} \epsilon^4 + O(\epsilon^5) \]

See also Figure 2 which displays \( e^{-nh_+(\epsilon)} \) as a function of \( \epsilon \) for \( p = 0.6 \) and \( n = 10, 20, 30 \).

Theorem 3 (Large Deviations Estimate). For \( 0 < \epsilon < 1-p \) and \( n \geq 1 \),

\[ \mathbb{P}_n \left[ \frac{S_n}{n} \geq p + \epsilon \right] \leq e^{-nh_+(\epsilon)}. \]

Proof. Fix \( t > 0 \) and then

\[ \mathbb{P}_n \left[ \frac{S_n}{n} \geq p + \epsilon \right] = \mathbb{P}_n \left[ e^{t(S_n - np - \epsilon n)} \geq 1 \right]. \]
Now apply Markov’s Inequality to this positive random variable:

\[
P_n \left[ e^{t(S_n - np - ne)} \geq 1 \right] \leq \mathbb{E} \left[ e^{t(S_n - np - ne)} \right] = e^{-nt(p+\epsilon)} \mathbb{E} \left[ e^{tS_n} \right] = e^{-nt(p+\epsilon)} \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k (1-p)^{n-k} = e^{-nt(p+\epsilon)} (1-p+pe^t)^n = e^{-n(t(p+\epsilon)-\ln(1-p+pe^t))}.
\]

Now to complete the estimation, we only need to find

\[
\sup_{t>0} (t(p+\epsilon) - \ln(1-p+pe^t)).
\]

For notation, let \( g(t) = (t(p+\epsilon) - \ln(1-p+pe^t)) \) for parameters \( 0 < p < 1 \) and \( 0 < \epsilon < 1 - p \). Then \( g(0) = 0 \) and \( g'(t) = p + \epsilon - pe^t/(1-p+pe^t) \), so \( g'(0) = \epsilon \). Furthermore, \( \lim_{t \to \infty} g'(t) = p + \epsilon - 1 < 0 \). Thus the supremum is attained at some strictly positive value. The derivative is zero only at

\[
t = \ln \left( \frac{-p + p^2 - \epsilon + \epsilon p}{p(p + \epsilon - 1)} \right) = \ln \left( \frac{(p + \epsilon)(1-p)}{p(1-p-\epsilon)} \right).
\]

Therefore, \( g(t) \) has a maximum value of \( h_+(\epsilon) \) since by plugging in the
critical point, we obtain:

\[
(p + \epsilon) \ln \left(\frac{p + \epsilon}{p}\right) + (p + \epsilon) \ln \left(\frac{1 - p}{1 - p - \epsilon}\right) - \ln \left(1 - p + p \frac{(p + \epsilon)(1 - p)}{p(1 - p - \epsilon)}\right) \\
= (p + \epsilon) \ln \left(\frac{p + \epsilon}{p}\right) + (p + \epsilon) \ln \left(\frac{1 - p}{1 - p - \epsilon}\right) - \ln \left(1 - p + \frac{(p + \epsilon)}{(1 - p - \epsilon)}\right) \\
= (p + \epsilon) \ln \left(\frac{p + \epsilon}{p}\right) + (p + \epsilon) \ln \left(\frac{1 - p}{1 - p - \epsilon}\right) - \ln \left(\frac{1 - p}{1 - p - \epsilon}\right) \\
= h_+ (\epsilon)
\]

Define \(h_-(\epsilon) = h_+ (-\epsilon)\) for \(0 < \epsilon < p\).

**Corollary 1.** For \(0 < \epsilon < p\) and \(n \geq 1\),

\[
\mathbb{P}_n \left[ \frac{S_n}{n} \leq p - \epsilon \right] \leq e^{-nh_-(\epsilon)}.
\]

**Proof.** Interchange the designation of success and failure, so now “success” occurs with probability \(1 - p\) and “failure” occurs with probability \(p\). Consider the probability that the proportion of “successes” in this complementary process, \(S_n^c / n\) exceeds the mean by \(\epsilon\):

\[
\mathbb{P}_n \left[ \frac{S_n^c}{n} > 1 - p + \epsilon \right] \leq e^{-nh^c_+ (\epsilon)}
\]

where \(h^c_+ (\epsilon) = (1 - p + \epsilon) \ln \left(\frac{1 - p + \epsilon}{1 - p}\right) + (p - \epsilon) \ln \left(\frac{p - \epsilon}{p}\right) = h_+ (-\epsilon) = h_-(\epsilon)\).

The inequality is then

\[
\mathbb{P}_n \left[ 1 - \frac{S_n^c}{n} < p - \epsilon \right] = \mathbb{P}_n \left[ \frac{S_n}{n} < p - \epsilon \right] \leq e^{-nh_-(\epsilon)}.
\]

**Corollary 2.** For \(0 < \epsilon < \min(p, 1 - p)\) and \(n \geq 1\),

\[
\mathbb{P}_n \left[ \left| \frac{S_n}{n} - p \right| \geq \epsilon \right] \leq e^{-nh_+(\epsilon)} + e^{-nh_-(\epsilon)}.
\]

See Figure 3. Note that this estimate is correct but empty for \(\epsilon = 0\) and a neighborhood of \(\epsilon = 0\).
The Large Deviations Estimate Cannot Be Improved

We will frequently use the following lemma to estimate the value of a binomial probability for large values of $n$ and proportionally large values of $k$.

**Lemma 4 (Binomial Asymptotics).** If $k_n \to \infty$ as $n \to \infty$ then

$$
\mathbb{P}_n [S_n = k_n] \sim \frac{1}{\sqrt{2\pi n}} \frac{n}{k_n(n-k_n)} \left( \frac{np}{k_n} \right)^{k_n} \left( \frac{n(1-p)}{n-k_n} \right)^{n-k_n},
$$

as $n \to \infty$.

**Proof.** We know that

$$
\mathbb{P}_n [S_n = k_n] = \frac{n!}{k_n!(n-k_n)!} p^{k_n} (1-p)^{n-k_n}.
$$

so

$$
1 = \lim_{n \to \infty} \frac{n!}{k_n!(n-k_n)!} p^{k_n} (1-p)^{n-k_n}.
$$

(1)

Use Stirling’s approximation in the form $m! \sim \sqrt{2\pi m^{m+1/2}} e^{-m}$. Find the asymptotics of the binomial probability by replacing each factorial with the Stirling approximation. That is

$$
1 = \lim_{n \to \infty} \sqrt{2\pi n} n^{n+1/2} e^{-n} \frac{n!}{k_n!(n-k_n)!} \left( \frac{np}{k_n} \right)^{k_n} \left( \frac{n(1-p)}{n-k_n} \right)^{n-k_n}
$$

$$
= \lim_{n \to \infty} \sqrt{2\pi n} n^{n+1/2} e^{-n} \frac{n!}{k_n!(n-k_n)!} \left( \frac{np}{k_n} \right)^{k_n} \left( \frac{n(1-p)}{n-k_n} \right)^{n-k_n}
$$

(2)
and
\[ 1 = \lim_{n \to \infty} \frac{(n - k_n)!}{\sqrt{2\pi(n - k_n)(n - k_n) + 1/2}e^{-(n - k_n)}} \]
\[ = \lim_{n \to \infty} \frac{(n - k_n)!}{\sqrt{2\pi \sqrt{n - k_n(n - k_n) + e^{-n + k_n}}}} \quad (3) \]

and
\[ 1 = \lim_{n \to \infty} \frac{(k_n)!}{\sqrt{2\pi(k_n)k_n + 1/2}e^{k_n}} \]
\[ = \lim_{n \to \infty} \frac{k_n!}{\sqrt{2\pi \sqrt{k_n(k_n - n - k_n) + e^{-k_n}}}}. \quad (4) \]

Then multiply the terms on the left and right sides of equations (1), (2), (3), and (4) and cancel the factorials, two of the \(\sqrt{2\pi}\) terms, the exponentials and combine the powers to obtain:
\[
\mathbb{P}_n \left[ S_n = k_n \right] \sim \frac{1}{\sqrt{2\pi}} \frac{n}{k_n(n - k_n)} \left( \frac{np}{k_n} \right)^{k_n} \left( \frac{n(1 - p)}{n - k_n} \right)^{n - k_n}.
\]

The estimate given by the Large Deviations result is as good as it can be in the following sense:

**Theorem 5.** For all \(0 < \epsilon < 1 - p\),
\[
\lim_{n \to \infty} \frac{1}{n} \ln \left( \mathbb{P}_n \left[ S_n \geq p + \epsilon \right] \right) = -h_+(\epsilon).
\]

**Proof.** First, the Large Deviations Estimate can be written as
\[
\frac{1}{n} \ln \left( \mathbb{P}_n \left[ S_n \geq p + \epsilon \right] \right) \leq -h_+(\epsilon)
\]
so immediately
\[
\limsup_{n \to \infty} \frac{1}{n} \ln \left( \mathbb{P}_n \left[ S_n \geq p + \epsilon \right] \right) \leq -h_+(\epsilon).
\]
Let $k_n = \lceil n(p + \epsilon) \rceil$ be the least integer greater than or equal to $n(p + \epsilon)$.

Then
\[ k_n - 1 < n(p + \epsilon) \leq k_n. \]

Therefore, $k_n \to \infty$ as $n \to \infty$. Also $k_n \sim n(p + \epsilon)$.

Note that
\[ \mathbb{P}_n [S_n \geq n(p + \epsilon)] \geq \mathbb{P}_n [S_n = k_n]. \]

We will show that
\[ \lim_{n \to \infty} \frac{1}{n} \ln(\mathbb{P}_n [S_n = k_n]) = -h_+(\epsilon). \]

Then
\[ \lim \inf_{n \to \infty} \frac{1}{n} \ln(\mathbb{P}_n [S_n \geq n(p + \epsilon)]) \geq -h_+(\epsilon) \]
and we will have established the theorem.

We already know from the binomial probability
\[ \mathbb{P}_n [S_n = k_n] \sim \frac{n!}{k_n!(n-k_n)!} p^{k_n} (1-p)^{n-k_n}. \tag{5} \]

Now $k_n$ goes to $+\infty$ as $n \to \infty$. Note that the probability on the left side goes to 0 as $n \to \infty$ by the Weak Law of Large of Numbers. By the Zero Asymptotic Limits Lemma in [Asymptotic Limits], the right side must also go to 0 as $n \to \infty$. Since the two sequences are asymptotic, by the Logarithm Lemma in [Asymptotic Limits] their logarithms must be asymptotic.

Since $k_n \sim n(p + \epsilon)$ and $n - k_n \sim n(1 - p - \epsilon)$, then
\[ \sqrt{\frac{n}{k_n(n-k_n)}} \sim \frac{1}{\sqrt{(p+\epsilon)(1-p-\epsilon)n}} \]
so
\[ \lim_{n \to \infty} \frac{1}{n} \ln \frac{1}{\sqrt{2\pi} \sqrt{\frac{n}{k_n(n-k_n)}}} = 0. \]

Next,\[
k_n \ln \left( \frac{np}{k_n} \right) = n(p+\epsilon) \ln \left( \frac{p}{p+\epsilon} \right) + n(p+\epsilon) \ln \left( \frac{n(p+\epsilon)}{k_n} \right) + (k_n - n(p+\epsilon)) \ln \left( \frac{np}{k_n} \right),\]
Now recall that $k_n - n(p + \epsilon)$ is bounded by 1, so
\[ \lim_{n \to \infty} \frac{1}{n} \ln \left( \left( \frac{np}{k_n} \right)^{k_n} \right) = (p + \epsilon) \ln \left( \frac{p}{p+\epsilon} \right). \tag{6} \]
Using the same basic calculations, we can show

\[
\lim_{n \to \infty} \frac{1}{n} \ln \left( \left( \frac{n(1 - p)}{n - k_n} \right)^{n-k_n} \right) = (1 - p - \epsilon) \ln \left( \frac{1 - p}{1 - p - \epsilon} \right).
\]  

(7)

Then in the binomial probability (5) replace the powers with the asymptotic approximations (6) and (7) and the theorem is established.

Sources

This section is adapted from: Heads or Tails, by Emmanuel Lesigne, Student Mathematical Library Volume 28, American Mathematical Society, Providence, 2005, Chapter 6, [1].

Some history and definitions are also adapted from the Wikipedia article on Large Deviations Theory.

Problems to Work for Understanding

Reading Suggestion:

References

Outside Readings and Links:

1. [Virtual Laboratories in Probability and Statistics & Binomial](#)

2. Wikipedia, [Large deviations theory](#)

I check all the information on each page for correctness and typographical errors. Nevertheless, some errors may occur and I would be grateful if you would alert me to such errors. I make every reasonable effort to present current and accurate information for public use, however I do not guarantee the accuracy or timeliness of information on this website. Your use of the information from this website is strictly voluntary and at your risk.

I have checked the links to external sites for usefulness. Links to external websites are provided as a convenience. I do not endorse, control, monitor, or guarantee the information contained in any external website. I don’t guarantee that the links are active at all times. Use the links here with the same caution as you would all information on the Internet. This website reflects the thoughts, interests and opinions of its author. They do not explicitly represent official positions or policies of my employer.

Information on this website is subject to change without notice.

Steve Dunbar’s Home Page, [http://www.math.unl.edu/~sdunbar1](http://www.math.unl.edu/~sdunbar1)
Email to Steve Dunbar, sdunbar1 at unl dot edu
Last modified: Processed from LATEX source on June 4, 2011