Appendix. Bounding of Large Deviations

We give here some basic bounds on large deviations that are useful when employing the probabilistic method. Our treatment is self-contained. Most of the results may be found in, or immediately derived from, the seminal paper of H. Chernoff (1952). While we are guided by asymptotic considerations the inequalities are proven for all values of the parameters in the specified region. The first result, while specialized, contains basic ideas found throughout the Appendix.

(A.1) **Theorem.** Let $X_i, 1 \leq i \leq n$, be mutually independent random variables with

$$\Pr[X_i = +1] = \Pr[X_i = -1] = \frac{1}{2}$$

and set, following the usual convention,

$$S_n = X_1 + \dots X_n$$

Let $a > 0$. Then

$$\Pr[S_n > a] < e^{-a^2/2n}$$

Remark. For large n the Central Limit Theorem implies that S_n is approximately normal with zero mean and standard deviation $\sqrt{n}$. In particular, for any fixed u

$$\lim_{n\to\infty} \Pr[S_n > u\sqrt{n}] = \int_{t=u}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

which one can show directly is less than $e^{-u^2/2}$. Our proof, we emphasize once again, is valid for all n and all $a > 0$.

We require Markov's inequality which states: Suppose that Y is an arbitrary nonnegative random variable, $\alpha > 0$. Then

$$\Pr[Y > \alpha E[Y]] < 1/\alpha$$

Proof. Fix n, a and let, for the moment, $\lambda > 0$ be arbitrary. For $1 \leq i \leq n$

$$E[e^{\lambda X_i}] = (e^{\lambda} + e^{-\lambda})/2 = \cosh(\lambda)$$

We require the inequality

$$\cosh(\lambda) \leq e^{\lambda^2/2}$$

valid for all $\lambda > 0$, the special case $\alpha = 0$ of Lemma A.5 below. (The inequality may be more easily shown by comparing the Taylor series of the two functions termwise.)

$$e^{\lambda S_n} = \prod_{i=1}^{n} e^{\lambda X_i}$$

Since the X_i are mutually independent so are the $e^{\lambda X_i}$, expectations multiply and

$$E[e^{\lambda S_n}] = \prod_{i=1}^{n} E[e^{\lambda X_i}] = [\cosh(\lambda)]^n < e^{\lambda^2 n/2}$$

We note $S_n > a$ if and only if $e^{\lambda S_n} > e^{\lambda a}$ and apply Markov's inequality so that

$$\Pr[S_n > a] = \Pr[e^{\lambda S_n} > e^{\lambda a}] \leq E[e^{\lambda S_n}]/e^{\lambda a} \leq e^{\lambda^2 n/2 - \lambda a}$$

We set $\lambda = a/n$ to optimize the inequality, $\Pr[S_n > a] < e^{-a^2/2n}$ as claimed.

By symmetry we immediately have

(A.2)**Corollary** Under the assumptions of Theorem A.1

$$\Pr[|S_n| > a] < 2e^{-a^2/2n}$$

Our remaining results will deal with distributions X of the following prescribed type.

(A.3)**Assumptions**

$$p_1, \ldots, p_n \in [0,1]$$

$$p = (p_1 + \ldots + p_n)/n$$

$$X_1, \ldots, X_n \text{ mutually independent with}$$

$$\Pr[X_i = 1 - p_i] = p_i$$

$$\Pr[X_i = -p_i] = 1 - p_i$$

$$X = X_1 + \ldots + X_n$$

Remark. Clearly $E[X] = E[X_i] = 0$. When all $p_i = 1/2$, X has distribution $S_n/2$. When all $p_i = p$, X has distribution $B(n,p) - np$ where $B(n,p)$ is the usual Binomial Distribution.

(A.4)**Theorem**. Under assumptions (A.3) and with $a > 0$

$$\Pr[X > a] < e^{-2a^2/n}$$

(A.5)**Lemma.** For all reals α, β with $|\alpha| \leq 1$

$$\cosh(\beta) + \alpha \sinh(\beta) \leq e^{\beta^2/2+\alpha\beta}$$

Proof. This is immediate if $\alpha = +1$ or $\alpha = -1$ or $|\beta| \geq 100$. If the Lemma were false the function

$$f(\alpha, \beta) = \cosh(\beta) + \alpha \sinh(\beta) - e^{\beta^2/2+\alpha\beta}$$

would assume a negative global minimum in the interior of the rectangle

$$R = \{(\alpha, \beta) : |\alpha| \leq 1, |\beta| \leq 100\}$$

Setting partial derivatives equal to zero we find

$$\sinh(\beta) + \alpha \cosh(\beta) = (\alpha + \beta)e^{\beta^2/2+\alpha\beta}$$

$$\sinh(\beta) = \beta e^{\beta^2/2+\alpha\beta}$$

and thus $\tanh(\beta) = \beta$ which implies $\beta = 0$. But $f(\alpha, 0) = 0$ for all α, a contradiction.

(A.6)**Lemma**. For all $\theta \in [0, 1]$ and all λ

$$\theta e^{\lambda(1-\theta)} + (1 - \theta)e^{-\lambda\theta} \leq e^{\lambda^2/8}$$

Proof. Setting $\theta = (1 + \alpha)/2$ and $\lambda = 2\beta$, Lemma A.6 reduces to Lemma A.5.

Proof of Theorem (A.4). Let, for the moment, $\lambda > 0$ be arbitrary.

$$E[e^{\lambda X_i}] = p_i e^{\lambda(1-p_i)} + (1 - p_i)e^{-\lambda p_i} \leq e^{\lambda^2/8}$$

by Lemma A.6. Then

$$E[e^{\lambda X}] = \prod_{i=1}^{n} E[e^{\lambda X_i}] \leq e^{\lambda^2 n/8}$$

Applying Markov's inequality

$$\Pr[X > a] = \Pr[e^{\lambda X} > e^{\lambda a}] \leq E[e^{\lambda X}]/e^{\lambda a} \leq e^{\lambda^2 n/8-\lambda a}$$

We set $\lambda = 4a/n$ to optimize the inequality: $\Pr[X > a] < e^{-2a^2/n}$ as claimed.

Again by symmetry we immediately have

(A.7)**Corollary**. Under assumptions (A.3) and with $a > 0$

$$\Pr[|X| > a] < 2e^{-2a^2/n}$$

Under assumptions (A.3) with $\lambda > 0$ arbitrary

$$E[e^{\lambda X}] = \prod_{i=1}^{n} E[e^{\lambda X_i}] = \prod_{i=1}^{n}[p_i e^{\lambda(1-p_i)}+(1-p_i)e^{-\lambda p_i}] = e^{-\lambda pn}\prod_{i=1}^{n}[p_i e^{\lambda}+(1-p_i)]$$

With λ fixed the function

$$f(x) = \ln[xe^{\lambda} + 1 - x] = \ln[Bx + 1] \text{ with } B = e^{\lambda} - 1$$

is concave and hence (Jensen's Inequality)

$$\sum_{i=1}^{n} f(p_i) \leq nf(p)$$

Exponentiating both sides

$$\prod_{i=1}^{n}[p_i e^{\lambda} + (1 - p_i)] \leq [pe^{\lambda} + (1 - p)]^n$$

so that
(A.8) **Lemma.** Under the assumptions (A.3)

$$E[e^{\lambda X}] \leq e^{-\lambda pn}[pe^{\lambda} + (1 - p)]^n$$

(A.9) **Theorem.** Under the assumptions (A.3) and with $a > 0$

$$\Pr[X > a] < e^{-\lambda pn}[pe^{\lambda} + (1 - p)]^n e^{-\lambda a}$$

for all $\lambda > 0$.
Proof. $\Pr[X > a] = \Pr[e^{\lambda X} > e^{\lambda a}] \leq E[e^{\lambda X}]/e^{\lambda a}$. Now apply Lemma (A.8).
Remark. For given p, n, a an optimal assignment of λ in Theorem A.7 is found by elementary calculus to be

$$\lambda = \ln[(\frac{1-p}{p})(\frac{a+np}{n-(a+np)})]$$

This value is oftentimes too cumbersome to be useful. We employ suboptimal λ to achieve more convenient results.

Setting $\lambda = \ln[1 + a/pn]$ Theorem (A.9) reduces to
(A.10) **Corollary**

$$\Pr[X > a] < e^{a - pn\ln(1+a/pn) - a\ln(1+a/pn)}$$

(A.11) **Theorem**.

$$\Pr[X > a] < e^{-a^2/2pn + a^3/2(pn)^2}$$

Proof. With $u = a/pn$ apply the inequality

$$\ln(1+u) \geq u - u^2/2,$$

valid for all $u \geq 0$, to Corollary (A.10).

When all $p_i = p$, X has variance $np(1-p)$. With $p = o(1)$ and $a = o(pn)$ this bound reflects the approximation of X by a Normal Distribution with variance $\sim np$. The bound of Theorem (A.11) hits a minimum at $a = 2pn/3$. For $a > 2pn/3$ we have the simple bound

$$\Pr[X > a] \leq \Pr[X > 2pn/3] < e^{-2pn/27}$$

This is improved by the following. (A.12) **Theorem**. For $\beta \geq 1$

$$\Pr[X \geq (\beta - 1)pn] < [e^{\beta-1}\beta^{-\beta}]^{pn}$$

Proof. Direct "plug in" to Corollary (A.10).

$X + pn$ may be interpreted as the number of successes in n independent trials when the probability of success in the i-th trial is p_i.
(A.13) **Theorem**. Under assumptions (A.3) and with $a > 0$

$$\Pr[X < -a] < e^{-a^2/2pn}$$

Note that one cannot simply employ "symmetry" as then the roles of p and $1 - p$ are interchanged.
Proof. Let $\lambda > 0$ be, for the moment, arbitrary. Then

$$E[e^{-\lambda X}] = \prod_{i=1}^{n} E[e^{-\lambda X_i}] = \prod_{i=1}^{n} [p_i e^{-\lambda(1-p_i)} + (1-p_i)e^{\lambda p_i}] = e^{\lambda pn} \prod_{i=1}^{n} [p_i e^{-\lambda} + (1-p_i)]$$

With λ fixed the function

$$f(x) = \ln[xe^{-\lambda} + (1-x)] = \ln[Bx + 1] \text{ with } B = e^{-\lambda} - 1$$

is concave. (That B is here negative is immaterial.) Thus

$$\sum_{i=1}^{n} f(p_i) \leq nf(p)$$

Exponentiating both sides gives

$$E[e^{-\lambda X}] \leq e^{\lambda pn}[pe^{-\lambda} + (1-p)]^n$$

analogous to (A.8). Then

$$\Pr[X < -a] = \Pr[e^{-\lambda X} > e^{\lambda a}] < e^{\lambda pn}[pe^{-\lambda} + (1-p)]^n e^{-\lambda a}$$

analogous to Theorem (A.9). We employ the inequality

$$1 + u \leq e^u,$$

valid for all u, so that

$$pe^{-\lambda} + (1-p) = 1 + (e^{-\lambda} - 1)p < e^{p(e^{-\lambda}-1)}$$

and

$$\Pr[X < -a] \leq e^{\lambda pn + np(e^{-\lambda}-1) - \lambda a} = e^{np(e^{-\lambda}-1+\lambda) - \lambda a}$$

We employ the inequality

$$e^{-\lambda} \leq 1 - \lambda + \lambda^2/2,$$

valid for all $\lambda > 0$. (Note: The analogous inequality $e^{\lambda} \leq 1 + \lambda + \lambda^2/2$ is *not* valid for $\lambda > 0$ and so this method, when applied to $\Pr[X > a]$, requires an "error" term such as found in Theorem (A.11).) Now

$$\Pr[X < -a] \leq e^{np\lambda^2/2 - \lambda a}$$

We set $\lambda = a/np$ to optimize the inequality: $Pr[X < -a] < e^{-a^2/2pn}$ as claimed.

For clarity the following result is often useful.

(A.14) Corollary. Let Y be the sum of mutually independent indicator random variables, $\mu = E[Y]$. For all $\epsilon > 0$

$$\Pr[|Y - \mu| > \epsilon\mu] < 2e^{-c_\epsilon \mu}$$

where $c_\epsilon > 0$ depends only on ϵ.
Proof. Apply (A.12),(A.13) with $Y = X + pn$ and

$$c_\epsilon = \min[-\ln(e^{\epsilon-1}\epsilon^{-\epsilon}), \epsilon^2/2]$$

The asymmetry between $\Pr[X < a]$ and $\Pr[X > a]$ given by Theorems (A.12), (A.13) is real. The estimation of X be a normal distribution with zero mean and variance np is roughly valid for estimating $\Pr[X < a]$ for any a and for estimating $\Pr[X > a]$ while $a = o(np)$. But when a and np are comparable or when $a >> np$ the Poisson "takes over" and $\Pr[X > a]$ *cannot* be accurately estimated by using the normal distribution.

We conclude with two large deviation results involving distributions other than sums of indicator random variables.
(A.15) Theorem. Let P have Poisson distribution with mean μ. For $\beta > 0$

$$\Pr[P \leq \mu(1-\beta)] \leq e^{-\beta^2\mu/2}$$

$$\Pr[P \geq \mu(\beta-1)] \leq \left[e^{\beta-1}\beta^{-\beta}\right]^\mu$$

Proof. For any s

$$\Pr[P = s] = \lim_{n\to\infty} \Pr[B(n, \frac{\mu}{n}) = s]$$

Apply Theorems (A.12), (A.13).
(A.16) Let X_i, $1 \leq i \leq n$, be mutually indpendent with all $E[X_i] = 0$ and all $|X_i| \leq 1$. Set $S = X_1 + \ldots + X_n$. Then

$$\Pr[S > a] < e^{-a^2/2n}$$

Proof. Set, as in the proof of (A.1), $\lambda = a/n$. Set

$$h(x) = \frac{e^\lambda + e^{-\lambda}}{2} + \frac{e^\lambda - e^{-\lambda}}{2}x$$

For $x \in [-1, 1]$, $e^{\lambda x} \leq h(x)$. ($y = h(x)$ is the chord through the points $x = \pm 1$ of the convex curve $y = e^{\lambda x}$.) Thus

$$E[e^{\lambda X_i}] \leq E[h(X_i)] = h(E[X_i]) = h(0) = \cosh\lambda$$

The remainder of the proof follows as in (A.1).