Theorem A.13. *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

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

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

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

Exponentiating both sides gives

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

analogous to Theorem 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

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

valid for all u, so that

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

and

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

We employ the inequality

$$e^{-\lambda} \le 1 - \lambda + \frac{\lambda^2}{2},$$

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

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

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.

Corollary A.14. *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 Theorems A.12 and A.13 with $Y = X + pn$ and

$$c_\epsilon = \min\left[-\ln(e^{\epsilon}(1+\epsilon)^{-(1+\epsilon)}), \frac{\epsilon^2}{2}\right].$$ ∎

The asymmetry between $\Pr[X < a]$ and $\Pr[X > a]$ given by Theorems A.12 and A.13 is real. The estimation of X by 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 $a \gg np$, the Poisson behavior "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.

Theorem A.15. *Let P have Poisson distribution with mean μ. For $\epsilon > 0$,*

$$\Pr[P \le \mu(1-\epsilon)] \le e^{-\epsilon^2\mu/2},$$

$$\Pr[P \ge \mu(1+\epsilon)] \le \left[e^{\epsilon}(1+\epsilon)^{-(1+\epsilon)}\right]^{\mu}.$$

Proof. For any s,

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

Apply Theorems A.12 and A.13. ∎