See "Results formulated", Sect. 2b, 2c.

1 Stirling's formula

The well-known Stirling's formula

$$n! \sim n^n \mathrm{e}^{-n}\sqrt{2\pi n} \quad \text{as } n \to \infty$$

has several proofs. Before giving a (purely analytic) proof I want to show a possible probabilistic intuition behind it.

The standard exponential distribution has the density $f_1(x) = \mathrm{e}^{-x}$ for $x > 0$ (and 0 for $x < 0$), expectation 1 and variance 1. The sum of n independent random variables, distributed as above, has the so-called gamma distribution with the density

$$f_n(x) = \frac{1}{(n-1)!} x^{n-1} \mathrm{e}^{-x}$$

for $x > 0$ (and 0 for $x < 0$), expectation n and variance n (thus, the mean square deviation $\sqrt{n}$). For large n this density should be close to the normal density with expectation n and variance n,

$$g_n(x) = \frac{1}{\sqrt{2\pi n}} \exp\Big(-\frac{(x-n)^2}{2n} \Big) .$$

The special case $f_n(n) \sim g_n(n)$, that is, $\frac{1}{(n-1)!} n^{n-1} \mathrm{e}^{-n} \sim \frac{1}{\sqrt{2\pi n}}$, is equivalent to Stirling's formula (check it).

Alternatively, the linear transformation $y = (x-n)/\sqrt{n}$ leads to the density

$$\tilde f_n(y) = \sqrt{n} f_n(y\sqrt{n} + n) ,$$

which should be close to the standard normal density

$$\tilde g(y) = \sqrt{n} g_n(y\sqrt{n} + n) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-y^2/2} .$$

A straightforward calculation gives $\tilde f_n \to \tilde g$, provided that Stirling's formula is used. Otherwise it gives rather (see Exercise 1.1 below) $c_n \tilde f_n \to \tilde g$, where c_n are defined by

$$\frac{1}{(n-1)!} n^{n-1} \mathrm{e}^{-n} c_n = \frac{1}{\sqrt{2\pi n}} .$$

Taking into account that $\int \tilde f_n(y)\,\mathrm{d}y = 1$ and $\int \tilde g(y)\,\mathrm{d}y = 1$ (Exercises 1.2, 1.3) we conclude (Exercises 1.4, 1.5) that $c_n \to 1$, which proves Stirling's formula.

In fact,

$$\frac{1}{12n+1} \le \ln \frac{n!}{n^n \mathrm{e}^{-n}\sqrt{2\pi n}} \le \frac{1}{12n}\,,$$

but I do not prove it.

1.1 Exercise. Prove that

$$c_n \tilde f_n(y) = \frac{1}{\sqrt{2\pi}} \exp\biggl(-(n-1)\Bigl(\frac{y}{\sqrt n} - \ln\Bigl(1+\frac{y}{\sqrt n}\Bigr)\Bigr) - \frac{y}{\sqrt n} \biggr) \to \tilde g(y)$$

as $n\to\infty$, for all $y\in\mathbb{R}$.

Hint: $\ln\bigl(1+\frac{y}{\sqrt n}\bigr) = \frac{y}{\sqrt n} - \frac{y^2}{2n} + o(\frac1n)$.

1.2 Exercise. Prove that

$$\int \tilde f_n(y)\,\mathrm{d}y = \int f_n(x)\,\mathrm{d}x = 1$$

for all n.

Hint: induction in n; integration by parts.

1.3 Exercise. Prove that

$$\int \tilde g(y)\,\mathrm{d}y = \int g_n(x)\,\mathrm{d}x = 1$$

for all n.

Hint: calculate $\iint \tilde g(y_1)\tilde g(y_2)\,\mathrm{d}y_1\mathrm{d}y_2$ in polar coordinates.

However, the poinwise convergence does not ensure convergence of integrals.

1.4 Exercise. Prove that

$$c_n \int_{-\sqrt n}^{\sqrt n} \tilde f_n(y)\,\mathrm{d}y \to 1 \quad \text{as } n\to\infty\,.$$

Hint: take $\varepsilon>0$ such that $a-\ln(1+a)\ge\varepsilon a^2$ for all $a\in(-1,1)$; apply it to $a=y/\sqrt n$; use Exercises 1.1, 1.3 and the dominated convergence theorem.

1.5 Exercise. Prove that

$$c_n \int_{\sqrt n}^{\infty} \tilde f_n(y)\,\mathrm{d}y \to 0 \quad \text{as } n\to\infty\,.$$

2 Asymptotic normality

Let $n \in \{1,2,\dots\}$ and $k \in \{-n,-n+2,\dots,n\}$. We have

$$\mathbb{P}\big(S_n = k\big) = 2^{-n} \frac{n!}{(\frac{n-k}{2})!(\frac{n+k}{2})!}$$

and

$$n! = n^n \mathrm{e}^{-n}\sqrt{2\pi n}\,\beta(n) = n^{n+0.5}\mathrm{e}^{-n}\sqrt{2\pi}\,\beta(n)\,, \quad \beta(n)\to 1\,.$$

Thus,

$$\begin{aligned}\mathbb{P}\big(S_n = k\big) &= 2^{-n}n^{n+0.5}\Big(\frac{n-k}{2}\Big)^{-(n-k+1)/2}\Big(\frac{n+k}{2}\Big)^{-(n+k+1)/2}\cdot \\ &\quad \cdot \exp\Big(-n+\frac{n-k}{2}+\frac{n+k}{2}\Big)\frac{1}{\sqrt{2\pi}}\frac{\beta(n)}{\beta(\frac{n-k}{2})\beta(\frac{n+k}{2})} = \\ &= \underbrace{2^{-n+(n-k+1)/2+(n+k+1)/2}}_{=2}\cdot\underbrace{n^{n+0.5-(n-k+1)/2-(n+k+1)/2}}_{=1/\sqrt{n}}\cdot \\ &\quad \cdot\Big(1-\frac{k}{n}\Big)^{-(n-k+1)/2}\Big(1+\frac{k}{n}\Big)^{-(n+k+1)/2}\frac{1}{\sqrt{2\pi}}\frac{\beta(n)}{\beta(\frac{n-k}{2})\beta(\frac{n+k}{2})}\,.\end{aligned}$$

The following relations hold as $n\to\infty$ uniformly in k as long as k^2/n is bounded:

$$\begin{gathered}\frac{n\pm k}{2}\to\infty\,; \quad \frac{\beta(n)}{\beta(\frac{n-k}{2})\beta(\frac{n+k}{2})}\to 1\,; \quad \frac{k}{n} = O(1/\sqrt{n}) = o(1)\,; \\ \ln\Big(1\pm\frac{k}{n}\Big) = \pm\frac{k}{n}-\frac{k^2}{2n^2}+o\Big(\frac{k^2}{n^2}\Big)\,; \\ (n\pm k+1)\ln\Big(1\pm\frac{k}{n}\Big) = \pm k+\frac{k^2}{n}-\frac{k^2}{2n}+o(1)\,; \\ \frac12\sum_{\pm}(n\pm k+1)\ln\Big(1\pm\frac{k}{n}\Big) = \frac{k^2}{2n}+o(1)\,; \\ \mathbb{P}\big(S_n=k\big)\sim\frac{2}{\sqrt{2\pi n}}\exp\Big(-\frac{k^2}{2n}\Big)\,,\end{gathered}$$

which proves Prop. 2b1 of "Results formulated".

It follows that

$$\sum_{a\sqrt{n}<k<b\sqrt{n}}\mathbb{P}\big(S_n=k\big) = \big(1+o(1)\big)\sum_{a\sqrt{n}<k<b\sqrt{n}}\frac{2}{\sqrt{2\pi n}}\exp\Big(-\frac{k^2}{2n}\Big) \quad \text{as } n\to\infty$$

whenever $-\infty < a < b < \infty$ (all k are such that $k+n$ is even). However,

$$\frac{2}{\sqrt{n}} \sum_{a\sqrt{n}<k<b\sqrt{n}} \exp\Big(-\frac{k^2}{2n}\Big) \to \int_a^b \mathrm{e}^{-u^2/2}\,\mathrm{d}u \quad \text{as } n\to\infty\,.$$

Theorem 2b2 of "Results formulated" follows easily.

3 Large deviations

It was shown in Sect. 2 that

$$\mathbb{P}\big(S_n=k\big) = \frac{2}{\sqrt{2\pi n}}\Big(1-\frac{k}{n}\Big)^{-(n-k+1)/2}\Big(1+\frac{k}{n}\Big)^{-(n+k+1)/2}\frac{\beta(n)}{\beta(\frac{n-k}{2})\beta(\frac{n+k}{2})}\,,$$

where $\beta(n)\to 1$ as $n\to\infty$. However,

$$\frac{n-k+1}{2}\ln\Big(1-\frac{k}{n}\Big)+\frac{n+k+1}{2}\ln\Big(1+\frac{k}{n}\Big) =$$
$$= \frac{n}{2}\bigg(\Big(1-\frac{k}{n}\Big)\ln\Big(1-\frac{k}{n}\Big)+\Big(1+\frac{k}{n}\Big)\ln\Big(1+\frac{k}{n}\Big)\bigg)+\frac{1}{2}\bigg(\ln\Big(1-\frac{k}{n}\Big)+\ln\Big(1+\frac{k}{n}\Big)\bigg)\,,$$

that is,

$$\Big(1-\frac{k}{n}\Big)^{-(n-k+1)/2}\Big(1+\frac{k}{n}\Big)^{-(n+k+1)/2} = \frac{1}{\sqrt{1-\frac{k^2}{n^2}}}\exp\bigg(-n\gamma\Big(\frac{k}{n}\Big)\bigg)\,,$$

where

$$\gamma(c) = \frac{1}{2}(1+c)\ln(1+c)+\frac{1}{2}(1-c)\ln(1-c)$$

(and $0\ln 0 = 0$, of course). We see that

$$\mathbb{P}\big(S_n=k\big) \sim \frac{2}{\sqrt{2\pi n}}\frac{1}{\sqrt{1-\frac{k^2}{n^2}}}\exp\bigg(-n\gamma\Big(\frac{k}{n}\Big)\bigg) \tag{3.1}$$

as $n-|k|\to\infty$. That is, for every $\varepsilon>0$ there exists $M<\infty$ such that

$$\frac{\mathbb{P}\big(S_n=k\big)}{\frac{2}{\sqrt{2\pi n}}\frac{1}{\sqrt{1-\frac{k^2}{n^2}}}\exp\Big(-n\gamma\big(\frac{k}{n}\big)\Big)} \in [1-\varepsilon,1+\varepsilon]$$

whenever $n-|k|\ge M$.

An explicit dependence between ε and M may be found via the inequality $\ln \beta(n) \in [1/(12n+1), 1/(12n)]$.

It follow that

$$\mathbb{P}\big(S_n = k\big) = \exp\bigg(-n\gamma\Big(\frac{k}{n}\Big) + o(n) \bigg) ,$$

$$\frac{1}{n} \ln \mathbb{P}\big(S_n = k\big) = -\gamma\Big(\frac{k}{n}\Big) + o(1) \tag{3.2}$$

as $n - |k| \to \infty$. In fact, this relation holds as $n \to \infty$, uniformly in $k \in \{-n, -n+2, \dots, n\}$ (but I do not prove it).

It is possible to continue toward $\mathbb{P}\big(S_n \ge k\big)$. However, all that works only for the binomial distribution. Other distributions can be investigated via a more general approach, shown below (on the binomial case, still).

The inequality

$$\mathbb{P}\big(S_n \ge k\big) \le \frac{\mathbb{E}\,\mathrm{e}^{\lambda S_n}}{\mathrm{e}^{\lambda k}} \quad \text{for } \lambda \ge 0 \tag{3.3}$$

is a special case of Markov's inequality, but anyway, is rather evident:

$$\mathbb{E}\,\mathrm{e}^{\lambda S_n} = \sum_j \mathrm{e}^{\lambda j} \mathbb{P}\big(S_n = j\big) \ge \sum_{j \ge k} \mathrm{e}^{\lambda j} \mathbb{P}\big(S_n = j\big) \ge \\ \ge \sum_{j \ge k} \mathrm{e}^{\lambda k} \mathbb{P}\big(S_n = j\big) = \mathrm{e}^{\lambda k} \mathbb{P}\big(S_n \ge k\big) .$$

It holds for all $\lambda \ge 0$, thus,

$$\mathbb{P}\big(S_n \ge k\big) \le \inf_{\lambda \ge 0} \frac{\mathbb{E}\,\mathrm{e}^{\lambda S_n}}{\mathrm{e}^{\lambda k}} .$$

However,

$$\mathbb{E}\,\mathrm{e}^{\lambda S_n} = \mathbb{E}\left(\mathrm{e}^{\lambda X_1} \dots \mathrm{e}^{\lambda X_n}\right) = \left(\mathbb{E}\,\mathrm{e}^{\lambda X_1}\right)^n = \Big(\frac{\mathrm{e}^{-\lambda} + \mathrm{e}^{\lambda}}{2}\Big)^n = \cosh^n \lambda ,$$

thus,

$$\mathbb{P}\big(S_n \ge k\big) \le \inf_{\lambda \in \mathbb{R}} \frac{\cosh^n \lambda}{\mathrm{e}^{\lambda k}} .$$

The function $\lambda \mapsto \mathrm{e}^{-\lambda k} \cosh^n \lambda$ has a single minimum at

$$\lambda = \frac{1}{2} \ln \frac{n+k}{n-k} \tag{3.4}$$

(check it); it appears that

$$\inf_{\lambda\ge 0} \frac{\cosh^n \lambda}{\mathrm{e}^{\lambda k}} = \Big(1-\frac{k}{n}\Big)^{-(n-k)/2} \Big(1+\frac{k}{n}\Big)^{-(n+k)/2} = \mathrm{e}^{-n\gamma(k/n)} \, ,$$

therefore

$$\mathbb{P}\big(S_n \ge k \big) \le \mathrm{e}^{-n\gamma(k/n)} \, .$$

We see that

$$\frac{1}{n} \ln \mathbb{P}\big(S_n \ge k \big) \le -\gamma\Big(\frac{k}{n}\Big) \, .$$

Is it exact? Is there another function $\tilde\gamma > \gamma$ such that $\frac1n \ln \mathbb{P}\big(S_n \ge k \big) \le -\tilde\gamma(k/n)$ for large n? No, γ is optimal. Indeed, (3.2) tells us that $\frac1n \ln \mathbb{P}\big(S_n \ge k \big) \ge \frac1n \ln \mathbb{P}\big(S_n = k \big) = -\gamma(k/n) + o(1)$ as $n - |k| \to \infty$, therefore

$$\frac{1}{n} \ln \mathbb{P}\big(S_n \ge k \big) = -\gamma\Big(\frac{k}{n}\Big) + o(1) \tag{3.5}$$

as $n - |k| \to \infty$. (In fact, as $n \to \infty$.) This is mysterious! The exponential inequality (3.3) is only one among many similar inequalities (for instance, $\mathbb{P}\big(S_n \ge k \big) \le (\mathbb{E}\, S_n^{2m})/k^{2m}$ for all m), however, it gives the exact rate function γ. Can we understand this fact within the general framework (without (3.2))? Yes, we can; see below.

The question is, why the inequality (3.3) is (roughly) tight for some λ. We have

$$1 - \frac{\mathrm{e}^{\lambda k}}{\mathbb{E}\, \mathrm{e}^{\lambda S_n}} \mathbb{P}\big(S_n \ge k \big) = \\ = \sum_{j<k} \frac{\mathrm{e}^{\lambda j}}{\mathbb{E}\, \mathrm{e}^{\lambda S_n}} \mathbb{P}\big(S_n = j \big) + \sum_{j \ge k} \big(1 - \mathrm{e}^{-\lambda(j-k)} \big) \frac{\mathrm{e}^{\lambda j}}{\mathbb{E}\, \mathrm{e}^{\lambda S_n}} \mathbb{P}\big(S_n = j \big) \, ;$$

the question is, why some λ makes both summands small.

The numbers $\frac{\mathrm{e}^{\lambda j}}{\mathbb{E}\, \mathrm{e}^{\lambda S_n}} \mathbb{P}\big(S_n = j \big)$ for $j \in \{-n, -n+2, \dots, n\}$ may be thought of as another probability distribution. Moreover, it is basically binomial! Indeed,

$$\mathrm{e}^{\lambda j} \mathbb{P}\big(S_n = j \big) = \mathrm{e}^{\lambda j} 2^{-n} \frac{n!}{(\frac{n-k}{2})! (\frac{n+k}{2})!} = \\ = \mathrm{const}(n) \cdot \frac{n!}{(\frac{n-k}{2})! (\frac{n+k}{2})!} p^{(n+j)/2} (1-p)^{(n-j)/2} \, ,$$

if p is chosen so that $p^{j/2}(1-p)^{-j/2} = \mathrm{e}^{\lambda j}$, that is,

$$\frac{p}{1-p} = \mathrm{e}^{2\lambda} \, ; \quad p = \frac{\mathrm{e}^{2\lambda}}{1+\mathrm{e}^{2\lambda}} \, ; \quad \lambda = \frac12 \ln \frac{p}{1-p} \, .$$

Therefore (since the sum must be 1…),

$$\frac{\mathrm{e}^{\lambda j}}{\mathbb{E}\,\mathrm{e}^{\lambda S_n}}\mathbb{P}\big(S_n = j\big) = \frac{n!}{(\frac{n-j}{2})!(\frac{n+j}{2})!}p^{(n+j)/2}(1-p)^{(n-j)/2} = \mathbb{P}\big(S_n^{(p)} = j\big)\,,$$

where $S_n^{(p)} = X_1^{(p)} + \dots + X_n^{(p)}$ and $X_1^{(p)}, \dots, X_n^{(p)}$ are independent identically distributed random variables,

$$\mathbb{P}\big(X_1^{(p)} = 1\big) = p\,, \quad \mathbb{P}\big(X_1^{(p)} = -1\big) = 1 - p\,.$$

We get

$$1 - \frac{\mathrm{e}^{\lambda k}}{\mathbb{E}\,\mathrm{e}^{\lambda S_n}}\mathbb{P}\big(S_n \ge k\big) = \\ = \sum_{j<k}\mathbb{P}\big(S_n^{(p)} = j\big) + \sum_{j\ge k}\bigg(1 - \Big(\frac{1-p}{p}\Big)^{(j-k)/2}\bigg)\mathbb{P}\big(S_n^{(p)} = j\big) = \mathbb{E}\,f(S_n^{(p)})\,,$$

where $f : \{-n, -n+2, \dots, n\} \to \mathbb{R}$ is defined by

$$f(j) = \begin{cases} 1 & \text{for } j < k, \\ 1 - \big(\frac{1-p}{p}\big)^{(j-k)/2} & \text{for } j \ge k. \end{cases}$$

The question is, why some p makes $\mathbb{E}\,f(S_n^{(p)})$ small.

The function f vanishes at k and can be small only in a right-side neighborhood of k. On the other hand, $\frac{1}{n}S_n^{(p)}$ is usually close to

$$\mathbb{E}\,\frac{1}{n}S_n^{(p)} = \mathbb{E}\,X_1^{(p)} = 2p - 1$$

by the weak low of large numbers. Choosing p such that

$$2p - 1 = \frac{k}{n}\,, \quad p = \frac{n+k}{2n}\,, \quad \lambda = \frac{1}{2}\ln\frac{n+k}{n-k}$$

(compare it with (3.4)…), we give to $f(S_n^{(p)})$ a good chance to be small.

However, we should not expect too nuch. According to (3.1), $\mathbb{P}\big(S_n = k\big) \ll \mathrm{e}^{-n\gamma(k/n)}$. And do not think that $\mathbb{P}\big(S_n \ge k\big) \gg \mathbb{P}\big(S_n = k\big)$. You see, $\mathbb{P}\big(S_n = k+2\big) = \frac{n-k}{n+k+1}\mathbb{P}\big(S_n = k\big) \approx \frac{1-p}{p}\mathbb{P}\big(S_n = k\big)$; assuming that $\frac{k}{n} \in (0,1)$ is not close to 0 and 1 we observe that also $\mathbb{P}\big(S_n = k+4\big) \approx \frac{1-p}{p}\mathbb{P}\big(S_n = k+2\big)$ and so on, thus, $\mathbb{P}\big(S_n \ge k\big) \approx \frac{p}{2p-1}\mathbb{P}\big(S_n = k\big)$ is not much larger than $\mathbb{P}\big(S_n = k\big)$. It means that the inequality (3.3) (for the optimal λ) is not really tight; rather,

$$\begin{gathered}\mathbb{P}\big(S_n \ge k\big) \ge \frac{\mathbb{E}\,\mathrm{e}^{\lambda S_n}}{\mathrm{e}^{\lambda k}} \mathrm{e}^{-o(n)}\,;\\ \frac{\mathrm{e}^{\lambda k}}{\mathbb{E}\,\mathrm{e}^{\lambda S_n}} \mathbb{P}\big(S_n \ge k\big) \ge \mathrm{e}^{-o(n)}\,;\\ \mathbb{E}\, f(S_n^{(p)}) \le 1 - \mathrm{e}^{-o(n)}\,.\end{gathered}$$

The expectation need not be really small, it only needs to be a bit less than 1 in order to explain (3.5).

Now we have (at least) three ways to proceed (assuming still that $\frac{k}{n} \in (0,1)$ is not close to 0 and 1). The first way:

$$\mathbb{P}\big(S_n^{(p)} = k\big) \ge \frac{\text{const}}{\sqrt{n}}$$

by the local limit theorem; therefore

$$\mathbb{E}\,\big(1 - f(S_n^{(p)})\big) \ge \mathbb{P}\big(S_n^{(p)} = k\big) = \mathrm{e}^{-o(n)}\,.$$

The second way:

$$\mathbb{P}\big(k \le S_n^{(p)} \le k + \text{const}\cdot\sqrt{n}\big) \ge \text{const} > 0$$

by the central limit theorem; therefore

$$\mathbb{E}\,\big(1 - f(S_n^{(p)})\big) \ge \mathbb{P}\big(k \le S_n^{(p)} \le k + \text{const}\cdot\sqrt{n}\big) \cdot \Big(\frac{1-p}{p}\Big)^{\text{const}\cdot\sqrt{n}} = \mathrm{e}^{-o(n)}\,.$$

The third way: for every $\varepsilon > 0$,

$$\mathbb{P}\big(k \le S_n^{(p+\varepsilon)} \le k + 4\varepsilon n\big) \to 1 \quad \text{as } n \to \infty$$

by the weak law of large numbers; therefore

$$\mathbb{E}\,\big(1 - f(S_n^{(p+\varepsilon)})\big) \ge \big(1 - o(1)\big) \cdot \Big(\frac{1-p-\varepsilon}{p+\varepsilon}\Big)^{4\varepsilon n}\,;$$

$$\frac{\mathrm{e}^{\lambda k}}{\mathbb{E}\,\mathrm{e}^{\lambda S_n}} \mathbb{P}\big(S_n \ge k\big) \ge \frac{\mathrm{e}^{\lambda_\varepsilon k}}{\mathbb{E}\,\mathrm{e}^{\lambda_\varepsilon S_n}} \mathbb{P}\big(S_n \ge k\big) \ge \exp\Big(-n \cdot 4\varepsilon \ln \frac{p+\varepsilon}{1-p-\varepsilon} - o(n)\Big)\,;$$

it holds for all ε, and we get $\mathrm{e}^{-o(n)}$.