Beginning of Oct. 7, 2022

For a sequence $b_n \to 0$ converging to 0, we say b_n decays like e^{-cn} if

$$-c = \lim \frac{1}{n} \log b_n,$$

or equivalently, for $\epsilon > 0$,

$$e^{-(c+\epsilon)n} \leqslant b_n \leqslant e^{-(c-\epsilon)n}$$

for sufficiently large n.

Similarly, we want to show that $\mathbb{P}(S_n/\mu > a)$ decays like $e^{-I(a)n}$ for some $I(a) > 0$. Question: what is

$$\gamma(a) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n/n > a)?$$

We define $\pi_a := \mathbb{P}(S_n/n \geqslant a)$. We claim that $\log \pi_n$ is **superadditive**: $\pi_{m+n} \geqslant \pi_m \pi_n$, so $\log \pi_{m+n} \geqslant \log \pi_n + \log \pi_m$. This is true because

$$\mathbb{P}(S_{m+n} \geqslant (m+n)a) \geqslant \mathbb{P}(S_n \geqslant na, S_{n+m} - S_n \geqslant ma) = \mathbb{P}(S_n \geqslant na, S_m \geqslant ma) = \mathbb{P}(S_n \geqslant na)\mathbb{P}(S_m \geqslant ma).$$

Lemma: D2.7.1

If γ_n is superadditive, then $\gamma/n \to \sup_m \gamma_m/m$.

Proof. Call the supremum limit c. It suffices to show $0 \leqslant \liminf \leqslant \limsup \leqslant c$.

$\limsup \leqslant c = \sup$ is trivial by definition.

Conversely, we need to show $\liminf \gamma_n/n \geqslant \gamma_m/m$ for all m. Induction says if $n = n_1 + \ldots + n_k$ then $\gamma_n \geqslant \gamma_{n_1} + \ldots + \gamma_{n_k}$. In particular, if we fix m, then we can write n as $n = km + \ell$ with $0 \leqslant \ell < m$. Then,

$$\frac{\gamma_n}{n} \geqslant \frac{k\gamma_m + \gamma_\ell}{km + \ell} = \frac{km}{km+\ell}\frac{\gamma_m}{m} + \frac{\gamma_\ell}{km+\ell}.$$

As $n \to \infty$ so $k \to \infty$, ℓ is bounded, so $km/(km+\ell) \to 1$. So does $\gamma_\ell/(km+\ell)$. Therefore,

$$\liminf_{n\to\infty} \frac{\gamma_n}{n} \geqslant \frac{\gamma_m}{m}.$$

□

Therefore, $\mathbb{P}(S_n/n \leqslant a) \leqslant e^{\gamma(a)n}$ in particular, since $\gamma(a) \geqslant n^{-1} \log \mathbb{P}(S_n/n \geqslant a)$ as shown above. That is, this exponential decay rate is also an upper bound for $\mathbb{P}(S_n/n \leqslant a)$.

Suppose MGF $\varphi(\theta) = \mathbb{E}e^{\theta X}$ is finite in $(-\delta, \delta)$. In this interval,

$$\frac{X^k e^{\theta X}}{e^{(\theta+\epsilon)X}} = X^k e^{-\epsilon X} \to 0$$

as $X \to \infty$. In particular, if $\theta \in (-\delta, \delta)$, so does the new quantity when ϵ is small, so $\mathbb{E}(X^k e^{\theta X})$ is finite, for all k. In particular, for $k = 1$,

$$\lim_{h\to 0} \mathbb{E}\left(\frac{e^{(\theta+h)X} - e^{\theta X}}{h}\right) = \lim_{h\to 0} \mathbb{E}\left(e^{\theta X}\frac{e^{hX} - 1}{h}\right).$$

Assuming h positive,

$$\left|\frac{e^{hX} - 1}{h}\right| \leqslant \begin{cases} |X| & \text{if } X < 0 \\ Xe^{hX} & \text{if } X \geqslant 0 \end{cases}$$

We have shown that Xe^{hX} is integrable for small h, so indeed we can apply DCT and obtain $\varphi'(\theta) = \mathbb{E}(Xe^{\theta X})$. Similarly, if we differentiate twice, we obtain $\varphi''(\theta) = \mathbb{E}(X^2 e^{\theta X})$, and so on. Also,

$$(\log\varphi)'(\theta) = \frac{\varphi'(\theta)}{\varphi(\theta)} = \frac{\int Xe^{\theta X}\, \mathrm{d}\mathbb{P}}{\int e^{\theta X}\, \mathrm{d}\mathbb{P}}.$$

Given $g \geqslant 0$, we can define a probability measure by

$$\nu(A) = \frac{\int_A g\, \mathrm{d}\mathbb{P}}{\int g\, \mathrm{d}\mathbb{P}},$$

"$\mathbb{P}$ weighted by g", and equivalently

$$\mathbb{E}_\nu 1_A = \frac{\int 1_A g\, \mathrm{d}\mathbb{P}}{\int g\, \mathrm{d}\mathbb{P}}.$$

Using standard measure theory argument, we obtain

$$\mathbb{E}_\nu f = \frac{\int fg\, \mathrm{d}\mathbb{P}}{\int g\, \mathrm{d}\mathbb{P}}.$$

Therefore, $(\log\varphi)'(\theta)$ can be thought of as $\mathbb{E}_{\nu_\theta} X$, under the tilted distribution of ν_θ.
Also,

$$(\log\varphi)''(\theta) = \frac{\varphi(\theta)\varphi''(\theta) - \varphi'(\theta)^2}{\varphi(\theta)^2} = \frac{\int X^2 e^{\theta X}\, \mathrm{d}\mathbb{P}}{\int e^{\theta X}\, \mathrm{d}\mathbb{P}} - \left(\frac{\int Xe^{\theta X}\, \mathrm{d}\mathbb{P}}{\int e^{\theta X}\, \mathrm{d}\mathbb{P}}\right)^2,$$

namely $\mathrm{var}_{\nu_\theta}(X)$, which is nonnegative. Therefore, $\log\varphi$ is convex. Also note $(\log\varphi)(0) = 0$ with $(\log\varphi)'(0) = \mathbb{E}X$.
What about MGF of sums S_n for i.i.d. random varaibles?

$$\varphi_{S_n}(\theta) = \mathbb{E}e^{\theta(X_1+\ldots+X_n)} = \varphi(\theta)^n.$$

Now we fix $\theta > 0$. Then

$$\mathbb{P}(S_n/n > a) = \mathbb{P}(e^{\theta S_n} > e^{\theta na}),$$

so by Markov, this is bounded from above by

$$\mathbb{P}(S_n/n > a) \leqslant \frac{\mathbb{E}e^{\theta S_n}}{e^{\theta na}} = \frac{\varphi(\theta)^n}{e^{\theta na}} = \exp(-(a\theta - \log\varphi(\theta))n).$$

To show exponential decay, it suffices to show the above exponent is positive. In particular, if

$$I(a) := \sup_{\theta>0}(a\theta - \log\varphi(\theta)) > 0$$

we are done. Indeed! $a > \mathbb{E}X$, and $\log\varphi$ is convex, so if we start at the origin and draw a line $a\theta$, it is steeper than $\log\varphi$ so it will go above the graph of $\log\varphi$, resulting in a positive supremum.