1

Chapter 2

Chebyshev tells us that if X is a random variable w/ mean μ and variance σ^2, then

$$\mathbb{P}\{|X-\mu| \geq t\} \leq \frac{\sigma^2}{t^2}.$$

Unfortunately, the bound above is often hopelessly weak:

Toss a fair coin N times. Then

$$\underbrace{\mathbb{E} S_N}_{\text{sum of the corresponding random variables}} = \frac{1}{2} \cdot N = \frac{N}{2} \text{ and } Var(S_N) = Np(1-p) = \frac{N}{4}.$$

By Chebyshev, $\mathbb{P}\{S_N \geq \frac{3}{4}N\}$

$$\leq \mathbb{P}\left\{\left|S_N - \frac{N}{2}\right| \geq \frac{N}{4}\right\} \leq \frac{4}{N}.$$

goes to 0, but not too quickly.

Can we do better? Let's try the <u>central</u> <u>limit</u> <u>theorem</u>.

$$S_N = \underbrace{\sum_{i=1}^{N} X_i}_{Ber(\frac{1}{2})}$$

Let $Z_N = \dfrac{S_N - \frac{N}{2}}{\sqrt{\frac{N}{4}}} \longrightarrow N(0,1)$

The idea is that

$$\mathbb{P}\left\{S_N \geq \frac{3N}{4}\right\} = \mathbb{P}\left\{Z_N \geq \sqrt{\frac{N}{4}}\right\} \underbrace{\approx}_{\text{to be made more precise}} \mathbb{P}\left\{\underbrace{g}_{g \sim N(0,1)} \geq \sqrt{\frac{N}{4}}\right\}$$

The point is that we must estimate the tails of the normal distribution.

Tails of the normal distribution: Let $g \sim N(0,1)$

Then for all $t>0$,

$$\left(\frac{1}{t} - \frac{1}{t^3}\right)\frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \leq \mathbb{P}\{g \geq t\} \leq \frac{1}{t}\frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

③

In particular,

$$\mathbb{P}\{g \geq t\} \leq \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

Proof:

$$\mathbb{P}\{g \geq t\} = \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{t^2}{2}} e^{-ty} e^{-\frac{y^2}{2}} dy$$

$$\overset{\parallel}{\frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-\frac{x^2}{2}} dx} \qquad \leq \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \int_0^\infty e^{-ty} dy$$

set $x = t + y$ since $e^{-\frac{y^2}{2}} \leq 1$

check!

Therefore,

$$\mathbb{P}\{g \geq t\} \leq \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \int_0^\infty e^{-ty} dy$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \frac{1}{t} \checkmark$$

④

The lower bound is based on a small miracle:

$$\frac{1}{\sqrt{2\pi}}\int_t^\infty e^{-\frac{x^2}{2}}dx \geq \frac{1}{\sqrt{2\pi}}\int_t^\infty e^{-\frac{x^2}{2}}\left(1-3x^{-4}\right)dx$$

$$=\left(\frac{1}{t}-\frac{1}{t^3}\right)e^{-\frac{t^2}{2}}$$ (please make this calculation)

How should we have known to do this? There is no good answer to this question. Sometimes one just needs to experiment and hope for the best.

What did this buy us as far as the original coin flip game goes?

$$\mathbb{P}\left\{\geq \frac{3N}{4} \text{ heads}\right\} \leq \frac{1}{\sqrt{2\pi}} e^{-\frac{N}{8}}$$

exponential decay!

EXCEPT THAT WE CHEATED!

The approximation by $N(0,1)$ is valid, but costly! We shall see why in a moment.

Berry-Esen theorem: Under the assumptions of the central limit theorem, for every N and every $t \in \mathbb{R}$,

$$\left| \mathbb{P}\{Z_N \geq t\} - \mathbb{P}\{g \geq t\} \right| \leq \frac{\rho}{\sqrt{N}},$$

where $\rho = \mathbb{E}|X_1 - \mu|^3 / \sigma^3$ & $g \sim N(0,1)$.

Sadly, $\frac{1}{\sqrt{N}}$ error kills the exponential decay estimate above.

Can this $\frac{1}{\sqrt{N}}$ approximation be improved?

In general, no! Take the coin (fair) toss game. Then

$$\mathbb{P}\left\{ S_N = \frac{N}{2} \right\} = 2^{-N} \binom{N}{N/2} \sim \frac{1}{\sqrt{N}}$$

why is this true?

De Moivre - Stirling:

$$n! \sim \sqrt{2\pi}\, n^{n+\frac{1}{2}} e^{-n}$$

can you prove it?

(6)

So, $2^{-N}\binom{N}{N/2} \sim 2^{-N}\dfrac{N!}{\left(\frac{N}{2}\right)!\left(\frac{N}{2}\right)!} \sim$

$$\frac{2^{-N}\,\cancel{\sqrt{2\pi}}\;N^{N+\frac{1}{2}}\;\cancel{e^{-N}}}{\cancel{\sqrt{2\pi}}\,\sqrt{2\pi}\left(\frac{N}{2}\right)^{\frac{N}{2}+\frac{1}{2}}\left(\frac{N}{2}\right)^{\frac{N}{2}+\frac{1}{2}}\cancel{e^{-\frac{N}{2}}}\,\cancel{e^{-\frac{N}{2}}}}$$

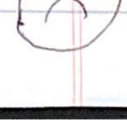
$$\sim 2^{-N}\,2^{N}\cdot 2\,N^{-\frac{1}{2}} \sim \frac{1}{\sqrt{N}},$$ as claimed.

It follows that $\mathbb{P}\{Z_N = 0\} \sim \dfrac{1}{\sqrt{N}}$

Since the normal distribution is continuous,
$\mathbb{P}\{g = 0\} = 0$, so the approximation has to be of order $\dfrac{1}{\sqrt{N}}$ and no better.

In summary, our attempt to improve Chebyshev for Ber(p) has failed!
Bernoulli w/ probability p

What's next?!



Definition: A random variable X has a symmetric Bernoulli distribution (Rademacher distribution) if it takes values ± 1 w/ probability $\frac{1}{2}$, i.e

$$\mathbb{P}\{X=-1\} = \mathbb{P}\{X=1\} = \frac{1}{2}.$$

Theorem: (Hoeffding's inequality) Let $X_1, X_2, \ldots, X_N$ independent symmetric Bernoulli random variables. Let $(a_1, a_2, \ldots, a_N) \in \mathbb{R}^N$. Then for any $t>0$,

$$\mathbb{P}\left\{\sum_{i=1}^{N} a_i X_i \geq t\right\} \leq \underbrace{\exp\left(\frac{-t^2}{2\|a\|_2^2}\right)}_{\text{this gives us an exponential drop-off}}$$

Before we prove this, observe that X has the usual Bernoulli distribution w/ parameter $\frac{1}{2}$ iff $Z = 2X-1$ has a symmetric Bernoulli distribution.

(8)

We may assume that $\|a\|_2 = 1$. Why? Because

$$\mathbb{P}\left\{\sum_{i=1}^{N} a_i X_i \geq t\right\} = \mathbb{P}\left\{\sum_{i=1}^{N} \frac{a_i}{\|a\|_2} X_i \geq \frac{t}{\|a\|_2}\right\}$$

If we let $\frac{a_i}{\|a\|_2} = b_i$, $\|b\|_2 = 1$

If we can prove that

$$\mathbb{P}\left\{\sum_{i=1}^{N} b_i X_i \geq t\right\} \leq \exp\left(-\frac{t^2}{2}\right),$$

then $\mathbb{P}\left\{\sum_{i=1}^{N} b_i X_i \geq \frac{t}{\|a\|_2}\right\} \leq \exp\left(-\frac{t^2}{2\|a\|_2^2}\right),$

which implies that

$$\mathbb{P}\left\{\sum_{i=1}^{N} a_i X_i \geq t\right\} \leq \exp\left(-\frac{t^2}{2\|a\|_2^2}\right)$$

Now that we have seen the value of scaling,

$$\mathbb{P}\left\{ \sum_{i=1}^{N} a_i X_i \geq t \right\} = \mathbb{P}\left\{ \exp\left(\underbrace{\lambda}_{\substack{\lambda>0 \\ \text{parameter}}} \sum_{i=1}^{N} a_i X_i \right) \geq \exp(\lambda t) \right\}$$

Markov

$$\leq e^{-\lambda t} \underbrace{\mathbb{E} \exp\left(\lambda \sum_{i=1}^{N} a_i X_i \right)}_{\text{moment generating function of } \sum_{i=1}^{N} a_i X_i}$$

Exercise: Independence implies that MGF of the sum is the product of MGFs of the terms (please work this out!)

It follows that

$$\mathbb{E} \exp\left(\lambda \sum_{i=1}^{N} a_i X_i \right) = \prod_{i=1}^{N} \mathbb{E} \exp\left(\lambda a_i X_i \right)$$

(10)

Since X_i takes values ± 1 w/ probability $\frac{1}{2}$ each,

$$\mathbb{E}\exp(\lambda a_i X_i) = \frac{\exp(\lambda a_i) + \exp(-\lambda a_i)}{2}$$

$$= \cosh(\lambda a_i)$$

<u>Claim:</u> $\cosh(x) \le \exp\left(\frac{x^2}{2}\right)$

$$\shortparallel$$

$$\frac{e^x + e^{-x}}{2} = \frac{1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \cdots + 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \cdots}{2}$$

$$= 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \frac{x^6}{6!} + \cdots$$

$$\le 1 + \frac{x^2}{2} + \frac{\left(\frac{x^2}{2}\right)^2}{2!} + \frac{\left(\frac{x^2}{2}\right)^3}{3!} + \cdots$$

$$\shortparallel$$

$$e^{\frac{x^2}{2}}$$

(11)

Using the bound we just proved,

$$\mathbb{E}\left(\exp(\lambda a_i X_i)\right) \leq \exp\left(\lambda^2 a_i^2/2\right)$$

It follows that

$$\mathbb{P}\left\{\sum_{i=1}^{N} a_i X_i \geq t\right\} \leq$$

$$e^{-\lambda t} \prod_{i=1}^{N} \exp\left(\lambda^2 a_i^2/2\right) =$$

$$\exp\left(-\lambda t + \frac{\lambda^2}{2}\underbrace{\left(\sum_{i=1}^{N} a_i^2\right)}_{=1 \text{ by our original reduction}}\right)$$

$$= \exp\left(-\lambda t + \frac{\lambda^2}{2}\right)$$

$$= \exp\left(\frac{1}{2}\left(\lambda^2 - 2\lambda t\right)\right)$$

$$= \exp\left(\frac{1}{2}\left(\left(\lambda - t\right)^2 - t^2\right)\right)$$

12

Since the bound holds for arbitrary λ,

$$\mathbb{P}\left\{\sum_{i=1}^{N} a_i X_i \geq t\right\} \leq \exp\left(-\frac{t^2}{2}\right)$$

as desired!