Sums of i.i.d. variables, Gaussians, CLT.

Motivation. In analysis of algorithms, you have something (alg?) which "succeeds" w.p. p. You run it n times indep. You want to understand the total # of successes.

Let $X_1, X_2, \cdots, X_n$ be i.i.d. (indep. & identically distrib.) r.v.'s with
$$\Pr[X_i = 1] = p, \quad \Pr[X_i = 0] = (1-p) \qquad (\text{"Bernoulli"})$$
Let $S = X_1 + X_2 + \cdots + X_n$ —— want to understand it.

$\mathbb{E}[S] = np. = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots \mathbb{E}[X_n]$ (linearity of expectation, didn't need independence)

$\mathrm{Var}[S] = np(1-p).$

Recall: • $\mathrm{Var}[Y] = \mathbb{E}(Y - \mathbb{E}Y)^2 = \mathbb{E}Y^2 - (\mathbb{E}Y)^2$

• $\mathrm{Var}[Y+Y'] = \mathrm{Var}[Y] + \mathrm{Var}[Y']$ if indep.

• $\mathrm{Var}[cY] = c^2 \mathrm{Var}[Y]$

• $\mathrm{Var}[X_i] = p - p^2 = p(1-p)$

(Rule of life: if you ever have a r.v., consider making it mean 0 & var. 1)

mean 0 : $\quad \dfrac{S_n - np}{\mathrm{Stddev}[S_n] = \sqrt{np(1-p)}}$

var. 1 :

I.e. let $Z_n = \dfrac{S_n - np}{\sqrt{np(1-p)}} = \dfrac{S_n - \mu}{\sigma}.$

(Lost no info:) $\quad \Pr[S_n \le u] = \Pr[\sigma Z_n + \mu \le u] = \Pr\left[Z_n \le \frac{u-\mu}{\sigma}\right]$

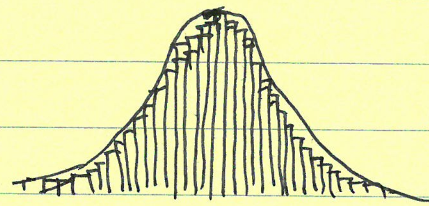
(So study Z_n)

Eg. $p = \frac{1}{2}$ (unbiased coin flips) $\quad Z_n = \dfrac{X_1 + X_2 + \cdots X_n - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}}$

$$= \frac{1}{\sqrt{n}}\Big(\underbrace{(2X_1 - 1)} + \underbrace{(2X_2 - 1)} + \cdots + \underbrace{(2X_n - 1)}\Big)$$

indep r.v.'s

Plot the histogram of Z_n:



For any p, even for $\Pr[X_i = 1] = p_i$ for different p_i's

Appears to be "converging" to a fixed continuous distrib. — the famous Bell curve/Gaussian/normal r.v.

<u>CLT.</u> For any i.i.d. $X_1, X_2, \cdots$ (not necessarily 0/1 valued).

$$Z_n \longrightarrow Z \text{ (standard Gaussian).}$$

in that $\forall u \in \mathbb{R}$, $\Pr[Z_n \le u] \xrightarrow{n\to\infty} \Pr[Z \le u]$

<u>Remark.</u> Practically useless. No info on speed of convergence. Also i.i.d. is too strong.

<u>Def.</u> "$Z \sim \mathcal{N}(\overset{\text{mean}}{0}, \overset{\text{stddev}}{1})$" $\equiv$ "Z is a std. normal/Gaussian", means Z is a continuous r.v. with pdf $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$.

<u>Fact.</u> Let $\vec{z} = (z_1, z_2, \cdots, z_d) \in \mathbb{R}^d$, where $z_1, z_2, \cdots, z_d \sim \mathcal{N}(0,1)$, i.i.d (random vector)

Then $\vec{z}$'s distrib is <u>rotationally symmetric</u>. (i.e. equally likely for $\vec{z}_1$ & $\vec{z}_2$ when $\|\vec{z}_1\| = \|\vec{z}_2\|$)

<u>Proof.</u> pdf of $\vec{z}$ @ $(z_1, z_2, \cdots, z_d) = \phi(z_1)\phi(z_2)\cdots\phi(z_d) = \left(\frac{1}{\sqrt{2\pi}}\right)^d e^{-\underbrace{(z_1^2+z_2^2+\cdots+z_d^2)}_{\|\vec{z}\|^2}/2}$

<u>Cor 1.</u> $\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} = 1$ (it's really a pdf)

<u>Proof</u> look at 2-dimensional Gaussian's pdf — homework.

<u>Cor 2</u> Sum of indep. Gaussian's is Gaussian. — homework.

<u>Def.</u> Let. $Z \sim \mathcal{N}(0,1)$. Let $\mu, \sigma \in \mathbb{R}$. $Y = \mu + \sigma z$ rem: $\mathbb{E}Y = \mu$. $\mathrm{Var}[Y] = \sigma^2$

We call Y a Gaussian (non-standard) too. write $Y \sim \mathcal{N}(\mu, \sigma^2)$.

<u>Fact.</u> Let $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$. indep. Let $Z = aX + bY$

Then $Z \sim \mathcal{N}(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$ — homework.

Berry-Esseen Thm. (CLT w/ error bounds).

Let $X_1, \cdots, X_n$ be indep. Assume $\mathbb{E}X_i = 0\ \forall i$, $\mathrm{Var}[X_i] (= \mathbb{E}X_i^2) = \sigma_i^2$, $\sum \sigma_i^2 = 1$ (WOLOG. why?)

Let $S = X_1 + X_2 + \cdots + X_n$ (So $\mathbb{E}S = 0$. $\mathrm{Var}[S] = 1$)

Then $\forall u \in \mathbb{R}$, $\left| \Pr[S \le u] - \Pr_{Z \sim N(0,1)}[Z \le u] \right| \le O(1) \cdot \beta$ ($O(1) \le .5514$ [Shevtsova '13])

where $\beta = \sum_{i=1}^n \mathbb{E}|X_i|^3$

Is this error small? Let's try an example.

Study n coins flips, let $X_i = \begin{cases} +1/\sqrt{n} & \text{w.p. } 1/2 \\ -1/\sqrt{n} & \text{w.p. } 1/2. \end{cases}$

mean: 0 ✓ $\sigma_i^2 = \mathbb{E}X_i^2 = \frac{1}{n}$. $\sum \sigma_i^2 = 1$ ✓

$\mathbb{E}|X_i|^3 = \frac{1}{n^{3/2}}$, so $\beta = \frac{1}{\sqrt{n}}$. $\Rightarrow \forall u \in \mathbb{R}$. $\left| \Pr[S \le u] - \Pr[Z \le u] \right| \le .56/\sqrt{n}$

Remark. $S = \frac{\#H - \#T}{\sqrt{n}}$ (say n even). $S = 0 \Leftrightarrow \#H = \#T = n/2$

$$\Pr[\#H = \#T] = \Pr[S=0] = \Pr[S \le 0] - \Pr[S \le -\varepsilon] \quad (\varepsilon \to 0^+)$$

$$\le \left(\Pr[S \le 0] - \Pr[Z \le 0]\right) - \left(\Pr[S \le -\varepsilon] - \Pr[Z \le -\varepsilon]\right)$$

$$\le .56/\sqrt{n} + .56/\sqrt{n} = \frac{1.12}{\sqrt{n}}.$$

On the other hand.

$$\frac{\binom{n}{n/2}}{2^n} = \frac{\frac{n!}{((n/2)!)^2}}{2^n} \approx \frac{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n}{2\pi \frac{n}{2} \left(\frac{n}{2e}\right)^n \cdot 2^n} = \frac{\sqrt{2}}{\sqrt{\pi n}} \sim \frac{.798}{\sqrt{n}}.$$

$\hookrightarrow$ Sterling's approx.

Chernoff / Tail Bounds

Motivation. Let $X_i = \begin{cases} +1 & \text{w.p. } 1/2 \\ -1 & \text{w.p. } 1/2 \end{cases}$ i.i.d. $S = \sum_{i=1}^{n} X_i$ $\boxed{\text{Scenario } (*)}$

recall: Berry-Esseen: $\Pr[S \geq \sqrt{n} \cdot t] \approx \Pr[G \geq t] \pm \frac{O(1)}{\sqrt{n}}$

$$\Pr[G \geq t] = \int_t^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \sim \Theta(1) \cdot e^{-t^2/2}$$

Let $t = 10\sqrt{\ln n}$. : $\Pr[S \geq \sqrt{n} \cdot 10\sqrt{\ln n}] \leq O(\frac{1}{n^{50}}) + O(\frac{1}{\sqrt{n}}) = O(\frac{1}{\sqrt{n}})$.

Remark The error term $O(\frac{1}{\sqrt{n}})$ dominates and prevents us from getting better bounds!

Goal Use Chernoff Bound to get something much better. (when the tail is small).

Bounding R.V.'s : more info = better bounds.

Markov Ineq (when only knowing the mean)

Assume $X \geq 0$: $\Pr[X \geq t \cdot \mathbb{E}X] \leq 1/t$.

Proof. $\mathbb{E}X \geq \Pr[X \geq \alpha] \cdot \alpha + \Pr[X < \alpha] \cdot 0 \Rightarrow \Pr[X \geq \alpha] \leq \frac{\mathbb{E}X}{\alpha}$

Take $\alpha = (\mathbb{E}X) \cdot t$.

Chebyshev Ineq. (know about mean & var.)

Let $\mathbb{E}X = \mu$. $\mathrm{Var}[X] = \sigma^2$ $(\sigma > 0)$. Then $\forall t > 0$, $\Pr[|X - \mu| \geq t \cdot \sigma] \leq \frac{1}{t^2}$.

Proof. Let $Y = (X-\mu)^2$ $\mathbb{E}Y = \sigma^2$. $Y \geq 0$.

Apply Markov to Y: $\Pr[Y \geq t^2 \cdot \mathbb{E}Y] \leq \frac{1}{t^2}$

$\hookrightarrow \Pr[(X-\mu)^2 \geq t^2 \cdot \sigma^2] = \Pr[|X - \mu| \geq t \cdot \sigma]$

Back to Scenario (*)

By Markov: Let $T = S + n \geq 0$. $\mathbb{E}T = n + \mathbb{E}S = n$

$$\Pr[S \geq 10\sqrt{n \ln n}] = \Pr[T \geq 10\sqrt{n \ln n} + n]$$

$$= \Pr\left[T \geq \mathbb{E}T \cdot \frac{10\sqrt{n \ln n} + n}{n}\right] \leq \frac{n}{n + 10\sqrt{n \ln n}} = 1 - \Theta\left(\sqrt{\frac{\ln n}{n}}\right) \quad :($$

By Chebyshev. $\mu = \mathbb{E}S = 0$. $\sigma^2 = \mathrm{Var}[S] = n$.

$$\Pr[S \geq 10\sqrt{n \ln n}] \leq \Pr[|S - \mu| \geq 10\sqrt{\ln n} \cdot \sigma] \leq \frac{1}{100 \ln n}$$

($\to 0$ at least, still, truth is $\sim \frac{1}{n^{50}}$).

Remark: Chebyshev doesn't need independence, just "pair-wise" independence.

$$\mathrm{Var}[S] = \mathrm{Var}[X_1+X_2+\cdots+X_n] = \mathbb{E}(X_1+\cdots+X_n)^2 - (\mathbb{E}(X_1+\cdots+X_n))^2$$

$$= \mathbb{E}(X_1+\cdots+X_n)^2 = \mathbb{E}X_1^2+\cdots+\mathbb{E}X_n^2 + \sum_{i\neq j} \mathbb{E}X_iX_j$$

$$\hookrightarrow \sum_{i\neq j}(\mathbb{E}X_i)(\mathbb{E}X_j) \text{ (if pairwise indep.)}$$

<u>Can we use stronger independence to get bounds better than Chebyshev?</u>

<u>"Fourth-moment method"</u> (works w/ 4-wise indep.).

$$\mathbb{E}S^4 = \mathbb{E}\Big(\sum_{i=1}^n X_i\Big)^4 = \sum_{i=1}^n \mathbb{E}X_i^4 + \cancel{\sum_{i=1}^n\sum_{j\neq i}}\ \cancel{\sum_{i\neq j}}$$

$$\underbrace{\mathbb{E}X_i^2X_j^2 \text{ terms}} + \underbrace{\mathbb{E}X_iX_j^3 \text{ terms} + \mathbb{E}X_iX_jX_k^2 \text{ terms} + \mathbb{E}X_iX_jX_kX_l \cdots}_{\to 0}$$

$$\hookrightarrow 1\cdot \text{how many?} = 1\cdot\binom{n}{2}\cdot\binom{4}{2} = \frac{n(n-1)}{2}\cdot 6 = 3n^2-3n.$$

$$= n+3n^2-3n = 3n^2-2n \le 3n^2.$$

<u>Apply Markov to S^4</u>: $\Pr[|S| \ge t\sqrt{n}] = \Pr[S^4 \ge t^4\cdot n^2] \le \cancel{\frac{t^4\cdot n^2}{\mathbb{E}S^4}}\ \frac{\mathbb{E}S^4}{t^4\cdot n^2} \le \frac{3}{t^4}$

take $t = 10\sqrt{\ln n}$: $\Pr[S \ge 10\sqrt{n\ln n}] \le \frac{3}{10000(\ln n)^{\cancel{4}2}}$

<u>Keep going?</u> $\mathbb{E}S^{2k} \le C_k\cdot n^k$. Markov on S^{2k}: $\Pr[|S| \ge t\sqrt{n}] \le \frac{C_k}{t^{2k}}$, optimize over k.
(Pretty painful.).

<u>"Chernoff Method."</u> Consider $e^{\lambda S}$ instead of S^{2k} ($\lambda > 0$).

$$\mathbb{E}e^{\lambda S} = \mathbb{E}e^{\lambda\cdot\sum X_i} = \mathbb{E}\prod_{i=1}^n e^{\lambda X_i} = \prod_{i=1}^n \mathbb{E}e^{\lambda X_i} \text{ (used full indep.).}$$

$$\mathbb{E}e^{\lambda X_i} = \tfrac{1}{2}e^{\lambda} + \tfrac{1}{2}e^{-\lambda}$$

$$= \tfrac{1}{2}\Big(1+\lambda+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}+\frac{\lambda^4}{4!}+\cdots\Big) + \tfrac{1}{2}\Big(1-\lambda+\frac{\lambda^2}{2!}-\frac{\lambda^3}{3!}+\frac{\lambda^4}{4!}-\cdots\Big)$$

$$= 1+\frac{\lambda^2}{2!}+\frac{\lambda^4}{4!}+\frac{\lambda^6}{6!}+\cdots \le e^{\lambda^2/2}$$

$$\Rightarrow\ \le e^{n\lambda^2/2}$$

<u>Apply Markov to $e^{\lambda S}$</u>: $\Pr[S \ge 10\sqrt{n\ln n}] = \Pr[e^{\lambda S} \ge e^{10\sqrt{n\ln n}\cdot\lambda}]$

$$\le \mathbb{E}e^{\lambda S}/e^{10\sqrt{n\ln n}\cdot\lambda} \le e^{n\lambda^2/2 - 10\sqrt{n\ln n}\cdot\lambda}$$

(take $\lambda = 10\sqrt{\frac{\ln n}{n}}$) $\le e^{\ln n\cdot 50 - 100\ln n} = \frac{1}{n^{50}}$ ☺

<u>Chernoff Bound</u>. Let $X_1, X_2, \cdots, X_n$ be indep. Bernoulli variables, $\mathbb{E}X_i = p_i$.

Let $X = \sum X_i$, $\mu = \mathbb{E}X$. Then for any $\delta > 0$.

1) $\Pr[X \geq (1+\delta)\mu] \leq \left[\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right]^{\mu}$. 2). $\Pr[X \leq (1-\delta)\mu] \leq \left[\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right]^{\mu}$.

<u>Remark</u> When $0 < \delta < 1$. 1) $\Rightarrow \Pr[X \geq (1+\delta)\mu] \leq e^{-\delta^2\mu/3}$) look nicer.

2) $\Rightarrow \Pr[X \leq (1-\delta)\mu] \leq e^{-\delta^2\mu/2}$)

<u>Proof</u>. Only prove 1). 2) is similar.

For any $\lambda > 0$ $\Pr[X \geq (1+\delta)\mu] = \Pr[e^{\lambda X} \geq e^{\lambda(1+\delta)\mu}]$

(Markov) $\leq \frac{\mathbb{E}e^{\lambda X}}{e^{\lambda(1+\delta)\mu}}$. $(*)$

where $\mathbb{E}e^{\lambda X} = \mathbb{E}\prod_{i=1}^{n} e^{X_i \cdot \lambda} = \prod_{i=1}^{n} \underbrace{\mathbb{E}e^{\lambda X_i}}$ (indep.)

$\hookrightarrow \underbrace{p_i e^{\lambda} + (1-p_i) = 1 + p_i(e^{\lambda}-1) \leq e^{p_i(e^{\lambda}-1)}}_{(1+x \leq e^x)}$

$\leq \prod_{i=1}^{n} e^{p_i(e^{\lambda}-1)}$

Therefore. $(*) \leq \frac{\prod_{i=1}^{n} e^{p_i(e^{\lambda}-1)}}{\prod_{i=1}^{n} e^{\lambda(1+\delta)p_i}} = e^{\sum_{i=1}^{n} p_i(e^{\lambda}-1) - \sum_{i=1}^{n} \lambda(1+\delta)p_i}$

$= e^{\sum_{i=1}^{n} p_i(e^{\lambda}-1-\lambda(1+\delta))}$

Set. $\lambda = \ln(1+\delta)$, we get.

$\Pr[X \geq (1+\delta)\mu] \leq e^{\sum_{i=1}^{n} p_i((1+\delta)-1-(1+\delta)\ln(1+\delta))} = e^{\mu \cdot (\delta - (1+\delta)\ln(1+\delta))}$

$= \left[\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right]^{\mu}$.