Central limit theorem and Tail bounds

1.1 Bernoulli variables

- Let $x_i = 1$ with prob $= p$
  $= 0$ " " $= 1-p$
- Let $x_1, x_2, \dots x_n$ be i.i.d. samples of x
- $S_n = \sum_{i=1}^{n} x_i$. Let S be the r.v. whose value lie in the interval $[0, n]$ and

$$P[S = S_n = k] = \binom{n}{k} p^k (1-p)^{n-k}$$

- $E(x_i) = p$, $Var(x_i) = p(1-p)$
- $E(S) = np$ $Var(S) = np(1-p)$
- Define $z_n = \dfrac{S_n - np}{\sqrt{np(1-p)}}$
- with $\mu = np$, $\sigma^2 = np(1-p)$: $z_n = \dfrac{S_n - \mu}{\sigma}$
- clearly: $P[S_n \le u] = P[\sigma z_n + \mu < u]$

$$= P\left[z_n \le \frac{u-\mu}{\sigma}\right]$$

Distribution of z_n: set $p = 1/2$

$$z_n = \frac{\sum_{i=1}^{n} x_i - \left(\frac{n}{2}\right)}{\frac{\sqrt{n}}{2}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (2x_i - 1) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \eta_i$$

$\eta_i = 2x_i - 1 = 1$ with prob $1/2$
$= -1$ " " $1/2$

- For any integer $k \in [0, n]$

$$P\left[Z_n \le \frac{2k-n}{\sqrt{n}} \right] = \binom{n}{k} \cdot \frac{1}{2^n} = P[S_n = k]$$

- A plot of this prob as $n \to \infty$ converges to the bell shaped Gaussian distribution: $N(0,1)$

1.2 Central limit theorem: Let $Z \sim N(0,1)$. For any i.i.d. $x_1, x_2 \dots x_n$ (not necessarily binary valued) as $n \to \infty$, we have $Z_n \to Z$ in the sense: $\forall\, u \in \mathbb{R}$

$$P[Z_n \le u] \to P[Z \le u].$$

More specifically, for every $\varepsilon > 0$, $\exists$ integer N: for all $n \geq N$ and for every $u \in \mathbb{R}$

$$\left| P[Z_n \le u] - P[Z \le u] \right| < \varepsilon$$

- Our goal is to under this error ε as a function of n.

- Recall: - $Z \sim N(0,1)$, $Y \sim N(\mu, \sigma^2)$: $Z = \frac{Y-\mu}{\sigma}$
- For $Z \in \mathbb{R}^d$, $\phi(Z)$ is circularly symmetric (ii) rotational symmetry.

2. Berry - Esseen Theorem: B-E-T strengthens CLT by providing rate of convergence

1940-42 **B-E-Theorem** Let $x_1 \dots x_n$ be independent and w.l.o.g, let $E(x_i) = 0$ and $Var(x_i) = \sigma_i^2$ and $\sum_{i=1}^{n} \sigma_i^2 = 1$ [obtained by multiplicative scaling]. Then for $\forall u \in R$, we have

$S_n = \sum_{i=1}^{n} x_i$

$$|P[S < u] - P[Z \leq u]| \leq O(1)\beta \to (1)$$

where $\beta = \sum_{i=1}^{n} E|x_i|^3$ and $Z \sim N(0,1)$

- The Constant (1) is small: 0.5514
- B-E does not require x_i's to be from identical distribution

How this bound works? Let $x_i = \frac{1}{\sqrt{n}}$ with $p = 1/2$

$= -1/\sqrt{n}$ " $(1-p) = 1/2$

Example
Rate of Convergence

be independent variables.

- $E(x_i) = 0$, $Var(x_i) = \left(\frac{1}{\sqrt{n}}\right)^2 \cdot \frac{1}{2} + \left(-\frac{1}{\sqrt{n}}\right)^2 1/2 = \frac{1}{n}$
- $Var(S_n) = \sum_{i=1}^{n} Var(x_i) = \frac{1}{n} \cdot n = 1$

- $E|x_i|^3 = \left|\frac{1}{\sqrt{n}}\right|^3 \cdot \frac{1}{2} + \left|-\frac{1}{\sqrt{n}}\right|^3 \cdot \frac{1}{2} = \frac{1}{n^{3/2}}$

$\therefore \beta = \sum E|x_i|^3 = n \frac{1}{n^{3/2}} = \frac{1}{\sqrt{n}}$

- According to B-E-T:

$$\boxed{\forall u \in R \quad |P[S \leq u] - P[Z \leq u]| \leq \frac{0.56}{\sqrt{n}}}$$
$\hookrightarrow (1)$

- That is, $\frac{0.56}{\sqrt{n}}$ is the rate of convergence

- **Question**: Can this rate be improved?

- Let n be even. Then $S = \dfrac{\#\text{Heads} - \#\text{Tails}}{\sqrt{n}}$

- When $S = 0 \Rightarrow \#\text{Heads} = \#\text{Tails} = \dfrac{n}{2}$

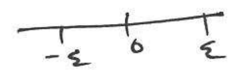
- Estimate the probability using (1)

- Let $\varepsilon > 0$. Then

$$P[\#H = \#T] = P[S = 0] = \underline{P[S \le 0] - P[S \le -\varepsilon]}$$

$$= \Big(P[S \le 0] - P[Z \le 0]\Big)$$

$$- \Big(P[S \le -\varepsilon] - P[Z \le -\varepsilon]\Big)$$



$$+ \Big(P(Z \le 0) - P[Z \le -\varepsilon]\Big)$$

$$\le \big| P[S \le 0) - P(Z \le 0) \big|$$

$$+ \big| P[S \le -\varepsilon] - P(Z \le -\varepsilon) \big|$$

$$+ \underbrace{P[-\varepsilon < Z \le 0]}_{\text{Tends to zero as } \varepsilon \to 0^+}$$

- Let $\varepsilon \to 0^+$. Then

$$P[\#H = \#T] \le \big| P(S \le 0) - P[Z \le 0) \big|$$

$$+ \big| P[S \le \varepsilon] \big| - P[Z \le \cancel{-}\varepsilon] \big|$$

$$\le \frac{0.56}{\sqrt{n}} + \frac{0.56}{\sqrt{n}} = \frac{1.12}{\sqrt{n}} \longrightarrow (2)$$

- But $P[\#H = \#T] = \dbinom{n}{n/2} \dfrac{1}{2^n}$

$n! \simeq \sqrt{2\pi n}\left(\dfrac{n}{e}\right)^n$

$\dbinom{n}{n/2} = \dfrac{n!}{\frac{n}{2}!\,\frac{n}{2}!}$

$= \dfrac{n!}{\left[\left(\frac{n}{2}\right)!\right]^2}$

$$= \frac{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n}{\left(2\pi \frac{n}{2}\right)\left(\frac{n}{2e}\right)^n} \cdot \frac{1}{2^n}$$

$$= \frac{\sqrt{2}}{\sqrt{\pi n}} \simeq \frac{0.798}{\sqrt{n}} \longrightarrow (3)$$

- This is exactly the bound given by B-E-T. Hence, it is tight.

3 Tail bounds: let

$$S = \sum_{i=1}^{n} x_i, \quad x_i \begin{cases} \to H = +1 \\ \to T = -1 \end{cases} \text{prob} = 1/2, \text{ Independent Tosses.}$$

- What is the probability that S deviates from its mean $E(S) = 0$?
- More specially: $P[S > t]$ – tail probability
- We expect this to be small for large t.
- Want to quantify it by deriving an upper bound.

$E(S) = 0$
$Var(S) = \sum Var(x_i) = n$
$\frac{S}{\sqrt{n}} \to G \sim N(0,1)$

By B-E-T:

$$\left| P[S \geq \sqrt{n}\, t] - P[G > t] \right| \leq \frac{O(1)}{\sqrt{n}}$$

- That is

$$P[S \geq \sqrt{n}\, t] = P[G \geq t] \pm \frac{O(1)}{\sqrt{n}} \to (4)$$

- Verify that

$$P[G \geq t] = \int_t^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \leq O(1)\, e^{-t^2/2} \to (5)$$

- choose $t = 10\sqrt{\log n}$. Combine (4) – (5):

$$P[S \geq \sqrt{n}\, t] \leq P[G \geq t] + O\left(\frac{1}{\sqrt{n}}\right)$$

$$= O\left(\exp\left[-\frac{(10\sqrt{\log n})^2}{2}\right]\right) + O\left(\frac{1}{\sqrt{n}}\right)$$

$$= \underbrace{O\left(\frac{1}{n^{50}}\right)}_{\text{Tail of the Gaussian}} + \underbrace{O\left(\frac{1}{\sqrt{n}}\right)}_{\text{B-E-T}} = O\left(\frac{1}{\sqrt{n}}\right)$$

$(10\sqrt{\log n})^2/2$
$= (100 \log n)/2$
$= 50 \log n$
$e^{-50\log n} = e^{\log \frac{1}{n^{50}}}$
$= \frac{1}{n^{50}}$

$\therefore$ B-E-T bound cannot be improved!

4. Markov inequality — uses only the mean

Tail bound

Thm 7: $\mathbb{P}[X \geq t\,E(X)] \leq \frac{1}{t}$ for $X \geq 0$

- Verify: $P[X \geq \alpha] \leq \frac{E(X)}{\alpha}$. $\boxed{\text{Set } \alpha = t\,E(X)}$

~~Example: Consider $S = \sum_{i=1}^{n} x_i$, $x_i = 1$ with p, $= 1$,, $1-p$, ≥ 0~~

Example: $x_i = \pm 1$ with $p = 1/2$ $\quad S = \sum_{i=1}^{n} x_i \in [-n, n]$

$\boxed{\begin{array}{l} E(S) = 0 \\ Var(S) = n \end{array}}$

- Consider $T = S + n \geq 0$

$$E(T) = E(S) + n = n$$

- Set $t = 10\sqrt{n \log n}$

$$P[S \geq t] = P[T \geq t + n]$$

$$= P\left[T \geq E(T) \cdot \frac{(t+n)}{n}\right]$$

$$\leq \frac{n}{t+n} = \frac{n}{n + 10\sqrt{n \log n}}$$

$$= \frac{1}{1 + \frac{10}{n}\sqrt{n \log n}}, \text{ a very bad bound}$$

$$\to 1 \quad \text{(does not converge to zero)}$$

$\frac{10}{n}\sqrt{n \log n}$
$= 10\sqrt{\frac{n \log n}{n^2}}$
$= 10\sqrt{\frac{\log n}{n}}$
$\to 0$ as $n \to \infty$

---

5. Chebyshev bound: uses mean and Variance

Theorem 8 $E(X) = \mu$, $Var(X) = \sigma^2 > 0$. For $t > 0$

$$\boxed{P[|X - \mu| \geq t \cdot \sigma] \leq \frac{1}{t^2}}$$

- Set $Y = (X - \mu)^2 \geq 0 \implies Var(Y) = \sigma^2 = E(X-\mu)^2$

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

$$= P[Y \geq t^2 \, E(Y)] \leq \frac{1}{t^2}$$

Example $x_i = \pm 1$ with $p = 1/2$

$$S = \sum_{i=1}^{n} x_i$$

$$\mu = E(S) = 0, \quad Var(S) = \sum_{i=1}^{n} Var(x_i) = n = \sigma^2$$

$$\therefore \sigma = \sqrt{n}$$

$$\therefore P[S \geq 10\sqrt{n \log n}] \leq P[|S| \geq 10\sqrt{n \log n}]$$

$$= P\left[|S| \geq \sigma \frac{10\sqrt{n\log n}}{\sigma}\right]$$

$$\leq \frac{\sigma^2}{(10\sqrt{n\log n})^2} = \frac{n}{100\, n \log n} = \frac{1}{100 \log n} \to 0$$

. This is not as good as expected but $\to 0$.

Note:- This derivation needs only pair wise independence of x_i's. For

$$Var(S) = Var(x_1 + x_2 \cdots x_n)$$

$$= E(x_1 + x_2 + \cdots x_n)^2 - \big(\underbrace{E(x_1 + x_2 \cdots + x_n)}_{=0}\big)^2$$

$$= \sum E(x_i^2) + \sum_{i \neq j} E(x_i x_j)$$

$\boxed{\text{Need only } x_i \text{ ind. of } x_j \;\; i \neq j}$

$\boxed{= 0 \text{ due to ind} + E(x_i) = 0}$

$$= \sum E[x_i^2] = n$$

---

6) 4th moment method

. let $S^4 \geq 0$. By Markov

$$P[S \geq 10\sqrt{n \log n}] \leq P[S^4 \geq (10\sqrt{n\log n})^4]$$

$$\leq \frac{E[S^4]}{10^4 n^2 \log^2 n} \longrightarrow (6)$$

Estimate $E(S^4)$:

$$E[S^4] = E\left[\left(\sum x_i\right)^4\right]$$

$$= \sum_i E[x_i^4] + \frac{1}{2}\binom{4}{2} \sum_i \sum_{j \neq i} E(x_i^2 x_j^2)$$

$$\frac{1}{2} \cdot \binom{4}{2} = \frac{1}{2} \cdot \frac{4!}{2!\,2!} = \frac{1}{2} \cdot \frac{\cancel{4} \times 3 \times \cancel{2} \times 1}{1 \cdot \cancel{2} \times 1 \cdot \cancel{2}} = 3$$

$$+ \binom{4}{1} \sum_i \sum_{j \neq i} E(x_i x_j^3)$$

$$+ \binom{4}{2} \sum_i \sum_{j \neq i} \sum_{k:\, k \neq i,\, k \neq j} E\, x_i x_j x_k^2$$

$$+ \sum_i \sum_{j \neq i} \sum_{k \neq i,\, k \neq j} \sum_{q \neq i,\, \neq j,\, \neq k} E(x_i x_j x_k x_q) \quad \rightarrow (7)$$

. Due to independence and $E(x_i) = 0$: Verify

$$E[x_i x_j^3] = E[x_i x_j x_k^2] = E[x_i x_j x_k x_q] = 0$$

Hence

$x_i = \pm 1$ with $1/2$; $E(x_i^4) = 1 \cdot 1/2 + 1 \cdot 1/2 = 1$

$$E[S^4] = \sum_i E(x_i^4) + 3 \sum_i \sum_{j \neq i} \underbrace{E(x_i^2 x_j^2)}_{E(x_i^2)\, E(x_j^2) = 1 \cdot 1 = 1}$$

$$= n + 3\, n(n-1) = 3n^2 - 2n$$

$$\leq 3n^2 \quad \longrightarrow (8)$$

. From (6) + (8) :

$$P\left[S \geq 10\sqrt{n \log n}\right] \leq \frac{3n^2}{10^4\, n^2 \log^2 n} = \frac{3}{10^4 \log^2 n}$$

. A better bound than Chebyshev.

Note:: We can extend it to S^{2k} and picking k to optimize the upper bound w. r. to k. But, this requires estimation of $E(S^{2k})$ which can be demanding.

7. <u>Chernoff bound</u>

- Instead of S^{2k} consider $e^{\lambda S}$ for λ positive

- Since e^{x} is $\uparrow$:

$$P[S \geq 10\sqrt{n \log n}] = P[\lambda S \geq 10\lambda\sqrt{n \log n}]$$
$$\leq P[e^{\lambda S} \geq e^{10\lambda\sqrt{n\log n}}]$$

- By Markov:

$$P[e^{\lambda S} > e^{10\lambda\sqrt{n \log n}}] \leq \frac{E(e^{\lambda S})}{e^{10\lambda\sqrt{n\log n}}} \longrightarrow (9)$$

- Due to independence of x_i's

$$E[e^{\lambda S}] = E[e^{\lambda \Sigma x_i}] = E\left(\prod_{i=1}^{n} e^{\lambda x_i}\right)$$
$$= \prod_i E(e^{\lambda x_i}) \longrightarrow (10)$$

- Since $x_i = \pm 1$ with $p = 1/2$

$$E[e^{\lambda x_i}] = \frac{1}{2}e^{(\lambda)} + \frac{1}{2}e^{(-\lambda)}$$
$$= \frac{1}{2}\left[1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots\right] + \frac{1}{2}\left[1 - \lambda + \frac{\lambda^2}{2!} - \frac{\lambda^3}{3!} + \cdots\right]$$
$$= 1 + \frac{\lambda^2}{2!} + \frac{\lambda^4}{4!} + \frac{\lambda^6}{6!} \leq e^{\lambda^2/2} \longrightarrow (*)$$

- Combine with (10):

$$E e^{\lambda S} \leq \prod e^{\lambda^2/2} = e^{\frac{n\lambda^2}{2}}$$

- From (9):

$$P[e^{\lambda S} \geq e^{10\lambda\sqrt{n\log n}}] \leq e^{\frac{n\lambda^2}{2} - 10\lambda\sqrt{n\log n}} \longrightarrow (11)$$

. Set $\lambda = 10\sqrt{\frac{\log n}{n}}$ that minimizes the R.H.S of (11) and we get

Verify!

$$P\left[e^{\lambda s} \geq e^{10\lambda\sqrt{n\log n}}\right] \leq e^{50\log n - 100\log n}$$

$$= \frac{1}{n^{50}}$$

- We can use the same pathway for general r.v.

8 Chernoff bound:

Thm 9: Let X_i i.i.d (Bernoulli) r.v., $E(X_i) = p_i$ $1 \leq i \leq n$

(i) $X_i = 1$ with p, $= 0$,, $1-p$. Let $X = \sum_{i=1}^{n} X_i$, $\mu = E(X)$

. Then for any $\delta > 0$:

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

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

Note: (i) This claim holds even for $X_i \in [0,1]$ instead of Bernoulli

(ii) For $\delta \in [0,1]$, this implies

$$P[X \geq (1+\delta)\mu] \leq e^{-\frac{\delta^2\mu}{3}}$$

$$P[X \geq (1-\delta)\mu] \leq e^{-\frac{\delta^2\mu}{3}}$$

Proof: Need only to prove the upper bound.

. For any $\lambda > 0$, by Markov:

$$P[X \geq (1+\delta)\mu] = P\left[e^{\lambda X} \geq e^{\lambda(1+\delta)\mu}\right] \leq \frac{E[e^{\lambda X}]}{e^{\lambda(1+\delta)\mu}} \quad \to (12)$$

· But

$$E[e^{\lambda x_i}] = p_i e^{\lambda} + (1+p_i) \quad \left[\because \begin{array}{l} x_i = 1 \to p_i \\ \quad = 0 \to 1-p_i \end{array}\right]$$

$$= 1 + p_i(e^{\lambda} - 1)$$

$$\leq \exp\left[p_i(e^{\lambda}-1)\right] \quad \left[\because 1+x \leq e^{x},\ x \in \mathbb{R} \text{ small}\right]$$

$$\therefore E[e^{\lambda X}] = \prod_{i=1}^{n} E[e^{\lambda x_i}] = \exp\left[\sum_i p_i(e^{\lambda}-1)\right]$$

$$= \exp\left[(e^{\lambda}-1) \underline{\underline{\Sigma p_i}}\right]$$

$$\because \boxed{E(X) = \Sigma p_i = \mu}$$

$$= \exp\left[(e^{\lambda}-1)\mu\right]$$

· Combine with (12)

$$P[X \geq (1+\delta)\mu] \leq \exp[\mu(e^{\lambda}-1)] / \exp[\lambda(1+\delta)\mu] \longrightarrow (*)$$

· Set $\lambda = \log(1+\delta) \geq 0$. $\Rightarrow e^{\lambda} = 1+\delta$ } index of Nr. of (*)

$\Rightarrow \mu(e^{\lambda}-1) = \underline{\mu\delta}$

· Also $\lambda(1+\delta)\mu = \mu(1+\delta)\log(1+\delta)$

$$= \log(1+\delta)^{\mu(1+\delta)}$$

$$\therefore e^{\lambda(1+\delta)\mu} = \underbrace{(1+\delta)^{\mu(1+\delta)}}$$

· Substituting:

$$P[X \geq (1+\delta)\mu] \leq \left[\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right]^{\mu} \longrightarrow \boxed{\text{upper bound}}$$

$\boxed{\text{Other Tail}}$

9) **Hoeffding** x_i i.i.d

$x_i \in [a_i, b_i]$. $X = \sum_i x_i$

$$P[X \geq t + E(X)] \leq \exp\left[\frac{-t^2}{\sum_i (b_i - a_i)^2}\right]$$

where $t > 0$

Bernstein x_i are i.i.d

$|x_i| \leq M$. $X = \Sigma x_i$, $E(x_i) = 0$

$$P[X \geq t] \leq \exp\left[\frac{-t^2}{\sum_i E(x_i^2) + \frac{1}{3}Mt}\right]$$