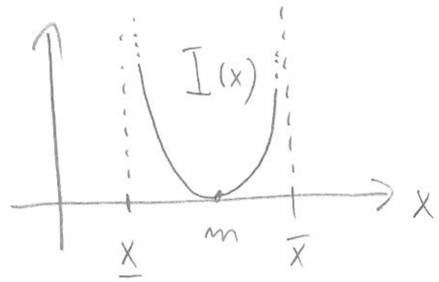
RECALL: $X_1, X_2, \ldots$ I.I.D., $S_n = X_1 + \ldots + X_n$,

$$Z(\lambda) = \mathbb{E}\left(e^{\lambda \cdot X_i}\right), \quad \hat{I}(\lambda) = \ln\left(Z(\lambda)\right)$$

$$I(x) = \max_{\lambda} \left\{ \lambda x - \hat{I}(\lambda) \right\}$$



CRAMÉR'S THM IMPLIES:

IF $x \in (\underline{x}, \overline{x})$ THEN

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \ln\left(\mathbb{P}\left(\frac{S_n}{n} \in [x-\varepsilon, x+\varepsilon]\right)\right) = -I(x)$$

HEURISTICALLY: $\boxed{\mathbb{P}\left(\frac{S_n}{n} \approx x\right) \approx e^{-n \cdot I(x)}}$

A HEU. EXPLANATION OF CONVEXITY OF $I(x)$:

WANT: $I\left(\frac{x_1 + x_2}{2}\right) \leq \frac{1}{2}\left(I(x_1) + I(x_2)\right)$

INDEED: $S_{2n} - S_n = X_{n+1} + \ldots + X_{2n}$

$$e^{-2n \cdot I\left(\frac{x_1+x_2}{2}\right)} \approx \mathbb{P}\left(\frac{S_{2n}}{2n} \approx \frac{x_1+x_2}{2}\right) =$$

$$\mathbb{P}\left(S_{2n} \approx n \cdot x_1 + n \cdot x_2\right) \geq \mathbb{P}\left(S_n \approx n \cdot x_1,\ S_{2n} - S_n \approx n \cdot x_2\right)$$

$$= \mathbb{P}\left(S_n \approx n \cdot x_1\right) \cdot \mathbb{P}\left(S_{2n} - S_n \approx n \cdot x_2\right) \approx$$

$$\approx e^{-n \cdot I(x_1)} \cdot e^{-n \cdot I(x_2)} = e^{-2n \cdot \frac{1}{2} \cdot \left(I(x_1) + I(x_2)\right)}$$

✓

A HEU. EXPLANATION OF THE FACT THAT INDEED $\hat{I}(\lambda) = \max\limits_{x} \{ \lambda \cdot x - I(x) \}$:

ASSUME THAT X_i ARE ABS. CONT. :

DENOTE BY $f_n(x)$ THE P.D.F. OF $\frac{S_n}{n}$

HEU. CRAMÉR $\Longrightarrow$ $\boxed{f_n(x) \underset{C}{\approx} e^{-n \cdot I(x)}}$

$$(z(\lambda))^n = \mathbb{E}\left(e^{\lambda \cdot S_n}\right) = \mathbb{E}\left(e^{n \cdot \lambda \cdot \frac{S_n}{n}}\right) =$$

$$= \int_{-\infty}^{\infty} e^{n \cdot \lambda \cdot x} \cdot f_n(x)\, dx \approx \int_{-\infty}^{\infty} e^{n \cdot \lambda \cdot x} \cdot e^{-n \cdot I(x)}\, dx \text{, THUS}$$

$$\hat{I}(\lambda) = \lim_{n \to \infty} \frac{1}{n} \ln\left(z(\lambda)^n \right) =$$

$$= \lim_{n \to \infty} \frac{1}{n} \ln\left(\int_{-\infty}^{\infty} e^{n \cdot (\lambda x - I(x))}\, dx \right) \overset{\text{LAPLACE PRINCIPLE HW 1.2}}{=}$$

$$= \sup_{x} \{ \lambda x - I(x) \} \checkmark$$

EX: $X_1 \sim GEO(p)$ (OPTIMISTIC) :

$$\mathbb{P}(X_1 = k) = (1-p)^{k-1} \cdot p, \quad k = 1, 2, 3, \ldots$$

$$z(\lambda) = \mathbb{E}\left(e^{\lambda \cdot X_1}\right) = \sum_{k=1}^{\infty} e^{\lambda \cdot k} \cdot (1-p)^{k-1} \cdot p =$$

$$= p\cdot e^{\lambda}\cdot \sum_{r=1}^{\infty}\left(e^{\lambda}\cdot(1-p)\right)^{r-1} = \begin{cases} +\infty & \text{IF } \lambda \geq \ln\left(\frac{1}{1-p}\right) \\ \dfrac{p\cdot e^{\lambda}}{1-e^{\lambda}\cdot(1-p)} & \text{IF } \lambda < \ln\left(\frac{1}{1-p}\right) \end{cases}$$

$$\hat{I}(\lambda) = \ln(z(\lambda)) = \ldots = -\ln\left(\frac{e^{-\lambda}-(1-p)}{p}\right), \text{ THUS}$$

$$\hat{I}'(\lambda) = \ldots = \frac{1}{1-e^{\lambda}\cdot(1-p)}$$

$$\hat{I}'(\lambda^*) = x \Rightarrow \ldots \Rightarrow \lambda^* = \ln\left(\frac{1-\frac{1}{x}}{1-p}\right) \quad \text{IF } x > 1$$

$$I(x) = \lambda^*\cdot x - \hat{I}(\lambda^*) = \ln\left(\frac{1-\frac{1}{x}}{1-p}\right)\cdot x + \ln\left(\frac{\frac{1-p}{1-1/x}-(1-p)}{p}\right)$$

$$= \ldots = (x-1)\cdot\ln\left(\frac{x-1}{1-p}\right) - x\cdot\ln(x) - \ln(p) \quad \text{IF } x \geq 1$$

BUT IF $x < 1$ THEN

$$I(x) = \sup_{\lambda}\{\lambda x - \hat{I}(\lambda)\} = +\infty \quad \text{SINCE}$$

$$\lim_{\lambda\to-\infty} \lambda x - \hat{I}(\lambda) = \lim_{\lambda\to-\infty} \lambda\cdot x + \ln\left(\frac{e^{-\lambda}-(1-p)}{p}\right) \geq$$

$$\lim_{\lambda\to-\infty} \lambda\cdot x - \ln\left(\frac{e^{-\lambda}}{p}\right) = +\infty \quad \text{IF } x < 1$$

EX: IF $X \sim GEO(p)$, WHAT IS THE DISTRIBUTION OF THE EXPONENTIALLY TILTED R.V. $X^{(\lambda)}$?

SOLUTION: LET $k = 1, 2, 3, \dots$

$$\mathbb{P}(X^{(\lambda)} = k) = \frac{1}{Z(\lambda)} \cdot e^{\lambda \cdot k} \cdot \mathbb{P}(X = k) =$$

$$= \frac{1 - e^{\lambda} \cdot (1-p)}{p \cdot e^{\lambda}} \cdot e^{\lambda \cdot k} \cdot (1-p)^{k-1} \cdot p =$$

$$= (1 - e^{\lambda} \cdot (1-p)) \cdot e^{\lambda \cdot (k-1)} \cdot (1-p)^{k-1} =$$

$$= (e^{\lambda} \cdot (1-p))^{k-1} \cdot (1 - e^{\lambda} \cdot (1-p)), \quad k = 1, 2, 3, \dots$$

THUS $\boxed{X^{(\lambda)} \sim GEO(1 - e^{\lambda} \cdot (1-p))}$

IF $\lambda < \ln\left(\frac{1}{1-p}\right)$

THUS IT FOLLOWS FROM CRAMÉR'S THM

THAT IF $X_1, X_2, \dots$ I.I.D. GEO(P) AND

$x > \mathbb{E}(X_i) = \frac{1}{p}$ THEN $\boxed{\mathbb{P}\left(\frac{S_n}{n} > x\right) \approx e^{-n \cdot I(x)}}$

WHERE $\boxed{I(x) = (x-1)\cdot \ln\left(\frac{x-1}{1-p}\right) - x \cdot \ln(x) - \ln(p)}$

NOTE THIS RELATION TO TO THE LARGE DEV. THM. ABOUT THE SUM OF I.I.D. BER(P) R.V.'S (SEE PAGE 2):

IF $Y_1, Y_2, \dots$ I.I.D. $\mathbb{P}(Y_i = 1) = 1 - \mathbb{P}(Y_i = 0) = p$

THEN

$$\begin{array}{|c|c|c|c|c|c|c|c}
Y_1 & Y_2 & Y_3 & Y_4 & Y_5 & Y_6 & Y_7 & \dots \\
0 & 0 & 1 & 1 & 0 & 1 & 1 &
\end{array}$$

$$\underbrace{0 \quad 0 \quad 1}_{\boxed{X_1 = 3}} \quad \underbrace{1}_{\boxed{X_2 = 1}} \quad \underbrace{0 \quad 1}_{\boxed{X_3 = 2}} \quad \underbrace{1}_{\boxed{X_4 = 1}}$$

THUS $e^{-n \cdot I(x)} \approx \mathbb{P}(X_1 + \dots + X_n > \lfloor n \cdot x \rfloor) =$

$\mathbb{P}($THE n'TH SUCCESS COMES AFTER $\lfloor n \cdot x \rfloor$'TH TRIAL$) =$

$\mathbb{P}(Y_1 + \dots + Y_{\lfloor n \cdot x \rfloor} < n) \approx \exp\left(-\lfloor n \cdot x \rfloor \cdot J\left(\frac{1}{x}\right)\right)$

(SEE PAGE 2)

WHERE $\boxed{J(x) = x \cdot \ln\left(\frac{x}{p}\right) + (1-x)\cdot \ln\left(\frac{1-x}{1-p}\right)}$

AND INDEED: $\boxed{I(x) = x \cdot \mathcal{F}\left(\frac{1}{x}\right)}$ ✓

LET $X_1, X_2, \ldots$ INDEPENDENT (BUT NOT NECESSARILY IDENTICALLY DISTRIBUTED.

$\boxed{S_n = X_1 + \ldots + X_n}$ CAN WE BOUND THE LARGE DEVIATION PROBABILITIES OF S_n IF WE DON'T KNOW TOO MUCH ABOUT THE DISTRIBUTION OF X_i ?

<u>THM</u> (HOEFFDING'S INEQUALITY, 1963)

IF $\mathbb{P}(a_i \leq X_i \leq b_i) = 1$ FOR $a_i, b_i \in \mathbb{R}$ THEN

$$\boxed{\mathbb{P}(S_n \geq \mathbb{E}(S_n) + t) \leq \exp\left(-\frac{2t^2}{(b_1-a_1)^2 + \ldots + (b_n-a_n)^2}\right)}$$

<u>PROOF</u>: W.L.O.G. WE MAY ASSUME $\boxed{\mathbb{E}(X_i) = 0}$ (I.E., THAT X_i IS <u>CENTERED</u>)

<u>LEMMA</u>: IF $\mathbb{E}(X) = 0$, $\mathbb{P}(a \leq X \leq b) = 1$

THEN $\boxed{Z(\lambda) \leq \exp\left(\frac{1}{8} \cdot \lambda^2 \cdot (b-a)^2\right)}$

PROOF OF LEMMA:

SUBLEMMA: $\boxed{\mathbb{P}(a \le Y \le b) = 1} \Rightarrow \boxed{Var(Y) \le \frac{(b-a)^2}{4}}$

PROOF OF SUBLEMMA: W.L.O.G. WE MAY ASSUME THAT $\boxed{a = -b}$ (JUST ADD AN APPROPRIATE CONSTANT TO Y). THEN

$$Var(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 \le \mathbb{E}(Y^2) \le b^2 = \frac{(b-a)^2}{4}$$

PROOF OF LEMMA USING SUBLEMMA:

WANT: $\hat{I}(\lambda) \le \frac{1}{8} \cdot \lambda^2 \cdot (b-a)^2$

KNOW: $\hat{I}(0) = 0$, $\hat{I}'(0) = \mathbb{E}(X) = 0$

$$\hat{I}''(\lambda) = Var(X^{(\lambda)}) \le \frac{(b-a)^2}{4}$$

SINCE $\mathbb{P}(a \le X^{(\lambda)} \le b) = 1$

THUS $\hat{I}'(\lambda) = \underbrace{\hat{I}'(0)}_{0} + \int_0^\lambda \hat{I}''(\mu)\, d\mu \le \lambda \cdot \frac{(b-a)^2}{4}$

$$\hat{I}(\lambda) = \underbrace{\hat{I}(0)}_{0} + \int_0^\lambda \hat{I}'(\mu)\, d\mu \le \int_0^\lambda \mu \cdot \frac{(b-a)^2}{4}\, d\mu = \frac{1}{8} \cdot \lambda^2 \cdot (b-a)^2$$

✓

PROOF OF HOEFFDING USING LEMMA:

$$\mathbb{E}\left(e^{\lambda\cdot S_n}\right) = \prod_{k=1}^{n} \mathbb{E}\left(e^{\lambda\cdot X_k}\right) \leq \prod_{i=1}^{n} \exp\left(\frac{1}{8}\cdot\lambda^2\cdot(b_i-a_i)^2\right) =$$

$$= e^{\frac{1}{2}\sigma^2\cdot\lambda^2}, \quad \text{WHERE } \sigma^2 = \frac{1}{4}\cdot\left((b_1-a_1)^2+\dots+(b_n-a_n)^2\right)$$

$$\mathbb{P}(S_n \geq t) = \mathbb{E}\left(e^{\lambda\cdot S_n} \geq e^{\lambda t}\right) \leq \frac{\mathbb{E}\left(e^{\lambda S_n}\right)}{e^{\lambda t}} \leq e^{\frac{1}{2}\sigma^2\cdot\lambda^2-\lambda t}$$

NOW $\min\limits_{\lambda\geq 0}\left\{\frac{1}{2}\sigma^2\cdot\lambda^2-\lambda t\right\} = -\frac{1}{2}\frac{t^2}{\sigma^2}$

SEE PAGE 14

EX: TOWN WITH 1000 HOUSEHOLDS, ONE GARBAGE CAN PER HOUSEHOLD. CAPACITY OF A GARBAGE CAN: 30 kg. THE AVERAGE WEEKLY GARBAGE OUTPUT MAY DIFFER FROM HOUSEHOLD TO HOUSEHOLD. AVERAGE WEEKLY OUTPUT OF TOWN IS 10^4 kg. CAPACITY OF ONE TRUCK: 10^3 kg. HOW MANY TRUCKS DO WE NEED IF WE WANT ALL GARBAGE TAKEN AWAY WITH 95% CHANCE?

SOLUTION: $a_i = 0$, $b_i = 30$

$$\mathbb{P}\left(S_{1000} \geq 10^4 + t\right) \leq \exp\left(-\frac{2t^2}{1000\cdot 30^2}\right) = \frac{1}{20} \implies$$

$$\implies t = \sqrt{\frac{1}{2}\ln(20)\cdot 9\cdot 10^5} \approx 1161$$

$\boxed{\text{WE NEED 12 TRUCKS.}}$

PAGE 32