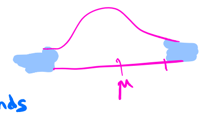
Concentration of measure (tail bounds)



The more you know about r.v. X, the better the bounds

① If only thing you know is mean & $X \geq 0$ (but not $= 0$)

Markov's Inequality

$$\boxed{Pr(X \geq t) \leq \frac{E(X)}{t}} \quad t \geq 0$$

- tight
- doesn't hold if r.v. can be -ve
- can't be used for $Pr(X \leq t)$

Pf $\quad Pr(X \geq t) = \int_t^\infty f(x)\,dx \leq \int_t^\infty \frac{x}{t} f(x)\,dx \leq \frac{E(X)}{t}$

$$\boxed{\left[Pr(X \geq c\,E(X)) \leq \frac{1}{c} \right] \quad c \geq 0}$$

Corollary

If $0 \leq X \leq B \implies \underbrace{Pr(\underbrace{B-X}_{Y} \geq \underbrace{B-t}_{})}_{Pr(X \leq t)} \leq \frac{E(B-X)}{B-t}$

② If you know mean μ and variance σ^2

Chebychev's Inequality (2nd moment method)

$\forall t > 0 \quad Pr(|X-\mu| \geq t\underbrace{\sigma}_{\text{std deviation}}) \leq \frac{1}{t^2}$

Markov's Ineq.

Pf: $\quad Pr(|X-\mu| \geq t\sigma) = Pr(\underbrace{(X-\mu)^2}_{Y} \geq t^2 \underbrace{\sigma^2}_{E(Y)}) \leq \frac{1}{t^2}$

$E[Y] = E[(X-\mu)^2] = \sigma^2$

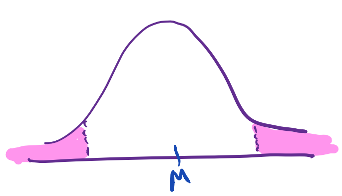
③ $X = X_1 + X_2 + \dots + X_n$

$X \sim Bin(n, p)$

X_i iid. $Ber(p)$

$\mu_X = np \quad \sigma_X^2 = np(1-p)$

$\sigma_X = \sqrt{np(1-p)}$



$Pr(X \geq k) = \sum_{i \geq k} \binom{n}{i} p^i (1-p)^{n-i}$

Law of Large Numbers

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} X_i}{n} = p$$

Central Limit Theorem

$$\frac{X - n\mu}{\sqrt{n}\,\sigma} \longrightarrow N(0,1)$$

specifically $Pr(|X-\mu| > \beta\sigma) \to \frac{1}{\sqrt{2\pi}} \int_{\beta}^{\infty} e^{-t^2/2}\, dt \approx \frac{1}{\sqrt{2\pi}\,\beta} e^{-\beta^2/2}$

Issue:

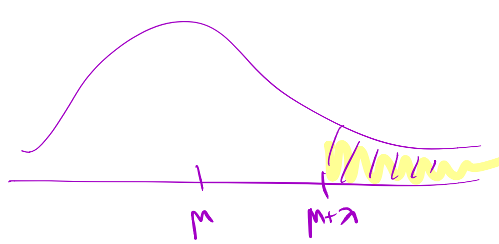
- CLT is asymptotic; no info about rate of convergence
- Applies to deviations on order of σ; want bounds for arbitrary deviations

try extending ideas using higher even moments

$$Pr(|X-\mu| \geq t) = Pr((X-\mu)^4 \geq t^4) \leq \frac{E[(X-\mu)^4]}{t^4}$$

$E(X^k)$ k^{th} moment of r.v. X

$E(X^2)$



$$\Pr(X-\mu \geq \lambda) = \Pr(X \geq \lambda+\mu) = \Pr(tX \geq t(\lambda+\mu))$$
$$= \Pr(e^{tX} \geq e^{t(\lambda+\mu)}) \leq \boxed{\frac{E[e^{tX}]}{e^{t(\lambda+\mu)}}}$$

$$\frac{d^k M_X(t)}{dt^k}\Big|_{t=0} = E[X^k]$$

check.

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

$M_X(t) \triangleq \boxed{E[e^{tX}]}$ called $\boxed{\text{moment generating fn of } X}$

$$= E\left(\sum_{k\geq 0} \frac{t^k}{k!} X^k\right) = \sum_{k\geq 0} \frac{t^k}{k!} E[X^k]$$

$X = X_1 + X_2 + \cdots + X_n$

$X_i \sim \mathrm{Ber}(p)$

$$E[e^{tX}] = E[e^{tX_1 + tX_2 + \cdots + tX_n}] = E[e^{tX_1} e^{tX_2} \cdots e^{tX_n}]$$
$$= E[e^{tX_1}] \cdots E[e^{tX_n}] = (1-p+pe^t)^n$$

X, Y indep

$E(XY) = E(X)E(Y)$

$E(f(X)g(Y)) = E[f(X)]E[g(Y)]$

$$E[e^{tX_i}] = e^{t0} \Pr(X_i=0) + e^{t1} \Pr(X_i=1)$$
$$= (1-p) + pe^t$$

$X_i \sim \mathrm{Ber}(p)$

$$\Pr(X-\mu \geq \lambda) \leq \frac{(1-p+e^t p)^n}{e^{t(\lambda+\mu)}}$$

use calculus to choose t to minimize this

optimal choice $\quad e^t = \boxed{\frac{q\,(p+\frac{\lambda}{n})}{p\,(q-\frac{\lambda}{n})}} \qquad \boxed{q = 1-p}$

Plug this back in get

$$\Pr(X-\mu \geq \lambda) \leq \left[\left(\frac{p}{p+\frac{\lambda}{n}}\right)^{p+\frac{\lambda}{n}} \left(\frac{q}{q-\frac{\lambda}{n}}\right)^{q-\frac{\lambda}{n}}\right]^n$$

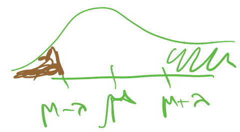
$$x^n = e^{n \ln x}$$

$$= e^{n\underbrace{\left[(p+\frac{\lambda}{n})\ln\left(\frac{p}{p+\frac{\lambda}{n}}\right) + (q-\frac{\lambda}{n})\ln\left(\frac{q}{q-\frac{\lambda}{n}}\right)\right]}_{-\text{ relative entropy}}}$$

relative entropy $(x, 1-x) \leftrightarrow (p, 1-p)$

$$H(x \| p) = x \ln\left(\frac{x}{p}\right) + (1-x) \ln\left(\frac{1-x}{1-p}\right)$$

$$\boxed{\Pr(X \geq \mu + \lambda) \leq e^{-n H\left(p + \frac{\lambda}{n} \| p\right)}}$$



Corollary $\Pr(X \leq \mu - \lambda) \leq e^{-n H\left(1-p+\frac{\lambda}{n} \| 1-p\right)}$

np $\qquad \frac{E(X)}{n} = p$

$$\Pr(\boxed{\underbrace{n - X}_{Y} \geq \underbrace{n - \mu}_{E(Y)} + \lambda})$$

$$\frac{E(Y)}{n} = 1 - p$$

$$X \leq \mu - \lambda$$

Now suppose $X = X_1 + X_2 + \dots + X_n$ where $X_i \sim Ber(p_i)$ and mutually independent

$\mu = E(X) = \sum_{i=1}^{n} p_i$

$p = \frac{\mu}{n}$

$$Pr(X - \mu \ge \lambda) = Pr(X \ge \lambda + \mu) = Pr(e^{tX} \ge e^{t(\lambda+\mu)}) \le \frac{E[e^{tX}]}{e^{t(\lambda+\mu)}}$$

$$E[e^{tX}] = E[e^{tX_1}]E[e^{tX_2}]\cdots E[e^{tX_n}] = \prod_{i=1}^{n}[(1-p_i)+p_ie^t] \le (e^t p + (1-p))^n$$

claim.

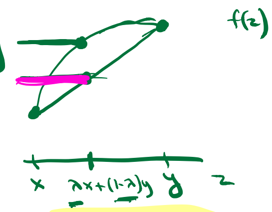
$$E[e^{tX_i}] = (1-p_i) + p_ie^t = 1 + (e^t-1)p_i$$

AMGM $a_1, \dots, a_n \ge 0$

$$\left(\prod_{i=1}^{n} a_i\right)^{1/n} \le \frac{1}{n}\sum_{i=1}^{n} a_i$$

apply concavity $f(z) = \ln((e^t-1)z+1)$ $[f''(z) < 0]$

and Jensen's Inequality $[f(E(X)) \ge E[f(X)]]$



$$\prod_{i=1}^{n}[(1-p_i)+p_ie^t] = e^{\sum_{i=1}^{n}\ln(1-p_i+p_ie^t)} \le e^{n\ln(1-p+pe^t)}$$

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n}\ln(1-p_i+p_ie^t)}_{\frac{1}{n}\sum_{i=1}^{n} f(p_i)} \le \ln(1-p+pe^t) = f\Big(\underbrace{\frac{1}{n}\sum_{i=1}^{n} p_i}_{p}\Big)$$

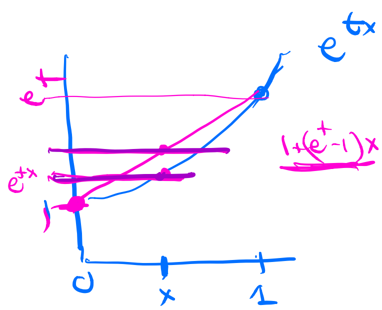
$E[f(X)]$ X takes value p_i with prob $\frac{1}{n}$

$f(\lambda x + (1-\lambda)y) \ge \lambda f(x) + (1-\lambda)f(y)$

$X = \begin{cases} x & \text{wp } \lambda \\ y & \text{wp } (1-\lambda) \end{cases}$

Finally, suppose $X_i \in [0,1]$ with $E(X_i) = p_i$

e^{tx} is convex fn of x



$$e^{tx} \le 1 + (e^t - 1)x$$

$$E[e^{tX_i}] \le E[1 + (e^t-1)X_i] = 1 + (e^t-1)p_i$$

Theorem $X = X_1 + X_2 + \dots + X_n$ where X_i's indep & $X_i \in [0,1]$ with $E(X_i) = p_i$

$\mu = E(X) = \sum_{i=1}^{n} p_i \qquad p = \frac{\mu}{n}$

$$Pr(X \ge \mu + \lambda) \le e^{-nH(p+\frac{\lambda}{n}\|p)} \quad (1)$$

$$Pr(X \le \mu - \lambda) \le e^{-nH(1-p+\frac{\lambda}{n}\|1-p)} \quad (2)$$

$H(x\|p) = x\ln\left(\frac{x}{p}\right) + (1-x)\ln\left(\frac{1-x}{1-p}\right)$ relative entropy

More useful forms (corollaries)

$$\boxed{\begin{array}{l} \Pr(X \geq \mu + \lambda) \\ \Pr(X \leq \mu - \lambda) \end{array} \leq e^{-\frac{2\lambda^2}{n}}} \quad (3)$$

Pf just show $\quad n\,H\left(p+\frac{\lambda}{n} \| p\right) \geq \frac{2\lambda^2}{n}$

$$\equiv \quad H(p+z \| p) \geq 2z^2 \qquad z = \frac{\lambda}{n}. \qquad \text{for } 0 \leq z \leq 1-p$$

$$\equiv \quad f(z) = (p+z)\ln\left(\frac{p+z}{p}\right) + (1-p-z)\ln\left(\frac{1-p-z}{1-p}\right) - 2z^2 \geq 0$$

for $0 \leq z \leq 1-p$

$$\begin{cases} f(0) = 0 \\ f'(z) \geq 0 \qquad 0 \leq z \leq 1-p \end{cases}$$

$\star$ $\quad p = \frac{\mu}{n}$

$$\boxed{\begin{array}{ll} \Pr(X \leq (1-\delta)\mu) \leq e^{-\frac{\delta^2 \mu}{2}} \qquad 0 \leq \delta \leq 1 & \begin{array}{l} X = X_1 + X_2 + \dots + X_n \\ X_i \in [0,1] \quad E(X_i) = p_i \\ \mu = \sum p_i \end{array} \\ \Pr(X \geq (1+\delta)\mu) \leq e^{-\mu[(1+\delta)\ln(1+\delta) - \delta]} = \begin{cases} e^{-\frac{\delta^2 \mu}{2+\delta}} & \delta > 0 \\ e^{-\frac{\delta^2 \mu}{3}} & 0 < \delta \leq 1 \end{cases} & \end{array}} \quad (4)$$

$\mu_L \leq \mu$

$\Pr(X \leq (1-\delta)\mu_L) \leq e^{-\frac{\delta^2 \mu_L}{2}}$

$\mu \leq \mu_H$

hold with μ replaced by μ_H

$$\Pr(X \leq (1-\delta)\mu) \leq e^{n\left[(1-p)\ln\left(\frac{1-p}{1-p+\delta p}\right) + (p-\delta p)\ln\left(\frac{p}{p-\delta p}\right)\right]}$$

simplifying the exponent using

$|x| < 1$

$$\boxed{\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} \dots}$$

$$\boxed{(1-x)\ln(1-x) \geq -x + \frac{x^2}{2}}$$

for $x > 0 \quad \ln(1-x) \leq -x$

More details for proof of (4)

$$\Pr(X \le \mu - \lambda) \le e^{-n \underbrace{H(1-p+\frac{\lambda}{n} \| 1-p)}}$$

$\lambda = \delta\mu \Rightarrow \frac{\lambda}{n} = \delta p$

$\mu = pn$

$$(1-p+\delta p)\ln\left(\frac{1-p+\delta p}{\delta p}\right) + (p-\delta p)\ln\left(\frac{p-\delta p}{p}\right)$$

$$= -\left[(1-p+\delta p)\ln\left(\frac{\delta p}{1-p+\delta p}\right) + (p-\delta p)\ln\left(\frac{1}{1-\delta}\right)\right]$$

Therefore,

$$\Pr(X \le (1-\delta)\mu) \le e^{n\left[(1-p+\delta p)\ln\left(1-\frac{\delta p}{1-p+\delta p}\right) + p(1-\delta)\ln\left(\frac{p}{p-\delta p}\right)\right]}$$

$|x| < 1 \quad \ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} \cdots$

$$\le e^{n[-\delta p - p(1-\delta)\ln(1-\delta)]}$$

$$= e^{-pn[\delta + (1-\delta)\ln(1-\delta)]}$$

$(1-x)\ln x \ge -x + \frac{x^2}{2}$

$$\le e^{-pn[\delta - \delta + \frac{\delta^2}{2}]}$$

$$= e^{-pn\frac{\delta^2}{2}}$$

$\boxed{\ln(1-x) \le -x}$

$(1-x)\ln(1-x)$

$= -(1-x)\left[x + \frac{x^2}{2} + \frac{x^3}{3} + \dots\right]$

$= -x - \frac{x^2}{2} - \frac{x^3}{3} \cdots$

$\quad + x^2 + \frac{x^3}{2} + \frac{x^4}{4} \cdots$

$\ge -x + \frac{x^2}{2}$

$$\Pr(X \ge (1+\delta)\mu) \le e^{-n H(p+\delta p \| p)}$$

$$= e^{n\left[(1-p-\delta p)\underbrace{\ln\left(\frac{1-p}{1-p-\delta p}\right)} + (p+\delta p)\underbrace{\ln\left(\frac{p}{p+\delta p}\right)}\right]}$$

$\ln\left(1 + \frac{\delta p}{1-p-\delta p}\right) \qquad -\ln(1+\delta)$

$\le \frac{\delta p}{1-p-\delta p}$

$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots \Rightarrow \boxed{\ln(1+x) \le x}$

$(1+\frac{x}{2})\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots$

$\quad + \frac{x^2}{2} - \frac{x^3}{4} + \frac{x^4}{6} - \dots$

$\ge x$

$\Rightarrow \boxed{\ln(1+x) \ge \frac{x}{1+\frac{x}{2}}}$

$$= e^{n[\delta p - p(1+\delta)\ln(1+\delta)]}$$

$$= e^{-np[-\delta + (1+\delta)\ln(1+\delta)]}$$

$$\le e^{-np\left[-\delta + (1+\delta)\frac{\delta \cdot 2}{(1+\frac{\delta}{2}) \cdot 2}\right]}$$

$$= e^{-np\left[\frac{\delta^2}{2+\delta}\right]}$$

$-\delta + \frac{(1+\delta)2\delta}{2+\delta}$

$-\frac{(2+\delta)\delta}{2+\delta} + \frac{(1+\delta)2\delta}{2+\delta} = \frac{-2\delta - \delta^2 + 2\delta + 2\delta^2}{2+\delta} = \frac{\delta^2}{2+\delta}$

$X = X_1 + X_2 + \dots + X_n$

X_i's indep. $\quad X_i$ takes values in $[a_i, b_i]$ $\quad E(X) = \mu$

$$\left.\begin{array}{l} \Pr(X \le \mu - \lambda) \\ \Pr(X \ge \mu + \lambda) \end{array}\right\} \le e^{-\frac{2\lambda^2}{\sum_{i=1}^{n} (b_i - a_i)^2}} \qquad (5)$$

A variance bound

$X = X_1 + \dots + X_n \quad X_i \in [0,1] \quad E(X_i) = p_i \quad Var(X_i) = \sigma_i^2$

$E(X) = \mu = \sum_{i=1}^{n} p_i \qquad Var(X) = \sum_{i=1}^{n} \sigma_i^2 \triangleq \sigma^2$

$$\Pr(X > \mu + \lambda) = \Pr\left(\sum_i (X_i - p_i) > \lambda \right)$$

$$= \Pr\left(e^{t \sum_i (X_i - p_i)} > e^{\lambda t} \right)$$

$$\le \frac{E\left[e^{t \sum_i (X_i - p_i)} \right]}{e^{\lambda t}}$$

$e^x \le 1 + x + x^2 \quad 0 \le |x| \le 1$

$1 + x \le e^x$

$$\le \frac{\prod_{i=1}^{n} E\left[e^{t(X_i - p_i)} \right]}{e^{\lambda t}}$$

$$\le \frac{\prod_{i=1}^{n} E\left[1 + t(X_i - p_i) + t^2 (X_i - p_i)^2 \right]}{e^{\lambda t}}$$

if $|t(X_i - p_i)| < 1 \doteq t \max(p_i, 1 - p_i) \le 1$

$$\le \frac{\prod_{i=1}^{n} (1 + t^2 \sigma_i^2)}{e^{\lambda t}} \le \underbrace{\frac{e^{t^2 \sigma^2}}{e^{\lambda t}}}_{\text{choose } t \text{ to min this}}$$

$t = \frac{\lambda}{2\sigma^2}$

$$\le e^{-\frac{\lambda^2}{4\sigma^2}}$$

Bernstein's Inequality

$X_1, \dots, X_n \qquad X_i - E(X_i) \le b \quad \forall i$

$X = X_1 + X_2 + \dots + X_n \qquad \sigma^2 = \sigma_1^2 + \dots + \sigma_n^2$

$$\Pr(X > E(X) + \lambda) \le e^{\left[-\frac{\lambda^2}{2\sigma^2 (1 + b\lambda / 3\sigma^2)} \right]}$$