Example: Gambling Game

$$\begin{cases} \text{you pay me \$10} & .99 \\ \text{I pay you \$1000} & .01 \end{cases}$$

X : your expected gain

$E(X) = 1000 \cdot 0.01 - 10 \cdot 0.99 = .1$ (10 cents)

$Pr(X \geq .1) = 0.01$

$= Pr(\text{you win} \geq E(X))$

$Y = \text{my winnings} = -X$

$E(Y) = -0.1$ $\qquad Pr(Y \geq -0.1) = 0.99$

<u>Conclusion:</u> r.v. might almost never be $\geq$ exp

or might almost always be $\geq$ exp.

$Q = \#$ comparisons made by rand QS

$E(Q) = 2n \ln n$

Could $\Pr(Q \geq cn^2)$ be high ?

$$E(Q) = \sum_{0 \leq i < cn^2} i \Pr(Q=i) + \sum_{i \geq cn^2} i \Pr(Q=i)$$

$$\geq \sum_{i \geq cn^2} i \Pr(Q=i) \geq cn^2 \Pr(Q \geq cn^2)$$

$$\therefore \Pr(Q \geq cn^2) \leq \frac{E(Q)}{cn^2} = O\left(\frac{\ln n}{n}\right) \longrightarrow 0$$

<u>Ex</u> $n = 2^{10} \Rightarrow \approx \frac{10}{2^{10}}$

Markov Inequality

$$\Pr(X \geq \alpha) \leq \frac{E(X)}{\alpha}$$

X nonnegative

Chebyshev Inequality

$$\Pr(|X - \mu| \geq \alpha) \leq \frac{Var(X)}{\alpha^2}$$

Example: $X \sim Bin(n, \frac{1}{2})$

$$\Pr\left(X \geq \frac{3}{4}n\right) \leq \frac{\frac{n}{2}}{\frac{3n}{4}} = \frac{2}{3}$$

Markov

$$\Pr\left(X \geq \frac{3}{4}n\right) \leq \Pr\left(\left|X - \frac{n}{2}\right| \geq \frac{n}{4}\right) \leq \frac{Var(X)}{\left(\frac{n}{4}\right)^2} = \frac{\frac{n}{4}}{\left(\frac{n}{4}\right)^2} = \frac{4}{n} \xrightarrow[n \to \infty]{} 0$$

Chebyshev

Sampling & Polling

What fraction of people approve of president?

<u>Poll</u>: call up n random people

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

Define $\bar{X} = \frac{X}{n}$ as our estimate

Questions:

What should n be? how confident are we?

How good an estimate?

can never say my polling estimate 100% guaranteed to be within $\pm 2\%$ of truth.

Question: Given θ, $1-\varepsilon$: how large does n need to be

so

$$\Pr(|\bar{X}-p| \le \underset{0.02}{\theta}) \ge 1 - \underset{0.95}{\varepsilon}$$

↑ margin of error ↑ confidence

$|\bar{X}-p| \le \theta$

$-\theta \le \bar{X} - p \le \theta$

$p - \theta \le \bar{X} \le p + \theta$

Apply Chernoff: $X_i \sim \mathrm{Ber}(p)$

$$\Pr(X > (1+\delta)pn) \le e^{-\delta^2\mu/2} = e^{-\frac{\delta^2 np}{2}}$$

$$pn - \theta n \le X \le pn + \theta n$$

$$\Pr(X < (1-\delta)pn) \le e^{-\frac{\delta^2 np}{3}}$$

$$\Rightarrow \Pr(|X - np| \ge \underbrace{\delta pn}_{\theta n}) \le 2e^{-\frac{\delta^2 np}{3}}$$

$$\Rightarrow \quad \delta p = \theta$$

$$\delta = \frac{\theta}{p}$$

$$\Pr(|X - np| \ge \delta pn) \le 2e^{-\frac{\delta^2 np}{3}}$$

$$\left(\frac{\theta}{p}\right)^2 \frac{np}{3} = \frac{\theta^2 n}{3p} > \frac{\theta^2 n}{3}$$

So we want $2e^{-\theta^2 n/3} \le \varepsilon$

$$\frac{2}{\varepsilon} \le e^{\frac{\theta^2 n}{3}}$$

$$\ln\left(\frac{2}{\varepsilon}\right) \le \frac{\theta^2 n}{3}$$

$$\boxed{\frac{3}{\theta^2}\ln\left(\frac{2}{\varepsilon}\right) \le n}$$

e.g. $\theta = 0.02$

$\varepsilon = 0.05$

Notes:

of samples n doesn't depend on size of total population

Ex $\theta = 0.02$ $1-\varepsilon = 0.95$

$$n \geq \frac{3}{(0.02)^2} \ln\left(\frac{2}{0.05}\right) \approx 186{,}000$$

really costly thing is high accuracy

confidence is cheap because of ln

100,000 computers

each indep sends packet w/ prob $q = 0.01$ each secs

Router processes its buffer each sec

How many packet buffers so it drops a packet:

never: 100,000

With prob $\leq 10^{-6}$ each hour?

$X_{it} = \begin{cases} 1 & \text{if computer } i \text{ sends a packet in } t^{th} \text{ second} \\ 0 & \text{otherwise} \end{cases}$

$X_t = \sum_{1 \leq i \leq n} X_{it}$ # of packets sent in a second

B: size of buffer needed so that in T secs prob of overflow $\leq 10^{-6}$

for what B is

$Pr(\exists t \; 1 \leq t \leq T \text{ s.t. } X_t \geq B) \leq \varepsilon$

First, $Pr(X_t \geq B)$

By Chernoff bound:

$$Pr(X > \underbrace{(1+\delta)\mu}_{\text{\# of buffers } B}) \leq e^{-\frac{\delta^2 \mu}{2}}$$

$$B = (1+\delta)\mu = (1+\delta)(1000)$$

$$Pr(\text{overflow in } T \text{ secs}) = Pr(\exists t, 1 \leq t \leq T \text{ s.t. } X_t > (1+\delta)\mu)$$

$$\leq T e^{-\frac{\delta^2 \mu}{2}}$$

$$\leq T e^{-\delta^2 \mu / 2}$$

We want $T e^{-\delta^2 \mu / 2} \leq \varepsilon$

where $\mu = 1000$

$\varepsilon = 10^{-6}$

T: # secs in year

Recipe: solve for δ

use that to determine $B = (1+\delta)\mu$

$$Te^{-\delta^2 \mu/2} \leq \varepsilon$$

$$e^{\frac{\delta^2 \mu}{2}} \geq \frac{T}{\varepsilon} \qquad \frac{\delta^2 \mu}{2} \geq \ln\left(\frac{T}{\varepsilon}\right)$$

$$\delta^2 \geq \frac{2}{\mu} \ln\left(\frac{T}{\varepsilon}\right)$$

$$\delta \geq \sqrt{\frac{2}{\mu} \ln\left(\frac{T}{\varepsilon}\right)}$$

$$B = (1+\delta)\mu$$

Example: $\mu = 1000$

$T = \underset{\text{min}}{60} \cdot \underset{\text{sec}}{60}$

$\varepsilon = 10^{-6}$

$$\delta = \sqrt{\frac{2}{1000} \ln\left(\frac{3600}{10^{-6}}\right)} = 0.2097$$

Buffer size $= 1.2097 \times 1000 \approx 1210$