Example: Gambling Game

\[
\begin{align*}
\{ & \text{you pay me $10} \quad .99 \\
& \text{I pay you $1000} \quad .01
\end{align*}
\]

\[X : \text{your expected gain}\]

\[
E(X) = 1000 \cdot 0.01 - 10 \cdot 0.99 = .1 \quad (10 \text{ cents})
\]

\[\Pr(X \geq .1) = 0.01\]

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

\[Y = \text{my winnings} = -X\]

\[
E(Y) = -0.1 \quad \Pr(Y \geq -0.1) = 0.99
\]

**Conclusion:** R.V. might almost never be $\geq \text{exp}$

or might almost always be $\geq \text{exp}$. 

$Q = \# \text{ comparisons made by rand QS}$

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

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

$E(Q) = \sum_{i = 1}^{cn^2} Pr(Q = i) + \sum_{i > cn^2} Pr(Q = i)$

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

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

$E\equiv n = 2^{10} \Rightarrow 10 \approx \frac{10}{2^{10}}$
Markov Inequality

\[ \Pr(X > x) \leq \frac{E(X)}{x} \]

\[ \downarrow \text{nonnegative} \]

Chebyshev Inequality

\[ \Pr(|Y-m| > x) \leq \frac{\text{Var}(Y)}{x^2} \]

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

\[ \Pr(X \geq \frac{3}{4}n) \leq \frac{n}{\frac{3}{4}n} = \frac{4}{3} \]

Markov

\[ \Pr(X \geq \frac{3}{4}n) \leq \Pr(|X - \frac{n}{2}| \geq \frac{n}{4}) \leq \frac{\text{Var}(X)}{\left(\frac{n}{4}\right)^2} = \frac{n}{\left(\frac{n}{4}\right)^2} = \frac{4}{n} \]

\[ \xrightarrow{n \to \infty} 0 \]

Chebyshev
 Sampling & Polling

What fraction of people approve of president?

Poll: call up n random people

\[ X = X_1 + X_2 + \ldots + X_n \]

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

Questions:

What should n be?

how confident are we?

How good an estimate?


Can now say my polling estimate 100% guaranteed to be within ± 2% of truth.

Question: Given Θ, 1-𝓁: how large does n need to be

So

\[ Pr(|X-p| ≤ Θ) ≥ 1-𝓁 \]

0.02 0.95

\[ \text{margin of error} \quad \text{confidence} \]

\[ -Θ ≤ X - p ≤ Θ \]

\[ p - Θ ≤ \bar{X} ≤ p + Θ \]

Apply Chernoff: \( X_i \sim \text{Ber}(p) \)
\[ \Pr(X > (1+\delta)pn) \leq e^{-\delta^2 pn} = e^{-\frac{\delta^2 np}{3}} \quad \text{for} \quad pn - \delta n \leq X \leq pn + \delta n \]

\[ \Pr(X < (1-\delta)pn) \leq e^{-\delta^2 np} \]

\[ \Rightarrow \Pr(|X - np| \geq \delta pn) \leq 2e^{-\frac{\delta^2 np}{3}} \]

\[ \text{So we want} \quad 2e^{-\frac{\delta^2 np}{3}} \leq \varepsilon \]

\[ \frac{\varepsilon}{3} \leq e^{\varepsilon/3} \]

\[ \ln \left( \frac{\varepsilon}{3} \right) \leq \frac{\varepsilon^2 n}{3} \]

\[ \frac{3}{\varepsilon^2} \ln \left( \frac{1}{\varepsilon} \right) \leq n \]

\[ \varepsilon \approx 0.02 \]

\[ \varepsilon = 0.05 \]
Notes:

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

Ex. $\Theta = 0.02 \quad 1 - 3 = 0.95$

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

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 sec

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 ?

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

$X_t = \sum_{i=1}^{n} X_{it}$  $\#$ 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

$\Pr(\exists t, 1 \leq t \leq T \text{ s.t. } X_t \geq B) \leq 3$
First, $\Pr(X_t \geq B)$

By Chernoff bound:

$$\Pr(X > (1+\delta)\mu) \leq e^{-\frac{\delta^2 \mu}{B}}$$

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

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

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

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

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

where $\mu = 1000$

$T: \# \text{ secs in year}$

$\varepsilon = 10^{-6}$
Recipe: solve for $\delta$

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

$Te^{-\delta^2 M} \leq \varepsilon$

$e^{\frac{\delta^2 M}{2}} \geq \frac{T}{\varepsilon}$

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

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

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

$B = (1 + \delta)M$

Example: $M = 1000$

$T = 60 \cdot 60$  (min sec)

$\varepsilon = 10^{-6}$

$\delta = \sqrt{\frac{2}{1000} \ln\left(\frac{3600}{10^{-6}}\right)} = 0.2097$
Buffer size = 1.2097 \times 1000 \approx 1210