Alex Psomas: Lecture 20.

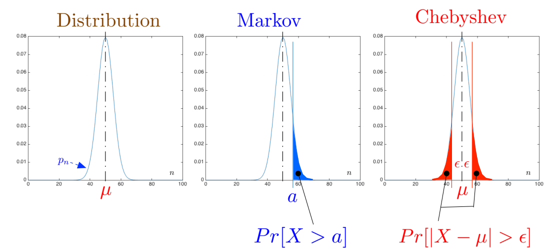
Chernoff and Erdős

1. Confidence intervals
2. Chernoff
3. Probabilistic Method

Reminders

- Quiz due tomorrow.
- Quiz coming out today.
- Midterm re-grade requests closing tomorrow.

Inequalities: An Overview



Confidence intervals example

You flip n coins. Each with probability p for H. p is unknown.

If you flip n coins, your estimate for p is $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

You many coins do you have to flip to make sure that your estimation $\hat{p}$ is within 0.01 of the true p, with probability at least 95%?

$E[\hat{p}] = E[\frac{1}{n}\sum_{i=1}^{n} X_i] = p$

$Var[\hat{p}] = Var[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n^2} Var[\sum_{i=1}^{n} X_i] = \frac{p(1-p)}{n}$

$$Pr[|\hat{p} - p| \geq \varepsilon] \leq \frac{Var[\hat{p}]}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2}$$

Confidence intervals example continued

Estimation $\hat{p}$ is within 0.01 of the true p, with probability at least 95%.

$$Pr[|\hat{p} - p| \geq \varepsilon] \leq \frac{p(1-p)}{n\varepsilon^2}$$

We want to make $Pr[|\hat{p} - p| \leq 0.01]$ at least 0.95.

Same as $Pr[|\hat{p} - p| \geq 0.01]$ at most 0.05.

It's sufficient to have $\frac{p(1-p)}{n\varepsilon^2} \leq 0.05$ or $n \geq \frac{20p(1-p)}{\varepsilon^2}$.

$p(1-p)$ is maximized for $p = 0.5$. Therefore it's sufficient to have $n \geq \frac{5}{\varepsilon^2}$.

For $\varepsilon = 0.01$ we get that $n \geq 50000$ coins are sufficient.

Chernoff

Markov: Only works for non-negative random variables.

$$Pr[X \geq t] \leq \frac{E[X]}{t}$$

Chebyshev:

$$Pr[|X - E[X]| \geq t] \leq \frac{Var[X]}{t^2}$$

Chernoff:
The good: Exponential bound
The bad: Sum of mutually independent random variables.
The ugly: People get scared the first time they see the bound.

Chernoff bounds

There are many different versions.

Today:
Theorem Let $X = \sum_{i=1}^{n} X_i$, where $X_i = 1$ with probability p_i and 0 otherwise , and all X_i are mutually independent. Let $\mu = E[X] = \sum_i p_i$. Then, for $0 < \delta < 1$:

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

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

$\#omg$ $\#ididntsignupforthis$

Proof idea

Markov: $Pr[X \geq a] \leq \frac{E[X]}{a}$

Apply Markov to e^{tX}!

$e^{\sum something} = \prod e^{something}$

Product of numbers smaller than 1 becomes small really fast!

$$Pr[X \geq a] = Pr[e^{tX} \geq e^{ta}] \leq \frac{E[e^{tX}]}{e^{ta}}$$

What is $E[e^{tX}]$?

Proof

What is $E[e^{tX}]$? $X = \sum_i X_i$, $\sum_i p_i = \mu$

X_i takes value 1 with prob. p_i, and 0 otherwise.

$$E[e^{tX_i}] = p_i e^{t \cdot 1} + (1-p_i)e^{t \cdot 0} = 1 + p_i(e^t - 1) \leq e^{p_i(e^t-1)}$$

Used that for all y, $1 + y \leq e^y$.

$$E[e^{tX}] = E\left[e^{t\sum_i X_i}\right] = E\left[\prod_{i=1}^{n} e^{tX_i}\right] = \prod_{i=1}^{n} E\left[e^{tX_i}\right]$$

$$\leq \prod_{i=1}^{n} e^{p_i(e^t-1)} = e^{\sum_i p_i(e^t-1)} = e^{(e^t-1)\sum_i p_i} = e^{(e^t-1)\mu}$$

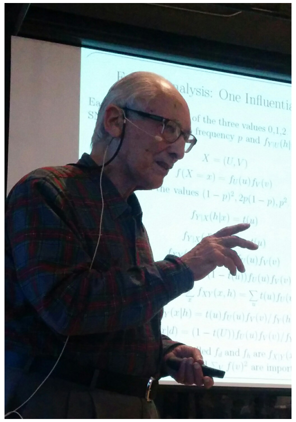
Proof

$$\begin{aligned} Pr[X \geq (1+\delta)\mu] &= Pr[e^{tX} \geq e^{t(1+\delta)\mu}] \\ &\leq \frac{E[e^{tX}]}{e^{t(1+\delta)\mu}} \\ &\leq \frac{e^{(e^t-1)\mu}}{e^{t(1+\delta)\mu}} = \left(\frac{e^{(e^t-1)}}{e^{t(1+\delta)}}\right)^{\mu} \end{aligned}$$

Since $\delta > 0$, we can set $t = \ln(1+\delta)$. Plugging in we get:

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

Herman Chernoff



With great proof comes great power

Flip a coin n times. Probability of H is p. X counts the number of heads.
X follows the Binomial distribution with parameters n and p.
$X \sim B(n,p)$.
$E[X] = np$. $Var[X] = np(1-p)$.
Say $n = 1000$ and $p = 0.5$. $E[X] = 500$. $Var[X] = 250$.

Markov says that $Pr[X \geq 600] \leq \frac{500}{600} = \frac{5}{6} \approx 0.83$

Chebyshev says that $Pr[X \geq 600] \leq 0.025$

Actual probability: < 0.000001

Chernoff:

$$Pr[X \geq (1+\delta)500] \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{500}$$

With great proof comes great power

Chernoff:

$$Pr[X \geq (1+\delta)500] \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{500}$$

$(1+\delta)500 = 600 \implies \delta = \frac{1}{5} = 0.2$:

$$Pr[X \geq 600] \leq \left(\frac{e^{0.2}}{(1+0.2)^{(1+0.2)}}\right)^{500} = 0.000083...$$

Chernoff Bounds come in many flavors:

- $Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$
- $Pr[X \geq (1+\delta)\mu] \leq e^{-\frac{\mu\delta^2}{3}}$
- $Pr[X \leq (1-\delta)\mu] \leq e^{-\frac{\mu\delta^2}{2}}$
- For $R > 6\mu$: $Pr[X \geq R] \leq 2^{-R}$

Better confidence intervals

You flip n coins. Each with probability p for H. p is unknown.

If you flip n coins, your estimate for p is $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

You many coins do you have to flip to make sure that your estimation $\hat{p}$ is within 0.01 of the true p, with probability at least 95%?

$E[n\hat{p}] = E[\sum_{i=1}^{n} X_i] = np$

$$Pr[p \notin [\hat{p}-\varepsilon, \hat{p}+\varepsilon]]$$

$$Pr[np \notin [n(\hat{p}-\varepsilon), n(\hat{p}+\varepsilon)]]$$

$$Pr[np \leq n(\hat{p}-\varepsilon)] + Pr[np \geq n(\hat{p}+\varepsilon)]$$

$$Pr\left[n\hat{p} \geq np(1+\frac{\varepsilon}{p})\right] + Pr\left[n\hat{p} \leq np(1-\frac{\varepsilon}{p})\right]$$

Confidence intervals example continued

Estimation $\hat{p}$ is within 0.01 of the true p, with probability at least 95%.

$$Pr\left[n\hat{p} \geq np(1+\frac{\varepsilon}{p})\right] + Pr\left[n\hat{p} \leq np(1-\frac{\varepsilon}{p})\right]$$

The first term is at most

$$e^{-\frac{\mu\delta^2}{3}} = e^{-\frac{np(\frac{\varepsilon}{p})^2}{3}} = e^{-\frac{n\varepsilon^2}{3p}}$$

The second term is at most

$$e^{-\frac{\mu\delta^2}{2}} = e^{-\frac{np(\frac{\varepsilon}{p})^2}{2}} = e^{-\frac{n\varepsilon^2}{2p}}$$

Confidence intervals example continued

$$Pr[p \notin [\hat{p}-\varepsilon, \hat{p}+\varepsilon]] \leq e^{-\frac{n\varepsilon^2}{3p}} + e^{-\frac{n\varepsilon^2}{2p}}$$

p is unknown... Bound gets worse as p increases, and $p \leq 1$. So just plug in $p = 1$:

$$Pr[p \notin [\hat{p}-\varepsilon, \hat{p}+\varepsilon]] \leq e^{-\frac{n\varepsilon^2}{3}} + e^{-\frac{n\varepsilon^2}{2}}$$

Confidence intervals example continued

$$Pr[p \notin [\hat{p}-\varepsilon, \hat{p}+\varepsilon]] \leq e^{-\frac{n\varepsilon^2}{3}} + e^{-\frac{n\varepsilon^2}{2}}$$

For our application: $\varepsilon = 0.01$. The bound should be smaller than .05

$$e^{-\frac{n0.01^2}{3}} + e^{-\frac{n0.01^2}{2}} \leq 0.05$$

Wolframalpha says: $n \geq 95436$. Worse than Chebyshev... Welcome to my life

Well, that was a waste of time...

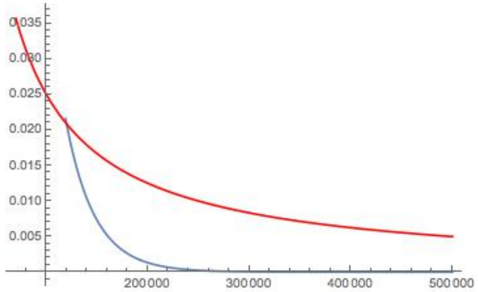
If you want the probability of failure to be smaller than 1%:

Chebyshev: $250,000$ coins.

Chernoff: $\approx 141,000$ coins.

Yay!

If you want to be within 0.01 of the truth:



x axis is number of coins. y-axis is probability of failure.

Red function is Chebyshev.

For a million coins: Chebyshev: 0.0025

Chernoff: $3.33824 * 10^{-15}$

Today's gig: The Probabilistic Method.

Gigs so far:

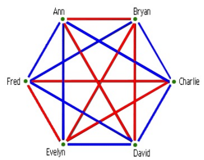
1. How to tell random from human.
2. Monty Hall.
3. Birthday Paradox.
4. St. Petersburg paradox.
5. Simpson's paradox.
6. Two envelopes problem.
7. Kruskal's Count.

Today: The Probabilistic Method

Proof techniques so far

- Direct
- Contrapositive
- Contradiction
- Induction

6 volunteers



Blue edge if they know each other.

Red edge if they don't know each other.

There is always a group of 3 that either all know each other, or all are strangers.

There always exists a monochromatic triangle.

How can we show that things exist?

Say I have a group of 1000 people.

Is there a "monochromatic" group of 3? What about 10? What about 20?

How big can these monochromatic cliques be???

And how would you prove it?

Try all colorings?? Good luck with that...

Number of colorings: $2^{\binom{1000}{2}} \approx 3.039 * 10^{150364}$.

Commonly accepted for the number of particles in the observable universe $\approx 10^{80}$.

How can we show that things exist?

Say I want to prove that there is a coloring for the clique with 1000 vertices such that there is no monochromatic clique of size, say, 20.

Trying all coloring is pointless.

Induction? Nah... It shouldn't be true if I replace 1000 with something much bigger.

Contradiction? Ok, say there exists a monochromatic clique. Now what?

.....

The probabilistic method

Step 1: Randomly color the graph. Each edge is colored red w.p. 0.5 and blue w.p. 0.5

Step 2: Compute an upper bound on the probability that there exists a monochromatic clique of size k.

Hey! I did this in a homework already!!!

Step 3: See if that probability is **strictly** smaller than 1.

If the probability that there exists a monochromatic clique is strictly less than 1, that means that the probability there isn't one is strictly bigger than 0.

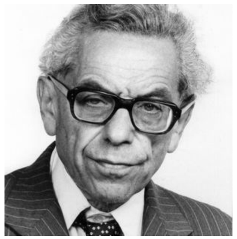
Well, that means that there is a coloring with no monochromatic clique of size k!

The probabilistic method

If I do something at random, and the probability I fail is strictly less than 1, that means that there is a way to succeed!!

The probabilistic method

Paul Erdős



Many quotes:
My brain is open!
Another roof, another proof.
It is not enough to be in the right place at the right time. You should also have an open mind at the right time.

Summary

Chernoff and Erdős

- Chernoff.
- The Probabilistic Method.