Lecture 6: Probability Review

Question:

USA — 300 mil

India — 1.2 billion

R & D parties

Goal: Estimate the percentage of R voters up to 1% error.

Method: randomly sample $\underline{\underline{S}}$ potential voters
report the fraction of R voters

How much larger, compared to the US, should the "S" be for India?

More abstract example:

n independent tosses of a fair coin.

What's the chance of seeing 51% heads?

→ our running example.

Probability Review

Sample Space: Set of all possible outcomes. Ω

In 2nd example: $\{H,T\}^n$

↪ = all possible sequences with H or T of length n.

Size $= 2^n$.

Each outcome $x \in \{H,T\}^n$ has a probability mass $p(x)$. Then, $\sum_{x\in\Omega} p(x) = 1$.

Events: subsets of Ω.

$$\forall S \subseteq \Omega, \quad \Pr[S] = p(S) = \sum_{x \in S} p(x)$$

Conditional Probability

$$\Pr[S \mid T] = \frac{\Pr[S \cap T]}{\Pr[T]}$$

(real valued)

Random Variables

neither random, nor variable

Def: A random variable is real-valued function $X : \Omega \to \mathbb{R}$.

Ex:

$$X_i = \begin{cases} 1 & \text{if } i^{th} \text{ toss is head} \\ 0 & \text{o/w} \end{cases}$$

Independent Random Variables

X, Y are indep if

$$\Pr[X = a \wedge Y = b] = \Pr[X = a] \cdot \Pr[Y = b]$$

also called indicator random var.

Ex.

$X =$ number of heads in n indep coin tosses

Note: $X = \sum_i X_i$

Expectation (or average!)

Def $$\mathbb{E}[X] := \sum_{\omega \in \Omega} \Pr[\omega \in \Omega] \cdot X(\omega)$$

e.g. $$\mathbb{E}[X_i] = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 0 = \frac{1}{2}$$

Linearity of Expectation

For every $X_1, \dots, X_n$ (not necessarily indep),

$$\mathbb{E}\left[\sum_i X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i]$$

e.g. $\mathbb{E}[\#\text{ heads in } n \text{ coin tosses}]$

$$= \sum_i \mathbb{E}[X_i]$$

$$= \frac{n}{2}.$$

Union Bound

$$\Pr[A_1 \vee A_2]$$

$$\leq \Pr[A_1] + \Pr[A_2]$$

$$\Pr[X \geq t \cdot X] \leq \frac{1}{t}$$

E.g. $\Pr[\#\text{ heads} \geq 51\%]$

$$= \Pr\left[\sum_i X_i \geq \frac{n}{2}(1+\delta)\right] \leq \frac{1}{1+\delta}$$

for $\delta = 2/100$

$$\sim \frac{1}{1+2/100} \sim 98\%.$$

not much better than saying can't get 51% or more heads ALL THE TIME.

Markov's Inequality is tight.

E.g. $X = 0$ w.p. $(1-1/k)$, $X = 1$ w.p. $1/k$.

$\mathbb{E}X = 1/k$. $\Pr[X \geq k \cdot \mathbb{E}X] = 1/k$.

(When) Can we do better?

functions not overly sensitive to any single coordinate

<u>Principle</u>: sum of indep random variables concentrate around the mean.

<u>Variance</u>

measure of deviation around the mean.

Def: $\mathrm{Var}(X) = \mathbb{E}(X - \mathbb{E}X)^2$

(average of squared deviation around the mean)

e.g. $X_i = \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{o/w} \end{cases}$

$\mathbb{E}X_i = p \qquad \mathrm{Var}(X) = \mathbb{E}(X_i - p)^2$

$$= (1-p)^2 \cdot p + p^2 \cdot (1-p)$$

$$= p + p^3 - 2p^2 + p^2 - p^3$$

$$= p - p^2 = p(1-p)$$

Lemma (Variance of a sum of indep R.V.s)

$$X = \sum_i X_i \qquad X_i \text{ indep.}$$

Then, $\mathrm{Var}(X) = \sum_i \mathrm{Var}(X_i)$.

<u>Pf</u>: $$X - \mathbb{E}X = \sum_i (X_i - \mathbb{E}X_i)$$

$$\mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}\left(\sum_i (X_i - \mathbb{E}X_i)\right)^2$$

$$= \sum_i \mathbb{E}(X_i - \mathbb{E}X_i)^2 + \sum_{i \neq j} \underbrace{\mathbb{E}(X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)}_{=0}$$

$$= \sum_i \mathrm{Var}(X_i)$$

E.g. $X = \sum_i X_i$, $X_i = \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{w.p. } 1-p \end{cases}$

$$\mathrm{Var}(X) = \sum_i \mathrm{Var}(X_i)$$
$$= n \cdot p \cdot (1-p).$$

For $p = 1/2$, $\mathrm{Var}(X) = n/4$.

Lemma (Chebyshev)

$$\Pr\left[|X - \mathbb{E}X| \geq t \cdot \sqrt{\mathrm{Var}(X)}\right] \leq 1/t^2$$

Pf: $\Pr\left[\,|X - \mathbb{E}[X]| \geq t\sqrt{\mathrm{Var}(X)}\right]$

$$= \Pr\left[(X - \mathbb{E}X)^2 \geq t^2 \cdot \mathrm{Var}(X)\right]$$

$$\leq 1/t^2 \quad \text{by Markov}$$

Simple example of moment method.

Can use any monotone function instead of the square.

Cor: $\Pr\left[\#\text{heads} \geq \frac{n}{2}(1+\delta)\right]$

$$= \Pr\left[\,\text{———} \geq \frac{n}{2} + \underbrace{\delta\sqrt{n}}\cdot\frac{\sqrt{n}}{2}\right] \leq \frac{1}{\delta^2 \cdot n}$$

significantly better ~~bound~~ than prev: $\frac{1}{1+\delta} \sim 1-\delta$

if n is large

For sum of indep r.v.s we'll use the moment method to derive a significantly tighter bound.

Theorem (Chernoff, Hoeffding)

$$X = \sum_i X_i \quad , \quad X_i \text{ indep ; } \begin{cases} 1 & \text{w.p. } p_i \\ 0 & \text{o/w} \end{cases}$$

$$\mu := \mathbb{E} X = \sum_i X_i = \sum_i p_i .$$

Then,
$$\Pr[X \geq (1+\delta)\mu] \leq e^{-\frac{\delta^2}{2+\delta} \cdot \mu} \quad \forall \delta > 0$$

$$\Pr[X \leq (1-\delta)\mu] \leq e^{-\frac{\mu \cdot \delta^2}{2}} \quad \forall \delta < 1$$

Cor: $$\Pr[|X - \mu| \geq \delta\mu] \leq 2 \cdot e^{-\mu \cdot \frac{\delta^2}{3}} \quad \forall \delta \leq 1.$$

Let's apply this in 2 examples to see how it's going.

In the fair coin toss example,

$$\mathbb{E} X_i = p = 1/2 \ \forall i.$$

So: $\mathbb{E} X = \frac{n}{2}$.

$$\Pr[X \geq (1+\overset{2/100}{\delta}) \cdot \mu] \leq e^{-\delta^2 \cdot \mu / 3}$$

$$e^{-\frac{4}{10^4 \cdot 6} \cdot n}$$

As n grows this falls exponentially fast.

Recall: Markov tail: $\frac{1}{1+\delta} \sim 1-\delta$

Chebyshev tail: $\frac{1}{\delta^2 \cdot n}$

→ "inverse poly"

Chernoff tail: "inverse exp"

Thm: (Hoeffding)

$X_i \in [0,1]$, indep; $\mathbb{E}X_i = p_i$.

$$\Pr[\,|X-\mu| \geq t\,] \leq 2 \cdot e^{-\frac{2t^2}{n}}.$$

Polling Question.

Suppose population has p frac R & $(1-p)$ frac D.

$$X_i := \begin{cases} 1 & \text{if } i^{th} \text{ sample is R} \\ 0 & \text{o/w.} \end{cases}$$

What's $\mathbb{E} X_i$? p. ✓

$\operatorname{Var} X_i$? $p \cdot (1-p)$.

Suppose we take n samples.

\# of R : $X = \sum_i X_i$ $\quad \mathbb{E} X = n \cdot p = \mu$

$$\Pr[\,|X-\mu| \geq \delta \cdot n] \leq 2 \cdot e^{-\frac{2 \cdot \delta^2 \cdot n^2}{n}}$$

$$= 2 \cdot e^{-2\delta^2 \cdot n}$$

$\delta = 2\%$ say.

Then: $2 \cdot e^{-2 \cdot \frac{4}{10^4} \cdot n} \leq \frac{1}{20} = 5\%$

if $n \geq 4600$ (5000) say.

A sample size of 5000 is enough to get a 2% error estimate with 95% confidence.

doesn't depend at all on the size of the population.

Proof of Chernoff Bound

$$\Pr[X \geq t]$$

$$= \Pr[e^{sX} \geq e^{s \cdot t}]$$

$$\leq \frac{\mathbb{E}[e^{sX}]}{e^{t \cdot s}} \quad \text{Markov}$$

exp moment
or
MGF

$$= \mathbb{E}\, e^{\sum_{i=1}^{n} sX_i}$$

$$= \prod_i \mathbb{E}\, e^{sX_i}$$

$$= \prod_i (e^s \cdot p_i + 1 - p_i)$$

$$\leq \prod_i e^{p_i(e^s-1)}$$

$1+y \leq e^y$

$$= e^{\mu \cdot (e^s-1)}$$

$$t = (1+\delta) \cdot \mu$$

$$s = \ln(1+\delta).$$