Mathematics for Computer Science

MIT 6.042J/18.062J

Deviation from the Mean

Example: IQ

IQ measure was constructed so that

average IQ = 100.

What fraction of the people can possibly have an IQ $\geq$ 300?

IQ Higher than 300?

Fraction f with IQ $\geq$ 300 adds $\geq$ 300f to average, so 100 = avg IQ $\geq$ 300f:

$$f \leq 100/300 = 1/3$$

IQ Higher than 300?

At most 1/3 of people have IQ $\geq$ 300

$$\Pr\{IQ \geq 300\} \leq \frac{E[IQ]}{300}$$

IQ Higher than x?

Besides mean = 100, we used only one fact about the distribution of IQ:

IQ is always nonnegative

Markov Bound

If R is nonnegative, then

$$\Pr\{R \geq x\} \leq \frac{E[R]}{x}$$

for $x \geq E[R]$

Markov Bound

- Weak
- Obvious
- Useful anyway

IQ ≥ 300, again

Suppose we are given that IQ is always ≥ 50?
Get a better bound using
(IQ – 50)
since this is now ≥ 0.

IQ ≥ 300, again

f contributes (300-50)f to the average of (IQ-50), so

$$50 = E[IQ-50] \geq 250f$$

$$f \leq 50/250 = 1/5$$

Better bound from Markov by shifting R to have 0 as minimum

Improving the Markov Bound

$$\Pr\{|R-\mu| \geq x\}$$

$$= \Pr\{(R-\mu)^2 \geq x^2\}$$

by Markov:

$$\leq \frac{E[(R-\mu)^2]}{x^2}$$

variance of R

Chebyshev Bound

$$\Pr\{|R-\mu| \geq x\} \leq \frac{\mathrm{Var}[R]}{x^2}$$

$$\mathrm{Var}[R] ::= E[(R-\mu)^2]$$

$$\sigma_R ::= \sqrt{\mathrm{Var}[R]}$$

Standard Deviation

$$\Pr\{|R-\mu| \geq x\} \leq \frac{\sigma^2}{x^2}$$

R probably not many σ's from μ:

further than	σ	Pr ≤ 1
	2σ	Pr ≤ 1/4
	3σ	Pr ≤ 1/9
	4σ	Pr ≤ 1/16

Variance of an Indicator

I an indicator with E[I]=p:

$$\mathrm{Var}[I] ::= E\left[(I-p)^2\right]$$
$$= E\left[I^2\right] - 2pE\left[I\right] + p^2$$
$$= E\left[I\right] - 2p\cdot p + p^2$$
$$= p - 2p^2 + p^2 = pq$$

Calculating Variance

$$\mathrm{Var}[aR+b] = a^2\,\mathrm{Var}[R]$$
$$\mathrm{Var}[R] = E[R^2] - (E[R])^2$$

simple proofs applying linearity of E[] to the def of Var[]

Calculating Variance

Pairwise Independent Additivity

$$\mathrm{Var}[R_1 + R_2 + \cdots + R_n]$$
$$= \mathrm{Var}[R_1] + \mathrm{Var}[R_2] + \cdots + \mathrm{Var}[R_n]$$

providing $R_1, R_2, \ldots, R_n$ are pairwise independent

again, a simple proof applying linearity of E[] to the def of Var[]

Mathematics for Computer Science

MIT 6.042J/18.062J

Deviation of Repeated Trials

Jacob D. Bernoulli (1659–1705)

Even the stupidest man —by some instinct of nature *per se* and by no previous instruction (this is truly amazing) —knows for sure that the more observations ...that are taken, the less the danger will be of straying from the mark.

---*Ars Conjectandi* (The Art of Guessing), 1713*

*taken from Grinstead \& Snell,
http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/book.html
Introduction to Probability, American Mathematical Society, p. 310.

Jacob D. Bernoulli (1659–1705)

It certainly remains to be inquired whether after the number of observations has been increased, the probability…of obtaining the true ratio…finally exceeds any given degree of certainty; or whether the problem has, so to speak, its own asymptote —that is, whether some degree of certainty is given which one can never exceed.

Repeated Trials

Random var R with mean μ

n independent observations

$$R_1, \cdots, R_n$$

Repeated Trials

take average:

$$A_n ::= \frac{R_1 + R_2 + \cdots + R_n}{n}$$

Bernoulli question: is it probably close to μ if n is big

$$\Pr\{|A_n - \mu| \le \delta\} = ?$$

Weak Law of Large Numbers

$$\lim_{n \to \infty} \Pr\{|A_n - \mu| \le \delta\} = 1$$

$$\lim_{n \to \infty} \Pr\{|A_n - \mu| > \delta\} = 0$$

Jacob D. Bernoulli (1659 – 1705)

Therefore, this is the problem which I now set forth and make known after I have pondered over it for twenty years. Both its novelty and its very great usefulness, coupled with its just as great difficulty, can exceed in weight and value all the remaining chapters of this thesis.

Weak Law of Large Numbers

will follow easily by Chebyshev & variance properties

$$\lim_{n \to \infty} \Pr\{|A_n - \mu| > \delta\} = 0$$

Repeated Trials

$$E[A_n] ::= E\left[\frac{R_1 + R_2 + \cdots + R_n}{n}\right]$$

$$= \frac{E[R_1] + E[R_2] + \cdots + E[R_n]}{n}$$

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

Weak Law of Large Numbers

So by Chebyshev

$$\Pr\{|A_n - \mu| > \delta\} \le \frac{\mathrm{Var}[A_n]}{\delta^2}$$

need only show

$\mathrm{Var}[A_n] \to 0$ as $n \to \infty$

Albert R Meyer, May 7, 2010 lec 13F.61

Repeated Trials

$$\mathrm{Var}[A_n] = \mathrm{Var}\left[\frac{(R_1 + R_2 + \cdots + R_n)}{n}\right]$$

$$= \frac{(\mathrm{Var}[R_1] + \mathrm{Var}[R_2] + \cdots + \mathrm{Var}[R_n])}{n^2}$$

$$\text{QED} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \longrightarrow 0$$

Albert R Meyer, May 7, 2010 lec 13F.63

Analysis of the Proof

proof only used that $R_1,\dots,R_n$ have

- same mean
- same variance
- & variances add

  — which follows from pairwise independence

Albert R Meyer, May 7, 2010 lec 13F.64

Pairwise Independent Sampling

Theorem:
Let $R_1,\dots,R_n$ be pairwise independent random vars with the same finite mean μ and variance σ^2. Let $A_n ::= (R_1 + R_2 + \cdots + R_n)/n.$ Then

$$\Pr\{|A_n - \mu| > \delta\} \le \frac{1}{n}\left(\frac{\sigma}{\delta}\right)^2$$

Albert R Meyer, May 7, 2010 lec 13F.65

Pairwise Independent Sampling

The punchline:
we now know how big a sample is needed to estimate the mean of
any* random variable within
any* desired tolerance with
any* desired probability

*variance $< \infty$, tolerance > 0, probability < 1

Albert R Meyer, May 7, 2010 lec 13F.66

Team Problems

Problems

1–3

Albert R Meyer, May 7, 2010 lec 13F.67