



Deviation from the Mean.II



Chebyshev's Theorem

Random Variable R ,
Given $E[R] = \mu$, $Var[R]$

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

Probability that I see a value further than x away from the expected value

is limited by the variance



Another Form of Chebyshev

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

Let $x = c \sigma$, where $c = \text{constant}$ and $Var[R] = \sigma^2$

$$\Pr\{|R - \mu| \geq c\sigma\} \leq \frac{Var[R]}{c^2 \sigma^2} = \frac{\cancel{\sigma^2}}{c^2 \cancel{\sigma^2}} = \frac{1}{c^2}$$



Interpreting Variance

$$\Pr\{|R - \mu| \geq c\sigma\} \leq \frac{1}{c^2}$$

Probability that

$$|R - \mu| \geq \sigma \text{ is } \leq 1$$

$$|R - \mu| \geq 2\sigma \text{ is } \leq 1/4 \quad (25\%)$$

$$|R - \mu| \geq 3\sigma \text{ is } \leq 1/9 \quad (11\%)$$

$$|R - \mu| \geq 4\sigma \text{ is } \leq 1/16 \quad (6\%)$$



Space Station Mir

Suppose that main computer has a probability p of failing every year

- Mean Time to Failure: when do we expect it to fail?

$$E[T] = 1/p$$

$$Var[T] = ?$$



Calculating Variance

We know $\Pr\{T=k\} = (1-p)^{k-1}p$
How do we compute $Var[T]$?

$$Var[T] = E[T^2] - (E[T])^2$$

- Define $Y ::= T^2$

$$• T = 1, 2, 3, \dots, k, \dots \infty$$

$$• Y = 1, 4, 9, \dots, k^2, \dots \infty$$

6	9	13	7
12	10	5	
3	1	4	14
15	8	11	2

Calculating Variance

$$\begin{aligned}
 E[T^2] &= E[Y] \\
 &= \sum_{i \in \text{range}(Y)} i \Pr\{Y = i\} \\
 &= \sum_{k=1}^{\infty} k^2 \Pr\{Y = k^2\} \\
 &= \sum_{k=1}^{\infty} k^2 \Pr\{T = k\} = \sum_{k=1}^{\infty} k^2 (1-p)^{k-1} p
 \end{aligned}$$

↓ Range of Y
↓ Same event

6	9	13	7
12	10	5	
3	1	4	14
15	8	11	2

Mean Time to Failure



$$E[T] = 1/p \quad \text{Var}[T] = \sigma^2 = \frac{1}{p} \left(\frac{1}{p} - 1 \right)$$

- $p = 1/6$ $E[T] = 6,$ $\sigma = 6$ (Dice)
- $p = 1/10$ $E[T] = 10,$ $\sigma = 10$ (Mir 1)
- $p = 1/1000$ $E[T] = 1000,$ $\sigma = 1000$ (Mir 2)

Chebyshev tells us that probability that Mir 1 will last for **more than 30 years** is **less than 25%**

6	9	13	7
12	10	5	
3	1	4	14
15	8	11	2

Class Exercise

$$\begin{aligned}
 \text{Var}[R] &::= E[(R - E[R])^2] \\
 &= E[R^2] - (E[R])^2
 \end{aligned}$$

$\text{Var}[5] = ?$

$\text{Var}[5R] = ?$

$\text{Var}[I] = ?$ where I is an indicator variable
with $\Pr\{I = 1\} = p$
and $\Pr\{I = 0\} = (1-p)$

6	9	13	7
12	10	5	
3	1	4	14
15	8	11	2

Birthday Pairs



Given a room of N people, how many pairs of people have the same birthday, if a year has Y days?

$$E[P] = N^2/2Y$$

(by linearity of expectations)

6	9	13	7
12	10	5	
3	1	4	14
15	8	11	2

Birthday Pairs



$E[P] \approx 29$ for this class (149) and planet

What is the $\Pr\{P = 23\}$?

Calculating $\Pr\{P = k\}$ is hard

However, we can still calculate the variance!

6	9	13	7
12	10	5	
3	1	4	14
15	8	11	2

Variance of Sums

$$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$$

if X, Y are **independent**.

Generally,

$$\text{Var}[X_1 + X_2 + \dots] = \text{Var}[X_1] + \text{Var}[X_2] + \dots$$

if X_i are **pairwise independent**.



Pairwise Independence

$U ::=$ Albert and Radhika have same bday

$V ::=$ Albert and Tina have same bday

any U and V are independent

But

$W ::=$ Radhika and Tina have same bday

$(U \wedge V)$, and W are NOT independent



Birthday Pairs

$X_i = 1$ if pair i has the same birthday

X_i are pairwise independent

• $E[X_i] = p$ where $p ::= 1/Y$

• $\text{Var}[X_i] = p(1-p) \approx 1/Y$

Number of pairs $M \approx N^2/2$

$$E[P] = E[X_1 + X_2 + \dots + X_M] \approx$$

$$\text{Var}[P] = \text{Var}[X_1 + X_2 + \dots + X_M] \approx$$



Birthday Predictions



For 146 students, $E[P] \approx 29$,

$\text{Var}[P] \approx 29$ (std dev 5.3)

Chebyshev: More than 75% of the time we expect to see 29 ± 10 pairs, that is, between 9 & 39 pairs

- What we saw: 23 pairs
- What if N is larger (larger class)?



In Class Problems

Additivity of Variance?

- Binomial Distribution YES
- Random Hat Check NO



In-Class Problems

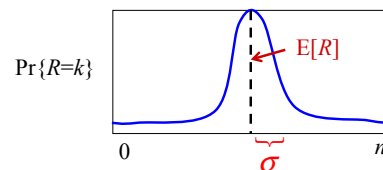
Problems 1,2,3



Binomial Distribution

$$E[R] = np$$

$$\text{Standard Deviation } \sigma = \sqrt{np(1-p)}$$





Binomial Distribution

as a fraction of n

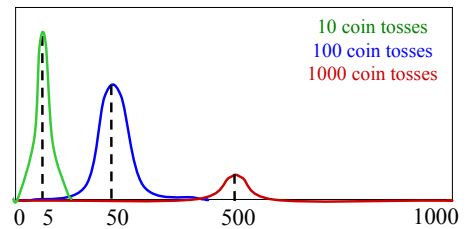
$$E[R] = np \quad p \quad 1/2$$

$$\text{StdDev}[R] = \sqrt{np(1-p)} \quad \sqrt{\frac{pq}{n}} \quad \frac{2}{\sqrt{n}}$$

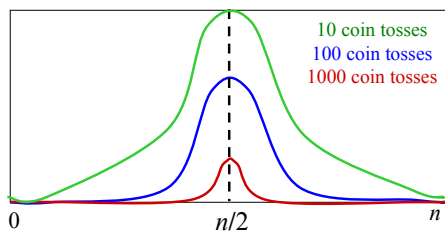
More times I toss, the smaller the variation from expected fraction of heads:
Probability of being close to the mean goes up!



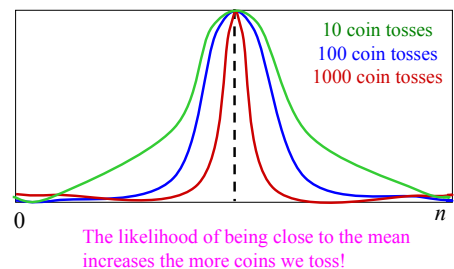
Fair Binomial Distribution



Normalize the x axis to be 0 to n



Normalize the y axis



Repeated Trials

The average value of repeated experiments is **likely to be close** to the expected value.

And it gets **more likely** as we do **more experiments**



Weak Law of Large Numbers

Let A_n be the average of n independent experiments, each with same μ and σ

$$\lim_{n \rightarrow \infty} \Pr\{|A_n - \mu| \leq \epsilon\} = 1$$



Pairwise Independent Sampling

Let A_n be the average of n independent experiments, each with same μ and σ

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

↑
CHEBYSHEV:
The probability
is limited



Applications



- Guess the bias of a coin
- Guess voting results from random polling
- Guess pollution levels from sampling Charles River water..