# Random Variables

Random Variables: $X : \Omega \to R$.

Distribution: $Pr[X = a] = \sum_{\omega : X(\omega) = a} Pr(\omega)$

$X$ and $Y$ independent $\iff$ all associated events are independent.
Expectation: $E[X] = \sum_a aPr[X = a] = \sum_{\omega \in \Omega} X(\omega)Pr(\omega)$.
Linearity: $E[X + Y] = E[X] + E[Y]$.

Variance: $Var(X) = E[(X - E[X])^2] = E[X^2] - (E(X))^2$
For independent $X, Y$, $Var(X + Y) = Var(X) + Var(Y)$.
Also: $Var(cX) = c^2 Var(X)$ and $Var(X + b) = Var(X)$.

Poisson: $X \sim P(\lambda)$ $\quad Pr[X = i] = e^{-\lambda} \frac{\lambda^i}{i!}$.
$E(X) = \lambda$, $Var(X) = \lambda$.
Binomial: $X \sim B(n, p)$ $\quad Pr[X = i] = \binom{n}{i} p^i (1 - p)^{n-i}$
$E(X) = np$, $Var(X) = np(1 - p)$
Uniform: $X \sim U\{1, \ldots, n\}$ $\quad \forall i \in [1, n], Pr[X = i] = \frac{1}{n}$.
$E[X] = \frac{n+1}{2}$, $Var(X) = \frac{n^2-1}{12}$.
Geometric: $X \sim G(p)$ $\quad Pr[X = i] = (1 - p)^{i-1} p$
$E(X) = \frac{1}{p}$, $Var(X) = \frac{1-p}{p^2}$

Note: Probability Mass Function $\equiv$ Distribution.

**Definition** The covariance of $X$ and $Y$ is

$$cov(X,Y) := E[(X-E[X])(Y-E[Y])].$$

**Definition** The correlation of $X, Y$, $Cor(X,Y)$ is

$$corr(X,Y) : \frac{cov(X,Y)}{\sigma(X)\sigma(Y)}.$$

Note: $|corr(X,Y)| \leq 1$.

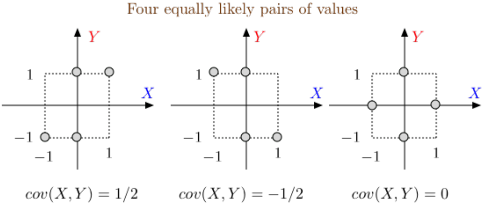
$corr(X,X)$? 1
$corr(X,-X)$? -1
$corr(X,X/2)$? 1
$corr(X,5X)$? 1
$corr(X,X+Y)$ with $var(X) = Var(Y)$, and $X, Y$ independent? $\frac{1}{\sqrt{2}}$

$cov(X,X+Y) = E[(X-E[X])(X-E[X]+Y-E[Y])] = Var(X) + cov(X,Y) = Var(X)$.

$corr(X,X+Y) = \frac{varX}{\sigma(X)\sigma(X+Y)} = \frac{varX}{\sigma(X)\sqrt{2}\sigma(X)} = \frac{1}{\sqrt{2}}$

$r^2 = corr(X,Y)^2$ is fraction of variance of $Y$ explained by $X$.

# Examples of Covariance



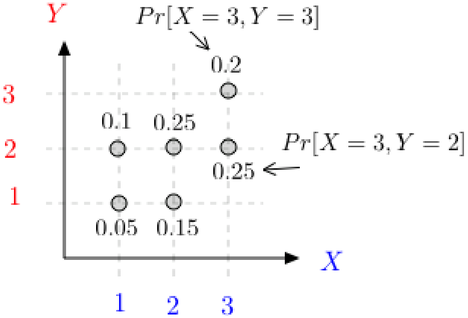
Note that $E[X] = 0$ and $E[Y] = 0$ in these examples. Then $cov(X, Y) = E[XY]$.

When $cov(X, Y) > 0$, the RVs $X$ and $Y$ tend to be large or small together. $X$ and $Y$ are said to be positively correlated.

When $cov(X, Y) < 0$, when $X$ is larger, $Y$ tends to be smaller. $X$ and $Y$ are said to be negatively correlated.

When $cov(X, Y) = 0$, we say that $X$ and $Y$ are uncorrelated.

# Examples of Covariance



$$E[X] = 1 \times 0.15 + 2 \times 0.4 + 3 \times 0.45 = 2.3$$

$$E[X^2] = 1^2 \times 0.15 + 2^2 \times 0.4 + 3^2 \times 0.45 = 5.8$$

$$E[Y] = 1 \times 0.2 + 2 \times 0.6 + 3 \times 0.2 = 2$$

$$E[Y^2] = 1 \times 0.2 + 4 \times 0.6 + 9 \times 0.2 = 4.4$$

$$E[XY] = 1 \times 0.05 + 1 \times 2 \times 0.1 + \cdots + 3 \times 3 \times 0.2 = 4.85$$

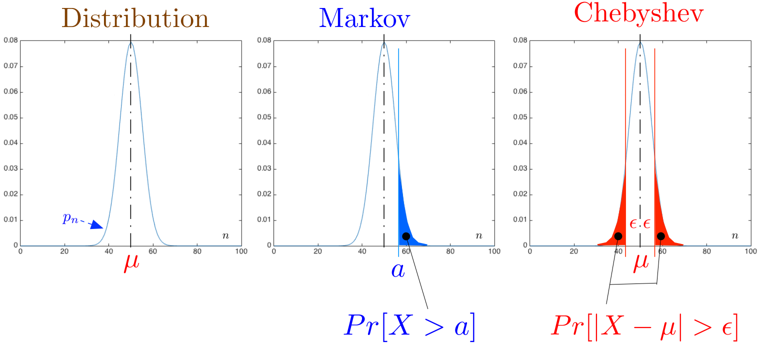
$$cov(X,Y) = E[XY] - E[X]E[Y] = .25$$

$$var[X] = E[X^2] - E[X]^2 = .51$$

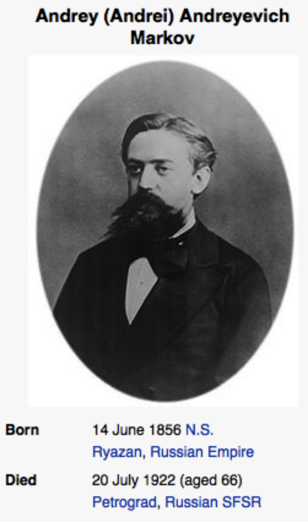
$$var[Y] = E[Y^2] - E[Y]^2 = .4$$

$$corr(X,Y) \approx 0.55$$

# Inequalities: An Overview



# Andrey Markov



**Andrey (Andrei) Andreyevich Markov**

| | |
|---|---|
| **Born** | 14 June 1856 N.S.<br>Ryazan, Russian Empire |
| **Died** | 20 July 1922 (aged 66)<br>Petrograd, Russian SFSR |

Andrey Markov is best known for his work on stochastic processes. A primary subject of his research later became known as Markov chains and Markov processes.

Pafnuty Chebyshev was one of his teachers.

Markov was an atheist. In 1912 he protested Leo Tolstoy's excommunication from the Russian Orthodox Church by requesting his own excommunication. The Church complied with his request.

# Markov's inequality

The inequality is named for Andrey Markov, though in work by Pafnuty Chebyshev. (Sometimes) called Chebyshev's first inequality.

**Theorem** Markov's Inequality
Assume $f : \Re \to [0, \infty)$ is nondecreasing. Then,

$$Pr[X \geq a] \leq \frac{E[f(X)]}{f(a)}, \text{ for all } a \text{ such that } f(a) > 0.$$

**Proof:**

Observe that

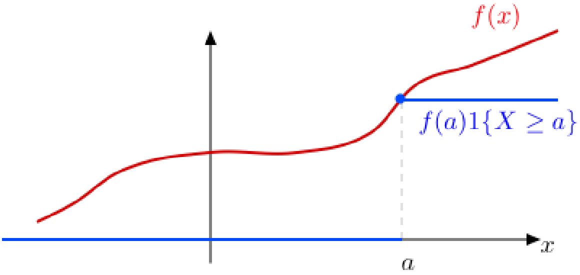
$$1\{X \geq a\} \leq \frac{f(X)}{f(a)}.$$

Indeed, if $X < a$, the inequality reads $0 \leq f(x)/f(a)$, which holds since $f(\cdot) \geq 0$. Also, if $X \geq a$, it reads $1 \leq f(x)/f(a)$, which holds since $f(\cdot)$ is nondecreasing.
Taking the expectation yields the inequality,
expectation of an indicator is the probability.
and expectation is monotone, e.g., weighted sum of points.

That is, $\sum_v Pr[X = v] 1\{v \geq a\} \leq \sum_v Pr[X = v] \frac{f(v)}{f(a)}$.

□

# A picture



$$f(a)1\{X \geq a\} \leq f(x) \Rightarrow 1\{X \geq a\} \leq \frac{f(X)}{f(a)}$$

$$\Rightarrow Pr[X \geq a] \leq \frac{E[f(X)]}{f(a)}$$

# Markov Inequality Example: G(p)

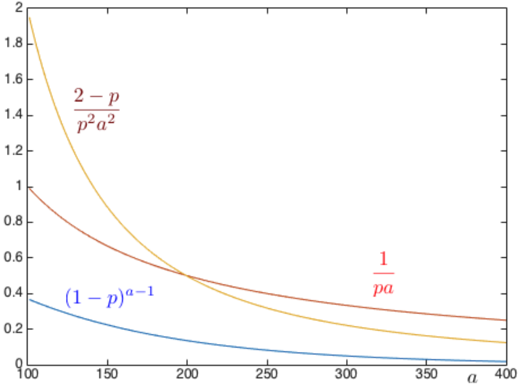
Let $X = G(p)$. Recall that $E[X] = \frac{1}{p}$ and $E[X^2] = \frac{2-p}{p^2}$.

Choosing $f(x) = x$, we get

$$Pr[X \geq a] \leq \frac{E[X]}{a} = \frac{1}{ap}.$$

Choosing $f(x) = x^2$, we get

$$Pr[X \geq a] \leq \frac{E[X^2]}{a^2} = \frac{2-p}{p^2 a^2}.$$



# Markov Inequality Example: $P(\lambda)$

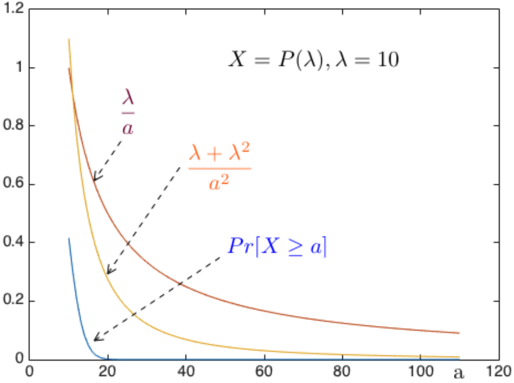
Let $X = P(\lambda)$. Recall that $E[X] = \lambda$ and $E[X^2] = \lambda + \lambda^2$.

Choosing $f(x) = x$, we get

$$Pr[X \geq a] \leq \frac{E[X]}{a} = \frac{\lambda}{a}.$$

Choosing $f(x) = x^2$, we get

$$Pr[X \geq a] \leq \frac{E[X^2]}{a^2} = \frac{\lambda + \lambda^2}{a^2}.$$



# Chebyshev's Inequality

This is Pafnuty's inequality:
**Theorem:**

$$Pr[|X - E[X]| > a] \leq \frac{var[X]}{a^2}, \text{ for all } a > 0.$$

**Proof:** Let $Y = |X - E[X]|$ and $f(y) = y^2$. Then,

$$Pr[Y \geq a] \leq \frac{E[f(Y)]}{f(a)} = \frac{var[X]}{a^2}.$$

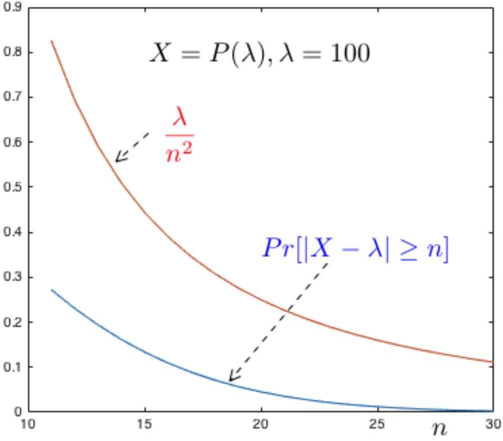
□

This result confirms that the variance measures the "deviations from the mean."

# Chebyshev and Poisson

Let $X = P(\lambda)$. Then, $E[X] = \lambda$ and $var[X] = \lambda$. Thus,

$$Pr[|X - \lambda| \geq n] \leq \frac{var[X]}{n^2} = \frac{\lambda}{n^2}.$$



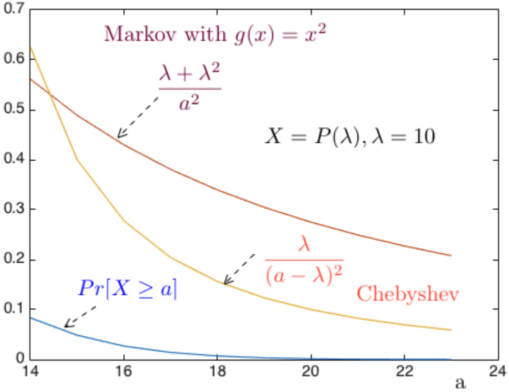
# Chebyshev and Poisson (continued)

Let $X = P(\lambda)$. Then, $E[X] = \lambda$ and $var[X] = \lambda$. By Markov's inequality,

$$Pr[X \geq a] \leq \frac{E[X^2]}{a^2} = \frac{\lambda + \lambda^2}{a^2}.$$

Also, if $a > \lambda$, then $X \geq a \Rightarrow X - \lambda \geq a - \lambda > 0 \Rightarrow |X - \lambda| \geq a - \lambda$.

Hence, for $a > \lambda$, $Pr[X \geq a] \leq Pr[|X - \lambda| \geq a - \lambda] \leq \frac{\lambda}{(a-\lambda)^2}$.



# Fraction of $H$'s

Here is a classical application of Chebyshev's inequality.

How likely is it that the fraction of $H$'s differs from 50%?

Let $X_m = 1$ if the $m$-th flip of a fair coin is $H$ and $X_m = 0$ otherwise.

Define

$$Y_n = \frac{X_1 + \cdots + X_n}{n}, \text{ for } n \geq 1.$$

We want to estimate

$$Pr[|Y_n - 0.5| \geq 0.1] = Pr[Y_n \leq 0.4 \text{ or } Y_n \geq 0.6].$$

By Chebyshev,

$$Pr[|Y_n - 0.5| \geq 0.1] \leq \frac{var[Y_n]}{(0.1)^2} = 100 var[Y_n].$$

Now,

$$var[Y_n] = \frac{1}{n^2}(var[X_1] + \cdots + var[X_n]) = \frac{1}{n} var[X_1] \leq \frac{1}{4n}.$$

$Var(X_i) = p(1 - lp) \leq (.5)(.5) = \frac{1}{4}$

# Fraction of *H*'s

$$Y_n = \frac{X_1 + \cdots + X_n}{n}, \text{ for } n \geq 1.$$

$$Pr[|Y_n - 0.5| \geq 0.1] \leq \frac{25}{n}.$$

For $n = 1,000$, we find that this probability is less than 2.5%.

As $n \to \infty$, this probability goes to zero.

In fact, for any $\varepsilon > 0$, as $n \to \infty$, the probability that the fraction of *H*s is within $\varepsilon > 0$ of 50% approaches 1:

$$Pr[|Y_n - 0.5| \leq \varepsilon] \to 1.$$

This is an example of the Law of Large Numbers.

We look at a general case next.

# Weak Law of Large Numbers

**Theorem** Weak Law of Large Numbers

Let $X_1, X_2, \ldots$ be pairwise independent with the same distribution and mean $\mu$. Then, for all $\varepsilon > 0$,

$$Pr[|\frac{X_1 + \cdots + X_n}{n} - \mu| \geq \varepsilon] \to 0, \text{ as } n \to \infty.$$

**Proof:**
Let $Y_n = \frac{X_1+\cdots+X_n}{n}$. Then

$$\begin{aligned} Pr[|Y_n - \mu| \geq \varepsilon] &\leq \frac{var[Y_n]}{\varepsilon^2} = \frac{var[X_1 + \cdots + X_n]}{n^2\varepsilon^2} \\ &= \frac{nvar[X_1]}{n^2\varepsilon^2} = \frac{var[X_1]}{n\varepsilon^2} \to 0, \text{ as } n \to \infty. \end{aligned}$$

□

# Summary

Variance; Inequalities; WLLN

- Variance: $var[X] := E[(X - E[X])^2] = E[X^2] - E[X]^2$
- Fact: $var[aX + b]a^2 var[X]$
- Sum: $X, Y, Z$ pairwise ind. $\Rightarrow var[X + Y + Z] = \cdots$
- Markov: $Pr[X \geq a] \leq E[f(X)]/f(a)$ where ...
- Chebyshev: $Pr[|X - E[X]| \geq a] \leq var[X]/a^2$
- WLLN: $X_m$ i.i.d. $\Rightarrow \frac{X_1 + \cdots + X_n}{n} \approx E[X]$

# Outline

Balls in Bins.

- Birthday.
- Coupon Collector.
- Load balancing.

Geometric Distribution: Memoryless property.
Poission Distribution: Sum of two Poission is Poission.
pause

Tail Sum for Expectation.

Regression (optional.)

# Confidence?

- You flip a coin once and get $H$.

  Do think that $Pr[H] = 1$?

- You flip a coin 10 times and get 5 $H$s.

  Are you sure that $Pr[H] = 0.5$?

- You flip a coin $10^6$ times and get 35% of $H$s.

  How much are you willing to bet that $Pr[H]$ is exactly 0.35?

  How much are you willing to bet that $Pr[H] \in [0.3, 0.4]$?

  Did different exam rooms perform differently? (6 afraid of 7?)

More generally, you estimate an unknown quantity $\theta$.

Your estimate is $\hat{\theta}$.

How much confidence do you have in your estimate?

# Confidence?

Confidence is essential is many applications:

- How effective is a medication?
- Are we sure of the milage of a car?
- Can we guarantee the lifespan of a device?
- We simulated a system. Do we trust the simulation results?
- Is an algorithm guaranteed to be fast?
- Do we know that a program has no bug?

As scientists and engineers, be convinced of this fact:

**An estimate without confidence level is useless!**

# Confidence Interval

The following definition captures precisely the notion of confidence.

**Definition: Confidence Interval**

An interval $[a,b]$ is a 95%-confidence interval for an unknown quantity $\theta$ if

$$Pr[\theta \in [a,b]] \geq 95\%.$$

The interval $[a,b]$ is calculated on the basis of observations.

Here is a typical framework. Assume that $X_1, X_2, \ldots, X_n$ are i.i.d. and have a distribution that depends on some parameter $\theta$.

For instance, $X_n = B(\theta)$.

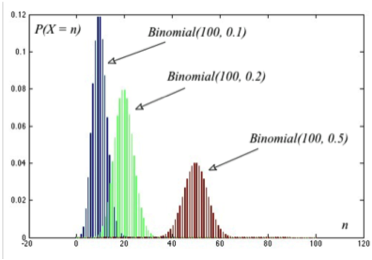
Thus, more precisely, given $\theta$, the random variables $X_n$ are i.i.d. with a known distribution (that depends on $\theta$).

- We observe $X_1, \ldots, X_n$
- We calculate $a = a(X_1, \ldots, X_n)$ and $b = b(X_1, \ldots, X_n)$
- If we can guarantee that $Pr[\theta \in [a,b]] \geq 95\%$, then $[a,b]$ is a 95%-CI for $\theta$.

# Confidence Interval: Applications

- We poll 1000 people.
    - Among those, 48% declare they will vote for Trump.
    - We do some calculations ....
    - We conclude that $[0.43, 0.53]$ is a 95%-CI for the fraction of all the voters who will vote for Trump.
- We observe $1,000$ heart valve replacements that were performed by Dr. Bill.
    - Among those, 35 patients died during surgery. (Sad example!)
    - We do some calculations ...
    - We conclude that $[1\%, 5\%]$ is a 95%-CI for the probability of dying during that surgery by Dr. Bill.
    - We do a similar calculation for Dr. Fred.
    - We find that $[8\%, 12\%]$ is a 95%-CI for Dr. Fred's surgery.
    - What surgeon do you choose?

# Coin Flips: Intuition



Say that you flip a coin $n = 100$ times and observe 20 Hs.

If $p := Pr[H] = 0.5$, this event is very unlikely.

Intuitively, if is unlikely that the fraction of Hs, say $A_n$, differs a lot from $p := Pr[H]$.

Thus, it is unlikely that $p$ differs a lot from $A_n$. Hence, one should be able to build a confidence interval $[A_n - \varepsilon, A_n + \varepsilon]$ for $p$.

The key idea is that $|A_n - p| \leq \varepsilon \Leftrightarrow p \in [A_n - \varepsilon, A_n + \varepsilon]$.

Thus, $Pr[|A_n - p| > \varepsilon] \leq 5\% \Leftrightarrow Pr[p \in [A_n - \varepsilon, A_n + \varepsilon]] \geq 95\%$.

It remains to find $\varepsilon$ such that $Pr[|A_n - p| > \varepsilon] \leq 5\%$.

One approach: Chebyshev.

# Confidence Interval with Chebyshev

- Flip a coin $n$ times. Let $A_n$ be the fraction of $H$s.
- Can we find $\varepsilon$ such that $Pr[|A_n - p| > \varepsilon] \leq 5\%$?

Using Chebyshev, we will see that $\varepsilon = 2.25\frac{1}{\sqrt{n}}$ works. Thus

$$[A_n - \frac{2.25}{\sqrt{n}}, A_n + \frac{2.25}{\sqrt{n}}] \text{ is a 95\%-CI for } p.$$

Example: If $n = 1500$, then $Pr[p \in [A_n - 0.05, A_n + 0.05]] \geq 95\%$.

In fact, $a = \frac{1}{\sqrt{n}}$ works, so that with $n = 1,500$ one has
$Pr[p \in [A_n - 0.02, A_n + 0.02]] \geq 95\%$.

# Confidence Intervals: Result

**Theorem:**
Let $X_n$ be i.i.d. with mean $\mu$ and variance $\sigma^2$.
Define $A_n = \frac{X_1 + \cdots + X_n}{n}$. Then,

$$Pr[\mu \in [A_n - 4.5\frac{\sigma}{\sqrt{n}}, A_n + 4.5\frac{\sigma}{\sqrt{n}}]] \geq 95\%.$$

Thus, $[A_n - 4.5\frac{\sigma}{\sqrt{n}}, A_n + 4.5\frac{\sigma}{\sqrt{n}}]]$ is a 95%-CI for $\mu$.

Example: Let $X_n = 1\{$ coin $n$ yields $H\}$. Then

$$\mu = E[X_n] = p := Pr[H]. \text{ Also, } \sigma^2 = var(X_n) = p(1-p) \leq \frac{1}{4}.$$

Hence, $[A_n - 4.5\frac{1/2}{\sqrt{n}}, A_n + 4.5\frac{1/2}{\sqrt{n}}]]$ is a 95%-CI for $p$.

# Confidence Interval: Analysis

We prove the theorem, i.e., that $A_n \pm 4.5\sigma/\sqrt{n}$ is a 95%-CI for $\mu$.

From Chebyshev:

$$Pr[|A_n - \mu| \geq 4.5\sigma/\sqrt{n}] \leq \frac{var(A_n)}{[4.5\sigma/\sqrt{n}]^2} = \frac{n}{20\sigma^2} var(A_n).$$

Now,

$$\begin{array}{rcl} var(A_n) & = & var(\frac{X_1 + \cdots + X_n}{n}) = \frac{1}{n^2} var(X_1 + \cdots + X_n) \\ & = & \frac{1}{n^2} \times n.var(X_1) = \frac{1}{n}\sigma^2. \end{array}$$

Hence,

$$Pr[|A_n - \mu| \geq 4.5\sigma/\sqrt{n}] \leq \frac{n}{20\sigma^2} \times \frac{1}{n}\sigma^2 = 5\%.$$

Thus,

$$Pr[|A_n - \mu| \leq 4.5\sigma/\sqrt{n}] \geq 95\%.$$

Hence,

$$Pr[\mu \in [A_n - 4.5\sigma/\sqrt{n}, A_n + 4.5\sigma/\sqrt{n}]] \geq 95\%.$$

□

# Confidence interval for $p$ in $B(p)$

Let $X_n$ be i.i.d. $B(p)$. Define $A_n = (X_1 + \cdots + X_n)/n$.

**Theorem:**

$$[A_n - \frac{2.25}{\sqrt{n}}, A_n + \frac{2.25}{\sqrt{n}}] \text{ is a 95\%-CI for } p.$$

**Proof:**

We have just seen that

$$Pr[\mu \in [A_n - 4.5\sigma/\sqrt{n}, A_n + 4.5\sigma/\sqrt{n}]] \geq 95\%.$$

Here, $\mu = p$ and $\sigma^2 = p(1-p)$. Thus, $\sigma^2 \leq \frac{1}{4}$ and $\sigma \leq \frac{1}{2}$.

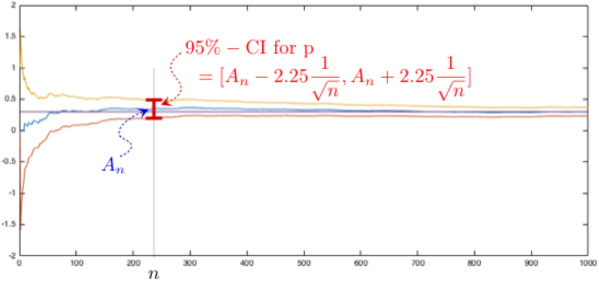
Thus,

$$Pr[\mu \in [A_n - 4.5 \times 0.5/\sqrt{n}, A_n + 4.5 \times 0.5/\sqrt{n}]] \geq 95\%.$$

□

# Confidence interval for *p* in *B*(*p*)

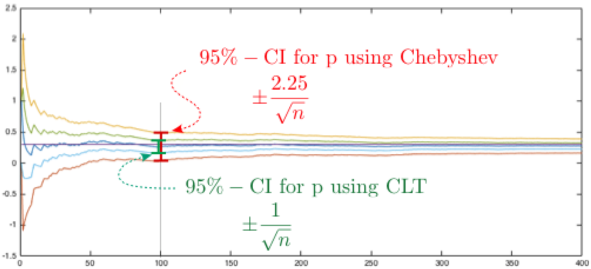
An illustration:



Good practice: You run your simulation, or experiment. You get an estimate. You indicate your confidence interval.

# Confidence interval for *p* in *B*(*p*)

Improved CI: In fact, one can replace 2.25 by 1.



95% – CI for p using Chebyshev

$\pm \frac{2.25}{\sqrt{n}}$

95% – CI for p using CLT

$\pm \frac{1}{\sqrt{n}}$

Quite a bit of work to get there: continuous random variables; Gaussian; Central Limit Theorem.

# Confidence Interval for $1/p$ in $G(p)$

Let $X_n$ be i.i.d. $G(p)$. Define $A_n = (X_1 + \cdots + X_n)/n$.

**Theorem:**

$$[\frac{A_n}{1+4.5/\sqrt{n}}, \frac{A_n}{1-4.5/\sqrt{n}}] \text{ is a 95\%-CI for } \frac{1}{p}.$$

**Proof:** We know that

$$Pr[\mu \in [A_n - 4.5\sigma/\sqrt{n}, A_n + 4.5\sigma/\sqrt{n}]] \geq 95\%.$$

Here, $\mu = \frac{1}{p}$ and $\sigma = \frac{\sqrt{1-p}}{p} \leq \frac{1}{p}$. Hence,

$$Pr[\frac{1}{p} \in [A_n - 4.5\frac{1}{p\sqrt{n}}, A_n + 4.5\frac{1}{p\sqrt{n}}]] \geq 95\%.$$

Now, $A_n - 4.5\frac{1}{p\sqrt{n}} \leq \frac{1}{p} \leq \frac{1}{p} \leq A_n + 4.5\frac{1}{p\sqrt{n}}$ is equivalent to

$$\frac{A_n}{1+4.5/\sqrt{n}} \leq \frac{1}{p} \leq \frac{A_n}{1-4.5/\sqrt{n}}.$$

□

**Examples:** $[0.7A_{100}, 1.8A_{100}]$ and $[0.96A_{10000}, 1.05A_{10000}]$.

# Which Coin is Better?

You are given coin $A$ and coin $B$. You want to find out which one has a larger $Pr[H]$. Let $p_A$ and $p_B$ be the values of $Pr[H]$ for the two coins.

**Approach:**

- Flip each coin $n$ times.
- Let $A_n$ be the fraction of Hs for coin $A$ and $B_n$ for coin $B$.
- Assume $A_n > B_n$. It is tempting to think that $p_A > p_B$. Confidence?

**Analysis:** Note that

$$E[A_n - B_n] = p_A - p_B \text{ and } var(A_n - B_n) = \frac{1}{n}(p_A(1-p_A) + p_B(1-p_B)) \leq \frac{1}{2n}.$$

Thus, $Pr[|A_n - B_n - (p_A - p_B)| > \varepsilon] \leq \frac{1}{2n\varepsilon^2}$, so

$$Pr[p_A - p_B \in [A_n - B_n - \varepsilon, A_n - B_n + \varepsilon]] \geq 1 - \frac{1}{2n\varepsilon^2}, \text{ and}$$

$$Pr[p_A - p_B \geq 0] \geq 1 - \frac{1}{2n(A_n - B_n)^2}.$$

**Example:** With $n = 100$ and $A_n - B_n = 0.2$, $Pr[p_A > p_B] \geq 1 - \frac{1}{8} = 0.875$.

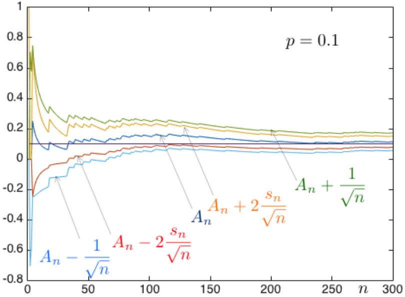
# Unknown $\sigma$

For $B(p)$, we wanted to estimate $p$. The CI requires $\sigma = \sqrt{p(1-p)}$. We replaced $\sigma$ by an upper bound: $1/2$.

In some applications, it may be OK to replace $\sigma^2$ by the following sample variance:

$$s_n^2 := \frac{1}{n} \sum_{m=1}^{n} (X_m - A_n)^2.$$

However, in some cases, this is dangerous! The theory says it is OK if the distribution of $X_n$ is nice (Gaussian). This is used regularly in practice. However, be aware of the risk.



# Summary

Confidence Intervals

1. Estimates without confidence level are useless!
2. $[a,b]$ is a 95%-CI for $\theta$ if $Pr[\theta \in [a,b]] \geq 95\%$.
3. Using Chebyshev: $[A_n - 4.5\sigma/\sqrt{n}, A_n + 4.5\sigma/\sqrt{n}]$ is a 95%-CI for $\mu$.
4. Using CLT, we will replace 4.5 by 2.
5. When $\sigma$ is not known, one can replace it by an upper bound.
6. Examples: $B(p), G(p)$, which coin is better?
7. In some cases, one can replace $\sigma$ by the empirical standard deviation.