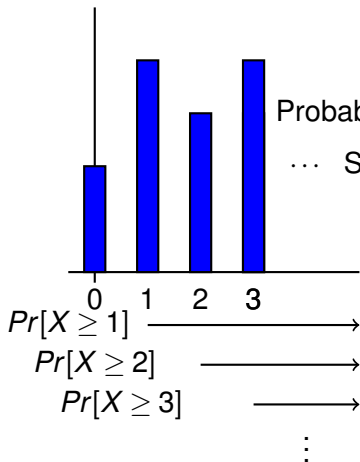


## Alex Psomas: Lecture 19.

1. Distributions
2. Tail bounds

**Theorem:** For a r.v.  $X$  that takes values in  $\{0, 1, 2, \dots\}$ , one has

$$E[X] = \sum_{i=1}^{\infty} \Pr[X \geq i].$$



Probability mass at  $i$ , counted  $i$  times.

... Same as  $\sum_{i=1}^{\infty} i \times \Pr[X = i]$ .

## A side step: Expected Value of Integer RV

**Theorem:** For a r.v.  $X$  that takes values in  $\{0, 1, 2, \dots\}$ , one has

$$E[X] = \sum_{i=1}^{\infty} Pr[X \geq i].$$

**Proof:** One has

$$\begin{aligned} E[X] &= \sum_{i=1}^{\infty} i \times Pr[X = i] \\ &= \sum_{i=1}^{\infty} i (Pr[X \geq i] - Pr[X \geq i+1]) \\ &= \sum_{i=1}^{\infty} (i \times Pr[X \geq i] - i \times Pr[X \geq i+1]) \\ &= \sum_{i=1}^{\infty} i \times Pr[X \geq i] - \sum_{i=1}^{\infty} i \times Pr[X \geq i+1] \\ &= \sum_{i=1}^{\infty} i \times Pr[X \geq i] - \sum_{i=1}^{\infty} (i-1) \times Pr[X \geq i] = \sum_{i=1}^{\infty} Pr[X \geq i]. \end{aligned}$$



# Geometric Distribution: Memoryless

Let  $X$  be  $Geom(p)$ . **Theorem**

$$Pr[X > n + m | X > n] = Pr[X > m], m, n \geq 0.$$

## Geometric Distribution: Memoryless

I flip a coin (probability of  $H$  is  $p$ ) until I get  $H$ .

What's the probability that I flip it exactly 100 times?  $(1 - p)^{99}p$

What's the probability that I flip it exactly 100 times if (given that) the first 20 were  $T$ ?

Same as flipping it exactly 80 times!

$(1 - p)^{79}p$ .

## Variance of geometric distribution.

$X$  is a geometrically distributed RV with parameter  $p$ .

Thus,  $\Pr[X = n] = (1 - p)^{n-1} p$  for  $n \geq 1$ . Recall  $E[X] = 1/p$ .

$$E[X^2] = (2 - p)/p^2 \quad (\text{tricks})$$

$$\text{var}[X] = E[X^2] - E[X]^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

$$\sigma(X) = \frac{\sqrt{1-p}}{p} \approx E[X] \text{ when } p \text{ is small(ish).}$$

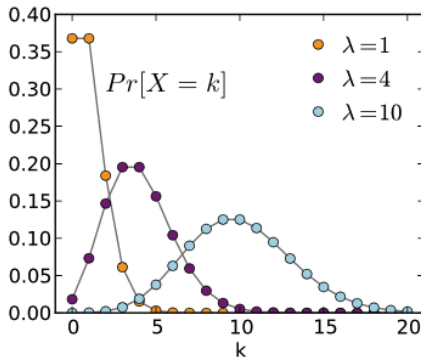
# Poisson

Experiment: flip a coin  $n$  times. The coin is such that

$$Pr[H] = \lambda/n.$$

Random Variable:  $X$  - number of heads. Thus,  $X = B(n, \lambda/n)$ .

**Poisson Distribution** is distribution of  $X$  “for large  $n$ .”



# Poisson

Experiment: flip a coin  $n$  times. The coin is such that

$$Pr[H] = \lambda/n.$$

Random Variable:  $X$  - number of heads. Thus,  $X = B(n, \lambda/n)$ .

**Poisson Distribution** is distribution of  $X$  “for large  $n$ .”

We expect  $X \ll n$ . For  $m \ll n$  one has

$$Pr[X = m] = \frac{\lambda^m}{m!} e^{-\lambda}.$$



# Poisson Distribution: Definition and Mean

**Definition** Poisson Distribution with parameter  $\lambda > 0$

$$X = P(\lambda) \Leftrightarrow \Pr[X = m] = \frac{\lambda^m}{m!} e^{-\lambda}, m \geq 0.$$

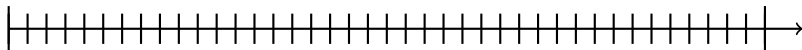
**Fact:**  $E[X] = \lambda$ .

# Poisson and Queueing.

Poisson: Distribution of how many events in an interval?

Average:  $\lambda$ .

What is the maximum number of customers you might see?



Idea: Cut into intervals so that “sum of Bernoulli (indicators)”.

$n = 10$  sub-intervals.

Binomial distribution, if only one event/interval!

Maybe more...

and more.

As  $n$  goes to infinity...analyze ...

$$\dots Pr[X = i] = \binom{n}{i} p^i (1 - p)^{n-i}.$$

derive simple expression.

...And we get the Poisson distribution!

## When to use Poisson

If an event can occur 0,1,2,... times in an interval,  
and the average number of events per interval is  $\lambda$   
and events are independent  
and the probability of an event in an interval is proportional to  
the interval's length,  
then it might be appropriate to use Poisson distribution.

$$Pr[k \text{ events in interval}] = \frac{\lambda^k}{k!} e^{-\lambda}$$

Examples: photons arriving at a telescope, telephone calls arriving in a system, the number of mutations on a strand of DNA per unit length...

# Simeon Poisson

The Poisson distribution is named after:

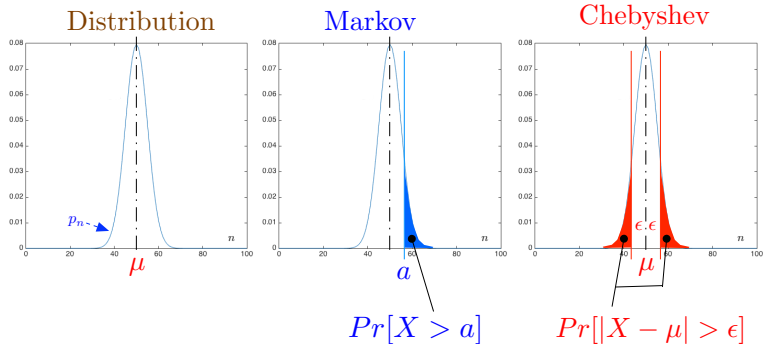


“Life is good for only two things: doing mathematics and teaching it.”

## Review: Distributions

- ▶  $Bern(p) : Pr[X = 1] = p;$   
 $E[X] = p;$   
 $Var[X] = p(1 - p);$
- ▶  $B(n, p) : Pr[X = m] = \binom{n}{m} p^m (1 - p)^{n-m}, m = 0, \dots, n;$   
 $E[X] = np;$   
 $Var[X] = np(1 - p);$
- ▶  $U[1, \dots, n] : Pr[X = m] = \frac{1}{n}, m = 1, \dots, n;$   
 $E[X] = \frac{n+1}{2};$   
 $Var[X] = \frac{n^2-1}{12};$
- ▶  $Geom(p) : Pr[X = n] = (1 - p)^{n-1} p, n = 1, 2, \dots;$   
 $E[X] = \frac{1}{p};$   
 $Var[X] = \frac{1-p}{p^2};$
- ▶  $P(\lambda) : Pr[X = n] = \frac{\lambda^n}{n!} e^{-\lambda}, n \geq 0;$   
 $E[X] = \lambda;$   
 $Var[X] = \lambda.$

# Inequalities: An Overview



# Andrey Markov

**Andrey (Andrei) Andreyevich  
Markov**



**Born** 14 June 1856 N.S.  
Ryazan, Russian Empire

**Died** 20 July 1922 (aged 66)  
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's inequality

The inequality is named after Andrey Markov, although it appeared earlier in the work of Pafnuty Chebyshev. It should be (and is sometimes) called Chebyshev's first inequality.

**Theorem** Markov's Inequality (the fancy version)

Assume  $f : \mathfrak{X} \rightarrow [0, \infty)$  is nondecreasing. Then, for a non-negative random variable  $X$

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

**Proof:**

Observe that

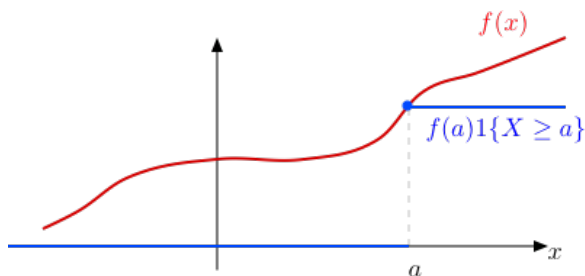
$$1_{\{X \geq a\}} \leq \frac{f(X)}{f(a)}.$$

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. Expectation is monotone: if  $X(\omega) \leq Y(\omega)$  for all  $\omega$ , then  $E[X] \leq E[Y]$ . Therefore,

$$E[1_{\{X \geq a\}}] \leq \frac{E[f(X)]}{f(a)}.$$



## 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 Note

A more common version of Markov is for  $f(x) = x$ :

**Theorem** For a non-negative random variable  $X$ , and any  $a > 0$ ,

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

# Markov Inequality Example: Geom(p)

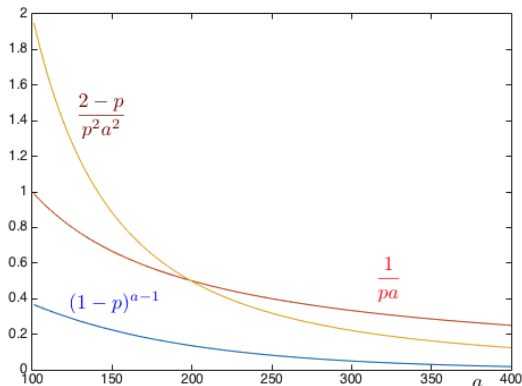
Let  $X \sim \text{Geom}(p)$ . Recall that  $E[X] = \frac{1}{p}$  and  $E[X^2] = \frac{2-p}{p^2}$ .

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's inequality example

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

What is a bound on the probability that a random  $X$  takes value  $\geq$  than twice its' expectation?

$\frac{1}{2}$ . It can't be that more than half of the people are twice above the average!

What is a bound on the probability that a random  $X$  takes value  $\geq$  than  $k$  times its' expectation?

$\frac{1}{k}$ .

## Markov's inequality example

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  $p$ .

$X \sim B(n, p)$ .

$E[X] = np$ . Say  $n = 1000$  and  $p = 0.5$ .  $E[X] = 500$ .

Markov says that  $Pr[X \geq 600] \leq \frac{1000 \cdot 0.5}{600} = \frac{5}{6} \approx 0.83$

Actual probability:  $< 0.000001$

Notice: Same bound for 10 coins and  $Pr[X \geq 6]$

# Chebyshev's Inequality

This is Pafnuty's inequality:

**Theorem:**

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

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

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

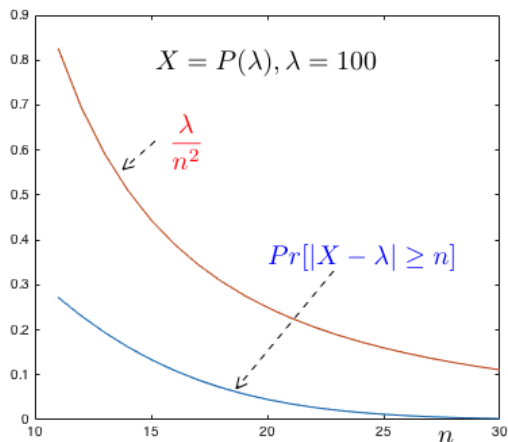


This result confirms that the variance measures the “deviations from the mean.”

# Chebyshev and Poisson

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

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



## Chebyshev's inequality example

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  $Pr[X \geq 600] = Pr[X - 500 \geq 100] \leq Pr[|X - 500| \geq 100] \leq \frac{250}{10000} = 0.025$

Actual probability:  $< 0.000001$

Notice: If we had 100 coins, the bound for  $Pr[X \geq 60]$  would be different.



## Chebyshev's inequality example continued

What if we had more coins?

Also, let's count the fraction of  $H$  instead of their number.  
 $p$  is still 0.5

Let  $X_i$  be the indicator random variable for the  $i$ -th coin.

Define

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

$$E[Y_n] = \frac{1}{n} E[\sum X_i] = \frac{1}{n} np = p = 0.5$$

$$\text{Var}[Y_n] = \frac{1}{n^2} \text{Var}[\sum X_i] = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n} = \frac{1}{4n}$$

Let's try to bound how likely it is that the fraction of  $H$ 's differs from 50%.

$$\Pr[|Y_n - 0.5| \geq \varepsilon]?$$

## Chebyshev's inequality example continued

$$E[Y_n] = 0.5, \text{Var}[Y_n] = \frac{1}{4n}.$$

$$Pr[|Y_n - 0.5| \geq \varepsilon] \leq \frac{\text{Var}[Y_n]}{\varepsilon^2} = \frac{1}{4n\varepsilon^2}$$

$$\text{For } \varepsilon = 0.01: Pr[|Y_n - 0.5| \geq 0.01] \leq \frac{2500}{n}$$

For  $n = 250,000$  this is 1%.

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

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

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

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

$$Pr\left[\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| \geq \varepsilon\right] \rightarrow 0, \text{ as } n \rightarrow \infty.$$

### Proof:

Let  $Y_n = \frac{X_1 + \dots + X_n}{n}$ . Then

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

(I used that variance is finite for this proof. More complicated proof without this assumption.)



## Confidence intervals example

Say  $p$  in the previous example was 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%?

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

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

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

## 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.

# Today's gig: ?

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.

Today: A magic trick.

# Summary

- ▶ Variance of Geometric.
- ▶ Markov's Inequality
- ▶ Chebyshev's Inequality.