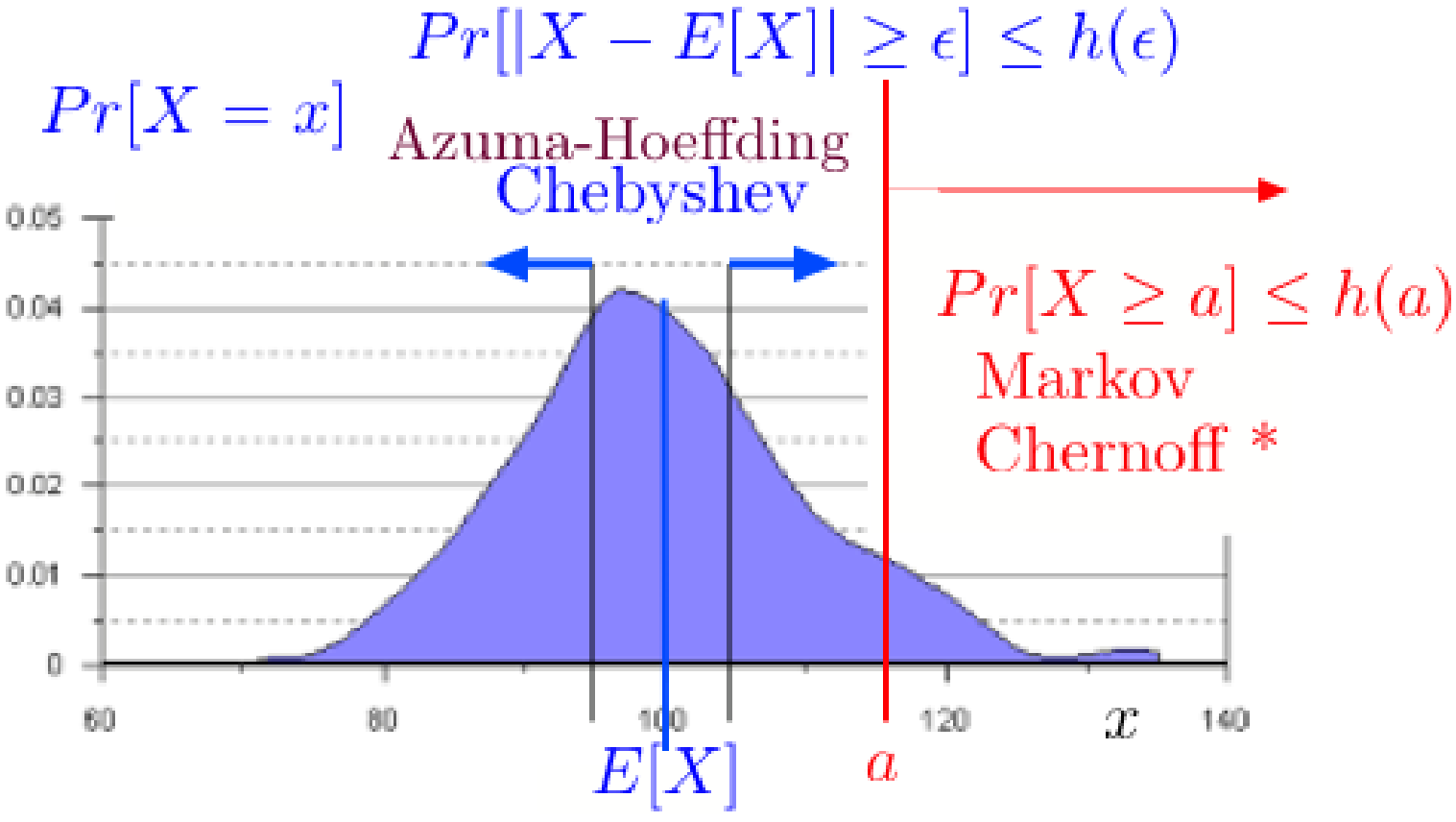
CS70: Jean Walrand: Lecture 30.

Markov & Chebyshev; A brief review

- Bounds:
  1. Markov
  2. Chebyshev
- A brief review
- A mini-quizz

Bounds: An Overview



* Chernoff: when $X = X_1 + \cdots + X_n$ (i.i.d.)

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 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.

Monotonicity

Let X be a RV. We write $X \geq 0$ if all the possible values of X are nonnegative. That is, $Pr[X \geq 0] = 1$.

It is obvious that

$$X \geq 0 \Rightarrow E[X] \geq 0.$$

We write $X \geq Y$ if $X - Y \geq 0$.

Then,

$$X \geq Y \Rightarrow X - Y \geq 0 \Rightarrow E[X] - E[Y] = E[X - Y] \geq 0.$$

Hence,

$$X \geq Y \Rightarrow E[X] \geq E[Y].$$

We say that expectation is monotone.
(Which means monotonic, but not monotonous!)

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

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:

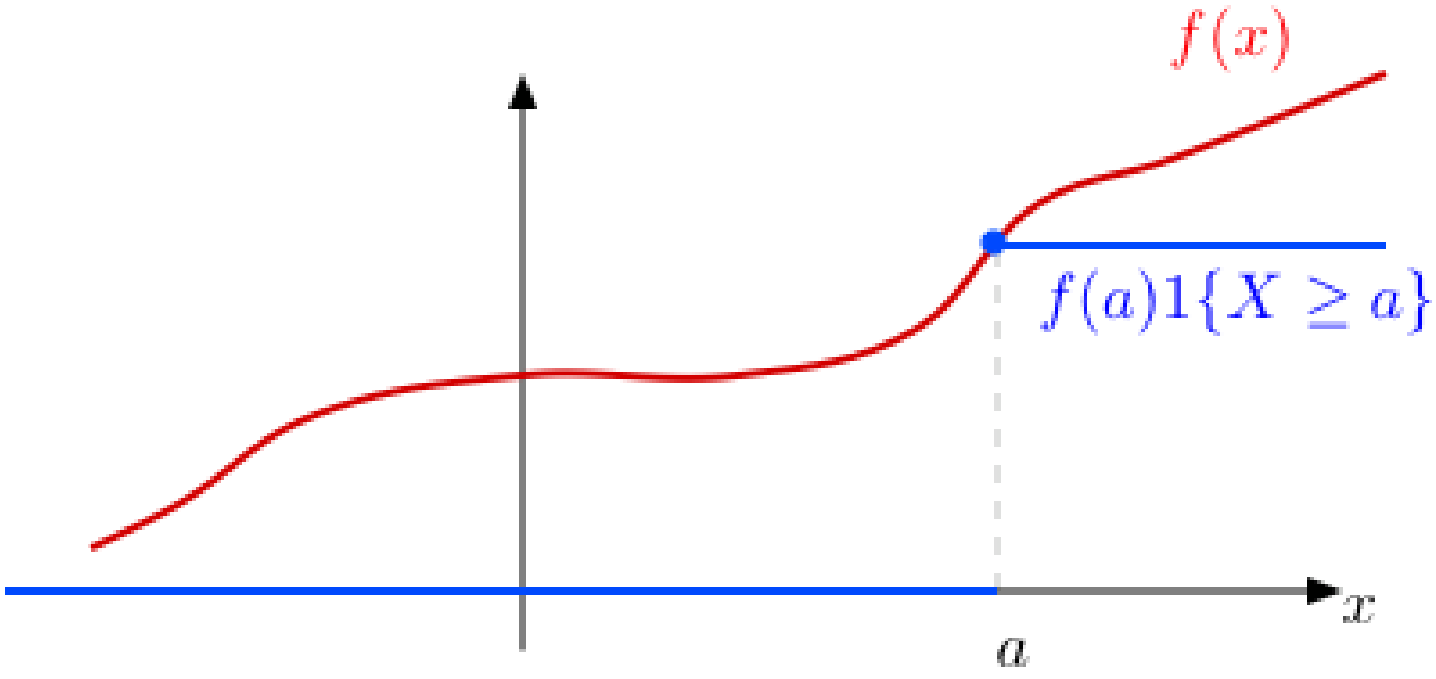
Observe that

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

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. □

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)

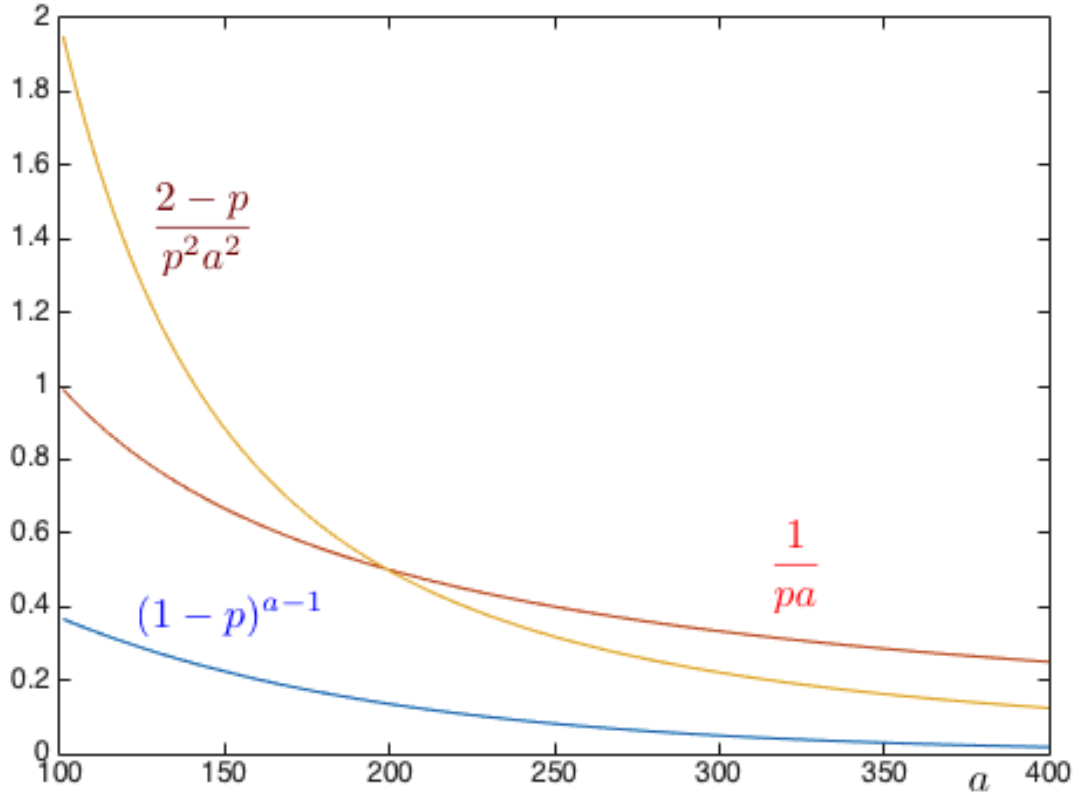
Let $X = G(p)$. Recall that $E[X] = \frac{1}{p}$ and $E[X^2] = \frac{2-p}{p^2}$.

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: $P(\lambda)$

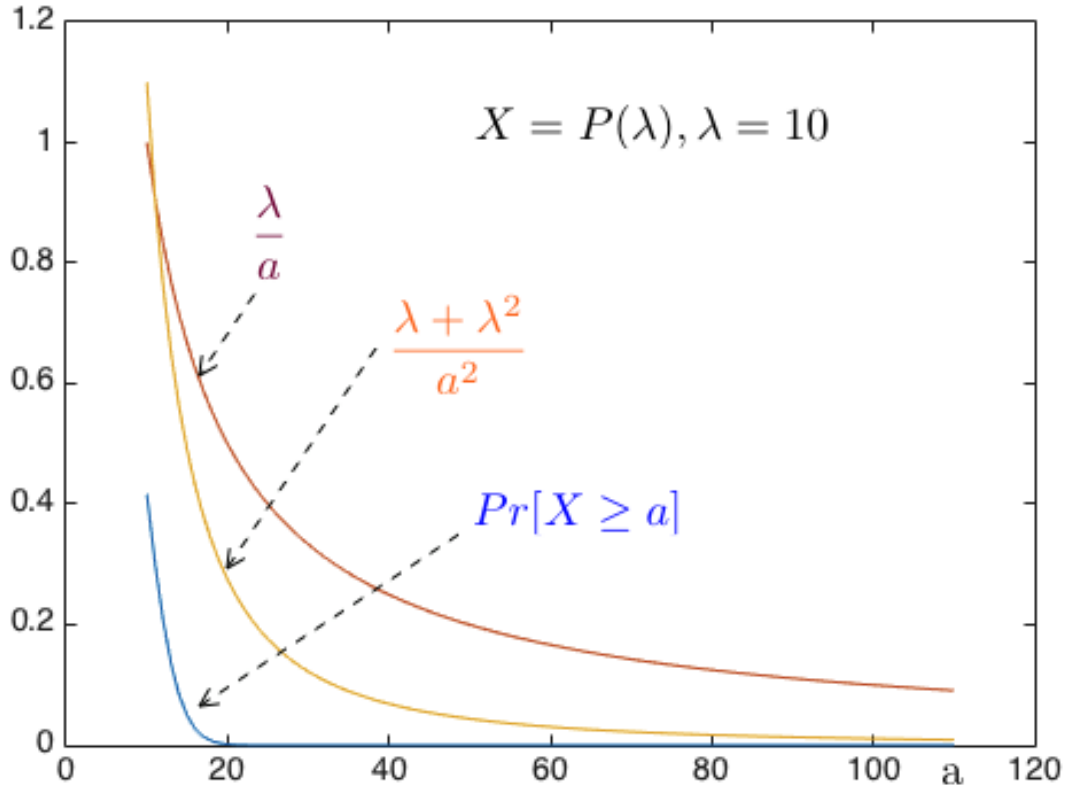
Let $X = P(\lambda)$. Recall that $E[X] = \lambda$ and $E[X^2] = \lambda + \lambda^2$.

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,

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

□

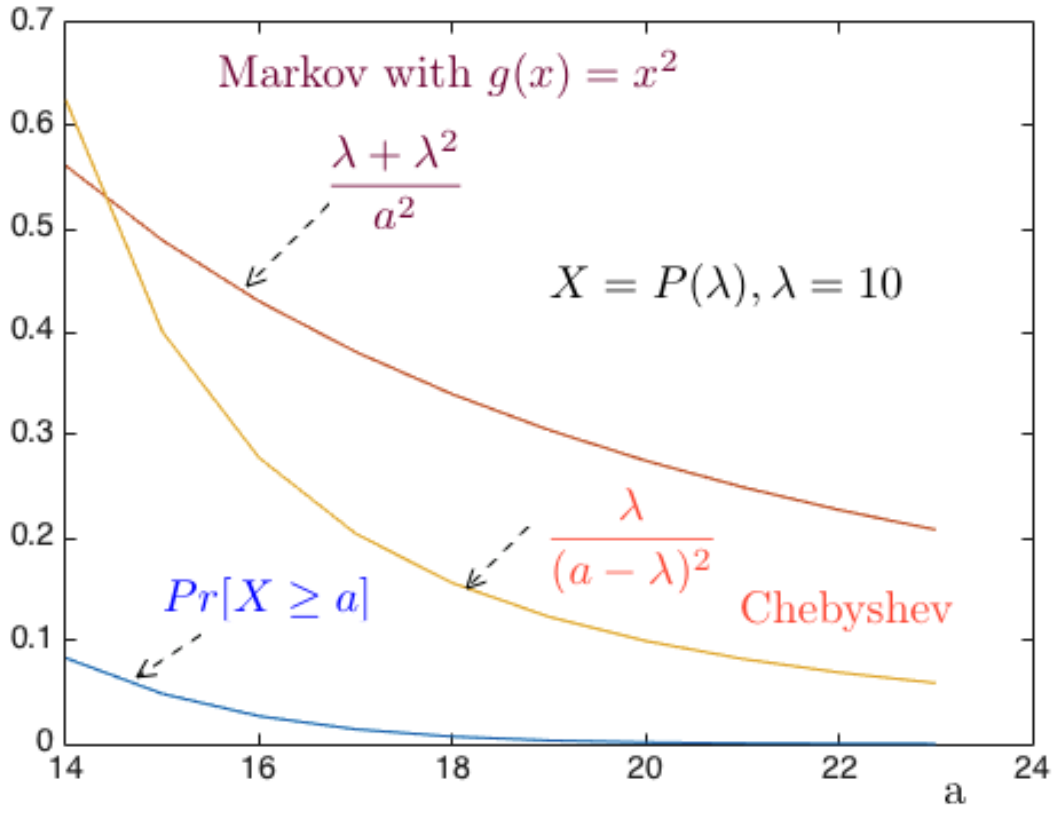
This result confirms that the variance measures the "deviations from the mean."

Chebyshev and $P(\lambda)$

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

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

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



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,

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

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.

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 independent with the same distribution. Then, for all $\varepsilon > 0$,

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

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

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

Summary

Bounds!

- Markov: $Pr[X \geq a] \leq E[f(X)]/f(a)$ where ...
- Chebyshev: $Pr[|X - E[X]| \geq a] \leq var[X]/a^2$

A brief review - 1

- Framework:

$$\Omega, Pr[\omega], P[A] = \sum_{\omega \in A} Pr[\omega]$$

- Conditional Probability and Bayes' Rule

$$Pr[A|B] := \frac{Pr[A \cap B]}{Pr[B]}; Pr[A_m|B] = \frac{Pr[A_m]Pr[B|A_m]}{\sum_k Pr[A_k]Pr[B|A_k]}$$

- Independence:

$$A, B \text{ ind. } \Leftrightarrow Pr[A \cap B] = Pr[A]Pr[B]$$

- Mutual Independence:

$$A, B, C, \ldots \text{ mutually independent } \Leftrightarrow \cdots$$

- Random Variable:

$$X : \Omega \to \Re$$

A brief review - 2

- Expectation:

$$E[X] := \sum_a a \times Pr[X = a] = \sum_\omega X(\omega) Pr[\omega].$$
$$E[aX + bY] = aE[X] + bE[Y].$$

- Independent RVs:

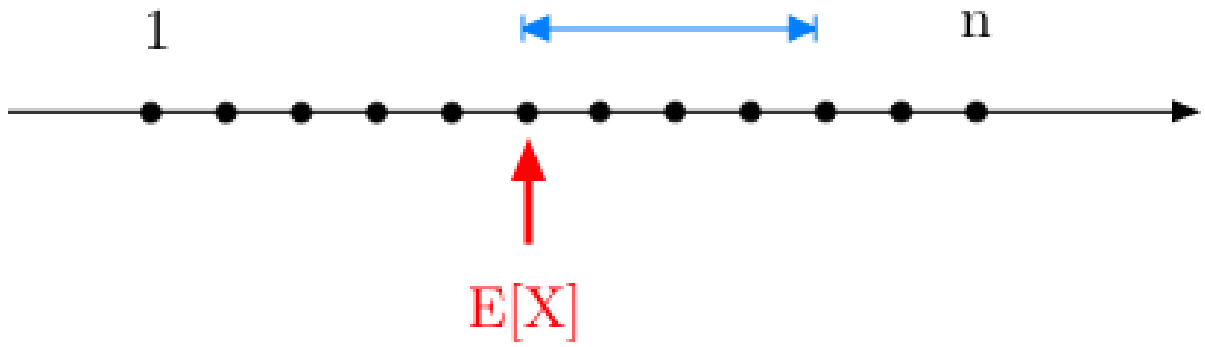
  X, Y are independent if and only if $Pr[X \in A, Y \in B] = \cdots$

- If independent, then
  - $f(X), g(Y)$ independent
  - $E[XY] = E[X]E[Y]$
  - $var[X + Y] = var[X] + var[Y]$

A brief review - 3

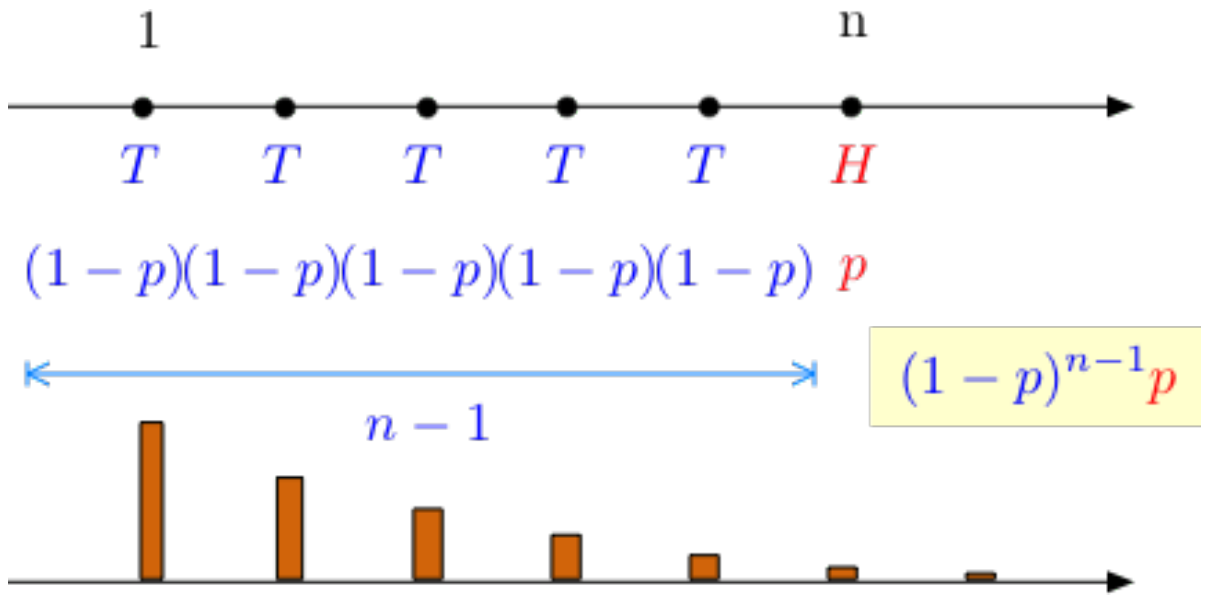
$$\sigma(X) = \sqrt{\frac{n^2-1}{12}}$$

$$\sigma(X) \approx n/3.46$$



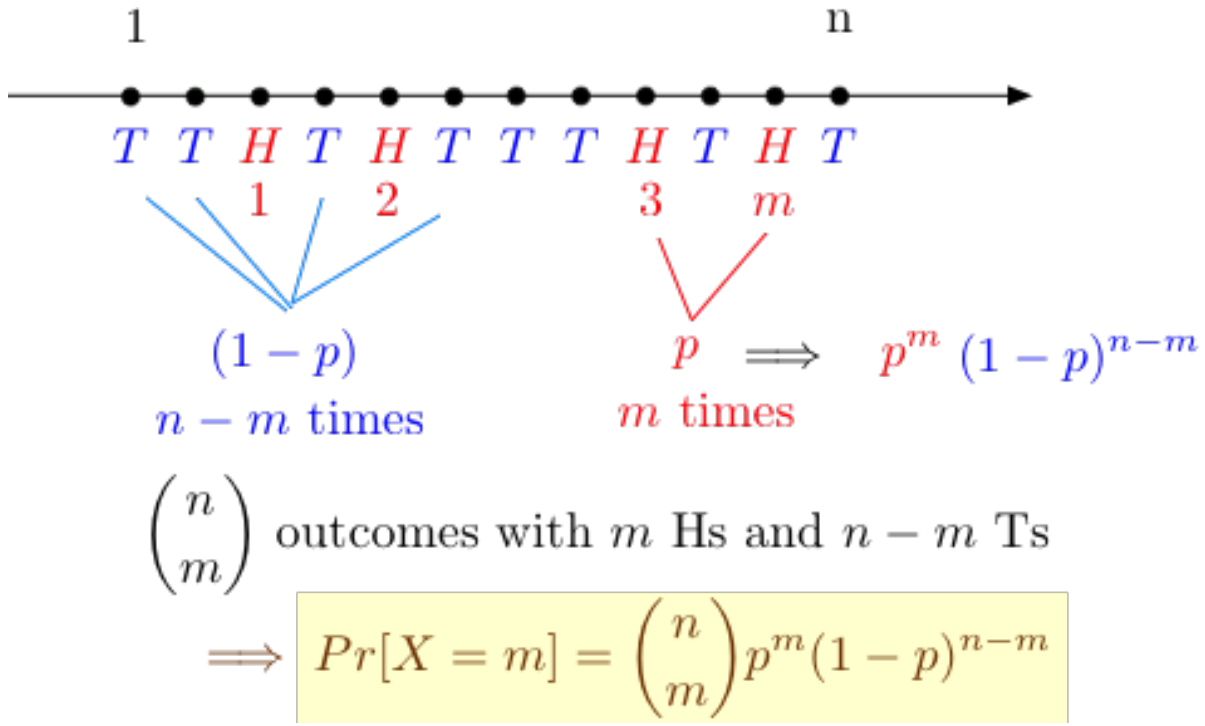
- Uniform distribution: $U[1,\ldots,n]$
  - Random experiment: Pick a number uniformly in $\{1,\ldots,n\}$
  - Definition: $Pr[X=m] = 1/n, m = 1,\ldots,n$;
  - Mean: $E[X] = (n+1)/2$; Calculation
  - Variance: $var[X] = (n^2-1)/12$; Calculation

A brief review - 4



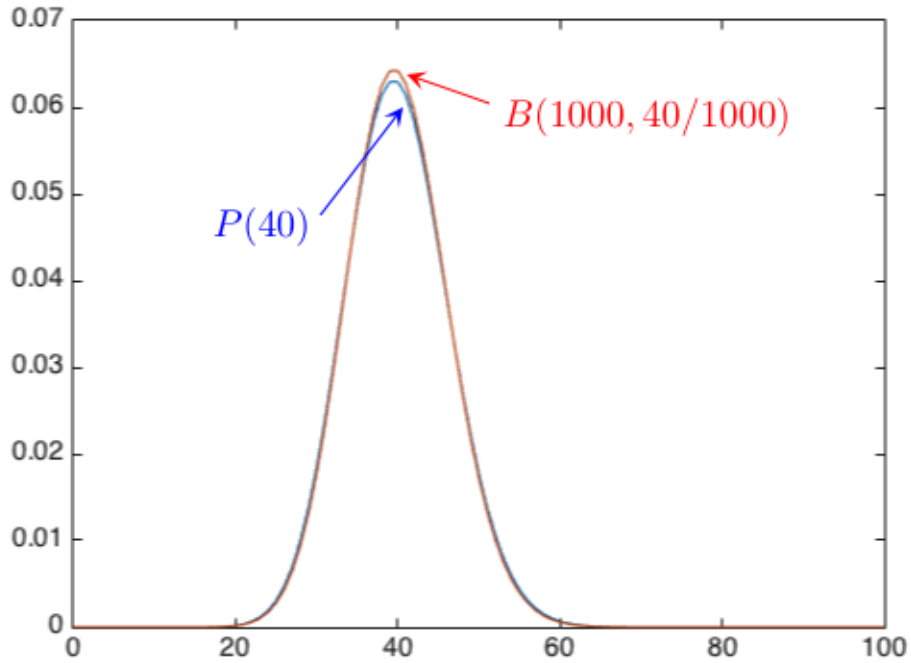
- Geometric distribution: $G(p), p \in [0,1]$
  - Random experiment: Number of flips until first H.
  - Definition: $Pr[X = m] = (1-p)^{m-1}p, m \geq 1$
  - Mean: $E[X] = 1/p$; "trick" or renewal
  - Variance: $var[X] = (1-p)/p^2$; "trick" or renewal
  - Renewal: $X = 1 + ZY, Y \equiv X, Pr[Z = 1] = q := 1-p$
    $E[X] = 1 + qE[X]; E[X^2] = 1 + 2qE[X] + qE[X^2]$.

A brief review - 5



- Binomial distribution: $B(n,p), n \geq 1, p \in [0,1]$
  - Random experiment: Number of heads out of n flips.
  - Definition: $Pr[X = m] = \binom{n}{m} p^m (1-p)^{n-m}, m = 0, 1, \ldots, n$
  - Mean: $E[X] = np$; $X = X_1 + \cdots + X_n$, $E[X] = nE[X_1]$, by linearity
  - Variance: $var[X] = np(1-p)$;
    $X = X_1 + \cdots + X_n$; $var[X] = nvar[X_1]$ by ind.

A brief review - 6



- Poisson distribution: $P(\lambda), \lambda > 0$
  - Random experiment: Number of H's out of $n \gg 1$ flips with $p = \lambda/n = B(n, \lambda/n)$.
  - Definition: $Pr[X = m] = \frac{\lambda^m}{m!} e^{-\lambda}, m \geq 0$
  - Mean: $E[X] = \lambda$; $E[X] = n(\lambda/n) = \lambda$
  - Variance: $var[X] = \lambda$;
    $var[X] = n(\lambda/n)(1 - \lambda/n) = \lambda - \lambda^2/n = \lambda$ since $n \gg 1$.

A mini-quizz

True or False:

- Probability Space:
  - $Pr[A \cup B] = Pr[A] + Pr[B]$. False True iff disjoint.
  - $Pr[A \cap B] = Pr[A]Pr[B]$. False True iff independent.
  - $A \cap B = \emptyset \Rightarrow A, B$ independent. False
  - For all A, B, one has $Pr[A|B] \geq Pr[A]$. False
  - $Pr[A \cap B \cap C] = Pr[A]Pr[B|A]Pr[C|B]$. False
- Random Variables:
  - For all X, Y, one has $E[XY] = E[X]E[Y]$. False True if (not iff) independent.
  - $var[3X + 1] = 3var[X]$. False $= 9var[X]$.
  - $var[2X + 3Y] = 4var[X] + 9var[Y]$. False True if (not iff) independent.
  - $E[2X + 3Y + 1] = 2E[X] + 3E[Y] + 1$. True

Inequalities: Some Applications

Below, Y is measured and used to construct a guess $g(Y)$ about X. Here are some possible goals.

- Detection: $Pr[X \neq g(Y)] \leq 10^{-8}$ (e.g., bit error rate)
- Estimation: $Pr[|X - g(Y)| > a] \leq 5\%$ (e.g., polling)
- Hypothesis Testing: $Pr[g(Y) = 1 | X = 0] \leq 5\%$ and $Pr[g(Y) = 0 | X = 1] \leq 5\%$ (e.g., fire alarm)
- Design: $Pr[\text{buffer overflow}] \leq 5\%$
- Control: $Pr[\text{crash}] \leq 10^{-6}$

Many situations where exact probabilities are difficult to compute but bounds can be found.