The central limit theorem

- Consider a sequence X_k of Bernoulli (p) trials.
- By the (strong) LLN we know that $\frac{\sum_1^n X_k}{n}$ will converge to p.
- In particular, the limit is deterministic, there is nothing random about it anymore.
- What is the limit of $\frac{\sum_1^n X_k - np}{n}$?
- This makes sense in light of what we discovered last time: $\sum_1^n X_k$ is concentrated in an interval of size $c\sqrt{n}$ about its mean np.
- A different type of limit is encountered if we normalize by $\sqrt{n}$ instead of n.
- The following graphs depict the pmf of

$$\hat{S}_n = \frac{\sum_{k=1}^n X_k - np}{\sqrt{np(1-p)}}.$$

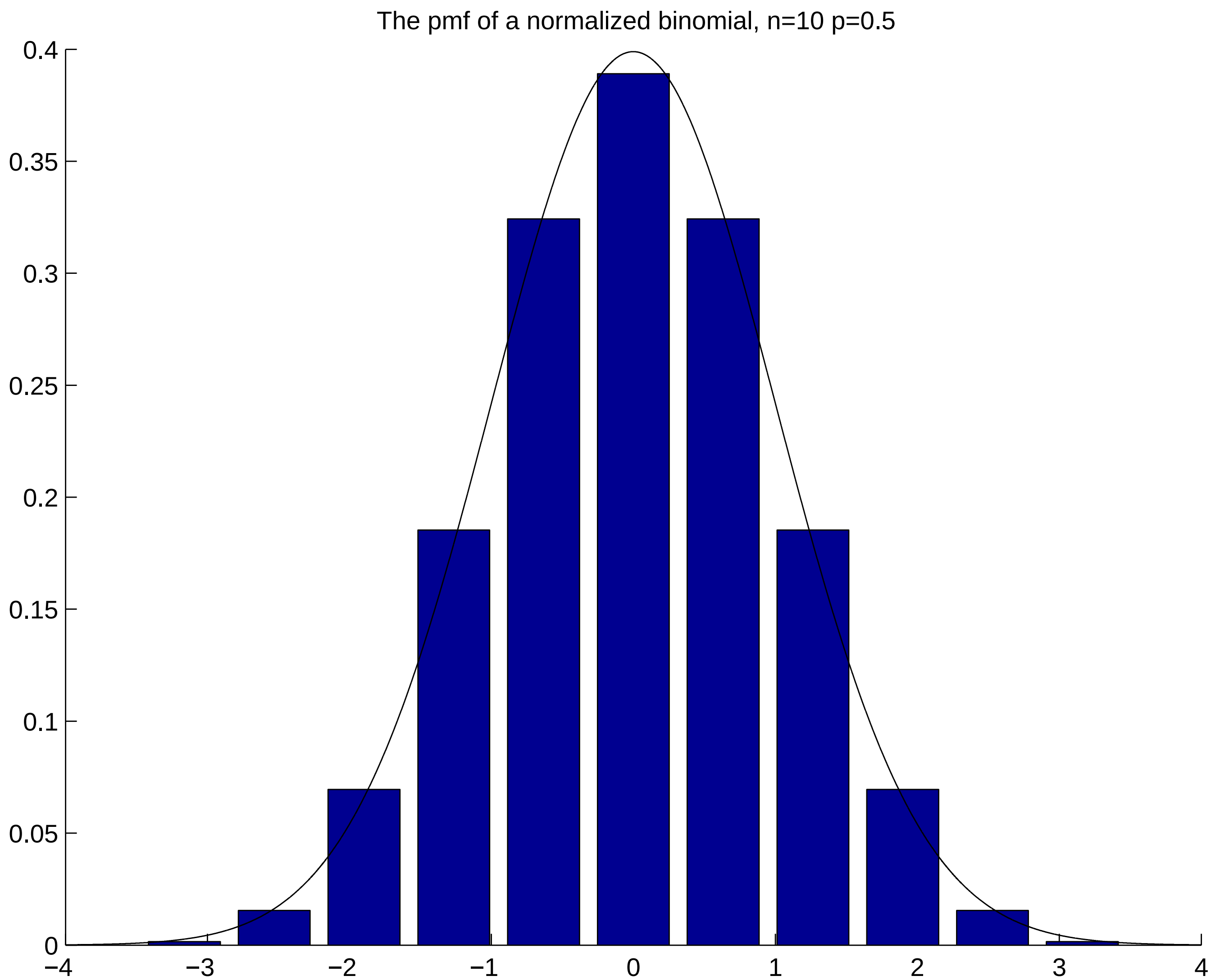

Figure 1: The pmf of $\hat{S}_{10,0.5}$

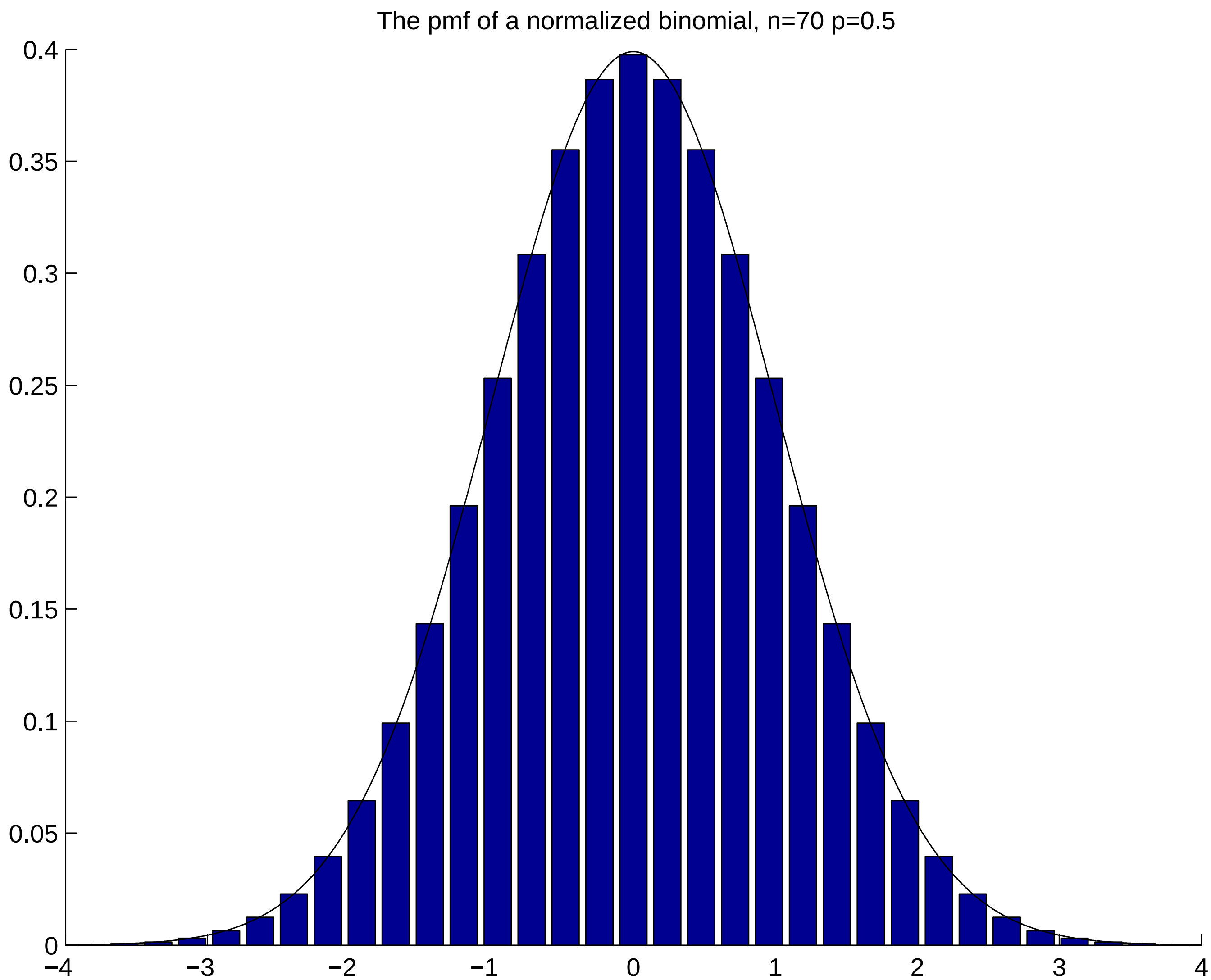

Figure 2: The pmf of $\hat{S}_{70,0.5}$

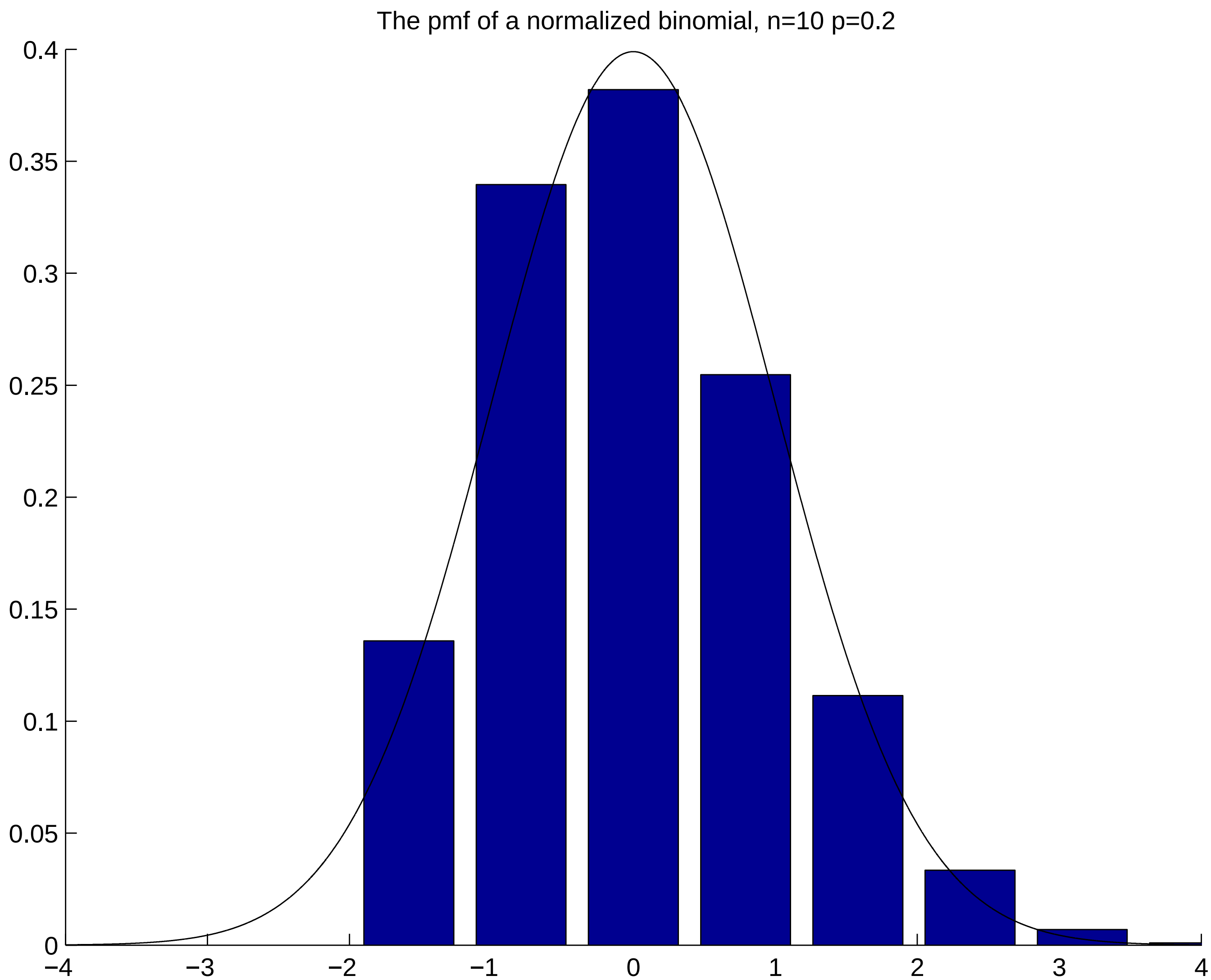

Figure 3: The pmf of $\hat{S}_{10,0.2}$

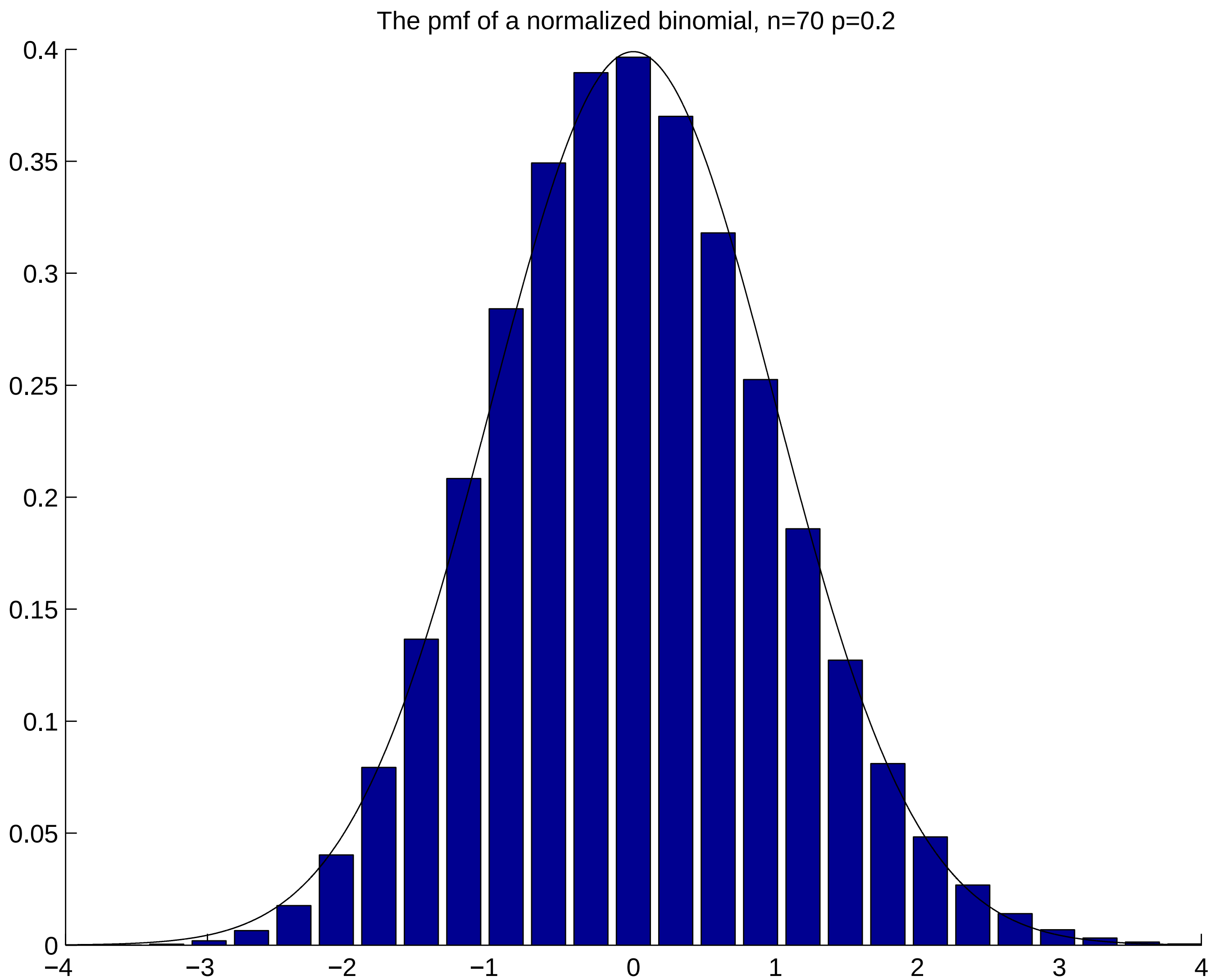

Figure 4: The pmf of $\hat{S}_{70,0.2}$

The central limit theorem

- The curve that you saw in all the graphs was that of $\varphi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$.
- If I tell you that this limit holds for a sequence of properly normalized Poisson iid random variables can you guess what the normalization is?
- **Theorem.** Suppose X_k are a sequence of iid random variables with mean μ and variance σ^2. Then

$$\lim_{n\to\infty} \Pr\left(\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n\sigma^2}} \leq \alpha\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{-x^2/2}.$$

- The *continuous* distribution on the rhs is called the normal or Gaussian distribution.

Conditional Expectation

- If X is the result of a fair die toss then $E(X) = 7/2$.
- Suppose we are given that X is even. Will that change the expected value?
- Now there are only three equally likely results: 2, 4, & 6, so the conditional expectation will be 4.
- **Def.** $E(X|A)$ the *conditional expectation* of X given A is defined for A with $\Pr(A) > 0$ by:

$$E(X|A) = \sum_x x \Pr(X = x|A).$$

- **Example.** Find $E(X|X \text{ is odd})$. Let $A = \{\omega : X(\omega) = 1, 3, 5\}$. Then

$$\begin{aligned} E(X|A) &= \sum_x x \Pr(X = x|A) \\ &= \sum_x x \frac{\Pr(\{X = x\} \cap A)}{\Pr(A)}. \\ &= 1 \cdot \frac{1/6}{1/2} + 3 \cdot \frac{1/6}{1/2} + 5 \cdot \frac{1/6}{1/2} = 3. \end{aligned}$$

- Note that in this example

$$E(X) = 7/2 = (3+4)/2 = E(X|A)/2 + E(X|\bar{A})/2.$$

Theorem: For all events A such that $\Pr(A), \Pr(\overline{A}) > 0$:

$$E(X) = E(X|A)\Pr(A) + E(X|\overline{A})\Pr(\overline{A})$$

Proof:

$$\begin{aligned}
E(X) &= \sum_x x \Pr(X = x) \\
&= \sum_x x \left[\Pr(\{X = x\} \cap A) + \Pr(\{X = x\} \cap \overline{A})\right] \\
&= \sum_x x \Big[\Pr(X = x|A)\Pr(A) \\
&\qquad\qquad + \Pr(X = x|\overline{A})\Pr(\overline{A})\Big] \\
&= \sum_x \left[x \Pr(X = x|A)\Pr(A)\right] \\
&\qquad\qquad + \left[x \Pr(X = x|\overline{A})\Pr(\overline{A})\right] \\
&= E(X|A)\Pr(A) + E(X|\overline{A})\Pr(\overline{A})
\end{aligned}$$

Example

- I toss a fair die. If it lands with 3 or more, I toss 5 times a coin with $\Pr(H) = p_1$. If it lands with less than 3, I toss 5 times a coin with $\Pr(H) = p_2$. What is the expected number of heads, X?
- Let A be the event that the die lands with 3 or more.
- Clearly, $\Pr(A) = 2/3$.
- What is $E(X|A)$?
- Conditioned on A, X is binomial B_{n,p_1} so

$$E(X|A) = np_1.$$

- Similarly for $\bar{A}$, so by the previous theorem,

$$\begin{aligned} E(X) &= E(X|A)\Pr(A) + E(X|\overline{A})\Pr(\overline{A}) \\ &= np_1 \cdot \frac{2}{3} + np_2 \cdot \frac{1}{3}. \end{aligned}$$

The Rabin-Miller Test

- **Input:**
  - $\cdot\ n = 2^s t + 1$ where t is odd and $s \in \mathbb{N}$
  - $\cdot\ b \in \{1, 2, \ldots, n-1\}$
- $\mathbf{T_{RM}}$**:** Does *exactly* one of the following hold?
  - $\cdot\ b^t \equiv 1 \pmod{n}$ or
  - $\cdot\ b^{2^j t} \equiv -1 \pmod{n}$ for one $0 \leq j \leq s-1$.
- **Claim.**
  - $\cdot$ If n is prime, $\mathrm{T}_{RM}(b, n)$ returns "yes" for all $b \in \{1, 2, \ldots, n-1\}$.
  - $\cdot$ If n is composite then for at least $3/4$ of those bs $\mathrm{T}_{RM}(b, n)$ returns "no" (i.e. n is a composite).
- Recall that a random primality test randomly draws numbers $b \in \{1, \ldots, n-1\}$ and asks whether b is a witness to n's primality, or whether $\mathrm{T}_{RM}(b, n)$ returns "yes".
- Suppose n is a composite and that we can truly create a uniform independent sample of bs.
- Let X count the number of tests till we hit a negative result. What is the distribution of X?

- X is a geometric random variable with $p \geq 3/4$ (success $= \mathrm{T}_{RM}(b, n)$ declares n is not a prime, or returns "no").
- What is the expected number of tests we'll perform before we get a negative one?
- $E(X) = 1/p = 4/3$.
- What is the probability that we will fail in our first 40 tests?

$$\underbrace{(1-p)\cdot(1-p)\ldots(1-p)}_{\text{40 times}} \leq \left(\frac{1}{4}\right)^{40} \sim 10^{-24}.$$

Contention Resolution

- One server, n unsaturable processes (the service can be bandwidth for example).
- Only one process can access the server at any round.
- If two or more processes try to gain access at the same time none gets it.
- How to share the resources without a central controller or inter-communication?
- Randomization is at the core of the "symmetry-breaking" protocol.
- At each round each process randomly tries to gain access with probability p independently of anything else.
- Let A_{it} be the event: the ith process attempts to access the server at round t.
- What is $\Pr(A_{it})$?
- p.
- What is the probability that the ith process will succeed in that attempt?

- Let S_{it} be the that event: $S_{it} = A_{it} \cap (\cap_{j \neq i} \bar{A}_{jt})$.
- By the independence,

$$\Pr(S_{it}) = \Pr(A_{it}) \prod_{j \neq i} \Pr(\bar{A}_{jt}) = p(1-p)^{n-1}.$$

- How can we maximize $\alpha = \Pr(S_{it})$?
- Consider $f(p) = p(1-p)^{n-1}$ for $p \in (0,1)$: it has a maximum at $p = 1/n$.
- $\alpha = \frac{1}{n}(1 - \frac{1}{n})^{n-1}$ is the maximal possible value for $\Pr(S_{it})$: this will now assumed to be the choice.
- $\frac{1}{e} \leq (1 - \frac{1}{n})^{n-1} \leq \frac{1}{2}$, so

$$\frac{1}{e} \cdot \frac{1}{n} \leq \Pr(S_{it}) \leq \frac{1}{2} \cdot \frac{1}{n}.$$

How long is the average wait?

- Let X_i denote the first round that i gains access to the server.
- What is the distribution of X_i?
- Geometric with $p = \Pr(S_{it}) = \frac{1}{n}(1 - \frac{1}{n})^{n-1}$.
- Since $\frac{1}{e} \leq (1 - \frac{1}{n})^{n-1} \leq \frac{1}{2}$, the expected waiting time for service, $E(X_i) = 1/p$, satisfies:

$$2n \leq E(X_i) \leq en.$$

- Compare that with an optimal strategy of round robin (requires a controller) where the expected waiting time is roughly $n/2$.

Average exhaustive service time

- What is the average waiting time for *all* the processes to be serviced?
- Let Y be the time (= number of rounds) it took for servicing all the processes.
- Let's order the processes according to their service time.
  - · Let Y_1 be the time (round) the first process was serviced.
  - · Let Y_2 be the *additional* time it took for the second process to be serviced.
  - · Note that the "second process was service" is not the same as the "second time a process gained access to the server" (why?).
  - · More generally, let Y_k be the time it took between the first servicing of the $k-1$st and the kth processes.
  - · What is the connection between Y and $Y_1, Y_2, \ldots, Y_n$?
  - · $Y = \sum_1^n Y_k$
  - · What is the distribution of Y_1?

· Geometric with

$$p_1 = \Pr(\cup_{i=1}^{n} S_{it}) = n \Pr(S_{it}) = n\frac{1}{n}(1-\frac{1}{n})^{n-1}.$$

· What is the distribution of Y_2?

· Geometric with

$$p_2 = \Pr(\cup_{n=2}^{n} S_{it}) = (n-1)\frac{1}{n}(1-\frac{1}{n})^{n-1}.$$

· What is the distribution of Y_k?

· Geometric with

$$p_k = \Pr(\cup_{i=k}^{n} S_{it}) = (n-k+1)\frac{1}{n}(1-\frac{1}{n})^{n-1}.$$

•

$$\Rightarrow E(Y) = \sum_{k=1}^{n} \frac{n}{(1-\frac{1}{n})^{n-1}} \frac{1}{n-k+1}$$
$$= \frac{n}{(1-\frac{1}{n})^{n-1}} \sum_{j=1}^{n} \frac{1}{j},$$

since $E(Y) = \sum_k E(Y_k)$, and $E(Y_k) = \frac{1}{p_k}$.

• **Def.** The nth *harmonic number* is $H(n) = \sum_{j=1}^{n} \frac{1}{j}$.

• By comparing $H(n)$ to $\int \frac{1}{x}$ one can show:

$$\log(n+1) < H(n) < 1 + \log n,$$
$$\Rightarrow 2n\log(n+1) < E(Y) < en(1+\log n).$$
$$\Rightarrow E(Y) = \Theta(n \log n).$$

Distribution of service waiting time

- What is the probability that the ith process will not gain access in the first t rounds?
- Let F_{it} be that event. Then $F_{it} = \cap_{r=1}^{t} \bar{S}_{ir}$, so

$$\Pr(F_{it}) = \left[1 - \frac{1}{n}(1 - \frac{1}{n})^{n-1}\right]^t.$$

- For $t = c\lceil ne \rceil$,

$$\begin{aligned}\left[1 - \frac{1}{n}(1 - \frac{1}{n})^{n-1}\right]^t &\leq \left[1 - \frac{1}{n}\frac{1}{e}\right]^t \\ &\leq \left[1 - \frac{1}{ne}\right]^{cne} \\ &= \left[\left(1 - \frac{1}{ne}\right)^{ne}\right]^c\end{aligned}$$

Using $(1 - 1/x)^x \leq 1/e$ for $x \geq 1$

$$\leq \frac{1}{e^c}.$$

- Choosing $c = \log n$, for $t = \log n \cdot \lceil ne \rceil$:

$$\Pr(F_{it}) \leq \frac{1}{e^{\log n}} = \frac{1}{n}.$$

Distribution time of servicing 'em all

- What is the probability that servicing all the processes would take more than t rounds?
- This is $\Pr\left(\cup_{i=1}^{n} F_{it}\right)$.
- By the inclusion-exclusion formula

$$\Pr\left(\cup_{i=1}^{n} F_{it}\right) = \sum_i \Pr(F_{it})$$

$$- \sum_{i<j} \Pr(F_{it} \cap F_{jt}) + \sum_{i<j<k} \Pr(F_{it} \cap F_{jt} \cap F_{kt}) - \ldots$$

$$\Pr(F_{it}) = \left[1 - \frac{1}{n}(1 - \frac{1}{n})^{n-1}\right]^t$$

similarly,

$$\Pr(F_{it} \cap F_{jt}) = \left[1 - \frac{2}{n}(1 - \frac{1}{n})^{n-1}\right]^t$$

and more generally,

$$\Pr(F_{i_1 t} \cap F_{i_2 t} \cap \ldots F_{i_k t}) = \left[1 - \frac{k}{n}(1 - \frac{1}{n})^{n-1}\right]^t.$$

- So,

$$\Pr\left(\cup_{i=1}^{n} F_{it}\right) = \sum_k (-1)^{k-1} \binom{n}{k} \left[1 - \frac{k}{n}(1 - \frac{1}{n})^{n-1}\right]^t.$$

- This can be computed for any values of n and t. However to get an idea about how this distribution looks like it pays to concentrate only on the first term of the inclusion-exclusion formula:
- For $t = m \log n \cdot \lceil ne \rceil$:

$$
\begin{aligned}
\Pr\big(\cup_{i=1}^{n} F_{it}\big) &\leq \sum_i \Pr(F_{it}) = n \Pr(F_{it}) \\
&= n\Big[1 - \frac{1}{n}(1 - \frac{1}{n})^{n-1}\Big]^t \\
&\leq n\Big[\big(1 - \frac{1}{ne}\big)^{ne}\Big]^{m \log n} \\
&\leq n \frac{1}{e^{\log n^m}} = \frac{1}{n^{m-1}}.
\end{aligned}
$$

- For example, for $t = 3 \log n \cdot \lceil ne \rceil$,

$$
\Pr\big(\cup_{i=1}^{n} F_{it}\big) \leq \frac{1}{n^2}.
$$

- What about the terms we neglected?
- First note that what we derived is a valid upper bound. Next consider for example,

$$
\begin{aligned}
\binom{n}{2}\Big[1 - \frac{2}{n}(1 - \frac{1}{n})^{n-1}\Big]^t &\leq \binom{n}{2}\Big[\big(1 - \frac{2}{ne}\big)^{ne/2}\Big]^{6 \log n} \\
&\leq \binom{n}{2} \frac{1}{e^{\log n^6}} < \frac{1}{2n^4}.
\end{aligned}
$$

- The "higher order" terms are going to be even smaller.
- On the other hand it's not difficult to prove that for $n \geq 2$

$$n\left[1 - \frac{1}{n}(1 - \frac{1}{n})^{n-1}\right]^t > \frac{1}{n^{3.1}},$$

so the first term indeed dominates the inclusion exclusion.

Finding the median

- Given a list of numbers $S = \{a_1, a_2, \ldots, a_{2m+1}\}$ find the median: the $m + 1$st largest element (if $n = 2m$ we look for the mth largest element).
- Simple solution: sort the list and report the median.
- Cost: sorting is at least $O(n \log n)$ (number of comparisons required).
- Can we do better?
- Yes, but we need to solve a more general problem.
- The function $\texttt{Select}(S, k)$ returns the kth smallest element in S.
- For $n = 2m+1$ what are: $\texttt{Select}(S, 1)$, $\texttt{Select}(S, m)$, $\texttt{Select}(S, n)$?
- To find the minimum and maximum we clearly do not need more than n comparisons.
- It is much less obvious that this is true in general for $\texttt{Select}(S, k)$.

$\texttt{Select}(S, k)$

- On input $S = \{a_1, a_2, \ldots, a_n\}$ and k:
  - · Randomly choose a splitter or pivot $a_i \in S$.
  - · Split S into $S^- := \{a_j : a_j < a_i\}$ and $S^+ := \{a_j : a_j > a_i\}$ (requires $n - 1$ comparisons).
  - · If $|S^-| = k - 1$ return a_i.
  - · Else if $|S^-| \geq k$ return $\texttt{Select}(S^-, k)$.
  - · Else return $\texttt{Select}(S^+, k - (|S^-| + 1))$.
- Note that the algorithm is called recursively with a strictly smaller set therefore it has to stop.
- Let $T(n)$ be the running time (number of comparisons) required by `Select` for an input of size n.
- Note that $T(n)$ is a random variable.
- How big can it be?
- cn^2: if we look for the median and keep choosing a pivot which is at either ends:

$$T(n) \geq n + (n-1) + (n-2) + \cdots + n/2.$$

- But we have to be extremely unfortunate for this to happen.

Average of $T(n)$

- We say the algorithm is in phase j if the size of the currently considered S is between $n(3/4)^j$ and $n(3/4)^{j+1}$.
- Let Y_j be the number of steps we spend at phase j.
- Clearly,

$$T(n) \leq \sum_{j=0}^{\lfloor \log_{3/4} n \rfloor} Y_j \cdot n(3/4)^j.$$

Therfore,

$$E[T(n)] \leq \sum_{j=0}^{\lfloor \log_{3/4} n \rfloor} n(3/4)^j \cdot E(Y_j).$$

- Choosing any number which is not in the first or last quadrants would leave us with both S^- and S^+ smaller than 3/4 the size of the current S thereby ending phase j.
- Thus, $E(Y_j) \leq \frac{1}{1/2} = 2$ and it follows that

$$E[T(n)] \leq 2n \sum_{j=0}^{\lfloor \log_{3/4} n \rfloor} (3/4)^j < 8n.$$

Logic

- Logic is a tool for formalizing reasoning.
- We want to be able to systematically analyze arguments like
  - Borogroves are mimsy whenever it is brillig.
  - It is now brillig and this thing is a borogrove.
  - Hence this thing is mimsy.
- Is this a valid conclusion?
- Is the following a valid argument: given that
  - All lions are fierce.
  - Some lions do not drink coffee.
- Can we conclude that some fierce creatures do not drink coffee?

Proposition Logic

- To formalize the reasoning process, we need to restrict the kinds of things we can say.
- Propositional logic is particularly restrictive.
- A proposition is a statement that is either true or false but not both.
- The *syntax* of propositional logic tells us what are legitimate formulas.
- We start with *primitive* or *atomic* propositions. Those are determined to be true or false from the context. For example,
  - · Washington D.C. is the capital of USA.
  - · $1 + 1 = 2$.
  - · 4 is odd.
  - · The empty set has 0 elements.
  - · Read this carefully - *not* a proposition.
- We can then form *compound* propositions using connectives like:

$$\neg : \text{ not} \qquad \wedge : \text{ and} \qquad \vee : \text{ or}$$

$$\rightarrow : \text{ implies} \qquad \longleftrightarrow : \text{ equivalent (if and only if)}$$

Negation operator (not)

- **Def.** Given a proposition p, the negation of p, denoted by $\neg p$ (read: "not p") is true if and only if p is false.
- Intuitively, $\neg p$ is the statement: "It is not the case that p".
- Example: if p =4 is odd, then $\neg p$ is the proposition "It is not the case that 4 is odd", or 4 is not odd.
  - Aside: Note that this does not necessarily imply that 4 is even unless we have more information such as: "every number is either odd or even" and that "4 is a number".
- Mathematically we can define the negation operator through its truth table:

p	$\neg p$
T	F
F	T

Conjunction

- **Def.** For propositions p and q, $p \wedge q$ ("p and q", "conjunction") is true if and only if both p and q are true.
- Example: the proposition ((1 + 1 = 2) $\wedge$ (Toronto is the capital of Canada) is true if and only if both propositions are true.
- The truth table of the conjunction operator is:

p	q	$p \wedge q$
T	T	T
T	F	F
F	T	F
F	F	F

Disjunction, and the "exclusive or"

- **Def.** For propositions p and q, $p \vee q$ ("p or q", "disjunuction") is false if and only if both p and q are false.
- The truth table of the disjuction operator is:

p	q	$p \vee q$
T	T	T
T	F	T
F	T	T
F	F	F

- Note that in English p or q might mean:
  - exclusive or, as in "Soup or salad comes with an entrée", or
  - inclusive or, as in "The prerequisites for this course are: Math100 or CS100".
- The logical or (disjunction) is inclusive, but we do have the exclusive or, $\oplus$, as well:

p	q	$p \oplus q$
T	T	F
T	F	T
F	T	T
F	F	F

Our first claim

Claim.

$$p \oplus q \text{ is equivalent to } (p \wedge \neg q) \vee (\neg p \wedge q).$$

Proof. Via truth tables:

p	q	$p \oplus q$
T	T	F
T	F	T
F	T	T
F	F	F

while

p	q	$\neg p$	$\neg q$	$p \wedge \neg q$	$q \wedge \neg p$	$(p \wedge \neg q) \vee (\neg p \wedge q)$
T	T	F	F	F	F	F
T	F	F	T	T	F	T
F	T	T	F	F	T	T
F	F	T	T	F	F	F