$\mathrm{Var}(cX)$

Claim: $\mathrm{Var}(cX) = c^2\mathrm{Var}(X)$

Proof:

$$\begin{aligned}
& \mathrm{Var}(cX) \\
= \; & E((cX)^2) - (E(cX))^2 \\
= \; & c^2(E(X^2) - c^2(E(X))^2 \\
= \; & c^2\mathrm{Var}(X)
\end{aligned}$$

Independent Random Variables

Claim: If X and Y are independent, $E(XY) = E(X)E(Y)$ and $\mathrm{Cov}(X,Y) = 0$.

Proof:

$$\begin{array}{rl}
& E(XY) \\
= & \Sigma_s (XY)(s) \Pr(s) \\
= & \Sigma_x \Sigma_y \Sigma_{\{s: X(s)=x\,,\,Y(s)=y\}} X(s) \cdot Y(s) \cdot \Pr(s) \\
= & \Sigma_x \Sigma_y \Sigma_{\{s: X(s)=x\,,\,Y(s)=y\}} x \cdot y \cdot \Pr(s) \\
= & \Sigma_x \Sigma_y x \cdot y \cdot \Pr(X = x \cap Y = y) \\
= & \Sigma_x \Sigma_y x \cdot y \cdot \Pr(X = x) \cdot \Pr(Y = y) \quad \text{[independence]} \\
= & \Sigma_x x \cdot \Pr(X = x) \Sigma_y y \cdot \Pr(Y = y) \\
= & E(X) \cdot E(Y).
\end{array}$$

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = 0$$

Corollary: If X and Y are independent

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

The variance of $B_{n,p}$

Corollary: If $X_1, \ldots, X_n$ are mutually independent then

$\mathrm{Var}(X_1+X_2+\ldots+X_n) = \mathrm{Var}(X_1)+\mathrm{Var}(X_2)+\ldots+\mathrm{Var}(X_n)$.

Proof: By induction. One subtlety: need to show that $X_1 + \ldots + X_{k-1}$ is independent of X_k.

Let X be a $B_{n,p}$ random variable. Then $X = \Sigma_{k=1}^{n} X_k$ where X_k are independent Bernoulli p random variables. So

$$\mathrm{Var}(X) = \mathrm{Var}(\Sigma_{k=1}^{n} X_k) = \Sigma_{k=1}^{n} \mathrm{Var}(X_k) = np(1-p).$$

- For a fixed p the variance increases with n.
- For a fixed n the variance is minimized for $p = 0, 1$ and maximized for $p = 1/2$.
    - ○ Note $p(1-p) \leq 1/4$ (by calculus)

Expectation and variance are two ways of compactly describing a distribution.

- They don't completely describe the distribution
- But they're still useful!

Markov's Inequality

Theorem: Suppose X is a nonnegative random variable and $\alpha > 0$. Then

$$\Pr(X \geq \alpha) \leq \frac{E(X)}{\alpha}.$$

Proof:

$$\begin{aligned} E(X) &= \Sigma_x x \cdot \Pr(X = x) \\ &\geq \Sigma_{x \geq \alpha} x \cdot \Pr(X = x) \\ &\geq \Sigma_{x \geq \alpha} \alpha \cdot \Pr(X = x) \\ &= \alpha \Sigma_{x \geq \alpha} \Pr(X = x) \\ &= \alpha \cdot \Pr(X \geq \alpha) \end{aligned}$$

Example: If X is $B_{100,1/2}$, then

$$\Pr(X \geq 100) \leq \frac{50}{100}.$$

This is not a particularly useful estimate. In fact, $\Pr(X \geq 100) = 2^{-100} \sim 10^{-30}$.

Chebyshev's Inequality

Theorem: If X is a random variable and $\beta > 0$, then

$$\Pr(|X - E(X)| \geq \beta) \leq \frac{\mathrm{Var}(X)}{\beta^2}.$$

Proof: Let $Y = (X - E(X))^2$. Then

$$|X - E(X)| \geq \beta \text{ iff } Y \geq \beta^2.$$

I.e.,

$$\{s : |X(s) - E(X)| \geq \beta\} = \{s : Y(s) \geq \beta^2\}.$$

In particular, the probabilities of these events are the same:

$$\Pr(|X - E(X)| \geq \beta) = \Pr(Y \geq \beta^2).$$

Since $Y \geq 0$ by Markov's inequality

$$\Pr(Y \geq \beta^2) \leq \frac{E(Y)}{\beta^2}.$$

Finally, note that $E(Y) = E[(X - E(X))^2] = \mathrm{Var}(X)$.

- Equivalent statement: $\Pr(|X - E(X)| \geq \beta\sigma_X) \leq \frac{1}{\beta^2}$.
- Intuitively, the probability of a random variable being k standard deviations from the mean is $\leq 1/k^2$.

Chebyshev's Inequality: Example

Chebyshev's inequality gives a lower bound on how well is X concentrated about its mean.

- Suppose X is $B_{100,1/2}$ and we want a lower bound on $\Pr(40 < X < 60)$.
- $E(X) = 50$ and

  $$40 < X < 60 \text{ iff } |X - 50| < 10$$

  so

  $$\begin{aligned} \Pr(40 < X < 60) &= \Pr(|X - 50| < 10) \\ &= 1 - \Pr(|X - 50| \geq 10). \end{aligned}$$

  Now

  $$\begin{aligned} \Pr(|X - 50| \geq 10) &\leq \tfrac{\mathrm{Var}(X)}{10^2} \\ &= \tfrac{100\cdot(1/2)^2}{100} \\ &= \tfrac{1}{4}. \end{aligned}$$

  So

  $$\Pr(40 < X < 60) \geq 1 - \frac{1}{4} = \frac{3}{4}.$$

This is not too bad: the correct answer is ~ 0.9611 (will calculate this using Central Limit Theorem).

The law of large numbers (LLN)

You suspect the coin you are betting on is biased. You would like to get an idea on the probability that it lands heads. How would you do that?

- Obvious answer: toss it n times and estimate p as $|\#H|/n$

Underlying assumption: as n grows bigger, the sample mean is a better and better approximation for the expected value.

- Is there a mathematical justification for this intuition?

LLN: Formal Statement

Theorem (Law of Large Numbers): Consider a sequence of n Bernoulli trials $X_1, \ldots, X_n$ with the same (but unknown) success probability p. Let $\overline{p_n} = (\Sigma_{k=1}^{n} X_k)/n$. Then for all $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr(|\overline{p_n} - p|) < \epsilon) = 1.$$

Proof: Let $Y_{n,p} = (\Sigma_{k=1} X_k)/n$.

- $E(Y_{n,p}) = p$
- $\mathrm{Var}(B_{n,p}/n) = \mathrm{Var}(B_{n,p})/n^2 = p(1-p)/n$

Chebyshev's Inequality says that

$$\Pr(|Y_{n,p} - E(Y_{n,p})| \geq \epsilon) \leq \frac{Var(Y_{n,p})}{\epsilon^2} = \frac{p(1-p)}{n\epsilon^2}.$$

So

$$\begin{array}{l} \lim_{n\to\infty} \Pr(|Y_{n,p} - p)| \geq \epsilon) = 0 \\ \lim_{n\to\infty} \Pr(|Y_{n,p} - p| < \epsilon) = 1 \end{array}$$

- $Y_{n,p} = \overline{p}$: the sample mean is a random variable

LLN can be generalized:

- Applies to arbitrary *iid* random variables:
  - ○ *independent* and *identically distributed*
  - ○ E.g., could be sequence of Poisson variables

Continuous Distributions

Suppose you wanted to describe the uniform distribution on the domain $[0,1] = \{x : 0 \leq x \leq 1\}$.

For all $x \in [0,1]$, the probability of choosing x is 0. So how can you describe this probability distribution:

- Using cumulative distribution:

$$F(x) = \Pr(X \leq x) = x$$

- Using a density function $f(x)$ such that

$$\textstyle\int_{-\infty}^{x} f(z)dz = F(x).$$

The Normal Distribution

The *normal distribution* is described by the density function

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

- It's symmetric around $y = 0$

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1$$

- $\frac{1}{\sqrt{2\pi}}$ is a normalization factor to make the integral 1.

The normal distribution is the famous "bell curve".

The Central Limit Theorem

The normal distribution = limit of normalized binomials.

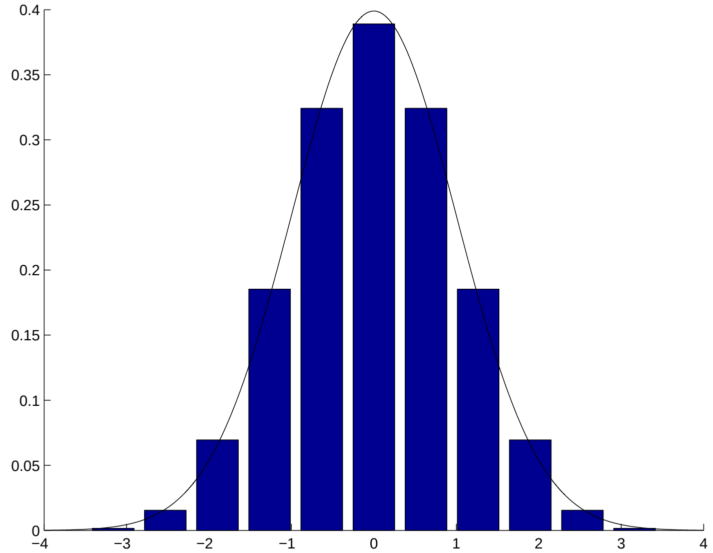
- Let $X_1, \ldots, X_n$ be iid Bernoulli with mean p
- Let $Y_{n,p} = (X_1 + \cdots + X_n)/n$. Recall
  - ○ $E(Y_{n,p}) = p$
  - ○ $\mathrm{Var}(Y_{n,p}) = p(1-p)/n$, so $\sigma_{Y_{n,p}} = \sqrt{p(1-p)/n}$
- Let $Z_{n,p} = (Y_{n,p} - p)/\sqrt{p(1-p)/n}$
- $Z_{n,p}$ is a "normalized binomial"
  - ○ $E(Z_{n,p}) = 0$; $\sigma_{Z_{n,p}} = 1$

Theorem (Central Limit Theorem): If N is the normal distribution, then for all p with $0 < p < 1$,

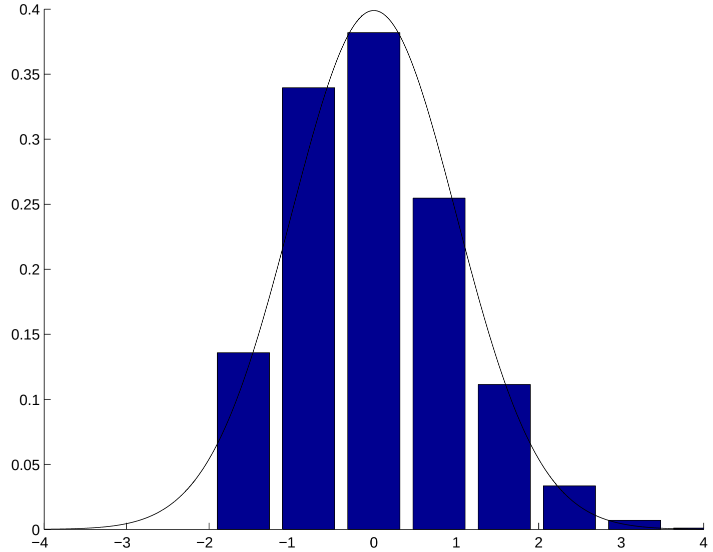
$$\lim_{n \to \infty} \Pr(c \leq Z_{n,p} \leq d) = \Pr(c \leq N \leq d).$$

Some Pictures

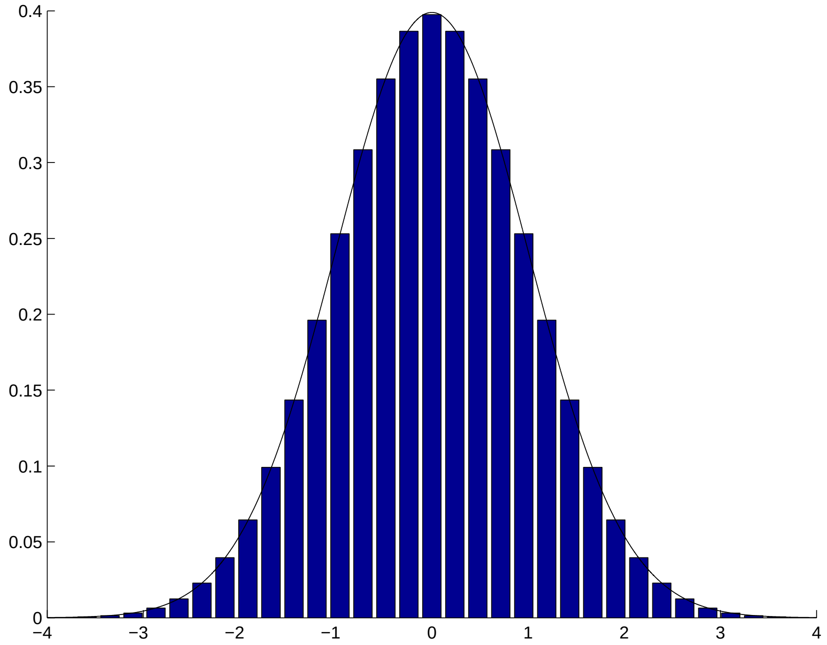
$n = 10$, $p = 0.5$:



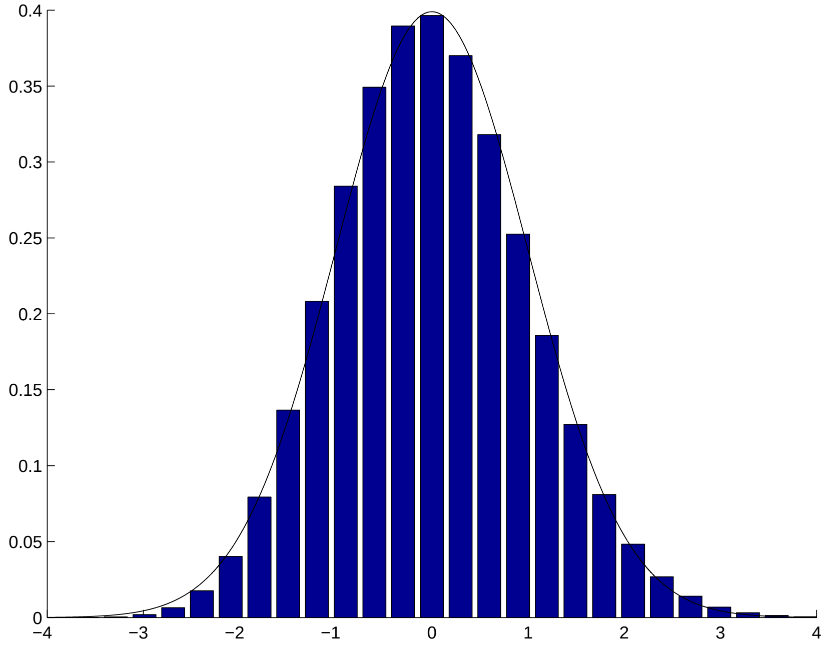
$n = 10$, $p = 0.2$:



$n = 70$, $p = 0.5$:



$n = 70$, $p = 0.2$:



CLT: Examples

Example 1: A fair die is rolled 600 times. What is the probability of getting 1 between 90 and 110 times.

Let $X_{600,1/6}$ be the random variable that describes the number of 1's in 600 tosses.

- $E(X_{600,1/6}/600) = 1/6$; $\sigma_{X_{600,1/6}/600} = \sqrt{(1/6)(5/6)/600}$
- By the CLT,
$$Z = \frac{X_{600,1/6}/600 - 1/6}{\sqrt{(1/6)(5/6)/600}} = \sqrt{\frac{6}{5}}\left(\frac{X_{600,1/6} - 100}{10}\right)$$
is approximately normally distributed
- $\Pr(90 \leq X_{600,1/6} \leq 110) = \Pr(-\sqrt{6/5} \leq Z \leq \sqrt{6/5})$
- $\sqrt{6/5} \approx 1.095$
- Table (DAM3, p. 581) says $\Pr(N \leq 1.09) = .8621$ and $\Pr(N \leq 1.10) = .8643$
  - ○ Split the difference; take $\Pr(N \leq 1.095) \approx .8632$

$$\begin{aligned}
& \Pr(-1.095 \leq N \leq 1.095) \\
= {} & \Pr(N \leq 1.095) - \Pr(N \leq -1.095) \\
= {} & \Pr(N \leq 1.095) - \Pr(N > 1.095) \quad \text{[by symmetry]} \\
= {} & .8632 - (1 - .8632) = .7264
\end{aligned}$$

Bottom line: the probability of getting 1 between 90 and 110 times is about .7264.

Polling

Example 2: 100 people are chosen at random and asked if they prefer B or K; 62 say K. What is the probability that between 52% and 72% actually support K?

Let X be the random variable that gives the number of 100 people that support K.

- In each state s, a different sample of 100 is chosen. $X(s)$ is # supporting K in the sample chosen in s.

X is distributed as $B_{p,100}$, where p is the actual fraction that support K. Define

$$Z = \frac{\frac{X}{100} - p}{\sqrt{p(1-p)/100}} = \frac{10}{\sqrt{p(1-p)}}\left(\frac{X}{100} - p\right)$$

Z is approximately normally distributed.

$$\Pr(|X/100 - p| \leq .1) = \Pr(|Z| \leq 1/\sqrt{p(1-p)})$$

Problem: we don't know p.

- But a little calculus shows $p(1-p) \leq 1/4$, so

$$\begin{aligned} & \Pr(|Z| \leq 1/\sqrt{p(1-p)}) \\ \geq\ & \Pr(|Z| \leq 2) \\ =\ & \Pr(Z \leq 2) - \Pr(Z \leq -2) \\ =\ & \Pr(Z \leq 2) - \Pr(Z > 2) \\ =\ & .9772 - (1 - .9772) \approx .954 \end{aligned}$$

Bottom line: With probability $> .95$, the sample mean is within .1 of the true mean, if the sample size is 100.

Example 3: How many people have to be polled to ensure that the probability that the sample mean is within .03 of the true mean is greater than .95?

- I.e., want to be almost certain that the error is ±3%.

Let X_n be sample mean (fraction of people who say K) in sample of size n. Define

$$Z = \frac{X_n/n - p}{\sqrt{p(1-p)/n}}$$

$$\begin{array}{rl} & \Pr(|X_n/n - p| \leq .03) \\ = & \Pr(|Z| \leq .03/\sqrt{p(1-p)/n}) \\ \geq & \Pr(|Z| \leq (.03)2\sqrt{n} \quad [\text{since } p(1-p) \leq 1/4] \end{array}$$

Want to choose n so that $\Pr(|Z| \leq .06\sqrt{n}) \geq .95$

- From table: $n = 1067$

Bottom line: No matter what the total population, a *random* sample of size 1067 gives you an error of ±3% with very high confidence.

- How do you know your sample is random?
- Telephone samples miss people with no telephone, people with weird hours.

CS Applications of Probability: Primality Testing

Recall idea of primality testing:

- Choose b between 1 and n at random
- Apply an easily computable (deterministic) test $T(b, n)$ such that
  - ○ $T(b, n) = 1$ (for all b) if n is prime.
  - ○ There are lots of b's for which $T(b, n) = 0$ if n is not prime.
    - ∗ In fact, for the standard test T, for at least 1/3 of the b's between 1 and n, $T(b, n) = 0$ if n is composite

So here's the algorithm:

```
Input $n$          [number whose primality is to be checked]
Output Prime              [Want Prime $= 1$ iff $n$ is prime]
Algorithm Primality
    for $k$ from 1 to 100 do
        Choose $b$ at random between 1 and $n$
        If $T(b, n) = 0$ return Prime $= 0$
    endfor
    return Prime $= 1$.
```

Probabilistic Primality Testing: Analysis

If n is composite, what is the probability that algorithm returns Prime = 1?

- $(2/3)^{100} < (.2)^{25} \approx 10^{-70}$
- I wouldn't lose sleep over mistakes!
- if 10^{-70} is unacceptable, try 200 random choices.

How long will it take until we find a witness

- Expected number of steps is ≤ 3

What is the probability that it takes k steps to find a witness?

- $(2/3)^{k-1}(1/3)$ (this is the geometric distribution)

Bottom line: the algorithm is extremely fast and almost certainly gives the right results.

An Average-Case Analysis

Remember this algorithm?

```
Input $n$                              [$n > 1$; number of items]
      $x_1, \ldots, x_n$                              [Items in set]
Output $m$                                      [Maximum value]
Algorithm MaxNumber
   $m \leftarrow x_1$
   for $k$ from $2$ to $n$ do
      if $x_k > m$ then $m \leftarrow x_k$
   endfor
```

How many times is m assigned a new value?

Let Y be the number of times is m assigned a new value

- Y is a random variable
- For each state (permutation) Y gives # assignments.

Let $X_k = 1$ if m is assigned in kth iteration; 0 otherwise

- $X_k = 1$ if $x_k > x_1, \ldots, x_{k-1}$
- $\Pr(X_k = 1) = 1/k$
- $Y = X_1 + \cdots + X_n$
- $E(Y) = \Sigma_{k=1}^{n} \frac{1}{k}$
- By calculus: $\log(n) - 1 < E(Y) < \log(n)$