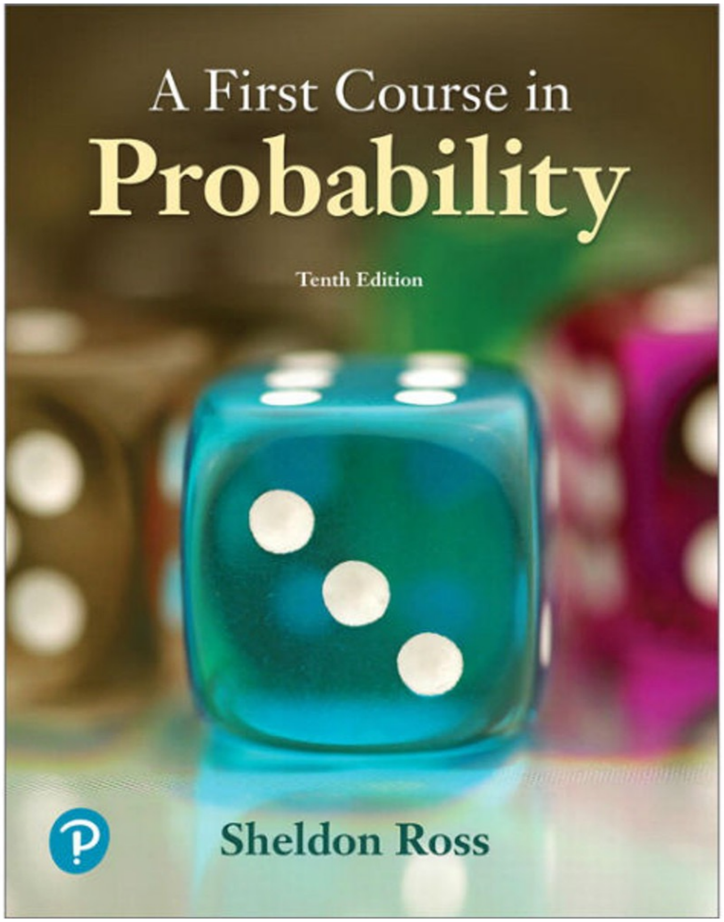
Lecture 12.2

INTERLUDE: Limit Theorems

PURDUE UNIVERSITY | Department of Mathematics



Today’s reading: 8.1-8.4

Next class: 6.3

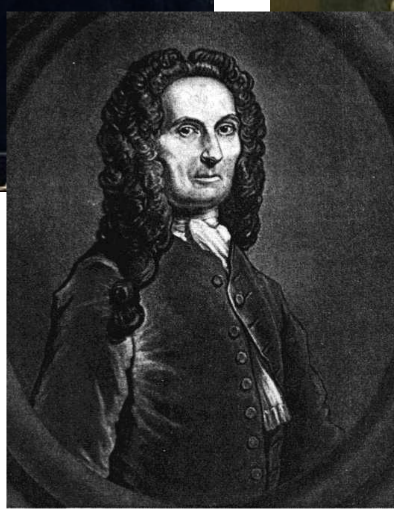
HW10 will be made available by end of day Friday. Will be due next Friday, 4/18.

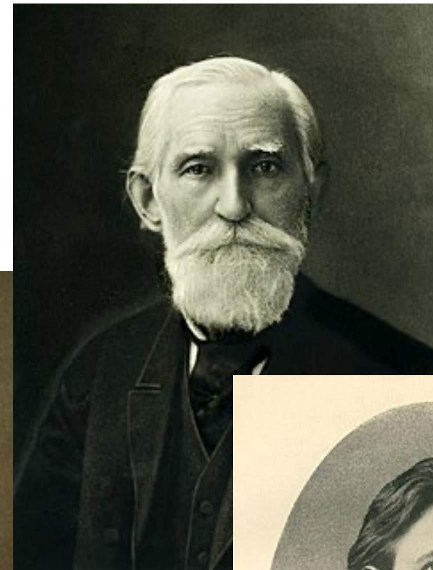
I will do my best to get graded MT2 back to you by Monday.

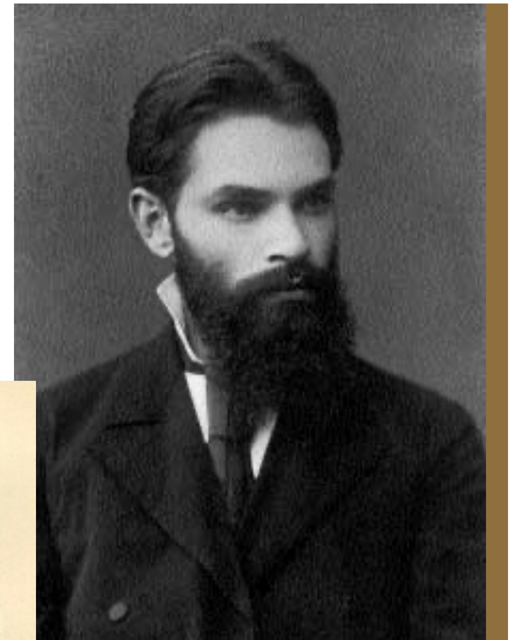
I will *definitely* get it back to you before next Friday, since that is the last day to drop (with W).

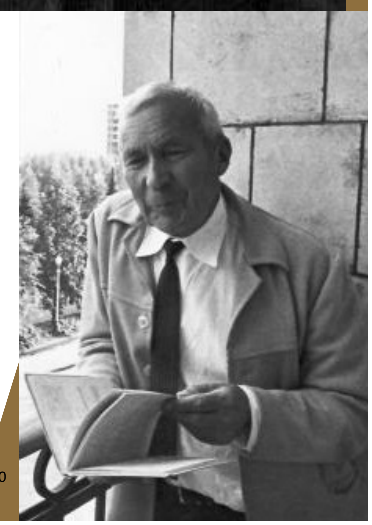
Remember: starting this Friday (4/11), we will have TWO draft problems every day, an “A” problem and a “B” problem. I will announce the A draftee lists and B draftee lists now.

Today we're going to peak ahead in the book because limit theorems are just too cool not to talk about as soon as possible. Moreover, there are some tools involved in proving them that are extremely useful in many settings.









Markov's inequality

Proposition 2.1

If X is a non-negative ($\mathbb{R}$-valued) random variable, then for any $a > 0$,

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

Why care? Markov's inequality shows that for ***ANY*** non-negative random variable, we can upper bound the probability of a "tail" event without needing to know *anything* except the mean of the random variable! (Of course, this bound will be quite bad for most specific choices of X**, but the usefulness of Markov's inequality comes from its universality**.)

Chebyshev's inequality

Proposition 2.2

If X is any ($\mathbb{R}$-valued) random variable with mean $\mu < \infty$ and variance σ^2, then for any $\epsilon > 0$

$$P\{|X - \mu| \geq \epsilon\} \leq \frac{\sigma^2}{\epsilon^2}$$

Why care? For similar reason as Markov's inequality. Chebyshev's inequality shows that for ***ANY*** random variable, we can upper bound the probability that it is "far away" from its mean without needing to know *anything* except the mean and the variance of the random variable! (Of course, this bound will be quite bad for most specific choices of X**, but the usefulness of Chebyshev's inequality comes from its universality**.)

A cute application of Chebyshev's inequality

Proposition 2.3

The only random variables with variance 0 are constant.
In other words: if $\mathrm{Var}(X) = 0$, then

$$P\{X = E[X]\} = 1$$

A more important application of Chebyshev's inequality: Weak law of large numbers

Theorem 2.1

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables ("i.i.d.") with $\mu = E[X_1] = E[X_2] = \cdots$. Then for any $\epsilon > 0$

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right\} \to 0 \qquad \text{as} \quad n \to \infty$$

Intuition: the probability that a *sample mean* is "far away" (that is, farther than epsilon away) from the *true mean* goes to 0 as we take more samples.

(One downside: this theorem doesn't tell us anything about how large n needs to be in order to guarantee the probability is small. With more effort, one can get such control.)

More is true! (Although we won't prove it)

Theorem 4.1 (Strong Law of Large Numbers):

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables ("i.i.d.") with $\mu = E[X_1] = E[X_2] = \cdots$. Then with probability 1

$$P\left\{\lim_{n\to\infty}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \mu\right\} = 1$$

Intuition: as we take more and more samples, the sample mean "almost certainly" converges to the true mean.

One of the "crown jewels" of probability theory

Theorem 3.1 (Central Limit Theorem)

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables ("i.i.d.") with mean $\mu = E[X_1] = E[X_2] = \cdots$ and variance $\sigma^2 = Var(X_1) = Var(X_2) = \cdots$. Then for every real number a

$$\lim_{n\to\infty} P\left\{\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq a\right\} = \Phi(a),$$

where $\Phi(a)$ is the CDF of the standard normal Z.

Intuition: if we "normalize" a sum of more and more i.i.d. random variables, then in the limit, the normalized CDF becomes. Put another way: if we "properly account" for the variance, then in the limit of taking more and more samples, the difference between the sample mean and the true mean is the standard normal bell curve.

Worth noting (if this means anything to you): while not stated this way, we can in fact show that we get ***uniform*** convergence of the CDFs.

A technical result needed to prove CLT

Lemma 3.2

Let $Z_1, Z_2, \ldots$ be a sequence of random variables having distribution functions F_{Z_n} and moment generating functions M_{Z_n}, $n \geq 1$, and let Z be a random variable having distribution function F_Z and moment generating function F_Z. If

$$\lim_{n\to\infty} M_{Z_n}(t) = M_Z(t)$$

for all t, then

$$\lim_{n\to\infty} F_{Z_n}(t) = F_Z(t)$$

for all t at which $F_Z(t)$ is continuous.

Intuition: if the "moment generating functions" of a sequence of random variables converges to the moment generating function of some other random variable, then the CDFs converge to the CDF.

Note: we'll talk about "moment generating functions" in a few weeks. After that, we'll circle back to proving CLT using this lemma (we won't prove this lemma).

Order in Apparent Chaos.—I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the "Law of Frequency of Error." The law would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement amidst the wildest confusion. The huger the mob, and the greater the apparent anarchy, the more perfect is its sway. It is the supreme law of Unreason. Whenever a large sample of chaotic elements are taken in hand and marshalled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to have been latent all along. The tops of the marshalled row form a flowing curve of invariable proportions; and each element, as it is sorted into place, finds, as it were, a pre-ordained niche, accurately adapted to fit it. If the measurement at any two specified Grades in the row are known, those that will be found at every other Grade, except towards the extreme ends, can be predicted in the way already explained, and with much precision.

Francis Galton, *Natural Inheritance* (1901)

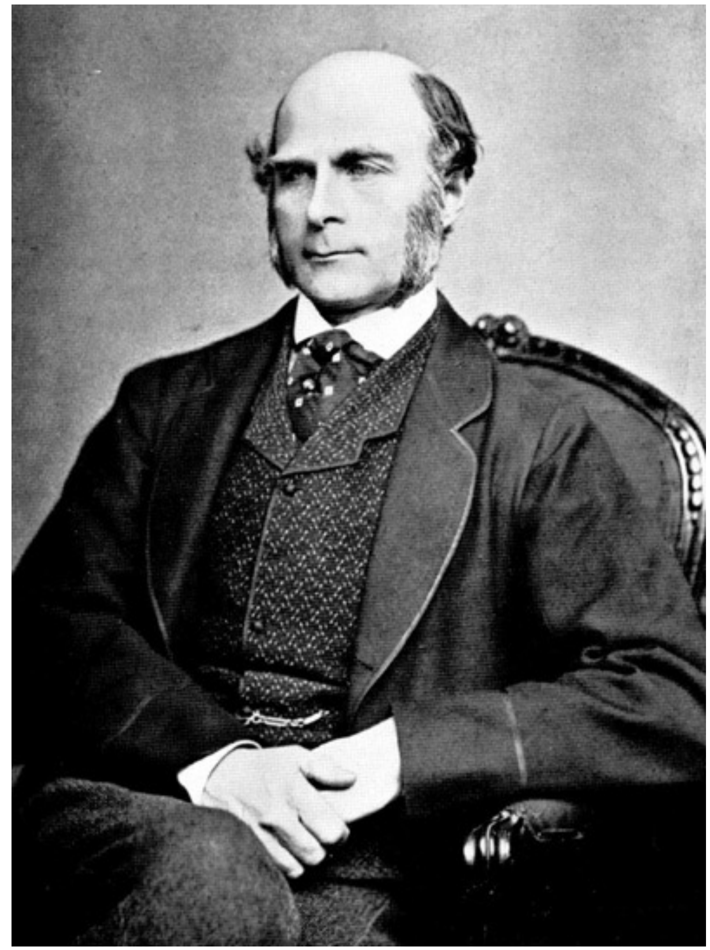
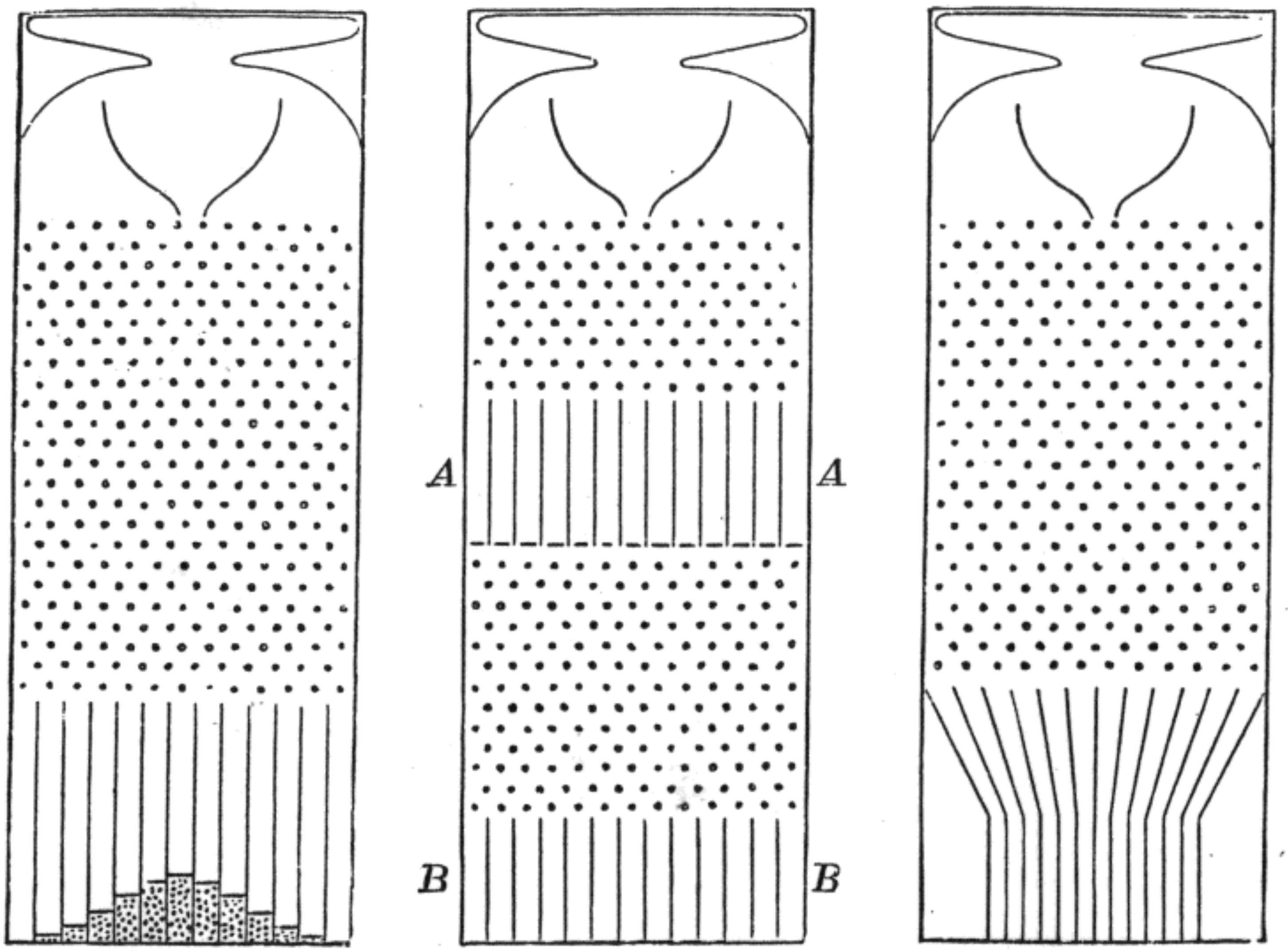
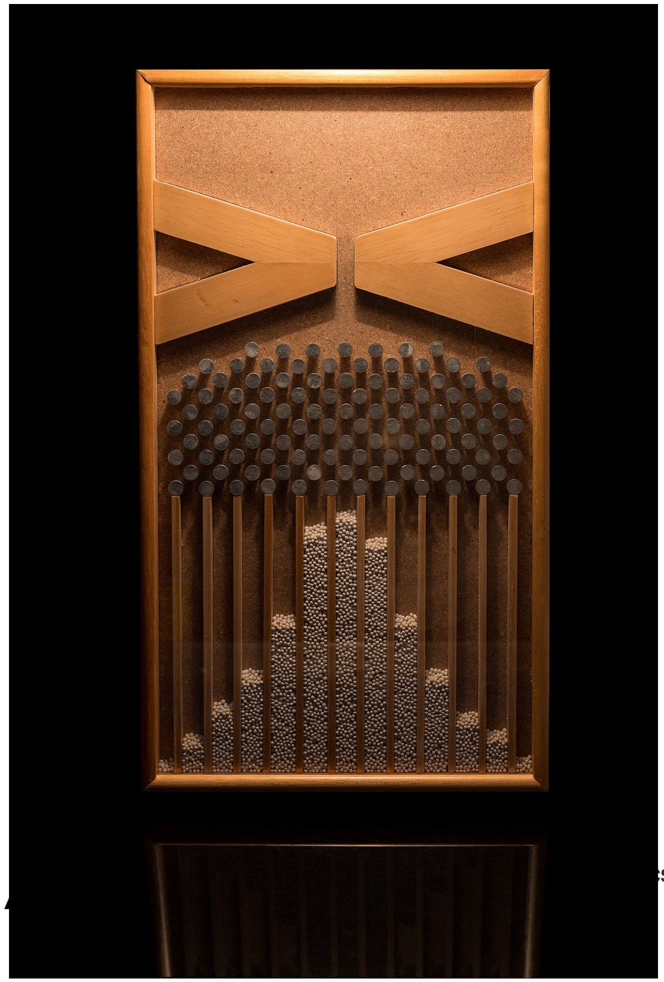


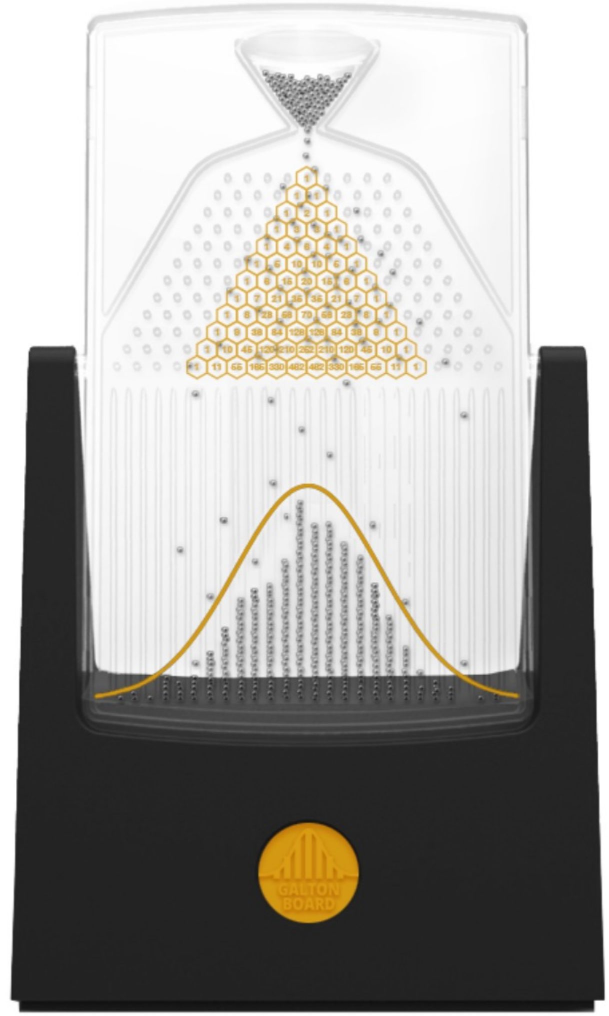
FIG.7. FIG.8. FIG.9.



By Mark Hebner - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=67111322

By Matemateca (IME USP) / (name of the photographer), CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=57045935





Friday's draft problem A

To be presented by Friday's A draftee.

Use Chebyshev's inequality to answer the following:

How many flips n of a biased coin with unknown probability of heads p does it take in order to be 90% certain that the ratio

$$\frac{\text{number of heads observed}}{n}$$

agrees with p to two decimal places?

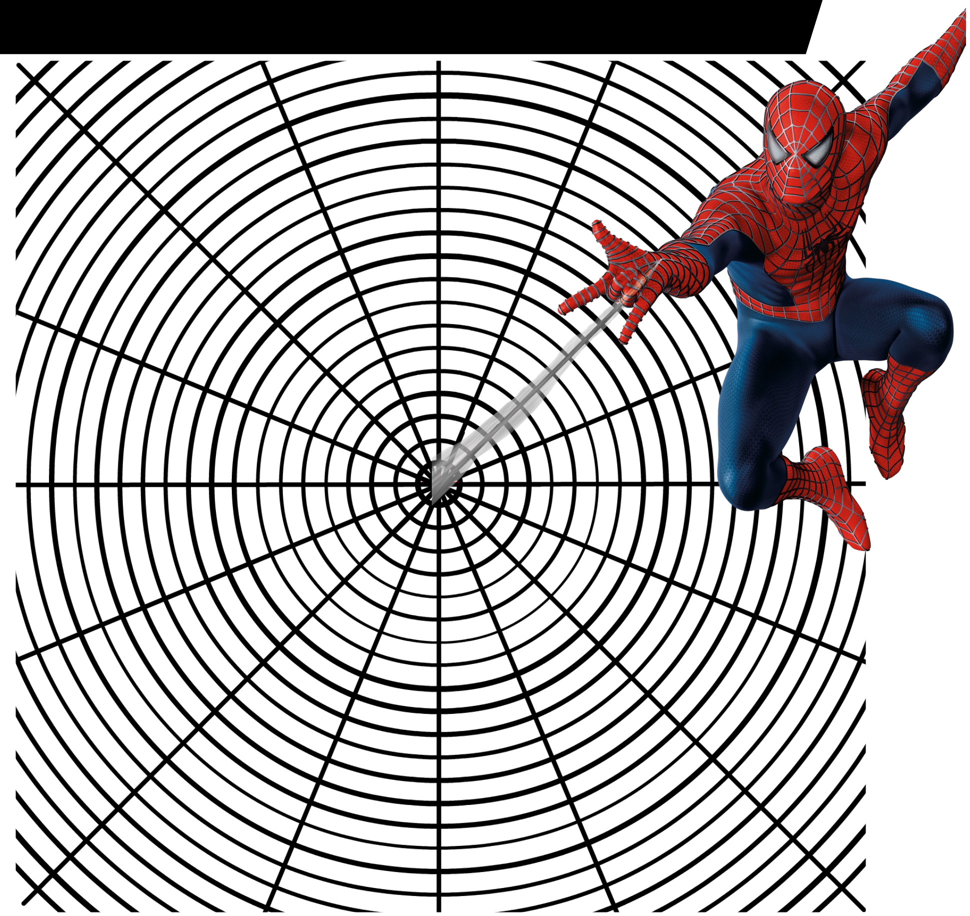
Put another way: if $0 \leq p \leq 1$ and X_n denotes a binomial random variable with parameters n and p where p is unknown, how large does n need to be in order to guarantee that

$$P\left\{\left|\frac{X_n}{n} - p\right| \geq 0.01\right\} \leq 0.1$$

Hint: $\left|\frac{X_n}{n} - p\right| \geq 0.01$ if and only if $|X_n - \mu_n| \geq 0.01n$, where $\mu_n = np = E[X_n]$. Now combine Chebyshev's inequality with the fact that $\sigma_n^2 = Var(X_n) \leq 0.25n$ (the key point is that this is independent of p).

Friday's draft problem B

To be presented by Friday's B draftee.



Consider two independent random variables X_1 and X_2 such that X_1 follows an exponential distribution with mean 2, and X_2 follows a uniform distribution on the interval $[0,2\pi]$. Let $Y_1 = \sqrt{X_1}\cos X_2$ and $Y_2 = \sqrt{X_1}\sin X_2$.

1. Show that Y_1 and Y_2 are independent.
2. Show that Y_1 and Y_2 are both standard normal variables.