## What are limit theorems?

- Limit theorems specify the probabilistic behavior of n random variables as $n \to \infty$
- Possible restrictions on RVs:
  - Independent random variables
  - Uncorrelated random variables
  - Have identical marginal CDFs/pdfs/pmfs
  - Have identical means and/or variances

  

## The average of n RVs

- n random variables $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ have finite expectations $\mu_1, \mu_2, \ldots, \mu_n$
- Let $\mathbf{Z} = (\mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_n)/n$
- What is $E[\mathbf{Z}]$?
- Expectation is a linear operator
- $E[\mathbf{Z}] = (E[\mathbf{X}_1] + E[\mathbf{X}_2] + \ldots + E[\mathbf{X}_n])/n$
- Expected value of average of n RVs = numerical average of their expectations
- If $E[\mathbf{X}_i] = \mu$ for all i, then $E[\mathbf{Z}] = \mu$ also

  

## The sample mean

- Model: An experiment is repeated n times
- $\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_n$ are the n observed values of a random variable $\mathbf{X}$ on the n independent trials of the experiment
- Random variable $\mathbf{X}$ has finite mean $\mu$
- The $\mathbf{X}_i$'s are said to be independent identically distributed (i.i.d. or iid) random variables
- $\mathbf{Z} = (\mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_n)/n$ is called the sample mean

  

## Variance of the sample mean

- i.i.d. RVs $\mathbf{X}_i$ with finite mean and variance
- Sample mean $\mathbf{Z} = (\mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_n)/n$
- $E[\mathbf{Z}] = E[\mathbf{X}] = \mu$
- $\mathrm{var}(\mathbf{Z}) = n^{-2} \bullet \mathrm{var}(\mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_n)$
  $= n^{-2} \bullet [\mathrm{var}(\mathbf{X}_1) + \mathrm{var}(\mathbf{X}_2) + \ldots + \mathrm{var}(\mathbf{X}_n)]$
  $= n^{-1} \bullet \mathrm{var}(\mathbf{X})$
- This is because the RVs are independent
- Hence, $\mathrm{cov}(\mathbf{X}_i, \mathbf{X}_j) = 0$ if $i \neq j$
- Also holds if the RVs are uncorrelated

  

## Variance decreases as n increases

- $\mathbf{Z}$ is the average of the n observed values of a random variable $\mathbf{X}$ with mean $\mu$ and variance $\mathrm{var}(\mathbf{X})$
- $E[\mathbf{Z}] = \mu$ and $\mathrm{var}(\mathbf{Z}) = n^{-1} \bullet \mathrm{var}(\mathbf{X})$
- If we wish to estimate the value of $\mu$, then the value of the sample mean $\mathbf{Z}$ is a much better estimator than the value of any individual observation $\mathbf{X}_i$
- Application to experimental results

  

## Confidence interval for mean

- $E[\mathbf{Z}] = \mu$ and $\mathrm{var}(\mathbf{Z}) = n^{-1} \bullet \mathrm{var}(\mathbf{X}) = n^{-1} \bullet \sigma^2$
- Assume $\mathrm{var}(\mathbf{X}) = \sigma^2$ is known
- Chebyshev inequality:
  $P\{|\mathbf{X} - \mu| \geq a\} \leq (\sigma/a)^2$
- $\mathbf{X} \pm 5\sigma$ is a 96% confidence interval for $\mu$
- $P\{|\mathbf{Z} - \mu| \geq a\} \leq a^{-2} n^{-1} \bullet \mathrm{var}(\mathbf{X}) = (\sigma/a\sqrt{n})^2$
- $\mathbf{Z} \pm 5\sigma/\sqrt{n}$ is 96% confidence interval for $\mu$
- Much smaller confidence interval with sample mean than with any individual $\mathbf{X}_i$

  

## Normality increases confidence!!

- Assume the $\mathbf{X}_i$ are Gaussian RVs
- For any confidence level, we get a much smaller confidence interval
- If the $\mathbf{X}_i$ are Gaussian RVs, then so is $\mathbf{Z}$
- $E[\mathbf{Z}] = \mu$ and $\text{var}(\mathbf{Z}) = n^{-1} \bullet \text{var}(\mathbf{X}) = n^{-1} \bullet \sigma^2$
- $\Phi(1.96) = 0.9750\ldots$
- $\mathbf{Z} \pm 1.96\sigma/\sqrt{n}$ is a 95% confidence interval for $\mu$
- $\mathbf{Z} \pm 5\sigma/\sqrt{n}$ is a 99.99997…% confidence interval for $\mu$



## The sample variance

- Sample mean $\mathbf{Z} = (\mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_n)/n$ is RV that is a function of the $\mathbf{X}_i$'s
- i-th deviation is $\mathbf{X}_i - \mathbf{Z}$. $\text{cov}(\mathbf{X}_i - \mathbf{Z}, \mathbf{Z}) = 0$
- Sample variance $\mathbf{S}^2 = (n-1)^{-1} \bullet \sum(\mathbf{X}_i - \mathbf{Z})^2$
- $\mathbf{S}^2$ is also a RV
- Distinguish between the sample variance $\mathbf{S}^2$ and the variance of the sample mean
- $\mathbf{S}^2$ is a random variable: the variance of the sample mean is the number $\text{var}(\mathbf{X})/n$



## Expectation of the sample variance

- $\mathbf{S}^2 = (n-1)^{-1} \bullet \sum(\mathbf{X}_i - \mathbf{Z})^2$
- $E[\mathbf{S}^2] = (n-1)^{-1} \bullet \sum E[(\mathbf{X}_i - \mathbf{Z})^2]$
- But, $\mathbf{X}_i - \mathbf{Z}$ is a zero-mean random variable (why?) and hence we have that
  $E[\mathbf{S}^2] = (n-1)^{-1} \bullet \sum \text{var}(\mathbf{X}_i - \mathbf{Z})$
  $= (n-1)^{-1} \bullet \sum \text{var}(\mathbf{X}_i) + \text{var}(\mathbf{Z}) - 2 \bullet \text{cov}(\mathbf{X}_i, \mathbf{Z})$
  $= (n-1)^{-1} \bullet \sum \text{var}(\mathbf{X}) + \text{var}(\mathbf{X})/n - 2 \bullet \text{var}(\mathbf{X})/n$
  since $\text{cov}(\mathbf{X}_i, n^{-1}\sum \mathbf{X}_j) = n^{-1} \bullet \text{cov}(\mathbf{X}_i, \mathbf{X}_i)$
- $E[\mathbf{S}^2] = (n-1)^{-1} \bullet \sum(1-1/n)\text{var}(\mathbf{X}) = \text{var}(\mathbf{X})$



## Variance of the sample variance $\mathbf{S}^2$

- If $\text{var}(\mathbf{X})$ is unknown, it can be estimated as $\mathbf{S}^2 = (n-1)^{-1} \bullet \sum(\mathbf{X}_i - \mathbf{Z})^2$
- $E[\mathbf{S}^2] = \text{var}(\mathbf{X})$, the unknown quantity
- The variance of the random variable $\mathbf{S}^2$ is
  $n^{-1} \bullet \{E[(\mathbf{X}-\mu)^4] - (n-3)(\text{var}(\mathbf{X}))^2/(n-1)\}$
  and also decreases as n increases
- Hence, we can use the Chebyshev inequality to find confidence intervals for the unknown variance $\text{var}(\mathbf{X})$



## Gaussian samples — I

- If the $\mathbf{X}_i$ are i.i.d. $\mathbf{N}(\mu,\sigma^2)$ RVs, then
  - $\mathbf{Z} = (\mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_n)/n$ is $\mathbf{N}(\mu,\sigma^2/n)$
  - If $\sigma^2$ is known, then $\mathbf{Z} \pm 1.96\sigma/\sqrt{n}$ is a 95% confidence interval for $\mu$
  - $\mathbf{S}^2 = (n-1)^{-1}\sum(\mathbf{X}_i - \mathbf{Z})^2$ is a gamma RV with parameters $((n-1)/2,(n-1)/2\sigma^2)$ with mean $\sigma^2$ and variance $2\sigma^4/(n-1)$
  - $\mathbf{Z}$ and $\mathbf{S}^2$ are independent RVs
  - Each deviation $\mathbf{X}_i - \mathbf{Z}$ is independent of the sample mean $\mathbf{Z}$



## Gaussian samples — II

- If the $\mathbf{X}_i$ are i.i.d. $\mathbf{N}(\mu,\sigma^2)$ RVs, then
  - $\mathbf{S}^2 = (n-1)^{-1}\sum(\mathbf{X}_i - \mathbf{Z})^2$ is a gamma RV with mean $\sigma^2$ and variance $2\sigma^4/(n-1)$
  - $(n-1)\mathbf{S}^2/\sigma^2 = \sigma^{-2} \bullet \sum(\mathbf{X}_i - \mathbf{Z})^2$ is a $\chi^2$ RV with n–1 degrees of freedom $= \Gamma((n-1)/2,1/2)$ RV with mean n–1
  - The value of the sample variance $\mathbf{S}^2$ is a good estimate of $\sigma^2$
  - Confidence intervals and levels can be set using known gamma pdf

## Gaussian samples — III

- If the $\mathbf{X}_i$ are i.i.d. $\mathbf{N}(\mu,\sigma^2)$ RVs, then
  - $\mathbf{S} = [(n-1)^{-1}\sum(\mathbf{X}_i - \mathbf{Z})^2]^{1/2}$
  - $(\mathbf{Z} - \mu)/(\mathbf{S}\sqrt{n})$ has Student's t distribution with n–1 degrees of freedom
  - pdf is $C \cdot [1 + u^2/(n-1)]^{n/2}$, $-\infty < u < \infty$ where $C = \Gamma(n/2)/\Gamma((n-1)/2)((n-1)\pi)^{1/2}$
  - If n > 2, mean 0 and variance n/(n–2)
  - When $\sigma^2$ is unknown, confidence levels and intervals for mean μ are set using the t distribution



## Weak Law of Large Numbers — I

- Weak Law of Large Numbers:

  If $\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_n, \ldots$ are i.i.d. RVs with finite mean μ, then for every ε > 0,

$P\{|(\mathbf{X}_1+\mathbf{X}_2+\ldots+\mathbf{X}_n)/n - \mu| \geq \varepsilon\} \to 0$ as $n \to \infty$

Equivalently

$P\{|(\mathbf{X}_1+\mathbf{X}_2+\ldots+\mathbf{X}_n)/n - \mu| \leq \varepsilon\} \to 1$ as $n \to \infty$

- Note that it is not necessary for the RVs to have finite variance
- But the proof is easier if variance is finite



## Weak Law of Large Numbers — II

- Weaker Law of Large Numbers:

  $\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_n, \ldots$ are i.i.d. with finite mean μ and finite variance $\sigma^2$. For every ε > 0,

$P\{|(\mathbf{X}_1+\mathbf{X}_2+\ldots+\mathbf{X}_n)/n - \mu| \leq \varepsilon\} \to 1$ as $n \to \infty$

- $E[(\mathbf{X}_1+\mathbf{X}_2+\ldots+\mathbf{X}_n)/n] = \mu$
- $\mathrm{var}((\mathbf{X}_1+\mathbf{X}_2+\ldots+\mathbf{X}_n)/n) = \sigma^2/n$
- Chebyshev inequality gives

  $P\{|(\mathbf{X}_1+\mathbf{X}_2+\ldots+\mathbf{X}_n)/n - \mu| \leq \varepsilon\} \geq 1 - (\sigma/\varepsilon)^2/n$

  which converges to 1 as $n \to \infty$



## Weak Law of Large Numbers — III

- Even More Weaker Law of Large Numbers

  $\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_n, \ldots$ are uncorrelated random variables with finite mean μ and finite variance $\sigma^2$. For every ε > 0,

$P\{|(\mathbf{X}_1+\mathbf{X}_2+\ldots+\mathbf{X}_n)/n - \mu| \leq \varepsilon\} \to 1$ as $n \to \infty$

- Weak law applies because the key result that $\mathrm{var}((\mathbf{X}_1+\mathbf{X}_2+\ldots+\mathbf{X}_n)/n) = \sigma^2/n$ requires only that the RVs be uncorrelated: it is not necessary that the RVs be independent



## Weak Law of Large Numbers — IV

- Weak Law of Large Numbers cannot be used if $E[\mathbf{X}_i]$ is not finite or is not defined
- Suppose that the $\mathbf{X}_i$ are i.i.d. Cauchy RVs
- What is the pdf of $(\mathbf{X}_1+\mathbf{X}_2+\ldots+\mathbf{X}_n)/n$?
- Characteristic function of $\mathbf{X}_i$ is $\exp(-|\omega|)$
- Characteristic function of $\mathbf{X}_1+\mathbf{X}_2+\ldots+\mathbf{X}_n$ is $\exp(-n|\omega|)$
- $(\mathbf{X}_1+\mathbf{X}_2+\ldots+\mathbf{X}_n)/n$ is also a Cauchy RV
- $P\{|(\mathbf{X}_1+\mathbf{X}_2+\ldots+\mathbf{X}_n)/n| \leq \varepsilon\}$ is same for all n!



## Strong Law of Large Numbers — I

- Strong Law of Large Numbers:

  If $\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_n, \ldots$ are i.i.d. RVs with finite mean μ, then

  $P\{\lim_{n\to\infty}(\mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_n)/n = \mu\} = 1$
- What's going on?
- Isn't that just you said the Weak Law of Large Numbers is?
- No: WLLN says $\lim_{n\to\infty} P\{\text{something}\} = 1$
- SLLN says $P\{\lim_{n\to\infty} \text{something}\} = 1$

### Strong Law of Large Numbers — II

- Strong Law of Large Numbers:

  If $\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_n, \ldots$ are i.i.d. RVs with finite mean μ, then

  $P\{\lim_{n\to\infty}(\mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_n)/n = \mu\} = 1$
- Experiment will be repeated infinitely often
- We observe that the RV **X** took on values $x_1, x_2, \ldots x_n, \ldots$ on these trials
- What can be said about $(x_1+x_2+\ldots+x_n)/n$?
- We think this should be approximately μ



### Strong Law of Large Numbers — III

- Experiment will be repeated infinitely often
- We observe that the RV **X** took on values $x_1, x_2, \ldots x_n, \ldots$ on these trials
- What can be said about the sequence whose n-th term is $(x_1+x_2+\ldots+x_n)/n$?
- There are three possibilities
  - Sequence converges to μ
  - or it converges to some other number
  - or it does not converge at all



### Strong Law of Large Numbers — IV

- The sequence whose n-th term is $(x_1+x_2+\ldots+x_n)/n$ either
  - converges to μ
  - or it converges to some other number
  - or it does not converge at all
- The Strong Law of Large Numbers says that $P\{(x_1+x_2+\ldots+x_n)/n$ converges to $\mu\} = 1$
- $P\{(x_1+x_2+\ldots+x_n)/n$ does not converge at all or converges to some other number$\} = 0$



### Strong Law of Large Numbers — V

- If the Strong Law of Large Numbers holds, then so does the Weak Law
- In fact, both require only that the RVs be i.i.d. with finite mean μ
- But, the Weak Law of Large Numbers might be applicable in cases when the Strong Law does not hold
- Example: Weak Law of Large Numbers still applies if the RVs are uncorrelated but not independent



### What's a limit, anyway?

- A sequence $t_1, t_2, \ldots t_n, \ldots$ converges to the limit L if for every choice of ε > 0, we can find an integer N such that

  $|t_n - L| < \varepsilon$ for all $n \ge N$
- Put another way, for each choice of ε, there can only be finitely many integers n for which the inequality $|t_n - L| < \varepsilon$ does not hold
- In fact, there are obviously fewer than N such integers



### Limit of a sequence of RVs

- Let $\mathbf{Z}_n = (\mathbf{X}_1+\mathbf{X}_2+\ldots+\mathbf{X}_n)/n$ where the $\mathbf{X}_i$ are i.i.d. RVs with finite mean μ
- The value of $\mathbf{Z}_n$ is approximately μ
- For each fixed choice of ε, there will be some (possibly infinitely many) values of n for which $\mathbf{Z}_n$ is close to μ, and the event $|\mathbf{Z}_n - \mu| < \varepsilon$ will have occurred
- For other (possibly infinitely many) values of n, the event $|\mathbf{Z}_n - \mu| > \varepsilon$ will have occurred

## What the Strong Law means

- Let $\mathbf{Z}_n = (\mathbf{X}_1+\mathbf{X}_2+\ldots+\mathbf{X}_n)/n$ where the $\mathbf{X}_i$ are i.i.d. RVs with finite mean $\mu$
- For each fixed choice of $\varepsilon$,
  - P{event $|\mathbf{Z}_n - \mu| < \varepsilon$ occurs for infinitely many choices of n} = 1
  - P{event $|\mathbf{Z}_n - \mu| > \varepsilon$ occurs for finitely many choices of n} = 1
  - P{event $|\mathbf{Z}_n - \mu| > \varepsilon$ occurs for infinitely many choices of n} = 0




## Strong Law versus Weak Law

- The Strong Law of Large Numbers says
  - P{event $|\mathbf{Z}_n - \mu| > \varepsilon$ occurs for finitely many choices of n} = 1
  - P{event $|\mathbf{Z}_n - \mu| > \varepsilon$ occurs for infinitely many choices of n} = 0
- The Weak Law does not require that the event $|\mathbf{Z}_n - \mu| > \varepsilon$ occurs only for finitely many choices of n — the event might occur infinitely often: just so long as

  $P\{|\mathbf{Z}_n - \mu| > \varepsilon\} \rightarrow 0$ as $n \rightarrow \infty$




## Strong Law and relative frequencies

- The Strong Law of Large Numbers justifies the estimation of probabilities in terms of relative frequencies
- If the $\mathbf{X}_i$ are i.i.d. Bernoulli RVs with parameter p (and hence, finite mean p), then the sample mean $\mathbf{Z}_n$ converges to p with probability 1 as $n \rightarrow \infty$
- The observed relative frequency of an event of probability p converges to p with probability 1




## The Central Limit Theorem — I

- The Strong Law of Large Numbers says that the sample mean $\mathbf{Z}_n$ converges to $\mu$ with probability 1
- The CDF of $\mathbf{Z}_n$ converges to a unit step function with $0 \rightarrow 1$ transition at the point $\mu$
- Suppose the RVs $\mathbf{X}_i$ have variance $\sigma^2$
- Then, $E[\mathbf{Z}_n] = \mu$, $\text{var}(\mathbf{Z}_n) = \sigma^2/n$
- $\mathbf{Y}_n = \sqrt{n}\bullet(\mathbf{Z}_n - \mu)/\sigma$ is a RV with mean 0 and variance 1
- What can we say about the CDF of $\mathbf{Y}_n$ ?




## The Central Limit Theorem — II

- $\mathbf{Y}_n = \sqrt{n}\bullet(\mathbf{Z}_n - \mu)/\sigma$

  $= (\mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_n - n\mu)/\sigma\sqrt{n}$

  is a RV with mean 0 and variance 1
- The Central Limit Theorem asserts for large values of n that the CDF of $\mathbf{Y}_n$ is well-approximated by the unit Gaussian CDF $\Phi(\bullet)$
- Formally, the Central Limit Theorem states that the CDF converges to $\Phi(\bullet)$




## The Central Limit Theorem — III

- Central Limit Theorem: Given i.i.d. RVs $\mathbf{X}_i$ with finite mean $\mu$ and finite variance $\sigma^2$, the CDF of the RV

  $(\mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_n - n\mu)/\sigma\sqrt{n}$

  converges to the unit Gaussian CDF $\Phi(\bullet)$, that is, for each choice of real number u,

  $P\{(\mathbf{X}_1+\ldots+\mathbf{X}_n-n\mu)/\sigma\sqrt{n} \le u\} \rightarrow \Phi(u)$ as $n \rightarrow \infty$
- The proof uses the moment-generating function (or characteristic function) of the $\mathbf{X}_i$ and will be omitted

## The Central Limit Theorem — IV

- The Central Limit Theorem does not claim that
  $Y_n = (X_1 + X_2 + \ldots + X_n - n\mu)/\sigma\sqrt{n}$
  is a unit Gaussian random variable
- All the Central Limit Theorem says is that $P\{Y_n \le u\}$ is approximately $\Phi(u)$ when n is large
- Thus, we have a computational tool for probabilities involving the sample mean



## The Central Limit Theorem — V

- In practical use of the Central Limit Theorem. we hardly ever use the RV
  $Y_n = (X_1 + X_2 + \ldots + X_n - n\mu)/\sigma\sqrt{n}$
- Instead, $X_1 + X_2 + \ldots + X_n$ is treated as if its CDF is approximately that of a $\mathbf{N}(n\mu, n\sigma^2)$ RV
- Thus, we compute
  $P\{X_1 + X_2 + \ldots + X_n \le u\} \approx \Phi((u-n\mu)/\sigma\sqrt{n})$
  which is effectively the same computation



## Use and Abuse of the CLT — I

- $P\{X_1 + X_2 + \ldots + X_n \le u\} \approx \Phi((u-n\mu)/\sigma\sqrt{n})$ gives a good approximation when $u \ge \mu$
- More generally,
  $P\{a \le X_1 + X_2 + \ldots + X_n \le b\}$
  $\approx \Phi((b-n\mu)/\sigma\sqrt{n}) - \Phi((a-n\mu)/\sigma\sqrt{n})$
  is a good approximation if $a < \mu < b$
- It is a very poor approximation if $a >> \mu$ or $b << \mu$
- Slow convergence in the tails of the CDF



## Use and Abuse of the CLT — II

- The RV $X_1 + X_2 + \ldots + X_n$ should not be treated as a Gaussian RV with all the rights and privileges appertaining thereto
- Example: $X_1 + X_2 + \ldots + X_n$
  and $X_{n+1} + X_{n+2} + \ldots + X_{2n}$
  are independent RVs, and each has a CDF that is approximately Gaussian, but their joint CDF is not jointly Gaussian CDF
- Such false premises lead to false conclusions



## How to Use the CLT

- $P\{a \le X_1 + X_2 + \ldots + X_n \le b\}$
  $\approx \Phi((b-n\mu)/\sigma\sqrt{n}) - \Phi((a-n\mu)/\sigma\sqrt{n})$
  is a good approximation if $a < \mu < b$
- It is a very poor approximation if $a >> \mu$ or $b << \mu$
- We saw this already for the case of binomial RVs with the DeMoivre-LaPlace theorem
- There are many versions of the CLT



## Summary — I

- We learned about the properties of sample means and sample variances
- We learned about the pdfs of the sample mean and sample variances for Gaussian samples
- Weak Law of Large Numbers:
  If $X_1, X_2, \ldots X_n, \ldots$ are i.i.d. RVs with finite mean $\mu$, then for every $\varepsilon > 0$,

$P\{|(X_1+X_2+\ldots+X_n)/n - \mu| \le \varepsilon\} \to 1$ as $n \to \infty$

## Summary — II

- Strong Law of Large Numbers: If $\mathbf{X}_1$, $\mathbf{X}_2$, … $\mathbf{X}_n$, … are i.i.d. RVs with finite mean μ,

  $P\{\lim_{n\to\infty}(\mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_n)/n = \mu\} = 1$
- If the Strong Law holds, then so does the Weak Law but the Weak Law may hold in cases where the Strong Law does not
- We attempted to understand the difference in what the two Laws are saying
- SLLN justifies estimation of probabilities in terms of relative frequencies



## Summary — III

- Central Limit Theorem: Given i.i.d. RVs $\mathbf{X}_i$ with finite mean μ and finite variance $\sigma^2$, the CDF of the RV

  $(\mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_n - n\mu)/\sigma\sqrt{n}$

  converges to the unit Gaussian CDF $\Phi(\bullet)$
- $P\{a \le \mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_n \le b\}$

  $\approx \Phi((b-n\mu)/\sigma\sqrt{n}) - \Phi((a-n\mu)/\sigma\sqrt{n})$

  is a good approximation if $a < \mu < b$ and a bad approximation if $a >> \mu$ or $b << \mu$



## Final Remarks

"We see that the theory of probability is at bottom only common sense reduced to calculation; it makes us appreciate with exactitude what reasonable minds feel by a sort of instinct, often without being able to account for it. … It is remarkable that this science, which originated in the consideration of games of chance, should become the most important object of human knowledge. … The most important questions of life are, for the most part, really only problems of probability."

Pierre Simon, Marquis de LaPlace, ***Analytical Theory of Probability***