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 \( X_1, X_2, \ldots, X_n \) have finite expectations \( \mu_1, \mu_2, \ldots, \mu_n \).
- Let \( Z = (X_1 + X_2 + \ldots + X_n)/n \).
- What is \( E[Z] \)?
  - Expectation is a linear operator
  - \( E[Z] = (E[X_1] + E[X_2] + \ldots + E[X_n])/n \)
  - Expected value of average of \( n \) RVs = numerical average of their expectations
  - If \( E[X_i] = \mu \) for all \( i \), then \( E[Z] = \mu \) also.

The sample mean

- Model: An experiment is repeated \( n \) times
- \( X_1, X_2, \ldots, X_n \) are the \( n \) observed values of a random variable \( X \) on the \( n \) independent trials of the experiment
- Random variable \( X \) has finite mean \( \mu \)
- The \( X_i \)'s are said to be independent identically distributed (i.i.d. or iid) random variables
- \( Z = (X_1 + X_2 + \ldots + X_n)/n \) is called the sample mean

Variance of the sample mean

- i.i.d. RVs \( X_i \) with finite mean and variance
- Sample mean \( Z = (X_1 + X_2 + \ldots + X_n)/n \)
- \( E[Z] = E[X] = \mu \)
- \( \text{var}(Z) = n^{-2} \cdot \text{var}(X_1 + X_2 + \ldots + X_n) = n^{-2} \cdot \text{var}(X_1) + \text{var}(X_2) + \ldots + \text{var}(X_n) = n^{-1} \cdot \text{var}(X) \)
- This is because the RVs are independent
- Hence, \( \text{cov}(X_i, X_j) = 0 \) if \( i \neq j \)
- Also holds if the RVs are uncorrelated

Variance decreases as \( n \) increases

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

Confidence interval for mean

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

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

Gaussian samples — I

- If the $X_i$ are i.i.d. $N(\mu, \sigma^2)$ RVs, then
  - $Z = (X_1 + X_2 + \ldots + X_n)/n$ is $N(\mu, \sigma^2/n)$
  - If $\sigma^2$ is known, then $Z \pm 1.96\sigma/\sqrt{n}$ is a 95% confidence interval for $\mu$
  - If $\sigma^2$ is unknown, then $Z \pm 1.96\sigma\sqrt{1/n}$ is a 95% confidence interval for $\mu$

Gaussian samples — II

- If the $X_i$ are i.i.d. $N(\mu, \sigma^2)$ RVs, then
  - $S^2 = (n-1)^{-1} \sum (X_i - \bar{X})^2$ is a gamma RV with parameters $(n-1)/2, (n-1)/(2\sigma^2)$
  - $Z$ and $S^2$ are independent RVs
  - Each deviation $X_i - \bar{X}$ is independent of the sample mean $\bar{X}$

Expectation of the sample variance

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

The sample variance

- Sample mean $Z = (X_1 + X_2 + \ldots + X_n)/n$ is a RV that is a function of the $X_i$’s
- $i$-th deviation is $X_i - Z$. $E[\text{cov}(X_i - Z, Z)] = 0$
- Sample variance $S^2 = (n-1)^{-1} \sum (X_i - Z)^2$
- $S^2$ is also a RV
- Distinguish between the sample variance $S^2$ and the variance of the sample mean
- $S^2$ is a random variable: the variance of the sample mean is the number $\text{var}(X)/n$

Normality increases confidence!!

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

- Even More Weaker Law of Large Numbers: $X_1, X_2, \ldots, X_n, \ldots$ are uncorrelated random variables with finite mean $\mu$ and finite variance $\sigma^2$. For every $\epsilon > 0$,
  \[ P\left(\left|\frac{1}{n}(X_1 + X_2 + \ldots + X_n) - \mu\right| \leq \epsilon\right) \to 1 \text{ as } n \to \infty \]
- Weak law applies because the key result that $\text{var}(\frac{1}{n}(X_1 + X_2 + \ldots + X_n)/n) = \frac{\sigma^2}{n}$ requires only that the RVs be uncorrelated: it is not necessary that the RVs be independent.

Weak Law of Large Numbers — I

- Weak Law of Large Numbers:
  \[ P\left(\left|\frac{1}{n}(X_1 + X_2 + \ldots + X_n) - \mu\right| \leq \epsilon\right) \to 1 \text{ as } n \to \infty \]
  \[ \text{Equivalently} \]
  \[ P\left(\left|\frac{1}{n}(X_1 + X_2 + \ldots + X_n) - \mu\right| \leq \epsilon\right) \to 1 \text{ as } n \to \infty \]
- 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

- Weak Law of Large Numbers:
  \[ X_1, X_2, \ldots, X_n, \ldots \text{ are i.i.d. RVs with finite mean } \mu \text{ and finite variance } \sigma^2. \]
  \[ \text{For every } \epsilon > 0, \]
  \[ P\left(\left|\frac{1}{n}(X_1 + X_2 + \ldots + X_n) - \mu\right| \leq \epsilon\right) \to 1 \text{ as } n \to \infty \]
  \[ \text{Equivalently} \]
  \[ P\left(\left|\frac{1}{n}(X_1 + X_2 + \ldots + X_n) - \mu\right| \leq \epsilon\right) \to 1 \text{ as } n \to \infty \]
- $\text{Chebyshev inequality gives} \quad P\left(\left|\frac{1}{n}(X_1 + X_2 + \ldots + X_n) - \mu\right| \geq \epsilon\right) \leq \frac{\sigma^2}{n}$
- $\text{which converges to 1 as } n \to \infty \quad \text{for } \epsilon > 0.$
**Strong Law of Large Numbers — II**

- **Strong Law of Large Numbers:**
  - If $X_1, X_2, \ldots, X_n, \ldots$ are i.i.d. RVs with finite mean $\mu$, then
  $$\Pr\left( \lim_{n \to \infty} \frac{X_1 + X_2 + \ldots + X_n}{n} = \mu \right) = 1$$
  - 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 $\mu$

**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 $\mu$
  - or it converges to some other number
  - or it does not converge at all
- The Strong Law of Large Numbers says
  $$\Pr\left( \frac{x_1 + x_2 + \ldots + x_n}{n} \text{ converges to } \mu \right) = 1$$
  $$\Pr\left( \frac{x_1 + x_2 + \ldots + x_n}{n} \text{ does not converge at all or converges to some other number} \right) = 0$$

**Strong Law of Large Numbers — IV**

- 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 $\mu$
- 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

**Strong Law of Large Numbers — V**

- 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 $\varepsilon > 0$, we can find an integer $N$ such that
    $$|t_n - L| < \varepsilon \text{ for all } n \geq N$$
  - Put another way, for each choice of $\varepsilon$
    - 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 $Z_n = (X_1 + X_2 + \ldots + X_n)/n$ where the $X_i$ are i.i.d. RVs with finite mean $\mu$
- The value of $Z_n$ is approximately $\mu$
- For each fixed choice of $\varepsilon$, there will be some (possibly infinitely many) values of $n$ for which $Z_n$ is close to $\mu$, and the event $|Z_n - \mu| < \varepsilon$ will have occurred
- For other (possibly infinitely many) values of $n$, the event $|Z_n - \mu| > \varepsilon$ will have occurred
Let \( Z_n = \frac{X_1 + X_2 + \ldots + X_n}{n} \) where the \( X_i \) are i.i.d. RVs with finite mean \( \mu \)

- For each fixed choice of \( \varepsilon \),
  - \( P(\text{event } |Z_n - \mu| < \varepsilon \text{ occurs for infinitely many choices of } n) = 1 \)
  - \( P(\text{event } |Z_n - \mu| > \varepsilon \text{ occurs for finitely many choices of } n) = 0 \)

What the Strong Law means

The Strong Law of Large Numbers says
- \( P(\text{event } |Z_n - \mu| > \varepsilon \text{ occurs for finitely many choices of } n) = 1 \)
- \( P(\text{event } |Z_n - \mu| > \varepsilon \text{ occurs for infinitely many choices of } n) = 0 \)

Strong Law versus Weak Law

The Weak Law does not require that the event \( |Z_n - \mu| > \varepsilon \) occurs only for finitely many choices of \( n \) — the event might occur infinitely often: just so long as \( P(|Z_n - \mu| > \varepsilon) \to 0 \text{ as } n \to \infty \)

The Central Limit Theorem — I

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

The Central Limit Theorem — II

- \( Y_n = \sqrt{n}(Z_n - \mu)/\sigma = (X_1 + X_2 + \ldots + X_n - n\mu)/\sigma\sqrt{n} \) is a RV with mean 0 and variance 1
- The Central Limit Theorem asserts for large values of \( n \) that the CDF of \( Y_n \) is well-approximated by the unit Gaussian CDF \( \Phi(\cdot) \)
- Formally, the Central Limit Theorem states that the CDF converges to \( \Phi(\cdot) \)

The Central Limit Theorem — III

- Central Limit Theorem: Given i.i.d. RVs \( X_i \) with finite mean \( \mu \) and finite variance \( \sigma^2 \), the CDF of the RV \( (X_1 + X_2 + \ldots + X_n - n\mu)/\sigma\sqrt{n} \) converges to the unit Gaussian CDF \( \Phi(\cdot) \), that is, for each choice of real number \( u \), \( P((X_1 + \ldots + X_n - n\mu)/\sigma\sqrt{n} \leq u) \to \Phi(u) \) as \( n \to \infty \)
- The proof uses the moment-generating function (or characteristic function) of the \( X_i \) and will be omitted
The Central Limit Theorem — IV

- The Central Limit Theorem does not claim that
  \( Y_n = \frac{(X_1 + X_2 + \ldots + X_n - n\mu)}{\sigma\sqrt{n}} \)
  is a unit Gaussian random variable
- All the Central Limit Theorem says is that
  \( P\{Y_n \leq u\} \) 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
  \( Y_n = \frac{(X_1 + X_2 + \ldots + X_n - n\mu)}{\sigma\sqrt{n}} \)
- Instead, \( X_1 + X_2 + \ldots + X_n \) is treated as if its CDF is approximately that of a \( N(n\mu, n\sigma^2) \) RV
- Thus, we compute
  \( P\{X_1 + X_2 + \ldots + X_n \leq u\} \) \( \approx \Phi((u - n\mu)/\sigma\sqrt{n}) \)
  which is effectively the same computation

Use and Abuse of the CLT — I

- \( P\{X_1 + X_2 + \ldots + X_n \leq u\} = \Phi((u-n\mu)/\sigma\sqrt{n}) \) gives a good approximation when \( u \geq \mu \)
- More generally,
  \( P\{a \leq X_1 + X_2 + \ldots + X_n \leq b\} = \Phi((b-n\mu)/\sigma\sqrt{n}) - \Phi((a-n\mu)/\sigma\sqrt{n}) \)
  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:
  \( X_1 + X_2 + \ldots + X_m \)
  and
  \( X_{m+1} + X_{m+2} + \ldots + X_{2n} \)
  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

- \( P\{a \leq X_1 + X_2 + \ldots + X_n \leq b\} = \Phi((b-n\mu)/\sigma\sqrt{n}) - \Phi((a-n\mu)/\sigma\sqrt{n}) \)
  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 \( X_1, X_2, \ldots, X_m \) are i.i.d. RVs with finite mean \( \mu \), then for every \( \varepsilon > 0 \),
  \( P\{|(X_1+X_2+\ldots+X_m)/n - \mu| \leq \varepsilon\} \rightarrow 1 \) as \( n \rightarrow \infty \)
Strong Law of Large Numbers: If $X_1, X_2, \ldots, X_n, \ldots$ are i.i.d. RVs with finite mean $\mu$,
$$P\left(\lim_{n \to \infty} \frac{X_1 + X_2 + \ldots + X_n}{n} = \mu\right) = 1$$

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

Central Limit Theorem: Given i.i.d. RVs $X_i$
with finite mean $\mu$ and finite variance $\sigma^2$,
the CDF of the RV
$$\frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}}$$
converges to the unit Gaussian CDF $\Phi(\cdot)$

$$P(a \leq X_1 + X_2 + \ldots + X_n \leq b) = \Phi\left(\frac{b-n\mu}{\sigma\sqrt{n}}\right) - \Phi\left(\frac{a-n\mu}{\sigma\sqrt{n}}\right)$$
is a good approximation if $a < \mu < b$ and a
bad approximation if $a >> \mu$ or $b << \mu$

"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