MATH 321: Mathematical Statistics

Lecture 5: The Central Limit Theorem

Chapter 5: Distributions of Functions of Random Variables (5.6 and 5.7)

The Central Limit Theorem (CLT)

Introduction

- The sample mean is one statistic whose large-sample behavior is quite important. In particular, we want to investigate its limiting distribution. This is summarized in one of the most important in statistics, the central limit theorem (CLT).
- In MATH 320, we introduced the CLT and thought about it as a sum of random variables.
- **Central Limit Theorem**: Let $X_1, \ldots, X_n$ be independent random variables, all of which have the same probability distribution and thus the same mean μ and variance σ^2. If n is large, the sum
$$S = X_1 + X_2 + \cdots + X_n$$
will be approximately normal with mean $n\mu$ and variance $n\sigma^2$.
- Written succinctly: If $X_i \overset{iid}{\sim} f(x)$ with mean μ and variance σ^2, then
$$S = \sum_{i=1}^{n} X_i \overset{approx}{\sim} \text{Normal}\left(n\mu, n\sigma^2\right) \quad \text{if } n \text{ is large}$$
- We used it to solve problems like this for example:

  Suppose the number of claims filed on for a particular policy follow a Poisson distribution with a mean of 2 claims per year and the company has a portfolio of 500 active policies this year, which are assumed to be independent.

  Find the distribution of the total number of filed claims for the entire portfolio.

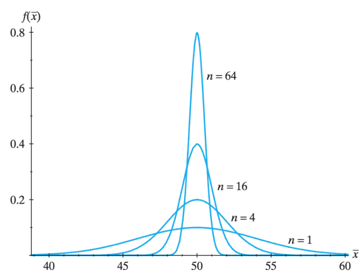
CLT - Different perspective

- Now we will think about the CLT from a convergence (in distribution) point of view.
- Build up to CLT / convergence in distribution.

  Let $X_i \stackrel{iid}{\sim} \text{Normal}\,(\mu, \sigma^2)$ and let $\bar{X}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$.

  1. For a fixed n:

  2. As $n \to \infty$:

     pdfs of means of samples from $N(50, 16)$

     The variance decreases until all probability is a single point.

  3. Suppose we "center the distribution" with $\bar{X}_n - \mu$:

     Even with the location adjustment, the variance of $\bar{X}_n$ disappears when the sample size n increases (without bound).

  4. We want to stop (or slow down) the "decay" of the variance, or we can think about this as spreading out the probability, so that when $n \to \infty$, $\bar{X}_n$ (and $\bar{X}_n - \mu$) does not converge to a constant, but rather distribution that still has some variation.

     To do this, we multiply the quantity of interest by a factor of n:

     Now this result doesn't converge to a constant because the variance doesn't depend on n and remains "in tact" when $n \to \infty$.

5. Then, we standardize the variance (adjusting the scale) by dividing by σ:

6. Lastly, it turns out that regardless of the distribution of X_i (so we are dropping the normal assumption), this result is always true!

   This is the CLT.

- Convergence in distribution
  - Definition: A sequence of random variables, $Y_1, Y_2, \ldots$, **converges in distribution** to a random variable Y if

    $$\lim_{n\to\infty} F_{Y_n}(y) = F_Y(y)$$

    at all points y where $F_Y(y)$ is continuous (notation: $Y_n \xrightarrow{d} Y$).

  - Although we talk of a sequence of random variables converging in distribution, it is really **the cdfs that converge, not the random variable (or statistic)**.

    So for the CLT, we technically have:

    $$\lim_{n\to\infty} \qquad F_{Y_n}(y) \qquad = \qquad F_Y(y)$$

Restating CLT

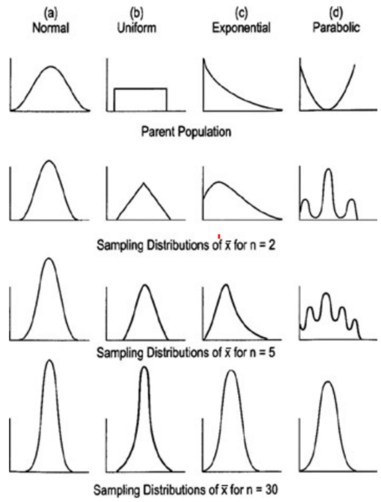
- **Theoretical result**

  Central Limit Theorem: Let $X_i \overset{iid}{\sim} f(x)$ with $E(X) = \mu$ and $V(X) = \sigma^2 > 0$. Then the distribution of

  $$W = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \text{Normal}\,(0, 1) \quad \text{as } n \to \infty$$

- **In practice**

  This means for **any** random variable X with $E(X) = \mu$ and $V(X) = \sigma^2 > 0$, as n gets larger the distribution of $W = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ can be more closely approximated by the standard normal distribution.



Distribution of sample mean, $\overline{X}$

- Using this practical result
  - Needed theorem: If $Z \sim \mathrm{N}(0,1)$, and μ and $\sigma > 0$ are constants, then

    $$X = \sigma Z + \mu \sim \mathrm{N}(\mu, \sigma^2)$$

    Proof:

  - Results

    (a) $\frac{\sigma}{\sqrt{n}} W + \mu = \bar{X}$ can be approximated by
    $\frac{\sigma}{\sqrt{n}} Z + \mu \sim \text{Normal}\,(\mu, \frac{\sigma^2}{n})$ for "large" n.

    (b) $n\bar{X} = X_1 + \ldots + X_n = S$ can be approximated by
    $(\sigma\sqrt{n})Z + n\mu \sim \text{Normal}\,(n\mu, n\sigma^2)$ for "large" n.

- How large must n be?
  - Although CLT gives us a useful general approximation, we have no automatic way of knowing how good the approximation is in general. In fact, the goodness of the approximation is a function of the original distribution, and so much be checked case by case.
  - The more the distribution of X (population distribution) is “like” a normal distribution (symmetric, unimodal, continuous, etc.), the smaller the n needed for $\bar{X}$ to be approximated well by a normal distribution.
  - **The rule $n \geq 30$ is a lie!!** But for “school” purposes, we can just use this rule of thumb as our check.
  - With the current availability of cheap, plentiful computing power, the importance of approximation like the CLT is somewhat lessened. However, despite its limitation, it is still a marvelous result.

Examples

1. Let $\bar{X}$ be the mean of a random sample of size 36 from Exp $(\lambda = 1/3)$. Approximate $P(2.5 \leq \bar{X} \leq 4)$.

2. Let $X_1, \ldots, X_{20}$ denote a random sample of size 20 from continuous Uniform $(2, 8)$. If $S = X_1 + \ldots + X_{20}$, approximate $P(S < 95)$.

t, Z, and the CLT

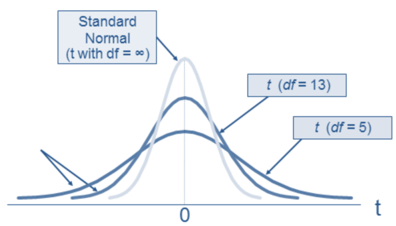
- Previously, we learned the following
  1. If $X_1, \ldots, X_n$ are a random sample for a $N(\mu, \sigma^2)$, we know that the quantity

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

  2. (Building on 1.) However, if σ is unknown, we substitute S then

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

  3. (Building on 2.) As $n \to \infty$, $t_{n-1} \to Z$



  4. If $X_1, \ldots, X_n$ are **not normal** random variables, when the sample size is large

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

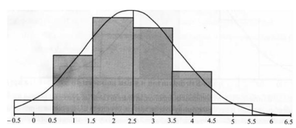
Approximations for discrete distributions

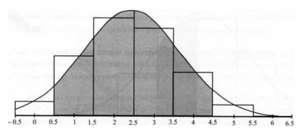
Continuity correction

- Motivation: Now we are going to use the CLT as an approximation tool when sampling from **discrete distributions**.

  Specifically, we will discuss a way to improve our approximations to account for the discrepancy created from using a continuous distribution / probability methods (integral to calculate area under curve) on originally discrete distributions.

- This is called the (half unit) **continuity correction** and is demonstrated below.





- Estimate the following probabilities using the continuity correction:

  (a) $P(1 \leq X \leq 4) \approx$

  (b) $P(X = 2) \approx$

  (c) $P(1 \leq X < 4) =$

  (d) $P(X > 2.5) =$

- In general, if X is the original discrete random variable of interest, S is the corresponding normal random variable based on the CLT, and a, b are some integers ($a \leq b$), then we can summarize the adjustments for the **continuity correction** with:

$$P(X = a) = P(a - 0.5 \leq S \leq a + 0.5)$$

$$P(a \leq X \leq b) = P(a - 0.5 \leq S \leq b + 0.5)$$

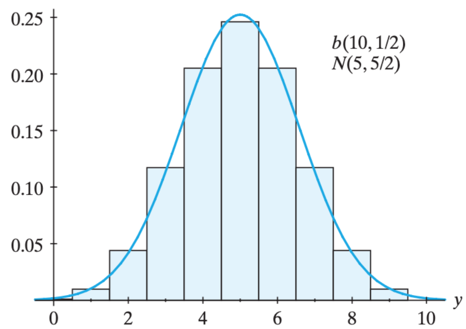
  Just need to take care to decide if we are want to include or exclude a or b (so can rewrite strict inequalities $<$ $>$ as inclusive $\leq$ $\geq$ and then use rule).

Normal approximation to the binomial distribution

- The most common scenario when applying the normal approximation is to the binomial distribution.
- Recall if $X \sim \text{Binomial}\,(n, p)$

- This means for "large n", $X =$

  "Rule of thumb" is that n is sufficiently large if $np \geq 5$ and $n(1-p) \geq 5$.
  We will discuss reasoning behind this after some examples.



- Examples→ Good use of the normal approximation

  1. Suppose that a multiple choice exam has 40 questions, each with 5 possible answers. A student feels that he has a probability of 0.55 of getting any particular question correct, with independence from one question to another.

     Approximate the probability of the student getting at least 25 correct.

     (a) With continuity correction:

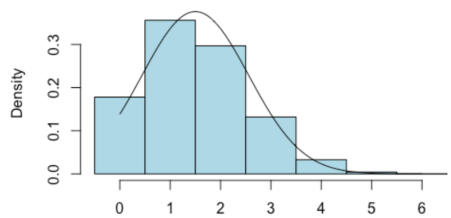
     (b) Without continuity correction:

     (c) Exact answer using binomial distribution:

- Bad example: Suppose we change the scenario in example 1 so that there are only 20 questions and the probability of getting any particular question correct is now 0.10.

  Compare the approximate answer and exact answer for $P(X \geq 3)$.

- Why is the approximation bad??
  - Lets take a look at the histogram and overlaid normal pdf for a similar scenario:



$X \sim Binomial(n = 6, p = 0.25)$

  - This illustrates the mismatch between the skewed probability histogram for and the symmetric pdf of the normal distribution. In order to do a good job of approximating the binomial distribution, the normal curve must have the bulk of its own distribution between legitimate outcomes for the Binomial distribution $[0, n]$.
  - How do we apply / check this: Based on the empirical rule, the central 95% of any normal distribution lies within two standard deviations of its mean.

  - Thus, as long as we ensure that ____________________________, the normal approximation to the a binomial distribution will be good.

    Contextually, this condition means that we must expect (expected value) the number of success (np) and failures (nq) to be at least 5.

- This relationship between a large sample binomial and normal is important for confidence intervals and hypothesis tests of population proportions which we will cover next.

Normal approximation to the Poisson distribution

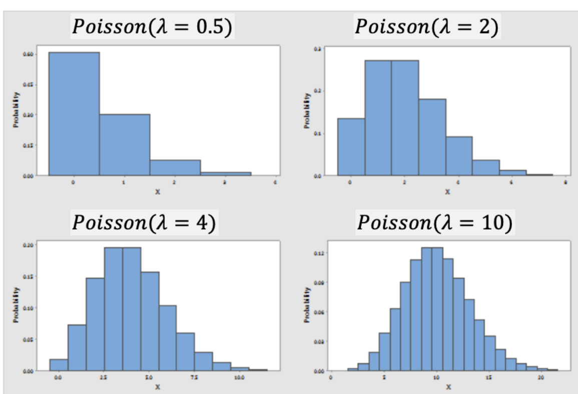
- A Poisson distribution with large enough mean can also be approximated with the use of a normal distribution.

  Let $X \sim \text{Poisson}\,(\lambda)$, with $E(X) = V(X) = \lambda$, where $\lambda = 1, 2, \ldots$ (in general, we just need $\lambda > 0$, but for demonstration lets assume λ is a positive integer).

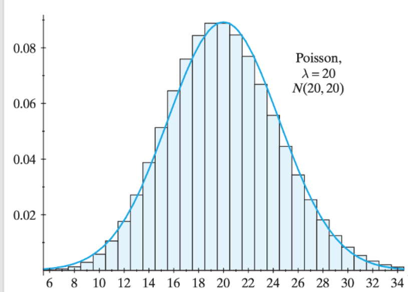
- We can rewrite X as a sum of Poisson random variables:

- This means for "large n", $X =$

  "Rule of thumb" is that n is sufficiently large if $\lambda \geq 10$ (doesn't need to be an integer).



Poisson probability histograms



- Example

  Let X equal the number of alpha particles emitted by barium-133 per second and counted by a Geiger counter. Assume that $X \sim \text{Poisson}\,(\lambda = 49)$.

  Approximate $P(45 \leq X < 60)$.

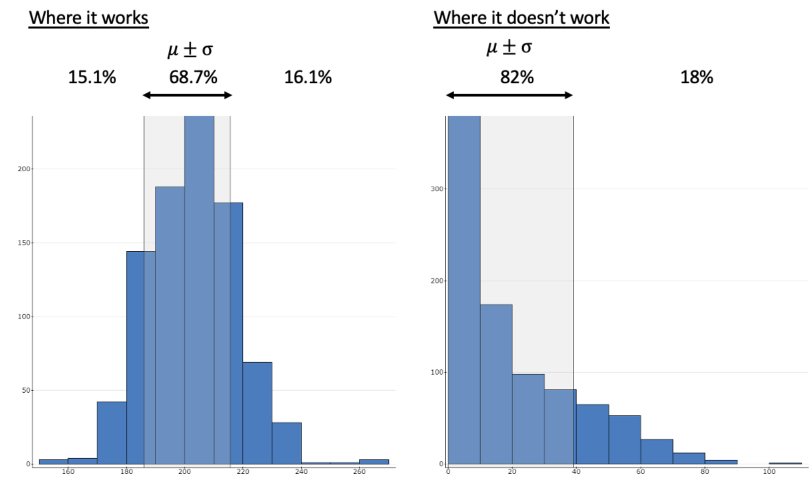
Summary of normal approximation to the binomial and Poisson distributions

- Suppose n is large and $a = 0, 1, \dots, n$
  (a) CLT:
  If $X \sim \text{Binomial}\,(n, p) \Longrightarrow X \approx S \sim \text{Normal}\,(\mu = np, \sigma = \sqrt{npq})$
  If $X \sim \text{Poisson}\,(\lambda) \Longrightarrow X \approx S \sim \text{Normal}\,(\mu = \lambda, \sigma = \sqrt{\lambda})$
  (b) Continuity correction:
  $P(X \leq a) \approx \text{Normalcdf(lower} = 0, \text{upper} = a + 0.5)$
  $P(X < a) \approx \text{Normalcdf(lower} = 0, \text{upper} = a - 0.5)$
- Final note: In practice, if you have technology / software, just compute discrete probabilities exactly. However, it is important to learn how to apply the central limit theorem.

Central interval probabilities

Empirical rule

- Motivation: Because the normal distribution can be used in so many scenarios due to the CLT, there are common generalizations that are made about **central interval** probabilities for distributions that are approximately bell-shaped.
- First, lets calculate these exactly for the standard normal curve. These will of course apply to any normal distribution X with mean μ and standard deviation σ because we can standardize to get Z.
  1. $P(-1 \leq Z \leq 1) =$
  2. $P(|Z| \leq 2) =$
  3. $P(|Z| \leq 3) =$
- Not all data is exactly normally distributed of course, but because of the CLT many distributions can be approximated by a normal distribution. So we can use the exact probabilities above to make generalizations about these distributions that have a similar shape.
- The **empirical rule** states that for approximately normal distribution:
  1. Approximately __________ of data falls within _____ standard deviation of the mean.
  2. Approximately __________ of data falls within _____ standard deviations of the mean.
  3. Approximately _______________ (nearly all) of data falls within _____ standard deviations of the mean.



- Example: Suppose that the scores on an achievement test are known to have, approximately, a normal distribution with mean $\mu = 64$ and standard deviation $\sigma = 10$.
  (a) Find the scores probability scores are between 54 and 74.

  (b) Find which two values lies the central 95%?

  (c) Find the percent of scores above 94.

- Thus, knowledge of the mean and the standard deviation gives us a fairly good picture of the frequency distribution of scores when the bell-shape is present (or assumed).