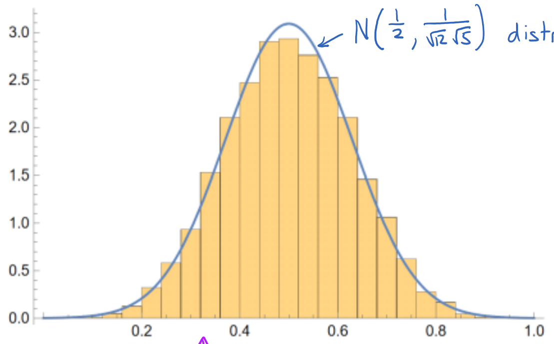
From last time:

5. Simulate 10,000 averages, each of k samples from a Unif[0,1] distribution. Make a histogram of the 10,000 averages. Start with $k = 1$ and then try larger values of k. How does the shape of the histogram depend on k?

larger $k \Longrightarrow$ closer to normal

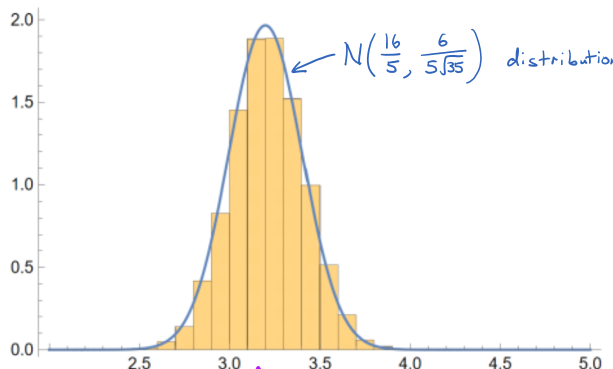


$N\left(\frac{1}{2}, \frac{1}{\sqrt{12}\sqrt{5}}\right)$ distribution

Histogram of 10,000 samples, each a sum of 5 Unif[0, 1] random variables.

6. Repeat the previous simulation, but now replace Unif[0,1] with a different distribution of your choice. What is the shape of the histogram? How does it depend on k?

again,
larger $k \Longrightarrow$ closer to normal



$N\left(\frac{16}{5}, \frac{6}{5\sqrt{35}}\right)$ distribution

Histogram of 10,000 samples, each a sum of 40 Hypergeometric$(n=8, M=20, N=50)$ random variables.

New problems today (limit theorems):

1. Let $X_1, X_2, \ldots, X_{300}$ be iid random variables with mean μ_X and standard deviation σ_X. Also let $T = X_1 + X_2 + \cdots + X_{300}$ and $\bar{X} = \frac{T}{300}$.

(a) What are μ_T, σ_T, $\mu_{\bar{X}}$, and $\sigma_{\bar{X}}$?

$\mu_T = 300\mu_X \qquad \sigma_T = \sigma_X\sqrt{300}$

$\mu_{\bar{X}} = \mu_X \qquad \sigma_{\bar{X}} = \dfrac{\sigma_X}{\sqrt{300}}$

(b) What distributions are good approximations for T and $\bar{X}$?

T is approx. $N(300\mu_X, \sigma_X\sqrt{300})$, $\bar{X}$ is approx $N\left(\mu_X, \frac{\sigma_X}{\sqrt{300}}\right)$

2. A farm packs tomatoes in crates. Individual tomatoes have mean weight of 10 ounces and standard deviation of 3 ounces. Estimate the probability that a crate of 40 tomatoes weighs between 380 and 410 ounces.

T_{40} is approximately $N(400, \underbrace{18.97}_{3\sqrt{40}})$

$P(380 < T_{40} < 410) \approx 0.555$

R: pnorm(410, 400, 18.97) − pnorm(380, 400, 18.97)

3. Let random variable X have one of the following distributions. For what distribution of iid random variables $Y_1, Y_2, \ldots, Y_n$ is it the case that $X = Y_1 + Y_2 + \cdots + Y_n$?

(a) $X \sim \text{Bin}(n, p)$ $Y_i \sim \text{Bernoulli}(p)$

X is approx. normal when $np \geq 10$ and $n(1-p) \geq 10$.

(b) $X \sim \text{Gamma}(\alpha = n, \beta)$ $Y_i \sim \text{Exp}\left(\lambda = \frac{1}{\beta}\right)$

X is approx. normal when α is large.

(c) $X \sim \text{Poisson}(\lambda = n)$ $Y_i \sim \text{Poisson}(1)$

X is approx. normal when λ is large.

(d) $X \sim \text{NegBin}(r = n, p)$ $Y_i \sim \text{Geom}(p)$

X is approx. normal when r is large.

4. Customers at a popular restaurant are waiting to be served. Waiting times are independent and exponentially distributed with mean $1/\lambda = 10$ minutes.

(a) What is the probability that the average wait time of the 50 customers is less than 12 minutes?

T_{50} is $\text{Gamma}(\alpha = 50, \beta = 10)$.

$\bar{X}_{50} = \frac{T_{50}}{50}$ $P(\bar{X}_{50} < 12) = P\left(\frac{T_{50}}{50} < 12\right) = P(T_{50} < 600) \approx 0.916$

R: pgamma(600, 50, $\frac{1}{10}$)

$\bar{X}_{50}$ is $\text{Gamma}\left(\alpha = 50, \beta = \frac{1}{5}\right)$. why? mgfs!

(b) Use a normal distribution to approximate the probability that the average wait time of 50 customers is less than 12 minutes. What limit theorem justifies this?

$\bar{X}_{50}$ is approx $N\left(10, \frac{10}{\sqrt{50}}\right)$, so $P(\bar{X}_n < 12) \approx P(Z < 12) \approx 0.921$

$Z \sim N\left(10, \frac{10}{\sqrt{50}}\right)$

R: pnorm(12, 10, 1.414)

5. Let $X_1, X_2, \ldots, X_n$ be iid random variables with an Exp($\lambda = 2$) distribution. Let $\mu = E(X_i)$.

(a) What is the distribution of T_n? What is the value of μ?

$T_n \sim \text{Gamma}\left(\alpha = n, \beta = \frac{1}{2}\right)$ $\qquad \mu = E(X_i) = \frac{1}{2}$

(b) In **R** or *Mathematica*, write a function that computes $P\left(\left|\frac{T_n}{n} - \mu\right| \geq \epsilon\right)$ for any given parameter values n and ϵ.

First: $P\left(\left|\frac{T_n}{n} - \mu\right| \geq \varepsilon\right) = 1 - P\left(\left|\frac{T_n}{n} - \frac{1}{2}\right| < \varepsilon\right) = 1 - P\left(-\varepsilon < \frac{T_n}{n} - \frac{1}{2} < \varepsilon\right)$

$= 1 - P\left(\frac{n}{2} - n\varepsilon < T_n < \frac{n}{2} + n\varepsilon\right)$

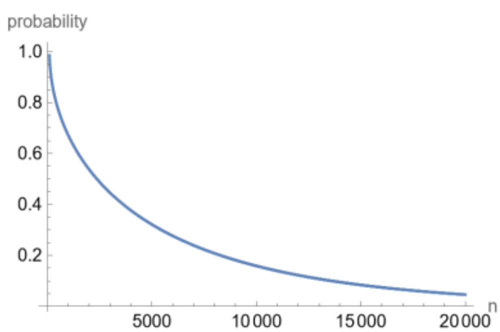
R:

```
wlln <- function(n, eps){
  1 - (pgamma(n/2 + n*eps, n, 2) - pgamma(n/2 - n*eps, n, 2))
}
```

Mathematica:

```
wlln[n_, ϵ_] := 1 - Probability[n/2 - n*ϵ < Tn < n/2 + n*ϵ, Tn ≈ GammaDistribution[n, 1/2]]
```

(c) Make a plot of $P\left(\left|\frac{T_n}{n} - \mu\right| \geq 0.01\right)$ for values of n between 1 and 10,000. What limit theorem does this plot illustrate?



$P\left(\left|\frac{T_n}{n} - \mu\right| \geq \frac{1}{100}\right)$ converges to zero, illustrating the weak law of large numbers.

— class ended here —

(d) What is the smallest n such that $P\left(\left|\frac{T_n}{n} - \mu\right| \geq 0.01\right) < 0.01$?

By trial and error, we find $n = 16{,}589$.

6. Suppose you flip a fair coin *lots* of times. What does the Law of Large Numbers say about the numbers of heads and tails you will observe?

7. Suppose that a certain casino game costs \$1 to play, and the expected winnings per game are \$0.98. What does the Law of Large Numbers say about your winnings if you play the game lots of times?

8. Fifty real numbers are each rounded to the nearest integer and then summed. If the indivdual round-off errors are uniformly distributed over $(-0.5, 0.5)$, then approximate the probability that the resultant sum differs from the exact sum by no more than 3.

9. Suppose that a fair coin is tossed 1000 times. If the first 100 tosses result in heads, what proportion of heads would you expect on the remaining 900 tosses? Interpret the statement "The law of large numbers swamps, but it does not compensate."