

Central Limit Theorem

ST551 Lecture 6

Charlotte Wickham

2017-10-02

Central Limit Theorem (CLT)

If the population distribution of a variable X has population mean μ and (finite) population variance σ^2 , then the *sampling distribution of the sample mean* becomes closer and closer to a Normal distribution as the sample size n increases.

We can write:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

for large values of n , where the symbol \sim means *approximately distributed as*.

We already figured out the mean and variance, the CLT adds the shape information

Central Limit Theorem (CLT)

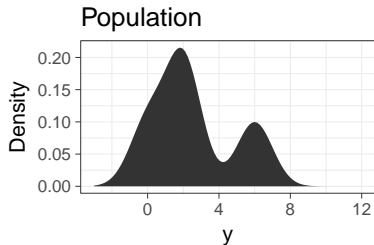
Slightly more formally..

Let X_1, X_2, \dots, X_n be an i.i.d. sample from some population distribution F with mean μ and variance $\sigma^2 < \infty$. Then as the sample size $n \rightarrow \infty$ we have

$$\frac{\bar{X} - \mu}{\sigma} \rightarrow_D N(0, 1)$$

Your Turn

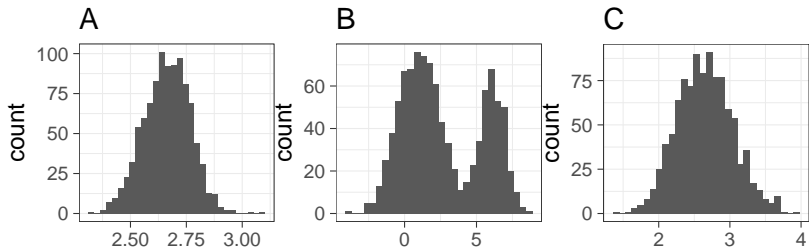
Consider the following population distribution.



The following three histograms represent:

1. A sample of size 1000 from the population
2. 1000 sample means of samples of size 10 from the population
3. 1000 sample means of samples of size 100 from the population

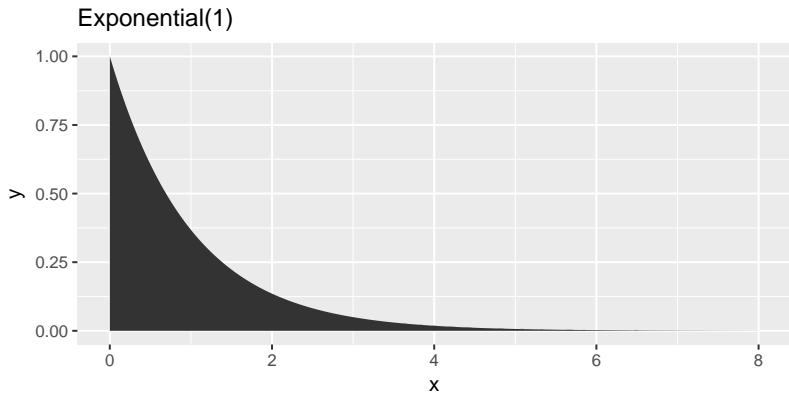
Which is which?



How big is big enough?

What value of n is *big enough* for the approximation to be *good enough*?

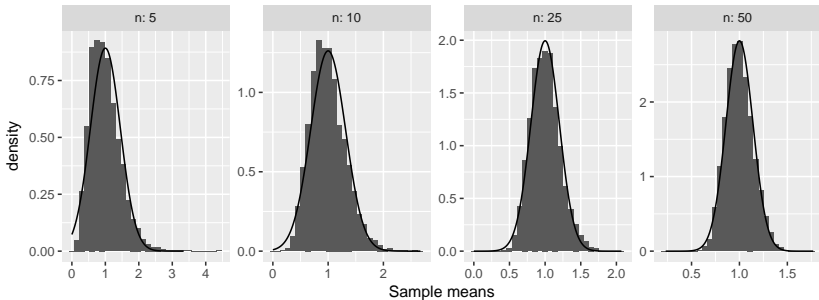
It depends on the population



How big is big enough?

What value of n is *big enough* for the approximation to be *good enough*?

It depends on the population



See more in lab...

Your turn

Using the CLT to approximation the sampling distribution

Population: $\sim (\mu = 20, \sigma^2 = 4)$

Sample: $n = 16$ i.i.d from population

Sample statistic: Sample mean

What is the approximate distribution for \bar{Y} ?

Applying CLT

Same setup: we have sample of size, $n = 16$ from a population with population mean $\mu = 20$ and population variance $\sigma^2 = 4$.

What is the probability the sample mean is less than 20.5?

CLT says:

Using a Standard Normal Probability Table

To use a Standard Normal table, we first need to transform our random variable (\bar{Y}) to a Standard Normal.

We can convert a probability for \bar{Y} to a standard Normal probability by:

1. Subtracting the mean of \bar{Y} from both sides, then
2. Dividing both sides by the standard deviation of \bar{Y} (square root of the variance)

$$P(\bar{Y} \leq 20.5) =$$

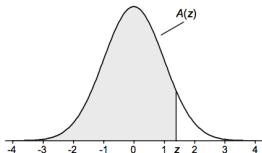
Using a Standard Normal Probability Table

Now look up $z = 1.00$ in Standard Normal Table.

$$P(\bar{Y} \leq 20.5) = 0.8413$$

TABLE A.1

Cumulative Standardized Normal Distribution



$A(z)$ is the integral of the standardized normal distribution from $-\infty$ to z (in other words, the area under the curve to the left of z). It gives the probability of a normal random variable not being more than z standard deviations above its mean. Values of z of particular importance:

z	$A(z)$	
1.645	0.9500	Lower limit of right 5% tail
1.960	0.9750	Lower limit of right 2.5% tail
2.326	0.9900	Lower limit of right 1% tail
2.576	0.9950	Lower limit of right 0.5% tail
3.090	0.9990	Lower limit of right 0.1% tail
3.291	0.9995	Lower limit of right 0.05% tail

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830

Using a Standard Normal Probability Table

What if z is negative?

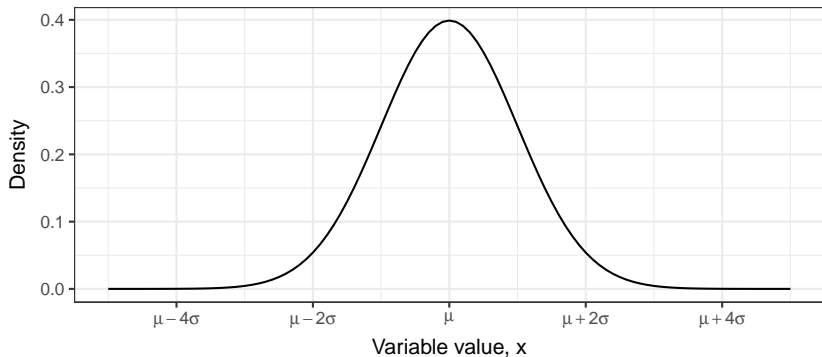
Standard Normal is symmetric around zero.

$$P(Z \leq -z) = 1 - P(Z \leq z)$$

Density

Density: height of probability density function at x : $f(x; \mu, \sigma)$

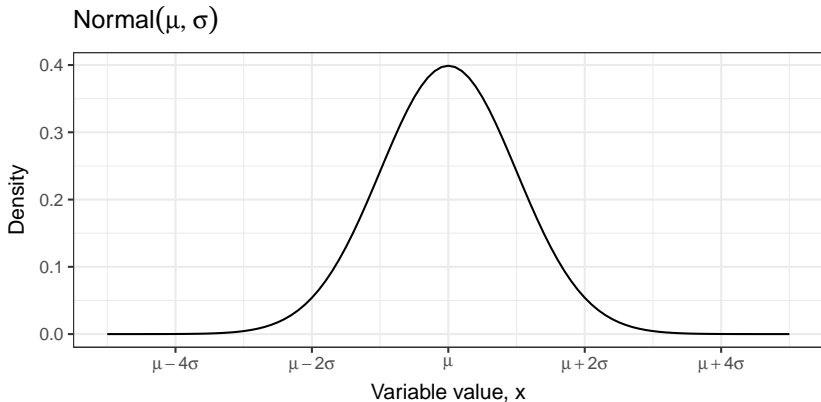
Normal(μ, σ)



In R: `dnorm(x, mean = mu, sd = sigma)`

Cumulative Probability

Cumulative Probability: the area under the probability density function to the right of x : $F(x; \mu, \sigma)$



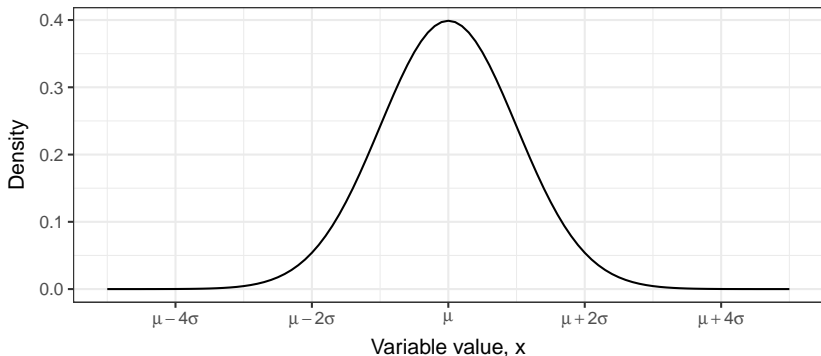
In R: `pnorm(q, mean = mu, sd = sigma)`

Area to left?

Quantiles

Quantile: The p th quantile is the value x , such that the area to the left of x under the density function is p . $F^{-1}(p; \mu = 0, \sigma^2 = 1)$

Normal(μ, σ)



```
qnorm(p, mean = mu, sd = sigma)
```

What should these return?

```
qnorm(0, mean = 0, sd = 1)
```

```
qnorm(1, mean = 0, sd = 1)
```

```
dnorm(-Inf)
```

```
dnorm(Inf)
```

```
pnorm(Inf)
```

```
pnorm(-Inf)
```


Exercise: Find probability

Return to example: $\mu = 20$ $\sigma^2 = 4$, $n = 16$

What is $P(\bar{Y} < 20.5)$? What is $P(\bar{Y} < 21)$

Exercise: Find sample size for specific variance in sample mean

$$\mu = 20 \quad \sigma^2 = 4.$$

What should n be so $\text{Var}(\bar{Y}) = 0.5$?

Exercise: Find sample size for an interval with desired probability

What should n be, so that $P(19.5 < \bar{Y} < 20.5) = 0.9$?