* Announcement:

① Exam-prep section: 2~3 PM

STAT 88: Lecture 25

Contents

Warm up: (Exercise 7.4.11) Each Data 8 student is asked to draw a random sample and estimate a parameter using a method that has chance 95% of resulting in a good estimate.

Suppose there are 1300 students in Data 8. Let X be the number of students who get a good estimate. Assume that all the students' samples are independent of each other.

(a) Find the distribution of X.

(b) Find $E(X)$ and $\text{SD}(X)$.

(c) Find the chance that more than 1250 students get a good estimate.

(a) $X \sim \text{Binom}(1300, 0.95)$

(b) $E(X) = 1300 \cdot (0.95)$, $\quad SD(X) = \sqrt{1300 \cdot (0.95)(0.05)}$

(c) $P(X > 1250) = \underbrace{\sum_{k=1251}^{1300} \binom{1300}{k} (0.95)^k (0.05)^{1300-k}}_{\text{Approximate using CLT (Central Limit Theorem)}}$

Last time

SD of sample sum:

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim}$ with mean μ and SD σ. Let $S_n = X_1 + X_2 + \cdots + X_n$. Then

$$E(S_n) = n\mu, \qquad \mathrm{SD}(S_n) = \sqrt{n}\sigma.$$

SD of sample mean:

Let $\bar{X}_n = S_n/n$. Then

$$E(\bar{X}_n) = \mu, \qquad \mathrm{SD}(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}.$$

The law of large numbers: For a fixed $c > 0$,

$$P(\mu - c < \bar{X}_n < \mu + c) = P(|\bar{X}_n - \mu| < c) \to 1 \text{ as } n \to \infty.$$

Today: How the shape of the distribution of S_n look like?

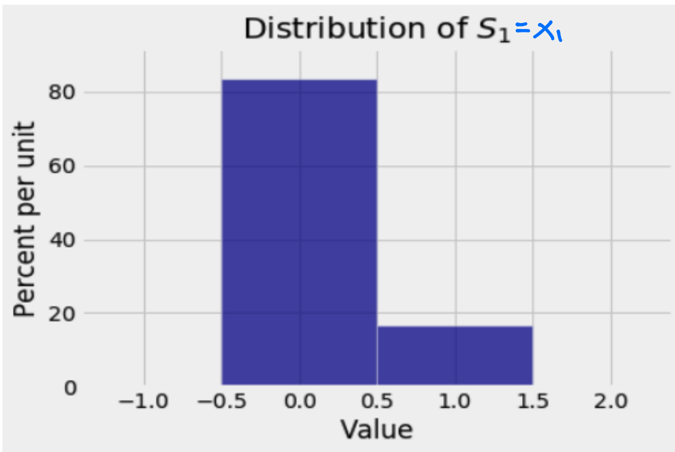
LLN: $\frac{S_n}{n} = \overline{X}_n$ concentrates around μ as $n \to \infty$

What is the distribution of $\overline{X}_n$ (or S_n)?

8.1. The Distribution of a Sample Sum

Sum of IID Indicators If $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$, then $S_n = X_1 + X_2 + \cdots + X_n$ has the Binomial(n, p) distribution. What the distribution of S_n look like?

$p = 1/6$



$n=1$

Lab 1: $X_1, \cdots, X_n \sim \text{Bernoulli}(p)$
Get $S_n = X_1 + X_2 + \cdots + X_n$

Lab 2: $X_1, \cdots, X_n \sim \text{Bernoulli}(p)$
Get $S_n = X_1 + X_2 + \cdots + X_n$

$\vdots$

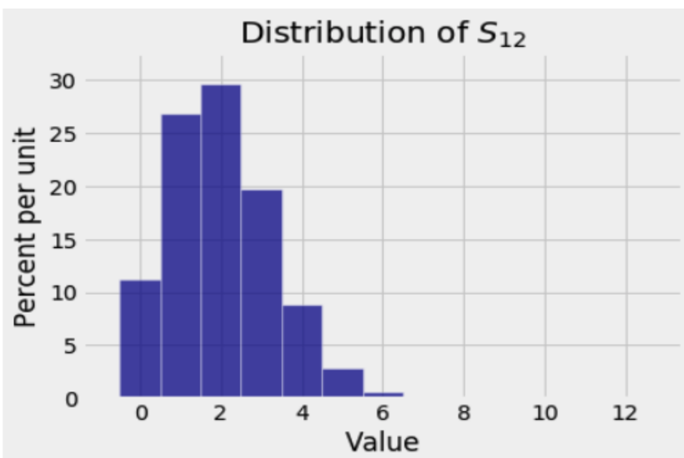
Lab 10,000: $X_1, \cdots, X_n \sim \text{Bernoulli}(p)$
Get $S_n = X_1 + X_2 + \cdots + X_n$

$\Downarrow$

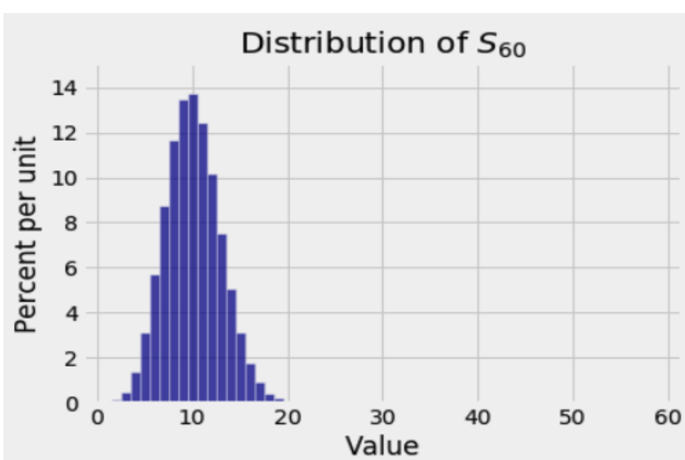
Distribution of S_n

$\hookrightarrow E(S_n) = n \cdot \mu$

$SD(S_n) = \sqrt{n} \cdot \sigma$

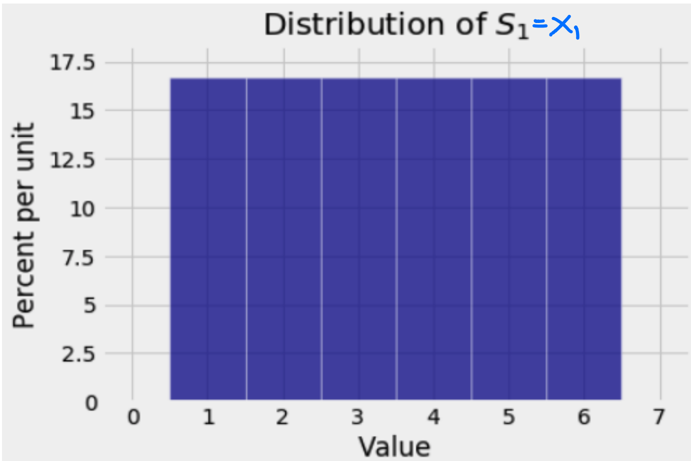


$n=12$

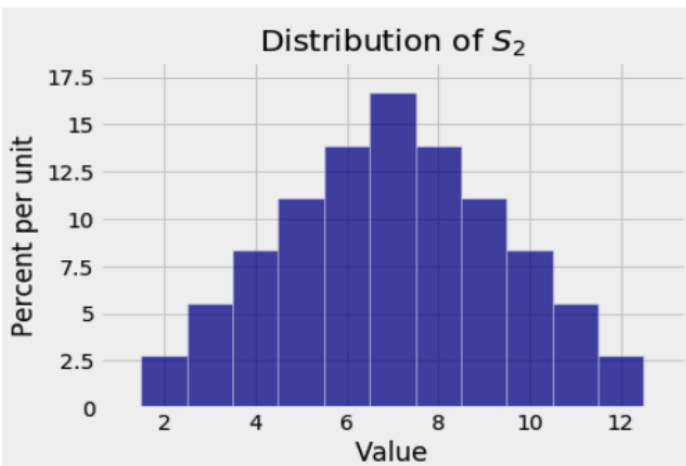


$n=60$

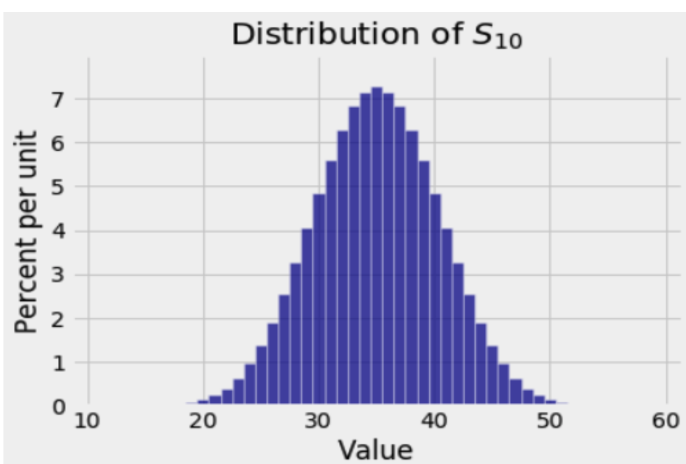
Sum of IID Uniform Random Variables Let $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Uniform}\{1,2,3,4,5,6\}$ and $S_n = X_1 + X_2 + \cdots + X_n$. What the distribution of S_n look like?



n=1

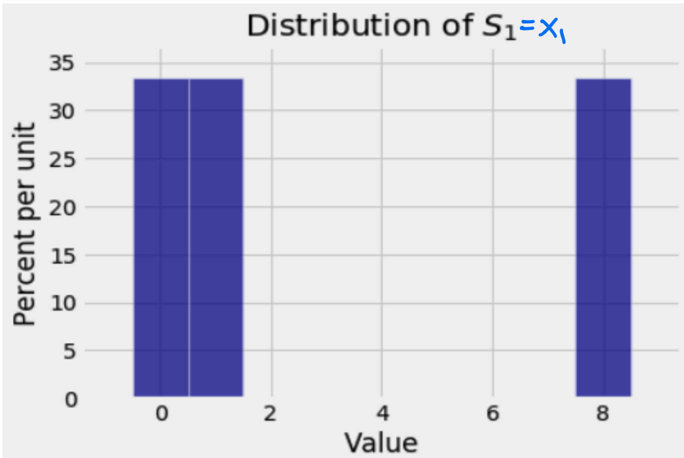


n=2

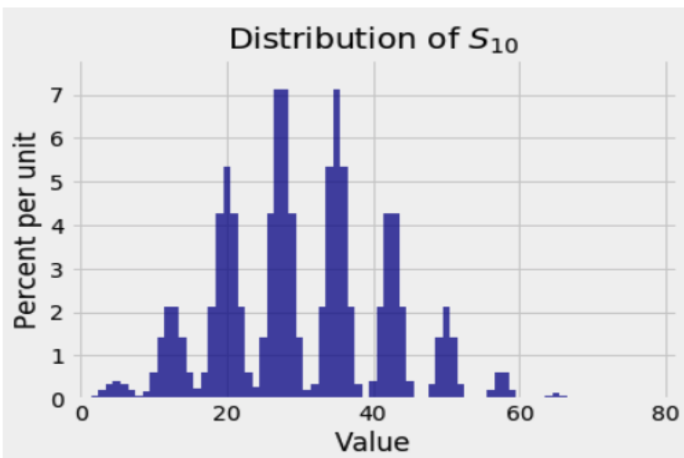


n=10

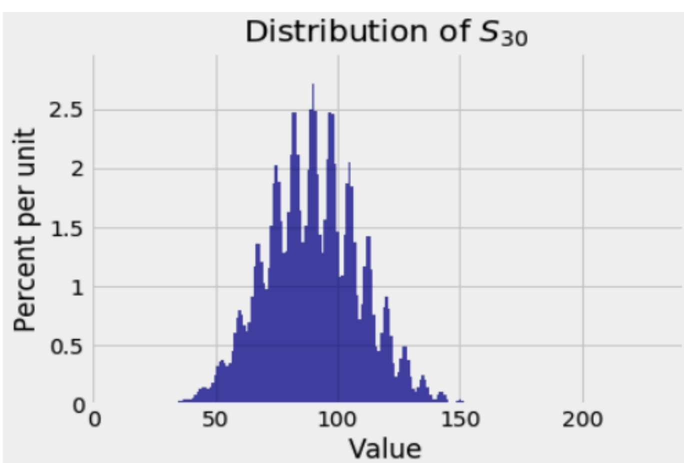
A Wild One Let $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Uniform}\{0, 1, 8\}$ and $S_n = X_1 + X_2 + \cdots + X_n$. What the distribution of S_n look like?



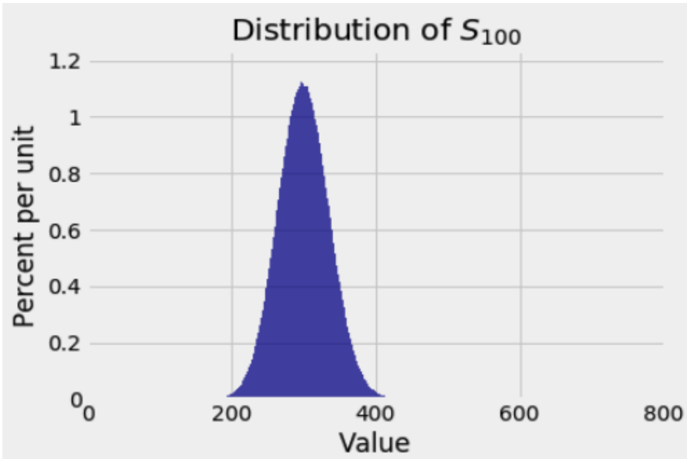
n=1



n=10

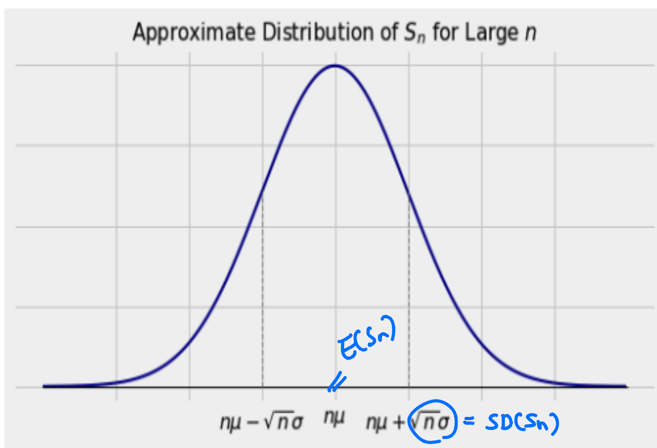


n=30

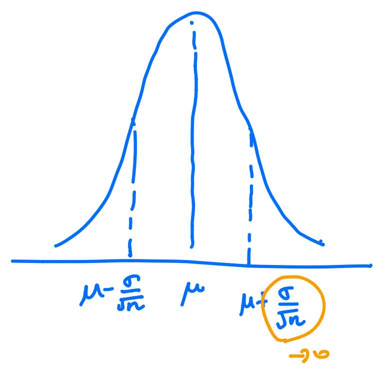


n=100

Central Limit Theorem Let $X_1, X_2, \dots, X_n$ be i.i.d. with $E(X_1) = \mu$ and $\mathrm{SD}(X_1) = \sigma$. Let $S_n = X_1 + X_2 + \cdots + X_n$ be the sample sum. If n is large, the distribution of S_n is approximately normal ≠(bell-shaped curve), regardless of the distribution of the X_i's.



Approximate Dist'n of $\bar{X}_n$ for Large n



Key idea: It is easier to approximate $P(X > 1250)$ using the fact that Binomial is almost Normal for large n.

$X \sim$ Binom(1300, 0.95)

$\approx$ Normal

Binomial ~ Normal { n is large, p is fixed }

Binomial ~ Poisson { n is large, p is small }

$p = \frac{1}{n}$, $np = 1$ "Poisson"

p = fixed, $np \to \infty$ "Normal"

- How fast dist'n of S_n converges to normal (= Bell-shaped) curve?

$|\text{Dist}(S_n) - \text{Normal curve}| \approx \frac{1}{\sqrt{n}} \cdot C$

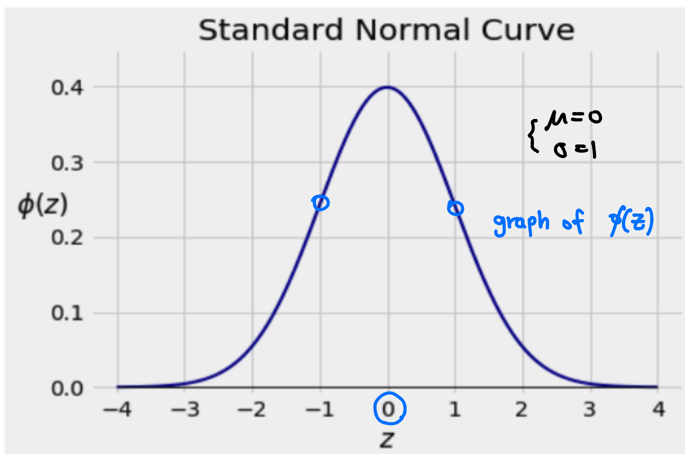
8.2. Standard Normal Curve

The normal or Gaussian curves are a family of bell-shaped curves named for the German mathematician and scientist Carl Friedrich Gauss.

The Standard Normal Curve

The standard normal curve is defined by

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \quad -\infty < z < \infty.$$



Properties:

- The curve is bell-shaped and symmetric about 0.
- The points of inflection are at $z = -1$ and $z = 1$.
- For $|z| > 3$, the curve is pretty close to 0.
- The total area under the curve is 1. $\rightsquigarrow$ approximation to a prob. distribution.

  $\left(\int_{-\infty}^{\infty} \phi(z)\, dz = 1 \right)$

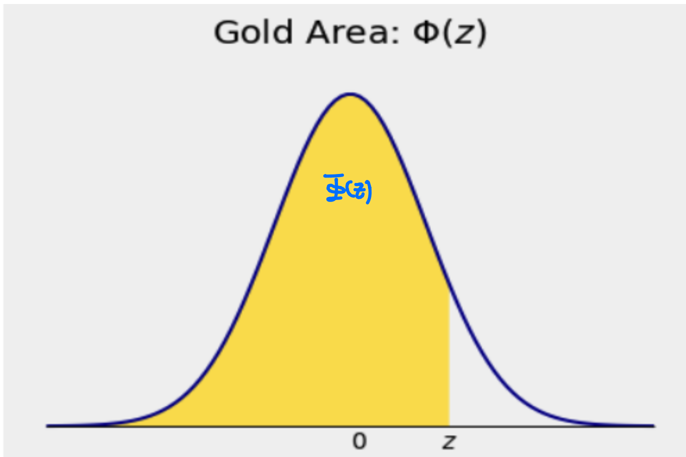
The Standard Normal 'CDF'

If you think of the standard normal curve as a probability histogram, then it is natural to think of areas under the curve as probabilities.

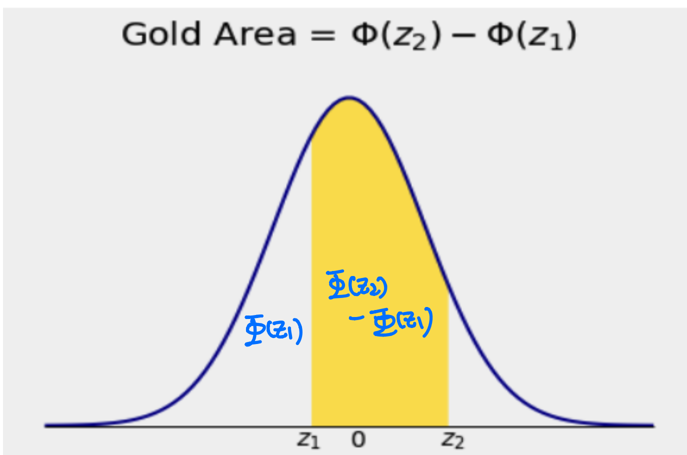
CDF : $F(x) = P(X \le x)$

$$\Phi(z) = \int_{-\infty}^{z} \phi(x)dx.$$

Φ gives all the area under the curve to the left of z:



The area under the curve over any interval (z_1, z_2) is then $\Phi(z_2) - \Phi(z_1)$:



Numerical Values of the Areas

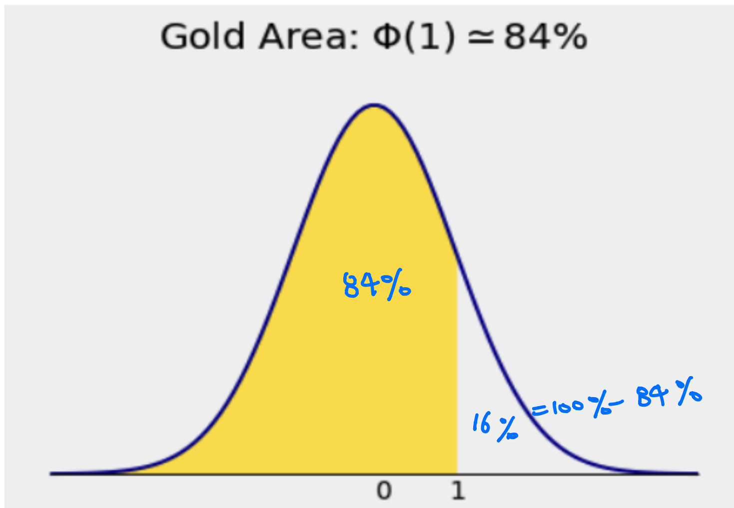
Calculating $\Phi(z)$ in Python:

```
stats.norm.cdf(1)
```

in scipy library

```
0.8413447460685429
```

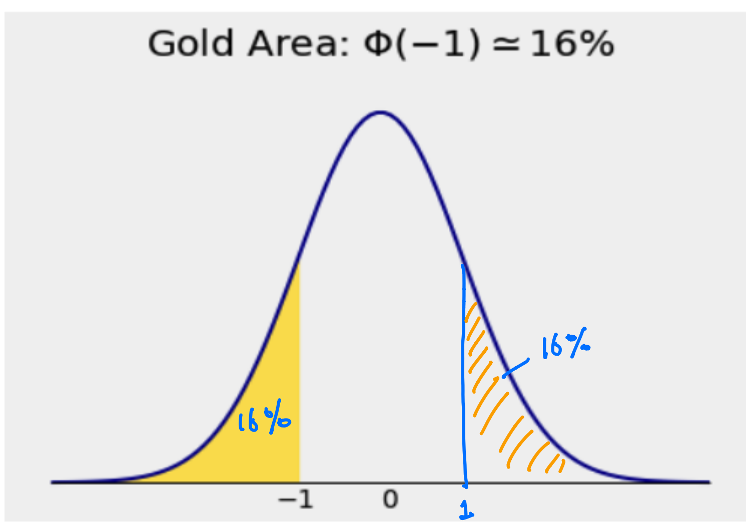
$= \Phi(1)$



By symmetry:

```
stats.norm.cdf(-1)
```

```
0.15865525393145707
```



→ For any z, $1 - \Phi(z) = \Phi(-z)$

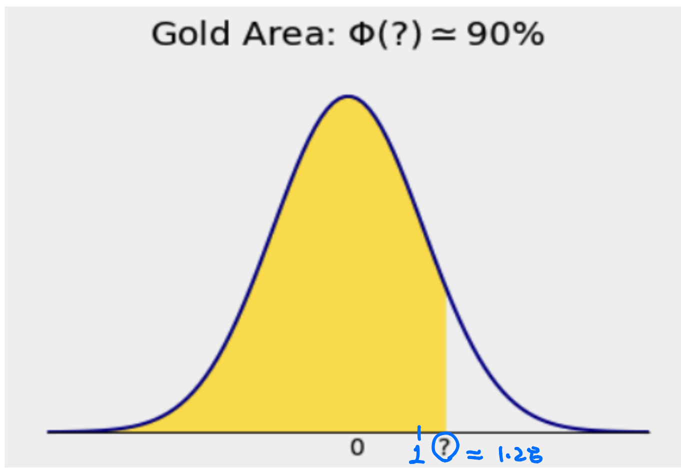
Percentiles

We saw the area under the curve to the left of 1 is about 84%

$$\Phi(1) \approx 84\%.$$

The point $z = 1$ is therefore called the 84th percentile of the curve. If you think of the curve as a probability histogram, then about 84% of the probability lies below $z = 1$.

The 90th percentile must be to the right of 1. But how far to the right?



We need to find the inverse of $\Phi(z)$. The 90th percentile is the point z such that $\Phi(z) = 0.9$, or

take Φ^{-1}

$$z = \Phi^{-1}(0.9).$$

inverse function of Φ.

Calculating $\Phi^{-1}(q)$ in Python:

Percent Point function

```
stats.norm.ppf(0.9)
```

//

```
1.2815515655446004
```

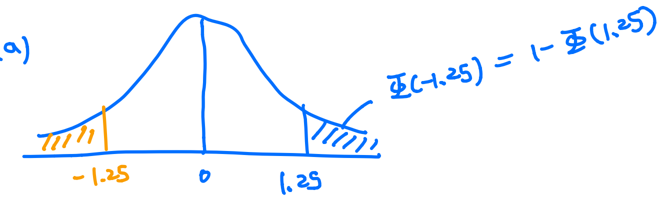
Example: Find the area

(a) to the right of 1.25.

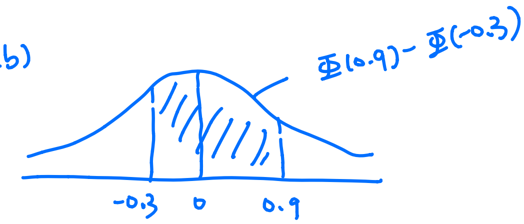
(b) between -0.3 and 0.9.

(c) Outside -1.5 and 1.5.

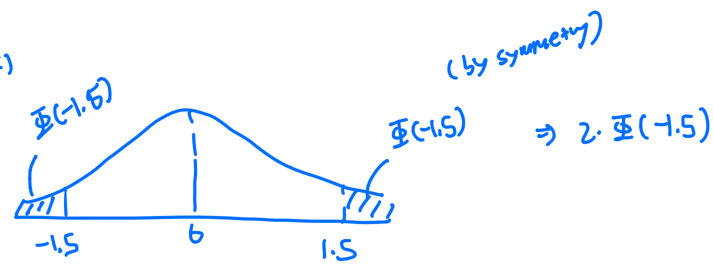
(a)



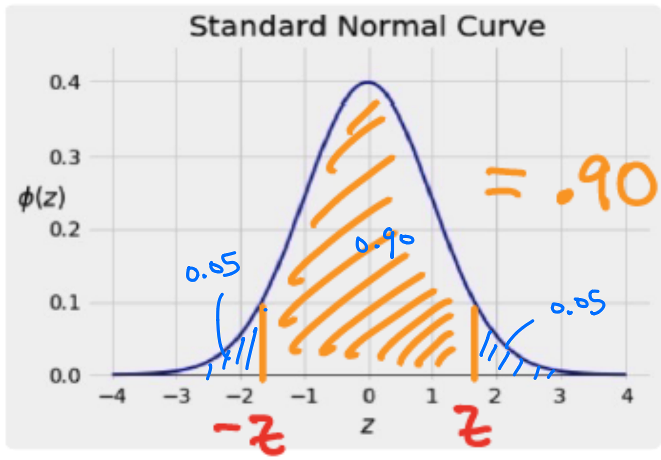
(b)



(c)



Example: The standard normal curve is sketched below. Solve for z.



$\Phi(-z) = 0.05 \iff \Phi(z) = 0.95$

$\Rightarrow -z = \Phi^{-1}(0.05) \qquad \Rightarrow z = \Phi^{-1}(0.95)$

$\Rightarrow z = -\Phi^{-1}(0.05)$

$-\Phi^{-1}(0.05) = \Phi^{-1}(0.95)$

$\Rightarrow$ for any z,

$-\Phi^{-1}(-z) = \Phi^{-1}(z)$