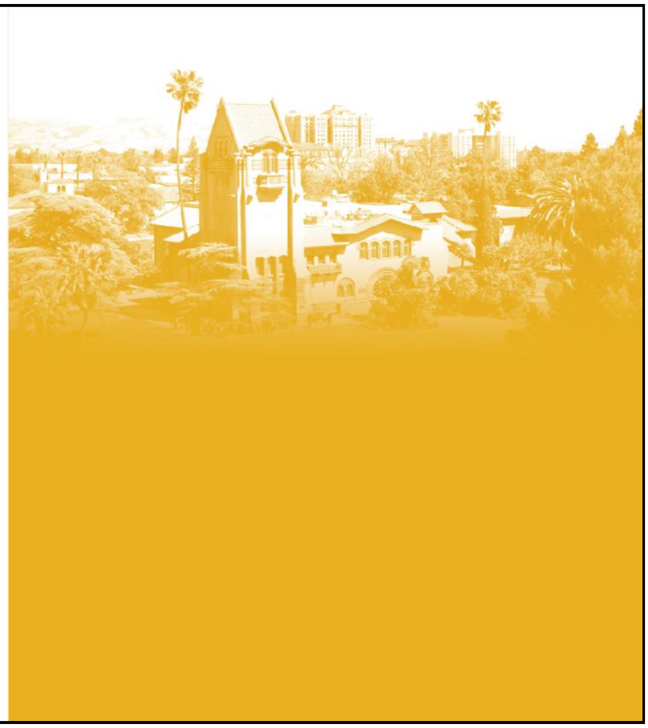
SJSU SAN JOSÉ STATE UNIVERSITY



Random Variables
& Discrete Distributions

1

SJSU SAN JOSÉ STATE UNIVERSITY

Agenda

- Continuous Random Variables & Probability Distributions
  - Uniform Distribution
  - Exponential Distribuion
  - Normal Distribution
  - Other Distributions (Gamma, Beta etc)

2

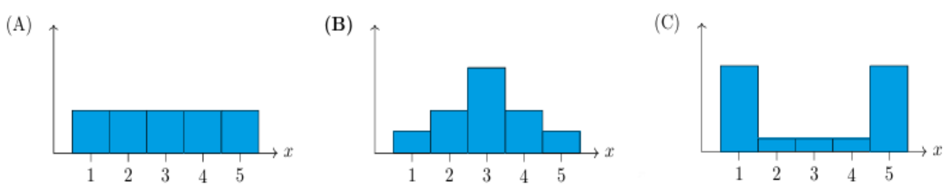
2

Review Questions

3

Variance of PMF’s

- Order the following PMF’s by size of standard deviation from biggest to smallest.



4

SJSU SAN JOSÉ STATE UNIVERSITY

Continuous Random Variables

5

SJSU SAN JOSÉ STATE UNIVERSITY

What is a Random Variable?

A random variable is a numeric quantity whose value depends on the outcome of a random event.

- We normally use a capital letter, like X, to denote a random variable
- The values of a random variable are denoted with a lowercase letter, in this case x
  - e.g. P(X = x) – denotes the probability of the (random variable) X having the value x.

6

6

Types of Random Variables

There are two types of random variables:

- Discrete random variables often take only integer values
- Continuous random variables take real (decimal) values

Feature	Continuous Random Variable	Discrete Random Variable
Possible Values	Infinite (uncountable)	Countable (finite or infinite)
Probability at a Single Value	0	Can be nonzero
Described by	Probability Density Function (PDF)	Probability Mass Function (PMF)
Example	Height, time, weight	Dice rolls, # of students in a class

7

7

Continuous Random Variables

- A continuous random variable can have any value within a given range.
- The value (not the probability) of a continuous random variable is given by its probability density function (PDF).

8

8

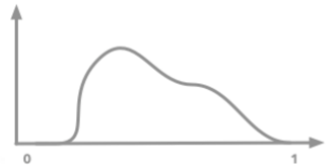
SJSU SAN JOSÉ STATE UNIVERSITY

Probability Distributions of Continuous Random Variables

A probability distribution of a continuous random variable can have any value within a given range and describes the likelihood of a continuous random variable taking on a specific value.

- It's also known as the probability density function (PDF).
- Unlike PMF, the probability of a single point is always zero:

$$P(X = x) = 0$$



- Instead, probability is determined over an interval:

$$P(a \leq x \leq b) = \int_a^b f(x)dx$$

9

9

SJSU SAN JOSÉ STATE UNIVERSITY

Probability Distributions of Continuous Random Variables

- This must satisfy two conditions:

$f(x) \geq 0$ for all x (non-negative)

$\int_{-\infty}^{\infty} f(x)dx = 1$ (total probability summed up to 1)

10

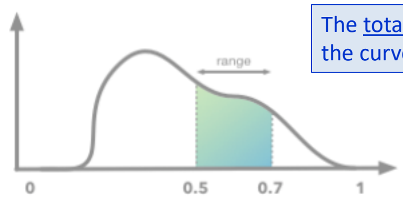
10

SJSU SAN JOSÉ STATE UNIVERSITY

Probability as the Area under the Curve

- If you graph a probability density function, the probability that the continuous random variable x has a value a ≤ x ≤ b is the area under the curve in the interval of a to b.

- Example: a = 0.5 & b = 0.7



The total area under the curve must equal 1.

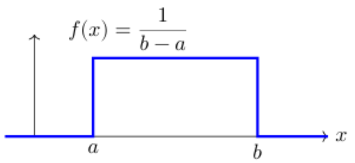
SJSU SAN JOSÉ STATE UNIVERSITY

Example: Uniform Distribution

- The PDF of the Uniform Distribution $X \sim U(a, b)$ is given by:

$$f(x) = \frac{1}{b-a}, \qquad a \leq x \leq b$$

- This PDF shows that each value in $[a, b]$ is equally likely.



Cumulative Distribution Function

The cumulative density function (CDF), denoted as $F(x)$, gives the probability that a random variable X is less than or equal to a given value

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt$$

- It's a continuous function with values between 0 and 1
- It accumulates the probability from $-\infty$ to x.
- The limits of $F(x)$: $\lim_{x \to -\infty} F(x) = 0, \qquad \lim_{x \to \infty} F(x) = 1$
- The PDF can also be computed from $F(x)$: $f(x) = \frac{d}{dx} F(x)$

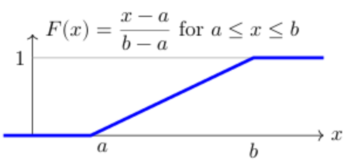
13

13

Example: Uniform Distribution (cont)

- The PDF of the Uniform Distribution is given by:

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases}$$



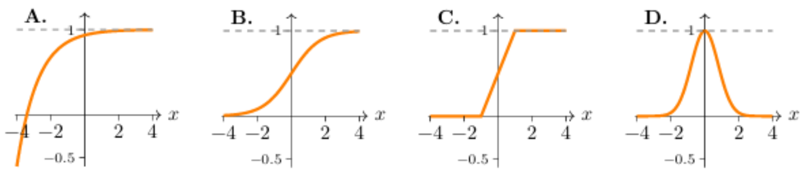
- It's a piecewise linear function increasing from 0 to 1.

14

14

Example

Which of the following are graphs of valid cumulative distribution functions?



15

15

Expected Value of a Continuous Random Variable

- If X is a continuous random variable with density curve $f(x)$, the **expected value** the **mean** of X is defined as the integral:

$$\mu_x = E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

- This is also known as a weighted average.

16

16

SJSU SAN JOSÉ STATE UNIVERSITY

Variance of a Continuous Random Variable

- The variance of a continuous random variable X is defined as the integral:

$$\sigma_X{}^2 = Var(X) = \int_{-\infty}^{\infty} (x - \mu_x)^2 \, f(x)\, dx = E(X^2) - (E(X))^2$$

- This measures the spread of the values of X.

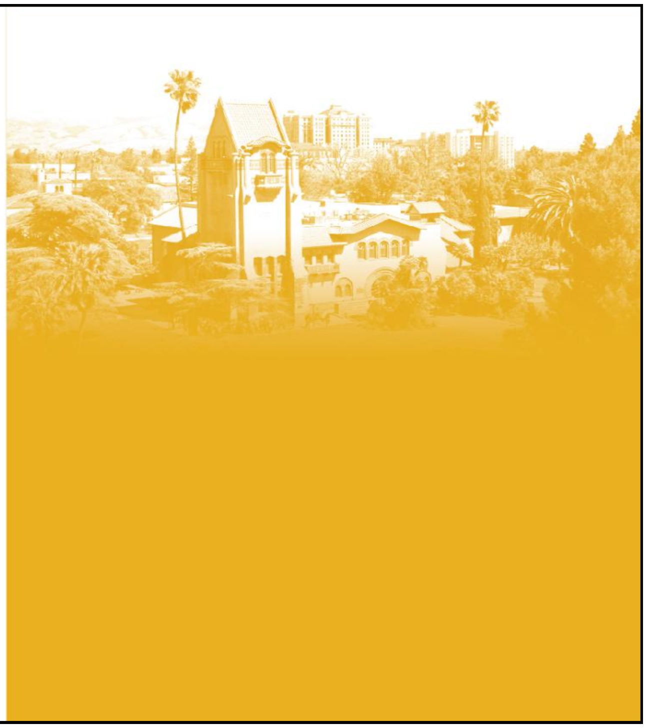
- The standard deviation of continuous random variable X is the square root of the variance:

17

17

SJSU SAN JOSÉ STATE UNIVERSITY

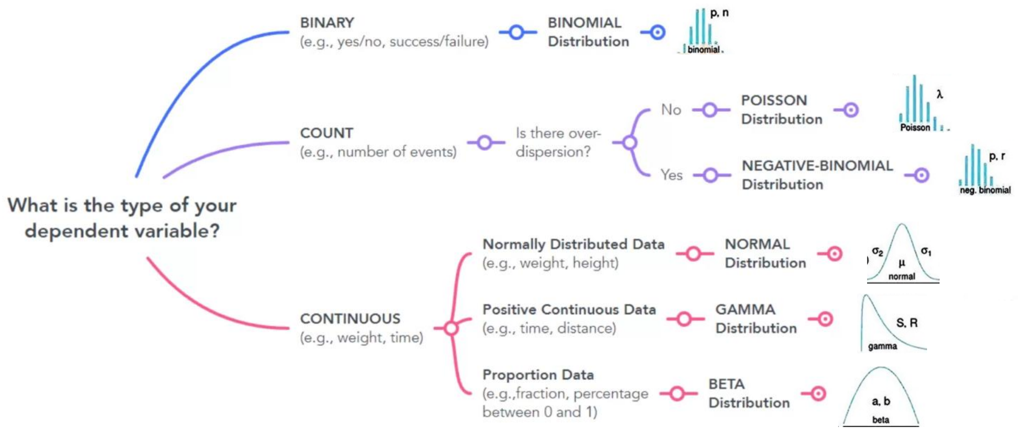
Distributions for Continuous Random Variables



18

SJSU SAN JOSÉ STATE UNIVERSITY

Random Variables & Distributions



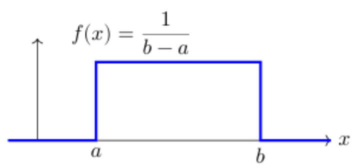
https://statisticseasily.com/generalized-linear-model-distribution-and-link-function/

19

SJSU SAN JOSÉ STATE UNIVERSITY

The Uniform Distribution

- The Uniform distribution $X \sim U(a, b)$ is a continuous probability distribution where all values within a given interval are equally likely.



- Mean (Expected Value):

$$E(X) = \frac{a+b}{2}$$

- Variance:

$$Var(X) = \frac{(b-a)^2}{12}$$

20

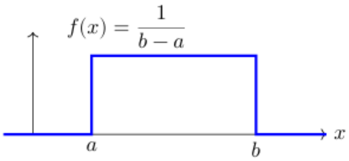
20

SJSU SAN JOSÉ STATE UNIVERSITY

PDF and CDF of the Uniform Distribution

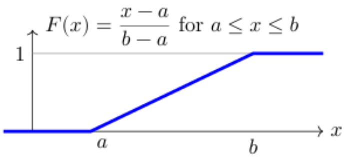
- The PDF of the Uniform Distribution $X \sim U(a,b)$ is given by:

$$f(x) = \frac{1}{b-a}, \qquad a \le x \le b$$



- This PDF shows that each value in $[a,b]$ is equally likely.
- The CDF of the Uniform Distribution $X \sim U(a,b)$ is given by:

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1, & x > b \end{cases}$$



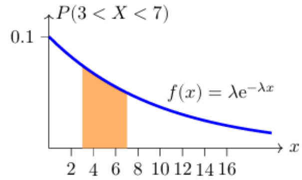
SJSU SAN JOSÉ STATE UNIVERSITY

Applications of the Uniform Distribution

- Random Number Generation: Used in computer simulations and Monte Carlo methods to generate random samples.
- Fair Lottery or Raffles: Assigning equal probabilities to participants in a drawing.
- Waiting Time in Randomized Algorithms: Used in randomized backoff protocols in networking.
- Sensor Noise Modeling: Used in engineering applications where measurement noise is assumed to be uniformly distributed.

The Exponential Distribution

The Exponential distribution $X \sim \exp(\lambda)$ models the time until an event occurs (e.g., failure time, waiting times).



- Range: $[0, \infty)$
- It's defined by <u>one</u> parameter:
  - Rate parameter λ
- It's a "memoryless" model - past events don't affect future probabilities:

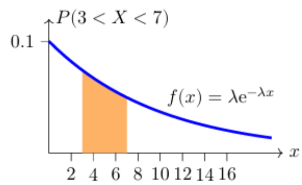
$$P(X > s + t \mid X > s) = P(X > t)$$

PDF and CDF of the Exponential Distribution

The probability density function (PDF) of an Exponential random variable X is:

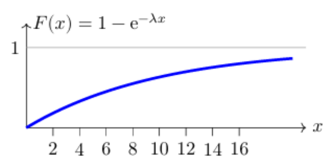
$$f(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \geq 0 \end{cases}$$

λ – rate



The cumulative density function (CDF) of an Exponential random variable X is:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} \lambda e^{-\lambda t}\, dt = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$



Mean and Variance of the Exponential Distribution

- Mean (Expected Value):

$$E(X) = \frac{1}{\lambda}$$ expected value or mean of X

- Variance

$$Var(X) = \frac{1}{\lambda^2}$$ the spread of X

Applications of the Exponential Distribution

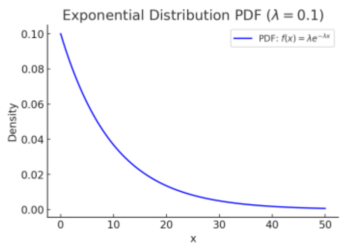
- Reliability Engineering: Models the failure time of electronic components and machinery.
- Queuing Theory: Models inter-arrival times of customers in a queue (e.g., call centers, banks).
- Radioactive Decay: Describes the time between decay events of radioactive particles.
- Poisson Processes: Used in modeling the time between rare events like earthquakes or accidents.

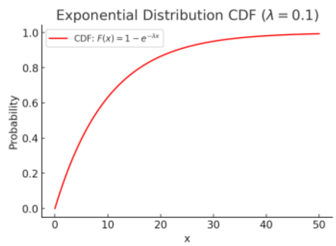
Example: Phone Booth Wait Times

Suppose that the length of a phone call in minutes is an exponential random variable with parameter λ = 1/10 . If someone arrives immediately ahead of you at a public telephone booth, find the probability that you will have to wait:

(a) more than 10 minutes;

(b) between 10 and 20 minutes.



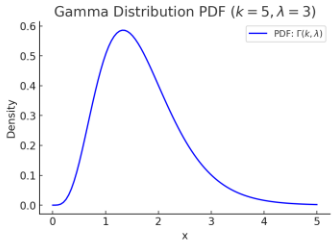


27

The Gamma Distribution

The Gamma distribution $X \sim \Gamma(k, \lambda)$ is a generalization of the exponential distribution for multiple independent exponential random variables.

- Range: $[0, \infty)$
- It's defined by two parameter:
  - Rate parameter λ
  - # of independent exponential random variables k
- It reduces to exponential distribution if $k = 1$



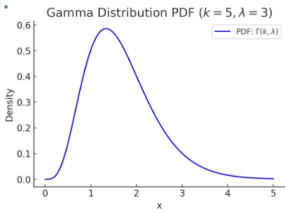
28

SJSU SAN JOSÉ STATE UNIVERSITY

PDF and CDF of the Gamma Distribution

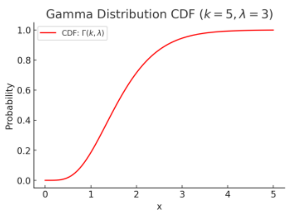
The probability density function (PDF) of a Gamma random variable X is:

$$f(x) = \frac{\lambda^k x^k e^{-\lambda x}}{\Gamma(k)}, \qquad x > 0$$



The cumulative density function (CDF) of a Gamma random variable X is:

$$F(x) = P(X \leq x) = \frac{1}{\Gamma(k)} \int_0^x \lambda^k t^k e^{-\lambda t}\, dt$$



SJSU SAN JOSÉ STATE UNIVERSITY

Mean and Variance of the Gamma Distribution

- Mean (Expected Value):

$$E(X) = \frac{k}{\lambda}$$ expected value or mean of X

- Variance

$$Var(X) = \frac{k}{\lambda^2}$$ the spread of X

SJSU SAN JOSÉ STATE UNIVERSITY

Applications of the Gamma Distribution

- Insurance Risk Modeling: Used to model claim sizes in actuarial science.
- Queue Waiting Times: Models the waiting time for multiple events to occur (e.g., service systems).
- Biological Processes: Models the distribution of blood clotting times or rainfall accumulation over time.
- Bayesian Statistics: Often used as a conjugate prior for the Poisson distribution in Bayesian inference.

SJSU SAN JOSÉ STATE UNIVERSITY

Example: Call Center

A call center receives an average of 2 calls per minute. Assuming the waiting time until 5 calls have arrived follows a Gamma distribution:

(a) What are the parameters of the Gamma distribution?

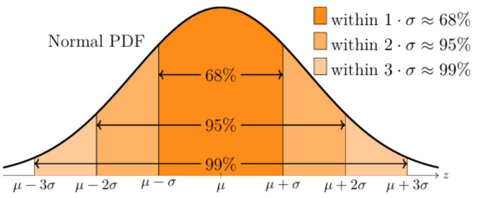
(b) What is the probability that the waiting time until 5 calls is more than 3 minutes?

SJSU SAN JOSÉ STATE UNIVERSITY

The Normal Distribution

The Normal or Gaussian distribution $N(\mu, \sigma^2)$ models many things. It's one of the most fundamental probability distributions in statistics and is widely used due to the Central Limit Theorem (CLT) ➔ sum of a large # of independent random variables tends to follow a normal distribution.

- Range: $(-\infty, \infty)$
- It's defined by two parameter:
  - Expected (mean) μ
  - Standard Deviation σ or Variance σ^2
- It's a bell-shaped curve.
- The 68-95-99.7 Rule (Empirical Rule for standard deviations).



SJSU SAN JOSÉ STATE UNIVERSITY

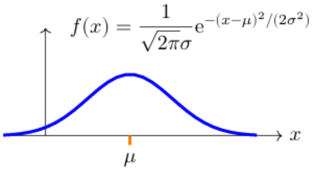
PDF and CDF of the Normal Distribution

The probability density function (PDF) of a Normal random variable X is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

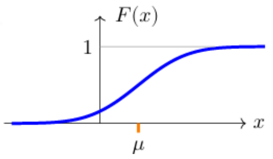
σ – standard deviation

μ – mean



The cumulative density function (CDF) of a Normal random variable X is:

$$F(x) = P(X \le x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$



SJSU SAN JOSÉ STATE UNIVERSITY

Mean and Variance of the Normal Distribution

- Mean (Expected Value):

$$E(X) = \mu$$ expected value or mean of X

- Variance

$$Var(X) = \sigma^2$$ the spread of X

SJSU SAN JOSÉ STATE UNIVERSITY

Applications of the Normal Distribution

- Human Traits Measurement: Heights, weights, IQ scores, and other biological characteristics follow a normal distribution.
- Stock Market Returns: Many financial models assume returns on assets are normally distributed.
- Measurement Errors: Used in quality control and experimental science to model errors.
- Central Limit Theorem Applications: In inferential statistics, sample means of independent variables tend to be normally distributed.

SJSU SAN JOSÉ STATE UNIVERSITY

Example: Exam Scores

The scores on a university entrance exam follow a normal distribution with a mean of 70 and a standard deviation of 10.

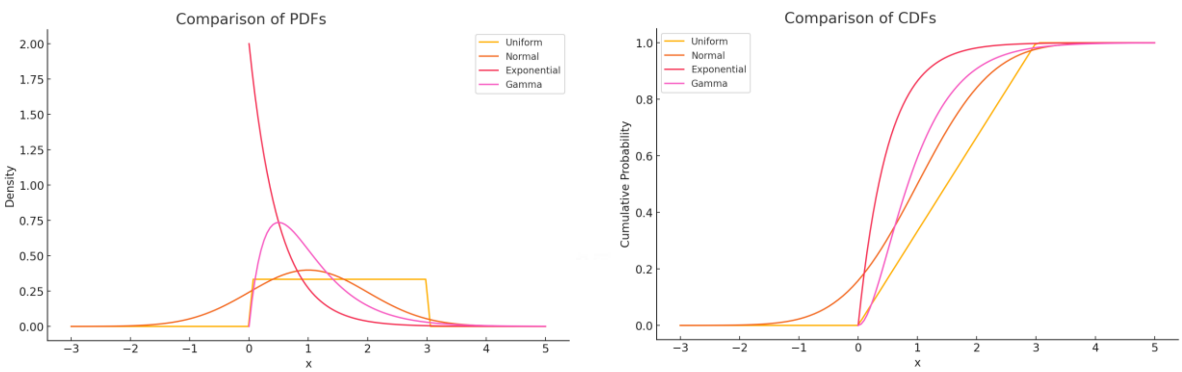
a) What proportion of students score above 85?

b) What proportion of students score between 60 and 80?

b) What is the 90th percentile score?

SJSU SAN JOSÉ STATE UNIVERSITY

Comparison of the Continuous Distributions



Summary Table of the 4 Continuous Distributions

Distribution	Definition	Range	Parameters	Memoryless
Uniform	Models all values within a given interval are equally likely	[a, b]	a, b	❌ No
Exponential	Models the time until an event occurs	$[0, \infty)$	λ	✅ Yes
Gamma	Models multiple independent exponential random variables	$[0, \infty)$	λ, k	❌ No
Normal	Models phenomena that depend on multiple independent factors	$(-\infty, \infty)$	μ, σ^2	❌ No

Summary Table of the 4 Continuous Distributions

Distribution	Parameters	PDF	CDF	Mean	Variance
Uniform	a, b	$f(x) = \frac{1}{b-a}$	$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1, & x > b \end{cases}$	$E(X) = \frac{a+b}{2}$	$Var(X) = \frac{(b-a)^2}{12}$
Exponential	λ	$f(x) = \lambda e^{-\lambda x}$	$F(x) = 1 - e^{-\lambda x}$	$E(X) = \frac{1}{\lambda}$	$Var(X) = \frac{1}{\lambda^2}$
Gamma	λ, k	$f(x) = \frac{\lambda^k x^k e^{-\lambda x}}{\Gamma(k)}$	$F(x) = \frac{1}{\Gamma(k)} \int_0^x \lambda^k t^k e^{-\lambda t}\, dt$	$E(X) = \frac{k}{\lambda}$	$Var(X) = \frac{k}{\lambda^2}$
Normal	μ, σ^2	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\, dt$	$E(X) = \mu$	$Var(X) = \sigma^2$

SJSU SAN JOSÉ STATE UNIVERSITY

Standard Distribution

41

SJSU SAN JOSÉ STATE UNIVERSITY

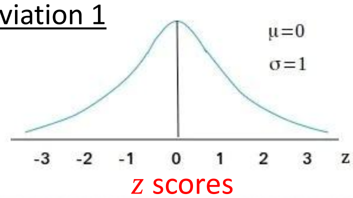
The Standard Normal Distribution

- Normally, you do **NOT** need to use integral calculus to compute those probabilities.
- First, convert your normal distribution of x values with mean μ and standard deviation σ to the standard normal distribution of z values.

$$z = \frac{x - \mu}{\sigma}$$

- The standard normal distribution has mean 0 and standard deviation 1

The standard normal distribution curve



42

42

SJSU SAN JOSÉ STATE UNIVERSITY

The Standard Normal Distribution

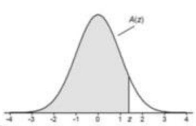
- The z values of a standard normal distribution are called z scores.
- For a standard normal distribution table, see

  https://www.math.arizona.edu/~rsims/ma464/standardnormaltable.pdf

SJSU SAN JOSÉ STATE UNIVERSITY

Standard Normal Distribution Probabilities

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
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.93 19
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817

Cumulative Distribution Function $F(x)$

The probability that a random value will be less than or equal to x.

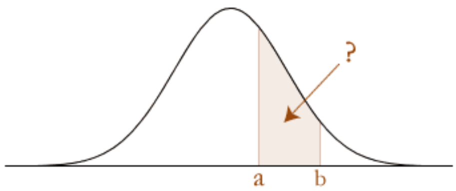
Therefore, for the standard normal distribution:

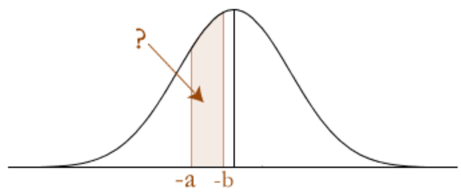
$$F(x) = P(z \le x)$$

Example: $F(0.35) = ???$

$\boldsymbol{P(z \le 0.35) = 0.6368}$

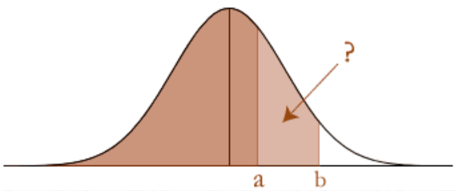
Standard Normal Distribution Probabilities (cont)

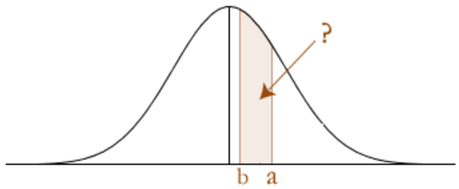




$P(a < z < b) = P(z < b) - P(z < a)$

$P(-a < z < -b) = P(b < z < a) = P(z < a) - P(z < b)$





Example: Standard Distribution

How do you compute the probability between a and b using standard z-tables?