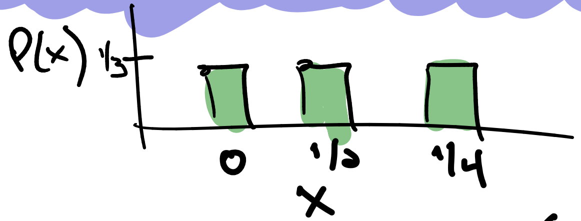
CS2810 Day 16

Mar 22

Admin:
Quiz2 grades release thurs (quiz2 mean .830 vs quiz 1 mean .832)

Content:
Continuous Distributions
Normal Distributions
Cumulative Distribution Functions
Central Limit Theorem

Discrete Distribution vs Continuous Distribution

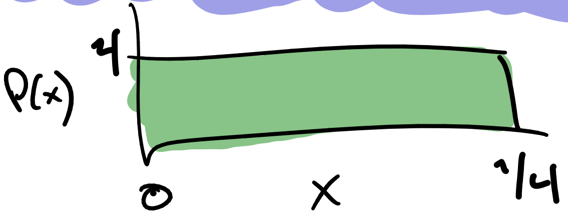


X has sample space $\{0, 1/2, 1/4\}$

$$\sum_i P(x) = 1$$

$$0 \le P(X=x) \le 1$$

Prob Mass Functions
PMF



X takes any value from 0 to 1/4

$$\int_{-\infty}^{\infty} P(x)\,dx = 1$$

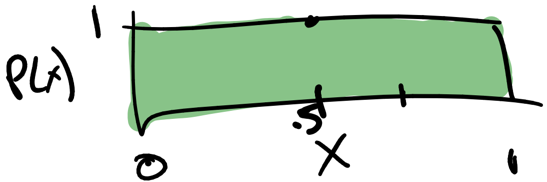
$$0 \le P(x)$$

Prob Density Function
PDF

THE $P(x) = X$ CONONDRUM

LET X BE A CONTINUOUS RANDOM VARIBLE ON [0,1]

(INCLUDES 0,1)



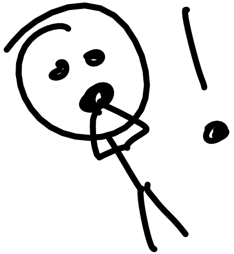
WHAT IS PROB THAT X IS EXACTLY .5?

$P(X=.5)=1$ x is always .5 (no prob left for any other outcome)

$P(X=.5)>0$ given there is an infinite amount of outcomes, each outcome can't have a positive probability (no number theory dive)

THE $P(x) = X$ CONONDRUM

$$P(x = 1/2) = 0$$



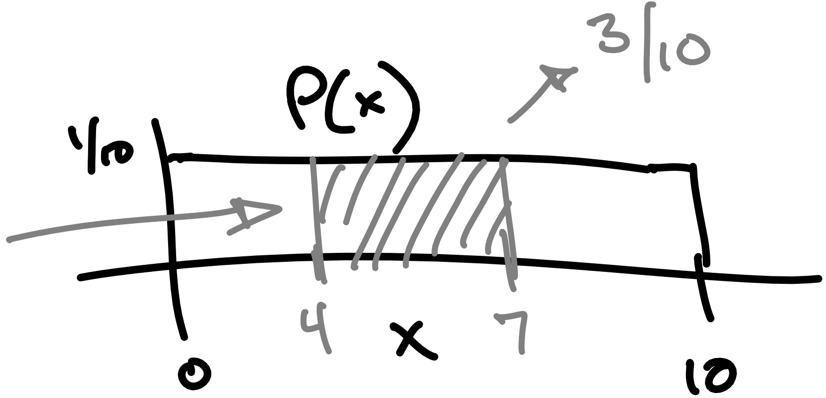
PROBABILITY OF OUTCOMES IN CONTINUOUS DISTRIBUTION CAN BE ZERO

SO WHATS PDF GOOD FOR THEN?

CONTINUOUS DISTRIBUTION

Continuous Distribution's Prob Distribution Function gives the probability an outcome falls in some range via integration:

GIVEN X HAS PDF



COMPUTE

AREA UNDER PDF FROM 4 TO 7

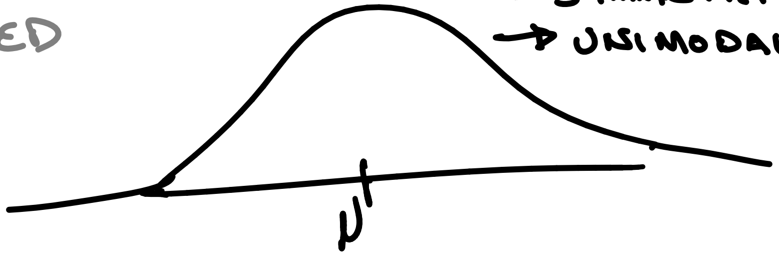
$$P(4 < x < 7) = \int_4^7 1/10 \, dx = \frac{x}{10}\Big|_4^7 = \frac{7}{10} - \frac{4}{10} = \frac{3}{10}$$

NORMAL | GAUSSIAN DISTRIBUTION

→ CONTINUOUS
→ UNBOUNDED
→ SYMMETRIC
→ UNIMODAL

X IS R.V. NORMALLY DISTRIBUTED

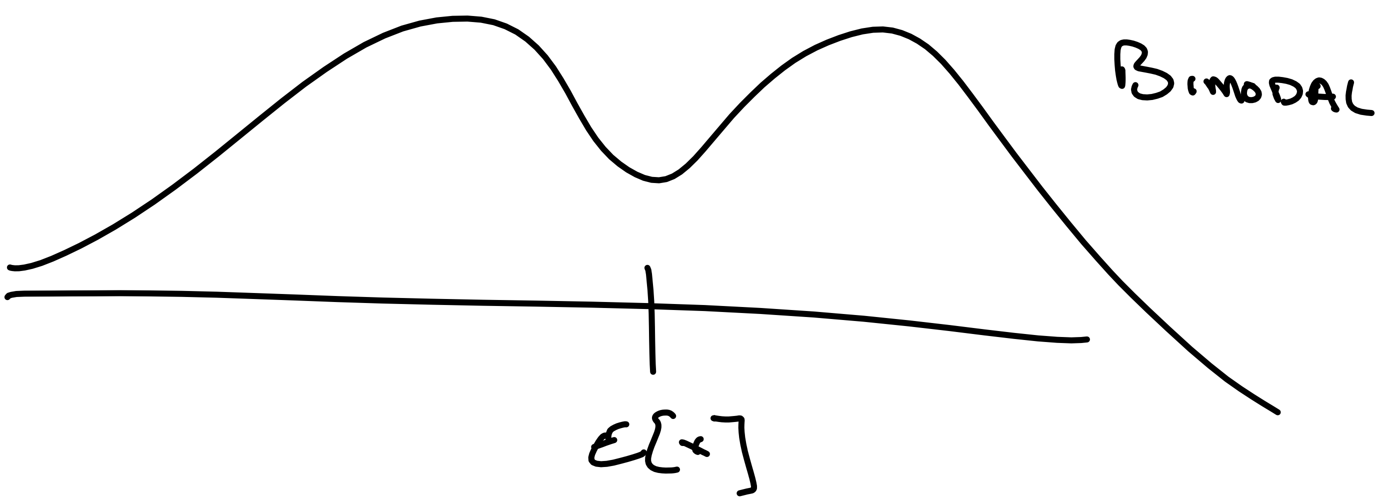
$$X \sim N(\mu, \sigma^2)$$



$$\Rightarrow P(X=x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2 \frac{(x-\mu)^2}{\sigma^2}}$$

$$E[X] = \mu$$

$$VAR(X) = \sigma^2$$


BIMODAL
E[x]

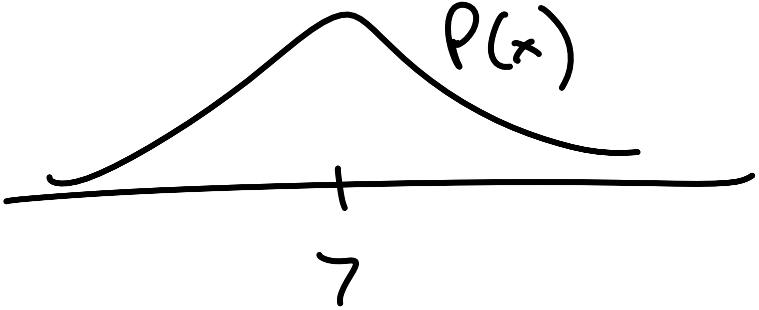
$X \sim N(7, 11)$

$\sigma^2 = VAR(X)$

$\mu = E[X]$

GIVEN $X \sim N(7, 11)$ COMPUTE PROB THAT

X IS BETWEEN 6 AND 10



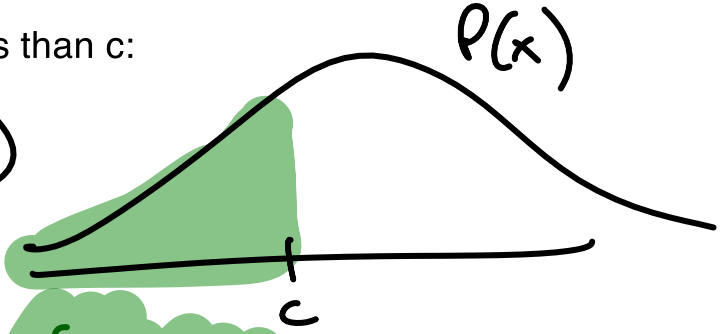
$$P(7 < X < 11)$$

$$= \int_{7}^{11} \frac{1}{\sqrt{11}\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-7)^2}{11}}$$

CUMULATIVE DISTRIBUTION FUNCTIONS

The CDF gives the probability that an outcome is less than c:

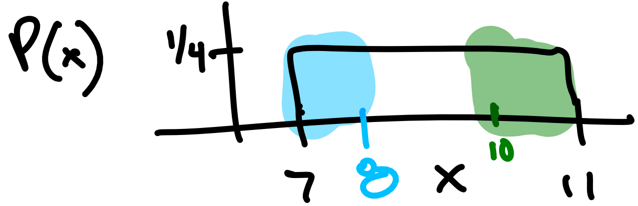
GIVEN RV x WITH PDF $P(x)$



$$CDF(c) = P(x \leq c) = \int_{-\infty}^{c} P(x)\,dx$$

CDF Demo

ICA



Given the PDF to the right, compute:

1. the probability that X is between 10 and 11
2. the probability X is less than 8
3. Express your answers for each of the above using only the CDF (not the PDF)

(+) As c gets smaller and smaller, describe the behavior of the CDF
(+) As c gets larger and larger, describe the behavior of the CDF

① $1/4$ AREA UNDER PDF FROM 10 TO 11

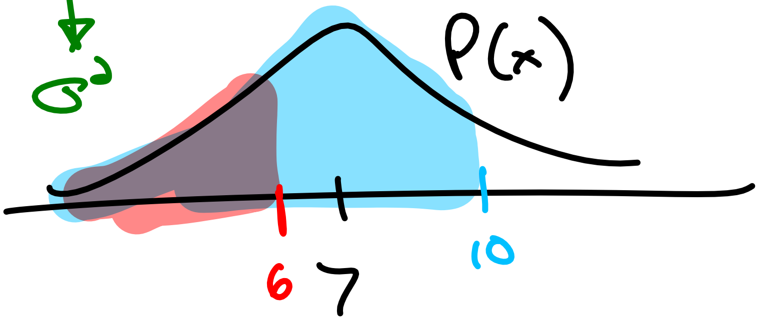
$CDF(11) - CDF(10)$

② $1/4$ AREA UNDER PDF FROM 7 TO 8

$CDF(8) - CDF(7) = CDF(8)$

GIVEN $X \sim N(\underset{\mu}{7}, \underset{\sigma^2}{11})$ COMPUTE PROB THAT

X IS BETWEEN 6 AND 10

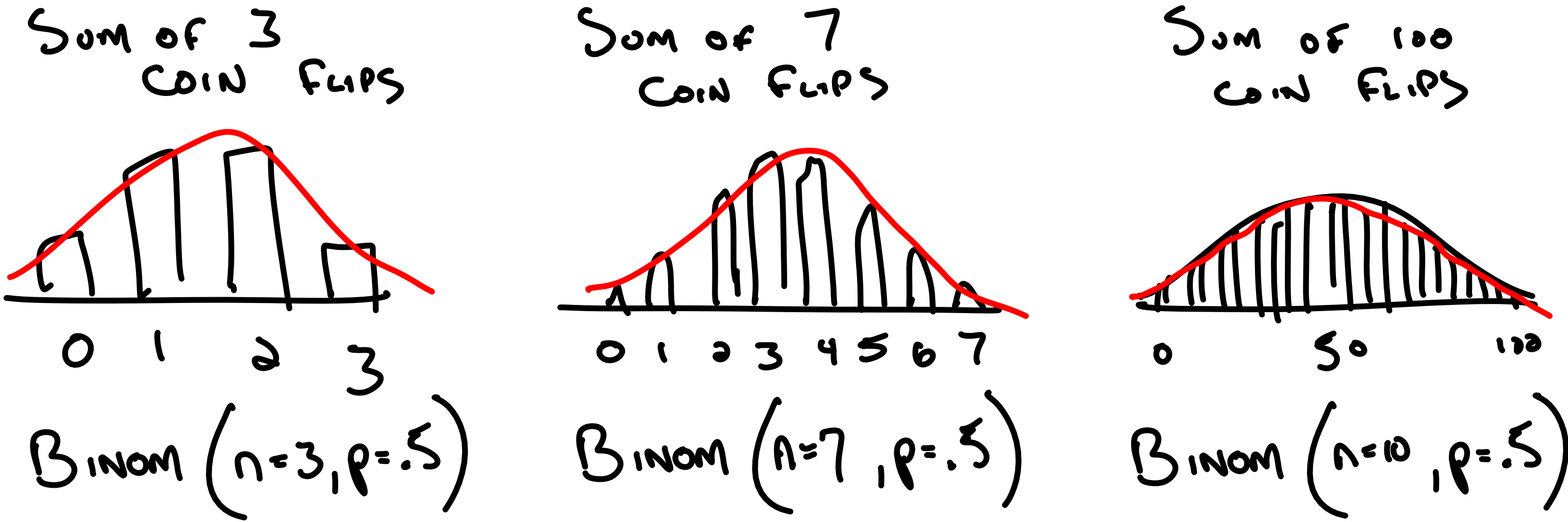


Demo: Updated Prob / Stats Calculator

$$\mathrm{CDF}(10) - \mathrm{CDF}(6) \approx .435$$

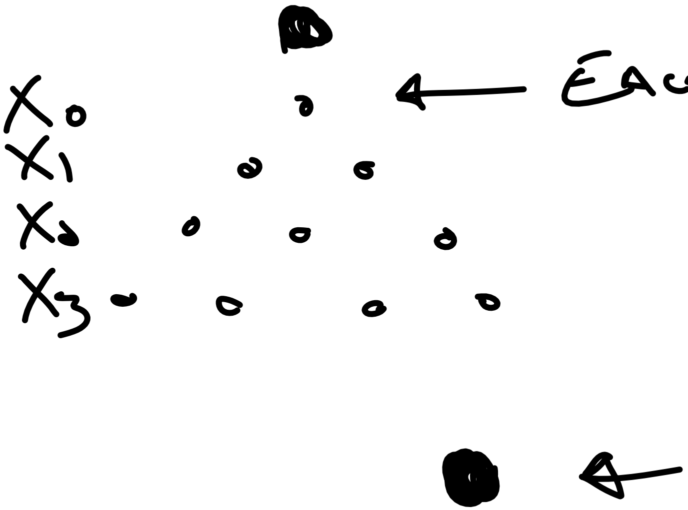
Central Limit Theorem

As we add more independent random variables together, the resulting sum gets closer and closer to normally distributed.



Demo: CLT

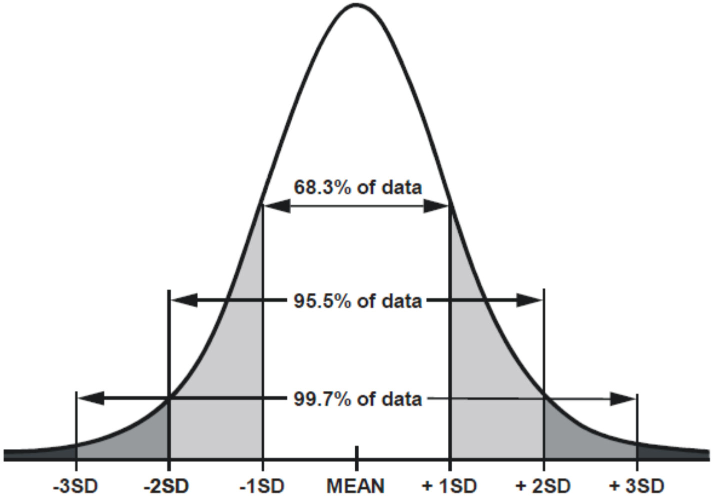
Galton Board on YouTube



EACH PEG IS A BINARY RANDOM VARIABLE X_i

FINAL POSITION OF BALL IS SOM OF X_i

NORMAL DISTRIBUTIONS USEFUL LANDMARKS



95% OF DATA LIES IN 1.96 STD DEVIATIONS

Why is the normal distribution so popular?

1. We often sum independent random variables

    CLT tells us this sum is (roughly) normally distributed

2. Linear Functions of Normal Random Variables are also Normal!

    And the linearity of expectation formulae can tell us the mean and variance of the sum!

ICA 2

$X_0 + X_1 + X_2 + \dots + X_{399}$

The following distribution gives the prices of ice creams bought at a shop.

X_0

x	\$1	\$2	\$10
$P(x)$	.5	.4	.1

What is the probability that the shop sells more than \$1000 of ice cream to 400 customers?
(make any assumptions you deem necessary)

→ EACH CUSTOMER PURCHASES INDEPENDENTLY

→ ASSUME TOTAL ICE CREAM SOLD IS NORMAL (CLT)

x	$1	$2	$10
P(x)	.5	.4	.1

$$E[x] = 1 \cdot .5 + 2 \cdot .4 + 10 \cdot .1$$
$$= .5 + .8 + 1 = 2.3$$

$$E[x^2] = 1^2 \cdot .5 + 2^2 \cdot .4 + 10^2 \cdot .1$$
$$= .5 + 1.6 + 10 = 12.1$$

$$VAR(x) = E[x^2] - E[x]^2$$
$$= 12.1 - 2.3^2$$

$$E\left[X_0 + X_1 + X_2 + \dots + X_{399}\right] = E[X_0] + E[X_1] + \dots + E[X_{399}]$$

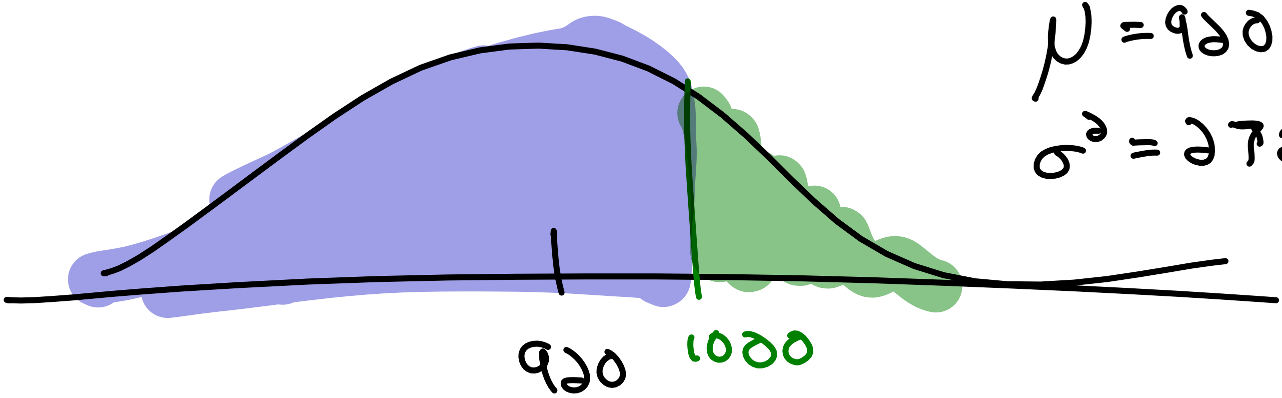
$$= 400 \cdot E[X] = 920$$

$$VAR\left(X_0 + X_1 + X_2 + \dots + X_{399}\right) = VAR(X_0) + VAR(X_1) + \dots$$

$$= 400\ VAR(X) \cong 2724$$

$\mu = 920$

$\sigma^2 = 2704$



$1 - \text{CDF}(1000)$

INVERSE OF CDF IS

PPF (AREA) = C WHERE

PERCENT POINT FUNCTION

$$VAR(X_0 + X_1 + X_2 + \dots + X_{399}) \neq VAR(400X)$$

SUM OF 400 CUSTOMER'S BILLS

1 CUSTOMER'S BILL × 400