Lecture 2 & 3. Basic Concepts and Review of Probability

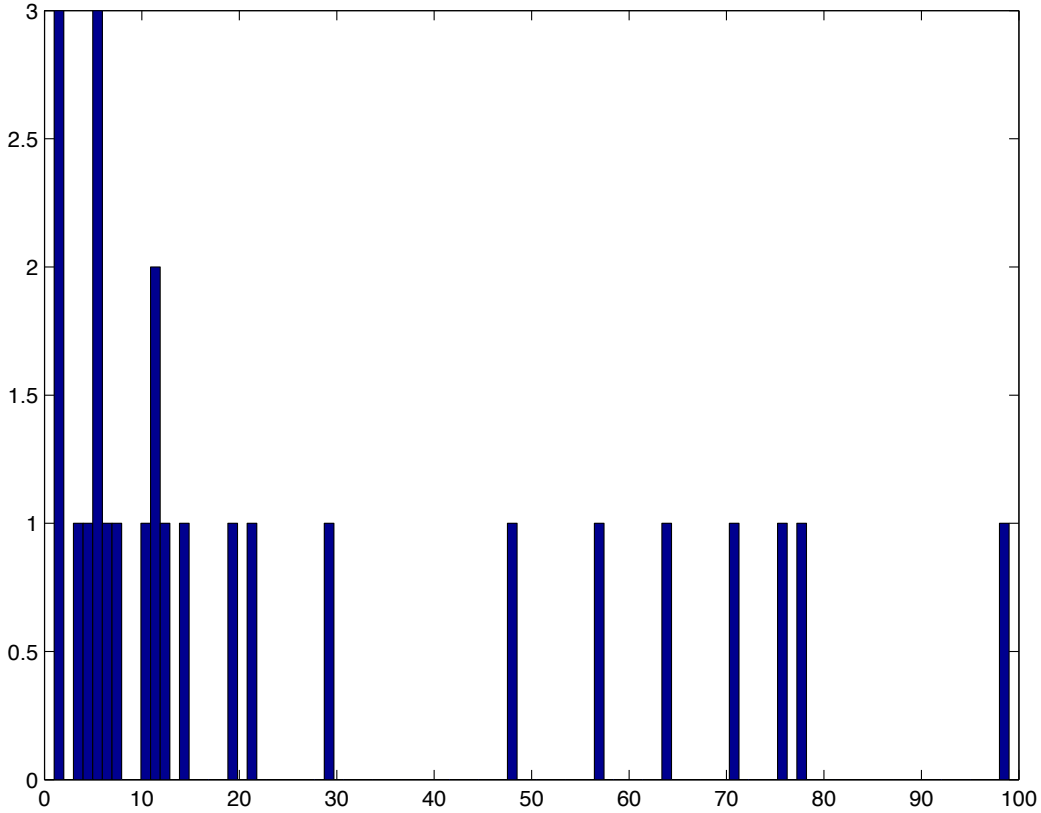
- Logistics
- Basic concepts
- Review of probability
- Use of R

- TA: Caglar Caglayan, office hour: W 12:00-1:00pm, Main Building 321

- Grading
  - Class Attendance 3%
  - Submitting Teaching Evaluation 2%
  - Homework - 15%
  - Computer Example 1 - 10%, Computer Exam 2 - 10%
  - Midterm 1 - 15%, Midterm 2 - 15%
  - Final - 30%

- Homework grades: $\{0, 1, 2\}$. Homeworks are aimed for practice. Students are responsible to check details with solutions.

Winning the number guessing



Winner: Mallory Weaver, Number: 3

List of movies

- Fight club
- Harry Potter and Hunger Games
- Game of thrones
- Dirty Dancing
- The Shawshank Redemption
- Coach Carter
- Money Ball
- 500 Days of Summer
- The Catcher in the Rye
- The Prestige
- The silence of the Lambs
- The Royal Tenenbaums
- The Shack, 1984
- Fahrenheit 451
- The Girl with the Dragon Tattoo
- Crazy Stupid Love
- Jurassic Park
- The Great Gatsby
- Forrest Gump

- The Giver
- Forrest Gump
- Howl's Moving Castle
- The Truman Show
- Equilibrium
- The Sherlock
- Infinite Jest

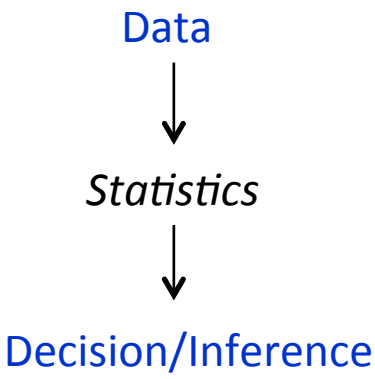
Statistics

The field of Statistics deals with the *collection*, *presentation*, *analysis*, and *use* of **data** to model systems, make decisions, solve problems, and design products and processes.

Statistics is the science of data

Examples: Statistics helps us in

- sports (ref: movie ``Money Ball'')
- stock market/finance
- weather forecast
- machine learning/computer/internet
- politics
- biology/medicine



Where data come from

- Retrospective study using historical data

e.g. historical temperature record of Atlanta

- Observational study

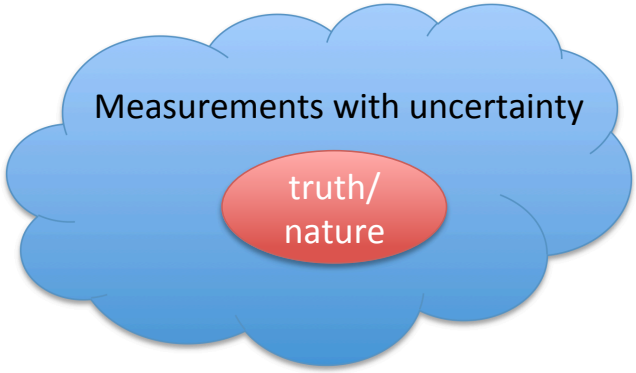
e.g. measurements of digital thermometer

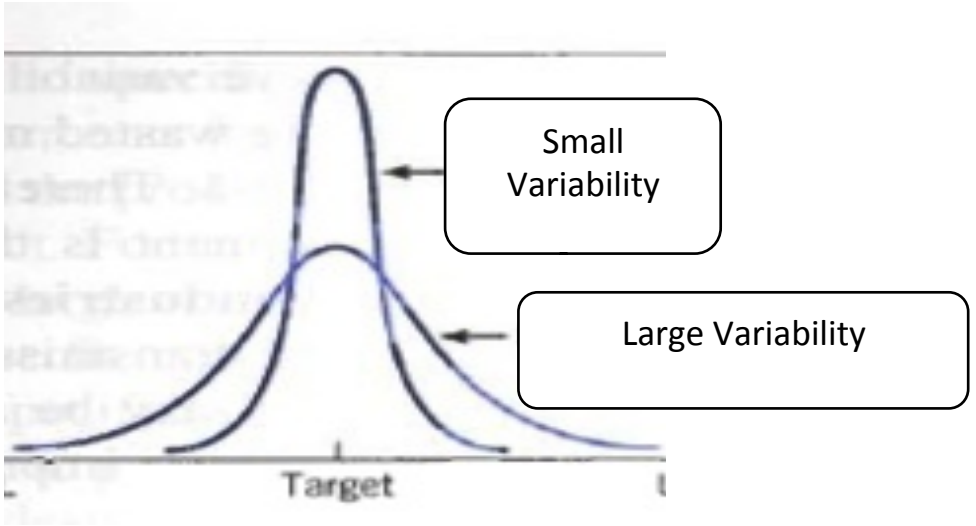
- A designed experiments

e.g. to test if a drug is effective

Statistical Methods

- Statistical methods are useful for describing and understanding variability.
- By variability, we mean successive observations of a system or phenomenon do *not* produce exactly the same result.
- Statistics gives us a framework for describing this variability and for learning about potential sources of variability.





Statistical Methods

- Point estimator
- Confidence Interval
- Hypothesis test
- Two sample / ANOVA
- Linear regression

Statistical Methods

- Point estimator

  e.g. estimating temperature from 5 digital thermometer measurements

- Confidence Interval

  e.g. giving an interval where true temperature is mostly likely to be within

- Hypothesis test

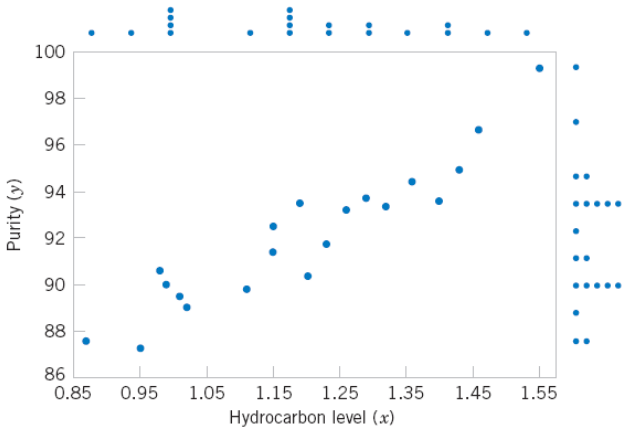
  e.g. testing good/bad the quality of a batch of laptops.

- Two sample test / ANOVA

  e.g. test effectiveness of a new drug

- Linear regression

  e.g. fitting a linear model from input/output measurements



Statistical Methods

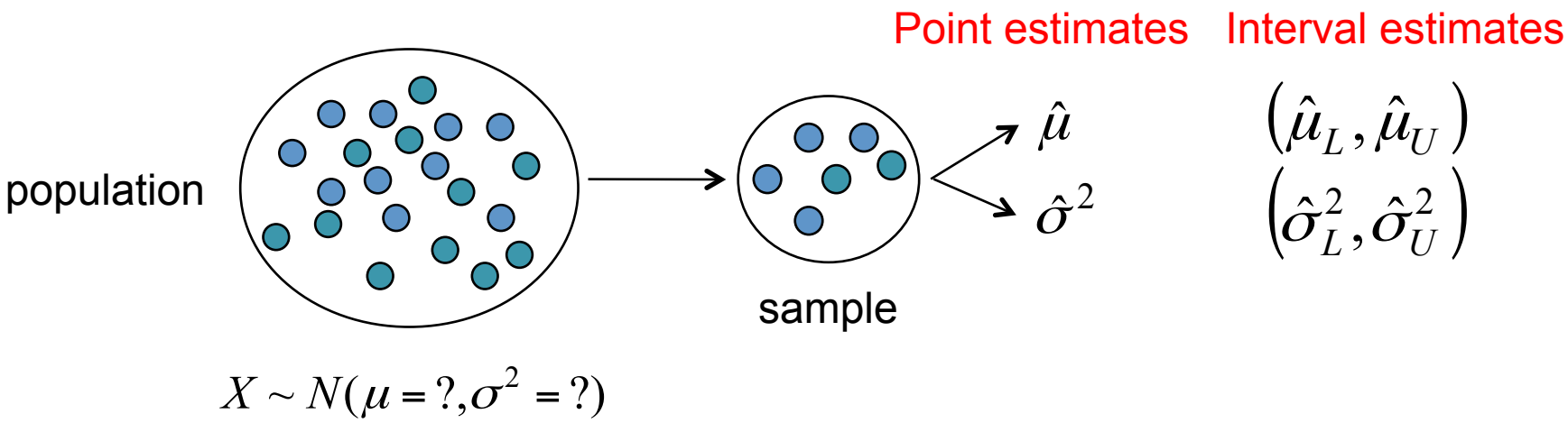
- Point estimator

estimate parameters of a distribution based on a random sample

e.g. estimating temperature from 5 digital thermometer measurements

- Confidence Interval

e.g. giving an interval where true temperature is mostly likely to be within



Statistical Methods

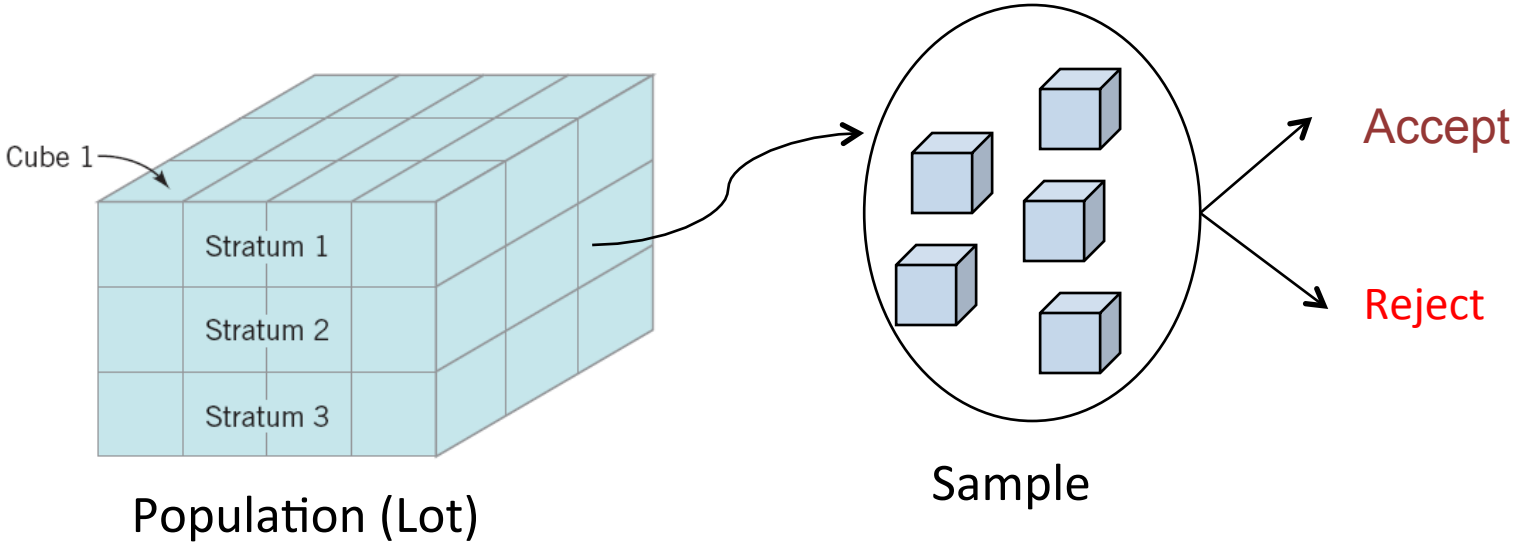
- Hypothesis test

Make a decision about a population based on a random sample

e.g. testing good/bad the quality of a batch of laptops.

- Two sample test / ANOVA

e.g. test effectiveness of a new drug

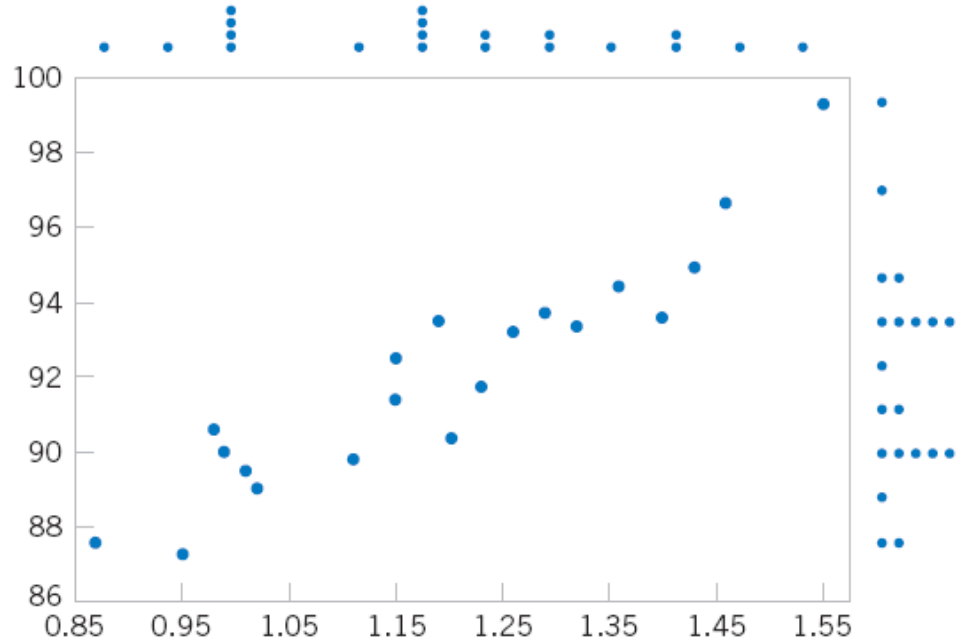


Statistical Methods

- Linear regression

Predict a response variable based on one or more predictor variables

e.g. fitting a linear model from input/output measurements, say voltage and current of a resistor



Role of Engineers

An engineer is someone who solves problems of interest to society by the efficient application of scientific principles through

- Refining existing methods or products
- Designing new methods or products

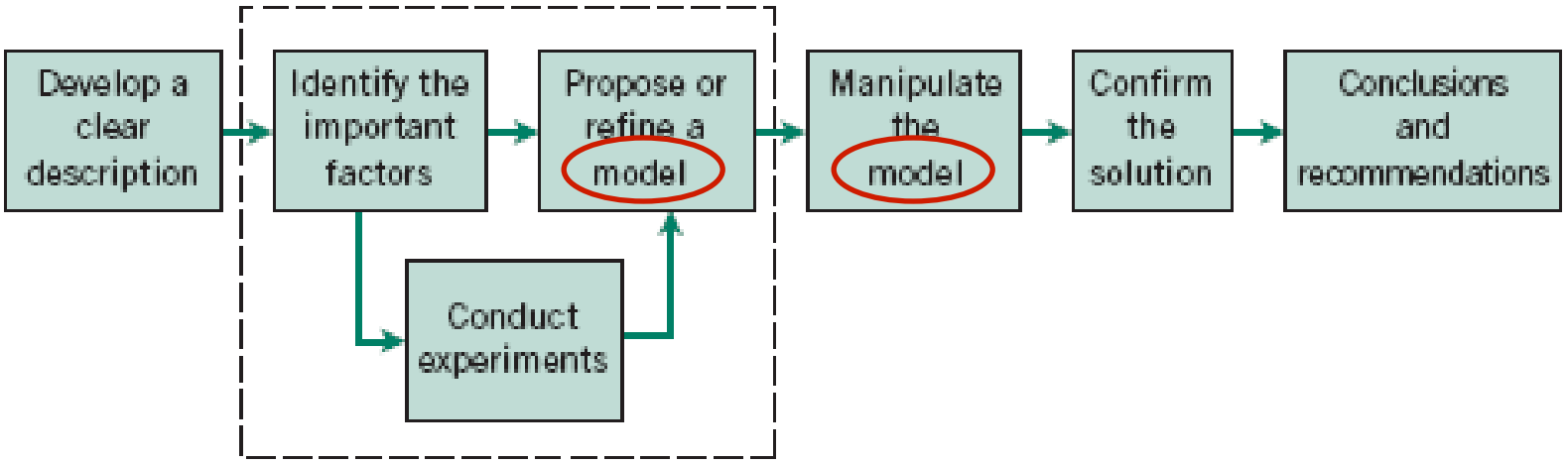


Figure 1.1. The engineering method

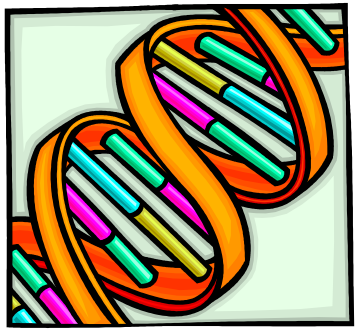
Model

Definition 1: (Merriam-Webster dictionary)

something (as a similar object or a construct) used to help visualize or explore something else (as the living human body) that cannot be directly observed or experimented on.

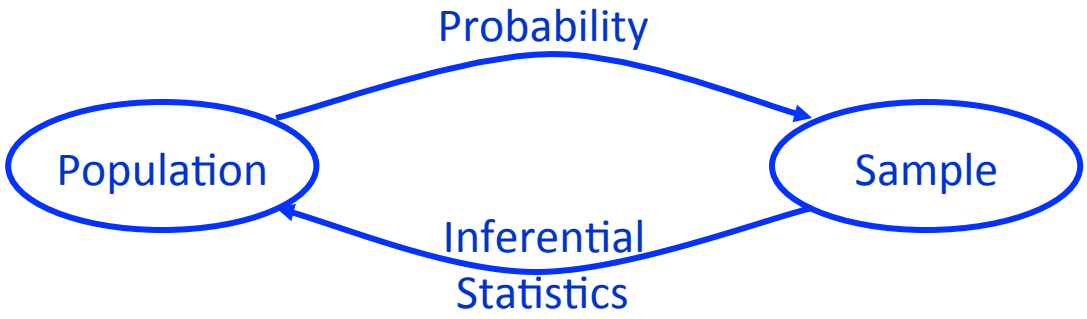
Definition 2: (Law and Kelton)

A representation of the system that is studied as a surrogate of the actual system



A. Law and D. Kelton, Simulation Modeling and Analysis, 3/e, McGraw-Hill, 2000

Probability Vs. Statistics



Probability: given the information in the pail, what is in your hand?

Statistics: given the information in your hand, what is in the pail?

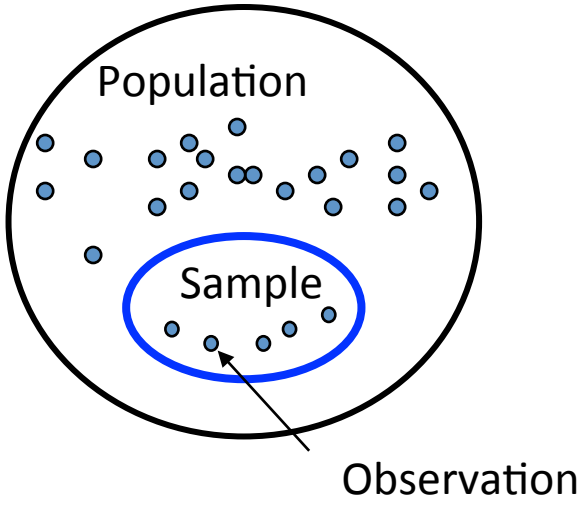
Review of Probability

- Key words
  - Population, Random Sample, Sample Space
  - Outcomes, Outcome Probability
  - Events, Compliments, Intersections, Unions
  - Conditional Probability, Independence
  - Law of Total Probability, Bayes Theorem

Population Vs. Sample

- Population: a finite well-defined group of ALL objects which, although possibly large, can be enumerated in theory
  (e.g. Atlanta population, Georgia Tech students).

- Sample: A sample is a SUBSET of a population
  (e.g. select 50 out of 1,000 bearings manufactured today).

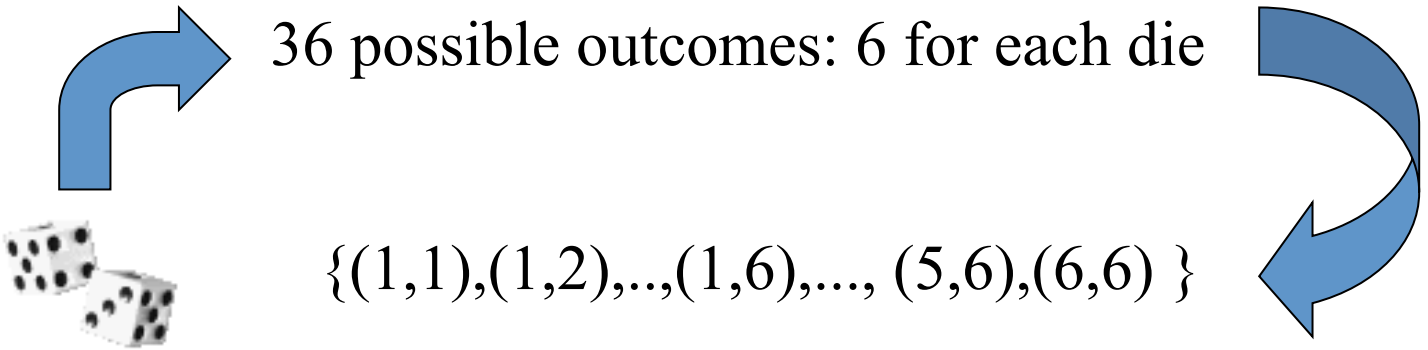


Other terms

- Random Sample = randomly selected sample
- Sample Size = number of object in the sample
- Experiment = measure characteristics of the sample/ population
- Outcome = possible values of the measurements
- Sample Space = space of all possible outcomes

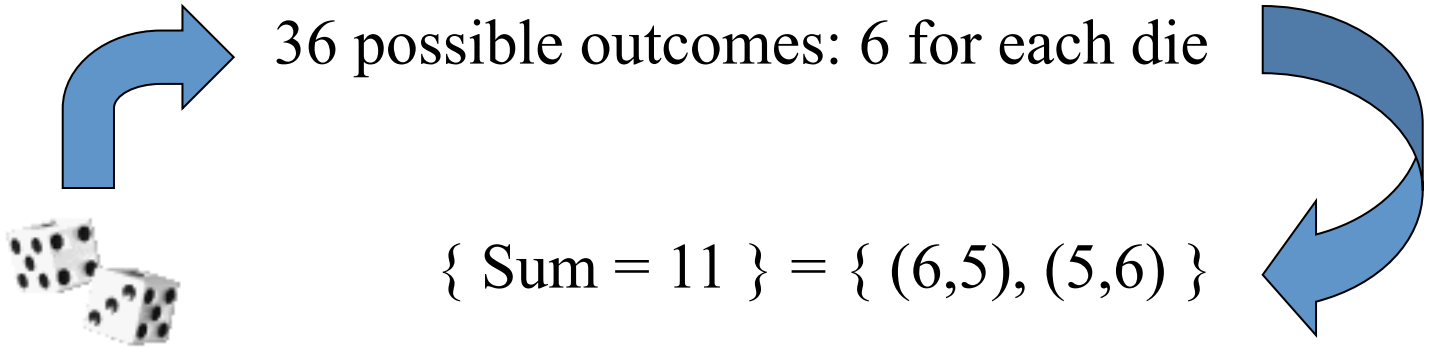
Example

- Experiment: Roll two dies simultaneously
- Sample space: {(1,1),(1,2),..,(1,6),..., (5,6), (6,6) }
- Outcome: (6,6)
- What is the population?



Outcomes and Events

- An Event is any collection of sample outcomes
- Simple Event = Collection of outcomes
- Compound Event = Collection of outcomes described as unions or intersections of other events



Outcome probability

- Probability = likelihood of the occurrence of an outcome in the sample space

$$S = \{o_1, \ldots, o_n\} \rightarrow \{p_1, \ldots, p_n\}$$

$$P(o_i) = p_i$$

$$0 \leq \text{probability} \leq 1$$

Example

Example: Roll a die twice.

1. The number of outcomes when at least one ‘6’ occurs in the combination of 2 dices:

A. 1 outcome **B. 6 outcomes** **C. 11 outcomes**

2. The probability for the event that {at least one ‘6’ occurs in the combination of 2 dices}:

A. 1/6 **B. 11/36** **C. 1/36**

Unions and Intersections

P(AUB) =P(Outcome contained in Event A OR Event B)

P(AB) = P(Outcome contained in both Event A AND Event B)

$$AB = A \bigcap B$$

$$P(AUB) = P(A) + P(B) - P(AB)$$

Conditional Probability

Probability that event A happens GIVEN that event B is known to happen

$$P(A \mid B) = \mathbf{P(AB) / P(B)}$$

Combination of Events: Example

At Georgia Tech, two major credit card providers want to investigate how many students have a Visa credit card and how many students have a MasterCard.

From our survey we find that the proportion of GT students holding a Visa credit card is 50%, the proportion of GT students holding a MasterCard is 40% and that the proportion of students holding both credit cards is 25%.

Let A be the event that a selected student has a Visa credit card and let B be the event that a selected student has a MasterCard.

Combination of Events: Example

1. What is the probability that the selected individual has at least one of the two types of cards?

2. What is the probability that the selected student has neither type of card?

3. What is the probability that the selected student has a Visa card but not a Master-Card?

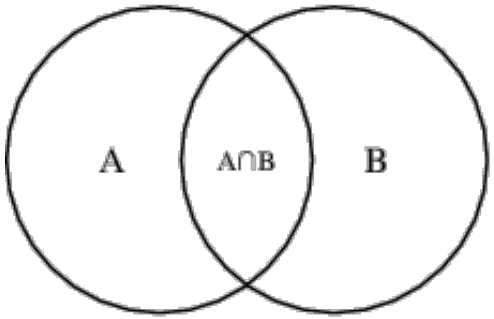
4. What is the probability that the selected student has a MasterCard, given that we know already that he/she has a Visa credit card?

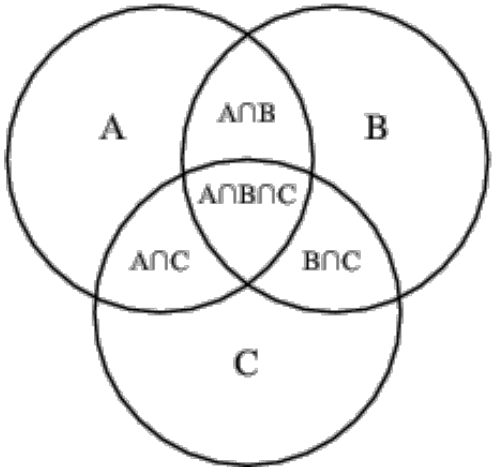
Combination of Events: Example

A = event that a selected student has a Visa Card (P(A)=.5)
B = event that a selected student has a Master Card (P(B)=.4)
Additional Information: P(AB) = .25

- P(AUB)
- P((AUB)’)
- P(A|B)
- P(B|A)

Use Venn Diagram





Independence

Events A and B are <u>independent</u> if

$$P(A \mid B) = P(A)$$

$$P(AB) = P(A)P(B)$$

Eg. A card is drawn at random from a pack of cards. Calculate the following probabilities:

P(King) = 4/52 = 1/13, and

P(King | Red) = 1/13.

Information about the card color did *not* effect the probability of drawing a King. This is the idea of independent events.

Independence - Definition

If events $A_1,\ldots,A_n$ are all independent of each other, then

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i)$$

Bayes' Theorem

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}$$

$$P(A_i \mid B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)P(B \mid A_i)}{\sum_{k=1}^{n} P(A_k)P(B \mid A_k)}$$

EXAMPLE: For the GT student credit card example

1. Calculate P(A|B) by

2. Calculate P(B|A) using P(A|B) calculated in 1 and Bayes' theorem

General Multiplication Rule

Multiplication Law: **P(AB) = P(A)P(B | A)**

Extension to three events:

P(ABC) = P(AB)P(C|AB) = P(A)P(B | A)P(C|AB)

Extension to n events $A_1,\ldots,A_n$

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) =$$
$$P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1A_2)\ldots P(A_n \mid A_1A_2\ldots A_{n-1})$$

Law of Total Probability

If $\{A_1,\ldots,A_n\}$ represents a Partition of the whole space, then these sets are mutually exclusive and

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) = 1$$

so that

$P(B) = P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + \ldots + P(A_n)P(B \mid A_n)$

Random Variables

In an experiment, only certain factors and outcomes are of interest to us.

Example: A researcher may test a sample of components from a production line and record only *the number of components* that have failed within 12 hours, rather than record individual component failure times.

A **Random Variable** is the number assigned to the (random) outcome of an experiment in order to summarize its meaning.

Random Variables

Definition: A **Random Variable** is an association rule or function from the *sample space* to the *state (range) space*:

$$\textbf{RV}: S \longrightarrow R$$

where the characteristic measurement of outcome **o** is **x**.

Example cont.: For each component, we observe 0's (failed) and 1's (not failed). What are possible outcomes? What are the possible values of the random variable?

Discrete Random Variables

Random variables (RV) have a *finite* or *countably infinite* range, and we can write the outcomes as $\{x_1, x_2, \ldots.\}$.

The set of probabilities associated with the outcomes

$$p_i = P(X = x_i) \text{ for each } x_i$$

Probability Mass Function (PMF) for the discrete random variable X:

$$0 \leq p_i \leq 1, \sum p_i = 1$$

Cumulative Distribution Function

The **Cumulative Distribution Function** (CDF) of a random variable is defined as

$$F_X(x) \text{ or } F(x) = P(X \le x) = \sum_{x_i \le x} P(X = x_i)$$

Note that $0 \le F(x) \le 1$ and $F(x)$ is increasing in x. In the dice example:

$F(x) = x/6$ for $x=1,2,3,4,5,6$.

Discrete vs. Continuous RVs

Discrete:
X takes discrete values

PMF: $p_i = P(X = x_i)$

Random Variables

Continuous:
X can be any (real) value in a given interval, so outcomes are not "countable"

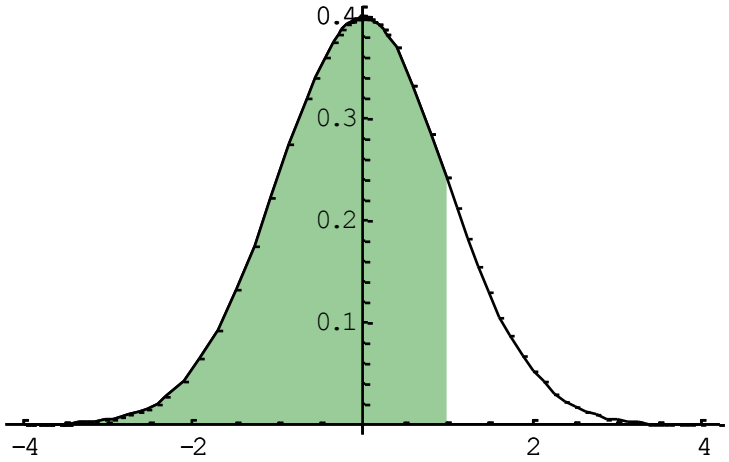
Probability Density Function (PDF) or density function for continuous random variables X

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

$$f(x) = \frac{d}{dx} F(x), F(x) = \int_{-\infty}^{x} f(t)\,dt$$

$$f(x) \ge 0$$

$$\int f(t)\,dt = 1$$



Example: Continuous Density

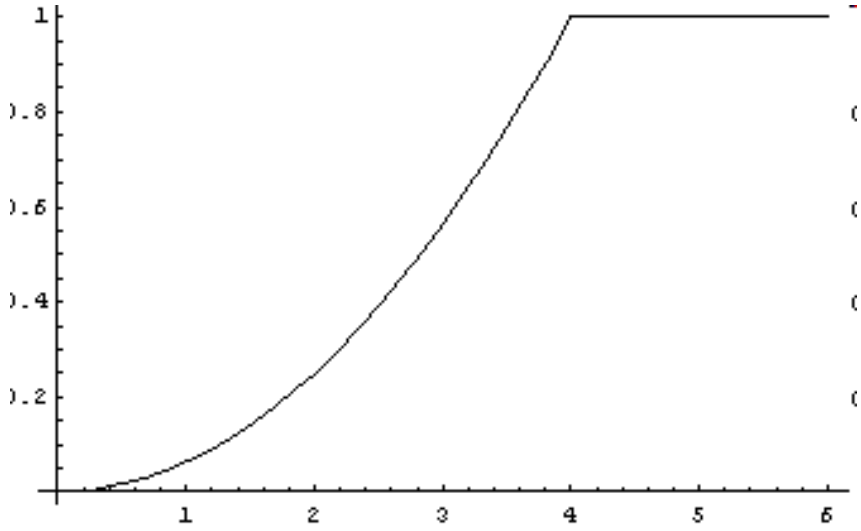
$F(x) = x^2/16, \quad 0 \le x \le 4$

$P(X \le 2) = F(2) = 4/16 = ¼$

$P(1 \le X \le 3) = P(X \le 3) - P(X \le 1)$

$= F(3) - F(1)$

$= 9/16 - 1/16 = 1/2$



$$f(x) = \frac{\partial}{\partial x} F_X(x) = x/8$$

Expectation

We are interested in summary values to measure the typical value of a RV (called the mean) or how much it varies (called variance).

Expectation/mean of X is E(X) = m:

1. Discrete Case $$E(X) = \sum_i x_i P(X = x_i)$$

2. Continuous Case $$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

Variance

$$Var(X) \equiv \sigma_X^2 = E\left(\left(X - E(X)\right)^2\right) = E(X^2) - E(X)^2$$

Standard Deviation is σ_x (the square root of the variance)

Linear Transform of RV

Given a RV X: with

$E(X) = \mu$

$Var(X) = \sigma^2$

Let another RV:

$Y = aX + b$

Then

$E(Y) = a\mu + b$

$Var(Y) = a^2\sigma^2$

Discrete RV

X = outcome of a roll of 6-sided die

$p_i = 1/6$, $1 \leq i \leq 6$, and $E(X) = \sum_{i=1}^{6} i \frac{1}{6} = \frac{1}{6}\left(\frac{6(6+1)}{2}\right) = 3.5$

$$E(X^2) = \sum_{i=1}^{6} x_i^2 P(X = x_i) = \sum_{i=1}^{6} \frac{i^2}{6} = \frac{1}{6}\left(\frac{6 \cdot 7 \cdot 13}{6}\right) = 15.167$$

$\text{Var}(X) = \sigma^2 = 15.167 - (3.5)^2 = 2.9167$, $\sigma = 1.71$

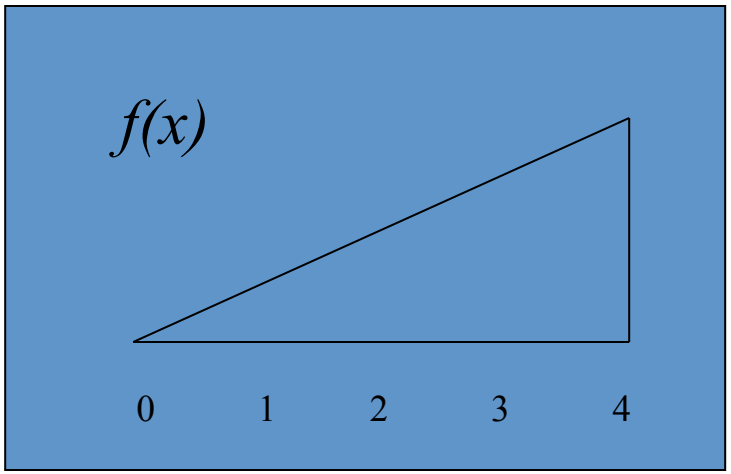
Continuous RV

$F(x) = x^2/16, \quad 0 < x < 4$

$$f(x) = \frac{x}{8},$$

$$E(X) = \int_0^4 \frac{x^2 dx}{8} = \frac{x^3}{24}\Big]_0^4 = 2.667,$$

$$E(X^2) = \int_0^4 \frac{x^3 dx}{8} = \frac{x^4}{32}\Big]_0^4 = 8$$

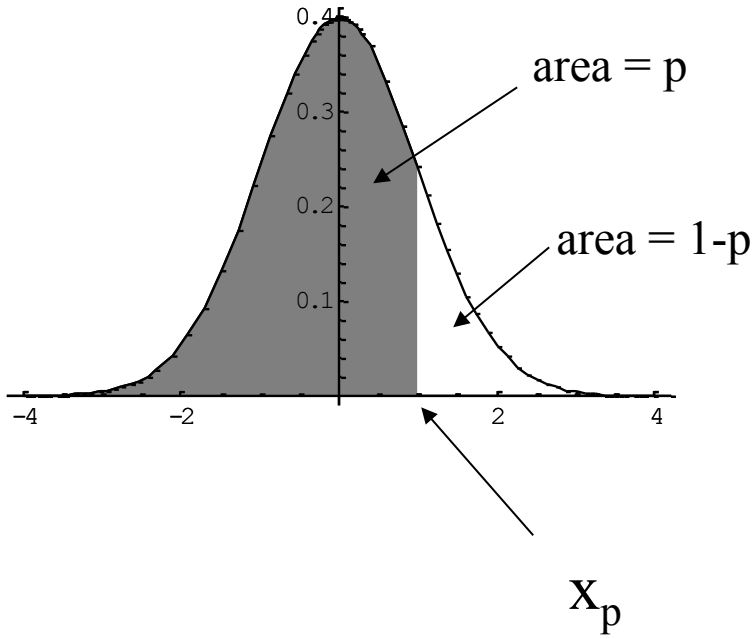


$$\sigma^2 = E(X^2) - E(X)^2 = 8 - (2.667)^2 = 0.89,$$

$$\sigma = \sqrt{0.89} = 0.94$$

Quartiles

- x_p is called the p^{th} *quartile* if $F(x_p) = p$.
- $x_{0.5}$ = Median
- Quartiles are useful in describing what values we might see for X.



Example continued

To solve for the 0.95 quantile, find t such that $F(t) = 0.95$:

$t^2/16 = 0.95$ means $t = 3.90$

Quartiles:

$x_{.25}$ = lower quartile: $(x_{.25})^2/16 = 0.25$ $\quad x_{.25} = 2$

$x_{.5}$ = median: $(x_{.5})^2/16 = 0.5$ $\quad x_{.5} = 2.82$

$x_{.75}$ = upper quartile: $(x_{.75})^2/16 = 0.75$ $\quad x_{.75} = 3.46$

Review: Discrete Distributions

- **Binomial Distribution**
- Geometric Distribution
- Negative Binomial Distribution
- Hypergeometric Distribution
- Poisson Distribution
- Multinomial Distribution

Binomial Distribution

Binomial random variables occur frequently in nature:

- n independent trials or events, n fixed
- Two outcomes in every trial (e.g., success vs. failure)
- P(success)=p is the same in every trial

Assign X = number of successes out of n trials

We say that X follows a ***Binomial distribution***.

- Write as X~B(n,p) or X~Bin(n,p)

• e.g. a batch of n = 12 laptops, the probability of each one of them being defective is p = 0.1, independent of each other, what's the probability that having X = 2 defective ones in this batch?

Binomial pmf

If P(*success*) = p, the P(X = k) = P(*k successes and n-k failures*). Each outcome like this has probability $p^k(1-p)^{n-k}$.

From counting rules, we know there are $\binom{n}{k}$ ways of choosing k successes from a group of k successes and n-k failures. As a result, we have

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

for k=0,1,2,…,n

Binomial – Expected Value and VAR

$$E(X) = \sum_{k=1}^{n} kP(X = k) = \sum_{k=1}^{n} k\left(\frac{n!}{k!n-k!}\right)p^k(1-p)^{n-k}$$

$$= \sum_{k=1}^{n} n\left(\frac{n-1!}{k-1!n-k!}\right)p^k(1-p)^{n-k} = n\sum_{j=0}^{n-1}\binom{n-1}{j}p^{j+1}(1-p)^{n-1-j} = np$$

$$E(X) = np$$

$$VAR(X) = n(1-p)p$$

Binomial Example

With a true/false test that has 20 questions, let X = no. correctly answered out of 20. For a grade of A, one needs X ≥ 19. What is the probability a student who guesses all the answers scores an A on this exam?

P(get an A) = P(X ≥ 19) = P(X=19) + P(X=20) where X~B(n=20, p= ½)

$$= \binom{20}{19}\left(\tfrac{1}{2}\right)^{19}\left(1-\tfrac{1}{2}\right)^{1} + \binom{20}{20}\left(\tfrac{1}{2}\right)^{20}\left(1-\tfrac{1}{2}\right)^{0} \approx 0.00002$$

In this case $E(X) = np = 10$, $Var(X) = np(1-p) = 5$

≥≤

Review: Continuous Distributions

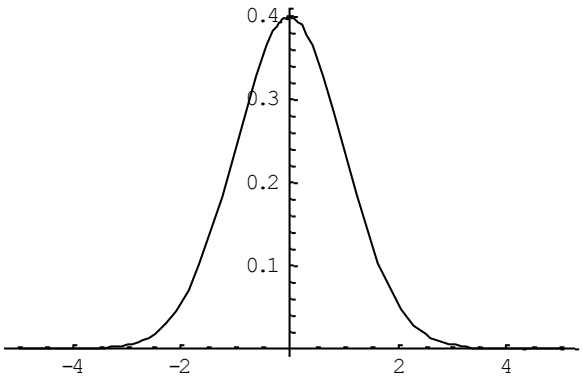
- Uniform Distribution
- Exponential Distribution
- Gamma Distribution
- Weibull Distribution
- Beta Distribution
- **Normal Distribution**

Normal Distribution

X is distributed as a normal random variable with mean μ and variance σ^2; i.e., X ~N(μ , σ^2)

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

Standard normal random variable Z: μ =0, σ^2 =1

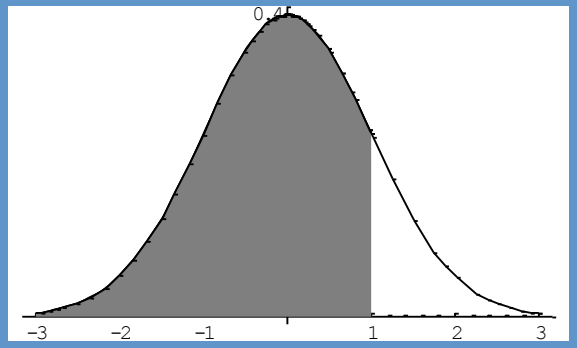


CDF for Normal Distribution

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(u)du \equiv \Phi_x(x)$$

Example :

$\Phi_z(1) = P(Z \le 1) = 0.8413$

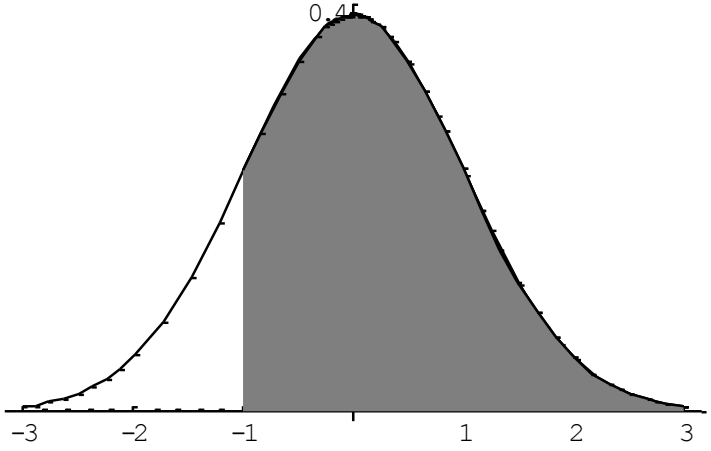


Properties of Standard Normal Distribution

- $\Phi_z(0) = 1/2,\ \Phi_z(\infty) = 1$
- $\Phi_z(t) = 1 - \Phi_z(-t)$

$P(Z \le t) = 1 - P(Z \le -t)$

- $P(Z \ge t) = P(Z \le -t)$



Example : $P(Z \ge -1) = P(Z \le 1)$

$= 0.8413$

Calculating Normal Probabilities

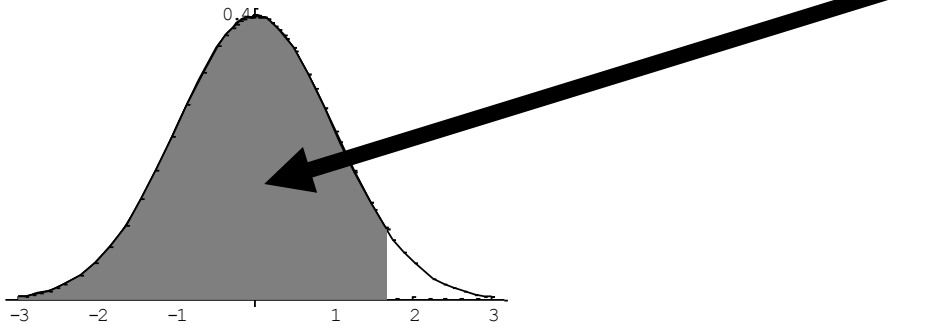
If $X \sim N(\mu, \sigma^2)$ then $Z = [(X-\mu)/\sigma] \sim N(0, 1)$

Example: Suppose $X \sim N(100, 9)$, so $\sigma = 3$

$P(X \leq 105) = P(X-\mu \leq 105 - \mu) =$

$= P((X-\mu)/\sigma \leq (105 - \mu)/\sigma) = P(Z \leq (105 - 100)/3)$

$= P(Z \leq 5/3) = P(Z \leq 1.67) = 0.9525$



Suppose again X ~N(100, 9):

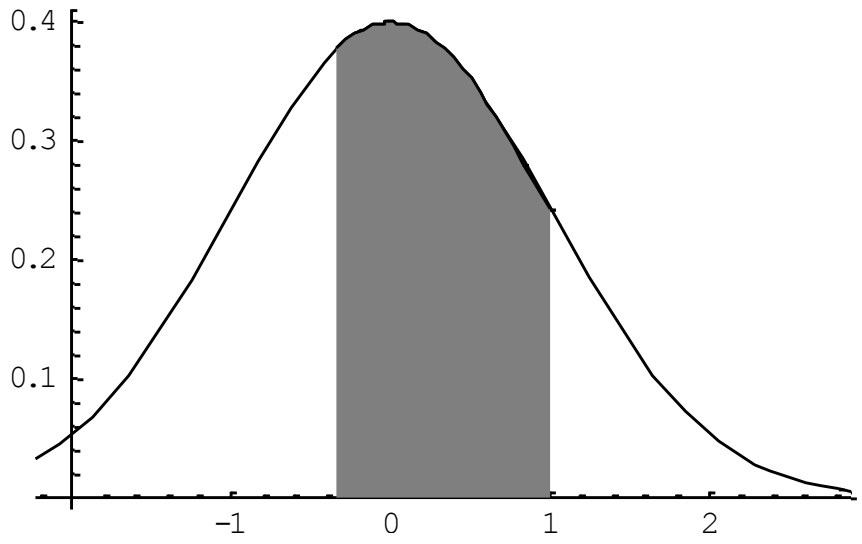
P(99< X <103) =

P((99 –100)/3 ≤(X - μ)/σ ≤(103 –100)/3) =

P(-1/3 <Z <1) =

P(Z<1) -P(Z<–1/3) =

0.8413 – 0.3707 = 0.4706

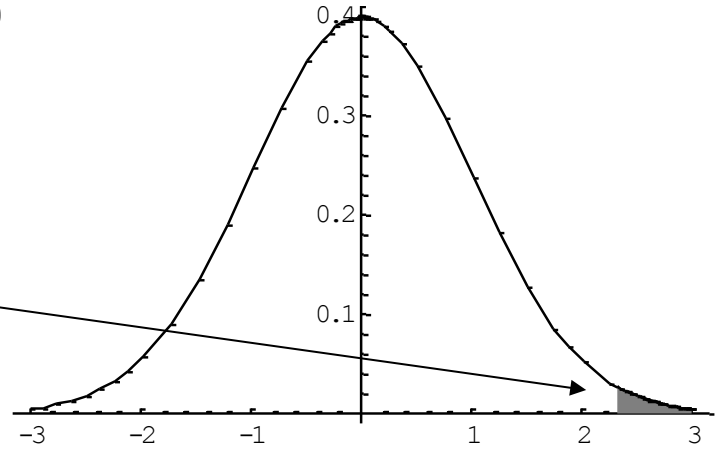


Computing Normal Quartiles

With X ~N(100, 9), what is the value X* such that
P(X >X*) =0.01 ? We will first find P(X<X*)(= 1-P(X>X*))

P(X <X*) =
P((X -μ)/σ <(X* –100)/3) = P(Z <Z*) = 0.99

Φ(2.33) =0.99



P(Z >2.33) =0.01 and therefore

Z* =(X*-100)/3 =2.33.

Solve for X* to get 3(2.33) +100 =106.99

Joint Discrete Distributions

Random variables can be multivariate – example: consider a *bivariate* random variables (X,Y)

Joint probability mass function of (X,Y):

$$X : \mathfrak{S}_X \to \mathfrak{R}_X$$

$$Y : \mathfrak{S}_Y \to \mathfrak{R}_Y$$

$$p_{(X,Y)} : \mathfrak{R}_X \times \mathfrak{R}_Y \to [0,1]$$

$$p_{(X,Y)}(x,y) = p(x,y) = P(X = x, Y = y)$$

$$\sum_{x \in \mathfrak{R}_X} \sum_{y \in \mathfrak{R}_Y} P(X = x, Y = y) = 1$$

Joint Discrete Distributions

Cumulative distribution function:

$F(x,y) = P(X \le x, Y \le y) = \sum_{s \le x} \sum_{t \le y} P(X = s, Y = t)$

$0 \le F(x,y) \le 1$

Marginal Probability functions:

$$p_X(x) = \sum_{y \in \mathfrak{R}_Y} p(x,y)$$

$$p_Y(y) = \sum_{x \in \mathfrak{R}_X} p(x,y)$$

Random Vector Example

Phone call center, for air conditioner maintenance:

X = service time (hours) spent for a request

Y = number of ACs in the unit being requested for service

	X=1	*X=2*	*X=3*	*X=4*
Y = 1	0.12	0.08	0.07	0.05
Y = 2	0.08	0.15	0.21	0.13
Y = 3	0.01	0.01	0.02	0.07
Y = 4	**0.21**	**0.24**	**0.30**	**0.25**

(x,y) = (1,1) means 12% of calls are for places with one AC, service time = 1 hr.

Example, continued

AC Maintenance example:

- $P[(X,Y)=(4,2)] = p_{42} = 0.13$
- $P[(X,Y) \leq (2,2)] = p_{11} + p_{12} + p_{21} + p_{22} = 0.43$
- $P[X \leq 2, Y=3] = p_{13} + p_{23} = 0.01 + 0.01 = 0.02$
- $P[X \leq 1] = p_{11}+p_{12}+p_{13} = 0.21$, $P[X \leq 2] = 0.45$, $P[X \leq 3] = 0.75$, $P[X \leq 4] = 1$

By "*averaging over*" Y, we have the CDF for X (called the **marginal distribution**)

Joint Continuous Distributions

If X and Y are two continuous random variables, define the joint probability density function of (X,Y):

$$X : \mathbb{S}_X \to \mathbb{R}_X, Y : \mathbb{S}_Y \to \mathbb{R}_Y$$

$$f_{(X,Y)} : \mathbb{R}_X \times \mathbb{R}_Y \to [0, \infty)$$

$$P(a \le X \le b, c \le Y \le d) = \int_a^b \int_c^d f(x, y) dy dx$$

$$\int_{\mathbb{R}_X} \int_{\mathbb{R}_Y} f(x, y) dx dy = 1$$

Linear Functions of Normal RVs

Linear Functions of Normal Random Variables are still Normal Random Variables.

This is *not true* with other distributions; if X ~Exponential, Y ~Exponential, that does NOT mean X +Y ~Exponential (it's not!)

- If $X \sim N(\mu_x, \sigma_x^2)$, then $aX + b \sim N(a\mu_x+b, a^2\sigma_x^2)$

- If $Y \sim N(\mu_y, \sigma_y^2)$ is **independent** of X, then
  $X+Y \sim N(\mu_x +\mu_y, \sigma_x^2 +\sigma_y^2)$

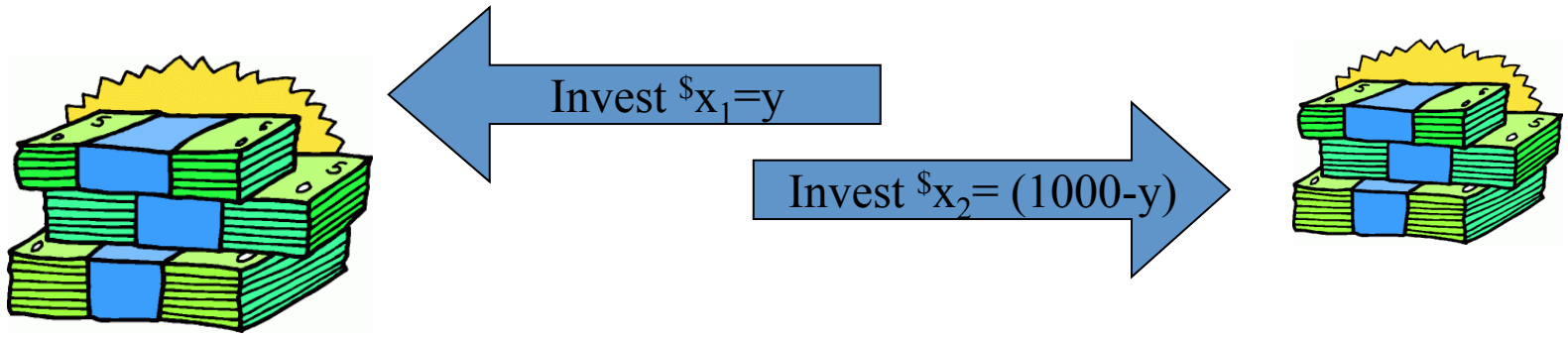
Example: Investing

We have $1000 to invest.

We put $Y in Mutual Fund I and the rest (1000-Y) into Mutual Fund II.

The **returns** are random and unknown, characterized with the normal distribution:

$R_1 \sim N(x_1, 0.0002x_1^2)$ $R_2 \sim N(1.05x_2, 0.0003x_2^2)$



Investment example, cont.

Let T be the amount we end up with, then

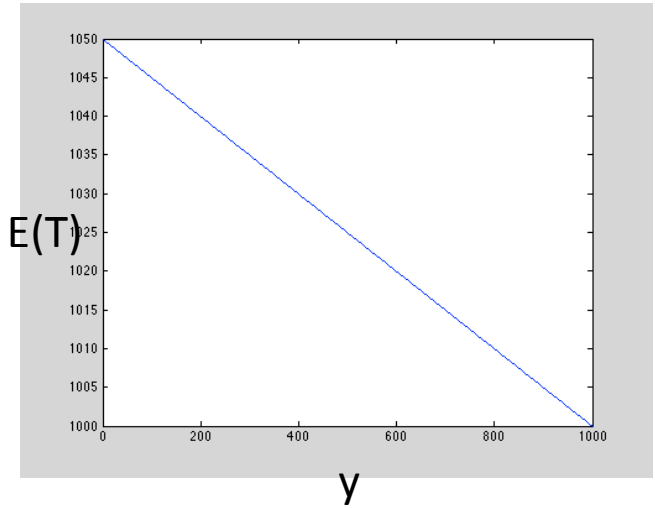
$T = R_1 + R_2$

What is the expected value after investment?

$E(T) = E(R_1 + R_2) = x_1 + 1.05x_2$

$= y + 1.05(1000 - y)$

$= 1050 - 0.05y$

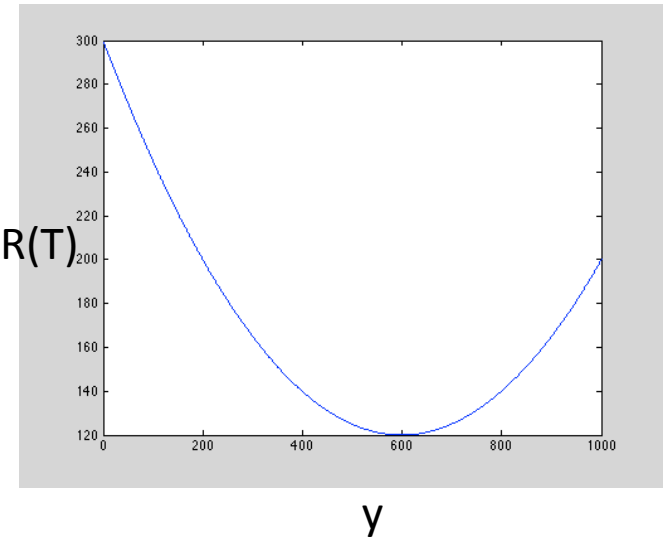


What is the variance of the final amount?

$Var(T) = Var(R_1) + Var(R_2)$

$= 0.0002x_1 + 0.0003\ x_2$

$= 0.0002y^2 + 0.0003(1000-y)^2$

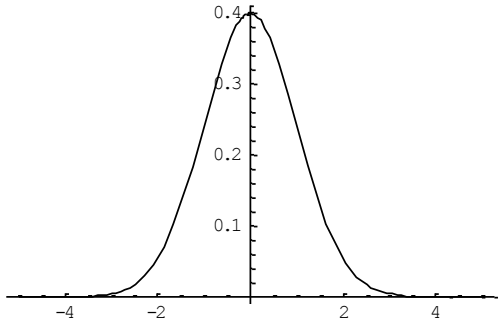


Central Limit Theorem

If $X_1,\ldots,X_n$ are independent and $E(X_i)=\mu$, $Var(X_i) = \sigma^2$, then if n is large

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$ is approximately Normally distributed, with

$$E[\bar{X}] = \mu, \quad V(\bar{X}) = \frac{\sigma^2}{n}$$



We do NOT require that the X_is are normally distributed. This approximation usually works well if n ≥ 30

Example - CLT

When a lithium batteries is prepared, the amount of a particular impurity in the battery is a random variable with mean value of 4.0 g and standard deviation 1.5 g. If 50 batteries are independently prepared, what is the (approximate) probability that the sample average amount of impurity is between 3.5 and 3.8 g?

The sample size n = 50 is large enough to use CLT to approximate the distribution of the average with a normal distribution. From CLT

$$\overline{X} \approx N(\mu = 4.0, \sigma^2 / n = .045)$$

We want to find the following probability $P(3.5 \le \overline{X} \le 3.8)$

$$\approx P\left(\frac{3.5-4}{\sqrt{.045}} \le Z \le \frac{3.8-4}{\sqrt{.045}}\right) = \Phi(-.94) - \Phi(-2.36) = .1645$$

Introduction to R

- Youtube video: An Introduction to R – A Brief Tutorial for R
  http://www.youtube.com/watch?v=LjuXiBjxryQ
- An Introduction to R:
  http://cran.r-project.org/doc/manuals/R-intro.html
- installation: http://cran.us.r-project.org/
- set directory: setwd("/directory/data")
- get help about a particular function: help(functionName)
- data $<-$ read.table(file = "Deflection.txt")

- index = data[,1]

  deflection = data[,2]

  plot(deflection, index,type='p')

  % type = 'p': plot points; type = 'l': plot a line