BUDT 730

Data, Models and Decisions

Lecture 03

Probability Distributions (I)

Prof. Sujin Kim

Outline - Probability Distributions (CH5)

- Probability & Random Variables
  - Understand the basic properties of probability
  - Understand the characteristics of random variables
- Sum of Random Variables
  - Calculate the mean and the variance of a sum of random variables

Introduction

- The key challenge in solving real business problems is dealing appropriately with uncertain outcomes
- There are many sources of uncertainty
  - For example, demand, time between customer arrivals, stock price returns, changes in interest rate
- Probability allows us to model this uncertainty and make decisions based on the ***best expected outcomes***

Probability & Random Variables

Practice - Comparing Two Projects

- A firm is selecting between two possible investment projects: project A or project B. The outcomes of two projects are uncertain.

- We would like to compare the two projects by analyzing the historical data.
- Explore `Two Projects.xlsx` data and compare the two projects.

Practice - Comparing Two Projects

- The firm assesses that by investing in A they can either lose 10 million or gain 90 million 190 million.
- For project B they assess that they can lose 10 million or gain 190 million.

- How to measure the likelihood of each outcome?

  Use Excel COUNTIF function: COUNTIF(range, condition):

  - The **COUNTIF** function counts the number of cells within a range that meet the given condition.

Practice - Comparing Two Projects

- The firm assesses that by investing in A they can either lose 10 million, with a 81% chance, gain 90 million with a 18% chance or gain 190 million with a 1% chance.
- For project B they assess that they can lose 10 million with a 90% chance or gain 190 million with a 10% chance.

In which project will you invest?

Basic Concepts of Probability

What is **probability**?

- A way to deal with and manage uncertainty
- An intuitive concept, e.g. the likelihood of losing 10 million.
- Experiment is any process whose outcome is not known in advance.
- *Sample space* is the set of all possible outcomes of an experiment.

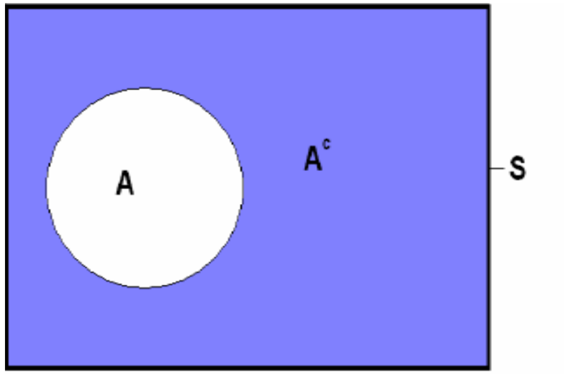
Facts about probabilities:

- All probabilities are between 0 and 1
- The closer its probability is to 1, the most likely it is to occur

Properties of Probability

Given that we can express probabilities for individual events, there are several important properties that describe how some events relate to others

- Rule of Complements: $P(\bar{A}\ or\ A^c) = \ 1 - P(A)$



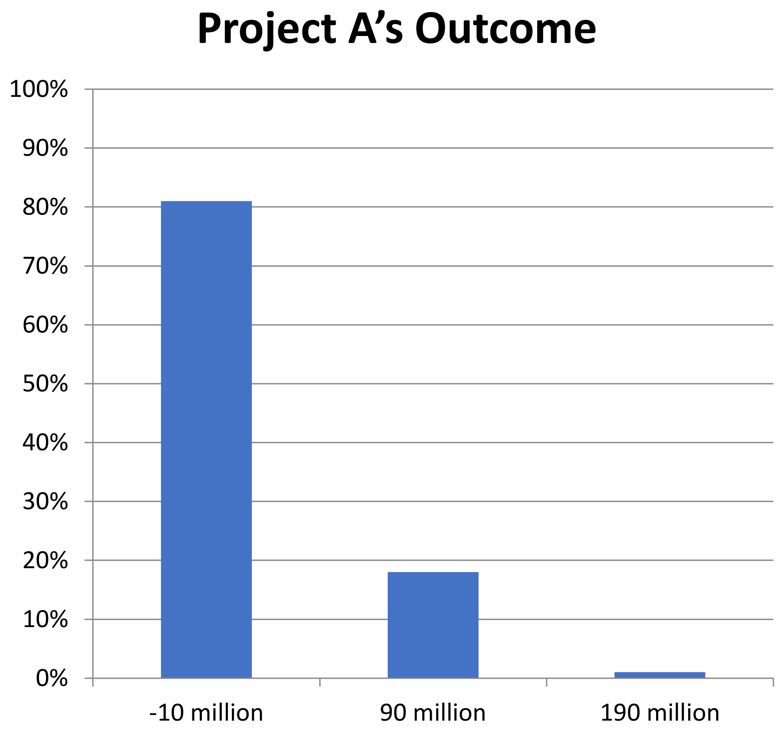
- Later (Decision Tree), we will study
  - Addition Rule
  - Conditional Probability and Multiplication Rule
  - Probabilistic Independence
  - Law of Total Probability

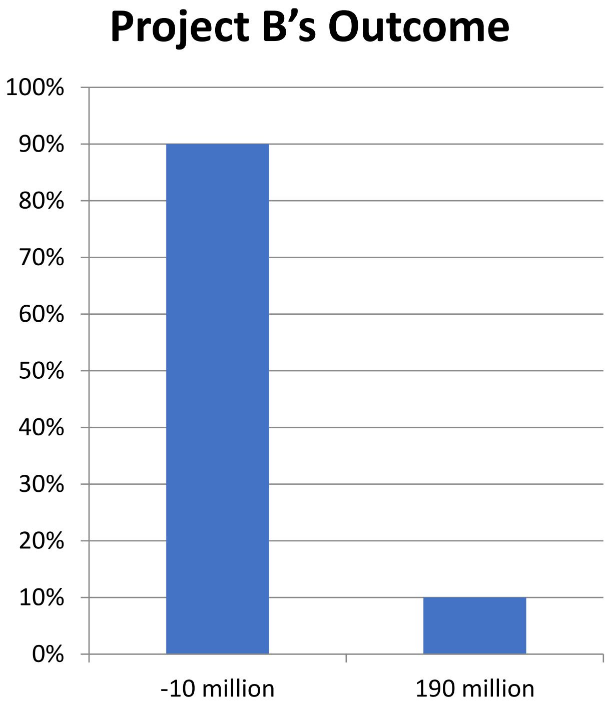
Comparing Two Projects

- A firm is selecting between two possible investment projects: project A or project B
  - The firm assesses that by investing in A they can either lose 10 million, with a 81% chance, gain 90 million with a 18% chance or gain 190 million with a 1% chance.
  - For project B they assess that they can lose 10 million with a 90% chance or gain 190 million with a 10% chance.

We can model the outcome of these projects as a discrete random variable, i.e. every possible outcome can be represented by a numerical value and an associated probability of occurrence

Probability Distribution of the Project's Outcomes





Random Variable (RV)

A *random variable (RV)* is a (numerical) variable that can take on different values, each with an associated probability.

A probability distribution describes the likelihood (probability) of a random variable.

- A probability distribution is not based on a data set (sample). It is a list of all possible outcomes (sample space, population) and their probabilities.

Remarks:

- Usually denoted RVs as capital letters, e.g., *X*, *Y*, *etc.*
- The set of possible values of X is called the *range* of X.

Discrete vs. Continuous RVs

There are two types of RVs, each of which is used to model different types of sample spaces

- A *discrete random variable* is a random variable which may take on only a countable number of distinct values such as 0,1,2,3,4,..
  - Ex: Number of customers who arrive at a bank on one day
- A *continuous random variable* takes on an infinite number of real values in some range
  - Range can be bounded on both sides [0, 1], one side $[0, \infty]$, or neither $[-\infty, \infty]$
  - Examples:
    - Manufacturing time for a particular car component
    - Time between customer arrivals at a bank

- Example: Daily profits of a newspaper stand – Discrete or Continuous RV?

Discrete Random Variables

- For a discrete random variable X , the *probability mass function*, often abbreviated ***pmf***, is

$$p(x) = P(X = x), for\ -\infty < x < \infty.$$

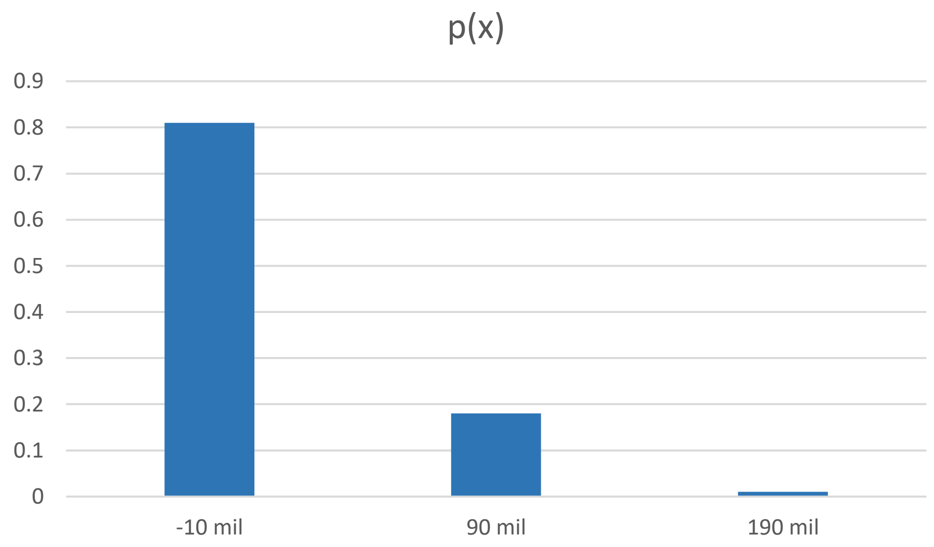
- A *cumulative probability distribution function (**cdf**)* describes the probability that the RV is less than or equal to some value
  - $F(x) = P(X \leq x)$

Example : pmf of Project A's outcome

Table

x	p(x)
-10 mil	0.81
90 mil	0.18
190 mil	0.01

Graph



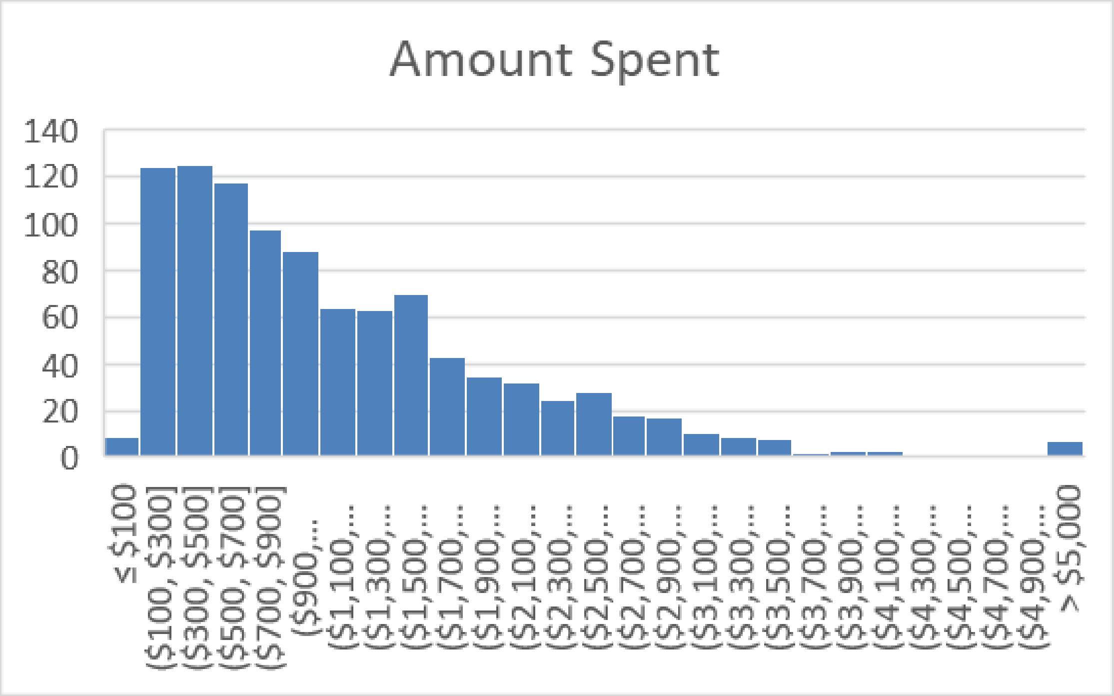
Example: cdf of Project A's outcome

x	p(x)	$F(x)$
-10 mil	0.81	0.81
90 mil	0.18	0.99
190 mil	0.01	1

$$F(x) = \begin{cases} 0 & x < -10\ mil \\ 0.81 & -10\ mil \leq x < 90\ mil \\ 0.99 & 90\ mil \leq x < 190\ mil \\ 1 & x \geq 190\ mil \end{cases}$$

HyTex Catalog Marketing Data

- How to describe the probability distribution of Amount Spent?



Continuous Random Variable

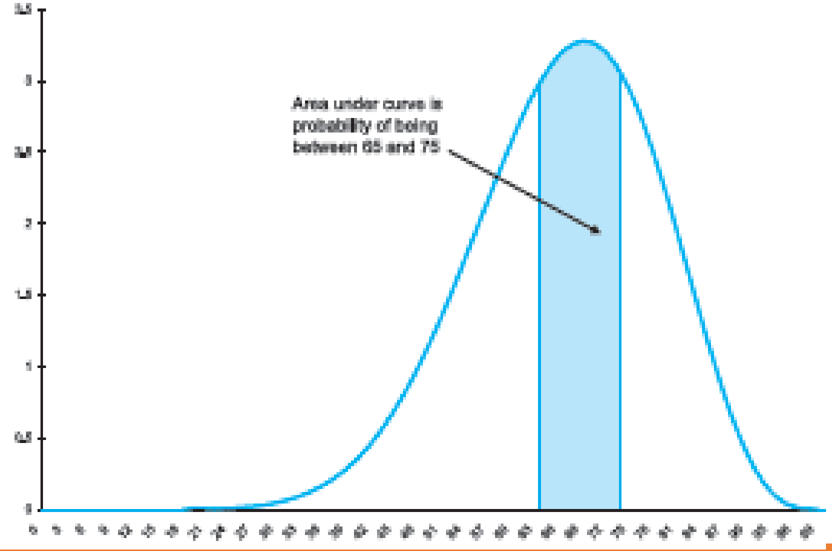
- For a continuous random variable X, the *probability density function f(x)*, often abbreviated ***pdf***, is the likelihood of x
- The cdf is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du.$$

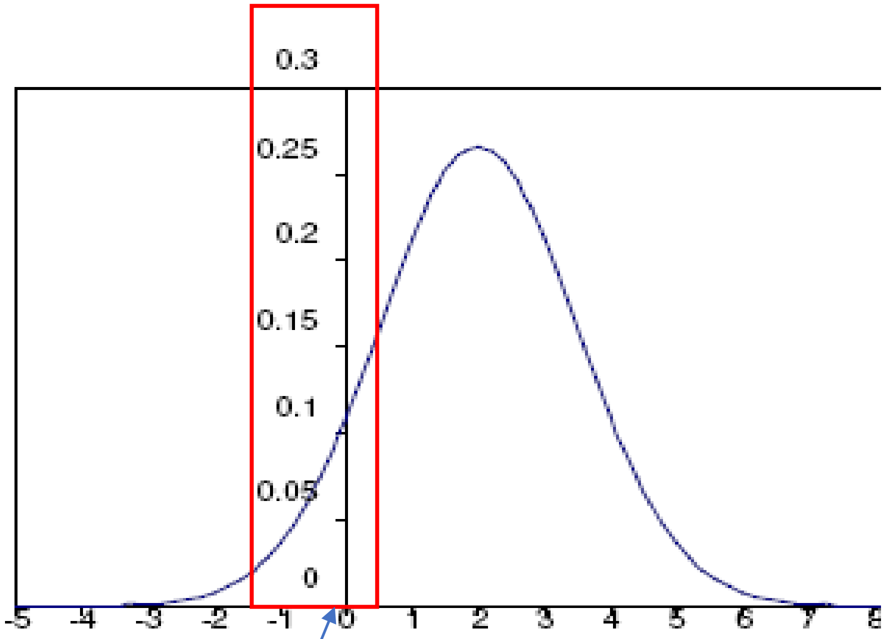
- Probabilities are found as areas under f(x) within a given interval
- If X is a continuous rv,

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

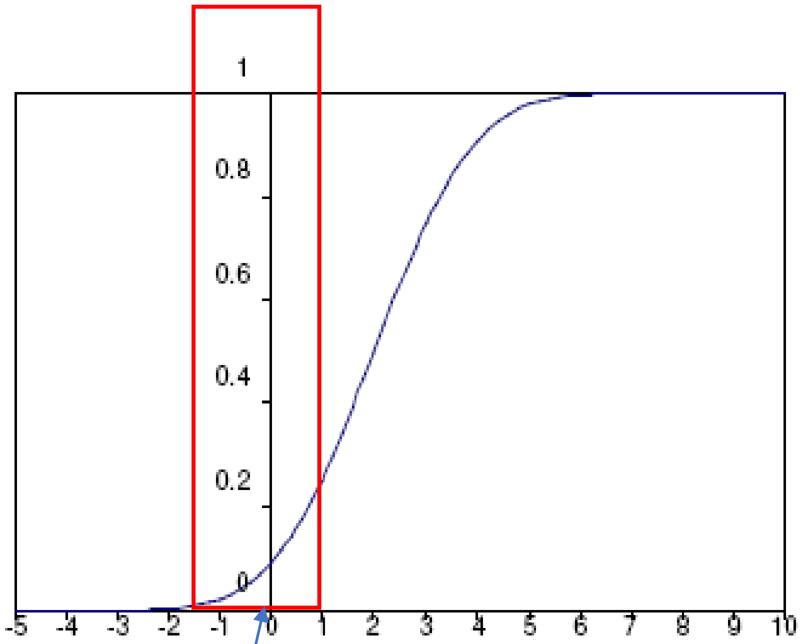
  for every real number *x.*



Example of PDF & CDF of Continuous RV



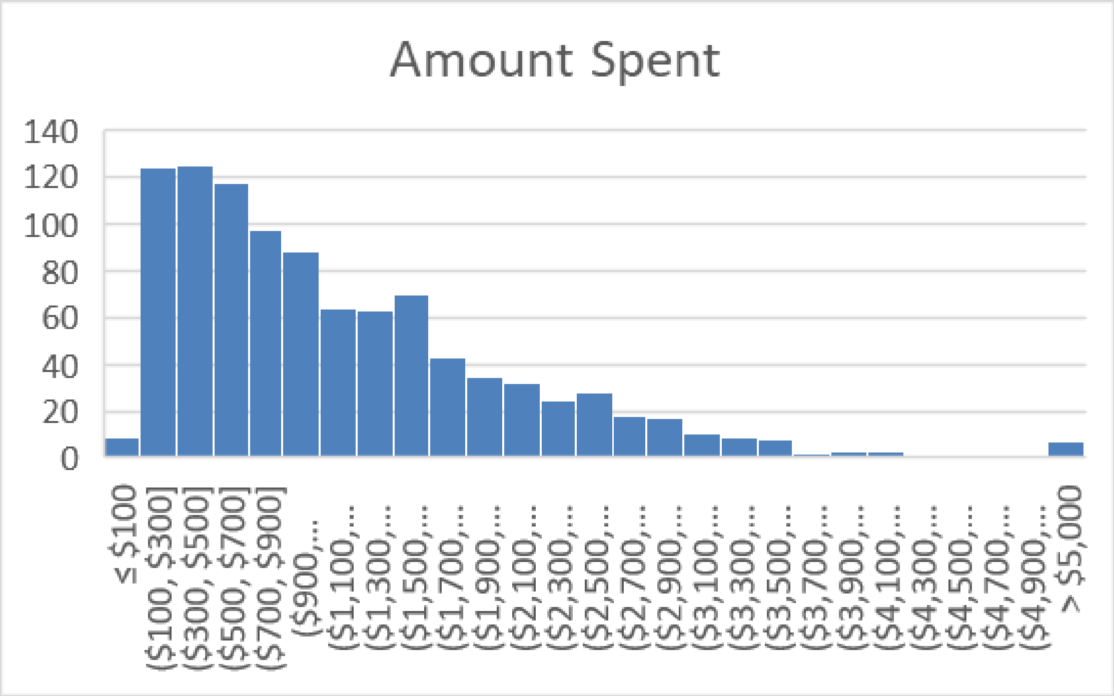
Likelihood- not a probability, unitless

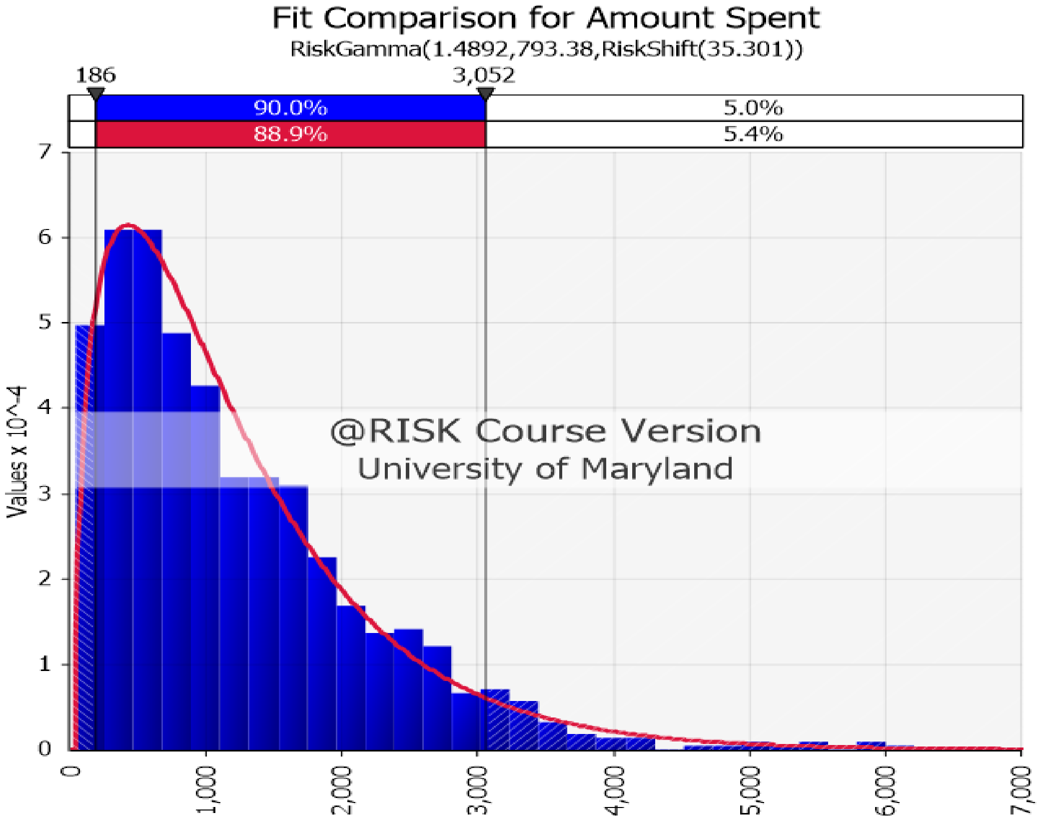


Probability

HyTex Catalog Marketing Data

- Amount Spent follows a Gamma distribution.





Expected Value of Random Variables

Expected Value of a Discrete Random Variable

- Let X be our discrete random variable that takes on values $x_1, \dots, x_k$ with probabilities $p_1, \dots, p_k$
- The ***expected value*** is the weighted average of all the k values, with their probabilities serving as weights

$$\boxed{\boldsymbol{\mu_X} = E(X) = \sum_{i=1}^{k} x_i \cdot p_i}$$

Expected Returns on Projects A and B

- Let X represent the random return of project A and Y represent the random return of project B
- The expected payback for project A:

Outcome:	-$10M	$90M	$190M
Probability:	81%	18%	1%

$$\mu_X = E[X] = (81\%)(-10) + (18\%)(90) + (1\%)(190)$$
$$= (0.81)(-10) + (.18)(90) + (0.01)(190) = \mathbf{10\ million}$$

- For project B:

Outcome:	-$10M	$190M
Probability:	90%	10%

$$\mu_Y = E[Y] = (.9)(-10) + (.1)(190) = \mathbf{10\ million}$$

Selection between Projects A and B

If the company would be need to select numerous times between independent projects exactly like projects A and B then

A. The company should always select type A projects
B. The company should always select type B projects
C. It depends
D. It doesn’t matter

On *average* the company will be equally well off with either strategy.

Interpreting the Expected Value

- The expected value is a long run average
- If the firm invests in a large number of projects with the same return as Project A or B, then we'd expect that the average return per project would be close to 10 million
- On *average* the company will be equally well off with either strategy.

Selection between Projects A and B

If projects A and B were one-time opportunities, in which project should you invest?

A. Project A
B. Project B
C. Neither
D. Impossible to say

The answer depends on the firms' risk preference:

- Can it handle a -$10M loss?
- If yes, is the company risk seeking or risk averse?
- Is getting a $190 a game changer for the company?

We will study the risk preference in Ch6.

Measuring Variability: **Variance**

- One way to quantify the variability of a random variable is to use the variance or the standard deviation
- For a discrete random variable, the **variance** measures the variability of the outcomes of X around its mean, weighted by their probabilities:

$$\boxed{\boldsymbol{\sigma_X^2} = Var(X) = \sum_{i=1}^{k}(x_i - \mu_X)^2 p_i}$$

- The **standard deviation** is

$$\boxed{\sigma_X = \sqrt{\sigma_X^2}}$$

Variability of Returns on Projects A and B

- What is the Standard Deviation of Project A?

$$\sigma_X = \sqrt{(-10 - \mathbf{10})^2 * .81 + (90 - \mathbf{10})^2 * .18 + (190 - \mathbf{10})^2 * .01} = 42.4$$

- Similarly you can find that the standard deviation of **project B is 60**.
- Which project would you invest in if you had the choice?
- The answer depends on your risk appetite:
  - Are you risk seeking, risk neutral, or risk averse? (we will discuss the risk preference later in Ch 6 Decision Tree)

Sum of Random Variables

Example: Supermarket Spending

- Historical data suggests that customer spending at a supermarket is randomly distributed with the mean of $85 and the standard deviation of $30.
- If 500 customers shop in a given day, what is the expected revenue of the supermarket on that day?
  - Intuitively, the answer is $85*500 = $42,500.
- What is the standard deviation?

Example: Supermarket Spending

- Let X_i be the amount spent by the ith customer
- Note that $E(X_i) = 85$ and $Var(X_i) = 30^2 = 900$
- The total amount spent by 500 customers is

$$Y_{500} = X_1 + X_2 + \cdots + X_{500}$$

Then,

$$\begin{aligned} E(Y_{500}) &= E(X_1 + X_2 + \cdots + X_{500}) \\ &= E(X_1) + E(X_2) + \cdots E(X_{500}) \\ &= 85 * 500 = \$42{,}500 \end{aligned}$$

Note: We assume that X_i's are **identically distributed**.

Example: Supermarket Spending

- What is the standard deviation?
- Suppose that $X_1, X_2, \dots, X_{500}$ are **independent**

$$
\begin{aligned}
Var(Y_{500}) &= Var(X_1 + X_2 + \cdots + X_{500}) \\
&= Var(X_1) + Var(X_2) + \cdots Var(X_{500}) \\
&= 900 * 500 = 450{,}000
\end{aligned}
$$

$$Stdev(Y_{500}) = \sqrt{450{,}000} = \$670.82$$

Note: We assume that X_i's are **independent and identically distributed (i.i.d.)**.

Properties of Expectation Function

- Expectation is a linear function.
  - $E(X+Y) = E(X) + E(Y)$

  a is a constant:
  - $E(a) = a$
  - $E(aX) = aE(X)$

Properties of Variance Function

- If X and Y are **independent**:
$$Var(X + Y) = Var(X) + Var(Y)$$

- If X and Y are **dependent**:
$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$$

- a is a constant:
  - $Var(a) = 0$
  - $Var(aX) = a^2 Var(X)$

Exercise: Sum of RVs

- Suppose that

$$E(X) = 0.5, Var(X) = 0.25,$$
$$E(Y) = 1, Var(Y) = 0.5$$

  and that X and Y are independent

  - $E(2X + Y + 3) =$
  - $Var(2X + Y + 3) =$

Exercise: Sum of RVs

- Suppose that

$$E(X) = 0.5, Var(X) = 0.25,$$
$$E(Y) = 1, Var(Y) = 0.5$$

  and that X and Y are independent

  - $E(2X + Y + 3) =$
  - $Var(2X + Y + 3) =$

  Solution:

  - $E(2X + Y + 3) = 2E(X) + E(Y) + 3 = 2 * 0.5 + 1 + 3 = 5$
  - $Var(2X + Y + 3) = 2^2 Var(X) + Var(Y)$
    $= 4 * 0.25 + 0.5 = 1.5$

Exercise - Two Projects

- The firm decided to invest in both project A & B.

Outcome:	-\$10M	\$90M	\$190M
Probability:	81%	18%	1%

Outcome:	-\$10M	\$190M
Probability:	90%	10%

- Compute
  - $E(Return\ of\ A + Return\ of\ B)$ and
  - $\mathrm{Var}\ (Return\ of\ A + Return\ of\ B)$

- Compute the probability distribution of $(Return\ of\ A + Return\ of\ B)$.