Business Analytics

Lecture 3

Random Variables

Dr Yufei Huang

UCL
SCHOOL OF
MANAGEMENT

Contents

- Discrete random variables
  - Probability distribution and cumulative distribution function
  - Expected Value, variance and standard deviation
  - Bernoulli random variable
  - The Binomial random variable

- Continuous random variables
  - The probability density function and cumulative distribution function
  - Uniform distribution
  - Normal Distribution

- Excel

Random Variables

- Definition. A variable is called random if its value depends on chance.

- Notation. We denote random variables by capital letters, e.g. X,Y,…, and the lower cases, x, y,…, denote the outcomes (actual values).

- Examples

1. Assign to the result of a coin toss experiment the numbers:

   X=0 if 'tail' is obtained, X=1 if 'head' is obtained

   The result of a coin toss depends on chance and is a random variable.

2. Let Y be the random variable: "service time when calling customer service"

   The values Y can take are: from 0 to ∞

   The actual result of Y depends on chance.

UCL SCHOOL OF MANAGEMENT

Types of Random Variables

1. Discrete Random variables

Definition. A discrete random variable is a variable which can assume at most a countable number of values.

Examples: coin toss, the result on the face of a die.

2. Continuous random variables

Definition. A continuous random variable may take on any value in an interval of numbers (i.e., its possible values are uncountably infinite).

Example: service time.

Quiz: Discrete or Continuous?

1. Number of Ford Focus sold per day

   Discrete

2. Number of online orders

   Discrete

3. Service time at KFC on Tottenham Court Road

   Continuous

4. Price of Iphone X

   Not a random variable

Are all outcomes equally likely?

Not always. The probability of some outcomes can be sometimes larger than the probability of others.

Discrete Random Variable

Probability Distribution

- Definition. A probability distribution for a discrete random variable is a mutually exclusive listing of all possible numerical outcomes for that variable such that a particular probability of occurrence is associated with each outcome.

- If x is a possible outcome of a discrete random variable X, we can denote the probability of x, by P(x) or P(X=x).

- The probability distribution of a discrete random variable X must satisfy the following two conditions.

  1. $P(x) \geq 0$ for all possible values x of a random variable X

  2. $\sum_{\text{all } x} P(x) = 1$

Cumulative Distribution

- We use F(x)=P(X≤x) to denote the probability of all outcomes from X that are smaller or equal to x. F(x) is called the cumulative distribution function.

- Example. x is the outcome of a fair die

x=outcome of a fair die	P(x)	F(x)	Reward
1	1/6	1/6	£6
2	1/6	2/6	£12
3	1/6	3/6	£18
4	1/6	4/6	£24
5	1/6	5/6	£30
6	1/6	1	£36

- *Question: What is the reward value that you are expecting?*

Expected Value

- The expected value of X is the mean of its outcomes, weighed by the probability of each outcome.

- Let X be a random variable. Denote its possible outcomes by $x_1, \ldots, x_N$. The expected value of X, μ, is given by:

$$\mu = E(X) = \sum_{i=1}^{N} x_i P(x_i).$$

- *Question: what's the difference between expected value and average value?*

Variance

- The variance of a random variables measures its spread.

- Definition. Let X be a discrete random variable. Denote its possible outcomes by $X_1, \ldots, X_N$. Let $\mu = E(X)$. Then the variance of X is given by

$$\sigma^2 = V(X) = E[(X-\mu)^2] = \sum_{i=1}^{N} (X_i - \mu)^2 P(X_i).$$

- It can also be written as:

$$\sigma^2 = V(X) = E[(X-\mu)^2] = E(X^2) - [E(X)]^2$$

Standard Deviation

- The standard deviation of X is given by

$$\sigma = SD(X) = \sqrt{V(X)} = \sqrt{E[(X-\mu)^2]} = \sqrt{\sum_{i=1}^{N}(X_i-\mu)^2 P(X_i)}$$

$$= \sqrt{E(X^2)-(E(X))^2}$$

- Remark. Both standard deviation and variance can be considered as risk measures.

Example

x=outcome of a fair die	P(x)	F(x)	Reward
1	1/6	1/6	£6
2	1/6	2/6	£12
3	1/6	3/6	£18
4	1/6	4/6	£24
5	1/6	5/6	£30
6	1/6	1	£36

- Expected reward:

$$\mu = E(X) = \sum_{i=1}^{N} x_i P(x_i)$$

$$= 6*(1/6) + 12*(1/6) + 18*(1/6) + 24*(1/6) + 30*(1/6) + 36*(1/6)$$

$$= 1 + 2 + 3 + 4 + 5 + 6 = 21$$

- Note. 21 is different from all possible values of reward.

Example

x=outcome of a fair die	P(x)	F(x)	Reward
1	1/6	1/6	£6
2	1/6	2/6	£12
3	1/6	3/6	£18
4	1/6	4/6	£24
5	1/6	5/6	£30
6	1/6	1	£36

- Variance of reward:

$$\sigma^2 = V(X) = E[(X-\mu)^2] = \sum_{i=1}^{N} (X_i - \mu)^2 P(X_i) = 105$$

$$\sigma^2 = V(X) = E[(X-\mu)^2] = E(X^2) - [E(X)]^2 = 105$$

UCL SCHOOL OF MANAGEMENT

Bernoulli Distribution

- Definition. A Bernoulli variable with parameter p is a random variable which has only two possible outcomes, usually denoted by 0 and 1, and satisfies:
  - $P(X = 0) = 1-p$, and
  - $P(X = 1) = p$.

- Notation. If X follows Bernoulli distribution with parameter p, we write: $X \sim BER(p)$. p can be considered the probability of success.

- Example. Coin toss follows $X \sim Ber(0.5)$.

- Remark. If X is a Bernoulli variable with parameter p then
  - $E(X)=p$ ⇒ $E(X)=1*p+0*(1-p)$
  - $E(X^2)=p$ ⇒ $E(X^2)=1^2*p+0^2*(1-p)=p$
  - $V(X)=p(1-p)$ ⇒ $V(X)=E(X^2)-E(X)^2=p-p^2=p(1-p)$.

Binomial Distribution

- Definition. A binomial random variable is a random variable that counts the number of successes in many independent, identical Bernoulli trials, that is:

$$X=X_1+X_2+\ldots+X_n,$$

  where X_i for i=1,…n is a Bernoulli trial with probability p.

- Independence. The results of each trial do not depend on each other.

- Notation: A binomial random variable is denoted by $X \sim B(n,p)$.

- Example. The number of defective computer chips in a sample of 100 chips.

- Remark. Binomial distribution measures the number of successes.

Binomial Distribution

- Theorem. Let X~B(n,p) be a binomial random variable. Then for x=0,1,2,…n:

$$1.\ P(X=x)=\binom{n}{x}p^x(1-p)^{n-x},$$

$$2.\ E(X)=np,$$

$$3.\ V(X)=np(1-p),$$

- Example. The probability that a chip is not defective is 0.6. What is the probability that when 5 chips are manufactured, there are 2 defective chips?

  Solution. P=0.6, n=5. 2 defects out of 5, so 3 good chips.

$$P(X=3)=\binom{5}{3}0.6^3(1\text{-}0.6)^2=\frac{5!}{3!2!}0.6^3(0.4)^2$$

$$=\frac{3!4\cdot 5}{3!2!}0.6^3(0.4)^2=10\cdot 0.6^3(0.4)^2=0.3456.$$

- Using Binomial Distribution Calculation Table

Binomial Distribution

- Exercise (Aczel, Ex. 3.39). A management graduate is applying for nine jobs, and believes that she has in each of the nine cases a constant and independent 0.48 probability of getting an offer. What is the probability that she will have at least three offers?

Solution. p=0.48, n=9. $P(X=x)=\binom{n}{x}p^x(1-p)^{n-x}$

$$P(X\geq 3)=1-P(X<3)=1-[P(X=0)+P(X=1)+P(X=2)]$$

$$=1-\binom{9}{0}0.48^0(1-0.48)^9-\binom{9}{1}0.48^1(1-0.48)^8-\binom{9}{2}0.48^2(1-0.48)^7$$

$$=1-1*1*0.52^9-\frac{9!}{8!1!}*0.48*0.52^8-\frac{9!}{7!2!}*0.48^2*0.52^7$$

$$=1-0.52^9-\frac{8!*9}{8!*1!}*0.48*0.52^8-\frac{7!*8*9}{7!*2!}*0.48^2*0.52^7$$

$$=1-0.52^9-9*0.48*0.52^8-4*9*0.48^2*0.52^7=0.8889$$

Continuous Random Variable

Continuous Random Variable

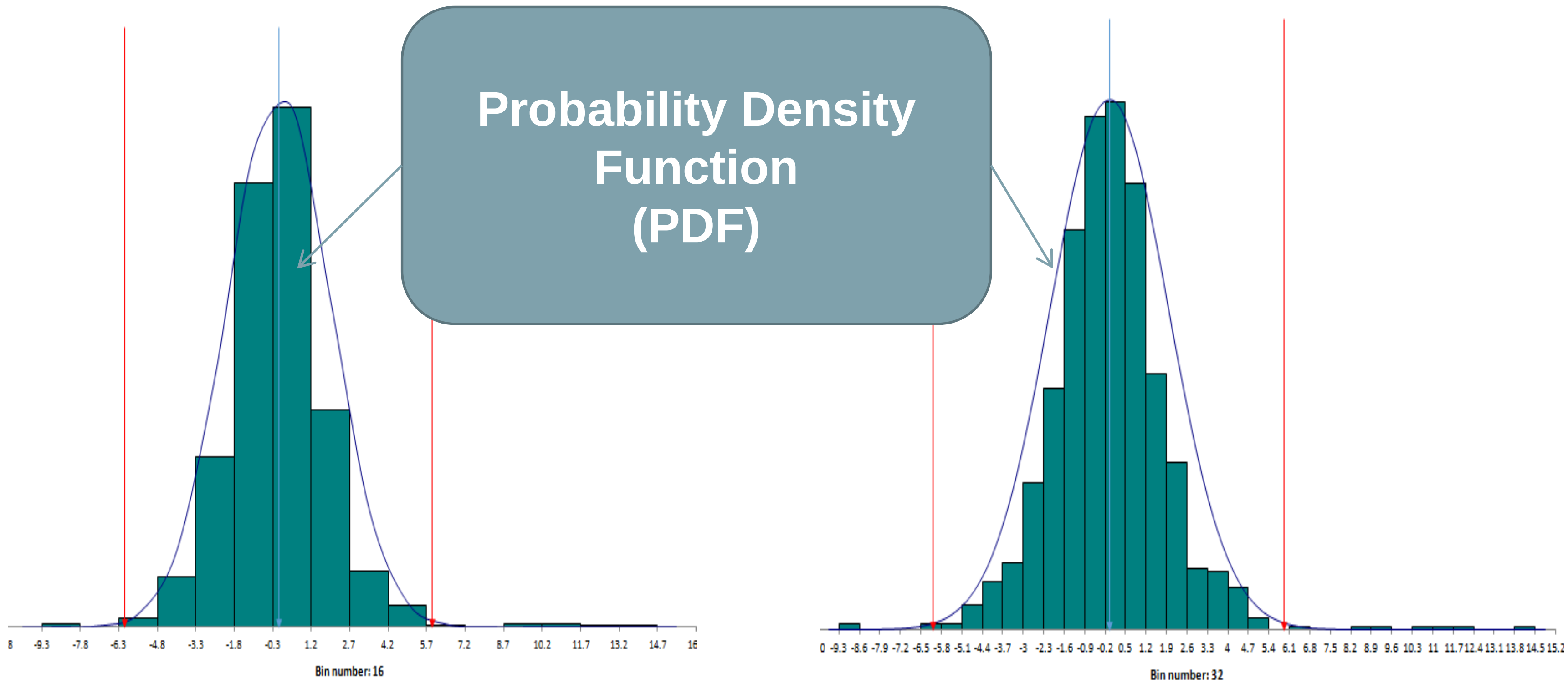
Definition. A continuous random variable is a random variable that can take on any value in an interval of numbers.

Examples:

- Time (e.g. time required to complete a stage in a project)
- Investment risk
- Profit (though one may argue that money has a finite resolution)

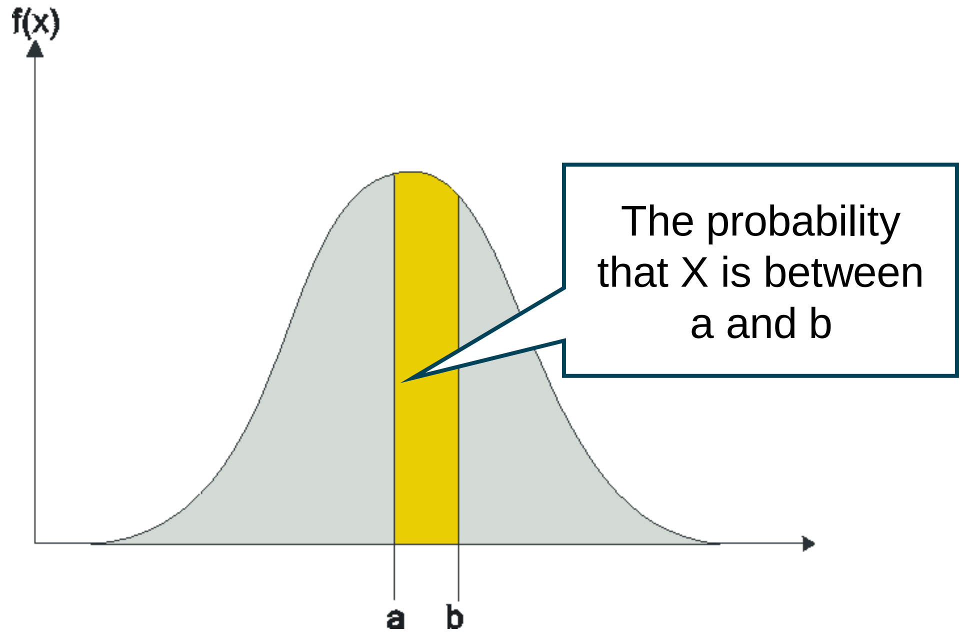
Probability Density Function

- As the number of outcomes increase, a probability distribution of the discrete random variable resembles the probability density function of a continuous variable



Probability Density Function

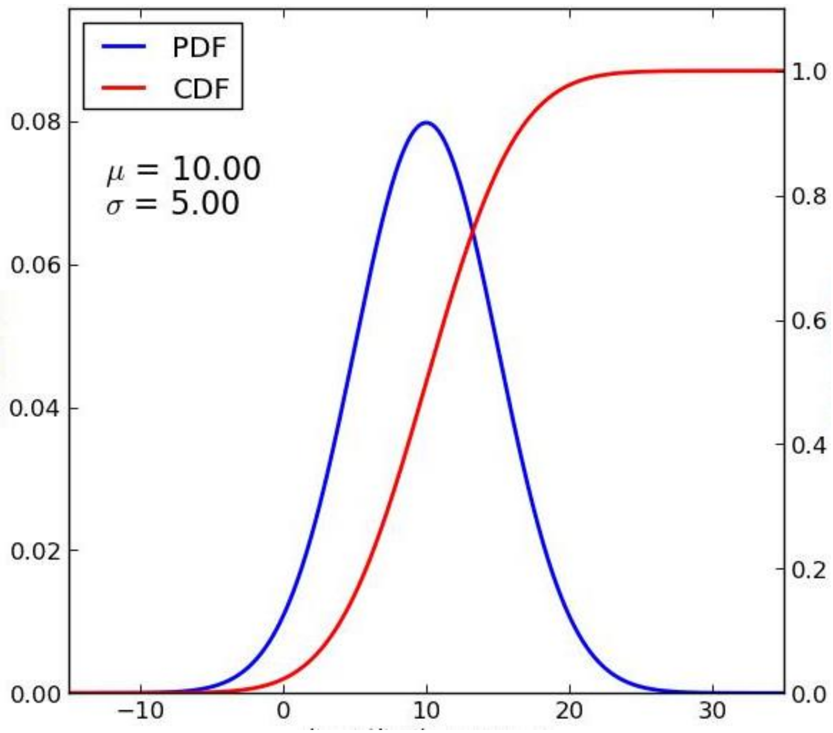
- Properties of the probability density function f(x):
  1. f(x)≥0 for all x
  2. The total area under f(x) is 1
  3. The probability that X is between a and b is the area under f(x) between a and b.



Cumulative Distribution Function

- Definition. The Cumulative Distribution Function (CDF) of a continuous random variable is F(x)=P(X ≤x).

- The meaning of the cumulative distribution function:
  - The area under f (x) between the smallest possible value of X and x
  - The probability that X ≤x .



UCL SCHOOL OF MANAGEMENT

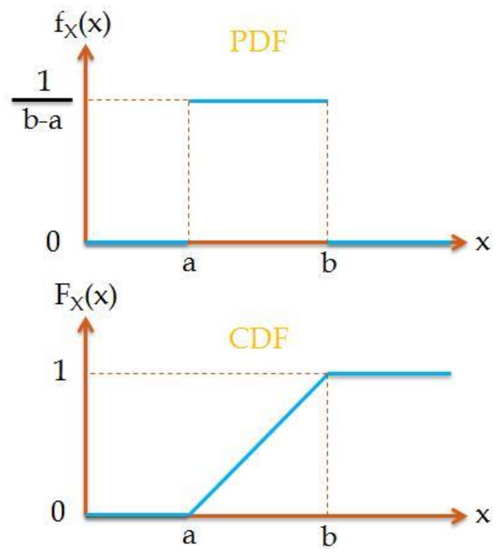
Uniform Distribution

- Definition. If X follows a uniform distribution between a and b, then we write: X~U(a,b).

- The probability density function of the uniform distribution:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{Otherwise} \end{cases}$$

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

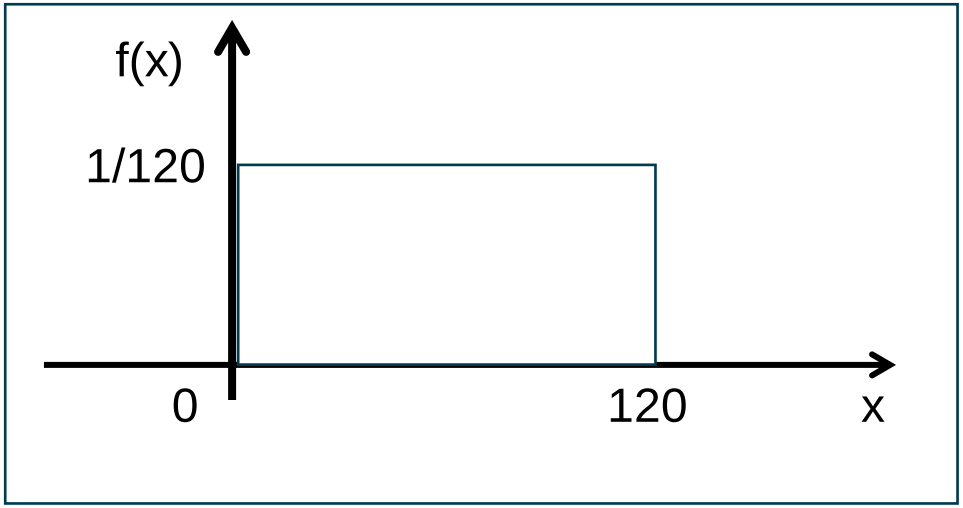


- Remark. If X~U(a,b) and a ≤ c ≤ d ≤ b then:
  1. P(c≤x ≤d)=(d-c)/(b-a)
  2. E(X)=(a+b)/2
  3. V(X)=(b-a)^2 /12.

Uniform Distribution

- Example. Assume that the time between two online orders of customers has a uniform distribution between 0 to 120 seconds. What is the probability that the time between two online orders is:

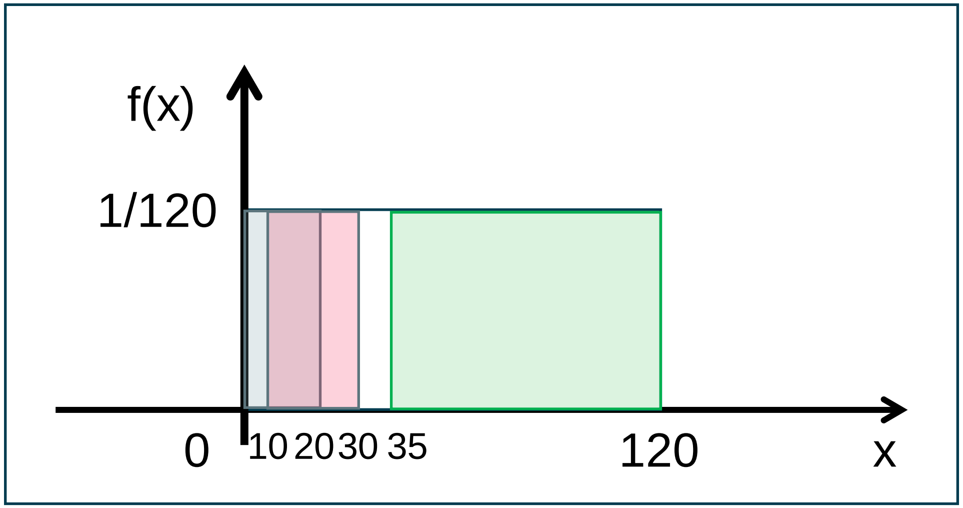
1. less than 20 seconds?
2. between 10 and 30 seconds?
3. more than 35 seconds?
4. What are the mean and standard deviation of the time between online orders?



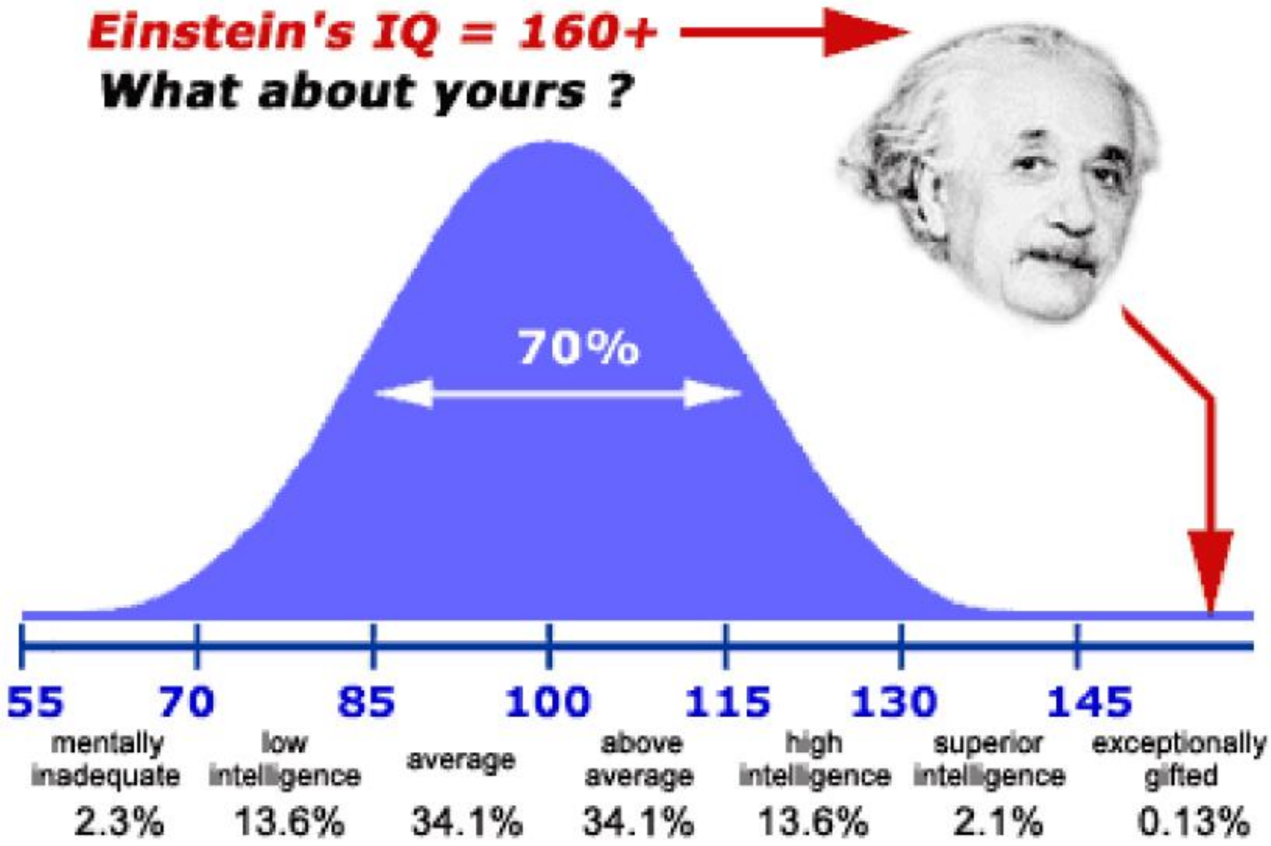
UCL SCHOOL OF MANAGEMENT

Uniform Distribution

- Solution.
  1. $P(X \leq 20) = P(0 \leq X \leq 20) = ((20-0) \cdot 1/120 = 1/6 = 0.1667.$
  2. $P(10 \leq X \leq 30) = (30-10) \cdot 1/120 = 1/6 = 0.1667.$
  3. $P(X \geq 35) = (120-35) \cdot 1/120 = 0.7083.$
  4. $E(X) = (a+b)/2 = (0+120)/2 = 60$. $V(X) = (b-a)^2 / 12 = (120-0)^2 / 12 = 1200$.
     $SD = 1200^{0.5} = 34.641$.



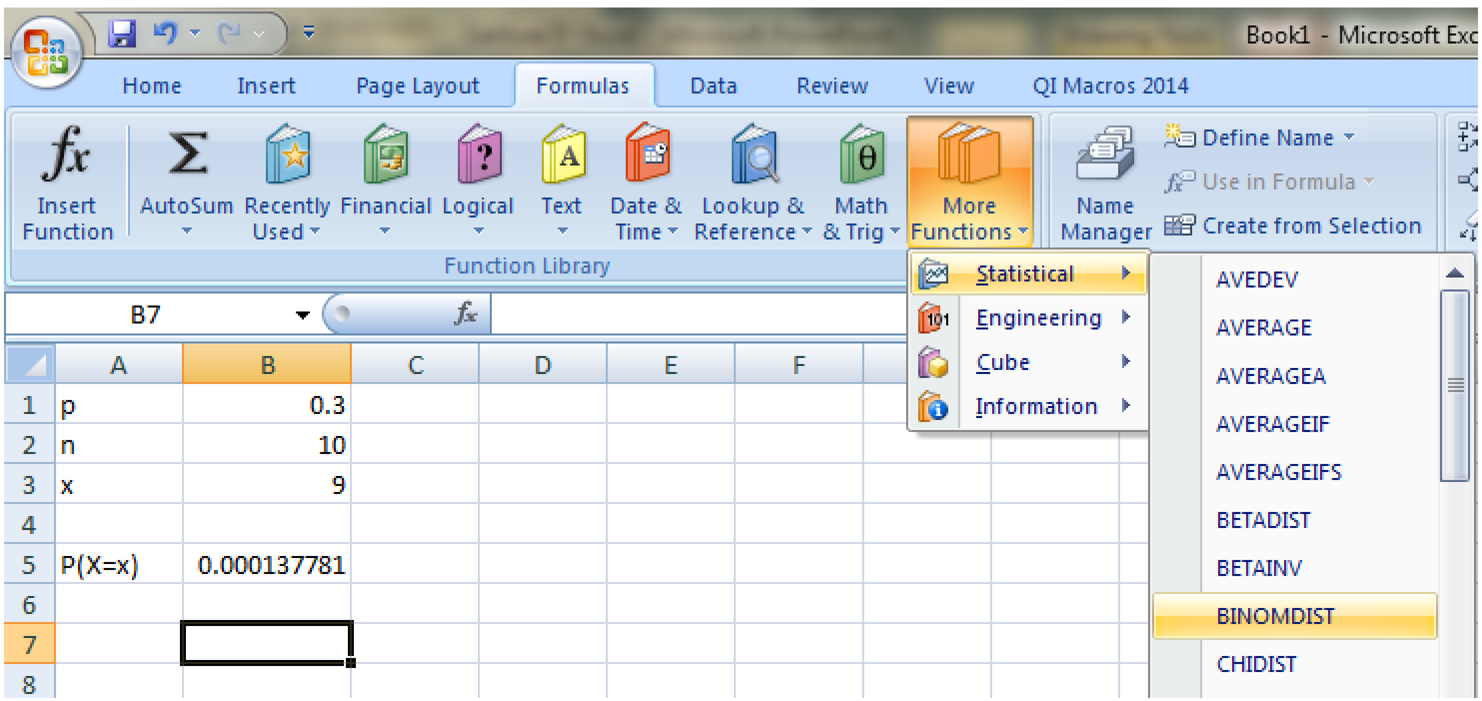
Normal Distribution (Next Lecture)



Excel: Calculation of Binomial Distribution

Excel Function: BinomDist

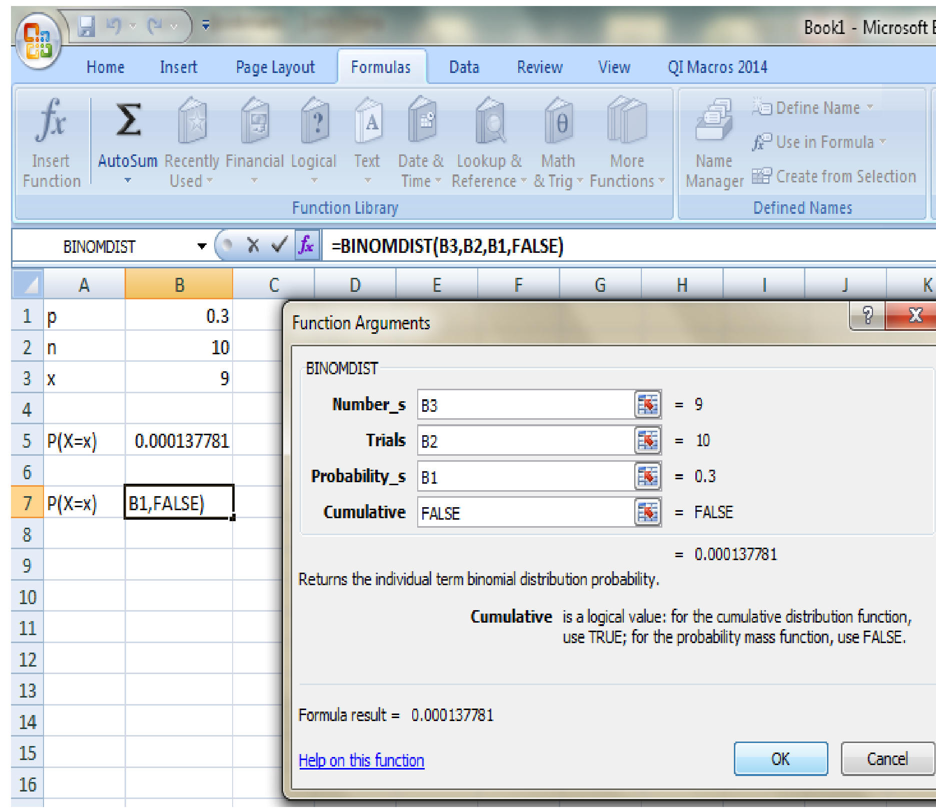
- To get to this function go to Formulas->More functions->Statistical->BinomDist
- We use X~B(10, 0.3) as an example.



UCL SCHOOL OF MANAGEMENT

Excel Function: BinomDist

- Number_s=number of successes (x)
- Trials=n
- Probability_s=probability of success
- Cumulative
  - FALSE for probability
  - TRUE for the cumulative probability



Excel Exercise

- Exercise (Aczel, Ex. 3.39). A management graduate is applying for nine jobs, and believes that she has in each of the nine cases a constant and independent 0.48 probability of getting an offer. What is the probability that she will have at least three offers?

- Solution in Excel.

Excel Exercise

In the final exam of MSIN0017, suppose for each student, the probability of passing the final exam is 0.9, and we have 100 students in the class, uses Excel to calculate the following:

- the probability of exactly 80 students pass the final exam
- the probability of all students pass the final exam
- the probability of at most 20 students fail the exam
- generate the probability distribution function and the cumulative distribution function, and draw a picture containing the two functions

Recommended reading

Chapter 3 of:

Aczel, A., & J. Sounderpandian. 2008. Complete Business Statistics. McGraw-Hill/Irwin, Seventh Edition.

Watch:

- https://www.youtube.com/watch?v=lHCpYeFvTs0

*The course leader thanks Dr. Daphne Sobolev for her help in developing the course materials.

Mathematical Background: Combinatorics

Definition. Let n be a natural number. $n!=1\cdot 2 \cdot \ldots \cdot n$.

Examples.

$3!=1\cdot 2\cdot 3=6$

$4!=1\cdot 2\cdot 3\cdot 4=24.$

$100!=9.3326e+157$

Theorem. The number of ways to choose k elements from a set of n elements is given by $\binom{n}{k}=\frac{n!}{k!(n-k)!}$

By definition, $\binom{n}{0}=1$

Example. Calculate $\binom{5}{3}$

Solution. $\binom{5}{3}=\frac{5!}{3!(5-3)!}=\frac{5!}{3!2!}=\frac{1\cdot 2\cdot 3\cdot 4\cdot 5}{1\cdot 2\cdot 3\cdot 1\cdot 2}=\frac{4\cdot 5}{1\cdot 2}=10.$

Example

In how many ways can a manager choose a team of 3 people from a department with 7 people to perform a certain task?

Solution.

N=7, k=3.

$$\binom{7}{3} = \frac{7!}{3!(7-3)!} = \frac{7!}{3!4!} = \frac{1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6\cdot 7}{1\cdot 2\cdot 3\cdot 1\cdot 2\cdot 3\cdot 4} = \frac{5\cdot 6\cdot 7}{1\cdot 2\cdot 3} = 5\cdot 7 = 35$$

Therefore, there are 35 possible teams of 3 people the manager should choose from.

Denote the employees by 1,2,..7.
The possible teams are:

1 2 3	1 2 4	1 2 5	1 2 6	1 2 7
1 3 4	1 3 5	1 3 6	1 3 7	
1 4 5	1 4 6	1 4 7		
1 5 6	1 5 7			
1 6 7				
2 3 4	2 3 5	2 3 6	2 3 7	
2 4 5	2 4 6	2 4 7		
2 5 6	2 5 7			
2 6 7				
3 4 5	3 4 6	3 4 7		
3 5 6	3 5 7			
3 6 7				
4 5 6	4 5 7			
4 6 7				
5 6 7				