

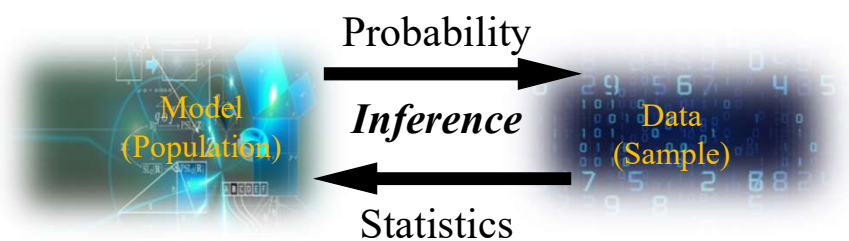
Introduction to Statistical Methods

Argon Chen
Graduate Institute of Industrial Eng'g
National Taiwan University

1

©Argon Chen

Probability and Statistics



2

©Argon Chen

An Example: Taiwan Big Lotto (大樂透)

- 6 winning numbers and 1 special number chosen from 49 numbers
- Mr. Chang chooses numbers randomly and never believes in any historical analysis of number appearance
- Mr. Fang chooses numbers that most frequently appear in the history
- Mr. Wang chooses numbers that most rarely appear in the history
- Mr. Yang chooses meaningful numbers, such as the date of birthday
- **Who is correct?**

3

©Argon Chen

Probability of Taiwan Big Lotto

- 6 numbers chosen from 49 numbers
- Chance of winning the first prize?
- Answer: $1 / C_6^{49}$
- Is this based on probability or statistics?
- What is the model and assumption behind the answer above?

Every number has the *identical* probability to be chosen *independently* each time! \Rightarrow *iid* assumption
- The event of winning is guessed (inferred) by the “model”.

4

©Argon Chen

Statistics of Taiwan Big Lotto

- What is “Statistics” in the Lotto problem?
- Is the model assumption correct?
- What is the real chance of winning? and the chance for each number to appear?
- “Statistics” is to estimate the appearance probability of each number. How?
by collecting and observing the data.
⇒ “Model” is **inferred** by the “sample data”
⇒ Statistical Inference

5

©Argon Chen

Taiwan Big Lotto (Cont’d)

- Back to Mr. Chang, Fang, and Wang
- Who is correct?
- What does Mr. Chang believe?
- What does Mr. Fang believe?
- What does Mr. Wang believe?
- What does Mr. Yang believe?

6

©Argon Chen

A note on “Models”

“All models are wrong, but some are useful.”
- George E. P. Box (1979)

7

©Argon Chen

Definition of “Experiment”

- An **experiment** is the **process** by which an **observation** (or measurement) is obtained.
- **Experiment:** Record an age
- **Experiment:** Toss a die
- **Experiment:** Record an opinion (yes, no)
- **Experiment:** Toss two coins
- **Experiment:** numbers of a lotto ticket

8

©Argon Chen

Definition of “Outcome”

- An **outcome** is observed on a single repetition of the experiment.
 - Basic element to which probability is applied.
 - One and only one outcome can occur when experiment is performed.
- An outcome is denoted by **O** with a subscript.

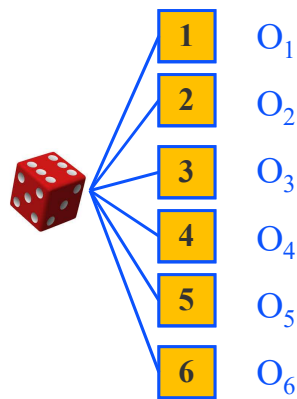
Definition of “Outcome Space”

- Each outcome is assigned a probability, measuring “how often” it occurs.
- Set of all outcomes of an experiment is called the **outcome space**, usually denoted by **S**.

Example of Experiment and Outcome

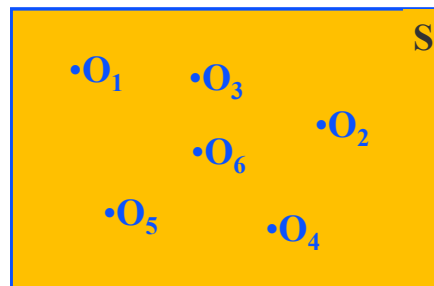
- The die toss experiment:

- Outcomes:



Outcome space:

$S = \{O_1, O_2, O_3, O_4, O_5, O_6\}$
(or $S = \{1, 2, 3, 4, 5, 6\}$)



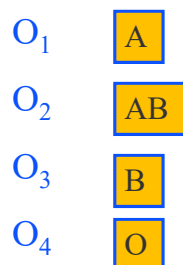
11

©Argon Chen

Example

- Record a person's blood type:

- Outcomes:



Outcome space:

$S = \{O_1, O_2, O_3, O_4\}$

$S = \{A, B, AB, O\}$

12

©Argon Chen

Example

- Record the numbers of a lotto ticket:
- Outcomes: Outcome space:
 $O_1 = [1, 2, 3, 4, 5, 6]$ $S = \{O_1, O_2, O_3, O_4, \dots,$
 $O_2 = [2, 3, 4, 5, 6, 7]$ $\dots\dots\dots, O_7\}$
 $O_3 = [3, 4, 5, 6, 7, 8]$
 \vdots
 \vdots
 $O_7 = [44, 45, 46, 47, 48]$

13

©Argon Chen

Definition of “Event”

- An **event** is a collection of one or more **outcomes**.

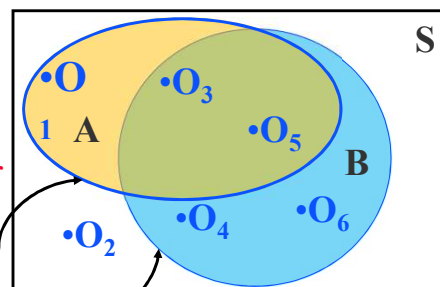
- The die toss events:

✓ Event A: an odd number

✓ Event B: a number > 2

$$A = \{O_1, O_3, O_5\}$$

$$B = \{O_3, O_4, O_5, O_6\}$$



Venn Diagram

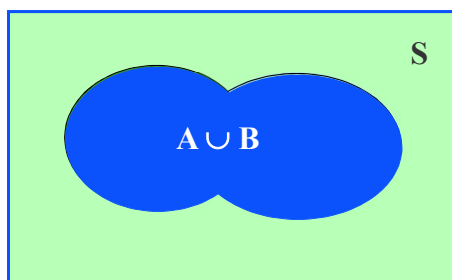
14

©Argon Chen

Event Relations - Union

The **union** of two events, A and B, is the event that either A **or** B **or both** occur when the experiment is performed. We write

$$A \cup B$$

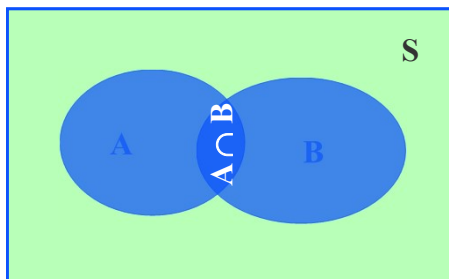


15

©Argon Chen

Event Relations-Intersection

The **intersection** of two events, A and B, is the event that both A **and** B occur when the experiment is performed. We write $A \cap B$.



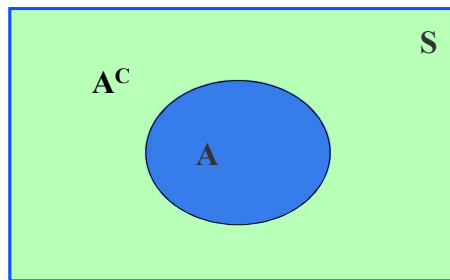
- If two events A and B are **mutually exclusive**, then $A \cap B = \emptyset$.

16

©Argon Chen

Event Relations - Complement

The **complement** of an event **A** consists of all outcomes of the experiment that do not result in event **A**. We write A^C (The event that event **A** doesn't occur).



17

©Argon Chen

Example

Select a student from a college

- **A**: student is colorblind
- **B**: student is biologically female
- **C**: student is biologically male

Mutually exclusive and $B = C^C$

• What is the relationship between events **B** and **C**?

• A^C : Student is not colorblind

• $B \cap C$: Student is both biologically male and female = \emptyset

• $B \cup C$: Student is either male or female = all students = S

18

©Argon Chen

Factorial and Sequences

Example: How many possible distinct sequences of 5 different numbers among 1, 2, 3, 4 and 5?

The “order” of the choices

$$5(4)(3)(2)(1) = 120$$

The number of distinct sequences you can arrange n distinct objects is n factorial denoted as:

$$n! = n(n-1)(n-2)\dots(2)(1) \text{ and } 0! \equiv 1.$$

Permutations

Example: How many 3-digit lock passwords can we make by using 3 **different numbers** among 1, 2, 3, 4 and 5?

The order of the choice is important!

$$5(4)(3) = 60$$

The number of ways you can arrange n distinct objects, taking them r at a time is

$$P_r^n = \frac{n!}{(n-r)!}$$

$$P_3^5 = \frac{5!}{(5-3)!} = \frac{5(4)(3)(2)(1)}{2(1)} = 60$$

Combinations

- The number of distinct combinations of n **distinct objects** that can be formed, **taking them r at a time** is $C_r^n = \frac{n!}{(n-r)!r!} = \frac{n!}{r!(n-r)!}$ since the number of permutations is $r!$ times that of combinations

Example: 3 members of a 5-person faculty must be chosen to form a committee. How many different committees could be formed?

The order of the choice is not important!

$$C_3^5 = \frac{5!}{3!(5-3)!} = \frac{5(4)(3)}{3(2)(1)} = \frac{5(4)}{(2)1} = 10$$

21

©Argon Chen

The Probability of an Event

- $P(A)$ must be between 0 and 1.
 - If event A can never occur, $P(A) = 0$.
 - If event A always occurs when the experiment is performed, $P(A) = 1$.
- Sum of probabilities for all simple events in S equals 1. $P(S) = 1$.
- **Probability of an event A is found by adding the probabilities of all outcomes contained in A**

22

©Argon Chen

Taiwan Big Lotto

- 6 numbers chosen from 49 numbers
- First prize: all six numbers

Event probability:

$$1/C_6^{49} = \frac{6!(49-6)!}{49!}$$

- Second prize: 5 numbers + 1 special number

Event probability?

$$C_5^6 / C_7^{49}$$

Why? Model assumptions?

23

©Argon Chen

Mutually Exclusive Events

- When A and B have no outcomes in common, they are said to be **mutually exclusive** or **disjoint** events
- Experiment: Toss a die

–A: observe an odd number

Not Mutually
Exclusive

–B: observe a number greater than 2

–C: observe a 6

–D: observe a 3

Mutually
Exclusive

A and C?
A and D?
B and C?

24

n

Conditional Probability

- For any two events A and B with $P(B) > 0$, the **conditional probability** of A given that B has occurred is defined by
$$P(A|B) = P(A \cap B) / P(B)$$

25

©Argon Chen

Independent Events

- Two events A and B are **independent** if $P(A|B) = P(A)$ and are dependent otherwise
- In other words,
$$P(A|B) = P(A \cap B) / P(B) = P(A)$$

$$\Rightarrow P(A \cap B) = P(A) P(B) \text{ if } A \text{ and } B \text{ are independent}$$

26

©Argon Chen

Basic Probability Rules

- Rule 1: Addition rule for **mutually exclusive** events $P(A \text{ or } B) = P(A) + P(B)$
- Rule 2: Multiplication rule for **independent** events $P(A \text{ and } B) = P(A)P(B)$
- Rule 3: General addition rule

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$
- Rule 4: General multiplication rule

$$P(A \text{ and } B) = P(A)P(B|A)$$

27

©Argon Chen

Sensitivity and Specificity

Test: disease, fire, quality, etc.
Testing performance?

		True Condition		
		Condition positive (CP)	Condition negative (CN)	
Predicted Condition	Test outcome positive (P)	True positive $(TP) = 20$ $\{P \cap CP\}$	False positive $(FP) = 180$ $\{P \cap CN\}$	Positive predictive value (PPV/Precision) $P(CP P)$ $= TP / (TP + FP)$ $= 20 / (20 + 180)$ $= 10\%$
	Test outcome negative (N)	False negative $(FN) = 10$ $\{N \cap CP\}$	True negative $(TN) = 1820$ $\{N \cap CN\}$	Negative predictive value (NPV) $P(CN N)$ $= TN / (FN + TN)$ $= 1820 / (10 + 1820) \approx 99.5\%$
		Sensitivity (Recall) $P(P CP)$ $= TP / (TP + FN)$ $= 20 / (20 + 10)$ $\approx 67\%$	Specificity $P(N CN)$ $= TN / (FP + TN)$ $= 1820 / (180 + 1820)$ $= 91\%$	

28

©Argon Chen

Meanings of Sensitivity and Specificity

Sensitivity (Recall): "I know my patient has the disease. What is the chance that the test will show that my patient has it?"

Specificity: "I know my patient doesn't have the disease. What is the chance that the test will show that my patient doesn't have it?"

Meanings of PPV and NPV

PPV (Precision/Risk): "I just got a positive test result back on my patient. What is the chance (risk) that my patient actually has the disease?"

NPV: "I just got a negative test result back on my patient. What is the chance that my patient actually doesn't have the disease?"

Total Probability and Bayes

- Let A_1, \dots, A_n be **mutually exclusive** and **exhaustive** events. Then for any other event B : (**law of total probability**)

$$P(B) = \sum_{i=1}^n P(B | A_i)P(A_i)$$

- Bayes Thm:** Let A_1, \dots, A_n be a collection of n **mutually exclusive** and **exhaustive** events with $P(A_i) > 0$ for $i=1, \dots, n$. Then for any other event B for which $P(B) > 0$

$$\underbrace{P(A_k | B)}_{\text{a posteriori inference}} = \frac{P(A_k \cap B)}{P(B)} = \frac{\underbrace{P(B | A_k)P(A_k)}_{\text{a priori knowledge}}}{\sum_{i=1}^n \underbrace{P(B | A_i)P(A_i)}}_{\text{a posteriori inference}}$$

31

©Argon Chen

Example 1

It is known that the H1N1 flu, a subtype of influenza type A also known as swine flu, prevalence is 10% (the proportion of population infected) in the pandemic period (2009~2010). A testing method has been developed to quickly test if a patient is infected with the influenza virus. This method is called Rapid Influenza Diagnostic Test (RIDT). Only 60% (sensitivity) of patients infected with H1N1 can be tested positive (40% of false negative) and 95% (specificity) of patients without H1N1 infection are tested negative. Only patients tested positive can be treated with Tamiflu administering. If a patient is not treated with Tamiflu administering, what is the probability that the patient is infected with H1N1 without treatment of Tamiflu?

32

©Argon Chen

Example 1 (Cont'd)

- a priori knowledge:
 - A_1 : infected with H1N1 $P(A_1)=10\%$
 - A_0 : without infection with $P(A_0)=1-P(A_1)$
 - B_1 : tested positive
 - B_0 : tested negative
 - $P(B_1|A_1)=0.6$; $P(B_0|A_0)=0.95$
- a posteriori question: $P(A_1 | B_0)=?$

33

©Argon Chen

Complete Testing Evaluation Table

		True condition			
		Condition positive	Condition negative	Prevalence = $\frac{\sum \text{Condition positive}}{\sum \text{Total population}}$	Accuracy (ACC) = $\frac{\sum \text{True positive} + \sum \text{True negative}}{\sum \text{Total population}}$
Predicted condition	Predicted condition positive	True positive	False positive, Type I error	Positive predictive value (PPV), Precision = $\frac{\sum \text{True positive}}{\sum \text{Predicted condition positive}}$	False discovery rate (FDR) = $\frac{\sum \text{False positive}}{\sum \text{Predicted condition positive}}$
	Predicted condition negative	False negative, Type II error	True negative	False omission rate (FOR) = $\frac{\sum \text{False negative}}{\sum \text{Predicted condition negative}}$	Negative predictive value (NPV) = $\frac{\sum \text{True negative}}{\sum \text{Predicted condition negative}}$
		True positive rate (TPR), Recall, Sensitivity, probability of detection, Power = $\frac{\sum \text{True positive}}{\sum \text{Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm = $\frac{\sum \text{False positive}}{\sum \text{Condition negative}}$	Positive likelihood ratio (LR+) = $\frac{\text{TPR}}{\text{FPR}}$	Diagnostic odds ratio (DOR) = $\frac{\text{LR+}}{\text{LR-}}$ F ₁ score = $2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$
		False negative rate (FNR), Miss rate = $\frac{\sum \text{False negative}}{\sum \text{Condition positive}}$	Specificity (SPC), Selectivity, True negative rate (TNR) = $\frac{\sum \text{True negative}}{\sum \text{Condition negative}}$	Negative likelihood ratio (LR-) = $\frac{\text{FNR}}{\text{TNR}}$	

34

©Argon Chen

Example 2

Only one in 1000 chips is defective by not conforming to a certain type of electrical test. The test is performed by a tester that detects the failure 99% of the time when the chip is actually defective. The tester misidentified a good chip only 2% of the time. If a randomly selected chip is tested defective, what is the probability that the chip is actually defective.

35

©Argon Chen

Variables with Random Nature

- **Definition:** For a given sample space of some experiment, a **random variable** is **any rule** that associates a number with each outcome in the sample space. A random variable is always denoted by a **capital letter** (e.g. X , Y , etc)
- Can you think of anything that is not random in nature? Can you think of two things that are exactly the same?

36

©Argon Chen

Random Variables

- What are random variables?
 - Life is a never ending experiment
 - **Define the rules of experiment with its outcome as a random variable**
- Types of random variables
 - **discrete** if its set of possible values is a discrete set
 - **continuous** if its set of possible values is an entire interval of numbers
- Examples: defects, c.d., thickness

37

©Argon Chen

Probability Distribution

- The **probability distribution** or **probability mass function (p.m.f)** of a **discrete** random variable is defined for every number x by $p(x) = P(X=x)$
- Let X be a **continuous random variable**. Then a probability distribution or **probability density function (p.d.f)** of X is a function $f(x)$ such that for any two numbers a and b with $a \leq b$

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

38

©Argon Chen

Cumulative Distribution Function (c.d.f)

- The **cumulative distribution function (c.d.f)** $F(x)$ of a discrete (continuous) random variable X with p.m.f $p(x)$ (p.d.f. $f(x)$) is defined for every number x by

$$F(x) = P(X \leq x) = \sum_{y: y \leq x} p(y)$$

$$(\quad = \int_{-\infty}^x f(y) dy)$$

39

©Argon Chen

Expected Value (Mean)

- The **expected (or mean value)** of a discrete (continuous) r.v. X with p.m.f $p(x)$ (p.d.f $f(x)$) is a **weighted average** weighted by $p(x)$ or $f(x)$:

$$E(X) = \mu_X = \sum_{x \in D} x \bullet p(x)$$

$$(\quad = \int_{-\infty}^{\infty} x \bullet f(x) dx)$$

where D represents the set of possible values for discrete r.v. X

40

©Argon Chen

Mean of a Function

- If the discrete (continuous) r.v. X has a set of possible values D and p.m.f (p.d.f) $p(x)$ ($f(x)$), then the **expected value** of any function $h(X)$, denoted by $E[h(X)]$ or, is computed by

$$E[h(X)] = \mu_{h(X)} = \sum_{x \in D} h(x) \cdot p(x) (= \int_D h(x) \cdot f(x) dx)$$

41

©Argon Chen

Variance

- The **variance** of a discrete (continuous) r.v. X with p.m.f. $p(x)$ (p.d.f $f(x)$) and mean value μ is

$$\begin{aligned} \sigma^2 = V(X) &= E[(X - \mu)^2] = \sum_{x \in D} (x - \mu)^2 \cdot p(x) \\ &= \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) dx \end{aligned}$$

- The standard deviation of X is $\sigma_X = \sqrt{V(X)}$

42

©Argon Chen

Rules of Mean and Variance

- **Rules of Expected Values:**

$$E(aX+b)=aE(X) + b$$

- **Shortcut Formula for $V(X)$:**

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

- **Rules of Variance:**

$$V(aX+b)=a^2V(X)$$

43

©Argon Chen

Family of Probability Distributions

- Suppose that $p(x)$ depends on a quantity that can be assigned any one of a number of possible values, with each different value determining a different probability distribution. Such a quantity is called a **parameter** of the distribution. The collection of all probability distributions for different values of the **parameter** is called a **family** of probability distributions

44

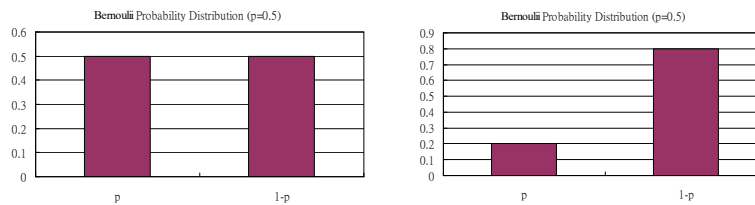
©Argon Chen

Discrete Distribution Models

- **Bernoulli Distribution:** tossing coin
r.v. $X = 0$ or 1
 $P(X=0)=1-p$; $P(X=1)=p$

One parameter: $0 < p < 1$

Example: tossing coin to get the head where the probability of head is p



Mean: p Variance: $p(1-p)$

45

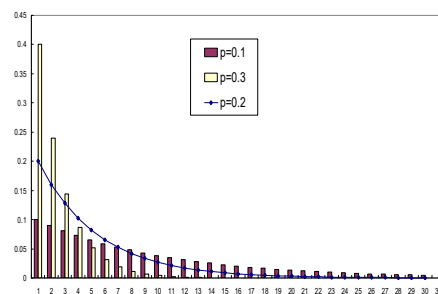
©Argon Chen

Geometric Distribution Family

- Tossing coin
r.v. X : number of tosses before first head appears where the probability of head is p

$$P(X = x) = p(1 - p)^{x-1}$$

$$P(X < x) = 1 - (1 - p)^{x-1}; P(X \geq x) = (1 - p)^{x-1}$$



Mean: ?

Variance: $(1-p)/p^2$

46

- Memoryless property: $P(X > m+n | X > n) = P(X > m)$

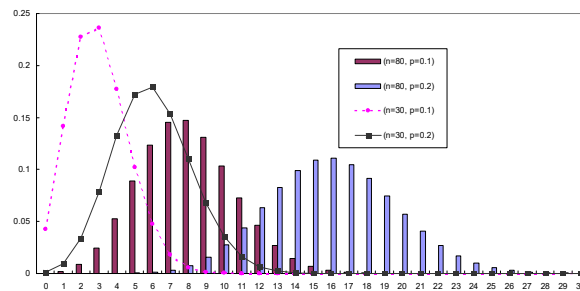
©Argon Chen

Binomial Distribution Family

- Tossing coin with **different r.v. definition**

X : number of heads in n tosses

$$b(x; n, p) = \begin{cases} C_x^n p^x (1-p)^{n-x} & x = 0, 1, 2, \dots, n \\ 0 & \text{otherwise} \end{cases}$$



Mean: ? Variance: $np(1-p)$

47

©Argon Chen

Derivation of Binomial Mean

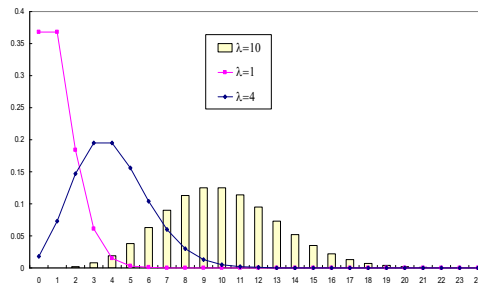
$$\begin{aligned} E(X) &= \sum_{x=0}^n x \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^n \frac{n!}{x!(n-x)!} x p^x (1-p)^{n-x} \\ &= \sum_{x=0}^n \frac{np(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x} \text{ let } k = x-1 \\ &= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k} = np \end{aligned}$$

48

©Argon Chen

Poisson Distribution Family

$$p(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!} \quad x = 0, 1, 2, \dots \text{ for some } \lambda > 0$$



- $E(X) = V(X) = \lambda$
- **Poisson ~ binomial with $n \rightarrow \infty$ and $p \rightarrow 0$**
- Example: defects on a wafer, communication network, car accidents

49

©Argon Chen

Derivation of Poisson Mean

$$E(X) = \sum_{x=0}^{\infty} x \frac{e^{-\lambda} \lambda^x}{x!}$$

$$= \sum_{x=1}^{\infty} \frac{e^{-\lambda} \lambda^x}{(x-1)!}$$

$$= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \quad \text{let } k = x-1$$

$$= \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{(k)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

50

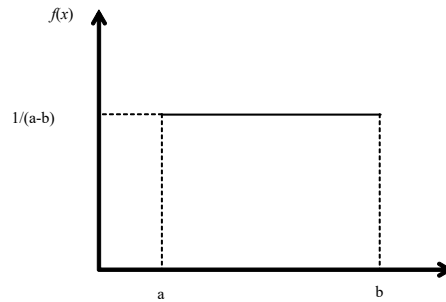
©Argon Chen

Continuous Distribution Models

- Uniform Distribution family:

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$$

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



- Mean? Variance?

51

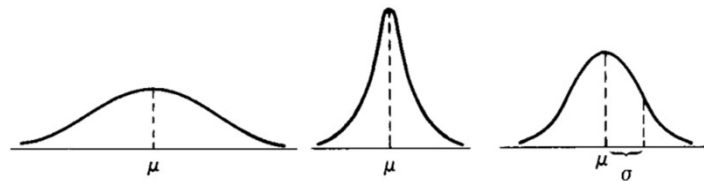
©Argon Chen

Normal Distribution Family

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

$$F(X = x; \mu, \sigma) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-(X-\mu)^2/2\sigma^2} dX \quad (\text{no close form})$$

- Mean: μ and Variance: σ^2



52

©Argon Chen

Standard Normal Distribution

- Normal distribution with mean 0 and standard deviation 1: $f(z;0,1)$; $F(x)=\Phi(x)$

$$\Phi(x) = F(X = x; 0, 1) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-X^2/2} dX \quad (\text{no close form})$$

- Proposition: If X follows a normal distribution with mean μ and standard deviation σ , then

$$Z = \frac{X - \mu}{\sigma}$$

is a standard normal random variable

- $F(x; \mu, \sigma) = \Phi(z = (x - \mu)/\sigma) \Rightarrow$ Table of $\Phi(x)$

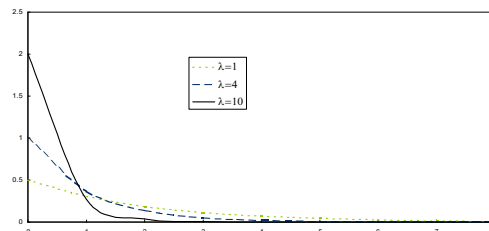
53

©Argon Chen

Exponential Distribution Family

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{where } \lambda > 0$$

- c.d.f. $F(x; \lambda) = \begin{cases} 0 & X < 0 \\ 1 - e^{-\lambda x} & X \geq 0 \end{cases}$; $P(X > x) = e^{-\lambda x}$



mean = $1/\lambda = \sigma$

variance = $1/\lambda^2$

- Memoryless property: $P(X > t + s | X > t) = P(X > s)$

54

©Argon Chen

Gamma Function

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx \quad \alpha > 0$$

- For $\alpha > 1$, $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$
- For any positive integer n , $\Gamma(n) = (n-1)!$
- $\Gamma(1/2) = \sqrt{\pi}$

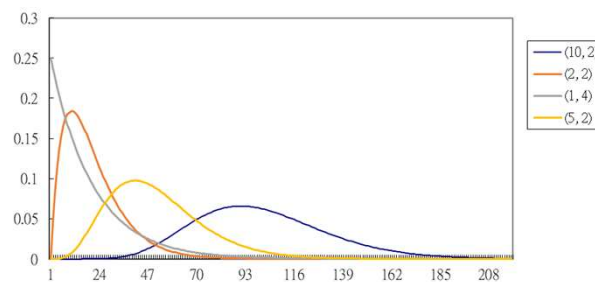
55

©Argon Chen

Gamma Distribution Family

$$f_{\Gamma}(x; \alpha, \beta) = \begin{cases} \frac{1}{\beta^{\alpha} \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

- $E(X) = \alpha\beta \quad V(X) = \alpha\beta^2$



56

©Argon Chen

Chi-square (χ^2) Distribution Family

$$f(x; \nu) = \begin{cases} \frac{1}{2^{\nu/2} \Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

- $\chi^2(\nu) \equiv f_{\Gamma}(\alpha=\nu/2, \beta=2)$
- $E(X)=? \quad V(X)=?$
- $X_i \sim \text{standard normal } N(0,1)$
 $\Rightarrow \sum_{i=1}^{\nu} X_i^2 \sim \chi^2(\nu)$

57

©Argon Chen

Discrete Joint Distribution and R.V.

- Let X and Y be two discrete random variables defined on a sample space of an experiment. The joint probability mass function $p(x,y)$ is defined for each pair of numbers (x, y) by $p(x, y) = P(X=x \text{ and } Y=y)$
- For any discrete set A consisting of (x,y) values,

$$P[(X, Y) \in A] = \sum_{(x,y) \in A} p(x, y)$$

58

©Argon Chen

Continuous Joint Distribution and R.V.

- Let X and Y be two continuous random variables. Then $f(x,y)$ is the joint probability density function for X and Y if for any two-dimensional set A

$$P[(X,Y) \in A] = \iint_A f(x,y) dx dy$$

- In particular,

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

59

©Argon Chen

Marginal Probability

- Marginal probability mass (density) functions:

$$p_X(x) = \sum_y p(x,y), \quad p_Y(y) = \sum_x p(x,y)$$

$$(f_X(x) = \int_{-\infty}^{\infty} f(x,y) dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y) dx)$$

60

©Argon Chen

Independence of Two R.V.'s

- Two discrete (continuous) random variables X and Y are said to be **independent** if for every pair of x and y values,

$$p(x, y) = p_X(x) \cdot p_Y(y)$$

$$f(x, y) = f_X(x) \cdot f_Y(y)$$

- If above is not satisfied for all (x, y) , X and Y are said to be **dependent**

61

©Argon Chen

Covariance of Two R.V.'s

- Variance of two discrete (continuous) random variables X and Y is defined:

$$\begin{aligned}\sigma(X, Y) &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY - XE[Y] - E[X]Y + E[X]E[Y]] \\ &= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y]\end{aligned}$$

- Then X and Y are **independent**:

$$E[XY] = E[X]E[Y] \text{ (why?)} \Rightarrow \sigma(X, Y) = 0$$

- However**, if $\sigma(X, Y) = 0$, X and Y may not be **independent unless they jointly normally distributed**.

62

©Argon Chen

Linear Combination of R.V.'s

- Linear Combination of r.v. X_i :

$$Y = \sum_{i=1}^n a_i X_i$$

- For independent X_i 's:

$$E\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i E(X_i)$$

$$V\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 V(X_i)$$

63

©Argon Chen

Average of Random Variables

- For independent and identically distributed (*iid*) X_i 's with μ and σ

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

- Mean and Variance:

$$E(\bar{X}) = \mu \text{ and } V(\bar{X}) = \frac{\sigma^2}{n}$$

64

©Argon Chen

Sum of R.V.'s

- Sum of r.v. X_i :

$$Y = \sum_{i=1}^n X_i$$

- For independent X_i 's:

$$E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i)$$

$$V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i)$$

- How about distribution of Y ? Not an easy answer!

65

©Argon Chen

Sum of Ind. Geometric RVs

- Sum of independent r.v. X_i : $Y = \sum_{i=1}^n X_i$

$$E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = n/p$$

$$V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i) = n(1-p)/p^2$$

- How about distribution of Y ? probability distribution function=?

– It's a Negative Binomial Distribution!

r.v. X =number of trials to n^{th} head

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

Why?

66

©Argon Chen

Convolution: Sum of Two Independent R.V.'s

- We already know that for two independent r.v. X and Y

$$E(X + Y) = E(X) + E(Y)$$

$$V(X + Y) = V(X) + V(Y)$$

- How about the distribution of $X+Y$?

That is, pmf or pdf $X+Y$: $p_{x+y}(X+Y)$ or $f_{x+y}(X+Y)$?

- In general, there is no easy answer for this even given the known distribution of X and Y ! In fact, the solution is to solve the convolution of functions:

$$p_{x+y}(X + Y = a) = \sum_{\forall y} p_x(X = a - y) p_y(Y = y)$$

$$f_{x+y}(X + Y = a) = \int_{-\infty}^{\infty} f_x(a - y) f_y(y) dy$$

67

©Argon Chen

Distribution of the Sum of Two Independent R.V.'s

- The answer could be as complicated as solving the convolution of functions. But could be quite simple too.....
- Example:

r.v. $X+Y$ takes possible values: 2,3,4,5

$$p_x(0) = .3 \quad p_x(1) = .2 \quad p_x(2) = .5$$

$$p_y(2) = .5 \quad p_y(3) = .5$$

Convolution:

$$p(x + y = 2) = p(x = 0)p(y = 2) = 0.15$$

$$p(x + y = 3) = p(x = 0)p(y = 3) + p(x = 1)p(y = 2) = 0.25$$

$$p(x + y = 4) = p(x = 1)p(y = 3) + p(x = 2)p(y = 2) = 0.35$$

$$p(x + y = 5) = p(x = 2)p(y = 3) = 0.25$$

$$\text{or in general } p(x + y = 5) = \sum_{y=2}^3 p(x = 5 - y)p(Y = y)$$

68

©Argon Chen

Distribution of the Sum of Two Independent R.V.'s

- The answer could be as complicated as solving the convolution of functions. But could be quite simple too.....
- Example:

r.v. $X+Y$ takes possible values: 2,3,4,5

$$p_x(0) = .3 \quad p_x(1) = .2 \quad p_x(2) = .5$$

$$p_y(2) = .5 \quad p_y(3) = .5$$

Convolution:

$$p(x+y=2) = p(x=0)p(y=2) = 0.15$$

$$p(x+y=3) = p(x=0)p(y=3) + p(x=1)p(y=2) = 0.25$$

$$p(x+y=4) = p(x=1)p(y=3) + p(x=2)p(y=2) = 0.35$$

$$p(x+y=5) = p(x=2)p(y=3) = 0.25$$

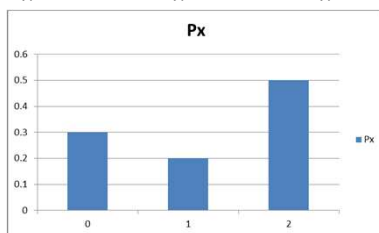
$$\text{or in general } p(x+y=5) = \sum_{y=2}^3 p(x=5-y)p(y)$$

69

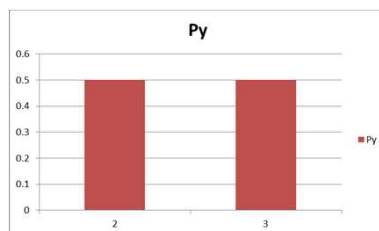
©Argon Chen

More on Sum of RV's

$$p_x(0) = .3 \quad p_x(1) = .2 \quad p_x(2) = .5$$

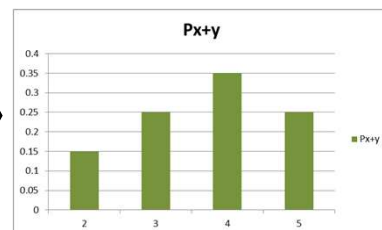


$$p_y(2) = .5 \quad p_y(3) = .5$$



$$p_{x+y}(2) = 0.15 \quad p_{x+y}(3) = 0.25$$

$$p_{x+y}(4) = 0.35 \quad p_{x+y}(5) = 0.25$$



70

©Argon Chen

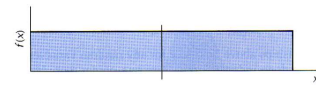
Central Limit Theorem

If X_1, X_2, \dots, X_n are outcomes of a sample of n independent observations of a random variable X with mean μ_x and variance σ_x^2 , then

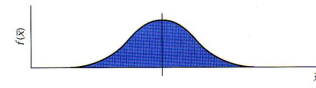
$$\Sigma X_i \sim N(n\mu_x, n\sigma_x^2)$$

and

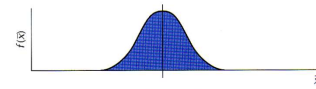
$$\Sigma X_i / n \sim N(\mu_x, \sigma_x^2 / n)$$



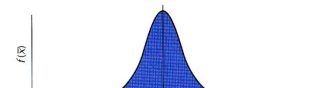
(a) Population of individuals.



(b) Sample means of $n = 4$.



(c) Sample means of $n = 9$.

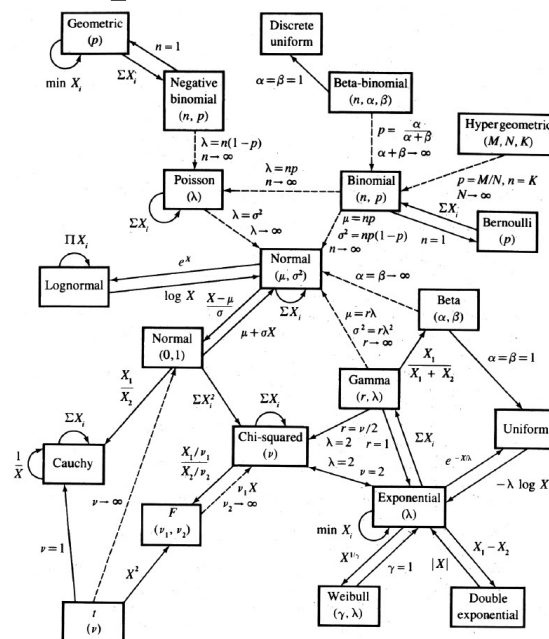


(d) Sample means of $n = 25$.

71

©Argon Chen

Relationships of Distribution Models



72

©Argon Chen