PMF, Expectation, Mean and Variance

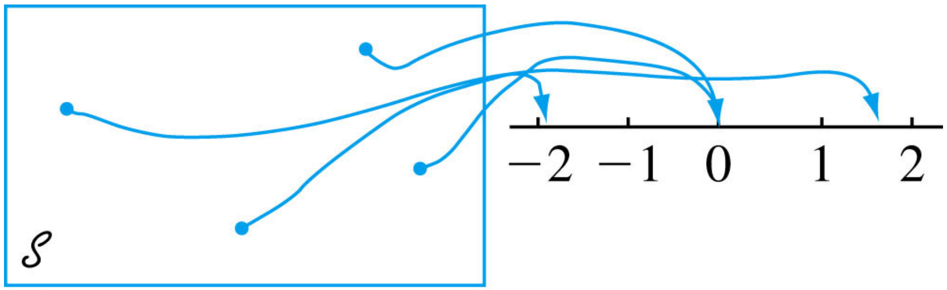
Prof Shou-de Lin

CSIE/GINM, NTU

sdlin@csie.ntu.edu.tw

Random Variable

- A random variable *X* is a function that maps *each* possible outcome of a random experiment to *one and only one* real number (i.e. assign a real value to each possible outcome).
- That is *X*: S --> *R*, where S is the set of all outcomes of an experiment and *R* is the set of real numbers.
- The space of X is the set of real numbers {*x*: *X*(s)= x, s $\in$ S   }



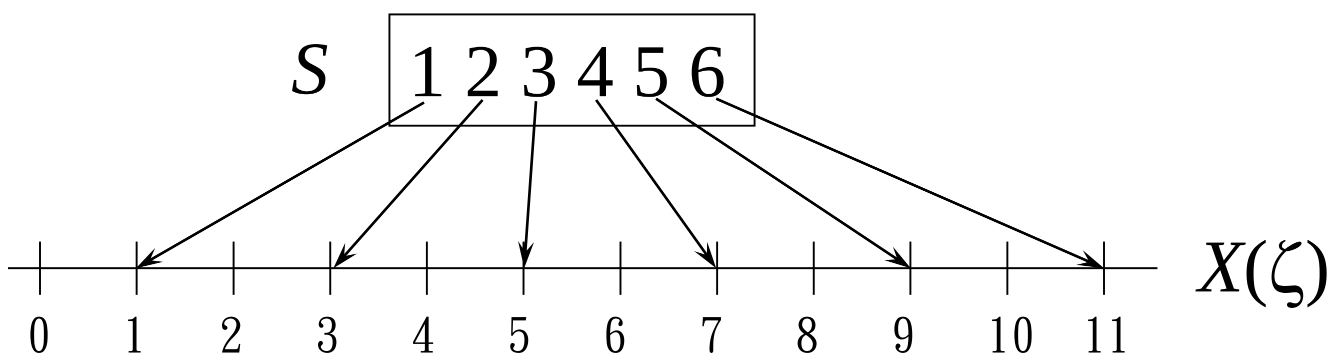
Example: coin toss

- Coin toss (once) has two possible outcomes: “head” and “tail”.
- One can define a random variable *X* that maps “head” to 1 and “tail” to 0.
- We can also define a random variable *Y* that maps “head” to 0 and “tail” to 1.
- The space of X is {0,1}, and so is the space of Y.

Notation: X → random variable,
x → assignment (or real value) of a random variable
X=1 ➔ outcome is head

Example: toss a die

- Example :
  S = {toss a dice} = {1, 2, 3, 4, 5, 6}
  X(s) = 2s - 1$\Rightarrow${X(s)} = {1, 3, 5, 7, 9, 11}



$$P\{s \mid X(s) \leq 8\} =$$

Discrete Random Variable vs. Continuous Random Variable

- Given a random variable X with space S.
- If *S* is a finite or countable infinite set, then *X* is said to be a discrete random variable.
- Otherwise X is said to be a continuous random variable

PMF(probability mass function)

- Definition: A function that maps a value of a discrete random variable to its relevant probability
- Notation: $P(X = x_i)$ or $f(x_i)$ or $p_X(x_i)$
- The pmf function should satisfy the following properties:
  - (a) $p_X(x_i) > 0$, $x_i \in S$;
  - (b) $\sum_{x_i \in S} p_X(x_i) = 1$
  - (c) $P(X \in A) = \sum_{x_i \in A} p_X(x_i)$, where $A \subset S$

Example of pmf

- RE(random experiment): Randomly select n=2 balls from a bag that contains N_1=5 red balls and N_2=4 blue balls,
- Event: x out of the 2 balls are red.
- S={zero, one, two balls}, X={0,1,2}
- f(x)=p(X=x) = $\dfrac{C_x^5 \times C_{2-x}^4}{C_2^9}$
- Note, f(x) is a function of x. This type of distribution is called Hypergeometric Distribution

曾暐翔

Hypergeometric Distribution

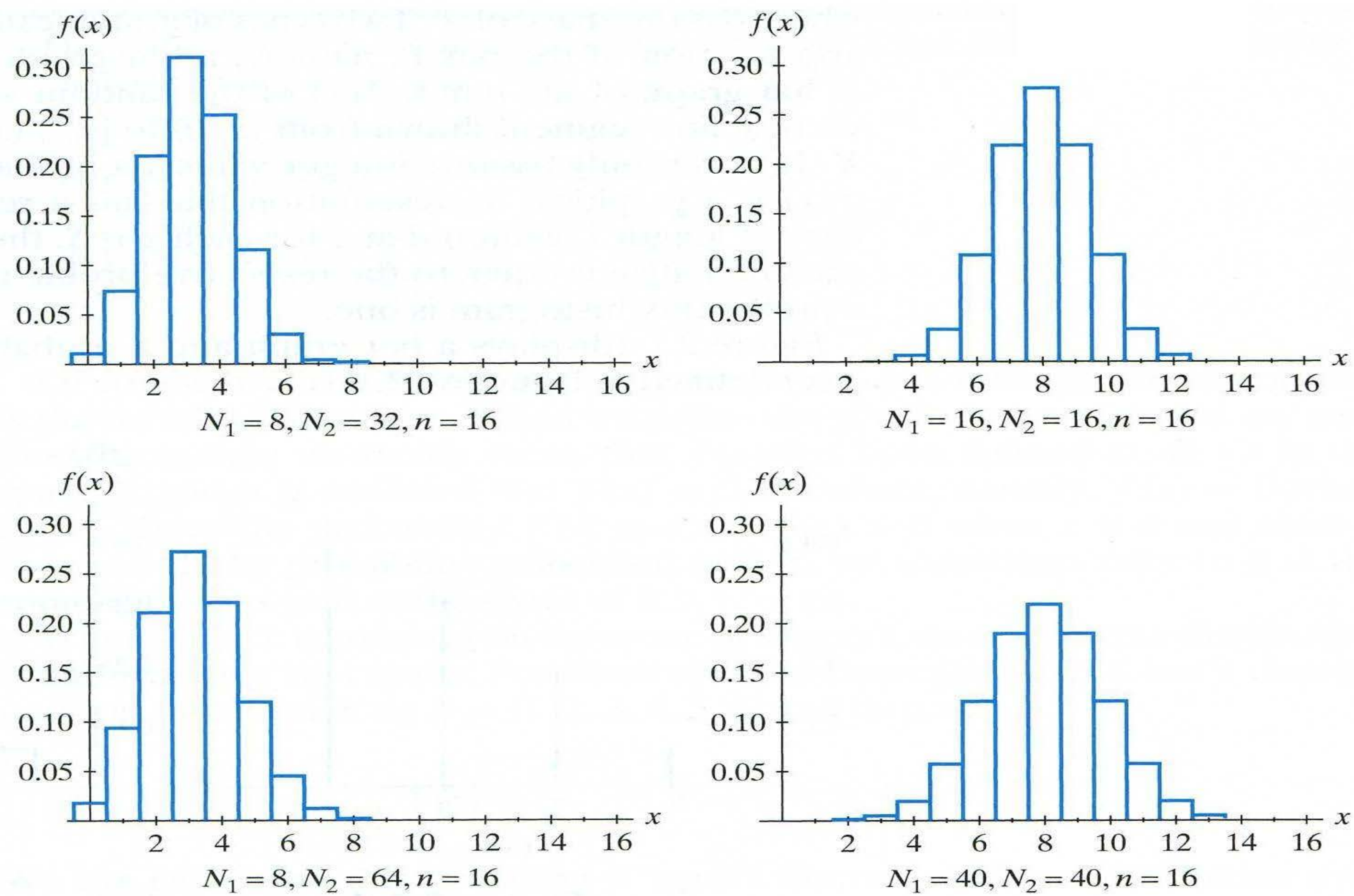
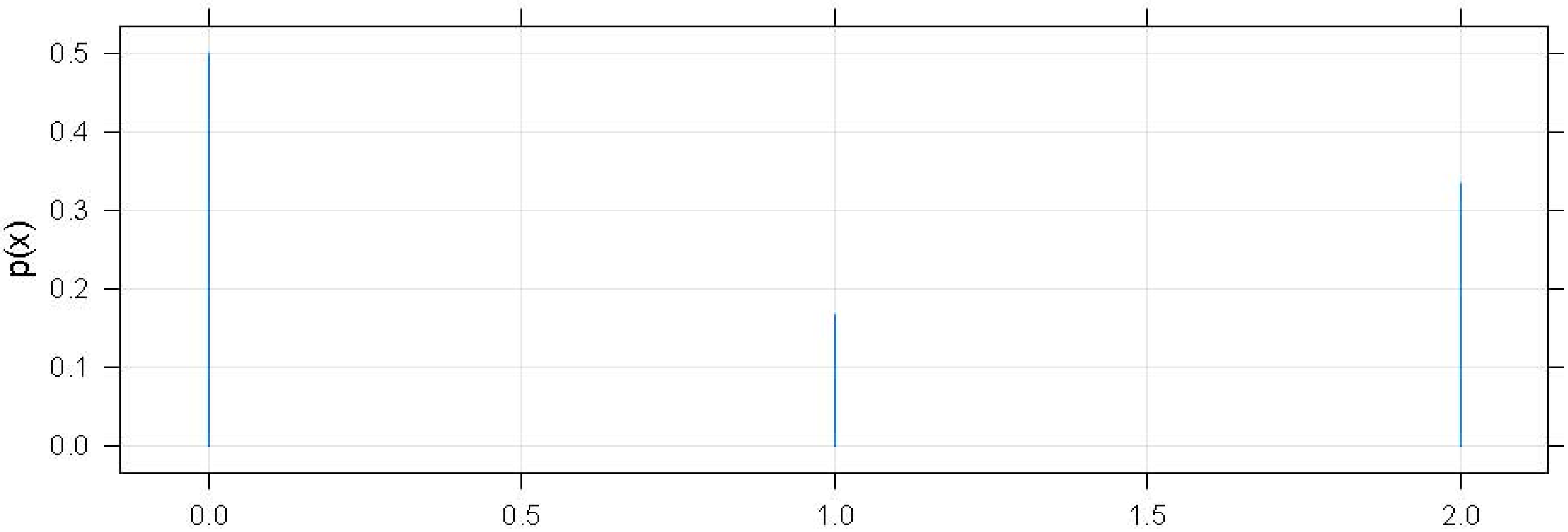


Figure 2.1-2: Hypergeometric probability histograms

Plotting the pmf

In Example 3.7 the pmf, in tabular form and as a plot, is

x	0	1	2
$p(x)$	0.5	0.167	0.333



Mathematical Expectation

- Def: if f(x) is the pmf of a random variable X, then the expected value of X, E[X], is $\sum_{x \in S} x f(x)$
- Def: if f(x) is the pmf of a random variable X with space S, then the expected value (or mathematical expectation) of the function u(X) is E[u(X)]= $\sum_{x \in S} u(x) f(x)$, e.g. E[X^2]= $\sum_{x \in S} x^2 f(x)$

Function of Random Variable

- X is a random variable on S$\Rightarrow$ u(x) is a random variable on S1, there are two ways to generate E[u(x)]
  - E[u(x)]= $\sum_{x \in S} u(x) f(x)$
  - Assuming u(x)=Y, then we can find out the pmf of Y as h(y), then E[u(x)]=E[Y]= $\sum_{y \in S1} y h(y)$

Example

- Let X be a random variable whose pmf is: $f(x)=1/3$, $x \in S$ where $S=\{-1,0,1\}$
- Let $u(X)=X^2$ then $E[u(X)]=\Sigma\, x^2 f(x) = ?$
- Or we can define $Y=X^2$, $S=\{0,1\}$, then ,h(y)=
  - $P(Y=0)=$ 1/3
  - $P(Y=1)=$ 2/3
- $\sum_{x \in S} yh(y) =$ 2/3

Property of Expectation

（a） If $u(X)=c$ is a constant, $E[u(X)]= c$

（b） If c is a constant and u is a function, then $E[c \cdot u(X)]=c \cdot E[u(X)]$

（c） If c_1 and c_2 are constants and u_1 and u_2 are functions, then $E[c_1 u_1(X)+c_2 u_2(X)]=c_1 E[u_1(X)]+c_2 E[u_2(X)]$

（c'） the above can be extended to more than two terms $c_1 \ldots c_n$

Example

- Let: $u(x)=(x-b)^{2,}$ where b is not a function of X
- Q1: What is $E[(X-b)^2]$?

E[X^2-2Xb+b^2]=E[X^2] -2E[X]b+E[b^2]=E[X^2]-2bE[X]+b^2

- Q2: find the value of b that minimize $E[(X-b)^2]$

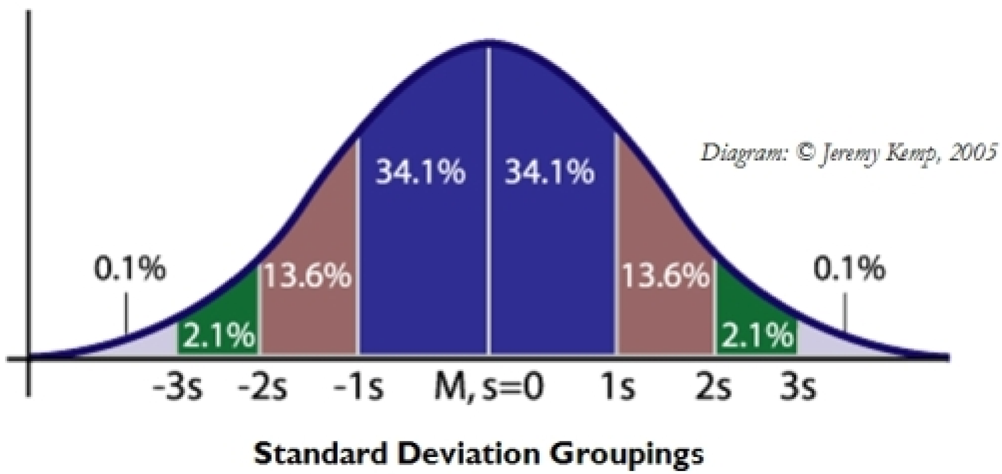
dE[(x-b)^2]/db=0-2E[X]+2b =0, b=E[X]

李信穎

Mean (μ), Variance (σ^2) and standard deviation (σ)

- The mean (μ) of a random variable X is E[X]
- The variance (σ^2) of a random variable is $E[(X-\mu)^2] = \sum_{x \in S} (x-\mu)^2 f(x) = \sum_{x \in S} x^2 f(x) - \mu^2$
  $= E[X^2] - E[X]^2$
- σ is called the standard deviation of X
- The variance of a random variable measures the deviation of its distribution from the mean.

Standard Deviation



Variance of a Discrete Random Variable

- Let X be a random variable with mean μ_X and variance σ_X^2. Let Y= aX+b, where a and b are constants. Then,
  - $E[Y]=a\mu_X+b$,
  - $Var[Y]=E[(aX+b- (a\mu_X+b))^2]=E[a^2(X-\mu_X)^2]= a^2\sigma_X^2$

Example: Lottery

- invoice: 1/1000 to win 200 dollars
- Lottery: $1/10^8$ to win $2*10^7$
- Mean: 0.2 for each case
- Variance: $\sum_{x\in S}(x-\mu)^2 f(x)$
  - Invoice=
  - Lottery=

李哲宇

What if you don’t know f(x)?

- $E[X]= \Sigma xf(x)$
- How to compute E[X] if we don’t know the pmf of X?
- Ans: Sampling

Sampling (or simulation)

- Sampling is a process of performing a random experiment n times and recording the results $(x_1,x_2,\ldots x_n)$ for each trial.
- The distribution of x's is called the empirical distribution since it is determined by the data.
- The average of $x_1 \ldots x_n$ is called the sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Sample Variance

- Assuming the *true mean* of a RE is μ, then the sample variance of the sampling (x_1,x_2,…x_n) is
$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$$

- However, μ is unknown (what we know is $\overline{x}$), the sample variance based on $\overline{x}$ is

$$\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2$$

$$\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2 > \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2$$

Is Sample Mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ unbiased ?

- Definition: Let $X_1 \ldots X_n$ be random samples from a distribution, $E[u(x_1, x_2, .. x_n)] = \theta$, then $u(x_1, x_2, .. x_n)$ is called an unbiased estimator of θ, otherwise biased.

$$E[\bar{x}] = E[\frac{1}{n}\sum_{i=1}^{n} x_i] = \frac{1}{n}\sum_{i=1}^{n} E[x_i] = \frac{n\mu}{n} = \mu$$

- Thus, $\bar{x}$ is an unbiased estimator of μ

Is $\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2$ an unbiased estimator for Var(X)?

E[(X-μ)[2]]= VAR(X)

$$E[\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2]=E[\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu-(\overline{x}-\mu))^2]$$

$$=E[\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2-\boxed{\frac{2}{n}\sum_{i=1}^{n}(x_i-\mu)(\overline{x}-\mu)}+\frac{1}{n}\sum_{i=1}^{n}(\overline{x}-\mu)^2]$$

$$=Var(X)-E[(\overline{x}-\mu)^2]=Var(X)-Var(\overline{X})=\frac{n-1}{n}Var(X)$$

$$E[\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2]=Var(X)$$

Note that, $Var(\overline{X})=Var(\sum_{i=1}^{n}x_i/n)=\frac{1}{n^2}*nVar(x)=\frac{1}{n}Var(X)$

Moment of a Distribution

- Let X be a random variable and *r* be a positive integer,
  - $E\left[(X-b)^r\right] = \sum_{x \in S} (x-b)^r f(x)$ is called the r^{th} moment of the distribution about b.
  - $E[X^r]$ is called the r^{th} moment of the distribution about origin.
  - What is variance (in terms of moment)? 李子于

The Moment-Generating Function $E[e^{tX}]$

- Let X be a discrete random variable with p.m.f f(x) and space S. If there is a positive number h such that $$E\left[e^{tX}\right]=\sum_{x_i \in S} e^{tx_i} f\left(x_i\right)$$

  exists and is finite for $-h < t < h$, then the function of t defined by $$M\left(t\right)=E\left[e^{tX}\right]$$

  is called the moment-generating function of X. and often abbreviated as m.g.f.

The Moment-Generating Function

- Let X and Y be two discrete random variable defined on the same space S.
  If their m.g.f. are the same (i.e., $E\left[e^{tX}\right]=E\left[e^{tY}\right]$,)
  then the probability mass functions of X and Y are equal.
- Q: If X and Y has different outcome space S, can them have the same m.g.f.?
- Assume that S= $\{b_1, b_2, \ldots, b_k\}$ contains only positive integers, and f(x), g(y) represents the pmf of X and Y, respectively, then
  $M_x(t)= e^{tb_1} f(b_1) + \ldots e^{tb_n} f(b_n)$
  $M_y(t)= e^{tb_1} g(b_1) + \ldots e^{tb_n} g(b_n)$

李昱翰

Example

- If X has the m.g.f M(t)= $e^{t}(\frac{3}{6})+e^{2t}(\frac{2}{6})+e^{3t}(\frac{1}{6})$ then, what is the pmf of X?

$$M(t)=\sum_{x_i \in S} e^{tx_i} f(x_i)$$

Ans: f(1)= 3/6 , f(2)= 2/6 , f(3)= 1/6

Why M(t) is called moment-generating function?

$$M(t) = E\left[e^{tX}\right] = \sum_{x \in S} e^{tx} f(x)$$

$$M'(t) = \sum_{x \in S} x e^{tx} f(x)$$

$$M''(t) = \sum_{x \in S} x^2 e^{tx} f(x)$$

$\sigma^2 = ?$

$$M^{(r)}(t) = \sum_{x \in S} x^r e^{tx} f(x)$$

$$M'(0) = \sum_{x \in S} x f(x) = \mu$$

$$M''(0) = \sum_{x \in S} x^2 f(x) = E[X^2]$$

$$M^{(r)}(0) = \sum_{x \in S} x^r f(x) = E[X^r]$$

Red-Envelop Trick

- You are given two red envelops, and told that one contains X dollars and the other contains 2X.
- An envelop is first picked by you, and you are given a chance to swap.
- Assuming the current envelop has Y dollars
  - if you swap, there is ½ chance you would obtain 2Y and ½ chance 0.5Y, therefore the expectation value of swapping is ½ (2Y+0.5Y)=1.25Y
  - If you don't swap, the expectation value is Y
  - So you HAVE to swap anyway~

李智源

- Is this statement correct? If not, can you explain why using the axioms of probability (i.e. random experiment, outcome, event, etc) ?