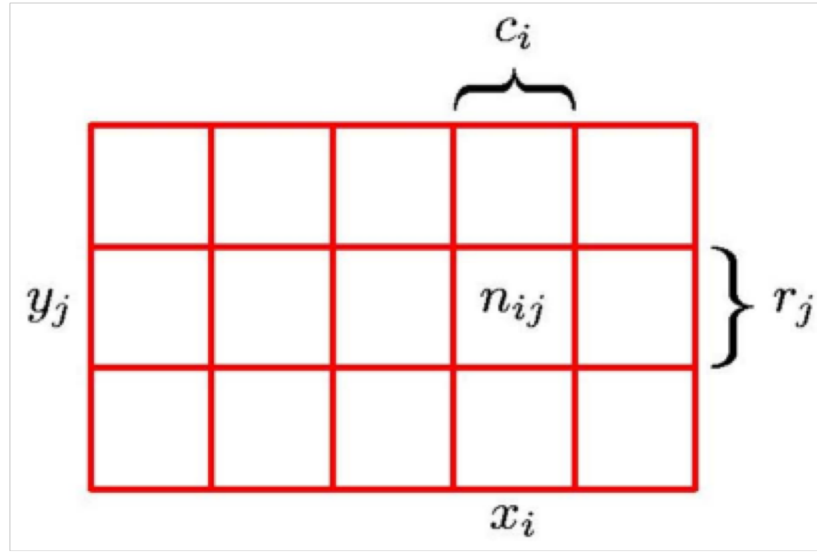
University at Albany
State University of New York

Lecture 31: Bayes Rules, Expected Value and Variance, and Binormal Distribution

Dr. Chengjiang Long
Computer Vision Researcher at Kitware Inc.
Adjunct Professor at SUNY at Albany.
Email: **clong2@albany.edu**

Recap Previous Lecture

- Conditional Probability
- Sum and Product Rules
- Bayes Rule



Sum Rule

$$p(X = x_i) = \frac{c_i}{N} = \frac{1}{N}\sum_{j=1}^{L} n_{ij}$$
$$= \sum_{j=1}^{L} p(X = x_i, Y = y_j)$$

Product Rule

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \cdot \frac{c_i}{N}$$
$$= p(Y = y_j | X = x_i) p(X = x_i)$$

Conditional Probability

$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$

$$P(x \mid y) = \frac{P(x, y)}{P(y)} = \frac{P(y \mid x) P(x)}{\sum_{x \in X} P(x, y)}$$

$$\text{posterior} = \frac{\text{likelihood} * \text{prior}}{\text{evidence}}$$

Outline

- Expected Value and Variance
- Binominal Distribution

Outline

- **Expected Value and Variance**
- Binominal Distribution

Expected Value and Variance

- All probability distributions are characterized by an expected value (mean) and a variance (standard deviation squared).

Expected value of a random variable

- Expected value is just the average or mean (μ) of random variable *x*.
- It's sometimes called a "weighted average" because more frequent values of X are weighted more highly in the average.
- It's also how we expect X to behave on-average over the long run ("frequentist" view again).

Expected value, formally

Discrete case:

$$E(X) = \sum_{\text{all } x} x_i \, p(x_i)$$

Continuous case:

$$E(X) = \int_{\text{all } x} x_i \, p(x_i) dx$$

- E(X) = μ. These symbols are used interchangeably

Example: expected value

- Recall the following probability distribution of ER arrivals:

x	10	11	12	13	14
P(x)	.4	.2	.2	.1	.1

$$\sum_{i=1}^{5} x_i\, p(x) = 10(.4) + 11(.2) + 12(.2) + 13(.1) + 14(.1) = 11.3$$

Sample Mean is a special case of Expected Value…

- Sample mean, for a sample of n subjects:

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \sum_{i=1}^{n} x_i \left(\frac{1}{n}\right)$$

The probability (frequency) of each person in the sample is 1/n.

Example: the lottery

- A certain lottery works by picking 6 numbers from 1 to 49. It costs $1.00 to play the lottery, and if you win, you win $2 million after taxes.

- *If you play the lottery once, what are your expected winnings or losses?*

Expected Value

- Expected value is an extremely useful concept for good decision-making!

Lottery

Calculate the probability of winning in 1 try:

$$\frac{1}{\binom{49}{6}} = \frac{1}{\frac{49!}{43!6!}} = \frac{1}{13{,}983{,}816} = 7.2 \text{ x } 10^{-8}$$

“49 choose 6”

Out of 49 numbers, this is the number of distinct combinations of 6.

The probability function (note, sums to 1.0):

x$	*p(x)*
-1	.999999928
+ 2 million	*7.2 x 10*$^{--8}$

Expected Value

<u>The probability function</u>

x$	*p(x)*
-1	.999999928
+ 2 million	$7.2 \times 10^{--8}$

<u>Expected Value</u>

E(X) = P(win)*$2,000,000 + P(lose)*-$1.00

$= 2.0 \times 10^{6} * 7.2 \times 10^{-8} + .999999928\ (-1)$

$= .144 - .999999928 = -\$.86$

Negative expected value is never good!

You shouldn't play if you expect to lose money!

Expected Value

If you play the lottery every week for 10 years, what are your expected winnings or losses?

520 x (-.86) = -$447.20

Variance/standard deviation

σ^2=Var(x) =E(x-μ)^2

"The expected (or average) squared distance (or deviation) from the mean"

$$\sigma^2 = Var(x) = E[(x-\mu)^2] = \sum_{\text{all } x} (x_i - \mu)^2 \, p(x_i)$$

Variance, continuous

Discrete case:

$$Var(X) = \sum_{\text{all } x} (x_i - \mu)^2 p(x_i)$$

Continuous case?:

$$Var(X) = \int_{\text{all } x} (x_i - \mu)^2 p(x_i) dx$$

- Var(X)= σ^2
- SD(X) = σ
- These symbols are used interchangeably.

Variance

$$\sigma^2 = \sum_{\text{all } x} (x_i - \mu)^2 p(x_i)$$

$$\sigma^2 = \sum_{\text{all } x} (x_i - \mu)^2 p(x_i) =$$

$$= (1 - 200{,}000)^2(.5) + (400{,}000 - 200{,}000)^2(.5) = 200{,}000^2$$

$$\sigma = \sqrt{200{,}000^2} = 200{,}000$$

Now you examine your personal risk tolerance…

Practice Problem

On the roulette wheel, X=1 with probability 18/38 and X= -1 with probability 20/38.

We already calculated the mean to be = -\$.053. What's the variance of X?

Answer

$$\sigma^2 = \sum_{\text{all x}} (x_i - \mu)^2 p(x_i)$$

$$= (+1 - -.053)^2 (18/38) + (-1 - -.053)^2 (20/38)$$

$$= (1.053)^2 (18/38) + (-1 + .053)^2 (20/38)$$

$$= (1.053)^2 (18/38) + (-.947)^2 (20/38)$$

$$= .997$$

$$\sigma = \sqrt{.997} = .99$$

Standard deviation is $.99. Interpretation: On average, you're either 1 dollar above or 1 dollar below the mean, which is just under zero. Makes sense!

Outline

- Expected Value and Variance
- **Binominal Distribution**

Binomial Probability Distribution

- A fixed number of observations (trials), n
  - e.g., 15 tosses of a coin; 20 patients; 1000 people surveyed
- A binary outcome
  - e.g., head or tail in each toss of a coin; disease or no disease
  - Generally called "success" and "failure"
  - Probability of success is p, probability of failure is 1 – p
- Constant probability for each observation
  - e.g., Probability of getting a tail is the same each time we toss the coin

Binomial distribution

Take the example of 5 coin tosses. What's the probability that you flip exactly 3 heads in 5 coin tosses?

Solution: One way to get exactly 3 heads: HHHTT

What's the probability of this exact arrangement?

P(H) x P(H) x P(H) x P(T) x P(T) = $(1/2)^3 \times (1/2)^2$

Another way to get exactly 3 heads: THHHT

Probability of this exact outcome = $(1/2)^1 \times (1/2)^3 \times (1/2)^1$

= $(1/2)^3 \times (1/2)^2$

Binomial distribution

- In fact, $(1/2)^3$ *x* $(1/2)^2$ is the probability of each unique outcome that has exactly 3 heads and 2 tails.

- So, the overall probability of 3 heads and 2 tails is:
  $(1/2)^3$ *x* $(1/2)^2$ + $(1/2)^3$ *x* $(1/2)^2$ + $(1/2)^3$ *x* $(1/2)^2$ + ….. for as many unique arrangements as there are—but how many are there?

Binomial distribution

$\binom{5}{3}$ ways to arrange 3 heads in 5 trials

Outcome	Probability
THHHT	$(1/2)^3 \, x \, (1/2)^2$
HHHTT	$(1/2)^3 \, x \, (1/2)^2$
TTHHH	$(1/2)^3 \, x \, (1/2)^2$
HTTHH	$(1/2)^3 \, x \, (1/2)^2$
HHTTH	$(1/2)^3 \, x \, (1/2)^2$
HTHHT	$(1/2)^3 \, x \, (1/2)^2$
THTHH	$(1/2)^3 \, x \, (1/2)^2$
HTHTH	$(1/2)^3 \, x \, (1/2)^2$
HHTHT	$(1/2)^3 \, x \, (1/2)^2$
THHTH	$(1/2)^3 \, x \, (1/2)^2$

10 arrangements $x \, (1/2)^3 \, x \, (1/2)^2$

The probability of each unique outcome (note: they are all equal)

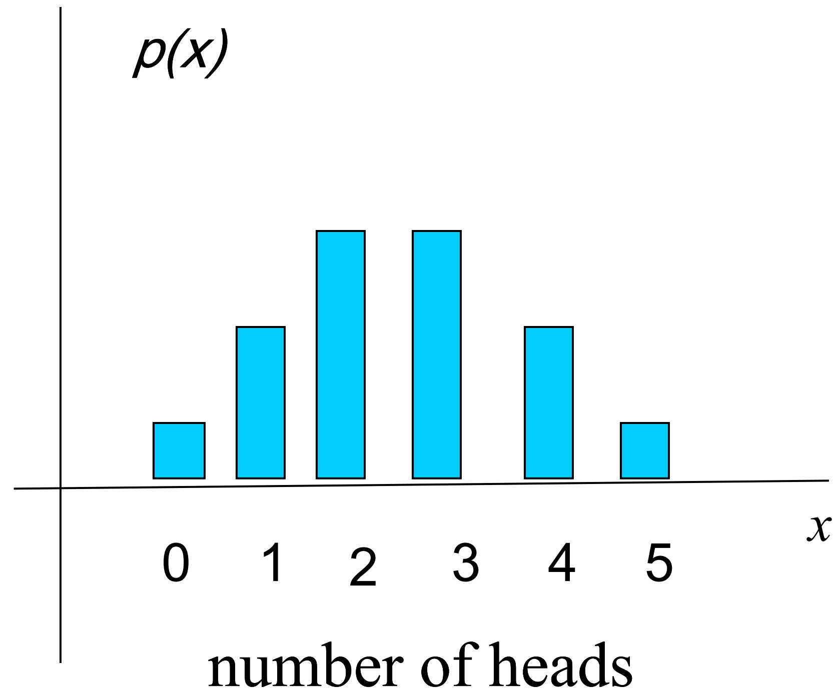
C(5,3) = 5!/3!2! = 10

Factorial review: n! = n(n-1)(n-2)…

Binomial distribution

$\therefore$ P(3 heads and 2 tails) = $\binom{5}{3}$ $x\ P(heads)^3\ x\ P(tails)^2 =$

$10\ x\ (½)^{5=}31.25\%$



Binomial distribution function:
X= the number of heads tossed in 5 coin tosses

Binomial distribution, generally

Note the general pattern emerging → if you have only two possible outcomes (call them 1/0 or yes/no or success/failure) in *n* independent trials, then the probability of exactly *X* "successes"=

n = number of trials

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

X = # successes out of *n* trials

p = probability of success

1-p = probability of failure

Example

- If I toss a coin 20 times, what's the probability of getting exactly 10 heads?

$$\binom{20}{10}(.5)^{10}(.5)^{10} = .176$$

Example

- If I toss a coin 20 times, what's the probability of getting of getting 2 or fewer heads?

$$\binom{20}{0}(.5)^0(.5)^{20} = \frac{20!}{20!0!}(.5)^{20} = 9.5x10^{-7} +$$

$$\binom{20}{1}(.5)^1(.5)^{19} = \frac{20!}{19!1!}(.5)^{20} = 20x9.5x10^{-7} = 1.9x10^{-5} +$$

$$\binom{20}{2}(.5)^2(.5)^{18} = \frac{20!}{18!2!}(.5)^{20} = 190x9.5x10^{-7} = 1.8x10^{-4}$$

$$= 1.8x10^{-4}$$

Expected Value and Variance for Binormal Distribution

- **All probability distributions are characterized by an expected value and a variance:**

If *X* follows a binomial distribution with parameters *n* and *p*: ***X ~ Bin (n, p)***

Then:

$E(X) = np$

$Var(X) = np(1-p)$

$SD(X) = \sqrt{np(1-p)}$

Note: the variance will always lie between

0*N-.25 *N

p(1-p) reaches maximum at p=.5

P(1-p)=.25

Example

- 1. You are performing a cohort study. If the probability of developing disease in the exposed group is .05 for the study duration, then if you (randomly) sample 500 exposed people, how many do you expect to develop the disease? Give a margin of error (+/- 1 standard deviation) for your estimate.

$$X \sim \text{binomial}\ (500, .05)$$

$$E(X) = 500\ (.05) = 25$$

$$Var(X) = 500\ (.05)\ (.95) = 23.75$$

$$StdDev(X) = \text{square root}\ (23.75) = 4.87$$

$$\therefore 25 \pm 4.87$$

Example

2. What's the probability that **at most** 10 exposed subjects develop the disease?

This is asking for a CUMULATIVE PROBABILITY: the probability of 0 getting the disease or 1 or 2 or 3 or 4 or up to 10.

P(X≤10) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4)+….+ P(X=10)=

$$\binom{500}{0}(.05)^0(.95)^{500}+\binom{500}{1}(.05)^1(.95)^{499}+\binom{500}{2}(.05)^2(.95)^{498}+\ldots+\binom{500}{10}(.05)^{10}(.95)^{490}<.01$$

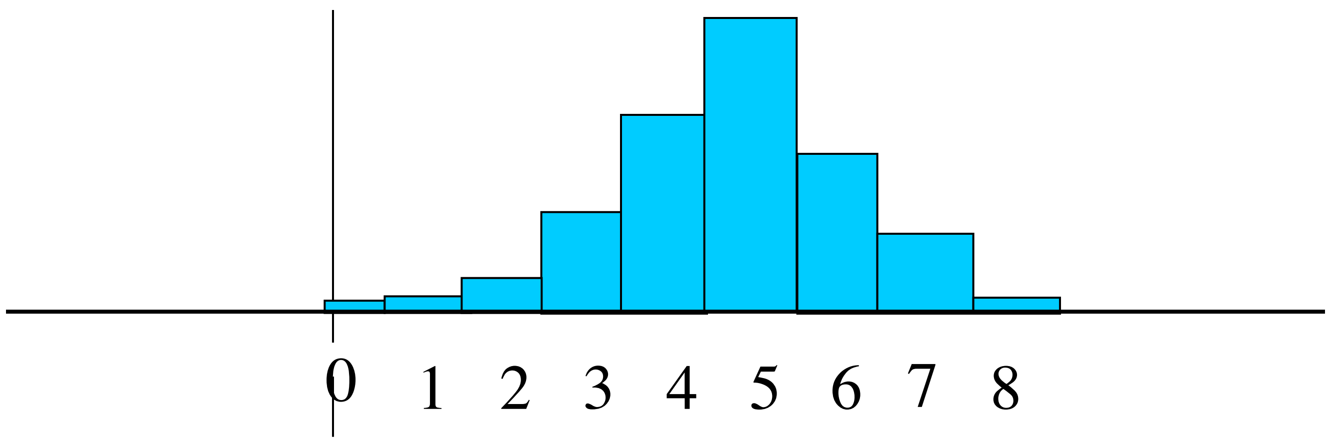
Practice Problem:

You are conducting a case-control study of smoking and lung cancer. If the probability of being a smoker among lung cancer cases is .6, what's the probability that in a group of 8 cases you have:

a. Less than 2 smokers?
b. More than 5?
c. What are the expected value and variance of the number of smokers?

Answer

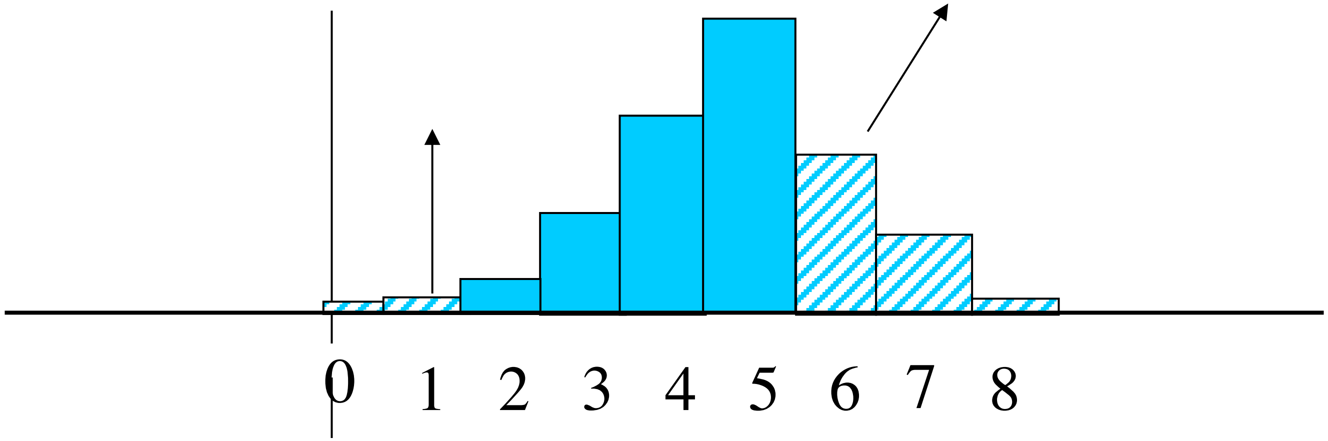
X	P(X)
0	$1(.4)^8$=.00065
1	$8(.6)^1$ $(.4)^7$=.008
2	$28(.6)^2$ $(.4)^6$=.04
3	$56(.6)^3$ $(.4)^5$=.12
4	$70(.6)^4$ $(.4)^4$=.23
5	$56(.6)^5$ $(.4)^3$=.28
6	$28(.6)^6$ $(.4)^2$=.21
7	$8(.6)^7$ $(.4)^1$=.090
8	$1(.6)^8$ =.0168



Answer, continued

P(<2)=.00065 + .008 = .00865

P(>5)=.21+.09+.0168 = .3168



E(X) = 8 (.6) = 4.8
Var(X) = 8 (.6) (.4) =1.92
StdDev(X) = 1.38

Proportions…

- The binomial distribution forms the basis of statistics for proportions.
- A proportion is just a binomial count divided by n.
  - For example, if we sample 200 cases and find 60 smokers, X=60 but the observed proportion=.30.
- Statistics for proportions are similar to binomial counts, but differ by a factor of n.

Stats for proportions

For binomial:

$$\mu_x = np$$

$$\sigma_x^{\ 2} = np(1-p)$$

$$\sigma_x = \sqrt{np(1-p)}$$

Differs by a factor of n.

Differs by a factor of n.

For proportion:

$$\mu_{\hat{p}} = p$$

$$\sigma_{\hat{p}}^{\ 2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

P-hat stands for "sample proportion."

It all comes back to normal…

- Statistics for proportions are based on a normal distribution, because the binomial can be approximated as normal if $np>5$

Next class

- Topic: Relations (1)
- Pre-class reading: Chap 9.1-9.2