UCLA STAT 13

Introduction to Statistical Methods for the Life and Health Sciences

- **Instructor**: **Ivo Dinov**,
  Asst. Prof. In Statistics and Neurology
- **Teaching Assistants:** Tom Daula and Kaiding Zhu,
  UCLA Statistics

University of California, Los Angeles, Fall 2002
http://www.stat.ucla.edu/~dinov/

Chapter 5: Discrete Random Variables

- Random variables
- Probability functions
- The Binomial distribution
- Expected values

Definitions

- An **experiment** is a naturally occurring phenomenon, a scientific study, a sampling trial or a test., in which an object (unit/subject) is selected at random (and/or treated at random) to *observe/measure* different outcome characteristics of the process the experiment studies.
- A ***random variable*** is a type of measurement taken on the outcome of a random experiment.

Definitions

- The ***probability function*** for a discrete random variable X gives **$P(X = x)$** [denoted $\mathrm{pr}(x)$ or $P(x)$]

  for every value x that the R.V. X can take
- E.g., number of heads when a coin is tossed twice

x	0	1	2
$\mathrm{pr}(x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Stopping at one of each or 3 children

Sample Space – **complete/unique description of the possible outcomes from this experiment.**

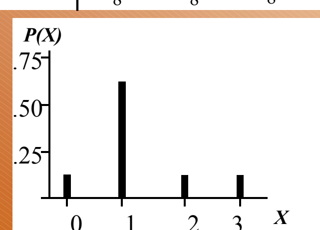
Outcome	GGG	GGB	GB	BG	BBG	BBB
Probability	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$

- For R.V. X = number of girls, we have

X	0	1	2	3
$\mathrm{pr}(x)$	$\frac{1}{8}$	$\frac{5}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

Plotting the probability function

X	0	1	2	3
$\mathrm{pr}(x)$	$\frac{1}{8}$	$\frac{5}{8}$	$\frac{1}{8}$	$\frac{1}{8}$



Tossing a biased coin twice

- For each toss, P(Head) = p → P(Tail) = P(comp(H))=1-p
- Outcomes: HH, HT, TH, TT
- Probabilities: $p.p$, $p(1\text{-}p)$, $(1\text{-}p)p$, $(1\text{-}p)(1\text{-}p)$
- Count X, the number of heads in 2 tosses

X	0	1	2
pr(x)	$(1-p)^2$	$2p(1-p)$	p^2

Hospital stays

Days stayed	x	4	5	6	7	8	9	10	Total
Frequency		10	30	113	79	21	8	2	263
Proportion	pr($X = x$)	0.038	0.114	0.430	0.300	0.080	0.030	0.008	1000
Cumulative Proportion	pr($X \le x$)	0.038	0.152	0.582	0.882	0.962	0.992	1.000	

Calculating Interval probabilities from cumulative probabilities

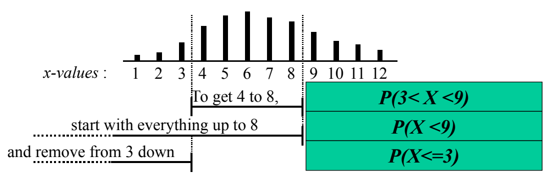


Figure 5.2.2 Interval probabilities from cumulative probabilities.[a]
[a][This Figure represents an arbitrary distribution, not the hospital distribution.]

How to find the upper-tail?

Review

- What is a random variable? What is a discrete random variable? (type of measurement taken on the outcome of random experiment)
- What general principle is used for finding P(X=x)? (Adding the probabilities of all outcomes of the experiment where we have measured the RV, X=x)
- What two general properties must be satisfied by the probabilities making up a probability function? (P(x)>=0; $\sum_x P(x) = 1$)
- What are the two names given to probabilities of the form P($X \le x$)? (cumulative & lower/left-tail)

Review

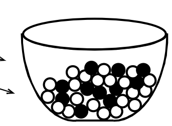
- How do we find an upper/right-tail probability from a cumulative probability? [P(X>x) = 1-P(X<= x)]
- When we use P($X \le 12$) – P($X \le 5$) to calculate the probability that X falls within an interval of values, what numbers are included in the interval? ([6:12])

The two-color urn model

N balls in an urn, of which there are

M black balls

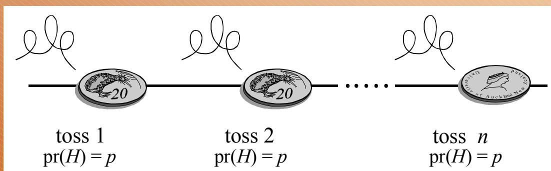
$N - M$ white balls



Sample n balls and count X = # black balls in sample

We will compute the probability distribution of the R.V. X

The biased-coin tossing model



Perform n tosses and count X = # heads

We also want to compute the probability distribution of this R.V. *X!*
Are the two-color urn and the biased-coin models <u>related</u>? How do we present the models in <u>mathematical terms</u>?

The answer is: Binomial distribution

- The distribution of the number of heads in n tosses of a biased coin is called the ***Binomial distribution***.

Binomial(N, p) – the probability distribution of the number of Heads in an N-toss coin experiment, where the probability for Head occurring in each trial is p.
E.g., Binomial(6, 0.7)

	x	0	1	2	3	4	5	6
Individual	$pr(X = x)$	0.001	0.010	0.060	0.185	0.324	0.303	0.118
Cumulative	$pr(X \le x)$	0.001	0.011	0.070	0.256	0.580	0.882	1.000

For example P(X=0) = P(all 6 tosses are Tails) =

$$(1-0.7)^6 = 0.3^6 = 0.001$$

Binary random process

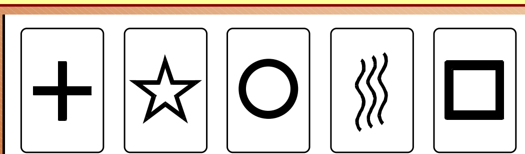
The ***biased-coin tossing model*** is a physical model for situations which can be characterized as a series of trials where:

- each trial has only two outcomes: *success* or *failure*;
- p = P(*success*) is the same for every trial; and
- trials are independent.

- The distribution of X = number of successes (heads) in N such trials is

Binomial(N, p)

Extra-Sensory Perception (ESP)



Hypnosis enhances your ESP? ***CumBin(324, 1500, 0.2) = 0.94***

A card is randomly selected from a shuffled deck. Two people under hypnosis participate, one staring at the card, the other trying to guess the card. 15 pairs of students tested, each doing the experiment 100 times. Total of <u>1,500 trials</u>. <u>325</u> correct guesses were recorded. Is there <u>evidence for ESP potentiation</u>?
Purely random guessing would yield expected 300 correct answers.

Sampling from a finite population – Binomial Approximation

If we take a sample of size n

- from a much larger population (of size N)
- in which a proportion p have a characteristic of interest, then the distribution of **X, the number in the sample with that characteristic**,
- is *approximately* Binomial(n, p).
  - (Operating Rule: Approximation is adequate if $n / N < 0.1$.)
- Example, polling the US population to see what proportion is/has-been married.

Odds and ends …

- For what types of situation is the urn-sampling model useful? For modeling binary random processes. When sampling with replacement, Binomial distribution is exact, where as, in sampling without replacement Binomial distribution is an approximation.
- For what types of situation is the biased-coin sampling model useful? Defective parts. Approval poll of cloning for medicinal purposes. Number of Boys in 151 presidential children (90).
- Give the three essential conditions for its applicability. (two outcomes; same p for every trial; independence)

Odds and ends …

- What is the distribution of the number of heads in n tosses of a biased coin?
- Under what conditions does the Binomial distribution apply to samples taken without replacement from a finite population? When interested in assessing the distribution of a R.V., X, the number of observations in the sample (of n) with one specific characteristic, where $n / N < 0.1$ and a proportion p have the characteristic of interest in the beginning of the experiment.

Binomial Probabilities – the moment we all have been waiting for!

- Suppose X ~ Binomial(n, p), then the probability

$$P(X = x) = \binom{n}{x} p^x (1-p)^{(n-x)}, \quad 0 \le x \le n$$

- Where the binomial coefficients are defined by

$$\binom{n}{x} = \frac{n!}{(n-x)!\ x!}, \qquad n! = 1 \times 2 \times 3 \times \ldots \times (n-1) \times n$$

n-factorial

Binomial Formula with examples

- Does the Binomial probability satisfy the requirements?

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

- Explicit examples for n=2, do the case n=3 at home!

$$\sum_{x=0}^{2} \binom{2}{x} p^x (1-p)^{(2-x)} = \Big\{ \textit{Three terms in the sum}$$

$$\binom{2}{0} p^0 (1-p)^2 + \binom{2}{1} p^1 (1-p)^1 + \binom{2}{2} p^2 (1-p)^0 =$$

$$1 \times 1 \times (1-p)^2 + 2 \times p \times (1-p) + 1 \times p^2 \times 1 =$$

$$(p + (1-p))^2 = 1$$

Usual quadratic-expansion formula

Expected values

- The game of chance: cost to play: \$1.50; Prices {\$1, \$2, \$3}, probabilities of winning each price are {0.6, 0.3, 0.1}, respectively.
- Should we play the game? What are our chances of winning/loosing?

Prize (\$)	x	1	2	3	
Probability	pr(x)	0.6	0.3	0.1	
What we would "expect" from 100 games					*add across row*
Number of games won		0.6×100	0.3×100	0.1×100	
\$ won		$1 \times 0.6 \times 100$	$2 \times 0.3 \times 100$	$3 \times 0.1 \times 100$	Sum

Total prize money = Sum; ***Average prize money*** = Sum/100
$= 1 \times 0.6 + 2 \times 0.3 + 3 \times 0.1$
$= 1.5$

***Theoretically** Fair Game: price to play EQ the expected return!*

TABLE 5.4.1 Average Winnings from a Game conducted N times

Number of games played (N)	Prize won in dollars (x): 1	2	3	Average winnings per game ($\bar{x}$)
	frequencies (Relative frequencies)			
100	64 (.64)	25 (.25)	11 (.11)	1.7
1,000	573 (.573)	316 (.316)	111 (.111)	1.538
10,000	5995 (.5995)	3015 (.3015)	990 (.099)	1.4995
20,000	11917 (.5959)	6080 (.3040)	2000 (.1001)	1.5042
30,000	17946 (.5982)	9049 (.3016)	3005 (.1002)	1.5020
∞	(.6)	(.3)	(.1)	1.5

So far we looked at the theoretical expectation of the game. Now we simulate the game on a computer to obtain random samples from our distribution, according to the probabilities {0.6, 0.3, 0.1}.

Definition of the expected value, in general.

- The expected value:

$$E(X) = \sum_{\text{all } x} x\,P(x) \left(= \int_{\text{all } X} x\,P(x)dx \right)$$

- = Sum of (value times probability of value)

Example

In the at least one of each or at most 3 children example, where X ={number of Girls} we have:

X	**0**	**1**	**2**	**3**
pr(x)	$\frac{1}{8}$	$\frac{5}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

$$E(X) = \sum_x x\,P(x)$$

$$= 0\times\frac{1}{8} + 1\times\frac{5}{8} + 2\times\frac{1}{8} + 3\times\frac{1}{8}$$

$$= 1.25$$

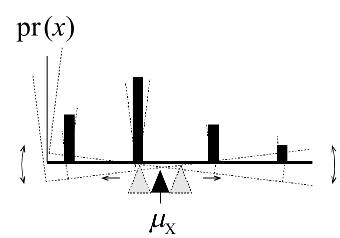
The expected value and population mean

$\mu_x = E(X)$ is called the ***mean*** of the distribution of X.

$\mu_X = E(X)$ is usually called the ***population mean***.

μ_x is the point where the bar graph of $P(X = x)$ balances.

The expected value as the point of balance



The mean μ_X is the balance point.

Population standard deviation

The ***population standard deviation*** is

$$sd(X) = \sqrt{E[(X - \mu)^2]}$$

Note that if X is a RV, then (X-μ) is also a RV, and so is $(X-\mu)^2$. Hence, the *expectation*, $E[(X-\mu)^2]$, makes sense.

For the Binomial distribution . . . mean

$E(X) = \boldsymbol{n\,p}$,

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

$$E(X) = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{(n-x)} =$$

E(X) = Sum(Value x Probability)

$$\sum_{x=1}^{n} x \binom{n}{x} p^x (1-p)^{(n-x)} =$$

If x=0, the entire term is zero

$$\sum_{x=0}^{n-1} (x+1) \binom{n}{x+1} p^{(x+1)} (1-p)^{(n-1-x)} =$$

Change variables: x → (x+1)

For the Binomial distribution . . . mean

$\mathrm{E}(\boldsymbol{X}) = \boldsymbol{n}\,\boldsymbol{p}$, $\mathrm{sd}(X) = \sqrt{np(1-p)}$

$$E(X) = \sum_{x=0}^{n-1} (x+1)\binom{n}{x+1} p^{(x+1)}(1-p)^{(n-1-x)} =$$

Expand the binomial coefficient

$$\sum_{x=0}^{n-1} (x+1)\frac{n\times(n-1)!}{(n-1-x)!(x+1)!}\, p\times p^{x}(1-p)^{(n-1-x)} =$$

$$n\times p \sum_{x=0}^{n-1} \frac{(n-1)!}{(n-1-x)!x!}\, p^{x}(1-p)^{(n-1-x)} =$$

$$n\times p\times\left(p+(1-p)\right)^{n-1} = n\times p$$

Remaining term is just the binomial formula – expectation of the constant 1, which is always 1

Slide **33** STAT 13, UCLA, Ivo Dinov

For the Binomial distribution . . . SD

$\mathrm{E}(\boldsymbol{X}) = \boldsymbol{n}\,\boldsymbol{p}$, $\mathrm{sd}(X) = \sqrt{np(1-p)}$

$$SD^2(X) = E\left((X-\boldsymbol{\mu})^2\right) = \underbrace{\sum_{x=0}^{n} (x-np)^2 \binom{n}{x} p^{x}(1-p)^{(n-x)}} =$$

$$\sum_{x=0}^{n} x^2 \binom{n}{x} p^{x}(1-p)^{(n-x)} +$$

E(X) = Sum(Value x Probability)

Expand the square term

$$\sum_{x=0}^{n} np^2 \binom{n}{x} p^{x}(1-p)^{(n-x)} - 2np \sum_{x=0}^{n} x \binom{n}{x} p^{x}(1-p)^{(n-x)} =$$

Slide **34** STAT 13, UCLA, Ivo Dinov

For the Binomial distribution . . . SD

$\mathrm{E}(\boldsymbol{X}) = \boldsymbol{n}\,\boldsymbol{p}$, $\mathrm{sd}(X) = \sqrt{np(1-p)}$

$$SD^2(X) = E\left((X-\boldsymbol{\mu})^2\right) = \sum_{x=0}^{n} x^2 \binom{n}{x} p^{x}(1-p)^{(n-x)} +$$

$$n^2p^2 \boxed{\sum_{x=0}^{n} \binom{n}{x} p^{x}(1-p)^{(n-x)}} - 2np \boxed{\sum_{x=0}^{n} x \binom{n}{x} p^{x}(1-p)^{(n-x)}} =$$

$$\sum_{x=0}^{n} x^2 \binom{n}{x} p^{x}(1-p)^{(n-x)} + n^2p^2 - 2\,n\,p\times E(X) =$$

$$\sum_{x=0}^{n} x^2 \binom{n}{x} p^{x}(1-p)^{(n-x)} + n^2p^2 - 2\,n^2\,p^2 =$$

Slide **35** STAT 13, UCLA, Ivo Dinov

For the Binomial distribution . . . mean

$\mathrm{E}(\boldsymbol{X}) = \boldsymbol{n}\,\boldsymbol{p}$, $\mathrm{sd}(X) = \sqrt{np(1-p)}$

$$SD^2(X) = \sum_{x=0}^{n} x^2 \binom{n}{x} p^{x}(1-p)^{(n-x)} + n^2p^2 - 2\,n^2\,p^2 =$$

$$\boxed{\sum_{x=0}^{n} x^2 \binom{n}{x} p^{x}(1-p)^{(n-x)}} - n^2p^2 =$$

This is simply the Expectation of X^2, $E(X^2)$ and we compute It exactly like E(X)

$$\sum_{x=1}^{n-1} (x+1)^2 \binom{n}{x+1} p^{x+1}(1-p)^{(n-1-x)} - n^2p^2 =$$

Change the summation index x → x+1

Slide **36** STAT 13, UCLA, Ivo Dinov

For the Binomial distribution . . . SD

$\mathrm{E}(\boldsymbol{X}) = \boldsymbol{n}\,\boldsymbol{p}$, $\mathrm{sd}(X) = \sqrt{np(1-p)}$

$$SD^2(X) =$$

As before, factor out n x p and do the math

$$np \sum_{x=1}^{n-1} (x+1) \binom{n-1}{x} p^{x}(1-p)^{(n-1-x)} - n^2p^2 =$$

Split off the (x+1) term

$$np\left[\sum_{x=1}^{n-1} x \binom{n-1}{x} p^{x}(1-p)^{(n-1-x)} + 1\right] - n^2p^2 =$$

$$np\left((n-1)p+1\right) - n^2p^2 = \cancel{n^2p^2} - np^2 + np - \cancel{n^2p^2} = np(1-p)$$

Binomial Formula and a bit of arithmetic yield the result

Slide **37** STAT 13, UCLA, Ivo Dinov

Linear Scaling (affine transformations) $\boldsymbol{aX + b}$

For any constants a and b, the expectation of the RV $\boldsymbol{aX + b}$ is equal to the sum of the product of a and the expectation of the RV X and the constant b.

$$\mathrm{E}(a\boldsymbol{X} + b) = \boldsymbol{a}\,\mathrm{E}(\boldsymbol{X}) + b$$

And similarly for the standard deviation (b, an additive factor, does not affect the SD).

$$\mathrm{SD}(a\boldsymbol{X} + b) = |\boldsymbol{a}|\,\mathrm{SD}(\boldsymbol{X})$$

Slide **38** STAT 13, UCLA, Ivo Dinov

Linear Scaling (affine transformations) $aX + b$

Why is that so?

$E(aX + b) = a\,E(X) + b \qquad SD(aX + b) = |a|\,SD(X)$

$$E(aX+b) = \sum_{x=0}^{n}(a\,x+b)\;P(X=x) =$$

$$\sum_{x=0}^{n} a\,x\,P(X=x) + \sum_{x=0}^{n} b\;P(X=x) =$$

$$a\sum_{x=0}^{n} x\,P(X=x) + b\sum_{x=0}^{n} P(X=x) =$$

$$aE(X) + b\times 1 = aE(X) + b.$$

Linear Scaling (affine transformations) $aX + b$

And why do we care?

$E(aX + b) = a\,E(X) + b \qquad SD(aX + b) = |a|\,SD(X)$

-completely general strategy for computing the distributions of RV's which are obtained from other RV's with known distribution. E.g., X~N(0,1), and Y=aX+b, then we need **not** calculate the mean and the SD of Y. We know from the above formulas that E(Y) = b and SD(Y) =|a|.

-These formulas hold for all distributions, not only for binomial.

Linear Scaling (affine transformations) $aX + b$

And why do we care?

$E(aX + b) = a\,E(X) + b \qquad SD(aX + b) = |a|\,SD(X)$

-E.g., say the rules for the game of chance we saw before change and the new pay-off is as follows: {\$0, \$1.50, \$3}, with probabilities of {0.6, 0.3, 0.1}, as before. What is the newly expected return of the game? Remember the old expectation was equal to the entrance fee of \$1.50, and the game was fair!

Y = 3(X-1)/2

{\$1, \$2, \$3} → {\$0, \$1.50, \$3},

E(Y) = 3/2 E(X) –3/2 = 3 / 4 = \$0.75

And the game became clearly biased. Note how easy it is to compute E(Y).

Review

- What does the expected value of X tell you about? (Expected outcome from an experiment regarding the characteristics measured by the RV X)
- Why is the expected value also called the population mean? [because for finite population E(X) is the ordinary mean (average)]
- What is the relationship between the population mean and the bar graph of the probability function? (balances the graph)
- What are the mean and standard deviation of the Binomial distribution? (np; np(1-p))
- Why is SD(X+10) = SD(X)?
- Why is SD(2X) = 2SD(X)? (Section 5.4.3)

Chapter 5 Summary

Summary

Random variable

- A type of measurement made on the outcome of a random experiment

Probability function

- P($X = x$) for every value X can take, abbreviated to P(x)

Expected value

Expected Value for a random variable X, denoted $E(X)$.

- Also called the population mean and denoted μ_X (abbreviated to μ).
- Is a measure of the long-run average of X-values in many repetitions of the experiment.
- Formula (for a discrete random variable):

$$\mu_x = E(X) = \sum x P(x)$$

Population standard deviation

- Standard deviation for a random variable X, denoted $SD(X)$ is :
  - also called the ***population standard deviation*** and denoted σ_X (abbreviated σ)
  - Is a measure of the variability of X-values.
  - Formula:

$$\sigma_x = SD(X) = \sqrt{E[(X-\mu)^2]}$$

  - for a discrete random variable X,

$$E[(X-\mu)^2] = \sum (x-\mu)^2 P(x)$$

Affine Transformations $aX + b$

For any constants a and b,

- $E(aX + b) = a\,E(X) + b$

and

- $SD(aX + b) = |a|\,SD(X)$

Sampling from a finite population

- The ***urn model*** is a physical model for situations in which we
  - sample n individuals at random from a finite population and
  - count X, the number of individuals with a characteristic of interest
- When $n/N < 0.1$, the distribution of X is approximately **Binomial(n, p)**
  - where p is the population proportion with the characteristic of interest

Observing a random process

The ***biased-coin tossing model*** is a physical model for situations which can be characterized as a series of trials where:

- each trial has only two outcomes: *success* and *failure*;
- p = P(*success*) is the same for every trial; and
- trials are independent.

- The distribution of X = number of successes (heads) in n such trials is

Binomial(n, p)

Binomial distribution

- The distribution of the number of successes in n trials (or the number of heads in n tosses) is Binomial (n, p)
- The Binomial distribution has

$$E(X) = \mu_x = np \qquad SD(X) = \sigma_x = \sqrt{np(1-p)}$$