# UCLA STAT 110A
## Applied Statistics


●**Instructor**: **Ivo Dinov,**
Asst. Prof. In Statistics and Neurology

●**Teaching Assistants:** Helen Hu, UCLA Statistics

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

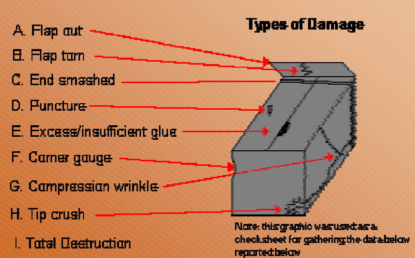



## Chapter 4: Discrete/Continuous Variables, Probabilities, CLT

- Density Histograms
- Probabilities
- Bernoulli trials
- Central Limit Theorem (CLT)
- Standardizing Transformations



## Frequency Distributions- damaged boxes






## Frequency Distributions- damaged boxes

| Type | Total Frequency | Relative Frequency | Percentage |
|---|---|---|---|
| A - Flap out | 16 | 0.0096 | 1 |
| B - Flap torn | 17 | 0.0102 | 1 |
| C - End smashed | 132 | 0.0793 | 8 |
| D – Puncture | 95 | 0.0571 | 6 |
| E - Glue problem | 87 | 0.0523 | 5 |
| F - Corner gouge | 984 | 0.5913 | 59 |
| G – Compr. wrinkle | 15 | 0.0090 | 1 |
| H - Tip crushed | 303 | 0.1821 | 18 |
| I - Tot. destruction | 15 | 0.0090 | 1 |
| Total | 1664 | 0.9999* | 100 |

(* the relative frequencies do not add to 1.0000 due to rounding)



## Frequency Distributions- damaged boxes

**Relative frequency for type A is:** $\frac{16}{1664} = 0.0096$

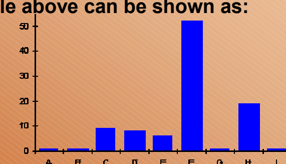
**Percentage for type A is:** $\frac{16}{1664} \times 100 = 0.96 \approx 1$ **percent.**

**The usefulness of relative frequencies and percentages is clear: for example, it is easily seen that corner gouge accounts for 59% of the total number of damages.**



## Frequency Distributions- damaged boxes

**The frequency distribution of a variable is often presented graphically as a bar-chart/bar-plot. For example, the data in the frequency table above can be shown as:**




**The vertical axis can be frequencies or relative frequencies or percentages. On the horizontal axis all boxes should have the same width leave gaps between the boxes (because there is no connection between them) the boxes can be in any order.**

## Experiments, Models, RV's

- 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.
- **Model** – generalized hypothetical description used to analyze or describe a phenomenon.
- 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 the chance that the observed value for the process equals a specific outcome, $x$.
  - $P(X = x)$ [denoted $pr(x)$ 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 |
|---|---|---|---|
| $pr(x)$ | $\frac{1}{4}$ | $\frac{1}{2}$ | $\frac{1}{4}$ |



## 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 |
|---|---|---|---|---|
| $pr(x)$ | $\frac{1}{8}$ | $\frac{5}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ |



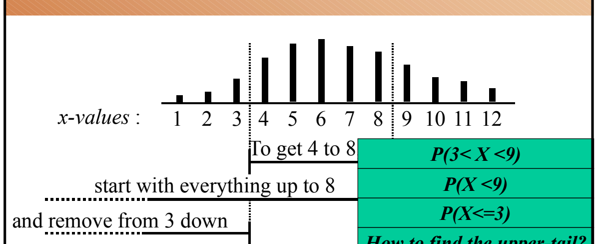
## Tossing a biased coin twice

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

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



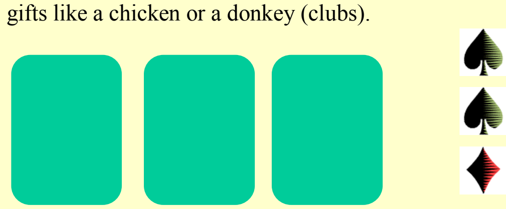
## Calculating Interval probabilities from cumulative probabilities






## Let's Make a Deal Paradox – aka, Monty Hall 3-door problem

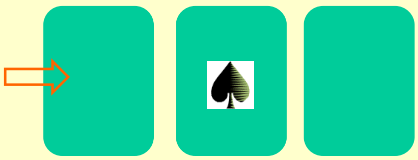
- This paradox is related to a popular television show in the 1970's. In the show, a contestant was given a choice of three doors/cards of which one contained a prize (diamond). The other two doors contained gag gifts like a chicken or a donkey (clubs).



## Let's Make a Deal Paradox.

- After the contestant chose an initial door, the host of the show then revealed an empty door among the two unchosen doors, and asks the contestant if he or she would like to switch to the other unchosen door. The question is should the contestant switch. Do the odds of winning increase by switching to the remaining door?

***Applets.dir/StatGames.exe***



1.Pick One card

2.Show one Club Card

3. Change 1st pick?



## Let's Make a Deal Paradox.

- The *intuition* of most people tells them that each of the doors, the chosen door and the unchosen door, are equally likely to contain the prize so that there is a 50-50 chance of winning with either selection? This, however, is **not the case**.
- The probability of winning by using the switching technique is 2/3, while the odds of winning by not switching is 1/3. The easiest way to explain this is as follows:



## Let's Make a Deal Paradox.

- The probability of picking the wrong door in the initial stage of the game is 2/3.
- If the contestant picks the wrong door initially, the host must reveal the remaining empty door in the second stage of the game. Thus, if the contestant switches after picking the wrong door initially, the contestant will win the prize.
- The probability of winning by switching then reduces to the probability of picking the wrong door in the initial stage which is clearly 2/3.



## Bernoulli Trials

- A Bernoulli trial is an experiment where only two possible outcomes are possible (0 / 1).
- Examples:
  - Coin tosses
  - Computer chip (0 / 1) signal.
  - Poll supporters/opponents; yes/no; for/against.

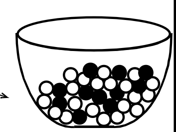


## The two-color urn model

$N$ 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


toss 1 pr($H$) = $p$   toss 2 pr($H$) = $p$   .....   toss $n$ pr($H$) = $p$

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($N$, $p$)



## 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.



## 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!

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

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

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

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

{ ***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* | | $0.6 \times 100$ | $0.3 \times 100$ | $0.1 \times 100$ | ↙ |
| *\$ 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!*



**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 |
| $\infty$ | (.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***.

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



## 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

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

***X~Binomial(n, p) ➔***

$$X = Y_1 + Y_2 + Y_3 + .. + Y_n,$$
*where* $Y_k \sim Bernoulli(p)$,
$$E(Y_1) = p \;➔$$
$$E(X) = E(Y_1 + Y_2 + Y_3 + .. + Y_n) = np$$



## For the Binomial distribution . . . SD

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

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

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

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

*E(X) = Sum(Value x Probability)*

*Expand the square term*



## For the Binomial distribution . . . SD

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

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

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

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

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



## For the Binomial distribution . . . mean

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

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

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

*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*

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

$$(x+1)^2 \frac{n \times (n-1)!}{(n-1-x)!\,(x+1)\times x!} p^x \times p = n \times p \frac{(x+1)^2}{(x+1)} \frac{(n-1)!}{(n-1-x)!\,x!} p^x$$



## For the Binomial distribution . . . SD

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

$SD^2(X) =$

*As before, factor out n×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*

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

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

*Binomial Formula and a bit of arithmetic yield the result*



## Linear Scaling (affine transformations) $aX + b$

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

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

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

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

## 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) $aX + b$

**Example:**

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

1. X={-1, 2, 0, 3, 4, 0, -2, 1}; P(X=x)=1/8, for each x
2. Y = 2X-5 = {-7, -1, -5, 1, 3, -5, -9, -3}
3. E(X)=
4. E(Y)=
5. Does **E(X) = 2 E(X) –5 ?**
6. Compute SD(X), SD(Y). **Does SD(Y) = 2 SD(X)?**



## 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 and Normal.



## 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).



## Means and Variances for (in)dependent Variables!

- **Means**:
  - Independent/Dependent Variables {X1, X2, X3, …, X10}
    - E(X1 + X2 + X3 + … + X10) = E(X1)+ E(X2)+ E(X3)+… + E(X10)
- **Variances**:
  - Independent Variables {X1, X2, X3, …, X10}, variances add-up
    Var(X1 +X2 + X3 + … + X10) = Var(X1)+Var(X2)+Var(X3)+…+Var(X1)
  - Dependent Variables {X1, X2}
    Variance contingent on the variable dependences,
    - E.g., If X2 = 2X1 + 5,

      Var(X1 +X2) =Var (X1 + 2X1 +5) =

      Var(3X1 +5) =Var(3X1) = 9Var(X1)



## For the Binomial distribution . . . SD

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

***X~Binomial(n, p) →***

***$X = Y_1 + Y_2 + Y_3 + \ldots + Y_n$,***

***where $Y_k \sim$ Bernoulli(p),***

***$Var(Y_1) = (1-p)^2 \times p + (0-p)^2 \times (1-p)$ →***

***$Var(Y_1) = (1-p)(p-p^2+p^2) = (1-p)p$ →***

***$Var(X) = Var(Y_1) + \ldots + Var(Y_n) = n(1-p)p$***

***SD(X)=Sqrt[Var(X)] = Sqrt[n(1-p)p]***

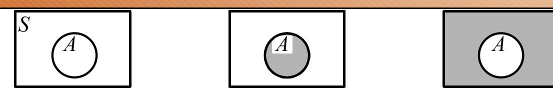
## Sample spaces and events

- A ***sample space***, *S*, for a random experiment is the set of all possible outcomes of the experiment.
- An ***event*** is a *collection* of outcomes.
- An event ***occurs*** if any outcome making up that event occurs.



## The complement of an event

- The complement of an event *A*, denoted $\bar{A}$ , occurs *if and only if A* does not occur.



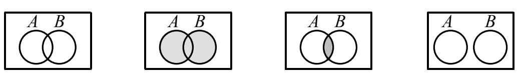

(a) Sample space containing event *A* (b) Event *A* shaded (c) $\bar{A}$ shaded

An event *A* in the sample space *S*.



## Combining events – *all statisticians agree on*

- *"A* **or** *B"* contains all outcomes in *A* **or** *B* (or both).
- *"A* **and** *B"* contains all outcomes which are **in both** *A* and *B*.




(a) Events *A* and *B* (b) "*A* or *B*" shaded (c) "*A* and *B*" shaded (d) Mutually exclusive events

Two events.

From *Chance Encounters* by C.J. Wild and G.A.F. Seber, © John Wiley & Sons, 2000.

Mutually exclusive events cannot occur at the same time.



## Probability distributions

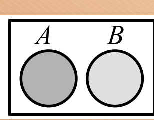
- Probabilities always lie between 0 and 1 and they sum up to 1 (across all simple events) .
- ***pr(A)*** can be obtained by adding up the probabilities of all the outcomes in *A*.

$$pr(A) = \sum_{E \text{ outcome in event } A} pr(E)$$



## Rules for manipulating Probability Distributions

For mutually exclusive events,
**pr(*A* or *B*) = pr(*A*) + pr(*B*)**






## Review

- If *A* and *B* are mutually exclusive, what is the probability that both occur? (0) What is the probability that at least one occurs? (sum of probabilities)
- If we have two or more mutually exclusive events, how do we find the probability that at least one of them occurs? (sum of probabilities)
- Why is it sometimes easier to compute *pr(A)* from *pr(A) = 1 - pr($\bar{A}$)*? (The complement of the even may be easer to find or may have a known probability. E.g., a random number between 1 and 10 is drawn. Let A ={a number less than or equal to 9 appears}. Find pr(A) = 1 – pr($\bar{A}$)). probability of $\bar{A}$ is pr({10 appears}) = 1/10 = 0.1. Also Monty Hall 3 door example!

## Sample vs. theoretical mean & varaince

- The **Expected value**: (population mean) $E(X) = \sum_{\text{all } x} x\,\mathrm{P}(x) \quad \left( = \int_{\text{all } x} x\,\mathrm{P}(x)dx \right)$
- **Sample mean** $\overline{X} = \frac{1}{N}\sum_{k=1}^{N} x_k$
- **(Theoretical) Variance**

$$Var(X) = \sum_{\text{all } x} (x-\mu_x)^2 \mathrm{P}(x) \quad \left( = \int_{\text{all } x} (x-\mu_x)^2\, \mathrm{P}(x)dx \right)$$

- **(Sample) variance**

$$Var(X) = \frac{1}{N-1}\sum_{k=1}^{N}(x_k - \overline{X})^2 = \sum_{k=1}^{N}(x_k - \overline{X})^2 \mathrm{P}(x)$$



## *Melanoma* – type of skin cancer – an example of laws of conditional probabilities

400 Melanoma Patients by Type and Site

| Type | Site: Head and Neck | Site: Trunk | Site: Extremities | Row Totals |
|---|---|---|---|---|
| Hutchinson's melanomic freckle | 22 | 2 | 10 | 34 |
| Superficial | 16 | 54 | 115 | 185 |
| Nodular | 19 | 33 | 73 | 125 |
| Indeterminant | 11 | 17 | 28 | 56 |
| Column Totals | 68 | 106 | 226 | 400 |

Contingency table based on Melanoma histological type and its location



## Conditional Probability

The ***conditional probability*** of ***A*** occurring ***given*** that *B* occurs is given by

$$\mathrm{pr}(A \mid B) = \frac{\mathrm{pr}(A \text{ and } B)}{\mathrm{pr}(B)}$$

Suppose we select one out of the 400 patients in the study and we want to find the probability that the cancer is on the *extremities* *given that* it is of type *nodular*: P = 73/125 = P(C. on Extremities | Nodular)

$$\frac{\#\text{nodular patients with cancer on extremities}}{\#\text{nodular patients}}$$



## Multiplication rule- what's the percentage of Israelis that are poor and Arabic?

$$\mathrm{pr}(P \,\&\, A) = \mathrm{pr}(P \mid A)\mathrm{pr}(A) = \mathrm{pr}(A \mid P)\mathrm{pr}(P)$$

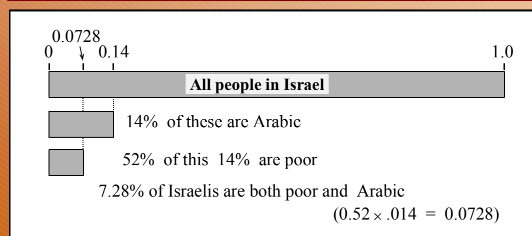
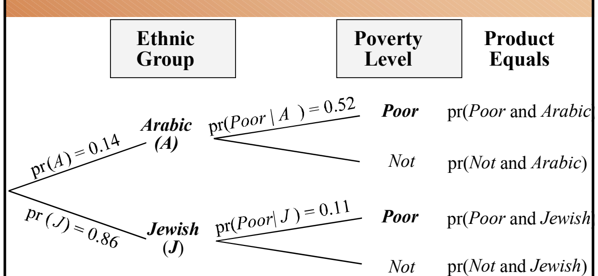



Illustration of the multiplication rule.



## Tree diagram for poverty in Israel






## Statistical independence

- Events *A* and *B* are *statistically independent* if knowing whether *B* has occurred gives no new information about the chances of *A* occurring,

  i.e. if **pr(*A* | *B*) = pr(*A*)**

- Similarly, P(*B* | *A*) = P(*B*), since

  P(B|A)=P(B & A)/P(A) = P(A|B)P(B)/P(A) = P(B)

- If *A* and *B* are *statistically independent*, then

$$\mathrm{pr}(A \text{ and } B) = \mathrm{pr}(A) \times \mathrm{pr}(B)$$

## People vs. Collins

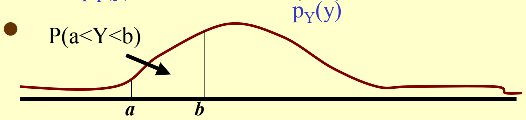
**Frequencies Assumed by the Prosecution**

| | | | |
|---|---|---|---|
| Yellow car | $\frac{1}{10}$ | Girl with blond hair | $\frac{1}{3}$ |
| Man with mustache | $\frac{1}{4}$ | Black man with beard | $\frac{1}{10}$ |
| Girl with ponytail | $\frac{1}{10}$ | Interracial couple in car | $\frac{1}{1000}$ |

- The first occasion where a conviction was made in an American court of law, largely on statistical evidence, 1964. A woman was mugged and the offender was described as a wearing dark cloths, with blond hair in a pony tail who got into a yellow car driven by a black male accomplice with mustache and beard. The suspect brought to trial were picked out in a line-up and fit all of the descriptions. Using the *product rule for probabilities* an expert witness computed the chance that a random couple meets these characteristics, as 1:12,000,000.

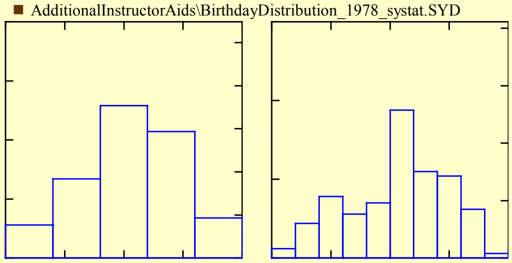


## Continuous RV's

- A RV is continuous if it can take on any real value in a non-trivial interval (a ; b).
- PDF, probability density function, for a cont. RV, Y, is a non-negative function $p_Y(y)$, for any real value y, such that for each interval (a; b), the probability that Y takes on a value in (a; b), P(a<Y<b) equals the area under $p_Y(y)$ over the interval (a: b).
- 

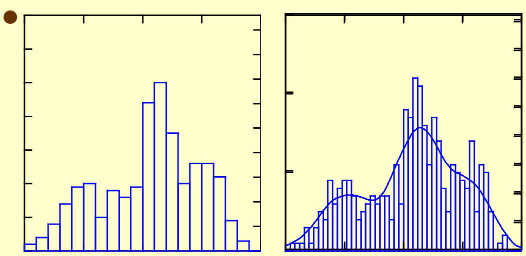



## Convergence of density histograms to the PDF

- For a continuous RV the density histograms converge to the PDF as the size of the bins goes to zero.
  - AdditionalInstructorAids\BirthdayDistribution_1978_systat.SYD
- 



## Convergence of density histograms to the PDF

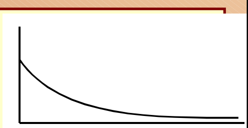
- For a continuous RV the density histograms converge to the PDF as the size of the bins goes to zero.
- 



## Computing Probabilities using PDFs

- $P(Y \in A) = \int_A p_Y(y)dy$

  $p_Y(y) = e^{-y}, y \ge 0$



(i) Exponential shape

- Example:

$$P(0 \le Y \le 3) = \int_0^3 p_Y(y)dy =$$

$$\int_0^3 e^{-y}dy = - \left. e^{-y} \right/_0^3 = 1 - e^{-3} \cong 1$$



## CDF (cumulative distribution function)

- $F_Y(y) = P(Y \le y) = \int_{-\infty}^{y} p_Y(y)dy$

  $p_Y(y) = e^{-y}, y \ge 0$

- Example:

$$F_Y(3) = P(Y \le 3) = \int_0^3 p_Y(y)dy =$$

$$\int_0^3 e^{-y}dy = - \left. e^{-y} \right/_0^3 = 1 - e^{-3} \cong 1$$

## Measures of central tendency/variability for Continuous RVs

- Mean
$$\mu_Y = \int_{-\infty}^{\infty} y \times p_Y(y)dy$$
- Variance $$\sigma_Y^2 = \int_{-\infty}^{\infty} (y-\mu_Y)^2 \times p_Y(y)dy$$
- SD $$\sigma_Y = \sqrt{\int_{-\infty}^{\infty} (y-\mu_Y)^2 \times p_Y(y)dy}$$



## Facts about PDF's of continuous RVs

- Non-negative
$$p_Y(y) \ge 0, \forall y$$
- Completeness $$\int_{-\infty}^{\infty} p_Y(y)dy = 1$$
- Probability
$$P(a < Y < b) = \int_a^b y \times p_Y(y)dy$$



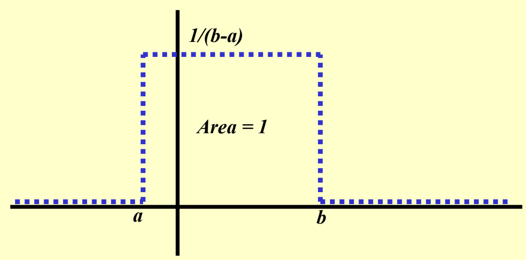
## Continuous Distributions

- Normal distribution
- Student's T distribution
- F-distribution
- Chi-squared ( $\chi^2$)
- Cauchy's distribution
- Exponential distribution
- Poisson distribution, …



## Uniform Distribution

- Uniform Distribution PDF: **Y~Uniform(a;b)** ←→ $p_Y(y)=1/(b-a)$, for each a<=y<=b, and $p_Y(y)=0$, otherwise.






## Uniform Distribution – CDF, mean, variance

- Uniform Distribution CDF:

$$F_Y(y) = \int_{-\infty}^{y} p_Y(x)dx = \int_a^{\min(y,b)} \frac{1}{b-a}dx =$$

$$\frac{x}{b-a}\Big|_a^y = \begin{cases} 0, & y < a \\ \frac{y-a}{b-a}, & a \le y \le b \\ 1, & b < y \end{cases}$$



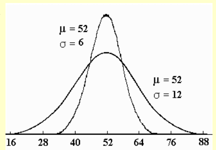
## Uniform Distribution – CDF, mean, variance

- Mean:
$$\mu_Y = \int_{-\infty}^{\infty} yp_Y(y)dy = \int_a^b \frac{y}{b-a}dy = \frac{y^2}{2(b-a)}\Big|_a^b = \frac{a+b}{2}$$
- Variance:
$$\sigma_Y^2 = \int_{-\infty}^{\infty} (y-\mu_Y)^2 p_Y(y)dy = \int_a^b \frac{(2y-(a+b))^2}{4(b-a)}dy = \frac{(b-a)^2}{12}$$
- SD:
$$\sigma_Y = \sqrt{\int_{-\infty}^{\infty} (y-\mu_Y)^2 p_Y(y)dy} = \frac{(b-a)}{\sqrt{12}}$$

## Continuous Distributions - Normal

- (General) Normal distribution

$$y = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$




- (Standard) Normal distribution (μ=0, σ=1)

$$y = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \qquad Z = \frac{Y-\mu}{\sigma}$$



## (General) Normal Distribution

- Normal Distribution PDF: **Y~Normal(μ, σ²)** ←→

$$p_Y(y) = \frac{e^{-\frac{(y-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}, \forall -\infty < y < \infty$$

$$F_Y(y) = \int_{-\infty}^{y} p_Y(x)dx = \int_{-\infty}^{y} \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} dx$$



## Standard Normal (Gaussian) Distribution

- Normal Distribution PDF: **Y~Normal(μ=0, σ²=1)** ←→

$$p_Y(y) = \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}, \forall -\infty < y < \infty$$

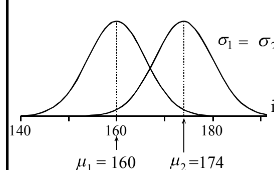
$$F_Y(y) = \int_{-\infty}^{y} p_Y(x)dx = \int_{-\infty}^{y} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx$$



## Effects of $\mu$ and $\sigma$ (on the graphs of Normal Distribution)

**(a) Changing** $\mu$

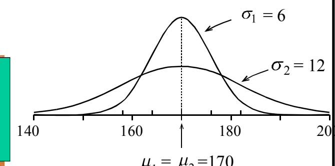
shifts the curve along the axis




Mean is a measure of … central tendency

**(b) Increasing** $\sigma$

increases the spread and flattens the curve



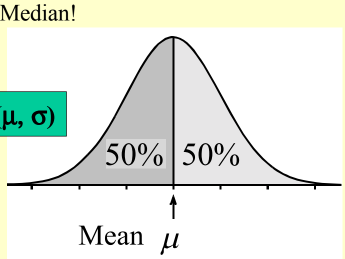

**Standard deviation** is a measure of … variability/spread



## The Normal distribution density curve

- Is symmetric about the mean! Bell-shaped and unimodal.
- Mean = Median!

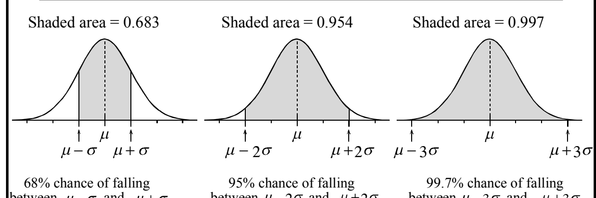
**N(μ, σ)**





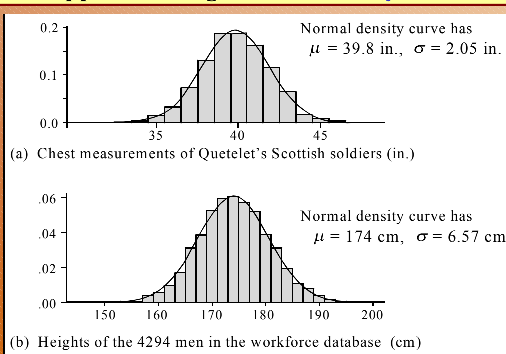

## Understanding the standard deviation: $\sigma$

Probabilities/areas and numbers of standard deviations for the Normal distribution




NormalCurveInteractive.html

## Two standardized histograms with approximating Normal density curve



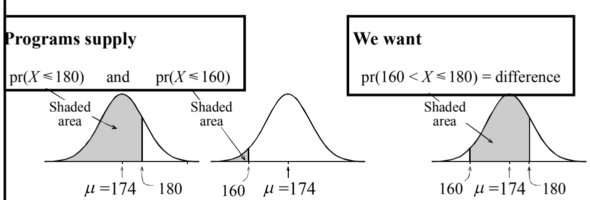



## Basic method for obtaining probabilities

- Sketch a Normal curve, marking the mean and other values of interest.
- Shade the area under the curve that gives the desired probability.
- Devise a way of getting the desired area from lower-tail areas.
- Obtain component lower-tail probabilities from a computer program



**(a) Computing pr(160 < $X$ ≤ 180)**




pr(160 < $X$ ≤ 180) = pr($X$ ≤ 180) − pr($X$ ≤ 160)



## Tabular representation of probabilities

**(c) More Normal probabilities**
*(values obtained from Minitab)*

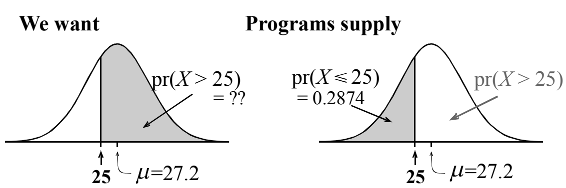
| b | pr($X$ ≤ b) | a | pr($X$ ≤ a) | pr(a < $X$ ≤ b) = difference |
|---|---|---|---|---|
| 167.6 | 0.165 | 152.4 | 0.001 | 0.164 |
| 177.8 | 0.718 | 167.6 | 0.165 | 0.553 |
| 177.8 | 0.718 | 152.4 | 0.001 | 0.717 |
| 182.9 | 0.912 | 167.6 | 0.165 | 0.747 |

**Note:** 152.4cm = 5ft, 167.6cm = 5ft 6in., 177.8cm = 5ft 10in., 182.9cm = 6ft

From *Chance Encounters* by C.J. Wild and G.A.F. Seber, © John Wiley & Sons, 2000.



**Obtaining an upper-tail probability**




Since total area under curve = 1, pr($X$ > 25) = 1 - pr($X$ ≤ 25)

Generally, **pr($X$ > $x$) = 1 − pr($X$ ≤ $x$)**



## Continuous Distributions – Student's T

- Student's T distribution [approx. of Normal(0,1)]
  - $Y_1, Y_2, \ldots, Y_N$ IID from a Normal($\mu$;$\sigma$)
  - Variance $\sigma^2$ is unknown
- In 1908, William Gosset (pseudonym Student) derived the exact sampling distribution of the following statistics

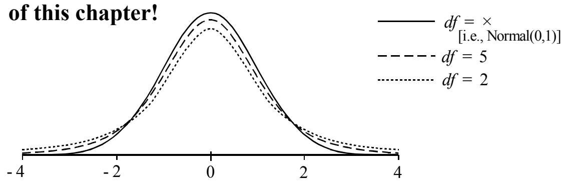
$$T = \frac{\bar{Y} - \mu_Y}{\hat{\sigma}_Y}$$

- T~Student(**df=N-1**), where

$$\hat{\sigma}_Y = \sqrt{\frac{\sum_{k=1}^{N} (Y_k - \bar{Y})^2}{N-1}}$$

## Density curves for Student's *t*

**We will come back to the T-distribution at the end of this chapter!**




Student(df) density curves for various *df*.



## Continuous Distributions – F-distribution

- F-distribution k-samples of different sizes.
- Snedecor's F distribution is most commonly used in tests of variance (e.g., ANOVA). The ratio of two chi-squares divided by their respective degrees of freedom is said to follow an F distribution

*k* {
- $\{Y_{1;1}, Y_{1;2}, \ldots\ldots\ldots\ldots, Y_{1;N1}\}$ IID from a Normal($\mu_1;\sigma_1$)
- $\{Y_{2;1}, Y_{2;2},.., Y_{2;N2}\}$ IID from a Normal($\mu_2;\sigma_2$)
- .,..
- $\{Y_{k;1}, Y_{k;2}, \ldots.., Y_{k;N2}\}$ IID from a Normal($\mu_2;\sigma_2$)
}

- $\sigma_1= \sigma_2= \sigma_3=\ldots \sigma_{n_k}= \sigma$. ($1/2 <= \sigma_k/\sigma_j<=2$)
- Samples are independent!



## Continuous Distributions – F-distribution

- F-distribution k-samples of different sizes

**TABLE 10.3.2 Typical Analysis-of-Variance Table for One-Way ANOVA**

| Source | Sum of squares | *df* | Mean sum of Squares[a] | *F*-statistic | *P*-value |
|---|---|---|---|---|---|
| Between | $\sum n_i(\bar{x}_i.-\bar{x}..)^2$ | $k$ - 1 | $s_B^2$ | $f_0=s_B^2/s_W^2$ | pr($F\geq f_0$) |
| Within | $\sum(n_i-1)s_i^2$ | $n_{tot}$ - $k$ | $s_W^2$ | | |
| Total | $\sum\sum(x_{ij}-\bar{x}..)^2$ | $n_{tot}$ - 1 | | | |

[a]Mean sum of squares = (sum of squares)/*df*

$$s_B^2 = \frac{\sum n_i(\bar{x}_i. - \bar{x}..)^2}{k-1}$$

$$s_W^2 = \frac{\sum(n_i-1)s_i^2}{n_{tot}-k}$$

- $s^2_B$ is a measure of variability of sample means, how far apart they are.
- $s^2_W$ reflects the avg. internal variability within the samples.



## Continuous Distributions – $\chi^2$ [Chi-Square]

- $\chi^2$ [Chi-Square] goodness of fit test:
  - Let $\{X_1, X_2, \ldots, X_N\}$ are IID N(0, 1)
  - $W = X_1^2 + X_2^2 + X_3^2 + \ldots + X_N^2$
  - **W ~ $\chi^2$(df=N)**
  - Note: If $\{Y_1, Y_2, \ldots, Y_N\}$ are **IID N($\mu$, $\sigma$),** then

$$SD(Y) = \frac{1}{N-1}\sum_{k=1}^{N}\left(Y_k - \bar{Y}\right)^2$$

  - And the Statistics **W ~ $\chi^2$(df=N-1)**

$$W = \frac{N-1}{\sigma^2}SD^2(Y)$$

$$X^2 = \sum_{k=1}^{N}\frac{(O_k - E_k)^2}{E_k} \sim \chi 2$$

  - E(W)=N; Var(W)=2N



## Continuous Distributions – Cauchy's

- Cauchy's distribution, X~Cauchy(t,s), t=location; s=scale
- PDF(X): $f(x) = \frac{1}{s\pi\left(1+((x-t)/s)^2\right)}$; x ∈ ***R*** (reals)
- PDF(Std Cauchy's(0,1)): $f(x) = \frac{1}{s\pi\left(1+x^2\right)}$
- The Cauchy distribution is (theoretically) important as an example of a *pathological case*. Cauchy distributions look similar to a normal distribution. However, they have much heavier tails. When studying hypothesis tests that assume normality, seeing how the tests perform on data from a Cauchy distribution is a good indicator of how sensitive the tests are to heavy-tail departures from normality. The mean and standard deviation of the Cauchy distribution are undefined!!! The practical meaning of this is that collecting 1,000 data points gives no more accurate of an estimate of the mean and standard deviation than does a single point.



## Continuous Distributions – Exponential

- Exponential distribution, X~Exponential($\lambda$)
- The exponential model, with only one unknown parameter, is the simplest of all life distribution models.

$$f(x) = \lambda e^{-\lambda x}; \quad x \geq 0$$

- E(X)=1/ $\lambda$; Var(X)=1/ $\lambda^2$;
- Another name for the exponential mean is the **Mean Time To Fail** or **MTTF** and we have MTTF = 1/ $\lambda$.
- If ***X*** is the time between occurrences of rare events that happen on the average with a rate l per unit of time, then ***X*** is distributed exponentially with parameter $\lambda$. Thus, the exponential distribution is frequently used to model the time interval between successive random events. Examples of variables distributed in this manner would be the gap length between cars crossing an intersection, life-times of electronic devices, or arrivals of customers at the check-out counter in a grocery store.

## Continuous Distributions – Exponential

- Exponential distribution, Example: ***By-hand vs. ProbCalc.htm***
- On weeknight shifts between 6 pm and 10 pm, there are an average of 5.2 calls to the UCLA medical emergency number. Let X measure the time needed for the first call on such a shift. Find the probability that the first call arrives (a) between 6:15 and 6:45 (b) before 6:30. Also find the median time needed for the first call ( 34.578%; 72.865% ).
  - We must first determine the correct average of this exponential distribution. If we consider the time interval to be 4x60=240 minutes, then on average there is a call every 240 / 5.2 (or 46.15) minutes. Then X ~ Exp(1/46), [E(X)=46] measures the time in minutes after 6:00 pm until the first call.



## Continuous Distributions – Exponential Examples

- Customers arrive at a certain store at an average of 15 per hour. What is the probability that the manager must wait at least 5 minutes for the first customer?
- The exponential distribution is often used in probability to model (remaining) lifetimes of mechanical objects for which the average lifetime is known and for which the probability distribution is assumed to decay exponentially.
- Suppose after the first 6 hours, the average remaining lifetime of batteries for a portable compact disc player is 8 hours. Find the probability that a set of batteries lasts between 12 and 16 hours.

**Solutions**:

- Here the average waiting time is 60/15=4 minutes. Thus X ~ exp(1/4). E(X)=4. Now we want P(X>5)=1-P(X <= 5). We obtain a right tail value of .2865. So around 28.65% of the time, the store must wait at least 5 minutes for the first customer.
- Here the remaining lifetime can be assumed to be X ~ exp(1/8). E(X)=8. For the total lifetime to be from 12 to 16, then the remaining lifetime is from 6 to 10. We find that P(6 <= X <= 10) = .1859.

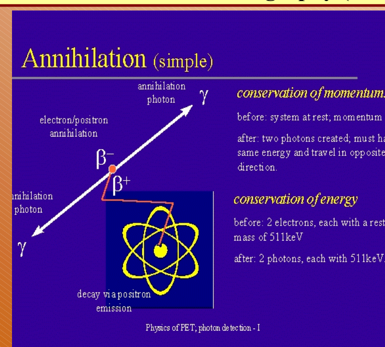


## Poisson Distribution – Definition

- Used to model counts – number of arrivals (k) on a given interval …
- The Poisson distribution is also sometimes referred to as the **distribution of rare events**. Examples of Poisson distributed variables are number of accidents per person, number of sweepstakes won per person, or the number of catastrophic defects found in a production process.

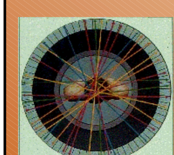


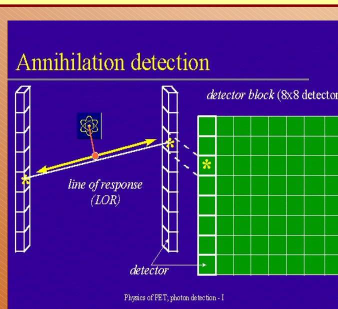
## Functional Brain Imaging – Positron Emission Tomography (PET)






## Functional Brain Imaging - Positron Emission Tomography (PET)

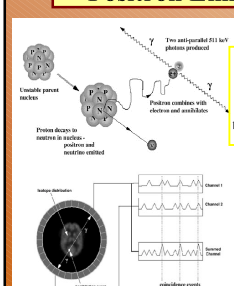





http://www.nucmed.buffalo.edu

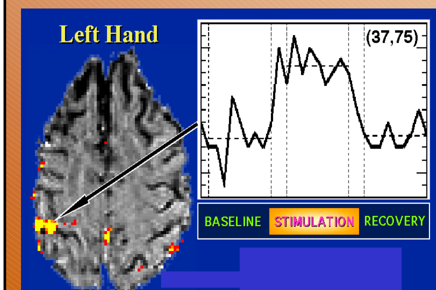


## Functional Brain Imaging – Positron Emission Tomography (PET)



| Isotope | Energy (MeV) | Range(mm) | 1/2-life | Appl. |
|---|---|---|---|---|
| $^{11}C$ | 0.96 | 1.1 | 20 min | receptors |
| $^{15}O$ | 1.7 | 1.5 | 2 min | stroke/activation |
| $^{18}F$ | 0.6 | 1.0 | 110 min | neurology |
| $^{124}I$ | ~2.0 | 1.6 | 4.5 days | oncology |

## Functional Brain Imaging – Positron Emission Tomography (PET)






## Poisson Distribution – Mean

- Used to model counts – number of arrivals (k) on a given interval …
- Y~Poisson($\lambda$), then P(Y=k) = $\frac{\lambda^k e^{-\lambda}}{k!}$ , k = 0, 1, 2, …
- Mean of Y, $\mu_Y = \lambda$, since

$$E(Y) = \sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{k\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} =$$

$$= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$



## Poisson Distribution - Variance

- Y~Poisson($\lambda$), then P(Y=k) = $\frac{\lambda^k e^{-\lambda}}{k!}$ , k = 0, 1, 2, …
- Variance of Y, $\sigma_Y = \lambda$, since

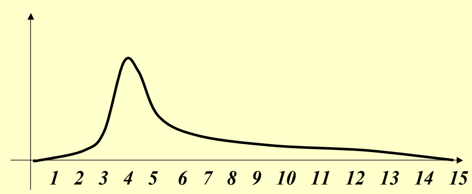
$$\sigma_Y^2 = Var(Y) = \sum_{k=0}^{\infty} (k-\lambda)^2 \frac{\lambda^k e^{-\lambda}}{k!} = ... = \lambda$$

- For example, suppose that Y denotes the number of blocked shots (arrivals) in a randomly sampled game for the UCLA Bruins men's basketball team. Then a Poisson distribution with mean=4 *may be* used to model Y .



## Poisson Distribution - Example

- For example, suppose that Y denotes the number of blocked shots in a randomly sampled game for the UCLA Bruins men's basketball team. Poisson distribution with mean=4 may be used to model Y .






## Poisson as an *approximation* to Binomial

- Suppose we have a sequence of Binomial(n, $p_n$) models, with $\lim(n\, p_n) \to \lambda$, as $n \to$ infinity.
- For each 0<=y<=n, if $Y_n$~ Binomial(n, $p_n$), then
  - P($Y_n$=y)= $\binom{n}{y} p_n^{\ y} (1-p_n)^{n-y}$
  - But this converges to:

$$\binom{n}{y} p_n^{\ y} (1-p_n)^{n-y} \xrightarrow[\substack{n \to \infty \\ n \times p_n \to \lambda}]{\text{WHY?}} \frac{\lambda^y e^{-\lambda}}{y!}$$

- Thus, Binomial(n, $p_n$) → Poisson($\lambda$)



## Poisson as an *approximation* to Binomial

- **Rule of thumb** is that approximation is good if:
  - **n>=100**
  - **p<=0.01**
  - **$\lambda$ =n p <=20**
- Then, Binomial(n, $p_n$) → Poisson($\lambda$)

## Example using Poisson approx to Binomial

- Suppose P(defective chip) = 0.0001=$10^{-4}$. Find the probability that a lot of 25,000 chips has > 2 defective!
- Y~ Binomial(25,000, 0.0001), find P(Y>2). Note that Z~Poisson($\lambda$ =n p =25,000 x 0.0001=2.5)

$$P(Z>2)=1-P(Z\le 2)=1-\sum_{z=0}^{2}\frac{2.5^z}{z!}e^{-2.5}=$$

$$1-\left(\frac{2.5^0}{0!}e^{-2.5}+\frac{2.5^1}{1!}e^{-2.5}+\frac{2.5^2}{2!}e^{-2.5}\right)=0.456$$



## Normal *approximation* to Binomial

- Suppose Y~Binomial(n, p)
- Then $Y=Y_1+Y_2+Y_3+\ldots+Y_n$, where
  - $Y_k$~Bernoulli(p) , $E(Y_k)$=p & $Var(Y_k)$=p(1-p) →
  - E(Y)=np & Var(Y)=np(1-p), SD(Y)= $(np(1-p))^{1/2}$
  - Standardize Y:
    - Z=(Y-np) / $(np(1-p))^{1/2}$
    - By CLT → Z ~ N(0, 1). So, Y ~ N [np, $(np(1-p))^{1/2}$]
- Normal Approx to Binomial is reasonable when np >=10 & n(1-p)>10 (p & (1-p) are NOT too small relative to n).



## Normal *approximation* to Binomial – Example

- Roulette wheel investigation:
- Compute P(Y>=58), where Y~Binomial(100, 0.47) –
  - The proportion of the Binomial(100, 0.47) population having more than 58 reds (successes) out of 100 roulette spins (trials).
  - Since np=47>=10 & n(1-p)=53>10 Normal approx is justified.
- Z=(Y-np)/Sqrt(np(1-p)) = 58 – 100*0.47)/Sqrt(100*0.47*0.53)=2.2
- P(Y>=58) ←→ P(Z>=2.2) = 0.0139
- True P(Y>=58) = 0.177, using SOCR (demo!)
- Binomial approx useful when no access to SOCR avail.

Roulette has 38 slots 18red 18black 2 neutral



## Normal *approximation* to Poisson

- Let $X_1$~Poisson($\lambda$) & $X_2$~Poisson($\mu$) → $X_1+X_2$~Poisson($\lambda+\mu$)
- Let $X_1, X_2, X_3, \ldots, X_k$ ~ Poisson($\lambda$), and independent,
- $Y_k = X_1 + X_2 + \cdots + X_k$ ~ Poisson($k\lambda$), $E(Y_k)=Var(Y_k)=k\lambda$.
- The random variables in the sum on the right are independent and each has the Poisson distribution with parameter $\lambda$.
- By CLT the distribution of the standardized variable $(Y_k - k\lambda) / (k\lambda)^{1/2}$ → N(0, 1), as *k* increases to infinity.
- So, for $k\lambda$ >= 100, $Z_k = \{(Y_k - k\lambda) / (k\lambda)^{1/2}\}$ ~ N(0,1).
- → $Y_k$ ~ N($k\lambda$, $(k\lambda)^{1/2}$).



## Poisson or Normal *approximation* to Binomial?

- Poisson Approximation (Binomial(n, $p_n$) → Poisson($\lambda$) ):

$$\binom{n}{y}p_n^{\,y}(1-p_n)^{n-y}\ \xrightarrow[\substack{n\longrightarrow\infty\\ n\times p_n\longrightarrow\lambda}]{\text{WHY?}}\ \frac{\lambda^y e^{-\lambda}}{y!}$$

  - n>=100 & p<=0.01 & $\lambda$ =n p <=20
- Normal Approximation

  (Binomial(n, p) → *N* ( np, $(np(1-p))^{1/2}$ ) )

  - np >=10 & n(1-p)>10



## Exponential *family* and arrival numbers/times

- First, let $T_k$ denote the time of the *k*'th arrival for *k* = 1, 2, ... The *gamma experiment* is to run the process until the *k*'th arrival occurs and note the time of this arrival.
- Next, let $N_t$ denote the number of arrivals in the time interval (0, *t*] for $t \ge 0$. The *Poisson experiment* is to run the process until time *t* and note the number of arrivals.
- How are $T_k$ & $N_t$ related?
- $N_t \ge k$ ←→ $T_k \le t$

*density function of the k'th arrival time is $f_k(t) = (rt)^{k-1}re^{-rt} / (k-1)!$, $t > 0$. This distribution is the gamma distribution with shape parameter k and rate parameter r. Again, 1/r is knows as the scale parameter. A more general version of the gamma distribution, allowing non-integer k,*

## Independence of continuous RVs

- The RV's $\{Y_1, Y_2, Y_3, \ldots, Y_n\}$ are independent if for any n-tuple $\{y_1, y_2, y_3, \ldots, y_n\}$

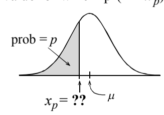
$$P(\{Y_1 \le y_1\} \cap \{Y_2 \le y_2\} \cap \{Y_3 \le y_3\} \cap \ldots \cap \{Y_n \le y_n\}) =$$
$$= P(Y_1 \le y_1) \times P(Y_2 \le y_2) \times P(Y_3 \le y_3) \times \ldots \times P(Y_n \le y_n)$$



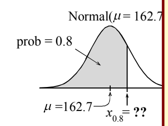
## The inverse problem – Percentiles/quantiles

(a) *p*-Quantile

**Programs supply $x_p$**
*x*-value for which $pr(X \le x_p) = p$




(b) **80th percentile** (0.8-quantile) **of women's heights**




80% of people have height below the 80th percentile. This is EQ to saying there's 80% chance that a random observation from the distribution will fall below the 80th percentile.

(c) **Further percentiles of women's heights**

| **Percent** | 1% | 5% | 10% | 20% | 30% | 70% | 80% | 90% | 95% |
|---|---|---|---|---|---|---|---|---|---|
| **Propn** | 0.01 | 0.05 | 0.1 | 0.2 | 0.3 | 0.7 | 0.8 | 0.9 | 0.95 |
| **Percentile** | *(or quantile)* | | | | | | | | |
| ***(cm)*** | 148.3 | 152.5 | 154.8 | 157.5 | 159.4 | 166.0 | 167.9 | 170.6 | 172.9 |
| ***(ft'in")*** | 4'10" | 5'0" | 5'0" | 5'2" | 5'2" | 5'5" | 5'6" | 5'7" | 5'8" |

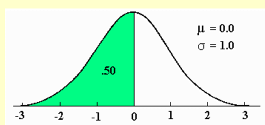
The inverse problem is what is the height for the 80th percentile/quantile? So far we studied given the height value what's the corresponding percentile?



## Standard Normal Curve

- The standard normal curve is described by the equation:

$$y = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$



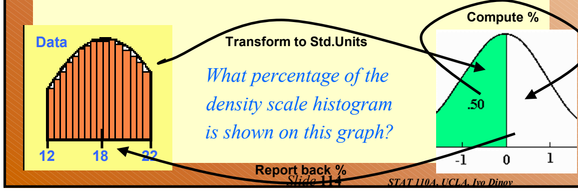

**Where remember, the natural number e ~ 2.7182…**

**We say: X~Normal(μ, σ), or simply X~N(μ, σ)**



## Standard Normal Approximation

- The standard normal curve can be used to estimate the percentage of entries in an interval for any process. Here is the protocol for this approximation:
  - Convert the interval (we need the assess the percentage of entries in) to standard units. We saw the algorithm already.
  - Find the corresponding area under the normal curve (from tables or online databases);



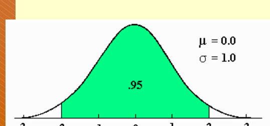


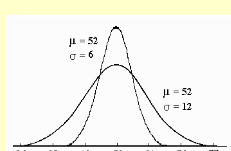

## General Normal Curve

- The general normal curve is defined by:
  - Where μ is the average of (the symmetric) normal curve, and σ is the standard deviation (spread of the distribution).

$$y = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$

  - Why worry about a standard and general normal curves?
  - How to convert between the two curves?

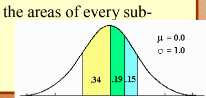





## Areas under Standard Normal Curve – Normal Approximation

- Protocol:
  - Convert the interval (we need to assess the percentage of entries in) to Standard units. Actually convert the end points in Standard units.
    - In general, the transformation X → (X-μ)/σ, standardizes the observed value X, where μ and σ are the average and the standard deviation of the distribution X is drawn from.
  - Find the corresponding area under the normal curve (from tables or online databases);
    - Sketch the normal curve and shade the area of interest
    - Separate your area into individually computable sections
    - Check the Normal Table and extract the areas of every sub-section
    - Add/compute the areas of all sub-sections to get the total area.



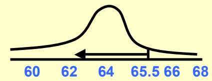
## Areas under Standard Normal Curve – **Example**

- Many histograms are similar in shape to the standard normal curve. For example, persons height. The height of all incoming female army recruits is measured for custom training and assignment purposes (e.g., very tall people are inappropriate for constricted space positions, and very short people may be disadvantages in certain other situations). The mean height is computed to be 64 *in* and the standard deviation is 2 *in*. Only recruits shorter than 65.5 in will be trained for tank operation and recruits within ½ standard deviations of the mean will have no restrictions on duties.
  - What percentage of the incoming recruits will be trained to operate armored combat vehicles (tanks)?
  - About what percentage of the recruits will have no restrictions on training/duties?



## Areas under Standard Normal Curve - **Example**

- The mean height is 64 *in* and the standard deviation is 2 *in*.
  - Only recruits shorter than 65.5 in will be trained for tank operation. What percentage of the incoming recruits will be trained to operate armored combat vehicles (tanks)?

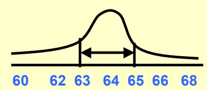



X → (X-64)/2

65.5 → (65.5-64)/2 = ¾

Percentage is 77.34%

  - Recruits within ½ standard deviations of the mean will have no restrictions on duties. About what percentage of the recruits will have no restrictions on training/duties?




X → (X-64)/2

65 → (65-64)/2 = ½

63 → (63-64)/2 = -½

Percentage is 38.30%



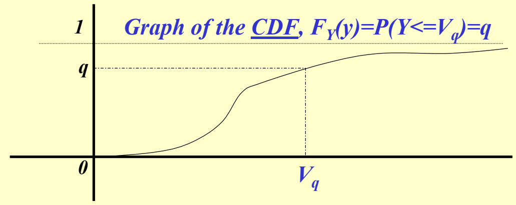
## Identifying Common Distributions – QQ plots

- **Plots** are useful for identifying candidate distribution model(s) in approximating a population (data) distribution.
- **Histograms**, can reveal much of the features of the data distribution.
- **Quantile-Quantile** plots indicate how well the model distribution agrees with the data.
- $q^{th}$ quantile, for 0<q<1, is the (data-space) value, $V_q$, at or below which lies a proportion q of the data.
- E.g., q=0.80, Y={1,2,3,4,5,6,7,8,9,10}. The $q^{th}$ quantile $V_q$= 8, since 80% of the data is at or below 8.



## Identifying Common Distributions – QQ plots

- **Quantile-Quantile** plots indicate how well the model distribution agrees with the data.
- $q^{th}$ quantile, for 0<q<1, is the (data-space) value, $V_q$, at or below which lies a proportion q of the data.






## Constructing QQ plots

- **Start off with data** $\{y_1, y_2, y_3, \ldots, y_n\}$
- **Order the data observations** $y_{(1)} <= y_{(2)} <= y_{(3)} <= \ldots <= y_{(n)}$
- Compute **quantile rank**, $q_{(k)}$, for each observation, $y_{(k)}$,

  $P(Y<= q_{(k)}) = (k-0.375) / (n+0.250)$, where

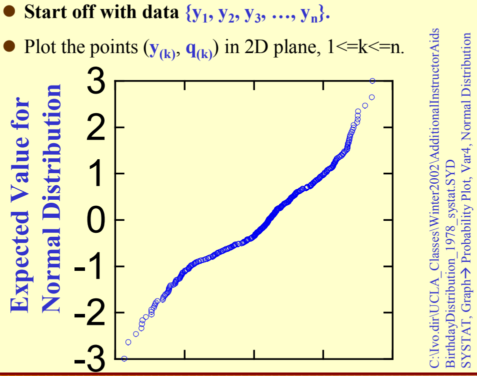
  Y is a RV from the (target) model distribution.
- Finally, plot the points $(y_{(k)}, q_{(k)})$ in 2D plane, 1<=k<=n.
- **Note:** Different statistical packages use slightly different formulas for the computation of $q_{(k)}$. However, the results are quite similar. This is the formulas employed in SAS.
- Basic idea: Probability that: **(model)Y<=(data)$y_1$ ~ 1/n;**

  **Y<=$y_2$ ~ 2/n; Y<=$y_3$ ~ 3/n; …**



## Example - Constructing QQ plots

- **Start off with data** $\{y_1, y_2, y_3, \ldots, y_n\}$.
- Plot the points $(y_{(k)}, q_{(k)})$ in 2D plane, 1<=k<=n.



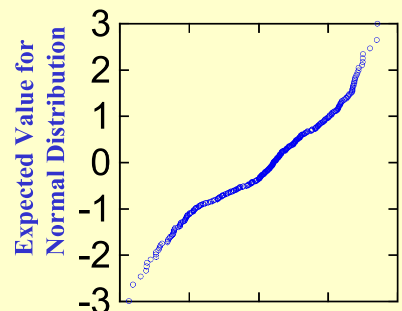
## Data transformations

- In practice oftentimes observed **data does not directly fit any of the models** we have available. In these cases transforming the raw data may provide/satisfy the requirements for using the distribution models we know.
- Common transformations: Y=T(X), X=raw data, Y=new
  - Data positively skewed to right use T(X)=Sqrt(X) or T(X)=log(X)
  - If data varies by more than 2 orders of magnitude
    - For X>0, use T(X)=log(X)
    - For any X, use T(X)= –1/X.
    - If X are counts (categorical var's), T(X)=Sqrt(X)
    - X=proportions & largest/ smallest Proportions >=2, use Logit transform: T(X) = log[X/(1-X)].



## Data transformations - Example

- For the BirthDay data:

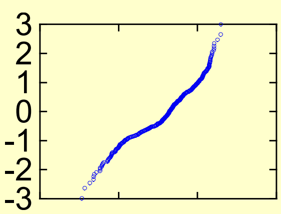



C:\Ivo.dir\UCLA_Classes\Winter2002\AdditionalInstructorAids
BirthdayDistribution_1978_systat.SYD
SYSTAT, Graph→ Probability Plot, Var4, Normal Distribution



## Data transformations - Example

- BirthDay data: C:\Ivo.dir\UCLA_Classes\Winter2002\AdditionalInstructorAids
BirthdayDistribution_1978_systat.SYD
SYSTAT, Graph→ Probability Plot, Var4, Normal Distribution

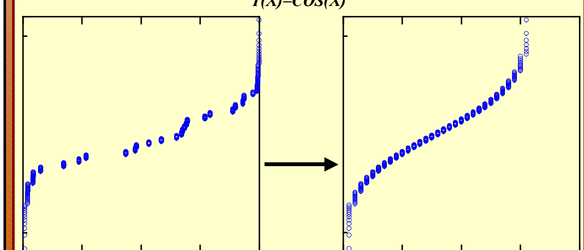





## Data transformations - Example

- BirthDay data: C:\Ivo.dir\UCLA_Classes\Winter2002\AdditionalInstructorAids
BirthdayDistribution_1978_systat.SYD
SYSTAT, Graph→ Probability Plot, COS(Var2), Normal Distribution

*T(X)=COS(X)*



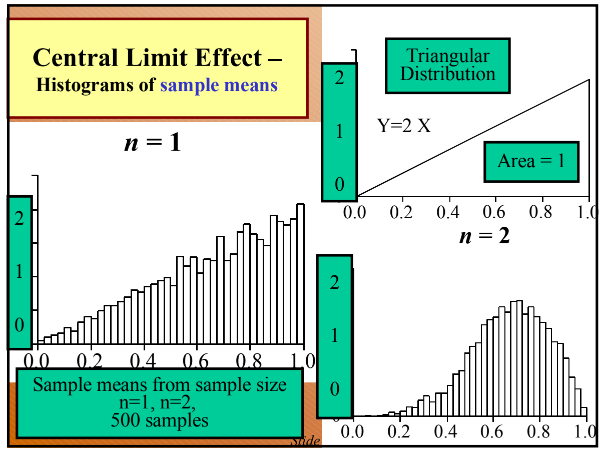


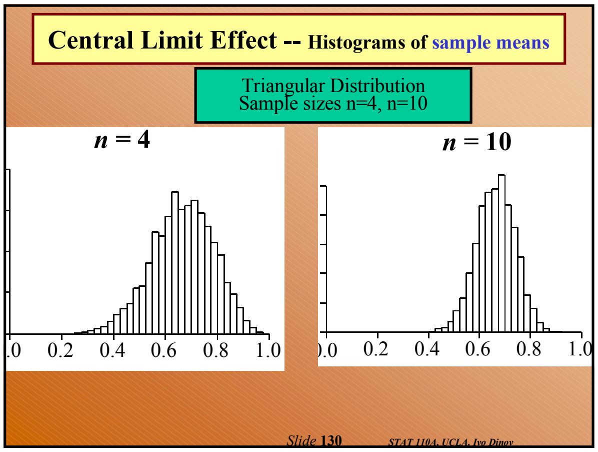
## Recall we looked at the sampling distribution of $\bar{X}$

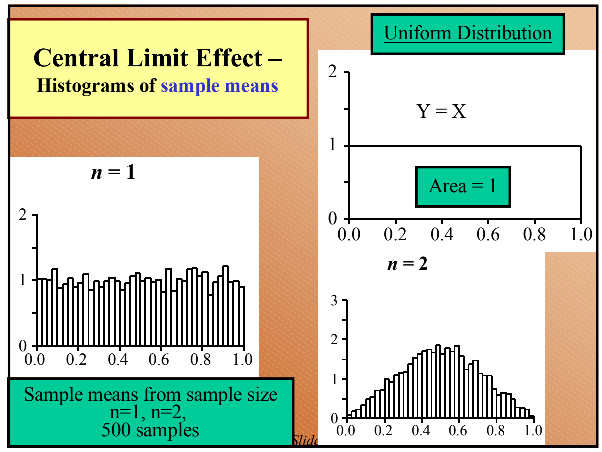
- For the sample mean calculated from a random sample, $E(\bar{X}) = \mu$ and $SD(\bar{X}) = \sigma/\sqrt{n}$, provided $\bar{X} = (X_1+X_2+ \dots + X_n)/n$, and $X_k \sim N(\mu, \sigma)$. Then
- $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$. And variability from sample to sample in the ***sample-means*** is given by the variability of the individual observations divided by the square root of the sample-size. In a way, averaging decreases variability.

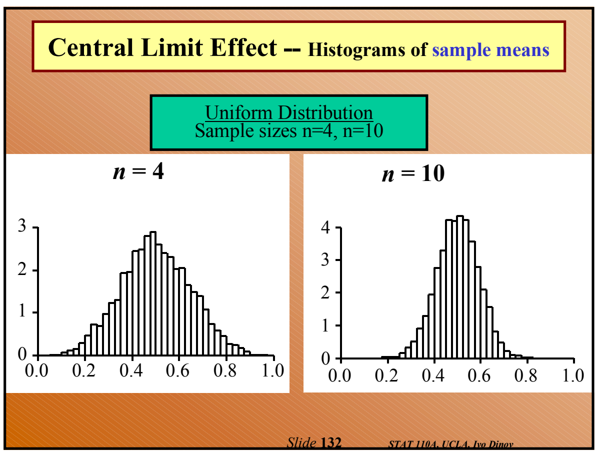


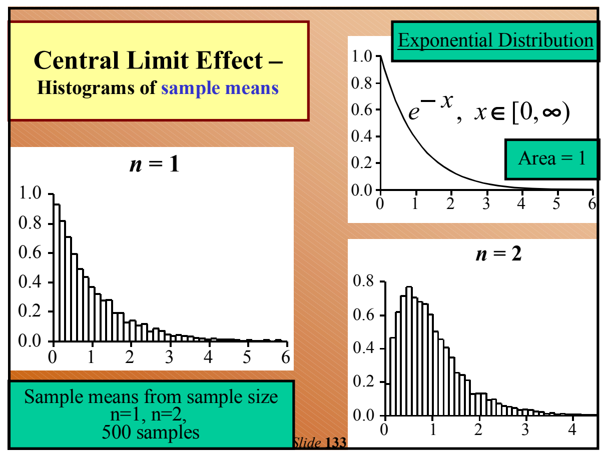
## Central Limit Effect – Histograms of sample means

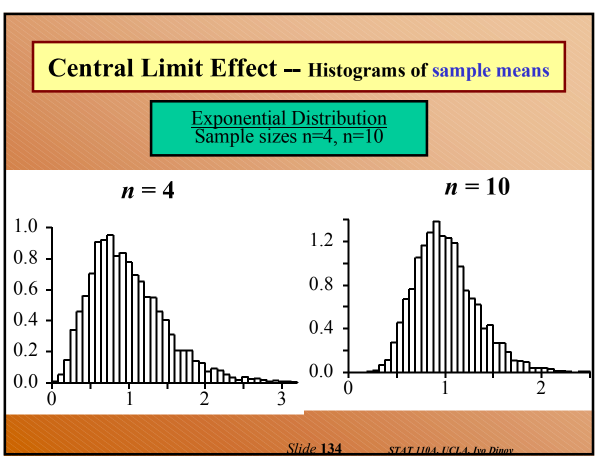


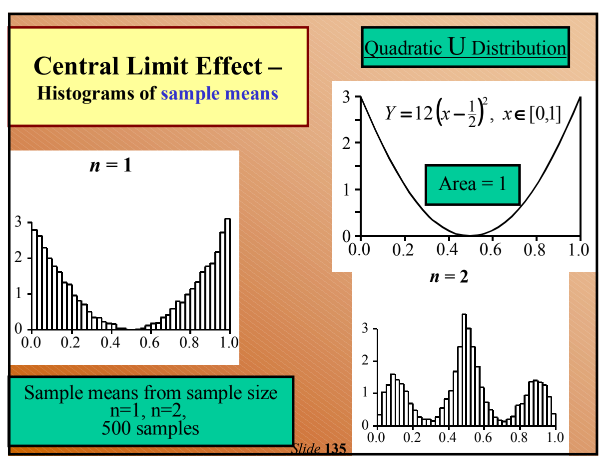

Central Limit Effect -- Histograms of sample means
Triangular Distribution
Sample sizes n=4, n=10
n = 4
n = 10
Slide 130
STAT 110A, UCLA, Ivo Dinov


Central Limit Effect –
Histograms of sample means
Uniform Distribution
Y = X
Area = 1
n = 1
n = 2
Sample means from sample size
n=1, n=2,
500 samples


Central Limit Effect -- Histograms of sample means
Uniform Distribution
Sample sizes n=4, n=10
n = 4
n = 10
Slide 132
STAT 110A, UCLA, Ivo Dinov

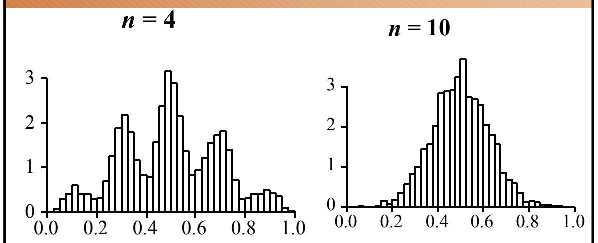

Central Limit Effect –
Histograms of sample means
Exponential Distribution
$e^{-x}$, $x \in [0, \infty)$
Area = 1
n = 1
n = 2
Sample means from sample size
n=1, n=2,
500 samples
Slide 133


Central Limit Effect -- Histograms of sample means
Exponential Distribution
Sample sizes n=4, n=10
n = 4
n = 10
Slide 134
STAT 110A, UCLA, Ivo Dinov


Central Limit Effect –
Histograms of sample means
Quadratic U Distribution
$Y = 12\left(x - \frac{1}{2}\right)^2$, $x \in [0,1]$
Area = 1
n = 1
n = 2
Sample means from sample size
n=1, n=2,
500 samples
Slide 135

## Central Limit Effect -- Histograms of sample means

Quadratic U Distribution
Sample sizes n=4, n=10






## Central Limit Theorem – heuristic formulation

**Central Limit Theorem:**
When sampling from almost any distribution,
$\overline{X}$ is approximately Normally distributed in large samples.

Show Sampling Distribution Simulation Applet:
file:///C:/Ivo.dir/UCLA_Classes/Winter2002/AdditionalInstructorAids/SamplingDistributionApplet.html



## Central Limit Theorem – theoretical formulation

Let $\{X_1, X_2, \ldots, X_k, \ldots\}$ be a sequence of independent observations from one specific random process. Let and $E(X) = \mu$ and $SD(X) = \sigma$ and both be finite ($0 < \sigma < \infty$; $|\mu| < \infty$). If $\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$, sample-avg,

Then $\overline{X}$ has a distribution which approaches N(μ, $\sigma^2$/n), as $n \to \infty$.



## Review

- What does the central limit theorem say? Why is it useful? (If the sample sizes are large, the mean in Normally distributed, as a RV)
- In what way might you expect the central limit effect to differ between samples from a symmetric distribution and samples from a very skewed distribution? (Larger samples for non-symmetric distributions to see CLT effects)
- What other important factor, apart from skewness, slows down the action of the central limit effect?

  (Heavyness in the tails of the original distribution.)



## Review

- When you have data from a moderate to small sample and want to use a normal approximation to the distribution of $\overline{X}$ in a calculation, what would you want to do before having any faith in the results? (30 or more for the sample-size, depending on the skewness of the distribution of *X*. Plot the data - non-symmetry and heavyness in the tails slows down the CLT effects).
- Take-home message: CLT is an application of statistics of paramount importance. Often, we are not sure of the distribution of an observable process. However, the CLT gives us a theoretical description of the distribution of the sample means as the sample-size increases (N(μ, $\sigma^2$/n)).



## The standard error of the mean – remember …

- For the sample mean calculated from a random sample, SD( $\overline{X}$ ) = $\frac{\sigma}{\sqrt{n}}$ . This implies that the variability from sample to sample in the *sample-means* is given by the variability of the individual observations divided by the square root of the sample-size. In a way, averaging decreases variability.
- Recall that for *known* SD(X)=σ, we can express the SD( $\overline{X}$ ) = $\frac{\sigma}{\sqrt{n}}$ . How about if SD(X) is *unknown*?!?

## The standard error of the mean

**The standard error of the sample mean is** an estimate of the *SD* of the sample mean

- i.e. a measure of the precision of the sample mean as an estimate of the population mean
- given by $SE(\bar{x}) = \frac{\text{Sample standard deviation}}{\sqrt{\text{Sample size}}}$

$$SE(\bar{x}) = \frac{s_x}{\sqrt{n}}$$

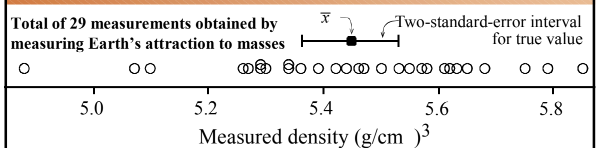
- Note similarity with
- $SD(\bar{X}) = \frac{\sigma}{\sqrt{n}}$ .



## Cavendish's 1798 data on mean density of the Earth, g/cm³, relative to that of $H_2O$

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 5.50 | 5.61 | 4.88 | 5.07 | 5.26 | 5.55 | 5.36 | 5.29 | 5.58 | 5.65 |
| 5.57 | 5.53 | 5.62 | 5.29 | 5.44 | 5.34 | 5.79 | 5.10 | 5.27 | 5.39 |
| 5.42 | 5.47 | 5.63 | 5.34 | 5.46 | 5.30 | 5.75 | 5.68 | 5.85 | |

Source: Cavendish [1798].




Newton's law of gravitation: $F = G m_1 m_2 / r^2$, the attraction force F is the ratio of the product (Gravitational const, mass of body1, mass body2) and the distance between them, r. Goal is to estimate G!



## Cavendish's 1798 data on mean density of the Earth, g/cm³, relative to that of $H_2O$

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 5.50 | 5.61 | 4.88 | 5.07 | 5.26 | 5.55 | 5.36 | 5.29 | 5.58 | 5.65 |
| 5.57 | 5.53 | 5.62 | 5.29 | 5.44 | 5.34 | 5.79 | 5.10 | 5.27 | 5.39 |
| 5.42 | 5.47 | 5.63 | 5.34 | 5.46 | 5.30 | 5.75 | 5.68 | 5.85 | |

Source: Cavendish [1798].

**Sample mean** $\bar{x} = 5.447931 \; g/cm^3$

**and sample SD =** $S_X = 0.2209457 \; g/cm^3$

**Then the standard error for these data is:**

$$SE(\bar{X}) = \frac{S_X}{\sqrt{n}} = \frac{0.2209457}{\sqrt{29}} = 0.04102858$$



## Student's *t*-distribution

- For random samples from a Normal distribution,

$$T = \frac{(\bar{X} - \mu)}{SE(\bar{X})}$$

Recall that for samples from N( μ , σ )

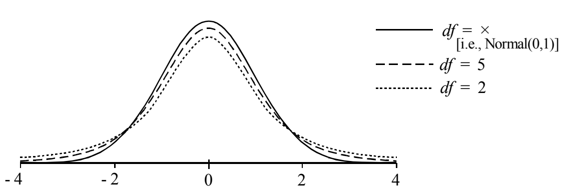
$$Z = \frac{(\bar{X} - \mu)}{SD(\bar{X})} = \frac{(\bar{X} - \mu)}{\sigma / \sqrt{n}} \sim N(0,1)$$

is exactly distributed as Student($df = n - 1$) ← Approx/Exact Distributions

- but methods we shall base upon this distribution for *T* work well even for small samples sampled from distributions which are quite non-Normal.
- *df* is number of observations –1, degrees of freedom.



## Density curves for Student's *t*

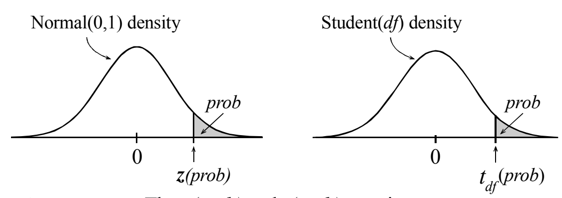



Student(df) density curves for various *df*.

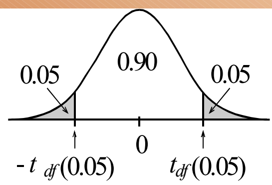


## Notation

- By $t_{df}(prob)$, we mean the number *t* such that when $T \sim$ Student(*df*), $P(T \geq t_{df}) = prob$; that is, the tail area above *t* (that is to the right of *t* on the graph) is *prob*.




The *z(prob)* and *t(prob)* notations.




The central 90% of the Student(*df*) distribution.



## Reading Student's *t* table

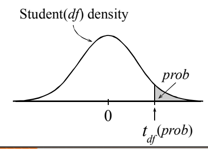



**TABLE 7.6.1 Extracts from the Student's t-Distribution Table**

*prob* (Desired *upper-tail prob*)

| df | .20 | .15 | .10 | .05 | .025 | .01 | .005 | .001 | .0005 | .0001 |
|---|---|---|---|---|---|---|---|---|---|---|
| 6 | 0.906 | 1.134 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | 5.208 | 5.959 | 8.025 |
| 7 | 0.896 | 1.119 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | 4.785 | 5.408 | 7.063 |
| 8 | 0.889 | 1.108 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | 4.501 | 5.041 | 6.442 |
| … | … | … | … | … | … | … | … | … | … | … |
| 10 | 0.879 | 1.093 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 4.144 | 4.587 | 5.694 |
| … | … | … | … | … | … | … | … | … | … | … |
| 15 | 0.866 | 1.074 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | 3.733 | 4.073 | 4.880 |
| … | | | | | | | | | | |
| ∞ | 0.842 | 1.036 | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 3.090 | 3.291 | 3.719 |

**Desired *df*** → 10; *t*-value → 2.228

Do we need an simulation of ***T*** and ***Z*** scores? Use the Online compute-engine …