UCLA STAT 110B

Applied Statistics for Engineering and the Sciences

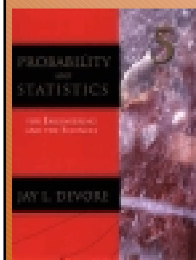
•**Instructor:** **Ivo Dinov,**
Asst. Prof. In Statistics and Neurology

•**Teaching Assistants:** **Brian Ng,** UCLA Statistics

University of California, Los Angeles, Spring 2003

http://www.stat.ucla.edu/~dinov/courses_students.html

Course Organization



Software: No specific software is required. SYSTAT, R, SOCR resource, etc.

Text: *Introduction to Probability and Statistics for Engineering and the Sciences 5th edition* -- Jay Devore

Course Description, Class homepage, online supplements, VOH's, etc.

http://www.stat.ucla.edu/~dinov/courses_students.html

Course Organization

- **Material Covered:** (Devore, Chapters 7-14)
  - Review of Key Concepts (ch 01-06)
  - Confidence Intervals (ch 07)
  - Single Sample Hypotheses testing (ch 08)
  - Inferences based on 2 samples (ch 09)
  - One- Two- and Three-Factor ANOVA (ch 10)
  - 2^k Factorial Designs (ch 11)
  - Linear Regression (ch 12)
  - Multiple & Nonlinear Regression (ch 13)
  - Goodness-of-Fit Testing (ch 14)

Overall Review

What is a statistic?

- Any quantity whose value can be calculated from sample data. It does not depend on any unknown parameter.
- Examples –

What are Random Variables?

- A function from the sample space to the real number line.

Before any data is collected, we view all observations and statistics as random variables

Properties of Expectation and Variance

- Let X be a random variable and a,b be constants. It follows that:

$$E[aX+b]=aE[X]+b$$

$$Var[aX+b]=a^2Var[X]$$

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

$$SD^2(X)=Var[X]$$

Linear Combinations of Random Variables

What if dependent?!?

Consider the collection of the independent random variables $X_1,\ldots,X_n$ where $E[X_i]=\mu_i$ and $Var[X_i]=\sigma_i^2$, and let $a_1,\ldots,a_n$ be constants. Define a random variable Y by

$$Y=a_1X_1+\ldots+a_nX_n$$

which is a linear combination of the X_i's. It follows that

$$E[a_1X_1+\ldots+a_nX_n]=a_1E[X_1]+\ldots+a_nE[X_n]=a_1\mu_1+\ldots+a_n\mu_n$$

$$Var[a_1X_1+\ldots+a_nX_n]=a_1^2Var[X_1]+\ldots+a_n^2Var[X_n]$$
$$=a_1^2\sigma_1^2+\ldots+a_n^2\sigma_n^2$$

Random Sample

$X_1,\ldots,X_n$ are an IID random sample of size n if:

1. The X_i's are **independent** random variables
2. Every X_i has the same (**identical**) probability distribution

These conditions are equivalent to the X_i's being independent and identically distributed (iid) random variables

Sample Mean and Total of a Random Sample

The sample mean is given by the random variable $\overline{X}$ defined as

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

The sample total is given by the random variable T_o defined as

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

Mean and Variance of T_o

For the total-sum random variable

$T_0 = X_1+\ldots+X_n$

$T_0 \sim N(n\mu, n\sigma^2)$.

Mean and Variance of $\overline{X}$

For the total-sum random variable

$\overline{X} = (1/n)(X_1+\ldots+X_n)$

$\overline{X} \sim N(\mu, \sigma^2/n)$.

Linear Combinations of Normal Random Variables from a Random Sample

Let $X_1,\ldots,X_n$ be a random sample from a normally distributed population with mean μ and variance σ^2, i.e. $X_i \sim N(\mu, \sigma^2)$. It follows that the random variable $Y = a_1X_1+\ldots+a_nX_n$ is normally distributed with mean $a_1\mu,\ldots,a_n\mu$ and variance $a_1^2\sigma^2+\ldots+a_n^2\sigma^2$. Hence, the sample mean and the sample total of the random sample will be normally distributed.

Central Limit Theorem

Arguably the most important theorem in Statistics (GUT theory)

The central limit theorem gives us information about the sample mean and the sample total for a "**large**" (n>30) random sample from a population that is not normally distributed. Specifically, it tells us that these will be approximately normally distributed. The larger n is, the better the approximation.

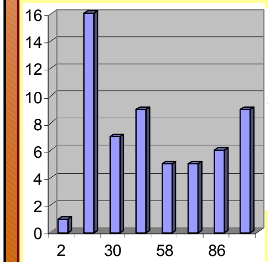
Example – Central Limit Theorem

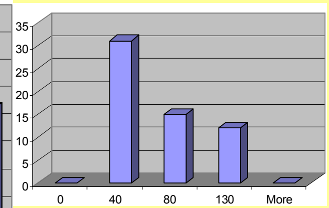
When a certain type of electrical resistor is manufactured, the mean resistance is 4 ohms with a standard deviation of 1.5 ohms. If 36 batches are independently produced, what is the probability that the **sample average resistance** of the batch is between 3.5 and 4.5 ohms. What is the probability that the sample **total resistance** is greater than 140 ohms?

Do **InteractiveNormalCurve** & CLT_Sampling Distribution Applets from SOCR resource

Uni- vs. Multi-modal histograms

- Number of clear humps on the frequency histogram plot determines the modality of a histogram plot.



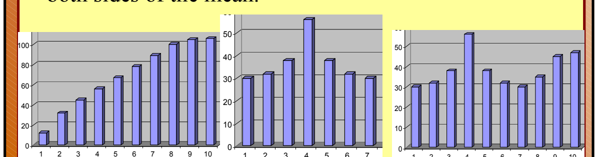


Skewness & Symmetry of histograms

- A histogram is **symmetric** is the bars (bins) to the left of some point (mean) are approximately mirror images of those to the right of the mean.

file:///C:/Ivo.dir/UCLA_Classes/Applets.dir/HistogramApplet.html

- Histogram is **skewed** if it is not symmetric, the histogram is heavy to the left or right, or non-identical on both sides of the mean.



Skewness & Kurtosis

- What do we mean by symmetry and positive and negative skewness? Kurtosis? Properties?!?

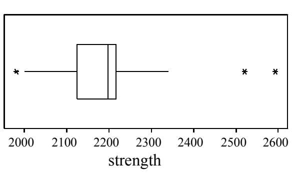
$$\text{Skewness} = \frac{\sum_{k=1}^{N}\left(Y_k - \overline{Y}\right)^3}{(N-1)SD^3}; \quad \text{Kurtosis} = \frac{\sum_{k=1}^{N}\left(Y_k - \overline{Y}\right)^4}{(N-1)SD^4}$$

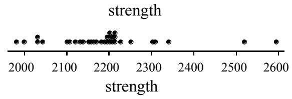
- Skewness is linearly invariant Sk(aX+b)=Sk(X)
- Skewness is a measure of **unsymmetry**
- Kurtosis is (also linearly invariant) a measure of **flatness**
- Both are used to quantify departures from StdNormal
- Skewness(StdNorm)=0; Kurtosis(StdNorm)=3

Comparing 3 plots of the same data

Stem-and-leaf of strength N = 33
Leaf Unit = 10

```
  1   19 8
  5   20 0334
  5   20
 10   21 00233
(8)   21 55668899
 15   22 000111112
  6   22 5
  5   23 014
  2   23
  2   24
  2   24
  2   25 2
  1   25 9
```





Three graphs of the breaking-strength data for}
gear-teeth in positions 4 & 10 (Minitab output).

Important points

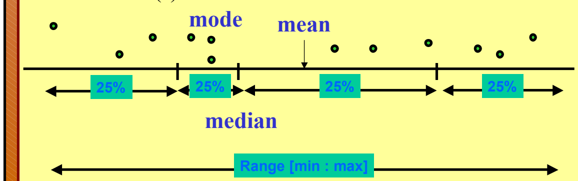
1. The distinction between a randomized experiment and an observational study is made at the time of result interpretation. The very same statistical analysis is carried for the two situations.
2. We've already stressed the importance of plotting data prior to stat-analysis. Plots have many important roles – prevent dangerous misconceptions from arising (data overlaps, clusters, outliers, skewness, trends in the data, etc.)

Analyzing Histogram Plots

- Modality – uni- vs. multi-modal (Why do we care?)
- Symmetry – how skewed is the histogram?
- Center of gravity for the Histogram plot – does it make sense?
- If center-of-gravity exists quantify the spread of the frequencies around this point.
- Strange patterns – gaps, atypical frequencies lying away from the center.

Measures of central tendency (location)

- **Mean** – sum of all observations divided by their number
- **Median** – (second quartile, Q_2) is the ***half-way-point*** for the distribution, 50% of all data are greater than it and 50% are smaller than Q_2.
- **Mode** – the (list of) most frequently occurring observation(s).



Measures of variability (deviation)

- **Mean Absolute Deviation (MAD)** –

$$MAD = \frac{1}{n-1}\sum_{i=1}^{n}|y_i - \bar{y}|$$

- **Variance** –

$$Var = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

- **Standard Deviation** –

$$SD = \sqrt{Var} = s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Measures of variability (deviation)

- **Example:**
- **Mean Absolute Deviation**– $MAD = \frac{1}{n-1}\sum_{i=1}^{n}|y_i - \bar{y}|$
- **Variance** – $Var = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$
- **Standard Deviation** – $SD = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}$

- X={1, 2, 3, 4}.

MAD=4/3=1.33

Var=5/3=1.67

SD=1.3

m=2.5

1 2 3 4

Trimmed, Winsorized means and Resistancy

- A data-driven **parameter estimate** is said to be **resistant** if it does not greatly change in the presence of outliers.
- **K-times trimmed mean**

$$\bar{y}_{tk} = \frac{1}{n-2k}\sum_{i=k+1}^{n-k} y_{(i)}$$

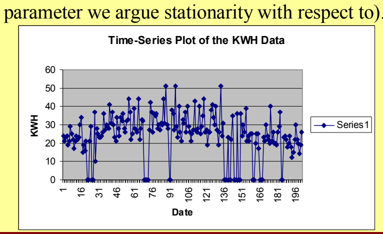
Order statistic

- **Winsorized k-times mean**:

$$\bar{y}_{wk} = \frac{1}{n}\left[(k+1)y_{(k+1)} + \sum_{i=k+2}^{n-k-1} y_{(i)} + (k+1)y_{(n-k)}\right]$$

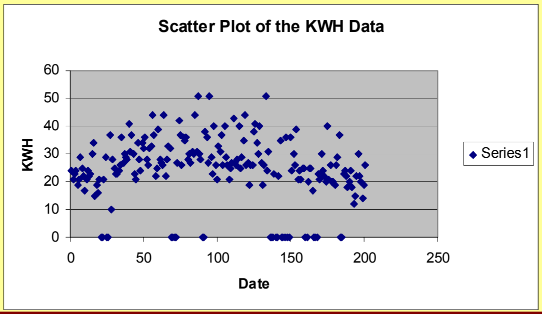
Stationary or Non-Stationary Process?

- **To assess stationarity:**
- **Rigorous assessment:** A stationary process has a **constant mean**, **variance**, and **autocorrelation** through time/place.
- **Visual assessment:** (Plot the data – observed vs. time/place – the parameter we argue stationarity with respect to).



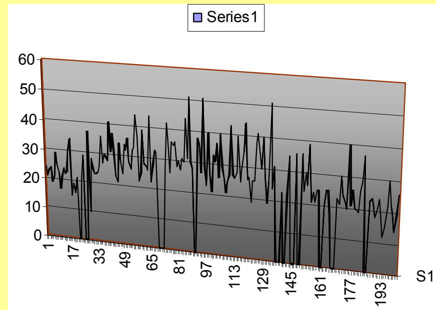
Stationary or **Non-Stationary** Process?

- **Visual assessment:** (Plot the data – observed vs. time/place, etc., – parameter we argue stationarity with respect to).



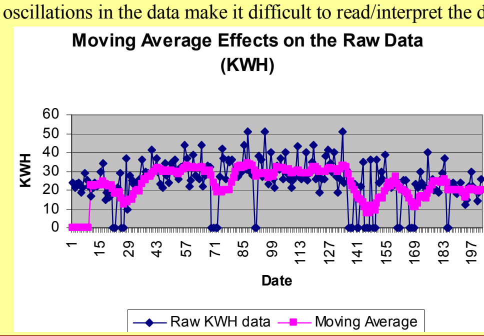
Moving Averages

- **Signal, Noise, Filtering:** Oftentimes high frequency oscillations in the data make it difficult to read/interpret the data.



Moving Averages – next 10 values are averaged

- **Signal, Noise, Filtering:** Oftentimes high frequency oscillations in the data make it difficult to read/interpret the data.



Properties of probability distributions

- A sequence of number $\{p_1, p_2, p_3, \ldots, p_n\}$ is a probability distribution for a sample space $S = \{s_1, s_2, s_3, \ldots, s_n\}$, if $pr(s_k) = p_k$, for each $1<=k<=n$. The two essential properties of a probability distribution $p_1, p_2, \ldots, p_n$?

$$p_k \geq 0; \quad \sum_k p_k = 1$$

- How do we get the probability of an event from the probabilities of outcomes that make up that event?
- If all outcomes are distinct & equally likely, how do we calculate $pr(A)$? If $A = \{a_1, a_2, a_3, \ldots, a_9\}$ and $pr(a_1)=pr(a_2)=\ldots=pr(a_9)=p$; then

$pr(A) = 9 \times pr(a_1) = 9p$.

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?

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

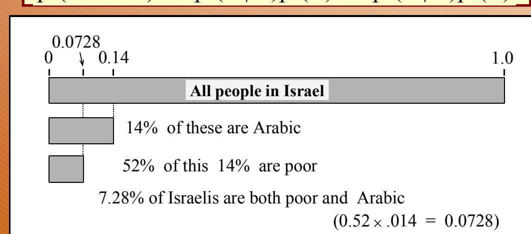


Illustration of the multiplication rule.

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

Permutation & Combination

Permutation: Number of **ordered** arrangements of **r** objects chosen from **n** ***distinctive*** objects

$$P_n^r = n(n-1)(n-2)\ldots(n-r+1)$$

$$P_n^n = P_n^{n-r} \cdot P_r^r$$

e.g. $P_6^3 = 6 \cdot 5 \cdot 4 = 120.$

Permutation & Combination

Combination: Number of **non-ordered** arrangements of r objects chosen from n ***distinctive*** objects:
$$C_n^r = P_n^r / r! = \frac{n!}{(n-r)!r!}$$

Or use notation of $\binom{n}{r} = C_n^r$

e.g. 3!=6 , 5!=120 , 0!=1

$$\binom{7}{3} = \frac{7!}{4!3!} = 35$$

Permutation & Combination

Combinatorial Identity:

$$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$$

Analytic proof: (expand both hand sides)

Combinatorial argument: Given n object focus on one of them (obj. 1). There are $\binom{n-1}{r-1}$ groups of size r that contain obj. 1 (since each group contains r-1 other elements out of n-1). Also, there are $\binom{n-1}{r}$ groups of size r, that do not contain obj1. But the total of all r-size groups of n-objects is $\binom{n}{r}$!

Permutation & Combination

Combinatorial Identity:

$$\binom{n}{r} = \binom{n}{n-r}$$

Analytic proof: (expand both hand sides)

Combinatorial argument: Given n objects the number of combinations of choosing any r of them is equivalent to choosing the remaining n-r of them (order-of-objs-not-important!)

Examples

1. Suppose car plates are 7-digit, like AB1234. If all the letters can be used in the first 2 places, and all numbers can be used in the last 4, how many different plates can be made? How many plates are there with no repeating digits?

Solution: a) $26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 \cdot 10$

b) $P_{26}^2 \cdot P_{10}^3 = 26 \cdot 25 \cdot 10 \cdot 9 \cdot 8 \cdot 7$

Examples

2. How many different letter arrangement can be made from the 11 letters of MISSISSIPPI?

Solution: There are: 1 M, 4 I, 4 S, 2 P letters.

Method 1: consider different permutations:

11!/(1!4!4!2!)=34650

Method 2: consider combinations:

$$\binom{11}{1}\binom{10}{4}\binom{6}{4}\binom{2}{2} = \ldots = \binom{11}{2}\binom{9}{4}\binom{5}{4}\binom{1}{1}$$

Examples

3. There are N telephones, and any 2 phones are connected by 1 line. Then how many lines are needed all together?

Solution: $C^2_N = N(N-1)/2$
If, N=5, complete graph with 5 nodes has C^2_5=10 edges.

Binomial theorem & multinomial theorem

Binomial theorem $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$

Deriving from this, we can get such useful formula (a=b=1)

$$\binom{n}{0}+\binom{n}{1}+\ldots+\binom{n}{n} = 2^n = (1+1)^n$$

Also from $(1+x)^{m+n}=(1+x)^m(1+x)^n$ we obtain:

$$\binom{m+n}{k} = \sum_{i=0}^{k} \binom{m}{i}\binom{n}{k-i}$$

On the left is the coeff of $1^k x^{(m+n-k)}$. On the right is the same coeff in the product of $(\ldots+\text{coeff}*x^{(m-i)}+\ldots)*(\ldots+\text{coeff}*x^{(n-k+i)}+\ldots)$.

Multinomial theorem

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

Generalization: Divide n distinctive objects into r groups, with the size of every group $n_1,\ldots,n_r$, and $n_1+n_2+\ldots+n_r = n$

$$(x_1+x_2+\ldots+x_r)^n = \sum \binom{n}{n_1,n_2,\ldots,n_r} x_1^{n_1} x_2^{n_2} \ldots x_r^{n_r}$$

where $\binom{n}{n_1,n_2,\ldots,n_r} = \binom{n}{n_1}\binom{n-n_1}{n_2}\ldots\binom{n-n_1-\ldots-n_{r-1}}{n_r} = \dfrac{n!}{n_1!n_2!\ldots n_r!}$

Sterling Formula for asymptotic behavior of n!

Sterling formula:

$$n! = \sqrt{\frac{2\pi}{n}} \times \left(\frac{n}{e}\right)^n$$

Probability and Venn diagrams

Proposition

$P(A_1 \cup A_2 \cup \ldots \cup A_n) =$

$$\sum_{i=1}^{n} P(A_i) - \sum_{1\le i1<i2\le n} P(A_{i1} \cap A_{i2}) + \ldots$$

$$+(-1)^{r+1} \sum_{1\le i1<j2<\ldots<ir\le n} P(A_{i1} \cap A_{i2} \cap \ldots \cap A_{ir}) + \ldots$$

$$+(-1)^{n+1} P(A_{i1} \cap A_{i2} \cap \ldots \cap A_{in})$$

Discrete Variables, Probabilities

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

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

n-factorial

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!

For the Binomial distribution . . . mean

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

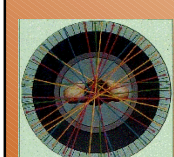
X~Binomial(n, p) ➔

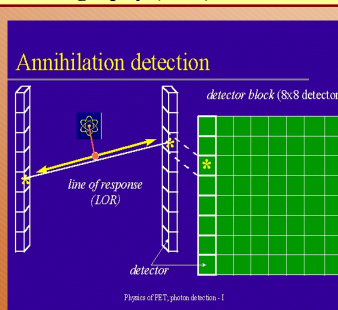
$X = Y_1 + Y_2 + Y_3 + .. + Y_n$,
where Y_k ~Bernoulli(p),
$E(Y_1) = p$ ➔
$E(X) = E(Y_1 + Y_2 + Y_3 + .. + Y_n) = np$

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)





http://www.nucmed.buffalo.edu

Poisson Distribution – Mean

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

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

- 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 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→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:

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

- Thus, Binomial(n, p_n) → Poisson(λ)

Poisson as an *approximation* to Binomial

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

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(λ =n p =25,000 x 0.0001=2.5)

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

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

Geometric, Hypergeometric, Negative Binomial

- X ~ Geometric(p), then the probability mass function is

Probability of first failure at x^{th} trial.

$$P(X=x) = (1-p)^{x-1}p; \quad E(X) = \frac{1-p}{p}; \quad Var(X) = \frac{1-p}{p^2}$$

- Ex: Stat dept purchases 40 light bulbs; 5 are defective.

  Select 5 components at random.

  Find: *P*(3rd bulb used is the first that does not work) = ?

Geometric, Hypergeometric, Negative Binomial

- Hypergeometric – X~HyperGeom(x; N, n, M)

Total objects: N. Successes: M. Sample-size: n (without replacement). X = number of Successes in sample

$$E(X) = n\frac{M}{N} \qquad P(X=x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

$$Var(X) = \frac{N-n}{N-1} \times n \times \frac{M}{N} \times \frac{N-M}{N}$$

Ex: 40 components in a lot; 3 components are defectives.

Select 5 components at random.

P(obtain one defective) = P(X=1) = ?

Hypergeometric Distribution & Binomial

- Binomial approximation to Hyperheometric
  - $\frac{n}{N}$ is small (usually < 0.1), then $\frac{M}{N} \approx p$

$$HyperGeom(x;N,n,M) \underset{M/N \approx p}{\overset{approaches}{\Rightarrow}} Bin(x;n,p)$$

Ex: 4,000 out of 10,000 residents are against a new tax. 15 residents are selected at random.

$P_{HyperGeom}$(at most 7 favor the new tax) = ? (0.78706)

Demo: **Applets.dir/ProbCalc.htm** (P_{Bin}(Y<=7)=0.7869]

HyperGeom(x; N=10^4, n=15, M=4x10^3) ➔ Bin(x;n=15,p=0.4)

Geometric, Hypergeometric, Negative Binomial

- Negative binomial pmf [X ~ NegBin(r, p), if r=1 ➔ Geometric (p)] $P(X=x)=(1-p)^{x-1}p$

Number of trials until the r^{th} success (negative, since number of successes (r) is fixed & number of trials (X) is random)

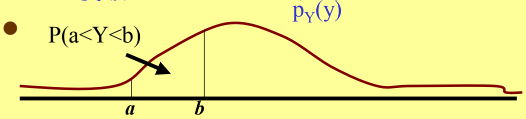
$$P(X=n)=\binom{n-1}{r-1}p^r(1-p)^{n-r}$$

$$E(X)=\frac{r}{p}; \quad Var(X)=\frac{r(1-p)}{p^2}$$

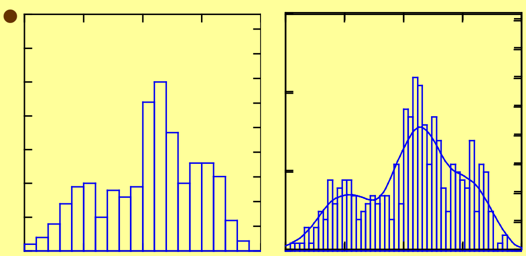
Find E(X) and Var(X)
X=# of times one must
Throw a dice until the
Outcome 1 occurs 4
Times:
X~NegBin(x;r=4,p=1/6)
E(X)=24; Var(X)=120

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).
- P(a<Y<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.
- 

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
$$p_Y(y) \geq 0, \forall y$$
- 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

- Uniform distribution
- Normal distribution
- Student's T distribution
- F-distribution
- Chi-squared (χ^2)
- Cauchy's distribution
- Exponential distribution
- Poisson distribution, …

(Continuous) Uniform Distribution

- $X \sim$ Uniform Distribution with parameters α and β if

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

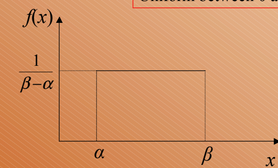
$$E(X) = \frac{\alpha+\beta}{2}, \quad Var(X) = \frac{(\beta-\alpha)^2}{12}$$

• random numbers follow Uniform between 0 and 1

Ex) Uniform, $\alpha = 2, \beta = 7$

(a) $P(X \geq 4) =$

(b) $P(3 < X < 5.5) =$



(General) Normal Distribution

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

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

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

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(μ;σ)
  - Variance σ^2 is unknown
- In 1908, William Gosset (pseudonym Student) derived the exact sampling distribution of the following statistics

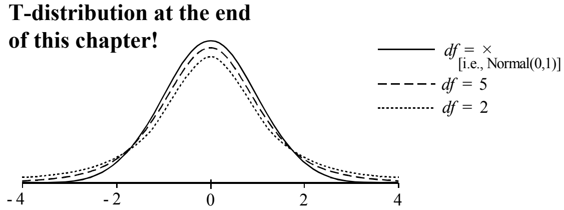
$$T = \frac{\bar{Y} - \mu_Y}{\hat{\sigma}_Y}$$

- 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*

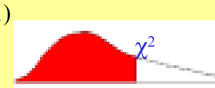
We will come back to the T-distribution at the end of this chapter!



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

Continuous Distributions – χ^2 [Chi-Square]

- χ^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 ~ χ^2(df=N)**
  - Note: If $\{Y_1, Y_2, \ldots, Y_N\}$ are **IID N(μ, σ²),** then

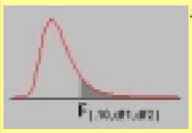


$$SD^2(Y) = \frac{1}{N-1}\sum_{k=1}^{N}(Y_k - \bar{Y})^2 \longrightarrow W = \frac{N-1}{\sigma^2}SD^2(Y)$$

  - And the Statistics **W ~ χ^2(df=N-1)**
  - E(W)=N; Var(W)=2N

Continuous Distributions – F-distribution

- F-distribution is the ratio of two χ^2 random variables.
- 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



$$SD^2(Y)=\frac{1}{N-1}\sum_{k=1}^{N}(Y_k-\bar{Y})^2;\quad SD^2(X)=\frac{1}{M-1}\sum_{l=1}^{M}(X_l-\bar{X})^2$$

$$W_Y=\frac{N-1}{\sigma_Y^2}SD^2(Y);\quad W_X=\frac{M-1}{\sigma_X^2}SD^2(X);$$

$$F_o=\frac{W_Y/(N-1)}{W_X/(M-1)}\sim \mathrm{F}(df_1=N-1, df_2=M-1)$$

Continuous Distributions – Cauchy's

- Cauchy's distribution, X~Cauchy(t,s), t=location; s=scale
- PDF(X): $f(x)=\frac{1}{s\pi\left(1+((x-t)/s)^2\right)};\quad x\in \boldsymbol{R}$ (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 (Cauchy=$T_{df=0}$→T_{df}→Normal).

Continuous Distributions – Exponential

- Exponential distribution, X~Exponential(λ)
- 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/ λ; Var(X)=1/ λ^2;
- Another name for the exponential mean is the **Mean Time To Fail** or **MTTF** and we have MTTF = 1/ λ.
- 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 λ. 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.

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.**

Roulette has 38 slots 18red 18black 2 neutral

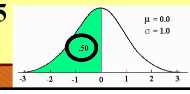
- **Z=(Y-np)/Sqrt(np(1-p)) = 58 – 100*0.47)/Sqrt(100*0.47*0.53)=2.2**
- **P(Y>=58) ⟵ ➔ P(Z>=2.2) = 0.0139**
- **True P(Y>=58) = 0.177, using SOCR (demo!)**
- **Binomial approx useful when no access to SOCR avail.**

Normal *approximation* to Poisson

- Let X_1~Poisson(λ) & X_2~Poisson(μ) $\rightarrow X_1 + X_2$~Poisson($\lambda+\mu$)
- Let $X_1, X_2, X_3, \ldots, X_k \sim$ Poisson(λ), and independent,
- $Y_k = X_1 + X_2 + \cdots + X_k \sim$ 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 λ.
- By CLT the distribution of the standardized variable $(Y_k - k\lambda) / (k\lambda)^{1/2} \rightarrow N(0, 1)$, as k increases to infinity.
- So, for $k\lambda >= 100$, $Z_k = \{(Y_k - k\lambda) / (k\lambda)^{1/2}\} \sim N(0,1)$.
- $\rightarrow Y_k \sim N(k\lambda, (k\lambda)^{1/2})$.

Normal *approximation* to Poisson – example

- Let X_1~Poisson(λ) & X_2~Poisson(μ) $\rightarrow X_1 + X_2$~Poisson($\lambda+\mu$)
- Let $X_1, X_2, X_3, \ldots, X_{200} \sim$ Poisson(2), and independent,
- $Y_k = X_1 + X_2 + \cdots + X_k \sim$ Poisson(400), $E(Y_k)=Var(Y_k)=400$.
- By CLT the distribution of the standardized variable $(Y_k - 400) / (400)^{1/2} \rightarrow N(0, 1)$, as k increases to infinity.
- $Z_k = (Y_k - 400) / 20 \sim N(0,1) \rightarrow Y_k \sim N(400, 400)$.
- $P(2 < Y_k < 400)$ = (std'z 2 & 400) =
- $P((2-400)/20 < Z_k < (400-400)/20) = P(-20 < Z_k < 0)$ = 0.5



Poisson or Normal *approximation* to Binomial?

- Poisson Approximation (Binomial(n, p_n) $\rightarrow$ Poisson(λ)):

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

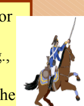
  - n>=100 & p<=0.01 & λ =n p <=20
- Normal Approximation

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

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

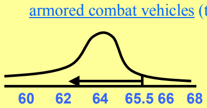
Areas under Standard Normal Curve – Example

- 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?



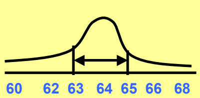
Areas under Standard Normal Curve - Example

- 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%

Gamma and Exponential Distributions

- **Gamma Distribution**
  - **Gamma function :**

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

Properties
$$\begin{cases} \Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1) & \\ \Gamma(n) = (n-1)! & \text{for positive integer } n \\ \Gamma(1) = 1 & \Gamma(0.5) = \sqrt{\pi} \end{cases}$$

- $X \sim$ Gamma with parameters α and β if

$$f(x) = \begin{cases} \dfrac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} & , x > 0 \\ 0 & , \text{otherwise} \end{cases}$$

where $\alpha>0$, $\beta>0$

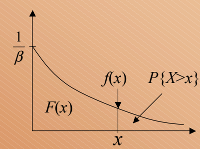
Gamma and Exponential Distributions

- **Exponential Distribution (cont'd)**

- CDF : $F(x) = P(X \le x) = \int_0^x \frac{1}{\beta} e^{-\frac{x}{\beta}} dx = 1 - e^{-\frac{x}{\beta}}, \ x > 0$

- Tail probability

$P(X > x) = 1 - F(x) = e^{-\frac{x}{\beta}}, \quad x > 0$



Ex 1) X = response time at a certain on-line computer terminal

$X \sim$ exponential with $E(X) = 5$(sec.).

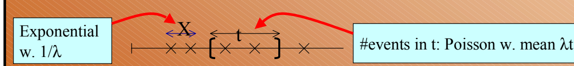
(a) $P(X \le 10) =$

(b) $P(5 \le X \le 10) =$

Gamma and Exponential Distributions

Relationship to the Poisson Process

<u># of events</u> in any time interval t has a Poisson distribution w/ parameter λt → the distribution of **the elapsed <u>time between</u> <u>two successive events</u>** is exponential with parameter $\beta = \frac{1}{\lambda}$



Why? Poisson : P(no events in t) = $P(0; \lambda t) = \frac{e^{-\lambda t}(\lambda t)^0}{0!} = e^{-\lambda t}$

Let X = time until the first event. $P(X > t) = e^{-\lambda t}$

Then P(no events in t) =

i.e., $P(0 \le X \le t) = 1 - e^{-\lambda t}$ = CDF of exponential with $\lambda = \frac{1}{\beta}$ or $\beta = \frac{1}{\lambda}$

Lognormal Distribution

- $X \sim$ lognormal with parameters μ and σ, if

$$\ln(X) \sim N(x; \mu, \sigma)$$

- $f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\,\sigma x} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} & x \ge 0 \\ 0 & , \text{ otherwise} \end{cases}$

- $E(X) = \exp(\mu + \sigma^2/2)$

$Var(X) = \exp(2\mu + \sigma^2)\{\exp(\sigma^2) - 1\}$

Ex) Let X ~ lognormal with parameter $\mu = 3.2$ and $\sigma, = 1$

$P(X > 8) =$

Weibull Distribution

- $X \sim$ Weibull Distribution with parameters α and β if

$f(x) = \begin{cases} \alpha\beta\, x^{\beta-1} e^{-\alpha x^{\beta}} & , x > 0 \\ 0 & , \text{ otherwise} \end{cases}$

- If $\beta = 1$; $f(x) = \alpha e^{-\alpha x}$ (exponential with parameter $\frac{1}{\alpha}$)

- $E(X) = \alpha^{-\frac{1}{\beta}} \Gamma(1 + \frac{1}{\beta})$ $\qquad F(x) = 1 - e^{-\alpha x^{\beta}}$

$Var(X) = \alpha^{-\frac{2}{\beta}} \{\Gamma(1 + \frac{2}{\beta}) - [\Gamma(1 + \frac{1}{\beta})]^2\}$

- Useful in Reliability, life testing problems

Beta Distribution

- Provides positive density only in an interval of finite length

$X \sim$ Beta Distribution with parameters α and β if

$f(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & , \ 0 < x < 1 \ (\alpha > 0, \ \beta > 0) \\ 0 & , \text{ otherwise} \end{cases}$

$E(X) = \frac{\alpha}{\alpha+\beta}$, $Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Ex)

X = proportion of TV sets requiring service during the first year

$\sim$ beta, $\alpha = 3$, $\beta = 2$.

P(at least 80% of the model sold this year will require service in 1 year)

Marginal & Joint PDF's Central Limit Theorem (CLT)

Joint probability mass function

- The joint probability mass function of the discrete random variables X and Y, denoted as $f_{XY}(x,y)$ satisfies:

$$(1) f_{XY}(x,y) \geq 0$$

$$(2) \sum_x \sum_y f_{XY}(x,y) = 1$$

$$(3) f_{XY}(x,y) = P(X = x, Y = y)$$

Joint probability mass function – example

The joint density, P{***X,Y***}, of the number of minutes waiting to catch the first fish, ***X*** , and the number of minutes waiting to catch the second fish, ***Y***, is given below.

P {***X*** = i,***Y*** = k }	k = 1	k = 2	k = 3	Row Sum P{ ***X*** = i }
i = 1	0.01	0.02	0.08	0.11
i = 2	0.01	0.02	0.08	0.11
i = 3	0.07	0.08	0.63	0.78
Column Sum P {***Y*** =k }	0.09	0.12	0.79	1.00

- The (joint) chance of **waiting 3 minutes to catch the first fish and 3 minutes to catch the second fish** is:
- The (marginal) chance of **waiting 3 minutes to catch the first fish** is:
- The (marginal) chance of **waiting 2 minutes to catch the first fish** is (circle all that are correct):
- The chance of **waiting at least two minutes to catch the first fish** is (circle none, one or more):
- The chance of **waiting at most two minutes to catch the first fish and at most two minutes to catch the second fish** is (circle none, one or more):

Marginal probability distributions

- Individual probability distribution of a random variable is referred to as its ***Marginal Probability Distribution.***
- Marginal probability distribution of X can be determined from the joint probability distribution of X and other random variables.
- Example: ***Marginal probability distribution of X is found by summing the probabilities in each column***, ***for y, summation is done in each row.***

Marginal probability distributions (Cont.)

- If X and Y are discrete random variables with joint probability mass function $f_{XY}(x,y)$, then the marginal probability mass function of X and Y are

$$f_X(x) = P(X = x) = \sum_{R_x} f_{XY}(X,Y)$$

$$f_Y(y) = P(Y = y) = \sum_{Ry} f_{XY}(X,Y)$$

where R_x denotes the set of all points in the range of (X, Y) for which $X = x$ and Ry denotes the set of all points in the range of (X, Y) for which $Y = y$

Mean and Variance

- If the marginal probability distribution of X has the probability function f(x), then

$$E(X) = \mu_X = \sum_x x f_X(x) = \sum_x x\left(\sum_{R_x} f_{XY}(x,y)\right) = \sum_x \sum_{R_x} x f_{XY}(x,y)$$

$$= \sum_R x f_{XY}(x,y)$$

$$V(X) = \sigma^2{}_X = \sum_x (x-\mu_X)^2 f_X(x) = \sum_x (x-\mu_X)^2 \sum_{R_x} f_{XY}(x,y)$$

$$= \sum_x \sum_{R_x} (x-\mu_X)^2 f_{XY}(x,y) = \sum_R (x-\mu_X)^2 f_{XY}(x,y)$$

- R = Set of all points in the range of (X,Y).
- Example.

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, σ²/n), as $n \to \infty$.

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 $\mathrm{SE}(\bar{x}) = \frac{\text{Sample standard deviation}}{\sqrt{\text{Sample size}}}$

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

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

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].

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:

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

Student's *t*-distribution

- For random samples from a Normal distribution,

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

Recall that for samples from N(μ , σ)

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

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.

Inference & Estimation

Parameters, Estimators, Estimates …

- E.g., We are interested in the population mean diameter (parameter) of washers the sample-average formula represents an estimator we can use, where as the value of the sample average for a particular dataset is the estimate (for the *mean* parameter).

$$\text{parameter} = \mu_Y; \qquad \text{estimator} = \overline{Y} = \frac{1}{N}\sum_{k=1}^{N} Y_k$$

$$\text{Data}: Y = \{0.1896,\ 0.1913,\ 0.1900\}$$

$$\text{estimate} = \bar{y} = \tfrac{1}{3}(0.1896 + 0.1913 + 0.1900)$$

$\bar{y} = 0.1903$. How about $\bar{y} = \tfrac{2}{3}(0.1896 + 0.1913 + 0.1900)$

A 95% confidence interval

- A type of interval that contains the true value of a parameter for 95% of samples taken is called a ***95% confidence interval*** for that parameter, the ends of the CI are called *confidence limits*.
- (For the situations we deal with) a confidence interval (CI) for the true value of a **parameter** is given by

estimate $\pm$ ***t standard errors (SE)***

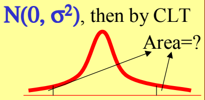
Value of the Multiplier, *t*, for a 95% CI

df :	7	8	9	10	11	12	13	14	15	16	17
t :	2.365	2.306	2.262	2.228	2.201	2.179	2.160	2.145	2.131	2.120	2.110
df :	18	19	20	25	30	35	40	45	50	60	∞
t :	2.101	2.093	2.086	2.060	2.042	2.030	2.021	2.014	2.009	2.000	1.960

(General) Confidence Interval (CI)

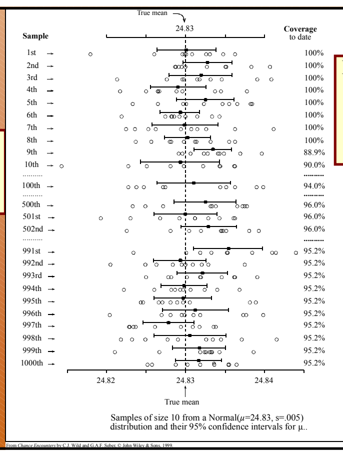
- A **level L confidence interval** for a parameter (θ), is an interval (θ_1^ , θ_2^), where θ_1^ & θ_2^, are estimators of θ, such that $P(\theta_1\hat{} < \theta < \theta_2\hat{}) = L$.
- E.g., **C+E model**: $Y = \mu+\varepsilon$. Where $\varepsilon \sim N(0, \sigma^2)$, then by CLT we have $Y\_bar \sim N(\mu, \sigma^2/n)$

➔ $n^{1/2}(Y\_bar - \mu)/\sigma \sim N(0, \sigma^2)$.



- $L = P\,(z_{(1-L)/2} < n^{1/2}(Y\_bar - \mu)/\sigma < z_{(1+L)/2})$, where z_q is the q^{th} quartile.
- E.g., $0.95 = P\,(z_{0.025} < n^{1/2}(Y\_bar - \mu)/\sigma < z_{0.975})$,

Most of the table



How many of the previous samples contained the true mean?

CI for population mean

Confidence Interval for the true (population) mean μ:

sample mean $\pm$ ***t standard errors***

or $\bar{x} \pm t\,\mathrm{se}(\bar{x})$, where $\mathrm{SE}(\bar{x}) = \frac{s_x}{\sqrt{n}}$ and $df = n-1$

Value of the Multiplier, *t*, for a 95% CI

df :	7	8	9	10	11	12	13	14	15	16	17
t :	2.365	2.306	2.262	2.228	2.201	2.179	2.160	2.145	2.131	2.120	2.110
df :	18	19	20	25	30	35	40	45	50	60	∞
t :	2.101	2.093	2.086	2.060	2.042	2.030	2.021	2.014	2.009	2.000	1.960

Effect of increasing the confidence level

99% CI, $\bar{x} \pm 2.576\ \mathrm{se}(\bar{x})$

95% CI, $\bar{x} \pm 1.960\ \mathrm{se}(\bar{x})$

90% CI, $\bar{x} \pm 1.645\ \mathrm{se}(\bar{x})$

80% CI, $\bar{x} \pm 1.282\ \mathrm{se}(\bar{x})$

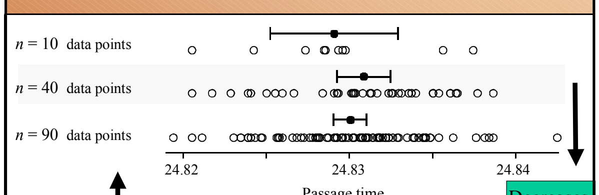
Confidence Level Increase

Increases the size of the CI

Why?

The greater the confidence level, the wider the interval

Effect of increasing the sample size



Increase Sample Size

Three random samples from a Normal(μ=24.83, s =.005) distribution and their 95% confidence intervals for μ.

Decreases the size of the CI

To ***double the precision*** we need ***four times*** as many observations.

Confidence intervals – non-symmetric case

- A marine biologist wishes to use male angelfish for an experiment and hopes their weights don't vary much. In fact, a previous random sample of n = 16 angelfish yielded the data below
- $\{y_1; \ldots ; y_n\}$ = { 5.1; 2.5; 2.8; 3.4; 6.3; 3.6; 3.9; 3.0; 2.7; 5.7; 3.5; 3.6; 5.3; 5.1; 3.5; 3.3}
- Sample statistics from these data include Avg. = 3.96 lbs, s^2 = 1.35 lbs, n = 16.
- **Problem**: Obtain a 100(1- α)% CI(σ^2).
- Point Estimator for σ^2? How about sample variance, s^2?
- Sampling theory for s^2? Not in general, but under Normal assumptions ...
- If a random sample $\{Y_1; \ldots; Y_n\}$ is taken from a normal population with mean μ and variance σ^2, then standardizing, we get a sum of squared N(0,1)

Confidence intervals – non-symmetric case

- $\{y_1; \ldots ; y_n\}$ = { 5.1; 2.5; 2.8; 3.4; 6.3; 3.6; 3.9; 3.0; 2.7; 5.7; 3.5; 3.6; 5.3; 5.1; 3.5; 3.3}
- **Problem**: Obtain a 100(1- α)% CI(σ^2).
- If a random sample $\{Y_1; \ldots; Y_n\}$ is taken from a normal population with mean μ and variance σ^2, then standardizing, we get a sum of squared N(0,1)

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{k=1}^{n}(Y_k - \bar{Y})^2}{\sigma^2} \sim \chi^2_{df=n-1}$$

For a=0.05, say. Need: 100(1- α)% CI(σ^2).

$$\Rightarrow 1-\alpha = P\left(\chi^2_{\left(n-1, 1-\frac{\alpha}{2}\right)} \le \frac{\sum_{k=1}^{n}(Y_k - \bar{Y})^2}{\sigma^2} \le \chi^2_{\left(n-1, \frac{\alpha}{2}\right)}\right)$$

Confidence intervals – non-symmetric case

- $\{y_1; \ldots ; y_n\}$ = { 5.1; 2.5; 2.8; 3.4; 6.3; 3.6; 3.9; 3.0; 2.7; 5.7; 3.5; 3.6; 5.3; 5.1; 3.5; 3.3}
- **Problem**: Obtain a 100(1- α)% CI(σ^2).

$$\frac{\sum_{k=1}^{n}(Y_k - \bar{Y})^2}{\chi^2_{\left(n-1, \frac{\alpha}{2}\right)}} \le \sigma^2 \le \frac{\sum_{k=1}^{n}(Y_k - \bar{Y})^2}{\chi^2_{\left(n-1, 1-\frac{\alpha}{2}\right)}}$$

- $\chi^2(15; 0.025)$=27:49 and $\chi^2(15; 0.975)$=6:26 →
- This yields the CI, the sample variance is **s^2=1.35**. Note the CI is NOT symmetric **(0.74 ; 3.24)**

Prediction vs. Confidence intervals

- **Confidence Intervals (for the population mean μ):**

$$\left(\bar{Y} - \frac{\hat{\sigma} \times t_{n-1, (1+L)/2}}{\sqrt{n}} \; ; \; \bar{Y} + \frac{\hat{\sigma} \times t_{n-1, (1+L)/2}}{\sqrt{n}}\right)$$

- **Prediction Intervals:** L-level prediction interval (**PI**) for a new value of the process Y is defined by:

$$\left(\hat{Y}_{new} - \hat{\sigma} \times t_{n-1, (1+L)/2} \; ; \; \hat{Y}_{new} + \hat{\sigma} \times t_{n-1, (1+L)/2}\right)$$

where the predicted value $\hat{Y}_{new} = \bar{Y}$, is obtained as an estimator of the unknown process mean μ.

Prediction vs. Confidence intervals – Differences?

- **Confidence Intervals (for the population mean μ):**

$$\left(\bar{Y} - \frac{\hat{\sigma} \times t_{n-1, (1+L)/2}}{\sqrt{n}} \; ; \; \bar{Y} + \frac{\hat{\sigma} \times t_{n-1, (1+L)/2}}{\sqrt{n}}\right)$$

$$\hat{\sigma} = \hat{\sigma}(\bar{Y}) = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}\left(y_k - \bar{y}\right)^2}$$

Which SD is bigger?!?

- **Prediction Intervals:**

$$\left(\hat{Y}_{new} - \hat{\sigma} \times t_{n-1, (1+L)/2} \; ; \; \hat{Y}_{new} + \hat{\sigma} \times t_{n-1, (1+L)/2}\right), \quad \text{where } \hat{Y}_{new} = \bar{Y}$$

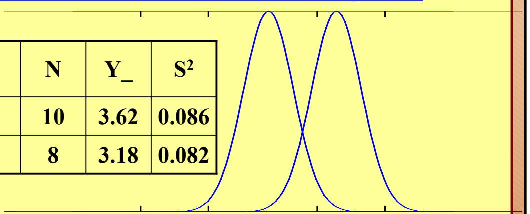
$$\hat{\sigma} = \hat{\sigma}(Y_{new} - \hat{Y}_{new}) = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}\left(y_k - \bar{y}\right)^2} \times \sqrt{1 + \frac{1}{n}}$$

Significance Testing –
Using Data to Test Hypotheses

Example – Carbon content in Steel

Percentage of ***C*** (Carbon) in 2 random samples taken from 2 steel shipments are measured and summarized below. The question is to **determine if there are statistically significant differences between the shipments**.

#	N	Y_	S²
1	10	3.62	0.086
2	8	3.18	0.082



Measuring the distance between the true-value and the estimate in terms of the SE's

- Intuitive criterion: Estimate is credible if it's not far-away from its hypothesized true-value!
- But how far is far-away?
- Compute the distance in **standard-terms**:

$$T = \frac{\text{Estimator} - \text{TrueParameterValue}}{\text{SE}}$$

- Reason is that the distribution of T is known in some cases (Student's t, or N(0,1)).
- The estimator (obs-value) is typical/atypical if it is close to the center/tail of the distribution.

Comparing CI's and significance tests

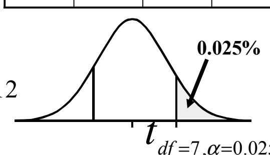
- These are different methods for coping with the uncertainty about the true value of a parameter caused by the sampling variation in estimates.
- **Confidence interval**: A fixed level of confidence is chosen. We determine ***a range of possible values*** for the parameter that are consistent with the data (at the chosen confidence level).
- **Significance test**: *Only one possible value* for the parameter, called the hypothesized value, is tested against the data. We determine the *strength of the evidence* (confidence) provided by the data against the proposition that the hypothesized value is the true value.

Review

- Are the **carbon** contents in the two steel shipments any different?

#	N	Y_	S²
1	10	3.62	0.086
2	8	3.18	0.082

$$t_0 = \frac{\text{Est\_1 - Est\_2 - 0}}{SE} = \frac{3.62 - 3.18}{SE(\hat{\mu}_1 - \hat{\mu}_2)} = \frac{0.44}{\sqrt{0.086/10 + 0.082/8}} = 3.12$$



Hypotheses

Guiding principles

We cannot **rule in** a hypothesized value for a parameter, we *can only* determine whether there is evidence, provided by the data, *to* **rule out** a hypothesized value.

The ***null hypothesis*** tested is typically a skeptical reaction to a ***research hypothesis***

The t-test

Using $\hat{\theta}$ to test H_0: $\theta = \theta_0$ versus some alternative H_1.

STEP 1 Calculate the ***test statistic***,

$$t_0 = \frac{\hat{\theta} - \theta_0}{se(\hat{\theta})} = \frac{\text{estimate - hypothesized value}}{\text{standard error}}$$

[This tells us how many standard errors the estimate is above the hypothesized value (t_0 positive) or below the hypothesized value (t_0 negative).]

STEP 2 Calculate the *P*-value using the following table.

STEP 3 Interpret the *P*-value in the context of the data.

The t-test

Alternative hypothesis	Evidence against H_0: $\theta > \theta_0$ provided by	P-value
$H_1: \theta > \theta_0$	$\hat{\theta}$ too much bigger than θ_0 (i.e., $\hat{\theta} - \theta_0$ too large)	$P = pr(T \geq t_0)$
$H_1: \theta < \theta_0$	$\hat{\theta}$ too much smaller than θ_0 (i.e., $\hat{\theta} - \theta_0$ too negative)	$P = pr(T \leq t_0)$
$H_1: \theta \neq \theta_0$	$\hat{\theta}$ too far from θ_0 (i.e., $\lvert\hat{\theta} - \theta_0\rvert$ too large)	$P = 2\,pr(T \geq \lvert t_0 \rvert)$

where $T \sim$ Student(df)

Interpretation of the p-value

Interpreting the Size of a *P*-Value

Approximate size of *P*-Value		Translation
> 0.12	(12%)	**No** evidence against H_0
0.10	(10%)	**Weak** evidence against H_0
0.05	(5%)	**Some** evidence against H_0
0.01	(1%)	**Strong** evidence against H_0
0.001	(0.1%)	**Very Strong** evidence against H_0

Is a second child gender influenced by the gender of the first child, in families with >1 kid?

1st Child	Second Child Gender: Male	Female	Total
Male	3,202	2,776	5,978
Female	2,620	2,792	5,412
Total	5,822	5,568	11,390

- Research hypothesis needs to be formulated first before collecting/looking/interpreting the data that will be used to address it. **Mothers whose 1st child is a girl are more likely to have a girl, as a second child, compared to mothers with boys as 1st children**.
- Data: 20 yrs of birth records of 1 Hospital in Auckland, NZ.

Analysis of the birth-gender data

- Samples are large enough to use Normal-approx. Since the two proportions come from totally diff. mothers they are independent → use formula 8.5.5.a

$$t_0 = \frac{\text{Estimate - HypothesizedValue}}{SE} = 5.49986 =$$

$$\frac{\hat{p}_1 - \hat{p}_2 - 0}{SE\left(\hat{p}_1 - \hat{p}_2\right)} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} =$$

$$P - value = \Pr(T \geq t_0) = 1.9 \times 10^{-8}$$

Analysis of the birth-gender data

- We have strong evidence to reject the H_0, and hence conclude mothers with first child a girl a more likely to have a girl as a second child.
- **Practical vs. Statistical significance**:
- How much more likely? A 95% CI:

CI (p_1- p_2) =[0.033; 0.070]. And computed by:

$$\text{estimate} \pm z \times \text{SE} = \hat{p}_1 - \hat{p}_2 \pm 1.96 \times SE\left(\hat{p}_1 - \hat{p}_2\right) =$$

$$\hat{p}_1 - \hat{p}_2 \pm 1.96 \times \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} =$$

$$0.0515 \pm 1.96 \times 0.0093677 = [3\%; 7\%]$$

Relation Among Various Continuous Distributions