Final Exam:
Thursday December 15th, 9am – 12pm
ESB 360 (here!)

40% of final grade: don't neglect to study!

Bring:
- **pen/pencil**
- **Calculator (not an internet enabled device, e.g., not an iPhone)**
- **1 sheet of notes (8.5x11", can be double-sided)**

Final Exam, likely content:

One problem on Quadrature

One problem on Nonlinear Regression

One problem on ODE Integration

One problem on Statistical testing

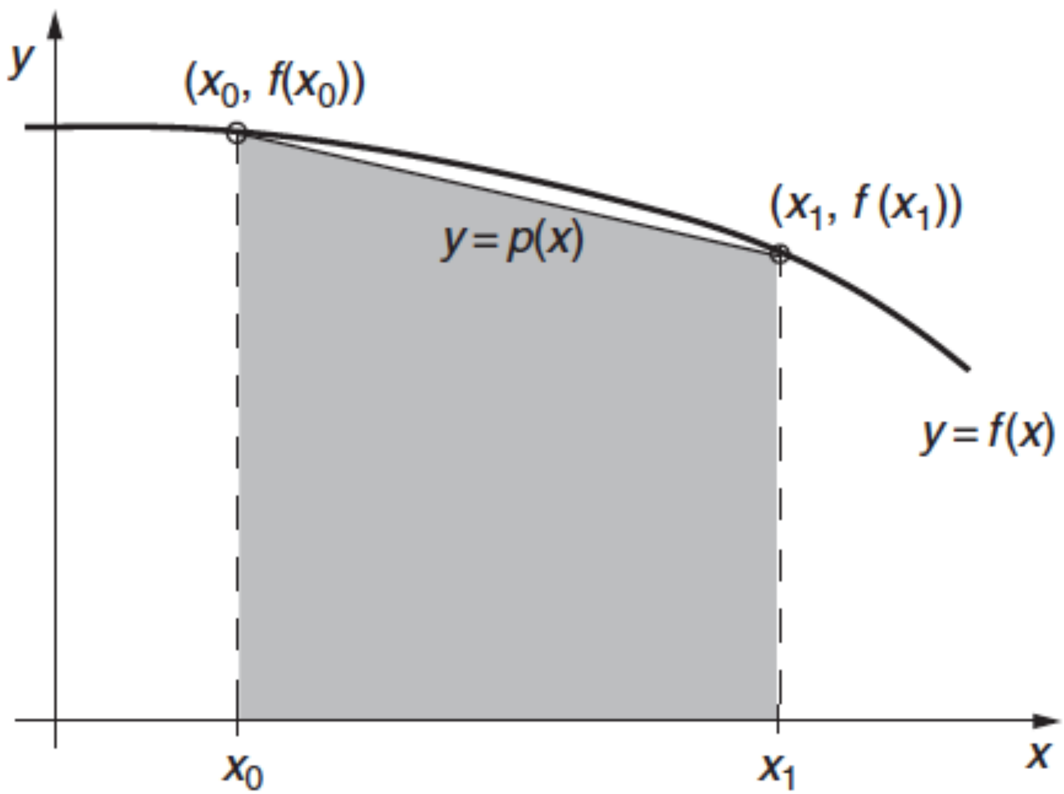
Short question involving Matlab

at least one problem involving plugging in numbers

Trapezoidal Rule

- What other rules are there? How about the trapezoidal rule?

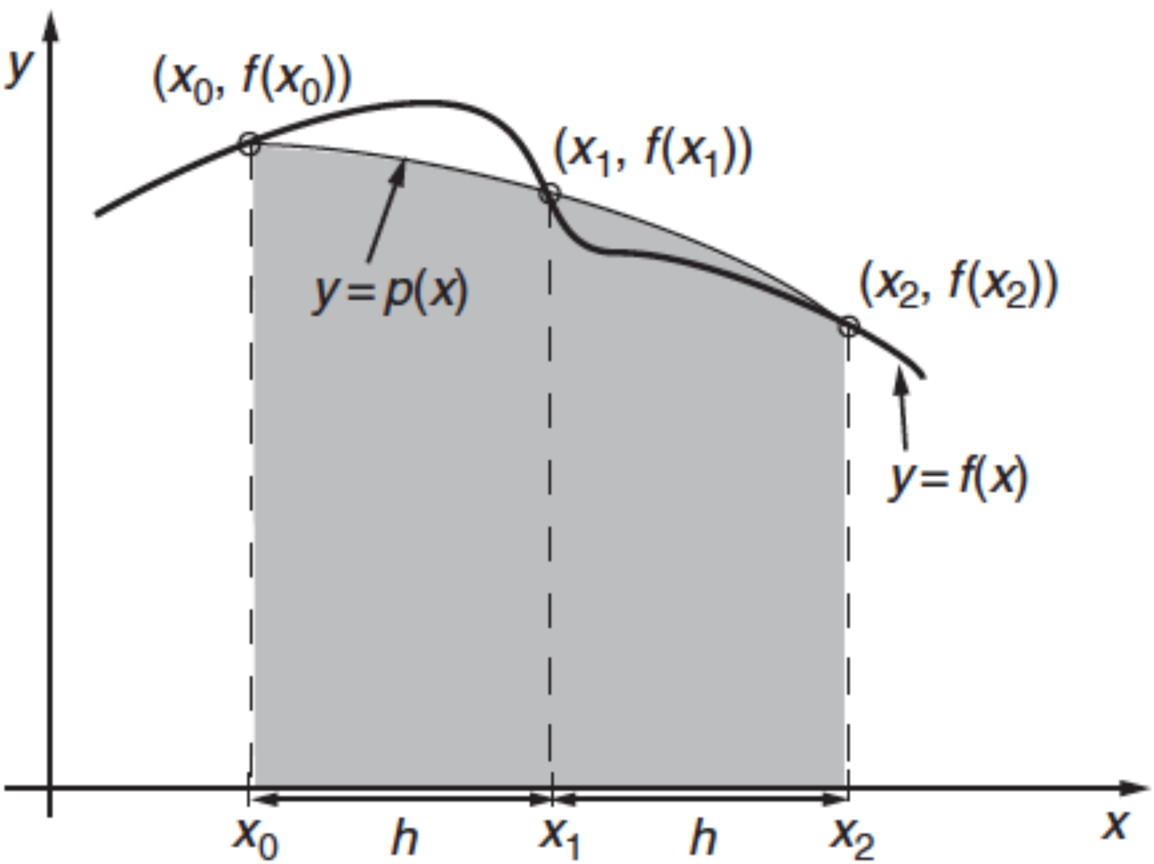
The trapezoidal rule for numerical integration. The shaded area is a trapezoid.



Simpson's Rule

- We can do better using Simpson's (1/3) Rule:

Graphical description of Simpson's 1/3 rule.



- We take a <u>pair</u> of panels with <u>three</u> function evaluations.
- We fit a parabola through these three points…

Gaussian Quadrature

- In Gaussian quadrature routines the weights and nodes are given based on the interval (-1,1). The nodes are thus <u>symmetric</u>:

For the four point rule:

x_i*	w_i*
+/- 0.8611363…	0.3478548…
+/- 0.3399810…	0.6521452…

Richardson Extrapolation

$$I = I_n + \lambda/n^2 + O(1/n^4)$$

next order in error term

- Note that odd powers in n are killed off due to symmetry.
- Suppose we took the n=2 and n=4 results:

$$I = I_2 + \lambda/4 + O(1/2^4)$$

$$I = I_4 + \lambda/16 + O(1/4^4)$$

- If we multiply the second equation by 4 and subtract from the first we can eliminate λ…

Ordinary Differential Equations

- We can also write a single n^{th} order ODE as a system of first order ODEs.
- Consider the damped pendulum:

$$2\pi ML(d^2\theta/dt^2) = -Mg \sin \theta - 6\pi\mu a(2\pi L)d\theta/dt$$

- The last term is the resistance due to viscosity of a sphere of radius a moving through a viscous liquid with velocity $(2\pi L)d\theta/dt$.
- We can rewrite this as a pair of first order equations:

$$y_1 = \theta, \quad y_2 = d\theta/dt$$

ODE Stability

- What if we had the equation:

$$dy/dt = -y$$

- In this case all of the family of solutions converge thus the error tends to be wiped out. Such equations are stable.
- The stability is determined by the Jacobian:

$$J = df/dy$$

- Note that the derivative is with respect to the dependent variable!

ODE Stability

If $J < 0$ the equation is <u>stable</u>

If $J > 0$ it's unstable

If $J = 0$ it's neutrally stable

- We may have an equation which is stable in places and unstable in others. For example:

$dy/dt = (1-t)\, y$ $\qquad\qquad$ $y(0)=1$

$J>0$ for $t<1$ $\quad$ unstable

$J<0$ for $t>1$ $\quad$ stable

ODE Stability

- We have the same for systems of equations. In this case the Jacobian is a <u>matrix</u>.

$$\mathbf{J} = J_{ij} = df_i/dy_j$$

- The stability of a system of equations is related to the <u>eigenvalues</u> of **J**.

ODE Integration

$[f(t_k,y_k) - f(t_k,y(t_k))]$

Jacobian

$= [f(t_k,y(t_k)) + (y_k - y(t_k))df/dy - f(t_k,y(t_k))]$

Expansion for f at y_k

$= (y_k - y(t_k))df/dy$

- Thus, the error at step k+1 is given by:

$$y^{EM}{}_{k+1} - y(t_{k+1}) = (y_k{}^{EM} - y(t_k))(1+h_kJ) - ½\, h_k{}^2y''(\xi)$$

- The term $½\, h_k{}^2y''(\xi)$ is the <u>local error</u> at each step.

ODE Integration

- The quantity $(1+hJ)$ is the amplification factor. Note that for all unstable ODEs this factor is greater than 1.
- The key result is that if J is very negative, then Euler's method may be unstable as well!
- If $hJ<-2$ then $|1+hJ|>1$ and our method is numerically unstable.
- Equations for which $J<<-1$ are termed stiff.

ODE Integration

- The interval of stability is given by:

$$-2 < hJ < 0$$

- If you have a stiff equation, your step size had better not be too large!
- For systems of equations, the Euler method will be stable if:

$$|1+h\lambda| < 1$$

for all eigenvalues. This is a bit more complex since λ may have an imaginary part. Thus the interval of stability is now a region. The method will be stable if all $h\lambda$ lie within a circle of radius 1 centered about $z=-1$ in the complex plane.

ODE Integration

- How do we deal with stiff problems?
- We use <u>implicit</u> methods. The simplest is the Backward Euler Method.

$$y^{BE}_{k+1} = y_k + h_k f(t_{k+1}, y_{k+1})$$

(note implicit dependence)

- A method is called <u>implicit</u> if the equation for y_{k+1} depends on a function of y_{k+1}.
- Okay what does error propagation look like in this case? We determine it the same way…

Runge-Kutta Integration

- Usually codes don't use 2-stage R-K rules, but rather 4-stage rules:

$$K_1 = h\, f(t_n, y_n)$$
$$K_2 = h\, f(t_n+0.5h, y_n+0.5K_1)$$
$$K_3 = h\, f(t_n+0.5h, y_n+0.5K_2)$$
$$K_4 = h\, f(t_n+h, y_n+K_3)$$
$$y_{n+1} = y_n + (1/6)(K_1+2K_2+2K_3+K_4)$$

- The local error is $O(h^5)$ and the overall rule is fourth order in the step size!

Higher order implicit rules

- Thus: $y^{TR}{}_{k+1} = (y^{EM}{}_{k+1} + y^{BE}{}_{k+1})½$

  $y_{k+1} = y_k + ½h(f(t_k,y_k) + f(t_{k+1},y_{k+1}))$

  which is <u>implicit</u>.
- The error is a bit difficult to calculate, but eventually you get:

$$y^{TR}{}_{k+1} – y(t_{k+1}) = (1+ ½hJ)/(1 – ½hJ)*(y_k – y(t_k)) + O(h^3)$$

Adaptive Step Size Control

$$(h_{opt}/h)^3 \sim (\tau/\Delta)$$

- Thus, $h_{opt} \sim L(\tau/\Delta)^{1/3}$
- In practice, we pick our new h to be a little smaller than this value, say:

$$h_{opt} = 0.9h(\tau/\Delta)^{1/3}$$

- We must also set an upper and lower limit to h. The upper limit will just be some fraction of the total interval of integration.

Hypothesis Testing (Ch. 4)

- Two Sample z-Test: We wish to test the null hypothesis that the means of two populations, estimated from two independent samples, are equal, when the samples are large.
- Sample 1: number of subjects n_1, mean $\overline{x}_1$, standard deviation s_1.
- Sample 2: number of subjects n_2, mean $\overline{x}_2$, standard deviation s_2.

Two Sample z-Test

- Assumptions:
  1. data are normally distributed
  2. data are independent
  3. samples are large (> ~30 for n_1, n_2)
- Calculate the <u>standard error of the difference between the means</u>:

$$SE(\overline{x}_1 - \overline{x}_2) = (s_1^2/n_1 + s_2^2/n_2)^{½}$$

$$z = (\overline{x}_1 - \overline{x}_2)/SE(\overline{x}_1 - \overline{x}_2)$$

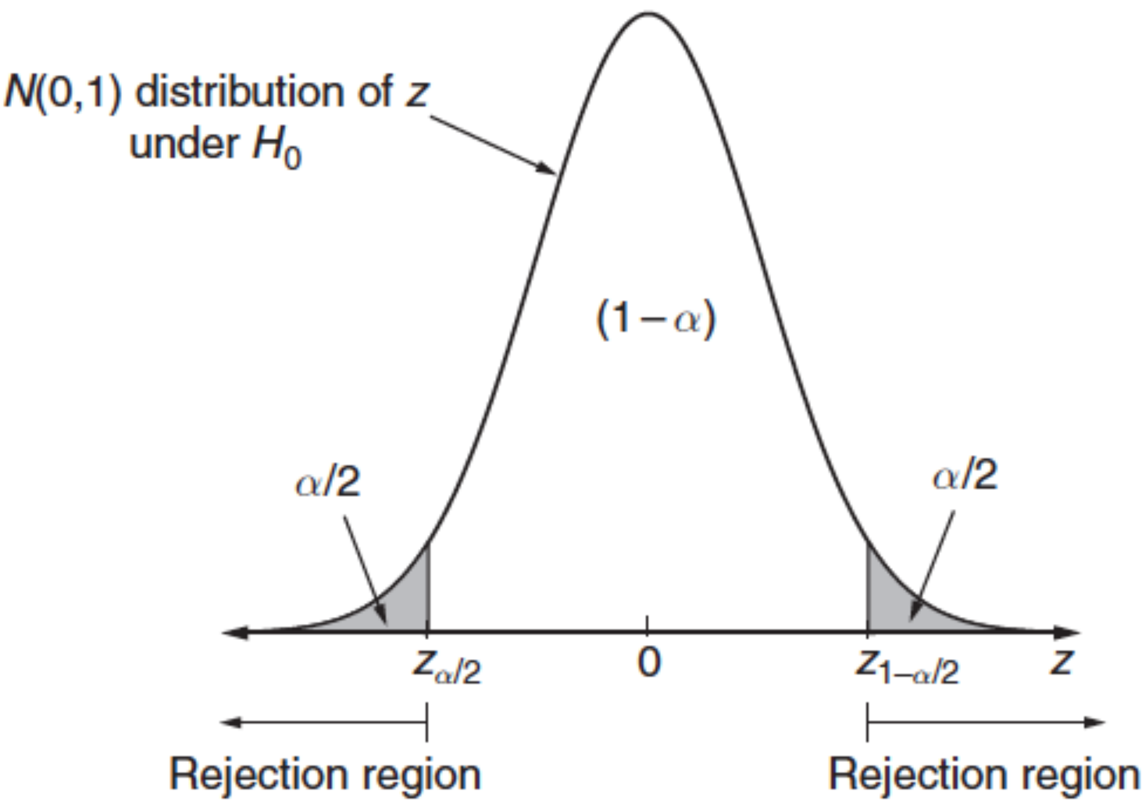
Two Sample z-Test

- Under the null hypothesis, z is distributed approximately as a normal distribution with mean = 0 and standard deviation = 1.
- A 95% confidence interval for the difference is:

$$(\overline{x}_1 - \overline{x}_2) \pm 1.96 \; SE(\overline{x}_1 - \overline{x}_2)$$

Two-sided test

Two-sided z test; z follows the $N(0,1)$ distribution curve when H_0 is true. The shaded areas are the rejection regions.



Two Sample z Test of Proportions

- Provided the sample sizes in the two groups are large, same method can be used on two proportions.
- Population mean $\mu_i \leftarrow\rightarrow$ sample mean x_i
- Population proportions of success: π_A, π_B

  ~ sample statistics: p_A, p_B
- Standard error for difference $p_1 - p_2$ given by:

$$SE(p_1 - p_2) = (p_1(1-p_1)/m + p_2(1-p_2)/n)^{1/2}$$

(Note difference from population means)

The p-value

- Suppose that z_α is now chosen such that the lower limit equals zero. That is, the confidence interval just includes the null hypothesis value of $\delta = \pi_A - \pi_B = 0$.

$$0.1861 - (z_\alpha \times 0.1175) = 0$$

$$z_\alpha = 0.1861/0.1175 = 1.58$$

- From the Normal distribution table, α=0.11 and $100(1 - \alpha)\% = 89\%$.

Statistical Inference

- Hypothesis Testing: method of deciding whether the data are consistent with the null hypothesis.
- Given a study with a single outcome measure and a statistical test, hypothesis testing can be summarized in three steps:
  1. Choose a significance level, α, of the test.
  2. Conduct the study, observe the outcome, and compute the p-value.
  3. …

Statistical Inference

3. If the p-value <= α $\rightarrow$ data are not consistent with the null hypothesis.

If p-value > α, do not reject the null hypothesis, and view it as "not yet disproven".

- Do not confuse the significance level and the p-value!
- If one rejects the null hypothesis when it is in fact true, then one makes a <u>Type I error</u>.
- The significance level α is the probability of making a Type I error. This is set <u>before</u> the test is carried out. The p-value is the result observed <u>after</u> the study is completed.

Small samples of continuous data

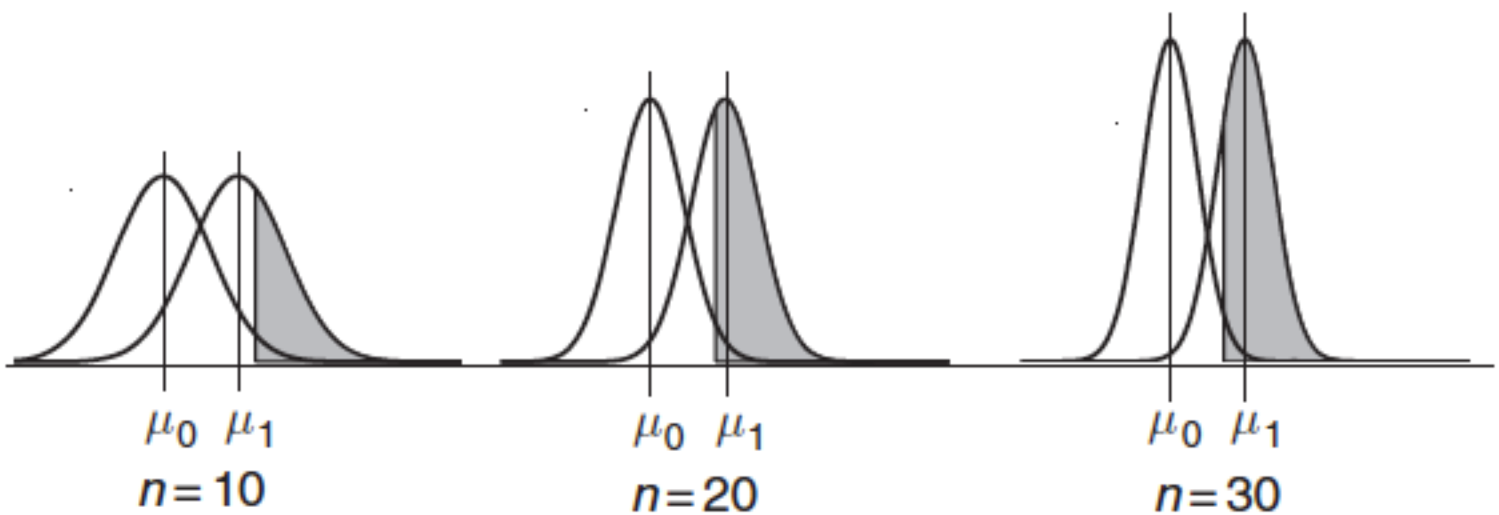
- Student's t-distribution: If samples are small, $\overline{x}$ and s will not be close to μ and σ.
- (In one way sample size is already taken into account through the calculation of the standard deviation of the mean, SE($\overline{x}$) when dividing by sqrt(n).)
- In small samples, however, values of s far from σ are not uncommon.

Two-Sample or Unpaired t-test

- The unpaired test (aka independent sample or unrelated test) arises when the data in two groups are not connected, as in:
  - Parallel group clinical trial: one group receives the test treatment, and one group receives the control treatment (randomized)
  - Unmatched case-control study: the control group not designed to “look like” the diseased population (e.g., med student control)

The **power** of the hypothesis test is given by $(1 - \beta)$. It is the probability of rejecting a false null hypothesis.

When we do not reject a false null hypothesis, we make a **type II error**. If the null hypothesis is false, the probability of making a type II error is β. In other words, β is the probability of obtaining a statistically insignificant result, even when the null hypothesis is incorrect.



The χ^2 (chi-squared) Test

- 2x2 Contingency Tables: lets modify the test for a comparison of proportions to cover the situations of small samples.
- Recall the peptic ulcer study of Familiari et al. (1981).

Drug	Healed	Not Healed	Total	% healed
A: Pirenzipine	23 (a)	7 (c)	30 (m)	76.67
B: Trithiozine	18 (b)	13 (d)	31 (n)	58.06
Total	41 (r)	20 (s)	61 (N)	

The χ^2 (chi-squared) Test

- $\chi_c^2 = N(|ad - bc|) - ½N)^2/(mnrs)$
- The subtraction of ½ N causes χ_c^2 to be smaller than χ^2
- ½N is called the Yates correction for continuity, and arises because we are dealing with discrete data but the χ^2 distribution used to calculate the p-value is continuous (like the Normal distribution).
- For the peptic ulcer example, $\chi_c^2 = 1.62$.

Chi-squared Test (χ^2) in 2x2 Tables

- $E_{ij} = R_i \times C_j/N$ where R_i is row total

  C_j is column total
- For example, $E_{11} = (a+c)(a+b)/N = mr/N$
- Then, $X^2 = \Sigma(O - E)^2/E$ O: observed

  E: expected
- With Yates correction,

$$X^2 = \Sigma(|O - E| - ½)^2/E$$

Analysis of Variance (ANOVA)

- In general,

		Group 1	Group 2	…	Group k
Population	mean	μ_1	μ_2		μ_k
	std. dev.	σ_1	σ_2		σ_k
Sample	mean	x_1	x_2		x_k
	std. dev.	s_1	s_2		s_k
	sample size	n_1	n_2		n_k

- Assume k populations are independent and normally distributed.

Analysis of Variance (ANOVA)

- Pooled estimate of common variance is:

$$s_w^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + (n_3-1)s_3^2}{n_1 + n_2 + n_3 - 3}$$

- If k=10, t-tests would become complicated, with $\binom{10}{2}$ = 45 different pairwise tests.

- More importantly, many two-sample t-tests are likely to lead to an incorrect conclusion.

Analysis of Variance (ANOVA)

- Null hypothesis: H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$
  for a set of k populations.
- Variability of individual observations around their population means.
- Let $n = n_1 + n_2 + \ldots + n_k$
- $$s_w^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + \ldots + (n_k-1)s_k^2}{n-k}$$
- Weighted average of k individual sample variances. **W: "within-groups" variability**

Analysis of Variance (ANOVA)

- Extent that population means vary around the overall mean.
- $s_B^{\,2} = \dfrac{n_1(\overline{x}_1 - \overline{x})^2 + n_2(\overline{x}_2 - \overline{x})^2 + \ldots + n_k(\overline{x}_k - \overline{x})^2}{k - 1}$
- $(\overline{x}_i - \overline{x})^2$: squared deviation of sample means $\overline{x}_i$ from the grand mean $\overline{x}$.
- <u>Grand mean</u>: overall average of n observations that make up the k different samples.

Analysis of Variance (ANOVA)

- Do sample means vary around the grand mean more than individual observations vary around the sample means?
- Yes? → corresponding population means

  are in fact different.
- Test statistic: $F = s_B^2/s_w^2$
- Under the null hypothesis both s_w^2 and s_B^2 estimate a common variance σ^2 → $F \sim 1$

Multiple Comparisons via ANOVA

- H_0: $\mu_i = \mu_j$
- Calculate $t_{ij} = \dfrac{\overline{x}_i - \overline{x}_j}{\mathrm{sqrt}(s_w^2(1/n_i + 1/n_j))}$
- Note that we use all the additional info from the k samples to estimate the common variance σ^2.
- Under H_0, t_{ij} has a t-distribution with n – k df.
- One drawback of Bonferroni correction: highly conservative, lacks statistical power.

Non-normal distributions and data transformations

- Parametric tests: (e.g., t-test and ANOVA) assumes various parameters and normal distribution.
- Nonparametric tests: no assumption of population distribution is made.
- Testing if your data meets normality assumptions…

The Sign Test

	REE (kcal/day)			
Pair	CF	Healthy	Difference	Sign
1	1153	996	157	+
2	1132	1080	52	+
3	1165	1182	-17	-
4	1460	1452	8	+
5	1634	1162	472	+
6	1493	1619	-126	-
7	1358	1140	218	+
8	1453	1123	330	+
9	1185	1113	72	+
10	1824	1463	361	+
11	1793	1632	161	+
12	1930	1614	316	+
13	2075	1836	239	+

Wilcoxon Signed-Rank Test

Subject	Placebo	Drug	Difference	Rank	Signed rank		
1	224	213	11	1	1		
2	80	95	-15	2		-2	
3	75	33	42	3	3		
4	541	440	101	4	4		
5	74	-32	106	5	5		
6	85	-28	113	6	6		
7	293	445	-152	7		-7	
8	-23	-178	155	8	8		
9	525	367	158	9	9		
10	-38	140	-178	10		-10	
11	508	323	185	11	11		
12	255	10	245	12	12		
13	525	65	460	13	13		
14	1023	343	680	14	14		
					86	-19	totals

Multidimensional Optimization

- Life get <u>much</u> more complex for higher (N>1) dimensional optimization.
- In general, we start from a given point, pick a <u>search direction</u>, and do a 1-D search in that direction.
- <u>Method of steepest descent</u>: If we want to get to a minimum, it makes sense to go downhill.

  Search direction = – grad(F)

Multidimensional Newton's Method

- The quadratic approximation is differentiated with respect to x and set equal to zero:

$$H(x_0)\Delta x = -\nabla f(x_0),$$

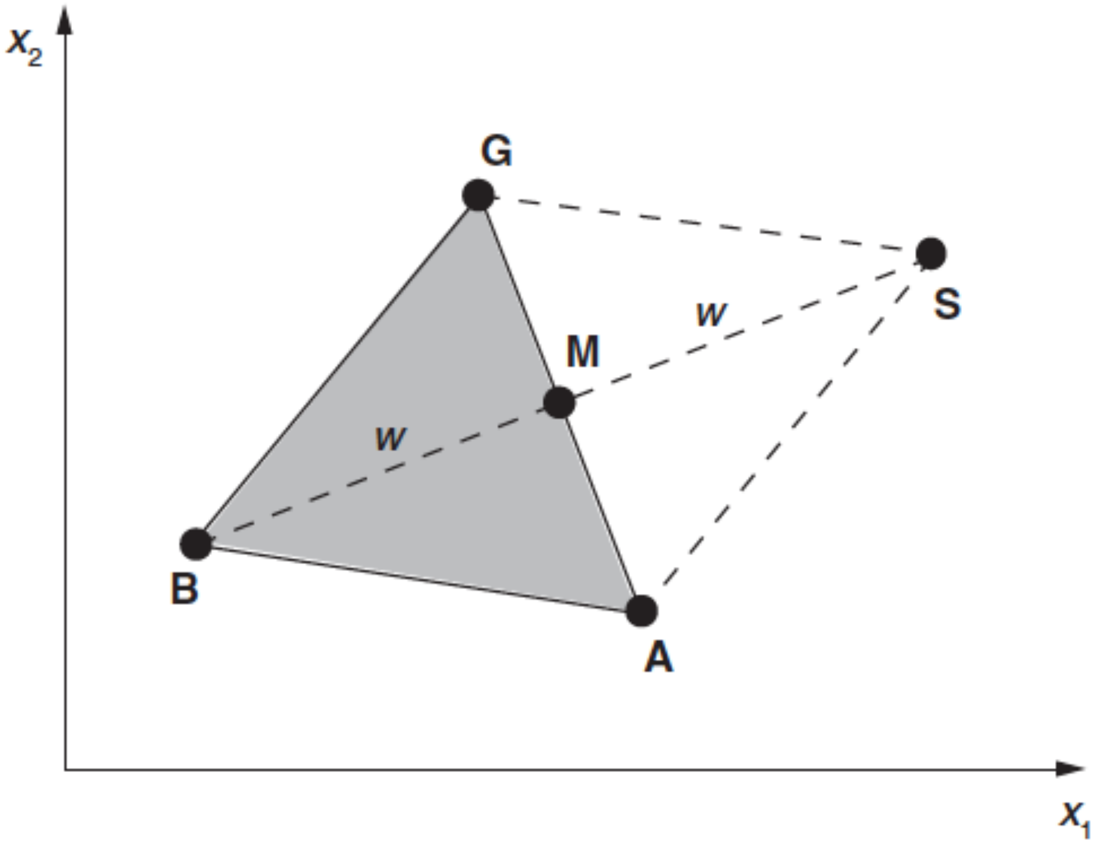
where $\Delta x = x - x_0$.

- If we truncate after this term, we get an equation for the next guess at the critical point!

$$x^{(k+1)} = x^{(k)} - \left[H\left(x^{(k)}\right)\right]^{-1} \nabla f\left(x^{(k)}\right).$$

- These are Newton's equations for this problem!

Construction of a new triangle **GSA** (simplex). The original triangle is shaded in gray. The line **BM** is extended beyond **M** by a distance equal to *w*



- We pick 3 points in the form of an equilateral triangle.
- We discard the largest point (when looking for a minimum) and pick a new point which is its <u>mirror image</u>.

- We then wander through space until the minimum is reached!

Lagrange Multipliers

- Here we have the augmented problem:

$$\nabla^* F^* = 0,$$

where

$$\nabla^* = \begin{pmatrix} \nabla_x \\ \nabla_\lambda \end{pmatrix}.$$

- This is because

$$\mathbf{grad}_\lambda F^* = \mathbf{g}(\mathbf{x}) = \mathbf{0}$$

or just the equality constraints!

Inequality Constraints

- Now for inequality constraints.
- We have: $\mathbf{g}(\mathbf{x}) <= \mathbf{0}$
- We can convert inequality constraints to equality constraints by adding slack variables.
- Let $\mathbf{g} + \mathbf{s} = 0$

  where $s_i = x^2_{n+1} >= 0$

  (convenient choice ensuring that $s_i >= 0$)

We then treat the problem using Lagrange multipliers again!

Penalty Functions

- Finally, we look at penalty functions.
- We wish to have an unconstrained optimization problem. We can do this even with constraints in an artificial manner.
- Suppose we have:

$$\min_{\mathbf{x}} F(\mathbf{x}), \quad \mathbf{g}(\mathbf{x}) = \mathbf{0}$$

- We may define $F^*(\mathbf{x}) = F(\mathbf{x}) + P||\mathbf{g}(\mathbf{x})||^2$

  where P is a positive number.