HRP 259: September 25th, 2006

Introduction/Math Review

I. Mathematical shorthand

$\sum$ = Summation

$\prod$ = Product

! = Factorial

Examples:

Take a set of 5 observations: X_1 to X_5 = { 5, 6, 7, 9, 10 }
(the observations are indexed by subscripts i):

$$\sum_{i=1}^{5} X_i = X_1 + X_2 + X_3 + X_4 + X_5 = 5 + 6 + 7 + 9 + 10 = 37$$

$$\sum_{i=0}^{10} i = 0 + 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10 = 55$$

$$\sum_{i=1}^{n} i = 1 + 2 + 3 + \ldots + + n-3 + n-2 + n-1 + n = n(n+1)/2$$

$$\sum_{i=1}^{4} 10i = 10(1) + 10(2) + 10(3) + 10(4) = 10\,[1+2+3+4] = 10\sum_{i=1}^{4} i$$

$$\sum_{i=1}^{n} ci = c\sum_{i=1}^{n} i$$

$$\prod_{i=1}^{5} X_i = 5(6)(7)(9)(10) = 18{,}900$$

$$\prod_{i=1}^{5} i = 1(2)(3)(4)(5) = 120 = 5!$$

$$\prod_{i=1}^{n} i = n!$$

5! = 5(4)(3)(2)(1)
n!=n(n-1)(n-2)(n-3)…(3)(2)(1)
<u>note:</u> 0! = 1 by convention

In-Class Exercises:

1. $\sum_{i=1}^{5} i \sum_{j=0}^{i-1} j^2 =$

2. $\prod_{i=1}^{5} i \sum_{j=1}^{i} j =$

3. $\sum_{j=1}^{5} j! =$

4. $\sum_{k=0}^{\infty} \frac{1}{k!} =$

<u>Answers:</u>

1. $\sum_{i=1}^{5} i \sum_{j=0}^{i-1} j^2 =$

1 (0^2) + 2 $(0^2 + 1^2)$ + 3 $(0^2 + 1^2 + 2^2)$ + 4 $(0^2 + 1^2 + 2^2 + 3^2)$ + 5 $(0^2 + 1^2 + 2^2 + 3^2 + 4^2)$ = 223

2. $\prod_{i=1}^{5} i \sum_{j=1}^{i} j =$

1 (1) * 2 (1 + 2) * 3 (1 + 2 + 3) * 4 (1 + 2 + 3 + 4) * 5 (1 + 2 + 3 + 4 + 5) = 324,000

3. $\sum_{j=1}^{5} j! =$

1 + 2(1) + (3)(2)(1) + (4)(3)(2)(1) + (5)(4)(3)(2)(1) = 153

4. $\sum_{k=0}^{\infty} \frac{1}{k!} =$

1 + 1 + ½ + 1/6 + 1/24 + 1/120 + 1/720……=2.718…= e

II. Functions and Calculus

f(x) or y = function: *x* maps to *f(x)* *(or y)*

e^x = the exponential function (*e* is just a constant=2.71828…)

ln(x) = the natural log function (inverse of exponential)

$\frac{df(x)}{dx}$ *or* $\frac{dy}{dx}$ *or y' or f'(x)* = the derivative of y with respect to x

$\int f(x)dx$ = integral of *f(x)* over *x*

Examples:
Some continuous functions: *f(x) = 2x* *[linear]*

$f(x) = \frac{x^3}{3} - \frac{x^2}{2} - 6x - 4$ *[polynomial]*

$y = x^2$ *[quadratic]*

$y = e^x$ *[exponential]*

A piecewise function:

$$y = \begin{cases} 1 & \text{if } x \le 0 \\ 0 & \text{if } x > 0 \end{cases}$$

A constant function: *f(x)* = 5

Manipulation of logs:

$e^{ln(x)} = x$

$\ln(e^x) = x$

ln 0 = undefined

$e^0 = 1$

1. "The log of a product is the sum of logs" → used often to LINEARIZE (turn a product to a sum); herein lies the real power of logs! It's much easier to deal with sums…

ln(a*b*c) = lna + lnb + ln c

$$\ln\left(\prod_{i=1}^{5} i\right) = \sum_{i=1}^{5} \ln i$$

2. "The log of a power is the power times the log": ✓ $\ln x^c = \sum_{i=1}^{c} \ln x = c \ln x$

3. "The log of a quotient is the difference of logs"
ln(a/b) = lna - lnb

In-Class Exercise:

Solve:. $\ln\left(\prod_{i=1}^{n} i^2\right) =$

Answer:

$$\ln(\prod_{i=i}^{n} i^2) = \ln(1^2 * 2^2 * 3^2 ...) = \ln 1^2 + \ln 2^2 + ... = \sum_{i=1}^{n} 2\ln i = 2\sum_{i=1}^{n} \ln i$$

Derivative = Slope!

f'(x) (or y')= The slope of the tangent line at the point *x*. How fast is *y* changing as *x* changes?:

$f(x) = 2x$: $\frac{d(2x)}{dx} = 2$ – i.e., the slope of the tangent line is 2 at every point since the function is a straight line.

$y = x^2$: $\frac{dy}{dx} = 2x$ -- i.e., the slope of the tangent line changes depending on where you are on the function.

$f(x) = 5$: $f'(x) = 0$; -- i.e. the tangent line is a horizontal line at every point since the function is a horizontal line.

$f(x) = \frac{x^3}{3} - \frac{x^2}{2} - 6x - 4$: $f'(x) = x^2 - x - 6 = (x-3)(x+2)$

$f(x) = ln(x)$: $\frac{d(\ln x)}{dx} = \frac{1}{x}$

$y = e^x$: $y' = e^x$

Recall How to Maximize a Function (max and mins occur where slope of the tangent line is 0):

--you'll need this later in the course for "Maximum Likelihood Estimation"

Procedure: (1) take derivative of f(x); (2) set f'(x) = 0; (3) solve for x ; examples:

$f'(x) = 2$; $2 = 0$; 2 can never equal 0, so *no* max or min
$y' = 2x$; $2x = 0$; minimum is at x=0
$y' = e^x$; $e^x = 0$; x = ln 0 is undefined, ∴ *no* max or min
$y' = (x-3)(x+2)$; $(x-3)(x+2) = 0$; *max is at x=3*

Integral = Area!

Integration gives the area underneath the curve between two x values. This is very important in probability.
A probability function is any function that integrates to 1.
Integration is akin to summation…

$\int_0^{10} 2x\ dx = x^2 \Big|_0^{10} = 100 - 0 = 100$ note: 100 = the area under the triangle with vertices (0,0),(10,0), and (0,20)

$\int_1^{3} 2\ dx = 2x \Big|_1^{3} = 6 - 2 = 4$ note: 4 = the area under the line y=2 from x=1 to x=3

$\int_1^{5} \frac{1}{x}\ dx = \ln(x) \Big|_1^{5} = 1.6 - 0 = 1.6$

$\int_0^{+\infty} e^{-x} = -e^{-x} \Big|_0^{+\infty} = 0 + 1 = 1$; ∴probability function

In-Class Exercises:

1. Maximize the following function
(**HINT**: First take the natural log of the function to linearize it; then maximize. If you maximize the natural log of the function, you are also maximizing the function.):

$f(x) = \frac{1}{k} e^{-\frac{1}{2}(x-2)^2}$; where k is some constant

2. Solve: $\int_0^3 \frac{x}{2} \; dx =$

Answers:

1. Maximize the following function

$f(x) = \frac{1}{k} e^{-\frac{1}{2}(x-2)^2}$; where k is some constant

1. Take natural log:

$ln(f(x)) = \ln[\frac{1}{k} e^{-\frac{1}{2}(x-2)^2}] = \ln 1 - \ln k + \ln e^{-\frac{1}{2}(x-2)^2} = \ln 1 - \ln k - \frac{1}{2}(x-2)^2$

2. Take the derivative:

$$\frac{d(\ln(f(x))}{dx} = 0 - 0 - 1(x-2)$$

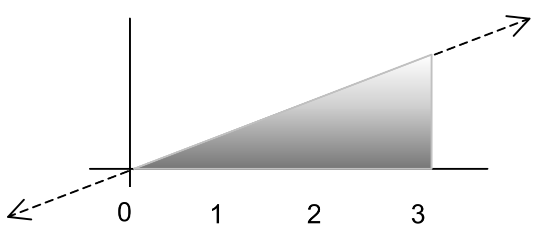
3. Set the derivative equal to 0:

$-(x-2) = 0$

4. Solve for x:

$x=2$

2. Solve: $\int_0^3 \frac{x}{2} \; dx =$



III. Some Statistical Symbols (keep this as a handy reference):

X = often used to indicate the independent (predictor) variable
Y = often used to indicate the dependent (or outcome) variable
X_i = the i^{th} observation of X
X_{ij} = in a table, the observation in row *i* and column *j*
μ_y = the true mean of Y
$\overline{Y}$ = the sample mean of Y
σ_x^2 = the true variance of random variable X / σ_x = the true standard deviation
s_x^2 = the sample variance of X / s_x = the sample standard deviation
E (X) = the expected value (or expectation) of random variable X
Var (Y) = another way to write the variance of Y
~ = "is distributed as", for example:
X ~bin (n, p) = "X *is distributed as* binomial with parameters n and p"
Y ~N(0,1) = "Y *is distributed as* normal with parameters 0 and 1"
^ = when placed over a term, means "estimated value", for example:
$\hat{\pi}$ = estimated proportion
π = true proportion
p = p-value
H_0 = null hypothesis
H_a or H_1 = alternative hypothesis
P( ) = "the probability that…", for example:
P(X=1) = "*the probability that* X=1"
P(A/B) = conditional probability; "the probability of A *given that* B is true"
${}_nC_r$ or $\binom{n}{r}$ = "n *choose* r" = the # of different combos of *r* objects that you can form out of *n* distinct objects
i.i.d. = "independently identically distributed" = two independent random variables that come from the same underlying distribution

Homework H1: due Wednesday September 27 at 4:15pm

Problems:

1. Using the set of n=5 observations: $x_{1\text{-}5}$ = {15 22 37 19 22}, calculate the following:

   a. $$\frac{\sum_{i=1}^{n}(x_i - \frac{\sum_{j=1}^{n} x_j}{n})^2}{n-1} =$$

   b. $$\prod_{i=1}^{n} \frac{x_i}{\sum_{j=1}^{n} x_j} \times \frac{(\sum_{j=1}^{n} x_j)!}{(\sum_{j=1}^{n} x_j - x_i)! x_i!} =$$

   c. $$\sum_{i=1}^{5} \sum_{j=1}^{i} ij =$$

2. Simplify:

   a. $\ln(\frac{5x^3}{z}) =$

   b. $e^{\ln x^2} =$

3. Maximize the function: $y = \ln[\frac{10!}{7! \times 3!} x^7 (1-x)^3]$

Bonus problem:

Prove that:

$$\sum_{i=1}^{n} i = n(n+1)/2$$