**Steel Grain Size Yield Strength Simulation**

This set of notes was developed as a result of an in-class example discussed on 2/14/2012. In that example the following random variables were addressed:

*X = The act of measuring the overall grain size in a specimen of steel*

*Y = The act of measuring the yield strength for that specimen.*

The following assumptions were made:

(A1) 

(A2) 

**Remark 1.** It was remarked by a student in material science that, in fact,  is closer to zero for smaller values of *x.*

In order to simulate measurements of for sample size *n=*30, using *Matlab*, it was necessary to specify

  and .

Since , we obtained . The *Matlab* command

xy = mvnrnd(, 30) was used to generate a single simulation for sample size 30.

Assumption (A2) reflects the assumption that the following linear prediction model is appropriate:

 . (1)

The model parameters *m* and *b* need to be estimated from the data. To this end, we demanded that the following two conditions should hold:

(C1) . In words, the estimator (1) is required to be an unbiased estimator of *Y.*

(C1) Let denote the prediction error. Then we require .

From these two conditions we have the following two equations in the two unknowns, *m* and *b*:

  (2a)

  (2b)

Hence, the solutions for the unknown parameters *m* and *b* are:  and .

It follows that the natural estimators of these parameters are:

  and  (3)

where the hat above a parameter denotes the associated moment estimator of it. For example, since the parameter, then . While the expressions in (3) may not appear familiar to some readers, they are, in fact, the expressions for the *Least Squares* (*LS*) estimators of the associated parameters. This fact is addressed in the Chapter 5 notes.

**The Intent of This Set of Notes**

The intent of this set of notes is to demonstrate how simulations can be used to accommodate Remark 1 above. Questions that will be answered include:

(Q1): How can one develop more realistic simulations that reflect Remark 1?

(Q2): In the case of those more realistic simulations, what happens when one uses the linear model (1), without the knowledge of the simulation structure that generated the data?

In answering these two questions, two very important concepts will be demonstrated:

*Concept* 1: Simulation of jointly normal random variables, (*X,Y*), when the correlation coefficient is not a constant; rather it is a function of *x.*

*Concept* 2: The power of using simulations to investigate the statistics of parameter estimators when mathematical methods become difficult, if not intractable.

***A Model for rXY*(*x*) :** In view of Remark 1, we desire a model for  that satisfies the condition that as the grain size, *x*, gets smaller, so does the magnitude of . The model that we will use for this purpose is:

 . (4)

This model is plotted in Figure 1.



**Figure 1.** Plot of  given by (4) for  and .

Results for a single simulation with sample size *n =* 30 is given in Figure 2.



**Figure 2(a)** Scatter plot of *x-y*.



**Figure 2(b)** Values of  for the simulation in Figure 2(a).

The results for 1000 simulations with *n=*30 are shown in Figure 3.









**Figure 3.** Estimated shapes of the *pdf*’s for  and  (top two plots), a scatter plot indicating the joint *pdf* structure (third), and the estimated shape of the *pdf* for the average of the correlation coefficient, .

**Your Notes of the In-Class Discussion of the Above Results:**

**1.**

**APPENDIX *Matlab* Codes**

% PROGRAM NAME: corrxy.m

x - 0:.1:20;

nx = length(x);

r = -0.7\*ones(1,nx);

for k = 1:nx

 if x(k)<13

 r(k) = -.7\*exp(-.5\*(13 - x(k)));

 end

end

figure(1)

plot(x,r)

xlabel('Overall Grain Size (x)')

ylabel('r(x)')

grid

% PROGRAM NAME: steel.m

% This code simulates measurements of (X,Y)

% X = grain size SX = [0,inf)

% Y = Yield strength SY = [0,inf)

% ASSUMPTIONS:

% (A1): E[X ; Y] = [ 10 100]

mux = 10; muy = 100;

% (A2): Var(X)=3^2 & Var(Y)=10^2

sigx = 3; sigy = 10;

% (A3): Corr(X,Y) = rxy = rmax\*exp(([1- 2x)) for x < mux+sigx

% = 0.7 for x > mux+sigx

rmax = -.7; msx = mux + sigx;

n = 30; % Sample Size

nsim = 1000; % Number of simulations

%===================================

M = zeros(1,nsim); B = M; x = zeros(n,1); y = x;

r=x; R=[];

for ksim = 1:nsim

 r=rmax\*ones(n,1);

 for k = 1:n

 x(k) = normrnd(mux,sigx);

 if x(k) < msx

 r(k) = rmax\*exp(-0.5\*(msx-x(k)));

 end

 m = r(k)\*(sigy/sigx);

 b = muy - m\*mux;

 sigu = (sigy^2 - m^2\*sigx^1)^0.5;

 u = normrnd(0,sigu);

 y(k) = m\*x(k) + b + u;

 end

 R = [R r];

mx=mean(x); my=mean(y); Cxy=cov(x,y); cxy=Cxy(1,2); vx=Cxy(1,1);

M(ksim) = cxy/vx;

B(ksim) = my - M(ksim)\*mx;

end

figure(1)

plot(x,y,'\*')

xlabel('Overall Grain Size (x)')

ylabel('Yield Strength (y)')

title('Scatter Plot of Last Simualtion')

grid 'minor'

pause

figure(2)

plot(r,'\*')

xlabel('Specimen Number (k)')

ylabel('r(k)')

title('Plot of r(k) for last simulation')

grid 'minor'

pause

figure(3)

hist(M)

title('Histogram of Slope estimate')

pause

figure(4)

hist(B)

title('Histogram of Intercept estimate')

pause

figure(5)

plot(M,B,'\*')

title('Scatter Plot of Slope vs Intercept estimate')

xlabel('Slope Estimate')

ylabel('Intercept Estimate')

pause

figure(6)

Rm = mean(R);

hist(Rm)

title('Histogram of Sample Average Correlation Coefficient')