Questions from: <http://www.math.wustl.edu/~jmding/math3200/chw/hw1.html>

1. (a) *SAS code labeled “Program for 1a”*. For uniformly distributed values of X, the theoretical mean is 0.5 and standard deviation of 1/(2sqrt(3))=0.289 [from sqrt((1-0)^2/12)]. In our random sample of 100 X\_i observations, the mean is 0.5050260 and the standard deviation is 0.3069817. For such a small sample size, the mean and standard deviation are very close—most of the difference probably due to sampling error.

Yes, the sample distribution of the 100 random integers (Y\_i) is close to 1/10 for each k with 1 le k le 10 for this relatively small sample. Although k=8 had a low frequency (6/100 = 0.6/10) and k=10 had a high frequency (14/100 = 1.4/10), each of the other k values had frequencies which were very close to 1/10. *See output pg 4.*

(b) *SAS code labeled “Program for 1b”.* In our random sample of 10,000 XX observations, the mean is 0.5055551 and the standard deviation is 0.2903252. Compared to the sample of 100 observations, the mean was higher and therefore further from the theoretical, although by a very small amount (0.0002951 away from the mean of our first random sample and 0.0055551 from the theoretical mean of 0.5). The standard deviation of this sample was slightly (0.0166565) lower than the first random sample and closer to (and only about 0.0013252 higher than) the theoretical mean. Looking at the standard deviation and considering sampling error, the results do somewhat improve.

As for the 10,000 random integers YY, the sample distribution is much closer to 1/10 for each k with 1 le k le 10. At most, the frequency is 1.071/10 and at least, the frequency is 0.935/10. So yes, the results for the r.v. YY did improve with a greater sample size. *See output pg 6.*

2. (a) *SAS code labeled “Program 2”. Proc means output included for checking purposes*. By defining K\_i as 0 when Y\_i < X\_i and 1 otherwise (if Y\_i ge X\_i), we see that Y\_i was less than X\_i 56.56% of the time (K\_i = 1 5656/10000 times). *See output pg. 8.* The engineer predicted that Y\_i would be less than X\_i 57.16% of the time. These probabilities are only 0.6% different—a difference which might easily be explained by sampling error. Therefore, the results are indeed consistent with the engineer’s statement

\*Program for 1a;

**data** mydata;

call streaminit(**123456**);

do i=**1** to **100**;

X\_i=rand('uniform');

Y\_i=**1**+floor(**10**\*X\_i);

put Y\_i;

output;

end;

TITLE 'Homework 1: 1a';

**proc** **print** data=mydata;

**proc** **means** data=mydata mean stddev;

var X\_i;

**proc** **freq** data=mydata;

table Y\_i;

**run**;

\*Program for 1b;

**data** mydata2;

call streaminit(**12345**);

do i=**1** to **10000**;

XX=rand('uniform');

YY=**1**+floor(**10**\*XX);

put YY;

output;

end;

TITLE "Homework 1:1b";

**proc** **means** mean stddev min max;

var XX;

**proc** **freq** data=mydata2;

table YY;

**run**;

\*Program for 2;

**data** normdata;

do i=**1** to **10000**;

X\_i=**0.526**+**0.0035**\*rannor(**92641**);

Y\_i=**0.525**+**0.0043**\*rannor(**23423**);

if Y\_i<X\_i then K\_i=**1**;

else K\_i=**0**;

output;

end;

TITLE "Homework 1:2";

**proc** **means** mean stddev;

var X\_i;

var Y\_i;

**proc** **freq** data=normdata;

tables K\_i;

**run**;