EE 376A/Stat 376A
Information Theory
TTh 11-12:15pm, 380-380C

Handout #6
Thursday, January 8, 2009
T. Cover

EE376A Random Numbers

(Written by William Wu, TA)

60 numbers were submitted on 1/6/2009. 47 of them lie in $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$; just for fun, let us call those submissions the "typical set". The remaining 13 submissions were:

-2.37 -1 $\pi^{-\pi}$ $\pi/4$ $-\pi$ 8.8888888 0 9.9 $e^{-\pi}$ -9.993781 6.24 7.63920001 6.723

A summary of the data is shown below.

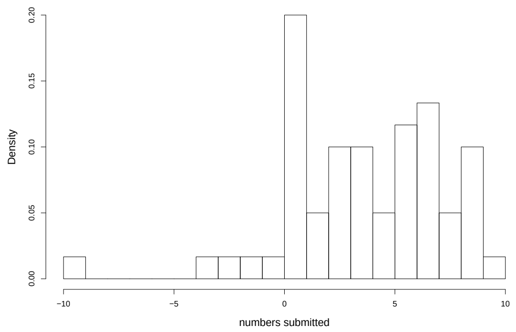

Figure 1: Histogram of all 60 submissions

	sample size	μ	σ
all	60	4.012	3.672
typical	47	4.617	2.650
atypical	13	1.826	5.723
male	49	4.214	3.261
female	8	2.906	5.635
Ph.D.	20	3.386	3.397
M.S.	35	4.794	3.148
B.S.	5	1.049	6.448

Figure 2: Counts, Means, Standard Dev.

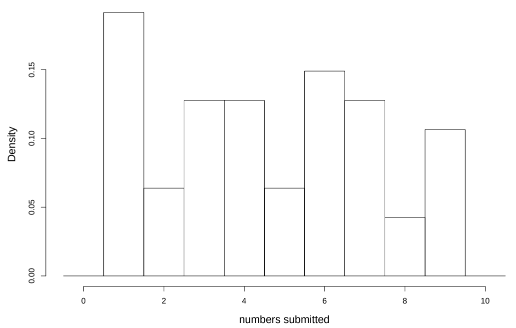

Figure 3: Histogram of typical set

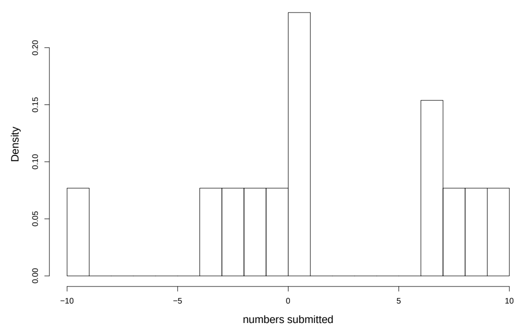

Figure 4: Histogram of atypical set

In the sequel, we will justify the following statements:

- **We cannot say that the typical numbers are not uniformly distributed.**
- **We can say that typical numbers are not just as likely as atypical numbers, and significantly more people chose typical numbers.**
- **We cannot say that males and females do not generate numbers from the same distribution.**
- **We cannot say that M.S. and Ph.D. students do not generate numbers from the same distribution.**
- **We cannot rule out the possibility that one's educational degree is independent of submitting an atypical number.**

1 Are the Numbers Uniformly Distributed?

We would like to know if the numbers could have been generated uniformly at random. We first address this question only to the typical set. A discrete Q-Q plot of the typical set against the the discrete uniform distribution over $\{1, 2, \ldots, 9\}$ is shown in Figure 5. The data points do not deviate too much from the 45 degree line, illustrating plausibility that the typical set could have been generated by uniform distribution.

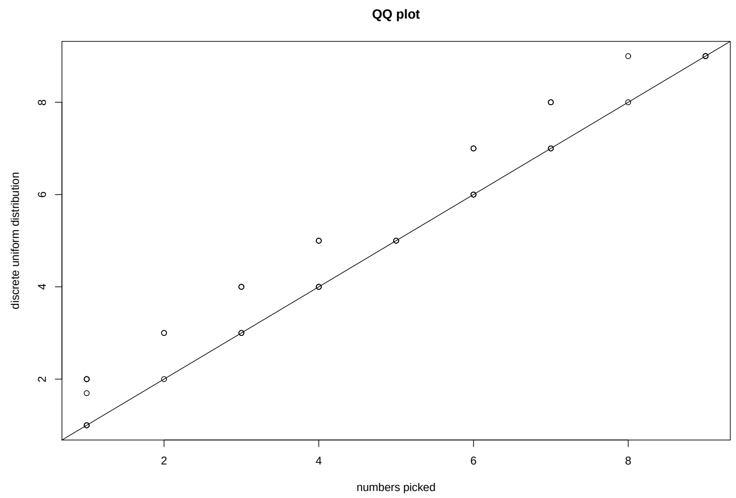

Figure 5: Q-Q plot of typical set against uniform distribution

We now use Pearson's chi-square test for goodness-of-fit.[1] The null hypothesis H_0 is that the typical numbers were generated by a discrete uniform distribution over $\{1, 2, \ldots, 9\}$. We then compute the chi-square statistic $\chi^2 = \sum_{i=1}^{9} \frac{(O_i - E_i)^2}{E_i} = 7.575$, where $E_i = 47/9$, the expected number of occurrences of the integer i under H_0, and O_i is the observed number of occurrences of integer i in our experiment. The number of degrees of freedom (df) is

$$df = (\text{number of bins}) - (\text{number of parameters to be fit}) - 1 = 9 - 0 - 1 = 8.$$

The p-value is $\Pr(\chi_8^2 \geq 7.575) = 0.476$, which is far above the 5% significance level. Hence, Pearson's chi-square test says that we **cannot reject the null hypothesis that the typical set was generated by a uniform distribution**.[2] We cannot reject it because under the null hypothesis, a χ^2 statistic greater than or equal to 7.575 can occur by chance alone almost half the time.

Now examining the atypical set, Figures 6 and 7 show Q-Q plots for both the atypical set and the entire dataset against continuous uniform distributions over intervals ranging between the minimum and maximum numbers in their respective sets. Considering the amount of wiggle away from the solid line, claiming that any of this data followed a uniform distribution would be a hard sell.

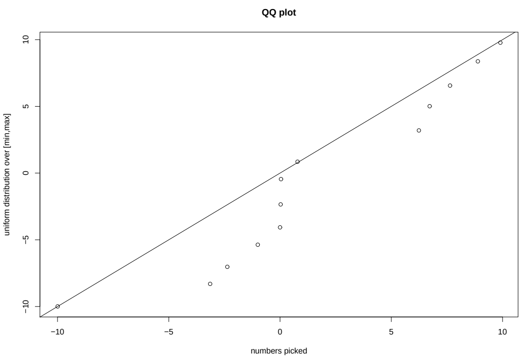

Figure 6: Q-Q plot of atypical set

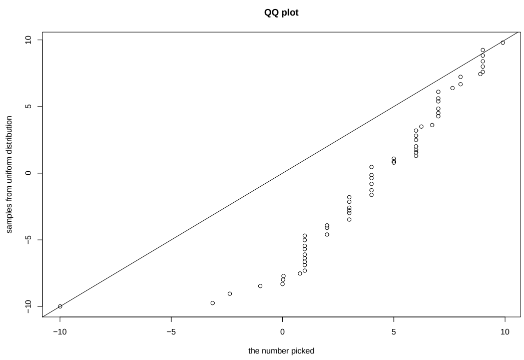

Figure 7: Q-Q plot of all data

2 Are "Typical Numbers" Typical?

We now give some justification for why the typical set might deserve its name. Again, we can run Pearson's chi-square test against the null hypothesis H_0 that typical and atypical

[1] An alternative approach to assessing goodness-of-fit is given by the relative entropy $D(O||E)$; the relation between this and χ^2 is explored in Problem 11.2 of Cover and Thomas.

[2] Note that this is *not* the same as saying that the data is generated from a uniform distribution. The only rigorous statement we can make is that we cannot reject the null hypothesis.

numbers are equally likely. E_t, the expected number of typical submissions under H_0, is $60/2 = 30$. Similarly, E_a, the expected number of atypical submissions, is also 30. The corresponding observed numbers of occurrences are $O_t = 47$ and $O_a = 13$. There are two bins and no parameters to estimate, so $df = 2 - 0 - 1 = 1$. We compute $\chi^2 = 19.2667$, and the p-value is 1.137e-05, well below 5%. **Thus students are not equally likely to choose between typical numbers and atypical numbers, and significantly more students chose typical numbers than atypical ones (a ratio of 47 to 13).**

3 Men Versus Women

We now address whether or not the numbers generated by males differ from the numbers generated by females in a statistically significant way. By analyzing names and photographs, we could definitively establish 8 submissions as having come from females, and 49 submissions as having come from males. The remaining 3 students with androgynous qualities were excluded. Figure 8 shows superimposed normalized histograms of numbers submitted by each sex, including both typical and atypical data. They seem comparable.

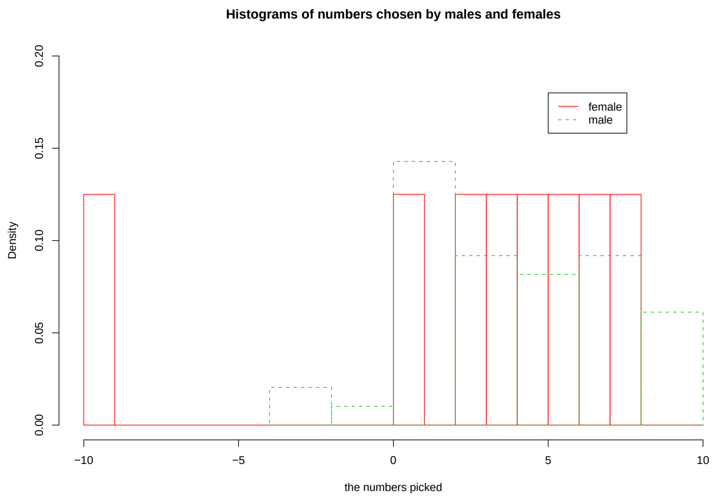

Figure 8: Histograms of male and female data

We apply the two-sample Kolmogorov-Smirnov (K-S) test against the null hypothesis that the male and female datasets come from the same underlying probability distribution. The K-S test does not require us to specify what that common distribution might be. The test statistic is $D = D_{ab} = \left(\frac{ab}{a+b}\right)^{1/2} \sup_{x \in \mathbf{R}} |E_m(x) - E_f(x)|$, where $a = 49$, $b = 8$, and E_m

and E_f are the empirical CDFs for the two datasets. It is known that D_{ab} converges in distribution to the Kolmogorov-Smirnov distribution as a and b both tend to infinity. In our dataset, $D = 0.1811$, and the p-value is $\Pr(D \geq 0.1811) = 0.978$, much higher than the 5% significance level. **Hence, we cannot reject the null hypothesis that males and females generate data from the same distribution.**

4 Ph.D. vs. M.S., And A Little B.S.

We now compare the Ph.D. students (20) with the M.S. students (35). Since there are only 5 units of B.S., we do not include them in the comparisons here because the sample size is too small. Figure 9 shows superimposed normalized histograms of the data from each group.

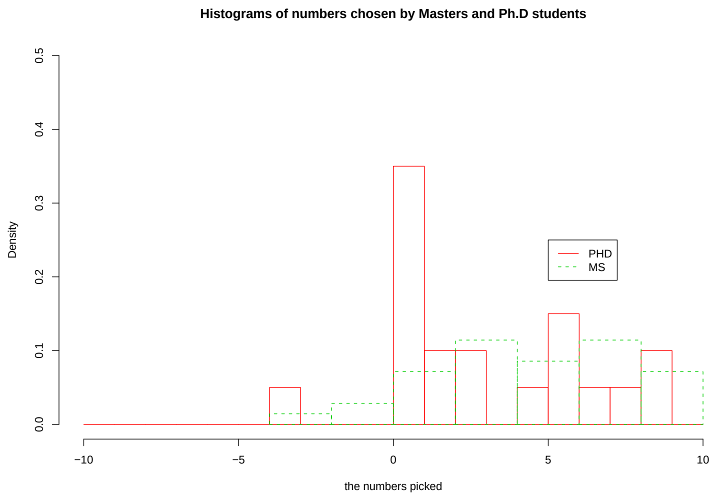

Figure 9: Histograms of M.S. and Ph.D. data

Using K-S again, we can argue that **we cannot reject the null hypothesis that Ph.D. and M.S. students generate data from the same distribution**. This is evidenced by $D = 0.271$ and a p-value of 0.3055.

Lastly, we investigate whether Ph.D. students are any more or less likely than M.S. students to submit an atypical (weird) number. Let the null hypothesis H_0 be that whether or not one submits an atypical number is independent of one's educational background. Applying Pearson's chi-squared test for independence, we tabulate our experimental observations in Table 1. E_{ij}, the expected number of observations in cell (i, j) of the table, is

Table 1: Pearson's chi-square test for independence of education and creativity

	Ph.D.	M.S.	row sums
typical	16	28	44
atypical	4	7	11
column sums	20	35	

$$E_{ij} = \frac{\text{(sum of row i)(sum of column j)}}{\text{(total number of samples in table)}}$$

and the number of degrees of freedom is

$$df = ((\text{number of rows}) - 1) \times ((\text{number of columns}) - 1) = (2-1) \times (2-1) = 1.$$

Plugging all these values in, we get $\chi^2 = 0.123$, and the p-value is 0.726, which is above the 5% significance level. **So apparently we cannot even rule out the sad possibility that educational background is independent of creative thought.**

Note that this conclusion is not the same as rejecting the alternative hypothesis, which is that there actually exists some dependence between educational background and creative thought. But then, one must wonder whether more education encourages thinking outside the box, or stifles it. 20% of all Ph.D.s and 20% of all M.S. students submitted an atypical number, whereas 40% of the B.S. submitted an atypical number. An undergraduate was also responsible for submitting the lowest negative number. Perhaps these facts can simply be disregarded due to small sample size. To be convincing, more B.S. is needed.