2.1 Give 10 challenge-response pairs. Each response should have at least 6 random-looking characters. You should store these pairs and check your memory in a day, a week, and a month from now. At any moment in time in this course, you may be asked to reply to 1 or more of your random challenges.

\frac{1}{2} \text{ point per challenge/response pair. (total 5 pts)}

Be ready to respond to your challenges!

2.2 Suggest a virtually infinite source of personal challenge - response pairs

Vague suggestions like "memories" or "books" get 2 points. More specific suggestions that had some problems (like requiring the human to write down a virtually infinite list of challenge/response pairs) get 4 points. (total 5 pts)

2.3 In a world with d days per year, what is the probability \( pr \) that no two people in a class of \( p \) people have the same birthday?

a. Give an exact formula

If each person’s birthday is chosen uniformly, then the probability \( P_r(p, d) \) that \( p \) people have no birthdays in common is:

\[
(1 - \frac{1}{d})(1 - \frac{2}{d}) \cdots (1 - \frac{p}{d}) = \prod_{i=1}^{p} (1 - \frac{i}{d}) = \frac{d!}{(d-p)!d^p}
\]

(total 5 pts)

b. Substitute \( p = \sqrt{d} \) and show that in the limit as \( d \to \infty \), the correct answer is quite pretty.

We’ll let

\[
Q(d) = \prod_{i=1}^{\sqrt{d}} (1 - \frac{i}{d})
\]

and we’ll evaluate

\[
P = \lim_{d \to \infty} \ln Q(d).
\]

Because \( \ln \) is continuous, it will follow that

\[
\lim_{d \to \infty} Q(d) = e^P.
\]

Now substituting the value of \( Q(d) \), we can see that

\[
\ln Q(d) = \sum_{i=1}^{\sqrt{d}} \ln (1 - \frac{i}{d})
\]
and we know from the Taylor expansion of $\ln(1 - x)$ that

$$
\sum_{i=1}^{\sqrt{d}} \ln(1 - \frac{i}{d}) = \sum_{i=1}^{\sqrt{d}} \sum_{j>0} \frac{(i/d)^j}{j}
$$

$$
= -\sum_{j>0} \frac{1}{j} \sum_{i=1}^{\sqrt{d}} \left( \frac{i}{d} \right)^j
$$

$$
= -\frac{\sqrt{d}}{d} - \frac{1}{2} \sum_{i=1}^{\sqrt{d}} \left( \frac{i}{d} \right)^2 - \cdots
$$

$$
\Rightarrow \lim_{d \to \infty} -\frac{1}{2} - \frac{1}{2\sqrt{d}} - \frac{1}{2} \sum_{i=1}^{\sqrt{d}} \left( \frac{i}{d} \right)^2 - \cdots = -\frac{1}{2}.
$$

So in the limit $Q(d) \to e^{-\frac{1}{2}}$. (total 5 pts)

c. Give an approximation that is easy to compute on a calculator for very large $d$ yet works “well” also for small numbers, like $d=10$.

A good approximation uses the first two terms of the Taylor series, i.e.

$$
Pr(p, d) \approx \exp\left(-\frac{p(p-1)}{2d} - \frac{p(p-1)(p-2)}{6d^2}\right)
$$

which predicts $Pr(5, 10) = .3024$ exactly. (total 5 pts)

2.4 (COUPON COLLECTOR’S PROBLEM)

Give an exact or very good approximate solution to this problem (see its statement below) that you can use on a simple calculator. Your solution $CC(n)$ should be correct in the limit as $n \to \infty$, in the sense that the ratio of your approximation to the actual value of $CC(n)$ should go to 1. In addition, your approximation should give good results for small $n$, like $n=10$.

QUESTION: A cereal box contains one of $n$ coupons, each coupon chosen uniformly at random (i.e. each coupon is equally likely to appear in a box). How many cereal boxes should one expect to buy in order to get all $n$ coupons?

Let the random variable $Y$ denote the number of cereal boxes it takes to get all $n$ coupons. Let $Y_k$ denote the number of cereal boxes it takes to go from having $k-1$ coupons to having $k$ coupons. Then

$$
Y = \sum_{i=1}^{n} Y_i
$$

and by linearity of expectation

$$
E[Y] = \sum_{i=1}^{n} E[Y_i].
$$

Since once we have $k-1$ coupons we will get a new coupon with probability $\frac{n-k+1}{n}$, it is clear that $E[Y_1] = \frac{n}{n-k+1}$, or that

$$
E[Y] = n \sum_{i=1}^{n} \frac{1}{i} = nH_n,
$$

where $H_n$ denotes the $n$th harmonic number. It can be shown by integration that $\ln n \leq H_n \leq \ln n + 1$, but this is not a very tight bound for $n = 10$, since $nH_n = 29$, $n \ln n = 23$, and $n(1 + \ln n) = 33$. Instead we use the tighter bound $H_n = \ln n + \gamma + o(1/n) \approx \ln n + 0.57$ where $\gamma = 0.57 \ldots$ is Euler’s constant. (total 5 pts)