Coupon Conspiracy

Heather Parsons
Dickens Nyabuti
Ben Speidel
Megan Silberhorn
Mark Osegard
Nick Thull
Introduction

Today we are going explain two coupon collecting examples that incorporate important statistical elements. We will first begin by using a well-known example of a coupon collecting-type of game.
We are going to use the McDonald’s Monopoly Best Chance Game to connect a real world example with our Coupon Collecting Examples.
### McDonald’s® Scam

**List of Prizes Allegedly Won By Fraudulent Means**

<table>
<thead>
<tr>
<th>Year</th>
<th>Prize Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>(1) $100,000 Prize, (2) $1,000,000 Prizes</td>
</tr>
<tr>
<td>1997</td>
<td>(1) $1,000,000 Prize</td>
</tr>
<tr>
<td>1998</td>
<td>$200,000</td>
</tr>
<tr>
<td>1999</td>
<td>(3) $1,000,000 Prizes</td>
</tr>
<tr>
<td>2000</td>
<td>(3) $1,000,000 Prizes</td>
</tr>
<tr>
<td>2001</td>
<td>(3) $1,000,000 Prizes</td>
</tr>
</tbody>
</table>
The Odds of Winning

Instant Win • 1 in 18,197
Collect & Win • 1 in 2,475,041
Best Buy Gift Cards • 1 in 55,000
Important Statistical Elements

• **Expected Value** – gives the average behavior of the Random Variable $X$
  Denoted as: $E[X] = \sum x_i P\{X = x_i\}$

• **Variance** – measures spread, variability, and dispersion of $X$
  Denoted as: $Var(X) = E[(X - E[X])^2]$ 

We will use these concepts in more detail in our fair coupon collecting examples later in our discussion.
Review

• **Sample space** – the set of all possible outcomes of a random experiment \((S)\)

• **Event** – any subset of the sample space

• **Probability** – is a set function defined on the power set of \(S\)

So if \(A \subseteq S\),

then \(P(A) = \text{probability of } A\)
Bernoulli Random Variable

- Toggles between 0 and 1 values
- Where 1 represents a success and 0 a failure
- Success probability = \( p \)
- Probability of Failure = \( (1-p) \)

Bernoulli Trials

- Experiments having two possible outcomes
- Independent sequences of Bernoulli RV’s with the same success probability

\[
E[X] = p \quad \text{Var}(X) = p(1-p)
\]
Review Continued

Geometric Random Variables

Performing Bernoulli Trials:
P = success probability
X = # of trials until 1st success
Assume 0 < P ≤ 1

\[ E[X] = \frac{1}{P} \]

\[ Var(X) = \frac{(1 - p)}{p^2} \]
Expectations of Sums of Random Variables

• For single valued R.V.’s:
  \[ E[X] = \sum_i x_i P\{X = x_i\} \]

• Suppose \( g : \mathbb{R}^2 \rightarrow \mathbb{R} \) (i.e. \( g(x,y) = g \))

For multiple R.V.’s:
\[ E[g(X, Y)] = \sum_x \sum_y g(x, y) P\{X = x, Y = y\} \]
Properties of Sums of Random Variables

– Expectations of Random Variables can be summed.

– Recall:
\[ E[g(X, Y)] = \sum_x \sum_y g(x, y) P\{X = x, Y = y\} \]

– Corollary:
\[ E[X + Y] = E[X] + E[Y] \]

– More generally:
\[ E\left[ \sum_{i=1}^{n} X_i \right] = \sum_{i=1}^{n} E[X_i] \]
Sums of Random Variables

- Sums of Random Variables can be summed up and kept in its own Random Variable

\[ Y = \sum_{i=0}^{\infty} X_i \]

Where \( Y \) is a R.V. and \( X_i \) is the \( i^{th} \) instance of the Random Variable \( X \).
Variance & Covariance of Random Variables

- Variance a measure of spread and variability
  \[
  \text{var}(X) = E[(X - E[X])^2] \\
  \text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2 \text{cov}(X, Y)
  \]

- Facts:
  - i) \( \text{var}(X) = E[X^2] - (E[X]^2) \)
  - ii) \( \text{var}(aX + b) = a^2 \text{Var}(X) \)

- Covariance
  - Measure of association between two r.v.’s
  \[
  \text{cov}(X, Y) = E[(X-E[X])(Y-E[Y])]
  \]
Properties of Covariance

Where X, Y, Z are random variables, C constant:

- \( \text{cov}(X, X) = \text{var}(X) \)
- \( \text{cov}(X, Y) = \text{cov}(Y, X) \)
- \( \text{cov}(cX, Y) = c \times \text{cov}(X, Y) \)
- \( \text{cov}(X, Y + Z) = \text{cov}(X, Y) + \text{cov}(X, Z) \)
McDonald’s®
Coupon Conspiracy I
Suppose there are $M$ different types of coupons, each equally likely.

Let

$$X = \text{number of coupons one needs to collect in order to get the entire set of coupons.}$$

**Problem:**

Find the expected value and variance of $X$

i.e $E \ [X]$

$Var \ (X)$
Solution:

Idea: Break $X$ up into a sum of simpler random variables

Let

$$X_i = \text{number of coupons needed after } i \text{ distinct types have been collected until a new type has been obtained.}$$

Note:

$$X = \sum_{i=1}^{m-1} X_i$$

the $X_i$ are independent so

$$E[X] = \sum_{i=1}^{m-1} E[X_i]$$

$$Var(X) = \sum_{i=1}^{m-1} Var(X_i)$$
Observe:

If we already have \( i \) distinct types of coupons, then using **Geometric Random Variables**, 

\[
P(\text{next is new}) = \frac{(m-i)}{m}
\]

Regarding each coupon selection as a trial

\( X_i = \) number of trials until the next success

and

\[
P(\text{next is new}) = \frac{(m-i)}{m}
\]
We know that the

\[
E[X_i] = \frac{1}{p} = \frac{m}{m-i}
\]

\[
Var(X_i) = \frac{1-p}{p^2} = \left( \frac{i}{m} \right) \left( \frac{m^2}{(m-i)^2} \right) = \frac{mi}{(m-i)^2}
\]

so

\[
E[X] = \sum_{i=0}^{m-1} E[X_i] = \sum_{i=0}^{m-1} \left( \frac{m}{m-i} \right)
\]

\[
Var(X) = \sum_{i=0}^{m-1} Var(X_i) = \sum_{i=0}^{m-1} \left( \frac{mi}{(m-i)^2} \right)
\]
so

\[ E[X] = \sum_{i=0}^{m-1} \left( \frac{m}{m-i} \right) = m \sum_{i=0}^{m-1} \frac{1}{m-i} \]

\[ = m \left[ \frac{1}{m} + \frac{1}{m-1} + \frac{1}{m-2} + \ldots + \frac{1}{2} + \frac{1}{1} \right] \]

reversing the expression above,

\[ = m \left[ 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{m} \right] \]

\[ = m \left[ \sum_{i=1}^{m} \frac{1}{i} \right] \]
Euler’s Constant

\[ \gamma = \lim_{m \to \infty} \left( \sum_{i=1}^{m} \frac{1}{i} - \log m \right) \]

so

\[ \sum_{i=1}^{m} \frac{1}{i} \approx \log m \]

\[ \therefore \quad E \left[ X \right] = m \left( \sum_{i=1}^{m} \frac{1}{i} \right) \approx m \log m \]
So,

\[ \text{Var}(X) = \sum_{i=0}^{m-1} \left( \frac{mi}{(m-i)^2} \right) \]

\[ = \sum_{i=1}^{m-1} \left( \frac{mi}{(m-i)^2} \right) \]

\[ = m \sum_{i=1}^{m-1} \left( \frac{i}{(m-i)^2} \right) \]

to simplify this, we will use a trick in the next slide
Trick: adding $m$ to and subtracting $m$ from the numerator of the sum.

\[
\frac{i}{(m-i)^2} = \frac{(i-m+m)}{(m-i)^2} = \frac{m-(m-i)}{(m-i)^2} = \frac{m}{(m-i)^2} - \frac{(m-i)}{(m-i)^2}
\]

\[
\therefore \frac{i}{(m-i)^2} = \frac{m}{(m-i)^2} - \frac{1}{(m-i)}
\]
Applying the trick from the previous slide to,

\[ m \sum_{i=1}^{m-1} \left( \frac{i}{(m-i)^2} \right) \]

we get,

\[ m \sum_{i=1}^{m-1} \left( \frac{m}{(m-i)^2} \right) - m \sum_{i=1}^{m-1} \left( \frac{1}{(m-i)} \right) \]

pulling out the m in the first sum,

\[ = m^2 \sum_{i=1}^{m-1} \left( \frac{1}{(m-i)^2} \right) - m \sum_{i=1}^{m-1} \left( \frac{1}{(m-i)} \right) \]
Expanding the sums from the previous slide

\[= m^2 \left[ \frac{1}{(m-1)^2} + \frac{1}{(m-2)^2} + \ldots + \frac{1}{1^2} \right] - m \left[ \frac{1}{(m-1)} + \frac{1}{(m-2)} + \ldots + \frac{1}{1} \right] \]

reversing the order of the sum above,

\[= m^2 \left[ \frac{1}{(1)^2} + \frac{1}{(2)^2} + \ldots + \frac{1}{(m-1)^2} \right] - m \left[ \frac{1}{(1)} + \frac{1}{(2)} + \ldots + \frac{1}{(m-1)} \right] \]

\[= m^2 \sum_{i=1}^{m-1} \frac{1}{i^2} - m \sum_{i=1}^{m-1} \frac{1}{i} \]

By the Basel series we will simplify this in the next slide.
Explaining the Basel Series

\[ \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \ldots = \frac{\pi^2}{6} \]

for a large \( m \), the Basel Series converges to \( \frac{\pi^2}{6} \)

\[ \text{Var}(X) \approx m^2 \left( \frac{\pi^2}{6} \right) - m \log(m) \approx m^2 \]
In Conclusion

So clearly the variance is:

\[ \text{Var}(X) \approx m^2 \]

and the expected value is:

\[ E \left[ X \right] \approx m \log m \]

These approximations are dependent on the fact that \( m \) is a large number approaching infinity. Where \( m \) is the number of different types of coupons.
McDonald’s®
Coupon Conspiracy II
The 2\textsuperscript{nd} Coupon Conspiracy

Given:

- Sample of \( n \) coupons
- \( m \) possible types
- \( X := \) number of distinct types of coupons

Find:

- \( E[X] \) (expected value of \( x \))
- \( \text{Var}(X) \) (variance of \( x \))
Redefining $X$

Decompose $X$ into a sum of Bernoulli indicators

Let $X_i = \begin{cases} 1 & \text{if a type } i \text{ is present} \\ 0 & \text{otherwise} \end{cases}$

$1 \leq i \leq m$

Note: $X = \sum_{i=1}^{m} X_i$

Note: $E[X] = \sum_{i=1}^{m} E[X_i]$

Remark: $X_1, X_2, \ldots, X_m$ are dependent

(knowing one coupon type occurred lowers the opportunity for the other coupons types to occur)
Redefining $\text{Var}(X)$

Recall: $\text{Var}(X) = \text{Cov}(X, X)$

\[ \text{Var}(X) = \text{Var}\left(\sum_{i=1}^{m} X_i\right) = \text{Cov}\left(\sum_{i=1}^{m} X_i, \sum_{i=1}^{m} X_i\right) \]

\[ = \sum_{i=1}^{m} \text{Cov}\left(X_i, \sum_{j=1}^{m} X_j\right) \]

\[ = \sum_{i=1}^{m} \sum_{j=1}^{m} \text{Cov}\left(X_i, X_j\right) \]

(by covariance bilinearity)
Summary thus far

\[ E \left[ X \right] = \sum_{i=1}^{m} E \left[ X_i \right] \]

\[ Var (X) = \sum_{i=1}^{m} \sum_{j=1}^{m} Cov (X_i, X_j) \]
Success Probability

Recall: \( X_i \) is Bernoulli, with success probability \( P \)

\[
P = P \{ X_i = 1 \} = 1 - P \{ X_i = 0 \}
\]

Note: \( P\{X_i = 0\} = P\{\text{type } i \text{ does not occur on } j^{\text{th}} \text{ trial}\} \)

\[
\prod_{j=1}^{n} P = 1 - \prod_{j=1}^{n} P \{ X_i = 0 \}
\]

\[
= 1 - \left( \frac{m - 1}{m} \right)^n
\]

Thus \( E [ X_i ] = 1 - \left( \frac{m - 1}{m} \right)^n \)
Recap

\[ E[X_i] = 1 - \left( \frac{m-1}{m} \right)^n \]

\[ \text{Var}[X_i] = \left( \frac{m-1}{m} \right)^n \left[ 1 - \left( \frac{m-1}{m} \right)^n \right] \]

So,

\[ E[X] = \sum_{i=1}^{m} E[X_i] \]

\[ = \sum_{i=1}^{m} \left[ 1 - \left( \frac{m-1}{m} \right)^n \right] \]

\[ = m \left[ 1 - \left( \frac{m-1}{m} \right)^n \right] \]
Var(X)

To calculate $\text{Var}(X)$, we need to find $\text{Cov}(X_i, X_j)$

$$\text{Cov}(X_i, X_j) = E[X_i X_j] - E[X_i]E[X_j]$$

In particular when $i \neq j$
$E \left[ X_i X_j \right]$

Let $A_k$ = event that type $k$ is present

$E[X_iX_j] = P\{X_iX_j = 1\}$

$= P\left( A_i \cap A_j \right) = 1 - P\left( A_i \cap A_j \right)^c$

By De Morgan, $\left( A_i \cap A_j \right)^c = A_i^c \cup A_j^c$

$= 1 - \left( A_i^c \cup A_j^c \right)$

Using the Inclusion/Exclusion rule:

$\{P\left(A \cup B\right) = P\left(A\right) + P\left(B\right) - P\left(A \cap B\right)\}$

$= 1 - \left[ P\left(A_i^c \right) + P\left(A_j^c \right) - P\left(A_i^c \cap A_j^c \right) \right]$
\[ E \left[ X_i X_j \right] \text{ cont.} \]

Recall: \[ P \left( A_i^c \right) = \left( \frac{m-1}{m} \right)^n, \quad P \left( A_j^c \right) = \left( \frac{m-1}{m} \right)^n \]

\[ P \left( A_i^c \cap A_j^c \right) = \left( \frac{m-2}{m} \right)^n \]

\[ E[X_i X_j] = 1 - \left[ P(A_i^c) + P(A_j^c) - P(A_i^c \cap A_j^c) \right] \]

\[ = 1 - \left[ \left( \frac{m-1}{m} \right)^n + \left( \frac{m-1}{m} \right)^n - \left( \frac{m-2}{m} \right)^n \right] \]

\[ = 1 - \left[ 2 \left( \frac{m-1}{m} \right)^n - \left( \frac{m-2}{m} \right)^n \right] \]
\[
\text{Cov}(X_i, X_j) = 1 - 2 \left( \frac{m-1}{m} \right)^n - \left( \frac{m-2}{m} \right)^n \right) - \left( 1 - \left( \frac{m-1}{m} \right)^n \right)^2 \\
= 1 - 2 \left( \frac{m-1}{m} \right)^n + \left( \frac{m-2}{m} \right)^n - 1 + 2 \left( \frac{m-1}{m} \right)^n - \left( \frac{m-1}{m} \right)^{2n} \\
= \left( \frac{m-2}{m} \right)^n - \left( \frac{m-1}{m} \right)^{2n}
\]
Back to the $\text{Var}(X)$

Recall: $\text{Var}(X_i) = \text{Cov}(X_i, X_j)$

$$\text{Var}(X) = \sum_{i=1}^{m} \sum_{j=1}^{m} \text{Cov}(X_i, X_j)$$

$$= \sum_{i=1}^{m} \text{Cov}(X_i, X_i) + 2 \sum_{i=1}^{m} \sum_{j<i}^{m} \text{Cov}(X_i, X_j)$$
Var(X) in terms of Covariance

\[
Var(X) = \sum_{i=1}^{m} Cov(X_i, X_i) + 2 \sum_{i=1}^{m} \sum_{j<i}^{m} Cov(X_i, X_j)
\]

Simplifying the summation, We get:

\[
= \sum_{i=1}^{m} Cov(X_i, X_i) + 2 \sum_{i=1}^{m} m (m - 1) Cov(X_i, X_j)
\]
Var(X) equals

\[ Var(X) = \sum_{i=1}^{m} Cov(X_i, X_i) + 2 \sum_{i=1}^{m} m(m-1) Cov(X_i, X_j) \]

Recall: \( Var(X_i) = m \left( \frac{m-1}{m} \right)^n \left( 1 - \left( \frac{m-1}{m} \right)^n \right) \)

\[ = m \left( \frac{m-1}{m} \right)^n \left( 1 - \left( \frac{m-1}{m} \right)^n \right) + (2m^2 - 2m) \left( \frac{m-2}{m} \right)^n - \left( \frac{m-1}{m} \right)^{2n} \]

Simplified,

\[ Var(X) = m \left( \frac{m-1}{m} \right)^n + m(m-1) \left( \frac{m-2}{m} \right)^n - m^2 \left( \frac{m-1}{m} \right)^{2n} \]
In Case you Forgot:

Independent Case:

\[ E \left[ X \right] \approx m \log m \]

\[ \text{Var}(X) \approx m^2 \]

Dependent Case:

\[ E \left[ X \right] = m \left[ 1 - \left( \frac{m-1}{m} \right)^n \right] \]

\[ \text{Var}(X) = m\left( \frac{m-1}{m} \right)^n + m(m-1)\left( \frac{m-2}{m} \right)^n - m^2\left( \frac{m-1}{m} \right)^{2n} \]
Analytic vs. Simulation Approaches

- For a simulation we will consider a fair set of the McDonald’s Coupons from Problem I.
- We will analytically find $E[X]$ and $\text{Var}(X)$ and compare these two values with the results from our computer simulation.
Analytic Approach

Given $M = 10$, compute

$$E \left[ X \right] = m \left( \sum_{i=1}^{m} \frac{1}{i} \right) \approx m \log m$$

$$Var(X) \approx m^2 \left( \frac{\pi^2}{6} \right) - m \log(m) \approx m^2$$
• References

- Sheldon Ross
- 2-3 pages from C++ books
- McDonald’s®
- Dr. Deckelman
Any Questions???