Time to collect coupons

- $X$-time to get $n$ coupons.
- $X_1$ - time to get first coupon. Note: $X_1 = 1$. $E(X_1) = 1$.
- $X_2$ - time to get second coupon after getting first.
- $Pr[\text{"get second coupon"|"got milk first coupon"]] = \frac{n-1}{n}$
- $E[X_2] = \frac{1}{n} \Rightarrow E[X_2] = n$.
- $Pr[\text{"getting $i$th coupon"|"got $i-1$st coupons"]] = \frac{n-i+1}{n}$
- $E[X_i] = \frac{1}{n} \Rightarrow E[X_i] = n$.
- $E[X] = E[X_1] + \cdots + E[X_n] = n \ln(n) + \gamma$.

Review: Harmonic sum

$H(n) = 1 + \frac{1}{2} + \cdots + \frac{1}{n} \approx \int_1^n \frac{1}{x} \, dx = \ln(n)$.

A good approximation is $H(n) \approx \ln(n) + \gamma$ where $\gamma \approx 0.58$ (Euler-Mascheroni constant).

Paradox

Consider this stack of cards (no glue!):

If each card has length 2, the stack can extend $H(n)$ to the right of the table. As $n$ increases, you can go as far as you want!
Stacking

The cards have width 2. Induction shows that the center of gravity after \( n \) cards is \( H(n) \) away from the right-most edge.

Geometric Distribution.

Experiment: flip a coin with heads prob. \( p \) until Heads. Random Variable \( X \): number of flips.

And distribution is:

(A) \( X \sim G(p) : \text{Pr}[X = i] = (1 - p)^{i-1}p \).
(B) \( X \sim B(p,n) : \text{Pr}[X = i] = \binom{n}{i}p^i(1-p)^{n-i} \).

(A) Distribution of \( X \sim G(p) : \text{Pr}[X = i] = (1 - p)^{i-1}p \).

Calculating \( E[g(X)] \): LOTUS

Let \( Y = g(X) \). Assume that we know the distribution of \( X \).
We want to calculate \( E[Y] \).

Method 1: We calculate the distribution of \( Y \):
\[
\text{Pr}[Y = y] = \text{Pr}[X \in g^{-1}(y)] \quad \text{where} \quad g^{-1}(x) = \{ x \in \mathbb{R} : g(x) = y \}.
\]
This is typically rather tedious!

Method 2: We use the following result.

Called “Law of the unconscious statistician.”

Theorem:
\[
E[g(X)] = \sum g(x)\text{Pr}[X = x].
\]

Proof:
\[
E[g(X)] = \sum_{m \in \text{Range}(X)} g(m)\text{Pr}[X = m] = \sum_{m \in \text{Range}(X)} g(m)\sum_{x \in g^{-1}(m)} \text{Pr}[X = x] = \sum_{x} g(x)\text{Pr}[X = x].
\]

Poll.

Which is LOTUS?

(A) \( E[X] = \sum_{x \in \text{Range}(X)} g(x)\text{Pr}[g(X) = g(x)] \)
(B) \( E[X] = \sum_{x \in \text{Range}(X)} g(x)\text{Pr}[X = x] \)
(C) \( E[X] = \sum_{x} \text{Range}(g) \text{Pr}[g(X) = x] \)

Geometric Distribution: Memoryless - Interpretation

\[
\text{Pr}[X > n + m | X > n] = \text{Pr}[X > m], m, n \geq 0.
\]

A': is \( m \) coin tosses before heads.
A:B: \( m \) 'more' coin tosses before heads.
The coin is memoryless, therefore, so is \( X \).

Independent coin: \( \text{Pr}[H' \text{ any previous set of coin tosses}] = p \)

Geometric Distribution: Memoryless by derivation.

Let \( X \) be \( G(p) \). Then, for \( n \geq 0, \)
\[
\text{Pr}[X > n] = \text{Pr}[\text{first } n \text{ flips are T}] = (1 - p)^n.
\]

Theorem
\[
\text{Pr}[X > n + m | X > n] = \text{Pr}[X > m], m, n \geq 0.
\]

Proof:
\[
\text{Pr}[X > n + m | X > n] = \frac{\text{Pr}[X > n + m \text{ and } X > n]}{\text{Pr}[X > n]} = \frac{\text{Pr}[X > n + m]}{\text{Pr}[X > n]} = \frac{(1 - p)^{n + m}}{(1 - p)^n} = (1 - p)^m = \text{Pr}[X > m].
\]
Variance and Standard Deviation

**Fact:** \( \text{var}[X] = E[X^2] - E[X]^2. \)

Indeed:

\[
\begin{align*}
\text{var}(X) & = E[(X - E[X])^2] \\
& = E[X^2] - 2E[X]E[X] + E[X]^2 \\
& = E[X^2] - 2E[X]E[X] + E[X]^2, \text{ by linearity} \\
& = E[X^2] - E[X]^2. \\
\end{align*}
\]

Example

Consider the random variable \( X \) such that

\[
X = \begin{cases} 
\mu - \sigma, & \text{w. p. } \frac{1}{2} \\
\mu + \sigma, & \text{w. p. } \frac{1}{2}.
\end{cases}
\]

Then, \( E[X] = \mu \) and \( (X - E[X])^2 = \sigma^2 \). Hence,

\[
\text{var}(X) = \sigma^2 \text{ and } \sigma(X) = \sigma.
\]

Exercise: How big can you make \( \sigma(X) \) if \( E[|X - E[X]|] \)?

Uniform

Assume that \( P[X = i] = \frac{1}{n} \) for \( i \in \{1, \ldots, n\} \). Then

\[
E[X] = \sum_{i=1}^{n} i \cdot P[X = i] = \frac{1}{n} \sum_{i=1}^{n} i \\
= \frac{1}{n} \left( \frac{n(n+1)}{2} \right) = \frac{n+1}{2}
\]

Also,

\[
E[X^2] = \sum_{i=1}^{n} i^2 \cdot P[X = i] = \frac{1}{n} \sum_{i=1}^{n} i^2 \\
= \frac{1}{n} \left( \frac{n(n+1)(2n+1)}{6} \right) = \frac{1+3n+2n^2}{6}, \text{ as you can verify.}
\]

This gives

\[
\text{var}(X) = \frac{1+3n+2n^2}{6} - \frac{(n+1)^2}{4} = \frac{n^2-1}{12}.
\]

(Sort of \( \int_0^1 x^2 \, dx = \frac{1}{3} \))

Variance of geometric distribution.

\( X \) is a geometrically distributed RV with parameter \( p \). Thus, \( P[X = n] = (1-p)^{n-1}p \) for \( n \geq 1 \). Recall \( E[X] = 1/p. \)

\[
E[X] = \sum_{i=1}^{n} i \cdot P[X = i] = \frac{1}{p}
\]

\[
E[X^2] = \sum_{i=1}^{n} i^2 \cdot P[X = i] = \frac{1}{p} \sum_{i=1}^{n} i^2 \\
= \frac{1}{p} \left( \frac{n(n+1)(2n+1)}{6} \right) = \frac{1+3n+2n^2}{6}, \text{ as you can verify.}
\]

This gives

\[
\text{var}(X) = \frac{1+3n+2n^2}{6} - \frac{(n+1)^2}{4} = \frac{n^2-1}{12}.
\]

\[
\text{var}(X) = \frac{n^2-1}{12} \approx E[X] \text{ when } p \text{ is small(ish)}.
\]
**Fixed points.**

Number of fixed points in a random permutation of $n$ items.

"Number of student that get homework back."

$X = X_1 + X_2 \cdots + X_n$

where $X_i$ is indicator variable for $i$th student getting hw back.

\[
E(X^2) = \sum_i E(X_i^2) + \sum_{i\neq j} E(X_iX_j).
\]

\[
= n \frac{1}{n} + (n)(n-1) \times \frac{1}{n(n-1)}
\]

\[= 1 + 1 = 2.
\]

\[
E(X^2) = 1 \times Pr[X_i = 1] + 0 \times Pr[X_i = 0]
\]

\[
E(X|X) = 1 \times Pr[X_i = 1 \cap X_j = 1] + 0 \times Pr[\text{"anything else"}]
\]

\[= 1 \times \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}
\]

\[
\text{Var}(X) = E(X^2) - (E(X))^2 = 2 - 1 = 1.
\]

**Poll: fixed points.**

What's true?

(A) $X_i$ and $X_j$ are independent.

(B) $E[X_i X_j] = Pr[X_i X_j = 1]$.

(C) $Pr[X_i X_j] = \frac{(n-2)!}{n!}$.

(D) $X_i^2 = X_i$.

**Variance: binomial.**

\[
E[X^2] = \sum \binom{n}{i} p^i (1-p)^{n-i}.
\]

Too hard!

Ok.. fine.

Let's do something else.

Maybe not much easier...but there is a payoff.

**Independent random variables.**

Independent: $P[X = a, Y = b] = Pr[X = a]Pr[Y = b]$


\[
E[XY] = \sum a \sum b a \times b \times Pr[X = a]Pr[Y = b]
\]

\[= ( \sum a \times Pr[X = a] ) ( \sum b \times Pr[Y = b] )
\]

\[= E[X]E[Y]
\]

\[
\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y).
\]

**Theorem:**

If $X$ and $Y$ are independent, then

\[
\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y).
\]

**Proof:**

Since shifting the random variables does not change their variance, let us subtract their means.

That is, we assume that $E(X) = 0$ and $E(Y) = 0$.

Then, by independence,

\[
E(XY) = E(X)E(Y) = 0.
\]

Hence,

\[
\text{Var}(X+Y) = \text{Var}(X+Y)^2 = \text{Var}(X^2 + 2XY + Y^2)
\]

\[= \text{Var}(X^2) + 2E(XY) + E(Y^2) = E(X^2) + E(Y^2)
\]

\[= \text{Var}(X) + \text{Var}(Y).
\]
Variance of sum of independent random variables

**Theorem:** If $X, Y, Z, \ldots$ are pairwise independent, then
$$\text{Var}(X + Y + Z + \cdots) = \text{Var}(X) + \text{Var}(Y) + \text{Var}(Z) + \cdots.$$  

**Proof:** Since shifting the random variables does not change their variance, let us subtract their means. That is, we assume that $E[X] = E[Y] = \cdots = 0$.

Then, by independence, $E[XY] = E[X]E[Y] = 0$. Also, $E[XZ] = E[YZ] = \cdots = 0$.

Hence,
$$\text{Var}(X + Y + Z + \cdots) = E[(X + Y + Z + \cdots)^2] - E[X + Y + Z + \cdots]^2$$
$$= E[X^2 + Y^2 + Z^2 + \cdots + 2XY + 2XZ + 2YZ + \cdots]$$
$$= E[X^2] + E[Y^2] + E[Z^2] + \cdots + 0 + \cdots + 0$$
$$= \text{Var}(X) + \text{Var}(Y) + \text{Var}(Z) + \cdots.$$  

**Covariance**

**Definition** The covariance of $X$ and $Y$ is
$$\text{cov}(X, Y) := E[(X - E[X])(Y - E[Y])].$$

**Fact**
$$\text{cov}(X, Y) = E[XY] - E[X]E[Y].$$

**Proof:** Think about $E[X] = E[Y] = 0$. Just $E[XY]$. For the sake of completeness.

$$= E[XY] - E[X]E[Y].$$

**Correlation**

**Definition** The correlation of $X, Y$, $\text{Cor}(X, Y)$ is
$$\text{corr}(X, Y) := \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$

**Theorem:** $-1 \leq \text{corr}(X, Y) \leq 1$.

**Proof:** Idea: $(a - b)^2 > 0 \iff a^2 + b^2 \geq 2ab$.


$\text{Cor}(X, Y) = E[XY]/\sqrt{E[X^2]}\sqrt{E[Y^2]}$.


$\Rightarrow E[XY] < 1.


Shifting and scaling doesn’t change correlation.

Variance of Binomial Distribution.

Flip coin with heads probability $p$.

$X$ - how many heads?

$X_i = \begin{cases} 1 & \text{if } i\text{th flip is heads} \\ 0 & \text{otherwise} \end{cases}$

$E(X_i) = p$ and $\text{Var}(X_i) = p(1-p)$

$E(X) = np$ and $\text{Var}(X) = np(1-p)$

$E(X^2) = E(X) + \text{Var}(X) = np + np(1-p) = np$.

Poisson Distribution: Variance.

**Definition** Poisson Distribution with parameter $\lambda > 0$

$X \sim \text{Poisson}(\lambda) \iff P[X = m] = \frac{e^{-\lambda} \lambda^m}{m!}, m \geq 0.$

Mean, Variance?

Ugh.

Recall that Poisson is the limit of the Binomial with $p = \lambda/n$ as $n \to \infty$.

**Variance:** $\text{Var}(X) = E(X^2) - (E(X))^2$

$E(X^2) = \lambda + \lambda^2$.

Examples of Covariance

Note that $E[X] = 0$ and $E[Y] = 0$ in these examples. Then $\text{cov}(X, Y) = E[XY]$.

When $\text{cov}(X, Y) > 0$, the RVs $X$ and $Y$ tend to be large or small together. $X$ and $Y$ are said to be positively correlated.

When $\text{cov}(X, Y) < 0$, when $X$ is larger, $Y$ tends to be smaller. $X$ and $Y$ are said to be negatively correlated.

When $\text{cov}(X, Y) = 0$, we say that $X$ and $Y$ are uncorrelated.
Examples of Covariance

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P(Y=1, X=1)</th>
<th>P(Y=1, X=2)</th>
<th>P(Y=1, X=3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.3</td>
<td>0.35</td>
<td>0.25</td>
<td>0.2</td>
</tr>
<tr>
<td>2</td>
<td>0.15</td>
<td>0.25</td>
<td>0.15</td>
<td>0.1</td>
</tr>
<tr>
<td>3</td>
<td>0.15</td>
<td>0.35</td>
<td>0.15</td>
<td>0.1</td>
</tr>
</tbody>
</table>

- `E[X] = 1.5 / 2 + 0.4 / 3 + 0.45 / 3 = 2.3`
- `E[X^2] = 1^2 * 0.15 + 2^2 * 0.4 + 3^2 * 0.45 = 5.8`
- `E[Y] = 1 * 0.2 + 2 * 0.6 + 3 * 0.2 = 2`
- `E[Y^2] = 1 * 0.2 + 4 * 0.6 + 9 * 0.2 = 4.4`
- `E[XY] = 1 * 1 * 0.05 + 1 * 2 * 0.1 + ... = 3 * 0.2 = 4.85`

- `var[X] = E[X^2] - E[X]^2 = 51`
- `corr(X,Y) = 0.55`

Properties of Covariance

- **Fact**
  - (a) `var[X] = cov(X,X)`
  - (b) `X,Y` independent ⇒ `cov(X,Y) = 0`
  - (c) `cov(aX + bY,cU + dV) = ac cov(X,U) + ad cov(X,V)`
  - (d) `cov(aX+bY,cU+dV) = ac cov(X,U) + ad cov(X,V) + bc cov(Y,U) + bd cov(Y,V)`

Proof:
- (a)-(b)-(c) are obvious.
- (d) In view of (c), one can subtract the means and assume that the RVs are zero-mean. Then,
  - `cov(aX+bY,cU+dV) = E[(aX+bY)(cU+dV)]`
  - `= ac E[XU] + ad E[XV] + bc E[YU] + bd E[YV]`
  - `= ac cov(X,U) + ad cov(X,V) + bc cov(Y,U) + bd cov(Y,V).`

Random Variables so far.

- Probability Space: `Ω, Pr : Ω → [0,1], ∑ω∈Ω Pr(w) = 1`.
- Random Variables: `X : Ω → R`.
- Associated event: `Pr[X = a] = ∑ω∈X(a)Pr(ω)`
- X and Y independent if all associated events are independent.
- Expectation: `E[X] = ∑a Pr[X = a] = ∑ω∈Ω X(ω)Pr(ω)`.
- Variance: `Var(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2`
- For independent `X,Y`, `Var(X+Y) = Var(X) + Var(Y)`.
- Poisson: `X ∼ P(λ) → E(X) = λ, Var(X) = λ`.
- Binomial: `X ∼ B(n,p) → E(X) = np, Var(X) = np(1-p)`.
- Uniform: `X ∼ U[1,...,n] → E(X) = n+1\,2, Var(X) = \frac{n^2-1}{12}`.
- Geometric: `X ∼ G(p) → E(X) = \frac{1}{p}, Var(X) = \frac{1-p}{p^2}`.

Summary

- **Variance**: `var[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2`
- **Fact**: `var[aX+b] = a^2 var[X]`
- **Sum**: `X, Y, Z pairwise ind. ⇒ var[X + Y + Z] = ...`