13: Statistics of Multiple RVs

Jerry Cain
April 26, 2021
## Quick slide reference

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>27</td>
</tr>
<tr>
<td>48</td>
</tr>
</tbody>
</table>
Expectation of Common RVs
Linearity of Expectation is useful

Expectation is a linear mathematical operation. If \( X = \sum_{i=1}^{n} X_i \) :

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

- Even if you don’t know the distribution of \( X \) (e.g., because the joint distribution of \( (X_1, ..., X_n) \) is unknown), you can still compute expectation of \( X \)!!

- Problem-solving key: Define \( X_i \) such that

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

Most common use cases:
- \( E[X_i] \) easy to calculate
- Or sum of dependent RVs
Expectations of common RVs: Binomial

\[ X \sim \text{Bin}(n, p) \quad E[X] = np \]

# of successes in \( n \) independent trials with probability of success \( p \)

Recall: \( \text{Bin}(1, p) = \text{Ber}(p) \)

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

Let \( X_i = \text{ith trial is heads} \quad X_i \sim \text{Ber}(p), E[X_i] = p \)

\[
E[X] = E \left[ \sum_{i=1}^{n} X_i \right] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p = np
\]
Expectations of common RVs: Negative Binomial

\[ Y \sim \text{NegBin}(r, p) \quad E[Y] = \frac{r}{p} \]

# of independent trials with probability of success \( p \) until \( r \) successes

Recall: \( \text{NegBin}(1, p) = \text{Geo}(p) \)

\[ Y = \sum_{i=1}^{?} Y_i \]

1. How should we define \( Y_i \)?

2. How many terms are in our summation?
Expectations of common RVs: Negative Binomial

\( Y \sim \text{NegBin}(r, p) \quad E[Y] = \frac{r}{p} \) # of independent trials with probability of success \( p \) until \( r \) successes

Recall: \( \text{NegBin}(1, p) = \text{Geo}(p) \)

\[
Y = \sum_{i=1}^{?} Y_i
\]

Let \( Y_i = \# \text{ trials to get } i\text{th success (after} \ \ \ (i-1)\text{th success)} \)

\( Y_i \sim \text{Geo}(p), \ E[Y_i] = \frac{1}{p} \)

\[
E[Y] = E \left[ \sum_{i=1}^{r} Y_i \right] = \sum_{i=1}^{r} E[Y_i] = \sum_{i=1}^{r} \frac{1}{p} = \frac{r}{p}
\]
Coupon Collecting Problems
Linearity of Expectation is useful

Expectation is a linear mathematical operation. If \( X = \sum_{i=1}^{n} X_i \):

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

- Even if you don’t know the distribution of \( X \) (e.g., because the joint distribution of \( (X_1, \ldots, X_n) \) is unknown), you can still compute expectation of the sum!!

- Problem-solving key: Define \( X_i \) such that

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

Most common use cases:
- \( E[X_i] \) easy to calculate
- Or sum of dependent RVs
Coupon collecting problems: Server requests

The **coupon collector’s problem** in probability theory:

- You buy boxes of cereal.
- There are $k$ different types of coupons
- For each box you buy, you ”collect” a coupon of type $i$.

1. How many coupons do you expect after buying $n$ boxes of cereal?

What is the expected number of utilized servers after $n$ requests?

* 52% of Amazon profits
** more profitable than Amazon’s North America commerce operations

source
Computer cluster utilization

Consider a computer cluster with $k$ servers. We send $n$ requests.

- Requests independently go to server $i$ with probability $p_i$
- Let $X = \#$ servers that receive $\geq 1$ request.

What is $E[X]$?
Computer cluster utilization

Consider a computer cluster with \( k \) servers. We send \( n \) requests.

- Requests independently go to server \( i \) with probability \( p_i \)
- Let \( X = \# \) servers that receive \( \geq 1 \) request.

What is \( E[X] \)?

1. Define additional random variables.
   
   Let: \( A_i = \) event that server \( i \) receives \( \geq 1 \) request
   \( X_i = \) indicator for \( A_i \)

   \[
P(A_i) = 1 - P(\text{no requests to } i) = 1 - (1 - p_i)^n
   \]

   Note: \( A_i \) are dependent!

2. Solve.
   
   \[
   E[X_i] = P(A_i) = 1 - (1 - p_i)^n
   \]
   \[
   E[X] = E\left[\sum_{i=1}^{k} X_i \right] = \sum_{i=1}^{k} E[X_i] = \sum_{i=1}^{k} \left(1 - (1 - p_i)^n\right)
   \]
   \[
   = \sum_{i=1}^{k} 1 - \sum_{i=1}^{k} (1 - p_i)^n = k - \sum_{i=1}^{k} (1 - p_i)^n
   \]
The coupon collector’s problem in probability theory:

- You buy boxes of cereal.
- There are $k$ different types of coupons
- For each box you buy, you ”collect” a coupon of type $i$.

1. How many coupons do you expect after buying $n$ boxes of cereal?

2. How many boxes do you expect to buy until you have one of each coupon?

What is the expected number of utilized servers after $n$ requests?

What is the expected number of strings to hash until each bucket has $\geq 1$ string?

Stay tuned for live lecture!
Covariance
Statistics of sums of RVs

For any random variables $X$ and $Y$,

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

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

But first... a new statistic!
Spot the difference

Compare/contrast the following two distributions:

Both distributions have the same $E[X]$, $E[Y]$, $\text{Var}(X)$, and $\text{Var}(Y)$

Difference: how the two variables vary with each other.

Assume all points are equally likely.

$P(X = x, Y = y) = \frac{1}{N}$
Covariance

The **covariance** of two variables $X$ and $Y$ is:

\[
\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] \\
= E[XY] - E[X]E[Y]
\]

Proof of second part:

\[
\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] \\
= E[XY] - E[X]E[Y]
\]

(linearity of expectation)

($E[X], E[Y]$ are scalars)
Covarying humans

What is the covariance of weight $W$ and height $H$?

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]
= E[XY] - E[X]E[Y]$$

<table>
<thead>
<tr>
<th>Weight (kg)</th>
<th>Height (in)</th>
<th>$W \cdot H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>57</td>
<td>3648</td>
</tr>
<tr>
<td>71</td>
<td>59</td>
<td>4189</td>
</tr>
<tr>
<td>53</td>
<td>49</td>
<td>2597</td>
</tr>
<tr>
<td>67</td>
<td>62</td>
<td>4154</td>
</tr>
<tr>
<td>55</td>
<td>51</td>
<td>2805</td>
</tr>
<tr>
<td>58</td>
<td>50</td>
<td>2900</td>
</tr>
<tr>
<td>77</td>
<td>55</td>
<td>4235</td>
</tr>
<tr>
<td>57</td>
<td>48</td>
<td>2736</td>
</tr>
<tr>
<td>56</td>
<td>42</td>
<td>2352</td>
</tr>
<tr>
<td>51</td>
<td>42</td>
<td>2142</td>
</tr>
<tr>
<td>76</td>
<td>61</td>
<td>4636</td>
</tr>
<tr>
<td>68</td>
<td>57</td>
<td>3876</td>
</tr>
</tbody>
</table>

$E[W] = 62.75$  
$E[H] = 52.75$  
$E[WH] = 3355.83$

Covariance > 0: one variable ↑, other variable ↑
Properties of Covariance

The covariance of two variables $X$ and $Y$ is:

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$
$$= E[XY] - E[X]E[Y]$$

Properties:

1. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
2. $\text{Var}(X) = E[X^2] - (E[X])^2 = \text{Cov}(X, X)$
3. Covariance of sums = sum of all pairwise covariances
   $$\text{Cov}(X_1 + X_2, Y_1 + Y_2) = \text{Cov}(X_1, Y_1) + \text{Cov}(X_2, Y_1) + \text{Cov}(X_1, Y_2) + \text{Cov}(X_2, Y_2)$$
4. Non-linearity (to be discussed in live lecture)

(proof left to you)
Variance of sums of RVs
Statistics of sums of RVs

For any random variables $X$ and $Y$, 

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

\[ \text{Var}(X + Y) = \text{Var}(X) + 2 \cdot \text{Cov}(X, Y) + \text{Var}(Y) \]
Variance of general sum of RVs

For any random variables $X$ and $Y$,

$$\text{Var}(X + Y) = \text{Var}(X) + 2 \cdot \text{Cov}(X, Y) + \text{Var}(Y)$$

Proof:

$$\text{Var}(X + Y) = \text{Cov}(X + Y, X + Y)$$

$$= \text{Cov}(X, X) + \text{Cov}(X, Y) + \text{Cov}(Y, X) + \text{Cov}(Y, Y)$$

$$= \text{Var}(X) + 2 \cdot \text{Cov}(X, Y) + \text{Var}(Y)$$

More generally:

$$\text{Var} \left( \sum_{i=1}^{n} X_i \right) = \sum_{i=1}^{n} \text{Var}(X_i) + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} \text{Cov}(X_i, X_j)$$

(proof in extra slides)
Statistics of sums of RVs

For any random variables $X$ and $Y$,

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

\[
\text{Var}(X + Y) = \text{Var}(X) + 2 \cdot \text{Cov}(X, Y) + \text{Var}(Y)
\]

For independent $X$ and $Y$,

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

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

(Lemma: proof in extra slides)
Variance of sum of independent RVs

For independent $X$ and $Y$,

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

Proof:


2. $\text{Var}(X + Y) = \text{Var}(X) + 2 \cdot \text{Cov}(X, Y) + \text{Var}(Y) = \text{Var}(X) + \text{Var}(Y)$

NOT bidirectional: $\text{Cov}(X, Y) = 0$ does NOT imply independence of $X$ and $Y$!
To simplify the algebra a bit, let $q = 1 - p$, so $p + q = 1$.

Let's instead prove this using independence and variance!

Let $X \sim \text{Bin}(n, p)$.

**Var(X) = np(1 - p)**

Let $Y \sim \text{Bin}(n', p)$.

For some independent $X$ and $Y$,

**Var(X + Y) = Var(X) + Var(Y)**

Definition of Binomial Distribution: $p + q = 1$

Factors of Binomial Coefficient: $\binom{n}{k} - \binom{n}{k-1}$

Change of limit: term is zero when $k = 0$

Putting $j = 1 - k$, $m = n - 1$

Splitting sum into two

Factors of Binomial Coefficient: $\binom{n}{m} = \binom{n}{n-m}$

Change of limit: term is zero when $j = 0$

Binomial Theorem

so $p + q = 1$

by algebra

Then:

**Var(X) = E(X^2) - [E(X)]^2**

**E(X^2) = E(X) \cdot E[X^2]**

**E(X) = np**

**E[X^2] = np + np^2**

as required.
Proving Variance of the Binomial

\[ X \sim \text{Bin}(n, p) \quad \text{Var}(X) = np(1 - p) \]

Let \[ X = \sum_{i=1}^{n} X_i \]

Let \( X_i = \) \( i \)th trial is heads \( X_i \sim \text{Ber}(p) \)
\[ \text{Var}(X_i) = p(1 - p) \]

\( X_i \) are independent (by definition)

\[ \text{Var}(X) = \text{Var} \left( \sum_{i=1}^{n} X_i \right) \]
\[ = \sum_{i=1}^{n} \text{Var}(X_i) \]
\[ = \sum_{i=1}^{n} p(1 - p) \]
\[ = np(1 - p) \]

\( X_i \) are independent, therefore variance of sum = sum of variance

Variance of Bernoulli