Alice & Bob are gambling (again). $X = $ Alice’s gain per flip:

$$X = \begin{cases} +1 & \text{if Heads} \\ -1 & \text{if Tails} \end{cases}$$

$E[X] = 0$

... Time passes ...

Alice (yawning) says “let’s raise the stakes”

$$Y = \begin{cases} +1000 & \text{if Heads} \\ -1000 & \text{if Tails} \end{cases}$$

$E[Y] = 0$, as before.

Are you (Bob) equally happy to play the new game?
E[X] measures the “average” or “central tendency” of X. What about its variability?

If E[X] = μ, then E[|x-μ|] seems like a natural quantity to look at: how much do we expect X to deviate from its average. Unfortunately, it’s a bit inconvenient mathematically; following is easier/more common.

**Definition**

The variance of a random variable X with mean E[X] = μ is Var[X] = E[(X-μ)^2], often denoted σ^2.

The standard deviation of X is σ = \sqrt{Var[X]}
what does variance tell us?

The *variance* of a random variable $X$ with mean $E[X] = \mu$ is $\text{Var}[X] = E[(X-\mu)^2]$, often denoted $\sigma^2$.

I:

Square always $\geq 0$, and exaggerated as $X$ moves away from $\mu$, so $\text{Var}[X]$ emphasizes *deviation* from the mean.

II:

Numbers vary a lot depending on exact distribution of $X$, but typically $X$ is

- within $\mu \pm \sigma$ $\sim 66\%$ of the time, and
- within $\mu \pm 2\sigma$ $\sim 95\%$ of the time.

(We’ll see the reasons for this soon.)
\( \mu = E[X] \) is about location; \( \sigma = \sqrt{\text{Var}(X)} \) is about spread
Alice & Bob are gambling (again). $X = \text{Alice's gain per flip}$:

$$X = \begin{cases} 
+1 & \text{if Heads} \\
-1 & \text{if Tails}
\end{cases}$$

$E[X] = 0$ \hspace{1cm} Var$[X] = 1$

\ldots \text{Time passes} \ldots$

Alice (yawning) says “let’s raise the stakes”

$$Y = \begin{cases} 
+1000 & \text{if Heads} \\
-1000 & \text{if Tails}
\end{cases}$$

$E[Y] = 0$, as before. \hspace{1cm} Var$[Y] = 1,000,000$

Are you (Bob) equally happy to play the new game?
Two games:

a) flip 1 coin, win $Y = $100 if heads, $-100$ if tails

b) flip 100 coins, win $Z = (\#(\text{heads}) - \#(\text{tails}))$ dollars

Same expectation in both: $E[Y] = E[Z] = 0$

Same extremes in both: max gain = $100; max loss = $100

But variability is very different:
\[ \text{Var}(X) = E[X^2] - (E[X])^2 \]
Example:

What is $\text{Var}[X]$ when $X$ is outcome of one fair die?

$$E[X^2] = 1^2 \left( \frac{1}{6} \right) + 2^2 \left( \frac{1}{6} \right) + 3^2 \left( \frac{1}{6} \right) + 4^2 \left( \frac{1}{6} \right) + 5^2 \left( \frac{1}{6} \right) + 6^2 \left( \frac{1}{6} \right)$$

$$= \left( \frac{1}{6} \right) (91)$$

$E[X] = 7/2$, so

$$\text{Var}(X) = \frac{91}{6} - \left( \frac{7}{2} \right)^2 = \frac{35}{12}$$
properties of variance

$$\text{Var}[aX+b] = a^2 \text{Var}[X]$$

\[
\begin{align*}
\text{Var}(aX + b) &= E[(aX + b - a\mu - b)^2] \\
&= E[a^2(X - \mu)^2] \\
&= a^2 E[(X - \mu)^2] \\
&= a^2 \text{Var}(X)
\end{align*}
\]

Ex:

\[X = \begin{cases} 
+1 & \text{if Heads} \\
-1 & \text{if Tails} 
\end{cases}\]

\[\text{E}[X] = 0 \quad \text{Var}[X] = 1\]

\[Y = \begin{cases} 
+1000 & \text{if Heads} \\
-1000 & \text{if Tails} 
\end{cases}\]

\[\text{E}[Y] = \text{E}[1000 X] = 1000 \text{ E}[x] = 0 \]

\[\text{Var}[Y] = \text{Var}[1000 X] = 10^6 \text{Var}[X] = 10^6\]
properties of variance

In general:

$$\text{Var}[X+Y] \neq \text{Var}[X] + \text{Var}[Y]$$

Ex 1:

Let $X = \pm 1$ based on 1 coin flip

As shown above, $E[X] = 0, \text{Var}[X] = 1$

Let $Y = -X$; then $\text{Var}[Y] = (-1)^2 \text{Var}[X] = 1$

But $X+Y = 0$, always, so $\text{Var}[X+Y] = 0$

Ex 2:

As another example, is $\text{Var}[X+X] = 2\text{Var}[X]$?
a zoo of (discrete) random variables
Bernoulli random variables

An experiment results in “Success” or “Failure”

$X$ is a random indicator variable ($1 =$ success, $0 =$ failure)

$P(X=1) = p$ and $P(X=0) = 1-p$

$X$ is called a **Bernoulli** random variable: $X \sim \text{Ber}(p)$

$E[X] = E[X^2] = p$

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

Examples:

- coin flip
- random binary digit
- whether a disk drive crashed
Consider \( n \) independent random variables \( Y_i \sim \text{Ber}(p) \)

\( X = \sum_i Y_i \) is the number of successes in \( n \) trials

\( X \) is a \textit{Binomial} random variable: \( X \sim \text{Bin}(n,p) \)

\[
P(X = i) = \binom{n}{i} p^i (1 - p)^{n-i} \quad i = 0, 1, \ldots, n
\]

By Binomial theorem,
\[
\sum_{i=0}^{n} P(X = i) = 1
\]

Examples

\# of heads in \( n \) coin flips
\# of 1’s in a randomly generated length \( n \) bit string
\# of disk drive crashes in a 1000 computer cluster

\[
E[X] = pn
\]

\[
\text{Var}(X) = p(1-p)n
\]  

\((\text{proof below, twice})\)
binomial pmfs

PMF for $X \sim \text{Bin}(10, 0.5)$

PMF for $X \sim \text{Bin}(10, 0.25)$

$k$

$P(X=k)$

$\mu \pm \sigma$
binomial pmfs

PMF for $X \sim \text{Bin}(30, 0.5)$

PMF for $X \sim \text{Bin}(30, 0.1)$
mean and variance of the binomial

\[
E[X^k] = \sum_{i=0}^{n} i^k \binom{n}{i} p^i (1 - p)^{n-i}
\]
\[
= \sum_{i=1}^{n} i^k \binom{n}{i} p^i (1 - p)^{n-i}
\]
\[
E[X^k] = np \sum_{i=1}^{n} i^{k-1} \binom{n-1}{i-1} p^{i-1} (1 - p)^{n-i}
\]
\[
= np \sum_{j=0}^{n-1} (j + 1)^{k-1} \binom{n-1}{j} p^{j} (1 - p)^{n-1-j}
\]
\[
= np E[(Y + 1)^{k-1}]
\]

where \(Y\) is a binomial random variable with parameters \(n - 1, p\).

\(k=1\) gives: \(E[X] = np\); \(k=2\) gives \(E[X^2] = np[(n-1)p+1]\)

hence:
\[
Var(X) = E[X^2] - (E[X])^2
\]
\[
= np[(n - 1)p + 1] - (np)^2
\]
\[
= np(1 - p)
\]
products of independent r.v.s

Theorem: If $X$ & $Y$ are independent, then $E[X \cdot Y] = E[X] \cdot E[Y]$

Proof:

Let $x_i, y_i, i = 1, 2, \ldots$ be the possible values of $X, Y$.

$$E[X \cdot Y] = \sum_i \sum_j x_i \cdot y_j \cdot P(X = x_i \land Y = y_j)$$

$$= \sum_i \sum_j x_i \cdot y_j \cdot P(X = x_i) \cdot P(Y = y_j)$$

$$= \sum_i x_i \cdot P(X = x_i) \cdot \left( \sum_j y_j \cdot P(Y = y_j) \right)$$

$$= E[X] \cdot E[Y]$$

Note: NOT true in general; see earlier example $E[X^2] \neq E[X]^2$
Theorem: If $X$ & $Y$ are independent, then

$$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$$

Proof: Let $\hat{X} = X - E[X]$ and $\hat{Y} = Y - E[Y]$

$E[\hat{X}] = 0$ and $E[\hat{Y}] = 0$

$\text{Var}[\hat{X}] = \text{Var}[X]$ and $\text{Var}[\hat{Y}] = \text{Var}[Y]$

$\text{Var}[X + Y] = \text{Var}[\hat{X} + \hat{Y}]$

$= E[(\hat{X} + \hat{Y})^2] - (E[\hat{X} + \hat{Y}])^2$

$= E[\hat{X}^2 + 2\hat{X}\hat{Y} + \hat{Y}^2] - 0$

$= E[\hat{X}^2] + 2E[\hat{X}\hat{Y}] + E[\hat{Y}^2]$

$= \text{Var}[\hat{X}] + 0 + \text{Var}[\hat{Y}]$

$= \text{Var}[X] + \text{Var}[Y]$

(Bienaymé, 1853)
mean, variance of binomial r.v.s

If $Y_1, Y_2, \ldots, Y_n \sim \text{Ber}(p)$ and independent,

then $X = \sum_{i=1}^{n} Y_i \sim \text{Bin}(n, p)$.

$E[X] = E[\sum_{i=1}^{n} Y_i] = nE[Y_1] = np$

$\text{Var}[X] = \text{Var}[\sum_{i=1}^{n} Y_i] = n\text{Var}[Y_1] = np(1 - p)$
A RAID-like disk array consists of $n$ drives, each of which will fail independently with probability $p$. Suppose it can operate effectively if at least one-half of its components function, e.g., by “majority vote.” For what values of $p$ is a 5-component system more likely to operate effectively than a 3-component system?

$X_5 = \# \text{ failed in 5-component system} \sim \text{Bin}(5, p)$
$X_3 = \# \text{ failed in 3-component system} \sim \text{Bin}(3, p)$
$X_5 = \# \text{ failed in 5-component system } \sim Bin(5, p)$

$X_3 = \# \text{ failed in 3-component system } \sim Bin(3, p)$

$P(5 \text{ component system effective}) = P(X_5 < 5/2)$

\[
\binom{5}{0} p^0 (1 - p)^5 + \binom{5}{1} p^1 (1 - p)^4 + \binom{5}{2} p^2 (1 - p)^3
\]

$P(3 \text{ component system effective}) = P(X_3 < 3/2)$

\[
\binom{3}{0} p^0 (1 - p)^3 + \binom{3}{1} p^1 (1 - p)^2
\]

**Calculation:**

5-component system is better iff $p < 1/2$
Goal: send a 4-bit message over a noisy communication channel.

Say, 1 bit in 10 is flipped in transit, independently.

What is the probability that the message arrives correctly?

Let $X = \#$ of errors; $X \sim \text{Bin}(4, 0.1)$

$P(\text{correct message received}) = P(X=0)$

$$P(X = 0) = \binom{4}{0}(0.1)^0(0.9)^4 = 0.6561$$

Can we do better? Yes: error correction via redundancy.

E.g., send every bit in triplicate; use majority vote.

Let $Y = \#$ of errors in one trio; $Y \sim \text{Bin}(3, 0.1)$; $P(\text{a trio is OK}) =$

$$P(Y \leq 1) = \binom{3}{0}(0.1)^0(0.9)^3 + \binom{3}{1}(0.1)^1(0.9)^2 = 0.972$$

If $X' = \#$ errors in triplicate msg, $X' \sim \text{Bin}(4, 0.028)$, and

$$P(X' = 0) = \binom{4}{0}(0.028)^0(0.972)^4 = 0.8926168$$
The Hamming(7,4) code:
Have a 4-bit string to send over the network (or to disk)
Add 3 “parity” bits, and send 7 bits total
If bits are $b_1b_2b_3b_4$ then the three parity bits are
$$\text{parity}(b_1b_2b_3), \text{parity}(b_1b_3b_4), \text{parity}(b_2b_3b_4)$$
Each bit is independently corrupted (flipped) in transit with probability 0.1
$$Z = \text{number of bits corrupted} \sim \text{Bin}(7, 0.1)$$
The Hamming code allow us to correct all 1 bit errors.
(E.g., if $b_1$ flipped, 1st 2 parity bits, but not 3rd, will look wrong; the only single bit error causing this symptom is $b_1$. Similarly for any other single bit being flipped. Some, but not all, multi-bit errors can be detected, but not corrected.)
$$P(\text{correctable message received}) = P(Z \leq 1)$$
Using Hamming error-correcting codes: $Z \sim \text{Bin}(7, 0.1)$

$$P(Z \leq 1) = \binom{7}{0} (0.1)^0 (0.9)^7 + \binom{7}{1} (0.1)^1 (0.9)^6 \approx 0.8503$$

Recall, uncorrected success rate is

$$P(X = 0) = \binom{4}{0} (0.1)^0 (0.9)^4 = 0.6561$$

And triplicate code error rate is:

$$P(X' = 0) = \binom{4}{0} (0.028)^0 (0.972)^4 = 0.8926168$$

Hamming code is nearly as reliable as the triplicate code, with $5/12 \approx 42\%$ fewer bits. (& better with longer codes.)