Independence, Variance, Bayes’ Theorem

Russell Impagliazzo and Miles Jones
Thanks to Janine Tiefenbruck

http://cseweb.ucsd.edu/classes/sp16/cse21-bd/
May 16, 2016
Resolving collisions with chaining

Hash Table

Each memory location holds a pointer to a linked list, initially empty.

Each linked list records the items that map to that memory location.

Collision means there is more than one item in this linked list.
Element Distinctness: HOW

Given list of positive integers $A = a_1, a_2, \ldots, a_n$, and $m$ memory locations available

$\text{ChainHashDistinctness}(A, m)$
1. Initialize array $M[1,\ldots,m]$ to null lists.
2. Pick a hash function $h$ from all positive integers to $1,\ldots,m$.
3. For $i = 1$ to $n$,
4. For each element $j$ in $M[h(a_i)]$,
5. If $a_j = a_i$ then return "Found repeat"
6. Append $i$ to the tail of the list $M[h(a_i)]$
7. Return "Distinct elements"
Element Distinctness: WHY

Given list of positive integers $A = a_1, a_2, \ldots, a_n$, and $m$ memory locations available

ChainHashDistinctness($A$, $m$)
1. Initialize array $M[1,\ldots,m]$ to null lists.
2. Pick a hash function $h$ from all positive integers to $1,\ldots,m$.
3. For $i = 1$ to $n$,
4. For each element $j$ in $M[h(a_i)]$,
5. If $a_j = a_i$ then return "Found repeat"
6. Append $i$ to the tail of the list $M[h(a_i)]$
7. Return "Distinct elements"

Correctness: Goal is
If there is a repetition, algorithm finds it ✔️
If there is no repetition, algorithm reports "Distinct elements" ✔️
Element Distinctness: MEMORY

Given list of positive integers $A = a_1, a_2, \ldots, a_n$, and $m$ memory locations available

**ChainHashDistinctness**($A$, $m$)
1. Initialize array $M[1,\ldots,m]$ to null lists.
2. Pick a hash function $h$ from all positive integers to $1,\ldots,m$.
3. For $i = 1$ to $n$,
   4. For each element $j$ in $M[h(a_i)]$,
      5. If $a_j = a_i$ then return "Found repeat"
   6. Append $i$ to the tail of the list $M[h(a_i)]$
4. Return "Distinct elements"

*What's the memory use of this algorithm?*
Element Distinctness: MEMORY

Given list of distinct integers $A = a_1, a_2, \ldots, a_n$, and $m$ memory locations available

**ChainHashDistinctness**($A$, $m$)
1. Initialize array $M[1,\ldots,m]$ to null lists.
2. Pick a hash function $h$ from all positive integers to $1,\ldots,m$.
3. For $i = 1$ to $n$,
4. For each element $j$ in $M[h(a_i)]$,
5. If $a_j = a_i$ then return "Found repeat"
6. Append $i$ to the tail of the list $M[h(a_i)]$
7. Return "Distinct elements"

*What's the memory use of this algorithm?*
Size of $M$: $O(m)$. Total size of all the linked lists: $O(n)$. Total memory: $O(m+n)$. 
Element Distinctness: WHEN

\[ \text{ChainHashDistinctness}(A, m) \]

1. Initialize array \( M[1,..,m] \) to null lists.  \( \Theta(1) \)
2. Pick a hash function \( h \) from all positive integers to \( 1,..,m \).
3. For \( i = 1 \) to \( n \),
4. \hspace{1em} For each element \( j \) in \( M[\ h(a_i) \] \),
5. \hspace{2em} If \( a_j = a_i \) then return "Found repeat"
6. \hspace{1em} Append \( i \) to the tail of the list \( M[\ h(a_i) \] \)
7. Return "Distinct elements" \( \Theta(1) \)
Element Distinctness: WHEN

\text{ChainHashDistinctness}(A, m)
1. Initialize array \text{M}[1,..,m] to null lists.
2. Pick a hash function \( h \) from all positive integers to 1,..,m.
3. For \( i = 1 \) to \( n \),
4. For each element \( j \) in \( \text{M}[ h(a_i) ] \),
5. If \( a_j = a_i \) then return "Found repeat"
6. Append \( i \) to the tail of the list \( \text{M}[ h(a_i) ] \)
7. Return "Distinct elements"

Worst case is when we don't find \( a_i \):
\[ O(1 + \text{size of list } \text{M}[ h(a_i) ] ) \]
ChainHashDistinctness(A, m)
1. Initialize array M[1,..,m] to null lists.
2. Pick a hash function $h$ from all positive integers to 1,..,m.
3. For i = 1 to n,
4. For each element j in M[ h(a_i) ],
5. If $a_j = a_i$ then return "Found repeat"
6. Append i to the tail of the list M [ h(a_i) ]
7. Return "Distinct elements"

Worst case is when we don't find $a_i$:
\[ O( 1 + \text{size of list } M[ h(a_i) ] ) \]
\[ = O( 1 + \# j<i \text{ with } h(a_j)=h(a_i) ) \]
**Element Distinctness: WHEN**

ChainHashDistinctness(A, m)
1. Initialize array M[1,..,m] to null lists.
2. Pick a hash function \( h \) from all positive integers to 1,..,m.
3. For \( i = 1 \) to \( n \),
4.    For each element \( j \) in \( M[ h(a_i) ] \),
5.       If \( a_j = a_i \) then return "Found repeat"
6.    Append \( i \) to the tail of the list \( M[ h(a_i) ] \)
7. Return "Distinct elements"

**Total time:** \( O(n + \sum_{i=1}^{n} \text{# collisions between pairs } a_i \text{ and } a_j, \text{ where } j<i ) ) \)

\[ = O(n + \text{total # collisions}) \]
Element Distinctness: WHEN

**Total time:** $O(n + \sum_{i=1}^{n} \text{# collisions between pairs } a_i \text{ and } a_j, \text{ where } j<i )$

$= O(n + \text{total # collisions})$

*What's the expected total number of collisions?*
Element Distinctness: WHEN

Total time: $O(n + \sum_{i=1}^{n} \# \text{collisions between pairs } a_i \text{ and } a_j, \text{ where } j<i)$

$= O(n + \text{total } \# \text{collisions})$

What's the expected total number of collisions?

For each pair (i,j) with j<i, define: $X_{i,j} = 1$ if $h(a_i)=h(a_j)$ and $X_{i,j}=0$ otherwise.

Total # of collisions = $\sum_{(i,j):j<i} X_{i,j}$
Element Distinctness: WHEN

**Total time**: $O(n + \sum_{i=1}^{n} \# \text{collisions between pairs } a_i \text{ and } a_j, \text{ where } j<i)$

$= O(n + \text{total # collisions})$

**What's the expected total number of collisions?**

For each pair (i,j) with j<i, define:

$X_{i,j} = 1$ if $h(a_i)=h(a_j)$ and $X_{i,j}=0$ otherwise.

Total # of collisions $= \sum_{(i,j): j<i} X_{i,j}$

So by **linearity of expectation**: $E(\text{total # of collisions}) = \sum_{(i,j): j<i} E(X_{i,j})$
Element Distinctness: WHEN

**Total time:** $O(n + \sum_{i=1}^{n} \# \text{collisions between pairs } a_i \text{ and } a_j, \text{ where } j<i)$

$= O(n + \text{total # collisions})$

*What's the expected total number of collisions?*

For each pair $(i,j)$ with $j<i$, define:

$X_{i,j} = 1$ if $h(a_i)=h(a_j)$ and $X_{i,j}=0$ otherwise.

Total # of collisions $= \sum_{(i,j) : j<i} X_{i,j}$

*What's $E(X_{i,j})$?*

A. $1/n$
B. $1/m$
C. $1/n^2$
D. $1/m^2$
E. None of the above.
Element Distinctness: WHEN

**Total time:** \( O(n + \sum_{i=1}^{n} \# \text{collisions between pairs } a_i \text{ and } a_j, \text{ where } j<i ) \)

\[= O(n + \text{total # collisions}) \]

**What's the expected total number of collisions?**

For each pair \((i,j)\) with \(j<i\), define:

\[X_{i,j} = 1 \text{ if } h(a_i) = h(a_j) \text{ and } X_{i,j} = 0 \text{ otherwise.} \]

Total # of collisions = \( \sum_{(i,j):j<i} X_{i,j} \)

How many terms are in the sum? That is, how many pairs \((i,j)\) with \(j<i\) are there?

A. \(n\)
B. \(n^2\)
C. \(C(n,2)\)
D. \(n(n-1)\)
Element Distinctness: WHEN

Total time: $O(n + \sum_{i=1}^{n} \# \text{collisions between pairs } a_i \text{ and } a_j, \text{ where } j<i)$

$= O(n + \text{total # collisions})$

What's the expected total number of collisions?

For each pair $(i,j)$ with $j<i$, define: $X_{i,j} = 1$ if $h(a_i)=h(a_j)$ and $X_{i,j}=0$ otherwise.

So by linearity of expectation:

$E(\text{total # of collisions}) = \sum_{(i,j): j<i} E(X_{i,j}) = \binom{n}{2} \frac{1}{m} = O(n^2/m)$
Element Distinctness: WHEN

**Total time:** $O(n + \sum_{i=1}^{n} \# \text{collisions between pairs } a_i \text{ and } a_j, \text{ where } j<i)$

\[
= O(n + \text{total } \# \text{ collisions})
\]

**Total expected time:** $O(n + n^2/m)$

In ideal hash model, as long as $m>n$ the total expected time is $O(n)$. 
Independent Events

Two events E and F are **independent** iff \( P( E \cap F ) = P(E) \cdot P(F) \).

**Problem:** Suppose

- E is the event that a randomly generated bitstring of length 4 starts with a 1
- F is the event that this bitstring contains an even number of 1s.

Are E and F independent if all bitstrings of length 4 are equally likely? Are they disjoint?

**First impressions?**

A. E and F are independent and disjoint.
B. E and F are independent but not disjoint.
C. E and F are disjoint but not independent.
D. E and F are neither disjoint nor independent.
Two events $E$ and $F$ are independent iff \( P( E \cap F ) = P(E) P(F) \).

**Problem:** Suppose

- $E$ is the event that a randomly generated bitstring of length 4 starts with a 1
- $F$ is the event that this bitstring contains an even number of 1s.

Are $E$ and $F$ independent if all bitstrings of length 4 are equally likely? Are they disjoint?
Independent Random Variables

Let $X$ and $Y$ be random variables over the same sample space. $X$ and $Y$ are called independent random variables if, for all possible values of $v$ and $u$,

$$P ( X = v \text{ and } Y = u ) = P ( X = v ) P( Y = u)$$

Which of the following pairs of random variables on sample space of sequences of H/T when coin is flipped four times are independent?

A. $X_{12} = \# \text{ of } H \text{ in first two flips, } X_{34} = \# \text{ of } H \text{ in last two flips.}$
B. $X = \# \text{ of } H \text{ in the sequence, } Y = \# \text{ of } T \text{ in the sequence.}$
C. $X_{12} = \# \text{ of } H \text{ in first two flips, } X = \# \text{ of } H \text{ in the sequence.}$
D. None of the above.
Independence

Theorem:
If X and Y are independent random variables over the same sample space, then

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

Note: This is not necessarily true if the random variables are not independent!
Say that you are conducting a poll about whether people use your product. You’ll ask random people whether they use your product, a yes/no question. (Assume your samples are random, even given the person actually answers, that people answer honestly and know whether they use the product...) How many people do you need to ask before you get, with high probability, a close estimate for the actual fraction of the population that use it?
How close (on average) will we be to the average / expected value?

Let \( X \) be a random variable with \( E(X) = E \).

The **unexpectedness** of \( X \) is the random variable

\[
U = |X - E|
\]

The **average unexpectedness** of \( X \) is

\[
AU(X) = E(|X - E|) = E(U)
\]

The **variance** of \( X \) is

\[
V(X) = E(|X - E|^2) = E(U^2)
\]

The **standard deviation** of \( X \) is

\[
\sigma(X) = (E(|X - E|^2))^{1/2} = V(X)^{1/2}
\]
Concentration

How close (on average) will we be to the average / expected value?

Let $X$ be a random variable with $E(X) = E$.

The **variance** of $X$ is

$$V(X) = E(|X - E|^2) = E(U^2)$$

**Example:** $X_1$ is a random variable with distribution

- $P(X_1 = -2) = 1/5$, $P(X_1 = -1) = 1/5$, $P(X_1 = 0) = 1/5$, $P(X_1 = 1) = 1/5$, $P(X_1 = 2) = 1/5$. $X_2$ is a random variable with distribution

- $P(X_2 = -2) = 1/2$, $P(X_2 = 2) = 1/2$.

Which is true?

A. $E(X_1) \neq E(X_2)$
B. $V(X_1) < V(X_2)$
C. $V(X_1) > V(X_2)$
D. $V(X_1) = V(X_2)$
$X_1$ is a random variable with distribution
\[ P( X_1 = -2 ) = 1/5, \ P( X_1 = -1 ) = 1/5, \ P( X_1 = 0 ) = 1/5, \ P( X_1 = 1 ) = 1/5, \ P( X_1 = 2 ) = 1/5. \]

$E(X_1) =$

$U$ is distributed according to:

$U^2$ is distributed according to:

$AU =$

$V =$
$X_2$ is a random variable with distribution

$P( X_2 = -2 ) = 1/2, \ P( X_2 = 2 ) = \frac{1}{2}$.

$E(X_2)=$

$U$ is distributed according to:

$U^2$ is distributed according to:

$AU=$

$V=$
How close (on average) will we be to the average / expected value?

Let \( X \) be a random variable with \( E(X) = E \).

The \textbf{unexpectedness} of \( X \) is the random variable

\[
U = |X - E|
\]

The \textbf{average unexpectedness} of \( X \) is

\[
AU(X) = E\left(|X - E|\right) = E(U)
\]

The \textbf{variance} of \( X \) is

\[
V(X) = E\left(|X - E|^2\right) = E(U^2)
\]

The \textbf{standard deviation} of \( X \) is

\[
\sigma(X) = \left(E\left(|X - E|^2\right)\right)^{1/2} = V(X)^{1/2}
\]

Weight all differences from mean equally

Weight large differences from mean more
Concentration

How close (on average) will we be to the average / expected value?

Let \( X \) be a random variable with \( E(X) = E \).

The variance of \( X \) is

\[
V(X) = E( |X - E|^2 ) = E( U^2 )
\]

**Theorem:** \( V(X) = E(X^2) - ( E(X) )^2 \)
Concentration

How close (on average) will we be to the average / expected value?
Let $X$ be a random variable with $E(X) = E$.
The variance of $X$ is

$$V(X) = E(|X - E|^2) = E(U^2)$$

**Theorem:** $V(X) = E(X^2) - (E(X))^2$

**Proof:**

$$V(X) = E((X-E)^2) = E(X^2 - 2XE + E^2) = E(X^2) - 2E(E(X)) + E^2$$

$$= E(X^2) - 2E^2 + E^2$$

$$= E(X^2) - (E(X))^2 

\smile$$

Linearity of expectation
Some consequences

Note that this means that the only time when $E[X^2] = (E[X])^2$ is when the variance is 0.

Since $V[X] = E[U^2]$ and $U^2 \geq 0$, this only happens when when $U = 0$ always, which is when the random variable $X$ is actually a fixed value with no randomness at all.
Some consequences

Also, since \( V[X] \) is the expectation of \( U^2 \geq 0 \), it follows that:

**Theorem**

*For any random variable \( X \), \( E[X^2] \geq (E[X])^2 \).*

We can also apply this to the variable \( U \).

\[
V[X] = E[U^2] \geq (E[U])^2 = (AU[X])^2.
\]

Taking the square root of both sides,

\[
\sigma[X] = (V[X])^{1/2} \geq AU[X].
\]

So the average unexpectedness is never bigger than the standard deviation.
The standard deviation gives us a bound on how far off we are likely to be from the expected value. It is frequently but not always a fairly accurate bound.
Theorem (Additive Property of Variance)

If $X$ and $Y$ are independent random variables,


Proof.
Coming Soon...
Say we have a set of independent random variables $\{X_1, \ldots, X_n\}$ all of which take values between 0 and 1. Then for each of these variables, the maximum unexpectedness is 1, so the maximum variance is 1. By additivity,

$$V[X_1 + \cdots + X_n] = V[X_1] + \cdots + V[X_n] \leq n.$$ 

Then

$$AU[X_1 + \cdots + X_n] \leq \sigma[X_1 + \cdots + X_n]$$

$$= V[X_1 + \cdots + X_n]^{1/2}$$

$$\leq n^{1/2}.$$
Since $\sqrt{n} \in o(n)$, this says sums of independent variables tend to stay close to their expectations.

On average, the unexpectedness becomes a smaller and smaller fraction of the total as the number of random variables $n$ increases. We'll see some other formalisations of this intuition later.
A detour we need to prove additivity of variance is important in its own right.

**Theorem (Multiplicative Property of Expectation)**

If $X$ and $Y$ are independent random variables,

$$E[X \times Y] = E[X] \times E[Y].$$
Theorem (Additive Property of Variance)

If $X$ and $Y$ are independent random variables,


Proof.

Let $X$ and $Y$ be independent random variables. Then,

$$V[X + Y] = E[(X + Y)^2] - (E[X + Y])^2$$

$$= E[(X^2 + 2XY + Y^2)] - (E[X] + E[Y])^2$$


$$+ (E[Y^2] - E[Y]^2)$$

$$= V[X] + V[Y].$$
In the polling example, we want to give some bound on the probability that the result of our poll is very far from its expectation.

Such a result is called a *concentration bound*, because it’s saying the probability distribution is concentrated near its expected value, with only small amounts far away.

Many concentration bounds are known and useful, depending on what we know about the random variable.
Garrison Keillor describes his fictional hometown, Lake Wobegon, as a place where “all the children are above average.”

How many children in a town can actually be above average, say in height? What fraction of children can be twice as tall as average?
Theorem

Let $X$ be a non-negative random variable with expectation $E$. Then $\text{Prob}[X > kE] < 1/k$

Proof: Let $p$ be the probability that $X > kE$. Then the conditional expectation given this event occurs is $> kE$, and the conditional expectation given that it doesn’t is $\geq 0$ since $X$ is non-negative. Thus, $E > p(kE) + (1 - p)0$, or $p < 1/k$. 
Markov’s inquality is pretty obvious, but we can apply it in more subtle ways. First, we can apply it to $U$ rather than $X$, since $U \geq 0$ being an absolute value.

**Theorem**

$\text{Prob}[U \geq k\text{AU}(X)] \leq 1/k$, so $\text{Prob}[U \geq k\sigma[X]] \leq 1/k$.

The first is Markov’s inequality applied to $U$. The second follows since $\text{AU}(X) \leq \sigma[X]$. 
But we can do better. Apply this to $U^2$ rather than $U$. If $U \geq k\sigma[X]$, then $U^2 \geq k^2 V[X]$. Since $V[X]$ is the expected value of $U^2$, Markov’s inequality tells us:

**Theorem**

$$\text{Prob}[U \geq k\sigma[X]] \leq 1/k^2.$$ 

In particular, remember that when we have a sum $S$ of $n$ independent variables in the range $[0, 1]$, $V[S] \leq n$, and $\sigma[S] \leq \sqrt{n}$. So it follows that $\text{Prob}[|S - E[S]| \geq k\sqrt{n}] \leq 1/k^2$. 
Applying this to polls

Remember our original example. We are polling \( n \) people, each answering yes independently with probability \( p \). Let \( S \) be the total number of yes answers. Our empirical estimate will be \( p' = S/n \), and we are worried that \( |p' - p| \geq \epsilon \). Multiplying through by \( n \), this is the same as \( |S - pn| \geq \epsilon n = \epsilon \sqrt{n} \sqrt{n} \).

We have the rule: \( \text{Prob}[|S - E[S]| \geq k \sqrt{n}] \leq 1/k^2 \). So we can apply the above with \( k = \epsilon \sqrt{n} \) to see that the probability of going wrong by an \( \epsilon \) amount is at most \( 1/k^2 = 1/((\epsilon^2)n) \). Reversing this, if we want this probability to be at most \( \delta \), it suffices to sample

\[
    n = \frac{1}{\delta \epsilon^2}
\]
Is this tight?

There are actually stronger concentration bounds which say that the probability of being off from the average drops exponentially rather than polynomially. Even with these stronger bounds, the actual number becomes $\Theta \left( \frac{\log(\frac{1}{\delta})}{\epsilon^2} \right)$ samples.

If you see the results of polling, they almost always give a margin of error which is obtained by plugging in $\delta = 0.01$ and solving for $\epsilon$. 
Recall: Conditional probabilities

Probability of an event may **change** if have additional information about outcomes.

Suppose E and F are events, and \(P(F)>0\). Then,

\[
P(E|F) = \frac{P(E \cap F)}{P(F)}
\]

i.e.

\[
P(E \cap F) = P(E|F)P(F)
\]

*Rosen p. 456*
Based on previous knowledge about how probabilities of two events relate to one another, how does knowing that one event occurred impact the probability that the other did?
Bayes' Theorem: Example 1

Rosen Section 7.3

A manufacturer claims that its drug test will detect steroid use 95% of the time. What the company does not tell you is that 15% of all steroid-free individuals also test positive (the false positive rate). 10% of the Tour de France bike racers use steroids. Your favorite cyclist just tested positive. What’s the probability that he used steroids?

Your first guess?
A. Close to 95%
B. Close to 85%
C. Close to 15%
D. Close to 10%
E. Close to 0%
A manufacturer claims that its drug test will detect steroid use 95% of the time. What the company does not tell you is that 15% of all steroid-free individuals also test positive (the false positive rate). 10% of the Tour de France bike racers use steroids. Your favorite cyclist just tested positive. What’s the probability that he used steroids?

Define events: we want \( P(\text{used steroids} \mid \text{tested positive}) \)
Bayes' Theorem: Example 1

A manufacturer claims that its drug test will detect steroid use 95% of the time. What the company does not tell you is that 15% of all steroid-free individuals also test positive (the false positive rate). 10% of the Tour de France bike racers use steroids. Your favorite cyclist just tested positive. What’s the probability that he used steroids?

Define events: we want \( P(\text{used steroids} | \text{tested positive}) \) so let

\[
P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|\overline{F})P(\overline{F})}
\]

\( E = \text{Tested positive} \)
\( F = \text{Used steroids} \)
Bayes' Theorem: Example 1

Rosen Section 7.3

A manufacturer claims that its drug test will detect steroid use 95% of the time. What the company does not tell you is that 15% of all steroid-free individuals also test positive (the false positive rate). 10% of the Tour de France bike racers use steroids. Your favorite cyclist just tested positive. What’s the probability that he used steroids?

Define events: we want $P(\text{used steroids | tested positive})$

$E = \text{Tested positive}$  \hspace{1cm}  $P( E | F ) = 0.95$

$F = \text{Used steroids}$
Bayes' Theorem: Example 1

A manufacturer claims that its drug test will detect steroid use 95% of the time. What the company does not tell you is that 15% of all steroid-free individuals also test positive (the false positive rate). 10% of the Tour de France bike racers use steroids. Your favorite cyclist just tested positive. What’s the probability that he used steroids?

Define events: we want \( P( \text{used steroids} \mid \text{tested positive}) \)

- \( E = \text{Tested positive} \)  \( P(E \mid F) = 0.95 \)
- \( F = \text{Used steroids} \)  \( P(F) = 0.1 \)  \( P(\overline{F}) = 0.9 \)
A manufacturer claims that its drug test will detect steroid use 95% of the time. What the company does not tell you is that 15% of all steroid-free individuals also test positive (the false positive rate). 10% of the Tour de France bike racers use steroids. Your favorite cyclist just tested positive. What’s the probability that he used steroids?

Define events: we want \( P( \text{ used steroids | tested positive} ) \)

\[
P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|\overline{F})P(\overline{F})}
\]

\( E = \text{Tested positive} \)  \( P( E | F ) = 0.95 \)  \( P( E | \overline{F} ) = 0.15 \)

\( F = \text{Used steroids} \)  \( P(F) = 0.1 \)  \( P( \overline{F} ) = 0.9 \)
Bayes' Theorem: Example 1

Rosen Section 7.3

A manufacturer claims that its drug test will detect steroid use 95% of the time. What the company does not tell you is that 15% of all steroid-free individuals also test positive (the false positive rate). 10% of the Tour de France bike racers use steroids. Your favorite cyclist just tested positive. What’s the probability that he used steroids?

Define events: we want \( P( \text{used steroids} \mid \text{tested positive}) \)

\[
P(F \mid E) = \frac{P(E \mid F)P(F)}{P(E \mid F)P(F) + P(E \mid \overline{F})P(\overline{F})} = \frac{0.95 \cdot 0.1}{0.95 \cdot 0.1 + 0.15 \cdot 0.9} = 0.41
\]

Plug in: 41%
Bayes' Theorem: Example 2

\[
P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|\bar{F})P(\bar{F})}
\]

Suppose we have found that the word “Rolex” occurs in 250 of 2000 messages known to be spam and in 5 out of 1000 messages known not to be spam. Estimate the probability that an incoming message containing the word “Rolex” is spam, assuming that it is equally likely that an incoming message is spam or not spam.
Bayes' Theorem: Example 2

\[ P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|\bar{F})P(\bar{F})} \]

Suppose we have found that the word “Rolex” occurs in 250 of 2000 messages known to be spam and in 5 out of 1000 messages known not to be spam. Estimate the probability that an incoming message containing the word “Rolex” is spam, assuming that it is equally likely that an incoming message is spam or not spam.

We want: \( P(\text{spam} | \text{contains } "Rolex") \). So define the events

\[ E = \text{contains } "Rolex" \]
\[ F = \text{spam} \]
Bayes' Theorem: Example 2

Suppose we have found that the word “Rolex” occurs in 250 of 2000 messages known to be spam and in 5 out of 1000 messages known not to be spam. Estimate the probability that an incoming message containing the word “Rolex” is spam, assuming that it is equally likely that an incoming message is spam or not spam.

We want: $P(\text{spam} \mid \text{contains "Rolex"})$. So define the events

$E = \text{contains "Rolex"}$
$F = \text{spam}$

What is $P(E\mid F)$?

A. 0.005
B. 0.125
C. 0.5
D. Not enough info
Suppose we have found that the word “Rolex” occurs in 250 of 2000 messages known to be spam and in 5 out of 1000 messages known not to be spam. Estimate the probability that an incoming message containing the word “Rolex” is spam, assuming that it is equally likely that an incoming message is spam or not spam.

We want: \( P( \text{spam} \mid \text{contains "Rolex"} ) \).

\[
P(F \mid E) = \frac{P(E \mid F)P(F)}{P(E \mid F)P(F) + P(E \mid \bar{F})P(\bar{F})}
\]

\( E = \text{contains "Rolex"} \) \quad \( P( E \mid F) = 250/2000 = 0.125 \) \quad \( P( E \mid \bar{F}) = 5/1000 = 0.005 \)
Bayes' Theorem: Example 2

Suppose we have found that the word “Rolex” occurs in 250 of 2000 messages known to be spam and in 5 out of 1000 messages known not to be spam. Estimate the probability that an incoming message containing the word “Rolex” is spam, assuming that it is equally likely that an incoming message is spam or not spam.

We want: \( P(\text{spam} | \text{contains } "\text{Rolex}" ) \).

\[
P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|\bar{F})P(\bar{F})}
\]

\( E = \text{contains } "\text{Rolex}" \)

\( P( E | F ) = 250/2000 = 0.125 \quad P( E | \bar{F} ) = 5/1000 = 0.005 \)

\( F = \text{spam} \)

\( P( F ) = P(\bar{F} ) = 0.5 \)
Bayes' Theorem: Example 2

Suppose we have found that the word “Rolex” occurs in 250 of 2000 messages known to be spam and in 5 out of 1000 messages known not to be spam. Estimate the probability that an incoming message containing the word “Rolex” is spam, assuming that it is equally likely that an incoming message is spam or not spam.

We want: $P(\text{spam} \mid \text{contains "Rolex"})$.

E = contains "Rolex" \hspace{1cm} P( E \mid F) = 250/2000 = 0.125 \hspace{1cm} P( E \mid \bar{F}) = 5/1000 = 0.005

F = spam \hspace{1cm} P( F ) = P( \bar{F} ) = 0.5

Rosen Section 7.3

$$P(F \mid E) = \frac{P(E \mid F)P(F)}{P(E \mid F)P(F) + P(E \mid \bar{F})P(\bar{F})} = \frac{0.125 \cdot 0.5}{0.125 \cdot 0.5 + 0.005 \cdot 0.5} = 0.96$$

Plug in: 96%