1 Events & Probability

1.1 Example: Comparing Polynomials

Slide 1

Example ①: Comparing Polynomials

Problem: compare two polynomials, $F(x) \stackrel{?}{=} G(x)$?? e.g.
$(x+1)(x-2)(x+3)(x-4)(x+5)(x-6) \stackrel{?}{=} x^6 - 7x^3 + 25$?

Solution: trying to avoid converting both to canonical form; instead: Let d = degree of the polynomial

1. Select a value r uniformly in the range $\{1, \ldots, 100d\}$;
2. Evaluate $F(r)$ and $G(r)$.
3. If $F(r) \neq G(r)$, then surely report $F(\cdot) \neq G(\cdot)$
4. If $F(r) = G(r)$, what should we report?

- we could have by chance found a root of $F(x) - G(x) = 0$.
- If $F(\cdot) \neq G(\cdot) \implies$ at most d solutions of $F(x) - G(x) = 0$;
- $\implies$ prob to select one of these roots by chance $\leq 1/100$.

Should we report $F(\cdot) = G(\cdot)$? Allowing for a possible error.

1.2 Axioms of Probability

Slide 2

Axioms of Probability

In general, a chance experiment is characterized by:

Definition 1.1.: Probability space

1. Sample space Ω = all possible outcomes.
   In the example, pairs of numbers $(F(r), G(r))$.
   The elements $\omega \in \Omega$ are the "elementary outcomes".
2. Allowable events, a family $\mathcal{F}$ of sets $E \subseteq \Omega$.
   E.g., E = a pair of matching values.
3. A probability function $\Pr : \mathcal{F} \to R$.
   E.g., any of the pairs $(F(1), G(1)), \ldots, (F(100d), G(100d))$ *is equally likely ("uniform").*

1.3 Interpreting probabilities

Slide 3

Interpreting Probabilities

Equally likely outcomes: *e.g. in the example, E = "selecting one* of the roots", let $N = |\Omega|$ and $n_E = |E|$. We used

$$\Pr(E) = \frac{n_E}{N}$$

and we argued $n_E \leq d \implies \Pr(E) \leq d/(100d)$

Subjective probabilities: E = "nuclear war in 21st century". $\Pr(E) = 0.001$ ($\implies$ $(E^c) = 0.999$, for E^c = not E). Here $\Pr(E)$ is my subjective judgement.

Long run frequencies: Weather in Austin, on 12/31 12pm

- $\Omega = \{$ "sunny", "cloudy", "rain", "snow"$\}$.
- Pr(sunny) = 80% (based on many past years)

here $\Pr(\cdot)$ is a long run frequency.

All interpretation fit into the same formal framework ...

Slide 4

Probability Function

In general, the probability function needs to satisfy some constraints. For example, Pr("something happening") better be 1 (some weather will surely happen!), be ≥ 0 etc.

Definition 1.2: Probability function $\Pr : \mathcal{F} \to R$, with

A1. $0 \leq \Pr(E) \leq 1$, for any $E \in \mathcal{F}$;

A2. $\Pr(\Omega) = 1$;

A3. For any (finite or countably infinite) sequence of pairwise mutually exclusive events $E_1, E_2, \ldots$

$$\Pr\left(\bigcup_{i\geq 1} E_i\right) = \sum_{i\geq 1} \Pr(E_i)$$

1.4 Addition Rule

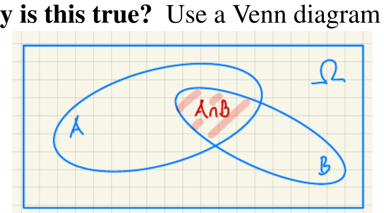
Slide 5

Addition Rule

Lemma 1.1: Addition Rule. For any two events A and B

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$

Why is this true? Use a Venn diagram to see



(this is not a proof - just a visualization)

$\to$ book for a proof (using $E_1 = A \cap B^c, E_2 = B \cap A^c, E_3 = A \cap B$)

Since all $\Pr(\cdot)$ are non-negative, we have

Lemma 1.2: Union bound. For any sequence of events $E_1, E_2, \ldots$

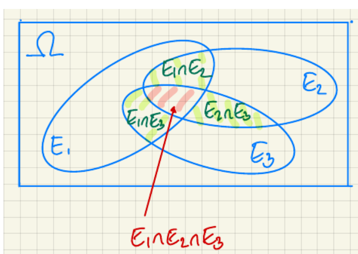
$$\Pr\left(\bigcup_{i\geq 1} E_i\right) \leq \sum_{i\geq 1} \Pr(E_i)$$

Slide 6

Lemma 1.3: Inclusion-exclusion principle. For any events $E_1, \ldots, E_n$,

$$\Pr(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} \Pr(E_i) - \sum_{i<j} \Pr(E_i \cap E_j) + \sum_{i<j<k} \Pr(E_i \cap E_j \cap E_k) - \ldots + (-1)^{\ell-1} \sum_{i_1<i_2\ldots<i_\ell} \Pr(E_{i_1} \cap \ldots E_{i_\ell})$$

Why is this true?

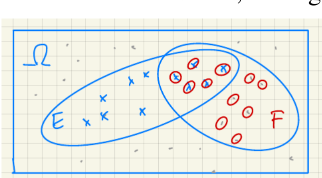


1.5 Conditional Probability

Slide 7

Conditional Probability

Equally likely outcomes: we define $\Pr(E \mid F) \equiv$ the relative # of outcomes that favor E, among all those that favor F:



$$\Pr(E \mid F) = \frac{n_{E\cap F}}{n_F}$$

$$\ldots = \frac{n_{E\cap F}/N}{n_F/N} = \frac{\Pr(E \cap F)}{\Pr(F)}.$$

General: this motivates the general definition

$$\Pr(E \mid F) = \frac{\Pr(E \cap F)}{\Pr(F)}.$$

Careful: $\Pr(E \mid F) \neq \Pr(F \mid E)$ is not symmetric.

Slide 8

Multiplication rule: immediate corollary

$$\begin{aligned}\Pr(E \cap F) &= \Pr(F)\Pr(E \mid F) \\ &= \Pr(E)\Pr(F \mid E)\end{aligned}$$

Careful: $\Pr(E \mid F) \neq \Pr(F \mid E)$ is not symmetric.

1.6 Independence

Slide 9

Independence

Example 1 (ctd.): Recall, we draw $r_1 \in \{1, \ldots, 100d\}$.
If $F(r_1) - G(r_1) \neq 0 \rightarrow$ report $F \neq G$ – easy.
If $F(r_1) - G(r_1) = 0 \rightarrow$ report $F = G$, with error prob $\leq \frac{1}{100}$.

Improved error bound: repeat with more random draws: $r_2 \in \{1, \ldots, 100d\}$ (and r_3, etc.)
Let $E_i = \{F(r_i) - G(r_i) = 0\}$ and $d_0 = \#$ roots.
If in fact $F(\cdot) \neq G(\cdot)$, then

$$\begin{aligned}\underline{\Pr(E_1 \cap E_2)} &= \frac{n_{E_1\cap E_2}}{N} = \frac{d_0 \cdot d_0}{100d \cdot 100d} \\ &= \frac{d_0}{100d} \cdot \frac{d_0}{100d} = \underline{\Pr(E_1)\Pr(E_2)} \leq \left(\frac{1}{100}\right)^2\end{aligned}$$

Independence: In general we define E, F are independent if

$$\underline{\Pr(E \cap F) = \Pr(E) \cdot \Pr(F)}$$

Note when $\Pr(F) > 0$, then $\ldots \iff \Pr(E \mid F) = \Pr(E)$.
"Knowing F does not change judgment on E."

Slide 10

Independence (ctd.)

Extend the same definition to multiple events:

k-events: $E_1, \ldots, E_k$ are *mutually independent* if for any subset $I \subseteq \{1, \ldots, k\}$

$$\Pr\left(\bigcap_{i\in I} E_i\right) = \prod_{i\in I} \Pr(E_i)$$

Note: pairwise ind. for all pairs $E_i, E_j \not\Rightarrow$ mutual ind.
For example, tossing two fair coins, $E_1 = \{HT, HH\}$ (head on 1st), $E_2 = \{TH, HH\}$ (H on 2nd), and $E_3 = \{HH, TT\}$ (same on both). Then E_1, E_2, E_3 are pairwise, but not mutually independent.

Slide 11

Example 1 (ctd.)

We can improve the error bound even more:

- Sample $r_i \in \{1, \ldots, 100d\}$, with replacement, $i = 1, \ldots, k$
- If in fact $F(\cdot) - G(\cdot) \neq 0$, then

$$\Pr(E_1 \cap \ldots E_k) = \prod_{i=1}^{k} \Pr(E_i) \leq \left(\frac{1}{100}\right)^k$$

- When $F(r_i) = G(r_i)$, $i = 1 \ldots k \to$ report "$F(\cdot) = G(\cdot)$", with error probability $\leq (1/100)^k$.

1.7 Examples

Slide 12

Examples

 # 1.2: Rolling two dice. Let A = "both dice show the same number", $B = \{a_1 > a_2\}$, where a_j = number on j-th die, and $C = \{a_1 + a_2 \geq 10\}$.

Find $\Pr(A), \Pr(B), \Pr(C), \Pr(C \mid A)$ and $\Pr(A \mid C)$. Are A and C independen?

Solution: There are $N = 36$ outcomes, $N_A = 6$ favor A $\implies \Pr(A) = 6/36$.

$N_0 = 6$ outcomes with $a_1 = a_2$, and therefore (by symmetry) $N_B = (36 - 6)/2 = 15$ with $a_1 > a_2 \implies \Pr(B) = 15/36$.

$n_C = 6 \implies \Pr(C) = \frac{6}{36} = \frac{1}{6}$,

$n_{A \cap C} = 2 \implies \Pr(C \mid A) = \frac{2}{6} = \frac{1}{3}$, $\Pr(A \mid C) = \frac{2}{6} = \frac{1}{3}$.

A and C are not independent because $\Pr(C) \neq \Pr(C \mid A)$. $\Pr(C \mid A) = \frac{1}{3} = \Pr(A \mid C)$ is a coincidence here – not true in general.

Slide 13

Before the next example, two useful results to count # outcomes:

1. *Permutation:* $n!$ = # of ordered arrangements of n items (in n positions).
2. *Combinations:* $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ = # of un-ordered subsets of size k of n elements ($k \leq n$).

   Note $\binom{n}{k} = \binom{n}{n-k}$ (since *not selecting* $(n-k)$ elements = selecting k elements).
3. *Basic principle of counting:* if experiment (step) 1 can result in n_1 outcomes, and experiment 2 in n_2 outcomes $\implies (n_1 \cdot n_2)$ outcomes of the joint experiment.

Slide 14

#1.3(e): Poker is played with a deck of $52 = 13 \times 4$ cards, with 13 different denomiations in each of 4 suits (diamonds, clubs, hearts and spades).
A 5-card poker hand is said to be a *full house* if it consists of 3 cards of the same denomination and 2 other cards of another denomination, i.e., a pattern "aaabb" where "a"and "b" are two different denominations. Let C denote the event that one is dealt a full house. Find $\Pr(C)$.

Solution: $\Pr(C) = n_C/N$, with $N = \binom{52}{5}$ and $n_C = \binom{4}{3} \cdot 13 \cdot \binom{4}{2} \cdot 12$, and therefore

$$\Pr(C) = \frac{\frac{4!}{3!} \cdot 13 \cdot \frac{4!}{2 \cdot 2} \cdot 12}{\frac{52!}{47!5!}} = 0.0014$$

1.8 Example: Verifying Matrix Multiplication

Slide 15

Example 2: Verifying Matrix Multiplication

Problem: Given $(n \times n)$ matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$, verify $\boldsymbol{AB} = \boldsymbol{C}$? Simple matrix multiplication needs order n^3 operations $(\Theta(n^3))$

A randomized algorithm: for a faster verification:

- Select a random vector $\boldsymbol{r} = (r_1, \ldots, r_n) \in \{0, 1\}^n$;
- compute $\boldsymbol{Br}$, $\boldsymbol{A}(\boldsymbol{Br})$ and $\boldsymbol{Cr}$ $(\Theta(n^2)$ operations);
- if $\boldsymbol{ABr} \neq \boldsymbol{Cr} \Rightarrow$ report $\boldsymbol{AB} \neq \boldsymbol{C}$.
  if $\boldsymbol{ABr} = \boldsymbol{Cr}$ report $\boldsymbol{AB} = \boldsymbol{C}$ (and we might be wrong ...).

Slide 16

Example 2 (ctd.)

Result: Sampling $\boldsymbol{r}$ uniformly in $\{0, 1\}^n$ is equivalent to sampling r_i independently, uniformly from $\{0, 1\}$

(we write $r_i \sim \text{Unif}(\{0, 1\})$, i.i.d.)

Proof: $\Pr(\boldsymbol{r} = \boldsymbol{x}) = \Pr(r_1 = x_1, \ldots, r_n = x_n) = \prod_{i=1}^n \Pr(r_i = x_i) = \prod_{i=1}^n \frac{1}{2} = \left(\frac{1}{2}\right)^n = \frac{1}{2^n}$

1.9 Law of Total Prob

Slide 17

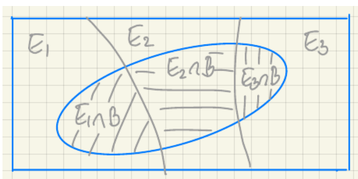
Law of Total Probability

Theorem 1.6: Law of Total Probability: If $E_1, \ldots, E_n$ are mutually disjoint events in Ω and $\bigcup_i E_i = \Omega$ (i.e., $\Omega = \bigcup E_i$ is a *partition*), then

$$\Pr(B) = \sum_{i=1}^{n} \Pr(B \cap E_i) = \sum_{i=1}^{n} \Pr(B \mid E_i)\ \Pr(E_i).$$

In words: the (marginal) probability $\Pr(B)$ is the average of the conditional probabilities $\Pr(B \mid E_i)$ under scenarios $E_1, \ldots, E_n$.

Why is that true?



Proof: $\rightarrow$ book (easy)

1.10 Verifying Matrix Mult (ctd.)

Slide 18

Example (2) (ctd.)

Recall the problem to verify $\boldsymbol{AB} = \boldsymbol{C}$. We sample $\boldsymbol{r}$ uniformly from $\{0,1\}^n$ (in short, $\boldsymbol{r} \sim \text{Unif}(\{0,1\}^n)$).

- If $\underbrace{(\boldsymbol{AB} - \boldsymbol{C})}_{\boldsymbol{D}}\boldsymbol{r} \neq 0 \rightarrow$ easy – report $\boldsymbol{AB} \neq \boldsymbol{C}$.
- If $\boldsymbol{Dr} = 0$, should we report $\boldsymbol{AB} = \boldsymbol{C}$?
  But even if $\boldsymbol{AB} \neq \boldsymbol{C}$, we could by chance have generated $\boldsymbol{r}$ as a solution of $\boldsymbol{Dr} = 0$.

  Next, we evaluate the chance of this happening.

Slide 19

Result: If in fact $\boldsymbol{AB} \neq \boldsymbol{C}$, and $\boldsymbol{r} = (r_1, r_2, \ldots, r_n) \sim \text{Unif}(\{0,1\}^n)$, then

$$\Pr(\boldsymbol{ABr} = \boldsymbol{Cr}) \leq \frac{1}{2}$$

Proof: first consider r_1 assuming we know $\boldsymbol{r}_{2:n} \equiv (r_2, \ldots, r_n)$. We can then bound $\Pr(\boldsymbol{ABr} - \boldsymbol{Cr} = 0)$ using the law of total probability.

- Let $\boldsymbol{D} = \boldsymbol{AB} - \boldsymbol{C}$. Then $\boldsymbol{ABr} = \boldsymbol{Cr} \iff \boldsymbol{Dr} = 0$.
  If $\boldsymbol{D} \neq 0$ then some elements of $\boldsymbol{D}$ are $\neq 0$ – assume that's $d_{11} \neq 0$ (w.l.o.g.).
- Then $\boldsymbol{Dr} = 0 \implies \sum_{j=1}^{n} d_{1j} r_j = 0$ or $r_1 = \frac{\sum_{j=2}^{n} d_{1j} r_j}{d_{11}}$.
  That is, for given $\boldsymbol{r}_{2:n} = \boldsymbol{x}_{2:n}$ we need $r_1 = \frac{\sum_{j=2}^{n} d_{1j} r_j}{d_{11}}$.
  But $r_1 \sim \text{Unif}(\{0,1\}) \implies \Pr(r_1 = \ldots \mid r_2, \ldots, r_n) \leq \frac{1}{2}$.

- Now we are ready for the proof: Assume $\boldsymbol{AB} \neq \boldsymbol{C}$. We will use the law of total probability.

Slide 20

Assume $\boldsymbol{AB} \neq \boldsymbol{C}$. We use the law of total probability, with $E_j = \{\boldsymbol{r}_{2:n} = \boldsymbol{x}_{2:n}\}$ for $\Pr(\boldsymbol{ABr} = \boldsymbol{Cr}) = \sum_j \Pr((\boldsymbol{ABr} = \boldsymbol{Cr}) \cap E_j)$:

$$\begin{aligned}
\Pr(\boldsymbol{ABr} = \boldsymbol{Cr}) &= \sum_{\boldsymbol{x}_{2:n}} \Pr\left((\boldsymbol{ABr} = \boldsymbol{Cr}) \cap \underbrace{(\boldsymbol{r}_{2:n} = \boldsymbol{x}_{2:n})}_{E_j} \right) \leq \\
&\leq \sum_{\boldsymbol{x}_{2:n}} \Pr\left(\left(r_1 = -\frac{\sum_{j=2}^{n} d_{1j} r_j}{d_{11}} \right) \cap (\boldsymbol{r}_{2:n} = \boldsymbol{x}_{2:n}) \right) \\
&= \sum_{\boldsymbol{x}_{2:n}} \Pr\left(r_1 = -\frac{\sum_{j=2}^{n} d_{1j} r_j}{d_{11}} \right) \cdot \Pr\left(\boldsymbol{r}_{2:n} = \boldsymbol{x}_{2:n}\right) \\
&\leq \sum_{\boldsymbol{x}_{2:n}} \frac{1}{2} \Pr\left(\boldsymbol{r}_{2:n} = \boldsymbol{x}_{2:n}\right) = \frac{1}{2}
\end{aligned}$$

1.11 Example: The Birthday Paradox

Slide 21

Example (3): The Birthday Paradox

Note: this is §5.1 in the book.

Question: In a class of $m = 30$, is

$$\Pr(\text{2 same b-day}) \overset{?}{>} \Pr(\text{no two same b-days})$$

It's easier to find $\Pr(\bar{A})$ for $\bar{A}$ = "no 2 same b-days" (and then use $\Pr(A) = 1 - \Pr(\bar{A})$).

Solution: assume that (i) b-days are selected uniform over $n = 365$ days, (ii) independently across the m people.

Count $N_{\bar{A}} = |\bar{A}|$ = # outcomes that favor $\bar{A}$, and N = # all possible outcomes, to get $\Pr(\bar{A}) = N_{\bar{A}}/N$.

- # ways to chose $m = 30$ birthdays = $\binom{365}{30}$, and # ways to assign those = 30!; therefore $N_{\bar{A}} = \binom{365}{30} \cdot 30!$
- $N = 365^{30}$ (365 choices for each of the 30 b-days)
- $\Pr(\bar{A}) = N_{\bar{A}}/N = 0.29 \implies \Pr(A) = 0.71$.

Slide 22

An Alternative Solution

Let E_i = (i-th b-day $\neq$ the first $i - 1$),

$$\Pr(E_i \mid E_1 \cap \ldots E_{i-1}) = \frac{365 - (i-1)}{365} = 1 - \frac{i-1}{365}$$

and

$$\Pr(\bar{A}) = \Pr(\bigcap_{i=2}^{m} E_i) = \prod_{i=2}^{m} \Pr(E_i \mid E_2 \cap \ldots E_{i-1}) = \prod_{i=2}^{m} \left(1 - \frac{i-1}{365}\right) = 0.29.$$

1.12 Bayes Theorem

Slide 23

Bayes Theorem

In Example ②, let $E_1 = \{\boldsymbol{AB} = \boldsymbol{C}\}$, $E_2 = \{\boldsymbol{AB} \neq \boldsymbol{C}\}$ and event $B = \{\boldsymbol{ABr} = \boldsymbol{Cr}\}$, i.e., "works for $\boldsymbol{r}$".

Did you notice that we answered the wrong question? We found $\Pr(B \mid E_2)$. But if we already know that the identity is wrong, it's silly to test it! We really want $\Pr(E_j \mid B)$!

Theorem 1.7: Bayes Theorem. Let $\bigcup_{i=1}^{k} E_i$ be a partition of the sample space. Then

$$\Pr(E_j \mid B) = \frac{\Pr(B \mid E_j)\Pr(E_j)}{\sum_i \Pr(B \mid E_i)\Pr(E_i)}$$

Bayes' Theorem "turns around the conditioning".

Proof: Use the definition of cond prob, the multiplication rule, and the law of total prob

$$\Pr(E_j \mid B) = \frac{\Pr(B \cap E_j)}{\Pr(B)} = \frac{\Pr(B \mid E_j)\Pr(E_j)}{\sum_i \Pr(B \mid E_i)\Pr(E_i)}$$

that's all!

Slide 24

Let $\bar{E}$ = not E denote the complement to E.
Bayes Theorem is often useful for the partition $(E, \bar{E})$:

$$\Pr(E \mid B) = \frac{\Pr(B \mid E)\Pr(E)}{\Pr(B \mid E)\Pr(E) + \Pr(B \mid \bar{E})(1 - \Pr(E))}.$$

Next, we will use this for Example ②, using

- E = identity is correct, $\boldsymbol{AB} = \boldsymbol{C}$,
- B = "works for $\boldsymbol{r}$" (i.e., $B = \{\boldsymbol{ABr} = \boldsymbol{Cr}\}$).

Slide 25

Example ② (ctd.)

- E = identity is correct, $\boldsymbol{AB} = \boldsymbol{C}$.
- B = "works for $\boldsymbol{r}$" (i.e., $B = \{\boldsymbol{ABr} = \boldsymbol{Cr}\}$).

Using Bayes' theorem we can find $\Pr(E \mid B)$.

- Start with $\Pr(E) = \frac{1}{2}$ (before the computer experiment with $\boldsymbol{r}$).
- We know $\Pr(B \mid E) = 1$ and just found $\Pr(B \mid \bar{E}) \leq \frac{1}{2}$.

$$\Longrightarrow \Pr(E \mid B) = \frac{\Pr(B \mid E)\Pr(E)}{\Pr(B \mid E)\Pr(E) + \Pr(B \mid \bar{E})(1 - \Pr(E))} \geq \frac{1 \cdot 1/2}{1 \cdot 1/2 + 1/2 \cdot 1/2} = \frac{2}{3}$$

When B is the experiment (evidence), we refer to

- $\Pr(E)$ as "prior probability", and
- $\Pr(E \mid B)$ as "posterior probability"

1.13 Example

Slide 26

Example: Monty Hall Problem

#1.12: game show,

- 3 doors: Behind one door is a car, behind the other two doors a goat.
  Let E_j denote the event "car behind door j".
- Initially $\Pr(E_j) = \frac{1}{3}$, and you randomly guess one door to have the car, say you guess E_1.
- I open one of the other two doors, say door 2, to show you a goat. This is the evidence, B.

You now have a chance to change your guess. Assuming you want the door with the car, should you change your guess? Find $\Pr(E_1 \mid B)$ and $\Pr(E_3 \mid B)$.

Slide 27

Solution:

We use Bayes theorem. Let B = "I show you a goat behind door 2". We assume

- $\Pr(B \mid E_1) = \frac{1}{2}$, since I could open door 2 *or* 3 to show you a goat,
- $\Pr(B \mid E_3) = 1$, since i can only open door 2 now, assuming that I don't want to show you the car behind door 3!

And we already have $\Pr(E_i) = 1/3$. Then

$$\Pr(E_1 \mid B) = \frac{\Pr(B \mid E_1)\Pr(E_1)}{\Pr(B \mid E_1)\Pr(E_1) + \Pr(B \mid E_3)\Pr(E_3)} = \frac{1/2}{3/2} = \frac{1}{3}$$

and therefore $\Pr(E_3 \mid B) = \frac{2}{3}$. Change your guess to E_3!