NLP - HW2

September 22, 2025

Sathvik Manikantan Napa Ugandhar - 333E9A

Q1

Given

Assume a probability space $(\mathcal{E},\, p)$ with axioms:
(i) $p(\mathcal{E}) = 1$
(ii) If $X \cap Y = \varnothing$, then $p(X \cup Y) = p(X) + p(Y)$.
For $p(Z) > 0$, define $p(X \mid Z) = \frac{p(X \cap Z)}{p(Z)}$.

Q1. a

To Prove: if $Y \subseteq Z$, then $p(Y) \leq p(Z)$

Using additive property,

$p(Z) = p(Y) + p(Z \backslash Y) \geq p(Y)$

so $p(Y) \leq p(Z)$

Thus, if $Y \leq Z$, then $p(Y) \leq p(Z)$ ■

Q1. b

To Prove: $0 \leq p(X \mid Z) \leq 1$**.**

Probabilities of all sets are defined in the range [0, 1]
$p(X \mid Z) \geq 0, \quad p(X \cap Z) \geq 0, \quad p(Z) \geq 0$

$p(X \mid Z) = \frac{p(X \cap Z)}{p(Z)} \geq 0$

Since $X \cap Z \subseteq Z$,

by (a) $p(X \cap Z) \leq p(Z)$ and $p(X \cap Z) \geq 0$.

Divide by $p(Z) > 0$:

$0 \leq \frac{p(X \cap Z)}{p(Z)} \leq 1$. ■

Q1. c

To Prove: if $p(\varnothing) = 0$

$p(\mathcal{E}) = p(\mathcal{E} \cup \varnothing) = p(\mathcal{E}) + p(\varnothing)$
$= 1 + p(\varnothing)$
We know that the range of probabilities lie within [0, 1].
Hence, $p(\varnothing) = 0$ ■

Q1. d

To Prove: $\bar{X} = \mathcal{E} - X$

$\mathcal{E} = X \cup \bar{X}$
p($\mathcal{E}$) = p($X \cup \bar{X}$)

Since X and $\bar{X}$ are mutually exclusive events,
$p(\mathcal{E}) = p(X) + p(\bar{X})$

$p(X) = 1 - p(\bar{X})$ ■

Q1. e

To Prove: $p(singing \cap rainy \mid rainy) = p(singing \mid rainy)$

Let S be the random variable representing singing and R be the random variable for rainy.
$p(S \cap R \mid R) = p((S \cap R) \cap R \mid R)$
$= \frac{p(S \cap R)}{R}$
$= p(S \mid R)$ ■

Q1. f

To Prove: $p(X \mid Y) = 1 - p(\bar{X} \mid Y)$

$$p(\bar{X} \mid Y) = \frac{p(\bar{X} \cap Y)}{p(Y)} = \frac{p(Y) - p(X \cap Y)}{p(Y)} = 1 - \frac{p(X \cap Y)}{p(Y)} = 1 - p(X \mid Y). \quad ■$$

Q1. g

Simplify $\left(p(X \mid Y)p(Y) + p(X \mid \bar{Y})p(\bar{Y})\right) \frac{p(Z|X)}{p(Z)}$

$p(X \mid Y)p(Y) + p(X \mid \bar{Y})p(\bar{Y}) = p(X) \quad \Rightarrow \quad \dfrac{p(X)\, p(Z \mid X)}{p(Z)} = \dfrac{p(X \cap Z)}{p(Z)} = p(X \mid Z).$

Q1. h

When singing and raining are mutually exclusive events and are disjoint events
$p(X \cap Z) = 0$

Q1. i

When singing and raining are conditionally independent events
$p(X \cap Z) = p(X).p(Z)$

Q1. j

To Prove: If $p(X \mid Y) = 0$**, then** $p(X \mid \bar{Y}) \geq p(X)$**.**

If $p(X \mid Y) = 0$

$$p(x \mid Y, Z) = \frac{X \cap Y \cap Z}{Y \cap Z}$$

We know that $\frac{p(X \cap Y)}{p(Y)} = 0$
p($X \cap Y$) = 0
$X \cap Y \cap Z \leq X \cap Y$
p($X \cap Y \cap Z$) = 0, since limit of probability is [0, 1]

if $P(Y \cap Z) \geq 0$,
$p(X \mid Y, Z) = \frac{p(X \cap Y \cap Z)}{p(Y \cap Z)} = 0$

Q1. k

If $p(W \mid Y) = 1$**, then** $p(W \mid \bar{Y}) \leq p(W)$**.**

Total probability gives

$$p(W) = 1 \cdot p(Y) + p(W \mid \bar{Y})p(\bar{Y}) \;\Rightarrow\; p(W \mid \bar{Y}) = \frac{p(W) - p(Y)}{1 - p(Y)} \leq p(W).$$

Q2

Given. Cars are either red or blue. The witness claims *blue*.
Witness reliability is 0.8 (correct color w.p. 0.8 regardless of the color).

Base rate: $p(\text{Actual} = \text{blue}) = 0.10$ (hence $p(\text{Actual} = \text{red}) = 0.90$).

Let A be the event "Actual = blue," A^c be "Actual = red," and C be "Claimed = blue."

Q2. a

By Bayes Theorem,

$$p(Actual = blue \mid Claimed = blue) = \frac{p(Claimed = blue \mid Actual = blue).p(Actual = blue)}{p(Claimed = blue)}$$

$$p(A \mid C) = \frac{p(C \mid A)\, p(A)}{p(C)} \quad \text{with} \quad p(C) = p(C \mid A)\, p(A) + p(C \mid A^c)\, p(A^c).$$

Q2. b

- Event as hypothesis: A ("Actual = blue").
- Event as evidence: C ("Claimed = blue").

For the three probabilities in 2(a):

$$\underbrace{p(A)}_{\text{prior}}, \qquad \underbrace{p(C \mid A)}_{\text{likelihood}}, \qquad \underbrace{p(A \mid C)}_{\text{posterior}}.$$

Q2. c

Which of the three probabilities matters to the judge?

$p(A) = 0.10, \qquad p(C \mid A) = 0.80, \qquad p(C \mid A^c) = 0.20$ (symmetry from 0.8 reliability).

Compute

$$p(C) = 0.8 \cdot 0.10 + 0.2 \cdot 0.90 = 0.08 + 0.18 = 0.26,$$
$$p(A \mid C) = \frac{0.8 \cdot 0.10}{0.26} = \frac{4}{13} \approx 0.3077\,.$$

The judge should care about th e posterior $p(A \mid C)$, i.e., the probability the car was actually blue given the claim.

Q2. d

To Prove:

$$\boxed{p(A \mid B, Y) = \frac{p(B \mid A, Y)\, p(A \mid Y)}{p(B \mid Y)}},$$

$$p(A \mid B, Y) = \frac{p(A, B \mid Y).p(A \cap B \mid Y)}{p(B \mid Y)}$$

$$= \frac{p(B \mid A, Y).p(A \mid Y)}{p(B \mid Y)} \quad \blacksquare$$

Q2. e

Apply law of total probability to **(Q2. d)**

$$\frac{p(B \mid A, Y).p(A \mid Y)}{p(B \mid A, Y)\, p(A \mid Y) + p(B \mid A^c, Y)\, p(A^c \mid Y)}$$

Q2. f

Solve Let Y = "city = Baltimore" With reliability 0.8 and base rate $p(A \mid Y)$ = 0.10:

$$p(B \mid A, Y) = 0.80, \qquad p(C \mid A^c, Y) = 0.20, \qquad p(A^c \mid Y) = 0.90.$$

Therefore

$$\boxed{p(\text{Actual=blue} \mid \text{Claimed=blue}, Y) = \frac{0.80 \cdot 0.10}{0.80 \cdot 0.10 + 0.20 \cdot 0.90} = \frac{0.08}{0.26} = \frac{4}{13} \approx 0.3077.}$$

Q3

Let cry $\in$ {bwa, bwee, kiki} and situation $\in$ {Predator, Timber, Help}. The given conditional table is

$p(\text{cry} \mid \text{situation})$	Predator	Timber	Help
bwa	0	0.1	0.8
bwee	0	0.6	0.1
kiki	1.0	0.3	0.1

Q3. a

For each fixed situation s,

$$\sum_{\text{cry} \in \{\text{bwa}, \text{bwee}, \text{kiki}\}} p(\text{cry} \mid \text{situation} = s) = 1.$$

E.g., $p(\text{bwa} \mid \text{Predator}) + p(\text{bwee} \mid \text{Predator}) + p(\text{kiki} \mid \text{Predator}) = 1$, and similarly for the other two columns.

Q3. b

Given $p(\text{Predator}) = 0.2$, $p(\text{Timber}) = 0$, so $p(\text{Help}) = 1 - 0.2 - 0 = 0.8$. Use

$$p(\text{cry}, \text{situation}) = p(\text{cry} \mid \text{situation})\, p(\text{situation}).$$

The completed joint table is:

$p(\text{cry}, \text{situation})$	Predator	Timber	Help	Total
bwa	$0 \times 0.2 = 0$	$0.1 \times 0 = 0$	$0.8 \times 0.8 = 0.64$	0.64
bwee	$0 \times 0.2 = 0$	$0.6 \times 0 = 0$	$0.1 \times 0.8 = 0.08$	0.08
kiki	$1.0 \times 0.2 = 0.20$	$0.3 \times 0 = 0$	$0.1 \times 0.8 = 0.08$	0.28
Total	0.20	0	0.80	1.00

Q3. c

i) The probability can be written as

$$p(\text{Predator} \mid \text{kiki})$$

ii) Can be rewritten as:

$$= \frac{p(\text{Predator} \cap \text{kiki})}{p(\text{kiki})}.$$

iii) From the joint table

$$= \frac{0.20}{0.28} = \frac{5}{7} \approx 0.714.$$

iv) By Bayes with total probability:

$$\boxed{\frac{p(\text{kiki} \mid \text{Predator})\, p(\text{Predator})}{p(\text{kiki} \mid \text{Predator})\, p(\text{Predator}) + p(\text{kiki} \mid \text{Timber})\, p(\text{Timber}) + p(\text{kiki} \mid \text{Help})\, p(\text{Help})}}$$

v) Plugging numbers:

$$\frac{1.0 \cdot 0.2}{1.0 \cdot 0.2 + 0.3 \cdot 0 + 0.1 \cdot 0.8} = \frac{0.20}{0.28} = \frac{5}{7} \approx 0.714,$$

Q4

Let a sentence be $\mathbf{w} = w_1 w_2 \cdots w_n$, with $w_{-1} = w_0 = BOS$ and $w_{n+1} = EOS$. A trigram LM assigns

$$p(\mathbf{w}) \;=\; \prod_{i=1}^{n+1} p(w_i \mid w_{i-2}, w_{i-1})\,.$$

Naive (unsmoothed MLE) parameters are

$$p(u \mid x, y) = \frac{c(x, y, u)}{c(x, y)},$$

where $c(\cdot)$ are corpus counts of bigrams/trigrams.

$$\boxed{\begin{aligned} p(w_1 w_2 w_3 w_4) &= p(w_1 \mid BOS, BOS)\, p(w_2 \mid BOS, w_1)\, p(w_3 \mid w_1, w_2)\, p(w_4 \mid w_2, w_3)\, p(EOS \mid w_3, w_4) \\ &= \frac{c(BOS, BOS, w_1)}{c(BOS, BOS)} \cdot \frac{c(BOS, w_1, w_2)}{c(BOS, w_1)} \cdot \frac{c(w_1, w_2, w_3)}{c(w_1, w_2)} \\ &\quad \cdot \frac{c(w_2, w_3, w_4)}{c(w_2, w_3)} \cdot \frac{c(w_3, w_4, EOS)}{c(w_3, w_4)}. \end{aligned}}$$

- $c(BOS, BOS)$: number of sentence starts (i.e., # sentences).
- $c(BOS, BOS, w_1)$: # sentences whose first token is w_1.
- $c(BOS, w_1)$: # times bigram (BOS, w_1) appears at sentence start.
- $c(w_1, w_2)$, $c(w_1, w_2, w_3)$, $c(w_2, w_3)$, $c(w_2, w_3, w_4)$: occurrences of those bigrams/trigrams anywhere.
- $c(w_3, w_4, EOS)$: # sentences ending with $w_3\, w_4$.

Q4. b

$$p(\text{do you think the}) = \cdots \times p(\text{the} \mid \text{you, think}) \times \boxed{p(EOS \mid \text{think, the})}.$$

In English, determiners like ("the") almost never end a sentence, so $p(EOS \mid \text{think, the}) \approx 0$ (unsmoothed, often exactly 0), drives the product to be extremely small. The factor $p(\text{the} \mid \text{you, think})$ may also be small, but the key parameter making the probability tiny is the *end-of-sentence* transition $\boxed{p(EOS \mid \text{think, the})}$.

Q4. c

Expressions:

(A) $p(\text{do})\, p(\text{you} \mid \text{do})\, p(\text{think} \mid \text{do, you})$
(B) $p(\text{do} \mid BOS)\, p(\text{you} \mid BOS, \text{do})\, p(\text{think} \mid \text{do, you})\, p(EOS \mid \text{you, think})$
(C) $p(\text{do} \mid BOS)\, p(\text{you} \mid BOS, \text{do})\, p(\text{think} \mid \text{do, you})$

Descriptions:

(1) : the first complete sentence you hear is *do you think*
(2) : the first three words you hear are *do you think*
(3) : the first complete sentence you hear starts with *do you think*

Answer.

$$\boxed{(A) \to (2), \qquad (B) \to (1), \qquad (C) \to (3).}$$

Reasoning: (A) lacks BOS/EOS, so it models hearing those three words in a row somewhere; (B) includes BOS and EOS, so it is exactly the whole sentence; (C) has BOS but no EOS, so it is the probability a sentence *begins* with that trigram and may continue.

Q4. d

Define the reversed model

$$p_{\text{rev}}(\mathbf{w}) \;=\; \prod_{i=0}^{n} p(w_i \mid w_{i+1}, w_{i+2})\,, \quad w_0 = BOS,\; w_{n+1} = w_{n+2} = EOS,$$

with the same MLE form $p(u \mid x, y) = c(u, x, y)/c(x, y)$.

Claim: For any sentence $\mathbf{w}$,

$$\boxed{p(\mathbf{w}) = p_{\text{rev}}(\mathbf{w})}$$

when both models use MLE estimates from the same corpus.

Proof: Write each as a product of count ratios (Like in the 'Federalist Paper Authorship' problem)

Forward:

$$p(\mathbf{w}) = \frac{\prod_{i=1}^{n+1} c(w_{i-2}, w_{i-1}, w_i)}{\prod_{i=1}^{n+1} c(w_{i-2}, w_{i-1})} = \frac{c(BOS, BOS, w_1)\cdots c(w_{n-1}, w_n, EOS)}{c(BOS, BOS)\; c(BOS, w_1)\cdots c(w_{n-1}, w_n)}.$$

Reversed:

$$p_{\text{rev}}(\mathbf{w}) = \frac{\prod_{i=0}^{n} c(w_i, w_{i+1}, w_{i+2})}{\prod_{i=0}^{n} c(w_{i+1}, w_{i+2})} = \frac{c(BOS, w_1, w_2)\cdots c(w_n, EOS, EOS)}{c(w_1, w_2)\cdots c(w_n, EOS)\; c(EOS, EOS)}.$$

Take the ratio and cancel all interior bigram/trigram counts; only endpoints remain:

$$\frac{p(\mathbf{w})}{p_{\text{rev}}(\mathbf{w})} = \frac{c(BOS, BOS, w_1)}{c(w_n, EOS, EOS)} \cdot \frac{c(w_n, EOS)\; c(EOS, EOS)}{c(BOS, BOS)\; c(BOS, w_1)}.$$

In any BOS–BOS / EOS–EOS padded corpus we have the identities

$$c(BOS, BOS, w) = c(BOS, w), \qquad c(w, EOS, EOS) = c(w, EOS), \qquad c(BOS, BOS) = c(EOS, EOS) = S,$$

where S is the number of sentences. Substituting gives

$$\frac{p(\mathbf{w})}{p_{\text{rev}}(\mathbf{w})} = \frac{c(BOS, w_1)}{c(w_n, EOS)} \cdot \frac{c(w_n, EOS)\, S}{S\, c(BOS, w_1)} = 1,$$

hence $p(\mathbf{w}) = p_{\text{rev}}(\mathbf{w})$. ■

Q5

Let a sentence be $\mathbf{w} = w_1 w_2 w_3 w_4$ and let $A \in \{1, \ldots, k\}$ denote the *topic* that (once chosen) persists for the whole sentence. Assume

$$p(A) = \pi_A, \qquad p(w_i \mid w_{i-1}, A) = \text{topic-conditioned bigram}.$$

Conditional independence assumption: given the topic A, each word depends only on the previous word: $w_i \perp\!\!\!\perp (w_{<i-1}) \mid (w_{i-1}, A)$.

Joint with the topic: By the chain rule and the assumption above (using $w_0 = BOS$),

$$p(w_1, w_2, w_3, w_4, A) = \pi_A\, p(w_1 \mid BOS, A)\, p(w_2 \mid w_1, A)\, p(w_3 \mid w_2, A)\, p(w_4 \mid w_3, A)\, [\times\, p(EOS \mid w_4, A)].$$

Marginal (topic unknown) Sum over the latent topic:

$$\boxed{p(w_1 w_2 w_3 w_4) = \sum_{a=1}^{k} \pi_a\, p(w_1 \mid BOS, a)\, p(w_2 \mid w_1, a)\, p(w_3 \mid w_2, a)\, p(w_4 \mid w_3, a)\, [\times\, p(EOS \mid w_4, a)]}$$

If the topic is known: For $A = a^{known}$,

p($w_1 w_2 w_3 w_4 \mid A = a^{known}$) = p($w_1 \mid BOS,$ a^{known}) $\times$ p(w$_2 \mid w_1, a^{known}$) $\times$ p(w$_3 \mid w_2, a^{known}$)$\times$ p(w$_4 \mid w_3, a^{known}$) $\times$ p(EOS$\mid w_4, a^{known}$).

Q7

Posted a thread on Piazza
Topic: Log-linear model for reasoning routing in LLM question-answering
https://piazza.com/class/meqs8zlj8z86mc/post/98https://piazza.com/class/meqs8zlj8z86mc/post/98

Q8

a. (9) Similar Words

seattle
seahawks dallas atlanta wichita tacoma lauderdale florida spokane chino dulles

dog
dogs badger cat hound puppy dachshund sighthound poodle rat keeshond

communist
socialist bolshevik communists comintern trotskyist leftist bolsheviks cominform stalinist leninist

jpg
png svg szczepanek buteo pix gif image galleria regnum fiav

the
in its of which entire within from a part second

google
webpage yahoo web search com blogging faq editable archived redirection

a. (10)

Patterns noticed with some examples that work well and some that don't

- proper noun examples like name 'Seattle', 'Dogs', 'California' seem to provide close enough examples.

- function words such as 'the', 'which', 'an', 'a' seem to provide a examples with mixed topic types.

- polysemy words (words with multiple meanings) such as 'head', 'set' does not seem to capture all meanings cleanly.

a. (11)

The lower dimensions seem to produce co-occurrance patterns than more informed similar words in the latter dimensions.
For example, with d=10 the word 'dog' has the following similar words
turnip coronets ass pig embroidered eyed cow unicorns haired melon

with d=20, the word 'dog' has the following similar words
cat dogs ass badger hound sighthound canine azawakh bikini quadruped

with d=50, the word 'dog' has the following similar words
dogs badger cat hound puppy dachshund sighthound poodle rat keeshond

with d=100, the word 'dog' has the following similar words
dogs hound badger inu komondor mastiff canine borzoi puppy keeshond

You can observe a marked difference in the similar words with more dimensionality. With lower dimensions, the similar words for dog had embroidered, eyed and bikini. For bigger dimensions $d \geq 50$, the similar words for dog produces dog breeds. With lower dimensions, we observe a mixed bowl of words whereas for larger dimensions, the similar words are much proximate in their meanings.

b. (12)

Few Analogies

- bank - river + money (works well)
- apple - fruit + company (works well)
- python - animal + programming (works well)
- nurse - woman + man (works poorly with lower dimensions)
- beautiful - more + most (works poorly with lower dimensions)

For the analogy **nurse - woman + man**,
the lower dimension $d < 50$ shows similar words that are not very closer to what we expect 'nurse' who is a male. for d = 10, the resolution for analogy is as follows:
'housewife girlfriend boyfriend gwyneth phoney housekeeper secombe huck bookish girlfriends'

For larger dimensions, the similar words are closer to the profession
'nurse doctor dentist technician apprentice retires peppard communicators master beregond bunter'. Note: for this example, we didn't exclude the word 'nurse' itself, since we expect the same word.

For the analogy **beautiful - more + most**,
in the lower dimensions $d < 50$ shows words that are far from what we expect (examples below)
'beltaine dawn supergroup goldsmiths tuileries whirling stockyards pomegranates michelangelo artworks'

This improves with increase in dimensionality $d \geq 50$
when $d = 50$, **'famous birthplace magnificent renowned finest famed fountains mausoleums pied immortalized'**

when $d = 100$, **'famous finest waterfalls magnificent renowned beauties beauty beaches fountains masterpiece'**

b. (13)

Word2Vec represents each word as a vector. These vectors act like coordinates and place the words in a high dimension representation. Assume each dimension to capture semblance in the words and their relationships. The words with similar meanings are placed closer in this space.

In the analogy **King - Man + Woman**, when we perform **King - Man**, we separate the concept of royalty from men. Now, the representation would loosely translate to ‘royalty‘.

When we add Woman to this concept of ‘royalty‘, **King - Man + Woman**, we add the female equivalent of ‘royalty‘. Hence, our ‘findsim.py‘ would find words similar to the female equivalent of ‘royalty‘.

b. (14)

Extended ‘findsim.py‘ code added to gradescope.

Q9 Extra Credit (15)

Can this approximation be justified by using the chain rule plus backoff? No, this approximation cannot be justified as a valid joint probability distribution. In general, we can not get it from the chain rule with ordinary backoff.

The chain rule on four variables gives
$p(A, B, C, D) = p(A)\, p(B \mid A)\, p(C \mid A, B)\, p(D \mid A, B, C)$.

Backoff for a linear Markov chain would replace
$p(C \mid A, B) \to p(C \mid B)$ and $p(D \mid A, B, C) \to p(D \mid C)$,
yielding a *line* $A \to B \to C \to D$, not a *ring/graph.* it does not produce the circular conditional probability $p(D \mid A)$.

Possible Fix:
We can implement a circular Markov chain. Let $T(w' \mid w)$ be a bigram transition and let π be a stationary prior ($\pi^\top T = \pi^\top$). Choose a start position $S \in \{A, B, C, D\}$ uniformly, move across the ring once, and trace back. Marginalizing S gives the normalized joint

$$\boxed{p(A, B, C, D) = \frac{1}{4} \sum_{n \in \{A,B,C,D\}} \pi(w_n)\, T(w_{n^+} \mid w_n)\, T(w_{n^{++}} \mid w_{n^+})\, T(w_{n^{+++}} \mid w_{n^{++}})\, T(w_n \mid w_{n^{+++}})}$$

where n^+ denotes the next node around the ring. This contains the desired wrap-around factor $T(A \mid D)$ and is properly normalized.

The bare product $p(A \mid B) \times p(B \mid C) \times p(C \mid D) \times p(D \mid A)$ lacks normalization and ignores half of each node's neighborhood. We need to treat it as a pseudolikelihood-style approximation, not an exact joint.

Q10 (Extra credit) (16)

Given

$$\text{(a) } p(\neg\text{shoe} \mid \neg\text{nail}) = 1,$$
$$\text{(b) } p(\neg\text{horse} \mid \neg\text{shoe}) = 1,$$
$$\text{(c) } p(\neg\text{race} \mid \neg\text{horse}) = 1,$$
$$\text{(d) } p(\neg\text{fortune} \mid \neg\text{race}) = 1,$$

prove

$$\text{(e)} \quad p(\neg\text{fortune} \mid \neg\text{nail}) = 1.$$

Lemma If $p(U \mid V) = 1$ and $p(V \cap W) > 0$, then $p(U \mid V \cap W) = 1$

Proof From $p(U \mid V) = 1$ we have $p(\neg U \cap V) = 0$. Since $V \cap W \subseteq V$, (Problem 1(a)) gives $p(\neg U \cap V \cap W) \le p(\neg U \cap V) = 0$.

Hence

$$p(U \mid V \cap W) = 1 - p(\neg U \mid V \cap W) = 1 - \frac{p(\neg U \cap V \cap W)}{p(V \cap W)} = 1. \qquad \blacksquare$$

Using the lemma repeatedly

$$p(\neg\text{horse} \mid \neg\text{shoe}, \neg\text{nail}) = 1,$$
$$p(\neg\text{race} \mid \neg\text{horse}, \neg\text{shoe}, \neg\text{nail}) = 1,$$
$$p(\neg\text{fortune} \mid \neg\text{race}, \neg\text{horse}, \neg\text{shoe}, \neg\text{nail}) = 1.$$

By the chain rule,

$$\begin{aligned} &p(\neg\text{fortune}, \neg\text{race}, \neg\text{horse}, \neg\text{shoe} \mid \neg\text{nail}) \\ &\quad = p(\neg\text{shoe} \mid \neg\text{nail})\, p(\neg\text{horse} \mid \neg\text{shoe}, \neg\text{nail})\, p(\neg\text{race} \mid \neg\text{horse}, \neg\text{shoe}, \neg\text{nail})\, p(\neg\text{fortune} \mid \neg\text{race}, \neg\text{horse}, \neg\text{shoe}, \neg \\ &\quad = 1 \cdot 1 \cdot 1 \cdot 1 \;=\; 1. \end{aligned}$$

Since $\neg\text{fortune} \cap \neg\text{race} \cap \neg\text{horse} \cap \neg\text{shoe} \subseteq \neg\text{fortune}$, (From Problem 1(a))

$$p(\neg\text{fortune} \mid \neg\text{nail}) \;\ge\; p(\neg\text{fortune} \cap \neg\text{race} \cap \neg\text{horse} \cap \neg\text{shoe} \mid \neg\text{nail}) = 1.$$

But conditional probabilities lie in $[0, 1]$ (Problem 1(b)), so

$$\boxed{p(\neg\text{fortune} \mid \neg\text{nail}) = 1}. \qquad \blacksquare$$