The Full Employment Theory in Probability Kinematics

Graduate Colloquium, UBC

Stefan Lukits

March 23, 2012

Probability Kinematics

You have a fully quantified probability assessment of an event space. You make an observation. What is the most rationally defensible way of updating your probabilities? Three cases:

1. The observation is an event.
2. The observation redistributes the probabilities of a partition of the event space.
3. The observation is an affine constraint.

Conditioning

Example: Hercule Poirot considers the probabilities as he investigates the murder of Roger Ackroyd. He observes that the culprit must have been male.

Mrs. Cecil Ackroyd, the sister-in-law	0.08	0.00
Flora Ackroyd, her daughter	0.08	0.00
Major Blunt, a big-game hunter	0.08	0.11
Geoffrey Raymond, the personal secretary	0.08	0.11
Ralph Paton, the stepson with heavy debts	0.52	0.68
Parker, the snooping butler	0.08	0.11
Ursula Bourne, the parlour maid	0.08	0.00

Jeffrey Conditioning

Example: Hercule Poirot considers the probabilities as he investigates the murder of Roger Ackroyd. He observes that the chance of the culprit being male is only 1/2.

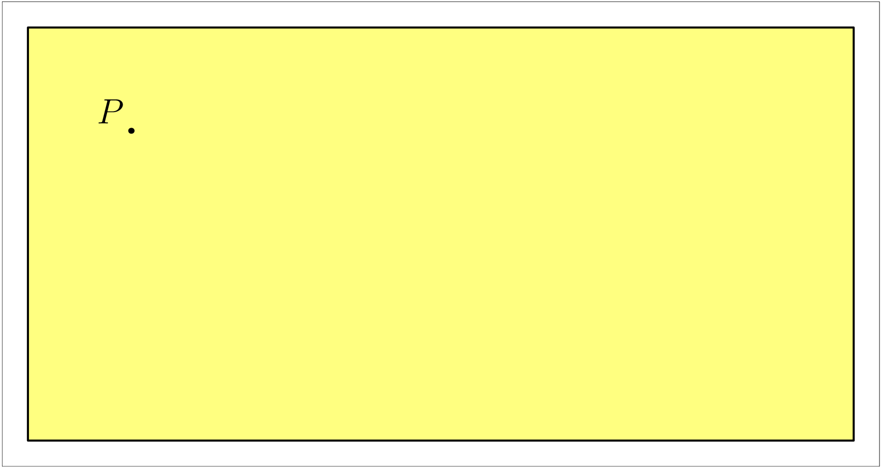
Mrs. Cecil Ackroyd, the sister-in-law	0.08	0.17
Flora Ackroyd, her daughter	0.08	0.17
Major Blunt, a big-game hunter	0.08	0.05
Geoffrey Raymond, the personal secretary	0.08	0.05
Ralph Paton, the stepson with heavy debts	0.52	0.34
Parker, the snooping butler	0.08	0.05
Ursula Bourne, the parlour maid	0.08	0.17

Affine Constraint

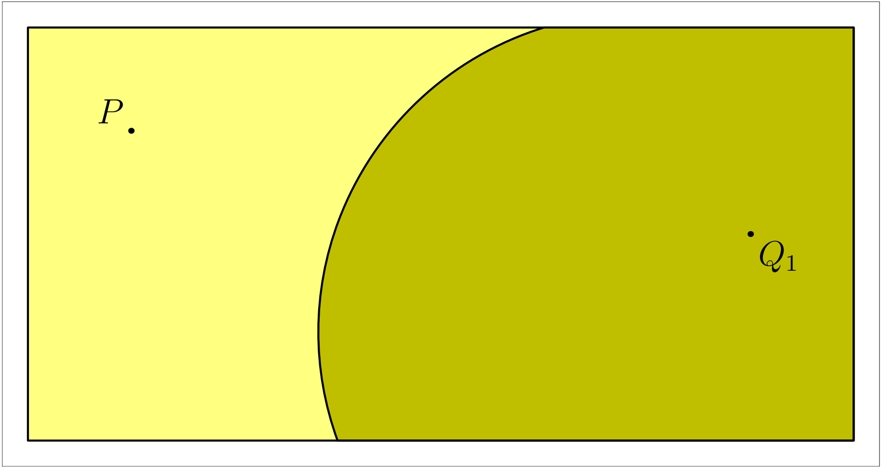
Example: Hercule Poirot considers the probabilities as he investigates the murder of Roger Ackroyd. He observes that the chance of Ralph Paton being the culprit is twice as high as the chance of Major Blunt being the culprit.

Mrs. Cecil Ackroyd, the sister-in-law	0.08	?.??
Flora Ackroyd, her daughter	0.08	?.??
Major Blunt, a big-game hunter	0.08	?.??
Geoffrey Raymond, the personal secretary	0.08	?.??
Ralph Paton, the stepson with heavy debts	0.52	?.??
Parker, the snooping butler	0.08	?.??
Ursula Bourne, the parlour maid	0.08	?.??

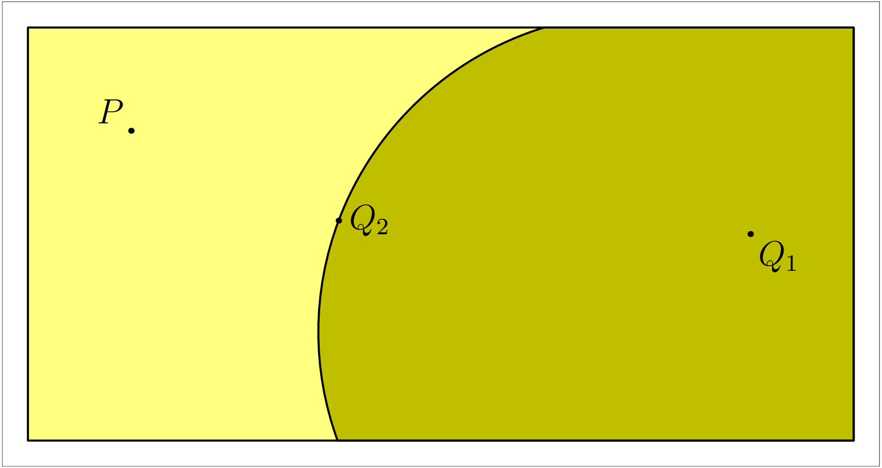
The Principle of Maximum Entropy



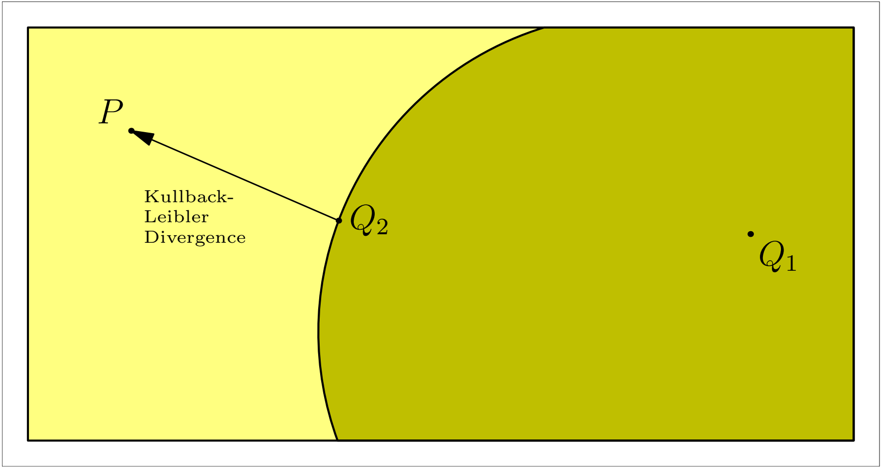
The Principle of Maximum Entropy



The Principle of Maximum Entropy



The Principle of Maximum Entropy



Principle of Maximum Entropy vs. Full Employment I

- Shore and Johnson (1980) show that, under certain assumptions, there is a unique solution to the problem of updating probabilities under an affine constraint. The Principle of Maximum Entropy: given an affine constraint, the updated probability assignment is the one that is both consistent with the constraint and informationally minimally distant from the prior probability assignment.

Principle of Maximum Entropy vs. Full Employment II

- Enter the Full Employment Theorem. Shore and Johnson's assumptions are not unanimously endorsed (see especially Uffink 1995). There are examples where the Principle of Maximum Entropy results in counterintuitive probability updating, most prominently the Judy Benjamin case.
- Consequently, there is an array of solutions and no objective method to decide between them. Only a subjective assessment of the situation can narrow down the solution space.
- We will undermine the notion that the results of the examples are counterintuitive, specifically with respect to independence, the powerset approach, epistemic entrenchment, and coarsening at random.

Principle of Maximum Entropy vs. Full Employment III

Principle of Maximum Entropy

"Jaynes's principle of maximum entropy and Kullbacks principle of minimum cross-entropy (minimum directed divergence) are shown to be uniquely correct methods for inductive inference when new information is given in the form of expected values."
(Shore and Johnson)

Full Employment

"The uniqueness proofs are flawed, or rest on unreasonably strong assumptions. A more general class of inference rules, maximizing the so-called Rényi entropies, is exhibited which also fulfill the reasonable part of the consistency assumptions." (Jos Uffink)

Principle of Maximum Entropy vs. Full Employment IV

Principle of Maximum Entropy

"We show that Skilling's method of induction leads us to a unique general theory of inductive inference, the maximum entropy method, and precisely how it is that other entropies such as those of Rényi or Tsallis are ruled out for problems of inference. We then explore the compatibility of Bayes and maximum entropy updating. We show that maximum entropy is capable of producing every aspect of orthodox Bayesian inference and prove the complete compatibility of Bayesian and entropy methods." (Adom Giffin)

Principle of Maximum Entropy vs. Full Employment V

Full Employment

"Maximum entropy and relative entropy have proved quite successful in a number of applications, from physics to natural-language modeling. Unfortunately, they also exhibit some counterintuitive behavior on certain applications. Although they are valuable tools, they should be used with care."
(Joseph Halpern)

Halpern's Full Employment Theorem

Choosing a representation:

- possible worlds
- probability measures
- lower and upper probabilities
- Dempster-Shafer belief functions
- possibility measures
- ranking functions
- relative likelihoods
- plausibility measures

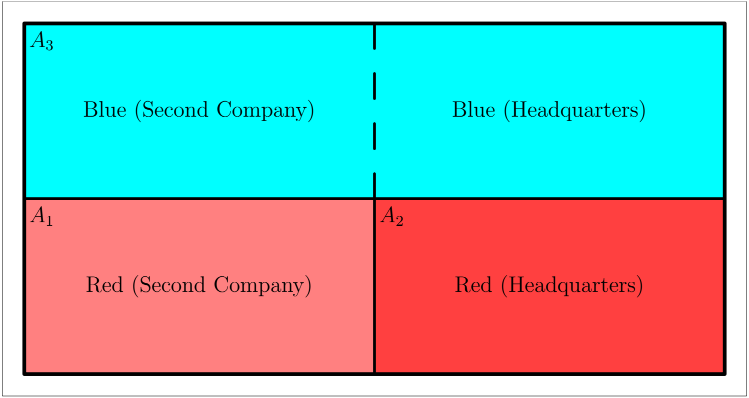
Grove and Halpern

"one must always think carefully about precisely what the information means"
("Probability Update . . . ," p6)

Joseph Halpern

"there is no escaping the need to understand the details of the application"
(*Reasoning* . . . , p423)

The Judy Benjamin Case I



The Judy Benjamin Case II

- (MAP) Judy has no idea where she is. She is on team Blue. Because of the map, her probability of being in Blue territory equals the probability of being in Red territory, and being on Red Second Company ground equals the probability of being on Red Headquarters ground.
- (HDQ) Headquarters inform Judy that in case she is in Red territory, her chance of being on their Headquarters ground is three times the chance of being on their Second Company ground.

$$2 \cdot P(A_1) = 2 \cdot P(A_2) = P(A_3) \tag{MAP}$$

$$q = P(A_2 | A_1 \cup A_2) = \frac{3}{4} \tag{HDQ}$$

Two Intuitions

(T1) HDQ should not affect $P(A_3)$. Let's call P the prior probability distribution, before receiving HDQ, and Q the posterior probability distribution, after receiving HDQ. Then $Q(A_3) = P(A_3)$.

(T2) If the value of q approaches 1 then $Q(A_3)$ should approach 2/3. HDQ would then be "if you are in Red territory you are almost certainly on Red Headquarters ground." Considering MAP, $Q(A_3)$ should approach 2/3. Continuity considerations pose a contradiction to T1.

Two Intuitions

(T1) HDQ should not affect $P(A_3)$. Let's call P the prior probability distribution, before receiving HDQ, and Q the posterior probability distribution, after receiving HDQ. Then $Q(A_3) = P(A_3)$.

(T2) If the value of q approaches 1 then $Q(A_3)$ should approach $2/3$. HDQ would then be "if you are in Red territory you are almost certainly on Red Headquarters ground." Considering MAP, $Q(A_3)$ should approach $2/3$. Continuity considerations pose a contradiction to T1.

The PME Applied to Judy Benjamin

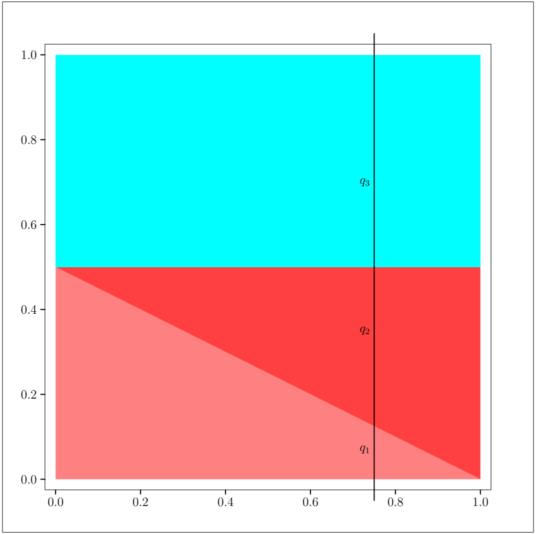
Intuition T1: use the space originally occupied by the affected probabilities and redistribute it:

$$v_1 = (0.125, 0.375, 0.500)$$

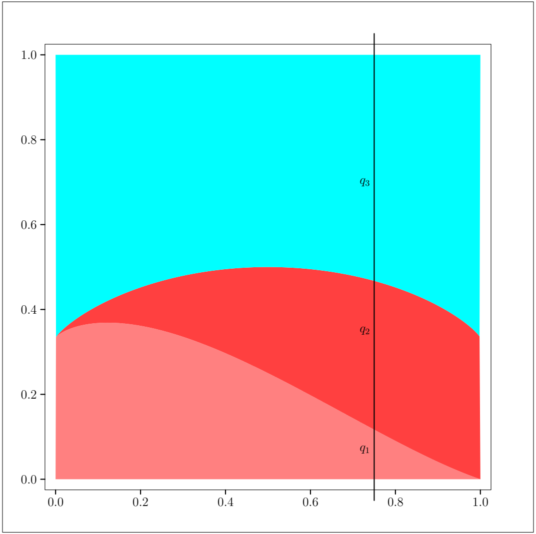
Intuition T2: use the constraint rule to determine which probability assignment is consistent with the the constraint while minimizing the informational divergence with the prior probability:

$$v_2 \approx (0.117, 0.350, 0.533)$$

Intuition T1



Intuition T2



Four Scenarios I

Here are four scenarios, in which Judy may have received (HDQ):

S1 Judy was dropped off by a pilot who flipped two coins. If the first coin landed H, then Judy was dropped off in Blue territory, otherwise in Red territory. If the second coin landed H, she was dropped off on Headquarters ground, otherwise on Second Company ground. Judy's headquarters find out that the second coin was biased $q : 1 - q$ toward H. The normalized odds vector is $v = (.125, .375, .5)$ and agrees with (T1), because the choice of Blue or Red is completely independent from the choice of Headquarters or Second Company.

Four Scenarios II

S2 Judy's Headquarters has divided the map into 24 congruent rectangles, A_3 into twelve, and A_1 and A_2 into six rectangles each. They have information that the only subsets of the 24 rectangles in which Judy Benjamin may be located are such that they contain three times as many A_2 rectangles than A_1 rectangles. Thus they relay (HDQ) to Judy, which she correctly interprets with the normalized odds vector $v = (.108, .324, .568)$ (evaluating the 16777216 subsets).

Four Scenarios III-IV

S3 The pilot randomly lands in any of the four quadrants and rolls a die. If she rolls an even number, she drops off Judy. If not, she takes her to another (or the same) randomly selected quadrant to repeat the procedure. Headquarters find out, however, that for A_1, the pilot requires a six to drop off Judy, not just an even number. Thus they relay (HDQ) to Judy, which she correctly interprets with the normalized odds vector $v = (.1, .3, .6)$.

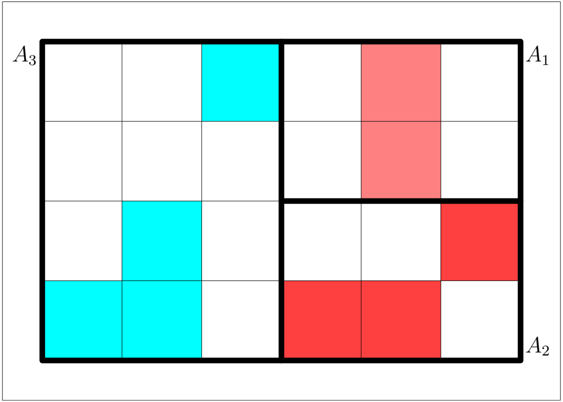
S4 Judy's headquarters know that Judy is not in A_3 and that her chance of being in A_2 is three times the chance of being in A_1. They only succeed, however, in informing Judy of the second part of the message. If Judy had all the information, her normalized odds vector would be $v = (.33, .67, 0)$.

Powerset Approach I

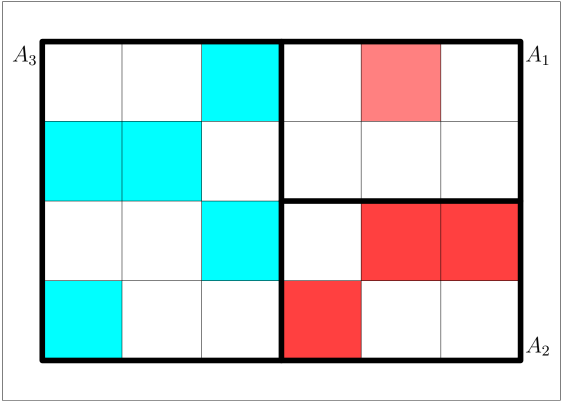
Reconsider Scenario S2: Judy's Headquarters has divided the map into 24 congruent rectangles, A_3 into twelve, and A_1 and A_2 into six rectangles each. They have information that the only subsets of the 24 rectangles in which Judy Benjamin may be located are such that they contain three times as many A_2 rectangles than A_1 rectangles. Thus they relay (HDQ) to Judy, which she correctly interprets with the normalized odds vector $v = (.108, .324, .568)$ (evaluating the 16777216 subsets).

Now let the grain of this partition become infinitely small. There are strong independence and uniformity assumptions at work here. It is reasonable to predict that the results of the powerset approach will support intuition T1.

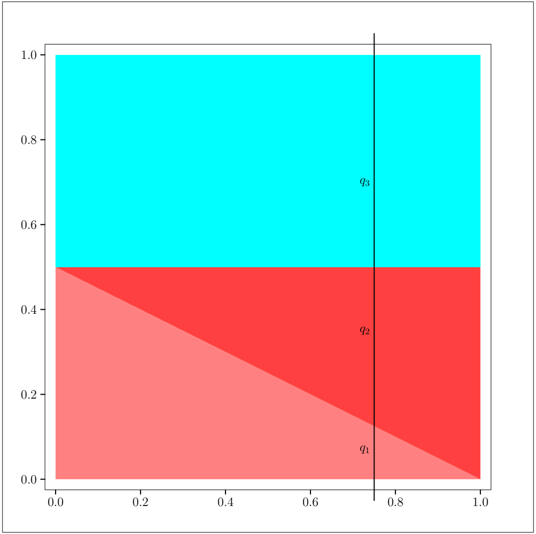
Powerset Approach II



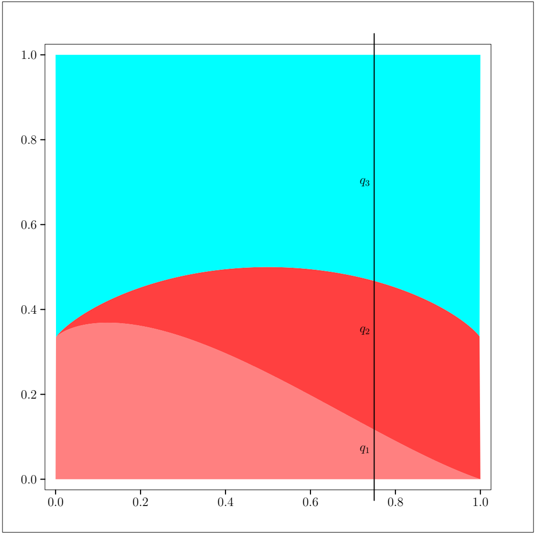
Powerset Approach III



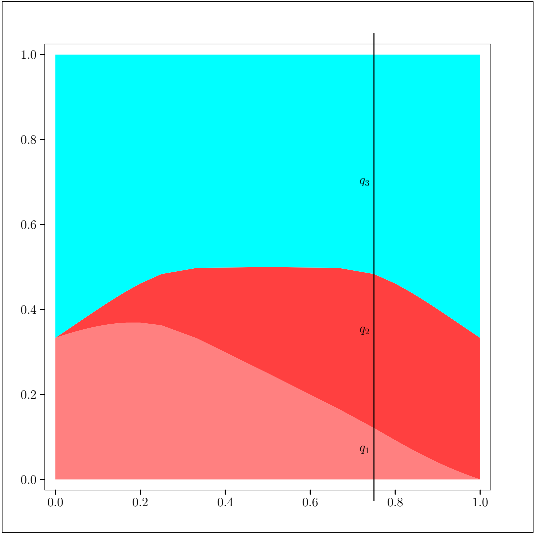
Intuition T1



Intuition T2



Powerset Approach



Epistemic Entrenchment I

Sarah and Marian have arranged to go for sundowners at the Westcliff hotel tomorrow. Sarah feels there is some chance that it will rain, but thinks they can always enjoy the view from inside. To make sure, Marian consults the staff at the Westcliff hotel and finds out that in the event of rain, the inside area will be occupied by a wedding party. So she tells Sarah: "If it rains tomorrow, we cannot have sundowners at the Westcliff." Upon learning this conditional, Sarah sets her probability for sundowners and rain to zero, but she does not adapt her probability for rain.

A jeweller has been shot in his store and robbed of a golden watch. However, it is not clear at this point what the relation between these two events is; perhaps someone shot the jeweller and then someone else saw an opportunity to steal the watch. Kate thinks there is some chance that Henry is the robber. On the other hand, she strongly doubts that he is capable of shooting someone, and thus, that he is the shooter. Now the inspector, after hearing the testimonies of several witnesses, tells Kate: "If Henry robbed the jeweller, then he also shot him." As a result, Kate becomes more confident that Henry is not the robber, while her probability for Henry having shot the jeweller does not change.

Epistemic Entrenchment III

Leaving the Antecedent Alone

"In most cases the learning of a conditional is or would be irrelevant to one's degree of belief for the conditional's antecedent . . . the learning of the relevant conditional should intuitively leave the probability of the antecedent unaltered."
(Douven and Romeijn)

Coarsening at Random I

The CAR (Coarsening at Random) condition specifies when conditioning in a naive space works. A well-known example for conditioning in a naive space is the Monty Hall puzzle. We will consider Martin Gardner's Three Prisoners Puzzle, which is very similar.

The Three Prisoners Puzzle

Three men (A, B, and C) are under sentence of death when the governor decides to pardon one of them. The warden of the prison knows which of the three men is pardoned, but none of the men do. In a private conversation, A says to the warden, Tell me the name of one of the others who will be executed—it will not give anything away whether I will be executed or not. The warden agrees and tells A that B will be executed.

Coarsening at Random II

observation	*update rule*	*coincidence*
event and pairwise disjoint	conditioning	always
event and arbitrary set	conditioning	iff CAR holds
vector and partition	Jeffrey conditioning	iff generalization of CAR holds
vector and no full partition	principle of MAXENT	only in degenerate case

Table: Coincidence of naive and sophisticated conditioning

Coarsening at Random III

CAR does not hold for Judy Benjamin

The set of observations for which [PME] conditioning corresponds to conditioning in the sophisticated space is a (Lebesgue) measure 0 set in the space of possible observations . . . Seidenfeld shows that, under very weak conditions, [PME] updating cannot coincide with sophisticated conditioning if the observations have the form "the conditional probability of U given V is α (as is the case in the Judy Benjamin problem). (Grünwald and Halpern)

Coarsening at Random IV

This puts us in an awkward position.

Coarsening at Random V

$$P(\text{`A is pardoned'}|\text{`B will be executed'}) =$$

$$\frac{P(\text{`A is pardoned'})}{P(\text{`A is pardoned'}) + P(\text{`C is pardoned'})} = \frac{1}{2} \text{ (incorrect)}$$

Coarsening at Random VI

$$P(\text{'A is pardoned'}|\text{'warden says B will be executed'}) =$$

$$\frac{P(\text{'A is pardoned' and 'warden says B will be executed'})}{P(\text{'warden says B will be executed'})} =$$

$$\frac{1/6}{1/2} = \frac{1}{3} \text{ (correct)}$$

Coarsening at Random VII

Three Prisoners Puzzle	*Judy Benjamin Problem*
naive space (informationally incorrect)	naive space
sophisticated space (informationally correct)	no sophisticated space without retrospective conditioning

Table: The disanalogy between the Three Prisoners and Judy Benjamin

End of Presentation