Basic Probability Theory (II)

Adrian Brasoveanu

[partly based on slides by Robert Henderson]

April 22, 2014

1. Probability: Frequency vs Reasonable Expectation

2. Generalizing Classical Logic

3. Patterns of Plausible Inference
    - Modus Ponens
    - Modus Tollens
    - Affirming the Consequent
    - Denying the Antecedent
    - Affirming the Consequent of Weaker/Plausible Implications

4. Probability Theory as the Logic of Data Analysis

Probability and the Structure of Plausible Inference

"Probability is not really about numbers but about the structure of reasoning." (Glenn Shafer, cited in Pearl 1988, p. 77)
Reasoning about what? We implicitly took a frequentist perspective:

- Probability is plausible reasoning about (hypothetical) repeated sampling under (basically) identical conditions and about the long-term frequencies of the sample statistics (i.e., the sample features of interest) that arise.
- E.g., repeated coin flips or repeated sequences thereof, repeated sampling of students to study their eye and hair color, repeated measurement of reading time for sentences with two quantifiers etc.

"The significant point is that the initial circumstances [under which we obtained the actual data/sample] are assumed to be capable of indefinite repetition." (Cox 1946, p. 1)

The Bayesian answer is that probability is reasoning about the plausibility of propositions/beliefs on their own and given other propositions/beliefs:

- E.g., a box contains 2 white balls and 1 black ball and we are exclusively concerned with a **single** trial in which a blindfolded man extracts a ball from the box. How confident are we that the proposition *A white ball is extracted* is true?
- A **reasonable expectation** about/confidence in the truth of this proposition is $2/3$, i.e., the same answer as the frequentist would give.
- Our corresponding confidence in the truth of the proposition *A black ball is extracted* is $1/3$.
- Our odds, i.e., the relative confidence in the truth of these propositions, are the ratio of our reasonable expectations about their truth, i.e., of our probabilities: $\frac{2/3}{1/3} = 2/1$.

"If it could be shown that every measure of reasonable expectation is also a frequency [. . .] and that every frequency [. . .] measures a reasonable expectation, then the choice of one or the other as the primary meaning of probability would not be very important." (Cox 1946, p. 2) But:

Probability and the Structure of Plausible Inference

"If it could be shown that every measure of reasonable expectation is also a frequency [...] and that every frequency [...] measures a reasonable expectation, then the choice of one or the other as the primary meaning of probability would not be very important." (Cox 1946, p. 2) But:

- "There are probabilities in the sense of reasonable expectations for which no [frequency] exists and [...] if one is conceived, it is clearly no more than a convenient mental artifice." (Cox 1946, p. 2) We can decide that probability theory should be used for these too – the Bayesian choice, or that it shouldn't – the frequentist choice.

Probability and the Structure of Plausible Inference

"If it could be shown that every measure of reasonable expectation is also a frequency [. . .] and that every frequency [. . .] measures a reasonable expectation, then the choice of one or the other as the primary meaning of probability would not be very important." (Cox 1946, p. 2) But:

- "There are probabilities in the sense of reasonable expectations for which no [frequency] exists and [. . .] if one is conceived, it is clearly no more than a convenient mental artifice." (Cox 1946, p. 2) We can decide that probability theory should be used for these too – the Bayesian choice, or that it shouldn't – the frequentist choice.
- "Moreover, there is so gradual a transition from the cases in which there is a discoverable [frequency] and those in which there is none that a theory that requires a sharp distinction between them [has] difficulties" (Cox 1946, p. 2) This is a conceptual argument for the Bayesian choice.

Probability and the Structure of Plausible Inference

Consider for example (again, all from Cox 1946):

1. the probability that the number of heads thrown in a certain number of tosses of an unbiased coin lies within certain limits—the number of heads varies from one trial to another and frequencies and reasonable expectations are basically identical here

   just as in the case of the box with 2 white balls and 1 black ball

Probability and the Structure of Plausible Inference

Consider for example (again, all from Cox 1946):

1. the probability that the number of heads thrown in a certain number of tosses of an unbiased coin lies within certain limits—the number of heads varies from one trial to another and frequencies and reasonable expectations are basically identical here

   just as in the case of the box with 2 white balls and 1 black ball

2. the probability that the true value of a physical constant lies within certain limits—the value of the constant is unique, we speak of probability here only because our knowledge is incomplete, i.e., only as reasonable expectation

Consider for example (again, all from Cox 1946):

1. the probability that the number of heads thrown in a certain number of tosses of an unbiased coin lies within certain limits—the number of heads varies from one trial to another and frequencies and reasonable expectations are basically identical here

   just as in the case of the box with 2 white balls and 1 black ball
2. the probability that the true value of a physical constant lies within certain limits—the value of the constant is unique, we speak of probability here only because our knowledge is incomplete, i.e., only as reasonable expectation

   although this probability might be equivalent to another probability that the error of the average of a number of measurements lies within certain limits—which can be easily understood in terms of frequencies

Finally, consider:

3. the case of a purely mathematical constant whose existence has been proved but the value determined only within certain limits, e.g., that large enough integers can be expressed as sums of a small number of cubes, namely 4, 5, 6, 7 or 8, but we don't know which;

Finally, consider:

3. the case of a purely mathematical constant whose existence has been proved but the value determined only within certain limits, e.g., that large enough integers can be expressed as sums of a small number of cubes, namely 4, 5, 6, 7 or 8, but we don't know which;

   evidence based on computations indicates that integers requiring 8 cubes drop out early in the progress to higher integers, those requiring 7 cubes disappear later on etc.

Finally, consider:

3. the case of a purely mathematical constant whose existence has been proved but the value determined only within certain limits, e.g., that large enough integers can be expressed as sums of a small number of cubes, namely 4, 5, 6, 7 or 8, but we don't know which;

   evidence based on computations indicates that integers requiring 8 cubes drop out early in the progress to higher integers, those requiring 7 cubes disappear later on etc.

   we can speak of the probability that the least number of cubes necessary for expressing large enough integers is 4;

Finally, consider:

3. the case of a purely mathematical constant whose existence has been proved but the value determined only within certain limits, e.g., that large enough integers can be expressed as sums of a small number of cubes, namely 4, 5, 6, 7 or 8, but we don't know which;

   evidence based on computations indicates that integers requiring 8 cubes drop out early in the progress to higher integers, those requiring 7 cubes disappear later on etc.

   we can speak of the probability that the least number of cubes necessary for expressing large enough integers is 4;

   this is a **reasonable expectation/belief in a proposition**; it's much harder to find an equivalent probability that can be readily understood in terms of repeated sampling and corresponding frequencies.

"It must be admitted that there is a kind of reasoning common to all these examples. The gambler in the first example, the physicist in the second [...] and the mathematician in the [third] are all using similar processes of inference. [...]

"It must be admitted that there is a kind of reasoning common to all these examples. The gambler in the first example, the physicist in the second [. . .] and the mathematician in the [third] are all using similar processes of inference. [. . .]

This difficulty of the frequency theory of probability may now be summarized. There is a field of probable inference which lies outside the range of that theory. The derivation of the rules of probability by ordinary algebra from the characteristics of [repeated sampling] cannot justify the use of these rules in this outside field.

Probability and the Structure of Plausible Inference

"It must be admitted that there is a kind of reasoning common to all these examples. The gambler in the first example, the physicist in the second [...] and the mathematician in the [third] are all using similar processes of inference. [...]

This difficulty of the frequency theory of probability may now be summarized. There is a field of probable inference which lies outside the range of that theory. The derivation of the rules of probability by ordinary algebra from the characteristics of [repeated sampling] cannot justify the use of these rules in this outside field.

The use of these rules in this field appears to be a fundamental part of our reasoning. Thus, the frequency theory [...] fails to justify what is conceived to be a legitimate use of its own rules." (Cox 1946, pp. 3-4)

Probability and the Structure of Plausible Inference

"In what follows, I [. . .] show that by employing the algebra of symbolic logic [i.e., classical sentential logic] it is possible to derive the rules of probability from two quite primitive notions, which are independent of the concept of [frequency] and which, as I think, appeal rather immediately to common sense." (Cox 1946, pp. 3-4)

We assume the following classical sentential operators:

- negation $\sim$, e.g., $\sim\varphi$
- conjunction $\cdot$, e.g., $\varphi \cdot \psi$
- disjunction $\vee$, e.g., $\varphi \vee \psi$

Their probabilistic interpretation satisfies the expected standard equivalences:

$$p({\sim}{\sim}\varphi) = p(\varphi)$$

$$p(\varphi \cdot \varphi) = p(\varphi) \qquad p(\varphi \vee \varphi) = p(\varphi)$$

$$p((\varphi \cdot \psi) \cdot \chi) = p(\psi \cdot (\varphi \cdot \chi)) \qquad p((\varphi \vee \psi) \vee \chi) = p(\psi \vee (\varphi \vee \chi))$$

$$p(\varphi \cdot \psi) = p(\psi \cdot \varphi) \qquad p(\varphi \vee \psi) = p(\psi \vee \varphi)$$

$$p({\sim}(\varphi \cdot \psi)) = p({\sim}\varphi \vee {\sim}\psi) \qquad p({\sim}(\varphi \vee \psi)) = p({\sim}\varphi \cdot {\sim}\psi)$$

$$p(\varphi \cdot (\varphi \vee \psi)) = p(\varphi) \qquad p(\varphi \vee (\varphi \cdot \psi)) = p(\varphi)$$

To this, we add the conditional probability operator $\mid$, e.g., $\varphi \mid \psi$.

To derive the rules of probability, Cox (1946) shows that we only need to make the following two assumptions about the meaning of conditionalization:

- $p(\chi \cdot \psi \mid \varphi)$ is uniquely determined by (is a function of) $p(\psi \mid \varphi)$ and $p(\chi \mid \psi \cdot \varphi)$
- $p(\sim\psi \mid \varphi)$ is uniquely determined by (is a function of) $p(\psi \mid \varphi)$

Probability and the Structure of Plausible Inference

It is intuitively clear that $p(\sim\psi \mid \varphi)$, i.e., the probability/reasonable expectation about the truth of $\sim\psi$ given φ, should be uniquely determined by the probability of ψ given φ.
Any other assumption would be at least as complicated.

Probability and the Structure of Plausible Inference

It is intuitively clear that $p(\sim\psi \mid \varphi)$, i.e., the probability/reasonable expectation about the truth of $\sim\psi$ given φ, should be uniquely determined by the probability of ψ given φ.
Any other assumption would be at least as complicated.

But what about the other assumption? Why not let $p(\chi \cdot \psi \mid \varphi)$ be uniquely determined solely by $p(\psi \mid \varphi)$ and $p(\chi \mid \varphi)$?

Probability and the Structure of Plausible Inference

It is intuitively clear that $p(\sim\psi \mid \varphi)$, i.e., the probability/reasonable expectation about the truth of $\sim\psi$ given φ, should be uniquely determined by the probability of ψ given φ.
Any other assumption would be at least as complicated.

But what about the other assumption? Why not let $p(\chi \cdot \psi \mid \varphi)$ be uniquely determined solely by $p(\psi \mid \varphi)$ and $p(\chi \mid \varphi)$?

Because ψ and χ might not be independent: "It is plausible that the next person you meet has a brown right eye. It is plausible that the next person you meet has a blue left eye. But it is not plausible at all that the next person you meet will have a brown right eye and a blue left eye." (Jaynes 2003)

Probability and the Structure of Plausible Inference

It is intuitively clear that $p(\sim\psi \mid \varphi)$, i.e., the probability/reasonable expectation about the truth of $\sim\psi$ given φ, should be uniquely determined by the probability of ψ given φ.
Any other assumption would be at least as complicated.

But what about the other assumption? Why not let $p(\chi \cdot \psi \mid \varphi)$ be uniquely determined solely by $p(\psi \mid \varphi)$ and $p(\chi \mid \varphi)$?

Because ψ and χ might not be independent: "It is plausible that the next person you meet has a brown right eye. It is plausible that the next person you meet has a blue left eye. But it is not plausible at all that the next person you meet will have a brown right eye and a blue left eye." (Jaynes 2003)

But if they are independent, the assumption can be automatically simplified: "It is plausible that the next person you meet has blue eyes. It is plausibile that the next person you meet has black hair. It is reasonably plausible that the next person you meet will have blue eyes and black hair." (Jaynes 2003) [ignoring what we learned about eye and hair color from Snee 1974]

Deductive Inference is Preserved

In the resulting probabilistic logic, deductive reasoning is preserved: it is the limiting form of reasoning under uncertainty.

Deductive Inference is Preserved

In the resulting probabilistic logic, deductive reasoning is preserved: it is the limiting form of reasoning under uncertainty.

- Modeling weak/plausible inference patterns doesn't mean we make the deductive part of our logic weaker, i.e., we reason over certainty (classical truth) and impossibility (classical falsity) in the same way.

Deductive Inference is Preserved

In the resulting probabilistic logic, deductive reasoning is preserved: it is the limiting form of reasoning under uncertainty.

- Modeling weak/plausible inference patterns doesn't mean we make the deductive part of our logic weaker, i.e., we reason over certainty (classical truth) and impossibility (classical falsity) in the same way.
- We retain propositional logic validities and deductive inference patterns.

Deductive Inference is Preserved

In the resulting probabilistic logic, deductive reasoning is preserved: it is the limiting form of reasoning under uncertainty.

- Modeling weak/plausible inference patterns doesn't mean we make the deductive part of our logic weaker, i.e., we reason over certainty (classical truth) and impossibility (classical falsity) in the same way.
- We retain propositional logic validities and deductive inference patterns.
- Classical logic is not inconsistent with how we reason probabilistically, just insufficient for modeling the entire space of inferences we make.

Deductive Inference is Preserved: Modus Ponens

Modus ponens

	$A \rightarrow B$	If Digby is a dog, then Digby likes french fries.
	A	Digby is a dog.
$\models$	B	Therefore, Digby likes french fries.

Deductive Inference is Preserved: Modus Ponens

Modus ponens

	$A \rightarrow B$	If Digby is a dog, then Digby likes french fries.
	A	Digby is a dog.
$\models$	B	Therefore, Digby likes french fries.

We let $A \rightarrow B := \sim A \vee B$.

- $p(A \cdot B \mid A \rightarrow B) = p(A \mid A \rightarrow B)p(B \mid A \cdot (A \rightarrow B))$ (mult. rule)
- hence: $p(B \mid A \cdot (A \rightarrow B)) = \frac{p(A \cdot B \mid A \rightarrow B)}{p(A \mid A \rightarrow B)} = \frac{p(A \cdot B \mid \sim A \vee B)}{p(A \mid \sim A \vee B)}$
- $p(A \cdot B \mid \sim A \vee B) = \frac{p(A \cdot B \cdot (\sim A \vee B))}{p(\sim A \vee B)} = \frac{p((A \cdot B \cdot \sim A) \vee (A \cdot B \cdot B))}{p(\sim A \vee B)} = \frac{p((A \cdot \sim A) \vee (A \cdot B))}{p(\sim A \vee B)} = \frac{p(A \cdot (\sim A \vee B))}{p(\sim A \vee B)} = p(A \mid \sim A \vee B)$
- so: $p(B \mid A \cdot (\sim A \vee B)) = p(B \mid A \cdot (A \rightarrow B)) = 1$

Deductive Inference is Preserved: Modus Tollens

Modus tollens

	$A \to B$	If Digby makes his own french fries, then Digby is smarter than the average dog.
	$\sim B$	Digby is not smarter than the average dog.
$\models$	$\sim A$	Therefore, Digby does not make his own french fries.

Deductive Inference is Preserved: Modus Tollens

Modus tollens

	$A \rightarrow B$	If Digby makes his own french fries, then Digby is smarter than the average dog.
	$\sim B$	Digby is not smarter than the average dog.
$\models$	$\sim A$	Therefore, Digby does not make his own french fries.

Again, we let $A \rightarrow B := \sim A \vee B$.

- $p(A \cdot \sim B \mid A \rightarrow B) = p(\sim B \mid A \rightarrow B) p(A \mid \ \sim B \cdot (A \rightarrow B))$ (mult. rule)
- hence: $p(A \mid \ \sim B \cdot (A \rightarrow B)) = \frac{p(A \cdot \sim B \mid A \rightarrow B)}{p(\sim B \mid A \rightarrow B)} = \frac{p(A \cdot \sim B \mid \sim A \vee B)}{p(\sim B \mid \sim A \vee B)}$
- $p(A \cdot \sim B \mid \ \sim A \vee B) = \frac{p(A \cdot \sim B \cdot (\sim A \vee B))}{p(\sim A \vee B)} = \frac{p((A \cdot \sim B \cdot \sim A) \vee (A \cdot \sim B \cdot B))}{p(\sim A \vee B)} = \frac{p((A \cdot \sim A) \vee (\sim B \cdot B))}{p(\sim A \vee B)} = \frac{0}{p(\sim A \vee B)} = 0$
- so: $p(A \mid \ \sim B \cdot (A \rightarrow B)) = \frac{0}{p(\sim B \mid \sim A \vee B)} = 0$

Patterns of Plausible Inference Beyond Classical Logic

We also get patterns of non-deductive, but **plausible** inference that we use in scientific or common-sense reasoning.

Patterns of Plausible Inference Beyond Classical Logic

We also get patterns of non-deductive, but **plausible** inference that we use in scientific or common-sense reasoning.

Two simple ones first:

- if $p(A \mid C \cdot D) > p(A \mid C)$ and $p(B \mid A \cdot C \cdot D) = p(B \mid A \cdot C)$, then $p(A \cdot B \mid C \cdot D) \geq p(A \cdot B \mid C)$ – if we get additional information D (relative to the current info C) that makes A more plausible, but the plausibility of B remains the same, then $A \cdot B$ can be no less plausible than it was before we learned D

Patterns of Plausible Inference Beyond Classical Logic

We also get patterns of non-deductive, but **plausible** inference that we use in scientific or common-sense reasoning.

Two simple ones first:

- if $p(A \mid C \cdot D) > p(A \mid C)$ and $p(B \mid A \cdot C \cdot D) = p(B \mid A \cdot C)$, then $p(A \cdot B \mid C \cdot D) \geq p(A \cdot B \mid C)$ – if we get additional information D (relative to the current info C) that makes A more plausible, but the plausibility of B remains the same, then $A \cdot B$ can be no less plausible than it was before we learned D

- if $p(A \mid C \cdot D) > p(A \mid C)$, then $p({\sim}A \mid C \cdot D) < p({\sim}A \mid C)$ – additional information that makes A more plausible automatically makes ${\sim}A$ less plausible.

Patterns of Plausible Inference Beyond Classical Logic

We also get patterns of non-deductive, but **plausible** inference that we use in scientific or common-sense reasoning.

Two simple ones first:

- if $p(A \mid C \cdot D) > p(A \mid C)$ and $p(B \mid A \cdot C \cdot D) = p(B \mid A \cdot C)$, then $p(A \cdot B \mid C \cdot D) \geq p(A \cdot B \mid C)$ – if we get additional information D (relative to the current info C) that makes A more plausible, but the plausibility of B remains the same, then $A \cdot B$ can be no less plausible than it was before we learned D

- if $p(A \mid C \cdot D) > p(A \mid C)$, then $p(\sim A \mid C \cdot D) < p(\sim A \mid C)$ – additional information that makes A more plausible automatically makes $\sim A$ less plausible.

  "When you have eliminated the impossible, whatever remains, however improbable, must be the truth": when we reduce belief in certain possibilities, we necessarily increase belief in the remaining ones.

Patterns of Plausible Inference Beyond Classical Logic

Affirming the Consequent

	$A \rightarrow B$	If it will start to rain at 10 AM tomorrow, the sky will be cloudy shortly before 10 AM.
	B	The sky is cloudy shortly before 10 AM.
$\models$	A	Therefore, (it is more plausible that) it will start to rain at 10 AM.

Affirming the Consequent

	$A \to B$	If it will start to rain at 10 AM tomorrow, the sky will be cloudy shortly before 10 AM.
	B	The sky is cloudy shortly before 10 AM.
$\models$	A	Therefore, (it is more plausible that) it will start to rain at 10 AM.

- $p(A \mid A \to B)p(B \mid A \cdot (A \to B)) =$
  $p(B \mid A \to B)p(A \mid B \cdot (A \to B))$ (mult. rule)
- hence: $\frac{p(A \mid B \cdot (A \to B))}{p(A \mid A \to B)} = \frac{p(B \mid A \cdot (A \to B))}{p(B \mid A \to B)}$
- $p(B \mid A \cdot (A \to B)) = 1$ (modus ponens) and
  $p(B \mid A \to B) \leq 1$
- so $p(A \mid B \cdot (A \to B)) \geq p(A \mid A \to B)$

Patterns of Plausible Inference Beyond Classical Logic

Denying the Antecedent

	$A \rightarrow B$	If Digby rolls over on command, Ryan gives him a treat.
	$\sim A$	Digby didn't roll over on command.
$\models$	$\sim B$	Therefore, Ryan didn't give him a treat.

Denying the Antecedent

	$A \to B$	If Digby rolls over on command, Ryan gives him a treat.
	$\sim A$	Digby didn't roll over on command.
$\models$	$\sim B$	Therefore, Ryan didn't give him a treat.

- $p(\sim A \mid A \to B)p(B \mid \sim A \cdot (A \to B)) =$ $p(B \mid A \to B)p(\sim A \mid B \cdot (A \to B))$ (mult. rule)
- hence: $\frac{p(B \mid \sim A \cdot (A \to B))}{p(B \mid A \to B)} = \frac{p(\sim A \mid B \cdot (A \to B))}{p(\sim A \mid A \to B)}$
- $p(A \mid B \cdot (A \to B)) \geq p(A \mid A \to B)$ (affirm. conseq.), hence $p(\sim A \mid B \cdot (A \to B)) \leq p(\sim A \mid A \to B)$
- therefore, $p(B \mid \sim A \cdot (A \to B)) \leq p(B \mid A \to B)$
- so $p(\sim B \mid \sim A \cdot (A \to B)) \geq p(\sim B \mid A \to B)$

Patterns of Plausible Inference Beyond Classical Logic

While **Affirming the Consequent** and **Denying the Antecedent** are not valid in classical sentential logic, they are intuitively justified.

Patterns of Plausible Inference Beyond Classical Logic

While **Affirming the Consequent** and **Denying the Antecedent** are not valid in classical sentential logic, they are intuitively justified.

They become even better when insert a probability adverbial in the consequent.

- "If it will start to rain at 10 AM tomorrow, the sky will be cloudy shortly before 10 AM. The sky is cloudy shortly before 10 AM. Therefore, it will **probably** rain."

Patterns of Plausible Inference Beyond Classical Logic

While **Affirming the Consequent** and **Denying the Antecedent** are not valid in classical sentential logic, they are intuitively justified.

They become even better when insert a probability adverbial in the consequent.

- “If it will start to rain at 10 AM tomorrow, the sky will be cloudy shortly before 10 AM. The sky is cloudy shortly before 10 AM. Therefore, it will **probably** rain.”
- “If Digby rolls over on command, Ryan gives him a treat. Digby didn’t roll over on command. Therefore, Ryan **probably** didn’t give him a treat.”

Patterns of Plausible Inference Beyond Classical Logic

While **Affirming the Consequent** and **Denying the Antecedent** are not valid in classical sentential logic, they are intuitively justified.

They become even better when insert a probability adverbial in the consequent.

- "If it will start to rain at 10 AM tomorrow, the sky will be cloudy shortly before 10 AM. The sky is cloudy shortly before 10 AM. Therefore, it will **probably** rain."
- "If Digby rolls over on command, Ryan gives him a treat. Digby didn't roll over on command. Therefore, Ryan **probably** didn't give him a treat."

We like even weaker inference patterns . . .

Affirming the Conseq. of Weaker/Plausible Implications

	$A \rightarrow_{\text{PLAUSIBLY}} B$	It's plausible that if a man has recently escaped from prison, he'll be wearing handcuffs and an orange jumpsuit.
	B	That man is wearing handcuffs and an orange jumpsuit.
$\models_{\text{PLAUSIBLY}}$	A	Therefore, he plausibly recently escaped from prison.

Patterns of Plausible Inference Beyond Classical Logic

Affirming the Conseq. of Weaker/Plausible Implications

	$A \rightarrow_{\text{PLAUSIBLY}} B$	It's plausible that if a man has recently escaped from prison, he'll be wearing handcuffs and an orange jumpsuit.
	B	That man is wearing handcuffs and an orange jumpsuit.
$\models_{\text{PLAUSIBLY}}$	A	Therefore, he plausibly recently escaped from prison.

Interpret $A \rightarrow_{\text{PLAUSIBLY}} B$ given background information C as:
$p(B \mid A \cdot C) > p(B \mid C)$.

Patterns of Plausible Inference Beyond Classical Logic

Affirming the Conseq. of Weaker/Plausible Implications

	$A \rightarrow_{\text{PLAUSIBLY}} B$	It's plausible that if a man has recently escaped from prison, he'll be wearing handcuffs and an orange jumpsuit.
	B	That man is wearing handcuffs and an orange jumpsuit.
$\models_{\text{PLAUSIBLY}}$	A	Therefore, he plausibly recently escaped from prison.

Interpret $A \rightarrow_{\text{PLAUSIBLY}} B$ given background information C as:
$p(B \mid A \cdot C) > p(B \mid C)$.
Interpret the conclusion of the argument given the premise B as:
$p(A \mid B \cdot C) > p(A \mid C)$. This is what we need to establish.

Affirming the Conseq. of Weaker/Plausible Implications

	$A \rightarrow_{\text{PLAUSIBLY}} B$	It's plausible that if a man has recently escaped from prison, he'll be wearing handcuffs and an orange jumpsuit.
	B	That man is wearing handcuffs and an orange jumpsuit.
$\models_{\text{PLAUSIBLY}}$	A	Therefore, he plausibly recently escaped from prison.

Interpret $A \rightarrow_{\text{PLAUSIBLY}} B$ given background information C as:
$p(B \mid A \cdot C) > p(B \mid C)$.
Interpret the conclusion of the argument given the premise B as:
$p(A \mid B \cdot C) > p(A \mid C)$. This is what we need to establish.

- $p(A \mid C) \cdot p(B \mid A \cdot C) = p(B \mid C) \cdot p(A \mid B \cdot C)$ (mult. rule)

Patterns of Plausible Inference Beyond Classical Logic

Affirming the Conseq. of Weaker/Plausible Implications

	$A \rightarrow_{\text{PLAUSIBLY}} B$	It's plausible that if a man has recently escaped from prison, he'll be wearing handcuffs and an orange jumpsuit.
	B	That man is wearing handcuffs and an orange jumpsuit.
$\models_{\text{PLAUSIBLY}}$	A	Therefore, he plausibly recently escaped from prison.

Interpret $A \rightarrow_{\text{PLAUSIBLY}} B$ given background information C as: $p(B \mid A \cdot C) > p(B \mid C)$.

Interpret the conclusion of the argument given the premise B as: $p(A \mid B \cdot C) > p(A \mid C)$. This is what we need to establish.

- $p(A \mid C) \cdot p(B \mid A \cdot C) = p(B \mid C) \cdot p(A \mid B \cdot C)$ (mult. rule)
- hence: $\frac{p(A \mid C)}{p(A \mid B \cdot C)} = \frac{p(B \mid C)}{p(B \mid A \cdot C)}$

Affirming the Conseq. of Weaker/Plausible Implications

	$A \rightarrow_{\text{PLAUSIBLY}} B$	It's plausible that if a man has recently escaped from prison, he'll be wearing handcuffs and an orange jumpsuit.
	B	That man is wearing handcuffs and an orange jumpsuit.
$\models_{\text{PLAUSIBLY}}$	A	Therefore, he plausibly recently escaped from prison.

Interpret $A \rightarrow_{\text{PLAUSIBLY}} B$ given background information C as:
$p(B \mid A \cdot C) > p(B \mid C)$.
Interpret the conclusion of the argument given the premise B as:
$p(A \mid B \cdot C) > p(A \mid C)$. This is what we need to establish.

- $p(A \mid C) \cdot p(B \mid A \cdot C) = p(B \mid C) \cdot p(A \mid B \cdot C)$ (mult. rule)
- hence: $\frac{p(A \mid C)}{p(A \mid B \cdot C)} = \frac{p(B \mid C)}{p(B \mid A \cdot C)}$
- since $p(B \mid C) < p(B \mid A \cdot C)$, we have that $p(A \mid C) < p(A \mid B \cdot C)$

While none of these inferences are classically valid, they are fine if our goal is inference to the best explanation and not logically validity.

- common sense = probabilistic inference + prior information

But not only common-sense:

- scientific investigation (e.g., data analysis) = probabilistic inference + prior info that a skeptical audience agrees with

Probability Theory as the Logic of Data Analysis

While none of these inferences are classically valid, they are fine if our goal is inference to the best explanation and not logically validity.

- common sense = probabilistic inference + prior information

But not only common-sense:

- scientific investigation (e.g., data analysis) = probabilistic inference + prior info that a skeptical audience agrees with

E.g., hierarchical/'random-effects' models are simply conjoined conditional probabilities of the form $p(A|B)p(B|C)$ in which we have an 'intermediate' reasoning layer B, e.g.:

- the unobserved differences between subjects/items B (the subject/item 'random-effects')
- ... that are conditional on the unobserved parameters C of the subject/item population
- ... and that condition the observed experimental data A

Cox, R. T. (1946). "Probability, Frequency and Reasonable Expectation". In: *American Journal of Physics* 14.1, pp. 1–13.

Jaynes, Edwin T. (2003). *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press.

Pearl, Judea (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufman.

Snee, R. D. (1974). "Graphical display of two-way contingency tables". In: *The American Statistician* 38, pp. 9–12.