Bruno de Finetti's Objectivity

Philip Dawid[1]

de Finetti Lecture, ISBA 2018, Edinburgh

25 June 2018

[1]University of Cambridge

Bruno de Finetti (1906–1985) in 1979

Objectivism vs. objectivity

Objectivism vs. objectivity

Objectivism?

- No!

Objectivism vs. objectivity

Objectivism?

- No!
- de Finetti rejects the idea of probability as a feature of the external world

(1931b) "Objective probability never exists"

Objectivism vs. objectivity

Objectivism?

- No!
- de Finetti rejects the idea of probability as a feature of the external world

  (1931b) "Objective probability never exists"

- Rather, it is a feature of my description of the world

Objectivism vs. objectivity

Objectivism?

- No!
- de Finetti rejects the idea of probability as a feature of the external world

  (1931b) "Objective probability never exists"

- Rather, it is a feature of my description of the world
- Subjective/Personal Probability

Objectivism vs. objectivity

Objectivism?

- No!
- de Finetti rejects the idea of probability as a feature of the external world

  (1931b) "Objective probability never exists"

- Rather, it is a feature of my description of the world
- Subjective/Personal Probability

Objective (objectionable?) Bayes?

- No!

Objectivism vs. objectivity

Objectivism?
- No!
- de Finetti rejects the idea of probability as a feature of the external world

  (1931b) "Objective probability never exists"

- Rather, it is a feature of my description of the world
- Subjective/Personal Probability

Objective (objectionable?) Bayes?
- No!
- That is a totally different enterprise, aiming to eliminate subjectivity

Objectivism vs. objectivity

Objectivism?
- No!
- de Finetti rejects the idea of probability as a feature of the external world

  (1931b) "Objective probability never exists"

- Rather, it is a feature of my description of the world
- Subjective/Personal Probability

Objective (objectionable?) Bayes?
- No!
- That is a totally different enterprise, aiming to eliminate subjectivity

Objectivity?
- Yes!

Objectivism vs. objectivity

Objectivism?
- No!
- de Finetti rejects the idea of probability as a feature of the external world

  (1931b) "Objective probability never exists"
- Rather, it is a feature of my description of the world
- Subjective/Personal Probability

Objective (objectionable?) Bayes?
- No!
- That is a totally different enterprise, aiming to eliminate subjectivity

Objectivity?
- Yes!
- "Subjective" does NOT mean "anything goes"

Objectivism vs. objectivity

Objectivism?
- No!
- de Finetti rejects the idea of probability as a feature of the external world

  (1931b) "Objective probability never exists"

- Rather, it is a feature of my description of the world
- Subjective/Personal Probability

Objective (objectionable?) Bayes?
- No!
- That is a totally different enterprise, aiming to eliminate subjectivity

Objectivity?
- Yes!
- "Subjective" does NOT mean "anything goes"
- My description of the world should take full account of what I know about the world

Objectivism vs. objectivity

Objectivism?

- No!
- de Finetti rejects the idea of probability as a feature of the external world

  (1931b) "Objective probability never exists"

- Rather, it is a feature of my description of the world
- Subjective/Personal Probability

Objective (objectionable?) Bayes?

- No!
- That is a totally different enterprise, aiming to eliminate subjectivity

Objectivity?

- Yes!
- "Subjective" does NOT mean "anything goes"
- My description of the world should take full account of what I know about the world
- . . . and aim to be a "good" description.

de Finetti's positivism

- "Probability" can only meaningfully be attached to observable events and quantities

de Finetti's positivism

- "Probability" can only meaningfully be attached to observable events and quantities
  - we can (in principle) and should compare our probability assessments with actual outcomes

de Finetti's positivism

- "Probability" can only meaningfully be attached to observable events and quantities
  - we can (in principle) and should compare our probability assessments with actual outcomes
  - we should not (directly) consider distributions over unobservable parameters

de Finetti's positivism

- "Probability" can only meaningfully be attached to observable events and quantities
  - we can (in principle) and should compare our probability assessments with actual outcomes
  - we should not (directly) consider distributions over unobservable parameters (!!!)

de Finetti's positivism

- "Probability" can only meaningfully be attached to observable events and quantities
  - we can (in principle) and should compare our probability assessments with actual outcomes
  - we should not (directly) consider distributions over unobservable parameters (!!!)
  - the principal aim of "statistical inference" should be the probabilistic prediction of as-yet-unobserved quantities, given our observations

de Finetti's positivism

- "Probability" can only meaningfully be attached to observable events and quantities
  - we can (in principle) and should compare our probability assessments with actual outcomes
  - we should not (directly) consider distributions over unobservable parameters (!!!)
  - the principal aim of "statistical inference" should be the probabilistic prediction of as-yet-unobserved quantities, given our observations
- Only finite combinations of events are meaningful

de Finetti's positivism

- "Probability" can only meaningfully be attached to observable events and quantities
  - we can (in principle) and should compare our probability assessments with actual outcomes
  - we should not (directly) consider distributions over unobservable parameters (!!!)
  - the principal aim of "statistical inference" should be the probabilistic prediction of as-yet-unobserved quantities, given our observations
- Only finite combinations of events are meaningful
  - finitely additive probability

Some quotes

(1962) Though maintaining the subjectivist idea that no fact can prove or disprove belief, I find no difficulty in admitting that any form of comparison between probability evaluations. . . and actual events may be an element influencing my further judgment.

Some quotes

(1962) Though maintaining the subjectivist idea that no fact can prove or disprove belief, I find no difficulty in admitting that any form of comparison between probability evaluations. . . and actual events may be an element influencing my further judgment.

(1974) Subjectivists believe that every evaluation of probability is based on available information, including objective data.

Some quotes

(1962) Though maintaining the subjectivist idea that no fact can prove or disprove belief, I find no difficulty in admitting that any form of comparison between probability evaluations... and actual events may be an element influencing my further judgment.

(1974) Subjectivists believe that every evaluation of probability is based on available information, including objective data.

(1974) Every probability evaluation essentially depends on two components:

(1). the objective component, consisting of the evidence of known data and facts; and

(2). the subjective component, consisting of the opinion concerning unknown facts based on known evidence.

Definition or evaluation?

The evaluation of probability should take into account all available evidence, including frequencies and symmetries. However, it would be a mistake to put these elements, which are useful ingredients of the evaluation of probability, at the basis of the definition of probability.

Individual probability

"Probability is defined as the degree of confidence of an individual, at a given instant and with a given set of information, regarding the occurrence of an event" (1970)

Individual probability

"Probability is defined as the degree of confidence of an individual, at a given instant and with a given set of information, regarding the occurrence of an event" (1970)

Weather forecaster: "The probability of precipitation tomorrow is 30%"

Individual probability

"Probability is defined as the degree of confidence of an individual, at a given instant and with a given set of information, regarding the occurrence of an event" (1970)

Weather forecaster: "The probability of precipitation tomorrow is 30%"

Angelina Jolie: "My doctors estimated that I had an 87% of breast cancer. . . although the risk is different in the case of each woman"

Individual probability

"Probability is defined as the degree of confidence of an individual, at a given instant and with a given set of information, regarding the occurrence of an event" (1970)

Weather forecaster: "The probability of precipitation tomorrow is 30%"

Angelina Jolie: "My doctors estimated that I had an 87% of breast cancer... although the risk is different in the case of each woman"

Obama on Osama: "At the end of the day, this was still a 55/45 situation"

Individual probability

"Probability is defined as the degree of confidence of an individual, at a given instant and with a given set of information, regarding the occurrence of an event" (1970)

Weather forecaster: "The probability of precipitation tomorrow is 30%"

Angelina Jolie: "My doctors estimated that I had an 87% of breast cancer... although the risk is different in the case of each woman"

Obama on Osama: "At the end of the day, this was still a 55/45 situation"

- How to interpret?

Individual probability

"Probability is defined as the degree of confidence of an individual, at a given instant and with a given set of information, regarding the occurrence of an event" (1970)

Weather forecaster: "The probability of precipitation tomorrow is 30%"

Angelina Jolie: "My doctors estimated that I had an 87% of breast cancer... although the risk is different in the case of each woman"

Obama on Osama: "At the end of the day, this was still a 55/45 situation"

- How to interpret?
- How to quantify?

Individual probability

"Probability is defined as the degree of confidence of an individual, at a given instant and with a given set of information, regarding the occurrence of an event" (1970)

Weather forecaster: "The probability of precipitation tomorrow is 30%"

Angelina Jolie: "My doctors estimated that I had an 87% of breast cancer... although the risk is different in the case of each woman"

Obama on Osama: "At the end of the day, this was still a 55/45 situation"

- How to interpret?
- How to quantify?
- How to assess?

Subjective probability and frequency

Only the subjective view explains WHY we should use "objective" frequencies to estimate probabilities:

Subjective probability and frequency

Only the subjective view explains WHY we should use "objective" frequencies to estimate probabilities: EXCHANGEABILITY

Subjective probability and frequency

Only the subjective view explains WHY we should use "objective" frequencies to estimate probabilities: EXCHANGEABILITY

- Infinite sequence of (binary) variables $X_1, X_2, \ldots$.

Subjective probability and frequency

Only the subjective view explains WHY we should use "objective" frequencies to estimate probabilities: EXCHANGEABILITY

- Infinite sequence of (binary) variables $X_1, X_2, \ldots$.
- Often natural to assume indifference to the order in which presented.

Subjective probability and frequency

Only the subjective view explains WHY we should use "objective" frequencies to estimate probabilities: EXCHANGEABILITY

- Infinite sequence of (binary) variables $X_1, X_2, \ldots$.
- Often natural to assume indifference to the order in which presented.
- Any joint distribution with this property is a mixture of Bernoulli sequences:

$$P(x_1, x_2, \ldots, x_n) = \int_0^1 \theta^r (1-\theta)^{n-r} dF(\theta) \qquad \left(r = \sum_{i=1}^{n} x_i \right)$$

Subjective probability and frequency

Only the subjective view explains WHY we should use "objective" frequencies to estimate probabilities: EXCHANGEABILITY

- Infinite sequence of (binary) variables $X_1, X_2, \ldots$.
- Often natural to assume indifference to the order in which presented.
- Any joint distribution with this property is a mixture of Bernoulli sequences:

$$P(x_1, x_2, \ldots, x_n) = \int_0^1 \theta^r (1-\theta)^{n-r} dF(\theta) \qquad \left(r = \sum_{i=1}^n x_i\right)$$

- With P-probability 1,

$$P(X_{n+1} = 1 \mid X_1, X_2, \ldots, X_n) - (R/n) \to 0$$

What is “the probability”?

Many (exchangeable) students take many (exchangeable) examinations. What is “the probability” that Thomas will fail his Statistics exam? (based on all other results)

What is “the probability”?

Many (exchangeable) students take many (exchangeable) examinations. What is “the probability” that Thomas will fail his Statistics exam? (based on all other results)

- Thomas's relative frequency α^T of failure on his other papers?

What is “the probability”?

Many (exchangeable) students take many (exchangeable) examinations. What is “the probability” that Thomas will fail his Statistics exam? (based on all other results)

- Thomas's relative frequency α^T of failure on his other papers?
- Other students' relative frequency β^S of failure on the Statistics paper?

What is “the probability”?

Many (exchangeable) students take many (exchangeable) examinations. What is “the probability” that Thomas will fail his Statistics exam? (based on all other results)

- Thomas's relative frequency α^T of failure on his other papers?
- Other students' relative frequency β^S of failure on the Statistics paper?
- Overall failure rate μ of all students on all papers?

What is “the probability”?

Many (exchangeable) students take many (exchangeable) examinations. What is “the probability” that Thomas will fail his Statistics exam? (based on all other results)

- Thomas's relative frequency α^T of failure on his other papers?
- Other students' relative frequency β^S of failure on the Statistics paper?
- Overall failure rate μ of all students on all papers?
- Something more complicated taking all the data into account?...

What is “the probability”?

Many (exchangeable) students take many (exchangeable) examinations. What is “the probability” that Thomas will fail his Statistics exam? (based on all other results)

- Thomas's relative frequency α^T of failure on his other papers?
- Other students' relative frequency β^S of failure on the Statistics paper?
- Overall failure rate μ of all students on all papers?
- Something more complicated taking all the data into account?...
  - $\alpha^T + \beta^S - \mu$

What is “the probability”?

Many (exchangeable) students take many (exchangeable) examinations. What is “the probability” that Thomas will fail his Statistics exam? (based on all other results)

- Thomas's relative frequency α^T of failure on his other papers?
- Other students' relative frequency β^S of failure on the Statistics paper?
- Overall failure rate μ of all students on all papers?
- Something more complicated taking all the data into account?...
  - $\alpha^T + \beta^S - \mu$???

What is “the probability”?

Many (exchangeable) students take many (exchangeable) examinations. What is “the probability” that Thomas will fail his Statistics exam? (based on all other results)

- Thomas's relative frequency α^T of failure on his other papers?
- Other students' relative frequency β^S of failure on the Statistics paper?
- Overall failure rate μ of all students on all papers?
- Something more complicated taking all the data into account?...
  - $\alpha^T + \beta^S - \mu$???
  - $\phi(u_i, v_j)$

What is "the probability"?

Many (exchangeable) students take many (exchangeable) examinations. What is "the probability" that Thomas will fail his Statistics exam? (based on all other results)

- Thomas's relative frequency α^T of failure on his other papers?
- Other students' relative frequency β^S of failure on the Statistics paper?
- Overall failure rate μ of all students on all papers?
- Something more complicated taking all the data into account?...
  - $\alpha^T + \beta^S - \mu$???
  - $\phi(u_i, v_j)$???

What is "the probability"?

Many (exchangeable) students take many (exchangeable) examinations. What is "the probability" that Thomas will fail his Statistics exam? (based on all other results)

- Thomas's relative frequency α^T of failure on his other papers?
- Other students' relative frequency β^S of failure on the Statistics paper?
- Overall failure rate μ of all students on all papers?
- Something more complicated taking all the data into account?. . .
  - $\alpha^T + \beta^S - \mu$???
  - $\phi(u_i, v_j)$???
- Thomas's failure rate on his previous attempts at Statistics?

Data array

?	0	1	0	1	0	1	0	0	1
1	0	1	0	1	1	0	1	0	0
0	0	0	1	1	0	0	0	1	1
1	1	0	1	0	1	1	0	1	0
0	0	1	1	1	0	1	0	0	1
1	0	1	0	0	1	0	1	1	0
0	0	1	0	1	1	0	1	0	1
0	1	0	0	0	0	1	0	1	1
1	0	1	1	0	1	0	1	0	1

Local pattern

?	**0**	**1**	0	1	0	1	0	0	1
1	**0**	**1**	0	1	1	0	1	0	0
0	**0**	**0**	1	1	0	0	0	1	1
1	**1**	**0**	1	0	1	1	0	1	0
0	**0**	**1**	1	1	0	1	0	0	1
1	0	1	0	1	1	0	1	1	0
0	0	1	0	1	1	0	1	0	1
0	1	0	0	0	0	1	0	1	1
1	0	1	1	0	1	0	1	0	1

Repeat pattern

?	**0**	**1**	0	1	0	1	0	0	1
1	**0**	**1**	0	1	1	0	1	0	0
0	**0**	**0**	1	1	0	0	0	1	1
1	**1**	**0**	1	0	1	1	0	1	0
0	**0**	**1**	**1**	**1**	0	1	0	0	1
1	0	1	**0**	**1**	1	0	1	1	0
0	0	**1**	**0**	**1**	1	0	1	0	1
0	1	**0**	**0**	**0**	0	1	0	1	1
1	0	**1**	**1**	**0**	1	0	1	0	1

Backwards and forwards

How should I construct/examine my probability distribution for X?

Backwards and forwards

How should I construct/examine my probability distribution for X?

Like the penal code, an approach to this should have a double purpose:

Backwards and forwards

How should I construct/examine my probability distribution for X?

Like the penal code, an approach to this should have a double purpose:

Forward looking Before observing X, to motivate me to think carefully about my behaviour (assess my distribution)

Backwards and forwards

How should I construct/examine my probability distribution for X?

Like the penal code, an approach to this should have a double purpose:

Forward looking Before observing X, to motivate me to think carefully about my behaviour (assess my distribution)

Backward looking After observing $X = x$, to reward or punish my behaviour (my assessed distribution)

Backwards and forwards

How should I construct/examine my probability distribution for X?

Like the penal code, an approach to this should have a double purpose:

Forward looking Before observing X, to motivate me to think carefully about my behaviour (assess my distribution)

Backward looking After observing $X = x$, to reward or punish my behaviour (my assessed distribution)

These aspects are related but distinct

Backwards and forwards

How should I construct/examine my probability distribution for X?

Like the penal code, an approach to this should have a double purpose:

Forward looking Before observing X, to motivate me to think carefully about my behaviour (assess my distribution)

Backward looking After observing $X = x$, to reward or punish my behaviour (my assessed distribution)

These aspects are related but distinct (*e.g.*, rabid dog)

Scoring rules

Let Q be my Quoted distribution for an observable X. After observing the value x of X, we want to contrast this outcome with my quoted "probability forecast" Q.

Scoring rules

Let Q be my *Q*uoted distribution for an observable X. After observing the value x of X, we want to contrast this outcome with my quoted "probability forecast" Q.

One way is by applying a *scoring rule* (real-valued loss function) $S(x, Q)$.

Scoring rules

Let Q be my Quoted distribution for an observable X. After observing the value x of X, we want to contrast this outcome with my quoted "probability forecast" Q.

One way is by applying a *scoring rule* (real-valued loss function) $S(x, Q)$.

- Suppose I assess that X has distribution P, but quote Q.

Scoring rules

Let Q be my Quoted distribution for an observable X. After observing the value x of X, we want to contrast this outcome with my quoted "probability forecast" Q.

One way is by applying a *scoring rule* (real-valued loss function) $S(x, Q)$.

- Suppose I assess that X has distribution P, but quote Q.
- My expected score is

$$S(P, Q) := \mathrm{E}_{X \sim P}\{S(X, Q)\}.$$

Scoring rules

Let Q be my Quoted distribution for an observable X. After observing the value x of X, we want to contrast this outcome with my quoted "probability forecast" Q.

One way is by applying a *scoring rule* (real-valued loss function) $S(x, Q)$.

- Suppose I assess that X has distribution P, but quote Q.
- My expected score is

$$S(P, Q) := \mathrm{E}_{X \sim P}\{S(X, Q)\}.$$

- S is termed a proper scoring rule if my expected score $S(P, Q)$ is minimised in Q at $Q = P$; and strictly proper if $S(P, Q) > S(P, P)$ for $Q \neq P$.

Scoring rules

Let Q be my *Q*uoted distribution for an observable X. After observing the value x of X, we want to contrast this outcome with my quoted "probability forecast" Q.

One way is by applying a *scoring rule* (real-valued loss function) $S(x, Q)$.

- Suppose I assess that X has distribution P, but quote Q.
- My expected score is

$$S(P, Q) := \mathrm{E}_{X \sim P}\{S(X, Q)\}.$$

- S is termed a proper scoring rule if my expected score $S(P, Q)$ is minimised in Q at $Q = P$; and strictly proper if $S(P, Q) > S(P, P)$ for $Q \neq P$.
- When S is proper, honesty is the best policy: If I believe $X \sim P$, I will minimise my expected score by quoting $Q = P$.

Probability score table (Brier)

A occurs	A does not occur	q
25.0	25.0	.500
22.6	27.6	.525
20.2	30.2	.550
18.1	33.1	.575
16.0	36.0	.600
14.1	39.1	.625
12.2	42.2	.650
10.6	45.6	.675
9.0	49.0	.700
7.6	52.6	.750
6.2	56.2	.725
5.1	60.1	.775
4.0	64.0	.800
3.1	68.1	.825
2.2	72.2	.850
1.6	76.6	.875
1.0	81.0	.900
0.6	85.6	.925
0.2	90.2	.950
0.1	95.1	.975
0.0	100.0	1 000

Probability score table (Brier)—reduced

A occurs	*A does not occur*
25.0	**25.0**
22.6	**27.6**
20.2	**30.2**
18.1	**33.1**
16.0	**36.0**
14.1	**39.1**
12.2	**42.2**
10.6	**45.6**
9.0	**49.0**
7.6	**52.6**
6.2	**56.2**
5.1	**60.1**
4.0	**64.0**
3.1	**68.1**
2.2	**72.2**
1.6	**76.6**
1.0	**81.0**
0.6	**85.6**
0.2	**90.2**
0.1	**95.1**
0.0	**100.0**

Probability assessment quiz

Mark in Column 1 the answer you consider correct. In Column 2, attach a number between 0.5 and 1 to indicate your personal probability that this is the correct answer. When the answers are revealed, indicate in Column 3 whether you were right or wrong, and insert the corresponding penalty score (from Table 1) in Column 4.

		1	2	3	4
1.	**Crystals of common salt are** a) octahedral b) cubical				
2.	**Shi'ism is a branch of** a) Islam b) Confucianism				
3.	**Which has the larger area?** a) France b) The Iberian Peninsula (Spain and Portugal)				
4.	**The blood in the pulmonary artery flows** a) from heart to lungs b) from lungs to heart				
5.	**An ibex is** a) a bird b) a goat				
6.	**Buddhism had its origins in** a) India b) China				
7.	**Sea-water freezes at** a) –3°C b) +2°C				
8.	**In Heraldry, "gules" refers to the colour** a) red b) blue				
9.	**An anticyclone is a region of** a) low pressure b) high pressure				
10	**The Venetian gondola is propelled by** a) a pole b) an oar				
	TOTAL SCORE				

Probability assessment quiz

Mark in Column 1 the answer you consider correct. In Column 2, attach a number between 0.5 and 1 to indicate your personal probability that this is the correct answer. When the answers are revealed, indicate in Column 3 whether you were right or wrong, and insert the corresponding penalty score (from Table 1) in Column 4.

		1	2	3	4
1.	**Crystals of common salt are** **a) octahedral** **b) cubical**	a			
2.	**Shi'ism is a branch of** **a) Islam** **b) Confucianism**				
3.	**Which has the larger area?** **a) France** **b) The Iberian Peninsula (Spain and Portugal)**				
4.	**The blood in the pulmonary artery flows** **a) from heart to lungs** **b) from lungs to heart**				
5.	**An ibex is** **a) a bird** **b) a goat**				
6.	**Buddhism had its origins in** **a) India** **b) China**				
7.	**Sea-water freezes at** **a) –3°C** **b) +2°C**				
8.	**In Heraldry, "gules" refers to the colour** **a) red** **b) blue**				
9.	**An anticyclone is a region of** **a) low pressure** **b) high pressure**				
10	**The Venetian gondola is propelled by** **a) a pole** **b) an oar**				
	TOTAL SCORE				

Probability assessment quiz

Mark in Column 1 the answer you consider correct. In Column 2, attach a number between 0.5 and 1 to indicate your personal probability that this is the correct answer. When the answers are revealed, indicate in Column 3 whether you were right or wrong, and insert the corresponding penalty score (from Table 1) in Column 4.

		1	2	3	4
1.	**Crystals of common salt are** a) octahedral b) cubical	a	0.650		
2.	**Shi'ism is a branch of** a) Islam b) Confucianism				
3.	**Which has the larger area?** a) France b) The Iberian Peninsula (Spain and Portugal)				
4.	**The blood in the pulmonary artery flows** a) from heart to lungs b) from lungs to heart				
5.	**An ibex is** a) a bird b) a goat				
6.	**Buddhism had its origins in** a) India b) China				
7.	**Sea-water freezes at** a) –3°C b) +2°C				
8.	**In Heraldry, "gules" refers to the colour** a) red b) blue				
9.	**An anticyclone is a region of** a) low pressure b) high pressure				
10	**The Venetian gondola is propelled by** a) a pole b) an oar				
	TOTAL SCORE				

Probability assessment quiz

Mark in Column 1 the answer you consider correct. In Column 2, attach a number between 0.5 and 1 to indicate your personal probability that this is the correct answer. When the answers are revealed, indicate in Column 3 whether you were right or wrong, and insert the corresponding penalty score (from Table 1) in Column 4.

		1	2	3	4
1.	**Crystals of common salt are** a) octahedral b) cubical	a	0.650	✖	
2.	**Shi'ism is a branch of** a) Islam b) Confucianism				
3.	**Which has the larger area?** a) France b) The Iberian Peninsula (Spain and Portugal)				
4.	**The blood in the pulmonary artery flows** a) from heart to lungs b) from lungs to heart				
5.	**An ibex is** a) a bird b) a goat				
6.	**Buddhism had its origins in** a) India b) China				
7.	**Sea-water freezes at** a) –3°C b) +2°C				
8.	**In Heraldry, "gules" refers to the colour** a) red b) blue				
9.	**An anticyclone is a region of** a) low pressure b) high pressure				
10	**The Venetian gondola is propelled by** a) a pole b) an oar				
	TOTAL SCORE				

Probability assessment quiz

Mark in Column 1 the answer you consider correct. In Column 2, attach a number between 0.5 and 1 to indicate your personal probability that this is the correct answer. When the answers are revealed, indicate in Column 3 whether you were right or wrong, and insert the corresponding penalty score (from Table 1) in Column 4.

		1	2	3	4
1.	**Crystals of common salt are** a) octahedral b) cubical	a	0.650	✖	42.2
2.	**Shi'ism is a branch of** a) Islam b) Confucianism				
3.	**Which has the larger area?** a) France b) The Iberian Peninsula (Spain and Portugal)				
4.	**The blood in the pulmonary artery flows** a) from heart to lungs b) from lungs to heart				
5.	**An ibex is** a) a bird b) a goat				
6.	**Buddhism had its origins in** a) India b) China				
7.	**Sea-water freezes at** a) –3°C b) +2°C				
8.	**In Heraldry, "gules" refers to the colour** a) red b) blue				
9.	**An anticyclone is a region of** a) low pressure b) high pressure				
10	**The Venetian gondola is propelled by** a) a pole b) an oar				
	TOTAL SCORE				

Construction of proper scoring rule S (binary case — extends much more generally)

$\mathcal{X} = \{0, 1\}$, $Q(X = 1) = q$. H a concave *entropy* function on $[0, 1]$.

Construction of proper scoring rule S (binary case — extends much more generally)

$\mathcal{X} = \{0, 1\}$, $Q(X = 1) = q$. H a concave *entropy* function on $[0, 1]$.

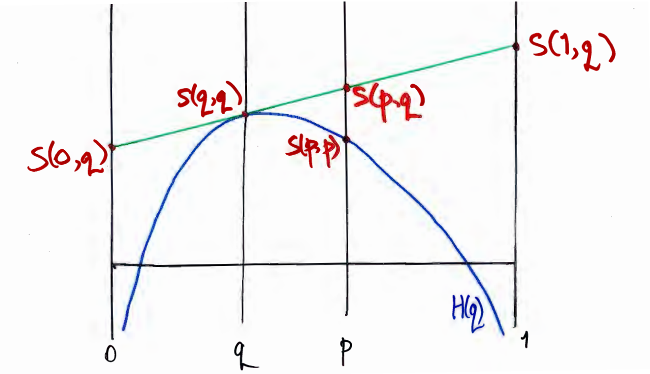
$$\begin{aligned} S(x, q) &= H(q) + (x - q)H'(q) \\ S(p, q) &= H(q) + (p - q)H'(q) \end{aligned}$$

Construction of proper scoring rule S
(binary case — extends much more generally)

$\mathcal{X} = \{0, 1\}$, $Q(X = 1) = q$. H a concave *entropy* function on $[0, 1]$.

$$\begin{aligned} S(x, q) &= H(q) + (x - q)H'(q) \\ S(p, q) &= H(q) + (p - q)H'(q) \end{aligned}$$

Then $S(p, p) = H(p)$. $D(p, q) := S(p, q) - H(p) \geq 0$ measures the discrepancy between p and q.



General construction

- Action space $\mathcal{A}$
- Loss function $L(x, a)$

General construction

- Action space $\mathcal{A}$
- Loss function $L(x, a)$
- Distribution P over $\mathcal{A}$ $\rightarrow$ expected loss $L(P, a) := \mathrm{E}_{X \sim P} L(X, a)$
- $\min_{a \in \mathcal{A}} L(P, a)$ $\rightarrow$ Bayes act a_P

General construction

- Action space $\mathcal{A}$
- Loss function $L(x, a)$
- Distribution P over $\mathcal{A}$ $\rightarrow$ expected loss $L(P, a) := \mathrm{E}_{X \sim P} L(X, a)$
- $\min_{a \in \mathcal{A}} L(P, a)$ $\rightarrow$ Bayes act a_P
- $S(x, Q) := L(x, a_q)$

General construction

- Action space $\mathcal{A}$
- Loss function $L(x, a)$
- Distribution P over $\mathcal{A}$ $\rightarrow$ expected loss $L(P, a) := \mathrm{E}_{X \sim P} L(X, a)$
- $\min_{a \in \mathcal{A}} L(P, a)$ $\rightarrow$ Bayes act a_P
- $S(x, Q) := L(x, a_q)$

Then $S(P, Q) = L(P, a_Q) \geq L(P, a_P) = S(P, P)$

Examples

Important special cases ($\mathcal{X} = \{0, 1\}$):

Examples

Important special cases ($\mathcal{X} = \{0, 1\}$):

Brier (quadratic) score:

$$\begin{aligned} S(x, q) &= (x - q)^2 \\ H(p) &= p(1 - p) \\ D(p, q) &= (p - q)^2 \end{aligned}$$

This rule was extensively used by de Finetti (theory, football)

Examples

Important special cases ($\mathcal{X} = \{0, 1\}$):

Brier (quadratic) score:

$$\begin{aligned} S(x, q) &= (x - q)^2 \\ H(p) &= p(1 - p) \\ D(p, q) &= (p - q)^2 \end{aligned}$$

This rule was extensively used by de Finetti
(theory, football)

log score (Good)

$$\begin{aligned} S(x, q) &= -\{x \log q + (1 - x) \log(1 - q)\} \\ &\quad \text{log likelihood} \\ H(p) &= -\{p \log p + (1 - p) \log(1 - p)\} \\ &\quad \text{Shannon entropy} \\ D(p, q) &= p \log(p/q) + (1 - p) \log\{(1 - p)/(1 - q)\} \\ &\quad \text{KL divergence} \end{aligned}$$

Independence of scoring rule

Suppose my quoted probability for an event E is:

- q_1 under scoring rule S_1

Independence of scoring rule

Suppose my quoted probability for an event E is:

- q_1 under scoring rule S_1
- q_2 under scoring rule S_2

Independence of scoring rule

Suppose my quoted probability for an event E is:

- q_1 under scoring rule S_1
- q_2 under scoring rule S_2

My overall penalty is the total (or average) of these.

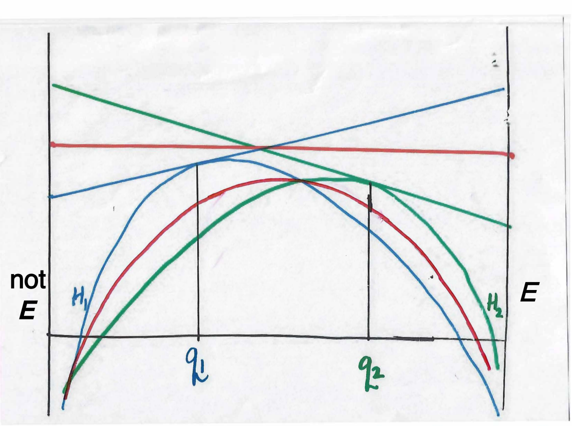
Independence of scoring rule

Suppose my quoted probability for an event E is:

- q_1 under scoring rule S_1
- q_2 under scoring rule S_2

My overall penalty is the total (or average) of these.

If $q_1 \neq q_2$, I could have done better, however E turns out.



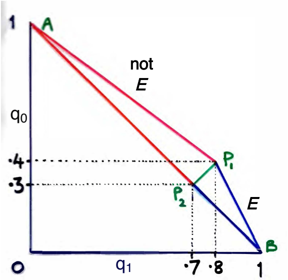
Coherence

I quote q_1 (say 0.8) as my probability for E, and q_0 (0.4) for not E.

Coherence

I quote q_1 (say 0.8) as my probability for E, and q_0 (0.4) for not E.
My total Brier score is

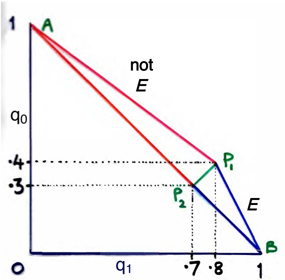
$$
\begin{array}{ll}
(1-q_1)^2 + q_0^2 & \text{if } E \text{ occurs} \\
& = \text{squared distance from } P_1 = (q_1, q_0) \text{ to } B = (1,0) \\
q_1^2 + (1-q_0)^2 & \text{if } E \text{ does not occur} \\
& = \text{squared distance from } P_1 \text{ to } A = (0,1)
\end{array}
$$



Coherence

I quote q_1 (say 0.8) as my probability for E, and q_0 (0.4) for not E. My total Brier score is

$$
\begin{array}{ll}
(1-q_1)^2 + q_0^2 & \text{if } E \text{ occurs} \\
 & = \text{squared distance from } P_1 = (q_1, q_0) \text{ to } B = (1,0) \\
q_1^2 + (1-q_0)^2 & \text{if } E \text{ does not occur} \\
 & = \text{squared distance from } P_1 \text{ to } A = (0,1)
\end{array}
$$



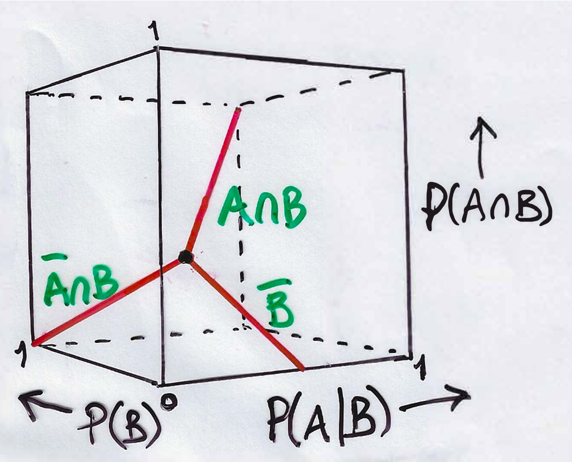
If I do not have $q_i \geq 0$, $q_1 + q_0 = 1$, I could have done better (*e.g.*, with $P_2 = (0.7, 0.3)$), however E turns out.

Compound probability

I assess separate values for $P(A \mid B)$, $P(B)$, $P(A \cap B)$.

Compound probability

I assess separate values for $P(A \mid B)$, $P(B)$, $P(A \cap B)$.
Unless red lines are coplanar—which requires $P(A \cap B) = P(A \mid B)\, P(B)$—I could have done better in all circumstances.



Assessing forecasts against outcomes

I have made a sequence of forecasts, Nature has determined the outcomes. How well did I do?

Assessing forecasts against outcomes

I have made a sequence of forecasts, Nature has determined the outcomes. How well did I do?

NB: I am trying to capture Nature's outcomes, not Her "data-generating process".

Assessing forecasts against outcomes

I have made a sequence of forecasts, Nature has determined the outcomes. How well did I do?

NB: I am trying to capture Nature's outcomes, not Her "data-generating process".

Probability p	0.4	0.6	0.3	0.2	0.6	0.3	0.4	0.5	0.6	0.2	0.6	0.4	0.3	0.5
Outcome	0	0	1	0	1	0	1	1	1	0	1	0	0	1

Assessing forecasts against outcomes

I have made a sequence of forecasts, Nature has determined the outcomes. How well did I do?

NB: I am trying to capture Nature's outcomes, not Her "data-generating process".

Probability p	0.4	0.6	0.3	0.2	0.6	0.3	0.4	0.5	0.6	0.2	0.6	0.4	0.3	0.5
Outcome	0	0	1	0	1	0	1	1	1	0	1	0	0	1

Group by assigned probability:

Probability π	0.2	0.3	0.4	0.5	0.6
Instances n	2	3	3	2	4
Successes r	0	1	1	2	3
Proportion ρ	0	0.33	0.33	1	0.75

Assessing forecasts against outcomes

I have made a sequence of forecasts, Nature has determined the outcomes. How well did I do?

NB: I am trying to capture Nature's outcomes, not Her "data-generating process".

Probability p	0.4	0.6	0.3	0.2	0.6	0.3	0.4	0.5	0.6	0.2	0.6	0.4	0.3	0.5
Outcome	0	0	1	0	1	0	1	1	1	0	1	0	0	1

Group by assigned probability:

Probability π	0.2	0.3	0.4	0.5	0.6
Instances n	2	3	3	2	4
Successes r	0	1	1	2	3
Proportion ρ	0	0.33	0.33	1	0.75

Calibration: ρ_j's close to π_j's

Assessing forecasts against outcomes

I have made a sequence of forecasts, Nature has determined the outcomes. How well did I do?

NB: I am trying to capture Nature's outcomes, not Her "data-generating process".

Probability p	0.4	0.6	0.3	0.2	0.6	0.3	0.4	0.5	0.6	0.2	0.6	0.4	0.3	0.5
Outcome	0	0	1	0	1	0	1	1	1	0	1	0	0	1

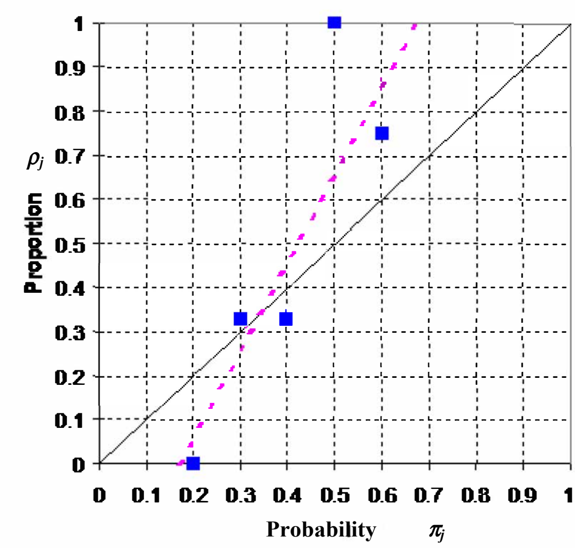
Group by assigned probability:

Probability π	0.2	0.3	0.4	0.5	0.6
Instances n	2	3	3	2	4
Successes r	0	1	1	2	3
Proportion ρ	0	0.33	0.33	1	0.75

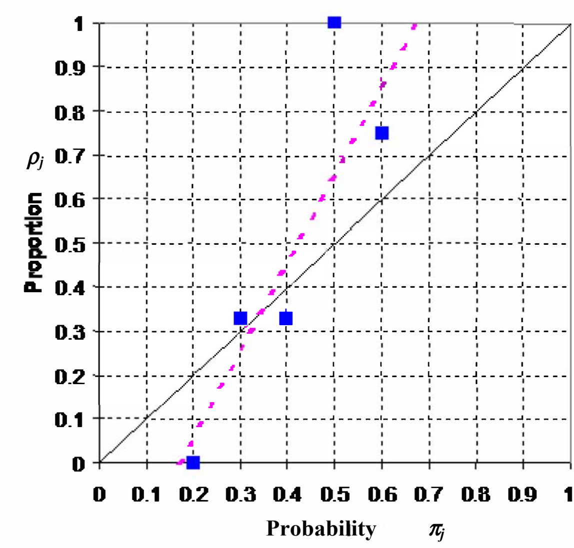
Calibration: ρ_j's close to π_j's

Refinement: ρ_j's close to 0 or 1

Calibration plot



Calibration plot



If forecasts are "good", the points should lie near the diagonal.

Scoring rule decomposition

Assess/compare forecasters by their total achieved Brier score:

$$S_+ = \sum_{i=1}^{14} (x_i - p_i)^2 = 2.764$$

Scoring rule decomposition

Assess/compare forecasters by their total achieved Brier score:

$$S_+ = \sum_{i=1}^{14} (x_i - p_i)^2 = 2.764$$

We can decompose:

$$S_+ = S_1 + S_2$$

where

$$\begin{aligned} S_1 &= \sum_{j=1}^{5} n_j (\pi_j - \rho_j)^2 = 0.687 \\ S_2 &= \sum_{j=1}^{5} n_j \rho_j (1 - \rho_j) = 2.077. \end{aligned}$$

Scoring rule decomposition

Assess/compare forecasters by their total achieved Brier score:

$$S_+ = \sum_{i=1}^{14} (x_i - p_i)^2 = 2.764$$

We can decompose:

$$S_+ = S_1 + S_2$$

where

$$\begin{aligned} S_1 &= \sum_{j=1}^{5} n_j (\pi_j - \rho_j)^2 = 0.687 \\ S_2 &= \sum_{j=1}^{5} n_j \rho_j (1 - \rho_j) = 2.077. \end{aligned}$$

- S_1 penalizes poor calibration

Scoring rule decomposition

Assess/compare forecasters by their total achieved Brier score:

$$S_+ = \sum_{i=1}^{14} (x_i - p_i)^2 = 2.764$$

We can decompose:

$$S_+ = S_1 + S_2$$

where

$$\begin{aligned} S_1 &= \sum_{j=1}^{5} n_j (\pi_j - \rho_j)^2 = 0.687 \\ S_2 &= \sum_{j=1}^{5} n_j \rho_j (1 - \rho_j) = 2.077. \end{aligned}$$

- S_1 penalizes poor calibration
- S_2 penalizes poor refinement

Scoring rule decomposition

Assess/compare forecasters by their total achieved Brier score:

$$S_+ = \sum_{i=1}^{14} (x_i - p_i)^2 = 2.764$$

We can decompose:

$$S_+ = S_1 + S_2$$

where

$$\begin{aligned} S_1 &= \sum_{j=1}^{5} n_j (\pi_j - \rho_j)^2 = 0.687 \\ S_2 &= \sum_{j=1}^{5} n_j \rho_j (1 - \rho_j) = 2.077. \end{aligned}$$

- S_1 penalizes poor calibration
- S_2 penalizes poor refinement

Similar decomposition for other proper scoring rules

Real weather forecasts

- Forecasts of precipitation occurrence have been routinely expressed in probabilistic terms on a nationwide basis in the US since 1965.
- A subjective precipitation probability forecast expresses the forecaster's "degree of belief" that a measurable amount of precipitation (≥ 0.01 in.) will occur during a specified period (generally 12 h) at a particular point in the forecast area (generally the official rain gauge).
- A typical forecast might state that "the precipitation probability for Denver (Colorado) today is 30 per cent".
- Millions of such forecasts have been formulated and issued to the public by the National Weather Service (NWS).

Reliability diagram

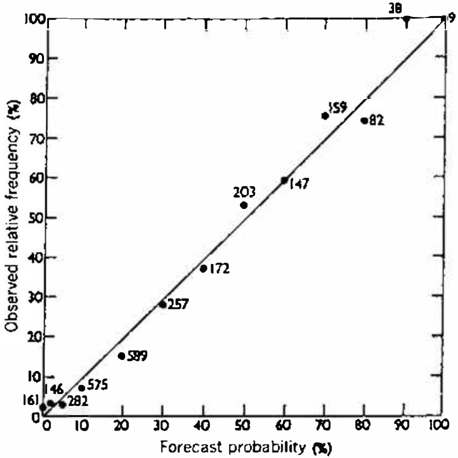


Figure: The reliability diagram for all of the precipitation probability forecasts formulated by NWS forecasters at Chicago during the period from July 1972 to June 1976.

Refining calibration

Sequential development of forecasts and outcomes:

Refining calibration

Sequential development of forecasts and outcomes:

day	1	2	3	$\cdots$
forecast	p_1			
rain?				

Refining calibration

Sequential development of forecasts and outcomes:

day	1	2	3	$\cdots$
forecast	p_1			
rain?	x_1			

Refining calibration

Sequential development of forecasts and outcomes:

day	1	2	3	$\cdots$
forecast	p_1	p_2		
rain?	x_1			

Refining calibration

Sequential development of forecasts and outcomes:

day	1	2	3	$\cdots$
forecast	p_1	p_2		
rain?	x_1	x_2		

Refining calibration

Sequential development of forecasts and outcomes:

day	1	2	3	$\cdots$
forecast	p_1	p_2	p_3	
rain?	x_1	x_2		

Refining calibration

Sequential development of forecasts and outcomes:

day	1	2	3	$\cdots$
forecast	p_1	p_2	p_3	
rain?	x_1	x_2	x_3	

Refining calibration

Sequential development of forecasts and outcomes:

day	1	2	3	$\cdots$
forecast	p_1	p_2	p_3	$\ldots$
rain?	x_1	x_2	x_3	

Refining calibration

Sequential development of forecasts and outcomes:

day	1	2	3	$\cdots$
forecast	p_1	p_2	p_3	$\ldots$
rain?	x_1	x_2	x_3	$\ldots$

Refining calibration

Sequential development of forecasts and outcomes:

day	1	2	3	$\cdots$
forecast	p_1	p_2	p_3	$\ldots$
rain?	x_1	x_2	x_3	$\ldots$

Select (computably) a subsequence of days, using previous information — *e.g.*, those prime-numbered weekdays that immediately follow two wet and two dry days, for which the probability forecast exceeds 0.6.

Refining calibration

Sequential development of forecasts and outcomes:

day	1	2	3	$\cdots$
forecast	p_1	p_2	p_3	$\ldots$
rain?	x_1	x_2	x_3	$\ldots$

Select (computably) a subsequence of days, using previous information — *e.g.*, those prime-numbered weekdays that immediately follow two wet and two dry days, for which the probability forecast exceeds 0.6.

Ask for asymptotic equality of average probability and relative frequency, for any such subsequence.

Refining calibration

Sequential development of forecasts and outcomes:

day	1	2	3	$\cdots$
forecast	p_1	p_2	p_3	$\ldots$
rain?	x_1	x_2	x_3	$\ldots$

Select (computably) a subsequence of days, using previous information — *e.g.*, those prime-numbered weekdays that immediately follow two wet and two dry days, for which the probability forecast exceeds 0.6.

Ask for asymptotic equality of average probability and relative frequency, for any such subsequence.

WHY?

Refining calibration

Sequential development of forecasts and outcomes:

day	1	2	3	$\cdots$
forecast	p_1	p_2	p_3	$\ldots$
rain?	x_1	x_2	x_3	$\ldots$

Select (computably) a subsequence of days, using previous information — *e.g.*, those prime-numbered weekdays that immediately follow two wet and two dry days, for which the probability forecast exceeds 0.6.

Ask for asymptotic equality of average probability and relative frequency, for any such subsequence.

WHY?

If the forecasts are conditional probabilities based on a joint distribution P, then this property holds almost surely under P.

Asymptotic uniqueness

Suppose two sequences of (computable) forecasts, (p_i) and (q_i), both satisfy this criterion.

Asymptotic uniqueness

Suppose two sequences of (computable) forecasts, (p_i) and (q_i), both satisfy this criterion.

Then

$$p_i - q_i \to 0.$$

Asymptotic uniqueness

Suppose two sequences of (computable) forecasts, (p_i) and (q_i), both satisfy this criterion.

Then

$$p_i - q_i \to 0.$$

- No scope for subjectivity?

Asymptotic uniqueness

Suppose two sequences of (computable) forecasts, (p_i) and (q_i), both satisfy this criterion.

Then

$$p_i - q_i \to 0.$$

- No scope for subjectivity?
- "Objective" forecasts?

Asymptotic uniqueness

Suppose two sequences of (computable) forecasts, (p_i) and (q_i), both satisfy this criterion.

Then

$$p_i - q_i \to 0.$$

- No scope for subjectivity?
- "Objective" forecasts?

Problem: Not computable.

Asymptotic uniqueness

Suppose two sequences of (computable) forecasts, (p_i) and (q_i), both satisfy this criterion.

Then

$$p_i - q_i \to 0.$$

- No scope for subjectivity?
- "Objective" forecasts?

Problem: Not computable.

Can elaborate to allow side information (*e.g.*, temperature).

Again have asymptotic uniqueness—but dependence on the information used

Asymptotic uniqueness

Suppose two sequences of (computable) forecasts, (p_i) and (q_i), both satisfy this criterion.

Then

$$p_i - q_i \to 0.$$

- No scope for subjectivity?
- "Objective" forecasts?

Problem: Not computable.

Can elaborate to allow side information (*e.g.*, temperature).

Again have asymptotic uniqueness—but dependence on the information used

No conflict between "deep determinism" and non-extreme probabilities

Asymptotic uniqueness

Suppose two sequences of (computable) forecasts, (p_i) and (q_i), both satisfy this criterion.

Then

$$p_i - q_i \to 0.$$

- No scope for subjectivity?
- "Objective" forecasts?

Problem: Not computable.

Can elaborate to allow side information (*e.g.*, temperature).

Again have asymptotic uniqueness—but dependence on the information used

No conflict between "deep determinism" and non-extreme probabilities

Deterministic Chaos?

Lessons?

- Asymptotic “objective” probabilities (relative to state of information) may exist

Lessons?

- Asymptotic “objective” probabilities (relative to state of information) may exist
- But we may never be able to discover them

Lessons?

- Asymptotic "objective" probabilities (relative to state of information) may exist
- But we may never be able to discover them
- "Perfect" probability assessment may be unattainable

Lessons?

- Asymptotic “objective” probabilities (relative to state of information) may exist
- But we may never be able to discover them
- “Perfect” probability assessment may be unattainable
- So we should just do the best we can…

Lessons?

- Asymptotic "objective" probabilities (relative to state of information) may exist
- But we may never be able to discover them
- "Perfect" probability assessment may be unattainable
- So we should just do the best we can...
- Subjective?

Probability assessment quiz: Solutions

Mark in Column 1 the answer you consider correct. In Column 2, attach a number between 0.5 and 1 to indicate your personal probability that this is the correct answer. When the answers are revealed, indicate in Column 3 whether you were right or wrong, and insert the corresponding penalty score (from Table 1) in Column 4.

		1	2	3	4
1.	**Crystals of common salt are** **a)** octahedral **b)** cubical	b			
2.	**Shi'ism is a branch of** **a)** Islam **b)** Confucianism	a			
3.	**Which has the larger area?** **a)** France **b)** The Iberian Peninsula (Spain and Portugal)	a			
4.	**The blood in the pulmonary artery flows** **a)** from heart to lungs **b)** from lungs to heart	a			
5.	**An ibex is** **a)** a bird **b)** a goat	b			
6.	**Buddhism had its origins in** **a)** India **b)** China	a			
7.	**Sea-water freezes at** **a)** –3°C **b)** +2°C	a			
8.	**In Heraldry, "gules" refers to the colour** **a)** red **b)** blue	a			
9.	**An anticyclone is a region of** **a)** low pressure **b)** high pressure	b			
10	**The Venetian gondola is propelled by** **a)** a pole **b)** an oar	b			
	TOTAL SCORE				

THANK YOU!

Acknowledgment

I wish to express my special appreciation to Maria Carla Galavotti, Emeritus Professor of Philosophy at the University of Bologna, for sharing her expert knowledge and understanding of de Finetti and his works.

References

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability.
Monthly Weather Review, 78:1–3.

Dawid, A. P. (1986). Probability forecasting.
In Kotz, S., Johnson, N. L., and Read, C. B., editors, *Encyclopedia of Statistical Sciences*, volume 7, pages 210–218. Wiley-Interscience, New York.

Dawid, A. P. (2004). Probability, causality and the empirical world: A Bayes–de Finetti–Popper–Borel synthesis.
Statistical Science, 19:44–57.

Dawid, A. P. and Galavotti, M. C. (2009). de Finetti's subjectivism, objective probability, and the empirical validation of probability assessments.
In Galavotti, M. C., editor, *Bruno de Finetti, Radical Probabilist*, pages 97–114. College Publications, London.

de Finetti, B. (1931b). Sul significato soggettivo della probabilità.
Fundamenta Mathematicae, 17:298–329.

de Finetti, B. (1931a). Probabilismo.
Logos, pages 163–219.
Reprinted in B. de Finetti, La Logica dell'Incerto, Milano, Il Saggiatore, 1989, pp. 3–70. English translation in Erkenntnis **31** (1989): pp. 169-223.

de Finetti, B. (1937). La prévision: Ses lois logiques, ses sources subjectives.
Annales de l'Institut Henri Poincaré, 7:1–68.
English translation in Kyburg and Smokler (eds.), Studies in Subjective Probability, New York-London: Wiley, 1964, pp. 95-158.

de Finetti, B. (1962). Does it make sense to speak of 'good probability appraisers'?
In Good, I. J. et al., editors, *The Scientist Speculates. An Anthology of Partly-Baked Ideas*, pages 357–364. Basic Books, New York.

de Finetti, B. (1974). The value of studying subjective evaluations of probability.
In Staël von Holstein, C.-A., editor, *The Concept of Probability in Psychological Experiments*, pages 1–14. Reidel, Dordrecht-Boston.

de Finetti, B. (1970). *Teoria delle Probabilità.*
Einaudi.
English edition: Theory of Probability, New York: Wiley, 1975.

Galavotti, M. C. (2001). Subjectivism, objectivism and objectivity in Bruno de Finetti's bayesianism.
In Corfield, D. and Williamson, J., editors, *Foundations of Bayesianism*, volume 24 of *Applied Logic Series*, pages 161–174. Springer Netherlands, Dordrecht.

Good, I. J. (1952). Rational decisions.
Journal of the Royal Statistical Society, Series B, 14:107–114.

Murphy, A. H. and Winkler, R. L. (1977). Reliability of subjective probability forecasts of precipitation and temperature.
Journal of the Royal Statistical Society. Series C (Applied Statistics), 26:41–47.