Title: Foundations and Interpretation of Quantum Theory - Lecture 7B

Date: Feb 04, 2010 03:30 PM

URL: http://pirsa.org/10020082

Abstract: After a review of the axiomatic formulation of quantum theory, the generalized operational structure of the theory will be introduced (including POVM measurements, sequential measurements, and CP maps). There will be an introduction to the orthodox (sometimes called Copenhagen) interpretation of quantum mechanics and the historical problems/issues/debates regarding that interpretation, in particular, the measurement problem and the EPR paradox, and a discussion of contemporary views on these topics. The majority of the course lectures will consist of guest lectures from international experts covering the various approaches to the interpretation of quantum theory (in particular, many-worlds, de Broglie-Bohm, consistent/decoherent histories, and statistical/epistemic interpretations, as time permits) and fundamental properties and tests of quantum theory (such as entanglement and experimental tests of Bell inequalities, contextuality, macroscopic quantum phenomena, and the problem of quantum gravity, as time permits).

Firefox File Edit View History Bookmarks Tools Window Help Thu 3:33 PM Robin Blume-Kohout

Robin's Notes – An ongoing quest for coherence

Robin's Notes – An o... | Bayesian probability ... | Probability axioms – ... | Quantum Theory – | Principia Mathematic...

But consider the consequences if you bet this way for a while. Each round, you either double your money or you lose it **all**. So, if we assume w/o l.o.g. that $p > \frac{1}{2}$, then after N rounds you have multiplied your wealth by 2^N, with probability p^N, or lost it all with probability $1 - p^N$. As N goes up, you almost certainly go broke! More generally, a more conservative strategy will almost certainly net you more money.

What this tells us is that the mean value of your winnings is misleading, because it can be distorted by a tiny probability of huge gain. Much better to try and maximize the *median* value of your winnings. We can do this by observing that your wealth increases geometrically (like compound interest, except with randomness), so its logarithm increases linearly in time. This is nice because

1. with probability p, your logmoney increases by $\log(2x)$
2. with probability $1 - p$, your logmoney increases by $\log(2(1 - x))$.

So, over repeated trials (bets), logmoney undergoes a nice random walk process. And the central limit theorem applies: after a while, your logmoney will be roughly Gaussian in distribution. The median of a Gaussian is the same as its mean... but the median value of logmoney is the log of the median money!

$$\mathrm{median}(money) = e^{\mathrm{median}(\log(money))}$$

Firefox File Edit View History Bookmarks Tools Window Help Thu 3:34 PM Robin Blume-Kohout

Robin's Notes – An ongoing quest for coherence

Robin's Notes – An o... | Bayesian probability ... | Probability axioms – ... | Quantum Theory – | Principia Mathematic...

But consider the consequences if you bet this way for a while. Each round, you either double your money or you lose it **all**. So, if we assume w/o l.o.g. that $p > \frac{1}{2}$, then after N rounds you have multiplied your wealth by 2^N, with probability p^N, or lost it all with probability $1 - p^N$. As N goes up, you almost certainly go broke! More generally, a more conservative strategy will almost certainly net you more money.

What this tells us is that the mean value of your winnings is misleading, because it can be distorted by a tiny probability of huge gain. Much better to try and maximize the *median* value of your winnings. We can do this by observing that your wealth increases geometrically (like compound interest, except with randomness), so its logarithm increases linearly in time. This is nice because

1. with probability p, your logmoney increases by $\log(2x)$
2. with probability $1 - p$, your logmoney increases by $\log(2(1 - x))$.

So, over repeated trials (bets), logmoney undergoes a nice random walk process. And the central limit theorem applies: after a while, your logmoney will be roughly Gaussian in distribution. The median of a Gaussian is the same as its mean... but the median value of logmoney is the log of the median money!

$$\text{median}(money) = e^{\text{median}(\log(money))}$$

Firefox File Edit View History Bookmarks Tools Window Help Thu 3:35 PM Robin Blume-Kohout

Robin's Notes – An ongoing quest for coherence

Robin's Notes – An o... | Bayesian probability ... | Probability axioms – ... | Quantum Theory – | Principia Mathematic...

So if we maximize the average logmoney, we maximize the median of money.

Expected logmoney is

$$\langle \log D \rangle_1 = p \log(2xD) + (1-p)\log(2(1-x)D)$$
$$= \log(2D) + p\log x + (1-p)\log(1-x).$$

Now, in information theory this thing has a name: the *cross-entropy* of p and x,

$$H(\vec{P}|\vec{X}) = -\sum_k p_k \log x_k$$
$$= p\log x + (1-p)\log(1-x).$$

So your expected logmoney after one bet is

$$\langle \log D \rangle_1 = \log D + \log 2 - H(\vec{P}|\vec{X}),$$

and since everything is additive, after N bets it's

$$\langle \log D \rangle_N = \log D + N\log 2 - NH(\vec{P}|\vec{X}).$$

So you want to minimize $H(\vec{P}|\vec{X})$.

So you want to minimize $H(\vec{P}|\vec{X})$.

Now, we know (and you can prove, using basic calculus) that the minimum value of $H(\vec{P}|\vec{X})$ is achieved by setting $x_k = p_k$, at which point you get

$$H(\vec{P}|\vec{P}) = H(\vec{P}) = -\sum_k p_k \log p_k.$$

Which means that:

1. You should bet your money *proportional to the probabilities*,
2. Even if you do that, you still lose money at a rate given by the *entropy* of $\vec{P}$.
3. If you don't, you lose money at a *faster* rate given by the *cross-entropy* between $\vec{P}$ and $\vec{X}$.

The difference is called the *relative entropy* of $\vec{P}$ with respect to $\vec{X}$,

$$D(\vec{P}|\vec{X}) \equiv H(\vec{P}|\vec{X}) - H(\vec{P})$$

.

Now, what's interesting is that frequentists and Bayesians can both follow this conversation. And both of them interpret the result as meaning the same thing: "When playing this game, I should bet my money proportional to the probabilities."

Firefox File Edit View History Bookmarks Tools Window Help Thu 3:39 PM Robin Blume-Kohout

Robin's Notes – An ongoing quest for coherence

Robin's Notes – An o... | Bayesian probability ... | Probability axioms – ... | Quantum Theory – | Principia Mathematic...

.

Now, what's interesting is that frequentists and Bayesians can both follow this conversation. And both of them interpret the result as meaning the same thing: "When playing this game, I should bet my money proportional to the probabilities."

The frequentist justifies this by saying, "Well, duh... we just calculated what is going to happen, and if I follow this strategy then with high probability I will win more money than anybody else."

The Bayesian says, "We calculated the consequence of my prior beliefs: since I believe $\vec{P}$ about the coin, I believe with $Pr \sim 1$ that this strategy will win more money than any other strategy."

And as a pignistic Bayesian, I say, "Please stop babbling at me and place your bets! Once I see your x, I will have measured your p." In other words, I believe that probability is an empirically measurable (physical) property of an agent in contact with a system. **If you think this is [conventional] physical probability, please read the sentence again: probability is not a property of the object.**

Firefox File Edit View History Bookmarks Tools Window Help Thu 3:39 PM Robin Blume-Kohout

Robin's Notes – An ongoing quest for coherence

Robin's Notes – An o... | Bayesian probability ... | Probability axioms – ... | Quantum Theory – | Principia Mathematic...

Now, what's interesting is that frequentists and Bayesians can both follow this conversation. And both of them interpret the result as meaning the same thing: "When playing this game, I should bet my money proportional to the probabilities."

The frequentist justifies this by saying, "Well, duh... we just calculated what is going to happen, and if I follow this strategy then with high probability I will win more money than anybody else."

The Bayesian says, "We calculated the consequence of my prior beliefs: since I believe $\vec{P}$ about the coin, I believe with $Pr \sim 1$ that this strategy will win more money than any other strategy."

And as a pignistic Bayesian, I say, "Please stop babbling at me and place your bets! Once I see your x, I will have measured your p." In other words, I believe that probability is an empirically measurable (physical) property of an agent in contact with a system. **If you think this is [conventional] physical probability, please read the sentence again: probability is not a property of the object.**

This may well be what DeFinetti had in mind all along, and (to me), it's the very conversation. And both of them interpret the result as meaning the same thing: "When playing this game, I should bet my money proportional to the probabilities."

The frequentist justifies this by saying, "Well, duh... we just calculated what is going to happen, and if I follow this strategy then with high probability I will win more money than anybody else."

The Bayesian says, "We calculated the consequence of my prior beliefs: since I believe $\vec{P}$ about the coin, I believe with $Pr \sim 1$ that this strategy will win more money than any other strategy."

And as a pignistic Bayesian, I say, "Please stop babbling at me and place your bets! Once I see your x, I will have measured your p." In other words, I believe that probability is an empirically measurable (physical) property of an agent in contact with a system. **If you think this is [conventional] physical probability, please read the sentence again: probability is not a property of the object.**

This may well be what DeFinetti had in mind all along, and (to me), it's the very root of Bayesian probability. But it's dreadfully obscured by the usual language about belief, which is really easy to misinterpret as solipsistic, philosophical, mystical, etc.

Statistical Inference

> "The chief object of this work is to provide a method of drawing inferences from observational data that will be self-consistent and can also be used in practice." *Jeffreys, preface*

> "The fundamental problem of scientific progress, and a fundamental one of everyday life, is that of learning from experience." *Jeffreys, 1*

So far, we have discussed "direct probability". This is shorthand for any question, calculation, or concept related to the idea that probability determines events (or beliefs about events). But how are we to get our hands on the probabilities in the first place? After all, we only see events. How can we deduce probabilities from events?

So far, we have discussed "direct probability". This is shorthand for any question, calculation, or concept related to the idea that probability determines events (or beliefs about events). But how are we to get our hands on the probabilities in the first place? After all, we only see events. How can we deduce probabilities from events?

This used to be called "inverse probability", back when Laplace naively thought you could just hurl Bayes' Theorem at it. Fischer, Von Mises, and company rightly smacked Laplace upside the head for this, and in the process sent statistics into a sort of frequentist Dark Ages for 100 years. (Actually, this may have been a good thing... and, after all, the Dark Ages gave us madrigals, Notre Dame, and Gregorian Chant.) These days, it's called *statistical inference*.

Frequentists think the basic problem of statistical inference is **to estimate the true value of something (e.g., a probability)**. The framework in which classical statistical inference operates is:

- There is a [possibly imaginary] *population* described by *population*

Frequentists think the basic problem of statistical inference is **to estimate the true value of something (e.g., a probability)**. The framework in which classical statistical inference operates is:

- There is a [possibly imaginary] *population* described by *population parameters*. These are real physical quantities, but not directly observable.
- You observe *samples*, which yields data, and you compute a *statistic* from the data. This is an estimate of the underlying parameter, and your goal is to get it as "close" as possible.

Bayesians believe that the basic problem of statistical inference is either:

1. **to infer the relative probabilities of various theories,** or
2. **to make good decisions about future events (i.e., predict the future).**

Their framework is less rigid and restricted, but a pretty common setup is:

- You have a set Θ of theories θ. Each theory assigns probabilities to all the

Firefox File Edit View History Bookmarks Tools Window Help Thu 3:48 PM Robin Blume-Kohout

Robin's Notes – An ongoing quest for coherence

Robin's Notes – An o... | Bayesian probability ... | Probability axioms – ... | Quantum Theory – | Principia Mathematic...

Frequentists think the basic problem of statistical inference is **to estimate the true value of something (e.g., a probability)**. The framework in which classical statistical inference operates is:

- There is a [possibly imaginary] *population* described by *population parameters*. These are real physical quantities, but not directly observable.
- You observe *samples*, which yields data, and you compute a *statistic* from the data. This is an estimate of the underlying parameter, and your goal is to get it as "close" as possible.

Bayesians believe that the basic problem of statistical inference is either:

1. **to infer the relative probabilities of various theories,** or
2. **to make good decisions about future events (i.e., predict the future).**

Their framework is less rigid and restricted, but a pretty common setup is:

- You have a set Θ of theories θ. Each theory assigns probabilities to all the possible observations D, i.e.

Firefox File Edit View History Bookmarks Tools Window Help Thu 3:49 PM Robin Blume-Kohout

Robin's Notes – An ongoing quest for coherence

Robin's Notes – An o... | Bayesian probability ... | Probability axioms – ... | Quantum Theory – | Principia Mathematic...

statistical inference operates is:

- There is a [possibly imaginary] *population* described by *population parameters*. These are real physical quantities, but not directly observable.
- You observe *samples*, which yields data, and you compute a *statistic* from the data. This is an estimate of the underlying parameter, and your goal is to get it as "close" as possible.

Bayesians believe that the basic problem of statistical inference is either:

1. **to infer the relative probabilities of various theories,** or
2. **to make good decisions about future events (i.e., predict the future).**

Their framework is less rigid and restricted, but a pretty common setup is:

- You have a set Θ of theories θ. Each theory assigns probabilities to all the possible observations D, i.e.

$$Pr_\theta(D),$$

or (more formally),

- There is a [possibly imaginary] *population* described by *population parameters*. These are real physical quantities, but not directly observable.
- You observe *samples*, which yields data, and you compute a *statistic* from the data. This is an estimate of the underlying parameter, and your goal is to get it as "close" as possible.

Bayesians believe that the basic problem of statistical inference is either:

1. **to infer the relative probabilities of various theories,** or
2. **to make good decisions about future events (i.e., predict the future).**

Their framework is less rigid and restricted, but a pretty common setup is:

- You have a set Θ of theories θ. Each theory assigns probabilities to all the possible observations D, i.e.

$$Pr_\theta(D),$$

or (more formally),

$$P(D|\theta)$$

Firefox File Edit View History Bookmarks Tools Window Help Thu 3:50 PM Robin Blume-Kohout

Robin's Notes – An ongoing quest for coherence

Robin's Notes – An o... | Bayesian probability ... | Probability axioms – ... | Quantum Theory – | Principia Mathematic...

2. **to make good decisions about future events (i.e., predict the future).**

Their framework is less rigid and restricted, but a pretty common setup is:

- You have a set Θ of theories θ. Each theory assigns probabilities to all the possible observations D, i.e.

$$Pr_\theta(D),$$

or (more formally),

$$P(D|\theta)$$

- You have some beliefs about Nature, represented by a *prior distribution* $P_0(\theta)$. Usually these beliefs are "I have no idea which theory is true!!!", and we work very hard at figuring out how to represent this mathematically!
- You make some observations, ending up with a particular set of *data* D'.
- You update your prior, by inserting the data into *Bayes' Theorem*, to obtain a *posterior*:

$$P_f(\theta) \equiv P(\theta|D') = \frac{P_0(\theta)P(D'|\theta)}{P(D')}$$

- You use your posterior to make an estimate:

Their framework is less rigid and restricted, but a pretty common setup is:

- You have a set Θ of theories θ. Each theory assigns probabilities to all the possible observations D, i.e.

$$Pr_\theta(D),$$

or (more formally),

$$P(D|\theta)$$

- You have some beliefs about Nature, represented by a *prior distribution* $P_0(\theta)$. Usually these beliefs are "I have no idea which theory is true!!!", and we work very hard at figuring out how to represent this mathematically!
- You make some observations, ending up with a particular set of *data* D'.
- You update your prior, by inserting the data into *Bayes' Theorem*, to obtain a *posterior*:

$$P_f(\theta) \equiv P(\theta|D') = \frac{P_0(\theta)P(D'|\theta)}{P(D')}$$

- You use your posterior to make an estimate:
  - (i) you're content to just write down the posterior distribution over

Firefox File Edit View History Bookmarks Tools Window Help Thu 3:51 PM Robin Blume-Kohout

Robin's Notes – An ongoing quest for coherence

Robin's Notes – An o... | Bayesian probability ... | Probability axioms – ... | Quantum Theory – | Principia Mathematic...

theories,

- (ii) you pick your favorite theory, using the posterior as a guide (e.g., the mean or the mode of the posterior)
- (iii) you use the posterior to calculate the *expected value* of some **cost function**, conditional on your course of action (e.g., what theory you pick), and then pick the theory that minimizes the expected cost.

The central quantity in statistical inference is the **Likelihood Function**:

$$\mathcal{L}(theory) = Pr(\text{observed data}|\text{theory}) = P(D|\theta)$$

There is a widely agreed-upon principle called the Likelihood Principle, which states that **Everything that the data can tell you about θ is contained in $\mathcal{L}(\theta)$.**

By and large, both frequentists and Bayesians buy into this. What they disagree about how you should deal with it. Frequentists basically believe you should use *nothing* but the likelihood — which, if you believe the Likelihood Principle, is equivalent to saying that your estimate of θ should be based entirely upon the data.

Firefox File Edit View History Bookmarks Tools Window Help Thu 3:51 PM Robin Blume-Kohout

Robin's Notes – An ongoing quest for coherence

Robin's Notes – An o... | Bayesian probability ... | Probability axioms – ... | Quantum Theory – | Principia Mathematic...

The central quantity in statistical inference is the **Likelihood Function**:

$$\mathcal{L}(theory) = Pr(\text{observed data}|\text{theory}) = P(D|\theta)$$

There is a widely agreed-upon principle called the Likelihood Principle, which states that **Everything that the data can tell you about θ is contained in $\mathcal{L}(\theta)$.**

By and large, both frequentists and Bayesians buy into this. What they disagree about how you should deal with it. Frequentists basically believe you should use *nothing* but the likelihood — which, if you believe the Likelihood Principle, is equivalent to saying that your estimate of θ should be based entirely upon the data.

Bayesians think this is silly for two reasons:

- What if you knew something about θ beforehand? Should you be required to ignore it, even if the data turn out to be worthless?
- This whole business about estimating θ is a total red herring! The point is to predict the future, or make a decision. We need to:
  - (i) specify how choosing a value of θ will influence your decision,
  - (ii) determine the consequences of making the right or wrong decision

Firefox File Edit View History Bookmarks Tools Window Help Thu 3:53 PM Robin Blume-Kohout

Robin's Notes – An ongoing quest for coherence

Robin's Notes – An o... | Bayesian probability ... | Probability axioms – ... | Quantum Theory – | Principia Mathematic...

By and large, both frequentists and Bayesians buy into this. What they disagree about how you should deal with it. Frequentists basically believe you should use *nothing* but the likelihood — which, if you believe the Likelihood Principle, is equivalent to saying that your estimate of θ should be based entirely upon the data.

Bayesians think this is silly for two reasons:

- What if you knew something about θ beforehand? Should you be required to ignore it, even if the data turn out to be worthless?
- This whole business about estimating θ is a total red herring! The point is to predict the future, or make a decision. We need to:
    - (i) specify how choosing a value of θ will influence your decision,
    - (ii) determine the consequences of making the right or wrong decision (**cost function**).

So Bayesians typically believe that you gotta include other factors besides just the data. To put it very concisely:

- **Frequentists seek the truth.**
- **Bayesians seek the best decision.** (and may not even believe that there *is* such a thing as "truth").

So Bayesians typically believe that you gotta include other factors besides just the data. To put it very concisely:

- **Frequentists seek the truth.**
- **Bayesians seek the best decision.** (and may not even believe that there *is* such a thing as "truth").

In practice, the central point of disagreement is *priors*. A prior is a probability distribution over theories, or "states of nature", $P_0(\theta)$.

Bayesians say:

- that this is totally legitimate, as you obviously have beliefs about nature, which can be represented by probabilities,
- that the whole point is to update this prior (and maybe use it to make a decision),
- and that they can pick "noninformative priors" to represent total ignorance

Frequentists say:

- that this is ridiculous, because Nature is not a random variable (and you can't assign probabilities to deterministic things),
- there is no such thing as a noninformative prior,

So Bayesians typically believe that you gotta include other factors besides just the data. To put it very concisely:

- **Frequentists seek the truth.**
- **Bayesians seek the best decision.** (and may not even believe that there *is* such a thing as "truth").

In practice, the central point of disagreement is *priors*. A prior is a probability distribution over theories, or "states of nature", $P_0(\theta)$.

Bayesians say:

- that this is totally legitimate, as you obviously have beliefs about nature, which can be represented by probabilities,
- that the whole point is to update this prior (and maybe use it to make a decision),
- and that they can pick "noninformative priors" to represent total ignorance

Frequentists say:

- that this is ridiculous, because Nature is not a random variable (and you can't assign probabilities to deterministic things),
- there is no such thing as a noninformative prior,

Firefox File Edit View History Bookmarks Tools Window Help Thu 3:54 PM Robin Blume-Kohout

Robin's Notes – An ongoing quest for coherence

Robin's Notes – An o... | Bayesian probability ... | Probability axioms – ... | Quantum Theory – | Principia Mathematic...

So Bayesians typically believe that you gotta include other factors besides just the data. To put it very concisely:

- **Frequentists seek the truth.**
- **Bayesians seek the best decision.** (and may not even believe that there *is* such a thing as "truth").

In practice, the central point of disagreement is *priors.* A prior is a probability distribution over theories, or "states of nature", $P_0(\theta)$.

Bayesians say:

- that this is totally legitimate, as you obviously have beliefs about nature, which can be represented by probabilities,
- that the whole point is to update this prior (and maybe use it to make a decision),
- and that they can pick "noninformative priors" to represent total ignorance

Frequentists say:

- that this is ridiculous, because Nature is not a random variable (and you can't assign probabilities to deterministic things),
- there is no such thing as a noninformative prior,
- and so the Bayesian's conclusions are inevitably contaminated by subjectivity

Firefox File Edit View History Bookmarks Tools Window Help Thu 3:54 PM Robin Blume-Kohout

Robin's Notes – An ongoing quest for coherence

Robin's Notes – An o... | Bayesian probability ... | Probability axioms – ... | Quantum Theory – | Principia Mathematic...

Oddly enough, Bayes' Theorem (Rule) is not controversial, whatever you may have heard. It's just that frequentists don't believe you can apply it to statistical inference, because you can't write down a prior. And without a prior, you **cannot** get a posterior — i.e., you cannot *end up* assigning probabilities to theories unless you *start* by assigning probabilities.

Bayes' Theorem is just a rewriting of conditional probability, which you can derive from Kolmogorov's axioms.

$$P(A,B) = P(B) \cdot P(A|B) = P(A) \cdot P(B|A)$$

so

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Statistical Inference meets Gambling

Firefox File Edit View History Bookmarks Tools Window Help Thu 3:54 PM Robin Blume-Kohout

Robin's Notes – An ongoing quest for coherence

Robin's Notes – An o... | Bayesian probability ... | Probability axioms – ... | Quantum Theory – | Principia Mathematic...

Statistical Inference meets Gambling

Let's conclude by taking a look at statistical inference in the case of our little gambling problem.

Suppose you need to bet on coin flips — N of them in succession. If you "know" the $\vec{P}$ describing each coin flip (whatever that means to you), then you know how to bet. But what if you *don't* know it?

All you know is that the coin flips are: (i) independent of each other, and (ii) all identical to each other. But you don't know $\vec{P}$. What do you do?

> Note: while it's easy for a frequentist to say "There is a true $\vec{P}$, but I do not know it," it's a bit harder for a Bayesian to do this. DeFinetti demonstrated that the most general way for a Bayesian to state this is by saying "To each number N of coin flips, I assign an

Statistical Inference meets Gambling

Let's conclude by taking a look at statistical inference in the case of our little gambling problem.

Suppose you need to bet on coin flips — N of them in succession. If you "know" the $\vec{P}$ describing each coin flip (whatever that means to you), then you know how to bet. But what if you *don't* know it?

All you know is that the coin flips are: (i) independent of each other, and (ii) all identical to each other. But you don't know $\vec{P}$. What do you do?

> Note: while it's easy for a frequentist to say "There is a true $\vec{P}$, but I do not know it," it's a bit harder for a Bayesian to do this. DeFinetti demonstrated that the most general way for a Bayesian to state this is by saying "To each number N of coin flips, I assign an *exchangeable* probability distribution." An exchangeable distribution

You realize that while you may take a loss on the first few flips, you can start using the *data* that you collect (i.e., n_{heads} and n_{tails}) to bet more wisely. In effect, you are going to use the first N flips as data to assign estimated probabilities $\hat{P}_N$.

So how should you do that?

One obvious idea is to go with the *empirical distribution*:

$$\hat{P} = \left(\frac{n_{heads}}{N}, \frac{n_{tails}}{N}\right).$$

This is actually what a hardcore frequentist would do — it is the *Maximum Likelihood Estimator*. But a moment's thought shows that leads to ruin:

- If the first N flips come up "heads", then you will assign $\hat{P}_N = (1, 0)$.
- If they come up "tails" then you will assign $\hat{P}_N = (0, 1)$.
- If the $(N+1)$th flip is different, you will lose all your money.
- So you lose all your money quite quickly.

There is an ingenuous solution to this. Consider all possible schemes. In each

You realize that while you may take a loss on the first few flips, you can start using the *data* that you collect (i.e., n_{heads} and n_{tails}) to bet more wisely. In effect, you are going to use the first N flips as data to assign estimated probabilities $\hat{P}_N$.

So how should you do that?

One obvious idea is to go with the *empirical distribution*:

$$\hat{P} = \left(\frac{n_{heads}}{N}, \frac{n_{tails}}{N}\right).$$

This is actually what a hardcore frequentist would do — it is the *Maximum Likelihood Estimator*. But a moment's thought shows that leads to ruin:

- If the first N flips come up "heads", then you will assign $\hat{P}_N = (1,0)$.
- If they come up "tails" then you will assign $\hat{P}_N = (0,1)$.
- If the $(N+1)$th flip is different, you will lose all your money.
- So you lose all your money quite quickly.

There is an ingenuous solution to this. Consider all possible schemes. In each

- So you lose all your money quite quickly.

There is an ingenuous solution to this. Consider all possible schemes. In each case, you will assign some estimate $\hat{P}_N$ *conditional* upon seeing flips $\{f_1 \ldots f_N\}$. We could write this as $P(f_{N+1}|f_1 \ldots N)$.

So, when faced with our first coin, we bet according to $P(f_1|\text{nothing})$. And our logmoney increases by

$$\sum_{f_1} p_{f_1} \log P(f_1|\text{nothing})$$

In the second round, we bet according to $P(f_2|f_1)$, and our logmoney increases by

$$\sum_{f_2} p_{f_1} \log P(f_1|f_2),$$

which means that over the first two flips, our logmoney increases by

$$\sum p_{f_1,f_2} [\log P(f_1|f_2) + \log P(f_1)],$$

There is an ingenuous solution to this. Consider all possible schemes. In each case, you will assign some estimate $\hat{P}_N$ *conditional* upon seeing flips $\{f_1 \ldots f_N\}$. We could write this as $P(f_{N+1}|f_1 \ldots N)$.

So, when faced with our first coin, we bet according to $P(f_1|\text{nothing})$. And our logmoney increases by

$$\sum_{f_1} p_{f_1} \log P(f_1|\text{nothing})$$

In the second round, we bet according to $P(f_2|f_1)$, and our logmoney increases by

$$\sum_{f_2} p_{f_1} \log P(f_1|f_2),$$

which means that over the first two flips, our logmoney increases by

$$\sum_{f_1,f_2} p_{f_1,f_2} [\log P(f_1|f_2) + \log P(f_1)],$$

which is equal to

$$\sum_{f_1,f_2} p_{f_1,f_2} \log\left[P(f_1|f_2)P(f_1)\right],$$

and that's equal to

$$\sum_{f_1,f_2} p_{f_1,f_2} \log\left[P(f_1,f_2)\right],$$

if we just *define* $P(f_1, f_2)$ as

$$P(f_1,f_2) \equiv P(f_1)P(f_2|f_1).$$

If you work through a few more rounds, you find that the pattern continues, and after N rounds, your logmoney increases by

$$\sum_{f_1 \dots f_N} p_{f_1 \dots f_N} \log\left[P(f_1 \dots f_N)\right].$$

Which means, essentially, that ***every* strategy for adaptive betting can be analyzed as a Bayesian one — which assigns probabilities ahead of time to every string of heads and tails.**

Firefox File Edit View History Bookmarks Tools Window Help Thu 3:59 PM Robin Blume-Kohout

Robin's Notes – An ongoing quest for coherence

Robin's Notes – An o... | Bayesian probability ... | Probability axioms – ... | Quantum Theory – | Principia Mathematic...

But there's more.

We still have a problem: there are lots of different Bayesian schemes, corresponding to different priors over $\vec{P}$. How can we choose between them? The Bayesian scheme seems to be entirely subjective, not preferring any belief to any other.

The answer involves a lot of math, but is simple to state. We can simply ask "Which strategy works best in the **worst** case?"

I.e., for each strategy (prior), consider all possible $\vec{P}$, identify the one that makes you lose as much money as possible, and take that to be the worst case. Then rank the various priors by their worst-case behavior, and pick the one that works the best!

The result is an estimation scheme that — while Bayesian — can be justified **objectively**, even to a frequentist.

Firefox File Edit View History Bookmarks Tools Window Help Thu 4:00 PM Robin Blume-Kohout

Robin's Notes – An ongoing quest for coherence

Robin's Notes – An o... | Bayesian probability ... | Probability axioms – ... | Quantum Theory – | Principia Mathematic...

> "It is hard to imagine that the current situation, with several competing foundations for statistics, will exist indefinitely. Assuming that a unified foundation is inevitable, what will it be?
>
> - ...First, the language of statistics will be Bayesian. Statistics is about measuring uncertainty...
> - ...this is not about subjectivity or objectivity; the Bayesian language can be used for either subjective or objective statistical analysis.
> - ...from a methodological perspective, it is becoming clear that both Bayesian and frequentist methodology is (sic) giong to be important.
> - ...In nonparametric analysis... Bayesian procedures can behave poorly from a frequentist perspective. Although poor frequentist performance is not necessarily damning to a Bayesian, it typically should be viewed as a warning sign that something is amiss

Firefox File Edit View History Bookmarks Tools Window Help Thu 4:00 PM Robin Blume-Kohout

Robin's Notes – An ongoing quest for coherence

Robin's Notes – An o... | Bayesian probability ... | Probability axioms – ... | Quantum Theory – | Principia Mathematic...

> from a conditional frequentist perspective.
>
> - ...I am *not* arguing for an eclectic attitude toward statistics here; indeed I think the general refusal in our field to strive for a unified perspective has been the single biggest impediment to its advancement." *Berger, JASA 2000, p. 1272*

> "The Bayesian `machine', together with Markov Chain Monte Carlo, is arguably the most powerful mechanism ever created for processing data and knowledge." *Berger, JASA 2000, p. 1273*

> "Statisticians should readily use both Bayesian and frequentist ideas. In Section 2 we discuss situations in which simultaneous frequentist and Bayesian thinking is essentially required. For the most part, however, the situations we discuss are situations in which it is simply extremely useful for Bayesians to use frequentist methodology or frequentists to use Bayesian methodology." *Berger and Bayarri, Stat. Sci 2004, pg. 58*

foundations for statistics, will exist indefinitely. Assuming that a unified foundation is inevitable, what will it be?

- ...First, the language of statistics will be Bayesian. Statistics is about measuring uncertainty...
- ...this is not about subjectivity or objectivity; the Bayesian language can be used for either subjective or objective statistical analysis.
- ...from a methodological perspective, it is becoming clear that both Bayesian and frequentist methodology is (sic) giong to be important.
- ...In nonparametric analysis... Bayesian procedures can behave poorly from a frequentist perspective. Although poor frequentist performance is not necessarily damning to a Bayesian, it typically should be viewed as a warning sign that something is amiss.
- ...there are an increasing number of examples in which frequentist arguments yield satisfactory answers quite directly, while Bayesian analysis requires a formidable amount of extra work... I believe that the frequentist answer can be accepted by Bayesians as an approximate Bayesian answer.
- ...It has long been known that "optimal" unconditional frequentist proceures must be Bayesian... there is growing evidence that this must be so even

Firefox File Edit View History Bookmarks Tools Window Help Thu 4:00 PM Robin Blume-Kohout

Robin's Notes – An ongoing quest for coherence

Robin's Notes – An o... | Bayesian probability ... | Probability axioms – ... | Quantum Theory – | Principia Mathematic...

"It is hard to imagine that the current situation, with several competing foundations for statistics, will exist indefinitely. Assuming that a unified foundation is inevitable, what will it be?

- ...First, the language of statistics will be Bayesian. Statistics is about measuring uncertainty...
- ...this is not about subjectivity or objectivity; the Bayesian language can be used for either subjective or objective statistical analysis.
- ...from a methodological perspective, it is becoming clear that both Bayesian and frequentist methodology is (sic) giong to be important.
- ...In nonparametric analysis... Bayesian procedures can behave poorly from a frequentist perspective. Although poor frequentist performance is not necessarily damning to a Bayesian, it typically should be viewed as a warning sign that something is amiss.
- ...there are an increasing number of examples in which frequentist arguments yield satisfactory answers quite directly, while Bayesian analysis requires a formidable amount of extra work... I believe that the frequentist answer can be accepted by Bayesians as an approximate Bayesian answer.
- ...It has long been known that "optimal" unconditional frequentist proceures

Firefox File Edit View History Bookmarks Tools Window Help Thu 4:00 PM Robin Blume-Kohout

Robin's Notes – An ongoing quest for coherence

Robin's Notes – An o... | Bayesian probability ... | Probability axioms – ... | Quantum Theory – | Principia Mathematic...

> "It is hard to imagine that the current situation, with several competing foundations for statistics, will exist indefinitely. Assuming that a unified foundation is inevitable, what will it be?
>
> - ...First, the language of statistics will be Bayesian. Statistics is about measuring uncertainty...
> - ...this is not about subjectivity or objectivity; the Bayesian language can be used for either subjective or objective statistical analysis.
> - ...from a methodological perspective, it is becoming clear that both Bayesian and frequentist methodology is (sic) giong to be important.
> - ...In nonparametric analysis... Bayesian procedures can behave poorly from a frequentist perspective. Although poor frequentist performance is not necessarily damning to a Bayesian, it typically should be viewed as a warning sign that something is amiss.
> - ...there are an increasing number of examples in which frequentist arguments yield satisfactory answers quite directly, while Bayesian analysis requires a formidable amount of extra work... I believe that the frequentist answer can be accepted by Bayesians as an approximate Bayesian answer.
> - ...It has long been known that "optimal" unconditional frequentist proceures must be Bayesian... there is growing evidence that this must be so even from a conditional frequentist perspective.
> - ...I am *not* arguing for an eclectic attitude toward statistics here; indeed I think the general refusal in our field to strive for a unified perspective has been the single biggest