Though A/B testing seems simple in that you pit page 'A' against page 'B' and see which one perfoms better, figuring out whether your results actually mean anything is quite complicated.

Luckily, great minds have been working on this problem for a long time and have developed data science techniques to help.

But to benefit from their work, marketers have to understand the problems and know where to find the solutions.

In the first two parts of this post, I explained [how to determine the sample size](https://econsultancy.com/blog/65672-using-data-science-with-a-b-tests-sample-sizing) required to run a significant A/B test and what to do if you cannot get enough samples [(chi-squared testing](https://econsultancy.com/blog/65711-using-data-science-with-a-b-tests-chi-squared-testing)).

But is there anything you can do if you face a very small sample size? Can you measure results that are not in the 1000s, 100s, but the 10s?

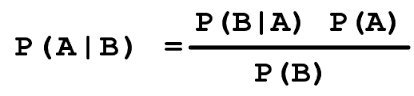


Yes you can. There is another approach to help in this case, though it's a bit more difficult to understand.

The Bayesian way

You can apply Bayesian analysis to your A/B testing, which is based on a formula devised by an English Presbyterian minister, who also happened to be a statistician: Thomas Bayes.

Here is the formula upon which the analysis is based:



Don't worry about the equation just yet.  Just know that it means is that when you make a decision about something, you can, mathematically, use*all of the useful informational available* - and not just the facts you have collected.

That is, when you're examining evidence you have to not only look at what's in front of you, but think about what is likely to be true as well.

Sounds reasonable, but how do you do that?  And how does this apply to A/B testing?

First, the previous approach

Well, with the previous tests, sample-sizing and chi-squared, you based your decision on whether 'A' beat 'B' only from the data in the test.  All other information is irrelevant as you are simply testing 'A' against 'B'.

And this sounds right.  We just want to know whether 'A' is better than 'B'. And nothing else is relevant, much like justice should be blind to outside beliefs.



The Bayesian approach

Well the Bayesian approach lets you to think a bit deeper about the problem. When you're testing 'A' against 'B' you actually do have some other information. You know what makes sense. And this is valuable information when making a decision.

So, sure, justice may be blind - but sometimes we need her to peek a bit and make sure what's on the scale makes sense!



For A/B testing, what this means is that you, the marketer, have to come up with what conversion rate 'makes sense'. That is, if you typically see a 10% conversion in 'A' you would not, during the test, expect to see it at 100%.

Then instead of only finding the winner in the test itself, Bayesian analysis will include your 'prior knowledge' into the test. That is, you can tell the test what you 'believe' the right answer to be - and then using that knowledge, or prior belief, the test can tell you whether 'A' beat 'B'.

And, because it uses more information than is in the test itself, it can give you a defensible answer as to whether 'A' beat 'B' from a remarkably small sample size.

The stats

The math behind Bayesian A/B testing is terrifying, and far beyond the scope of this post. You should, however, rest assured that there is a lot of confidence in Bayesian methods among statisticians.

And the best bit for us, the marketers, is that it works well even with minimal results.

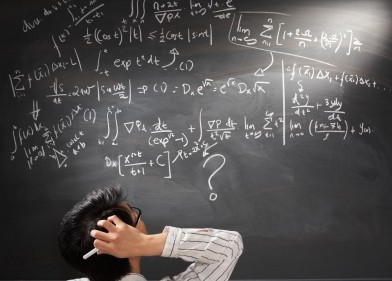
If you're still curious, there is a lot of material on the web which give real examples of how Bayes analysis has been used - with medical tests and A/B tests. [Here is one of my favorites](http://yudkowsky.net/rational/bayes), but [do try the Quora thread as well](http://www.quora.com/What-is-an-intuitive-explanation-of-Bayes-Rule).

The Catch

So, what's the catch? Why don't marketers just use Bayesian all the time?

Well, [some do](http://dataorigami.net/blogs/napkin-folding/17987495-the-class-imbalance-problem-in-a-b-testing), but most do not because the domain knowledge for Bayesian testing is a lot more important than for the other tests.

See, you need to come up with an estimate for what you *believe your conversion rate to be*(say 5%) and *how likely it is to deviate from that number* (say ±2%) and then graph it.



Say what??

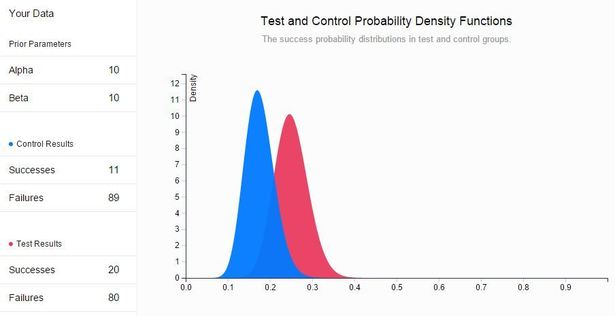
Yeah, this is the hard bit. You might know the conversion percentage, but it's a bit tough to come up with the 'deviation' magnitude - and even tougher to come up with the number.  So most people revert to the previous tests.

But should you be able to come up with a good deviation figure, then the end result is that Bayesian analysis can tell you - after comparing *any* A and B test results - how likely it is that B is actually better than A.

OK, that's still confusing.

Perhaps it's good to look at a concrete example.  Click over to [this really great Bayesian A/B calculator](http://developers.lyst.com/bayesian-calculator/) - and let's have a look.

Like the previous example, we're going to look at the difference between an 'A' test which had 11 conversions out of 100 and a 'B' test which had 20 conversions out of 100.



OK this is more familiar territory. Here, you can input your 'successes' and 'failures' in the same way that you did in previous A/B testing calculators. In this example the 'A' test had 11 successes and 89 failures. The 'B' test had 20 successes and 80 failures.

This makes sense, but...

...What are these new 'alpha' and 'beta' parameters for 'prior belief'. How do you come up with those?

Enter hacking

Well, again the math is complicated, but with this tool you can use your hacking skills, come up with a few numbers and decide whether they look right. Let's run through an example.

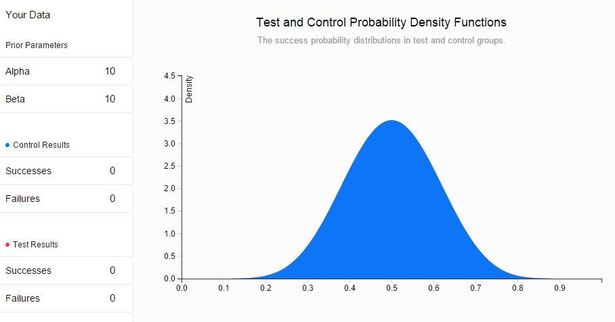
How to find alpha and beta

Consider a site with a typical 5% conversion rate which moves around a bit, but not a lot - and never gets past 20%.

To get your alpha and beta numbers for that scenario, first clear your samples and recalculate.

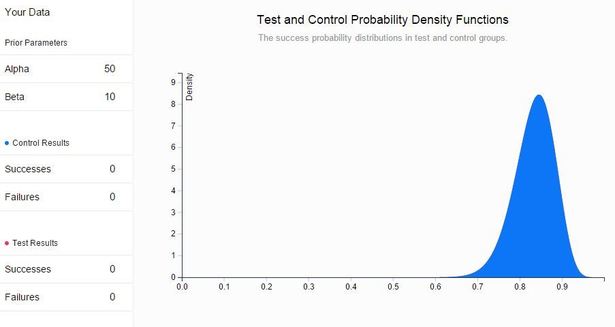
Then hack away with a few numbers and try to get the blue graph to match your intuition.

Look at the example below: Alpha=10 and Beta=10.



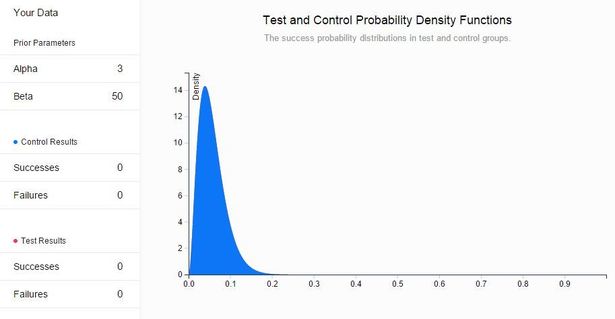
OK, that gives you a conversion rate of 50% with a wide deviation of results. That is, it's 50% on average but it can vary a lot.  Sometimes it's as low as 20%, sometimes as high as 80%.  That's not right.

So, fiddle with the parameters - let's move alpha to 50 and keep beta at 10.



Whoa - that's totally wrong!  That would be a 'prior belief' that your conversion rate was typically 85% and occasionally moved near to 100%.

OK, now let's use ones I prepared earlier.  Alpha=3, Beta=50.

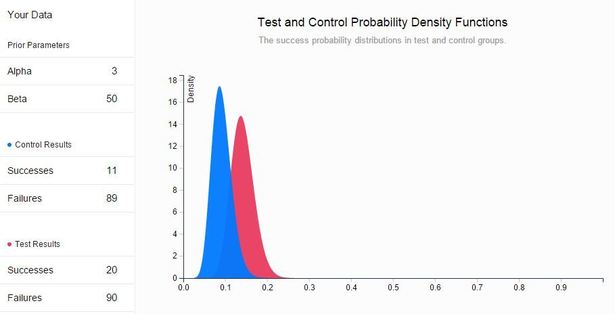


There! That looks right.  The conversion is most likely to be around 5% (say per day) with some, but not much, deviation around that number.

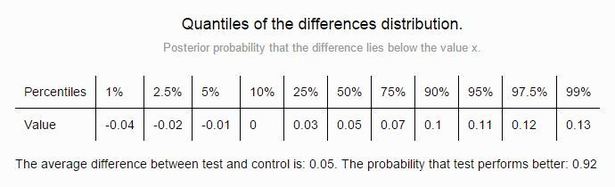
So now we have alpha and beta...

And then you can run the test. And you get - as predicted - 5 successes vs. 95 failures for the control results (the 'A'). The test (the 'B') produces 10 successes vs 90 failures.

Finger in the air, you'd say that was a success -you've jumped from 5% conversion to 10% conversion. What does out test say?



Bayes testing largely agrees.  In fact it says that it's 92% likely that 'B' performed better than 'A'.



If you can live with that uncertainty, and most marketers probably can, then you have an answer from Bayes A/B testing that you couldn't get with the original A/B test(which, for those keeping score, is known as Frequentist).

So... (TL;DR)

Bayesian analysis of A/B tests allow you to include your domain knowledge *in the test itself* so that you can get an accurate - and defensible result - from a remarkably few test samples.

For this reason, I suspect that this method of analysis will become more popular over time - so it's worth understanding both the theory and the practice.

One pitfall, of course, is that the results are only as good as your domain knowledge or 'prior belief'.  This isn't the method to try on a new ad or campaign with no track record, nor should you entrust the 'prior belief' figure to someone who does not have intimate knowledge of previous results.

That said, it's worthwhile trying out on everything as you will almost certainly learn something about how good - or how poor - your test results are when analyzed properly.

End of the series

Hopefully this series on making A/B tests more bulletproof with data science has been useful to you. I think applying real statistics to digital marketing analytics is becoming more popular now, but it does take some effort to get right. And since we have the data and the tools at our disposal, it makes sense to both learn and do the analysis.

Good luck with your tests - and do let me know of any other statistical methods or tests you may use in the comments!

[Jeff Rajeck](https://econsultancy.com/blog/authors/jeff-rajeck/)

Published 17 November, 2014 by [Jeff Rajeck](https://econsultancy.com/blog/authors/jeff-rajeck/)

Jeff Rajeck is the APAC Research Analyst for Econsultancy . You can follow him on [Twitter](https://twitter.com/jrajeck) or connect via [LinkedIn](https://sg.linkedin.com/in/jeffrajeck).

[196 more posts from this author](https://econsultancy.com/blog/authors/jeff-rajeck/)

 126 shares

[7 comments](https://econsultancy.com/blog/65755-using-data-science-with-a-b-tests-bayesian-analysis/#comments)

[Data](https://econsultancy.com/blog/tags/data/), [A/B Testing](https://econsultancy.com/blog/tags/a-b-testing/), [Bayesian](https://econsultancy.com/blog/tags/bayesian/), [Data & Analytics](https://econsultancy.com/search/?editorial_metadata=topics%3Adata_and_analytics&locale=uk&only=BlogPost)

You might be interested in

[Emarsys Deepens AI and Personalization Capabilities,...](https://www.econsultancy.com/press-releases/8367-emarsys-deepens-ai-and-personalization-capabilities-transforming-websites-into-a-truly-integrated?ite=15505&ito=876&itq=6fa2d8cb-b579-4a47-9980-c4d6252924f3&itx%5Bidio%5D=1715428)

[Emarsys Deepens AI and Personalization Capabilities,...](https://www.econsultancy.com/press-releases/8367-emarsys-deepens-ai-and-personalization-capabilities-transforming-websites-into-a-truly-integrated-?ite=15434&ito=876&itq=6fa2d8cb-b579-4a47-9980-c4d6252924f3&itx%5Bidio%5D=1715428)

[Emarsys Deepens AI and Personalization Capabilities,...](https://www.econsultancy.com/press-releases/8367-emarsys-deepens-ai-and-personalization-capabilities-transforming-websites-into-a-truly-integrated-marketing-channel?ite=15278&ito=876&itq=6fa2d8cb-b579-4a47-9980-c4d6252924f3&itx%5Bidio%5D=1715428)

[Eight steps to landing page success](https://www.econsultancy.com/blog/69612-eight-steps-to-landing-page-success?ite=16178&ito=876&itq=6fa2d8cb-b579-4a47-9980-c4d6252924f3&itx%5Bidio%5D=1715428)

[Why the butterfly effect is killing the conversion...](https://www.econsultancy.com/blog/69495-why-the-butterfly-effect-is-killing-the-conversion-optimisation-industry?ite=14864&ito=876&itq=6fa2d8cb-b579-4a47-9980-c4d6252924f3&itx%5Bidio%5D=1715428)

[Emojis in email subject lines: smiley face, or smiley poop?](https://www.econsultancy.com/blog/69595-emojis-in-email-subject-lines-smiley-face-or-smiley-poop?ite=15977&ito=876&itq=6fa2d8cb-b579-4a47-9980-c4d6252924f3&itx%5Bidio%5D=1715428)

Comments (7)

Top of Form

No-profile-pic



Bottom of Form

Avatar-blank-50x50

[Hayden Sutherland](http://www.idealinterface.co.uk/)

If you need proof that Marketing is now officially a Science (and not an art form) then this posting should provide this.   
Thanks for posting and for explaining the statistical methods behind a complex & useful area of site optimisation.

about 3 years ago

Pete Austin

Pete Austin, CINO at Fresh Relevance

Re: "Bayesian analysis will include your 'prior knowledge' into the test. That is, you can tell the test what you 'believe' the right answer to be - and then using that knowledge, or prior belief, the test can tell you whether 'A' beat 'B'."

This is true and reinforces the biggest issue with A:B testing.

(1) If you are doing a one-off test, it's all good. You can use really small sample sizes and get good results - for example when comparing subject lines to use for a single email campaign.

(2) But if you are going to use these results long term, for example to fine tune lots of aspects of your website, there's a really big problem, because as you proceed more and more of your 'knowledge' is based on previous A:B testing. Meaning it's not necessarily correct (and is increasingly unlikely to be correct as time passes and the assumptions you made in your Baysian analysis become outdated). These errors accumulate and the results of each successive test you do are increasingly unlikely to be correct.

So if you want to use your test results long term, you need to be much more careful: repeat all experiments and insist on highly significant results (see my comment against the first post in this series for more on this same issue).

about 3 years ago

Avatar-blank-50x50

Dan

Any chance of fixing the link to the Bayesian calculator used in this example please? It's pointing to a non-existent page at the moment.

about 3 years ago

Avatar-blank-50x50

Ben Hurst, Senior E-communications Executive at Nectar

Interesting article, embarassing how much has disappeared since school!

One of the calculator links is broken:   
"this really great Bayesian A/B calculator "

Thanks

about 3 years ago

[Jeff Rajeck](https://econsultancy.com/blog/authors/jeff-rajeck/)

[Jeff Rajeck](https://econsultancy.com/blog/authors/jeff-rajeck/), Research Analyst at EconsultancySmall Business

Oops sorry guys. I'll ask them to fix that. In the meantime here it is

<http://developers.lyst.com/bayesian-calculator/>

about 3 years ago

Avatar-blank-50x50

[Kristian Petterson](http://www.searchmarketinggroup.co.uk/)

Great Article.

Agreed with Pete Austin earlier - further, there is a challenge of generating "prior knowledge" if you're starting from a low dataset position (eg: something not previously investigated).

What worked a few years ago may be completely irrelevant now so using prior knowledge could actually be a hindrance to studies.

The more traditional sciences like medicine and chemistry tend to get away with it as we don't expect things like the human body to change its workings dramatically over 5/10/20 years. Can that really be said for marketing or SEO?

Anyway really enjoyed this article (and the series) - thanks!

about 3 years ago

[Jeff Rajeck](https://econsultancy.com/blog/authors/jeff-rajeck/)

[Jeff Rajeck](https://econsultancy.com/blog/authors/jeff-rajeck/), Research Analyst at EconsultancySmall Business

Great comments - thanks!

I am not a expert, but there are quite a few people in the data science community who give reasons why Bayesian A/B tests are superior to traditional tests (the ones with p-values, etc.)

Here are 3 from a post by Bayesian Witch (link below)

1) It’s far easier to interpret the results   
2) You can peek as often as you like   
3) You can alter your test material in the middle of the test

In regards to starting with bad prior information and thereby ruining all future tests, I think 3) addresses that. Your prior automatically gets adjusted as you run tests - thereby adjusting the 'right' result according to how the world has changed.

Anyway, I'm glad to get the discussion started and will try to do more on this topic as it has piqued so much interest.

<http://www.bayesianwitch.com/blog/2014/bayesian_ab_test.html>

about 3 years ago