

## Re: A simple but confusing question

**Source:** <http://sci.tech-archive.net/Archive/sci.stat.math/2004-10/0087.html>

---

**From:** Alex Strashny ([alex\\_strashny\\_at\\_yahoo.com](mailto:alex_strashny_at_yahoo.com))

**Date:** 10/06/04

Date: 6 Oct 2004 00:05:08 -0700

szhao@darkwing.uoregon.edu (Shanyu Zhao) wrote in message news:<96a39245.0409270026.63fd76c@posting.google.com>...

> *Here is the naive question:*

>

> *There are a large number of balls in a bucket, the white color balls  
> occupy  $p$  and the black balls  $(1-p)$ . If  $p$  is unknown, when you pick 100  
> balls from the bucket, find that all of them are white. Then you pick  
> the 101st ball, what is the probability that the ball still a white  
> one?*

The answer can be anything from 0 to 1. (See below.) The answer that I would consider reasonable is 101/102.

> *Is this problem a parameter estimation or hypothesis testing? If we  
> use parameter estimation, clearly  $p=1$ , which means the probability is  
> 100%. But this is not true.*

This is not right.

There are a couple of different definitions of probability. The frequentist definition states: if you pick  $n$  balls, and  $x$  of them are white, then probability of a white ball is  $p = (x/n)$  \*\*as  $n$  goes to infinity\*\*. People sometimes forget this last part, and it often does not matter; it matters here. 100 is not equal to infinity, so you cannot apply the frequentist definition to calculate probability.

This is an estimation problem.  $p$ , the probability, is a parameter that you want to estimate. A model consists of two parts: a likelihood, which expresses the interaction between the parameters and the data; and a prior, which expresses your prior beliefs about the parameters.

For this particular problem, the likelihood is Binomial( $x | n, p$ ). If the prior is Beta( $a, b$ ), then the best estimator of  $p$  (under quadratic loss) is  $p_{\text{hat}} = (x+a)/(n+a+b)$ .

What should  $a$  and  $b$  be? In principle, they could be anything. They express your beliefs about  $p$  before you drew any of the balls. That is

sci.stat.math: Re: A simple but confusing question

why I said that your estimate of  $p$  could be anything between 0 and 1.  
If  $a = 1$ ,  $b \rightarrow \text{Infinity}$ , then  $\hat{p} \rightarrow 0$ . But is such a prior reasonable? Probably not.

If you have no prior information about  $p$ , then I would say, set  $a = b = 1$ .  $\text{Beta}(1,1)$  is the  $\text{Uniform}(0,1)$ . This prior says that, a priori,  $p$  is equally likely to be anything between 0 and 1. Under this prior,  $\hat{p} = 101/102$ .

Some people would say that if you have no prior information, set  $a = b = 0.5$ . (This is the Jeffrey's prior.) I would disagree, but in practice it doesn't matter. Under Jeffrey's prior,  $\hat{p} = 100.5 / 101$ .