

Re: A simple but confusing question

Source: <http://sci.tech-archive.net/Archive/sci.stat.math/2004-10/0088.html>

From: Ian Jermyn (*Ian.Jermyn_at_sophia.inria.fr*)

Date: 10/06/04

Date: Wed, 6 Oct 2004 10:37:33 +0200

This is not an estimation problem. No loss function need be introduced to solve it, as my derivation in a previous post makes clear. The proportion of white balls in the bucket is a nuisance parameter, and should be integrated out of the expression for the probability. To describe the problem as an estimation problem risks giving a false impression of how to proceed in problems more complex than this one. In this particular case, because

$$\Pr(w_{n+1} | W_n, n, N) = \int dp \Pr(w_{n+1} | p, W_n, n, N) \Pr(p | W_n, n, N)$$

$$= \int dp \Pr(w_{n+1} | p) \Pr(p | W_n, n, N)$$

$$= \int dp p \Pr(p | W_n, n, N)$$

$$= \langle p \rangle$$

under the posterior distribution for p , i.e. because the likelihood is linear in the parameter p , the MMSE estimate for p happens to agree with the integration over the nuisance parameter. Other estimates do not so agree, however, and this is the situation in general (although in certain other simple cases similar coincidences do occur). The elimination of nuisance parameters by integration is in general not the same as estimating the parameter and substituting it into the likelihood.

Ian.

--

Ian Jermyn
ianjermyn@wanadoo.fr
"Alex Strashny" <alex_strashny@yahoo.com> a écrit dans le message de
news:e3b3c584.0410052305.1fb14d91@posting.google.com...
> 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

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

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