Opinionated Lessons in Statistics

by Bill Press

#5 Bernoulli Trials

Where we are:

$$P(A|S_B I) = \int_x P(A|S_B x I)\, p(x|I)\, dx$$
$$= \int_x \frac{1}{1+x}\, p(x|I)\, dx$$

We are trying to estimate a parameter

$$x = P(S_B|BC), \quad (0 \le x \le 1)$$

The form of our estimate is a (Bayesian) probability distribution (of the parameter, itself here just happening to be a probability)

This is a sterile exercise if it is just a debate about priors. What we need is data! Data might be a previous history of choices by the jailer in identical circumstances.

BCBCCBCCCBBCBCBCCCCBBCBCCCBCBCBBCCB

$$N = 35, \quad N_B = 15, \quad N_C = 20$$

(What's wrong with: x=15/35=0.43? Hold on…)

We hypothesize (might later try to check) that these are i.i.d. "Bernoulli trials" and therefore informative about *x*

"independent and identically distributed"

As good Bayesians, we now need $P(\mathrm{data}|x)$

$P(\text{data}|x)$ means different things in frequentist vs. Bayesian contexts, so this is a good time to understand the differences (we'll use both ideas as appropriate)

Frequentist considers the universe of what might have been, imagining repeated trials, even if they weren't actually tried, and needs no prior:

since i.i.d. only the $\mathcal{N}$'s can matter (a so-called "sufficient statistic").

$$P(\text{data}|x) = \underbrace{\binom{N}{N_B}}_{\text{no. of equivalent arrangements}} \overbrace{x^{N_B}(1-x)^{N_C}}^{\text{prob. of exact sequence seen}} \qquad \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Bayesian considers only the exact data seen, and has a prior:

$$P(x|\text{data}) \propto x^{N_B}(1-x)^{N_C}\, p(x|I)$$

but we might first suppose that the prior it is **uniform**

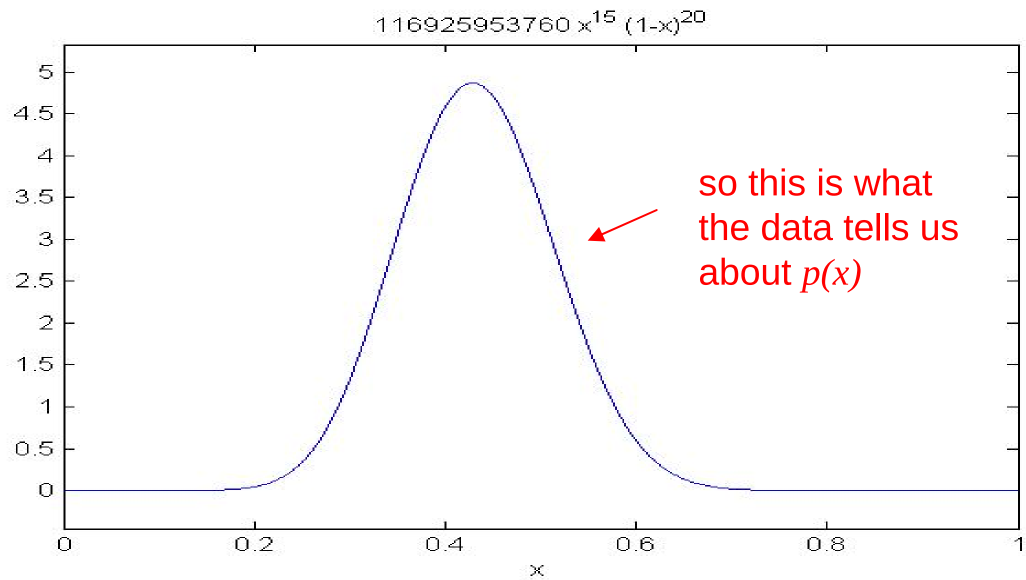
No binomial coefficient, both conceptually and also since independent of x and absorbed in the proportionality. Use only the data you see, not "equivalent arrangements" that you didn't see. This issue is one we'll return to, not always entirely sympathetically to Bayesians (e.g., goodness-of-fit).

Bayes numerator and denominator are:

$$P(x|\text{data}) \propto x^{N_B}(1-x)^{N-N_B} \times 1$$

$$\int_0^1 P(x|\text{data}) = \int_0^1 x^{N_B}(1-x)^{N-N_B}dx = \frac{\Gamma(N_B+1)\Gamma(N-N_B+1)}{\Gamma(N+2)}$$

Plot of numerator over denominator for N=35, N_B = 15:



You should learn to do calculations like this in MATLAB or Mathematica:

```
syms nn nb x
num = x^nb * (1-x)^(nn-nb)
num =
x^nb*(1-x)^(nn-nb)
denom = int(num, 0, 1)
denom =
gamma(nn-nb+1)*gamma(nb+1)/gamma(nn+2)
p = num / denom
p =
x^nb*(1-x)^(nn-nb)/gamma(nn-
nb+1)/gamma(nb+1)*gamma(nn+2)
ezplot(subs(p,[nn,nb],[35,15]),[0,1])
```

In[7]:= `num = x^nb (1 - x)^(nn - nb)`

Out[7]= $(1-x)^{-nb+nn}\, x^{nb}$

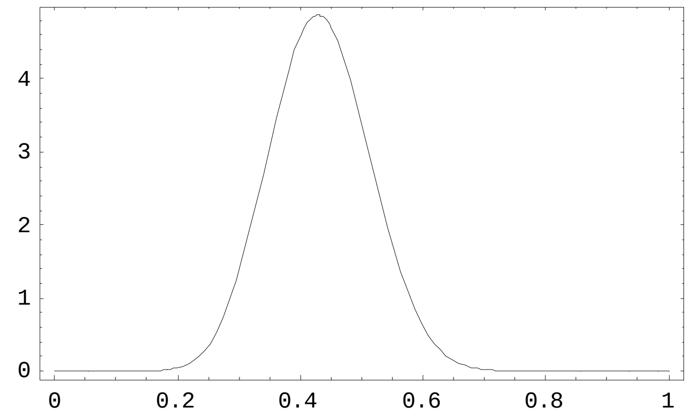
In[8]:= `denom = Integrate[num, {x, 0, 1}, GenerateConditions → False]`

Out[8]= $\frac{\text{Gamma}[1+nb]\ \text{Gamma}[1-nb+nn]}{\text{Gamma}[2+nn]}$

In[9]:= `p[x_] = num / denom`

Out[9]= $\frac{(1-x)^{-nb+nn}\, x^{nb}\ \text{Gamma}[2+nn]}{\text{Gamma}[1+nb]\ \text{Gamma}[1-nb+nn]}$

In[12]:= `Plot[p[x] /. {nn → 35, nb → 15}, {x, 0, 1}, PlotRange → All, Frame → True]`



Out[12]= - Graphics -

What we are illustrating is called Bernoulli trials:

- two possible outcomes
- i.i.d. events
- a single parameter x (the probability of one outcome)
- a sufficient statistic is the pair of numbers N and N_B

Jacob and Johann Bernoulli

$$P(\mathrm{data}|x) = x^{N_B}(1-x)^{N-N_B}$$ (in the Bayesian sense)

$$P(x|\mathrm{data}) \propto x^{N_B}(1-x)^{N-N_B} \times P(x|I)$$

for uniform prior, the Bayes denominator is, as we've seen, easy to calculate:

$$\int_0^1 P(x|\mathrm{data}) = \int_0^1 x^{N_B}(1-x)^{N-N_B}dx = \frac{\Gamma(N_B+1)\Gamma(N-N_B+1)}{\Gamma(N+2)}$$

Find the mean, standard error, and mode of our estimate for x

$$P(x|\text{data}) \propto x^{N_B}(1-x)^{N-N_B}$$

$$\frac{dP(x|\text{data})}{dx} = 0 \;\Rightarrow\; x = \frac{N_B}{N}$$

"maximum likelihood" (ML) answer is to estimate x as exactly the fraction seen

$$\langle x \rangle = \int_0^1 x P(x|\text{data}) dx = \frac{N_B + 1}{N + 2}$$

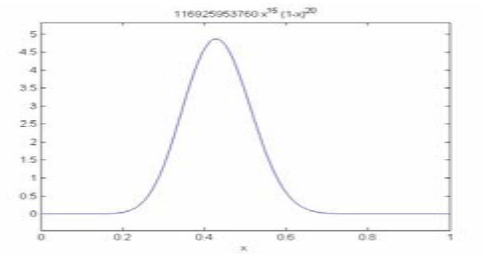
mean is the 1st moment
notice it's different from ML!

variance involves the 2nd moment,

$$\text{Var}(x) = \left\langle x^2 \right\rangle - \langle x \rangle^2 = \int_0^1 x^2 P(x|\text{data}) dx - \langle x \rangle^2 = \frac{(N_B + 1)(N - N_B + 1)}{(N+2)^2(N+3)}$$

This shows how *p(x)* gets narrower as the amount of data increases.

Are there any other mathematical forms for the prior that would still leave the Bayes denominator easy to calculate?



Yes! try

$$P(x|I) \propto x^{\beta}(1-x)^{\alpha}$$

Choose α and β to make any desired center and width.

$$P(x|\text{data}) = x^{N_B}(1-x)^{N-N_B} \times x^{\beta}(1-x)^{\alpha}$$

$$\int_0^1 P(x|\text{data}) = \int_0^1 x^{N_B+\beta}(1-x)^{N-N_B+\alpha}dx$$

$$= \frac{\Gamma(N_B+\beta+1)\Gamma(N-N_B+\alpha+1)}{\Gamma(N+\alpha+\beta+2)}$$

Priors that preserve the analytic form of p(x) are called “conjugate priors”. There is nothing special about them except mathematical convenience.

If you start with a conjugate prior, you’ll also be able to assimilate new data trivially, just by changing the parameters of your estimate. This is because every posterior is in the right analytic form to be the new prior!

By the way, if I show a special love of Bernoulli trials, it might be because I am an academic descendent of the Bernoulli brothers!

Actually, this is not a very exclusive club: Gauss and the Bernoullis each have ~50,000 recorded descendents in the Mathematics Genealogy database, and probably many times more unrecorded.

The probability of getting n events in N tries, each with i.i.d. probability p is

$$\mathrm{bin}(n, N, p) = \binom{N}{n} p^n (1-p)^{N-n}$$