Elementary parameter estimation

Prior 6.14 and formula 6.70

Suppose we draw N balls from a Bernoulli distribution of probability p (i.e. each ball is red with probability g and white with probability $1 - g$). The probability we get R reds is $\binom{N}{R} g^R (1-g)^{(N-R)}$ (this is the "binomial monkey prior" of section 6.7). Now, if our prior probability for p is uniform, then the probability of getting R reds is $\int_0^1 \binom{N}{R} g^R (1-g)^{(N-R)} dg$. Bayes 1763 paper tells us this is $\frac{1}{N+1}$ (see "Bayes' billiards", for example Story 8.3.2 in Blitzstein-Hwang, "Introduction to Probability"); in Jaynes' book this is formula 6.70. So indeed, 6.14 is an uniformed prior in this sense as well.

Remark: This is very different from "binomial monkey prior" because in this model data informs us about g, whereas in the "binomial monkey prior" g is fixed and can not be influenced by data.

Remark: This works in multicolor setting as well: Suppose N balls drawn from a large vat of balls in which there are balls of K colors in total, with unknown fractions $p_1, \ldots, p_K$ of balls of each color. Then the probability of getting N_1 balls of the first color, N_2 balls of the second color etc. - averaged over all possible tuples $p_1, \ldots, p_K$ - is always the same, no matter what the numbers N_i are. Since there are $\binom{N+K-1}{K-1}$ such tuples, each one has probability $\binom{N+K-1}{K-1}^{-1}$; for $K = 2$ this is $\frac{1}{N+1}$ as before.

6.15

No computation needed: with $p(R|NI)$ independent of R, the only term in 6.13 left dependent on R is $p(D|NRI_0)$, so $p(R|DNI) \sim p(D|NRI_0) \sim \binom{R}{r}\binom{N-R}{n-r}$.

Summation formula 6.16

To choose $n + 1$ balls from $N + 1$: first choose the number $R + 1$ of the $r + 1$st chosen ball; then choose r balls from the first R; and, finally, $n - r$ from the last $N - R$.

Remark: This also follows from the more obvious Vandermonde identity $\sum_r \binom{R}{r}\binom{N-R}{n-r} = \binom{N}{n}$ by "upper index negation", see here.

Most probable value 6.21

This is the same as derivation of formulas 3.26 and 3.27 in Chapter 3 (see notes for that chapter).

6.24

$$\langle R \rangle = \sum_{R=0}^{N} R \binom{N+1}{n+1}^{-1} \binom{R}{r} \binom{N-R}{n-r}$$

$$\langle R \rangle = \binom{N+1}{n+1}^{-1} \sum_{R=0}^{N} R \binom{R}{r} \binom{N-R}{n-r}$$

$$\langle R \rangle + \binom{N+1}{n+1}^{-1} \sum_{R=0}^{N} \binom{R}{r} \binom{N-R}{n-r} = \binom{N+1}{n+1}^{-1} \sum_{R=0}^{N} R \binom{R}{r} \binom{N-R}{n-r} + \binom{N+1}{n+1}^{-1} \sum_{R=0}^{N} \binom{R}{r} \binom{N-R}{n-r}$$

Due to (6.16) we have:

$$\langle R \rangle + \binom{N+1}{n+1}^{-1} \binom{N+1}{n+1} = \binom{N+1}{n+1}^{-1} \left\{ \sum_{R=0}^{N} R \binom{R}{r} \binom{N-R}{n-r} + \sum_{R=0}^{N} \binom{R}{r} \binom{N-R}{n-r} \right\}$$

Since $(R+1)\binom{R}{r} = R\binom{R}{r} + \binom{R}{r}$, we have:

$$\langle R \rangle + 1 = \binom{N+1}{n+1}^{-1} \left\{ \sum_{R=0}^{N} (R+1) \binom{R}{r} \binom{N-R}{n-r} \right\}$$

Due to (6.23) we have:

$$\langle R \rangle + 1 = \binom{N+1}{n+1}^{-1} \left\{ \sum_{R=0}^{N} (r+1) \binom{R+1}{r+1} \binom{N-R}{n-r} \right\}$$

$$\langle R \rangle + 1 = \binom{N+1}{n+1}^{-1} (r+1) \left\{ \sum_{R=0}^{N} \binom{R+1}{r+1} \binom{N-R}{n-r} \right\}$$

Set $R_q = R+1$ and $r_q = r+1$

$$\langle R \rangle + 1 = \binom{N+1}{n+1}^{-1} (r+1) \left\{ \sum_{R_q=1}^{N+1} \binom{R_q}{r_q} \binom{N-(R_q-1)}{n-(r_q-1)} \right\}$$

$$\langle R \rangle + 1 = \binom{N+1}{n+1}^{-1} (r+1) \left\{ \sum_{R_q=1}^{N+1} \binom{R_q}{r_q} \binom{N+1-R_q}{n+1-r_q} \right\}$$

Apply (6.19) again

$$\langle R \rangle + 1 = \binom{N+1}{n+1}^{-1} (r+1) \binom{N+2}{n+2}$$

6.29

$$p(R_{n+1}|DNI_0) = \sum_{R=0}^{N} \frac{R-r}{N-n} \binom{N+1}{n+1}^{-1} \binom{R}{r} \binom{N-R}{n-r}$$

$$p(R_{n+1}|DNI_0) = \frac{1}{N-n} \sum_{R=0}^{N} (R-r) \binom{N+1}{n+1}^{-1} \binom{R}{r} \binom{N-R}{n-r}$$

$$p(R_{n+1}|DNI_0) = \frac{1}{N-n} \sum_{R=0}^{N} \left\{ R \binom{N+1}{n+1}^{-1} \binom{R}{r} \binom{N-R}{n-r} - \sum_{R=0}^{N} r \binom{N+1}{n+1}^{-1} \binom{R}{r} \binom{N-R}{n-r} \right\}$$

$$p(R_{n+1}|DNI_0) = \frac{1}{N-n} \left\{ \sum_{R=0}^{N} R \binom{N+1}{n+1}^{-1} \binom{R}{r} \binom{N-R}{n-r} - \binom{N+1}{n+1}^{-1} r \sum_{R=0}^{N} \binom{R}{r} \binom{N-R}{n-r} \right\}$$

Due to 6.16, we have:

$$p(R_{n+1}|DNI_0) = \frac{1}{N-n} \left\{ \sum_{R=0}^{N} R \binom{N+1}{n+1}^{-1} \binom{R}{r} \binom{N-R}{n-r} - r \right\}$$

As already proven for 6.24, we get.

$$p(R_{n+1}|DNI_0) = \frac{1}{N-n} \frac{(N+2)(r+1)}{(n+2)} - 1 - r$$

$$p(R_{n+1}|DNI_0) = \frac{r+1}{n+2}$$

Laplace rule: 6.25, 6.29 and 6.73

$$E\left[\frac{R-r}{N-n}\right] = \frac{E[R+1] - (r+1)}{N-n} =$$

$$\frac{\frac{(N+2)(r+1)}{n+2} - (r+1)}{N-n} = \frac{r+1}{n+2}$$

In light of our remarks about the Prior 6.14 this **is** equivalent the fact that uniform is the Beta$(1,1)$, conjugate prior to Bernoulli with the two parameters $\alpha = 1$ and $\beta = 1$ equal to the number of prior imaginary successes and failures, which is another instance of Laplace rule of succession, formula 6.73

Formula 6.44

"Some calculation" is $\binom{N+1}{n+1} - \binom{N}{n} = \binom{N}{n+1}$, leading to the correct formula 6.44

$$p(R|r = 0, NI_1) = \binom{N}{n+1}^{-1} \binom{N-R}{n}$$

This is noted in the unofficial errata.

Some of Exercise 6.1

As in Section 6.5

$$P(R|r = n, NI_1) = S^{-1} \binom{R}{n}$$

for $R = 1, \ldots, N - 1$, and

$$S = \binom{N+1}{n+1} - \binom{0}{n} - \binom{N}{n} = \binom{N+1}{n+1} - \binom{N}{n} = \binom{N}{n+1}$$

$$p(R|r = n, DI_1) = \binom{N}{n+1}^{-1} \binom{R}{n}$$

for $R = 1, \ldots, N - 1$, identical with $p(R|r = n, DI_1) = \binom{N+1}{n+1}^{-1} \binom{R}{n}$ in that range, just renormailzed.

Expectation of R:

Using $(R + 1)\binom{R}{n} = (n + 1)\binom{R+1}{n+1}$ get

$$\sum_{R=1}^{N-1}(R + 1)\binom{R}{n} = \sum_{R=1}^{N-1}(n + 1)\binom{R+1}{n+1} = (n + 1)\binom{N+1}{n+2}$$

$$E[R + 1] = (n + 1)\binom{N}{n+1}^{-1}\binom{N+1}{n+2} = \frac{(N + 1)(n + 1)}{n + 2}$$

Expectation of R^2:

Using $(R + 2)(R + 1)\binom{R}{n} = (n + 2)(n + 1)\binom{R+2}{n+2}$ get

$$\sum_{R=1}^{N-1}(R+2)(R+1)\binom{R}{n} = \sum_{R=1}^{N-1}(n+2)(n+1)\binom{R+2}{n+2} = (n+2)(n+1)\binom{N+2}{n+3}$$

$$E[(R+2)(R+1)] = \frac{n+1}{n+3}(N+1)(N+2)$$

All the formulas for this case are the same as for uniform prior (with $r = n$) with N replaced by $N-1$.

So with $p = \frac{r+1}{n+2} = \frac{n+1}{n+2}$ as before, the mean is $m = n + (N-1-n)p$, variance $v = \frac{p(1-p)}{n+3}(N+1)(N-1-n)$.

Concave prior on R via improper β prior on p.

Suppose we still fill the urn by random tossing, but our prior over g is now non-uniform, but rather is given by (imporper, concave) prior

$$\text{Beta}(0,0)(g) \sim g^{-1}(1-g)^{-1}.$$

Then the probability that from N balls we have R red ones is proportional to

$$\int_0^1 \binom{N}{R} g^R(1-g)^{(N-R)} g^{-1}(1-g)^{-1} dg =$$

$$\binom{N}{R} \frac{1}{\binom{N-2}{R-1}(N-1)} \sim \frac{1}{R(N-R)}$$

which is the "concave prior" 6.49.

In general, if we start with $\text{Beta}(\alpha, \beta)$ and observe r reds on n draws, our posterior over the **remainder of the bin** is the same as the prior we would've gotten with $\text{Beta}(\alpha + r, \beta + n - r)$. This is why the formula 6.52 looks like 6.17 with N and n reduced by 2 and R and r reduced by 1. This also gives an alternative solution to Exercise 6.2 – under hypothesis $r \geq 1$, $n - r \geq 1$ we can update on one of the red and one of the white samples first, so that posterior distribution of R under concave prior and data of r reds on n draws is the same as the posterior distribution of $R-1$ from uniform prior on urn with $N-2$ balls with data of $r-1$ reds on $n-2$ draws.

Exercise 6.3

$$p(R|r = 0, NI_{00}) = \frac{A}{R(N-R)}\binom{N-R}{n}$$

This would still be infinite if $R = 0$, but is ok if we condition on/only consider $1 \leq R \leq N-1$. Then $A^{-1} = \sum_{R=1}^{N-1} \frac{1}{R(N-R)}\binom{N-R}{n}$.

Similarly

$$p(R|r = n, NI_{00}) = \frac{A}{R(N-R)}\binom{R}{n}$$

This would be infinite if $R = N$.

Now consider $n = 1$, $r = 0$ case. Then $p(R|r = 0, NI_{00}) = \frac{A}{R}$ with

$$A = \sum_{R=1}^{R=N-1} \frac{1}{R} \approx \ln N$$

(famously) and $E[R] \approx \frac{N}{\ln N}$, so expected fraction of red balls is approximately $\frac{1}{\ln N}$.

Exercise 6.4

$$P(R|DI)P(D|I) = P(D|RI)P(R|I).$$

Writing this for $R = R_1$, $R = R_2$ and multiplying we get

$$P(R_1|DI)P(R_2|DI)P(D|I)^2 = P(D|R_1I)P(D|R_2I)P(R_1|I)P(R_2|I)$$

Writing the same for $R = R_1R_2$ and using $P(R_1R_2|I) = P(R_1|I)P(R_2|I)$ we get

$$P(R_1R_2|DI)P(D|I) = P(D|R_1R_2I)P(R_1R_2|I) = P(D|R_1R_2I)P(R_1|I)P(R_2|I).$$

We see that $P(R_1R_2|DI) = P(R_1|DI)P(R_2|DI)$ is equivalent to

$$P(D|R_1R_2I)P(D|I) = P(D|R_1I)P(D|R_2I)$$

or

$$\frac{P(D|R_1R_2I)}{P(D|R_2I)} = \frac{P(D|R_1I)}{P(D|I)}.$$

Any further insight into the meaning of this condition would be appreciated.

Discussion of 6.9.1

TO BE EXTENDED.

This is related to Chapter 4, Exercise 4.4 – suppose I decide to keep getting samples until I accept hypothesis A. If A is false (B is true) I may still end up accepting A (in exercise 4.4 the probability of this is low, but that's because the evidence cutoff is chosen to be high).

Relevant: Anscombe, 1954 Fixed-Sample-Size Analysis of Sequential Observations, p.92, and the Anscombe-Armitage discussion.

Also relevant.

Formula 6.86

$$\sum_{n=c}^{\infty} \frac{n!}{c!(n-c)!} \phi^c (1-\phi)^{(n-c)} \frac{\exp\{-s\}s^n}{n!} =$$

$$\sum_{i=0}^{\infty} \frac{1}{c!i!} \phi^c (1-\phi)^i \exp\{-s\} s^i s^c = \frac{1}{c!} (\phi s)^c \exp\{-s\} \exp\{s(1-\phi)\} =$$

$$\frac{(s\phi)^c \exp\{-s\phi\}}{c!}$$

Formula 6.89

$$\frac{\text{Pois}_s(n)\text{Bin}_{n,\phi}(c)}{\text{Pois}_{s\phi}(c)} = \frac{\exp\{-(1-\phi)s\}s^n \phi^c (1-\phi)^{n-c}}{(s\phi)^c (n-c)!} = \text{Pois}_{s(1-\phi)}(n-c)$$

This can be interpreted as saying that in addition to c detected particles there were some undetected ones, distributed by poisson with rate $s(1-\phi)$.

Possible start of Exercise 6.5

As usual, the problem statement is ambiguous. One formalization is as follows: after a particle is registered, for time Δ no further registrations are possible, after which probability of registration for an incident particle reverts to (known) ϕ. We suppose that the counter has registered c particles in time $T = 1s$, with the emission still driven by Poisson process with rate s so that for any set of times B with total duration t we have $p_B(n) = \frac{\exp\{-st\}(st)^n}{n!}$. For more sophisticated versions and solutions see the referred paper of Takács and references therein.

Intuitively, we expect $p(n|\phi cs) = \text{Pois}_{s(1-\phi)(1-c\Delta)+sc\Delta}(n-c)$ (the "undetected" rate being $s(1-\phi)$ during the non-locked phases and s during the locked phases). Let's see if this holds up.

Bayes fhtagn: $p(n|\phi cs) = p(n|s)\frac{p(c|n\phi s)}{p(c|\phi s)}$

Here $p(n|s) = \text{Pois}_s(n)$ as before, and $p(c|\phi s)$ is as usual a normalization constant independent of n and the main point is to compute $p(c|n\phi s)$.

Conditional on n the Poisson arrival times $\vec{t} = (t_1, \ldots, t_n)$ are equivalent to being order statistics of iid uniform r.v.s., and are uniformly distributed over the simplex $0 \leq t_1, \ldots, t_n \ldots 1$. The problem suggests dividing this simplex into regions depending on which inequalities $t_i + \Delta > t_{i+s}$ hold, and obtain $p(c|n\phi\vec{t})$ for each region separately, and then marginalize over $\vec{t}$. This seems not entirely straightforward (the combinatorics is somewhat manageable, but unless one gets insight/miracle the marginalization would be tough). One could: a) consider Δ to be small, justifying ignoring cases where more than one of $t_i + \Delta < t_{i+s}$ hold and dividing only into two regions, plus possible further simplifications from ignoring terma that are higher-order in Δ b) do something clever, possibly withe MGFs, possibly borrowed from the counter literature (Feller, Takács and others) or c) abandon this line of attack.

We choose c) for now but hope to return to this at a later occasion.

Formula 6.108

$$\frac{1}{(1-x)^{a+1}} = \left(\sum_{i=0}^{\infty} x^i\right)^{a+1} = \sum_{m=0}^{\infty} c_m x^m$$

where c_m is the number of ways of partitioning m into $a+1$ ordered parts, which is done by selecting a dividers among $m+a$ places, and so $c_m = \binom{m+a}{a} = \binom{m+a}{m}$. Now $(x\frac{d}{dx})x^m = mx^m$, so $(x\frac{d}{dx}^n)x^m = m^n x^m$, hence all together

$$\left(x\frac{d}{dx}^n\right)\frac{1}{(1-x)^{a+1}} = \sum_{m=0}^{\infty} \binom{m+a}{m} m^n x^m$$

as wanted.

Exercise 6.6

$$S(N) = \sum_{n=N}^{\infty} p(c|\phi n) = \sum_{n=N}^{\infty} \binom{n}{c} \phi^c (1-\phi)^{n-c}$$

Let $1 - \phi = x$. Then the cth derivative $(x^n)^{(c)} = c!\binom{n}{c}x^{n-c}$ and

$$S(N) = \frac{\phi^c}{c!} \left(\frac{x^N}{1-x}\right)^{(c)}$$

This reproduces $S(0) = \frac{\phi^c}{c!}\frac{c!}{(1-x)^{c+1}} = \frac{1}{\phi}$ and generally gives, using the "higher derivative of product" rule $(fg)^{(c)} = \sum_{k=0}^{c} \binom{c}{k} f^{(c-k)} g^{(k)}$,

8

$$S(N) = \frac{\phi^c}{c!} \sum_{k=0}^{c} \binom{c}{k} (c-k)!(1-x)^{-(c-k+1)} \frac{N!}{(N-k)!} x^{N-k} =$$

$$\frac{1}{\phi} \sum_{k=0}^{c} \binom{N}{k} \phi^k (1-\phi)^{N-k}$$

This is a subsum of binomial distribution and one can apply for large N the Poisson approximation to binomial:

$$\binom{N}{k} \phi^k (1-\phi)^{N-k} \approx \frac{\exp\{-N\phi\}(N\phi)^k}{k!}$$

For large N either before or after approximating one sees that the term is with $k = c$ dominates all the other ones, so indeed

$$\frac{S(N)}{S(0)} = \sum_{k=0}^{c} \binom{N}{k} \phi^k (1-\phi)^{N-k} \approx \frac{\exp\{-N\phi\}(N\phi)^c}{c!}$$

as wanted.

Pluggin in $\phi = 0.1, c = 10, N = 270$ gives $\frac{S(N)}{S(0)} \approx 10^{-4}$, etc.

6.121, 6.122

$$p(n_1|I_J) = \int_0^s \frac{1}{s} \frac{s^{n_1} \exp\{-s\}}{n_1!} ds = \frac{1}{n_1} = \int_0^s \frac{s^{n_1-1} \exp\{-s\}}{(n_1-1)!} ds = \frac{1}{n_1}$$

With help of 6.108 we also compute:

$$p(c_1|I_J) = \sum_n p(c_1|n_1 I_J) p(n_1|I_J) = \sum_{n_1=c_1}^{\infty} \binom{n_1}{n_1 - c_1} \phi^{c_1} (1-\phi)^{n_1-c_1} \frac{1}{n_1}$$

$$= \frac{\phi^{c_1}}{c_1} \sum_{n_1=c_1}^{\infty} \binom{n_1-1}{c_1-1} (1-\phi)^{n_1-c_1} = \frac{\phi^{c_1}}{c_1} \sum_{m=0}^{\infty} \binom{m+c_1-1}{m} (1-\phi)^m =$$

$$= \frac{\phi^{c_1}}{c_1} \frac{1}{\phi^{c_1}} = \frac{1}{c_1}$$

And so

9

$$p(n_1|c_1\phi I_J) = p(n_1|I_J)\frac{p(c_1|n_1\phi I_J)}{p(c_1|\phi I_J)} = \frac{c_1}{n_1}p(c_1|n_1\phi I_J)$$

confirming 6.121.

Now

$$p(n_1|c_1\phi I_J) = \frac{c_1}{n_1}\binom{n_1}{n_1-c_1}\phi^{c_1}(1-\phi)^{n_1-c_1} = \binom{n-1}{c_1-1}\phi^{c_1}(1-\phi)^{n_1-c_1}$$

So $\hat{n}_J$ is found by Jaynes' favorite method of equating

$$\binom{n-2}{c_1-1}\phi^{c_1}(1-\phi)^{n_1-1-c_1} = \binom{n-1}{c_1-1}\phi^{c_1}(1-\phi)^{n_1-c_1}$$

$$(n-c_1+1) = (n-1)(1-\phi)$$

$$n\phi = \phi + c_1 - 1$$

$$\hat{n}_J = \frac{c_1 - 1 + \phi}{\phi}$$

And, finally, the expectation is (again, using 6.108)

$$\sum_{n_1=c_1}^{\infty} n_1 \frac{c_1}{n_1}\binom{n_1}{n_1-c_1}\phi^{c_1}(1-\phi)^{n_1-c_1} =$$

$$c_1\phi^{c_1}\sum_m \binom{m+c_1}{m}(1-\phi)^m = \frac{c_1}{\phi}$$

as wanted.

6.140

Formula 6.112 says $p(n_1|\phi c_1 I_A) = \binom{n_1}{c_1}\phi^{c_1+1}(1-\phi)^{n_1-c_1}$ and 6.116 says $E_A[n_1] = \frac{c_1+1-\phi}{\phi} = c_1 + \frac{(c_1+1)(1-\phi)}{\phi}$.

Continuing in same vein

$$E_A[(n_1 - c_1)(n_1 - c_1 - 1)] =$$

$$\phi^{c_1+1}(1-\phi)^2(c_1+1)(c_1+2)\sum_{n_1}\binom{n_1}{n_1-c_1-2}(1-\phi)^{n_1-c_1-2} =$$

$$\phi^{c_1+1}(1-\phi)^2(c_1+1)(c_1+2)\frac{1}{\phi^{c_1+3}} = (c_1+1)(c_1+2)\frac{(1-\phi)^2}{\phi^2}$$

$$E[(n-c_1)^2] = (c_1+1)(c_1+2)\frac{(1-\phi)^2}{\phi^2} + \frac{(c_1+1)(1-\phi)}{\phi}$$

$$\mathrm{var}(n_1) = \mathrm{var}(n_1-c_1) = E[(n-c_1)^2] - E[(n-c_1)]^2 =$$

$$(c_1+1)[\frac{(1-\phi)^2}{\phi^2} + \frac{(1-\phi)}{\phi}]$$

$$= (c_1+1)\frac{(1-\phi)}{\phi^2}$$

as wanted.

6.141

Formula 6.129 says

$$p(n_1|\phi c_2 c_1 I_B) = \binom{n_1+c_2}{c_1+c_2}\left(\frac{2\phi}{1+\phi}\right)^{c_1+c_2+1}\left(\frac{1-\phi}{1+\phi}\right)^{n_1-c_1}$$

and 6.133 says $E_B[n_1] = \frac{c_1+c_2+1-\phi}{2\phi} + \frac{c_1-c_2}{2} = c_1 + (c_1+c_2+1)\left(\frac{1-\phi}{2\phi}\right)$

$$E[(n_1-c_1)(n_1-c_1-1)] =$$

$$\left(\frac{2\phi}{1+\phi}\right)^{c_1+c_2+1}\left(\frac{1-\phi}{1+\phi}\right)^2(c_1+c_2+1)(c_1+c_2+2)\sum_{m=n_1-c_1-2=0}^{\infty}\binom{m+(c_1+c_2+2)}{m}\left(\frac{1-\phi}{1+\phi}\right)^m$$

$$\left(\frac{2\phi}{1+\phi}\right)^{c_1+c_2+1}\left(\frac{1-\phi}{1+\phi}\right)^2(c_1+c_2+1)(c_1+c_2+2)\left(\frac{2\phi}{1+\phi}\right)^{-(c_1+c_2+3)}$$

$$= (c_1+c_2+1)(c_1+c_2+2)\left(\frac{1-\phi}{2\phi}\right)^2$$

11

$$E[(n_1 - c_1)^2] = (c_1 + c_2 + 1)(c_1 + c_2 + 2)\left(\frac{1-\phi}{2\phi}\right)^2 + (c_1 + c_2 + 1)\left(\frac{1-\phi}{2\phi}\right)$$

$$\text{var}(n_1) = \text{var}(n_1 - c_1) = E[(n_1 - c_1)^2] - E[(n_1 - c_1)]^2$$

$$(c_1 + c_2 + 1)[\left(\frac{1-\phi}{2\phi}\right)^2 + \left(\frac{1-\phi}{2\phi}\right)] = (c_1 + c_2 + 1)\frac{1-\phi^2}{4\phi^2}$$

as wanted.

6.142

We have from 6.121 $p(n_1|c_1\phi I_J) = \frac{c_1}{n_1}\binom{n}{c_1}\phi^{c_1}(1-\phi)^{n_1-c_1} = \binom{n-1}{c_1-1}\phi^{c_1}(1-\phi)^{n_1-c_1}$
and from 6.122 also $E_J[n_1] = \frac{c_1}{\phi}$

We get

$$E[n(n+1)] = c_1(c_1+1)\phi^{c_1}\sum_{n_1=c_1}^{\infty}\binom{n_1+1}{c_1+1}(1-\phi)^{n_1-c_1} =$$

$$c_1(c_1+1)\phi^{c_1}\sum_{m=n_1-c_1=0}^{\infty}\binom{m+c_1+1}{m}(1-\phi)^m = \frac{c_1(c_1+1)}{\phi^2}$$

so

$$\text{var}(n_1) = \frac{c_1}{\phi^2} - \frac{c_1}{\phi} = \frac{c_1(1-\phi)}{\phi^2}$$

as wanted.

Exercise 6.9

Formula 6.164 says

$$p(\alpha|DI) = \frac{(n-1)\alpha^{-n}}{\alpha_0^{1-n} - \alpha_1^{1-n}}$$

for $\alpha_0 \leq \alpha \leq \alpha_1$, $n > 1$.

As $\alpha_1 \to \infty$ this converges to $p(\alpha|DI) = \frac{(n-1)\alpha^{-n}}{\alpha_0^{1-n}}$ on $\alpha_0 \leq \alpha$.

For $n = 1$ 6.165 gives

$$p(\alpha|DI) = \frac{\alpha^{-1}}{\log \alpha_1 - \log \alpha_0}$$

For $n = 0$ of course one is back to the improper uniform distribution on $[\alpha_0, \infty)$.

In all cases the posterior is proportional to α^{-n} on $[\alpha_0, \infty)$, and has $n - 2$ first moments finite. The cases of 0 or 1 observations are most troublesome - no median or quartiles, and certainly no finite mean or standard deviation. The cases of two observations has well-defined median and quartiles, but infinite expectation. Three observations give finite mean, but infinite variance. C'est la vie.

Exercise 6.8

With one observation above the posterior is proportional to α^{-1} on $[\alpha_0, \alpha_1]$. If $n = 1$ the CDF is $F(\alpha) = \frac{\ln \alpha - \ln \alpha_0}{\ln \alpha_1 - \ln \alpha_0}$, so the median m is the geometric mean of α_0 and α_1, and the interquartile range is between geometric mean of m and α_0 and of m and α_1. If $\alpha_1 \approx 53333 = 3^{-1}2^4 10^4$ and $\alpha_0 = 27$ the estimate is $m = 1200$ and quartiles are 180 and 8000.

The mean, meanwhile is $\frac{\alpha_1 - 27}{\ln \alpha_1 - \ln 27} \approx 7025$, variance $\frac{\alpha_1^2 - 27^2}{2(\ln \alpha_1 - \ln 27)} - E[\alpha]^2 \approx$ 138071335, standard deviation 11750. Since the distribution is very overweighted on the left, the mean is far to the right of the median (the torgue due to far-flung right part of the distribution is high, making fulcrum/center of mass/mean move to the right of the "half-mass line"/median to balance.)

Both of these are sensitive to the α_1, but, at least the median, in a power one-half way, so a hundredford change in maximum α_1 gives a tenfold change in posterior median, reasonably consistent with weakness of data.

If $n = 2$ or, equivalently, one uses Jeffrey's prior the posterior is proportional to $\frac{1}{\alpha^2}$. In terms of density and median/quartiles this is insensitive to α_1 and for large α_1 the CDF is $F(\alpha) = 27(\frac{1}{27} - \frac{1}{\alpha})$, so $m = 54$, and quartiles are 36 and 108, a substantially narrower range.

The mean is, for lagre α_1, about $27 \ln(\alpha_1/27)$. In our $\alpha_1 = 53333$ case this is about 205, with $E(\alpha^2) \approx 27 \cdot 53333 = 1440000$ and $\text{var}(\alpha) \approx 1440000$, standard deviation 1200. The standard deviation is reduced from order of magnitude of $\alpha_1 / \ln \alpha_1$ to order of magnitude $\alpha_1^{1/2}$, but is still much larger than interquartile range which is now independent of α_1.

With higher n both mean and variance become insensitive to α_1.