

Typos from the 'Og

Christian P. Robert

August 19, 2011

The Bayesian Choice

1. (*Thanks to Bastien Boussau, Berkeley*) In Example 1.1.3 (or Example 1.3), on page 3, I consider an hypergeometric $\mathcal{H}(30, N, 20/N)$ distribution, while in Appendix A, I denote hypergeometric distributions as $\mathcal{H}(N; n; p)$, inverting the role of the population size and of the sample size.
2. (*Thanks to Thong Pham, Ritsumeikan University, Japan*) In Example 1.3.2 (or Example 1.11), the term σ^n in the expression at the top of page 15 should be σ^{n-1} .
3. (*Thanks to Cristiano Passerini, Pontecchio Marconi, Italy*). In Example 4.3.3 (or Example 4.19), page 184, further involves a typo related with the hypergeometric distribution $\mathcal{H}(N; n; p)$. The ratio should be

$$\frac{\binom{n_1}{n_{11}} \binom{n-n_1}{n_2-n_{11}} / \binom{n}{n_2} \pi(N=n)}{\sum_{k=36}^{50} \binom{n_1}{n_{11}} \binom{k-n_1}{n_2-n_{11}} / \binom{k}{n_2} \pi(N=k)}.$$

4. (*Thanks to Peng Yu*) On pages 344-345, (7.1.1) should designate the general setting of a collection of models, i.e. the last formula of page 344, rather than the mixture example.
5. (*Thanks to Benjamin Remy Holcblat*) In formula (10.4.3), on page 480, (formula (10.15), page 493 in the paperback), the denominator should be $p_{x_{n+1}}(x_1, \dots, x_n)$ instead of $p_{x_{n+1}}(x_1, \dots, x_n) + 1$.
6. (*Thanks to Anthony Lee*) In Definition 10.2.1, formula (10.2.1), the last integrand should be θ_n , not θ_{n+1} . (*I somehow thought this had been corrected already!*)
7. (*Thanks to Stefan Webb*) The density

$$f(x|p) = \frac{\binom{pn}{x} \binom{(1-p)N}{n-x}}{\binom{N}{n}} \mathbb{I}_{\{n-(1-p)N, \dots, pN\}}(x) \mathbb{I}_{\{0, 1, \dots, n\}}(x).$$

should be

$$f(x|p) = \frac{\binom{pN}{x} \binom{(1-p)N}{n-x}}{\binom{N}{n}} \mathbb{I}_{\{n-(1-p)N, \dots, pN\}}(x) \mathbb{I}_{\{0, 1, \dots, n\}}(x).$$

8. (*Thanks to Peng Yu*) On page 578, the reference West (1992) is a phantom reference in that it is not quoted in the book.

Bayesian Core

1. (*Thanks to Bo Jacoby*) On page 70, the summation starting with $i = 1$ should start with $c = 1$.
2. (*Thanks to Francisco Garcìa, Santiago, Chile*). In Algorithm 6.5, page 169, the last item of the algorithm should be numbered $2p$ not $2p - 1$. In addition, the last ratio in the acceptance probability should be

$$\frac{\pi(x_{2p-1}^{(t)})}{\pi_{\alpha_1}(x_{2p-1}^{(t)})}$$

the power used on page 169 being a special case of indexing in the tempering sequence. Tempering requires an even number of steps to cancel the normalising constants of the tempered targets.

3. (*Thanks to Reza Seirafi, Virginia Tech*) On page 178, there is a typo in the reversible jump acceptance probability for the mixture model, in the formula for the acceptance probability of the split move. It is missing the density of the auxiliary parameters u_1 , u_2 , and u_3 , not necessarily equal to 1. It should be

$$\min \left(\frac{\tilde{\pi}_{(k+1)k}}{\tilde{\pi}_{k(k+1)}} \frac{\varrho(k+1)}{\varrho(k)} \frac{\pi_{k+1}(\theta_{k+1})\ell_{k+1}(\theta_{k+1})}{p_i(u_2)\pi_k(\theta_k)\ell_k(\theta_k)} \frac{p_{jk}}{(1-u_1)^2} \sigma_{jk}^2, 1 \right)$$

since u_2 is distributed from a non-uniform density. The physical reason for this typo is that I cut-and-pasted the L^AT_EX code from *Monte Carlo Statistical Methods* (page 439) where u_2 is a uniform variate! There also is a minor typo a few lines above: when defining

$$\mu_{(j+1)(k+1)} = \mu_{jk} - \frac{p_{j(k+1)}u_2}{p_{jk} - p_{j(k+1)}},$$

it should be

$$\mu_{(j+1)(k+1)} = \mu_{jk} - \frac{p_{j(k+1)}u_2}{p_{jk} - p_{j(k+1)}}.$$

Obviously, this could also affect the associated R code but I checked it and found the line

```
jacob=(kprop-2)*log(1-propp[kprop]) - singleprior(propmix,kprop)
```

which I think is correct for the normal proposal.

Monte Carlo Statistical Methods

1. (*Thanks to Amir Alipour*) On page 4, line 9, in $(\theta_1, \dots, \theta_n, p_1, \dots, p_n)$, the index should be k instead of n .

2. (*Thanks to Amir Alipour*) On page 4, Example 1.3, in the last line, $n > q$, should be $n \geq q$.
3. (*Thanks to Amir Alipour*) On page 5, $\sigma^{-(n+q)}$ should be $\sigma^{-(n+q+1)}$.
4. (*Thanks to Amir Alipour*) On page 8, in formula (1.10), the gradient symbol ∇ is used with no definition, which is only found on page 19.
5. (*Thanks to Amir Alipour*) On page 8, Example 1.6, the cumulant function should involve $\log(-1/(2\theta_2))$
6. (*Thanks to Amir Alipour*) On page 10, in the definition of the modified Bessel function, the series in z should be in t .
7. (*Thanks to Amir Alipour*) On page 10, the last formula of Example 1.9, the likelihood function is inversely proportional to a power of the displayed product and the power of σ should be $2n/(p+1)$.
8. (*Thanks to Ping Yu*) On page 31, the reference Robert (1994a) should be Robert (2001), referring to *The Bayesian Choice*.
9. (*Thanks to Gholamhossein Gholami*) On page 37, Example 2.2, $X_n \in [0, 1]$. Then $2X_n - 1 \in [-1, 1]$ should have the same behaviour as a sequence of random numbers distributed according to the arcsine distribution not X_n itself. Furthermore, in Figure 2.1, the function $y_n = F(x_n)$ should be defined as the arcsine transform.
10. (*Thanks to Gholamhossein Gholami*) On page 52, Example 2.19, line 26, the power of $1 - b$ should be $a - \alpha$, and we are looking at the minimum of M in b , not the maximum.
11. (*Thanks to Gholamhossein Gholami*) On page 53, Example 2.20, line 10, the ratio should be $f/g_\alpha(z) = e^{\alpha(z-\mu)} e^{-z^2/2}$.
12. On page 53, Example 2.20, line 15, Eqn. (2.11) should be

$$\alpha^* = \frac{\mu + \sqrt{\mu^2 + 4}}{2}.$$

13. (*Thanks to Edward Kao, University of Houston*) In Problem 3.21, page 114, $\mathcal{G}a(y, 1)$ should be $\mathcal{G}a(1, y)$.
14. (*Thanks to Edward Kao, University of Houston*) On page 149, Problem 4.5, an h is missing in the expectation of the last line of question (a). And in question (c), $Bin(y)$ should be $Bin(n, y)$
15. (*Thanks to Doug Rivers, Stanford University, and to Liaosa Xu, Virginia Tech.*) On page 175, in Example 5.14, the last displayed equation should be

$$L(\theta|\mathbf{y}) = \int L^p(\theta|\mathbf{y}) f(\mathbf{z}|\mathbf{y}, \theta) d\mathbf{z}$$

where $L^P(\theta|\mathbf{y})$ is the part of the likelihood only involving y , instead of

$$L(\theta|y) = \mathbb{E}[L^c(\theta|y, Z)].$$

The expectation, as written is incorrect: the z_i 's are truncated in a. Their density is therefore a renormalised version of $f(z - \theta)$. The whole example should be rewritten because it starts as if m observations were uncensored out of n , only to move to a fixed censoring bound a . While both likelihoods are proportional when $a = y_m$, this confusion should not be present.

16. (*Thanks to Edward Kao, University of Houston*) Example 5.18 on page 179 studies a missing data phone plan model and its EM resolution. First, the customers in area i should be double-indexed, i.e.

$$Z_{ij} \sim \mathcal{M}(1, (p_1, \dots, p_5))$$

which implies in turn that

$$T_i = \sum_{j=1}^{n_j} Z_{ij}.$$

Then the summary \mathbf{T} should be defined as

$$\mathbf{T} = (T_1, T_2, \dots, T_n)$$

and W_5 as

$$W_5 = \sum_{i=m+1}^n T_{i5},$$

given that the first m customers have the fifth plan missing.

17. (*Thanks to Edward Kao, University of Houston*) On page 211, the kernel is called P instead of K twice, including in Definition 6.8.
18. (*Thanks to Tor Andre Myrvoll*) On page 211, Lemma 6.7, *through some y on the n th step* should be *through some y on the n th step* to be coherent with the above decomposition (even though, technically, it is not wrong).
19. (*Thanks to Edward Kao, University of Houston*) On page 212, the inequality in $P(\tau_2 > \tau_3)$ should be inverted in $P(\tau_2 < \tau_3)$.
20. (*Thanks to Edward Kao, University of Houston*) On page 213, in Example 6.14, the state size K should be M .
21. (*Thanks to Edward Kao, University of Houston*) On page 215, Example 6.20, second displayed equation, $\theta x_n \bar{w} \sigma^{-2}$ should be $2\theta x_n \bar{w}$

22. (Thanks to Edward Kao, University of Houston) On page 250, Problem 6.22 (d) should be phrased as “show that the invariant distribution is also the limiting distribution of the chain” instead of the obvious “show that the invariant distribution is also the stationary distribution of the chain”. And the solution to (c) given in the solution manual has a mistake (see Kao’s own solution in his *Introduction to Stochastic Processes*, Example 4.3.1).
23. (Thanks to Edward Kao, University of Houston) On page 250, Problem 6.24 (a) should ask to show aperiodicity, not periodicity. And in part (b) the transition in 0 should be defined in the same way as in Problem 6.22.
24. (Thanks to Gholamhossein Gholami) On page 583, in the Negative Binomial distribution, $n + x + 1$ should be $n + x - 1$

Introducing Monte Carlo Methods with R

1. (Thanks to Jerry Sin) On page 11, matrix summation in the matrix commands of Figure 1.2 should be matrix multiplication.
2. (Thanks to Liaosa Xu from Virginia Tech) On page 20, when we mention the uniform over the set

$$\left\{ (a, b) : y_i(a + bx_i) > \log \frac{u_i}{1 - u_i} \right\}$$

this set is missing (a) an intersection sign before the curly bracket and (b) a $(-1)^{y_i}$ instead of the y_i . It should be

$$\bigcap_{i=1}^n \left\{ (a, b) : (-1)^{y_i} (a + bx_i) > \log \frac{u_i}{1 - u_i} \right\}$$

3. (Thanks to Matthieu Gomez) In formula (2.1), the transform for the gamma $\mathcal{G}(\alpha, \beta)$ distribution assumes β is a scale parameter, while the remainder of the book takes the opposite convention.
4. In Exercise 2.17, page 58, question d. should be *d. Show that the maximum of $b^{-a}(1 - b)^{a-\alpha}$ is attained at $b = a/\alpha$.*
5. In Exercise 2.21, page 59, in item (ii), $\|\theta\|$ should be replaced by $\sqrt{\lambda}$ and question b. should be removed.
6. On page 71, due to the late inclusion of an extra-exercise in the book, the *above exercise* in Exercise 3.5 actually means Exercise 3.3.
7. (Thanks to Brad McNeney, Simon Fraser University) The end of Example 3.6 (page 75) is missing a marginal estimate, i.e. there is a $x(1 - x)$ missing from $m(x)$. It should have been obvious from the estimates we derived, 19 and 16, which do not even appear on the support of the posterior distribution represented on Figure 3.5. The R code is given as

```

> mean(y[,1]*apply(y,1,f)/den)/mean(apply(y,1,h)/den)
[1] 19.33745
> mean(y[,2]*apply(y,1,f)/den)/mean(apply(y,1,h)/den)
[1] 16.54468

```

and should have been view source

```

> mean(y[,1]*f(y)/den)/mean(f(y)/den)
[1] 94.08314
> mean(y[,2]*f(y)/den)/mean(f(y)/den)
[1] 80.42832

```

A similar modification applies to the remark after eqn. (3.7) (page 76):

```
mean(apply(y,1,h)/den)
```

should be

```
mean(f(y)/den)
```

8. In Exercise 3.11, page 86, question c, a line got commented by mistake in the \LaTeX file and it should read *Explore the gain in efficiency from this method. Take $a = 4.5$ in part (a) and run an experiment to determine how many normal $\mathcal{N}(0, 1)$ random variables would be needed to calculate $P(Z > 4.5)$ to the same accuracy obtained from using 100 random variables in this importance sampler.*
9. (Thanks to Edward Kao, University of Houston) In Exercise 3.17, page 88, question b, it should be $X|Y \sim \mathcal{G}(1, y)$, not $X|Y \sim \mathcal{G}(y, 1)$.
10. (Thanks to Ashley) On page 128, running

```

> for (i in 1:(nlm(like,sta)$it)){
+ mmu=rbind(mmu,nlm(like,sta,iter=i)$est)}

```

produces the error message

```
Error in f(x, ?) : unused argument(s) (iter = 1)
```

as the latest versions of R, e.g. 2.12.1, calling nlm requires using `iterlim` instead of the abbreviation `iterr`:

```

> args(nlm)
function (f, p, ..., hessian = FALSE, tysize = rep(1, length(p)),
  fscale = 1, print.level = 0, ndigit = 12, gradtol = 1e-06,
  stepmax = max(1000 * sqrt(sum((p/tysize)^2)), 1000), steptol = 1e-06,
  iterlim = 100, check.analyticals = TRUE)

```

In addition, the function `nlm` produces the minimum of the first argument `f` and `like` should thus be defined as

```
> like=function(mu){
+   -sum(log(.25*dnorm(da-mu[1])+.75*dnorm(da-mu[2])))}
```

to end up with local maxima as on Figure 5.2.

11. (*Thanks to Ashley*) Another mistake on page 128, namely that we used

```
> da=rbind(rnorm(10^2),2.5+rnorm(3*10^2))
```

instead of

```
> da=c(rnorm(10^2),2.5+rnorm(3*10^2))
```

to create a sample. Since both two normal samples have different sizes, `rbind` induces a duplication of the smaller sample, not what we intended!

12. In Exercise 4.5, page 101, the \bar{X}_k should not be in bold fonts.
 13. In Exercise 4.9, page 108, I commented too many lines when revising and thus the variance terms vanished. It should read

$$\begin{aligned}\mathbb{E}[\exp -X^2|y] &= \frac{1}{\sqrt{2\pi\sigma^2/y}} \int \exp\{-x^2\} \exp\{-(x-\mu)^2 y/2\sigma^2\} dx \\ &= \frac{1}{\sqrt{2\sigma^2/y+1}} \exp\left\{-\frac{\mu^2}{1+2\sigma^2/y}\right\}\end{aligned}$$

14. In Exercise 4.13, page 122, following the removal of one exercise, Exercise 4.2 should read Exercise 4.1
 15. In Exercise 4.15, page 122, $Bin(y)$ should be $Bin(n, y)$ (as in Problem 4.5 of Monte Carlo Statistical Methods)
 16. (*Thanks to Ashley*) In Example 5.2, page 128, to simulate a sample of 400 observations from the mixture, we use

```
> da=rbind(rnorm(10^2),2.5+rnorm(3*10^2))
```

and need to replace `rbind` with `+c+`:

```
> da=sample(c(rnorm(10^2),2.5+rnorm(3*10^2)))
```

Note that `da` is not a sample from a mixture, because the proportion from each component is fixed.

17. On page 131, Exercise 5.3 has no simple encompassing set and the constraint should be replaced by

$$x^2(1 + \sin(y/3) \cos(8x)) + y^2(2 + \cos(5x) \cos(8y)) \leq 1$$

18. (*Thanks to Ashley*) In Example 5.5, page 134, in the R code

```
+ apply(z,1,mean)}
```

should be

```
+ apply(z,1,prod)}
```

as we are not working with log-densities.

19. (cont'ed) Example 5.5 (pages 133–135) is completely off-the-mark! Not only the representation of the likelihood should have used `prod` instead of `mean`, not only the constant should call the `val` argument of `integrate`, not only `integrate` uses `lower` and `upper` rather than `from` and `to`, but also the approximation is missing a scale factor of 10, squared root of the sample size. The corrected R code is thus

```
> cau=rcauchy(10^2)
> mcau=median(cau)
> rcau=diff(quantile(cau,c(.25,.75)))/sqrt(10^2)
> f=function(x){
+   z=dcauchy(outer(x,cau,FUN="-"))
+   apply(z,1,prod)}
> fcst=integrate(f,lower=-20,upper=20)$val
> ft=function(x){f(x)/fcst}
> g=function(x){dt((x-mcau)/rcau,df=49)/rcau}
> curve(ft,from=-1,to=1,xlab="",ylab="",lwd=2)
> curve(g,add=T,lty=2,col="steelblue",lwd=2)
```

and Figure 5.5 (page 134) is therefore modified. Note that the fit by the *tt* distribution is not as perfect as before. A normal approximation would do better.

20. (*Thanks to Gilles Guillot, Technical University of Denmark*) On page 137, second equation from bottom

$$h(\theta + \beta\zeta) - h(\theta + \beta\hat{\zeta})$$

should be

$$h(\theta + \beta\zeta) - h(\theta - \beta\hat{\zeta})$$

21. (Thanks to Gilles Guillot, Technical University of Denmark) On page 138, Example 5.7, the denominator in the gradient should be `2*beta` [the error actually occurs twice. And once again in the R code.] In addition, the first paragraph of page 138 is missing details: the conditions on α and β are necessary and sufficient. Furthermore, `demo(Chapter.5)` triggers an error message [true, the shortcut `max=TRUE` instead of `maximise=TRUE` in `optimise` does not work with the latest versions of R] This is what occurs with version 2.11.1:

```
demo(Chapter.5)
-----

Type <Return> to start :

> # Section 5.1, Numerical approximation
>
> ref=rcauchy(5001)

> f=function(y){-sum(log(1+(x-y)^2))}

> mi=NULL

> for (i in 1:400){
+   x=ref[1:i]
+   aut=optimise(f,interval=c(-10,10),max=T)
+   mi=c(mi,aut$max)
+ }
Error in f(arg, ?) : unused argument(s) (max = TRUE)

> optimise(f,interval=c(-10,10),maximum=T)
$maximum
[1] -2.571893

$objective
[1] -6.661338e-15
```

22. (Thanks to Robin Ryder, CREST) In Example 5.14, page 153, the sentence between parentheses should end up with *equal to 0*. And *step* should be replaced with *iteration* in the first paragraph.
23. (Thanks to Robin Ryder, CREST) On page 156, in Example 5.15, *likelihood surface* should be *log-likelihood surface*;
24. On page 158, Example 5.16 has a typo in that the EM sequence should be

$$\hat{\theta}_1 = \left\{ \frac{\theta_0 x_1}{2 + \theta_0} + x_4 \right\} / \left\{ \frac{\theta_0 x_1}{2 + \theta_0} + x_2 + x_3 + x_4 \right\}$$

And *or, equivalently, by* should be *or, equivalently, with*;

25. On page 162, in the caption of Figure 5.14, *LE estimator* should be *MLE*;
26. In Exercise 5.15, page 163, the *Z*'s in the formula should be in capital letters, namely

$$P(Z_i = 1) = 1 - P(Z_i = 2)$$

27. (*Thanks to Pierre Jacob and Robin Ryder, CREST*) In Exercise 5.17, page 164, the function ϕ should be written φ to be coherent with the notation of Example 5.14
28. Exercise 5.21, page 165, should have been removed as it duplicates Exercise 5.11 (but this redundancy is not going to confuse anyone!)
29. (*Thanks to Boris Vasiliev, Ottawa*) In Exercise 6.4, page 176, the second parameter of the gamma distribution $\mathcal{G}(a, b)$ is a rate parameter and not a scale parameter.
30. (*Thanks to Pierre Jacob and Robin Ryder, CREST*) In Exercise 6.13, page 197, both α and β must use a double exponential proposal in the Metropolis-Hastings algorithm of question b.
31. (*Thanks to Pierre Jacob and Robin Ryder, CREST*) In Exercise 6.15, page 197, the $\mathcal{L}(0, \omega)$ distribution should be a $\mathcal{N}(0, \omega)$ normal distribution,
32. In Example 7.3, page 204, part of the code is wrong: it should be

```
> sigma2=theta=rep(0,Nsim)                #init arrays
> sigma2[1]=1/rgamma(1,shape=a,rate=b)     #init chains
> B=sigma2[1]/(sigma2[1]+n*tau2)
> theta[1]=rnorm(1,m=B*theta0+(1-B)*xbar,sd=sqrt(tau2*B))
```

instead of

```
> sigma=theta=rep(0,Nsim)                 #init arrays
> sigma{1}=1/rgamma(1,shape=a,rate=b)     #init chains
> B=sigma2{1}/(sigma2{1}+n*tau2)
> theta{1}=rnorm(1,m=B*theta0+(1-B)*xbar,sd=sqrt(tau2*B))
```

33. In Algorithm 8, page 206, there are two commas before *given*;
34. in Example 7.6, page 210, I forgot to include the truncation probability $\Phi(a - \theta)^{n-m}$ in the likelihood and the notations are not completely coherent with Example 5.13 and 5.14 in that the *x*'s became *y*'s.
35. In the caption of Figure 7.6, page 2114, these are the allele probabilities, not the genotype probabilities;
36. in Exercise 7.21, page 235, `rtnorm` is missing `sigma` as one of its arguments.

37. In Exercise 7.22 c, page 235, the matrix is positive definite if and only if the condition is satisfied;
38. Exercise 7.23, page 235, has nothing wrong per se but it is rather a formal (mathematical) exercise.
39. in Exercise 7.25, page 235, the $x + a$ in question a should be $(x + a)$ to avoid any confusion.
40. On page 239, *the slower chain* should be *the slowest chain*;
41. On page 241, *you can load all coda functions* instead of *download*;
42. In the caption of Figure 8.2, page 244, the upper quantile is a 97.5
43. (*Thanks to Thomas Clerc, Fribourg*) On page 247, the complex number ι should be defined as the squared root of -1 instead of 1 !
44. In Exercise 8.1, page 241, we compare an estimator δ_1 with δ_k , not δ_2 . We need a bit more stability in the Markov chain to ensure that it has a finite variance. The assumption that both first moments are finite is not enough. Thomas Clerc from Fribourg also pointed out that the lazy R programming of the form `beta=c(beta,betaprop)` should not be encouraged!
45. In Example 8.9, page 264, $Y_k \sim \mathcal{N}(\theta_i, \sigma^2)$ should be $Y_i \sim \mathcal{N}(\theta_i, \sigma^2)$ and the θ_i 's are normal, not the μ_i 's.
46. In Exercise 8.7, page 265, it is the distribution on α that is closed-form, not the one on μ .
47. In Exercise 8.11, page 268, the y_i at the end of page 267 should be d_i ;
48. In Exercise 8.16, page 268, jpeg is mistakenly qualified as open.