Rejection sampling and RC-dating

You may want to compare timings: if so use function `system.time`.

This exercise illustrates rejection sampling. Ideas of shrinkage and Bayesian hypothesis tests come into it also.

Specimen $i = 1, 2, ..., K$ has a true age $\theta_i$ measured in years BP. Let $\theta = (\theta_1, ..., \theta_K)$.

Radiocarbon measurements $Y_i \sim N(r(\theta_i), \sigma_i^2)$ are assumed independent normal with laboratory-supplied standard deviation $\sigma_i$. This model or something very close is widely used in the literature. The non-linear function $r(t)$ is called the radiocarbon calibration curve. It is a non-linear function of its argument, and the likelihood often has several maxima, depending on the data $Y$.

```
#7 radiocarbon measurements on charcoal fragments from
#the SM/C Dune sequence at Shag River mouth (Anderson 1996)

#uncalibrated radiocarbon dates
y=c(580,600,537,670,646,630,660)
K=length(y)
#measurement errors
s=c(47,50,44,47,47,35,46)
plot(1:K,y,pch=16,ylim=c(400,800),xlab='measurement index',ylab='uncalibrated RC date')
arrows(1:K, y-s, 1:K, y+s, code=3, angle=90, length=0.1)

#In order to convert uncalibrated RC dates to calendar years BP
#(before the present, where present=1950) we use a calibration curve.
#This is effectively a function mapping calendar years to RC-years.
#It goes back to around 50000 years BP.
mu=read.table("intcal09.14c.txt",header=T,sep="",skip=10)

head(mu)
plot(mu$CAL.BP,mu$X14C.age,xlab='calendar years BP',ylab='RC measurement')
tail(mu)

#we will need only the most recent part of the calibration curve mapping
MAX.AGE=1000
i=which(mu$CAL.BP<=MAX.AGE)
mu=mu[i,]
head(mu)
n=dim(mu)[1]

#We make a "function" rc[y] that maps 'year' BP to RC year
#measurements by interpolating the calibration curve at annual
#intervals - this is convenient and quite fine enough given
#the measurement errors involved. The interpolation is linear.
#In the literature splines are commonly used.
rc=approx(x=mu$CAL.BP,y=mu$X14C.age,xout=1:MAX.AGE)$y
```
#the observation model for dating a specimen with true age theta
#is y ~ N(rc(theta),s^2) with s assumed known (reported by lab with y).

\[ t_{sq} = 2s^2 \]

\[
\log.lkd<-function(th) {
    \#Log-likelihood for vector 'th' of calendar dates BP
    -sum((rc[round(th)]-y)^2/tsq)
}\]

The final command gives a plot illustrating the likelihood of a single date.

We will carry out a Bayesian analysis in order to reconstruct the unknown true ages, \( \theta \). We will simulate these posterior distributions using rejection sampling. In general for Bayesian inference

\[
p(\theta | y) = \frac{L(\theta; y)p(\theta)}{m(y)}
\]

with \( p(\theta | y) \) the posterior, \( L(\theta; y) \) the likelihood, \( p(\theta) \) the prior and \( m(y) \) a normalizing constant which corresponds to the marginal likelihood. Let

\[
\tilde{p}(\theta | y) = \exp(-\sum_i (r(\theta_i) - y_i)^2/2\sigma_i^2)p(\theta)
\]

denote the un-normalised posterior.

Recall the rejection algorithm for sampling a density \( f(\theta) \): given a proposal density \( q(\theta) \) and a bound \( M \geq f(\theta)/q(\theta) \), simulate \( \theta \sim q \) and accept \( \theta \) with probability \( f(\theta)/Mq(\theta) \), otherwise try again. In our case the target \( f(\theta) \) is the un-normalized posterior \( \tilde{p}(\theta | y) \). If we take the prior \( p \) as our proposal density \( q \), then \( f/q = \tilde{p}(\theta | y)/p(\theta) \) is equal \( \exp(...) \) (since the prior \( p \) cancels). This is bounded by \( M = 1 \) since \( -\sum_i (r(\theta_i) - y_i)^2 \) is at most zero. Our rejection algorithm sampling the posterior is therefore

1) Simulate \( \theta \sim p(\theta) \) according to the prior. Simulate \( u \sim U(0, 1) \).

2) if \( u \leq \exp(-\sum_i (r(\theta_i) - y_i)^2/2\sigma_i^2) \) then accept \( \theta \) as a sample, otherwise return to 1).

Make sure you understand why this is a correct rejection algorithm before you continue.

Our data are the RCD’s \( y \) and our unknown parameters are the ages \( \theta \). We will consider two different priors \( p_1(\theta) \) and \( p_2(\theta) \) in order to make a sensitivity analysis. The posterior \( p_j(\theta | y) \) for prior \( j \) has the form

\[
p_j(\theta | y) = \frac{L(\theta; y)p_j(\theta)}{m_j(y)}.
\]

In the setting in which the data were gathered the key question was how long the site had been occupied. This is called the ‘span’ of settlement. We will estimate span using the span function \( s(\theta) = \max(\theta) - \min(\theta) \). We will assume that it is known from separate sources that the site was occupied after \( U = 700 \text{BP} \) and deserted before \( L = 500 \text{BP} \), so \( L \leq \theta_i \leq U \) for \( i = 1, 2, \ldots, K \).
The first prior we will consider is the uniform prior on \([L, U]^K\), \(p_1(\theta) = (U - L)^{-K}\). This prior is easy to simulate. We simply simulate \(\theta_i \sim U(L, U)\) for \(i = 1 : K\).

To simulate one set of \(\theta_1,...,\theta_K\) from \(p_1(\theta)\):

```plaintext
rp1<-function(K) {return(runif(min=L,max=U,n=K))}
```

It is not too hard to show that the marginal density \(p_S(s)\) of \(s(\theta) = \max(\theta) - \min(\theta)\) in this prior is \(p_S(s) \propto s(\theta)^{K-2}(U - L - s(\theta))^{-1}\). This means that a span of 2\(s\) years is about \(2^5\) times more probable a priori than a span of \(s\) years in our case (\(K = 7\)). This kind of bias is not justified by prior knowledge.

The second prior has the property that the marginal distribution of the span statistic is uniform:

\[ p_2(\theta) \propto s(\theta)^{2-K}(U - L - s(\theta))^{-1}. \]

This is also easy to simulate. We simulate a span \(s \sim U(0, U - L)\) then simulate the lowest aged date \(\theta'_1 \sim U(L, U - s)\) so that the span will fit in the interval \([L, U]\). The largest aged date must be \(\theta'_K = \theta'_1 + s\). Now just throw the rest of the \(K - 2\) dates down between \(\theta'_1\) and \(\theta'_K\), \(\theta'_i \sim U(\theta'_1, \theta'_K), i = 2, ..., K - 1\). The permutation must be random (we dont want \(y_1\) always to be associated with the least age etc, all specimen are equally likely a priori to be youngest) so we obtain \(\theta \sim p_2(\theta)\) by randomly permuting the \(\theta'\) values.

To sample for prior with uniform prior on span:

```plaintext
rp2<-function(K) {
    #simulate a span uniformly from 0 to R
    sp=runif(1,0,R)
    #simulate the lowest date uniformly from L to U-span
    ps=c(0,0)
    ps[1]=runif(1,min=L,max=U-sp)
    #the largest date must be the lowest plus the span
    #simulate the rest of the dates between the lowest and the highest
    th=c(ps,runif(min=ps[1],max=ps[2],n=K-2))
    #randomly permute so that each y is equally likely to be assigned given age
    return(sample(th))
}
```

Ex 1

(a) Use \(rp1(K)\) and \(rp2(K)\) to simulate the priors \(p_1\) and \(p_2\) and investigate the marginal density of the span by making histograms.

(b) Explain why the density of \(\theta\) returned by the algorithm \(rp2(K)\) is indeed \(p_2(\theta)\).

(c) Prove that the generic rejection algorithm above, which samples a posterior by simulating proposals according to the prior, and accepting with probability equal the ratio of the likelihood to its upper bound, is correct.

We can now implement our rejection samplers for the two target distributions.

```plaintext
#rejection sampler for posterior
#takes argument rp() a function simulating th~p(th)
#It carries out rejection sampling, proposing th~rp(K) with 'th'
```
#K-component vector, and accepting ‘th’ w.p. exp(...). It returns
#list(theta=th,psi=c(min(th),max(th)),count=ct), where count is the
#number of trials it took for acceptance

}  #1000 samples - focus on span of dates
J=1000
ps1=ps2=matrix(0,J,2)
for (k in 1:J) {ps1[k,]=rpost(rp1)$psi}
plot(density(diff(t(ps1))),xlim=c(0,300),
        xlab='span=max-min ages',ylab='estimated posterior density',main='')
for (k in 1:J) {ps2[k,]=rpost(rp2)$psi}
lines(density(diff(t(ps2))),col=2)

Ex 2
(a) Implement a rejection sampler as indicated above. Your function should take as input
the sampler function rp1 or rp2. It should carry out rejection sampling, proposing
θ ∼ pi(θ) and accepting θ with probability exp(...). Your function should returns a list
list(theta=th,psi=c(min(th),max(th)),count=ct) where count is the number of
trials it took for acceptance.

(b) How does the choice of prior impact the posterior distribution of the span?

(c) show that E(N) (the expectation of N, the number of trials we make in our loops through
steps 1) and 2)) in the rejection algorithm is E(N) = c/m(y) with c = ∏i(2πσi2)−1/2. If
N is the average number of trials per acceptance in multiple simulations of θ ∼ P(θ|y),
show that n̂ = c/N is a consistent estimator for the marginal likelihood m(y).

(d) The Bayes factor B = m2(y)/m1(y) can be used to test evidence for the ’shrinkage’ prior
2 over the ’uniform’ prior 1. Let N(i) be the average number of trials it takes to make one
draw from the posterior when we use prior i. Use the estimator B̂ = N(1)/N(2) to carry
out Bayesian model selection.

Ex 3  Suppose we were interested in the posterior mean span. We can use samples from the
two posteriors to estimate these spans.

> mean(diff(t(ps1)))
[1] 110.8006
> mean(diff(t(ps2)))
[1] 58.67977

Consider using un-normalised importance sampling to estimate the posterior mean span for
posterior p1(θ|y) using samples from p2(θ|y)

1. Calculate the weight function w(θ) = p1(θ)/p2(θ) for self normalized importance sam-
pling and show that

\[ w(θ) = s(θ)^5(U - L - s(θ)) \]

are suitable weights. Make an estimate \[ \sum_k s(θ(k))w_k / \sum_k w_k \] of E(s(θ)|y, p1) using
samples θ(k) ∼ p2(θ|y), k = 1, 2, ..., J.

2. Why might this particular importance sampling scheme be unreliable? Investigate the
stability of your estimate for E(s(θ)|y, p1).