Aims

- Posterior simulation and integration
- Direct simulation
  - by calculating at a grid of points
  - Rejection Sampling
- Numerical Integration
  - Importance Sampling
  - Importance Resampling (SIR)

Posterior simulation and integration

- Complicated model such as hierarchical models are most conveniently summarized by random draws from the posterior distribution of the model parameters
- We should report
  - percentiles of the posterior distribution of univariate estimands
  - scatterplot of simulations
  - contour plot of the density functions of the posterior distribution of the multivariate estimands

Direct approximation by calculating at a grid of points

1. Compute the unnormalized density, $q(\theta \mid y)$, at a set of evenly spaced values $\theta_1, \ldots, \theta_N$, that cover the range of parameter space that is of interest
2. Approximate $p(\theta \mid y)$ by a discrete density at $\theta_1, \ldots, \theta_N$ with probabilities $q(\theta_i)/\sum_{i=1}^{N} q(\theta_i \mid y)$
3. the method will work also with an unnormalized density $q(\theta \mid y)$
4. Once a grid of density values is computed, a random draw from $p(\theta \mid y)$ is obtained by drawing a random sample $u \sim U[0,1]$ then transforming by the inverse cdf method to obtain a sample from the discrete approximation

Rejection Sampling

1. Technique for simulating draws directly from $p(\theta \mid y)$
2. let $q(\theta \mid y) \propto L(\theta)p(\theta)$, e.g. an unnormalized density
3. let $g(\theta)$ a positive function defined for all θ such that:

   we can draw from $g(\theta)$

   $\exists M : \; q(\theta \mid y) \leq Mg(\theta)$ $\forall\theta$ such that $q(\theta \mid y) > 0$

   $g(\theta)$ must have a finite integral

1. draw θ_j from $g(\theta)$
2. draw u_j from $U[0,1]$
3. if $Mu_j g(\theta_j) < L(\theta_j)p(\theta_j) \;\rightarrow$ accept θ_j otherwise reject θ_j
4. *in other words we accept θ_j with prob* $\frac{L(\theta_j)p(\theta_j)}{Mg(\theta_j)}$

Advantages

1. if $g \propto p$ with a suitable value of M we can accept every draw with probability 1
2. if g it is not nearly proportional to p, M must be so large that almost all the samples obtained in step 1. will be rejected in step 2.
3. the method is self monitoring — if the method is not working efficiently, very few simulated draws will be accepted

figure

Illustration of the rejection sampling: the top curve is an approximation function, $Mg(\theta)$, and the bottom curve is the target density $q(\theta \mid y)$. As required $Mg(\theta) \geq q(\theta \mid y)$. The vertical line indicates a single random draw θ^* from $g(\theta)$. The probability that a sampled θ^* is accepted is the ratio of the height of the lower curve to the height of the higher curve in the value θ^*

Importance Sampling and Numerical Integration

- Aim: to estimate $E[h(\theta) \mid y] = \int h(\theta)p(\theta \mid y)d\theta$
- Problem: a closed form of $p(\theta \mid y)$ is not available
- Let $g(\theta)$ be a normalized density from which we can generate random draws
- we can write

$$\begin{aligned} E[h(\theta) \mid y] &= \int h(\theta)\frac{p(\theta|y)}{g(\theta)}g(\theta)d\theta \\ &= c^{-1}\int h(\theta)w(\theta)g(\theta)d\theta \text{ where} \\ w(\theta) &= \frac{q(\theta|y)}{g(\theta)} \text{ and } c = \int q(\theta \mid y)d\theta \end{aligned}$$

- draw $\theta^1, \ldots, \theta^L$ from $g(\theta)$
- calculate the importance ratios $w(\theta^l) = \frac{q(\theta^l|y)}{g(\theta^l)}$
- estimate $E[h(\theta) \mid y]$ by $\frac{\frac{1}{L}\sum_{l=1}^{L} h(\theta^l)w(\theta^l)}{\frac{1}{L}\sum_{l=1}^{L} w(\theta^l)}$

Importance Sampling, cont.

- Unlike the rejection sampling, the approximating density $g(\theta)$ must be normalized
- Importance sampling is not a very useful method if the importance ratios vary substantially
- estimates will be poor if the largest ratios are too large relative to the others.
- for example using a t_3 as an approximation of the normal (good idea)
- using a normal as an approximation of the t_3 (bad idea)

Sampling-Importance Resampling SIR

- Importance weights can be used to get a sequence of draws that approximate the target distributions by using the SIR method
- $g(\theta$ can be unnormalized
- if the ratio $q(\theta \mid y)/g(\theta)$ is bounded, then we can use rejection sampling also
- in SIR we sample $\theta^1, \ldots, \theta^L$ from $g(\theta)$, a sample $k < L$ draws from a better approximation of $p(\theta \mid y)$ can be simulated as follows:
  1. Sample a value θ from the set $\left\{\theta^1, \ldots, \theta^L\right\}$ where the probability of sampling each θ^l is proportional

to the weight

$$w(\theta^l) = \frac{q(\theta^l \mid y)}{g(\theta^l)}, \; \sum_{l=1}^{L} w(\theta^l) = 1$$

  2. sample a second value using the same procedure, but excluding the already sampled value from the set
  3. repeatedly sample without replacement $k-2$ more times

- Example: Program simulations, using t_3 as approximation of the normal (ex:10.9)
- in other words... we draw θ^* from the discrete distribution over $\{\theta^1, \ldots, \theta^l, \ldots, \theta^L\}$ with probabilities $w(\theta^l)$ (*weighted bootstrap*)

Why sample without replacement?

- If the importance weights do not vary much, sampling with or without replacement gives similar results
- now consider a bad case, with a few large values and many small values
- sampling with replacement will pick the same few values of θ again and again
- sampling without replacement yields a more desirable approximation somewhere between the starting and the target densities
- histogram of the logarithms of the largest importance ratios to check that there are no extremely high values that would unduly influence the distribution.


```
#importance sampling when the importance weights are well  behaved
 importance_function(w=0,L,K,mmm=0,sss=2,DF=3,xl=c(-5,5)){
  if (w == 1) postscript("/home/biostats/fdominic/course/SIRgood.ps")
  par(mfrow=c(2,1))
  par(oma=c(0,0,2,0))
  theta_NULL
  theta.better_NULL
  weight_NULL
  for(l in 1:L){
    theta[l]_rt(1,DF)
    weight[l]_dnorm(theta[l],mean=mmm,sd=sss)/dt(theta[l],df=DF)
  }
  hist(log(weight),nclass=100,xlab="log importance ratios",density=-1)
  post.mean_mean(theta*weight)
  post.var_1/L*sum(weight*(theta-mean(theta))^2)
  theta.better_sample(theta, K,prob=weight)
  hist(theta.better,xlab="theta",nclass=100,
       xlim=c(xl[1],xl[2]) ,density=-1,prob=T,yaxt="n")
  true   _density(theta.better,width=width,from=xl[1],to=xl[2])
  lines(true$x,true$y,type="l", lty=2)
  abline(v=mean(theta.better))
  abline(v=mean(theta.better)-sqrt(var(theta.better)),lty=2)
  abline(v=mean(theta.better)+sqrt(var(theta.better)),lty=2)
  par(mfrow=c(1,1))
  par(oma=c(0,0,0,0))
  if (w == 1) dev.off()
  return(post.mean,post.var)
}
```

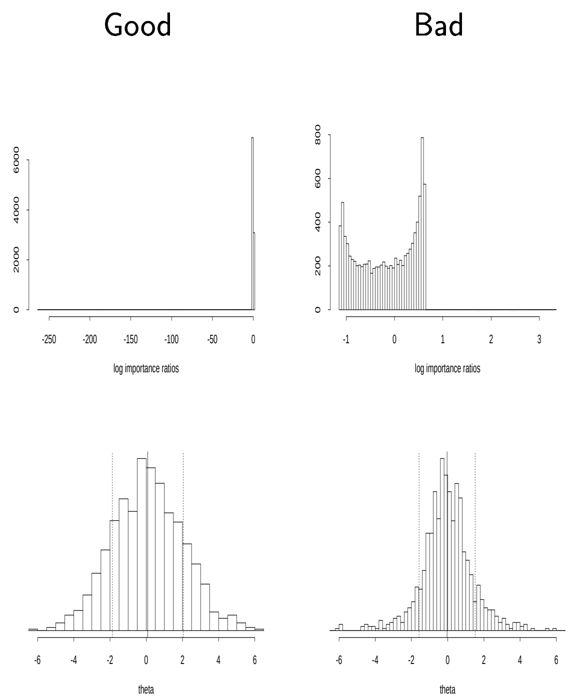

Figure 1: Importance sampling when the weights are good and bad behaved