Monte Carlo Methods

Ricardo Ehlers
ehlers@icmc.usp.br

Departamento de Matemática Aplicada e Estatística
Universidade de São Paulo

Ch. 3 in Robert & Casella
Ch. 3 in Gamerman & Lopes
Ch. 10 in Gelman et. al.

Introduction to Monte Carlo

Simple Monte Carlo

Suppose we want to compute the following integral,

$$I = \int_a^b g(\theta) d\theta.$$

which can be rewritten as,

$$I = (b - a) \int_a^b g(\theta) \; \frac{1}{b - a} \; d\theta = (b - a) \; E[g(\theta)].$$

where $\theta \sim U(a, b)$.

So, we transformed the problem of evaluating the integral into the statistical problem of estimating a mean, $E[g(\theta)]$.

Given a random sample $\theta_1, \ldots, \theta_n$ from the uniform distribution in the interval (a, b) we also have a sample of values $g(\theta_1), \ldots, g(\theta_n)$ of the function $g(\theta)$ and the integral above can be approximated by,

$$\hat{I} = (b - a)\ \frac{1}{n} \sum_{i=1}^{n} g(\theta_i).$$

This estimate is unbiased since,

$$E(\hat{I}) = \frac{(b-a)}{n} \sum_{i=1}^{n} E[g(\theta_i)] = (b-a)E[g(\theta)] = \int_a^b g(\theta) d\theta.$$

Then,

1. Generate $\theta_1, \ldots, \theta_n \sim U(a, b)$.
2. Compute $g(\theta_i)$, $i = 1, \ldots, n$.
3. Compute $\bar{g} = \sum_{i=1}^{n} g(\theta_i)/n$
4. Compute $\hat{I} = (b - a)\ \bar{g}$.

Example. An R function for $g(\theta) = e^{-\theta}$.

```
> int.exp <- function(n,a,b){
+  x = runif(n,a,b)
+  y = exp(-x)
+  int.exp = (b-a)*mean(y)
+  return(int.exp)
+ }

> int.exp(n=10,a=0,b=1)

[1] 0.6405642
```

Example. Repeting the simulations 20 times with $n = 10$, $a = 1$ and $b = 3$ we note a considerable variation in the results.

```
> m = NULL
> for (i in 1:20) m = c(m,int.exp(n=10,a=1,b=3))
> write(file="",round(m,4))
```

```
0.3634 0.2536 0.3366 0.2979 0.2738
0.2441 0.3109 0.3087 0.2866 0.2998
0.2364 0.3277 0.4164 0.2754 0.3677
0.3811 0.3121 0.3713 0.2333 0.3307
```

```
> summary(m)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.2333  0.2750  0.3098  0.3114  0.3433  0.4164
```

This is called Monte Carlo error and it decreases as we increase the number of simulations.

Example. Repeting the previous example with $n = 1000$ we notice a much smaller variation.

```
> m = NULL
> for (i in 1:20){m = c(m,int.exp(1000,1,3))}
```

Estimates,

```
> write(file="",round(m,4))

0.3149 0.3222 0.3154 0.3198 0.3163
0.316 0.3215 0.3136 0.3239 0.3131
0.335 0.3262 0.3174 0.3165 0.3202
0.3268 0.3118 0.3223 0.3134 0.3044
```

```
> summary(m)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.3044  0.3146  0.3170  0.3185  0.3222  0.3350
```

Evolution of the Monte Carlo error with the number of simulations.

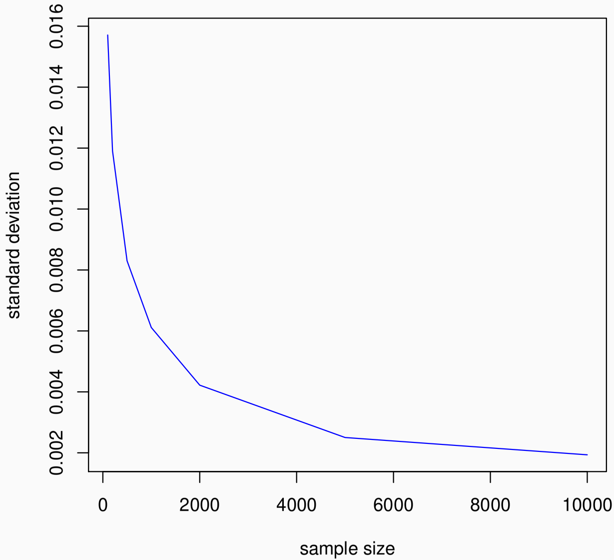
```
> n = c(100,200,500,1000,2000,5000,10000)
> y = matrix(0,ncol=length(n),nrow=2)
> for (j in 1:length(n)){
+   m=NULL
+   for (i in 1:50) m = c(m,int.exp(n[j],1,3))
+   y[1,j] = mean(m)
+   y[2,j] = sd(m)
+ }
```

Estimates and standard deviations,

```
> round(y,4)

       [,1]   [,2]   [,3]   [,4]   [,5]   [,6]   [,7]
[1,] 0.3163 0.3178 0.3166 0.3167 0.3182 0.3186 0.3177
[2,] 0.0157 0.0119 0.0083 0.0061 0.0042 0.0025 0.0019
```

Standard deviations across sample sizes.



This method can be generalized to compute the expectation of a function $g(\theta)$ which density function is $\pi(\theta)$.

$$I = \int g(\theta)\pi(\theta)d\theta = E[g(\theta)]. \tag{1}$$

1. Generate $\theta_1, \ldots, \theta_n$ from the distribution $\pi(\theta)$.
2. Compute $g(\theta_i), \quad i = 1, \ldots, n$.
3. Compute $\hat{I} = \frac{1}{n}\sum_{i=1}^{n} g(\theta_i)$.

By the strong law of large numbers, $\hat{I}$ converges to I almost surely, that is

$$\frac{1}{n}\sum_{i=1}^{n} g(\theta_i) \quad \overset{n\to\infty}{\longrightarrow} \quad E[g(\theta)],$$

or equivalently,

$$P\left(\lim_{n\Rightarrow\infty} |\hat{I} - I| < \epsilon\right) = 1.$$

Example. Computing the mean and variance of $\theta \sim \text{Exp}(1)$.

$$E(\theta) = \int_0^{\infty} \theta e^{-\theta} d\theta.$$

In this case $g(\theta) = \theta$ and $\pi(\theta) = e^{-\theta}$. Likewise, for $E(\theta^2)$ we take $g(\theta) = \theta^2$.

1. Generate independent values $\theta_1, \ldots, \theta_n \sim Exp(1)$.
2. Compute $\sum_{i=1}^{n} \theta_i / n$.
3. Compute $\sum_{i=1}^{n} \theta_i^2 / n$.

```
> n = 1000
> x = rexp(n, rate=1)
> mean(x)

[1] 1.029486

> mean(x^2) - mean(x)^2

[1] 1.142981
```

Example. Computing probabilities for $\theta \sim \text{Exp}(1)$.

$$
\begin{aligned}
P(a < \theta < b) &= \int_a^b e^{-\theta} d\theta \\
&= \int_0^\infty I(\theta \in (a, b))\ e^{-\theta} d\theta = E[I(\theta \in (a, b))],
\end{aligned}
$$

with $g(\theta) = I(\theta \in (a, b))$ and $\pi(\theta) = e^{-\theta}$.

To compute $P(a < \theta < b)$ we simply need to compute the relative frequency of simulated values that fall in this interval.

1. Generate independent values $\theta_1, \ldots, \theta_n \sim Exp(1)$.
2. Compute the relative frequency of θ_i's $\in (a, b)$.

```
> n = 1000
> x = rexp(n,1)
> sum (x > 1 & x < 3) / n
```

```
[1] 0.327
```

Example. Suppose we want to compute $P(\theta < 1)$ where $\theta \sim N(0,1)$.

$$\begin{aligned} P(\theta < 1) &= \int_{-\infty}^{1} \frac{1}{\sqrt{2\pi}} \exp(-\theta^2/2) d\theta \\ &= \int_{-\infty}^{\infty} I(\theta < 1) \frac{1}{\sqrt{2\pi}} \exp(-\theta^2/2) d\theta = E[I(\theta < 1)] \end{aligned}$$

1. Generate independent values $\theta_1, \ldots, \theta_n \sim N(0,1)$.
2. Compute the relative frequency of θ_i's less than 1.

$$P(\theta < 1) \approx \frac{1}{n} \sum_{i=1}^{n} I(\theta_i < 1).$$

Example. This method might not be efficient to compute very small probabilities. For example if $\theta \sim N(0,1)$ probabilities of extreme values as $P(\theta > 4)$ will be too small and it would be necessary to simulate more values.

```
> q = c(-4,-3,-2,-1,0)
> n = c(1000,2000,5000,10000)
> prob = matrix(NA,length(q),length(n))
> for (j in 1:length(n)) {
+     zi= rnorm(n[j])
+     for (i in 1:length(q)) prob[i,j] = mean(zi < q[i])
+ }
> colnames(prob)=n
> rownames(prob)=q
> prob

    1000   2000   5000  10000
-4 0.000 0.0000 0.0000 0.0000
-3 0.001 0.0010 0.0024 0.0010
-2 0.020 0.0215 0.0238 0.0223
-1 0.163 0.1610 0.1506 0.1541
0  0.479 0.4950 0.4864 0.4978
```

Example. Suppose we need to calculate $P(a < \theta < b)$ where $\theta = X/\sqrt{Y}$, $X \sim N(0,1)$ and $Y \sim N(0,1)$.

The distribution of θ is unknown and while asymptotic approximations do exist it is easier and more straightforward to use simulation.

If $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ are simulated values from a $N(0,1)$ then $\theta_i = x_i/\sqrt{y_i}$, $i = 1, \ldots, n$ form a sample from the distribution of θ.

1. Generate independent values $x_1, \ldots, x_n, y_1, \ldots, y_n \sim N(0,1)$.
2. Compute $\theta_i = x_i/\sqrt{y_i}$, $i = 1, \ldots, n$
3. Compute the relative frequency of θ_i's $\in (a, b)$.

Computing the Monte Carlo error

Suppose that,

1. $\theta_1, \ldots, \theta_n$ were generated and,
2. we computed $g(\theta_1), \ldots, g(\theta_n)$.

Since the simulations are independent,

$$Var\left(\frac{1}{n}\sum_{i=1}^{n} g(\theta_i)\right) = \frac{\sigma_g^2}{n}$$

where $\sigma_g^2 = Var(g(\theta_i))$, $i = 1, \ldots, n$,

This is in turn estimated as,

$$\frac{1}{n}\sum_{i=1}^{n}(g(\theta_i) - \bar{g})^2,$$

where

$$\bar{g} = \frac{1}{n}\sum_{i=1}^{n} g(\theta_i).$$

Then, the variance of the Monte Carlo estimator is approximated as,

$$\widehat{Var}\left(\frac{1}{n}\sum_{i=1}^{n} g(\theta_i)\right) = \frac{1}{n^2}\sum_{i=1}^{n}(g(\theta_i) - \bar{g})^2.$$

The Monte Carlo error (standard error of the estimate) is the square root of this expression.

So,

$$\hat{\sigma} = \sqrt{\frac{1}{n^2}\sum_{i=1}^{n}(g(\theta_i) - \overline{g})^2},$$

where $\sigma^2 = Var[g(\theta)]$.

- The central limit theorem also applies. For n large it follows that,

$$\frac{\overline{g} - E[g(\theta)]}{\hat{\sigma}}$$

has approximate $N(0, 1)$ distribution.

- We can use this result to assess convergence and construct intervals $\overline{g} \pm z_{\alpha/2}\hat{\sigma}$.

- For a random vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)'$ with density function $\pi(\boldsymbol{\theta})$ the simulated values will also be vectors $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n$ and the Monte Carlo estimate is,

$$\hat{I} = \frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{\theta}_i).$$

Example. Computing the Monte Carlo error for $P(1 < \theta < 3)$, $\theta \sim \mathit{Exp}(1)$.

```
> z = (x-mean(x))**2/n
> ep.mc = sqrt(mean(z))
> ep.mc
```

```
[1] 0.02020395
```

Importance Sampling

It can be either too costly or even impossible to sample values from $\pi(\theta)$.

The integral of interest can however be rewritten as,

$$I = \int_{-\infty}^{\infty} g(\theta)\pi(\theta)d\theta = \int_{-\infty}^{\infty} \frac{g(\theta)\pi(\theta)}{q(\theta)} q(\theta)d\theta = \int_{-\infty}^{\infty} w(\theta)q(\theta)d\theta.$$

That is,

$$E_\pi[g(\theta)] = E_q[w(\theta)]$$

Given a random sample $\theta_1, \ldots, \theta_n$ from the distribution q a Monte Carlo estimate of the integral above is,

$$\hat{I} = \frac{1}{n} \sum_{i=1}^{n} \frac{g(\theta_i)\pi(\theta_i)}{q(\theta_i)}.$$

Again, when $n \to \infty$,

$$\hat{I} \longrightarrow \int_{-\infty}^{\infty} g(\theta)\pi(\theta) d\theta.$$

1. Generate values $\theta_1, \ldots, \theta_n$ from $q(\cdot)$.
2. Compute the weights,

$$w(\theta_i) = \frac{g(\theta_i)\pi(\theta_i)}{q(\theta_i)}, i = 1, \ldots, n.$$

3. Compute

$$\hat{I} = \frac{1}{n}\sum_{i=1}^{n} w(\theta_i).$$

So, the algorithm produces a weighted average which is used to approximate $E_\pi[g(\theta)]$.

- $q(\theta)$ is usually called *importance function* and should be easy to sample from.
- The procedure is commonly called *importance sampling*.
- If $\pi(\theta)$ is completely known the estimator uses all generated values with the same weights $1/n$.
- In practice, better results are obtained when $q(\theta)$ is a good approximation for $\pi(\theta)$.
- If some few weights are much larger than all the others they will dominate the estimate. To try to avoid this, $q(\theta)$ should in general have heavier tails than $\pi(\theta)$, that is $\pi(\theta) \leq q(\theta)$ in the tails.

Example. Suppose we want to compute $P(\theta > 4)$ when $\theta \sim N(0, 1)$.

Applying the usual Monte Carlo method, we would generate a sample $\theta_1, \ldots, \theta_n \sim N(0, 1)$ and compute,

$$P(\theta > 4) \approx \frac{1}{n} \sum_{i=1}^{n} I(\theta_i > 4).$$

Since $P(\theta > 4)$ is too small there will be many terms $I(\theta_i > 4) = 0$ and $P(\theta > 4) \approx 0$.

It is more efficient to generate θ_i from a distribution $q(\theta)$ with higher probabilities in the tails and apply the approximation,

$$P(\theta > 4) \approx \frac{1}{n} \sum_{i=1}^{n} I(\theta_i > 4) \frac{\pi(\theta_i)}{q(\theta_i)}.$$

Suppose now that $\pi(\theta) = kf(\theta)$, where k is an unknown constant.

Then,

$$\pi(\theta) = \frac{f(\theta)}{\int f(\theta) d\theta}$$

and we still want to compute $E[g(\theta)]$.

However,

$$
\begin{aligned}
E[g(\theta)] &= \frac{1}{\int f(\theta)d\theta} \int g(\theta)\ f(\theta)d\theta \\
&= \frac{1}{\int \frac{f(\theta)}{q(\theta)} q(\theta)d\theta} \int \frac{g(\theta)\ f(\theta)}{q(\theta)} q(\theta)d\theta \\
&= \frac{E_q[g(\theta)w(\theta)]}{E_q[w(\theta)]}.
\end{aligned}
$$

where,

$$
w(\theta) = \frac{f(\theta)}{q(\theta)}.
$$

- So, when π is only known up to a constant, expectations are ratios of two integrals.
- The approximation is then based on the ratio of two Monte Carlo estimators of the integrals (using the same importance density).

1. Generate $\theta_1, \ldots, \theta_n$ from q.
2. Compute, $w(\theta_i) = f(\theta_i)/q(\theta_i),\ i = 1, \ldots, n$.
3. Approximate the expectation as,

$$\widehat{E[g(\theta)]} = \frac{1}{\sum_{i=1}^{n} w(\theta_i)} \sum_{i=1}^{n} g(\theta_i) w(\theta_i).$$

This estimator is still strongly consistent,

$$\widehat{E[g(\theta)]} \overset{n \to \infty}{\longrightarrow} E[g(\theta)],\ a.s.$$

and asymptotically unbiased.

Example. Let $X \sim N(\theta, 1)$ and $\theta \sim Cauchy(0, 1)$.

We have that,

$$\begin{aligned} p(x|\theta) &\propto \exp(-(\theta - x)^2/2) \\ p(\theta) &= \frac{1}{\pi(1+\theta^2)} \\ \pi(\theta) = p(\theta|x) &= \frac{p(x|\theta)p(\theta)}{p(x)} = \frac{p(x|\theta)p(\theta)}{\int p(x|\theta)p(\theta)d\theta}. \end{aligned}$$

and we need an approximation for,

$$E[\theta|x] = \int \theta p(\theta|x) d\theta$$

After the associated substitutions it follows that,

$$E[\theta|x] \quad = \quad \frac{\displaystyle\int \frac{\theta}{1+\theta^2} \exp[-(x-\theta)^2/2] \; d\theta}{\displaystyle\int \frac{1}{1+\theta^2} \exp[-(x-\theta)^2/2] \; d\theta}.$$

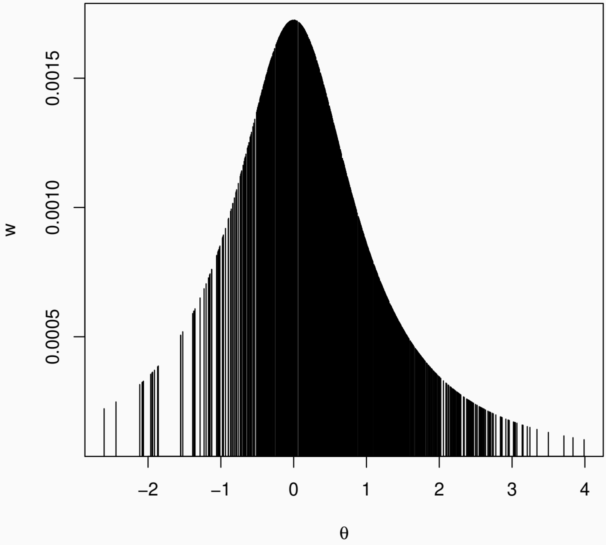
The approximate solutions is,

1. generate $\theta_1, \ldots, \theta_n$ independently from the $N(x, 1)$ distribution;
2. compute $w_i = \dfrac{1}{1+\theta_i^2}$.
3. compute, $\hat{E}(\theta|\mathbf{x}) = \dfrac{\sum_{i=1}^n \theta_i w_i}{\sum_{i=1}^n w_i}$.

1000 simulated values of $\theta_i \sim N(x, 1)$ with $X \sim N(2, 1)$ and normalized weights.

```
> set.seed(1234)
> x = rnorm(1,2,1)
> n = 1000
> theta = rnorm(n,mean=x,sd=1)
> w = dcauchy(theta)
> sw= sum(w)
> w= w/sw
```

Normalized weights.



Estimating θ,

```
> g = theta/(1+theta^2)
> g1= 1/(1+theta^2)
> est = sum(g)/sum(g1)
> est
```

```
[1] 0.4208462
```

Example. In the previous example we can generalize for k observations $x_1, \ldots, x_k$.

In this case,

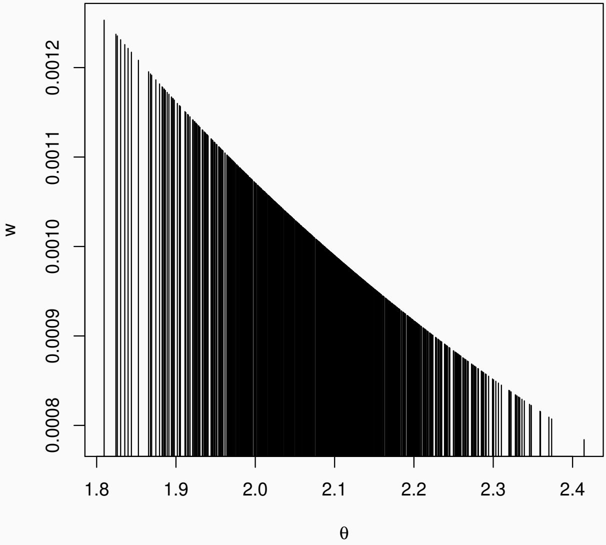
$$p(\theta|\mathbf{x}) \propto \frac{1}{1+\theta^2} \exp[-k(\bar{x}-\theta)^2/2].$$

and an approximation for $E(\theta|\mathbf{x})$ is obtained with the same algorithm generating $\theta_1, \ldots, \theta_n$ independently from a $N(\bar{x}, 1/k)$ distribution.

1000 simulated values of $\theta_i \sim N(\bar{x}, 1/k)$ with $X_1, \ldots, X_k \sim N(2, 1)$, $k = 100$ and normalized weights.

```
> set.seed(123)
> k = 100
> x = rnorm(k,2,1)
> xbar = mean(x)
> n = 1000
> theta = rnorm(n,mean=xbar,sd=sqrt(1/k))
> w = dcauchy(theta)
> sw = sum(w)
> w= w/sw
```

Normalized weights.



Estimating θ,

```
> g = theta/(1+theta^2)
> g1= 1/(1+theta^2)
> est = sum(g)/sum(g1)
> est

[1] 2.084481
```

Example. Suppose we need to compute $P(0.1 < \theta < 0.5)$ where θ has density function given by,

$$\pi(\theta) \propto \theta(1-\theta)^3, \theta \in (0,1).$$

Then,

$$P(0.1 < \theta < 0.5) = \frac{\int_0^1 I[\theta \in (0.1, 0.5)]\ \theta(1-\theta)^3 d\theta}{\int_0^1 \theta(1-\theta)^3 d\theta}$$

$$\frac{1}{\int_0^1 \frac{\theta(1-\theta)^3}{q(\theta)} q(\theta) d\theta} \int_0^1 I[\theta \in (0.1, 0.5)]\ \frac{\theta(1-\theta)^3}{q(\theta)} q(\theta) d\theta.$$

Using the previous algorithm with $q(\theta) = 1$,

1. Generate $\theta_1, \dots, \theta_n \sim U(0, 1)$.
2. For $i = 1, \dots, n$, $w(\theta_i) = \theta_i(1 - \theta_i)^3$.
3. Compute,

$$P(0.1 < \theta < 0.5) \approx \frac{\sum_{i=1}^{n} I[\theta \in (0.1, 0.5)] w(\theta_i)}{\sum_{i=1}^{n} w(\theta_i)}.$$

```
> set.seed(1234)
> x= runif(1000,0,1)
> y= x[x > 0.1 & x < 0.5]
> wx= x * (1-x)^3
> wy= y * (1-y)^3
> sum(wy)/sum(wx)
```

```
[1] 0.7265098
```

The normalizing constant is obtained as a by product.

```
> 1/mean(wx)
```

```
[1] 20.241
```

Example. A random variable θ has a Laplace (or double exponential) distribution with mean zero and variance 1 if,

$$p(\theta) = \frac{1}{\sqrt{2}} \exp(-\sqrt{2}\ |\theta|),\ \theta \in \mathbb{R}.$$

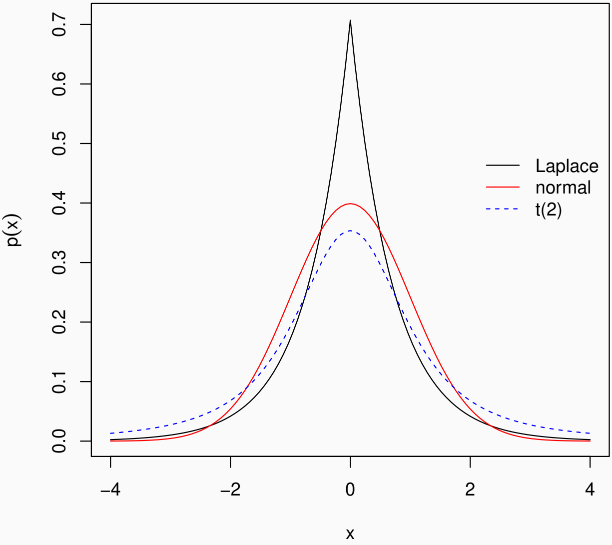
We can use the standard normal as an auxiliary distribution to compute,

$$\begin{aligned} P(\theta \leq q) &= \int_{-\infty}^{q} p(\theta) d\theta = \int_{-\infty}^{\infty} I_{(-\infty,q)}(\theta)\ p(\theta) d\theta \\ &= \int_{-\infty}^{\infty} I_{(-\infty,q)}(\theta)\ \frac{p(\theta)}{\phi(\theta)} \phi(\theta) d\theta \\ &= E_{\phi}\left[I_{(-\infty,q)}(\theta)\ \frac{p(\theta)}{\phi(\theta)} \right] \end{aligned}$$

1. Generate $\theta_1, \ldots, \theta_n \sim N(0, 1)$.
2. For $i = 1, \ldots, n$, set $w(\theta_i) = p(\theta_i)/\phi(\theta_i)$.
3. Compute,

$$P(\theta \leq q) \approx \frac{1}{n} \sum_{i=1}^{n} w(\theta_i) I_{(-\infty, q)}(\theta_i).$$

Density functions, Normal, Student-t and Laplace.



```
> laplace <- function (x) {
+   (1/sqrt(2))*exp(-sqrt(2)*abs(x))
+ }

> prob.laplace <- function(n=1000,q) {
+   x= rnorm(n,0,1)
+   y= x[x<q]
+   w= laplace(y)/dnorm(y,0,1)
+   sum(w)/n
+   }
> prob.laplace(n=1000,q=0)

[1] 0.5059664

> prob.laplace(n=1000,q=2)

[1] 0.9581785
```

Bayes Theorem and Rejection Methods

Given a sample $x_1, \ldots, x_k$ we want to simulate values from $\pi(\theta)$ where,

$$\begin{aligned}\pi(\theta) = p(\theta|\mathbf{x}) &= k\ p(\mathbf{x}|\theta)p(\theta) \\ &= \frac{1}{\int p(\mathbf{x}|\theta)p(\theta)d\theta}\ p(\mathbf{x}|\theta)p(\theta)\end{aligned}$$

A rejection method:

1. Select an auxiliary density $q(\theta)$ and specify a constant $A < \infty$ such that,
$$\pi(\theta) < Aq(\theta).$$
2. Generate values $\theta_1, \ldots, \theta_n$ from the ausiliary distribution.
3. Accept each generated value with probability,
$$\frac{\pi(\theta)}{A\ q(\theta)} \propto \frac{p(\mathbf{x}|\theta)\ p(\theta)}{A\ q(\theta)}.$$

Step 3 is equivalent to generate $u \sim U(0,1)$ and accept if,
$$u \leq \frac{\pi(\theta)}{A\ q(\theta)}.$$

The global acceptance probability is,

$$
\begin{aligned}
P\left(U \leq \frac{\pi(\theta)}{A\ q(\theta)}\right) &= \int P\left(U \leq \frac{\pi(\theta)}{A\ q(\theta)}|\theta\right) q(\theta)d\theta \\
&= \int \frac{\pi(\theta)}{A\ q(\theta)} q(\theta)d\theta = \frac{1}{A}
\end{aligned}
$$

which is in fact independent from the simulated value θ.

The number of proposed candidates until acceptance has Geometric distribution with parameter $1/A$.

Taking $q(\theta) = p(\theta)$ the acceptance probability is,

$$\frac{p(\mathbf{x}|\theta)\ p(\theta)}{A\ p(\theta)} = \frac{p(\mathbf{x}|\theta)}{A}.$$

One possible approach would be setting $A = p(\mathbf{x}|\hat{\theta})$ where $\hat{\theta}$ is the maximum likelihood estimate of θ.

In this case,

1. Generate values $\theta_1, \ldots, \theta_n$ from the prior distribution.
2. Accept each generated value with probability,

$$\frac{p(\mathbf{x}|\theta_i)}{p(\mathbf{x}|\hat{\theta})}.$$

In the general case we have the following steps,

1. generate a value θ^* from the auxiliary distribution;
2. generate $u \sim U(0, 1)$;
3. if $u < \dfrac{p(\theta^*|\mathbf{x})}{Aq(\theta^*)}$,
    - set $\theta^{(j)} = \theta^*$,
    - set $j = j + 1$
    - return to step 1.

   otherwise return to step 1.

Example. Let $X_1, \ldots, X_n \sim N(\theta, 1)$ and $\theta \sim \text{Cauchy}(0, 1)$. The likelihood function is,

$$p(\mathbf{x}|\theta) \propto \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2\right] \propto \exp\left\{-\frac{n}{2}(\overline{x} - \theta)^2\right\}.$$

The maximum likelihood estimate is $\hat{\theta} = \bar{x}$.

1. Generate $\theta^* \sim \text{Cauchy}(0, 1)$.
2. Compute the acceptance probability,

$$p = \exp[-n(\bar{x} - \theta^*)^2/2].$$

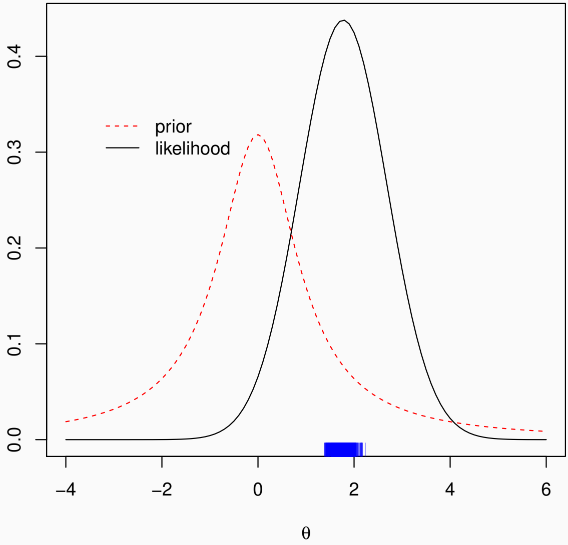
3. Generate $u \sim U(0, 1)$.
4. If $u < p$ set $\theta^{(j)} = \theta^*$, set $j = j + 1$, return to step 1. Otherwise return to step 1.

```
> rej <- function(x,m,location=0) {
+   total = 0
+   theta = NULL
+   x.bar = mean(x)
+   n = length(x)
+   for (i in 1:m) {
+     accept = FALSE
+     while (!accept) {
+       total=total+1
+       theta.new = rcauchy(1,location,1)
+       prob = exp(-0.5*n*(theta.new-x.bar)^2)
+       u = runif(1,0,1)
+       if (u < prob) {
+         theta = c(theta,theta.new)
+         accept = TRUE
+       }
+     }
+   }
+   cat("\nAcceptance rate",round(m/total,4),"\n","Number of sim
+   return(list(theta=theta, total= total))
+ }
```

Example. In the previous example, for $X_1, \dots, X_n \sim N(2, 1)$ with $n = 50$ we obtained 1000 simulated values of θ.

```
Acceptance rate 0.0278
 Number of simulations : 35915
```

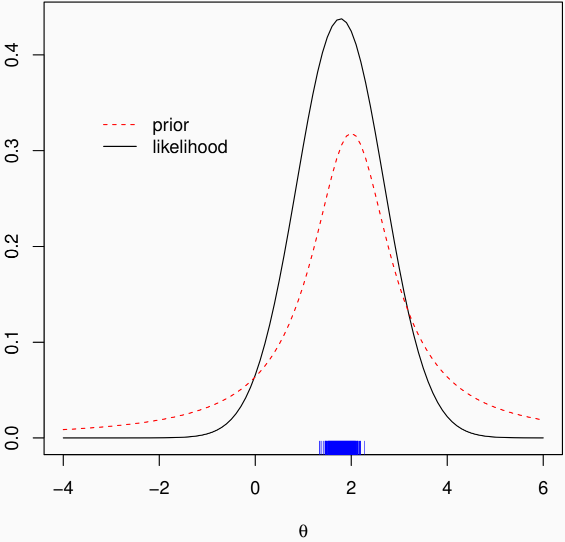
1000 simulated values of θ via rejection method.



Example. Changing the prior distribution to $\theta \sim \text{Cauchy}(2, 1)$,

$$p(\theta) = \frac{1}{\pi[1 + (\theta - 2)^2]}.$$

```
Acceptance rate 0.1049
 Number of simulations : 9531
```



prior
likelihood
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4
6
θ

Bayes theorem and Resampling

Suppose we draw a sample $\theta_1, \ldots, \theta_n \sim q(\cdot)$. A sample from $\pi(\theta)$ is obtained as follows.

1. Compute the weights,

$$w_i^* = \frac{\pi(\theta_i)}{q(\theta_i)}.$$

2. Compute the normalized weights,

$$w_i = \frac{w_i^*}{\sum_{j=1}^n w_j^*}.$$

3. Sample m values $\theta_1^*, \ldots, \theta_m^*$ from $\theta_1, \ldots, \theta_n$ with probabilities $w_1, \ldots, w_n$ with replacement.

Example. $X_1, \ldots, X_n \sim N(\theta, 1)$ and $\theta \sim \text{Cauchy}(0, 1)$. Then,

$$p(\mathbf{x}|\theta) \propto \exp\left\{-\frac{n}{2}(\overline{x} - \theta)^2\right\}.$$

We can use the prior as an auxiliary distribution.

1. Generate $\theta_1, \ldots, \theta_n$ from a Cauchy(0,1).
2. Compute the normalized weights,

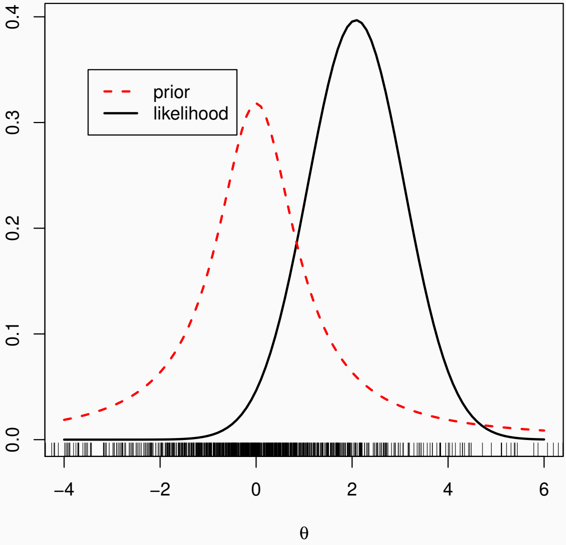
$$w_i = \frac{p(\mathbf{x}|\theta_i)}{\sum_{j=1}^n p(\mathbf{x}|\theta_j)},\ \ i = 1, \ldots, n.$$

3. Resample θ with probabilities $w_1, \ldots, w_n$.

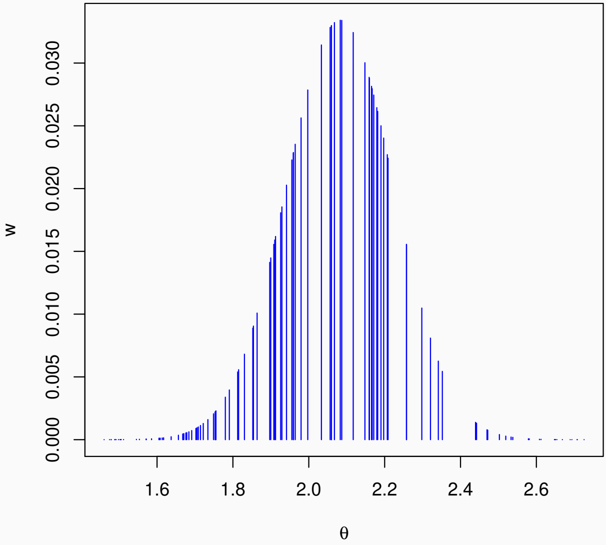
```
> resample <- function(x,n,m) {
+   x.bar= mean(x)
+   nobs = length(x)
+   theta= rcauchy(n,0,1)
+   w    = exp(-0.5*nobs*(theta-x.bar)**2)
+   aux = sum(w)
+   w   = w/aux
+   theta.star= sample(theta,size=m,replace=TRUE,prob=w)
+   return(list(sample=theta,w=w,resample=theta.star))
+ }

> x = rnorm(n=50,mean=2,sd=1)
> m = resample(x,n=1000,m=500)
```

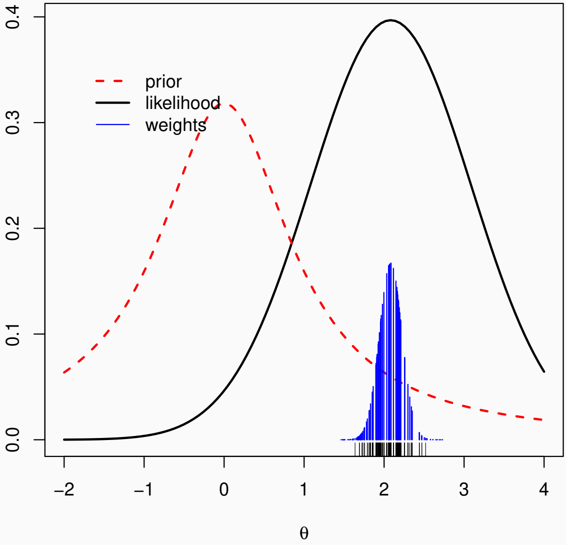
1000 simulated values from the prior distribution.



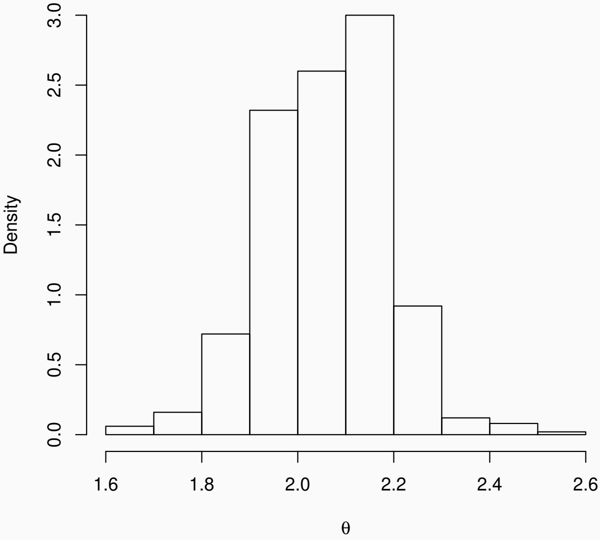
Weights of the 1000 simulated values.



500 resampled values with their weights.



Histogram of the 500 resample values.



Example. Simple linear regression with known variance,

$$\begin{aligned} y_i &= \beta x_i + \epsilon_i,\ \epsilon_i \sim N(0,1).\ i = 1, \ldots, n \\ \beta &\sim N(0,4) \end{aligned}$$

We want to estimate β based on data (y_i, x_i), $i = 1, \ldots, n$. By Bayes theorem,

$$p(\beta|\mathbf{y},\mathbf{x}) \propto p(\mathbf{y}|\beta,\mathbf{x})p(\beta)$$

where,

$$\begin{aligned} p(\mathbf{y}|\beta,\mathbf{x}) &\propto \exp(-0.5(\mathbf{y}-\beta\mathbf{x})'(\mathbf{y}-\beta\mathbf{x})) \\ &\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(y_i-\beta x_i)^2\right\} \\ p(\beta) &\propto \exp(-0.5\beta^2) \end{aligned}$$

Using the prior of β as auxiliary distribution.

1. Generate $\beta_1, \dots, \beta_N \sim N(0, 4)$.
2. Compute $w_j = p(\mathbf{y}|\beta_j, \mathbf{x})$, $j = 1, \dots, N$.
3. Compute the normalized weights, $w_j^* = w_j / \sum_{j=1}^N w_j$.
4. Generate the resample $\beta_1^*, \dots, \beta_M^*$ with probabilities $w_1^*, \dots, w_M^*$.

```
> resample.reg <- function(y,x,N,M){
+   beta = matrix(rnorm(n= N, mean=0, sd=2), nrow = N)
+   l = matrix(NA, nrow = N)
+   for (j in 1:N)
+       l[j] =  exp(-0.5 * t(y-beta[j]*x) %*% (y-beta[j]*x))
+   sw= sum(l)
+   w = l/sw
+   resample = sample(beta,size=M,replace=T, prob=w)
+   return(list(resample = resample, sample = beta))
+   }
```

Example. For illustration use the following data $\mathbf{y} = (-2, -1, 0, 1, 2)$ and $\mathbf{x} = (-2, 0, 0, 0, 2)$.

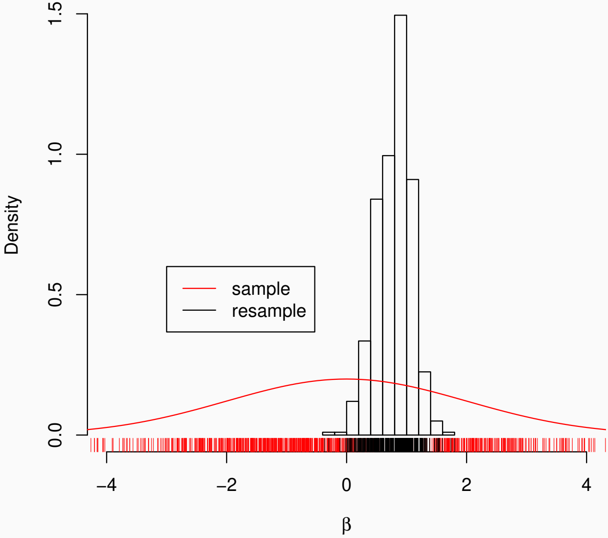
```
> x = c(-2,-1,0,1,2)
> y = c(-2, 0,0,0,2)
> out = resample.reg(y,x,N=1000,M=1000)
```

An estimate of β is given by $E[\beta|\mathbf{y}, \mathbf{x}]$ which can be approximated by,

$$\frac{1}{M}\sum_{j=1}^{M}\beta_j^*.$$

```
[1] 0.7947247
```

Sample and resample of β in the regression model.



Example. Simple linear regression with unknown variance,

$$
\begin{aligned}
y_i &= \beta x_i + \epsilon_i,\ \epsilon_i \sim N(0,\sigma^2).\ i = 1,\ldots,n \\
\beta|\sigma^2 &\sim N(0,\sigma^2) \\
\sigma^{-2} &\sim Gamma(0.1, 0.1)
\end{aligned}
$$

We need to estimate β e σ^2 based on data (y_i, x_i), $i = 1,\ldots,n$.

By Bayes theorem,

$$p(\beta,\sigma^2|\mathbf{y},\mathbf{x}) \propto p(\mathbf{y}|\beta,\sigma^2,\mathbf{x})p(\beta|\sigma^2)p(\sigma^2).$$

where,

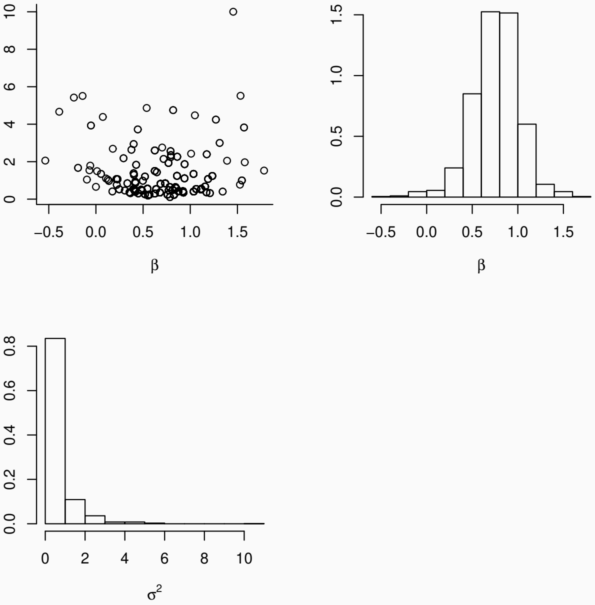
$$
\begin{aligned}
p(\mathbf{y}|\beta,\sigma^2,\mathbf{x}) &\propto \sigma^{-n}\exp(-0.5(\mathbf{y}-\beta\mathbf{x})'(\mathbf{y}-\beta\mathbf{x})/\sigma^2) \\
p(\beta|\sigma^2) &\propto \sigma^{-1}\exp(-0.5\beta^2/\sigma^2)
\end{aligned}
$$

Using the prior of (β, σ^2) as auxiliary distribution.

1. Generate values $\sigma_1^{-2}, \dots, \sigma_N^{-2} \sim \text{Gamma}(0.1, 0.1)$.
2. Set $\sigma_j^2 = 1/\sigma_j^{-2}$, $j = 1, \dots, N$.
3. Generate values $\beta_j | \sigma_j^2 \sim N(0, \sigma_j^2)$, $j = 1, \dots, N$.
4. Compute the weights $w_j = p(\mathbf{y} | \beta_j, \sigma_j^2, \mathbf{x})$, $j = 1, \dots, N$.
5. Compute the normalized weights, $w_j^* = w_j / \sum_{j=1}^{N} w_j$.
6. Resample M values of (β, σ^2) with probabilities, $w_1^*, \dots, w_M^*$.

```
> resample.reg1 <- function(y,x,N,M){
+   n=length(y)
+   inv.sigma2= rgamma(n=N, 0.1, 0.1)
+   sigma2= 1/inv.sigma2
+   beta= matrix(rnorm(n=N,mean=0,sd=sqrt(sigma2)),nrow=N)
+   l = matrix(NA, nrow = N)
+   for (i in 1:N)
+     l[i]=sigma2[i]^(-n/2)*exp(-0.5*t(y-beta[i]*x)%*%(y-beta[i]
+   sw = sum(l)
+   w = l/sw
+   ind = sample(1:N, size = M, replace = TRUE, prob = w)
+   return(cbind(beta[ind],sigma2[ind]))
+ }
```

Correlogram and histograms of the resampled values of (β, σ^2) using the same data in the previous example.



Example. Simple linear regression with Student-t errors.

$$\begin{aligned} y_i &= \beta x_i + \epsilon_i,\ \epsilon_i \sim t(0, \sigma, \nu),\ i = 1, \ldots, n \\ \beta|\sigma^2 &\sim N(0, \sigma^2) \\ \sigma^{-2} &\sim Gamma(a, b) \end{aligned}$$

and ν known.

$$\begin{aligned} p(\mathbf{y}|\beta, \sigma^2, \mathbf{x}) &\propto \prod_{i=1}^{n} \sigma^{-1} \left[1 + \frac{(y_i - \beta \mathbf{x}_i)^2}{\nu \sigma^2}\right]^{-(\nu+1)/2} \\ p(\beta|\sigma^2) &\propto \sigma^{-1} \exp(-0.5\beta^2/\sigma^2) \\ p(\sigma^{-2}) &\propto (\sigma^{-2})^{a-1} \exp(-b/\sigma^2) \end{aligned}$$

Using the prior of (β, σ^2) as auxiliary distribution.

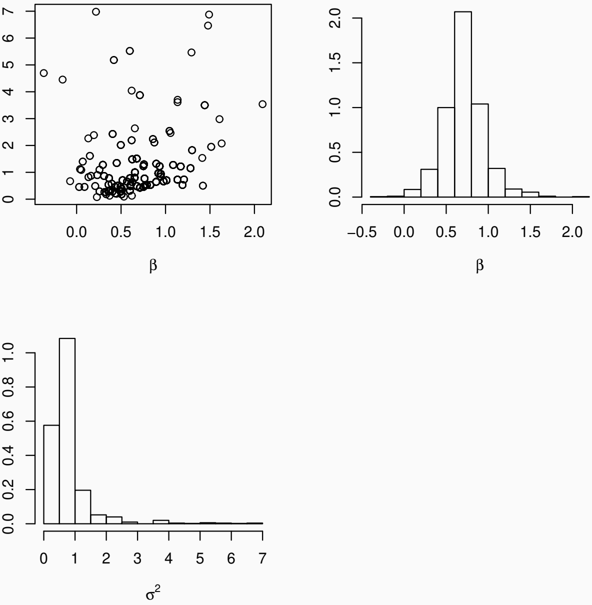
1. Generate values $\sigma_1^{-2}, \ldots, \sigma_N^{-2} \sim \text{Gamma}(a, b)$.
2. Set $\sigma_j^2 = 1/\sigma_j^{-2}$ and sample $\beta_j|\sigma_j^2 \sim N(0, \sigma_j^2)$, $j = 1, \ldots, N$.
3. Compute the weights,

$$w_j = \prod_{i=1}^{n} \sigma_j^{-1} \left[1 + \frac{(y_i - \beta_j \mathbf{x}_i)^2}{\nu \sigma_j^2}\right]^{-(\nu+1)/2}, \; j = 1, \ldots, N.$$

4. Compute the normalized weights, $w_j^* = w_j / \sum_{j=1}^{N} w_j$.
5. Resample M values of (β, σ^2) with probabilities $w_1^*, \ldots, w_M^*$.

```
> resample.regt <- function(y,x,nu,a,b,N,M){
+   n = length(y)
+   inv.sigma2 = rgamma(n=N,a,b)
+   sigma2= 1/inv.sigma2
+   beta = matrix(rnorm(n=N,mean=0,sd=sqrt(sigma2)),nrow=N)
+   L = matrix(NA, nrow = N)
+   for (j in 1:N)
+     L[j]=-(n/2)*log(sigma2[j])-((nu+2)/2)*sum(log(1+(y-beta[j]
+   l = exp(L)
+   sw= sum(l)
+   w = l/sw
+   ind = sample(1:N, size = M, replace = T, prob = w)
+   return(cbind(beta[ind],sigma2[ind]))
+ }
```

Correlogram and histograms of resampled values of (β, σ^2) with the same data in the previous example.



Example. Suppose the random variable Y represents count data which can present underdispersion, overdispersion, or equidispersion.

A flexible distribution to model such data is the Conway-Maxwell-Poisson (COM-Poisson) distribution for which the probability mass function is given by,

$$p(y|\mu,\nu) = \left(\frac{\mu^y}{y!}\right)^\nu \frac{1}{Z(\mu,\nu)}, \; \mu > 0, \nu > 0, y = 0, 1, \ldots$$

where,

$$Z(\mu,\nu) = \sum_{y=0}^{\infty} \left(\frac{\mu^\nu}{y!}\right)^\nu.$$

This is an intractable normalising constant, having no closed form representation.