Exercise 1 (chapter 3.5): The accept–reject method is used to simulate from a distribution $F_X(x)$ with unnormalized density $f_X^*(x)$ by using the proposal density $g_X(x)$ and majorizing constant M. However, the majorizing condition

$$f_X^*(x) \leq M g_X(x)$$

does not hold in some region of the space. Consequently, the accept–reject method does not simulate from the distribution corresponding to $f_X^*(x)$ but from another distribution. Write down the unnormalized density for the distribution that is simulated by the accept–reject method.

Solution: The accept–reject method uses the following algorithm

$$\begin{aligned} &\text{Simulate } X' \sim g_X \text{ and } U \sim \text{Unif}(0,1) \\ &\text{Compute } Y = M g_X(x')U \\ &\text{Accept } X' \text{ if } Y < f_X^*(x') \text{ and set } X = X' . \end{aligned}$$

The pair (X', Y) is uniformly distributed under the graph of Mg_X. Upon acceptance of X', the distribution of (X, Y) is uniform under the graph of $\min(f_X^*, Mg_X)$. If the majorizing condition does not hold in some region of the space, then the marginal distribution of X is described by the unnormalized density $\min(f_X^*, Mg_X)$ instead of f_X^*.

Exercise 2 (chapter 3.5): Let $\{y_i\}_{i=1}^n$ be conditionally independent observations from $\mathrm{N}\big(y_i \,|\, 0, \theta^{-1}\big)$, where $\theta > 0$ is the reciprocal of the variance parameter. The prior of θ is the half–Cauchy distribution. The density of the half-Cauchy distribution is

$$p(\theta) = \frac{2}{\pi(1+\theta^2)} \qquad \theta \geq 0 .$$

1. Find the normalized likelihood, that is, calculate the likelihood and normalize it so that it becomes a familiar density.
2. Suppose that $n = 1000$ and $\overline{y^2} = n^{-1} \sum_{i=1}^n y_i^2 = 0.96$. Draw a histogram from sample of the posterior which you obtained by using the accept–reject method and normalized likelihood as the proposal distribution.
3. It would also be feasible to use the prior as the proposal distribution, because the maximum-likelihood estimate can be found analytically and the half–Cauchy distribution can be simulated by taking the absolute value of a random number drawn from the ordinary Cauchy distribution. However, the acceptance probability would be rather low: about 3.5% as compared to 48% from the method of part 2. Can you explain why?

Solution: The likelihood is

$$p(y \mid \theta) = \prod_{i=1}^{n} \sqrt{\frac{\theta}{2\pi}} \exp\left\{-\frac{\theta}{2} y_i^2\right\} \propto \theta^{n/2} \exp\left\{-\frac{\theta}{2} \sum_{i=1}^{n} y_i^2\right\},$$

which is the kernel of the density of a $\text{Gamma}\big(n/2+1, \sum_{i=1}^{n} y_i^2/2\big)$ distribution. Since the prior density $p(\theta)$ is bounded, the least upper bound is

$$h(\theta) = \frac{p^*(y \mid \theta) p(\theta)}{p^*(y \mid \theta)} \leq \max p(\theta) \leq M\,,$$

where $p^*(y \mid \theta)$ is the normalized Gamma likelihood. The acceptance condition with $M = \max p(\theta) = 2/\pi$ is therefore

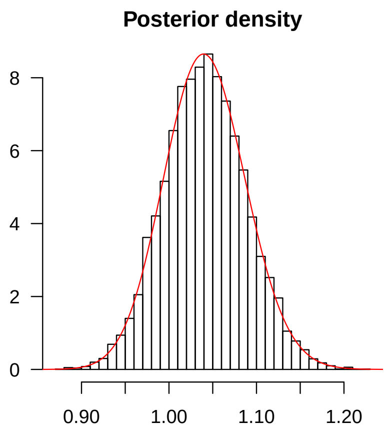
$$U \leq \frac{p^*(y \mid \theta) p(\theta)}{M p^*(y \mid \theta)} = \frac{p(\theta)}{\max p(\theta)} = \frac{1}{1+\theta^2}$$

yielding the following algorithm

$$\text{Simulate } \theta \sim \text{Gamma}\Big(n/2+1, n\overline{y^2}/2\Big) \text{ and } U \sim \text{Unif}(0,1)$$

$$\text{Accept } \theta \text{ if } U \leq \frac{1}{1+\theta^2}\,.$$

```
  n <- 1000; ySquareBar <- 0.96 ; nSamples <- 10000 ; nProposed <- 0
  x <- numeric( nSamples ); ii <- 0
  while( ii < nSamples ) {
     nProposed <- nProposed + 1
     xProposed <- rgamma( 1, 0.5 * n + 1, 0.5 * n * ySquareBar )
     if( runif( 1 ) < 1 / ( 1 + xProposed ^ 2 ) ) {
        x[ ii <- ii + 1 ] <- xProposed
     }
  }
# Estimated acceptance probability:  0.4776
```



The mode of the Gamma distribution and also the maximum-likelihood estimate of θ is

$$\hat{\theta}_{\mathrm{MLE}} = \frac{\alpha - 1}{\beta} = \frac{1}{\overline{y^2}} .$$

Using the prior as the proposal distribution, the least upper bound is

$$h(\theta) = \frac{p^*(y \,|\, \theta)p(\theta)}{p(\theta)} \leq p^*(y \,|\, \hat{\theta}_{\mathrm{MLE}}) \leq M .$$

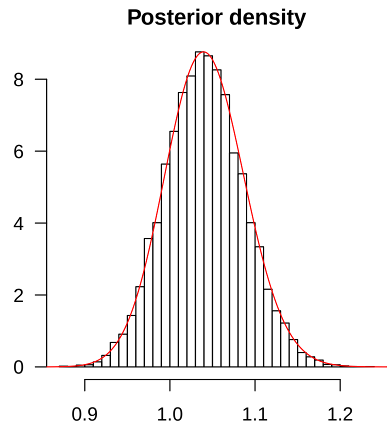
The acceptance condition with $M = p(y \,|\, \hat{\theta}_{\mathrm{MLE}})$ is therefore

$$U \leq \frac{p^*(y \,|\, \theta)p(\theta)}{Mp(\theta)} = \frac{\mathrm{Gamma}\Big(\theta \,|\, n/2 + 1, n\overline{y^2}/2\Big)}{\mathrm{Gamma}\Big(\hat{\theta}_{\mathrm{MLE}} \,|\, n/2 + 1, n\overline{y^2}/2\Big)}$$

yielding the following algorithm

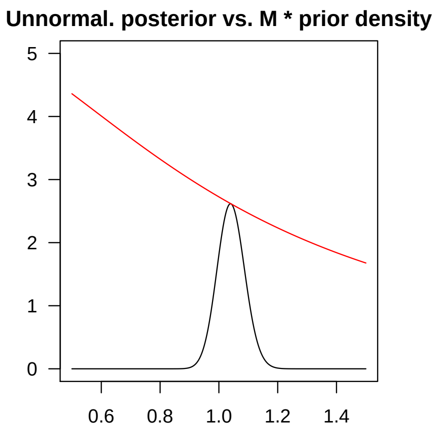
$$\begin{aligned} &\text{Simulate } \theta \sim \text{Half-Cauchy distribution and } U \sim \mathrm{Unif}(0,1) \\ &\text{Accept } \theta \text{ if } U \leq \frac{\mathrm{Gamma}\Big(\theta \,|\, n/2 + 1, n\overline{y^2}/2\Big)}{\mathrm{Gamma}\Big(\hat{\theta}_{\mathrm{MLE}} \,|\, n/2 + 1, n\overline{y^2}/2\Big)} . \end{aligned}$$

```
   normLikehood <- function( x ) { dgamma( x, 0.5 * n + 1, 0.5 * n * ySquareBar ) }
   ySquareBar <- 0.96 ; mle <- 1 / ySquareBar ;
   n <- 1000; nSamples <- 10000 ; nProposed <- 0
   x <- numeric( nSamples ); ii <- 0
   while( ii < nSamples ) {
      nProposed <- nProposed + 1
      xProposed <- abs( rcauchy( 1 ) )
      if( runif( 1 ) < normLikehood( xProposed  ) / normLikehood( mle ) ) {
         x[ ii <- ii + 1 ] <- xProposed
      }
   }
# Estimated acceptance probability:  0.0352
```



The acceptance probability is rather low, because the prior is flat and the unnormalized posterior density highly peaked around the maximum likelihood estimate. This leads to a large amount of rejections and a low acceptance probability.

```
n <- 1000 ; ySquareBar <- 0.96 ; mle <- 1 / ySquareBar
grid <- seq( 0.5, 1.5, by = 0.001 )
prior <- 2 / pi / ( 1 + grid ^ 2 )
normLikehood <- dgamma( grid, 0.5 * n + 1, 0.5 * n * ySquareBar )
M <- dgamma( mle, 0.5 * n + 1, 0.5 * n * ySquareBar )
par( mar = c( 3, 3, 2, 3 ), las = 1 )
plot( grid, prior * normLikehood, type = 'l', ylim = c( 0, 5 ),
   main = 'Unnormal. posterior vs. M * prior density' )
lines( grid, M * prior, col = 'red' )
```



Exercise 3 (chapter 3.8): Let the random vector $\boldsymbol{X} = [X_1, X_2, \ldots, X_n]^T$ follow a d-dimensional multivariate Student's–t distribution $\mathrm{St}_d(\boldsymbol{x} \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ with location parameter $\boldsymbol{\mu}$, symmetric and positive definite $d \times d$ scale matrix $\boldsymbol{\Sigma}$ and $\nu > 0$ degrees of freedom. The density of the multivariate Student's–t distribution is

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{\Gamma((\nu+d)/2)}{\nu^{d/2}\pi^{d/2}\Gamma(\nu/2)\det(\boldsymbol{\Sigma})^{1/2}} \left[1 + \frac{1}{\nu}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]^{-(\nu+d)/2}.$$

Suppose that the factorization $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ is available. Design an algorithm without using any other matrix factorizations and in which random numbers are only drawn Gamma and (univariate) standard normal distributions.

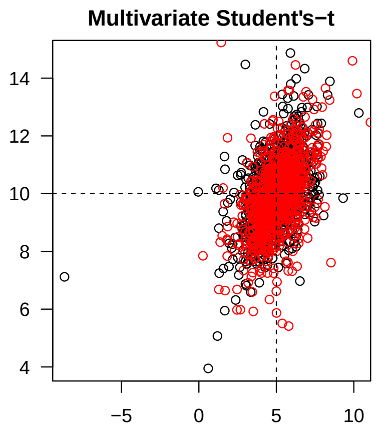
Solution: A pair $(\boldsymbol{X}, Y)$ with $\boldsymbol{X} = [X_1, \dots, X_p]^{\mathrm{T}}$ being a p-dimensional random vector follows a multivariate Normal–Gamma distribution with location parameter $\boldsymbol{\mu}$, symmetric and positive definite $d \times d$ scale matrix $\boldsymbol{\Sigma}$ and $\nu > 0$ degrees of freedom, if $\boldsymbol{X} \mid Y \sim \mathrm{MVN}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}/y)$ and $Y \sim \mathrm{Gamma}(\nu/2, \nu/2)$. The marginal distribution of $\boldsymbol{X}$ is a multivariate Student's–t distribution because

$$
\begin{aligned}
f_{\boldsymbol{X}}(\boldsymbol{x}) &= \int f_{\boldsymbol{X} \mid Y}(\boldsymbol{x} \mid y) f_Y(y) \,\mathrm{d}y \\
&\propto \int y^{(\nu+d)/2-1} \exp\left\{ -\left[\frac{\nu}{2} + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu}) \right] y \right\} \mathrm{d}y \\
&\propto \left[1 + \frac{1}{\nu}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu}) \right]^{-(\nu+d)/2},
\end{aligned}
$$

which is the kernel of a multivariate $\mathrm{St}_d(\boldsymbol{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ distribution. Simulation from the multivariate Student's–t distribution can be implemented by using the composition rule

$$
\begin{aligned}
\text{Simulate } y &\sim \mathrm{Gamma}(\nu/2, \nu/2) \\
\text{Simulate } [Z_1, \dots, Z_p] &\overset{\text{Ind}}{\sim} \mathrm{Normal}(0, 1) \\
\text{Set } [X_1, \dots, X_d]^{\mathrm{T}} &= \mu + \boldsymbol{A}\boldsymbol{Z}/\sqrt{Y}\,.
\end{aligned}
$$

```
mu <- c( 5, 10 ) ; Sigma <- matrix( c( 1, 0.5, 0.5, 1 ), 2, 2 ) ; nu <- 6
A <- t( chol( Sigma ) )
nSamples <- 1000
x1 <- mvtnorm::rmvt( nSamples, Sigma, nu, mu )
x2 <- matrix( 0, nSamples, 2 )
for( ii in seq_len( nSamples ) ) {
   y <- rgamma( 1, 0.5 * nu, 0.5 * nu )
   x2[ ii, ] <- mu + A %*% rnorm( 2 ) / sqrt( y )
}
```



Exercise 4 (chapter 5.4): Instead of the Inverse Gamma distribution, many authors use the scaled inverse chi-square distribution for a variance parameter σ^2 of a Normal distribution. See for instance the book by Gelman et al. with the title "Bayesian Data Analysis". The authors define the scaled inverse chi–square distribution Inv–$\chi^2\big(\sigma^2 \mid \nu, \sigma_0^2\big)$ with scale parameter $\sigma_0^2 > 0$ and degrees of freedom $\nu > 0$ as

$$Y = \frac{\sigma_0^2 \nu}{X} \qquad \text{when } X \sim \chi^2_\nu .$$

The density of the scaled inverse chi-square distribution is

$$f_Y(y) = \frac{\big(\sigma_0^2\nu/2\big)^{\nu/2}}{\Gamma(\nu/2)} y^{-\nu/2-1} \exp\left\{-\frac{\sigma_0^2\nu}{2y}\right\}.$$

1. Derive the density of Y from X.
2. If the variance parameter σ^2 follows a scaled inverse chi–square distribution Inv–$\chi^2\big(\sigma^2 \mid \sigma_0^2, \nu\big)$, then the precision parameter $\psi = 1/\sigma^2$ follows a Gamma distribution. What are its (hyper)parameters? Remember that if $X \sim \chi^2_\nu$ and $a > 0$, then X/a has a certain Gamma distribution.

Solution: The density of $Y = 1/X$ is

$$\begin{aligned} f_Y(y) = f_X(x)\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| &= \frac{1}{2^{\nu/2}\Gamma(\nu/2)}\left(\frac{\sigma_0^2\nu}{y}\right)^{\nu/2-1} \exp\left\{-\frac{\sigma_0^2\nu}{2y}\right\}\frac{\sigma_0^2\nu}{y^2} \\ &= \frac{\big(\sigma_0^2\nu/2\big)^{\nu/2}}{\Gamma(\nu/2)} y^{-\nu/2-1} \exp\left\{-\frac{\sigma_0^2\nu}{2y}\right\} \qquad y > 0 , \end{aligned}$$

which is the density of an Inverse–Gamma$\big(\nu/2, \sigma_0^2\nu/2\big)$ distribution. The density of $\psi = 1/\sigma^2$ is

$$\begin{aligned} p(\psi) = p\big(\sigma^2\big)\left|\frac{\mathrm{d}\sigma^2}{\mathrm{d}\psi}\right| &= \frac{\big(\sigma_0^2\nu/2\big)^{\nu/2}}{\Gamma(\nu/2)}\left(\frac{1}{\psi}\right)^{-\nu/2-1} \exp\left\{-\frac{\sigma_0^2\nu}{2}\psi\right\}\frac{1}{\psi^2} \\ &= \frac{\big(\sigma_0^2\nu/2\big)^{\nu/2}}{\Gamma(\nu/2)} \psi^{\nu/2-1} \exp\left\{-\frac{\sigma_0^2\nu}{2}\psi\right\} \qquad \psi > 0 , \end{aligned}$$

which is the density of a Gamma$\big(\nu/2, \sigma_0^2\nu/2\big)$ distribution.

Exercise 5 (chapter 5.5): Consider the simple linear regression model, where

$$\begin{aligned} p(\boldsymbol{y} \mid \boldsymbol{\beta}, \tau) &= \mathrm{MVN}_n\big(\boldsymbol{y} \mid \boldsymbol{X}\boldsymbol{\beta}, \tau^{-1}\boldsymbol{I}\big) \\ p(\boldsymbol{\beta}, \tau) &= \mathrm{MVN}_p(\boldsymbol{\beta} \mid \boldsymbol{\mu}, \boldsymbol{Q}^{-1})\mathrm{Gamma}(\tau \mid a, b) . \end{aligned}$$

Here $\boldsymbol{X}$ is a known $n \times p$ matrix of explanatory variables, $\tau > 0$ a scalar precision parameter of the error distribution and $\boldsymbol{\beta}$ a coefficient vector of length p. Note that $\boldsymbol{\beta}$ and τ are assumed to be independent in their joint prior distribution.

1. Write down the joint density $p(\boldsymbol{y}, \boldsymbol{\beta}, \tau)$ including all normalizing constants. Notice that $\det(c\boldsymbol{A}) = c^n\det(\boldsymbol{A})$ for a scalar c and $n \times n$ matrix $\boldsymbol{A}$.

2. Derive the full conditional distribution $p(\boldsymbol{\beta} \,|\, \tau, \boldsymbol{y})$ either from first principles or by using the theory in chapter 5.5.2.
3. Derive the full conditional distribution $p(\tau \,|\, \boldsymbol{\beta}, \boldsymbol{y})$ by finding a useful formula when starting from first principles or alternatively by extending the theory of chapter 5.4.2 to the present situation.

Solution: The joint density is proportional to

$$\begin{aligned} p(\boldsymbol{y}, \boldsymbol{\beta}, \tau) \propto (2\pi)^{-n/2} \tau^{n/2} \exp\Big\{-\frac{\tau}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\Big\} \times \\ (2\pi)^{-p/2} \det(\boldsymbol{Q})^{1/2} \exp\Big\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{Q}(\boldsymbol{\beta} - \boldsymbol{\mu})\Big\} \times \\ \frac{b^a}{\Gamma(a)} \tau^{a-1} \exp\{-b\tau\} \end{aligned}$$

The full conditional density of $\boldsymbol{\beta}$ is

$$p(\boldsymbol{\beta} \,|\, \tau, \boldsymbol{y}) \propto \exp\Big\{-\frac{1}{2}\Big[\tau(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) + (\boldsymbol{\beta} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{Q}(\boldsymbol{\beta} - \boldsymbol{\mu})\Big]\Big\},$$

where the quadratic form is equal to

$$\begin{aligned} q &= \tau(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) + (\boldsymbol{\beta} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{Q}(\boldsymbol{\beta} - \boldsymbol{\mu}) \\ &= \boldsymbol{\beta}^{\mathrm{T}} \tau \boldsymbol{X}^{\mathrm{T}} \boldsymbol{X}\boldsymbol{\beta} - 2\boldsymbol{\beta}^{\mathrm{T}} \tau \boldsymbol{X}^{\mathrm{T}} \boldsymbol{y} + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Q}\boldsymbol{\beta} - 2\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Q}\boldsymbol{\mu} + \text{const.} \\ &= \boldsymbol{\beta}^{\mathrm{T}} \tilde{\boldsymbol{Q}}\boldsymbol{\beta} - 2\boldsymbol{\beta}^{\mathrm{T}} \tilde{\boldsymbol{Q}}\tilde{\boldsymbol{\mu}} + \text{const.}, \end{aligned}$$

with $\tilde{\boldsymbol{Q}} = \boldsymbol{Q} + \tau \boldsymbol{X}^{\mathrm{T}} \boldsymbol{X}$ and $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{Q}}^{-1}(\boldsymbol{Q}\boldsymbol{\mu} + \tau \boldsymbol{X}^{\mathrm{T}} \boldsymbol{y})$. The full conditional of $\boldsymbol{\beta}$ is therefore a $\mathrm{MVN}_p\Big(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{Q}}\Big)$ distribution. The full conditional density of τ is

$$p(\tau \,|\, \boldsymbol{\beta}, \boldsymbol{y}) \propto \tau^{a+n/2-1} \exp\Big\{-\Big[b + \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\Big]\tau\Big\},$$

which is the kernel of a $\mathrm{Gamma}\Big(a + n/2, b + \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\Big)$ distribution.