$$
\begin{aligned}
&= \prod_{i=1}^{n} f(y_i|\beta, \sigma^2) \\
&\leq \prod_{i=1}^{n} f(y_i|\hat{\beta}_{MLE}, \hat{\sigma^2}_{MLE}) := M \,.
\end{aligned}
$$

*So an accept-reject sampler is possible to implement. However, we note that the dimensionality of the problem will certainly impede efficiency.*

If we are able to implement an accept-reject algorithm, then the inference mechanism is still the same. You obtain samples $X_1, X_2, \dots X_T$ from $\pi$ and then estimate the posterior mean and upper and lower quantiles.

## 16.4 Linchpin variable Accept-Reject

As we have discussed plenty of times now, it is difficult to implement AR when the target is high-dimensional or when the upper bound is hard to get. In the first case, a *linchpin variable* trick can be very useful. Suppose the target density is

$$\pi(x, y) \,.$$

Then as we have done multiple times before, we can split the joint distribution as the product of conditional times marginal. That is

$$\pi(x, y) = \pi(x|y)\, \pi(y) \,.$$

If $X|Y$ is known in closed form and we can sample from it, then we may try and get marginal samples from $y$. This is beneficial since the dimension of $y$ is smaller than $(x, y)$. So the algorithm would be

- Generate $Y \sim \pi(y)$
- Generate $X \sim X|Y$
- Output $(X, Y)$.

The variable $Y$ is called the linchpin variable, and $\pi(Y)$ is the target distribution. Let's see an example

**Example 41** (Bayesian linear regression)**.** *Recall the Bayesian linear regression model. The likelihood is*

$$y_1, \dots, y_n \mid \beta, \sigma^2 \overset{iid}{\sim} N(X_i\beta, \sigma^2) \,.$$

*The parameters of interest are* $\beta$ *and* $\sigma^2$*, just like regular MLE. We assume priors:*

$$\beta \sim N_p(0, \sigma^2) \quad \text{and} \quad \sigma^2 \sim IG(a, b)\,,$$

*We know that the posterior distribution is*

$$\pi(\beta, \sigma^2|y) = \left(\sigma^2\right)^{-n/2-p/2-a-1} \exp\left\{-\frac{(y - X\beta)^T(y - X\beta)}{2\sigma^2} - \frac{\beta^T\beta}{2\sigma^2} - \frac{b}{\sigma^2}\right\}$$

*First, note that we prefer* $\sigma^2$ *to be the linchpin variable since it is univariate, and* $\beta$ *is p-variate. So we need to find the distribution* $\beta|\sigma^2$ *and the marginal distribution of* $\sigma^2$*. Let* $A = (X^TX + I)$*.*

$$
\begin{aligned}
&\int \pi(\beta, \sigma^2|y)d\beta \\
&= \int \left(\sigma^2\right)^{-n/2-p/2-a-1} \exp\left\{-\frac{y^Ty - 2\beta^TX^Ty + \beta^TX^TX\beta}{2\sigma^2} - \frac{\beta^T\beta}{2\sigma^2} - \frac{b}{\sigma^2}\right\} d\beta \\
&= \left(\sigma^2\right)^{-n/2-p/2-a-1} \exp\left\{-\frac{y^Ty}{2\sigma^2} - \frac{b}{\sigma^2}\right\} \int \exp\left\{-\frac{\beta^TX^TX\beta - 2\beta^TX^Ty}{2\sigma^2} - \frac{\beta^T\beta}{2\sigma^2}\right\} d\beta \\
&= \left(\sigma^2\right)^{-n/2-p/2-a-1} \exp\left\{-\frac{y^Ty}{2\sigma^2} - \frac{b}{\sigma^2}\right\} \int \exp\left\{-\frac{\beta^T(X^TX + I)\beta - 2\beta^TX^Ty}{2\sigma^2}\right\} d\beta \\
&= \left(\sigma^2\right)^{-n/2-p/2-a-1} \exp\left\{-\frac{y^Ty}{2\sigma^2} - \frac{b}{\sigma^2}\right\} \int \exp\left\{ - \frac{\beta^TA\beta - 2\beta^TAA^{-1}X^Ty}{2\sigma^2}\right. \\
&\quad \left. - \frac{(A^{-1}X^Ty)^TA(A^{-1}X^Ty)}{2\sigma^2} + \frac{(A^{-1}X^Ty)^TA(A^{-1}X^Ty)}{2\sigma^2}\right\} d\beta \\
&= \left(\sigma^2\right)^{-n/2-p/2-a-1} \exp\left\{-\frac{y^Ty}{2\sigma^2} - \frac{b}{\sigma^2} + \frac{y^TXA^{-1}AA^{-1}X^Ty}{2\sigma^2}\right\} \\
&= \quad \times \int \exp\left\{-\frac{\beta^TA\beta - 2\beta^TAA^{-1}X^Ty + y^TXA^{-1}AA^{-1}X^Ty}{2\sigma^2}\right\} \\
&= \left(\sigma^2\right)^{-n/2-p/2-a-1} \exp\left\{-\frac{y^Ty}{2\sigma^2} - \frac{b}{\sigma^2} + \frac{y^TXA^{-1}X^Ty}{2\sigma^2}\right\} \int \exp\left\{-\frac{(\beta - A^{-1}X^Ty)^TA(\beta - A^{-1}X^Ty)}{2\sigma^2}\right\}
\end{aligned}
$$

*So* $\beta|\sigma^2$ *is a multivariate normal distribution*

$$\beta|\sigma^2 \sim N_p\left(A^{-1}X^Ty, \sigma^2A^{-1}\right)\,,$$

*and the integral integrates to a known constant.*

$$\int \pi(\beta, \sigma^2|y)d\beta) \propto \left(\sigma^2\right)^{-n/2-p/2-a-1} \exp\left\{-\frac{y^T(I - XA^{-1}X^T)y}{2\sigma^2} - \frac{b}{\sigma^2}\right\} \cdot \left(\sigma^2\right)^{p/2} \det(A)^{p/2}$$
$$\propto \left(\sigma^2\right)^{-n/2-a-1} \exp\left\{-\frac{y^T(I - XA^{-1}X^T)y}{2\sigma^2} - \frac{b}{\sigma^2}\right\}.$$

*So the marginal distribution of $\sigma^2|y$ is*

$$\sigma^2|y \sim IG\left(\frac{n}{2} + a, \frac{y^T(I - XA^{-1}X^T)y}{2} + b\right).$$

*Thus, we can estimate marginal means and quantiles of $\sigma^2$. But to estimate $\beta$, we do the following*

1. *Generate $\sigma^2 \sim IG$ as indicated above*
2. *Generate $\beta|\sigma^2 \sim$ Normal distribution as indicated above.*
3. *$(\beta, \sigma^2)$ is one draw from the posterior. Repeat for many draws, and estimate posterior mean and quantiles.*

**Example 42** (Weibull - Gamma)**.** *Consider a Bayesian reliability model, where the observed failure times of a lamp is distributed a s*

$$T_1, \ldots, T_n \mid \lambda, \beta \sim \textit{Weibull}(\lambda, \beta).$$

*We assume priors*

$$\beta \sim \textit{Gamma}(a_0, b_0) \qquad \textit{and } \lambda \sim \textit{Gamma}(a_1, b_1).$$

*We are interested in the posterior distribution of $(\lambda, \beta)$.*

$$\pi(\beta, \lambda|t) \propto \pi(\beta)\pi(\lambda)\prod_{i=1}^{n} f(T_i \mid \beta, \lambda)$$
$$= \beta^{a_0-1}\exp\{-b_0\beta\} \cdot \lambda^{a_1-1}\exp\{-b_1\lambda\}\prod_{i=1}^{n}\lambda\beta\,(t_i)^{\beta-1}\exp\left\{-\lambda t_i^{\beta}\right\}$$
$$= \beta^{n+a_0-1}\exp\{-b_0\beta)\}\left[\prod_{i=1}^{n} t_i\right]^{\beta-1}\lambda^{n+a_1-1}\exp\left\{-\lambda(b_1 + \sum_i t_i^{\beta})\right\}.$$

*Note that $\lambda|t \sim \textit{Gamma}(n + a_1, (b_1 + \sum_i t_i^{\beta}))$. So*

$$
\begin{aligned}
&\int \pi(\beta, \lambda|t) d\lambda \\
&\propto \beta^{n+a_0-1} \exp\{-b_0\beta)\} \left[\prod_{i=1}^{n} t_i\right]^{\beta-1} \int \lambda^{n+a_1-1} \exp\left\{-\lambda \sum_i t_i^{\beta}\right\} d\lambda \\
&\propto \beta^{n+a_0-1} \exp\{-b_0\beta)\} \left[\prod_{i=1}^{n} t_i\right]^{\beta-1} \\
&\quad \times \int \frac{\Gamma(n+a_1)}{(b_1+\sum_i t_i^{\beta})^{n+a_1}} \frac{(b_1+\sum_i t_i^{\beta})^{n+a_1}}{\Gamma(n+a_1)} \lambda^{n+a_1-1} \exp\left\{-\lambda(b_1+\sum_i t_i^{\beta})\right\} d\lambda \\
&\propto \beta^{n+a_0-1} \exp\{-b_0\beta)\} \left[\prod_{i=1}^{n} t_i\right]^{\beta-1} (b_1+\sum_i t_i^{\beta})^{-(n+a_1)}
\end{aligned}
$$

*So, we have that the marginal distribution of* $\beta \mid T$ *is*

$$
f(\beta|T) \propto \beta^{n+\alpha_\beta-1} e^{-\beta\theta_\beta} \left[\prod_{i=1}^{n} t_i^{\beta-1}\right] \left(\theta_\lambda + \sum t_i^{\beta}\right)^{-(n+\alpha_\lambda)} .
$$

*We want to find an appropriate proposal distribution to implement an accept-reject sampler. First we will do some algebra tricks: Note that*

$$
\left[\prod t_i\right]^{\beta-1} = \exp\left\{(\beta-1)\log\left(\prod t_i\right)\right\} = \exp\left\{\beta \sum \log t_i\right\} \exp\left\{-\sum \log t_i\right\} .
$$

*Also,*

$$
\begin{aligned}
&\theta_\lambda + \sum t_i^{\beta} \geq \sum t_i^{\beta} \geq n \min_i\{t_i^{\beta}\} = n\left(\min_i\{t_i\}\right)^{\beta} := n\, m_t^{\beta} \\
\Rightarrow &\left(\theta_\lambda + \sum t_i^{\beta}\right)^{-(n+\alpha_\lambda)} \leq n^{-(n+\alpha_\lambda)} m_t^{-\beta(n+\alpha_\lambda)} = n^{-(n+\alpha_\lambda)} \exp\{-\beta(n+\alpha_\lambda)\log m_t\} .
\end{aligned}
$$

*Using these two tricks, we get,*

$$
\begin{aligned}
f(\beta|T) &\propto \beta^{n+\alpha_\beta-1} e^{-\beta\theta_\beta} \left[\prod_{i=1}^{n} t_i^{\beta-1}\right] \left(\theta_\lambda + \sum t_i^{\beta}\right)^{-(n+\alpha_\lambda)} \\
&= \beta^{n+\alpha_\beta-1} e^{-\beta\theta_\beta} \exp\left\{\beta \sum \log t_i\right\} \exp\left\{-\sum \log t_i\right\} \left(\theta_\lambda + \sum t_i^{\beta}\right)^{-(n+\alpha_\lambda)} \\
&\leq \beta^{n+\alpha_\beta-1} e^{-\beta\theta_\beta} \exp\left\{\beta \sum \log t_i\right\} \exp\left\{-\sum \log t_i\right\} n^{-(n+\alpha_\lambda)} \exp\{-\beta(n+\alpha_\lambda)\log m_t\}
\end{aligned}
$$

$$= \exp\left\{-\sum \log t_i\right\} n^{-(n+\alpha_\lambda)}\ \beta^{n+\alpha_\beta-1} \exp\left\{-\beta\left(\theta_\beta + (n+\alpha_\lambda)\log m_t - \sum \log t_i\right)\right\}.$$

*As long as* $\theta_\beta + (n+\alpha_\lambda)\log m_t - \sum \log t_i > 0$, *the right hand side above is a proper density of a Gamma distribution, which can be your proposal distribution. So that*

$$\frac{\tilde{\pi}(\beta|t)}{\tilde{g}(\beta)} \le \exp\left\{-\sum \log t_i\right\}.$$

**Example 43** (Bayesian hierarchical models)**.** *Consider a Bayesian hierarchical model*

$$Y_i \mid \theta_i \overset{ind}{\sim} N(\theta_i, a_0)$$

*Each observation has it's own mean*

$$\theta_i \mid \mu, \lambda \sim N(\mu, \lambda)$$

*Such models are used when each observation can potentially have a completely different mean. This model has been useful for baseball data batting averages. The priors are*

$$\lambda \sim IG(b_0, c_0) \quad \text{and } f(\mu) \propto 1$$

*The posterior distribution is*

$$\begin{aligned}
\pi(\theta, \mu, \lambda|y) &\propto \pi(\theta, \mu, \lambda) \prod_{i=1}^n f(y_i \mid \theta_i, \mu, \lambda) \\
&= \pi(\mu)\,\pi(\lambda) \prod_{i=1}^n f(y_i \mid \theta_i, \mu, \lambda)\,\pi(\theta_i|\mu, \lambda) \\
&= \lambda^{-n/2} \exp\left\{-\frac{\sum_{i=1}^n (\theta_i - \mu)^2}{2\lambda} - \frac{\sum_{i=1}^n (y_i - \theta_i)^2}{2a_0}\right\} \lambda^{b_0-1} e^{-c_0/\lambda}.
\end{aligned}$$

*We will use a linchpin variable sampler with linchpin variable* $\lambda$. *Specifically, we will decompose*

$$\pi(\theta, \mu, \lambda|y) = \pi(\theta, \mu|\lambda, y)\,\pi(\lambda|y).$$

*Do we know* $\pi(\theta, \mu|\lambda, y)$*? Well we will decompose*

$$\pi(\theta, \mu|\lambda, y) = \pi(\theta|\mu, \lambda, y)\pi(\mu|\lambda, y).$$

*Similar to the Bayesian linear regression example, we can obtain that*

$$\theta_i|\mu, \lambda, y \overset{ind}{\sim} N\left(\frac{\lambda y_i + a\mu}{\lambda + a}, \frac{a\lambda}{\lambda + a}\right)$$
$$\mu|\lambda, y \sim N\left(\frac{1}{n}\sum_{i=1}^{n} y_i, \frac{\lambda + a}{n}\right)$$

*We also get*

$$\pi(\lambda|y) \propto \frac{1}{\lambda^{b_0+1}(\lambda + a_0)^{(n-1)/2}} \exp\left\{-\frac{c_0}{\lambda} - \frac{s^2}{2(\lambda + a_0)}\right\}$$

*where*

$$s^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 .$$

*We need to implement accept-reject to sample from $\pi(\lambda|y)$. Consider proposal distribution to be $IG(b_0, c_0)$. Then*

$$\begin{aligned}\frac{\tilde{\pi}(\lambda|y)}{g(\lambda)} &= \frac{1}{g(\lambda)\lambda^{b_0+1}(\lambda + a_0)^{(n-1)/2}} \exp\left\{-\frac{c_0}{\lambda} - \frac{s^2}{2(\lambda + a_0)}\right\} \\ &= \frac{1}{(\lambda + a_0)^{(n-1)/2}} \exp\left\{-\frac{s^2}{2(\lambda + a_0)}\right\}\end{aligned}$$

*The maximum for the above ratio occurs at*

$$\hat{\lambda} = \max\left\{0, \frac{s^2}{n-1} - a_0\right\}$$

*So plug it back into $\tilde{\pi}/g$ and obtain $M$, and implement accept-reject.*

# 17 Importance Sampling

## 17.1 Basic/simple importance sampling

Suppose we are interested in estimating the expectation of a function $h$ with respect to a distribution with density $\pi$ (known fully). That is, we want to estimate

$$\theta = \int_{\mathsf{X}} h(x)\pi(x)\, dx .$$