STA 250/MTH 342 Intro to Mathematical Statistics
Lab Session 6 / March 02, 2015 / Handout

In this session we focus on the Gibbs sampling algorithm. The motivation is to draw random samples from some complicated multivariate distributions.

See: `https://stat.duke.edu/courses/Spring15/sta250/labs/` for links to source code and data. Submit lab solutions via email to: `sta250@stat.duke.edu`. Any plots should be included in postscript form as attachments. The email subject must be "STA250 ..." with "..." replaced by your name.

1: Gibbs sampling.

Let $f(w, v)$ be a bivariate pdf from which we want to draw a sample $(w_1, v_1), \cdots, (w_M, v_M)$. Let $f_1(w \mid v)$ and $f_2(v \mid w)$ denote the two conditional pdfs associated with $f(w, v)$. A Gibbs sampling draws these samples iteratively as follows:

- Start with an arbitrary (w_0, v_0) at which $f(w_0, v_0) > 0$.
- For $i = 1, \cdots, M$ iterate the following
  - sample w_i from the conditional pdf $f_1(w \mid v = v_{i-1})$ [using v_{i-1} from previous step]
  - sample v_i from the conditional pdf $f_2(v \mid w = w_i)$ [using the new w_i].

This algorithm is fairly easy to implement provided it is easy to sample from $f_1(w \mid v)$ and $f_2(v \mid w)$. Some advanced probability theory shows that the samples we draw (perhaps after discarding some initial draws) well represent the bivariate pdf $f(w, v)$. Consequently (and maybe more usefully), the samples $w_1, \cdots, w_M$ well represent the marginal pdf $f_1(w)$ and the samples $v_1, \cdots, v_M$ well represent the marginal pdf $f_2(v)$.

Example[Bivariate normal]. Consider the bivariate pdf

$$f(w, v) = \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2(1-\rho^2)}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(w-\mu_1)^2}{\sigma_1^2} + \frac{(v-\mu_2)^2}{\sigma_2^2} - 2\frac{\rho(w-\mu_1)(v-\mu_2)}{\sigma_1\sigma_2}\right]\right\},$$

defined over $-\infty < w, v < \infty$. This is known as the bivariate normal pdf with means $\mu_1, \mu_2 \in (-\infty, \infty)$, variances $\sigma_1^2, \sigma_2^2 \in (0, \infty)$ and correlation $\rho \in (-1, 1)$. The following facts are easy to derive:

- $f_1(w) = \mathsf{No}(\mu_1, \sigma_1^2)$,
- $f_2(v) = \mathsf{No}(\mu_2, \sigma_2^2)$.
- $f_1(w \mid v) = \mathsf{No}(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(v - \mu_2), \sigma_1^2(1-\rho^2))$.
- $f_2(v \mid w) = \mathsf{No}(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(w - \mu_1), \sigma_2^2(1-\rho^2))$.

We shall use Gibbs sampler to draw samples of (w, v) and then compare their empirical joint/marginal distributions with $f(w, v)$, $f_1(w)$ and $f_2(v)$. The code below implements the Gibbs sampler for a total $B + M$ iterations, where B is the number of initial samples to be discarded (they are referred to as the burn-in samples). We start by setting $B = 0$ and will later consider some actual discarding. The first example uses $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$ and $\rho = 0.5$.

Here in Figure 1, we compare two sample sets: one is drawn directly from the joint distribution, the other is drawn by Gibbs sampling with $B = 0$. We see that the point clouds are similar. In Figure 2, we illustrate some steps of the sampling process. On the website shown above, we give a movie of the process.

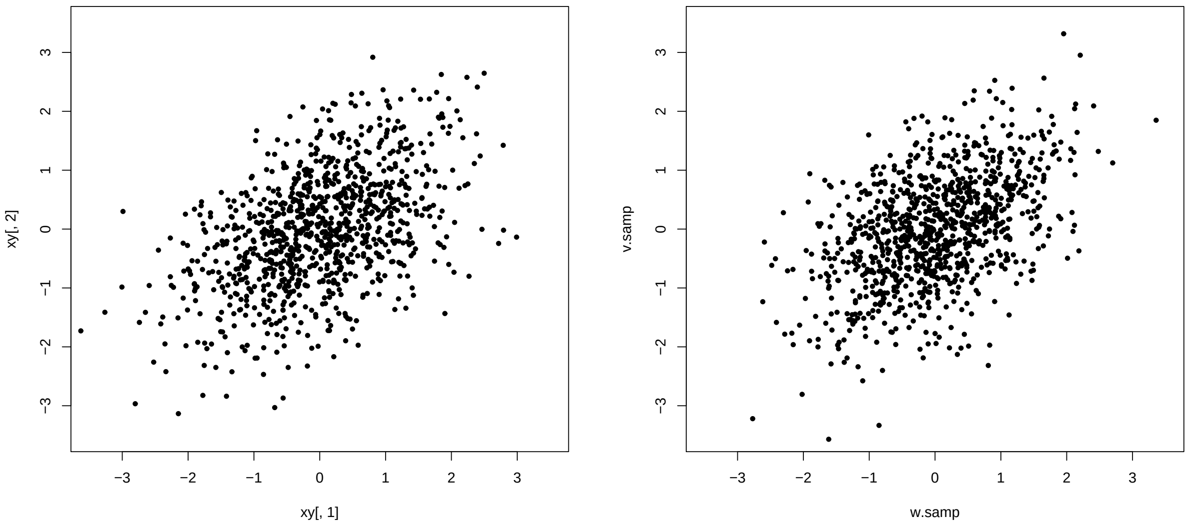

Figure 1: The comparison of direct sampling and Gibbs sampling. Left: i.i.d. samples directly drawn from the joint distribution; Right: output of Gibbs sampling algorithm (prepared by the script file "`gibbs.samp.R`").

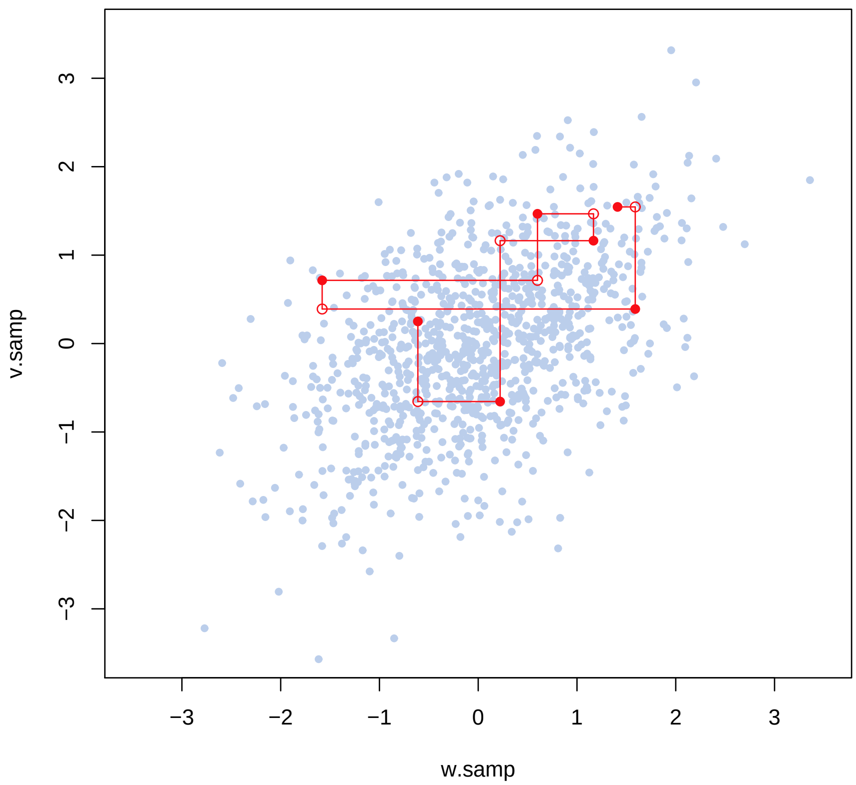

Figure 2: Some steps of the Gibbs sampling process. Prepared by the script file “`gibbs.samp.R`”.

```
# pdf parameters
mu.1 <- 0; mu.2 <- 0; sigma.1 <- 1; sigma.2 <- 1; rho <- 0.5

# inital values
w <- 0
v <- 0

# prepare vector to retain samples
M <- 1e3
B <- 0 ## number of initial samples to discard
w.samp <- rep(NA, B + M)
v.samp <- rep(NA, B + M)

# run Gibbs sampler
for(i in 1:(B + M)){
  w <- rnorm(1, mu.1 + rho * sigma.1 / sigma.2 * (v - mu.2), sigma.2 * sqrt(1 - rho^2))
  v <- rnorm(1, mu.2 + rho * sigma.2 / sigma.1 * (w - mu.1), sigma.1 * sqrt(1 - rho^2))
  w.samp[i] <- w
```

```
  v.samp[i] <- v
}

# discard the initial part
w.samp <- w.samp[B + 1:M]
v.samp <- v.samp[B + 1:M]
```

Next we visually compare the samples we obtained against the pdf $f(w, v)$. For this example, we could evaluate the pdf, or more usefully, its logarithm, on a grid of values over the range of w and v. It is convenient to write the log-pdf up to a constant

$$\log f(w, v) = const - \frac{1}{2(1-\rho^2)} \left\{ \frac{(w-\mu_1)^2}{\sigma_1^2} + \frac{(v-\mu_2)^2}{\sigma_2^2} - 2\frac{\rho(w-\mu_1)(v-\mu_2)}{\sigma_1\sigma_2} \right\}.$$

While plotting, we would not care about the constant. In fact, we would shift the values by a constant amount so that the maximum equals zero. These are just techniques to improve plots. The code below gives the details. The plot obtained is at Figure 3.

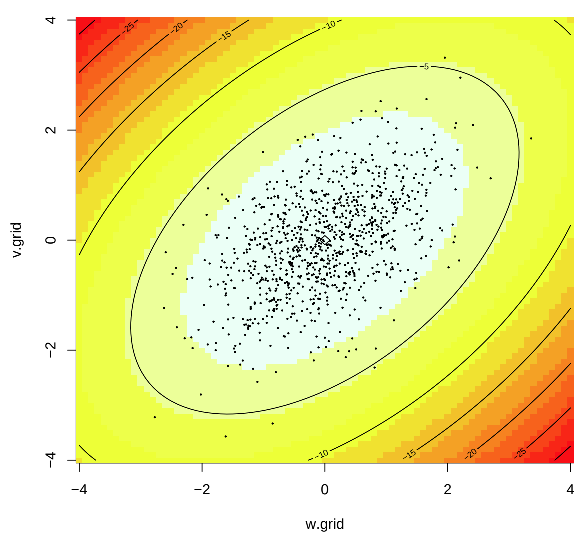

Figure 3: Comparison of the samples and the joint distribution by heat plot. Prepared by the script file "heat.plot.R".

```
# grids over ranges of w and v
w.grid <- mu.1 + sigma.1 * seq(-4,4,.1)
v.grid <- mu.2 + sigma.2 * seq(-4,4,.1)

# function to calculate log-pdf
log.pdf <- function(w, v) {
  z.1 <- (w - mu.1) / sigma.1
```

```
  z.2 <- (v - mu.2) / sigma.2
  return(-0.5 * (z.1^2 + z.2^2 - 2 * rho * z.1 * z.2) / (1 - rho^2))
}

# contour type plots
lf.grid <- outer(w.grid, v.grid, log.pdf)
lf.grid <- lf.grid - max(lf.grid) ## don't care about constant subtraction
image(w.grid, v.grid, lf.grid)
contour(w.grid, v.grid, lf.grid, add = TRUE)
points(w.samp, v.samp, pch = 20, cex = 0.3)
```

TASK 1 Draw separate histograms for samples of w and v. Compare them against the plots of the marginal pdf's $f_1(w)$ and $f_2(v)$.

TASK 2 Run the sampler again and draw the heat plot as in Figure 3 with $\rho = 0.9$. Compare the histogram of samples of w with its marginal distribution.

TASK 3 For $\rho = 0.9$, run the sampler again with starting values $w = 10$, $v = 10$. Draw the histogram of w together with the curve of the corresponding marginal distribution.

TASK 4 Now run the sampler again ($\rho = 0.9$) but allow burn-in $B = 100$. Make the histogram of w together with the curve of the marginal distribution. Does the histogram improve?

Remark. You are suggested to save the code into some files. When one task is done, you may simply make a copy of the file, and just change some parameters to finish the next task. Some model solutions are put in Figure 4.

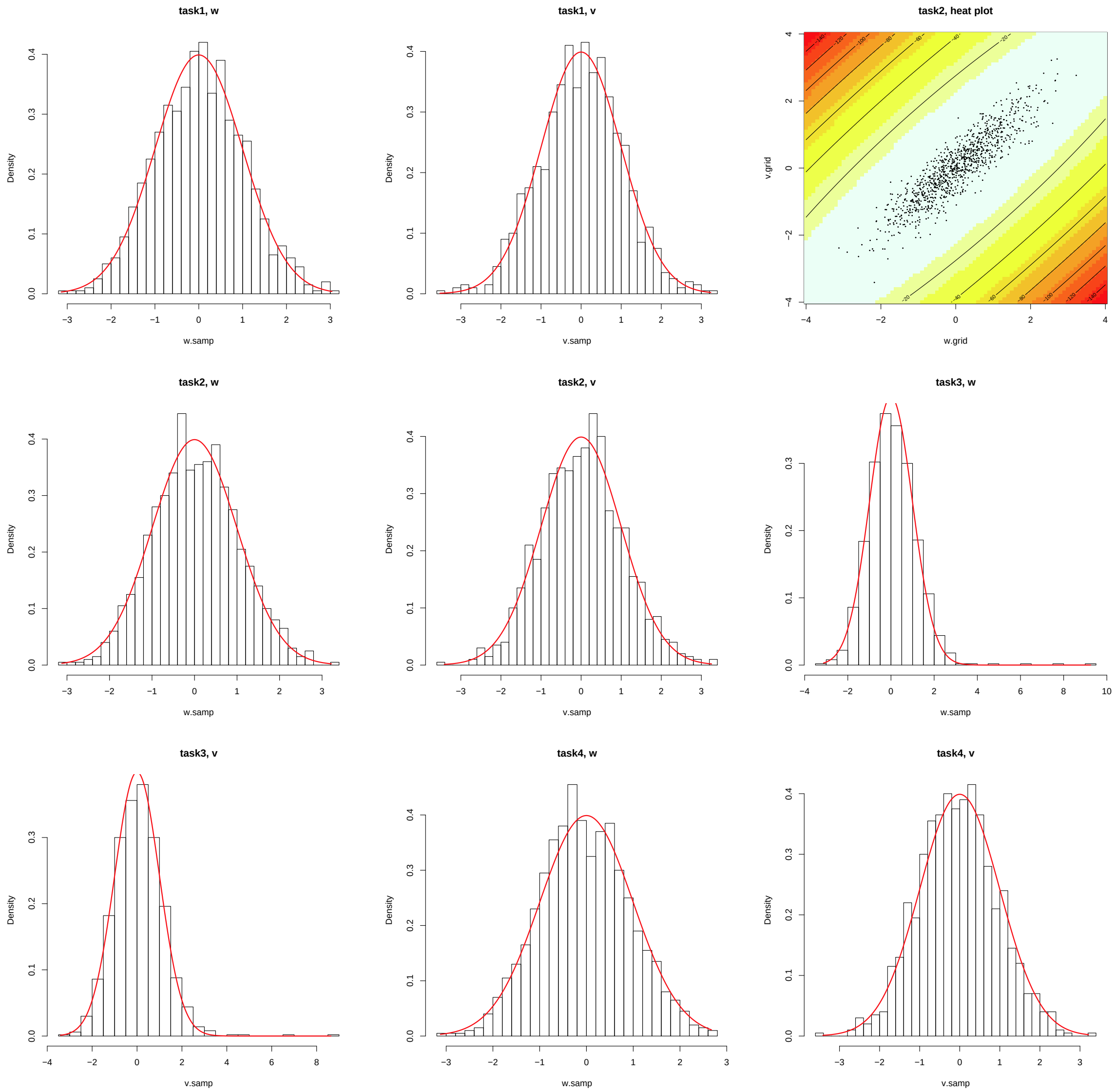

Figure 4: Model solutions.

∼∼END∼∼