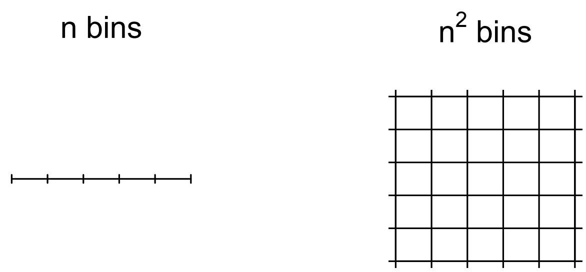
STA 114: Statistics

Lab 5

In Lab 3 we saw that we could sample from any arbitrary univariate pdf by discretizing its range into small bins. This becomes difficult to do for multivariate pdfs, because you need many more bins to cover a square on a plane than what you need to cover a line segment. A line segments of unit length can be covered by n many bins of width $1/n$. Whereas a square of unit side length needs n^2 square-shaped bins of side length $1/n$ for a coverage.



However, there are other methods to sample from a pdf, which often prove more efficient in the multivariate case. This lab overview a very popular approach, known as the Gibbs sampling.

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 $f_1(w|v)$ and $f_2(v|w)$ denote the two conditional pdfs associated with $f(w, v)$. A Gibbs sampling draws these samples iteratively as follows:

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

This program is fairly easy to run provided it is easy to sample from $f_1(w|v)$ and $f_2(v|w)$. Some fairly advanced probability theory shows that the samples we generate (perhaps after discarding some initial draws) well represent the bivariate pdf $f(w, v)$. Consequently (and 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 known:

- $f_1(w) = \textsf{Normal}(\mu_1, \sigma_1^2)$,
- $f_2(v) = \textsf{Normal}(\mu_2, \sigma_2^2)$.
- $f_1(w|v) = \textsf{Normal}(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(v - \mu_2), \sigma_1^2(1 - \rho^2))$.
- $f_2(v|w) = \textsf{Normal}(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(w - \mu_1), \sigma_2^2(1 - \rho^2))$.

We shall use Gibbs sampler to generate samples of (w, v) and then we shall compare them against $f(w, v)$, $f_1(w)$ and $f_2(v)$. The code below runs the sampler for a total of $B + M$ iterations where B is the number of initial samples to be discarded (these samples are referred to as the burn-in samples). We start by setting $B = 0$ and later consider some actual discarding. The first illustration is done wit h $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$ and $\rho = 0.5$.

```
# 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 generated 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

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

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 mere techniques to improve plots. The code below gives details of these.

```
# 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. Get separate histograms for samples of w and samples of v. Compare them against the plots of the marginal pdf $f_1(w)$ and $f_2(v)$. You can use the above grids.

TASK 2. Run the sampler again and generate the three plots above with $\rho = 0.9$.

TASK 3. For $\rho = 0.9$, run the sampler again with starting values $w = 10$, $v = 10$. Generate the three plots and comment on the histogram comparisons.

TASK 4. Now run the sampler again ($\rho = 0.9$) but allow burn-in $B = 100$. Do the histograms improve?

Example (Normal and inverse-chi-square)**.** Now consider the following bivariate pdf

$$f(w, v) = \text{const.} \times v^{-\frac{r+n+2}{2}} \exp\left[-\frac{rs + (n-1)s_x^2 + n(\bar{x} - w)^2}{2v} - \frac{(w-a)^2}{2b^2}\right]$$

defined over $w \in (-\infty, \infty)$ and $v \in (0, \infty)$. It is fairly difficult to calculate the constant terms that makes this function a pdf. But it is relatively easy to show that

- $f_1(w|v) = \mathsf{Normal}(\frac{nb^2\bar{x}+va}{nb^2+v}, \frac{vb^2}{nb^2+v})$ [note: the second parameter is the variance (not standard deviation) of the normal pdf]
- $f_2(v|w) = \chi^{-2}(r+n, \frac{rs+(n-1)s_x^2+n(\bar{x}-w)^2}{r+n})$. [Note $V \sim \chi^{-2}(r', s')$ means $(r's')/V \sim \chi^2(r')$ which is the same as the $\mathsf{Gamma}(r'/2, 1/2)$ distribution.]

TASK 5. Use the above two conditionals to run a Gibbs sampler and generate samples of (w, v) from $f(w, v)$. Take $a = 0, b = r = s = 1$, $n = 10$, $\bar{x} = 1.38$ and $s_x = 0.33$. Generate contour type plots to compare the samples with the log-pdf. Also generate histograms of w samples and separately for v samples.

TASK 6. From the samples generated above, give five quantile summaries for $f_1(w)$ and $f_2(v)$. Use the following grids: `w.grid <- x.bar + seq(-4,4,.1) / sqrt(n); v.grid <- s.x*s.x * seq(0.5, 5, 0.1)`