STA3127 Statistical Computing

Jeong Geonwoo

DUE Monday, September 25

Q1

i) Using the change of variables $u = x/2$ and $v = y - 2x$, rewrite the integral in (1) as an integration with respect to u and v. [You must properly specify the region of your integral and the integrand.]

Since $u = \frac{x}{2}$, $v = y - 2x$, $x = 2u$, $y = v + 4u$, $|J| = \begin{vmatrix} \frac{dx}{du} & \frac{dx}{dv} \\ \frac{dy}{du} & \frac{dy}{dv} \end{vmatrix} = \begin{vmatrix} 2 & 0 \\ 4 & 1 \end{vmatrix} = |-2| = 2$,

$$\begin{aligned}
\int_0^2 \int_{2x}^{2x+1} sin(x^2+y^2)dydx &= \int_0^2 \int_0^5 sin(x^2+y^2)\mathbf{1}(0 \le x \le 2)\mathbf{1}(2x \le y \le 2x+1)dydx \\
&= \int_0^1 \int_0^1 sin((2u)^2+(v+4u)^2)\mathbf{1}(0 \le u \le 1)\mathbf{1}(4u \le v+4u \le 4u+1)|J|dvdu \\
&= \int_0^1 \int_0^1 sin((2u)^2+(v+4u)^2)\mathbf{1}(0 \le u \le 1, 0 \le v \le 1)|J|dvdu \\
&= \int_0^1 \int_0^1 2sin(20u^2+8uv+v^2)\mathbf{1}(0 \le u \le 1, 0 \le v \le 1)dvdu
\end{aligned}$$

ii) Draw a contour plot of the integrand obtained in i) on the corresponding region of the integral.

```r
rm(list=ls())
set.seed(2018122062)
```

```r
u_grid <- seq(0, 1, length.out=1001)
v_grid <- seq(0, 1, length.out=1001)

integrand1 <- function(u,v){
  value <- 2*sin(20*u^2+8*u*v+v^2) * 1*(0<=u & u<=1) * 1*(0<=v & v<=1)
  return (value)
}

z_values1 <- matrix(NA, nrow = length(u_grid), ncol = length(v_grid))
```

```
for (i in 1:length(u_grid)) {
  for (j in 1:length(v_grid)) {
    z_values1[i, j] <- integrand1(u_grid[i], v_grid[j])
  }
}

contour(u_grid, v_grid, z_values1, xlab = "u", ylab = "v")
```

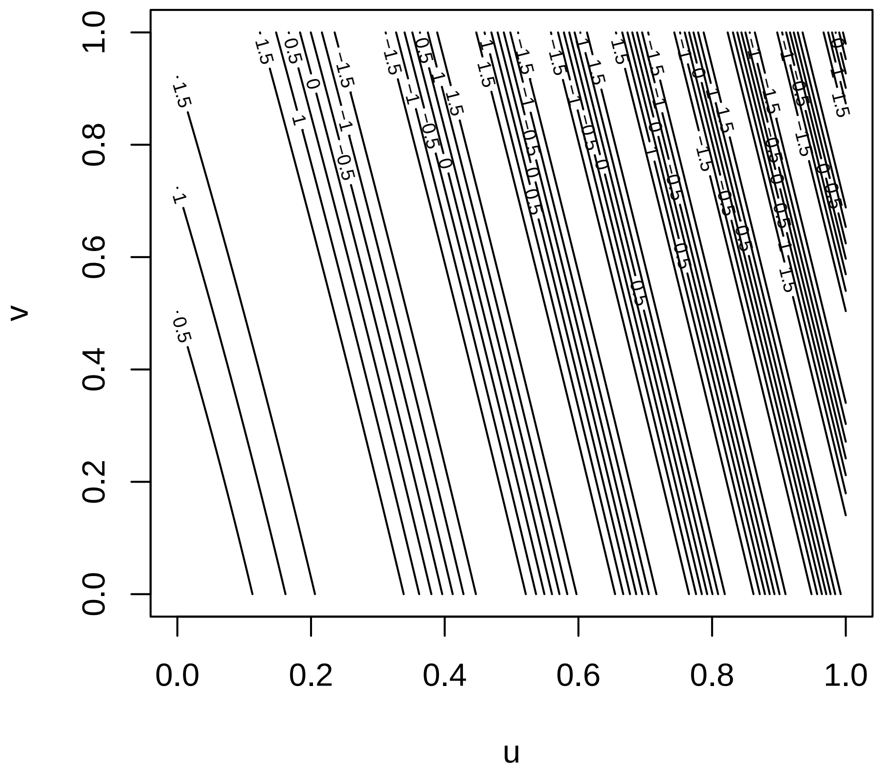

figure 1 : Contour plot of the integrand obtained in i)

iii) Using the change of variables u = x/2 and v = y/5, rewrite the integral in (2) as an integration with respect to u and v. [You must properly specify the region of your integral and the integrand.]

Since $u = \frac{x}{2}$, $v = y/5$, $x = 2u$, $y = 5v$, $|J| = \begin{vmatrix} \frac{dx}{du} & \frac{dx}{dv} \\ \frac{dy}{du} & \frac{dy}{dv} \end{vmatrix} = \begin{vmatrix} 2 & 0 \\ 0 & 5 \end{vmatrix} = 10$,

$$\begin{aligned}
&\int_0^2 \int_0^5 sin(x^2 + y^2)\mathbf{1}(0 \leq x \leq 2, 2x \leq y \leq 2x + 1)dydx \\
&= \int_0^1 \int_0^1 sin((2u)^2 + (5v)^2)\mathbf{1}(0 \leq u \leq 1, \frac{4}{5}u \leq v \leq \frac{4}{5}u + \frac{1}{5})|J|dvdu \\
&= \int_0^1 \int_0^1 10sin(4u^2 + 25v^2)\mathbf{1}(0 \leq u \leq 1, \frac{4}{5}u \leq v \leq \frac{4}{5}u + \frac{1}{5})dvdu
\end{aligned}$$

iv) Draw a contour plot of the integrand obtained in iii) on the corresponding region of the integral.

```
integrand2 <- function(u,v){
  value <- 10*sin(4*u^2+25*v^2) * 1*(0<=u & u<=1) * 1*(4/5*u<=v & v<=4/5*u+1/5)
  return (value)
}
z_values2 <- matrix(NA, nrow = length(u_grid), ncol = length(v_grid))
for (i in 1:length(u_grid)) {
  for (j in 1:length(v_grid)) {
    z_values2[i, j] <- integrand2(u_grid[i], v_grid[j])
  }
}
contour(u_grid, v_grid, z_values2, xlab = "u", ylab = "v")
```

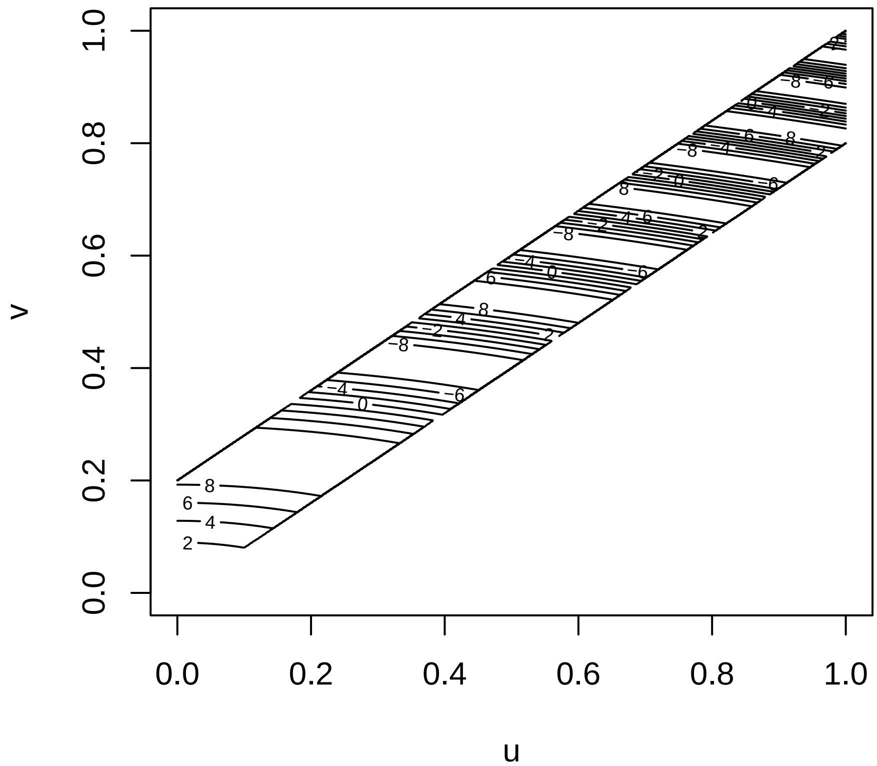

figure 2 : Contour plot of the integrand obtained in iii)

v) By generating 100 pairs of random numbers on (0, 1), estimate the integrals in i) and iii). Repeat this process 1000 times, thereby obtaining 1000 estimates for each of i) and iii) using 1000 distinct sets of 100 pairs of random numbers. Calculate the average and variance of your 1000 estimates for both i) and iii).

```
integral1_list <- c()
integral2_list <- c()
for (i in c(1:1000)){

  integral1 = 0
  integral2 = 0
  for (i in 1:100) {
    u <- runif(1, 0, 1)
    v <- runif(1, 0, 1)
    integral1 = integral1 + integrand1(u, v)
    integral2 = integral2 + integrand2(u, v)
  }
  integral1 = integral1 / 100
  integral2 = integral2 / 100
  integral1_list <- c(integral1_list, integral1)
  integral2_list <- c(integral2_list, integral2)
}
```

```r
cat("integral1 mean:", mean(integral1_list),
    ", integral1 variance:", var(integral1_list))
```

```
## integral1 mean: 0.2609244 , integral1 variance: 0.0195579
```

```r
cat("integral2 mean:", mean(integral2_list),
    ", integral2 variance:", var(integral2_list))
```

```
## integral2 mean: 0.2413366 , integral2 variance: 0.09756303
```

vi) Compare the computed averages and variances of the two estimates, and draw a conclusion about which method is better, supported by a rationale.

The two estimates are, after all, the result of integrating the same function, so the computed average is approximately the same. However, the variance of integral2 is larger than variance of integral1 because the region of integrand1 is $0 \leq u, v \leq 1$ while integrand2 is zero where outside the $0 \leq u \leq 1, \frac{4}{5}u \leq v \leq \frac{4}{5}u + \frac{1}{5}$. Therefore, it is better to estimate with the method of integral1.

Q2

i) Note that the probability of interest is the expected value of an indicator variable. Find the number of (pseudo) random numbers to generate one realization of the indicator variable.

We need k+1 (pseudo) random numbers. (because we can't use the sample() function)

We need k random numbers to randomly permute the location of bombs, and we need 1 random number to determine which cell to randomly click on.

ii) For any values of $p \geq 1$, $q \geq 1$, and $k \geq 0$, write down your R function that estimates the probability of interest as the average of n realizations of the indicator variable. Your function should take p, q, k, and n as input parameters. Using your R function, estimate the probabilities with a common choice of $n = 1000$ for the following scenarios: (a) $p = q = 1$ and $k = 1$, (b) $p = q = 3$ and $k = 1$, and (c) $p = 20, q = 15$, and $k = 40$.

Below is a function "prob_no_bomb" that estimates the probability of the first random click reveals zero bombs in the vicinity of the selected cell.

```
prob_no_bomb<- function(p, q, k, n){
  ########################################
  # input
  # p : the number of rows of map
  # q : the number of columns of map
  # k : the number of bombs in the map
  # n : the number of realizations of the indicator variable

  # output
  # success_ratio : estimated expectation(probability) of
  #                 zero bombs in the vicinity of the selected cell.
  ########################################

  count_success = 0
  # if clicked cell and surrounding 8 cells have no bomb,
  # then "count_success" increases by 1

  for (realization in c(1:n)){
    # Empty map
    map <- matrix(0, nrow = p, ncol = q)

    # Location of bombs : Initializing
    bomb_location <- c(1:(p*q))
    # Location of bombs : Random Permutation
    for (index in c(1:k)){
      I = floor(p*q*runif(1,0,1))+1
      temp <- bomb_location[I]
      bomb_location[I] <- bomb_location[index]
      bomb_location[index] <- temp
    }
    # Location of bombs : Select k-location
    bomb_location = bomb_location[1:k]

    # Placing the bombs
    for (location in bomb_location) {
```

```
      row_idx <- (location - 1) %/% q + 1
      col_idx <- (location - 1) %% q + 1
      map[row_idx, col_idx] <- 1
    }
    # Initial click
    click <- floor(p*q*runif(1,0,1))+1
    click_p <- (click-1) %/% q + 1
    click_q <- (click-1) %% q + 1

    # Count the number of bombs in surrounding cells
    count_bombs = 0
    for (i in c(-1,0,1)){
      for (j in c(-1,0,1)){
        if ((1 <= click_p+i & click_p+i <= p)
            & (1<= click_q+j & click_q+j <= q)){
          count_bombs = count_bombs + map[click_p+i, click_q+j]
        }
      }
    }
    # if there is no bomb in 9-cell, count_success increases by 1
    if (count_bombs == 0){
      count_success = count_success + 1
    }
  }
  success_ratio = count_success / n
  return (success_ratio)
}
```

Estimated probability (average of indicator variable)

```
prob_no_bomb(1,1,1,1000)
```

```
## [1] 0
```

```
prob_no_bomb(3,3,1,1000)
```

```
## [1] 0.379
```

```
prob_no_bomb(20,15,40,1000)
```

```
## [1] 0.312
```

Comparison : calculated probability

If the clicked cell is **inside** cell, then there should be no bombs in the **nine** cells. (clicked cell and surrounding cell) If the clicked cell is **side** cell, then there should be no bombs in the **six** cells. If the clicked cell is **corner** cell, then there should be no bombs in the **four** cells.

(a) $p = q = 1$ and $k = 1$

Since the cells you randomly click on and the cells with bombs are always the same, the probability that there are no bombs around the clicked cell is **zero**.

(b) $p = q = 3$ and $k = 1$

P(click **inside** cell) * P(zero bombs in the vicinity of cell)
+ P(click **side** cell) * P(zero bombs in the vicinity of cell)
+ P(click **corner** cell) * P(zero bombs in the vicinity of cell)

$= \frac{1}{9} * 0 + \frac{4}{9} * \frac{3}{9} + + \frac{4}{9} * \frac{5}{9} = 0.3950$

(c) $p = 20, q = 15$, and $k = 40$.

P(click **inside** cell) * P(zero bombs in the vicinity of cell)
+ P(click **side** cell) * P(zero bombs in the vicinity of cell)
+ P(click **corner** cell) * P(zero bombs in the vicinity of cell)

$= \frac{234}{300} * \frac{{}_{291}\mathrm{C}_{40}}{{}_{300}\mathrm{C}_{40}} + \frac{62}{300} * \frac{{}_{294}\mathrm{C}_{40}}{{}_{300}\mathrm{C}_{40}} + \frac{4}{300} * \frac{{}_{296}\mathrm{C}_{40}}{{}_{300}\mathrm{C}_{40}} = 0.3055$

We can expect that as the number of trials increases, the estimated probability to be closer to the calculated probability.