R handout: Integration and Transformation

Yuhui Yao

1. Monte Carlo Method

   Suppose X is a random variable with a p.d.f. f and a c.d.f. F over a domain D. Also, we have a sample $X_s = \{x_1, x_2, ..., x_n\}$ from X with n observations.

   $$E(X) = \int_D x f(x) dx \approx \bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

   Besides, because of Central Limit Theorem, $\bar{X}$ asymptotically follows $N(\mu, \frac{\sigma^2}{n})$ where $\mu = E(X)$ is the mean of X and σ^2 is the variance of X. So, when $n \to \infty$, $\bar{X} \to E(X)$. Hence, intuitively, the result is more precise with a larger sample size.

2. Example. $Y = |X|$ where $X \sim N(0, 1)$. Find out $E(Y)$

   (a) Direct Integration

   As we know, the p.d.f. of Y

   $$f_Y(y) = 2\phi(y), y \geq 0$$

   where ϕ is the p.d.f. of the standard normal distribution.

   - R code

   ```
   # define the p.d.f. of Y
   Y.pdf <- function(y) 2 * dnorm(y)

   # define the integrand of the expecation of Y
   E.Y <- function(y) y * Y.pdf(y)

   # Integrate the integrand
   integrate(E.Y, lower = 0, upper = Inf)
   # Result: 0.7978846
   ```

(b) Monte Carlo Method with a transformation

Because

$$E(Y) = E(|X|) = \int_{-\infty}^{\infty} |x|\phi(x)dx \approx \frac{1}{n}\sum_{i=1}^{n}|x_i|$$

where n is the number of simulation and $X_s = \{x_1, ..., x_n\}$ is a sample from $N(0,1)$.

- R code

```
# set seed to make this process repeatable
set.seed(12345)

# set the number of simulation
n <- 1000

# simulate a sample from the standard normal distribution
X.s <- rnorm(n)

# calculate the mean of the absoluate values
mean(abs(X.s))
# Result: 0.7944
```

(c) Monte Carlo Method with another transformation

Because $exp(1)$ has a same domain as Y,

$$E(Y) = \int_{0}^{\infty} yf_Y(y)dy = \int_{0}^{\infty} \frac{yf_y(y)}{e^{-y}}e^{-y}dy = E(\frac{yf_Y(y)}{e^{-y}}) \approx \frac{1}{n}\sum_{i=1}^{n}\frac{y_i f_Y(y_i)}{e^{-y_i}}$$

where n is the number of simulation and $Y_s = \{y_1, ..., y_n\}$ is a sample from $exp(1)$.

- R code

```
# set seed to make this process repeatable
set.seed(12345)

# set the number of simulation
n <- 1000

# define the p.d.f. of Y
Y.pdf <- function(y) 2 * dnorm(y)
```

```
# simulate a sample from the standard normal distribution
Y.s <- rexp(n)

# calculate the mean of the absoluate values
mean(Y.s * Y.pdf(Y.s) / dexp(Y.s))
# Result: 0.7729
```

3. Summary: Steps of Monte Carlo Method

   (a) Check the domain

   (b) Simulate an appropriate sample over the domain

   (c) Calculate the mean via an appropriate transformation

4. Practice: suppose $Y = 3(ln(X)^{1/3} - 1)$, where $X \sim exp(2)$. Use Monte Carlo Method to find $E(Y)$.

   Hint: You need to show the analytical part and the numerical part in R. Also, you may need to use $rexp$ and $help(rexp)$ in R