STA3127 Statistical Computing

Jeong Geonwoo

DUE Friday, October 06

Q1

i) Show that the cumulative distribution function is

$$F(x) = \begin{cases} \frac{\kappa^2}{1+\kappa^2} \exp\left(\frac{\lambda(x-m)}{\kappa}\right), & x < m, \\ 1 - \frac{1}{1+\kappa^2} \exp(-\kappa\lambda(x-m)), & x \geq m. \end{cases}$$

case 1) $x < m$, that is, $sign(x-m) = -1$

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(t; m, \lambda, \kappa) \\ &= \left(\frac{\lambda}{\kappa + 1/\kappa}\right) \int_{-\infty}^{x} e^{\lambda(t-m)/\kappa} dt \\ &= \left(\frac{\lambda}{\kappa + 1/\kappa}\right) \left[\frac{\kappa}{\lambda} e^{\lambda(t-m)/\kappa}\right]_{-\infty}^{x} \\ &= \left(\frac{\lambda}{\kappa + 1/\kappa}\right) \frac{\kappa}{\lambda} e^{\lambda(x-m)/\kappa} \\ &= (\frac{\kappa^2}{1+\kappa^2}) e^{\lambda(x-m)/\kappa} \end{aligned}$$

case 2) $x >= m$, that is, $sign(x-m) = 1$

$$\begin{aligned} F(x) &= \left(\frac{\lambda}{\kappa + 1/\kappa}\right) \int_{-\infty}^{x} e^{-\kappa\lambda(t-m)} dt \\ &= 1 - \left(\frac{\lambda}{\kappa + 1/\kappa}\right) \int_{x}^{\infty} e^{-\kappa\lambda(t-m)} dt \\ &= 1 - \left(\frac{\lambda}{\kappa + 1/\kappa}\right) \left[-\frac{1}{\kappa\lambda} e^{-\kappa\lambda(t-m)}\right]_{t}^{\infty} \\ &= 1 - \left(\frac{\lambda}{k + 1/k}\right) \frac{1}{\kappa\lambda} e^{-\kappa\lambda(x-m)} \\ &= 1 - \frac{1}{1+\kappa^2} e^{-\kappa\lambda(x-m)} \end{aligned}$$

Thus, $F(x) = \begin{cases} \frac{\kappa^2}{1+\kappa^2} \exp\left(\frac{\lambda(x-m)}{\kappa}\right), & x < m, \\ 1 - \frac{1}{1+\kappa^2} \exp(-\kappa\lambda(x-m)), & x \geq m. \end{cases}$

ii) For U $\sim$ Unif(0,1), find a map g such that $X \overset{\text{d}}{=} g(U)$ using the inverse transform method.

if $F(x) = y$, $F^{-1} = g$, then $F(g(y)) = y$

case 1) $x < m$, that is, $y < \frac{\kappa^2}{1+\kappa^2}$

$$
\begin{aligned}
y &= \frac{\kappa^2}{1+\kappa^2} e^{\lambda(x-m)/\kappa} \\
\frac{1+\kappa^2}{\kappa^2} y &= e^{\lambda(x-m)/\kappa} \\
\log\left(\frac{1+\kappa^2}{\kappa^2} y\right) &= \lambda(x-m)/\kappa \\
\frac{\kappa}{\lambda}\log\left(\frac{1+\kappa^2}{\kappa^2} y\right) + m &= x
\end{aligned}
$$

case 2) $x \geq m$, that is, $y \geq \frac{\kappa^2}{1+\kappa^2}$

$$
\begin{aligned}
y &= 1 - \frac{1}{1+\kappa^2} e^{-\kappa\lambda(x-m)} \\
\left(1+\kappa^2\right)(1-y) &= e^{-\kappa\lambda(x-m)} \\
\log\left[\left(1+\kappa^2\right)(1-y)\right] &= -\kappa\lambda(x-m) \\
-\frac{1}{\kappa\lambda}\left[\log\left(1+\kappa^2\right)(1-y)\right] + m &= x
\end{aligned}
$$

From inverse transform method, we can get the samples from asymmetric Laplace distribution which are $g(U) = F^{-1}(U) = \begin{cases} \frac{\kappa}{\lambda}\log\left(\frac{1+\kappa^2}{\kappa^2}U\right) + m & , U < \frac{\kappa^2}{1+\kappa^2} \\ -\frac{1}{\kappa\lambda}\left[\log\left(1+\kappa^2\right)(1-y)\right] + m = x & , U \geq \frac{\kappa^2}{1+\kappa} \end{cases}$

iii) Generate 10^6 random numbers from AL(1, 1, 2), and overlay a histogram (with a suitable bin size) with the target density curve f.

[Note: To properly overlay a histogram with a density curve, you may need to specify '**freq=FALSE.**']

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

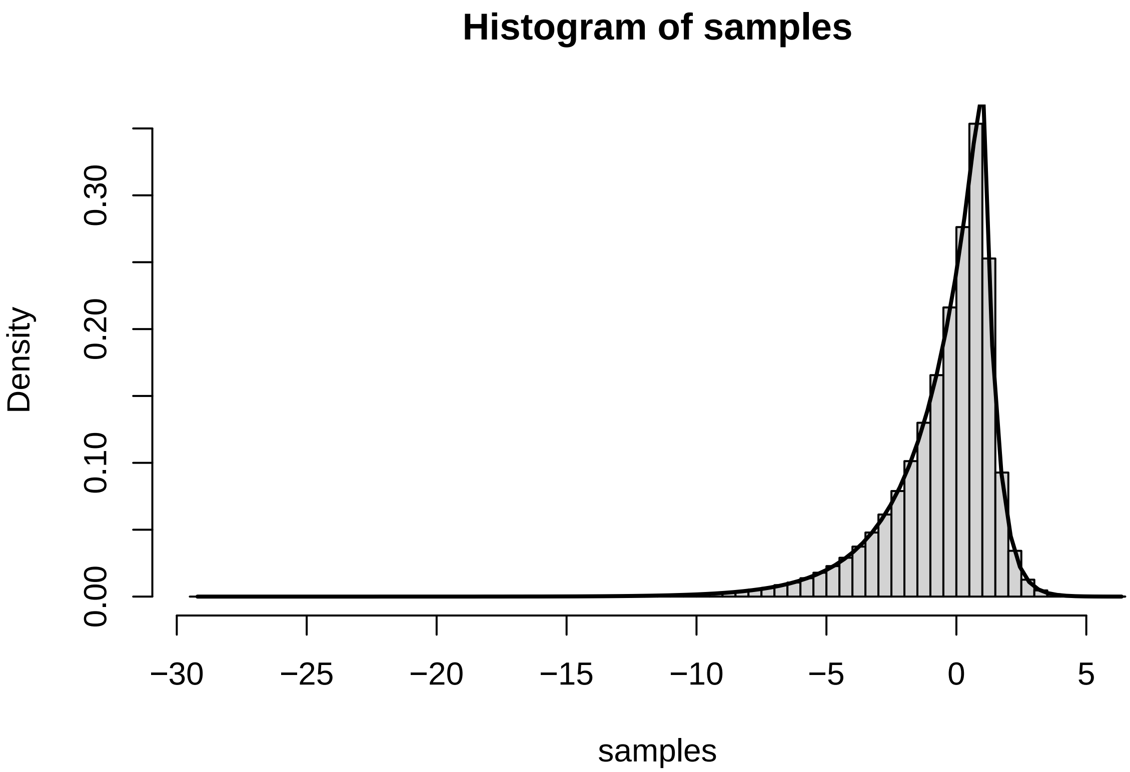
Below are the functions that define the PDF(probability Density Function) and inverse CDF(inverse Cumulative Density Function) of Asymmetric Laplace distribution.

```
AL_pdf <- function(x, m, lambda, kappa) {
  pdf <- (lambda / (kappa + 1/kappa)) * exp(-lambda * (x-m) * sign(x-m) * kappa^{sign(x-m)})
  return (pdf)
}

AL_cdf_inv <- function(U, m, lambda, kappa) {
  sapply(U, function(Ui) {
    if (Ui < kappa^2 / (1 + kappa^2)) {
      cdf_inv <- kappa / lambda * log((1 + kappa^2) / kappa^2 * Ui) + m
    } else {
      cdf_inv <- - 1 / (kappa * lambda) * log((1 + kappa^2)*(1-Ui)) + m
```

```
    }
    return(cdf_inv)
  })
}
```

```
U <- runif(10^6)
samples <- AL_cdf_inv(U, m=1, lambda=1, kappa=2)
hist(samples, breaks = 100, freq=FALSE)
curve(AL_pdf(x, m=1, lambda=1, kappa=2), from = min(samples), to = max(samples), add=TRUE, lwd=2)
```



We can generate the samples from asymmetric Laplace distribution by using the inverse CDF(inverse Cumulative Density Function) of asymmetric Laplace distribution.

Q2

i) Show that

$$F_Y(y) \equiv \Pr\{Y \le y\} = 1 - (1-q)^{\max\{0,\lfloor y \rfloor\}}, \quad -\infty < y < \infty$$

where $\lfloor \cdot \rfloor$ is the floor function.

Even if $y \in (-\infty, \infty)$ is continuous variable, by the definition of geometric distribution (which is "the **number of iid Bernoulli trials** until the first success"), we can consider y as $y = 1, 2, 3, ...$ discrete variable. In the sense that $y \ge 1$ (if $y < 1$, $\Pr\{Y < y\} = \Pr\{Y \le 0\} = 0$), we can express $\max\{0, y\}$ instead of y, and in the sense that y is the number of trials(positive integer), we can express $\Pr\{Y \le y\} = 1 - (1-q)^{\max\{0,\lfloor y \rfloor\}}$ instead of $\Pr\{Y \le y\} = 1 - (1-q)^{\max\{0,y\}}$. (e.g. $\Pr\{Y \le 3.5\} = \Pr\{Y \le 3\}$). Formula expression is following:

$$\begin{aligned} F_Y(y) &\equiv \Pr\{Y \le y\} \\ &= \Pr\{Y \le \max\{0, \lfloor y \rfloor\}\} \\ &= \sum_{i=1}^{\max\{0,\lfloor y \rfloor\}} \Pr\{Y = i\} \\ &= 1 - (1-q)^{\max\{0,\lfloor y \rfloor\}}, \quad -\infty < y < \infty \end{aligned}$$

ii) We define the generalized inverse h^{-1} of a function $h : \mathcal{X} \to \mathcal{Y}$ as $h^{-1}(y) = \inf\{x \in \mathcal{X} : y \le h(x)\}, y \in \mathcal{Y}$. Show that

$$F_Y^{-1}(y) = \left\lceil \frac{\log(1-y)}{\log(1-q)} \right\rceil, \quad 0 < y < 1,$$

where $\lceil \cdot \rceil$ is the ceiling function.

The CDF(Cumulative Density Function) is $y = F(x) = 1 - (1-q)^{\max\{0,\lfloor x \rfloor\}}$. To find the inverse CDF $F^{-1}(y) = x$, we can consider the general case.

$$\begin{aligned} y &= 1 - (1-q)^x \\ x &= log_{(1-q)}(1-y) \end{aligned}$$

As the discreteness of the geometric distribution, we know x is an integer. By the definition of geometric distribution (which is "the **number of iid Bernoulli trials** until the first success"), we need to round $x = log_{(1-q)}(1-y)$ up to the nearest larger integer, using ceiling function. (e.g, If $log_{(1-q)}(1-y) = 3.5$, that means that until the first success, 3 trials are not enough, so 4-th trial is required.)

$$\begin{aligned} F^{-1}(y) &= \left\lceil \log_{(1-q)}(1-y) \right\rceil \\ &= \left\lceil \frac{log(1-y)}{log(1-q)} \right\rceil, \quad 0 < y < 1 \end{aligned}$$

iii) Let $Y_1, ..., Y_r$ be iid copies of Y. Show that X can be represented by a linear function of $Y_1, ..., Y_r$, specifying the relationship between p and q. [Hint: Use the moment generating functions.]

As the result of independent trial is success or failure, p is equal to $1 - q$. The MGF (Moment Generating Function) of Negative binomial distribution is following :

$$\begin{aligned}
f_X(x) &= \begin{pmatrix} x+r-1 \\ x \end{pmatrix} (1-p)^r p^x, \quad x = 0, 1, 2 \ldots \\
M_X(t) &= E\left[e^{tX}\right] \\
&= \sum_{x=0}^{\infty} e^{tx} \begin{pmatrix} x+r-1 \\ x \end{pmatrix} (1-p)^r p^x \\
&= (1-p)^r \sum_{x=0}^{\infty} \begin{pmatrix} x+r-1 \\ x \end{pmatrix} \left(pe^t\right)^x \\
&= (1-p)^r \left(1-pe^t\right)^{-r} \quad (\textit{by negative binomial series}) \\
&= \left(\frac{q}{1-pe^t}\right)^r
\end{aligned}$$

Let $Z = Y - 1$ is the number of failures until the first success with success probability q. Then the MGF of Z is following :

$$\begin{aligned}
f_Z(z) &= (1-q)^z q, \quad z = 0, 1, 2.\ldots \\
M_Z(z) &= E\left[e^{tZ}\right] \\
&= \sum_{z=0}^{\infty} e^{tz}(1-q)^z q \\
&= q \cdot \sum_{z=0}^{\infty} \left(e^t(1-q)\right)^z \\
&= q \cdot \frac{1}{1-e^t(1-q)} \\
&= \frac{q}{1-pe^t}
\end{aligned}$$

From $M_X(t) = (M_Z(t))^r$, X can be represented the sum of Z. That means X can be represented by a linear function of $Y_1, ..., Y_r$, which is $X = Z_1 + ... + Z_r = Y_1 + ... + Y_r - r$

iv) Based on the result above, determine a map g such that

$$X \overset{d}{=} g\left(\log\left(U_1\right), \ldots, \log\left(U_r\right)\right), \quad U_i \overset{\text{iid}}{\sim} \text{Unif}(0,1)$$

In ii), we get $F_Y^{-1}(y) = \left\lceil \frac{\log(1-y)}{\log(1-q)} \right\rceil, \quad 0 < y < 1$. As $Z = Y - 1$, we know that $F_Z^{-1}(z) = \left\lceil \frac{\log(1-z)}{\log(1-q)} \right\rceil - 1$. Therefore, $g(log(U_1), ..., log(U_r)) = (\left\lceil \frac{\log(1-U_1)}{\log(1-q)} \right\rceil - 1) + ... + (\left\lceil \frac{\log(1-U_r)}{\log(1-q)} \right\rceil - 1)$

v) Using this framework, estimate $E(sin(X))$ by generating 10^6 random numbers from a negative binomial distribution with $r = 10$ and $p = 0.7$.

Defining the CDF of Geometric distribution (Y)

```
Geometric_cdf <- function(x, q) {
  sapply(x, function(xi) {
    cdf <- 1-(1-q)^{max(0, floor(xi))}
    return(cdf)
  })
}
```

```
Geometric_cdf_inv <- function(x, q) {
  return(ceiling(log(1-x) / log(1-q)))
}
```

Setting the success probability (p) and the number of failures (r)

```
p = 0.7
q = 1-p
r = 10
```

Sampling the random variable X (which is *neg_binorm* in below code) from the negative binomial distribution (using inverse CDF of Geometric distribution) and estimating $E(sin(X))$

```
n_samples=10^6
U = runif(n_samples,0,1)
samples <- rep(NA, n_samples)


for (i in 1:n_samples){
  sum_geometric <- sum(Geometric_cdf_inv(runif(r), q))
  neg_binorm <- sum_geometric - r
  samples_i <- sin(neg_binorm)
  samples[i] <- samples_i
}
```

Estimated $E(sin(X))$

```
mean(samples)
```

```
## [1] -0.0002791598
```