STA3127 Statistical Computing

Jeong Geonwoo

DUE Monday, Nov 20

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

Q1.

A negative binomial distribution models the number of successes in a sequence of iid Bernoulli trials until a specified number of failures occurs. Consider a random variable X that has a negative binomial distribution, that is,

$$\Pr\{X = k\} = \binom{k+r-1}{k}(1-p)^r p^k, \quad k = 0, 1, 2, \ldots,$$

where r is a pre-specified number of failures, p is a success probability, and k is the number of successes. In Homework 3, we have shown that

$$X \overset{d}{=} \sum_{k=1}^{r} \left\lceil \frac{\log U_k}{\log p} \right\rceil - r, \quad U_k \overset{\text{iid}}{\sim} \text{Unif}(0,1).$$

We want to estimate $\theta = E[b(X)]$ for $b(x) = \log(e^x + 1)$.

i) Let h be the function such that $b(X) \overset{d}{=} h(U_1, \ldots, U_r)$. Show that h is a monotone decreasing function of each of its coordinates, that is, show that $h(u_1, \ldots, u_k, \ldots, u_r) \geq h(u_1, \ldots, u_k', \ldots, u_r)$ for every $0 < u_k \leq u_k' < 1, k = 1, \ldots, r$. This shows that an antithetic estimator can be constructed using

$$V \equiv \frac{h(U_1, \ldots, U_r) + h(1 - U_1, \ldots, 1 - U_r)}{2}$$

Step 1. $\frac{log\ U_k}{log\ p}$ is decreasing function of U_k

$\log U_k < 0$ where $U \sim Unif(0,1)$ is increasing function of U_k. $\log p$ is also less than 0 $(0 < p < 1)$ but given in this case. From that, we can say that $\frac{log\ U_k}{log\ p}$ is decreasing function of U_k. Because the U_k is larger, the absolute value of $log\ U_k$ is smaller.

Step 2. $X \overset{d}{=} \sum_{k=1}^{r} \left\lceil \frac{\log U_k}{\log p} \right\rceil - r$ is decreasing function of U_k

The ceiling function is monotone increasing function of input value. that means when input value is smaller than before, ceiling function can not be the large value than before. From the Step 1, we know that $\frac{log\ U_k}{log\ p}$ is decreasing function of U_k. Therefore, $X \overset{d}{=} \sum_{k=1}^{r} \left\lceil \frac{\log U_k}{\log p} \right\rceil - r$ is decreasing function of U_k

Setp 3. e^x **and** $log\ x$ **is increasing function of** x**.**

We can know that without any proof. Step 3 means that $e^{\sum_{k=1}^{r}\lceil\frac{\log U_k}{\log p}\rceil - r}$ and $b(X) = log(e^{\sum_{k=1}^{r}\lceil\frac{\log U_k}{\log p}\rceil - r}+1)$ is decreasing function of U_k.

From Step 1~3, we can say $h(u_1,...,u_r)$ is a monotone decreasing function of any $u_k,\ k \in \{1,...,r\}$ with fixed $u_j,\ \forall j \in \{1,...,r \mid j \neq k\}$ That means with fixed $u_j,\ \forall j \in \{1,...,r \mid j \neq k\}$, we can say $h(u_1,...,u_k,...u_r) \geq h(u_1,...,u'_k,...u_r)$ for every $0 < u_k \leq u'_k < 1, k = 1,\ldots,r$.

ii) A reasonable (not the best) controlled estimator can be constructed using

$$W \equiv h\left(U_1,\ldots,U_r\right) - \frac{\sum_{i=1}^{r}\log\left(U_i\right)+r}{\log p}$$

Looking at the function h**, explain why this control variable is a reasonable choice.**

if W can be a reasonable controlled estimator of θ, $E(W) = E(h\left(U_1,\ldots,U_r\right) - \frac{\sum_{i=1}^{r}\log(U_i)+r}{\log p}) = \theta$. It is enough to show $E(\frac{\sum_{i=1}^{r}\log(U_i)+r}{\log p}) = 0$.

$$\begin{aligned}
E(log\ U) &= \int_0^1 log\ u\ du \\
&= [ulog\ u - u]_0^1 \\
&= -1
\end{aligned}$$

$$\begin{aligned}
\therefore E(\frac{\sum_{i=1}^{r}\log\left(U_i\right)+r}{\log p}) &= \frac{\sum_{i=1}^{r}E(\log U_i)+r}{\log p} \\
&= \frac{-r+r}{\log p} \\
&= 0
\end{aligned}$$

$$\begin{aligned}
W &\equiv h\left(U_1,\ldots,U_r\right) - \frac{\sum_{i=1}^{r}\log\left(U_i\right)+r}{\log p} \\
&= h\left(U_1,\ldots,U_r\right) - (\frac{\sum_{i=1}^{r}\log\left(U_i\right)+r}{\log p} - 0) \\
&= h\left(U_1,\ldots,U_r\right) - (\frac{\sum_{i=1}^{r}\log\left(U_i\right)+r}{\log p} - E(\frac{\sum_{i=1}^{r}\log\left(U_i\right)+r}{\log p}))
\end{aligned}$$

$-(\frac{\sum_{i=1}^{r}\log(U_i)+r}{\log p} - E(\frac{\sum_{i=1}^{r}\log(U_i)+r}{\log p}))$ can reduce the variance of estimator.

iii) We want to estimate θ **with** $r = 10$ **and** $p = 0.3$**. Let** X_i **be iid copies of** X**. Based on the central limit theorem, estimate** θ **with** $\bar{B}_n = n^{-1}\sum_{i=1}^{n} b\left(X_i\right)$ **for** n **such that** $\Pr\left\{\left|\bar{B}_n - \theta\right| \leq 0.01\right\} \approx 0.95$**. Use the successive computation. Report your** n **and** $\bar{B}_n$**.**

$$\begin{aligned}
&\frac{\bar{B}_n - E\left(\bar{B}_n\right)}{s^2(b(X))/n} \sim N(0,1), \\
&\text{where } s^2(b(X)) = \frac{1}{n_b}\sum_{i=1}^{n_b}\left(b\left(X_i\right)-\bar{B}_n\right)^2 \\
&P\left(\left|\bar{B}_n - \theta\right| \leqslant 0.01\right) \\
=&P\left(\frac{-0.01}{\sqrt{s^2(b(X))/n}} \leqslant \frac{\bar{B}_n - E\left(\bar{B}_n\right)}{\sqrt{s^2(b(X))/n}} \leqslant \frac{0.01}{\sqrt{s^2(b(X))/n}}\right)
\end{aligned}$$

Therefore, I will just compare $\frac{0.01}{\sqrt{S^2(b(X))/n}}$ to $\Phi^{-1}(0.975) = 1.959964$. If $\frac{0.01}{\sqrt{S^2(b(X))/n}}$ is larger than $\Phi^{-1}(0.975)$, we can say $\Pr\left\{\left|\bar{B}_n - \theta\right| \le 0.01\right\} \approx 0.95$ enough. Else, the sample size is not enough. So that case, I increased the sample size by 1.

```
### iii) ###
r=10; p=0.3

X <- function(u.vec, p=0.3){
  r <- length(u.vec)
  x <- sum( ceiling( log(u.vec) / log(p) ) ) - r
  return (x)}

b <- function(X){return (log(exp(X)+1))}

# I will start with 1st sample
n.b <- 1
Bn.bar <- b(X(runif(r)))
Bn.s2 <- 0
while (TRUE){
  # 2rd sample is always needed.
  # Why? 1 : CLT is established at n > 30
  # Why? 2 : If 2rd sample is not needed, sample variance = 0
  #          then the conditional statement below cannot be calculated.
  next_b <- b(X(runif(r)))
  # successive computation for computation speed
  new.Bn.bar <- Bn.bar + (next_b - Bn.bar) / (n.b+1)
  Bn.s2 <- (1-1/n.b) * Bn.s2 + (n.b+1) * (new.Bn.bar - Bn.bar)^2
  Bn.bar <- new.Bn.bar
  n.b <- n.b + 1
  # 3,4,...th samples is conditionally needed.
  # If sample size is enough, stop sampling
  if (0.01 / sqrt(Bn.s2/(n.b)) > 1.959964){break}}

cat("n of B", n.b, "\n", "Bn_bar", Bn.bar)
```

```
## n of B 219814
##  Bn_bar 4.368933
```

iv) Let V_i be iid copies of V and $\bar{V}_n = n^{-1}\sum_{i=1}^{n} V_i$ be the antithetic estimator. Repeat iii) and report n and $\bar{V}_n$.

```
### iv) ###
V <- function(u.vec, p=0.3){
   return ( ( b(X(u.vec)) + b(X(1-u.vec)) ) / 2 )
}

# I will start with 1st sample
n.v<- 1
Vn.bar <- V(runif(r))
Vn.s2 <- 0
while (TRUE){
  # 2rd sample is always needed.
  # Why? 1 : CLT is established at n > 30
  # Why? 2 : If 2rd sample is not needed, sample variance = 0
```

```
  #          then the conditional statement below cannot be calculated.
  next_V <- V(runif(r))
  # successive computation for computation speed
  new.Vn.bar <- Vn.bar + (next_V - Vn.bar) / (n.v+1)
  Vn.s2 <- (1-1/n.v) * Vn.s2 + (n.v+1) * (new.Vn.bar - Vn.bar)^2
  Vn.bar <- new.Vn.bar
  n.v <- n.v + 1
  # 3,4,...th samples is conditionally needed.
  # If sample size is enough, stop sampling
  if (0.01 / sqrt(Vn.s2/(n.v)) > 1.959964){break}}

cat("n of B", n.v, "\n", "Bn_bar", Vn.bar)
```

```
## n of B 77372
##  Bn_bar 4.356637
```

v) Let W_i be iid copies of W and $\bar{W}_n = n^{-1}\sum_{i=1}^{n} W_i$ be the controlled estimator. Repeat iii) and report n and $\bar{W}_n$.

```
### v) ###
W <- function(u.vec, p=0.3){
  return ( b(X(u.vec)) - (sum(log(u.vec))+r)/log(p) )
}

# I will start with 1st sample
n.w<- 1
Wn.bar <- W(runif(r))
Wn.s2 <- 0
while (TRUE){
  # 2rd sample is always needed.
  # Why? 1 : CLT is established at n > 30
  # Why? 2 : If 2rd sample is not needed, sample variance = 0
  #          then the conditional statement below cannot be calculated.
  next_W <- W(runif(r))
  # successive computation for computation speed
  new.Wn.bar <- Wn.bar + (next_W - Wn.bar) / (n.w+1)
  Wn.s2 <- (1-1/n.w) * Wn.s2 + (n.w+1) * (new.Wn.bar - Wn.bar)^2
  Wn.bar <- new.Wn.bar
  n.w <- n.w + 1
  # 3,4,...th samples is conditionally needed.
  # If sample size is enough, stop sampling
  if (0.01 / sqrt(Wn.s2/(n.w)) > 1.959964){break}}

cat("n of B", n.w, "\n", "Bn_bar", Wn.bar)
```

```
## n of B 30559
##  Bn_bar 4.365208
```

From above result : By using antithetic variable or control variable, we can reduce the sample size which is need for same level of certainty.

Q2.

i) For any random variable $X \sim F$ with its distribution function F and the generalized inverse F^{-1}, show that

$$h(X) \mid (a < X \leq b) \stackrel{d}{=} h\left(F^{-1}\left(U_i^*\right)\right), \quad U_i^* \sim \operatorname{Unif}(F(a), F(b))$$

where $-\infty \leq$ a<b $\leq \infty$

By definition of the distribution function F, for $a < x \leq b$, $P(a < X \leq b) = F(b) - F(a)$. Let $Y = F(X)$. Then $P(F(a) < Y \leq F(b)) = F(b) - F(a)$, meaning Y is uniform on $(F(a), F(b)]$. We can define a random variable U^* that is uniform on $(F(a), F(b)]$, then $F^{-1}(U^*)$ has the same distribution as X on the $(a, b]$. Therefore, we can say $P(a < X \leq b) = P(F(a) < F(X) \leq F(b)) = P(F(a) < U^* \leq F(b))$. Since $F^{-1}(U^*)$ is equal in distribution to X, applying the function h to both will maintain this equality in distribution: $h(X) \mid (a < X \leq b) \stackrel{d}{=} h\left(F^{-1}\left(U_i^*\right)\right)$

ii) Now suppose $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$. Show that

$$E(X \mid a < X \leq b) = \mu - \sigma \frac{\phi(\beta) - \phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)}$$

$$\operatorname{Var}(X \mid a < X \leq b) = \sigma^2\left[1 - \frac{\beta\phi(\beta) - \alpha\phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)} - \left(\frac{\phi(\beta) - \phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)}\right)^2\right]$$

where $\alpha = (a - \mu)/\sigma$ and $\beta = (b - \mu)/\sigma$, and ϕ and Φ are the density and distribution functions of the standard normal distribution, receptively.

Step 1. $E[X \mid a < X \leq b]$

$$\text{From } f'(x) = \left(-\frac{x - \mu}{\sigma^2}\right) f(x),$$

$$E[X \mid a < X \leqslant b] = \frac{\int_a^b x f(x) dx}{\int_a^b f(x) dx}$$

$$\begin{aligned} \int_a^b x f(x) dx &= -\sigma \int_a^b -\frac{(x - \mu)}{\sigma^2} f(x) dx + \mu \int_a^b f(x) dx \\ &= -\sigma \int_a^b f'(x) dx + \mu[\Phi(\beta) - \Phi(\alpha)] \\ &= -\sigma[\phi(\beta) - \phi(\alpha)] + \mu[\Phi(B) - \Phi(\alpha)] \end{aligned}$$

$$\int_a^b f(x) dx = \Phi(\beta) - \Phi(\alpha)$$

$$\therefore E[X \mid a < X \leqslant b] = \mu - \sigma \frac{\phi(\beta) - \phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)}$$

Step 2. $E[X^2 \mid a < X \le b]$

$$
\begin{aligned}
E\left[X^2 \mid a<X \leqslant b\right] &= \frac{\int_a^b x^2 f(x) d x}{\int_a^b f(x) d x} \\
&= \frac{-\sigma^2}{\int_a^b f(x) d x} \int_a^b x\left(-\frac{x-\mu}{\sigma^2}\right) f(x) d x+\frac{\mu}{\int_a^b f(x) d x} \int_a^b x f(x) d x \\
&= \frac{-\sigma^2}{\int_a^b f(x) d x} \int_a^b x f^{\prime}(x) d x+\mu\left[\mu-2 \sigma \frac{\phi(\beta)-\phi(\alpha)}{\Phi(\beta)-\Phi(\alpha)}\right] \\
&= \frac{-\sigma^2}{\int_a^b f(x) d x}\left[\left.x f(x)\right|_a ^b-\int_a^b f(x) d x\right]+\mu^2-2 \mu \sigma \frac{\phi(\beta)-\phi(\alpha)}{\Phi(\beta)-\Phi(\alpha)} \\
&= \frac{-\sigma^2}{\int_a^b f(x) d x}\left[\beta \phi(\beta)-\alpha \phi(\alpha)-\int_a^b f(x) d x\right]+\mu^2-2 \mu \sigma \frac{\phi(\beta)-\phi(\alpha)}{\Phi(\beta)-\Phi(\alpha)} \\
&= -\sigma^2\left(\frac{\beta \phi(\beta)-\alpha \phi(\alpha)}{\int_a^b f(x) d x}-1\right)+\mu^2-2 \mu \sigma \frac{\phi(\beta)-\phi(\alpha)}{\Phi(\beta)-\Phi(\alpha)} \\
&= \mu^2+\sigma^2-2 \mu \sigma \frac{\phi(\beta)-\phi(\alpha)}{\Phi(\beta)-\Phi(\alpha)}-\sigma^2 \frac{\beta \phi(\beta)-\alpha \phi(\alpha)}{\Phi(\beta)-\Phi(\alpha)}
\end{aligned}
$$

Step 3. $Var[X \mid a < X \le b]$

$$
\begin{aligned}
\operatorname{Var}(X \mid a<X \leqslant b) &= E\left[X^2 \mid a<X \leqslant b\right]-E[X \mid a<X \leqslant b]^2 \\
&= \mu^2+\sigma^2-2 \mu \sigma \frac{\phi(\beta)-\phi(\alpha)}{\Phi(\beta)-\Phi(\alpha)}-\sigma^2 \frac{\beta \phi(\beta)-\alpha \phi(\alpha)}{\Phi(\beta)-\Phi(\alpha)} \\
&\quad -\left(\mu^2-2 \mu \sigma \frac{\phi(\beta)-\phi(\alpha)}{\Phi(\beta)-\Phi(\alpha)}+\sigma^2\left(\frac{\phi(\beta)-\phi(\alpha)}{\Phi(\beta)-\Phi(\alpha)}\right)^2\right) \\
&= \sigma^2\left(1-\frac{\beta \phi(\beta)-\alpha \phi(\alpha)}{\Phi(\beta)-\Phi(\alpha)}-\left(\frac{\phi(\beta)-\phi(\alpha)}{\Phi(\beta)-\Phi(\alpha)}\right)^2\right)
\end{aligned}
$$

iii) Suppose that we estimate $\theta = E[h(X)]$ for some h using stratification on intervals, that is,

$$
\hat{\theta}=\sum_{i=1}^n \operatorname{Pr}\left\{a_{i-1}<X \leq a_i\right\} \hat{E}\left[h(X) \mid a_{i-1}<X \leq a_i\right]
$$

where $-\infty = a_0 < a_1 < \cdots < a_{n-1} < a_n = \infty$. Assume that we estimate each $E\left[h(X) \mid a_{i-1} < X \le a_i\right]$ using only one random draw from $X \mid (a_{i-1} < X \le a_i)$. Explain why it is difficult to find the values of a_i that lead to the optimal allocation even when h is a linear function. [Your answer must be based on mathematical expressions.]

Let $a_{i-1} < X_i \leqslant a_i$. X_i is a sample from each stratum.

For the optimal allocation, we need to draw $n_i = n\frac{p_i\sigma_i}{\sum_{i=1}^n p_i\sigma_i} \approx \frac{p_i s_i}{\sum_{i=1}^n p_i s_i}$ samples from i_{th} stratum where $\sum_{i=1}^n n_i = n,\ p_i = P(Y = y_i)$. (Neyman allocation)

$$
\begin{aligned}
n_i &= n \frac{p_i \sigma_i}{\sum_{i=1}^n p_i \sigma_i} \approx n \frac{p_i s_i}{\sum_{i=1}^n p_i s_i} \\
&= n \frac{p_i\left(1-\frac{\beta \phi(\beta)-\alpha \phi(\alpha)}{\Phi(\beta)-\Phi(\alpha)}-\left(\frac{\phi(\beta)-\phi(\alpha)}{\Phi(\beta)-\Phi(\alpha)}\right)^2\right)}{\sum_{i=1}^n p_i\left(1-\frac{\beta \phi(\beta)-\alpha \phi(\alpha)}{\Phi(\beta)-\Phi(\alpha)}-\left(\frac{\phi(\beta)-\phi(\alpha)}{\Phi(\beta)-\Phi(\alpha)}\right)^2\right)}
\end{aligned}
$$

That means if each stratum has a different conditional variance, then the sample size which is needed each stratum also different. If we draw only 1 sample from each $X \mid (a_{i-1} < X \leqslant a_i)$ as problem, $p_i \left(1 - \frac{\beta\phi(\beta) - \alpha\phi(\alpha)}{\Phi(\beta)-\Phi(\alpha)} - \left(\frac{\phi(\beta)-\phi(\alpha)}{\Phi(\beta)-\Phi(\alpha)}\right)^2\right)$ of each stratum should be same (for optimal allocation). But the closed form of that equation is absent. Therefore, finding an $(a_{i-1}, a_i]$ with the same conditional variance is difficult.

iv) Now, we instead consider the strategy of determining the intervals such that $\Pr\{a_{i-1} < X \leq a_i\} = 1/n$ for all $i = 1, \ldots, n$. Let $X \sim \mathrm{N}\left(-1, 2^2\right)$. Using the fact in (i), estimate $E\left[\log\left(e^X + 1\right)\right]$ with $n = 10^4$. By repeating the procedure 10^5 times, estimate the variance of this estimator. Compare this variance with that of the usual average estimator of $n = 10^4$ obtained by 10^5 simulations.

```
### iv) ###
estimate_usual <- function(num_simulation, n, mean, sd){
  # I will define "usual" function which is
  # estimator E(log(exp(X)+1)) one time
  usual <- function(n, mean, sd){
    normal_samples <- mean + sd * qnorm(runif(n))
    estimator <- mean(log(exp(normal_samples) + 1))
    return (estimator)
  }
  # I repeated usual function "num_simulation" times and get variance
  estimators <- replicate(num_simulation, usual(n, mean, sd))
  result <- list(mean = mean(estimators), var = var(estimators))
  return (result)
}

estimate_stratified <- function(num_simulation, n, mean, sd) {
  # I divided (0,1) into 1e4 intervals
  #   and generated one number from each interval using runif(interval)
  break_points <- seq(0, 1, length.out = n + 1)
  estimators <- replicate(num_simulation, {
    unif_samples <- runif(n, min = head(break_points, -1), max = tail(break_points, -1))
    normal_samples <- mean + sd * qnorm(unif_samples)
    # normal_samples is the samples from N(-1,2^2).
    mean(log(exp(normal_samples) + 1))
  })

  result <- list(mean = mean(estimators), var = var(estimators))
  return (result)
}
```

```
ptm <- proc.time()
result.usual <- estimate_usual(1e5, 1e4, -1, 2)
elapsed.time.usual <- proc.time() - ptm

ptm <- proc.time()
result.stratified <- estimate_stratified(1e5, 1e4, -1, 2)
elapsed.time.stratified <- proc.time() - ptm

cat("The usual average estimator|",
    "mean:", result.usual$mean, ", variance:", result.usual$var, "\n",
    "The stratified estimator|",
```

```
    "mean:", result.stratified$mean, ", variance:", result.stratified$var)
```

```
## The usual average estimator| mean: 0.6424874 , variance: 6.619651e-05
##  The stratified estimator| mean: 0.6424956 , variance: 2.34033e-09
```

```
elapsed.time.usual; elapsed.time.stratified
```

```
##    user  system elapsed
##  33.934   1.838  35.815
```

```
##    user  system elapsed
##  32.256   4.349  36.659
```

If the U (which is used for *qnorm* and generating X) is sampled from 1e4 divided interval, and each interval has same normal density, variance of estimator can be reduced.