Quantiles

---

29 marks

1. Suppose we have a continuous random variable X with distribution function $F_X(x) = Pr(X \le x)$ and quantile function $Q_X(p) = F_X^{-1}(p)$. That is $p = F_X(x) = Pr(X \le x)$ and $p = Pr(X \le Q_X(p)) = F_X(Q_X(p)) = F_X(F_X^{-1}(p)) = p$.

   a. *(4 marks)* Suppose $Y = aX + b$ for some constants $a > 0$ and b. **Prove** that a plot of the parametric curve $(Q_X(p), Q_Y(p))$ for $p \in (0, 1)$ must follow a straight line.

      Give the equation of that line.

   b. *(3 marks)* When $F_X(x)$ and $Q_X(p)$ are the cumulative distribution and quantile functions of the continuous random variable X, show that if $U \sim U(0, 1)$, then

      $$Pr(Q_X(U) \le x) = F_X(x).$$

   c. The above result implies that we could generate n independently and identically distributed (i.i.d.) random realizations X from $F_X(x)$ by generating n i.i.d. random realizations U from $U(0, 1)$ and defining $X = Q_X(U)$.

      In R the function `runif()` will generate uniform pseudo-random numbers.

      (Similarly, `dunif()`, `punif()`, and `qunif()` will return the density, the distribution, and the quantile functions, respectively, for a uniform random variable. See `help("runif")` for details.

      i. *(1 mark)* Write an R function

      ```
      r_unifgenFx <- function(n, qfunction = qnorm) {
                          # Insert your code here
                           }
      ```

      which will generate and return `n` pseudo random observations from the distrubution whose quantile function is the value of the argument `qfunction`. Show your code.

      ii. *(2 marks)* Execute the following code snippets to illustrate your code

      ```
      # make sure we all get the same result
      set.seed(1234567)
      # save the currentgraphical parameters and set `mfrow`
      oldPar <- par(mfrow = c(1,2))

      hist(r_unifgenFx(1000))  # Standard normal

      hist(r_unifgenFx(1000, qfunction = runif))'
      par(oldPar) # Return to original graphical parameters
      ```

      iii. *(2 marks)* Generate a sample of 1000 pseudo-random observations from a Student-t distribution on 3 degrees of freedom generated using `r_unifgen()` (unchanged) and the quantile function of the Student-t. Plot a histogram (appropriately labelled) of the results.

   d. Consider the `quantile()` function in R .

i. *(2 marks)* Explain the values returned by `quantile(mtcars$qsec))`. That is, what does `quantile()` do?

ii. *(2 marks)* Show how `quantile()` could be used to generate 1000 observations from the estimated distribution of `mtcars$qsec`.

iii. *(2 marks)* Would this work for `mtcars$cyl`? Why? Or, why not?

iv. *(4 marks)* Draw side by side (nicely labelled) histograms of `mtcars$qsec` and a sample of 1000 observations drawn from the estimated distribution of `mtcars$qsec`. Comment on how these compare.

v. *(3 marks)* Draw a (nicely labelled) `qqplot()` comparing the above two sets of observations. What do you conclude about their empirical distributions? Why?

vi. *(4 marks)* Suppose interest lay in producing a bootstrap distribution for some estimator $\widetilde{\theta}$. Instead of bootstrapping, how might `quantile()` be used? Which would you recommend and why?