Order Statistics and Quantiles

Dr. Syring

February 13, 2020

Order Statistics

The order statistics of a sample $X_1, X_2, \ldots, X_n$ are the smallest to largest, indicated by $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$.

Distribution of Order Statistics

Suppose X_i are iid for $i = 1, \ldots, n$ with distribution function (CDF) F . Then, the probability that the smallest order statistics is larger than a certain value x can be expressed using F :

$$P(X_{(1)} > x) = \prod_{i=1}^{n} P(X_i > x) = (1 - F(x))^n.$$

Then,

$$F_{X_{(1)}}(x) = P(X_{(1)} < x) = 1 - (1 - F(x))^n.$$

Similarly,

$$F_{X_{(n)}}(x) = P(X_{(n)} < x) = F(x)^n.$$

And, for a general $1 \leq j \leq n$ the cumulative distribution of $X_{(j)}$ is

$$F_{X_{(j)}}(x) = \binom{n}{j} F(x)^j (1 - F(x))^{n-j}.$$

Density function of order statistics

Density of $X_{(j)}$. A counting argument: There are n ways to choose which X_i is $X_{(j)}$. There are $n-1$ left and $j-1$ are smaller than $X_{(j)}$, so

$$f_{X_{(j)}}(x) = nf(x)\binom{n-1}{j-1}F(x)^{j-1}(1-F(x))^{n-j}$$

$$= \frac{n!}{(j-1)!(n-j)!}f(x)F(x)^{j-1}(1-F(x))^{n-j}$$

You can generalize this argument (carry it forward) to find joint densities of 2 or more order statistics.

Example: Uniform

Suppose we have n iid Uniform $(0, 1)$ r.v.'s. Then,

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f(x) F(x)^{j-1} (1 - F(x))^{n-j}$$

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} \times 1 \times x^{j-1} \times (1 - x)^{n-j}$$

which is a Beta distribution with shape parameters $\alpha = j$ and $\beta = n - j + 1$.

Example: Range

The range is $Y_1 = X_{(n)} - X_{(1)}$, and let $Y_2 = X_{(n)}$. The joint density of $(X_{(1)}, X_{(n)})$ is

$$f(x_1, x_n) = n(n-1)f(x_1)f(x_n)[F(x_n) - F(x_1)]^{n-2}.$$

The Jacobian from $(X_{(1)}, X_{(n)})$ to (Y_1, Y_2) is -1 . Then,

$$f(y_1, y_2) = n(n-1)f(y_2 - y_1)f(y_2)[F(y_2) - F(y_2 - y_1)]^{n-2}.$$

Suppose the r.v.'s are iid Unif(0,1), then

$$f(y_1) = \int_{y_1}^{\infty} n(n-1)f(y_2 - y_1)f(y_2)[F(y_2) - F(y_2 - y_1)]^{n-2} dy_2$$

$$f(y_1) = \int_{y_1}^{1} n(n-1)[y_2 - (y_2 - y_1)]^{n-2} dy_2$$

$$= n(n-1)y_1^{n-2}(1 - y_1)$$

which is $Beta(\alpha = n-1, \beta = 2)$.

Quantiles and sample quantiles

The p^{th} quantile of a r.v. X with distribution function F is $x_p = F^{-1}(p)$. One estimator is $X_{(k)}$ where $k = \lfloor p(n+1) \rfloor$. Then,

$$EF(X_{(k)}) = \int_a^b F(x_k) f_{X_{(k)}}(x_k) dx_k$$

$$EF(X_{(k)}) = \int_a^b F(x_k) \frac{n!}{(k-1)!(n-k)!} F(x_k)^{k-1} (1 - F(x_k))^{n-k} f(x_k) dx_k$$

Let $u = F(x_k)$ with $d_u = f(x_k) dx_k$ so that

$$EF(X_{(k)}) = \int_0^1 u \frac{n!}{(k-1)!(n-k)!} u^{k-1} (1-u)^{n-k} du$$

which is the expectation of $Beta(\alpha = k, \beta = n - k + 1)$ which is $\frac{\alpha}{\alpha+\beta} = \frac{k}{n+1}$.
Since $p \approx \frac{k}{n+1}$ $X_{(k)}$ is a nearly unbiased estimator.

Nonparametric CI for pop. quantile

We want to find a confidence interval for x_p the p^{th} quantile of the distribution of a continuous r.v. X . Consider the event $X_{(j)} < x_p < X_{(k)}$ for $j < p(n+1) < k$. Since $F(x_p) = p$ we have

$$P(X_{(j)} < x_p < X_{(k)}) = \sum_{\ell=j}^{k-1} \binom{n}{\ell} p^{\ell}(1-p)^{n-\ell}.$$

For a given coverage $(1-\alpha)$ we can compute the above probability and use sample quantiles $(x_{(j)}, x_{(k)})$ as the interval estimate.

```
data <- c(2,  3,  4,  6, 10, 13, 17, 22, 39, 43)
p <- 1/2
n<-10
binom.int <- function(n,j,k,p){
  s.seq <- seq(from = j, to = k-1, by = 1)
  return(sum(dbinom(s.seq,n,p)))
}
binom.int(10,2,8,1/2)
```

```
## [1] 0.9345703
```

```
binom.int(10,3,9,1/2)
```

```
## [1] 0.9345703
```

```
binom.int(10,4,9,1/2)
```

```
## [1] 0.8173828
```

```
binom.int(10,3,8,1/2)
```

```
## [1] 0.890625
```

```
c(data[2], data[8])
```

```
## [1]  3 22
```

```r
c(data[3], data[9])
```

```
## [1]  4 39
```