STK4011 and STK9011
Autumn 2016

Order statistics

Covers (most of) the following material from chapter 5:

- Section 5.4: pages 226-231

Ørnulf Borgan
Department of Mathematics
University of Oslo

Random sample

The random variables $X_1, X_2, ..., X_n$ are called a random sample from the population $f_X(x)$ if $X_1, X_2, ..., X_n$ are mutually independent and the marginal pmf or pdf of each X_i is $f_X(x)$

Alternatively, we may say that $X_1, X_2, ..., X_n$ are independent and identically distributed (iid) random variables with pmf or pdf $f_X(x)$

The joint pmf or pdf is

$$f(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} f_X(x_i)$$

Order statistics

The order statistics of a random sample $X_1, X_2, ..., X_n$ are the sample values placed in ascending order

The order statistics are denoted $X_{(1)}, X_{(2)}, ..., X_{(n)}$ and they satisfy $X_{(1)} \leq X_{(2)} \leq .... \leq X_{(n)}$

Some statistics defined by the order statistics:

Sample range: $R = X_{(n)} - X_{(1)}$

Sample median: $$M = \begin{cases} X_{((n+1)/2)} & \text{if } n \text{ is odd} \\ \frac{1}{2}\left(X_{(n/2)} + X_{(n/2+1)}\right) & \text{if } n \text{ is even} \end{cases}$$

(100*p*)th sample percentile: $$\begin{cases} X_{(\{np\})} & \text{if } \frac{1}{2n} < p < 0.5 \\ X_{(n+1-\{n(1-p)\})} & \text{if } 0.5 < p < 1 - \frac{1}{2n} \end{cases}$$

{*b*} is *b* rounded to the nearest integer

Other definitions of sample percentiles exist (the command "quantile" in R has 9 options)

Lower quartile: $Q_1 = 25\text{th percentile}$

Upper quartile: $Q_3 = 75\text{th percentile}$

Interquartile range: $Q_3 - Q_1$

We will look at the distribution of the order statistics

Let $X_1, \ldots, X_n$ be a random sample from a continuous distribution with cdf $F_X(x)$ and pdf $f_X(x)$

We start out by considering $X_{(n)} = \max X_i$ and $X_{(1)} = \min X_i$

The cumulative distribution of $X_{(n)}$ is given by

$$F_{X_{(n)}}(x) = P\left(\max X_i \le x\right) = P\left(\bigcap_{i=1}^{n}\{X_i \le x\}\right)$$

$$= \prod_{i=1}^{n} P\left(X_i \le x\right) = \left[F_X(x)\right]^n$$

Hence the pdf becomes

$$f_{X_{(n)}}(x) = n f_X(x)\left[F_X(x)\right]^{n-1}$$

The cumulative distribution of $X_{(1)}$ is given by

$$F_{X_{(1)}}(x) = P\left(\min X_i \le x\right) = 1 - P\left(\min X_i > x\right)$$

$$= 1 - P\left(\bigcap_{i=1}^{n}\{X_i > x\}\right) = 1 - \prod_{i=1}^{n} P\left(X_i > x\right)$$

$$= 1 - \prod_{i=1}^{n}\left[1 - P\left(X_i \le x\right)\right] = 1 - \left[1 - F_X(x)\right]^n$$

Hence the pdf becomes

$$f_{X_{(1)}}(x) = n f_X(x)\left[1 - F_X(x)\right]^{n-1}$$

We then look at the *j*-th order statistic

Theorem 5.4.4

Let $X_1, \ldots, X_n$ be a random sample from a continuous distribution with cdf $F_X(x)$ and pdf $f_X(x)$.

Then the *j*-th order statistic is has cdf

$$F_{X_{(j)}}(x) = \sum_{k=j}^{n}\binom{n}{k}\left[F_X(x)\right]^k\left[1 - F_X(x)\right]^{n-k}$$

and pdf

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!\,(n-j)!} f_X(x)\left[F_X(x)\right]^{j-1}\left[1 - F_X(x)\right]^{n-j}$$

Example: uniform order statistic

Let $X_1, \ldots, X_n$ be iid and uniform(0,1)

Then $f_X(x) = 1$ and $F_X(x) = x$ for $0 < x < 1$

Hence

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!\,(n-j)!} f_X(x)\left[F_X(x)\right]^{j-1}\left[1 - F_X(x)\right]^{n-j}$$

$$= \frac{n!}{(j-1)!\,(n-j)!} 1 \cdot x^{j-1}(1-x)^{n-j}$$

$$= \frac{\Gamma(n+1)}{\Gamma(j)\,\Gamma(n-j+1)} x^{j-1}(1-x)^{(n-j+1)-1}$$

Thus $X_{(j)} \sim \text{beta}(j, n-j+1)$

A result on the joint pdf of two order statistics:

Theorem 5.4.6

Let $X_1,\dots,X_n$ be a random sample from a continuous distribution with cdf $F_X(x)$ and pdf $f_X(x)$.

Then the joint pdf of the *i*-th and *j*-th order statistic is given by ($1 \le i < j \le n$)

$$f_{X_{(i)},X_{(j)}}(u,v)$$
$$= \frac{n!}{(i-1)!\,(j-1-i)!\,(n-j)!}$$
$$\times f_X(u)\,f_X(v)\left[F_X(u)\right]^{i-1}\left[F_X(v)-F_X(u)\right]^{j-1-i}\left[1-F_X(v)\right]^{n-j}$$

(A formal proof is outlined in exercise 5.26)

Range for a uniform distribution (example 5.4.7)

Let $X_1,\dots,X_n$ be iid and uniform(0,1)

Then $f_X(x)=1$ and $F_X(x)=x$ for $0<x<1$

The joint distribution of $X_{(1)}$ and $X_{(n)}$ is given by (for $0<x_1<x_n<1$)

$$f_{X_{(1)},X_{(n)}}(x_1,x_n)=n(n-1)\left(x_n-x_1\right)^{n-2}$$

Introduce the range

$$R=X_{(n)}-X_{(1)}$$

and the midrange

$$V=\left(X_{(1)}+X_{(n)}\right)/2$$

The inverse transformation is given by

$$X_{(1)}=V-R/2 \qquad X_{(n)}=V+R/2$$

The Jacobi equals -1 and the joint pdf of (*R*,*V*) becomes (for $0<r<1,\ r/2<v<1-r/2$)

$$f_{R,V}(r,v)=n(n-1)r^{n-2}$$

The marginal pdf for *R* is given by

$$f_R(r)=\int_{r/2}^{1-r/2} n(n-1)r^{n-2}dv \;=n(n-1)r^{n-2}(1-r)$$

Thus $R\sim \text{beta}(n-1,2)$

We also have a result on the joint pdf of all n order statistics:

Let $X_1,\dots,X_n$ be a random sample from a continuous distribution with pdf $f_X(x)$

Then the joint pdf of all the order statistic is given by

$$f_{X_{(1)},\dots,X_{(n)}}(x_1,\dots,x_n)=\begin{cases} n!\,f_X(x_1)\cdot\ldots\cdot f_X(x_n) & \text{if } x_1<\ldots<x_n \\ 0 & \text{otherwise}\end{cases}$$