ORDER STATISTICS AND RELATED MODELS

Director: Fernando López Blázquez

Dpto. Estadística e I.O.

Facultad de Matemáticas

Universidad de Sevilla

Tarfia, s/n, 41012 Sevilla, Spain

e-mail: lopez@us.es

N. Balakrishnan

Department of Mathematics and Stattistics

McMaster University

Hamilton, Ontario. Canada L8S 4K1

e-mail: bala@univmail.cis.mcmaster.ca

url: http://www.math.mcmaster.ca/bala/bala.html

Erhard Cramer and Udo Kamps

Institut für Statistik und Wirtschaftsmahematik

RWTH Aachen D-52056 Aachen. Germany

e-mail: erhard.cramer@rwth-aachen.de

kamps@stochastik.rwth-aachen.de

url: http://www.stochastik.rwth-aachen.de/index.php

A Course on Order Statistics and Related Models

by N. Balakrishnan, E. Cramer, U. Kamps and F. López Blazquez,

Sala de Juntas de la Facultad de Matemáticas.
Universidad de Sevilla.

Monday, Nov 27 (12:00h)

1. Welcome address
2. Order Statistics: Basic Properties and Distributions (60 min, F. López Blázquez))
3. Other Models for Ordered Data (45 min, Udo Kamps)

Monday, Nov 27 (16:00h)

4. Inferential Issues with Order Statistics (60 min, Balakrishnan)
5. Extreme Value Theory (30 min, E. Cramer)

Tuesday, Nov 28 (12:00h)

6. Progressively Censored Order Statistics (90 min, E. Cramer)

Tuesday, Nov 28 (16:00h)

7. Experimental Design in Progressive Censoring (45 min, E. Cramer)
8. Step-stress Models (45 min, Balakrishnan)

Wednesday, Nov 29 (12:00h)

9. Sequential Order Statistics (90 min, Udo Kamps)

Wednesday, Nov 29 (16:00h)

10. Ordered Random Variables and Aging Properties (30 min, Udo Kamps)
11. Recent Developments and Future Work (30 min, Balakrishnan)
12. Discussion (30 min, all)

Contents

Bibliography

Chapter 1

Order statistics. Basic distribution theory

1.1 Introduction

Definition 1.1.1 *Let $X_1, \ldots, X_n$ be n random variables. Suppose that they are arranged in increasing order to obtain*

$$X_{1:n} \leq \ldots \leq X_{i:n} \leq \ldots \leq X_{n:n}.$$

*Then the random variable (rv) $X_{i:n}$ is called the **ith-order statistics** of the sample of size n.*

In the following we will assume that $X_1, \ldots, X_n$ are independent and identically distributed random variables with common cumulative distribution function (cdf) F. We will distinguish two cases: the absolutely continuous case and the discrete case.

1.2 Distribution theory in the absolutely continuous case

Let $X_1, \ldots, X_n$ be iid rv's from a cdf F and with density function (df) f.

1.2.1 Distribution of a single order statistics

The distribution function of a single order statistics $X_{i:n}$, denoted by $F_{i:n}$, is obtained as follows:

$$\begin{aligned} F_{i:n}(x) &= P(X_{i:n} \leq x) \\ &= P(\text{at least } i \text{ observations among } X_1, \ldots, X_n \text{ are } \leq x) \\ &= \sum_{k=i}^{n} \binom{n}{k} F^k(x)\,(1 - F(x))^{n-k} \end{aligned} \tag{1.1}$$

and the density function is then:

$$f_{i:n}(x) = F'_{i:n}(x) = C_{i:n} F^{i-1}(x)\,(1 - F(x))^{n-i} f(x),$$

where

$$C_{i:n} = \frac{n!}{(i-1)!(n-i)!} = \frac{1}{\beta(i, n-i+1)}.$$

Particular cases of some interest in the applications are:

- The **maximum**, $X_{n:n}$, with distribution and density functions respectively:

$$F_{n:n}(x) = F^n(x);$$
$$f_{n:n}(x) = nF^{n-1}(x)f(x).$$

- The **minimum**, $X_{1:n}$, with distribution and density functions respectively:

$$F_{1:n}(x) = 1 - (1 - F(x))^n;$$
$$f_{1:n}(x) = n(1 - F(x))^{n-1} f(x).$$

- In particular, if $U_1, \ldots, U_n$ are iid from an uniform distribution in the unit interval, ($\mathcal{U}(0,1)$ distribution), then the ith order statistics, $U_{i:n}$, follows a beta distribution with parameters i and $n-i+1$, i.e., $U_{i:n} \sim Be(i, n-i+1)$, with density

$$f_{i:n}(x) = \frac{1}{\beta(i, n-i+1)} x^{i-1}(1-x)^{n-i},\ x \in (0,1),$$

and its distribution function can be obtained using the so called **incomplete beta function:**

$$\begin{aligned} F_{i:n}(x) &= IBe(x; i, n-i+1) \\ &= \frac{1}{\beta(i, n-i+1)} \int_0^x s^{i-1}(1-s)^{n-i} ds,\ x \in (0,1). \end{aligned}$$

1.2.2 Joint density of two order statistics

- It can be checked that the joint density of two order statistics $(X_{i:n}, X_{j:n})$ with $1 \leq i < j \leq n$ is

$$f_{i,j:n}(x,y) = C_{i,j:n} F^{i-1}(x) \left(F(y) - F(x)\right)^{j-i-1} (1 - F(y))^{n-j} f(x) f(y),$$

  for $x < y$, with

$$C_{i,j:n} = \frac{n!}{(i-1)!(j-i-1)!(n-j)!}.$$

- In particular the **joint density of the maximum and the minimum,** $(X_{1:n}, X_{n:n})$ is

$$f_{1,n:n}(x,y) = n(n-1) \left(F(y) - F(x)\right)^{n-2} f(x) f(y), \text{ for } x < y.$$

- The **joint density of two consecutive order statistics,** $(X_{i:n}, X_{i+1:n})$ is:

$$f_{i,i+1:n}(x,y) = \frac{n!}{(i-1)!(n-i-1)!} F^{i-1}(x)(1 - F(y))^{n-i-1} f(x) f(y), \text{ for } x < y.$$

1.2.3 Joint density of all the order statistics

- The joint density of all the order statistics is:

$$f_{1,2,\ldots,n:n}(x_1, x_2, \ldots, x_n) = n! \prod_{i=1}^{n} f(x_i), \text{ for } x_1 < x_2 < \ldots < x_n \tag{1.2}$$

  From this joint density we could have obtained, by integration, the density of a single order statistics, or the joint of two order statistics.

- In general, the joint density of k order statistics, $1 \leq i_i < \cdots < i_k \leq n$, from a sample of size n is:

$$f_{i_i,i_2,\ldots,i_k:n}(x_{i_1}, x_{i_2}, \ldots, x_{i_k}) =$$

$$C_{i_i,\cdots,i_k:n} F^{i_1-1}(x_{i_1}) \prod_{j=2}^{k} \left(F(x_{i_j}) - F(x_{i_j-1})\right)^{i_j - i_{j-1} - 1} (1 - F(x_{i_k}))^{n-i_k} \prod_{j=1}^{k} f(x_{i_j})$$

  for $x_{i_1} < x_{i_2} < \ldots < x_{i_k}$, with

$$C_{i_i,\cdots,i_k:n} = \frac{n!}{(i_1 - 1)! \prod_{j=2}^{k} (i_j - i_{j-1} - 1)!(n - i_k)!}.$$

1.3 Some properties of order statistics

First, we introduce the quantile function (or generalized inverse):

Definition 1.3.1 *Let F be a distribution function. The quantile function is:*

$$u \in (0,1],\ Q_F(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}.$$

For any cdf F, Q_F is non-decreasing and right-continuous. If F is continuous, then Q_F is continuous. If F is strictly increasing, then Q_F is the inverse function F^{-1}. To our purposes, the most important property of the quantile function is:

Theorem 1.3.2 *(Quantile transformation). Let X be a rv with cdf F. Let $U \sim \mathcal{U}(0,1)$. Then, the cdf of the rv $Q_F(U)$ is F, or in other words*

$$X \stackrel{d}{=} Q_F(U).$$

- Let $X_{1:n}, \ldots, X_{n:n}$ be the order statistics of n iid observations from a rv X with distribution function F. Consider the transformed random variable $Y = g(X)$, with g a Borel-measurable function. As the order is preserved by non-decreasing functions, we have:

$$(Y_{1:n}, \ldots, Y_{n:n}) \stackrel{d}{=} (g(X_{1:n}), \ldots, g(X_{n:n})),$$

  for any non-decreasing measurable function g.

- In particular, the order statistics from a $\mathcal{U}(0,1)$-distribution are intimately connected with the order statistics from another random variable $X \sim F$:

$$(X_{1:n}, \ldots, X_{n:n}) \stackrel{d}{=} (Q_F(U_{1:n}), \ldots, Q_F(U_{n:n})). \tag{1.3}$$

- From (1.3), we have:

$$X_{i:n} \stackrel{d}{=} Q_F(U_{i:n}),$$

  and consequently the cdf of $X_{i:n}$ can be written as

$$F_{i:n}(x) = IBe(F(x); i, n-i+1).$$

- Let g be a non-increasing measurable function and consider the transformed random variable $Y = g(X)$. As the order is inverted,

$$(Y_{1:n}, \ldots, Y_{n:n}) \stackrel{d}{=} (g(X_{n:n}), \ldots, g(X_{1:n})) \,. \tag{1.4}$$

- From (1.4),

$$Y_{i:n} \stackrel{d}{=} g(X_{n-i+1:n}).$$

- For instance,

$$-X_{i:n} \stackrel{d}{=} (-X)_{n-i+1:n},$$

- Moreover, if X is a symmetric rv (with respect to the origin), that is to say $X \stackrel{d}{=} -X$, we have

$$X_{i:n} \stackrel{d}{=} -X_{n-i+1:n}.$$

- Similarly, for symmetric random variables

$$(X_{i:n}, X_{j:n}) \stackrel{d}{=} (-X_{n-i+1:n}, -X_{n-j+1:n}).$$

- If the rv is symmetric with respect to other point rather than the origin, similar arguments as before can be made with slight changes.

1.4 Conditional distributions in the absolutely continuous case

Definition 1.4.1 *Let F be a cdf and $x_0 \in \mathbb{R}$, the distribution F truncated on the right at x_0 is the distribution function*

$$F^{(x_0^+)}(y) = \begin{cases} \frac{F(y)}{F(x_0)}, & y < x_0 \\ 1, & y \geq x_0. \end{cases} \tag{1.5}$$

- Let $X \sim F$. We will denote by $X^{(x_0^+)}$ a rv having cdf $F^{(x_0^+)}$ defined as in (1.5). For instance, the rv X conditional to the event $\{X \leq x_0\}$ has cdf $F^{(x_0^+)}$ and we denote this fact

$$\{X \mid X \leq x_0\} \sim F^{(x_0^+)}, \quad \text{or} \quad \{X \mid X \leq x_0\} \stackrel{d}{=} X^{(x_0^+)}.$$

Similarly,

Definition 1.4.2 *Let F be a cdf and $x_0 \in \mathbb{R}$, the distribution F truncated on the left at x_0 is the distribution function*

$$F^{(x_0^-)}(y) = \begin{cases} \frac{F(y)-F(x_0)}{1-F(x_0)}, & y > x_0 \\ 0, & y \leq x_0. \end{cases} \quad (1.6)$$

- Let $X \sim F$. We will denote by $X^{(x_0^-)}$ a rv having cdf $F^{(x_0^-)}$ defined as in (1.6). For instance, the rv X conditional to the event $\{X > x_0\}$ has cdf $F^{(x_0^-)}$ and we denote this fact

$$\{X \mid X > x_0\} \sim F^{(x_0^-)}, \quad \text{or} \quad \{X \mid X > x_0\} \stackrel{d}{=} X^{(x_0^-)}.$$

Theorem 1.4.3 *Let $X_{1:n}, \ldots, X_{n:n}$ be the order statistics from an absolutely continuous distribution F. Then, the conditional distribution of the vector $(X_{1:n}, \ldots, X_{j-1:n})$ given $X_{j:n} = x_j$ is the same as the joint distribution of the $j-1$ order statistics of a sample of size $j-1$ from the population $X^{(x_j^+)}$, that is to say:*

$$\{(X_{1:n}, \ldots, X_{j-1:n}) \mid X_{j:n} = x_j\} \stackrel{d}{=} \left(X_{1:j-1}^{(x_j^+)}, \ldots, X_{j-1:j-1}^{(x_j^+)} \right).$$

- As a consequence of the previous result, for $1 \leq i < j \leq n$,

$$\{X_{i:n} \mid X_{j:n} = x_j\} \stackrel{d}{=} X_{i:j-1}^{(x_j^+)},$$

that is to say, the conditional density of $X_{i:n}$ given $X_{j:n} = x_j$ is:

$$f_{i|X_{j:n}=x_j}(x_i) = C_{i:j-1} \left(\frac{F(x_i)}{F(x_j)} \right)^{i-1} \left(\frac{F(x_j) - F(x_i)}{F(x_j)} \right)^{j-i-1} \frac{f(x_i)}{F(x_j)}, \; x_i < x_j.$$

Theorem 1.4.4 *Let $X_{1:n}, \ldots, X_{n:n}$ be the order statistics from an absolutely continuous distribution F. Then, the conditional distribution of the vector $(X_{i+1:n}, \ldots, X_{n:n})$ given $X_{i:n} = x_i$ is the same as the joint distribution of the $n-i$ order statistics of a sample of size $n-i$ from the population $X^{(x_i^-)}$, that is to say:*

$$\{(X_{i+1:n}, \ldots, X_{n:n}) \mid X_{i:n} = x_i\} \stackrel{d}{=} \left(X_{1:n-i}^{(x_i^-)}, \ldots, X_{n-i:n-i}^{(x_i^-)}\right).$$

- As a consequence of the previous result, for $1 \leq i < j \leq n$,

$$\{X_{j:n} \mid X_{i:n} = x_i\} \stackrel{d}{=} X_{j-i:n-i}^{(x_i^-)},$$

that is to say, the conditional density of $X_{j:n}$ given $X_{i:n} = x_i$ is:

$$f_{j|X_{i:n}=x_i}(x_j) = C_{j-i:n-i}\left(\frac{F(x_j)-F(x_i)}{1-F(x_i)}\right)^{j-i-1}\left(\frac{1-F(x_j)}{1-F(x_i)}\right)^{n-j}\frac{f(x_j)}{1-F(x_i)},$$

for $x_j > x_i$.

1.5 Markov property

When the original iid variables $X_1, \ldots, X_n \sim F$ are ordered, the corresponding order statistics, $X_{1:1}, \ldots, X_{n:n}$ are no longer independent. For absolutely continuous distributions, the dependence structure can be described by a Markov Chain.

Theorem 1.5.1 *If F is absolutely continuous, the order statistics, $X_{1:1}, \ldots, X_{n:n}$, form a (discrete time) Markov chain with transition densities:*

$$f_{i+1|i}(y \mid x) = (n-i)\left(\frac{F(y)-F(x)}{1-F(x)}\right)^{n-i-1} \frac{f(y)}{1-F(x)},$$

for $y > x$; $i = 1, \ldots, n-1$.

Although the order statistics are dependent, they satisfy the following conditional independence property:

Theorem 1.5.2 *Let F be absolutely continuous. For any $1 < k < n$, the random vectors*

$$\boldsymbol{X}^{(1)} = (X_{1:n}, \ldots, X_{k-1:n}) \text{ and } \boldsymbol{X}^{(2)} = (X_{k+1:n}, \ldots, X_{n:n})$$

are conditionally independent given that $X_{k:n} = x_k$, that is to say

$$P[\boldsymbol{X}^{(1)} \in B_1,\ \boldsymbol{X}^{(2)} \in B_2 \mid X_{k:n} = x_k] =$$

$$P[\boldsymbol{X}^{(1)} \in B_1 \mid X_{k:n} = x_k] P[\boldsymbol{X}^{(2)} \in B_2 \mid X_{k:n} = x_k],$$

for any $B_1 \in \mathcal{B}(\mathbb{R}^{k-1})$ and $B_2 \in \mathcal{B}(\mathbb{R}^{n-k})$.

1.6 Distribution of order statistics from discrete distributions

Let X be a discrete random variable with probability function $p_x = P[X = x]$. We define,

$$u_x = P[X \leq x] = F(x), v_x = P[X < x] = F(x^-)$$
$$\text{and } w_x = P[X > x] = 1 - F(x).$$

Consider $X_1, \ldots, X_n$ iid rv distributed as X and the corresponding order statistics $X_{1:1} \leq \ldots \leq X_{n:n}$. The distribution function of the ith order statistics can be obtained using the same argument as in (1.1) to obtain:

$$F_{i:n}(x) = \sum_{k=i}^{n} \binom{n}{k} u_x^k w_x^{n-k},$$

from which the probability function of $X_{i:n}$ is given by:

$$P[X_{i:n} = x] = F_{i:n}(x) - F_{i:n}(x^-).$$

There are some other arguments that permits to obtain the probability function of $X_{i:n}$:

- **Integral representation:** Let Q_F be the quantile function of the discrete random variable X. Then,

$$P[X_{i:n} = x] = P[Q_F(U_{i:n}) = x] = P[F(x^-) < U_{i:n} \leq F(x)] =$$
$$= \frac{1}{\beta(i, n-i+1)} \int_{F(x^-)}^{F(x)} s^{i-1}(1-s)^{n-i} ds.$$

- **Multinomial argument:** Given a $X_1, \ldots, X_n$ from a discrete distribution F, we can define a random vector (I_1, I_2, I_3), where I_i, $i = 1, 2, 3$, represents the number of observations in the sample $X_1, \ldots, X_n$ which are respectively (strictly) less than x, equal to x and (strictly) large than x. It is clear that

$$(I_1, I_2, I_3) \sim \mathcal{M}(n; v_x, p_x, w_x).$$

Then,

$$P[X_{i:n} = x] = P[I_1 < i,\ I_1 + I_2 \geq i] =$$
$$= \sum_{k=0}^{i-1} \sum_{j=i-k}^{n} \binom{n}{k, j, n-k-j} v_x^k p_x^j w_x^{n-k-j}.$$

- The joint probability function of two order statistics, let's say $X_{i:n}$ and $X_{j:n}$ with $1 \leq i < j \leq n$ can be obtained using the multinomial argument. For $x < y$, define the random vector $(I_1, I_2, I_3, I_4, I_5)$, where I_i, $i = 1, \ldots, 5$, represents the number of observations in the sample $X_1, \ldots, X_n$ which are respectively (strictly) less than x, equal to x, strictly between x and y, equal to y and (strictly) large than y. In this case,

$$(I_1, I_2, I_3, I_4, I_5) \sim \mathcal{M}(n; v_x, p_x, v_y - u_y, p_y, w_y).$$

Then,

$$P[X_{i:n} = x,\, X_{j:n} = y] =$$
$$P[I_1 < i,\, I_1 + I_2 \geq i,\, I_1 + I_2 + I_3 < j, I_1 + I_2 + I_3 + I_4 \geq j].$$

Example 1.6.1 *Suppose that X is a discrete rv taking non-negative integer values with pmf: $p_k = P[X = k]$, $k \geq 0$. For an iid sample of size $n = 2$,*

$$P[X_{1:2} = j, X_{2:2} = k] = \begin{cases} 2p_j p_k, & \text{for } 0 \leq j < k \\ p_j^2, & \text{for } j = k \geq 0 \\ 0, & \text{otherwise.} \end{cases} \tag{1.7}$$

From (1.7), it can be easily obtained

$$P[X_{1:2} = j] = p_j^2 + 2p_j w_j, \;\; j \geq 0,$$
$$P[X_{2:2} = k] = p_k^2 + 2p_k v_k, \;\; k \geq 0.$$

- The dependence structure between order statistics in the discrete case is different to the absolutely continuous case. In fact, if the discrete random variable X has more than three points in its support, it can be shown, (see Arnold, Balakrishnan, Nagaraja, (1992), p.48) that:

$$P[X_{i+1:n} = x \mid X_{i:n} = y, X_{i-1:n} = z] < P[X_{i+1:n} = x \mid X_{i:n} = y].$$

- This non-Markovian behavior is due to the fact that in the discrete case ties between observations are possible. By extending the state space, it is possible to obtain a Markov process, for instance, Rüschendorf (1985) has shown that $\{(X_{i:n}, M_i)\}_{i=1}^{n}$ with M_i=*the number of* $X_{k:n}$ *´s with* $k \leq i$ *that are tied with* $X_{i:n}$, is a bivariate Markov process. Also Nagaraja (1986) has shown that conditioning on the event that all the X_i ´s are different, the order statistics of discrete distributions behave as a Markov Chain.

1.7 Order statistics in sampling without replacement

Up to this point we have assumed that $X_1, \ldots, X_n$ are iid random variables. When we choose a sample, $X_1, \ldots, X_n$ of size n without replacement from a finite population of size $N \geq n$, say $x_1 < x_2 < \cdots < x_N$, it is well known that the X_i ´s are identically distributed but not independent. We will determine the distribution of $X_{i:n}$ under this sampling scheme.

- For that, we use a **hypergeometric argument**: for x_k (one of the elements of the finite population), we consider the vector (I_1, I_2, I_3), where I_i, $i = 1, 2, 3$, represents the number of observations in the sample $X_1, \ldots, X_n$ which are respectively (strictly) less than x_k, equal to x_k and (strictly) large than x_k. It follows that,

$$(I_1, I_2, I_3) \sim \mathcal{H}(n; k-1, 1, N-k),$$

  then

$$P[X_{i:n} = x_k] = P[I_1 = i-1,\ I_2 = 1,\ I_3 = n-i] = \\ = \frac{\binom{k-1}{i-1}\binom{N-k}{n-i}}{\binom{N}{n}}.$$

- Similarly, the joint probability function of two order statistics is for $i < j$ and $k < l$:

$$P[X_{i:n} = x_k,\ X_{j:n} = x_l] = \\ = \frac{\binom{k-1}{i-1}\binom{l-k-1}{j-i-1}\binom{N-l}{n-j}}{\binom{N}{n}}.$$

1.8 Order statistics from some specific distributions

1.8.1 Exponential distribution

A random variable X has the exponential distribution with parameter $\lambda > 0$ if its density function is

$$f(x;\lambda) = \lambda \exp(-\lambda x),\ x > 0.$$

Then, we denote $X \sim Exp(\lambda)$. We will restrict our study to the standard exponential distribution, which is obtained when $\lambda = 1$, because it is easy to show that $X \sim Exp(\lambda)$ iff $W = \lambda X \sim Exp(1)$.

Theorem 1.8.1 *Let $W_{1:n},\ldots,W_{n:n}$ be the order statistics corresponding to n iid rv's from a standard exponential distribution. Then*

$$(W_{1:n},\ldots,W_{n:n}) \stackrel{d}{=} \left(\frac{Z_1}{n},\frac{Z_1}{n}+\frac{Z_2}{n-1},\ldots,\frac{Z_1}{n}+\frac{Z_2}{n-1}+\cdots+Z_n\right),$$

where Z_i, $i=1,\ldots,n$ are iid standard exponential random variables.

Some consequences of the previous theorem are:

- $W_{k:n} \stackrel{d}{=} \frac{Z_1}{n}+\frac{Z_2}{n-1}+\cdots+\frac{Z_k}{n-k+1}$.
- $W_{1:n} \stackrel{d}{=} \frac{Z_1}{n}$, that is to say, the minimum of (iid) standard exponential variables follows a exponential distribution with parameter n (or with mean $1/n$).
- It is also immediate that

$$(W_{1:n},\, W_{2:n}-W_{1:n},\ldots,W_{n:n}-W_{n-1:n}) \stackrel{d}{=} \left(\frac{Z_1}{n},\frac{Z_2}{n-1},\ldots,Z_n\right),$$

and

$$(nW_{1:n},\,(n-1)(W_{2:n}-W_{1:n}),\ldots,W_{n:n}-W_{n-1:n}) \stackrel{d}{=} (Z_1,\,Z_2,\ldots,Z_n), \tag{1.8}$$

- In general a difference between two order statistics is called a **spacing** $S_k = W_{k:n} - W_{k-1:n},\ k = 1, \dots, n$, (we agree that $W_{0:n} = 0$).

- So, the previous results show that spacings between consecutive order statistics from a exponential distribution are independent.

- Moreover, $(n - k + 1)S_k, k = 1, \dots, n$ are *iid* standard exponential rv's.

1.8.2 Uniform distribution

A random variable X has the uniform distribution on the interval (a, b), $a < b$, if its density function is

$$f(x; a, b) = \frac{1}{b - a}, \; x \in (a, b).$$

Then, we denote $X \sim \mathcal{U}(a, b)$.

We will restrict our study to the uniform distribution on the unit interval, which is obtained when $a = 0$ and $b = 1$, because it is easy to show that $X \sim \mathcal{U}(a, b)$ iff $U = (X - a)/(b - a) \sim \mathcal{U}(0, 1)$.

Let $U_{1:n}, \ldots, U_{n:n}$ denote the order statistics corresponding to n iid rv's from a uniform distribution on the unit interval.

- Recall that $U_{i:n} \sim Be(i, n-i+1)$.

- As the uniform distribution on the unit interval is symmetric with respect to 1/2, (that is $U \stackrel{d}{=} 1-U$),

$$U_{i:n} \stackrel{d}{=} 1 - U_{n-i+1:n}$$

- The joint distribution of the order statistics $U_{1:n}, \ldots, U_{n:n}$ is described in the following theorem:

Theorem 1.8.2 *Let $U_{1:n}, \ldots, U_{n:n}$ be the order statistics corresponding to n iid rv's from a uniform distribution on the unit interval. Then*

$$(U_{1:n}, \ldots, U_{n:n}) \stackrel{d}{=} \left(\frac{S_1}{S_{n+1}}, \frac{S_2}{S_{n+1}}, \ldots, \frac{S_n}{S_{n+1}}\right),$$

where $S_m = Z_1 + Z_2 + \cdots + Z_m$, $m = 1, \ldots, n$ and Z_i, $i = 1, \ldots, n$ are iid standard exponential random variables.

Consider the spacings between consecutive order statistics: $T_k = U_{k:n} - U_{k-1:n},\ k = 1, \ldots, n+1$, (here, we agree that $U_{0:n} = 0$ and $U_{n+1:n} = 1$). So, easily, from Theorem 1.8.2, we have,

$$(T_1, \ldots, T_{n+1}) \stackrel{d}{=} \left(\frac{Z_1}{S_{n+1}}, \frac{Z_2}{S_{n+1}}, \ldots, \frac{Z_{n+1}}{S_{n+1}} \right).$$

As a consequence:

- Spacings are equally distributed.
- Although they are not independent, they are symmetrically dependent in the sense that the structure of dependence within any pair (T_i, T_j) is always the same, i.e., does not depend on i, j $(i \neq j)$. More generally, all the possible k-dimensional subvectors of the form $(T_{i_1}, \ldots, T_{i_k})$, with $i_j \neq i_l$ for $j \neq l$ are equally distributed.
- Suppose that the $(n+1)$ spacings are ordered to obtain: $T_{1:n+1} \leq \ldots \leq T_{n+1:n+1}$. Then,

  $$T_{k:n+1} \stackrel{d}{=} \frac{Z_{k:n+1}}{S_{n+1}}$$

  and from the results of subsection (1.8.1):

  $$T_{k:n+1} \stackrel{d}{=} \frac{\frac{Z_1}{n+1} + \frac{Z_2}{n} + \cdots + \frac{Z_k}{n-k+2}}{S_{n+1}}.$$

- Due to the quantile transformation, the order statistics from a (standard) exponential distribution are related to the order statistics of the (unit) uniform by

$$(W_{n:n},\ldots,W_{1:n}) \stackrel{d}{=} (-\log U_{1:n},\ldots,-\log U_{n:n}). \tag{1.9}$$

  We showed before, see (1.8), that

$$(nW_{1:n},\,(n-1)(W_{2:n}-W_{1:n}),\ldots,W_{n:n}-W_{n:n-1}) \stackrel{d}{=} (Z_1,\,Z_2,\ldots,Z_n). \tag{1.10}$$

  Combining (1.9) and (1.10), we get

$$\left(\frac{U_{1:n}}{U_{2:n}},\left(\frac{U_{2:n}}{U_{3:n}}\right)^2,\ldots,(U_{n:n})^n\right) \stackrel{d}{=} (V_1,\ldots,V_n), \tag{1.11}$$

  where $V_1,\ldots,V_n$ are iid $\mathcal{U}(0,1)$.

- From (1.11),

$$U_{k:n} \stackrel{d}{=} \prod_{j=k}^{n} V_j^{1/j}, \tag{1.12}$$

  where $V_1,\ldots,V_n$ are iid $\mathcal{U}(0,1)$.

- It is not difficult to show that for $\alpha\in(0,1)$, and $U\sim\mathcal{U}(0,1)$, the transformed random variable $B=U^\alpha$ follows a $Be(1/\alpha,1)$ distribution. Then, from (1.12)

$$U_{k:n} \stackrel{d}{=} \prod_{j=k}^{n} B_j,$$

  where B_j, $j=1,\ldots,n$ are independent $Be(j,1)$ distributions.

1.9 Moments of order statistics

Let $X_1, \ldots, X_n$ be iid from a common distribution F. We showed in section 1.3 that

$$(X_{1:n}, \ldots, X_{n:n}) \stackrel{d}{=} (Q_F(U_{1:n}), \ldots, Q_F(U_{n:n})), \tag{1.13}$$

where $(U_{1:n}, \ldots, U_{n:n})$, represents the order statistic from n iid $\mathcal{U}(0,1)$ rv's. This fact permits to write the moments of $X_{i:n}$, provided that they exists, as:

$$\mu_{i:n}^{(r)} = EX_{i:n}^r = \frac{1}{\beta(i, n-i+1)} \int_0^1 Q_F^r(s) s^{i-1} (1-s)^{n-i} ds.$$

Sen (1959) gave a sufficient condition for the existence of moments of order statistics:

Theorem 1.9.1 *If $E|X|^\alpha < \infty$ for some $\alpha > 0$, then the moment $\mu_{i:n}^{(r)}$ exists for all i such that*

$$\frac{r}{\alpha} \leq i \leq n + 1 - \frac{r}{\alpha}.$$

An easy instance of Theorem (1.9.1) occurs when $\alpha = r$; in this case we can conclude that if $E|X|^r < \infty$ for some $r > 0$, then $\mu_{i:n}^{(r)}$ exists for all $1 \leq i \leq n$.

The variance of $X_{i:n}$ will be denoted $\sigma^2_{i:n}$. Besides the ordinary moments and variances we can consider multivariate moments as:

$$\mu_{i,j:n} = EX_{i:n}X_{j:n},$$

covariances:

$$\sigma_{i,j:n} = Cov(X_{i:n}, X_{j:n}) = \mu_{i,j:n} - \mu_{i:n}\mu_{j:n},$$

and correlation coefficients:

$$\rho_{i,j:n} = \frac{\sigma_{i,j:n}}{\sigma_{i:n}\sigma_{j:n}}.$$

Example 1.9.2 *Let $Z_{1:n} \leq \ldots \leq Z_{n:n}$ be the order statistics from a standard exponential distribution, then:*

$$\begin{aligned}
\mu_{i:n} &= \frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{n-i+1}, \\
\sigma^2_{i:n} &= \frac{1}{n^2} + \frac{1}{(n-1)^2} + \cdots + \frac{1}{(n-i+1)^2}, \\
\mu_{i,j:n} &= \sigma^2_{i:n}, \ \textit{for } j \geq i.
\end{aligned}$$

Example 1.9.3 *Let $U_{1:n} \leq \ldots \leq U_{n:n}$ be the order statistics from a unit uniform distribution, then:*

$$\begin{aligned}
\mu^{(\alpha)}_{i:n} &= \frac{\beta(i+\alpha, n-i+1)}{\beta(i, n-i+1)}. \\
\mu_{i:n} &= \frac{i}{n+1}, \\
\sigma^2_{i:n} &= \frac{i(n-i+1)}{(n+1)^2(n+2)}, \\
\mu_{i,j:n} &= \frac{i(n-j+1)}{(n+1)^2(n+2)}, \ \textit{for } j \geq i. \\
\rho^2_{i,j:n} &= \frac{i(n-j+1)}{j(n-i+1)}, \ \textit{for } j \geq i.
\end{aligned}$$

1.10 Limit distributions of extreme order statistics

1.10.1 Limits for maxima

Let $\{X_i\}_{i\geq 1}$ be a sequence of iid rv's from a cdf F. The problem here is to find sequences of real numbers $\{a_n\}_{n\geq 1}$ and $\{b_n\}_{n\geq 1}$ with $b_n > 0$ such that $(X_{n:n} - a_n)/b_n$ converges in distribution to a non-degenerate distribution.

A distribution function G is said to be $F-$type, with F another cdf, if: for certain $\mu \in \mathbb{R}$ and $\sigma > 0$

$$G(x) = F\left(\frac{x-\mu}{\sigma}\right), \text{ for all } x \in \mathbb{R}.$$

The following theorem is due to Gnedenko:

Theorem 1.10.1 *The set of all non-degenerate limit distributions of* $(X_{n:n} - a_n)/b_n$ *consists only of distributions that belong to the types:*

1. $\Lambda(x) = \exp(-\exp(-x)),\ x \in \mathbb{R}$
2. $\Phi_\alpha(x) = \exp(-x^{-\alpha})$, *for* $x > 0$ *and* $\alpha > 0$.
3. $\Psi_\alpha(x) = \exp(-(-x)^{\alpha})$, *for* $x < 0$ *and* $\alpha > 0$.

A distribution function F for which the limit distribution of $(X_{n:n} - a_n)/b_n$ is $\Lambda-$type (or Φ_α-type or Ψ_α-type) is said to belong to the domain of attraction of Λ (resp. Φ_α or Ψ_α) and this fact is denoted $F \in \mathcal{D}(\Lambda)$ (resp. $F \in \mathcal{D}(\Phi_\alpha)$ or $F \in \mathcal{D}(\Psi_\alpha)$).

There are necessary and sufficient conditions to establish when a given distribution F belong to one of the domains of attraction above described, see Galambos (1978, 1987). For practical purposes we present only sufficient conditions.

Theorem 1.10.2 *Let F be a cdf with positive derivative for all $x \geq x_0$. If for some $\alpha > 0$*

$$\lim_{x \to \infty} \frac{xF'(x)}{1 - F(x)} = \alpha,$$

then $F \in \mathcal{D}(\Phi_\alpha)$. The normalizing constants can be taken as

$$a_n = Q_F\left(1 - \frac{1}{n}\right), \quad b_n = 0.$$

Theorem 1.10.3 *Let F be a cdf with positive derivative in some interval (x_0, x_1) and $F'(x) = 0$ for all $x \geq x_1$. If for some $\alpha > 0$*

$$\lim_{x \to x_1} \frac{(x_1 - x)F'(x)}{1 - F(x)} = \alpha,$$

then $F \in \mathcal{D}(\Psi_\alpha)$. The normalizing constants can be taken as

$$a_n = x_1 - Q_F\left(1 - \frac{1}{n}\right), \quad b_n = x_1.$$

Theorem 1.10.4 *Let F be a cdf with negative second derivative in some interval (x_0, x_1) and $F'(x) = 0$ for all $x \geq x_1$. If for some $\alpha > 0$*

$$\lim_{x \to x_1} \frac{F''(x)\,(1 - F(x))}{(F'(x))^2} = -1,$$

then $F \in \mathcal{D}(\Lambda)$. The normalizing constants can be taken as

$$a_n = \frac{1 - F(b_n)}{F'(b_n)}, \quad b_n = Q_F\left(1 - \frac{1}{n}\right).$$

1.10.2 Limits for minima

Note that

$$X_{1:n} \stackrel{d}{=} -Y_{n:n},$$

with $Y = -X$. Also note that if $Y \sim F$, then the distribution function of X is $\widetilde{F}(x) = 1 - F(-x^+)$. Then, the domains of attraction for a sequence of minima conveniently normalized are:

1. $\widetilde{\Lambda}(x) = 1 - \Lambda(-x) = 1 - \exp(-\exp(x)),\ x \in \mathbb{R}$
2. $\widetilde{\Phi}_\alpha(x) = 1 - \Phi_\alpha(-x) = 1 - \exp(-(-x)^{-\alpha})$, for $x < 0$ and $\alpha > 0$.
3. $\widetilde{\Psi}_\alpha(x) = 1 - \exp(-x^\alpha)$, for $x > 0$ and $\alpha > 0$.

And the relation between the domains of attraction is:

$$F \in \mathcal{D}(\widetilde{\Lambda}) \Leftrightarrow \widetilde{F} \in \mathcal{D}(\Lambda)$$
$$F \in \mathcal{D}(\widetilde{\Phi}_\alpha) \Leftrightarrow \widetilde{F} \in \mathcal{D}(\Phi_\alpha)$$
$$F \in \mathcal{D}(\widetilde{\Psi}_\alpha) \Leftrightarrow \widetilde{F} \in \mathcal{D}(\Psi_\alpha)$$

1.10.3 Limits for kth largest

In this case, we study the limit distribution (conveniently normalized) of the kth largest observation $X_{n-k+1:n}$ when n goes to infinity and k is fixed.

Note that

$$F_{n-k+1:n}(x) = \sum_{m=0}^{k-1} \binom{n}{m} F(x)^{n-m} (1 - F(x))^m .$$

So, the following result holds:

Theorem 1.10.5 *For fixed k and $n \to \infty$,*

$$\frac{X_{n:n} - a_n}{b_n} \xrightarrow{D} W$$

with W a rv with cdf $H(x)$ if and only if

$$\frac{X_{n-k+1:n} - a_n}{b_n} \xrightarrow{D} W_k$$

with W_k a rv with cdf

$$H_k(x) = \sum_{m=0}^{k-1} \frac{H(x)\left(-\log H(x)\right)^m}{m!}.$$

A similar result can be obtained considering the sequence of kth smallest observation, $X_{k:n}$ when n goes to infinity and k is fixed.

For the limit behavior of the joint distribution function of order statistics, see Finner and Roters (1994).

Chapter 2

Records. Basic distribution theory

2.1 Ordinary records

Let $\{X_n\}_{n\geq 1}$ be a sequence of iid rv's with common distribution F. Let us define the **record time** sequence as:

$$\begin{aligned} T_0 &= 1, \\ T_n &= \min\{j > T_{n-1} : X_j > X_{T_{n-1}}\},\ n \geq 1, \end{aligned}$$

then the *nth* **record** of the sequence is the random variable

$$R_n = X_{T_n}.$$

Some remarks should be taken into account:

- The sequence of record times $\{T_n\}_{n\geq 0}$ is not well defined if the distribution function F has an atom at its right end, that is to say, if there exists $x_0 < \infty$ such that $F(x_0) - F(x_0^-) > 0$ and $F(x_0) = 1$. Consequently, in this case the sequence of records $\{R_n\}_{n\geq 0}$ is not well defined. Appropriate definitions of record times and records will be given later to handle such cases.

- In some applications it is more convenient to look at the sequence of minima rather than the sequence of maxima. So, we can consider lower records: Let $\{X_n\}_{n\geq 1}$ be a sequence of iid rv's with common distribution F. Let us define the **lower record time** sequence as:

$$\begin{aligned} L_0 &= 1, \\ L_n &= \min\{j > T_{n-1} : X_j < X_{L_{n-1}}\},\ n \geq 1, \end{aligned}$$

then the *nth* **lower record** of the sequence is the random variable

$$V_n = X_{L_n}.$$

- We will develop the distribution theory for upper records.

2.2 Distribution theory in the absolutely continuous case

2.2.1 Density of a single record

- Let $\{X_n\}_{n\geq 1}$ be a sequence of iid rv's with common distribution F. We will assume that F is absolutely continuous with density f. By definition, it is clear that

$$R_0 \stackrel{d}{=} X_1,$$

then, the density of the $0th$ record is

$$f_{R_0}(x) = f(x).$$

- Note that for any $n \geq 0$ and x such that $0 < F(x) < 1$:

$$P(R_{n+1} > y \mid R_n = x) = \sum_{j=0}^{\infty} F^j(x)(1 - F(y)) = \frac{1 - F(y)}{1 - F(x)},\ y > x. \tag{2.1}$$

Note also that (2.1) is independent of n. Then, the conditional density of the $(n+1)th$ record given that $R_n = x$ is

$$f_{R_{n+1}|R_n=x}(y) = \begin{cases} \dfrac{f(y)}{1 - F(x)}, & y > x. \\ 0, & y \leq x \end{cases} \tag{2.2}$$

- From (2.2), and standard properties of conditional densities, we obtain the recurrent formula:

$$f_{R_{n+1}}(y) = \int_{-\infty}^{y} \frac{f(y)}{1 - F(x)} f_{R_n}(x)dx,\ n \geq 0,$$

and by induction, it can be checked that

$$f_{R_n}(x) = \frac{1}{n!} \{-\log(1 - F(x))\}^n f(x). \tag{2.3}$$

2.2.2 Joint density

- Note that the occurrence of the nth record depends only on the value of R_{n-1}. So, we can say that the sequence $\{R_n\}_{n\geq 0}$ is a Markov chain with transition densities as given in (2.2). This Markovian property allows us to obtain the joint density of $\mathbf{R} = (R_0,\, R_1, \ldots, R_m)$:

$$f_{\mathbf{R}}(x_0,\, x_1, \ldots, x_m) = f_{R_0}(x_0) f_{R_1|R_0=x_0}(x_1) \cdots f_{R_m|R_{m-1}=x_{m-1}}(x_m)$$
$$= f(x_m) \prod_{i=0}^{m-1} \frac{f(x_i)}{1 - F(x_i)}, \tag{2.4}$$

  for $x_0 < x_1 < \ldots < x_m$.

- The joint density of two records, say $(R_n,\, R_m)$ with $n < m$, can be derived from the joint density of $(R_0,\, R_1, \ldots, R_m)$ and is, for $x_n < x_m$ with $0 < F(x_n), F(x_m) < 1$:

$$f_{(R_n,\, R_m)}(x_n, x_m) = \frac{\left(-\log \overline{F}(x_n)\right)^n}{n!} \frac{\left(\log \overline{F}(x_n) - \log \overline{F}(x_m)\right)^{m-n-1}}{(m-n-1)!} \frac{f(x_n) f(x_m)}{\overline{F}(x_n)},$$

  where $\overline{F}(x) = 1 - F(x)$.

- In particular, for consecutive records, i.e. $m = n + 1$:

$$f_{(R_n,\, R_{n+1})}(x_n, x_{n+1}) = \frac{\left(-\log \overline{F}(x_n)\right)^n}{n!} \frac{f(x_n) f(x_{n+1})}{\overline{F}(x_n)},$$

  for $0 < F(x_n) < F(x_{n+1}) < 1$.

2.3 Conditional densities

- If F is an absolutely continuous distribution with density f:

$$f_{R_{n+1},\ldots,R_{n+j}|R_n=r_n}(x_{n+1},\ldots,x_{n+j}) = \prod_{i=0}^{j-1} \frac{f(x_{n+i+1})}{\overline{F}(x_{n+i})}, \tag{2.5}$$

  for $x_n < x_{n+1} < \ldots < x_{n+j}$.

- From (2.5), for $m > n$, and $x_m > x_n$,

$$f_{R_m|R_n=x_n}(x_m) = \frac{\left(\log \overline{F}(x_n) - \log \overline{F}(x_m)\right)^{m-n+1}}{(m-n+1)!} \frac{f(x_m)}{\overline{F}(x_m)}.$$

- On the other hand,

$$f_{R_{0,\ldots,}R_{n-1}|R_n=x_n}(x_0,\ldots,x_{n-1}) = n!\left(-\log\overline{F}(x_n)\right)^{-n}\prod_{i=0}^{n-1}\frac{f(x_i)}{\overline{F}(x_i)}, \tag{2.6}$$

  for $x_0 < x_1 < \ldots < x_n$.

- Note that if F is a continuous cdf, then $G = -\log\overline{F}$ is a non-decreasing function. The truncated distribution on the right:

$$H(x) = \begin{cases} \frac{G(x)}{G(r_n)}, & x < r_n \\ 1, & x \geq r_n, \end{cases} \tag{2.7}$$

  has density $h(x) = \left(-\log\overline{F}(r_n)\right)^{-1}\dfrac{f(x)}{\overline{F}(x)}$, for $x < r_n$. Then, comparing (2.6) with the joint density of order statistics, we obtain:

Theorem 2.3.1 *For a continuous distribution F, the conditional distribution of records $(R_{0,} \ldots, R_{n-1})$ given $R_n = r_n$ is the same as the n order statistics of a sample of size n of iid random variables having the truncated distribution given in (2.7).*

- As a consequence of the previous theorem, the distribution of R_j given $R_n = x$, for $0 \leq j < n$ is equal to the distribution of the $(j+1)$th order statistic of a sample of size n from the truncated (2.7), that is:

$$f_{R_{j|}R_n=y}(x) =$$
$$C_{j+1:n}\left(-\log \overline{F}(y)\right)^{-n}\left(-\log \overline{F}(x)\right)^{j}\left(\log \overline{F}(x) - \log \overline{F}(y)\right)^{n-j-1}\frac{f(x)}{\overline{F}(x)},$$

  for $x < y$. In particular, for consecutive records:

$$f_{R_{n-1|}R_n=y}(x) = n\left(-\log \overline{F}(y)\right)^{-n}\left(-\log \overline{F}(x)\right)^{n-1}\frac{f(x)}{\overline{F}(x)},$$

  for $x < y$.

2.4 Some properties of records

- Let $\{R_n\}_{n\geq 0}$ be the sequence of records of an iid sequence $\{X_n\}_{n\geq 1}$ from an absolutely continuous distribution F. If the original sequence $\{X_n\}_{n\geq 1}$ is transformed by a strictly increasing function g, as the strict order is preserved, we obtain that the records of the transformed sequence $\{g(X_n)\}_{n\geq 1}$ are distributed as $\{g(R_n)\}_{n\geq 0}$.

- Observe that if F is strictly increasing in the set $S = \{x \in \mathbb{R} : 0 < F(x) < 1\}$ then the function $G(x) = -\log(1 - F(x))$ is strictly increasing in S. It is easy to check that if $X \sim F$ then, the transformed rv $G(X) \sim Exp(1)$.

- The conclusion of the above arguments is that the records $\{R_n\}_{n\geq 0}$ from a strictly increasing distribution F can be related to the records of a sequence of iid standard exponential variables, denoted hereafter $\{R_n^*\}_{n\geq 0}$, by means of the relationship
$$\{G(R_n)\}_{n\geq 0} \stackrel{d}{=} \{R_n^*\}_{n\geq 0}.$$

2.5 Record indicators

The sequence of **record indicators** is defined as follows:

$$I_1 = 1,$$
$$I_n = I\left\{X_n > X_{(n-1:n-1)}\right\}, \; n \geq 2.$$

The meaning of this sequence is that $I_n = 1$ if and only if the observation X_n is a record.

Theorem 2.5.1 *Let $\{X_j\}_{j\geq 1}$ be a sequence of iid rv's with continuous distribution F. Then, $\{I_n\}_{n\geq 1}$ is a sequence of independent random variables with $I_n \sim Be(\frac{1}{n})$, $n \geq 1$.*

2.6 The number of records

The **record counting process** is $\{N_n\}_{n\geq 1}$ with

$$N_n\text{=number of records among } X_1, \ldots, X_n.$$

It is obvious that:

$$N_n = \sum_{j=1}^{n} I_j,$$

where $\{I_j\}_{j\geq 1}$ is the sequence of record indicators.

For $n \geq 1$, consider the polynomial $Q_n(x) = x(x+1)\ldots(x+n-1)$. If this polynomial is expanded in powers of x, we obtain

$$Q_n(x) = \sum_{k=1}^{n} S_{n,k} x^k.$$

The coefficients $S_{n,k}$, with $1 \leq k \leq n$ are called **Stirling numbers of the first kind.** These numbers can be obtained by the recurrence formulas:

$$\begin{aligned} S_{n+1,k} &= nS_{n,k} + S_{n,k-1} \\ S_{1,1} &= 1, \end{aligned}$$

with the agreement that $S_{n,k} = 0$ for $k < 1$ or $k > n$.

Theorem 2.6.1 *The probability function of N_n is*

$$P(N_n = k) = \frac{S_{n,k}}{n!},\ k = 1, \ldots, n.$$

Some properties of the record counting process are:

- $EN_n = \sum_{j=1}^{n} \frac{1}{j}$.

- $Var(N_n) = \sum_{j=1}^{n} \frac{1}{j}\left(1 - \frac{1}{j}\right)$.

- For large n,

$$EN_n \approx \log n + \gamma; \quad Var(N_n) \approx \log n + \gamma - \frac{\pi^2}{6},$$

  where $\gamma = 0.5772...$ is the **Euler's constant**. So, we can say that records are rather uncommon. For instance in a sequence of 1000 observations we expect to observe approximately 7 records.

- From the Strong Law of Large Numbers for sequences with uniformly bounded variance

$$\frac{N_n}{\log n} \xrightarrow{a.s.} 1.$$

- From the Liapounov condition for the Central Limit Theorem

$$\frac{N_n - \log n}{\sqrt{\log n}} \xrightarrow{d.} N(0,1).$$

2.7 The distribution of record times

Let $\{X_j\}_{j\geq 1}$ be a sequence of iid rv's with continuous distribution F.

Once we have obtained the distribution of the record indicators, one can obtain the joint distribution of the record times. For $1 < n_1 < n_2 < \cdots < n_m$

$$P(T_1 = n_1, T_2 = n_2, \ldots, T_m = n_m) = P(I_2 = 0, \ldots, I_{n_1-1} = 0,\ I_{n_1} = 1, \ldots, I_{n_m} = 1) = \\ = \{(n_1 - 1)(n_2 - 1) \cdots (n_m - 1)n_m\}^{-1}. \quad (2.8)$$

The marginal distribution of T_1 is:

$$P(T_1 = n_1) = \frac{1}{n_1(n_1 - 1)},\ n_1 \geq 2,$$

and it can be easily checked that

$$ET_1 = \infty.$$

The distribution of the kth record time can be obtained from the joint given in (2.8) or alternatively using the record counting process:

$$P(T_k = n) = P(N_n = k + 1,\ N_{n-1} = k) = P(I_n = 1,\ N_{n-1} = k) = \\ = \frac{S_{n-1,k}}{n!}$$

Some properties of the sequence of record times are:

- $ET_k = \infty$; $E\left(T_k - T_{k-1}\right) = \infty$
- $\dfrac{\log T_k}{k} \overset{a.s.}{\longrightarrow} 1$, $k \longrightarrow \infty$.
- $\dfrac{\log T_k - k}{\sqrt{k}} \overset{d}{\longrightarrow} N(0,1)$, $k \longrightarrow \infty$.

2.8 Markov properties

There are several Markov chains related to a sequence of records from a continuous distribution:

- The counting process $\{N_n\}_{n\geq 1}$ is a non-homogeneous Markov process with transition probabilities:

$$P\left(N_n = j \mid N_{n-1} = i\right) = \begin{cases} \dfrac{n-1}{n}, & j = i \\ \dfrac{1}{n}, & j = i+1. \end{cases}$$

- The sequence of record times $\{T_n\}_{n\geq 1}$ is a homogeneous Markov process with transition probabilities:

$$P\left(T_n = j \mid T_{n-1} = i\right) = \frac{i}{j(j-1)}, \; j > i.$$

- The sequence of records $\{R_n\}_{n\geq 1}$ is a Markov process with transitions given by the conditional densities:

$$f_{R_n \mid R_{n-1}=x}(y) = \frac{f(y)}{1-F(x)}, \;\; y > x$$

2.9 Records from discrete distributions

- In section 2.1 we pointed out that records are not well defined for distributions with an atom at is upper end. For that reason, Vervaat (1973) introduced the concept of weak records:

- Let $\{X_n\}_{n\geq 1}$ be a sequence of iid rv's with common distribution F. Let as define the **weak record time** sequence as:

$$\widetilde{T}_0 = 1,$$
$$\widetilde{T}_n = \min\{j > \widetilde{T}_{n-1} : X_j \geq X_{\widetilde{T}_{n-1}}\},\ n \geq 1,$$

  then the *nth* **weak record** of the sequence is the random variable

$$W_n = X_{\widetilde{T}_n}.$$

- Note that the difference with ordinary records is that ties with previous records are also considered as records. Obviously, for continuous distributions weak records coincide (almost surely) with ordinary records. In the case of discrete distributions ordinary and weak records lead to different definitions.

For distributions with an atom at its upper end the sequence of weak records is well defined while ordinary records are not.

Suppose that X is a discrete rv with pmf $P(X = x_k) = p_k > 0$, for $k = 1, 2, \ldots, N$, where N is a natural number and possibly $N = \infty$. Let $q_k = P(X \geq x_k)$. For simplicity, we will assume that $x_1 < x_2 < \cdots < x_N$.

- For an iid sequence $\{X_n\}_{n\geq 1}$, the joint distribution of weak records is given by:

$$P\left(W_0 = x_{k_0}, W_1 = x_{k_1}, \ldots, W_n = x_{k_N}\right) = p_{k_n} \prod_{j=0}^{n-1} \frac{p_{k_j}}{q_{k_j}}, \tag{2.9}$$

  for $x_1 \leq x_{k_0} \leq x_{k_1} \ldots \leq x_{k_N} \leq x_N$.

- For ordinary records a bounded support is not permitted, then necessarily $N = \infty$ and the joint distribution is given by

$$P\left(R_0 = x_{k_0}, R_1 = x_{k_1}, \ldots, R_n = x_{k_N}\right) = p_{k_n} \prod_{j=0}^{n-1} \frac{p_{k_j}}{q_{k_j+1}}, \tag{2.10}$$

  for $x_1 \leq x_{k_0} < x_{k_1} \ldots < x_{k_N}$.

2.10 Records from some specific distributions

2.10.1 Exponential distribution

- The density of R_n^* is

$$f_{R_n^*}(x) = \frac{1}{n!} x^n \exp(-x), x > 0,$$

  then $R_n^* \sim Ga(n+1, 1)$.

- The joint density of two records $(R_n^*,\, R_m^*)$ with $n < m$ is

$$f_{(R_n^*, R_m^*)}(x, y) = \frac{1}{n!(m-n-1)!} x^n (y-x)^{m-n-1} \exp(-y),\ \ 0 < x < y.$$

- In particular, if they are consecutive, i.e. $m = n+1$,

$$f_{(R_n^*, R_{n+1}^*)}(x, y) = \frac{1}{n!} x^n \exp(-y),\ \ 0 < x < y.$$

- The joint density of $\mathbf{R}^* = (R_0^*, R_1^*, \ldots, R_m^*)$ is

$$f_{\mathbf{R}}(x_0, x_1, \ldots, x_m) = \exp(-x_m),\ x_0, < x_1 < \ldots < x_m. \tag{2.11}$$

- Note that the density given in (2.11) coincides with the density function of $(S_1, S_2, \ldots, S_m)$, where $S_k = \sum_{j=1}^k \xi_j$, with $\{\xi_j\}_{j \geq 1}$ a sequence of iid standard exponential random variables, that is to say

$$(R_0^*, R_1^*, \ldots, R_m^*) \stackrel{d}{=} (S_1, S_2, \ldots, S_{m+1}). \tag{2.12}$$

- Then, from (2.12),

$$(R_0^*, R_1^* - R_0^*, \ldots, R_m^* - R_{m-1}^*) \stackrel{d}{=} (\xi_1, \xi_2, \ldots, \xi_m),$$

  which means that spacings between consecutive records are iid and distributed as standard exponentials.

- Another consequence of (2.12) is that $\{R_m^* - (m+1)\}_{m \geq 0}$ is a martingale, i.e.,

$$E\left(R_m^* - (m+1) \mid R_{m-1}^*\right) = R_{m-1}^* - m,\ m \geq 0.$$

2.10.2 Geometric distributions

Weak records

A discrete random variable X follows a geometric distribution with parameter $p \in (0,1)$, if

$$P(X = k) = q^k p, \;\; k = 0, 1, 2, \ldots \text{with } q = 1 - p.$$

We denote $X \sim Ge(p)$. For this distribution $q_k = P[X \geq k] = q^k$, $k \geq 0$.

- The joint distribution of weak records is

$$P(W_0 = k_0, W_1 = k_1, \ldots, W_n = k_n) = q^{k_n} p^{n+1}, \tag{2.13}$$

  for $0 \leq k_0 \leq k_1 \leq \cdots \leq k_n$.

- It can be checked that (2.13) is also the probability function of $(S_1, S_2, \ldots, S_n)$, where $S_k = \sum_{j=1}^{k} \xi_j$, with $\{\xi_j\}_{j\geq 1}$ a sequence of iid $Ge(p)$ random variables, that is to say,

$$(W_0, W_1, \ldots, W_n) \stackrel{d}{=} (S_1, S_2, \ldots, S_{m+1}). \tag{2.14}$$

Some consequences of the representation given in (2.14) are:

- $W_k \sim NB(k+1, p)$, $k \geq 0$, i.e.,

$$P(W_k = j) = \binom{j+k}{j} q^j p^{k+1},\ j = 0, 1, \ldots$$

- Spacings between consecutive records are iid $Ge(p)$ random variables:

$$(W_0, W_1 - W_0, \ldots, W_m - W_{m-1}) \stackrel{d}{=} (\xi_1, \xi_2, \ldots, \xi_m),$$

- The sequence $\{W_n - (n+1)q/p\}_{n\geq 0}$ is a martingale:

$$E(W_n - (n+1)q/p \mid W_{n-1}) = W_{n-1} - nq/p,\ n \geq 1.$$

Geometric distribution. Ordinary records

- The joint distribution of ordinary records is

$$P(R_0 = k_0,\, R_1 = k_1, \ldots, R_n = k_n) = q^{k_n - n} p^{n+1}, \tag{2.15}$$

for $0 \leq k_0 \leq k_1 \leq \cdots \leq k_n$.

- It can be checked that (2.15) is also the probability function of $(S_1,\, S_2 + 1,\, S_3 + 2,\, \ldots, S_n + (n-1))$, where $S_k = \sum_{j=1}^{k} \xi_j$, with $\{\xi_j\}_{j\geq 1}$ a sequence of iid $Ge(p)$ random variables, that is to say,

$$(R_0,\, R_1, \ldots, R_n) \stackrel{d}{=} (S_1,\, S_2 + 1,\, S_3 + 2,\, \ldots, S_n + (n-1)). \tag{2.16}$$

Some consequences of the representation given in (2.16) are:

- $R_k - k \sim NB(k+1, p)$, $k \geq 0$, i.e.,

$$P(R_k = j) = \binom{j}{j-k} q^{j-k} p^{k+1},\ j = k,\, k+1, \ldots$$

- For spacings between consecutive ordinary records:

$$(R_0,\, R_1 - R_0 - 1, \ldots, R_n - R_{n-1} - 1) \stackrel{d}{=} (\xi_1,\, \xi_2,\, \ldots, \xi_m),$$

- The sequence $\{R_n - n/p\}_{n\geq 0}$ is a martingale:

$$E(R_n - n/p \mid R_{n-1}) = R_{n-1} - (n-1)/p,\ n \geq 1.$$

2.11 k-th records

- In the definition of ordinary (upper) records, we look at the sequence of largest observations, but it can be also interesting to keep track of the second largest, or more generally the kth largest observation. Thus, the concept of kth record appear:

- Let $\{X_n\}_{n\geq 1}$ be a sequence of independent and identically distributed (iid) random variables (rv's) from a distribution F. Let us consider the kth-*record times* defined recurrently as:

$$T_o^{(k)} = k,$$
$$T_{n+1}^{(k)} = \min\left\{j : j > T_n^{(k)},\ X_j > X_{T_n^{(k)}-k+1\,:\,T_n^{(k)}}\right\},\ n \geq 0,$$

and the nth *ordinary kth-record* from F as

$$R_n^{(k)} = X_{T_n^{(k)}-k+1\,:\,T_n^{(k)}},\ n \geq 0, \tag{2.17}$$

where $X_{i:m}$ denotes the ith order statistics of a sample of size m.

- Note that definition (2.17) does not make sense for a distribution with an atom at the right endpoint of its support. Following Vervaat (1973), this problem can be avoided with a slight change. So, we introduce firstly the *weak* kth-*record times* recurrently as:

$$U_o^{(k)} = k;$$
$$U_{n+1}^{(k)} = \min\left\{j : j > U_n^{(k)},\ X_j \geq X_{U_n^{(k)}-k+1\,:\,U_n^{(k)}}\right\},\ n \geq 0,$$

  and the n*th weak* kth-*record* from F as

$$W_n^{(k)} = X_{U_n^{(k)}-k+1\,:\,U_n^{(k)}},\ n \geq 0.$$

- Note that the difference between both definitions is that in the later a new observation from the sequence is labeled as a kth-record even in the case in which it takes the same value of the actual kth-record. Of course, if the underlying distribution F is continuous there is no difference between ordinary and weak kth-records. According to Wesolowski and Ahsanullah (2001) the use of weak records rather than ordinary records is more natural in the iid setting due to the fact that there is no preference between tied observations, moreover the mathematical theory for weak records seems to be richer than the one for ordinary records.

- In pag. 43 of the book *Records* by Arnold, Balakrishnan and Nagaraja (1998), one can read: *The sequence* $\{R_n^{(k)}, n \geq 0\}$ *from a (discrete or continuous) cdf. F is identical in distribution to a first record sequence* $\{R_n^{(1)}, n \geq 0\}$ *from the cdf.* $F_{1:k} = 1 - (1 - F)^k$.

- We denote this fact by

$$\left\{R_n^{(k)}(F), n \geq 0\right\} \stackrel{d}{\equiv} \left\{R_n^{(1)}\left(F_{1:k}\right), n \geq 0\right\}. \tag{2.18}$$

- We will show later that (2.18) is not true for discrete distributions.

- In the absolutely continuous case (2.18) holds, and consequently the theory of kth records can be reduced to the theory of (first) record. So, we have the following results in the absolutely continuous case:

- The joint density of $\mathbf{R}^{(k)} = (R_0^{(k)}, R_1^{(k)}, \ldots, R_m^{(k)})$:

$$f_{\mathbf{R}^{(k)}}(x_0, x_1, \ldots, x_m) =$$
$$= k^{m+1}\overline{F}^{k-1}(x_m)f(x_m)\prod_{i=0}^{m-1}\frac{\overline{F}^{k-1}(x_i)f(x_i)}{\overline{F}^{k}(x_i)},$$

for $x_0 < x_1 < \ldots < x_m$, with $\overline{F}(x) = 1 - F(x)$.

- The density of $R_n^{(k)}$ is:

$$f_{R_n^{(k)}}(x) = \frac{k^{n+1}}{n!}\left\{-\log\overline{F}(x)\right\}^n\overline{F}^{k-1}(x)f(x).$$

- The sequence of kth-records $\left\{R_n^{(k)}\right\}_{n\geq 1}$ is a Markov process with transitions given by the conditional densities:

$$f_{R_n|R_{n-1}=x}(y) = k\frac{(1-F(y))^{k-1}f(y)}{(1-F(x))^k},$$

for $y > x$.

For the standard exponential distribution, we have the following results:

- The density of $R_n^{(k)}$ is

$$f_{R_n^{(k)}}(x) = \frac{k^{n+1}}{n!} x^n \exp(-kx), x > 0,$$

  then $R_n^{(k)} \sim Ga(n+1, k)$.

- The joint density of $\mathbf{R}^* = (R_0^*, R_1^*, \ldots, R_m^*)$ is

$$f_{\mathbf{R}}(x_0, x_1, \ldots, x_m) = k^{n+1} \exp(-kx_m), \; x_0 < x_1 < \ldots < x_m. \tag{2.19}$$

- Note that the density given in (2.19) coincides with the density function of $(S_1/k, S_2/k, \ldots, S_m/k)$, where $S_k = \sum_{j=1}^{k} \xi_j$, with $\{\xi_j\}_{j\geq 1}$ a sequence of iid standard exponential random variables, that is to say

$$k(R_0^{(k)}, R_1^{(k)}, \ldots, R_m^{(k)}) \stackrel{d}{=} (S_1, S_2, \ldots, S_{m+1}). \tag{2.20}$$

- Then, from (2.20),

$$k(R_0^{(k)}, R_1^{(k)} - R_0^{(k)}, \ldots, R_m^{(k)} - R_{m-1}^{(k)}) \stackrel{d}{=} (\xi_1, \xi_2, \ldots, \xi_m),$$

  which means that spacings between consecutive kth records from standard exponential distributions are iid and distributed as exponential distributions with mean $1/k$.

- Another consequence of (2.20) is that $\left\{R_m^{(k)} - (m+1)/k\right\}_{m\geq 0}$ is a martingale, i.e.,

$$E[R_m^{(k)} - (m+1)/k \mid R_{m-1}^{(k)}] = R_{m-1}^{(k)} - m/k, \; m \geq 0.$$

2.12 kth records from discrete distributions

Property

$$\left\{R_n^{(k)}(F),\, n \geq 0\right\} \stackrel{d}{\equiv} \left\{R_n^{(1)}\left(F_{1:k}\right),\, n \geq 0\right\}.$$

does not hold for discrete distributions

For simplicity, consider discrete random variables with support on a set of integers of the form $S = \{0, 1, 2, ..., N\}$ with N possibly infinite. Let us suppose that the probability mass function (pmf) of X is $p_j = P(X = j) > 0$ and $q_j = P(X \geq j),\ j \in S$.

- We consider $k = 2$ and we will show that the sequence $\{W_n^{(2)}(F),\, n \geq 0\}$ does not have the same distribution as the sequence $\{W_n^{(1)}(F_{1:2}),\, n \geq 0\}$.

- Note that for $n = 0$, it is immediate that

$$W_0^{(2)}(F) \stackrel{d}{\equiv} W_0^{(1)}(F_{1:2}) \stackrel{d}{\equiv} X_{1:2}. \tag{2.21}$$

- We will show that for $n = 1$,

$$W_1^{(2)}(F) \stackrel{d}{\not\equiv} W_1^{(1)}(F_{1:2}), \tag{2.22}$$

  and so, we will prove that property (2.18) does not hold for weak records. The case of ordinary records will be discussed later.

- In order to prove (2.22), taking into account (2.21), it is enough to show that the conditional distribution

$$W_1^{(2)}(F) \mid W_0^{(2)}(F) = j, \;\; j \in S \tag{2.23}$$

  and

$$W_1^{(1)}(F_{1:2}) \mid W_0^{(1)}(F_{1:2}) = j, \;\; j \in S \tag{2.24}$$

  are different.

- Firstly, using formulas for the conditional pmf of ordinary weak records:

$$P\left(W_1^{(1)}\left(F_{1:2}\right) = m \mid W_0^{(1)}\left(F_{1:2}\right) = j\right) = \frac{q_m^2 - q_{m+1}^2}{q_j^2},\ m \geq j. \tag{2.25}$$

- On the other hand:

$$P\left(W_1^{(2)}\left(F\right) = m \mid W_0^{(2)}\left(F\right) = j\right) = \begin{cases} 2\dfrac{p_j}{q_j}\dfrac{q_m^2 - q_{m+1}^2}{q_j^2 - q_{j+1}^2}, & m > j \\ \\ \dfrac{p_j^2}{q_j^2 - q_{j+1}^2}, & m = j \end{cases} \tag{2.26}$$

see López Blazquez, Salamanca Miño and Dembinska (2005).

- Following similar arguments it can be shown that property (2.18) does not hold even if ordinary records are used. In fact, for $m, j \in S$:

$$P\left(R_1^{(2)}(F) = m \,|\, R_0^{(2)}(F) = j\right) = \begin{cases} 2\dfrac{p_j}{q_{j+1}}\dfrac{q_m^2 - q_{m+1}^2}{q_j^2 - q_{j+1}^2}, & m > j \\ \\ \dfrac{p_j^2}{q_j^2 - q_{j+1}^2}, & m = j, \end{cases}$$

while

$$P\left(R_1^{(1)}\left(F_{1:2}\right) = m \mid R_0^{(1)}\left(F_{1:2}\right) = j\right) = \frac{q_m^2 - q_{m+1}^2}{q_{j+1}^2},\ m > j.$$

- For more information about the distribution theory of k-th records in the discrete case, consult Dembinska, López-Blazquez (2005a).

2.13 Moments of records

- Let F be an absolutely continuous distribution function with density f. The moments of records, provided that they exists, can be obtained as:

$$ER_n^m = \frac{1}{n!}\int_{-\infty}^{\infty} x^m \{-\log(1-F(x))\}^n f(x)dx.$$

- If F is strictly increasing in the set $\{x : 0 < F(x) < 1\}$, and Q_F is its quantile function, the following alternative formulas can be used:

$$ER_n^m = \frac{1}{n!}\int_0^1 Q_F^m(u)\left\{-\log\left(\frac{1}{1-u}\right)\right\}^n du,$$
$$ER_n^m = \frac{1}{n!}\int_0^{\infty} Q_F^m(1-\exp(-t))t^n \exp(-t)dt.$$

- Conditions for the existence of moments of records are given in Nagaraja (1978):

Theorem 2.13.1 *If $X \sim F$ strictly increasing and $E|X|^p < \infty$ for some $p > 1$, then ER_n exists for all n.*

In the case of non-negative random variables, the following necessary and sufficient condition can be given:

Theorem 2.13.2 *If X is an unbounded positive and absolutely continuous random variable, a necessary and sufficient condition for the existence of ER_n is that $E(X\log(X))^n$ exists.*

Note that if X is bounded, R_n is also bounded and then, all its moments exists.

2.14 Limit distributions for records

The basic result in this area is due to Resnick (1973). In short, there exist three possible domains of attractions for $(R_n - A(n))/B(n)$, as n goes to infinite for suitable chosen sequences $A(n)$ and $B(n) > 0$ of real numbers.

Theorem 2.14.1 *To each cdf F associate the cdf*

$$\widetilde{F}(x) = 1 - \exp\left\{-\sqrt{-\log(1 - F(x))}\right\},$$

and let $N(n) = [exp(n)^{1/2}]$.
Then, $\widetilde{F}$ belongs to the domain of atraction of the limiting distribution H for maxima if and only if F belongs to the domain of attraction of the limiting distribution

$$G(x) = \Phi\left(-\log\left(-\log H(x)\right)\right)$$

for record values, where Φ denotes the cdf of a $\mathcal{N}(0,1)$ rv. The centering and normalizing constants for maxima ($a(n)$ and $b(n)$) and for records ($A(n)$ and $B(n)$) are related by

$$A(n) = a\left(N(n)\right), B(n)) = b\left(N(n)\right).$$