Avd. Matematisk statistik

KTH Teknikvetenskap

Sf 2955: Computer intensive methods :
ORDER STATISTICS FOR INDEPENDENT EXPONENTIAL VARIABLES / Timo Koski

1 Order Statistics

Let $X_1, \ldots, X_n$ be I.I.D. random variables with a continuous distribution. The order statistic of $X_1, \ldots, X_n$ is the ordered sample:

$$X_{(1)} < X_{(2)} < \ldots < X_{(n)},$$

Here

$$X_{(1)} = \min\left(X_1, \ldots, X_n\right)$$
$$X_{(n)} = \max\left(X_1, \ldots, X_n\right)$$

and

$$X_{(k)} = \text{ } k\text{th smallest of } X_1, \ldots, X_n \ .$$

The variable $X_{(k)}$ is called the kth order variable. The following theorem has been proved in, e.g., Allan Gut: *An Intermediate Course in Probability. 2nd Ed.*, Springer Verlag, Dordrecht e.t.c., 2009, ch. 4.3., theorem 3.1..

Theorem 1.1 Assume that $X_1, \ldots, X_n$ are I.I.D. random variables with the density f. The joint density of the order statistic is

$$f_{X_{(1)},X_{(2)},\ldots,X_{(n)}}\left(y_1, \ldots, y_n\right) = \begin{cases} n! \prod_{k=1}^{n} f\left(y_k\right) & \text{if } y_1 < y_2 < \ldots < y_n, \\ 0 & \text{elsewhere.} \end{cases} \tag{1.1}$$

■

2 Exponential Order Variables

Let $X_1, \ldots, X_n$ be I.I.D. random variables with distribution $\mathrm{Exp}(\theta)$. Thus the density function of each X_i is

$$f(x;\theta) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0. \end{cases} \tag{2.2}$$

We are interested in the differences of the order variables

$$X_{(1)}, X_{(i)} - X_{(i-1)}, \quad i = 2, \ldots, n.$$

Note that we may consider $X_{(1)} = X_{(1)} - X_{(0)}$, if $X_{(0)} = 0$. We shall next show the following theorem.

Theorem 2.1 Assume that $X_1, \ldots, X_n$ are I.I.D. random variables under $\mathrm{Exp}(1)$. Then

(a)

$$X_{(1)} \in \mathrm{Exp}\left(\frac{1}{n}\right), X_{(i)} - X_{(i-1)} \in \mathrm{Exp}\left(\frac{1}{n+1-i}\right),$$

(b) $X_{(1)}$, $X_{(i)} - X_{(i-1)}$ for $i = 2, \ldots, n$, are n independent random variables.

Proof: We define Y_i for $i = 1, \ldots, n$ by

$$Y_1 = X_{(1)}, \quad Y_i = X_{(i)} - X_{(i-1)}.$$

Then we introduce

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ -1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -1 & 1 \end{pmatrix}. \tag{2.3}$$

so that if

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} X_{(1)} \\ X_{(2)} \\ X_{(3)} \\ \vdots \\ X_{(n)} \end{pmatrix},$$

we have

$$\mathbf{Y} = \mathbf{A}\mathbf{X}.$$

It is clear that the inverse matrix $\mathbf{A}^{-1}$ exists, because we can uniquely find $\mathbf{X}$ from $\mathbf{Y}$ by

$$X_{(1)} = Y_1, \quad X_{(i)} = Y_i + Y_{i-1} + \ldots + Y_1.$$

We write these lastmentioned equalities in matrix form by

$$\mathbf{X} = \mathbf{A}^{-1}\mathbf{Y}.$$

Then we have the well known change of variable formula (se Gut work cited chap I)

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}\left(\mathbf{A}^{-1}\mathbf{y}\right)\frac{1}{|\det \mathbf{A}|}. \tag{2.4}$$

But now we evoke (1.1) to get

$$f_{\mathbf{X}}\left(\mathbf{A}^{-1}\mathbf{y}\right) = n! f(y_1) f(y_1 + y_2) \cdots f(y_1 + y_2 + \ldots + y_n), \tag{2.5}$$

since $y_1 < y_1 + y_2 < \ldots < y_1 + y_2 + \ldots + y_n$. As $f(x) = e^{-x}$, we get

$$f(y_1) f(y_1 + y_2) \cdots f(y_1 + y_2 + \ldots + y_n) = e^{-y_1} e^{-(y_1+y_2)} \cdots e^{-(y_1+y_2+\ldots+y_n)}$$

and rearrange and use $y_1 = x_{(1)}$ and $y_i = x_{(i)} - x_{(i-1)}$,

$$= e^{-ny_1} e^{-(n-1)y_2} \cdots e^{-2y_{n-1}} e^{-y_n}$$

$$= e^{-nx_{(1)}} e^{-(n-1)(x_{(2)}-x_{(1)})} \cdots e^{-(x_{(n)}-x_{(n-1)})}.$$

Hence, if we insert the last result in (2.4) and distribute the factors in $n! = n(n-1)\cdots 3 \cdot 2 \cdot 1$ into the product of exponentials we get

$$f_{\mathbf{Y}}(\mathbf{y}) = ne^{-nx_{(1)}}(n-1)e^{-(n-1)(x_{(2)}-x_{(1)})} \cdots e^{-(x_{(n)}-x_{(n-1)})}\frac{1}{|\det \mathbf{A}|} \tag{2.6}$$

Since $\mathbf{A}$ in (2.3) is a triangular matrix, its determinant equals the product of its diagonal terms, c.f. L. Råde and B. Westergren: *Mathematics Handbook for Science and Engineering*, Studneetlitteratur, Lund, 2009, p. 93. Hence from (2.3) we get $\det \mathbf{A} = 1$. In other words, we have obtained

$$f_{X_{(1)},X_{(2)}-X_{(1)},\ldots,X_{(n)}-X_{(n-1)}}\left(x_{(1)}, x_{(2)} - x_{(1)}, \ldots, x_{(n)} - x_{(n)}\right)$$

$$= ne^{-nx_{(1)}}(n-1)e^{-(n-1)(x_{(2)}-x_{(1)})}\cdots 2e^{-2(x_{(n-1)}-x_{(n-2)})}e^{-(x_{(n)}-x_{(n-1)})}. \quad (2.7)$$

But, checking against (2.2), $ne^{-nx_{(1)}}$ is the probabilty density of Exp $\left(\frac{1}{n}\right)$, $(n-1)e^{-(n-1)(x_{(2)}-x_{(1)})}$ is nothing but the probability density of Exp $\left(\frac{1}{n-1}\right)$, and so on, the generic factor in the product in (2.7) being $(n+1-i)e^{-(n+1-i)(x_{(i)}-x_{(i-1)})}$, which is the density of Exp $\left(\frac{1}{n+1-i}\right)$.

Hence we have that the product in (2.7) is a product of the respective probability densities for the variables $X_{(1)} \in$ Exp $\left(\frac{1}{n}\right)$ and for $X_{(i)} - X_{(i-1)} \in$ Exp $\left(\frac{1}{n+1-i}\right)$. Thus we have established the cases (a) and (b) in the theorem as claimed. ■

As is well known, there is also a more intuitively appealing way of seeing this result. First one shows that

$$X_{(1)} = \min(X_1, \ldots, X_n) \in \text{Exp}\left(\frac{1}{n}\right)$$

(which is also seen above), if $X_1, \ldots, X_n$ are I.I.D. random variables under Exp(1). Then one can argue by independence and the memorylessness property of the exponential distribution that $X_{(i)} - X_{(i-1)}$ is the minimum of lifetimes of $n+1-i$ independent Exp(1) -distributed random variables.