Notes 7 (p.1) Math 481/581 Mathematical Statistics I

Order Statistics.

Let $X_1, X_2, \ldots, X_n$ be independent r.v.'s each with the density $f(x)$. We introduce $X_{(1)} = \min(X_1, \ldots, X_n)$ and $X_{(n)} = \max(X_1, \ldots, X_n)$.

and in general we define the order statistics

$$X_{(1)} < X_{(2)} < \cdots < X_{(n-1)} < X_{(n)}.$$

!. Thus to obtain order stats we sort the values $X_1, X_2, \ldots, X_n$ but the distribution of $X_{(1)}$ is not just the distribution of X_1. The sorting is a non-trivial operation and must be done each time we generate the sample of size n; so the distribution of $X_{(1)}$ also depends on the sample size, n.

3.7 Example A Let $T_1, T_2, \ldots, T_n$ be independent exponential r.v.'s each with parameter λ.
Find the density of $T_{(1)} = \min(T_1, \ldots, T_n)$.
The variable $T_{(1)}$ may be realized as the time until failure of a system of n independent components connected in series where each component has an exponential lifetime with parameter λ.

n components. Lifetime of system = time until first failure among the n-components.

The distribution function of $T_{(1)}$ satisfies:

survival prob.

survival occurs $\Leftrightarrow$ each comp survives

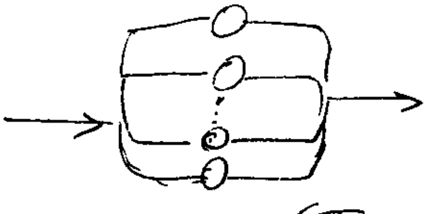
$$F_{T_{(1)}}(t) = P(T_{(1)} \le t) = 1 - \boxed{P(T_{(1)} > t)}$$

$$= 1 - P(\min(T_1, \ldots, T_n) > t) = 1 - P(T_1 > t) \cdots P(T_n > t) = 1 - (e^{-\lambda t})^n$$

$$= 1 - e^{-n\lambda t}. \quad \therefore\ f_{T_{(1)}}(t) = n\lambda e^{-n\lambda t},\ t \ge 0. \ (\text{exponential } (n\lambda))$$

Notes 7

(p. 2) This basic trick can be used to find the distribution of the lifetime $T_{(n)}$ of a system of n independent components (this time) connected in parallel, again with individual component lifetimes distributed as exponential(λ). $T_{(n)} = \max(T_1, \ldots, T_n)$



$$F_{T_{(n)}}(t) = P(T_{(n)} \le t) = P(\max(T_1, \ldots, T_n) \le t)$$

$$= P(T_1 \le t, T_2 \le t, \ldots, T_n \le t) = (1 - e^{-\lambda t})^n.$$

$$\therefore \quad f_{T_{(n)}}(t) = \lambda n (1 - e^{-\lambda t})^{n-1}$$

Note that this example and the last one are special cases of Thm A, p. 105 where the density of $X_{(k)}$ is shown: if $k=1$ we have $f_{(1)}(x) = \frac{n!}{(n-1)!} f(x) F^0(x) (1-F(x))^{n-1}$

(for exponential case) $= n \cdot \lambda e^{-\lambda x} \cdot 1 \cdot (1-(1-e^{-\lambda x}))^{n-1} = n \cdot \lambda e^{-\lambda x} \cdot e^{-(n-1)\lambda x}$

$= n \lambda e^{-n\lambda x}$; $x > 0$.

If $k = n$ we have $f_{(n)}(x) = \frac{n!}{(n-1)!\,0!} f(x) F^{n-1}(x) (1-F(x))^0$

(for exponential) $= n \cdot \lambda e^{-\lambda x} \cdot (1 - e^{-\lambda x})^{n-1} \cdot 1$, $x > 0$.

For the general case of k, Thm A can be seen by a direct calculation: The event $X_{(k)} \in [x, x+dx]$ holds $\iff$ there are $k-1$ indices j with $X_j < x$, 1 index i with $X_i \in [x, x+dx]$ and $n-k$ indices j' with $X_{j'} > x$. The probability of any such combination of these conditions is $F(x)^{k-1} f(x)dx\,(1-F(x))^{n-k}$. Further there are $\binom{n}{k-1,\,1,\,n-k}$ $= \binom{n}{k-1}\binom{n-k+1}{1}\binom{n-k}{n-k} = \frac{n!}{(k-1)!\,(n-k+1)!} \cdot (n-k+1) \cdot 1 = \frac{n!}{(k-1)!\,(n-k)!}$ such combinations of indices. ∎

Notes 7 3.7. Example C.

(p.3) Note that for $f(x) = \begin{cases} 1 & 0<x<1 \\ 0 & \text{else} \end{cases}$, Thm A

gives that the density of the k^{th} order statistic is

$$f_{(k)}(x) = \frac{n!}{(k-1)!(n-k)!} x^{k-1} \cdot 1 \cdot (1-x)^{n-k}, \quad 0<x<1$$

$$= \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)} x^{k-1} (1-x)^{(n-k+1)-1}, \quad 0<x<1.$$

[Beta(k, n-k+1)].

Problem 3.70. Let U_1, U_2, U_3, U_4, U_5 be independent r.v.'s each distributed as Uniform(0,1). The probability that all 5 of those values lie in the interval $(\frac{1}{4}, \frac{3}{4})$ is:

$$P(1/4 < U_{(1)} < U_{(5)} < 3/4)$$

$$= P(\tfrac{1}{4} < U_1 < \tfrac{3}{4})^5 = (\tfrac{1}{2})^5.$$

This can also be obtained by using that the joint density of $V := U_{(1)}$ and $U := U_{(n)}$ is

given by: $f(u,v)\,du\,dv = n(n-1)(u-v)^{n-2}(1du)(1dv)$, $0 \le v \le u \le 1$ (p.106)

(minimum at v, maximum at u, and all other variables between v & u)

So, with $n=5$, $P(\frac{1}{4} < V, U < \frac{3}{4}) = \int_{\frac{1}{4}}^{3/4} \left(\int_{v=1/4}^{u} 20(v-u)^3 dv \right) du$

$$= \int_{\frac{1}{4}}^{3/4} 20\left[-\frac{(u-v)^4}{4}\right]_{\frac{1}{4}}^{u} du = \int_{\frac{1}{4}}^{3/4} 5\,(u-\tfrac{1}{4})^4\, du$$

$$= (u-\tfrac{1}{4})^5 \Big]_{\frac{1}{4}}^{3/4} = (\tfrac{1}{2})^5.$$

Distribution of the Range

Define $R := X_{(n)} - X_{(1)}$. This is the range of the sample $X_1, \ldots, X_n$. Consider the Uniform case (over)