Order Statistics and Applications

Rosemary Smith
Introduction to Order Statistics

- Unordered Statistics and Observations

\[ x_1, x_2, \ldots, x_n \]
\[ X_1, X_2, \ldots, X_n \]

- Ordered Statistics and Observations

\[ x^{(1)} \leq x^{(2)} \leq \ldots \leq x^{(n)} \]
\[ X^{(1)} \leq X^{(2)} \leq \ldots \leq X^{(n)} \]
What kinds of problems are we trying to solve?

- Actuarial Science
  - Joint Life Insurance - A policy for a couple pays out when the first of the spouses dies. You want to know the distribution of $X_{min}$, which is the random variable defined to be the minimum of two lifespans of the couple.
  
  - Insurance Risk - If an insurance company holds 100 policies of which you have cash-at-hand to pay 50. You want to know the distribution of the variable $X_{(50)}$ the 50th occurrence of a pay-out.
What kinds of problems are we trying to solve?

- Industry

  ○ A machine may run on 10 batteries and shuts off when the 5th battery dies. You will want to know the distribution of $X_{(5)}$.

  ○ The same machine becomes less efficient when the third battery dies and costs you money every day it runs that way. You then want to know the distribution of $W_{3,5}$, the range between the third and fifth occurrences.
Further Definitions

- Probability distribution
  \[ p(x) \]

- Cumulative distribution
  \[ P(x) = \Pr\{X \leq x\} \]

- Probability Distribution of \( x^{(r)} \)
  \[ f_r(x) \]

- Cumulative Distribution of \( x^{(r)} \)
  \[ F_r(x) \]
Further Definitions

- Joint distribution of $x(r)$ and $x(s)$
  \[ f_{rs}(x, y) \]
- Cumulative Joint Distribution
  \[ F_{rs}(x, y) \]
- The Range
  \[ w_{rs}(x) = x(s) - x(r) \]
- Expected value of $x(r)$
  \[ E_r \]
Deriving $F_r$

We will assume the variables are i.i.d.

- First to derive $F_n(x)$

\[ F_n(x) = Pr\{\text{all } x_i \leq x\} \]
\[ = (Pr\{X \leq x\})^n \]
\[ = P^n(x) \]

- Now to derive $F_1(x)$

\[ F_1(x) = Pr\{\text{all } x_i \geq x\} \]
\[ = (Pr\{X \geq x\})^n \]
\[ = (1 - P(x))^n \]
General Formula

- Formula for $F_k(x)$

$$F_k(x) = Pr\{X_k \leq x\}$$

$$= Pr\{\text{at least } k \ x_i \leq x\}$$

$$= \sum_{j=k}^{n} Pr\{\text{exactly } j \ x_i \leq x\}$$

$$= \sum_{j=k}^{n} \binom{n}{j} P^j(x)(1 - P(x))^{n-j}$$
Deriving $f_{rs}$

- By thinking about where we need the variables to land in order to get a specified $x$ and $y$ value we can understand where the formula for the joint distribution comes from.

\[
f_{rs}(x, y) = C_{rs} P^{r-1}(x)p(x)[P(y) - P(x)]^{s-r-1}p(y)[1 - P(y)]^{n-s}
\]

where $C_{r,s} = \frac{n!}{(r - 1)!(s - r - 1)!(n - s)!}$
Deriving $W_{rs}$

- Logically we want to integrate over all of the points that give us the desired range. This we can do since we have a formula for the joint distribution of two order statistics.

\[
 f(W_{r,s}) = \int_{-\infty}^{\infty} f_{rs}(x, x + W_{r,s})
\]
Formula For $E_r$

- We start with the standard formula for expected value

$$E_r = \int_{-\infty}^{\infty} x f_r(x)$$

- Many people take the below formula and change variables to simplify.

$$E_r = \frac{n!}{(r-1)! (n-r)!} \int_{-\infty}^{\infty} x P^{r-1}(x)p(x)[1 - P(x)]^{n-r}$$
Example: Batteries

- Consider a machine that uses 10 batteries. Because it has so many, it doesn’t shut off until half of the batteries are dead.
  - Given the batteries all have a uniform distribution on the interval \([.5, 1]\) years, what is the probability that the machine dies before .75 years? What is the expected time when the fifth battery will die?
- Running the machine with 3 dead batteries lowers efficiency and costs $1 a day.
  - How much money will the company spend between the 3rd battery death and the 5th battery death? This entails finding the range \(W_{3,5}\).
Example: Batteries

Here \( n=10 \), \( k=5 \)

\[
F_5(x) = \sum_{j=5}^{10} \binom{10}{j} P^j(x)[1 - P(x)]^{10-j}
\]

\[
F_5(0.75) = \sum_{j=5}^{10} \binom{10}{j} P^j(0.75)[1 - P(0.75)]^{10-j}
\]

\[
F_5(0.75) = \sum_{j=5}^{10} \binom{10}{j}(0.5)^{10}
\]

\[
F_5(0.75) = 62.3\%
\]
Example: Batteries

To calculate expected value we need:

\[ p(x) = 2, \quad 0.5 \leq x \leq 1 \quad \text{and} \quad P(x) = 0, \quad x < 0.5 \]
\[ p(x) = 0 \text{ otherwise} \quad \text{and} \quad P(x) = 2x - 1, \quad 0.5 \leq x \leq 1 \]
\[ P(x) = 1, \quad x > 1 \]

\[ E(X_{(5)}) = 10 \binom{9}{4} \int_{-\infty}^{\infty} x P^4(x) [1 - P(x)]^5 p(x) \, dx \]
\[ E(X_{(5)}) = 10 \binom{9}{4} \int_{0.5}^{1} x (2x - 1)^4 [2 - 2x]^5 (2) \, dx \]
\[ = 0.7272 \]
Example: Batteries

What is the expected value of $W_{3,5}$?

$$E(W_{3,5}) = \int W_{3,5} f(W_{3,5})$$

$$f(W_{3,5}) = C_{3,5} \int_{0.5}^{1-W_{3,5}} (2x-1)^2(2)(2W_{3,5})(2)(2-2x-2W_{3,5})^5 dx$$

$$E(W_{3,5}) = (15120)(2^8) \int_0^{0.5} \int_{0.5}^{1-y} (2x-1)^2y^2(1-x-y)^5 dxdy$$

$$E(W_{3,5}) = \frac{1}{11} \text{ of one year}$$

This costs the company approx. $34
Resources

- Sean Ballentine
- *Order Statistics* by H. A. David
- *Probability and Statistical Inference* by Robbert V. Hogg and Elliot A. Tanis