Advanced Statistics

7. Order Statistics

Christian Aßmann

Chair of Survey Statistics and Data Analysis – Otto-Friedrich-Universität Bamberg

In this course we discussed fundamental ideas of probability theory and the theory of distributions.

There we considered the probability space of a random experiment, given by the 3-tuple

$$\{S, Y, P\},$$

where

- S: sample space being the set of all outcomes of the experiment
- Y: event space being the set of all events (typically a sigma-algebra on S)
- P: probability set function having domain Y used to assign probabilities to events.

There a typical question was:

> *Given the probability space, what can we say about the characteristics and properties of outcomes of an experiment?*

Example

Consider the experiment of tossing a fair coin 50 times. Assume that we are interested in the number of heads, say X. We know that this rv has a binomial distribution with parameters $n = 50$ and $p = 0,5$.
Hence, we have a completely specified probability space with a sample space $S = \{0, 1, \ldots, 50\}$, an event space Y consisting of all subsets of S, and a probability set function P characterized by the pdf of a binomial distribution.
From this we can deduce the characteristics of the outcomes of X like the expected number of heads ($np = 25$) or the shape of the pdf of X.

We now turn to the question of the probability theory and the theory of distribution around:

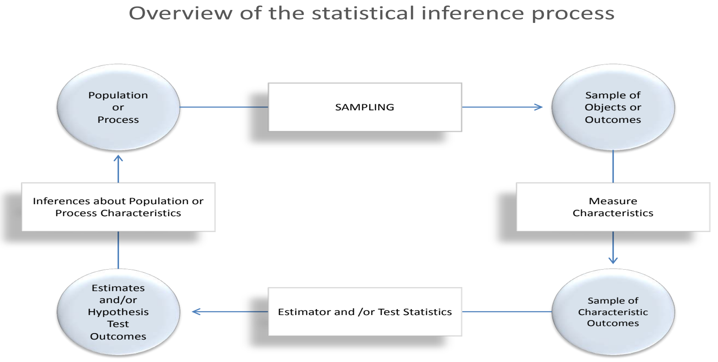
Given the observed characteristics and properties of outcomes of an experiment, what can we say (infer) about the probability space?

Example

Assume that we have a sample of daily returns observed for the German stock index DAX, which we interpret as the outcomes of a random process / experiment.
As a financial analyst we might be interested in finding the probability distribution (i.e. the probability space), which can be used to describe or approximate the observed behavior of returns.

Problems associated with this kind of question are addressed by the methods of **statistical inference**.

In general, the term statistical inference refers to the inductive process of generating information about characteristics of a population or process, by analyzing a sample of objects or outcomes from the population or process.
Figure 5 provides an overview of the process of statistical inference.[1]



[1] Source: Mittelhammer 1996, Fig. 6-1.

A typical problem of statistical inference is as follows

- Let X be a rv that represents the population under investigation, and let $f(x, \theta)$ denote the parametric family of pdfs of X. The set of possible parameter values is denotes by Ω.
- Then the job of the statistician is to decide on the basis of a sample randomly drawn from the population, say $\{X_i, i = 1, \dots, n\}$, which member of the family of pdfs $\{f(x, \theta), \theta \in \Omega\}$ can represent the pdf of X.

Example

Consider a random sample of daily DAX returns, say $\{X_1, X_2, \dots, X_n\}$. Assume that the returns represent a random sample from a normal distribution, i.e.

$$X_i \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2), \quad i = 1, \dots, n,$$

where μ and σ^2 are unknown parameters.
Our task is to generate statistical inferences based upon the random sample about the population values for μ and σ^2.

In statistical inference, we use **functions of the random sample** $X_1, \ldots, X_n$ to map / transform sample information into inferences regarding the population characteristics of interest. The functions used for this mapping are called **statistics**, defined as follows.

Definiton (Statistic)

Let $X_1, \ldots, X_n$ be a random sample from a population and let $T(x_1, \ldots, x_n)$ be a real-valued function, which does not depend on unobservable quantities. Then the random variable

$$Y = T(X_1, \ldots, X_n)$$

is called a (sample) statistic.

Often used statistics are

- the sample mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$,
- the rth order non-central sample moment $M_r' = \frac{1}{n}\sum_{i=1}^{n} X_i^r$,
- the sample variance $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$.

If statistics have certain qualifying statistical properties they can be used for the estimation of population parameters or hypothesis testing. Then they are called *estimators* or *test statistics*, respectively.

The most popular statistic is the **sample mean** $\bar{X}_n$ which has a lot of useful statistical properties, some of which are summarized as follows.

Summary (Sample mean $\bar{X}_n$)

Let $X_1, \ldots, X_n$ be a random sample from a population with expectation $E[X] = \mu$ and variance $Var[X] = \sigma^2$. Then the sample mean has the following properties

- $E[\bar{X}_n] = \mu$,
- $Var[\bar{X}_n] = \sigma^2/n$,
- $plim \bar{X}_n = \mu$, *which follows from the WLLN,*
- $\bar{X}_n \overset{asy}{\sim} \mathcal{N}(\mu, \sigma^2/n)$, *which follows from the CLT of Lindberg-Levy.*

In the next section we will discuss a special sample statistic, the so called *order statistic*.

Situations arise in practice, where we are interested in the largest or smallest value in a random sample rather than the average value. Examples are

- When constructing a dike we are interested in the highest flood water level rather than the average water level;
- As a risk manager of a portfolio of risky assets we are interested in the smallest portfolio return, whose expected values represents the expected maximal loss.

The largest and the smallest value in a random sample are examples of order statistics.

Definiton (Order Statistic)

Let $X_1, X_2, \dots, X_n$ be a random sample. Then $X_{[1]} \leq X_{[2]} \leq \dots \leq X_{[n]}$, where the $X_{[i]}$s are the X_is arranged in order of increasing magnitudes, are the **order statistics** *of the sample $X_1, X_2, \dots, X_n$. The variable $X_{[i]}$ is called the ith order statistic.*

Remark

Note that the order statistics are indeed statistics, since they are defined as functions of the random sample.

$\diamondsuit$

Example

Let the rv X be the return of a portfolio of risky assets. Then the 1st oder statistic $X_{[1]} = \min\{X_1, \dots, X_n\}$ is a critical value for a risk manager. He might be interested in the probability

$$Pr(X_{[1]} \leq -10\%).$$

Note that we need the sampling distribution of $X[1]$ in order to compute this probability.

The following theorem establishes the cdf for the sampling distribution of the kth order statistic for a random sample from a populaton distribution (i.e. a sample consisting of iid random variables).

Theorem (1.1)

Let $(X_1, \ldots, X_n)$ denote a random sample from a population distribution with cdf F, and let $X_{[k]}$ denote the kth order statistic. Then the cdf of $X_{[k]}$ is given by

$$F_{X_{[k]}}(b) = \sum_{j=k}^{n} \binom{n}{j} F(b)^j [1 - F(b)]^{n-j}.$$

Proof

For a given b, we define the rv $Y_i = I_{(-\infty,b]}(X_i)$, which is Bernoulli distributed with $p = Pr(Y_i = 1) = Pr(X_i \leq b) = F(b)$. Since the Y_is are iid, it follows that

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} I_{(-\infty,b)}(X_i) \sim \mathcal{B}in(n, p = F(b)).$$

Now note the equivalance of the following two events

$$\underbrace{\{x_{[k]} \leq b\}}_{\text{(event that the } k\text{th largest outcome is less or equal to } b)} = \underbrace{\{\sum_{i=1}^{n} I_{(-\infty,b]}(X_i) \geq k\}}_{\text{(event that at least } k \text{ outcomes are less or equal to } b)}$$

(continues)

Proof (continued)

Since both events are equivalent they have the same probability. Thus the cdf of $X_{[k]}$ is obtained as

$$F_{X_{[k]}}(b) \stackrel{(def.)}{=} Pr(X_{[k]} \le b) = Pr(\sum_{i=1}^{n} y_i \ge k) = \sum_{j=k}^{n} \underbrace{\binom{n}{j} F(b)^j [1 - F(b)]^{n-j}}_{\text{(the pdf of } \sum_{i=1}^{n} y_i\text{)}} .$$

□

The cdf of the smallest and largest order statistic, $X_{[1]}$ and $X_{[n]}$, are obtained as special cases of Theorem 1.1.

Corollary (1.1)

Under the conditions of Theorem 1.1 the cdfs of $X_{[1]}$ and $X_{[n]}$ are given by

$$F_{X_{[1]}}(b) = 1 - [1 - F(b)]^n, \quad \text{and} \quad F_{X_{[n]}}(b) = F(b)^n.$$

Proof

From Theorem 1.1 we have

$$\begin{aligned}
F_{X_{[1]}}(b) &= \sum_{j=1}^{n} \binom{n}{j} F(b)^j [1-F(b)]^{n-j} \\
&= \underbrace{\sum_{j=0}^{n} \binom{n}{j} F(b)^j [1-F(b)]^{n-j}}_{\text{(sum of binomial pdf over its support = 1)}} - \binom{n}{0} F(b)^0 [1-F(b)]^{n-0} \\
&= 1 - [1-F(b)]^n.
\end{aligned}$$

Also,

$$F_{X_{[n]}}(b) = \sum_{j=n}^{n} \binom{n}{j} F(b)^j [1-F(b)]^{n-j} = \binom{n}{n} F(b)^n [1-F(b)]^{n-n} = F(b)^n.$$

□

Remark

Note that the particular analytical form of the cdf of the order statistics $F_{X_{[k]}}(b)$ *depends upon the analytical form of the cdf of the parent distribution* F*, i.e. the population distribution of the random sample.*

$\diamondsuit$

Example

Suppose that the life of a certain light bulb measured in hours and denoted by X *is exponentially distributed with pdf*

$$f(x;\theta) = \frac{1}{\theta}\exp\{-x/\theta\}I_{(0,\infty)}(x), \quad \textit{with} \quad E[X] = \theta = 1000 \textit{ hours.}$$

In a random sample of $n = 10$ *such light bulbs, what is the distribution of the life of the bulb that fails first and what is its expected life time?*

(continues)

Example (continued)

First recall that the cdf of X is $F(b) = 1 - \exp\{x/\theta\}$, then the cdf for the life of the bulb that fails first $X_{[1]}$ is

$$F_{X_{[k]}}(b) \overset{\text{(Cor. 1.1)}}{=} 1 - [1 - F_X(x)]^n = 1 - [1 - 1 + \exp\{-x/\theta\}]^n = 1 - \exp\{-xn/\theta\},$$

which is the cdf of an exponential distribution with parameter θ/n.
Thus $X_{[1]}$ is exponentially distributed with mean $E[X_{[1]}] = \theta/n = 1000/10 = 100$.

Remark

Having established the cdf of an order statistic we can use the duality between cdfs and pdfs in order to obtain the pdf of the order statistic.

- *If the random sample is from a discrete population distribution with support* $x_1 < x_2 < \ldots < x_n$, *then the pdf of* $X_{[k]}$ *is obtained as*

$$f_{X_{[k]}}(x_i) = F_{X_{[k]}}(x_i) - F_{X_{[k]}}(x_{i-1}) \quad \text{for } i \geq 2, \quad \text{with} \quad f_{X_{[k]}}(x_1) = F_{X_{[k]}}(x_1).$$

- *In case of a continuous population distribution, the pdf is obtained by differentiation of the cdf as*[a]

$$f_{X_{[k]}}(x) = \frac{dF_{X_{[k]}}(x)}{dx} = \frac{n!}{(k-1)!(n-k)!} f(x) F(x)^{(k-1)} [1 - F(x)]^{(n-k)}.$$

$\diamondsuit$

[a] For details of the algebraic manipulations see Casella and Berger (2002, p. 229).

Example

Let $\{X_1, \ldots, X_n\}$ *be a random sample from a uniform (0,1) distribution with pdf* $f(x) = 1$ *for* $x \in (0,1)$ *and cdf* $F(x) = x$*. Then the pdf of the* k*th order statistic is*

$$\begin{aligned} f_{X_{[k]}}(x) &= \frac{n!}{(k-1)!(n-k)!} f(x) F(x)^{k-1} [1 - F(x)]^{n-k} \\ &= \frac{n!}{(k-1)!(n-k)!} x^{k-1} (1-x)^{n-k} \\ &= \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)} x^{k-1} (1-x)^{(n-k+1)-1}. \end{aligned}$$

Thus, the k*th order statistic from a uniform(0,1) random sample has a beta*(α, β) *distribution with* $\alpha = k$ *and* $\beta = n - k + 1$*.*

Sofar, we considered the marginal distribution of one order statistic. The joint sampling distribution of any pair of order statistic is given in the following theorem.

Theorem (1.2)

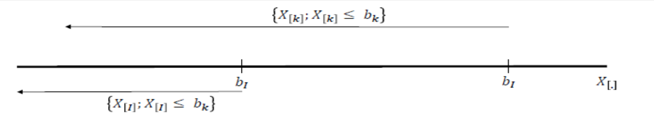
Let $(X_1, \ldots, X_n)$ be a random sample from a population distribution cdf F, and let $X_{[k]}$ and $X_{[\ell]}$, with $k < \ell$ denote the kth and ℓth order statistic. Then their joint pdf is given as

$$F_{X_{[k]},X_{[\ell]}}(b_k, b_\ell) = \begin{cases} F_{X_{[\ell]}}(b_\ell), & \text{for } b_k \geq b_\ell, \\ \sum_{i=k}^{n} \sum_{j=\max\{0,\ell-i\}}^{n-i} \frac{n!}{i!(n-i-j)!} F(b_k)^i & \\ \quad [F(b_\ell) - F(b_k)]^j [1 - F(b_\ell)]^{n-i-j}, & \text{for } b_k < b_\ell. \end{cases}$$

Proof

Case $b_k \geq b_\ell$: Given $X_{[k]} \leq X_{[\ell]}$, it follows that event $\{x_{[\ell]} \leq b_\ell\}$ implies event $\{X_{[k]} \leq b_k\}$, see figure 2. Hence we have

$$F_{X_{[k]},X_{[\ell]}}(b_k, b_\ell) = Pr(X_{[k]} \leq b_k, X_{[\ell]} \leq b_\ell) = Pr(X_{[\ell]} \leq b_\ell) = F_{X_{[\ell]}}(b_\ell).$$



(continuous)

Proof (continued)

Case $b_k \leq b_\ell$: Note that the event $\{X_{[k]} \leq b_k, X_{[\ell]} \leq b_\ell\}$ corresponds to the event {at least k of the X_is $\leq b_k$ and at least ℓ of the X_is $\leq b_\ell$}. This event can be represented as

$$\{X_{[k]} \leq b_k, X_{[\ell]} \leq b_\ell\} = \bigcup_{(i,j)\in I} \{i \text{ of the } X_i\text{s} \leq b_k \text{ and } j \text{ of the } X_j\text{s are such that } b_k < X_i \leq b_\ell\},$$

where $I = \{(i,j) : \max\{0, \ell - i\} \leq n - i; k \leq i < n\}$. Note that the events involved in the union operation are disjoint. In order to assign probabilities to those events we can use the multinomial distribution as follows. Categorize the outcomes of the X_is into one of the three types,

$$\{X_i \leq b_k\}, \quad \{b_k < X_i \leq b_\ell, \quad \{X_i > b_\ell\},$$

which occur with probability

$$F(b_k), \quad F(b_\ell - F(b_k), \quad 1 - F(b_\ell),$$

respectively.

(continuous)

Proof (continued)

Then directly applying the multinomial distribution yields

$$Pr(\{i \text{ of the } X_i s \leq b_k \text{ and } j \text{ of the } X_j s \text{ are such that } b_k < X_i \leq b_\ell\}) = \frac{n!}{i!(n-i-j)!} F(b_k)^i [F(b_\ell) - F(b_k)]^j [1 - F(b_\ell)]^{n-i-j}.$$

Finally, summing those probabilities of all the disjoint events in the union operation yields the second part of the theorem.

□

Remark

From the joint cdf of two order statistics we derive their joint pdf. If the random sample is discrete, the pdf would be obtained via differencing of the pdf.
In the continuous case the pdf is obtained by differentiation of the cdf, which yields

$$f_{X_{[k]},X_{[\ell]}}(x_k, x_\ell) = \frac{\partial^2 F_{X_{[k]},X_{[\ell]}}(x_k, x_\ell)}{\partial x_k \partial x_\ell} = \frac{n!}{(k-1)!(\ell-1-k)!(n-\ell)!}$$
$$f(x_k)f(x_\ell)F(x_k)^{k-1}[F(x_\ell)-F(x_k)]^{\ell-1-k}[1-F(x_\ell)]^{n-\ell},$$

for $x_k < x_\ell$. *This result for the joint pdf for two order statistics can be extended to derive the joint pdf of all order statistics for a continuous random sample, which is given by*

$$f_{X_{[1]},\ldots,X_{[n]}}(x_1, \ldots, x_n) = n! \prod_{i=1}^{n} f(x_i), \quad \textit{for} \quad x_1 < x_2 < \ldots < x_n.$$

$\diamondsuit$

In the previous section, we derived the marginal and joint distribution of order statistics. In this section, we will find the distribution of functions of order statistics, such as the sample median, the sample range, and the sample midrange.

Definiton (Sample Median)

Let $X_{[1]}, \ldots, X_{[n]}$ be the order statistics of a random sample of size n. Then the sample median is defined as

$$M = \begin{cases} X_{[k]}, & \text{if } n \text{ is odd with } n = 2k - 1, k \in \mathbb{N}, \\ (X_{[k]} + X_{[k+1]})/2, & \text{if } n \text{ is even with } n = 2k. \end{cases}$$

Hence the sample median is the middle order statistic (i n is odd) or the average of the middle two order statistics (if n is even) and is a measure if the center of the empirical distribution.

Definiton (Sample Range and Sample Midrange)

Let $X_{[1]}, \ldots, X_{[n]}$ be the order statistics of a random sample of size n. Then the sample range and the sample midrange are defined as

$$R = X_{[n]} - X_{[1]} \quad \text{and} \quad MR = (X_{[1]} + X_{[n]})/2,$$

respectively.

The sample midrange represents a further measure of the center of the empirical distribution, while the sample range is one measure for the dispersion of the empirical distribution.

- When the sample size n is odd with $n = 2k - 1$, the pdf of the sample median corresponds directly to the pdf of the kth order statistic, i.e.

$$f_M(m) = f_{X_{[k]}}(x_k).$$

- When the sample size n is even with $n = 2k$, the median is a function of two rvs. Then its pdf can be obtained by using the change-of-variable approach based opon the joint pdf of the two rvs.
  When the random sample is from a continuous distribution, the joint pdf of $X_{[k]}$ and $X_{[k+1]}$ is

$$f_{X_{[k]},X_{[k+1]}}(x_k, x_{k+1}) = \frac{n!}{(k-1)!(o)!(n-k-1)!} f(x_k) f(x_{k+1})$$
$$F(x_k)^{k-1}[F(x_{k+1}) - F(X_k)]^0 [1 - F(X_{k+1})]^{n-k-1},$$

  for $x_k < x_{k+1}$.

Now define the (2×1) vector function g as

$$\begin{pmatrix} v \\ m \end{pmatrix} = \begin{pmatrix} g_1(x_k, x_{k+1}) \\ g_2(x_k, x_{k+1}) \end{pmatrix} = \begin{pmatrix} x_k \\ (x_k + x_{k+1})/2 \end{pmatrix},$$

where $g_2(\cdot)$ is an auxiliary function introduced to allow for the use of the change-of-variable approach.
The function g is continuously differentiable with an inverse function g^{-1} arising by solving for x_k and x_{k+1} as

$$\begin{pmatrix} x_k \\ (x_k + x_{k+1})/2 \end{pmatrix} = \begin{pmatrix} g_1^{-1}(v, m) \\ g_2^{-1}(v, m) \end{pmatrix} = \begin{pmatrix} v \\ 2m - v \end{pmatrix}.$$

The Jacobian of the inverse g^{-1} is thus

$$J = \begin{pmatrix} \frac{\partial g_1^{-1}(\cdot)}{\partial v} & \frac{\partial g_1^{-1}(\cdot)}{\partial v} \\ \frac{\partial g_2^{-1}(\cdot)}{\partial v} & \frac{\partial g_2^{-1}(\cdot)}{\partial m} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 2 \end{pmatrix}, \quad \text{with} \quad |\det(J)| = 2.$$

Then by the multivariate change-of-variable approach the joint pdf of $V == X_{[k]}$ and $M = \frac{1}{2}(X_{[k]} + X_{[k+1]})$ is given by

$$\begin{aligned} f_{V,M}(v,m) &= |\det(J)| f_{X_{[k]},X_{[k+1]}}(\underbrace{g_1^{-1(v,m)}}_{X_{[k]}}, \underbrace{g_2^{-1}(v,m)}_{X_{[k+1]}}) \\ &= \frac{2n!}{(k-1)!(n-k-1)!} f(v) f(2m-v) F(v)^{k-1} [1 - F(2m-v)]^{n-k-1}. \end{aligned}$$

Finally, the marginal pdf of the sample median M is obtained by integrating v out from the joint pdf $f_{V,M}(v,m)$, i.e.

$$f_M(m) = \int f_{V,M}(v,m) dv.$$

Note that the result, i.e. the functional form of $f_M(m)$, depends upon the population distribution of the sample and the functional forms of F and f.

The sampling distributions of the sample range $R = X_{[n]} - X_{[1]}$ and the sample distribution of the midrange $MR = (X_{[1]} + X_{[n]})/2$ can be obtained in the same way we derived the distribution of the median. For further details see Mood, Graybill and Boes (1974, Chap. VI, 5.2) and Casella and Berger (2002, Chap. 5.4).

Theorem

Let $u_1, \ldots, u_{n-1}$ denote a random sample from a uniform (0,1) distribution with pdf $f(u) = I_{(0,1)}(u)$ and cdf $F(u) = u$. Denote the corresponding order statistics as $u_{[1]}, \ldots, u_{[n-1]}$. Then the pdf of

$$z = (z_1, \ldots, z_n) = (u_{[1]}, u_{[2]} - u_{[1]}, u_{[3]} - u_{[2]}, \ldots, u_{[n-1]} - u_{[n-2]}, 1 - u_{[n-1]})$$

is given as a Dirichlet distribution defined as

$$f(z) = \underbrace{\frac{\Gamma(\sum_{i=1}^{n} \alpha_i)}{\prod_{i=1}^{n} \Gamma(\alpha_i)}}_{\text{multinomial Beta function}} \prod_{i=1}^{n} z_i^{\alpha_i - 1},$$

where $\alpha_1, \ldots, \alpha_n > 0$, $z_i \in (0,1)$, $\sum_{i=1}^{n} z_i = 1$, and $\alpha_1 = \ldots, \alpha_n = 1$.

Proof

Note that the joint pdf of the $n-1$ order statistics $u_{[1]}, \ldots, u_{[n-1]}$ for a random sample with $n-1$ elements is given in general as

$$(n-1)! \prod_{i=1}^{n-1} f(u_{[i]}) I_{(u_{[1]} \leq u_{[2]} \leq \cdots \leq u_{[n-1]})}(u_{[1]}, \ldots, u_{[n-1]}).$$

(continuous)

Proof (continued)

Hence, in this context we have

$$(n-1)!I_{(0<u_{[1]}\leq u_{[2]}\leq\cdots\leq u_{[n-1]}<1)}(u_{[1]},\ldots,u_{[n-1]}).$$

We derive the joint distribution of z by the change-of-variables technique. In order to have a one-to-one transformation we augment the vector of $n-1$ order statistics by a degenerate random variable x with $f(x)=I_{\{1\}}(x)$. We then have the joint distribution of z given as

$$f(z)=(n-1)!I_{(0<z_1<z_2+z_1<\ldots<\sum_{i=1}^{n-1}z_i<1)}I_{\{1\}}(\sum_{i=1}^{n}z_i),$$

since $(u_{[1]},\ldots,u_{[n-1]},x)=g^{-1}(z)=(z_1,z_2+z_1,\ldots,\sum_{i=1}^{n-1}z_i,\sum_{i=1}^{n}z_i)$ *and*

$$|\det\left(\frac{\partial g^{-1}(z)}{\partial z}\right)|=|\det\left(\begin{pmatrix}1&0&&\cdots&0\\1&1&0&&\\1&1&1&0&\\&&&&0\\1&1&1&1&1\end{pmatrix}\right)|=1.$$

(continuous)

Proof (continued)

Note that $f(z)$ is a Dirichlet distribution with $\alpha_1 = \alpha_2 = \ldots = \alpha_n = 1$ implying $\Gamma(\sum_{i=1}^{n} \alpha_i) = (n-1)!$.

$\diamondsuit$

Remark

To highlight the statistical properties of each z_i it is more convenient to start with the bivariate joint density of two order statistics. We have

$$\begin{aligned} f(u_{[1]}) &= (n-1)(1-u_{[1]})^{n-2}, \\ f(u_{[n-1]}) &= (n-1)u_{[n-1]}^{n-2}, \\ f(u_{[i-1]}, u_{[i]}) &= \frac{(n-1)!}{(i-2)!(i-1-i+1)!(n-1-i)!} u_{[i-1]}^{i-2}(1-u_{[i]})^{n-i}. \end{aligned}$$

(continuous)

Remark (continued)

This yields

$$E[z_1] = E[u_{[1]}] = \frac{1}{n}, \quad Var[z_1] = Var[u_{[1]}] = \frac{n-1}{n^2(n+1)},$$
$$E[z_n] = E[1 - u_{[n-1]}] = \frac{1}{n}, \quad Var[z_n] = Var[u_{[n-1]}] = \frac{n-1}{n^2(n+1)},$$
$$E[z_i] = E[u_{[i]} - u_{[i-1]}] = \frac{1}{n}.$$

Note that the Dirichlet distribution plays a crucial role within the Bayesian bootstrap.