ORDER STATISTICS

Ashwin, Emma, Leland

1. Introduction

1.1. What are Order Statistics?

In 1599, Edward Wright authored the book *Certaine Errors in Navigation* wherein he has a section discussing the use of compasses to determine a location. Here he outlined methods of estimating "true values" through noise. It is not entirely clear that he used the median here, however scholars believe he may have, writing that "the middlemost [point] is likest to come nearest the truth"[1].

Later in 1699, in a more definitive use, brothers Chistiaan and Lodewijk Huygens had discussions about life expectancy. Here they made the distinction between the median of a distribution and the mean[1]. The two statistics offered equally intuitive, but distinct results: the mean is the average of expected lifetime, while the median captures half of the density, meaning someone would have a 50% chance to live to this age. For the problems of the time, however, the mean turned out to be a much more useful tool.

These two examples are almost surely not the first time people considered using medians for statistical analyses, but they offer cogent examples for the intuitive usefulness of such statistics. Medians are a particular instance of a type of summary computed from data (under particular conditions, which we'll discuss shortly), more generally known as order statistics.

Order statistics are the observed values in a sample arranged in an increasing order. For example, if you have a list of test scores, the smallest score would be the minimum—the "first order" statistic. The largest score would be the maximum—the "last order" statistic. More generally, the k-th order statistic is the k-th value in order from the start, i.e. the k-th smallest.

The median, defined as the center-most value, is then in the context of order statistics defined only for odd sized samples, where we can find a value so that there are an equal number of observations greater and less than it. In the case of even sized samples, the median is often defined as the average between the two center-most values–however, this value would not represent an order statistic as it is not itself actually observed in the data. Finally, we can define the median of a distribution as the 0.5-th quantile. Here we will generally only work with finite samples and on individual order statistics, meaning that we will disregard the latter two definitions, though the notion of quantiles will prove relevant later.

For this report, we drew from a variety of sources. Most notably, drawing from structure or using following similar methods to various lecture notes[2,3,4]. We also used results from various articles and sections of texts for particular results[5,6,7,8]. Finally, we had a few other sources[9,10,11] which helped guide us, but we did not concretely pull results from.

It is also worth noting that in writing this, we, in absence of checking the rubric, overwrote. We did not want to remove this extra work so we relegated many proofs to appendix 1. These are not necessary for this report to be complete though.

1.2. **Definitions, Notation, Significance.**

In order to treat order statistics in a rigorous manner and derive theoretical properties like distributional forms or asymptotic behavior, we will first need a coherent environment and explicit definitions for order statistics.

Definition 1.2.1. (Notation) Let i take on any value in $i \in \{1, \ldots, n\}$. Let X_i be individual realizations from a probability distribution $p(X)$. That is, $X_i \overset{iid}{\sim} p(X)$ for i. In the absence of any subscript, call $f(x)$ the pdf of X and $F(x)$ the cdf of X. Call $\mathbf{X^n} = (X_1, \ldots, X_n)$ the collection of X_i. We will generally also assume that the distribution $p(X)$ is continuous unless otherwise stated.

Definition 1.2.2. (Order Statistic). For a collection of observations $\mathbf{X^n}$. We define the order statistics generally $X_{(k)}$ the k-th smallest value in $\mathbf{X^n}$. We call $\mathbf{X^{(n)}} = (X_{(1)}, \ldots, X_{(n)})$ the collection of order statistics.

With this definition, we see the familiar min and max represented by $X_{(1)}$ and $X_{(n)}$ respectively. Furthermore, when n is odd, the usual median of $\mathbf{X^n}$ is $X_{(\frac{n+1}{2})}$, leaving $\frac{n-1}{2}$ values greater than and less than $X_{(\frac{n+1}{2})}$ as hoped.

Example 1.2.3. (Order statistics of uniform draws) Consider ten *iid* draws from a standard uniform distribution. We do not expect successive draws to be in ascending order, however we can sort them and compute the order statistics. Figure 1.2.1 shows the ten such draws plotted with order statistics labeled below and the order in which they were drawn labeled above.

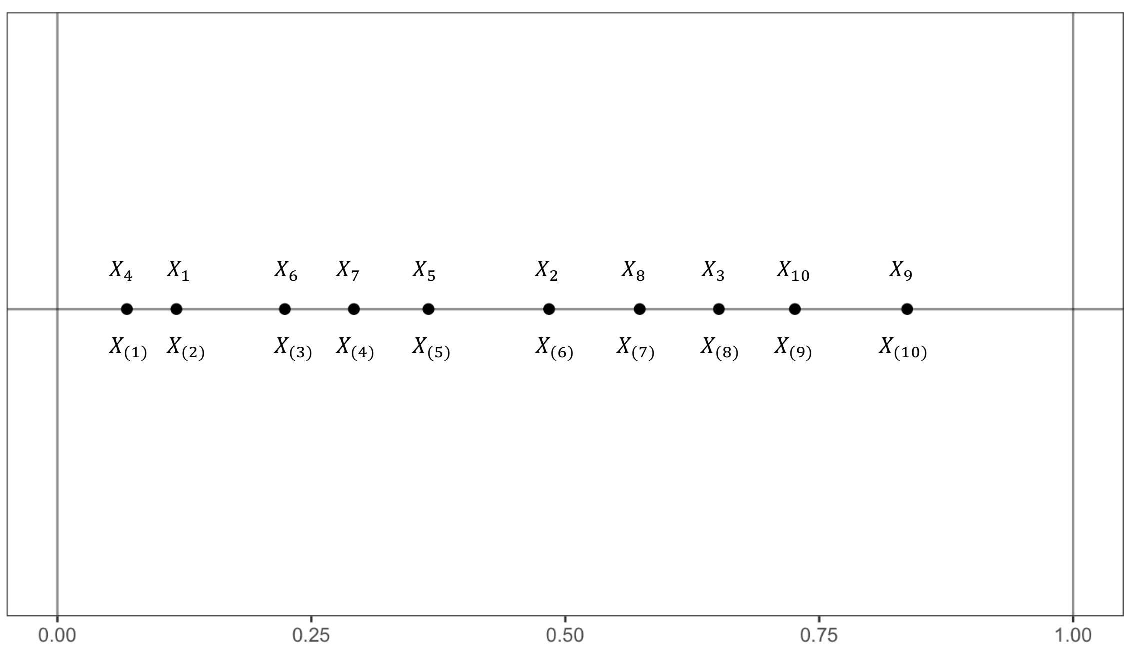

Figure 1.2.1: Draws from a standard uniform are plotted, as well as labels to indicate the order in which they were drawn and what order statistic they are.

What if we believe that an order statistic is unusually low? Can we use inference to test this claim? If we work with a general n, as n gets large, some tools from earlier in this course provide the promise of a solution for the maximum:

Claim 1.2.4. (an attempt to characterize the asymptotics of $X_{(n)}$; instant failure) Note that $X_{(n)} = \max\{\mathbf{X^n}\}$. So under the assumption that $\mathbf{X^n}$ comes from the $Unif(0,\theta)$ distribution, $X_{(n)}$ is $\hat{\theta}_{MLE}$. Further, MLEs are asymptotically normal with mean θ and variance $1/I(\theta)$. So we can make confidence intervals and hypothesis tests on $X_{(n)}$ using these asymptotics.

Although the above claim is well-meaning, it is unfortunately not true given the way we formulated the asymptotics of MLEs. The main problem with this is that the regularity conditions we have used for the asymptotic normality of MLEs do not hold—that is, the information does not exist for a uniform distribution, and so the parameters of the normal distribution are incoherent.

In order to determine exactly how order statistics behave asymptotically, we will need to put in some work to determine their pdfs before we can use previous results. To do this, we will start with the joint pdf of all the order statistics.

2. Theoretical Foundations

We will build the statement of distributional form up from a generic continuous probability density function. There are methods of dealing with order statistics coming from discrete probability distributions, however we will not consider these problems here. The derivation of distribution functions is inspired by University of Colorado Boulder lecture notes[2].

2.1. **Distribution Functions.**

We cannot start to derive the density of an order statistic on their own, so we will build up theory for the joint pdf of all order statistics. We will then provide methods to find the density of a subset of the order statistics.

Theorem 2.1.1. (Joint pdf of all $X_{(i)}$) The joint pdf of all $X_{(i)}$ is

$$f_{X_{(1)},\dots,X_{(n)}}(x_1,\dots,x_n) = n!\Big[f(x_1)\cdot\dots\cdot f(x_n)\Big].$$

Proof.
To begin, we will show that the joint cdf $F_{\mathbf{X^{(n)}}}(x_1,\dots,x_n) = n!F_{\mathbf{X^n}}(x_1,\dots,x_n)$; for simplicity call $(x_1,\dots,x_n) = \mathbf{x}$. Then we can quote independence and take derivatives to get our result.

To find the cdf of $\mathbf{X^{(n)}}$ observe that the cdf is the probability of the event

$$E_{(n)} : X_{(1)} < x_1, \dots, X_{(n)} < x_n, \quad \text{or,} \quad \mathbf{X^{(n)}} \overset{\text{coordinatewise}}{<} \mathbf{x}.$$

Notice that $\mathbf{X^{(n)}}$ is a permutation of $\mathbf{X^n}$. For distinct permutations $\pi_1,\dots,\pi_{n!}$ of $\mathbf{X^n}$ the probability of $E_{(n)}$ is equal to the sum of the probability of events:

$$E_1 : \pi_1(\mathbf{X^n}) \overset{\text{coordinatewise}}{<} \mathbf{x},$$

$$\vdots$$

$$E_{n!} : \pi_{n!}(\mathbf{X^n}) \overset{\text{coordinatewise}}{<} \mathbf{x}.$$

Once can see that these events are disjoint, because if all coordinates of $\pi_i(\mathbf{X^n})$ are less than $\mathbf{x}$, then we cannot have all coordinates of $\pi_j(\mathbf{X^n})$ be necessarily less than $\mathbf{x}$ since at least two coordinates in $\pi_j(\mathbf{X^n})$ must be displaced, and thus do not have the same order.

Now since X_i are *iid* from $p(X)$, we do not favor any of the E_i more than others, and so we can claim that

$$P(E_{(n)}) = \sum_{i=1}^{n!} P(E_i) = n!p(E_j), \text{ for any } j \in \{1, \dots, n!\}.$$

Going forward, we choose the j so that π_j is the identity, meaning that $P(E_j)$ is the the cdf of $\mathbf{X^n}$, just as $P(E_{(n)})$ is the cdf of $\mathbf{X^{(n)}}$. Hence

$$F_{\mathbf{X^{(n)}}}(\mathbf{x}) = n!F_{\mathbf{X^n}}(\mathbf{x}) \overset{X_i\ iid}{=} n![F(x_1)\dots F(x_n)].$$

Now, note that in order to work with the pdf, we have to take the $x_1, \dots, x_n$ mixed partial of the cdf. So compute that

$$\begin{aligned} f_{X_{(1)},\dots,X_{(n)}}(x_1, \dots, x_n) &= n!\frac{\partial}{\partial x_1}\cdots\frac{\partial}{\partial x_n}\left[F(x_1)\dots F(x_n)\right], \quad \text{(Definition of pdf/cdf)} \\ &= n!\left[\frac{\partial}{\partial x_1}F(x_1)\dots\frac{\partial}{\partial x_n}F(x_n)\right], \\ &= n!\Big[f(x_1)\dots f(x_n)\Big]. \end{aligned}$$

□

Having the joint pdf of $\mathbf{X^{(n)}}$ is powerful because it allows us to find the pdf for a subset of $\mathbf{X^{(n)}}$. Generally, all we need to do is integrate off the $X_{(i)}$ and corresponding x_i that we don't want from the joint pdf. While this may seem challenging, the process simplifies greatly with when considered in some generality. The next result provides some identities to do this.

Corollary 2.1.2. (Joint pdf of $\{X_{(i)}\}_{i\in I}$) Let I, J be normally ordered disjoint subsets of $\{1, \dots, n\}$ with size $|I| = \ell$, $|J| = m$. Enumerate values in I and J as i_k or j_k being the k-th value in I or J respectively. Then the following three identities hold:

(1) If for all $j \in J$, j is less than any $i \in I$, and the joint pdf of $\{X_{(i)}\}_{i\in I\cup J}$ is

$$c\Big[F(x_{j_1})^k f(x_{j_1})\cdots f(x_{j_m})f(x_{i_1})\cdots f(x_{i_\ell})g(x_{i_\ell})\Big],$$

then the joint pdf of $\{X_{(i)}\}_{i\in I\cup J\setminus\{j_1\}}$ is

$$\frac{c}{k+1}\Big[F(x_{j_2})^{k+1} f(x_{j_2})\cdots f(x_{j_m})f(x_{i_1})\cdots f(x_{i_\ell})g(x_{i_\ell})\Big]$$

for $k \in \mathbb{N}$, $c \in \mathbb{R}$, and generic function $g(x_{i_\ell})$, best thought of as a cdf exponentiated to some term.

(2) If for all $i \in I$, i is less than $j \in J$, and the joint pdf of $\{X_{(i)}\}_{i\in I\cup J}$ is

$$c\Big[g(x_{i_1})f(x_{i_1})\cdots f(x_{i_\ell})f(x_{j_1})\cdots f(x_{j_m})[1 - F(x_{j_m})]^k\Big],$$

then the joint pdf of $\{X_{(i)}\}_{i\in I\cup J\setminus\{j_m\}}$ is

$$\frac{c}{k+1}\Big[g(x_{i_1})f(x_{i_1})\cdots f(x_{i_\ell})f(x_{j_1})\cdots f(x_{j_m-1})[1-F(x_{j_{m-1}})]^{k+1}\Big]$$

once again for integer k, constant c, and generic function g.

(3) If all $j\in J$ there are two values $i_a, i_b\in I$ so that $i_a<j<i_b$, and the joint pdf of $\{X_{(i)}\}_{i\in I\cup J}$ is

$$c\Big[g(x_{i_1})f(x_{i_1})\cdots f(x_{i_a})f(x_{j_1})\cdots f(x_{j_m})[F(x_{i_b})-F(x_{j_m})]^k f(x_{i_b})\cdots f(x_{i_\ell})f(x_{i_\ell})\Big],$$

then the joint pdf of $\{X_{(i)}\}_{i\in I\cup J\setminus\{j_m\}}$ is

$$\frac{c}{k+1}\Big[g_1(x_{i_1})f(x_{i_1})\cdots f(x_{i_a})f(x_{j_1})\cdots f(x_{j_{m-1}})[F(x_{i_b})-F(x_{j_{m-1}})]^{k+1}f(x_{i_b})\cdots f(x_{i_\ell})g_2(x_{i_\ell})\Big],$$

where now g_1 and g_2 are similar but possibly distinct functions.

See section 1 of appendix 1 for an outline of the proof. The form of corollary 2.1.2 does not seem useful at first, however upon inspection, one can use this to compute any joint pdf of order statistics, defining I and J as necessary. The following example shows how one might do this by computing the distribution of just one order statistic:

Example 2.1.3. (pdf of one order statistic) Let $X_i\sim f(X)$, then we can compute $f_{X_{(k)}}(x)$ using corollary 2.1.2:

First, define $I=\{k,\dots,n\}$ and $J=\{1,\dots,k-1\}$. The joint pdf of all the $\mathbf{X^{(n)}}$ is (by theorem 2.1.1)

$$f_{\mathbf{X^{(n)}}}(x_1,\dots,x_n)=n!f(x_1)\cdots f(x_n).$$

So when $c=n!$, $m=0$, and $g(x)=1$, we have that this is in the form of (1) from corollary 2.1.2, ie

$$f_{\mathbf{X^{(n)}}}(x_1,\dots,x_n)=c\Big[F(x_1)^m f(x_1)\cdots f(x_n)g(x)\Big],$$

meaning that

$$f_{X_{(2)},\dots,X_{(n)}}(x_2,\dots,x_n)=\frac{n!}{1}\Big[F(x_2)f(x_2)\cdots f(x_n)\Big].$$

Now redefine $J=\{2,\dots,k-1\}$, and repeat, now with the same $c=n!$ and new $m=1$. This can be repeated as necessary, after removing $X_{(p)}$, we would have $c=\frac{n!}{m}$, with $m=p$. After repeating this as necessary and removing all of the $\{X_{(i)}\}_{i=1}^{k-1}$, we will have

$$f_{X_{(k)},\dots,X_{(n)}}(x_k,\dots,x_n)=\frac{n!}{(k-1)!}F(x_k)^{k-1}f(x_k)\cdots f(x_n).$$

Now notice that this is in the form of (2) from corollary 2.1.2 with $I=\{x_k\}$, $J=\{k+1,\dots,n\}$, $c=\frac{n!}{(k-1)!}$, $m=0$, and $g(x_k)=F(x_k)^{m-1}$. This means that we can move down from x_n down to x_{k+1}, removing all of these from the joint pdf. Doing so gives us the pdf

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

This repeated application of Corollary 2.1.2 allows us to compute any joint pdf of order statistics. Of course, we did not use the case of (3) in the above example, however the reasoning is identical to what we showed. Since we have the singular pdf, we can apply this to a distribution to get a real density function:

Example 2.1.4. (pdf of a standard uniform order statistic) Let $X_i \sim Unif(0,1)$. Then the pdf of X is 1 and the cdf is the identity. So from example 2.1.3, we have

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} x^{n-1}(1-x)^{n-k}.$$

Definition 2.1.5. (Beta distribution) Recall that the Beta(α, β) distribution, defined strictly on the unit interval has pdf

$$f_{\text{Beta}(\alpha,\beta)}(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}.$$

The Beta distribution has mean $\frac{\alpha}{\alpha+\beta}$ and variance $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$. Figure 2.1.1 shows the pdf for various α and β:

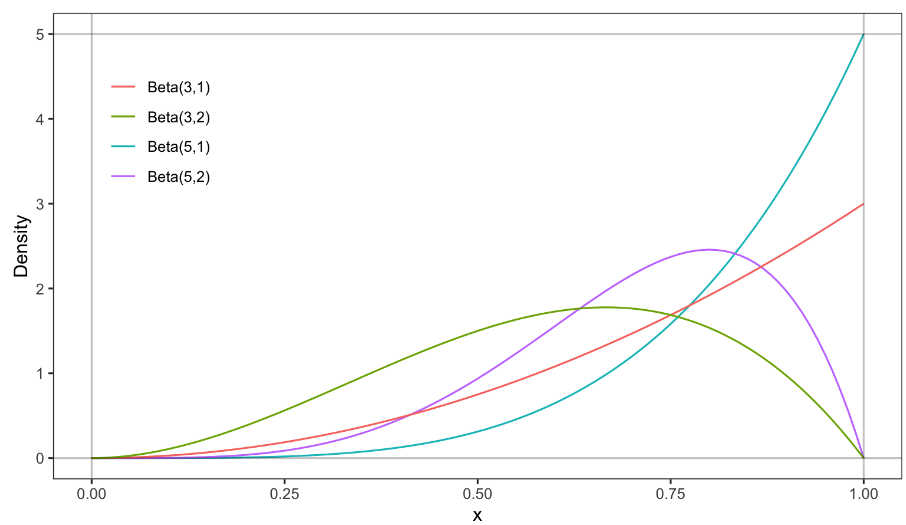

Figure 2.1.1: Densities of different Beta Distributions.

So the k-th order statistic $X_{(k)}$ of the standard uniform is Beta$(k, n-k+1)$ distributed.

This is a clean result, but the only reason we get this is because we are working with a standard uniform distribution—can this result be generalized past the uniform? Recall from homework 2 that for a random variable $X_i \sim p(X)$ with cdf F_X, the distribution $F_X(X) \sim$ Unif$(0,1)$. So, for a standard uniform random variable, U, we have the equivalent relation $F_X^{-1}(U) \sim p(X)$. So the upshot is that studying order statistics from the standard uniform is sufficiently generalizable for any probability distribution.

Another reasonable question to ask now that we have this result is how do we expect this to behave as our sample size gets larger and larger. What types of inference can we conduct?

2.2. **Asymptotics.**

Since $X_{(k)} \sim Beta(k, n-k+1)$–where $X_{(k)}$ is from the uniform–if we derive the asymptotics of $X_{(k)}$, using the relation of uniform to nonuniform random variables we should be able to have the asymptotics for most frequently used distributions.

The following derivations follow in the footsteps of a few different sources[4,5], and quote some identities[6,8]. We will list identities when they are relevant.

Many proofs here have been moved to the appendix since we need to build up quite a bit of supplementary machinery which are admittedly quite digressive before we can prove the results we want to.

We will start with a useful relation of the uniform order statistics and the beta distribution to some more frequently used random variables.

Lemma 2.2.1. Let $Y_k = \sum_{i=1}^{k} W_i$ and $Z_k = \sum_{i=k+1}^{n+1} W_i$ where our $W_i \overset{iid}{\sim} Exp(1)\ 1, \ldots, n$. Then $Q_k = Y_k/(Y_k + Z_k) \sim Beta(k, n+1-k)$.

The proof of this has been moved to section 2 of appendix 1. This demonstrates order statistics are equal in distribution to a quotient involving the sums of *iid* Exponential variables. This connection leads us to the following:

Theorem 2.2.2. Standard Uniform order statistics converge in distribution to a Normal. More precisely, for a given quantile q and a uniform order statistic corresponding closely to that quantile, $U_{(\lceil nq \rceil)}$, we have: $\sqrt{n}(U_{(\lceil nq \rceil)} - q) \xrightarrow{d} N\left(0, q(1-q)\right)$.

The proof of this result is lengthy so we have moved it to section 3 of appendix 1. Moving on, recall that studying order statistics from a uniform allows us to make statements about more general probability distributions. So we can generalize this.

Corollary 2.2.3. Given that $\sqrt{n}(U_{(\lceil nq \rceil)} - q) \xrightarrow{d} N\left(0, q(1-q)\right)$, we can claim that $\sqrt{n}(X_{(\lceil nq \rceil)} - F^{-1}(q)) \xrightarrow{d} N\left(\frac{q(1-q)}{[f(F^{-1}(q))]^2}\right)$.

Lemma 2.2.4. (delta method) Suppose that $\sqrt{n}(X_n - \theta) \xrightarrow{d} N(0, \sigma^2)$. Then for a differentiable function g, $\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow{d} N(0, \sigma^2[g'(\theta)]^2)$.

We have moved the proof of this to section 4 of appenxix 1. Moving swiftly on, we may now prove corollary 2.2.3.

Proof. (Corollary 2.2.3)
Given that we have $\sqrt{n}(U_{(\lceil nq \rceil)} \xrightarrow{d} N(0, q(1-q))$, we can apply the delta method with $g(x) = F^{-1}(x)$—the quantile function. Then, we have that

$$\sqrt{n}(U_{(\lceil nq \rceil)} - q) \xrightarrow{d} N(0, q(1-q)) \qquad \text{(Theorem 2.2.2)}$$

$$\sqrt{n}(F^{-1}(U_{(\lceil nq \rceil)}) - F^{-1}(q)) \xrightarrow{d} N\left(0, q(1-q)\left[\frac{d}{dq}F^{-1}(q)\right]\right). \qquad \text{(Lemma 2.2.4)}$$

At this point we make note of two things. First, the random variable obtained by evaluating the inverse cdf of random variable X at a uniform random variable is equal in distribution to X. Secondly, by the inverse function theorem,

$$\frac{d}{dx}f^{-1}(x) = \frac{1}{f'(f^{-1}(x))}.$$

Noting that the pdf is the derivative of the cdf, we get $\frac{d}{dq}F^{-1}(q) = \frac{1}{f(F^{-1}(q))}$. These identities fit well into the above statement of distribution:

$$\sqrt{n}(X_{(\lceil nq \rceil)} - x_q) \xrightarrow{d} N\left(0, \frac{q(1-q)}{[f(F^{-1}(q))]^2}\right).$$

□

Example 2.2.5. Consider the median, $X_{(\frac{n+1}{2})}$ for odd n, which will here be called $X_{(m)}$, as an estimator for the mean μ of a Normal distribution $N(\mu, \sigma^2)$. Note that $X_{(m)}$ corresponds to the 0.5-th quantile of the normal by symmetry.

We know $f(\mu) = \frac{1}{\sqrt{2\pi\sigma^2}}$, so by the above asymptotic normality we have:

$$\sqrt{n}(X_{(m)} - \mu) \xrightarrow{d} N\left(0, \frac{0.5(1-0.5)}{\frac{1}{\sqrt{2\pi\sigma^2}}^2}\right) = N\left(0, \frac{\pi\sigma^2}{2}\right).$$

Which is to say, the median is an asymptotically unbiased estimator for the mean of a Normal. Although, its variance is a constant scaled larger than the sample mean.

3. Practical Application

To complement the theory we have introduced regarding the distributions and limiting distributions of order statistics, we will apply results to theory. We will do this through confirmations that order statistics follow specified distributions. This will also provide visualizations for joint distributions. We will use simulated order statistics to do classical inference.

3.1. **Visualizations.**

To set up our simulation, we assume that each random variable X_i is drawn *iid* from the standard uniform distribution, i.e., $X_i \overset{iid}{\sim} \text{Uniform}(0,1)$. For our simulated sample, we draw $n = 10$ such observations, and repeat this 20,000 times. This constitutes our entire population model.

As discussed previously, the probability density function of the k-th uniform order statistic from a sample of size n has the $\text{Beta}(k, n+1-k)$ distribution. It can be shown, through repeated applications of corollary 2.1.2, that when we have two order statistics, i and j where $i < j$, the joint density function of $(X_{(i)}, X_{(j)})$ is

$$f_{X_{(i)},X_{(j)}}(x_i, x_j) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} x_j^{i-1}[x_j - x_i]^{j-i-1}[1-x_j]^{n-j}.$$

We now examine the joint behavior of order statistics by focusing on an interior pair: the 4th and 6th smallest observations, $(X_{(4)}, X_{(6)})$. Figure 3.1.2 shows sampled data.

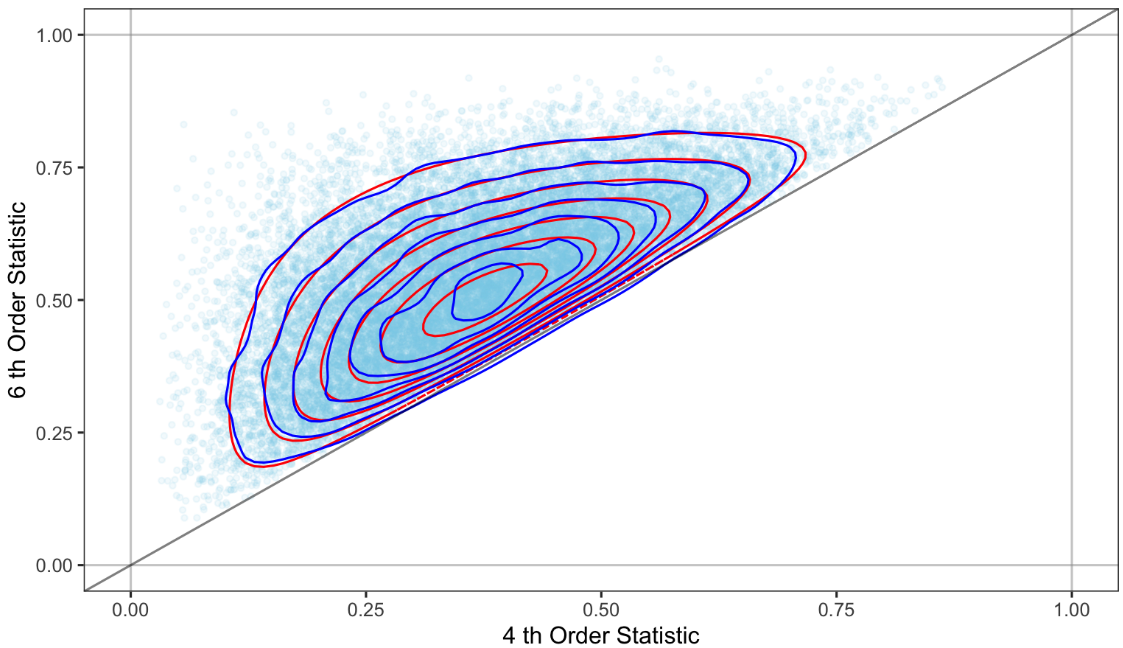

Figure 3.1.2 Sampled points with joint density contours for the vector $(X_{(4)}, X_{(6)})$. Blue points show particular sampled points. Dark blue lines and red lines are empirical and theoretical density contours respectively. The line through the slope marks off the region where we can have sampled order statistics from the region where we cannot.

The red contours overlay this theoretical density of $(X_{(4)}, X_{(6)})$, computed from the known formula for the joint distribution of order statistics. Meanwhile, the blue contours show the joint sampled density of the two order statistics. In general, the scatter cloud is tightly clustered and approximately symmetric around the -45-degree line, reflecting the fact that the 4th and 6th order statistics are close in rank and tend to fall near the center of the unit interval.

It's also worth noting that all points lie above the line with slope one and intercept at the origin. This reflects how order statistics have an inherent order, meaning that higher valued order statistics will be greater than lower valued ones. While the empirical density contours do extend below this line, this is likely a fragment of the function which creates the contours.

3.2. **Simulation Study: Median of a Normal Population.**

To check how well the large–sample theory works in practice for a non-uniform parent distribution, we simulated the scenario in which

$$X_i \overset{\text{iid}}{\sim} N(0,1), \qquad i = 1, \dots, 101,$$

and recorded the sample median $X_{(51)}$. Since mean μ is equal to the median for the normal by symmetry, the pdf of the normal at the mean is $\frac{1}{\sqrt{2\pi}}$. We can use our asymptotic theory developed in section 2.2 to claim that for odd n,

$$\sqrt{n}\Big(X_{(m)} - \mu\Big) \xrightarrow{d} N\left(0, \frac{0.5(1-0.5)\sigma^2}{\frac{1}{2\pi}}\right) = N\Big(0, \frac{\pi\sigma^2}{2}\Big), \qquad m = \frac{n+1}{2}.$$

With our simulation where $n = 101$ this gives the concrete approximation

$$X_{(51)} \;\approx\; N\Big(0, \tfrac{\pi\sigma^2}{202}\Big).$$

We tested the approximation with $25,000$ Monte-Carlo replications. Each replication generated 101 draws from standard normals and stored the median. Figure 3.1.3 shows the theoretical and empirical densities of the median. The near perfect overlap demonstrates that a sample size of roughly one hundred observations is already sufficient for the asymptotic normal limit to give an accurate description of the median.

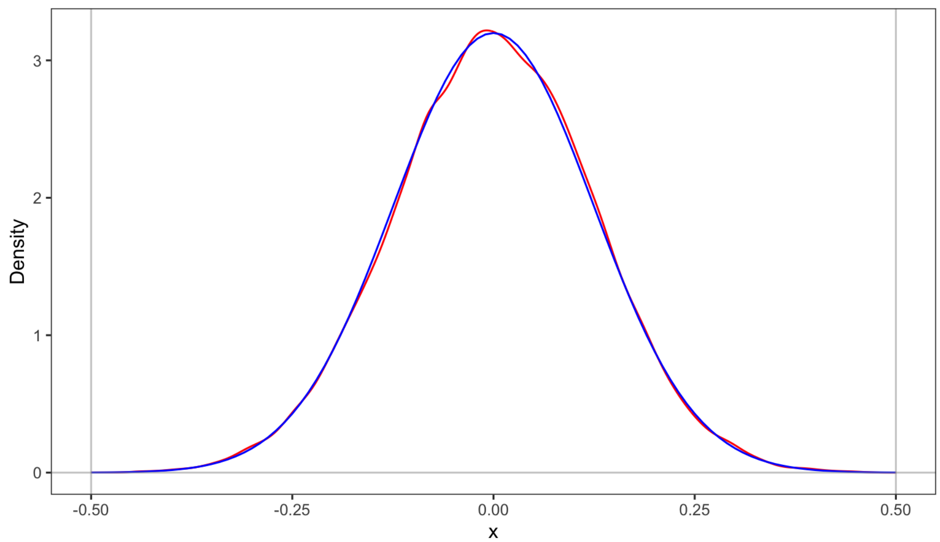

Figure 3.1.3: Theoretical (blue) vs Emperical (red) sampled densities of the median of $n = 101$ standard normal draws and 25,000 replicates.

3.3. **Classical Inference Using Only the Median.**

Because the approximate large–sample distribution of the sample median is known, we can perform full classical inference even when the median is the only statistic available.

Set-up: For the standard normal population used in Section 3.2, the asymptotic result

$$\sqrt{n}\Big(X_{(m)} - \mu\Big) \;\xrightarrow{d}\; N\Big(0, \tfrac{\pi\sigma^2}{2}\Big)$$

implies, with $n = 101$,

$$X_{(51)} \;\approx\; N\Big(\mu, \tfrac{\pi\sigma^2}{2\cdot 101}\Big).$$

Since the normal distribution is symmetric, the median $X_{(m)}$ is an *unbiased* estimator of the mean μ.

t–**statistic and confidence interval:** Replacing σ^2 by its usual unbiased estimator $\hat{\sigma}^2$ and dividing by the extra variance term means that

$$\frac{\sqrt{101}\Big(X_{(51)} - \mu\Big)}{\sqrt{\pi\hat{\sigma}^2/2}} \;\approx\; t_{100}.$$

Notice that this is exactly the result we use for a mean, but since the median has a different form of variance, we have extra terms. A typical 95 % confidence interval for μ based only on the median is then

$$\widehat{CI}_{0.95}(\mu) = X_{(51)} \ \pm \ t_{100}^{-1}(0.975)\sqrt{\frac{\pi\hat{\sigma}^2}{2\cdot 101}}.$$

Simulation check: Using the same simulation, we generated confidence intervals using the above theory for each observation of a median. Across all of the 25,000 intervals, 95.292% captured the mean, matching the nominal 95 % level.

4. Conclusions

In modern data analysis, the median has become a popular tool due to its natural robustness. While the sample mean is useful because it minimizes squared error and performs best when data are normally distributed, it is sensitive to extreme observations. Just one unusually large or small value can significantly skew the mean, especially with smaller sample sizes. The median, on the other hand, minimizes absolute error and can tolerate up to half of the data being extreme or incorrect before it breaks down. This makes the median particularly useful in real-world situations such as analyzing incomes, housing prices, or internet latencies, where extreme observations frequently occur. Additionally, our results show that the median, after appropriate scaling by $\sqrt{n}$, is asymptotically normal. This means that traditional statistical inference methods, such as confidence intervals and t-tests, can still be reliably used without requiring assumptions about variance or second moments.

Of course though, these results establish theoretical properties for more than just medians. We can use many of these tools to construct similar procedures for any order statistic. Many of the results we found for medians are true in some form for other order statistics. Broadly speaking, working with quantiles instead of explicit aggregations of the data allows for robustness since an order statistic is a single value.

Beyond robustness, order statistics underlie many of today's applied tools. Engineers use order statistics in reliability analysis[11], modeling the time to the first or k^{th} component failure, and in radio-frequency design, where the distribution of the strongest received signal guides antenna selection. In machine learning, quantile regression relies on piece-wise constant loss functions that are optimized at empirical medians and means[9]. Even everyday visual analytics tools, such as box-plots and violin plots, use order statistics (quartiles, deciles) to convey distributional shape at a glance.

5. References

1. Bakker, A., & Gravemeijer, K. P. E. (2006). An Historical Phenomenology of Mean and Median. Educational Studies in Mathematics, 62(2), 149–168. https://doi.org/10.1007/s10649-006-7099-8

2. "Order Statistics." University of Colorado Boulder. Lecture. https://www.colorado.edu/amath/sites/default/files/attached-files/order_stats.pdf

3. Colin Rundel. (2012). Lecture 15: Order Statistics. https://www2.stat.duke.edu/courses/Spring12/sta104.1/Lectures/Lec15.pdf

4. Chen Hung. (n.d.). Order Statistics. Nanyang Technological University. https://www.math.ntu.edu.tw/∼hchen/teaching/LargeSample/notes/noteorder.pdf

5. Ferguson, T. S. (2017). A Course in Large Sample Theory (1st ed.). Routledge. https://doi.org/10.1201/9781315136288

6. Curtiss, J. H. (1941). On the Distribution of the Quotient of Two Chance Variables. The Annals of Mathematical Statistics, 12(4), 409–421. https://doi.org/10.1214/aoms/1177731679

7. Mosteller, F. (1946). On Some Useful "Inefficient" Statistics. The Annals of Mathematical Statistics, 17(4), 377–408. https://doi.org/10.1214/aoms/1177730881

8. Casella, G., & Berger, R. (2024). Statistical Inference (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781003456285

9. IBM. (2024). Quantile Regression. IBM. https://www.ibm.com/docs/en/spss-statistics/saas?topic=regression-quantile

10. SteadyStoffel. (2023, March 30). Order Statistics. Rpubs. https://rpubs.com/SteadyStoffel/order_statistics1

11. Reliability Analysis. (n.d.). [Online post]. Science Direct. https://www.sciencedirect.com/topics/earth-and-planetary-sciences/reliability-analysis

6. Appendix 1: Assorted Derivations and Proofs

6.1. **Proof: Corollary 2.1.2.**

Proof. (Outline) In order to remove a variable from the pdf, we must integrate over all possible values it can take on, so we will do that for the above three cases. Figuring out the bounds of this integral though is is what sets each question apart.

For (1), we are working with the smallest order statistic in the joint pdf and so the bounds go from $-\infty$ to the next order statistic. For (2), we are working with the largest order statistic, and so the bounds go from the next largest to ∞. Finally, for (3), there are the two neighboring order statistics meaning we have the appropriate bounds ranging between the two order statistics.

The work of integrating this within the context of the larger expression is both largely unnecessary and obscures the mechanics of why these identities hold, so we will only compute the integrals with non constant terms relative to the integrand and show that they give the appropriate form, resulting in the expressions in Corollary 2.1.2. The remaining terms could be removed from the integral and reintroduced later.

Note that, as in theorem 2.1.1, we have imposed an order that $x_1 < x_2 < \cdots < x_n$. This means that integrating X_{j_1} out of the joint pdf means computing that

$$\int_{-\infty}^{x_{j_2}} f(x_{j_1})F(x_{j_1})^k \, dx_{j_1} = \int_{-\infty}^{x_{j_2}} F'(x_{j_1})F(x_{j_1})^k \, dx_{j_1} = \frac{F(x_{j_1})^{k+1}}{k+1}\bigg|_{-\infty}^{x_{j_2}} = \frac{F(x_{j_2})^{k+1}}{k+1}.$$

Here we used the fact that the pdf is the derivative of the cdf. We will continue to use this without explicitly stating the step.

Similarly, integrating X_{j_m} out of the joint pdf for (2) means (noting that $f(x)$ is the negative derivative of $1 - F(x)$) computing that

$$\int_{x_{j_{m-1}}}^{\infty} f(x_{j_m})[1 - F(x_{j_m})]^k \, dx_{j_m} = -\frac{[1 - F(x_{j_m})]^{k+1}}{k+1}\bigg|_{x_{j_{m-1}}}^{\infty} = \frac{[1 - F(x_{j_{m-1}})]^{k+1}}{k+1}.$$

Finally, integrating X_{j_m} out of the joint pdf for (3) means (noting that now $f(x)$ is the negative derivative of $F(u) - F(x)$), computing that

$$\int_{x_{j_{m-1}}}^{x_{i_b}} f(x_{j_m})[F(x_{i_b}) - F(x_{j_m})]^k \, dx_{j_m} = -\frac{[F(x_{i_a}) - F(x_{j_m})]^{k+1}}{k+1}\bigg|_{x_{j_{m-1}}}^{x_{i_a}} = \dots$$

$$\cdots = \frac{[F(x_{i_a}) - F(x_{j_{\ell-1}})]^{k+1}}{k+1}.$$

These are the general forms of the integrals needed to prove Corollary 2.1.2. The only step which remains is mechanical, substituting these into the larger expressions and verifying equality. We will not do this here as it takes too much room and is tough to follow as a reader.

$\square$

6.2. **Proof: Lemma 2.2.1:**

Proof. To begin, notice that Q_k must only have support on the unit interval, equaling zero when X_k equals zero, and one when Y_k equals zero (these events have probability zero, but as a hollistic these are useful). Furthermore, since Y_k and Z_k are sums of *iid* exponentials, they are respectively $Gamma(k, 1)$ and $Gamma(n+1-k, 1)$ distributed; these are also then independent.

A necessary identity for our formulation of this proof is as follows[6]: for the quotient X/Z of two random variables, the pdf of X/Z is

$$f_{X/Z}(w) = \int_{-\infty}^{\infty} |x| f_{X,Z}(wx, x)dx.$$

We can specialize this to Q_k. By definition of Q_k, and since it only has support on the positive reals, the pdf of Q_k is given by

$$f_{Q_k}(w) = \int_{0}^{\infty} x f_{Y_k, Y_k+Z_k}(wx, x)dx.$$

From here on, we will drop the subscripts on all random variables. Now consider the event

$$E_x : Y = wx \text{ and } Y + Z = x.$$

Since Y and Z are independent, we can say that this event is equal to the more palatable event

$$E'_x : Y = wx \text{ and } Z = (1 - w)x.$$

These events are technically have probability zero, but these are like densities. Therefore $f_{Y,Y+Z}(wx, x) = f_{Y,Z}(wx, (1-w)x)$. It is worth noting that the factor $(1-w)$ will not make the second argument negative since w is contained in the unit interval, so Z stays non-negative. Since Y and Z are independent, computing the pdf of Q becomes a matter of evaluating integrals:

$$\begin{aligned}
f_Q(w) &= \int_0^\infty x f_{Y,Z}\Big(wx, (1-w)x\Big)dx, \\
&= \int_0^\infty x f_Y(wx) f_Z\Big((1-w)x\Big)dx, && (Y, Z \text{ are independent}) \\
&= \frac{1}{\Gamma(k)\Gamma(n+1-k)} \int_0^\infty x(wx)^{k-1}\Big((1-w)x\Big)^{n-k} e^{-x(w+1-w)}dx, \\
&= \frac{1}{\Gamma(k)\Gamma(n+1-k)} w^{k-1}(1-w)^{n-k} \int_0^\infty x^n e^{-x}dx, \\
&= \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)} w^{k-1}(1-w)^{n-k}. && \text{(Recognize Gamma function)}
\end{aligned}$$

This is the pdf of the $Beta(k, n+1-k)$ distribution, completing the proof.

□

6.3. **Proof: Theorem 2.2.2.**

Proof.
Since the k-th uniform order statistic is equal in distribution to Q_k in the above lemma, we have to show that

$$\sqrt{n}\left(\frac{X}{X+Y}-q\right) \xrightarrow{d} N(0,q(1-q)).$$

Call $k = \lceil nq \rceil$ for now. Observe that we can expand the left hand side of this expression as follows:

$$\begin{aligned}\sqrt{n}\left(\frac{X}{X+Y}-q\right) &= \sqrt{n}\left(\frac{(1-q)X-qY}{X+Y}\right), \\ &= \frac{1}{\sqrt{n}}\left(\frac{(1-q)\sum_{i=1}^{k} W_i - q\sum_{i=k+1}^{n+1} W_i}{\frac{1}{n}\sum_{i=1}^{n+1} W_i}\right)\end{aligned}$$

Now note that n is asymptotically equivalent to $n+1$ meaning we can rewrite the denominator as $\bar{W}$ under the limit.

$$\approx \frac{1}{\sqrt{n+1}}\left(\frac{(1-q)\sum_{i=1}^{k} W_i - q\sum_{i=k+1}^{n+1} W_i}{\bar{W}}\right).$$

Now we will subtract relevant terms from both sums in the numerator to standardize random variables. Since the both terms are gamma distributions, they have expectations equal to the number of W_i's summed. Therefore, we get that

$$\begin{aligned}= \frac{1}{\sqrt{n+1}}\Bigg(&\frac{(1-q)\left[\sum_{i=1}^{k} W_i - k\right] - q\left[\sum_{i=k+1}^{n+1} W_i - (n+1-k)\right]}{\bar{W}}+ \\ &+\frac{(1-q)k-q(n+1-k)}{\bar{W}}\Bigg),\end{aligned}$$

Although since $k = \lceil nq \rceil$, and nq are asymptotically equal, we get that

$$\frac{(1-q)k-q(n+1-k)}{\sqrt{n+1}} = \frac{k-nq-q}{\sqrt{n+1}} = \frac{\lceil nq \rceil - nq - q}{\sqrt{n+1}} \to \frac{q}{\sqrt{n+1}} \to 0.$$

So our expression simplifies:

$$= \frac{1}{\bar{W}}\left(\frac{(1-q)\left[\sum_{i=1}^{k} W_i - k\right] - q\left[\sum_{i=k+1}^{n+1} W_i - (n+1-k)\right]}{\sqrt{n+1}}\right).$$

Now, separating the terms in the numerator, we may note that asymptotically,

$$\frac{1}{\sqrt{n+1}} \to \frac{1}{\sqrt{n}} \to \frac{1}{\sqrt{\frac{\lceil nq \rceil}{q}}} \text{ and } \frac{1}{\sqrt{n+1}} \to \frac{1}{\sqrt{n}} = \frac{1}{\sqrt{\frac{n(1-q)}{1-q}}} \to \frac{1}{\sqrt{\frac{n-\lceil nq \rceil}{1-q}}} \to \frac{1}{\sqrt{\frac{n+1-\lceil nq \rceil}{1-q}}}$$

This means that we may substitute the $\sqrt{n+1}$ term for the quantities listed above:

$$= \frac{1}{\bar{W}}\left(\sqrt{q}(1-q)\frac{\sum_{i=1}^{k} W_i - k}{\sqrt{k}} - q\sqrt{1-q}\frac{\sum_{i=k+1}^{n+1} W_i - (n+1-k)}{\sqrt{n+1-k}}\right).$$

Now, by the central limit theorem, we have that these two sums converge in distribution to normals. Further, since $\frac{1}{W}$ converges in probability to some constant by the LLN, we can use slutsky's theorems to say that

$$\begin{aligned}
&\xrightarrow{d} \frac{1}{E[W]}\left(N\Big(0, q(1-q)^2\Big) - N\Big(0, q^2(1-q)\Big)\right), \\
&= \frac{1}{E[W]} N\Big(0, q(1-q)(1-q+q)\Big), \\
&= N(0, q(1-q)).
\end{aligned}$$

The second equality we obtained by adding the variances and noting that there was no covariance between these two normals as they are independent. The third, we got through simplification and noting that $E[W] = 1$ since $W \sim Exp(1)$. Returning to where we started, we have shown that

$$\sqrt{n}(U_{(\lceil nq \rceil)} - q) \xrightarrow{d} N(0, q(1-q)).$$

□

6.4. **Proof: Lemma 2.2.4:**

Proof.
Observe that by assumption, $X_n \to \theta$. By the mean value theorem, there is some $\tilde{\theta} \in (\theta, X_n)$ so that

$$\begin{aligned}
g'(\tilde{\theta}) &= \frac{g(X_n) - g(\theta)}{(X_n - \theta)} \\
g'(\tilde{\theta})(X_n - \theta) &= g(X_n) - g(\theta).
\end{aligned}$$

Now $\sqrt{n}(X_n - \theta) \xrightarrow{d} N(0, \sigma^2)$, so

$$\begin{aligned}
\sqrt{n}(g(X_n) - g(\theta)) &= g'(\tilde{\theta})\sqrt{n}(X_n - \theta), && \text{(swapping sides of equality)} \\
\sqrt{n}(g(X_n) - g(\theta)) &\xrightarrow{d} g'(\tilde{\theta}) N(0, \sigma^2), \\
\sqrt{n}(g(X_n) - g(\theta)) &\xrightarrow{d} N(0, \sigma^2 [g'(\tilde{\theta})]^2)
\end{aligned}$$

The proof is completed by noting that $\tilde{\theta}$ is squeezed to be equal to θ under the limit, and so $g'(\tilde{\theta}) \to g'(\theta)$.

□

This method of finding the asymptotic distribution is called the delta method[8]. We technically only proved the univariate case. For the multivariate case, we proceed similarly, instead of using the mean value theorem, taking the first order Taylor expansion about θ, and then considering the variance of the expansion. Ultimately, we get the a similar result, replacing the variance term in the normal with

$$[\nabla g(\theta)]^T \Sigma [\nabla g(\theta)]$$

It is worth noting, that we could technically do the same thing here, however we did not need the extra machinery so we did not explicitly use it. The first expression from the mean value theorem though could be discovered equally from the Taylor series expansion:

$$g(X_n) \approx g(\theta) + (X_n - \theta)g'(\theta).$$

This is true only in approximation, but since we are taking the limit, any higher order terms vanish, giving us the same result as the mean value theorem.

There are also ways to deal with functions with possibly zero derivatives, although since we are working with cdfs we don't worry about this.

7. Appendix 2: R code

```
```{r Load Packages, message = F}
# Load necessary libraries
library(ggplot2)
library(magrittr)
library(tidyverse)
library(dplyr)
```

```{r example of order statistics, warning = F, message = F}
set.seed(19)  #cherrypick a seed with nice samples
n = 10        #10 draws
X = runif(n)  #draw from the standard uniform
samp = as.data.frame(X) %>%
  mutate(N = 1:n) %>% #add in indicators of the order of draw,
  arrange(X) %>%      #order the statistics
  mutate(k = 1:n)     #and then values for ordering

## -- Plot: 10 observations with draw order and order --
samp %>%
  ggplot(aes(X)) +
  geom_dotplot(binwidth = 0.001, dotsize = 10) +
  theme_bw() +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        axis.ticks.y = element_blank(),
        axis.text.y = element_blank()) +
  ylim(-0.25, 0.25) +
  geom_hline(yintercept=0.0045, linewidth=0.5, alpha=0.5) +
  geom_vline(xintercept = c(0,1), alpha = 0.5) +
  labs(x = "", y = "", title = "")

samp    # Values with sample order and order statistic order
```

```{r beta distribution figure}
## Create a few beta densities which only depend on x
beta51 = function(x) dbeta(x,5,1)
beta52 = function(x) dbeta(x,5,2)
beta31 = function(x) dbeta(x,3,1)
beta32 = function(x) dbeta(x,3,2)
beta13 = function(x) dbeta(x,1,3)
beta15 = function(x) dbeta(x,1,5)
beta25 = function(x) dbeta(x,2,5)

## -- Plot: Beta Densities for different alpha and beta --
data.frame(x = seq(0,1,0.001)) %>%
  ggplot(aes(x)) +
  theme_bw() +
```
```

````
  xlim(0,1) +
  stat_function(aes(color = "Beta(5,1)"), fun = beta51) +
  stat_function(aes(color = "Beta(5,2)"), fun = beta52) +
  stat_function(aes(color = "Beta(3,1)"), fun = beta31) +
  stat_function(aes(color = "Beta(3,2)"), fun = beta32) +
  theme(legend.position = "inside",
        legend.position.inside = c(0.125, y = 0.775),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank()) +
  labs(y = "Density", title = "", color = "") +
  geom_hline(yintercept = c(0,5), alpha = 0.25) +
  geom_vline(xintercept = c(0,1), alpha = 0.25)
```

```{r min and max figure}
## --- simulation parameters ----
n         <- 10       # sample size
num_sim   <- 1e4      # number of Monte-Carlo replications
set.seed(123)

## --- simulate min and max order statistics ----
sims      <- replicate(num_sim, runif(n)) # n-by-num_sim matrix
min_stats <- apply(sims, 2, min)          # X_(1)
max_stats <- apply(sims, 2, max)          # X_(n)

## --- tidy data for ggplot ----
df <- data.frame(
  value      = c(min_stats, max_stats),
  order_type = factor(rep(c("Min  (X[1])",
                            "Max  (X[n])"), each = num_sim),
                      levels = c("Min  (X[1])",
                                 "Max  (X[n])"))
)

## --- theoretical Beta densities on a fine grid ----
grid <- seq(0, 1, length.out = 1000)
beta_df <- data.frame(
  x          = c(grid, grid),
  density    = c(dbeta(grid, 1, n),      #Beta(1,n) for minimum
                 dbeta(grid, n, 1)),     #Beta(n,1) for maximum
  order_type = factor(rep(c("Min  (X[1])", "Max  (X[n])"),
                          each = length(grid)),
                      levels = c("Min  (X[1])", "Max  (X[n])"))
)

## -- histograms + theoretical curves, faceted side-by-side --
ggplot(df, aes(x = value)) +
  geom_histogram(aes(y = ..density..), bins = 30,
                 fill="skyblue", colour="black", alpha=0.7) +
  geom_line(data = beta_df, aes(x = x, y = density),
````

````
               colour = "red", size = 1.2) +
  facet_wrap(~order_type, ncol = 2) +
  labs(title = "",
       x = "Order-statistic value",
       y = "Density") +
  theme_bw() +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank())
```

```{r joint order statistics}
# Load necessary libraries
library(ggplot2)
library(dplyr)

# Set simulation parameters
n <- 10           # sample size
k <- 4            # first order statistic index
j <- 6            # second order statistic index
num_sim <- 20000 # number of simulations

# Function to pull the k and j-th order statistics from samples
get_two_order_stats <- function(sample, k, j) {
  s <- sort(sample)
  c(s[k], s[j])
}

# Simulate num_sim samples and extract the two order statistics
set.seed(123)  # for reproducibility
two_stats <- replicate(num_sim,
                       get_two_order_stats(runif(n), k, j))
two_stats <- as.data.frame(t(two_stats))
colnames(two_stats) <- c("Xk", "Xj")

# Theoretical joint density for Uniform order statistics:
joint_density <- function(x, y, n, k, j) {
  if (x > y || x < 0 || y > 1) return(0)
  coef <- factorial(n) *
          1/(factorial(k-1)*factorial(j-k-1)*factorial(n-j))
  coef * (x^(k-1)) * ((y - x)^(j - k - 1)) * ((1 - y)^(n - j))
}

# Create a grid of (x,y) values to evaluate the density
grid_points <- expand.grid(x = seq(0, 1, length.out = 100),
                           y = seq(0, 1, length.out = 100)) %>%
  filter(x <= y)  # Only valid region: x <= y

# Compute theoretical density on the grid
grid_points$density <- mapply(
    joint_density,
````

````
    grid_points$x,
    grid_points$y,
    MoreArgs = list(n = n, k = k, j = j)
)

# Plot the simulated draws and overlay density contours
ggplot() +
  # Scatter plot of simulated values
  geom_point(data = two_stats, aes(x = Xk, y = Xj),
               alpha = 0.1, color = "skyblue", size = 1) +
  # Overlay contour lines for the theoretical joint density
  geom_contour(data = grid_points, aes(x = x, y = y, z = density),
                 color = "red", bins = 8, linewidth = 0.5, size = 1) +
  # Overlay contour lines for the simulated joint density
  geom_density2d(data = two_stats, aes(x = Xk, y = Xj),
                   color = "blue", bins = 8) +
  # Adjust presentation of the plot
  labs(title = "",
       x = paste(k, "th␣Order␣Statistic"),
       y = paste(j, "th␣Order␣Statistic")) +
    theme_bw() +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank()) +
  geom_hline(yintercept=c(0,1), color="grey60", alpha=0.5) +
  geom_vline(xintercept=c(0,1), color="grey60", alpha=0.5) +
  geom_abline(slope = 1, intercept = 0, alpha = 0.5)
```

```{r Empirical vs Theoreitcal Densities}
set.seed(243)   # For reproducability
n = 101
I = 25000
## Initialize a dataframe so we can store Median observations
median_data = data.frame(x = c(0))

for(i in 1:I) {
  X = rnorm(n)              # Simulate n standard normal draws
  median_data[i,] = median(X) # Store the median
}

## Make standard normal density function for the median
density = function(x) dnorm(x, 0, sqrt(pi/(2*n)))

## -- Plot: Theoretical and sampled densities of the median --
median_data %>%
  ggplot(aes(x)) +
  xlim(-0.5, 0.5) +
  geom_density(color = "red") +
  stat_function(fun = density, color = "blue") +
  labs(y = "Density", title = "") +
````

```
  theme_bw() +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank()) +
  geom_hline(yintercept = 0, color = "grey60", alpha = 0.5) +
  geom_vline(xintercept = c(-0.5, 0.5),
             color = "grey60", alpha = 0.5)
```

```{r Confidence Intervals}
## Compute variance statistics. This will include pi/2 constant
s2hat = median_data %$% var(x) # Compute Sample Variance
shat = sqrt(s2hat)            # Standard error
tinv = qt(0.975, n-1)         # Quantile of the t

## Use these quantities to create Confidence Intervals
median_data %<>%
  mutate(CIl = x - tinv * shat,
         CIu = x + tinv * shat)

median_data %<>%
  mutate(capture = case_when(
    CIl < 0 & CIu > 0 ~ T,
    .default = F
    )) #mutate in a variable to list if we capture mean

## --- Validating Confidence Intervals ---
# Raw proportion of Confidence Intervals which captured mean
median_data %$% capture %!>% sum()/I

# -- Plot: First 500 Confidence Intervals plotted --
median_data %>%
  # Grab the first 500 plots
  mutate(i = 1:25000) %>%
  filter(i <= 500) %>%
  # Now plot the Confidence intervals
  ggplot(aes(x,i, color = capture)) +
  geom_point(size = 0.5) +
  geom_errorbar(aes(xmin = CIl, xmax = CIu, color = capture),
                linewidth = 0.5, alpha = 0.5) +
  # Style the plot
  theme_bw() +
  theme(axis.text.y = element_blank(),
        axis.ticks.y = element_blank(),
        axis.title.y = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        legend.position = "none") +
  labs(x = "x       ", y = "",
       color = "Captured true mean?",
       title =  "") +
```

```
  geom_vline(xintercept = 0, alpha = 0.5) +
  geom_hline(yintercept = c(0,501),
             color = "grey30", alpha = 0.5)
```
```
```