Course Notes for Math 162: Mathematical Statistics
The Sample Distribution of the Median

Adam Merberg and Steven J. Miller

February 15, 2008

Abstract

We begin by introducing the concept of order statistics and finding the density of the r^{th} order statistic of a sample. We then consider the special case of the density of the median and provide some examples. We conclude with some appendices that describe some of the techniques and background used.

Contents

1 Order Statistics

Suppose that the random variables $X_1, X_2, \dots, X_n$ constitute a sample of size n from an infinite population with continuous density. Often it will be useful to reorder these random variables from smallest to largest. In reordering the variables, we will also rename them so that Y_1 is a random variable whose value is the smallest of the X_i, Y_2 is the next smallest, and so on, with Y_n the largest of the X_i. Y_r is called the r^{th} **order statistic** of the sample.

In considering order statistics, it is naturally convenient to know their probability density. We derive an expression for the distribution of the r^{th} order statistic as in [MM].

Theorem 1.1. *For a random sample of size n from an infinite population having values x and density $f(x)$, the probability density of the r^{th} order statistic Y_r is given by*

$$g_r(y_r) \;=\; \frac{n!}{(r-1)!(n-r)!}\left[\int_{-\infty}^{y_r} f(x)\,dx\right]^{r-1} f(y_r)\left[\int_{y_r}^{\infty} f(x)\,dx\right]^{n-r}. \tag{1.1}$$

Proof. Let h be a positive real number. We divide the real line into three intervals: $(-\infty, y_r)$, $[y_r, y_r+h]$, and (y_r+h, ∞). We will first find the probability that Y_r falls in the middle of these three intervals, and no other value from the sample falls in this interval. In order for this to be the case, we must have $r-1$ values falling in the first interval, one value falling in the second, and $n-r$ falling in the last interval. Using the multinomial distribution, which is explained in Appendix A, the probability of this event is

$$\begin{aligned}&\text{Prob}(Y_r \in [y_r, y_r+h] \text{ and } Y_i \neq [y_r, y_r+h] \text{ if } i \neq r)\\ &=\; \frac{n!}{(r-1)!1!(n-r)!}\left[\int_{-\infty}^{y_r} f(x)\,dx\right]^{r-1}\left[\int_{y_r}^{y_r+h} f(x)\,dx\right]^{1}\left[\int_{y_r+h}^{\infty} f(x)\,dx\right]^{n-r}.\end{aligned} \tag{1.2}$$

We need also consider the case of two or more of the Y_i lying in $[y_r, y_r + h]$. As this interval has length h, this probability is $O(h^2)$ (see Appendix B for a review of big-Oh notation such as $O(h^2)$). Thus we may remove the constraint that exactly one $Y_i \in [y_r, y_r + h]$ in (1.2) at a cost of at most $O(h^2)$, which yields

$$\text{Prob}(Y_r \in [y_r, y_r + h]) \;=\; \frac{n!}{(r-1)!1!(n-r)!}\left[\int_{-\infty}^{y_r} f(x)\,dx\right]^{r-1}\left[\int_{y_r}^{y_r+h} f(x)\,dx\right]^{1}\left[\int_{y_r+h}^{\infty} f(x)\,dx\right]^{n-r} + O(h^2). \quad (1.3)$$

We now apply the Mean Value Theorem[1] to find that for some c_{h,y_r} with $y_r \le c_{h,y_r} \le y_r + h$, we have

$$\int_{y_r}^{y_r+h} f(x)\,dx \;=\; h \cdot f(c_{h,y_r}). \quad (1.6)$$

We denote the point provided by the mean value theorem by c_{h,y_r} in order to emphasize its dependence on h and y_r.

We can substitute this result into the expression of (1.3). We divide the result by h (the length of the middle interval $[y_r, y_r + h]$), and consider the limit as $h \to 0$:

$$\begin{aligned}
\lim_{h\to 0}\frac{\text{Prob}(Y_r \in [y_r, y_r + h])}{h} &= \lim_{h\to 0}\frac{\frac{n!}{(r-1)!1!(n-r)!}\left[\int_{-\infty}^{y_r} f(x)\,dx\right]^{r-1}\left[\int_{y_r+h}^{\infty} f(x)\,dx\right]^{1}\left[\int_{y_r+h}^{\infty} f(x)\,dx\right]^{n-r} + O(h^2)}{h}\\
&= \lim_{h\to 0}\frac{\frac{n!}{(r-1)!1!(n-r)!}\left[\int_{-\infty}^{y_r} f(x)\,dx\right]^{r-1} h\cdot f(c_{h,y_r})\left[\int_{y_r+h}^{\infty} f(x)\,dx\right]^{n-r}}{h}\\
&= \lim_{h\to 0}\frac{n!}{(r-1)!1!(n-r)!}\left[\int_{-\infty}^{y_r} f(x)\,dx\right]^{r-1} f(c_{h,y_r})\left[\int_{y_r+h}^{\infty} f(x)\,dx\right]^{n-r}\\
&= \frac{n!}{(r-1)!1!(n-r)!}\left[\int_{-\infty}^{y_r} f(x)\,dx\right]^{r-1} f(y_r)\left[\int_{y_r}^{\infty} f(x)\,dx\right]^{n-r}. \quad (1.7)
\end{aligned}$$

Thus the proof is reduced to showing that the left hand side above is $g_r(y_r)$. Let $g_r(y_r)$ be the probability density of Y_r. Let $G_r(y_r)$ be the cumulative distribution function of Y_r. Thus

$$\text{Prob}(Y_r \le y) \;=\; \int_{-\infty}^{y} g_r(y_r)dy_r \;=\; G_r(y), \quad (1.8)$$

and $G_r'(y) = g_r(y)$. Thus the left hand side of (1.7) equals

$$\lim_{h\to 0}\frac{\text{Prob}(Y_r \in [y_r, y_r + h])}{h} \;=\; \lim_{h\to 0}\frac{G_r(y_r + h) - G_r(y_r)}{h} \;=\; g_r(y_r), \quad (1.9)$$

where the last equality follows from the definition of the derivative. This completes the proof. □

Remark 1.2. *The technique employed in this proof is a common method for calculating probability densities. We first calculate the probability that a random variable Y lies in an infinitesimal interval $[y, y + h]$. This probability is $G(y + h) - G(y)$, where g is the density of Y and G is the cumulative distribution function (so $G' = g$). The definition of the derivative yields*

$$\lim_{h\to 0}\frac{\text{Prob}(Y \in [y, y + h])}{h} \;=\; \lim_{h\to 0}\frac{G(y + h) - G(y)}{h} \;=\; g(y). \quad (1.10)$$

2 The Sample Distribution of the Median

In addition to the smallest (Y_1) and largest (Y_n) order statistics, we are often interested in the **sample median**, $\tilde{X}$. For a sample of odd size, $n = 2m + 1$, the sample median is defined as Y_{m+1}. If $n = 2m$ is even, the sample median is defined as $\frac{1}{2}(Y_m + Y_{m+1})$. We will prove a relation between the sample median and the **population median** $\tilde{\mu}$. By definition, $\tilde{\mu}$ satisfies

$$\int_{-\infty}^{\tilde{\mu}} f(x)\,dx \;=\; \frac{1}{2}. \quad (2.11)$$

[1] If F is an anti-derivative of f, then the Mean Value Theorem applied to F,

$$\frac{F(b) - F(a)}{b - a} \;=\; F'(c) \quad (1.4)$$

is equivalent to

$$\int_a^b f(x)dx \;=\; (b - a)\cdot f(c). \quad (1.5)$$

It is convenient to re-write the above in terms of the cumulative distribution function. If F is the cumulative distribution function of f, then $F' = f$ and (2.11) becomes

$$F(\tilde{\mu}) \;=\; \frac{1}{2}. \tag{2.12}$$

We are now ready to consider the distribution of the sample median.

Median Theorem. *Let a sample of size $n = 2m+1$ with n large be taken from an infinite population with a density function $f(\tilde{x})$ that is nonzero at the population median $\tilde{\mu}$ and continuously differentiable in a neighborhood of $\tilde{\mu}$. The sampling distribution of the median is approximately normal with mean $\tilde{\mu}$ and variance $\frac{1}{8f(\tilde{\mu})^2 m}$.*

Proof. Let the median random variable $\tilde{X}$ have values $\tilde{x}$ and density $g(\tilde{x})$. The median is simply the $(m+1)^{\text{th}}$ order statistic, so its distribution is given by the result of the previous section. By Theorem 1.1,

$$g(\tilde{x}) \;=\; \frac{(2m+1)!}{m!m!}\left[\int_{-\infty}^{\tilde{x}} f(\tilde{x})\,dx\right]^m f(\tilde{x})\left[\int_{\tilde{x}}^{\infty} f(x)\,dx\right]^m. \tag{2.13}$$

We will first find an approximation for the constant factor in this equation. For this, we will use Stirling's approximation, which tells us that $n! = n^n e^{-n}\sqrt{2\pi n}(1+O(n^{-1}))$; we sketch a proof in Appendix D. We will consider values sufficiently large so that the terms of order $1/n$ need not be considered. Hence

$$\frac{(2m+1)!}{m!m!} \;=\; \frac{(2m+1)(2m)!}{(m!)^2} \;\approx\; \frac{(2m+1)(2m)^{2m}e^{-2m}\sqrt{2\pi(2m)}}{(m^m e^{-m}\sqrt{2\pi m})^2} \;=\; \frac{(2m+1)4^m}{\sqrt{\pi m}}. \tag{2.14}$$

As F is the cumulative distribution function, $F(\tilde{x}) = \int_{-\infty}^{\tilde{x}} f(x)\,dx$, which implies

$$g(\tilde{x}) \;\approx\; \frac{(2m+1)4^m}{\sqrt{\pi m}}\left[F(\tilde{x})\right]^m f(\tilde{x})\left[1-F(\tilde{x})\right]^m. \tag{2.15}$$

We will need the Taylor series expansion of $F(\tilde{x})$ about $\tilde{\mu}$, which is just

$$F(\tilde{x}) \;=\; F(\tilde{\mu}) + F'(\tilde{\mu})(\tilde{x}-\tilde{\mu}) + O((\tilde{x}-\tilde{\mu})^2). \tag{2.16}$$

Because $\tilde{\mu}$ is the population median, $F(\tilde{\mu}) = 1/2$. Further, since F is the cumulative distribution function, $F' = f$ and we find

$$F(\tilde{x}) \;=\; \frac{1}{2} + f(\tilde{\mu})(\tilde{x}-\tilde{\mu}) + O((\tilde{x}-\tilde{\mu})^2). \tag{2.17}$$

This approximation is only useful if $\tilde{x}-\tilde{\mu}$ is small; in other words, we need $\lim_{m\to\infty}|\tilde{x}-\tilde{\mu}| = 0$. Fortunately this is easy to show, and a proof is included in Appendix C.

Letting $t = \tilde{x}-\tilde{\mu}$ (which is small and tends to 0 as $m \to \infty$), substituting our Taylor series expansion into (2.15) yields[2]

$$g(\tilde{x}) \;\approx\; \frac{(2m+1)4^m}{\sqrt{\pi m}}\left[\frac{1}{2} + f(\tilde{\mu})t + O(t^2)\right]^m f(\tilde{x})\left[1-\left(\frac{1}{2} + f(\tilde{\mu})t + O(t^2)\right)\right]^m. \tag{2.18}$$

By rearranging and combining factors, we find that

$$g(\tilde{x}) \;\approx\; \frac{(2m+1)4^m}{\sqrt{\pi m}} f(\tilde{x})\left[\frac{1}{4} - (f(\tilde{\mu})t)^2 + O(t^3)\right]^m \;=\; \frac{(2m+1)f(\tilde{x})}{\sqrt{\pi m}}\left[1 - \frac{4m(f(\tilde{\mu})t)^2}{m} + O(t^3)\right]^m. \tag{2.19}$$

Remember that one definition of e^x is

$$e^x \;=\; \exp(x) \;=\; \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n; \tag{2.20}$$

see Appendix E for a review of properties of the exponential function. Using this, and ignoring higher powers of t for the moment, we have for large m that

$$g(\tilde{x}) \;\approx\; \frac{(2m+1)f(\tilde{x})}{\sqrt{\pi m}}\exp\left(-4mf(\tilde{\mu})^2t^2\right) \;\approx\; \frac{(2m+1)f(\tilde{x})}{\sqrt{\pi m}}\exp\left(-\frac{(\tilde{x}-\tilde{\mu})^2}{1/(4mf(\tilde{\mu})^2)}\right). \tag{2.21}$$

[2]Actually, the argument below is completely wrong! The problem is each term has an error of size $O(t^2)$. Thus when we multiply them together there is also an error of size $O(t^2)$, and this is the same order of magnitude as the secondary term, $(f(\tilde{\mu})t)^2$. The remedy is to be more careful in expanding $F(\tilde{x})$ and $1-F(\tilde{x})$. A careful analysis shows that their t^2 terms are equal in magnitude but opposite in sign. Thus they will cancel in the calculations below. In summary, we really need to use $F(\tilde{x}) = \frac{1}{2} + f(\tilde{m}u)(\tilde{x}-\tilde{m}u) + \frac{f'(\tilde{m}u)}{2}(\tilde{x}-\tilde{\mu})^2$ (and similarly for $1-F(\tilde{x})$).

Since, as shown in Appendix C, $\tilde{x}$ can be assumed arbitrarily close to $\tilde{\mu}$ with high probability, we can assume $f(\tilde{x}) \approx f(\tilde{\mu})$ so that[3]

$$g(\tilde{x}) \;\approx\; \frac{(2m+1)f(\tilde{\mu})}{\sqrt{\pi m}} \;\exp\left(-\frac{(\tilde{x}-\tilde{\mu})^2}{1/(4mf(\tilde{\mu})^2)}\right). \tag{2.23}$$

Looking at the exponential part of the expression for $g(\tilde{x})$, we see that it appears to be a normal density with mean $\tilde{\mu}$ and $\sigma^2 = 1/(8mf(\tilde{\mu})^2)$. If we were instead to compute the variance from the normalization constant, we would find the variance to be

$$\frac{m}{2(2m+1)^2 f(\tilde{\mu})^2}$$

We see that the two values are asymptotically equivalent, thus we can take the variance to be $\sigma^2 = 1/(8mf(\tilde{\mu})^2)$. Thus to complete the proof of the theorem, all that we need to is prove that we may ignore the higher powers of t and replace the product with an exponential in passing from (2.19) to (2.21). We have

$$\left(1 - \frac{4m(f(\tilde{\mu})t)^2}{m} + O(t^3)\right)^m \;=\; \exp\left(m \log\left(1 - 4(f(\tilde{\mu})t)^2 + O(t^3)\right)\right). \tag{2.24}$$

We use the Taylor series expansion of $\log(1-x)$:

$$\log(1-x) \;=\; -x + O(x^2); \tag{2.25}$$

we only need one term in the expansion as t is small. Thus (2.24) becomes

$$\begin{aligned}\left(1 - \frac{4m(f(\tilde{\mu})t)^2}{m} + O(t^3)\right)^m &= \exp\left(-m \cdot 4(f(\tilde{\mu})t)^2 + O(mt^3)\right) \\ &= \exp\left(-\frac{(\tilde{x}-\tilde{\mu})^2}{1/(4mf(\tilde{\mu})^2)}\right) \cdot \exp(O(mt^3)).\end{aligned} \tag{2.26}$$

Using the methods of Appendix C one can show that as $m \to \infty$, $mt^3 \to 0$. Thus the $\exp(O(mt^3))$ term above tends to 1, which completes the proof. □

Remark 2.1. *Our justification of ignoring the higher powers of t and replacing the product with an exponential in passing from (2.19) to (2.21) is a standard technique. Namely, we replace some quantity $(1-P)^m$ with $(1-P)^m = \exp(m\log(1-P))$, Taylor expand the logarithm, and then look at the limit as $m \to \infty$.*

3 Examples and Exercises

Example 3.1. Consider the case of a normal population. The normal density is symmetric about the mean $\tilde{\mu}$, hence $\tilde{\mu} = \mu$. Furthermore, we have

$$\begin{aligned} f(\tilde{\mu}) &= f(\mu) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\mu-\mu)^2}{2\sigma^2}\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}, \end{aligned} \tag{3.27}$$

which implies that

$$\frac{1}{8mf(\tilde{\mu})^2} \;=\; \frac{\pi\sigma^2}{4m} \tag{3.28}$$

For large n, we therefore see that the distribution of the median (from a normal distribution with mean μ and variance σ^2) will be approximately normal with mean μ and variance $\pi\sigma^2/4m$.

Exercise 3.2. *Find the distribution of the median of a large sample from an exponential population with parameter θ.*

[3] To prove that there is negligible error in replacing $f(\tilde{x})$ with $f(\tilde{\mu})$, we use the mean value theorem and find

$$f(\tilde{x}) - f(\tilde{\mu}) \;=\; f'(c_{\tilde{x},\tilde{\mu}}) \cdot (\tilde{x} - \tilde{\mu}); \tag{2.22}$$

here we have written the constant as $c_{\tilde{x},\tilde{\mu}}$ to emphasize the fact that we evaluate the first derivative in the interval $[\tilde{x}, \tilde{\mu}]$. As we have assumed f is continuously differentiable and $|\tilde{x} - \tilde{\mu}|$ is small, we may bound $f'(c_{\tilde{x},\tilde{\mu}})$ Thus we may replace $f(\tilde{x})$ with $f(\tilde{\mu})$ at a cost of $O(t)$, where $t = \tilde{x} - \tilde{\mu}$ tends to zero with m.

A The Multinomial Distribution

We can use a binomial distribution to study a situation in which we have multiple trials with two possible outcomes with the probabilities of each respective outcome the same for each trial and all of the trials independent.

A generalization of the binomial distribution is the **multinomial distribution**. Like the binomial distribution, the multinomial distribution considers multiple independent trials with the probabilities of respective outcomes the same for each trial. However, the multinomial distribution gives the probability of different outcomes when we have more than two possible outcomes for each trial. This is useful, for example, in proving the distribution of order statistics, where we take the different trials to be the sample data and the outcomes to be the three intervals in the real line in which these data can fall.

Suppose that we have n trials and k mutually exclusive outcomes with probabilities $\theta_1, \theta_2, \ldots, \theta_k$. We will let $f(x_1, x_2, \ldots, x_k)$ be the probability of having x_i outcomes of each corresponding type, for $1 \le i \le k$. Obviously, we must have $x_1 + x_2 + \cdots + x_k = n$. To compute $f(x_1, x_2, \ldots, x_k)$, we first note that the probability of getting these numbers of outcomes in some particular order is $\theta_1^{x_1}\theta_2^{x_2}\cdots\theta_k^{x_k}$. We now compute the number of orders in which our combination of numbers of outcomes is attainable. The x_1 outcomes of the first type can be chosen in $\binom{n}{x_1}$ ways, the x_2 outcomes of the second type can be chosen in $\binom{n-x_1}{x_2}$ ways, and so on up to the x_k outcomes of type k which can be chosen in $\binom{n-x_1-x_2-\cdots-x_{k-1}}{x_k}$ ways. The total number of orderings is therefore

$$\binom{n}{x_1}\binom{n-x_1}{x_2}\cdots\binom{n-x_1-\cdots-x_{k-1}}{x_k} = \frac{n!}{(n-x_1)!x_1!}\cdot\frac{(n-x_1)!}{(n-x_1-x_2)!x_2!}\cdots\frac{(n-x_1-...-x_{k-1})!}{(n-x_1-...-x_k)!x_k!}. \tag{A.29}$$

The product telescopes and we are left with

$$\frac{n!}{x_1!x_2!\cdots x_k!}. \tag{A.30}$$

The expression (A.30) is called a **multinomial coefficient** and is often denoted

$$\binom{n}{x_1, x_2, \ldots, x_k}. \tag{A.31}$$

Using the multinomial coefficient, we can see that

$$f(x_1, x_2, ..., x_n) = \frac{n!}{x_1!x_2!\cdots x_k!}\theta_1^{x_1}\theta_2^{x_2}\cdots\theta_k^{x_k}. \tag{A.32}$$

This is the multinomial distribution. We often write $f(x_1, x_2, ..., x_n; \theta_1, \theta_2, \ldots, \theta_k)$ to emphasize the dependence on the parameters.

Remark A.1. *One can derive the multinomial distribution by repeated uses of the binomial theorem. For example, if $k = 3$ there are three outcomes, say A, B and C. We may amalgamate B and C and consider the case of two outcomes: A and not A. If we let θ_1 equal the probability of A and $1-\theta_1$ the probability of not A, we find the probability of x_1 outcomes being A and $n-x_1$ outcomes being not A is just*

$$\binom{n}{x_1}\theta_1^{x_1}(1-\theta_1)^{n-x_1}. \tag{A.33}$$

Let θ_2 be the probability of outcome B, and θ_3 the probability of outcome C. Given A does not occur, the probability that B occurs is $\frac{\theta_2}{\theta_2+\theta_3}$; the probability that C occurs is $\frac{\theta_3}{\theta_2+\theta_3}$. Thus the probability that x_1 outcomes are A, x_2 are B and $x_3 = n - x_1 - x_2$ are C is

$$\binom{n}{x_1}\theta_1^{x_1}\left[\binom{n-x_1}{x_2}\left(\frac{\theta_2}{\theta_2+\theta_3}\right)^{x_2}\left(\frac{\theta_3}{\theta_2+\theta_3}\right)^{n_1-x_1-x_2}\right](1-\theta_1)^{n-x_1}; \tag{A.34}$$

however, as $1-\theta_1 = \theta_2+\theta_3$ and $\binom{n}{x_1}\binom{n-x_1}{x_2} = \frac{n!}{x_1!x_2!x_3!}$, the above simplifies to

$$\frac{n!}{x_1!x_2!x_3!}\theta_1^{x_1}\theta_2^{x_2}\theta_3^{n_1-x_1-x_2}, \tag{A.35}$$

which agrees with what we found above.

B Big-Oh Notation

Definition B.1 (Big-Oh Notation)**.** *$A(x) = O(B(x))$, read "$A(x)$ is of order (or big-Oh) $B(x)$", means there is a $C > 0$ and an x_0 such that for all $x \geq x_0$, $|A(x)| \leq C\, B(x)$. This is also written $A(x) \ll B(x)$ or $B(x) \gg A(x)$.*

Big-Oh notation is a convenient way to handle lower order terms. For example, if we write $F(x) = x^5 + O(x^2)$, this means that as x tends to infinity, the main term of $F(x)$ grows like x^5, and the correction (or error) terms are at most some constant times x^2.

Exercise B.2. *Prove for any $r, \epsilon > 0$, as $x \to \infty$ we have $x^r = O(e^x)$ and $\log x = O(x^\epsilon)$. Let $F(x) = x^2/2$ and $G(x) = \sum_{n \leq x} n$. Prove $F(x) \sim G(x)$.*

C Proof That With High Probability $|\tilde{X} - \tilde{\mu}|$ is Small

In proving the Median Theorem, we assume that we can ignore higher powers of $t = \tilde{X} - \tilde{\mu}$. We are able to do this because, with high probability, t is small. Here we provide a more formal statement of this fact, as well as a proof.

Lemma C.1. *Suppose $f(x)$ is a continuously differentiable function in some neighborhood of $\tilde{\mu}$. Then for any $c > 0$, we have*

$$\lim_{m \to 0} \text{Prob}(|\tilde{X} - \tilde{\mu}| \geq \ c) \ = \ 0. \tag{C.36}$$

Proof. This is equivalent to proving that

$$\lim_{m \to 0} \text{Prob}(\tilde{X} \ \leq \tilde{\mu} - c) \ = \ 0 \quad \text{and} \quad \lim_{m \to 0} \text{Prob}(\tilde{X} \ \geq \tilde{\mu} + c) \ = \ 0. \tag{C.37}$$

We will prove only the first of these two statements as the proof of the second is very similar.

By (2.15), we can approximate the density of the median as

$$g(\tilde{x}) \ \approx \ \frac{(2m+1)4^m f(\tilde{x})}{\sqrt{\pi m}} \left([F(\tilde{x})]\,[1 - F(\tilde{x})]\right)^m. \tag{C.38}$$

We consider the factor $([F(\tilde{x})]\,[1 - F(\tilde{x})])^m$. It is convenient to write $\theta = F(\tilde{x})$ and consider the function $h(\theta) = \theta(1-\theta)$. This function will attain its maximum for the same value of $\theta = F(\tilde{x})$ as $([F(\tilde{x})]\,[1 - F(\tilde{x})])^m$, and it is a simple exercise in calculus to show that this value is $\theta = \frac{1}{2}$. This condition holds only for $\tilde{x} = \tilde{\mu}$. We furthermore note that for $\theta < \frac{1}{2}$, $h'(\theta) = 1 - 2\theta > 0$, so h is increasing. Since $F(\tilde{x}) = \frac{1}{2}$ precisely when $\tilde{x} = \tilde{\mu}$, this means that for $\tilde{x} \leq \tilde{\mu} - c$, the maximum value of $g(\theta)$ occurs for $\tilde{x} = \tilde{\mu} - c$. We therefore have for $\tilde{x} \leq \tilde{\mu} - c$,

$$(F(\tilde{x})\,[1 - F(\tilde{x})])^m \ \leq \ (F(\tilde{\mu} - c)\,[1 - F(\tilde{\mu} - c)])^m \ < (F(\tilde{\mu})\,[1 - F(\tilde{\mu})])^m = \frac{1}{4^m}. \tag{C.39}$$

We choose α so that $\frac{\alpha}{4} = F(\tilde{\mu} - c)\,(1 - F(\tilde{\mu} - c))$. Equation (C.39) then tells us that for $\tilde{x} \leq \tilde{\mu} - c$,

$$(F(\tilde{x})\,[1 - F(\tilde{x})])^m \ \leq \ \left(\frac{\alpha}{4}\right)^m \ \leq \ \frac{1}{4^m}. \tag{C.40}$$

In particular, we note that $\alpha < 1$.

We now begin to look at the probability that $\tilde{X}$ is at most $\tilde{\mu} - c$. We have

$$\begin{aligned} \text{Prob}(\tilde{X} \leq \tilde{\mu} - c) \ &= \ \int_{-\infty}^{\tilde{\mu} - c} g(\tilde{x})\, d\tilde{x} \\ &\approx \ \int_{-\infty}^{\tilde{\mu} - c} \frac{(2m+1)4^m}{\sqrt{\pi m}} f(\tilde{x}) F(\tilde{x})^m (1 - F(\tilde{x}))^m \, d\tilde{x} \\ &< \ \frac{(2m)4^m}{\sqrt{m}} \int_{-\infty}^{\tilde{\mu} - c} f(\tilde{x}) F(\tilde{x})^m (1 - F(\tilde{x}))^m \, d\tilde{x}. \end{aligned} \tag{C.41}$$

In the last step, we use the fact that for m sufficiently large ($m > 1$, in fact), $2m < \frac{2m+1}{\sqrt{\pi}}$. This simplifies the expression as a factor of $2m$ is easier to work with than the factor of $2m + 1$. We now apply our bound on $F(\tilde{x})(1 - F(\tilde{x}))$ to find that

$$\text{Prob}(\tilde{X} \leq \tilde{\mu} - c) \ < \ \frac{(2m)4^m}{\sqrt{m}} \int_{-\infty}^{\tilde{\mu} - c} f(\tilde{x}) \left(\frac{\alpha}{4}\right)^m d\tilde{x} \ = \ \frac{(2m)4^m}{\sqrt{m}} \left(\frac{\alpha}{4}\right)^m \int_{-\infty}^{\tilde{\mu} - c} f(\tilde{x})\, d\tilde{x} \ < \ 2\alpha^m \sqrt{m} \int_{-\infty}^{\tilde{\mu}} f(\tilde{x})\, d\tilde{x} \tag{C.42}$$

In obtaining the rightmost expression, we have used the fact that $f(\tilde{x})$ is nonnegative everywhere and positive in a neighborhood of $\tilde{\mu}$, so that $\int_{-\infty}^{\tilde{\mu}-c} f(\tilde{x})\,d\tilde{x} \;<\; \int_{-\infty}^{\tilde{\mu}} f(\tilde{x})\,d\tilde{x}$. Since $\tilde{\mu}$ is the median of the population, by definition, we have $\int_{-\infty}^{\tilde{\mu}} f(\tilde{x})\,d\tilde{x} = \frac{1}{2}$, so that

$$\mathrm{Prob}(\tilde{X} \leq \tilde{\mu} - c) \;<\; \alpha^m \sqrt{m}. \tag{C.43}$$

Since $\alpha < 1$, it follows that the right side of this inequality must converge to 0 as m goes to infinity, so the probability on the right side must likewise converge to 0. □

Exercise C.2. *Let $\alpha \in (0,1)$. Prove*

$$\lim_{m\to\infty} \alpha^m \sqrt{m} \;=\; 0. \tag{C.44}$$

In fact, this expression tends to zero exponentially fast. Let $\delta = 1 - \alpha$. Show that for m sufficiently large,

$$\alpha^m \sqrt{m} \;\leq\; A\left(1 - \frac{\delta}{2}\right)^m \;=\; Ae^{-Bm}, \tag{C.45}$$

where A and B are constants (with $B = \log\left(1 - \frac{\delta}{2}\right)$.

D Stirling's Approximation Formula for $n!$

Exact computations involving factorials of large numbers can be very difficult. Fortunately, there is an approximation formula which can greatly simplify the computations.

Stirling's Formula.

$$n! \;=\; n^n e^{-n}\sqrt{2\pi n}\left(1 + \left(O\left(\frac{1}{n}\right)\right)\right) \tag{D.46}$$

Proof. For a proof, see [WW]. We show (D.46) is a reasonable approximation. It is often easier to analyze a product by converting it to a sum; this is readily accomplished by taking logarithms. We have

$$\log n! \;=\; \sum_{k=1}^{n} \log k \;\approx\; \int_1^n \log t dt \;=\; (t\log t - t)|_1^n. \tag{D.47}$$

Thus $\log n! \approx n\log n - n$, or $n! \approx n^n e^{-n}$. □

Exercise D.1. *Use the integral test to bound the error in* (D.47)*, and then use that to bound the error in the estimate of* $n!$.

E Review of the exponential function

These notes in this section are taken from [MT-B].

In this section we study some of the basic properties of the number e (see [Rud] for more properties and proofs). One of the many ways to define the number e, the base of the natural logarithm, is to write it as the sum of the following infinite series:

$$e \;=\; \sum_{n=0}^{\infty} \frac{1}{n!}. \tag{E.48}$$

Denote the partial sums of the above series by

$$s_m \;=\; \sum_{n=0}^{m} \frac{1}{n!}. \tag{E.49}$$

Hence e is the limit of the convergent sequence s_m. This representation is one of the main tool in analyzing the nature of e.

Exercise E.1. *Define*

$$e^x \;=\; \sum_{n=0}^{\infty} \frac{x^n}{n!}. \tag{E.50}$$

Prove $e^{x+y} = e^x e^y$. Show this series converges for all $x \in \mathbb{R}$; in fact, it makes sense for $x \in \mathbb{C}$ as well. One can define a^b by $e^{b\ln a}$. Hint: *Use the series expansion of* (E.50) *for e^x, e^y and e^{x+y}, and use the binomial theorem to expand the factors of $(x+y)^n$.*

Exercise E.2. *An alternate definition of* e^x *is*

$$e^x \;=\; \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n. \tag{E.51}$$

Show this definition agrees with the series expansion, and prove $e^{x+y} = e^x e^y$*. This formulation is useful for growth problems such as compound interest or radioactive decay; see for example [BoDi].* Hint: *Show*

$$\left(1+\frac{x}{n}\right)\cdot\left(1+\frac{y}{n}\right) \;=\; \left(1+\frac{x+y}{n}\right)\cdot\left(1+\frac{c(x,y,n)}{n^2}\right); \tag{E.52}$$

for n *large,* $|c(x,y,n)| \le |2xy|$*. For* $n \ge N$ *the* n^{th} *power of the second factor satisfies*

$$\left|1+\frac{c(x,y,n)}{n^2}\right|^n \;\le\; \left(1+\frac{|2xy/N|}{n}\right)^n, \tag{E.53}$$

and the limit as $n\to\infty$ *of the right hand side is* $e^{|2xy/N|}$*. As* N *was an arbitrary large number, this shows that the* n^{th} *power of the second factor in* (E.52) *does not contribute in the limit.*

Exercise E.3. *Prove* $\frac{d}{dx}e^x = e^x$*. As* $e^{\ln x} = x$*, the chain rule implies* $\frac{d}{dx}\ln x = \frac{1}{x}$ *(*$\ln x$ *is the inverse function to* e^x*).*

From the functions e^x and $\ln x$, we can interpret a^b for any $a > 0$ and $b \in \mathbb{R}$: $a^b = e^{b\ln a}$. Note the series expansion for e^x makes sense for all x, thus we have a well defined process to determine numbers such as $3^{\sqrt{2}}$. We cannot compute $3^{\sqrt{2}}$ directly because we do not know what it means to raise 3 to the $\sqrt{2}$-power; we can only raise numbers to *rational* powers.

Exercise E.4. *Split* 100 *into smaller integers such that each integer is two or more and the product of all these integers is as large as possible. Suppose now* N *is a large number and we wish to split* N *into smaller pieces, but all we require is that each piece be positive. How should we break up a large* N*?*
Hint: *For the second part, for each* n *consider* $N = a_1 + \cdots + a_n$*. Maximize the product* $a_1 \cdots a_n$*, and denote this value by* $f(n)$*. Though initially only defined for* n *integral, we may extend* f *to all positive real numbers at least 1; this extension is differentiable, and we can then use calculus to maximize* $f(n)$*. Investigate the sign of the first derivative, and deduce the largest value of* $f(n)$ *for* integral n *is either* $n = 2$ *or* $n = 3$*;* 3 *turns out to be superior to* 2 *because* 3 *is the closest integer to* e *(*2 *is the second closest). An interesting application arises in computer science: using a natural measure of storage cost, the most efficient computers for storing information are those in base* e*; the most efficient with an integral base are those base* 3*; however, the storage efficiency of base* 3 *is not large enough to overcome the enormous advantages of working base* 2*. See [Ha] for more details.*

Exercise E.5. *Without using a calculator or computer, determine which is larger:* e^π *or* π^e*.* Hint: *One approach is to study the function* $x^{1/x}$ *(take the* $e\pi$ *root of both sides to reduce the problem to comparing* $e^{1/e}$ *and* $\pi^{1/\pi}$*. Use calculus to find the maximum value. One could also study* $f(x) = e^x - x^e$ *and try to show* $f(x) > 0$ *when* $x > e$*; however, it is hard to analyze all the critical points. It is easier to study* $g(x) = e^{x/e} - x$*, and show* $g(x) > 0$ *for* $x > e$*.*

Alternatively, taking logarithms of both sides yields π *versus* $e\log\pi$*, or equivalently* $\frac{\pi}{e}$ *versus* $\log\pi$*. But*

$$\log\pi \;=\; \log\left(e\cdot\frac{\pi}{e}\right) \;=\; 1+\log\left(1+\frac{\pi-e}{e}\right). \tag{E.54}$$

The proof is completed by using the Taylor series expansion for $\log(1+x)$ *for* $x \in (0,1)$ *(the Taylor series is* $\sum_{n=1}^{\infty}\frac{(-1)^{n+1}x^n}{n}$*):*

$$\begin{aligned} \log\pi &= 1+\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}\left(\frac{\pi-e}{e}\right)^n \\ &= 1+\frac{\pi-e}{e}-\frac{1}{2}\left(\frac{\pi-e}{e}\right)^2+\cdots \\ &= \frac{\pi}{e}-\frac{1}{2}\left(\frac{\pi-e}{e}\right)^2+\cdots, \end{aligned} \tag{E.55}$$

which is less than $\frac{\pi}{e}$ *as we have an alternating series with decreasing absolute values. Note we multiplied by* 1 *in a clever way so as to be able to exploit the Taylor series expansion for* $\log(1+x)$*. Using alternating series where the terms have decreasing absolute values is a common technique to obtain bounds. If we stop at a positive term, we have an upper bound; if we stop at a negative term, we have a lower bound.*

References

[BoDi] W. Boyce and R. DiPrima, *Elementary differential equations and boundary value problems*, 7th edition, John Wiley & Sons, New York, 2000.

[Ha] B. Hayes, *Third Base: Three cheers for base 3!*, American Scientist **89** (2001), no. 6, 490–494.

[MM] I. Miller and M. Miller, *John E. Freund's Mathematical Statistics with Applications*, seventh edition, Prentice Hall, 2004.

[MT-B] Steven J. Miller and R. Takloo-Bighash, *An Invitation to Modern Number Theory*, Princeton University Press, Princeton, NJ, 2006.

[Rud] W. Rudin, *Principles of Mathematical Analysis*, third edition, International Series in Pure and Applied Mathematics, McGraw-Hill Inc., New York, 1976.

[WW] E. Whittaker and G. Watson, *A Course of Modern Analysis*, 4th edition, Cambridge University Press, Cambridge, 1996.