CALCULUS AND PROBABILITY BASED STATISTICS

MARK ANDREW GERADS

1. Finite statistics

1.1. **Introduction.** The purpose of this text is to have statistics built on a solid foundation of probability and calculus.

Definition 1.1. A is a repeatable event.

Definition 1.2. p is the probability that A will happen when A is given a chance to happen. For simplicity, we assume that p is the same every time A is given a chance to happen.

Definition 1.3. $n \in \mathbb{W}$ is the number of times A had a chance to happen.

Definition 1.4. $k \in \mathbb{W}$ is the number of times A actually happened.

Remark 1.5. $k \leq n$.

Definition 1.6. f_p is the probability is the probability distribution for p.

Remark 1.7. $1 = \int_0^1 f_p \mathrm{d}p$

Definition 1.8. m is the maximum value of f_p.

Definition 1.9. (a, b) is the prediction interval of p such that $b - a$ is minimized.

Definition 1.10. c is the desired certainty for the prediction interval of p.

Remark 1.11. $0 \leq c = \int_a^b f_p \mathrm{d}p \leq 1$

Remark 1.12. $0 \leq a \leq b \leq 1$

Now that we have the definitions, here is the problem: n, k, c are known, and we must solve for a, b.

Remark 1.13. The probability of k events given n chances is $\binom{n}{k} p^k (1-p)^{n-k}$, but we must scale it to satisfy 1.7:

$$f_p = \frac{\binom{n}{k} p^k (1-p)^{n-k}}{\int\limits_0^1 \binom{n}{k} p^k (1-p)^{n-k} \mathrm{d}p} = \frac{p^k (1-p)^{n-k}}{\int\limits_0^1 p^k (1-p)^{n-k} \mathrm{d}p} = (n+1)\binom{n}{k} p^k (1-p)^{n-k} \tag{1.1}$$

1.2. **The 4 cases of finding the interval.**

1.2.1. *Case $n = k = 0$.* This case is self-symmetrical.

Remark 1.14. We have

$$(a, b) = \left(\frac{1-c}{2}, \frac{1+c}{2}\right) \tag{1.2}$$

Date: February 19, 2016.

1.2.2. *Case* $n > k = 0$. This case is symmetrical to case 1.2.3.

Remark 1.15. We can solve

$$\int_0^b (n+1)(1-p)^n \mathrm{d}p = c = 1-(1-b)^{n+1} \tag{1.3}$$

Remark 1.16. We have

$$(a,b) = \left(0, 1-(1-c)^{\frac{1}{n+1}}\right) \tag{1.4}$$

Example 1.17. Now we can analyze the old saying "Third time's the charm.". $n = 2$.

$$c = 0.5 \Rightarrow b \approx 0.2062994740, c = 0.95 \Rightarrow b \approx 0.6315968501 \tag{1.5}$$

We can now see that if you fail at something twice, then success before the 8th try is worth approximately a coin flip.

Example 1.18. Suppose we were repeatedly trying a difficult task. How many times of trying without success must there be for us to conclude with $c = 0.99$ that $b \leq 0.01$?

$$n = \frac{2\log(3)+\log(11)}{2\log(2)-2\log(3)+2\log(5)-\log(11)} \approx 457.2105766 \tag{1.6}$$

We can see "If at first you don't succeed, try, try again." is sensible for things that we want to be certain are probable before giving up.

1.2.3. *Case* $n = k > 0$. This case is symmetrical to case 1.2.2.

Remark 1.19. We can solve

$$\int_a^1 (k+1)p^k \mathrm{d}p = c = 1-a^{k+1} \tag{1.7}$$

Remark 1.20. We have

$$(a,b) = \left((1-c)^{\frac{1}{k+1}}, 1\right) \tag{1.8}$$

1.2.4. *Case* $n > k > 0$. This case is self-symmetrical.

Remark 1.21. We desire to solve

$$\int_a^b (n+1)\binom{n}{k}p^k(1-p)^{n-k}\mathrm{d}p = c = (n+1)\binom{n}{k}\left(B_b(k+1,-k+n+1)-B_a(k+1,-k+n+1)\right) \tag{1.9}$$

where B is the incomplete beta function.

Remark 1.22. We can use the hint

$$a^k(1-a)^{n-k} = b^k(1-b)^{n-k} \tag{1.10}$$

Remark 1.23. By solving

$$\frac{\partial f_p}{\partial p} = 0 \tag{1.11}$$

, we get

$$\lim_{c\to 0}(a,b) = \left(\frac{k}{n}, \frac{k}{n}\right), m = (n+1)\binom{n}{k}\left(\frac{k}{n}\right)^k\left(1-\frac{k}{n}\right)^{n-k} \tag{1.12}$$

318 Thomas Drive, Marshalltown, Iowa, 50158
E-mail address: MGerads11@winona.edu, nazgand@gmail.com