SIMULTANEOUS TREATMENT OF DISCRETE AND CONTINUOUS PROBABILITY BY USE OF STIELTJES INTEGRALS

By

William Dowell Baten

The object of this paper is to present several theorems pertaining to the probability that certain functions lie within certain intervals. The first theorem is a generalization of Markoff's Lemma, which is proven for the discrete and continuous cases by use of the accumulative frequency function and Stieltjes integrals. Tchebycheff's Theorem is obtained as a corollary to a very general theorem, the proof of which is based upon the first theorem. Other corollaries are given.

Three theorems, due to Guldberg, which follow are concerned with the probability that a non-negative chance variable be less than certain functions of the expected value of the variable. These are proved for the discrete and continuous cases by employing accumulative frequency functions and Stieltjes integrals. This is the first time, as far as the writer knows, the discrete and continuous cases for these theorems have been included in a single proof.

Theorem 1. If A denotes the expected value of the non-negative variable x and t is any number greater than 1, then the probability that $x \leq At^2$ is greater than $1-1/t^2$.

Proof: If x is a discrete variable with values at x_i, ($i = 1, 2, \cdots, n$) with corresponding probabilities p_i, then it is understood that the probability that x takes other values is zero. If x is a continuous variable having a probability function defined over the interval (a, b), then it is understood that the probability that x lies outside of (a, b) is zero in case (a, b) is different from ($-\infty, +\infty$). In both cases x is a continuous variable in the interval ($-\infty, +\infty$). Let the probability that x lies in the interval ($-\infty, x$) be $F(x)$, with $F(-\infty) = 0$ and $F(+\infty) = 1$ Then the probability that x lies in the interval (x_1, x_2) is

$$F(x_2)-F(x_1)+\frac{1}{2}\left\{F(x_2+0)-F(x_2-0)\right\}+\frac{1}{2}\left\{F(x_1+0)-F(x_1-0)\right\}$$

where the last two limits are different from zero when there is probability different from zero at x_1 and x_2. This exists since $F(x)$ is a non-decreasing function over the interval $(-\infty, +\infty)$. In the special case when $F(x) = \int_{-\infty}^{x} f(x)dx$ where $f(x)$ is summable, $f(x)dx$ represents the probability that x lies in the interval $(x, \; x+dx)$.

In either case, by definition

$$A = \int_{-\infty}^{\infty} x \cdot dF(x)$$

$x > At^2$ in the interval $(At^2 + \epsilon, \infty)$, where ϵ approaches 0, hence

$$A > \int_{At^2+0}^{\infty} x \cdot dF(x)$$

But

$$\int_{At^2+0}^{\infty} x\,dF(x) = \lim_{\epsilon \to 0} \int_{At^2+\epsilon}^{\infty} x\,dF(x) = \lim_{\epsilon \to 0} z_\epsilon \int_{At^2+\epsilon}^{\infty} dF(x) = \lim_{\epsilon \to 0} Z_\epsilon \cdot \lim \int_{At^2+\epsilon}^{\infty} dF(x)$$

by the first theorem of the mean, which holds for Stieltjes integrals in this case. Here $z > At^2 + \epsilon$, hence $\lim_{\epsilon \to 0} z_\epsilon \geq At^2$, therefore $A > At^2 \int_{At^2 0}^{\infty} dF(x)$ But $\int_{At^2 0}^{\infty} dF(x)$ is the probability P that x is greater that At^2, hence

$$A > At^2 P, \quad Q > 1 - 1/t^2$$

where Q is the probability that $x \leq At^2$.

This theorem is a generalization of Markoff's Lemma[1], which he proved for the discrete case. The above proof takes care of the discrete case, the continuous case and the case which is a combination of the discrete and continuous.

Theorem 2. If $f(x_1, x_2, \cdots \cdots x_n)$ is a function of n independent variables, then the probability that

$$|f - k| \leq t\sqrt{E(f^2) - 2kE(f) + k^2}$$

1. "Wahrscheinlichkeitsrechnung," by Markoff. 1912. Page 54.

is greater than $1-1/t^2$; where E represents the expected value, k is a constant and $t>1$

Proof: Let

$$y=\{f(x_1, x_2, \cdots x_n)-k\}^2 \quad \text{then}$$
$$E(y)=E(f^2)-2kE(f)+k^2$$

By theorem 1 the probability that

$$|f-k| \leq t\sqrt{E(f^2)-2kE(f)+k^2}$$

is greater than $1-1/t^2$.

Corollary: If $f(x_1, x_2, \cdots x_n)=x_1+x_2+\cdots x_n$ and $k=\sum_{i=1}^{n} E(x_i)$, then theorem 2 becomes the famous Tchebycheff theorem[1]. This theorem is: If $x_1, x_2, \cdots x_n$ be n independent variables, then the probability that

$$\left|\sum_{i=1}^{n} x_i-\sum_{i=1}^{n} E(x_i)\right| \leq t\sqrt{\sum_{i=1}^{n} E(x_i^2)-2\sum_{\substack{i,j=1\\ i\neq j}}^{n} E(x_i x_j)+\{\sum_{i=1}^{n} E(x_i)\}^2},$$

is greater than $1-1/t^2$.

This proof is by far simpler than that given by Tchebycheff, while it is similar to that given by Markoff.

In the corollary if $k=\sum_{i=1}^{n} E(x_i)$, $E(x_i)=a$ $E(x_i^2)=A$ then the probability that

$$\left|\frac{\sum x_i}{n}-a\right| \leq \frac{t}{\sqrt{n}}\sqrt{A-a}$$

is greater than $1-1/t^2$.

If $f(x_1, x_2, \cdots\cdots x_n)=\sum_{i=1}^{n} x_i^s$ then the probability that

$$\left|\sum_{i=1}^{n} x_i^s-k\right| \leq t\sqrt{\sum_{i=1}^{n} A_{i,2s}+2\sum_{i=1}^{n} a_{i,s}\, a_{j,s}-2k\sum_{i=1}^{n} a_{i,s}+k^2}$$

is greater than $1-1/t^2$, where the variables are independent, $E(x_i^{2s})=A_{i,2s}$; $E(x_i^s)=a_{i,s}$. If s is negative it is under-

1. "Des Valeurs Moyennes," by Tchebecheff, Journal de Math. 1867 (2). Vol. 12.

stood that x can not take on the value zero, if $s = a/b$, where a is odd and b is even, it is understood that x is non-negative.

Other interesting results may be obtained from this theorem if $f(x_1, x_2, \cdots \cdots x_n)$ represents various functions of the n independent variables and k be given different values. If f is the sum of the variables, theorem 2 is a more general theorem than Tchebycheff's theorem because of the constant k which may have values other than $\sum_{i=1}^{n} E(x_i)$

Let x_i be the result of an individual throw of a coin, $x_i = 1$ if a head is thrown and $x_i = 0$ if a tail is thrown; then $E(x_i^2) = p \cdot 1 + q \cdot 0$, where p is the probability of a head and q is the probability of a tail. Let m represent the number of heads thrown in n throws and let $k = np \pm \sqrt{n - npq}$, then the probability that

$$\left| m - (np \pm \sqrt{n - npq}) \right| \leq t\sqrt{n}, \text{ or that}$$

$$\left| \frac{m}{n} - \left(p \pm \sqrt{\frac{1}{n} - \frac{pq}{n}} \right) \right| \leq \frac{t}{\sqrt{n}} \quad \text{, is greater than } 1 - 1/t^2 .$$

Let $t = \sqrt[4]{n}$, then the probability that

$$-\frac{1}{\sqrt[4]{n}} \pm \sqrt{\frac{1}{n} - \frac{pq}{n}} \leq \frac{m}{n} - p \leq \frac{1}{\sqrt[4]{n}} \pm \sqrt{\frac{1}{n} - \frac{pq}{n}}$$

is greater than $1 - 1/t^2$ or $1 - 1/\sqrt{n}$, which approaches unity as n increases. It is near unity for large values of n. This shows that the empirical probability approaches the true probability p as the number of throws increases, and the advantage of k.

Theorem 3: Let $u'_{n:x}$ be the expected value of the non-negative variable x raised to the power n and t any number greater than 1, then the probability that $x \leq t\sqrt[n]{u'_{n:x}}$ is greater than $1 - 1/t^n$.

Proof: *Let* $c > \sqrt[n]{u'_{n:x}}$, and let $F(x)$ be the probability that x lies in the interval $(-\infty, x)$, then by definition

$$u'_{n:x} = \int_{-\infty}^{\infty} x^n dF(x); \quad and \quad \frac{u'_{n:x}}{c} = \int_{-\infty}^{\infty} x^n dF(x)/c^n$$

Now

$$\frac{u'_{n:x}}{c^n} > \int_{-\infty}^{\infty} x^n dF(x)/c^n = \lim_{\epsilon \to 0} \int_{c+\epsilon}^{\infty} x^n dF(x)/c^n$$

$$= \lim_{\epsilon \to 0} \frac{(z_e)^n}{c^n} \cdot \lim_{\epsilon \to 0} \int_{c+\epsilon}^{\infty} dF(x) > i \cdot \int_{c+0}^{\infty} dF(x),$$

by the first theorem of the mean, and since $\lim\limits_{\epsilon \to 0} \frac{(z_\epsilon)^n}{c^n} \geq 1.$

Since $\int_{c+0}^{\infty} dF(x)$ is the probability P that x is greater than c,

$$\frac{u'_{n:x}}{c^n} > p \quad or \quad Q > 1 - \frac{u_{n:x}}{c^n}$$

where Q is the probability that $x \leq c$.

Let $t = \frac{c}{\sqrt[n]{u'_{n:x}}}$, then $\frac{1}{t^n} = \frac{u'_{n:x}}{c^n}$, hence

$$Q > 1 - 1/t^n \ .$$

But Q becomes the probability that $x \leq \sqrt[n]{u'_{n:x}}$, since c was any number greater than $\sqrt[n]{u'_{n:x}}$.

Let $y = |x - k|$, then theorem 3 becomes: If $u'_{n \cdot y}$ is the expected value of $|x-k|^n$ and t is greater than 1, then the probability that $|x-k|$ does not surpass the multiple $t\sqrt[n]{u'_{n \cdot y}}$, is greater than $1 - 1/t^n$, where k is a constant.

If $k = \int_{-\infty}^{\infty} x \cdot dF(x)$, then $u'_{n \cdot y}$ becomes $u'_{n:\bar{x}}$ and theorem 3 states that the difference $|x-k|$ does not surpass the multiple $t\sqrt[n]{u'_{n:\bar{x}}}$, is greater than $1 - 1/t^n$. In this special case theorem 3 becomes Guldberg's theorem[1], but this is more general than his theorem, for it includes the continuous case, the discrete case and the case which is a combination of the discrete and continuous.

If $y = |f(x) - k|$ is used for the variable, a more general theorem is obtained. Here $f(x)$ is a function of x . Of course, the probability law for $f(x)$ must be secure from that of x if the continuous case is under consideration. Certain restrictions must be placed upon $f(x)$ concerning continuity, summability and concerning the inverse.

---

1. "Sur un théorèmede M. Markoff," by Alf. Guldberg. Compte Rendue, Vol. 175. (1922) page 679.

Theorem 4. The probability that the difference $|x-m|$ is not greater than the multiple $t\,u_{r:x}$, $t>1$, is greater than $1-\left(\sqrt[n]{u'_{n:x}}/u'_{r:x}\right)^n \left(1/t^n\right)$, where $u_{r.x}$ is the expected value of $|x-m|^r$, and m is the expected value of x.

Theorem 5. The probability that the positive quantity x does not surpass the multiple tm, $(t>1)$, is greater than $1-\left(\frac{\sqrt[n]{u'_{n:x}}}{m}\right)^n \cdot \frac{1}{(t-1)^n}$,

where $u_{n.x}$ is the expected value of $|x-m|^n$, and $m=E(x)$.

These last two theorems are due to Guldberg[1] for the discrete case. By the method used in theorem 3 these can be proven for the continuous case, the discrete case, and the case which is a combination of the discrete and continuous.

---

1. "Sur quelques inéqualités des le calcul de probabilités," by Guldberg. Comp. Rend. Vol. 175 (1922), p. 1382.
   "Sur le théorème de, Tchebecheff," by Guldberg. Comptes Rendue, Vol. 175 (1922), p. 418.