

Proofs of Binomial and Geometric Series Expansions Under a Probabilistic Microscope with Real-Life Examples

Nitis Mukhopadhyay

Department of Statistics, University of Connecticut
Storrs, CT 06269-4120, U.S.A.
nitis.mukhopadhyay@uconn.edu

Received: 3rd July 2016 / Revised: 2nd August 2016 / Accepted: 2nd August 2016

©IAppStat-SL2016

ABSTRACT

This paper begins with a new and complete probabilistic proof of the well-known binomial theorem. Then, it broadens the same spirit for handling a useful infinite series of which the well-known geometric series forms a special case. Also included are applications of the theorems and real-life examples.

Keywords: binomial distribution, geometric distribution, independence, infinite series, negative binomial distribution.

1. Introduction

While teaching probability and statistics, it has been my experience that both students and teachers feel more comfortable in employing probabilistic arguments to justify non-probabilistic mathematical conclusions. I have observed first-hand how probabilistic derivations of some well-known mathematical results tend to energize in my classrooms. More generally, such an approach may help in building a bridge between teaching and research at many levels by enriching both.

Professor Sujit K. Mitra, a legendary figure in statistics and mathematics, often quoted problems from linear algebra which he used to prove using probabilistic and statistical concepts. He had a unique style of teaching. One may find my interview piece (Mukhopadhyay 1997) with him interesting to read. I felt totally mesmerized every time Professor Mitra gave brilliant probabilistic arguments to make his students think outside the box.

I recall that Professor Mitra once quoted the following mathematical problem along with his ingenious probabilistic argument to prove it: Suppose A, B are $p \times p$ real symmetric and positive semi definite (p.s.d.) matrices. Then, show that $A+B$ is a $p \times p$ p.s.d. matrix. Professor Mitra would argue as follows: Suppose that we have \mathbf{X} which is $N_p(\mathbf{0}, A)$, \mathbf{Y} which is $N_p(\mathbf{0}, B)$, and they are distributed independently. Then, $\mathbf{X} + \mathbf{Y}$ is $N_p(\mathbf{0}, A+B)$. Since a dispersion

matrix must be p.s.d., the matrix $A + B$ must be p.s.d.

This proof is as crisp and pretty as one gets. It brings back memories and it continues to energize me to create probabilistic arguments to prove non-probabilistic mathematical conclusions when possible. This can be challenging and refreshing at the same time.

Rosalsky (2007a) proved the binomial theorem using elementary probabilistic tools. Mukhopadhyay's (2007) letter to the editor and a response from Rosalsky (2007b) may be of interest to readers. Such a probabilistic treatment clearly remains important and it has drawn attention from other researchers recently. For example, a recent paper of Peterson (2013) testifies to this sentiment.

Binomial Theorem: For fixed real numbers a, b and a fixed positive integer n , the following holds:

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}. \quad (1)$$

There is ample room to develop a complete proof of (1) via probabilistic arguments. In Section 2, I do so with a new and complete probabilistic proof. Then, in Section 3, I broaden the same spirit by handling a useful infinite series of which the well-known geometric series forms a special case. I include applications of the results briefly in Section 2.1 and Section 3.3. Additionally, real-life examples are given in Section 2.2 and Section 3.4.

2. A Proof of the Binomial Theorem

For precision and accuracy, I subdivide our proof into a number of cases. These steps are important to keep us acutely conscious about couple of issues: I must (i) use simple probabilistic tools only; (ii) not use the binomial theorem itself indirectly in any shape or form; and (iii) not use mathematical induction. I interpret 0^j as 0 whatever be the positive integer j and $0^0 = 1$. Now, (1) obviously holds when $a = 0$ or $b = 0$ or both. So, without any loss of generality, I let $a \neq 0$ and $b \neq 0$.

Suppose that X_1, \dots, X_n are *independent and identically distributed* (i.i.d.) as Bernoulli(p) random variables, $0 < p < 1$. Then, $T = \sum_{i=1}^n X_i$ has the Binomial(n, p) distribution given as follows:

$$P(T = k) = \binom{n}{k} p^k q^{n-k}, k = 0, 1, \dots, n \text{ and } q = 1 - p. \quad (2)$$

For the probability distribution (2), I must have the sum of the probabilities of all

possible values of T , that is, $\sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = 1$. Thus, I can immediately claim:

$$(p + q)^n \equiv 1 = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} \text{ for all fixed } 0 < p < 1, \tag{3}$$

$q = 1 - p$ and positive integer n .

Case 1: Suppose that a and b are any two non-zero real numbers with the same sign.

Then, using (3), I write:

$$1 \equiv \frac{(a + b)^n}{(a + b)^n} = (p + q)^n \text{ with } p = \frac{a}{a + b}, q = 1 - p = \frac{b}{a + b}$$

so that I have:

$$\frac{(a + b)^n}{(a + b)^n} = \sum_{k=0}^n \binom{n}{k} \left(\frac{a}{a + b}\right)^k \left(\frac{b}{a + b}\right)^{n-k} \Rightarrow (a + b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}.$$

Now, Case 1 is done. ■

Case 2: Suppose that a and b are any two non-zero real numbers with different signs. Equivalently, I consider $(b - a)^n$ where $a > 0, b > 0$. Let X_1, \dots, X_n be i.i.d. Bernoulli(p) random variables with $p = \frac{a}{a + b}, q = 1 - p = \frac{b}{a + b}$. Observe that $E[(-1)^{X_k}] = q - p$ for all $k = 1, \dots, n$. Then, I can express:

$$(q - p)^n = \prod_{k=0}^n E[(-1)^{X_k}] = E[(-1)^{\sum_{k=0}^n X_k}] = \sum_{k=0}^n \binom{n}{k} (-p)^k q^{n-k},$$

which implies:

$$(b - a)^n = \sum_{k=0}^n \binom{n}{k} (-a)^k b^{n-k}.$$

The proof of (1) is now complete. ■

Remark 2.1. Case 2 could alternatively be handled as follows. I may define the following partition for the set $\{0,1\}^n$, that is the collection of all n -tuples filled only with 0's and 1's: Denote

$$A_k = \{\mathbf{x} = (x_1, \dots, x_n) \in \{0,1\}^n : t = t(\mathbf{x}) \equiv \sum_{i=1}^n x_i = k\}, \text{ for } k = 0, 1, \dots, n.$$

Observe that if $\mathbf{x} \in A_k$, then we must have $(1-2x_1)(1-2x_2)\dots(1-2x_n) = -1$ or 1 according as k is odd or even, respectively. Next, I may express $(q-p)^n$ as follows:

$$\begin{aligned} (1-2p)^n &= \sum_{k=0}^n \sum_{\mathbf{x} \in A_k} (1-2x_1)(1-2x_2)\dots(1-2x_n) \binom{n}{k} p^k q^{n-k} \\ &= \sum_{k=0}^n \binom{n}{k} (-p)^k q^{n-k}. \end{aligned}$$

2.1. Applications

Suppose that T is distributed as Binomial(n, p). Then, we have:

$$\begin{aligned} E[T] &= \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} = np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} q^{n-k} \\ &= np \{p + (1-p)\}^{n-1}, \text{ in view of (1),} \end{aligned}$$

which means $E[T] = np$ since $p + (1-p) = 1$. Also, we have:

$$V[T] = E[T(T-1)] + np - n^2 p^2,$$

but

$$\begin{aligned} E[T(T-1)] &= \sum_{k=0}^n k(k-1) \binom{n}{k} p^k q^{n-k} = n(n-1) p^2 \sum_{k=2}^n \binom{n-2}{k-2} p^{k-2} q^{n-k} \\ &= n(n-1) p^2 \{p + (1-p)\}^{n-2}, \text{ in view of (1),} \end{aligned}$$

so that $V[T] = np(1-p)$. These are well-known results that rely upon binomial theorem (1).

2.2. A Real-Life Example

Having observed a value t for T , one may estimate p by

$$\hat{p} = t/n \tag{4}$$

with its estimated bound of error:

$$B = 2\{\hat{p}(1 - \hat{p})/n\}^{1/2}. \quad (5)$$

In the light of Scheaffer et al. (2012, pp. 90-92), suppose that a sample of 100(= n) college seniors selected randomly from 1000(= N) college seniors, 60(= t) responded that they have held part-time jobs. One is interested in estimating the fraction(= p) of seniors in the whole college who have held part-time jobs.

Under random sampling with replacement, the scenario will correspond to binomial sampling and I obtain from (4)-(5):

$$\hat{p} = 60/100 = 0.6 \text{ with } B = 2(0.6(1 - 0.6)/100)^{1/2} \approx 0.09798.$$

Thus, with approximately 95% confidence, one may estimate p to lie between 50.2% and 69.8% which respectively correspond to $\hat{p} - B$ and $\hat{p} + B$ with B coming from (5).

In survey sampling, however, it is more customary to gather observations under random sampling without replacement. If one assumes that one had observed the same responses under random sampling without replacement, then I should report:

$$\hat{p} = t/n \quad (6)$$

with its estimated bound of error:

$$B = 2\{t(1-t)/(n-1)\}^{1/2}\{1 - nN^{-1}\}^{1/2}. \quad (7)$$

The extra expression $\{1 - nN^{-1}\}^{1/2}$ in B from (7) is referred to as the finite population correction. One may refer to Scheaffer et al. (2012, Eq (4.16), pp. 90-92).

Hence, I immediately obtain from (6)-(7):

$$\hat{p} = 60/100 = 0.6 \text{ with } B = 2(0.6(1 - 0.6)/99)^{1/2}(1 - 0.1)^{1/2} \approx 0.09342.$$

Thus, with approximately 95% confidence, one may estimate p to lie between 50.7% and 69.3% which respectively correspond to $\hat{p} - B$ and $\hat{p} + B$ with B coming from (7).

The two sets of answers are very similar to each other because the sampling fraction, namely n/N , is 0.1 which is rather small. That renders random sampling from a finite population, whether implemented using with replacement or without replacement, nearly equivalent.

3. Infinite Series Expansion Including the Geometric Series

Along the line of Section 2, I develop briefly proofs of some useful infinite

series expansions by exploiting simple probabilistic arguments only.

An Infinite Series Expansion: For a fixed real number ρ , $-1 < \rho < 1$, the following holds:

$$(1 - \rho)^{-r} = \sum_{k=1}^{\infty} a_{k-1,r} \rho^{k-1} \quad \text{with } a_{0,r} \equiv 1 \quad \text{and} \quad (8)$$

$$a_{k-1,r} = \frac{1}{(k-1)!} r(r+1)\dots(r+k-2) \quad \text{for } k = 2, 3, \dots$$

for all fixed positive integer r .

The expansion of $(1 - \rho)^{-r}$ laid out in (8) when $r = 1$ is referred to as the well-known **geometric series**. I will verify the infinite series expansion in (8) using only simple probabilistic steps in Section 3.1.

3.1. Probabilistic Derivation of an Infinite Series

The infinite series (8) obviously holds for any arbitrary but fixed positive integer r when $\rho = 0$. In what follows, I consider the two cases $0 < \rho < 1$ and $-1 < \rho < 0$ separately.

Case 1: $0 < \rho < 1$

Let me pretend having an infinite sequence of independent Bernoulli(p) trials with $p \equiv P(S) = 1 - \rho$ and $q \equiv P(F) = \rho$ where “ S ” , “ F ” respectively denotes a success or a failure. Now, I define the following random variable:

$Y =$ minimum number of trials required to observe the r^{th} success, S .

I will surely need at least r trials to observe the r^{th} S .

Now, for $y \geq r$, the event “ $Y = y$ ” is equivalent to the event where I observe $(r-1)$ successes (S 's) and $(y-r)$ failures (F 's) in the trial #'s $1, \dots, (y-1)$ along with one S at trial # y . Thus, I have the following probability distribution for Y :

$$P(Y = y) = \left\{ \binom{y-1}{r-1} p^{r-1} q^{y-r} \right\} p = \binom{y-1}{y-r} p^r q^{y-r}, \quad y = r, r+1, \dots \quad (9)$$

Observe that the expression within $\{ \}$ in (9) corresponds to binomial probability from (2).

Now, $\sum_{y=r}^{\infty} \binom{y-1}{y-r} p^r q^{y-r}$ being the sum of the probabilities of all

possible values of Y , we must have $\sum_{y=r}^{\infty} \binom{y-1}{y-r} p^r q^{y-r} \equiv 1$. Thus, I immediately claim:

$$(1-\rho)^{-r} \equiv p^{-r} = \sum_{y=r}^{\infty} \binom{y-1}{y-r} q^{y-r} = \sum_{y=r}^{\infty} \binom{y-1}{y-r} \rho^{y-r}. \quad (10)$$

One can see from (10) that in this infinite series expansion on the right-hand side, the 1st term is 1, the 2nd term is $r\rho$, the 3rd term is $\frac{1}{2}r(r+1)\rho^2$, the 4th term is $\frac{1}{6}r(r+1)(r+2)\rho^3$, and so on and so forth. Now, Case 1 is done. ■

Case 2: $-1 < \rho < 0$

We suppose that $\rho = -\rho'$ so that $0 < \rho' < 1$, and we want to expand $(1-\rho)^{-r}$ or equivalently $(1+\rho')^{-r}$. Since $0 < \rho^2 = \rho'^2 < 1$, observe that I can express:

$$(1-\rho^2)^{-r} \equiv (1-\rho'^2)^{-r} = 1 + a_{1,r}\rho'^2 + a_{2,r}\rho'^4 + a_{3,r}\rho'^6 + \dots, \text{ by Case 1,}$$

which implies:

$$(1+\rho')^{-r} = (1-\rho')^r \left(1 + a_{1,r}\rho'^2 + a_{2,r}\rho'^4 + a_{3,r}\rho'^6 + \dots \right) \quad (11)$$

Next, using the binomial theorem on $(1-\rho')^r$, I rewrite $(1+\rho')^{-r}$ from (11) as follows:

$$\begin{aligned} & \left\{ 1 - \binom{r}{1}\rho' + \binom{r}{2}\rho'^2 - \binom{r}{3}\rho'^3 + \dots + (-1)^r \binom{r}{r}\rho'^r \right\} \\ & \quad \times \left(1 + a_{1,r}\rho'^2 + a_{2,r}\rho'^4 + a_{3,r}\rho'^6 + \dots \right) \quad (12) \\ & = 1 - \binom{r}{1}\rho' + \left\{ \binom{r}{2} + a_{1,r} \right\} \rho'^2 - \left\{ \binom{r}{3} + \binom{r}{1}a_{1,r} \right\} \rho'^3 \\ & \quad + \left\{ \binom{r}{4} + \binom{r}{2}a_{1,r} + a_{2,r} \right\} \rho'^4 - \dots \end{aligned}$$

Now, let me examine carefully the coefficients of $-\rho'$, ρ'^2 , $-\rho'^3$, ρ'^4 , ... found in (12).

The coefficient of $-\rho'$ is: $r \equiv a_{1,r}$;

The coefficient of ρ'^2 is:

$$\binom{r}{2} + a_{1,r} = \frac{1}{2}r(r-1) + r = \frac{1}{2}r(r+1) \equiv a_{2,r};$$

The coefficient of $-\rho'^3$ is:

$$\begin{aligned} \binom{r}{3} + \binom{r}{1}a_{1,r} &= \frac{1}{6}r(r-1)(r-2) + r^2 \\ &= \frac{1}{6}r(r^2 - 3r + 2 + 6r) = \frac{1}{3!}r(r+1)(r+2) \equiv a_{3,r}; \end{aligned}$$

The coefficient of ρ'^4 is:

$$\begin{aligned} \binom{r}{4} + \binom{r}{2}a_{1,r} + a_{2,r} &= \frac{1}{24}r\{(r^3 - 6r^2 + 11r - 6) \\ &+ 12r^2 + 12\} = \frac{1}{4!}r(r+1)(r+2)(r+4). \end{aligned}$$

Proceeding along this line, (12) gives:

$$\begin{aligned} (1 + \rho')^{-r} &= 1 - a_{1,r}\rho' + a_{2,r}\rho'^2 - a_{3,r}\rho'^3 + a_{4,r}\rho'^4 - \dots \\ \Rightarrow (1 - \rho)^{-r} &= 1 + a_{1,r}\rho + a_{2,r}\rho^2 + a_{3,r}\rho^3 + a_{4,r}\rho^4 + \dots \end{aligned}$$

Now, my derivation of (8) is complete. ■

3.2. An Alternative Derivation of (8) Under Case 2

Again, let me assume that $-1 < \rho < 0$. Combining (11) and (12) for $r=1$, with $\rho' = -\rho$, I first rewrite:

$$(1 + \rho')^{-1}$$

$$\begin{aligned}
 &= (1-\rho')(1+a_{1,1}\rho'^2+a_{2,1}\rho'^4+a_{3,1}\rho'^6+\dots) \\
 &= (1+a_{1,1}\rho'^2+a_{2,1}\rho'^4+a_{3,1}\rho'^6+\dots) \\
 &\quad -\rho'(1+a_{1,1}\rho'^2+a_{2,1}\rho'^4+a_{3,1}\rho'^6+\dots) \\
 &= 1-\rho'+a_{1,1}\rho'^2-a_{1,1}\rho'^3+a_{2,1}\rho'^4-a_{2,1}\rho'^5+\dots
 \end{aligned} \tag{13}$$

But, it is easy to verify from (13) that $a_{1,1}=a_{2,1}=a_{3,1}=a_{4,1}=\dots\equiv 1$. This agrees with the expansion of $(1-\rho)^{-1}$:

$$(1-\rho)^{-1}=1+\rho+\rho^2+\rho^3+\rho^4+\dots, \tag{14}$$

and my proof of (8) under Case 2 is complete when $r=1$.

At this point, let me differentiate (with respect to ρ) both sides from (14) successively $r-1$ times and write:

$$\begin{aligned}
 &(r-1)!(1-\rho)^{-r} \\
 &= \frac{\partial^{r-1}}{\partial \rho'^{r-1}}(1+\rho+\rho^2+\rho^3+\rho^4+\dots) \\
 &= \frac{\partial^{r-1}}{\partial \rho'^{r-1}}(\rho^{r-1}+\rho^r+\rho^{r+1}+\dots) \\
 &= (r-1)!\dagger r!\rho+\frac{(r+1)!}{2!}\rho^2+\frac{(r+2)!}{3!}\rho^3+\dots
 \end{aligned} \tag{15}$$

From the first and the last steps in (15), I immediately note:

$$\begin{aligned}
 (1-\rho)^{-r} &= \frac{1}{(r-1)!}\left\{(r-1)!\dagger r!\rho+\frac{(r+1)!}{2!}\rho^2+\frac{(r+2)!}{3!}\rho^3+\dots\right\} \\
 &= 1+r\rho+\frac{(r+1)r}{2!}\rho^2+\frac{(r+2)(r+1)r}{3!}\rho^3+\dots,
 \end{aligned}$$

which agrees with the expansion (8). Case 2 is done. ■

Remark 3.1. In many texts, when $r=1$, the distribution from (9) and the observation Y are respectively referred to as a geometric distribution and a

geometric random variable. In general, this distribution is called a negative binomial distribution. See, for example, Casella and Berger (2001, pp. 30-31), Johnson and Kotz (1969, pp. 123-124 and Chapter 5), Mukhopadhyay (2000, pp. 35-37), Ross (1997, pp. 29, 82), Scheaffer (1995, pp. 97-102), Wackerly et al. (2008, section 3.5).

Remark 3.2. In order to find the mean μ and variance σ^2 of a negative binomial distribution or a geometric distribution, one customarily uses the result from (8) with appropriate positive integral values of r .

3.3. Applications

Let Y^* denote a random variable that has the same distribution as that of Y defined by (9) corresponding to $r = 1$. Then, we have:

$$\begin{aligned} E[Y^*] &= \sum_{y=1}^{\infty} y p q^{y-1} = p \{1 + 2q + 3q^2 + \dots\} \\ &= p(1-q)^{-2}, \end{aligned}$$

in view of (8), which means $E[Y^*] = p^{-1}$ since $1-q = p$. That is, $E[Y] = rE[Y^*] = rp^{-1}$.

Also, we have:

$$V[Y^*] = E[Y^*(Y^* - 1)] + p^{-1} - p^{-2},$$

but

$$\begin{aligned} E[Y^*(Y^* - 1)] &= \sum_{y=1}^{\infty} y(y-1) p q^{y-1} = 2pq \{1 + 3q + 4q^2 + \dots\} \\ &= 2pq(1-q)^{-3}, \end{aligned}$$

in view of (8), so that $V[Y^*] = 2pq(1-q)^{-3} + p^{-1} - p^{-2} = qp^{-2}$.

Thus, $V[Y] = rV[Y^*] = rqp^{-2}$. These are well-known results that rely upon geometric series (8).

3.4. A Real-Life Example

When a large cyclone or hurricane lands and passes, the administration of forestry (filled largely with coconut trees) in a certain country may want to know the number of coconut trees which are still standing after the calamity. That is an important problem given the agricultural base of this country and its economic dependence on ample production of coconut.

However, it will be nearly impossible to count ($= N$) every single standing coconut tree. But, it will be possible to estimate the population size, N , by means

of what is called **inverse sampling** (Scheaffer et al. 2012, Section 10.3) by implementing capture-recapture methodology.

The capture-recapture methodology consists of creating an observable random variable Y (= the minimum number of trials required to observe the r^{th} **success**) as defined in Section 3.1. But, what should correspond to a **success** (S)?

The forest manager's team would go around, throughout the forest, ideally in all conceivable directions, and tie m yellow ribbons around the trunks of m standing coconut trees each of which is then called a marked or tagged coconut tree. When randomly sampled, if one observes a marked or tagged coconut tree, that will be recorded as a **success**, otherwise it will be recorded as a **failure** with $p = m/N$.

Then, one-by-one random sampling (of size = y) of standing coconut trees are recorded, but terminating sampling as soon as r marked or tagged (and consequently $y - r$ not tagged) coconut trees are observed. The positive integers m and r should be predetermined by taking into account the total cost and budget for the completion of such a survey.

Reusing the previous expressions from Section 3.3, after some calculations, I claim:

$$E[Y] = r/p \text{ and } E[(Y + Y^2)/(r + 1)] = r/p^2 \text{ so that}$$

$$E[Y(Y - r)/(r + 1)] = r(p^{-2} - p^{-1}) = V[Y]$$

with $p = m/N$.

Thus, one can easily verify:

$$\hat{N} = my/r \text{ and } \hat{V}[\hat{N}] = m^2 y(y - r)/(r^2(r + 1)) \tag{16}$$

are respectively unbiased estimators of N and $V[\hat{N}]$. These explicit estimators are found in Scheaffer et al. (2012, p.316) without a proof of their unbiasedness.

Now, having observed a value y for Y , I may estimate N unbiasedly by

$$my/r \tag{17}$$

with its estimated bound of error:

$$B = 2\{m^2 y(y - r)/(r^2(r + 1))\}^{1/2}. \tag{18}$$

Suppose that $m = 100$, and I observed $y = 300$ standing coconut trees to locate the 35th marked ($r = 35$) standing coconut trees, and then I quit sampling right there. Then, from this data, I obtain using (16)-(18):

$$\hat{N} = 300(100)/35 \approx 857.14 \text{ with}$$

$$B = 2\left(100^2 \times 300(300 - 35)/(35^2(35 + 1))\right)^{1/2} \approx 268.53$$

Hence, with approximately 95% confidence, I estimate the number of standing coconut trees to lie between 589 and 1126 which respectively correspond to $\hat{N} - B$ and $\hat{N} + B$. Notably, however, such \hat{N} often comes with a large variance associated with it.

4. Epilogue

I have taught elementary probability and statistics courses for both undergraduate and first-year masters (MS) students from standard texts such as Scheaffer (1995), Ross (1997), Mukhopadhyay (2000), Casella and Berger (2001), and Wackerly et al. (2008) over many years. I have tested bits and pieces taken from what has been described in this unified paper in classrooms and during students' seminars as pilots with success. My students got a sense of what a research process might entail at their level. Such an approach helps in building a small bridge between teaching and research by enriching each other.

Acknowledgements

I am grateful to the Editor-in-Chief for the feedback on an earlier version which convinced me to add real-life examples. I also heartily thank Mrs. Yan Zhuang and Mr. Sudeep R. Bapat for their generous help in the final stages of preparing the required WORD file.

References

1. Casella G. and Berger R. L. (2001). *Statistical Inference*. 2nd Edition, California: Duxbury.
2. Johnson N. L. and Kotz S. (1969). *Distributions in Statistics: Discrete Distributions*. New York: Wiley.
3. Mukhopadhyay N. (1997). A conversation with Sujit Kumar Mitra. *Statistical Science*, 12:61-75. DOI:10.1214/ss/1029963263.
4. Mukhopadhyay N. (2000). *Probability and Statistical Inference*. New York: Dekker.
5. Mukhopadhyay N. (2007). Letter to the Editor (2007), "Rosalsky, A. (2007), A simple and probabilistic proof of the binomial theorem, *American Statistician*, 61:161-162." *American Statistician*, 61:283. DOI:10.1198/000313007X223739.
6. Peterson J. (2013). A probabilistic proof of a binomial identity. *American Mathematical Monthly*, 120:558-562. DOI: 10.4169/amer.math.monthly.120.06.

558.

7. Rosalsky A. (2007a). A simple and probabilistic proof of the binomial theorem. *American Statistician*, 61:161-162. DOI: 10.1198/000313007X188397.
8. Rosalsky A. (2007b). Response (to Mukhopadhyay's (2007) letter to the editor). *American Statistician*, 61:284. DOI:10.1198/000313007X223739.
9. Ross S. M. (1997). *Introduction to Probability Models*. 6th Edition. New York: Academic Press.
10. Scheaffer R. L. (1995). *Introduction to Probability and Its Applications*. 2nd Edition. California: Duxbury.
11. Scheaffer R. L., Mendenhall III W., Ott R. L. and Gerow K. G. (2012). *Elementary Survey Sampling*. 7th Edition. Boston: Brooks/Cole.
12. Wackerly D. D., Mendenhall III W. and Scheaffer R. L. (2008). *Mathematical Statistics with Applications*. 8th Edition. California: Duxbury.