Printed: 11 July 2021

Chapter 4

Poisson-Binomial

SECTION 4.1 *introduces the idea that the Binomial distribution is more spread out than each Poisson-Binomial distribution with the same number of trials and the same expected value.*

SECTION 4.2 *explains the identification of the Poisson-Binomial as the set of distributions whose probability generating functions are polynomials with all their roots real.*

SECTION 4.3 *presents some classical facts about polynomials with real roots.*

SECTION 4.4 *shows that all the members of a large family of distributions defined by sampling without replacement, which includes the hypergeometric, have Poisson-Binomial distributions.*

SECTION 4.5 *identifies the mode (or modes) of the Binomial distribution, as a warm-up for the more difficult task of identifying the mode (or modes) of a Poisson-Binomial distribution.*

SECTION 4.6 *shows that Poisson-Binomial distributions are (almost) unimodal, with modes close to their expected values. To avoid a lot of fiddling with special cases the results are first derived under a simplifying assumption.*

SECTION 4.7 *shows that the exact tail probabilities (as opposed to their upper bounds calculated by the MGF method) for the Poisson-Binomial are smaller than the corresponding Binomial tails.*

SECTION 4.8 *establishes the amazing fact that the median of a Poisson-Binomial is very close to its expected value. The argument makes a cunning application of the comparison bound from Section 4.7 followed by an appeal to the Central Limit Theorem.*

4.1 The extreme Binomial

Recall that the Poisson-Binomial distribution, $\text{PB}\textsc{in}(p_1, \dots, p_n)$, is the probability measure on the set $\{0, 1, \dots, n\}$ defined as the distribution of a sum $S = X_1 + \cdots + X_n$ of independent random variables with $X_i \sim \textsc{Ber}(p_i)$, for $i = 1, \dots, n$. As a special case, if $p_j = \theta$ for all j then $S \sim \textsc{Bin}(n, \theta)$.

Sections 3.6 and 3.7 from the previous chapter presented several examples of distributions (including the Poisson-Binomial and hypergeometric) whose MGF's are smaller than the MGF of a Binomial distribution with the same expected value. Application of the MGF method from Section 3.1 gave tail bounds for those distributions that were smaller than the corresponding tail bounds for the Binomial, inequalities that raised the question of whether the actual tail probabilities are really smaller than the actual tail probabilities for the Binomial. This Chapter will show that they are: in a strong sense, the Binomial is an an extreme member of the set of Poisson-Binomials distributions. In addition, Section 4.4 will show that the hypergeometric is pecial case of the the Poisson-Binomial, an assertion that I initially found hard to believe.

Along the way to these conclusions you will learn some surprising facts about the mode and median of the Poisson-Binomial, both of which lie within a distance one of their expected values.

> **Remark.** I first encountered the surprising fact about the median of the Binomial in a small paper on empirical processes by Lucien Le Cam (Le Cam, 1983). I was in awe that anyone should know such an amazing fact. Being young and energetic, I wanted to know why the result was true. After working diligently through the literature I finally realized that, for Le Cam's purposes, the Chebyshev inequality would have sufficed. Nevertheless, I was pleased to have gained a deeper understanding of the humble Poisson-Binomial. I hope you will feel the same way by the end of this Chapter.

4.2 The PGF of the Poisson-Binomial

Suppose $S \sim \text{PB}\textsc{in}(p_1, \dots, p_n)$. Some of the most interesting properties of S lie concealed within its probability generating function (PGF),

$$g(z) = g(z, \mathbf{p}) = \mathbb{P}z^S = \prod\nolimits_{j \le n} (1 - p_j + p_j z) = \sum\nolimits_{k=0}^{n} \mathbb{P}\{S = k\} z^k.$$

If $p_j = 0$ for some j then the corresponding X_j contributes a 0 to S and a 1 to the PGF. Thus there is usually no loss of generality in assuming that $p_j > 0$ for all j, which ensures that $g(z)$ is a polynomial of degree n, with roots $-(1-p_j)/p_j$, for $j = 1, \dots, n$. Note that all these roots are real numbers.

The value 0 can be a root of $g(z, \mathbf{p})$ only if at least one of the p_j's is equal to 1, contributing a factor z to the PGF and adding 1 to S. On a first pass I found it helpful to ignore the minor complications caused by p_j's in $\{0, 1\}$ by requiring $0 < p_j < 1$ for all j. However, there are sometimes good reasons to allow 0's and 1's, as you will see in Sections 4.7 and 4.8.

In general, a polynomial $h(z) = a_0 + a_1 z + \dots + a_n z^n$ of degree n (that is, $a_n \neq 0$) has exactly n roots $r_1, \dots, r_n$ in the complex plane,

$$h(z) = a_n \prod\nolimits_{j=1}^{n} (z - r_j).$$

The roots need not all be distinct. If the distinct values are $R_1, \dots, R_t$ with multiplicities $n_1, \dots, n_t$ then $\sum_{j=1}^{t} n_j = n$ and $h(z) = a_n \prod_{j=1}^{t} (z - R_j)^{n_j}$.

The all-real-roots property characterizes the Poisson-Binomial.

<1> **Lemma.** *Suppose a random variable X takes values in $\{0, 1, \dots, n\}$ with $\mathbb{P}\{X = n\} \neq 0$ and PGF $g(z) = \mathbb{P} z^X$. If all n roots $r_1, \dots, r_n$ of g are real then X has a* $\text{PBin}(p_1, \dots, p_n)$ *distribution with $p_j = (1 - r_j)^{-1}$.*

Proof. Factorize g as

$$g(z) := \mathbb{P} z^X = \sum\nolimits_{k=0}^{n} z^k \mathbb{P}\{X = k\} = a_n \prod\nolimits_{j=1}^{n} (z - r_j)$$

where $a_n = \mathbb{P}\{X = n\}$ and each r_j real. Strict monotonicity of g on $\mathbb{R}^+$ and $g(0) = \mathbb{P}\{X = 0\} \geq 0$ ensure that $r_j \leq 0$ for each j. Defining $p_j = 1/(1 - r_j)$ we then have $r_j = -(1 - p_j)/p_j$ and

$$g(z) = a_n \prod\nolimits_{j=1}^{n} (z + (1 - p_j)/p_j) = \frac{a_n}{\prod_{j=1}^{n} p_j} \prod\nolimits_{j=1}^{n} (1 - p_j + p_j z).$$

The equality $g(1) = 1$ forces $a_n = \prod_{j=1}^{n} p_j$.

Remark. The proof actually establishes a slightly stronger assertion: If h is a polynomial of degree n with $h(1) = 1$ and all its roots $r_1, \dots, r_n$ are nonpositive (≤ 0) real numbers then h is the PGF of the $\text{PBin}(p_1, \dots, p_j)$ distribution with $p_j = 1/(1 - r_j)$.

<2> **Example.** Suppose W is a random variable that takes values in $\{0, 1, 2\}$ with $\mathbb{P}\{W = k\} = a_k$ for $k = 0, 1, 2$, with $a_2 \neq 0$. Then W has PGF

$$g(z) = \mathbb{P} z^W = a_0 + a_1 z + a_2 z^2.$$

To decide (without the benefit of Lemma <1>) whether the distribution of W is Poisson-Binomial we need to determine whether the equations

<3> $$p_1p_2 = a_2$$

<4> $$p_1 + p_2 - 2p_1p_2 = p_1(1-p_2) + (1-p_1)p_2 = a_1$$

<5> $$1 - (p_1+p_2) + p_1p_n = (1-p_1)(1-p_2) = a_0$$

have a solution with $p_1, p_2 \in [0,1]$. If there were such a solution then we would have

$$p_1 + p_2 = a_1 + 2a_2 = 2\theta := \mathbb{P}W,$$

which would force $p_1 = \theta - t$ and $p_2 = \theta + t$ for some real t. Equation <3> then becomes $\theta^2 - t^2 = a_2$, which has a real solution for t if and only if

<6> $$\theta^2 - a_2 \geq 0.$$

Let me assume that <6> holds. Taking t to be the (positive) square root of $\theta^2 - a_2$ and $p_1 = \theta - t$ and $p_2 = \theta + t$, we then have $p_1p_2 = a_2$ and

$$\begin{aligned}(1-p_1)(1-p_2) &= (1-\theta+t)(1-\theta-t)\\ &= (1-\theta)^2 - t^2\\ &= 1 - 2\theta + \theta^2 - t^2\\ &= (a_0+a_1+a_2) - (a_1+2a_2) + a_2 = a_0,\\ p_1(1-p_2) + (1-p_1)p_2 &= 2\theta - 2(\theta^2 - t^2) = a_1 + 2a_2 - 2a_2 = a_1.\end{aligned}$$

Moreover, both p_i's belong to $[0,1]$ because

$$0 \leq \theta - \sqrt{\theta^2 - a_2} = p_1 \leq p_2 = \theta + \sqrt{(1-\theta)^2 - a_0} \leq 1.$$

In short, inequality <6> is necessary and sufficient for W to have a Poisson-Binomial distribution, a fact that might surprise you because the PGF g has real roots if and only if $a_1^2 \geq 4a_0a_2$. Fortunately,

$$\begin{aligned}\theta^2 - a_2 &= \left(\tfrac{1}{2}a_1 + a_2\right)^2 - a_2\\ &= \tfrac{1}{4}a_1^2 + (1-a_0-a_2)a_2 + a_2^2 - a_2 = \tfrac{1}{4}\left(a_1^2 - 4a_0a_2\right).\end{aligned}$$

That is, <6> holds if and only if g has only real roots. Hooray!

You might check that the recipe in Lemma <1> for generating the probabilities p_1 and p_2 from the roots of g leads to $\theta \pm t$ for $\{p_1, p_2\}$. It took me a goodly amount of algebra to convince myself directly of this fact.

4.3 Polynomials with real roots

There is an extensive literature, going back several centuries, regarding the roots of polynomials. For example, see Hardy, Littlewood, and Pólya (1989, Sections 2.2 and 4.5) and Pólya and Szegö (1976, Part 5, Chap 1) for a result due to Isaac Newton, which is derived in Example <10>. For a more exhaustive treatment see Prasolov (2004). Here are some results related to real roots that are useful for Poisson-Binomial purposes.

To avoid constant repetition of the clumsy "all-real-roots", for $n \ge 1$ define $\mathcal{P}_n$ to be the set of all polynomials of degree n with real coefficients such that all n of its roots are real. Write $\mathcal{P}$ for $\cup_{n\ge 0}\mathcal{P}_n$, where $\mathcal{P}_0$ consists of all constant functions. If $f \in \mathcal{P}_n$ with $n \ge 1$ then

<7> $$f(z) = a_0 + a_1 z + \cdots + a_n z^n = a_n \prod\nolimits_{j=1}^{n} (z - r_j) = a_n \prod\nolimits_{j=1}^{t} (z - R_j)^{n_j}$$

with all the a_j's and r_j's real and $a_n \ne 0$. Here the r_j's are the n roots of f and $R_1, \dots, R_t$ are the distinct roots with multiplicities n_j, that is, $n_j = |\{i \in [n] : r_i = R_j\}|$.

<8> **Lemma.**

(i) If $f \in \mathcal{P}_n$ and $g \in \mathcal{P}_m$ then $fg \in \mathcal{P}_{n+m}$.

(ii) If $f \in \mathcal{P}$ then its derivative Df also belongs to $\mathcal{P}$.

(iii) If $f \in \mathcal{P}_n$ has representation <7> then the polynomial

$$(\tau f)(z) := a_n + a_{n-1} z + \cdots + a_1 z^{n-1} + a_0 z^n$$

belongs to $\mathcal{P}_{n-\ell}$, where ℓ is the multiplicity (possibly zero) of 0 as a root of f. The roots of τf are the reciprocals of the nonzero R_j's with multiplicities n_j.

(iv) If $f \in \mathcal{P}_n$ then $f(\alpha + \beta z) \in \mathcal{P}_n$ for all real constants α and $\beta \ne 0$.

Proof. For (i): If $g(z)$ has roots $s_1, \dots, s_m$ then the polynomial fg has roots $r_1, \dots, r_n, s_1, \dots, s_m$.

For (ii): Without loss of generality suppose $R_1 < R_2 < \cdots < R_t$ in <7>. By Rolle's theorem, Df has at least one zero in each of the $t - 1$ open intervals (R_j, R_{j+1}). The expansion

$$Df(z) = a_n \sum\nolimits_{j=1}^{t} n_j (z - R_j)^{n_j - 1} \prod\nolimits_{i:j\ne i} (z - R_i)^{n_i}$$

shows that R_j is a root of f with multiplicity at least $n_j - 1$. That gives at least $(t-1) + \sum_{j=1}^{t}(n_j - 1) = n - 1$ real roots, counting multiplicities, for Df.

For (iii): If $\ell > 0$ suppose $r_1, \dots, r_\ell$ are the zero roots. First note that, for $z \neq 0$,

$$\begin{aligned}(\tau f)(z) &= z^n(a_0 + a_1/z + \cdots + a_n/z^n)\\ &= z^n f(1/z) = a_n z^n z^{-\ell}\prod\nolimits_{j=\ell+1}^{n}(z^{-1} - r_j)\\ &= a_n \prod\nolimits_{j=\ell+1}^{n}(1 - r_j z).\end{aligned}$$

The final expression is a polynomial (of degree $n-\ell$) in z that agrees with the polynomial $(\tau f)(z)$ for all $z \neq 0$, so it must also agree at $z = 0$. It follows that τf has roots $1/r_j$ for $\ell < j \leq n$.

For (iv): $\prod_j(\alpha + \beta z - r_j)$ has roots $z = (r_j - \alpha)/\beta$.

The next two Examples make use of the ***falling factorial*** function, which is defined for real x as $(x)_0 = 1$ and

$$(x)_k = \prod\nolimits_{j=0}^{k-1}(x-j) = x(x-1)\dots(x-k+1) \qquad \text{for } k \in \mathbb{N}.$$

If x is a positive integer then $(x)_k = 0$ for $k \geq x+1$.

<9> **Example.** Let me show you one way to use Lemma <8> to create a new member of $\mathcal{P}$ from a given polynomial

$$g(z) = a_0 + a_1 z + \cdots + a_n z^n$$

in $\mathcal{P}_n$ for which $a_k > 0$ for all k. The result will be most useful in Section 4.4.

> **Remark.** I spent a lot of energy trying, without success, to relax the strict positivity assumption on the coefficients. Eventually I gave up because a stronger result was not needed for the argument in Section 4.4. My failures led me to think that this result is not as elementary as it seems.

Suppose s and N are integers for which $N \geq n$ and $1 \leq s < N$. Define $b_j := a_j(N-j)_s$ for $0 \leq j \leq n$. Note that $(N-j)_s$ is nonzero iff $N - j \geq s$. That is, a b_j is nonzero iff $0 \leq j \leq m := \min(N-s, n)$. I assert that the polynomial $f(z) = b_0 + \cdots + b_m z^m$ belongs to $\mathcal{P}_m$.

Note that $N - s - m = \kappa := \max(0, N - s - n) \geq 0$. The following assertions are all justified by Lemma <8>.

$$g_1(z) := z^{N-n}(\tau g)(z) = a_n z^{N-n} + \cdots + a_0 z^N \in \mathcal{P}_N,$$

Differentiate s times, noting that $D^s z^{N-j} = 0$ if $j > N - s$.

$$\begin{aligned} g_2(z) := D^s g_1(z) &= \sum\nolimits_{j=0}^{N-s} a_j (N-j)_s z^{N-j-s} \\ &= \sum\nolimits_{j=0}^{(N-s)\wedge n} b_j z^{N-j-s} \\ &= b_0 z^{m+\kappa} + \cdots + b_m z^\kappa \\ &= z^\kappa (b_m + \cdots + b_0 z^m) \in \mathcal{P}_{m+\kappa}. \end{aligned}$$

If $\kappa > 0$ then the z^κ contributes κ zero roots to g_2, with the other real roots coming from $g_3(z) := b_m + \cdots + b_0 z^m$. Thus all m of the roots of g_3 must be real. One more application of τ then gives

$$f(z) = \tau g_3(z) = b_0 + \cdots + b_m z^m \in \mathcal{P}_m,$$

as asserted.

The knowledge that a polynomial in $\mathcal{P}$ has only real roots tells us something about the coefficients of the polynomial, information that can be extracted by repeated appeals to Lemma <8>. The short story is that the result is due to Isaac Newton; see the Notes for the long story.

<10> **Example.** Suppose $g(z) = a_0 + \cdots + a_n z^n \in \mathcal{P}_n$ with $a_j \neq 0$ for all j. I claim that, for $0 \leq k \leq n-2$,

<11> $$a_{k+1}^2 \geq \frac{(n-k)(k+2)}{(n-k-1)(k+1)} a_k a_{k+2} > a_k a_{k+2}.$$

Remark. The inequality $a_{k+1}^2 > a_k a_{k+2}$ also holds for $k \in \{-1, n-1\}$ if $a_{-1} = a_{n+1} = 0$.

The k-fold derivative of g is a polynomial of degree $n-k$,

$$D^k g(z) = b_0 + b_1 z + \cdots + b_{n-k} z^{n-k} \qquad \text{where } b_j := a_{k+j}(k+j)_k,$$

which belongs to $\mathcal{P}_{n-k}$. Focus your attention on the terms involving b_0, b_1, b_2, that is, the terms involving a_k, a_{k+1}, a_{k+2}. The operator τ then transforms $D^k g$ to

$$b_0 z^{n-k} + b_1 z^{n-k-1} + b_2 z^{n-k-2} + \cdots + b_{n-k} \in \mathcal{P}_{n-k}.$$

Another $\ell := n-k-2$ differentiations leaves $D^\ell \tau D^k g(z)$ as a quadratic in $\mathcal{P}_2$:

$$\begin{aligned} &b_0 z^2(\ell+2)_\ell + b_1 z(\ell+1)_\ell + b_2(\ell)_\ell \\ &\quad = (\ell+2)_\ell (k)_k a_k z^2 + (\ell+1)_\ell (k+1)_k a_{k+1} z + +(\ell)_\ell (k+2)_k a_{k+2} \\ &\quad = \ell! k! (Az^2 + Bz + C) \end{aligned}$$

where

$$\begin{aligned} A &= (\ell+1)(\ell+2) \times a_k/(1\times 2) = \tfrac{1}{2}(n-k-1)(n-k) \times a_k \\ B &= (\ell+1)(k+1) \times a_{k+1} = (n-k-1)(k+1) \times a_{k+1} \\ C &= (k+1)(k+2) \times a_{k+1}/(1\times 2). \end{aligned}$$

By Lemma <8>, this quadratic has real roots, which happens if and only if $B^2 \geq 4AC$, which is equivalent to the first inequality in <11>.

4.4 Sampling and the Poisson-Binomial

The main result in this Section is a tiny variation on an amazing result due to Vatutin and Mikhailov (1983).

<12> **Theorem.** *Suppose U is a finite set of size N and F and G are independent random members of 2^U, the collection of all 2^N subsets of U. Suppose also that $|F|$, the size of F, has a Poisson-Binomial distribution and, for a fixed m,*

$$\mathbb{P}\{G = J\} = 1/\binom{N}{m} \qquad \text{for each } J \in 2^U \text{ of size } m.$$

Then the size $X = |F \cap G|$ also has a Poisson-Binomial distribution.

Remark. Vatutin and Mikhailov (1983) actually assumed (in my notation) that $F_1, \dots, F_n$ were independent random subsets of U of sizes $|F_j| = N_j = N - s_j$. They showed that $N - |\cap_j F_j|$ has a PBin distribution, from which they deduced several interesting limit theorems.

The VM result follows by repeated appeals to Theorem <12> starting from the fact that $\mathbb{P}z^{|F_1|} = z^{N_1}$, a polynomial with only real roots. This fact was my motivation for the modification.

My modification of the theorem allows different ways to generate F_1. For example, it could be degenerate at a fixed subset of size N_1 or it could be generated by Poisson sampling, whereby the points of U are included independently in F_1 with $\mathbb{P}\{u \in F_1\} = \pi_u$ for a fixed set of values $\{\pi_u : u \in U\}$. It could even be generated by some combination of the two methods, whereby some subset U_0 is always included in F_1 and other points are included randomly. I don't have any striking examples in mind.

As an immediate consequence of Theorem <12> we learn that the hypergeometric a is special case of the Poisson-Binomial. The case where the sample size equals 2 gave me some insights into the subtlety of the Theorm.

<13> **Example.** If a sample of size n is taken without replacement from an urn (the set U) containing R red balls and $B = N - R$ black balls then W, the number of red balls in the sample, has distribution $\text{HYPER}(n, R, B)$, that is,

$$a_k := \mathbb{P}\{W = k\} = \binom{R}{k}\binom{B}{n-k}\Big/\binom{N}{n} \qquad \text{for } k = 0, 1, \ldots, n.$$

Here F_1 is taken to be degenerate at the set of all red balls and F_2 is a random subset of size n.

Remark. The properties of binomial coefficients ensure that $\mathbb{P}\{W = k\}$ is zero if $k > n$ or $n - k > B$.

If the sample for F_2 is selected one ball at a time then W can be written as a sum $\xi_1 + \cdots + \xi_n$, with $\xi_i = 1$ if the ith ball in the sample is red and $\xi_i = 0$ if it is black. Marginally speaking, each ξ_i has a $\text{Ber}(p)$ distribution, with $p = R/N$, but they are not independent, as would be required to show directly that W has a PBin distribution. Nevertheless, Theorem <12> tells us that there must be another decomposition of the distribution of W that corresponds to a sum of independent Bernoulli variables.

Consider the simplest case, where $n = 2$. First note that,

$$\mathbb{P}\{\xi_2 = 1 \mid \xi_1 = y_1\} = \begin{cases} (R-1)/(N-1) & \text{if } y_1 = 1 \\ R/(N-1) & \text{if } y_1 = 0 \end{cases}.$$

Clearly ξ_1 and ξ_2 are not independent. If both R and B were large then ξ_1 and ξ_2 would be close to independent, in some sense, but for an extreme case like $R = 1$ and $B = 2$ they would be far from independent.

The distribution of W is determined by the joint distribution of ξ_1 and ξ_2: for $y_1, y_2 \in \{0, 1\}$,

$$\mathbb{P}\{\xi_1 = y_1, \xi_2 = y_2\} = \frac{1}{N(N-1)} \begin{cases} R(R-1) & \text{if } y_1 = y_2 = 1 \\ B(B-1) & \text{if } y_1 = y_2 = 0 \\ RB & \text{if } y_1 + y_2 = 1 \end{cases}.$$

It follows that $h(z) = a_0 + a_1 z + a_2 z^2$, where

$$a_0 = \frac{B(B-1)}{N(N-1)}, \qquad a_1 = \frac{2RB}{N(N-1)}, \qquad a_2 = \frac{R(R-1)}{N(N-1)}.$$

The quadratic h has two real roots, $\{r_1, r_2\}$ given by

$$\frac{-RB \pm \sqrt{R^2B^2 - BR(B-1)(R-1)}}{R(R-1)} = \frac{-RB \pm \sqrt{RB(N-1)}}{R(R-1)},$$

both of them negative. Thus $W \sim \text{PB}\textsc{in}(p_1, p_2)$ for some values p_1, p_2 in $[0, 1]$.

By the method from Example <2> we can take $p_1 = \theta - t$ and $p_2 = \theta + t$ where $\theta = \frac{1}{2}\mathbb{P}W = R/N$ and

$$t^2 = \theta^2 - a_2 = \frac{R^2}{N^2} - \frac{R(R-1)}{N(N-1)} = \frac{RB}{N^2(N-1)},$$

which gives

$$\{p_1, p_2\} = \frac{R}{N} \pm \sqrt{\frac{RB}{N^2(N-1)}}.$$

Remark. Of course you should now feel obliged to check that the values $\{p_1, p_2\}$ do correspond to $\{(1-r_1)^{-1}, (1-r_2)^{-1}\}$, as promised by Lemma <1>. (I found that it cost me a good bit of algebra.)

For example, if $R = B = N/2$ then

$$\{p_1, p_2\} = \frac{1}{2} \pm \frac{1}{2\sqrt{N-1}}.$$

In this case, for sampling with replacement we would toss two fair coins and count the number of heads. For sampling without replacement we would toss two coins, one slightly biased towards red and the other equally biased towards black.

Proof (of Theorem <12>). To simplify notation I'll assume that $\mathbb{P} = P \otimes Q$ on the product space $2^U \times 2^U$, with (F, G) as the generic point. Thus, under $\mathbb{P}$, the coordinate map F has distribution P and G has distribution Q. Under Q, the distribution of G^c is uniform over the collection of all subsets of U of size $s = N - m$.

Here is the main idea. By assumption $|F| \sim \text{PBIN}(p_1, \dots, p_t)$ where, without loss of generality, we may assume $p_j > 0$ for each j. Define

$$h(z) = \mathbb{P}z^{|F|} = \sum\nolimits_{j=0}^{t} \mathbb{P}\{|F| = j\}z^j \quad ,$$
$$f(z) = \mathbb{P}z^X.$$

We need to show that f has only real roots. This fact will follow from the representation

<14> $$f_1(z) := f(1+z) = C_m \sum\nolimits_{k=0}^{t} a_k (N-k)_s z^k$$

for some positive constant C_m, where the a_j's are defined by

$$h(1+z) = \sum\nolimits_{j=0}^{t} a_j z^j.$$

If we can also show that $a_k > 0$ for all k then the result from Example <9> will imply that $f_1 \in \mathcal{P}$. As the roots of f are just the roots of f_1 shifted up by 1, it will then follow that $f \in \mathcal{P}$ so that X also has a PBIN distribution.

Let me start by proving the strict positivity of the a_j's and then turn to the representation <14>.

If L of the p_j's were equal to 1 then the probabilities $\mathbb{P}\{|F| = j\}$ would equal 0 for $j < L$. However we always have $\mathbb{P}\{|F| = t\} = \prod_{j=1}^{t} p_j > 0$. The $1+z$ takes care of any difficulties with the coefficients of h:

$$\begin{aligned} h(1+z) &= \sum\nolimits_{j=0}^{t} \mathbb{P}\{|F| = j\} \left(\sum\nolimits_{k=0}^{t} \binom{j}{k} z^k \right) \\ &= \sum\nolimits_{k=0}^{t} z^k \left(\sum\nolimits_{j=0}^{t} \mathbb{P}\{|F| = j\} \binom{j}{k} \right) \end{aligned}$$

Thus $a_k = \sum_{j=0}^{t} \mathbb{P}\{|F| = j\}\binom{j}{k} \geq \binom{t}{k}\mathbb{P}\{|F| = t\} > 0$ for $0 \leq k \leq t$. We could also expand $h(1+z)$ as

$$\mathbb{P}(1+z)^{|F|} = \sum\nolimits_{k=0}^{t} \mathbb{P}\binom{|F|}{k} z^k,$$

which shows that $a_k = P\binom{|F|}{k}$.

Now comes the really clever part. Because $X \leq |F| \leq t$ we have the expansion

$$f(1+z) = \mathbb{P}(1+z)^X = \sum\nolimits_{k=0}^{t} \mathbb{P}\binom{X}{k} z^k.$$

The binomial coefficient $\binom{X}{k}$ counts the number of subsets of $F \cap G$ of size k, which can written as a sum of $\binom{N}{k}$ indicator functions:

$$\binom{X}{k} = \sum\nolimits_{|J|=k} \{J \subseteq F \cap G\} = \sum\nolimits_{|J|=k} \{J \subseteq F\}\{J \subseteq G\}.$$

Take expectation with respect to Q, the marginal distribution of G:

$$Q\binom{X}{k} = \sum\nolimits_{|J|=k} \{J \subseteq F\} Q\{J \subseteq G\}.$$

The event $\{J \subseteq G\}$ is the same as $\{J^c \supseteq G^c\}$. Remember that the distribution of G^c is uniform over the collection of all subsets of U of size $s = N - m$. Thus

$$Q\{J \subseteq G\} = \binom{|J^c|}{s} \Big/ \binom{N}{s} = C_m \times (N-k)_s \qquad \text{with } C_m^{-1} = s!\tbinom{N}{s}.$$

Now take expectation with respect to P.

$$\begin{aligned} \mathbb{P}\binom{X}{k} &= C_m \times (N-k)_s P \sum\nolimits_{|J|=k} \{J \subseteq F\} \\ &= C_m \times (N-k)_s P \binom{|F|}{k} \\ &= C_m \times (N-k)_s \times a_k. \end{aligned}$$

Summation over k then gives the asserted representation <14>.

4.5 Ratios and mode(s)

You could regard this Section as mainly a warm-up for an analogous, but more delicate, argument for the general Poisson-Binomial.

Recall that if a random variable Y takes only integer values then $\text{MODE}(Y)$ can be defined as $\{k : \mathbb{P}\{Y = k\} = M\}$ where $M = \max_k \mathbb{P}\{Y = k\}$. If $\text{MODE}(Y)$ is a singleton, $\{\nu\}$, then ν is called 'the mode' of the distribution; otherwise the members of $\text{MODE}(Y)$ are often referred to as modes.

For example, if $S \sim \text{PBIN}(p_1, \dots, p_n)$ with $0 < p_j < 1$ for all j and $a_j = \mathbb{P}\{S = j\}$ then inequality <11> (with k replaced by $k-1$) implies

$$a_k/a_{k-1} > a_{k+1}/a_k \qquad \text{for } 0 < k < n.$$

(For a more direct coupling proof see Problem [1].) That is, the sequence a_k/a_{k-1} decreases as k increases. If K denotes the largest value for which

$a_K/a_{K-1} > 1$ then we have $a_{k-1} < a_k$ for $k \le K$. It might happen that $a_{K+1}/a_K = 1$ and then $a_k > a_{k+1}$ for $k \ge K+1$, in which case $\text{MODE}(S) = \{K, K+1\}$. If $a_{K+1}/a_K < 1$ then K is the unique mode.

It takes a fair amount of work to determine the value K for the general Poisson-Binomial because the distribution of S involves an intimidating multinomial in the components of $\mathbf{p} = (p_1, \dots, p_n)$:

<15> $$f(k) := f(k, \mathbf{p}) := \mathbb{P}\{S = k\} = \sum\nolimits_{|J|=k} \left(\prod\nolimits_{j\in J} p_j\right) \left(\prod\nolimits_{j\notin J} (1-p_j)\right),$$

where $\sum_{|J|=k}$ denotes summation over all $\binom{n}{k}$ subsets J of $[n] := \{1, 2, \dots, n\}$ of size k. The probabilities are strictly positive for $k = 0, 1, \dots, n$.

Remark. The function $f(k, \mathbf{p})$ is symmetric in the variables $p_1, \dots, p_n$. Its value does not depend on the order in which the p_j's are listed in the vector $\mathbf{p}$.

The task is much easier for the special case of the Binomial.

<16> **Example.** If $X \sim \text{Bin}(n, p)$ then

$$b(j) := \mathbb{P}\{X = j\} = \binom{n}{j} p^j q^{n-j} \qquad \text{for } j = 0, 1, \dots, n.$$

Here I again write q for $1-p$ and assume that $0 < p < 1$. As in the general PBin case, the ratio of consecutive $b(j)$'s is a strictly decreasing function of j with

$$\frac{b(j+1)}{b(j)} = \frac{n!(n-j)!j!p^{j+1}q^{n-j-1}}{(j+1)!(n-j-1)!n!p^j q^{n-j}} = \frac{(n-j)p}{(j+1)q} \qquad \text{for } 0 \le j < n.$$

It follows that

$$b(j+1) \gtreqless b(j) \text{ if } j+1 \lesseqgtr np+p \qquad \text{for } 0 \le j < n.$$

If $K := \lceil np+p \rceil - 1$, the unique integer for which $K < np+p \le K+1$, the pattern of inequalities is: **For** $K < np+p = K+1$**,**

<17> $$b(0) < b(1) < \cdots < b(K) = b(K+1) > \cdots > b(n-1) > b(n).$$

For $K < np+p < K+1$**,**

<18> $$b(0) < b(1) < \cdots < b(K) > b(K+1) > \cdots > b(n-1) > b(n).$$

That is, K and $K+1$ are both modes if $np+p$ is an integer, and otherwise K is the unique mode. In particular if $p = 1/2$ and n is even then $n/2$ is the unique mode but if n is odd then the modes are $(n-1)/2$ and $(n+1)/2$.

4.6 Mode(s) of the Poisson-Binomial

Throughout this section suppose $S = X_1 + \cdots + X_n \sim \text{PB}\textsc{in}(p_1, \ldots, p_n)$ and $f(k) := f(k, \mathbf{p}) := \mathbb{P}\{S = k\}$. To avoid tedious special cases, for most of the Section assume $0 < p_j < 1$ for all j and the p_j's are not all equal to $\overline{p}$. Only at the end of the Section (in subsection 4.6.1) will I discuss what happens if $p_j = 0$ or $p_j = 1$ for some j. The main result is given by the following Theorem, that shows $\mathbb{P}\{S = k\}$ is maximized by k close to $\mathbb{P}S$.

<19> **Theorem.** *Suppose $S \sim \text{PB}\textsc{in}(p_1, \ldots, p_n)$ with*

<20> $$0 < p_{min} := \min_{1 \le j \le n} p_j < \max_{1 \le j \le n} p_j =: p_{max} < 1.$$

Define $L := \lfloor p_{min} + \mathbb{P}S \rfloor$ and $K := \lceil p_{max} + \mathbb{P}S \rceil - 1$. Then there are only two possibilities:

(i) $L = K$ and

$$f(0) < \cdots < f(K) > f(K+1) > \cdots > f(n),$$

which makes K the unique mode;

(ii) $L + 1 = K$ and

$$f(0) < \cdots < f(L) \quad \text{AND} \quad f(K) > f(K+1) > \cdots > f(n),$$

with no assertion regarding the comparison between $f(L)$ and $f(K)$. That is, MODE(S) *is either $\{K\}$ or $\{K+1\}$ or $\{K, K+1\}$.*

Remark. Compare the upper tails with the situation where $X \sim \textsc{Bin}(n, \overline{p})$ with $b(j) = \mathbb{P}\{X = j\}$.

$$\begin{array}{ll} f(K_{pb}) > f(K_{pb}+1) > \ldots f(n) & \text{if } K_{pb} < p_{max} + \mathbb{P}S \le K_{pb} + 1 \\ b(K_b) = b(K_b+1) > \cdots > b(n) & \text{if } K_b < \overline{p} + \mathbb{P}X = K_b + 1, \\ b(K_b) > b(K_b+1) > \cdots > b(n) & \text{if } K_b < \overline{p} + \mathbb{P}X < K_b + 1. \end{array}$$

The gap between p_{min} and p_{max} has subtle consequences. In the next Section it will be most helpful not having to worry about a possible equality between $\mathbb{P}\{S = K_{pb}\}$ and $\mathbb{P}\{S = K_{pb} + 1\}$. Strict inequalities are sensitive to very small changes.

The proof of the Theorem depends on two Lemmas, which are stated next. Then comes the proof of the Theorem, followed by the proofs of the Lemmas.

<21> **Lemma.** *Under assumption <20>, for each j the ratio $f(k, \mathbf{p})/f(k-1, \mathbf{p})$ is strictly increasing in p_j, for fixed values of the other p_i's.*

<22> **Lemma.** *Under assumption <20>, $f(k, \mathbf{p}) > f(k-1, \mathbf{p})$ for each k in the range $1 \le k \le p_{min} + \mathbb{P}S$.*

Proof (of Theorem <19>). From Lemma <22> and the definition of L as the largest integer that is $\le p_{min} + \mathbb{P}S$ we have

$$f(0) < f(1) < \cdots < f(L).$$

For the other part of the range note that $n - S$ has a $\text{PB}\textsc{in}(n, q_1, \ldots, q_n)$ distribution for $q_i = 1 - p_i$ and $\mathbb{P}\{n - S = k\} = f(n-k)$. Also

$$\min_i q_i + \mathbb{P}(n - S) = 1 - p_{max} + n - \mathbb{P}S \ge (n+1) - (K+1).$$

Again from Lemma <22> it follows that

$$\mathbb{P}\{n - S = n - K\} > \mathbb{P}\{n - S = n - K - 1\} > \cdots > \mathbb{P}\{n - S = 0\},$$

which implies $f(K) > f(K+1) > \cdots > f(n)$.

Relationship between L and K The integers L and K are defined by the inequalities

$$L \le p_{min} + \mathbb{P}S < L + 1 \quad \text{AND} \quad K < p_{max} + \mathbb{P}S \le K + 1.$$

The inequalities $0 < p_{min} < p_{max} < 1$ give

$$L \le p_{min} + \mathbb{P}S < p_{max} + \mathbb{P}S \le K + 1.$$

The strict inequality then ensures that $L \le K$. Similarly,

$$p_{max} + \mathbb{P}S < p_{min} + \mathbb{P}S + 1 < (L+1) + 1,$$

which implies $L + 2 \ge \lceil p_{max} + \mathbb{P}S \rceil = K + 1$, so that $L + 1 \ge K$.

Proof (of Lemma <21>). The assertion is that $f(k, \mathbf{p})/f(k-1, \mathbf{p})$ is strictly increasing in each p_j. For notational simplicity consider only the case where $j = 1$.

Define $S_0 = \sum_{i=2}^n X_i$ and $g(k) := g(k, p_2, \ldots, p_n) = \mathbb{P}\{S_0 = k\}$, for which $g(k) > 0$ for $0 \le k \le n-1$. By the weak form of Newton's inequality from Example <10>,

<23> $$g^2(k-1) > g(k)g(k-2) \qquad \text{for } 1 \le k \le n.$$

(The inequalities for $k = 1$ and $k = n$ hold for the trivial reason that $g(-1) = g(n) = 0$.) Decompose according to the value of X_1:

$$\frac{f(k,\mathbf{p})}{f(k-1,\mathbf{p})} = \frac{(1-p_1)g(k) + p_1 g(k-1)}{(1-p_1)g(k-1) + p_1 g(k-2)} \qquad \text{for } 1 \le k \le n.$$

The denominator is strictly positive for $0 < p_1 < 1$. The partial derivative with respect to p_1 equals

$$\Big(f(k-1,\mathbf{p})\,[g(k-1) - g(k)] - f(k,\mathbf{p})\,[g(k-2) - g(k-1)]\Big)/f(k-1)^2$$

The numerator is linear in p_1 and approaches

$$\begin{aligned}
&g(k-1)\,[g(k-1) - g(k)] - g(k)\,[g(k-2) - g(k-1)] \\
&\quad = g^2(k-1) - g(k)g(k-2) \qquad \text{as } p_1 \to 0, \\
&g(k-2)\,[g(k-1) - g(k)] - g(k-1)\,[g(k-2) - g(k-1)] \\
&\quad = g^2(k-1) - g(k)g(k-2) \qquad \text{as } p_1 \to 1.
\end{aligned}$$

Thus the numerator is constant at the value $g^2(k-1) - g(k)g(k-2) > 0$. The p_1 derivative of $f(k,\mathbf{p})/f(k-1,\mathbf{p})$ is strictly positive.

Now comes the real work. Lemma <21> will play a crucial role.

Proof (of Lemma <22>). The assertion is: if $1 \le k \le p_{min} + \mathbb{P}S$ then $f(k) > f(k-1)$.

Define $p_0 = p_{min}$, so that $\sum_{i=0}^n p_i \ge k$. Choose $X_0 \sim \text{Ber}(p_0)$ independently of the other X_i's. Define $S^* = X_0 + S$ and

$$T_i^* = S^* - X_i = \sum\nolimits_{j=0}^{n} \{j \ne i\} X_j \qquad \text{for } 0 \le i \le n.$$

Note that $T_0^* = S$ and $T_i^* \sim \text{PBin}(p_1, \ldots, p_{i-1}, p_0, p_{i+1}, \ldots, p_n)$. Symmetry of PBin in its arguments lets me write the distribution in this way, which emphasizes that the change in distribution from T_i^* to S is caused by an increase from p_0 to p_i. Writing $f_i(j)$ for $\mathbb{P}\{T_i^* = j\}$ we then have, via Lemma <21>, the inequalities

<24> $$f(k)/f(k-1) = f_0(k)/f_0(k-1) \ge f_i(k)/f_i(k-1) \qquad \text{for } 1 \le i \le n,$$

with strict inequality if $p_0 < p_i$.

You should check the following equalities,

$$\begin{aligned}
\{S^* = k\} &= X_i\{T_i^* = k-1\} + (1 - X_i)\{T_i^* = k\} \qquad \text{for each } i, \\
k\{S^* = k\} &= \sum\nolimits_{i=0}^{n} X_i\{T_i^* = k-1\},
\end{aligned}$$

by calculating the values on both sides of each equality when $S^* = j$, for $0 \le j \le n+1$. (For example, the second equality corresponds to the fact that $S^* = k$ iff exactly k of the X_i's equal 1 and $1 + T_i^* = k$ for those i.) Then take expectations, using independence of X_i and T_i^*, to deduce that

<25> $$\mathbb{P}\{S^* = k\} = p_i f_i(k-1) + q_i f_i(k) \qquad \text{for each } i,$$

<26> $$k\mathbb{P}\{S^* = k\} = \sum\nolimits_{i=0}^{n} p_i f_i(k-1).$$

Combine. From the inequality $k \le \mathbb{P}S + p_{min} = \sum_{j=0}^{n} p_j$ we have

$$\begin{aligned}\sum\nolimits_{i=0}^{n} p_i f_i(k-1) &= k\mathbb{P}\{S^* = k\} \\ &\le \left(\sum\nolimits_{i=0}^{n} p_i\right)\mathbb{P}\{S^* = k\} \qquad \text{from} <26> \\ &= \sum\nolimits_{i=0}^{n} p_i \left(p_i f_i(k-1) + q_i f_i(k)\right) \qquad \text{by } <25>.\end{aligned}$$

Subtract.

$$\begin{aligned}0 &\le \sum\nolimits_{i=0}^{n} p_i \Big[p_i f_i(k-1) + q_i f_i(k) - (p_i + q_i) f_i(k-1)\Big] \\ &= \sum\nolimits_{i=0}^{n} p_i q_i \Big[f_i(k) - f_i(k-1)\Big] \\ &= \sum\nolimits_{i=0}^{n} p_i q_i f_i(k-1)\Big[f_i(k)/f_i(k-1) - 1\Big] \\ &\le \sum\nolimits_{i=0}^{n} p_i q_i f_i(k-1)\Big[f(k)/f(k-1) - 1\Big] \qquad \text{by } <24>.\end{aligned}$$

In fact the last inequality must be strict because <24> is strict for at least one i (such as the i for which $p_{max} = p_i$). That is,

$$\Big[f(k)/f(k-1) - 1\Big] \times (\text{something} > 0) > 0.$$

It follows that $f(k)/f(k-1) > 1$ and $f(k) > f(k-1)$.

Remark. It is worth noting that strict inequality in <24> was only needed to deduce the strict inequality for the final conclusion. If <24> might not be strict, as in the case where $p_{min} = p_{max}$, the Lemma would still apply but with the conclusion weakened to $f(k) \ge f(k-1)$.

4.6.1 Zeros and ones

Theorem <19> assumed that

$$0 < p_{min} := \min\{p_j : j \in [n]\} < \max\{p_j : j \in [n]\} =: p_{max} < 1.$$

What annoyances could result from the violation of this assumption?

To avoid total triviality, I still assume that

<27> $$0 < \overline{p} := n^{-1} \sum\nolimits_{i=1}^{n} p_i < 1.$$

Define

$$\begin{aligned} I_0 &= \{i : p_i = 0\} \qquad \text{with size } n_0 \\ M &= \{i : 0 < p_i < 1\} \qquad \text{with size } m \\ I_1 &= \{i : p_i = 1\} \qquad \text{with size } n_1. \end{aligned}$$

If $m = 0$ then $\mathbb{P}\{S = n_1\} = 1$, so that the mean, mode, and median of S are all equal to n_1.

If $m > 0$ things are little more interesting/complicated. Write $\boldsymbol{\theta}$ for the vectors of length m obtained by deleting all the 0's and 1's from $\mathbf{p}$. Define

$$\theta_{min} := \min \boldsymbol{\theta}, \quad \overline{\theta} = m^{-1} \sum\nolimits_{j \in M} \theta_j, \quad \theta_{max} := \max \boldsymbol{\theta}.$$

The sum $T := \sum_{j \in M} X_j$ has a $\text{PB}\textsc{in}(\boldsymbol{\theta})$ distribution and $S = T + n_1$ so that

$$\begin{aligned} \mathbb{P}S &= n_1 + \mathbb{P}T = n_1 + m\overline{\theta} \\ \mathbb{P}\{S = n_1 + j\} &= f(n_1 + j) = g(j) := \mathbb{P}\{T = j\} \qquad \text{for } 0 \le j \le m. \end{aligned}$$

The behavior of S is just the behavior of T translated by m to the right. In particular,

$$\textsc{mode}(S) = \{n_1 + j : j \in \textsc{mode}(T)\}.$$

It is now just a matter of applying the results from inequalities <17> and <18> or Theorem <19>, with minor notational changes.

In both of the following cases $\mathbb{P}\{S = k\} = 0$ for integers k that lie outside the range $\{n_1 + j : 0 \le j \le m\}$.

Case: $m \ge 1$ and $\theta_j = \overline{\theta}$ for all j Here $T \sim \textsc{Bin}(m, \overline{\theta})$. For $K = \lceil \overline{\theta} + \mathbb{P}S \rceil - 1 = n_1 + \lceil \overline{\theta} + \mathbb{P}T \rceil - 1$, the integer defined by the inequalities $K < \overline{\theta} + \mathbb{P}S \le K + 1$, we have two possibilities.

(i) If $\overline{\theta} + \mathbb{P}S$ is an integer then $K = \overline{\theta} + \mathbb{P}S$ and

$$\begin{aligned} \mathbb{P}\{S = n_1\} < \ldots < \mathbb{P}\{S = K\} \\ = \mathbb{P}\{S = K + 1\} > \ldots \mathbb{P}\{S = n_1 + m\}. \end{aligned}$$

(ii) If $\overline{\theta} + \mathbb{P}S$ is not an integer then $K < \overline{\theta} + \mathbb{P}S < K + 1$ and

$$\begin{aligned}\mathbb{P}\{S = n_1\} < \ldots < \mathbb{P}\{S = K\}\\ > \mathbb{P}\{S = K+1\} > \cdots > \mathbb{P}\{S = n_1 + m\}.\end{aligned}$$

Remark. For $m = 1$ we have $T \sim \text{Ber}(\overline{\theta})$ and the results reduce to

$$\mathbb{P}\{S = n_1\} = 1 - \overline{\theta} \quad \gtreqless \quad \mathbb{P}\{S = n_1 + 1\} = \overline{\theta} \qquad \text{if } \overline{\theta} \quad \lesseqgtr \quad 1/2,$$

as can be verified by direct calculation.

Case: $m \geq 2$ **and** $\theta_{min} < \theta_{max}$ For integers L and K defined by

$$L \leq \theta_{min} + \mathbb{P}S < L + 1 \quad \text{AND} \quad K < \theta_{max} + \mathbb{P}S \leq K + 1,$$

we have only two possibilities.

(i) $L = K$ and

$$\begin{aligned}\mathbb{P}\{S = n_1\} < \ldots < \mathbb{P}\{S = K\}\\ > \mathbb{P}\{S = K+1\} > \cdots > \mathbb{P}\{S = n_1 + m\}\end{aligned}$$

(ii) $L + 1 = K$ and

$$\begin{aligned}&\mathbb{P}\{S = n_1\} < \cdots < \mathbb{P}\{S = K - 1\}\\ &\mathbb{P}\{S = K\} > \mathbb{P}\{S = K+1\} > \cdots > \mathbb{P}\{S = n_1 + m\},\end{aligned}$$

with no assertion regarding the comparison between $\mathbb{P}\{S = K - 1\}$ and $\mathbb{P}\{S = K\}$.

4.7 Binomial tails versus Poisson-Binomial tails

Suppose $S \sim \text{PBin}(\mathbf{p})$ for $\mathbf{p} = (p_1, \ldots, p_n) \in [0, 1]^n$. Elementary calculations show that

$$\mathbb{P}S = n\overline{p} := \sum\nolimits_{j \leq n} p_j \quad \text{AND} \quad \text{var}(S) = \sum\nolimits_{j \leq n} p_j(1 - p_j).$$

By Jensen's inequality, for a given $\overline{p}$ the variance of S is maximized when $p_j = \overline{p}$ for all j. That is, $\text{var}(S) \leq \text{var}(X)$ where $X \sim \text{Bin}(n, \overline{p})$, an inequality that adds support to the idea that the distribution of X is more spread out than the distribution of S. The following Theorem adds rigor to this intuition.

<28> **Theorem.** *Suppose* $S \sim \text{PBin}(\mathbf{p})$ *with* $\mathbf{p} \in [0,1]^n$*. and* $\mu := \mathbb{P}S \in (0,n)$*. Let* k *be an integer with* $n \geq k \geq 1+\mu$*. Then*

(i) $\mathbb{P}\{S \geq k\} \leq \mathbb{P}\{X \geq k\}$ *where* $X \sim \text{Bin}(n, \overline{p})$ *with* $\overline{p} = \mu/n$

(ii) The inequality in (i) is strict if not all p_i*'s are equal to* $\overline{p}$*.*

The Theorem holds for trivial reasons if $p_j \in \{0,1\}$ for all j: in the notation of subsection 4.6.1, we would have $S = n_1 = \mu$ with probability 1, so that $\mathbb{P}\{S \geq 1+\mu\} = 0$. Enough said. For the rest of this Section, I assume that $\{j : 0 < p_j < 1\} \neq \emptyset$.

The main idea in the proof is easiest to understand when $n = 2$ and $0 < \mu \leq 1$, so that the only k with $1+\mu \leq k \leq 2$ is $k = 2$. If $p_1+r = \overline{p} = p_2-r$ for some $r \neq 0$ then we have

$$\mathbb{P}\{S=2\} = p_1p_2 = \overline{p}^2 - r^2 < \overline{p}^2 = \mathbb{P}\{X=2\}.$$

The general case, with $n \geq 3$, involves the same idea, applied repeatedly, transforming the initial $\mathbf{p}$ through a sequence of vectors $\boldsymbol{a}^{(1)}, \boldsymbol{a}^{(2)}, \dots$ in $[0,1]^n$ that converges to $\overline{p}\mathbb{1}$, the vector of probabilities for which $\text{PBin}(\overline{p}\mathbb{1})$ reduces to the $\text{Bin}(n,\overline{p})$. (It can also happen that some $\boldsymbol{a}^{(j)}$ is equal to $\overline{p}\mathbb{1}$.) At each iteration two component $a_\alpha^{(j)}$ and $\boldsymbol{a}_\beta^{(j)}$ are replaced by $(a_\alpha^{(j)} + a_\beta^{(j)})/2$, for a pair of distinct values α, β that might depend on what has happened up to step j. I denote this operation by the procedure $\text{smooth}_{\alpha,\beta}$. It keeps the sum of the components $(a_1^{(j)}, \dots, a_n^{(j)})$ of $\boldsymbol{a}^{(j)}$ equal to μ. For the specification of α and β it helps to have the components of $\boldsymbol{a}^{(j)}$ sorted into increasing order, which I denote by the procedure sort. In short, the iteration scheme is described by

$$
\begin{aligned}
&\boldsymbol{a}^{(0)} \leftarrow \text{sort}(\mathbf{p})\\
&\text{while } (\boldsymbol{a}^{(j)} \neq \overline{p}\mathbb{1})\\
&\qquad \boldsymbol{a}^* \leftarrow \text{smooth}_{\alpha,\beta}\left(\boldsymbol{a}^{(j)}\right) \qquad \text{with } (\alpha,\beta) \text{ depending on past?}\\
&\qquad \boldsymbol{a}^{(j+1)} \leftarrow \text{sort}\left(\boldsymbol{a}^*\right)
\end{aligned}
$$

For the sake of a clean notation, it helps to think of the X_j's as the coordinate maps on $\{0,1\}^n$ with $\mathbb{P}_{\boldsymbol{a}} = \otimes_{j\in[n]}\text{Ber}(a_j)$, the product of the the $\text{Ber}(a_j)$ measures on $\{0,1\}$. That is, for each $x = (x_1,\dots,x_n)$ in $\{0,1\}^n$ we have

$$\mathbb{P}_{\boldsymbol{a}}\{x\} = \prod\nolimits_{j=1}^n \Big(a_j\{x_j=1\} + (1-a_j)\{x_j=0\}\Big).$$

With that interpretation, the random variable S is always the same function, $\sum_{j\le n} X_j = \sum_{j\le n} x_j$, on $\{0,1\}^n$; only the underlying $\mathbb{P}_{\boldsymbol{a}}$ changes when $\boldsymbol{a}$ changes.

The following Lemma captures the important consequences of replacing a generic $\boldsymbol{a}$ by $\boldsymbol{a}^* = \textsc{smooth}_{\alpha,\beta}(\boldsymbol{a})$.

<29> **Lemma.** *Suppose* $\boldsymbol{a} = (a_1,\dots,a_n)$ *with* $a_1 \le \dots \le a_n$ *and at least one* a_j *satisfying* $0 < a_j < 1$. *For a pair* $\alpha,\beta \in [n]$ *with* $1 \le \alpha < \beta \le n$ *define* $\boldsymbol{a}^*$ *by*

$$a_j^* = \begin{cases} t & \text{if } j \in \{\alpha,\beta\} \\ a_j & \text{otherwise} \end{cases} \qquad \text{where } t := (a_\alpha + a_\beta)/2 \quad .$$

Also define $S_{\alpha,\beta} = \sum_{j\in[n]\setminus\{\alpha,\beta\}} X_j = S - X_\alpha - X_\beta$. *Then:*

(i) $\mathbb{P}_{\boldsymbol{a}^*}\{S \ge k\} - \mathbb{P}_{\boldsymbol{a}}\{S \ge k\} = r^2\Delta$ *where* $r := (a_\beta - a_\alpha)/2$ *and*

$$\Delta := \mathbb{P}_{\boldsymbol{a}}\{S_{\alpha,\beta} = k-2\} - \mathbb{P}_{\boldsymbol{a}}\{S_{\alpha,\beta} = k-1\}.$$

(ii) $\sum_{i\le n}(a_i - \overline{p})^2 = \sum_{i\le n}(a_i^* - \overline{p})^2 + 2r^2$.

(iii) if $\beta = n$ *and* $k \ge 1 + \mathbb{P}_{\boldsymbol{a}} S$ *then* $\Delta \ge 0$, *which ensures that the inequality* $\mathbb{P}_{\boldsymbol{a}^*}\{S \ge k\} \ge \mathbb{P}_{\boldsymbol{a}}\{S \ge k\}$ *holds.*

Proof. Decompose the event $\{S \ge k\}$ as

$$X_\alpha X_\beta\{S_{\alpha,\beta} \ge k-2\} + [X_\alpha(1-X_\beta) + (1-X_\alpha)X_\beta]\{S_{\alpha,\beta} \ge k-1\} \\ + (1-X_a)(1-X_\beta)\{S_{\alpha,\beta} \ge k\}$$

then take expected values under $\mathbb{P}_{\boldsymbol{a}}$ to get

$$\begin{aligned} &\mathbb{P}_{\boldsymbol{a}}\{S \ge k\} \\ &\quad = a_\alpha a_\beta \mathbb{P}_{\boldsymbol{a}}\{S_{\alpha,\beta} \ge k-2\} + \Big(a_\alpha + a_\beta - 2a_\alpha a_\beta\Big)\mathbb{P}_{\boldsymbol{a}}\{S_{\alpha,\beta} \ge k-1\} \\ &\qquad + \Big(1 - a_\alpha - a_\beta + a_\alpha a_\beta\Big)\mathbb{P}_{\boldsymbol{a}}\{S_{\alpha,\beta} \ge k\} \\ &\quad = a_\alpha a_\beta\Big(\mathbb{P}_{\boldsymbol{a}}\{S_{\alpha,\beta} \ge k-2\} - 2\mathbb{P}_{\boldsymbol{a}}\{S_{\alpha,\beta} \ge k-1\} + \mathbb{P}_{\boldsymbol{a}}\{S_{\alpha,\beta} \ge k\}\Big) \\ &\qquad + 2t\Big(\mathbb{P}_{\boldsymbol{a}}\{S_{\alpha,\beta} \ge k-1\} - \mathbb{P}_{\boldsymbol{a}}\{S_{\alpha,\beta} \ge k\}\Big) + \mathbb{P}_{\boldsymbol{a}}\{S_{\alpha,\beta} \ge k\} \\ &\quad = a_\alpha a_\beta\Big(\mathbb{P}_{\boldsymbol{a}}\{S_{\alpha,\beta} = k-2\} - \mathbb{P}_{\boldsymbol{a}}\{S_{\alpha,\beta} = k-1\}\Big) \\ &\qquad + 2t\mathbb{P}_{\boldsymbol{a}}\{S_{\alpha,\beta} = k-1\} + \mathbb{P}_{\boldsymbol{a}}\{S_{\alpha,\beta} \ge k\}. \end{aligned}$$

Under $\mathbb{P}_{\boldsymbol{a}^*}$ the calculation is almost the same because $S_{\alpha,\beta}$ has the same distribution under $\mathbb{P}_{\boldsymbol{a}}$ and $\mathbb{P}_{\boldsymbol{a}^*}$ and $a_\alpha + a_\beta = 2t = \alpha^*_\alpha + a^*_\beta$. The only change is replacement of $a_\alpha a_\beta$ by $a^*_\alpha a^*_\beta$; all other terms stay the same. Many terms cancel when $\mathbb{P}_{\boldsymbol{a}}\{S \geq k\}$ is subtracted from $\mathbb{P}_{\boldsymbol{a}^*}\{S \geq k\}$, leaving

$$\left(a^*_\alpha a^*_\beta - a_\alpha a_\beta\right)\Delta = (t^2 - (t-r)(t+r))\Delta_k = r^2\Delta,$$

the equality asserted by (i).

The argument for (ii) is similar, with a cancellation of all except the terms with $j \in \{\alpha, \beta\}$:

$$\begin{aligned}
&\sum\nolimits_j (a_j - \overline{p})^2 - \sum\nolimits_j (a^*_j - \overline{p})^2 \\
&\quad = (a_\alpha - \overline{p})^2 + (a_\beta - \overline{p})^2 - (a^*_\alpha - \overline{p})^2 - (a^*_\beta - \overline{p})^2 \\
&\quad = (t + r - \overline{p})^2 + (t - r - \overline{p})^2 - 2(t - \overline{p})^2 = 2r^2.
\end{aligned}$$

Assertion (iii) is the trickiest part of the proof. It involves an appeal to the inequalities from Section 4.6.1 applied to $S_{\alpha,\beta}$ instead of S. In general, the vector $\boldsymbol{a}$ looks like

<30> $$a = \boxed{a_1}\ \overset{n_0}{\boxed{0\cdots0}}\ \underbrace{\overset{m}{\boxed{a_{2+n_0}\cdots a_{2+n_0+m-1}}}}_{M}\ \overset{n_1}{\boxed{1\cdots1}}\ \boxed{a_n}\quad .$$

Here $0 < a_j < 1$ for $j \in M$, with $m \geq 1$ by assumption; the coordinate a_1 must equal 0 if $n_0 > 0$; the coordinate a_n must equal 1 if $n_1 > 0$. Define

$$\theta_{min} = \min_{j\in M} a_j = a_{2+n_0} \quad \text{AND} \quad \theta_{max} = \max_{j\in M} a_j = a_{2+n_0+m-1}.$$

Define $K := \lceil \theta_{max} + \mathbb{P}_{\boldsymbol{a}} S_{\alpha,\beta} \rceil - 1$. (Note that if $m = 1$ then $\theta_{max} = \theta_{min} = \overline{\theta} := m^{-1}\sum_{j\in M} a_j$.) Corresponding to both the nontrivial cases from Section 4.6.1 we have

<31> $$\begin{aligned}
\mathbb{P}_{\boldsymbol{a}}\{S_{\alpha,\beta} = K\} \geq \mathbb{P}_{\boldsymbol{a}}\{S_{\alpha,\beta} = K+1\} > \dots \mathbb{P}_{\boldsymbol{a}}\{S_{\alpha,\beta} = n_1 + m\} > 0,\\
\text{and } \mathbb{P}_{\boldsymbol{a}}\{S_{\alpha,\beta} = k\} = 0 \text{ for } k > n_1 + m,
\end{aligned}$$

Moreover, the first inequality in <31> is strict if $m \geq 2$ and $\theta_{min} < \theta_{max}$ or if $a_j = \overline{\theta}$ for $j \in M$ and $(m+1)\overline{\theta}$ is not an integer. The conditions for strict inequality are not needed for part (i) of Theorem <28> but they are needed for part (ii).

If we choose $\beta = n$ then $a_\beta \ge \theta_{max}$ and

<32> $$k \ge 1 + \mathbb{P}_{\boldsymbol{a}} S = 1 + \mathbb{P}_{\boldsymbol{a}} S_{\alpha,\beta} + a_\alpha + a_\beta \ge 1 + \mathbb{P}_{\boldsymbol{a}} S_{\alpha,\beta} + \theta_{max},$$

which implies that the integer $k-2$ is $\ge \theta_{max} + \mathbb{P}S_{\alpha,\beta} - 1$, thereby ensuring that

$$k - 2 \ge \lceil \theta_{max} + \mathbb{P}S_{\alpha,\beta} \rceil - 1 = K.$$

That is, $k-2$ and $k-1$ lie in the range where the inequalities <31> apply. Assertion (iii) follows.

With the results from the Lemma it is now easy to prove the first part of the Theorem.

Proof (for part (i) of Theorem <28>). At each step j choose $\alpha = 1$ and $\beta = n$. If some $\boldsymbol{a}^{(j)}$ has all of its components the same (namely, all equal to $\overline{p}$ because $\sum_i a_i^{(j)} = n\overline{p}$) then we have reached the Binomial and no more averaging is needed. Otherwise we get an infinite sequence for which

<33> $$\mathbb{P}_{\mathbf{p}}\{S \ge k\} = \mathbb{P}_{\boldsymbol{a}^{(0)}}\{S \ge k\} \le \dots \le \mathbb{P}_{\boldsymbol{a}^{(j)}}\{S \ge k\} \le \mathbb{P}_{\boldsymbol{a}^{(j+1)}}\{S \ge k\} \le \dots$$

From equality (ii) of the Lemma we have

$$\sum\nolimits_{i \le n} (a_i^{(j)} - \overline{p})^2 - \sum\nolimits_{i \le n} (a_i^{(j+1)} - \overline{p})^2 = 2r_j^2 \qquad \text{where } r_j := a_n^{(j)} - a_1^{(j)}.$$

where $r_j = a_n^{(j)} - a_1^{(j)}$. Sum over $j = 0, \dots, N$ for an arbitrarily large N to deduce that

$$\sum\nolimits_{i \le n} (p_i - \overline{p})^2 - \sum\nolimits_{i \le n} (a_i^{(N+1)} - \overline{p})^2 = 2\sum\nolimits_{j \le N} r_j^2 \qquad \text{where } r_j := a_n^{(j)} - a_1^{(j)}.$$

The series $\sum_{j=0}^\infty r_j^2$ must converge, implying $a_n^{(j)} - a_1^{(j)} \to 0$ as $j \to \infty$. It follows that $\max_i |a_i^{(j)} - \overline{p}| \to 0$ and $\boldsymbol{a}^{(j)} \to \overline{p}\mathbb{1}$. Continity (compare with representation <15>) of the map $\boldsymbol{a} \mapsto \mathbb{P}_{\boldsymbol{a}}\{S \ge k\}$ then leads to the desired conclusion,

$$\mathbb{P}_{\mathbf{p}}\{S \ge k\} \le \mathbb{P}_{\overline{p}\mathbb{1}}\{S \ge k\} = \mathbb{P}\{X \ge k\}$$

where $X \sim \text{Bin}(n\overline{p})$.

Now for the complication. In order to prove part (ii) of the Theorem, starting from a $\mathbf{p}$ that is not equal to $\overline{p}\mathbb{1}$, it would suffice to have

$$\mathbb{P}_{\boldsymbol{a}^{(j+1)}}\{S \geq k\} > \mathbb{P}_{\boldsymbol{a}^{(j)}}\{S \geq k\}$$

for even a single j. That is, for the $\mathbb{P}_{\boldsymbol{a}^{(j)}}$ analog of <31> we would need to satisfy one of the conditions for strict inequality in the first $\geq$. How hard can that be? I must admit that I had some trouble coming up with a watertight argument—special cases kept causing difficulties. The following example shows where the problem lies.

<34> **Example.** Suppose $\mathbf{p} = (0, 1/2, 1/2, 1/2, 1)$, with $\mathbb{P}_{\mathbf{p}}S = 5/2$ and $\overline{p} = 1/2$ and $K = \lceil 1/2 + 5/2 \rceil - 1 = 2$. We need to consider $k \geq 1 + 5/2$, that is, values of k in the set $\{4, 5\}$. Unfortunately, under $\mathbb{P}_{\mathbf{p}}$, the random variable $S_{1,5}$ has a $\text{Bin}(3, 1/2)$ distribution, which is one of those cases that produces an equality in <31>. For $k - 2 = 2 = K$, the very first step in the procedure used for proving part (i) would jump straight to the $\text{Bin}(5, 1/2)$ with an equality in the tail bound.

It is easy to work around the difficulty in this particular example: start with

$$\mathbf{p}^{(1)} = \text{sort}\left(\text{smooth}_{2,5}(\mathbf{p})\right) = (0, 1/2, 1/2, 3/4, 3/4)$$

then continue as in the proof for part (i) with $\alpha = 1$ and $\beta = 5$. The Lemma gives $\mathbb{P}_{\mathbf{p}^{(1)}}\{S \geq 4\} \geq \mathbb{P}_{\mathbf{p}}\{S \geq 4\}$ and then the second step gives a strict improvement.

In general I found it difficult to keep track of all the special cases, so I opted for the following multistep approach. The idea is to leave p_1 unchanged while replacing each of $p_2, \dots, p_n$ by the same value. That precaution ensures that we don't transform prematurely to the Binomial case, as in Example <34>. Thereafter a few a well chosen pairwise smoothings lead to the desired increase in the tail probabilities while en route to the $\text{Bin}(n, \overline{p})$.

Proof (for part (ii) of Theorem <28>). Consider an initial vector $\mathbf{p}$ with $0 \leq p_1 \leq \dots \leq p_n \leq 1$ and $p_1 < p_n$. We must have $p_1 < \overline{p}$, for otherwise we could not have $\sum_i p_i = n\overline{p}$. (If $p_n > p_1 \geq \overline{p}$ then $\sum_i p_i > n\overline{p}$.) Argue as in the proof for part (i) but with $\alpha = 2$ and $\beta = n$. In the limit we get a $\boldsymbol{b} = (p_1, b \dots, b)$ with $p_1 + (n-1)b = n\overline{p}$ and

$$\mathbb{P}_{\boldsymbol{b}}\{S \geq k\} \geq \mathbb{P}_{\mathbf{p}}\{S \geq k\}.$$

It must be true that $p_1 < \overline{p} < b$ and $r := (p_1 + b)/2 < \overline{p} < b$. Why? Then

$$\boldsymbol{c} = \textsc{sort}(\textsc{smooth}_{1,n}(\boldsymbol{b})) = (r, r, b, \dots, b) \qquad \text{with } r < \overline{p} < b\ .$$

If $n \geq 4$ or $n = 3$ and $r \neq 1/2$ then a $\textsc{smooth}_{1,n}$ would give a strict increase in the tail probability. And so on, as for part (i). If $n = 3$ and $r = 1/2$ then $S_{1,3} \sim \textsc{Bin}(1, 1/2)$ under $\mathbb{P}_{\boldsymbol{c}}$, another bad case. However a $\textsc{sort}(\textsc{smooth}_{1,3}(\boldsymbol{c}))$ would leave a newvector $(1/2, u, u)$ for which $u = (1/2 + b)/2 > 1/2$. Another $\textsc{sort}(\textsc{smooth}_{1,3})$ would then give a strict increase in the tail probability. And so on, as for part (i).

4.8 The median of the Poisson-Binomial

Theorem <28> provides the tool for proving the result about the median described at the end of Section 4.1. The following argument repackages a clever idea of Jogdeo and Samuels (1968).

Recall that a median of a random variable Y is a value M for which

$$\mathbb{P}\{Y \leq M\} \geq 1/2 \quad \textsc{and} \quad \mathbb{P}\{Y \geq M\} \geq 1/2.$$

Such an M always exists but it need not be unique. In fact, for each random variable Y there exist two medians, $m_0(Y) \leq m_1(Y)$ such that the set of all medians equals the closed interval $[m_0(Y), m_1(Y)]$. If $m_0(Y) = m_1(Y)$ then the median is unique. If $m_0(Y) < m_1(Y)$ then

$$\mathbb{P}\{Y \leq m_0(Y)\} = 1/2 = \mathbb{P}\{Y \geq m_1(Y)\},$$

so that $\mathbb{P}\{m_0(Y) < Y < m_1(Y)\} = 0$. All of these results, and more, are derived in Problem [5]. If any of these facts are new to you it would be a good idea to glance at that Problem before reading the proof of the next Theorem.

In the proof I rely on changing subscripts on the random variables instead of adding subscripts to the $\mathbb{P}$ as in the previous Section.

<35> **Theorem.** *Suppose $S \sim \textsc{PBin}(\mathbf{p})$ with $\mathbf{p} \in [0,1]^n$ and $\mu = \mathbb{P}S \in (0, n)$. Then $\lfloor \mu \rfloor \leq m_0(S) \leq m_1(S) \leq \lceil \mu \rceil$.*

Proof. From Problem [5], both endpoints $m_0(S)$ and $m_1(S)$ for the set of medians take values in the set $\{0, 1, \dots, n\}$.

For $m_1(S)$ it is enough to show that

$$\mathbb{P}\{S \geq k\} < 1/2 \qquad \text{where } k := 1 + \lceil \mu \rceil.$$

That inequality and the fact that $\mathbb{P}\{S \geq m_1(S)\} \geq 1/2$ would give

$$m_1(S) \leq k - 1 = \lceil \mu \rceil.$$

Remark. If $\lceil \mu \rceil = n$ then the inequality $m_1(S) \leq \lceil \mu \rceil$ holds for trivial reasons. Thus we need only consider the case where $k \leq n$.

Control of $\mathbb{P}\{S \geq k\}$ via Theorem <28> By definition, $k \geq 1 + \mu$. For $\theta_n := \mu/n$, the Theorem gives

$$\mathbb{P}\{S \geq k\} \leq \mathbb{P}\{W_n \geq k\} \qquad \text{where } W_n \sim \text{Bin}(n, \theta_n).$$

Remark. The inequality is strict if S is not actually Binomially distributed.

Define $S_{n+2} = 0 + W_n + 1$, which has a $\text{PBin}(\boldsymbol{a})$ distribution for $\boldsymbol{a} = (0, \theta_n, \ldots, \theta_n, 1)$, with the θ_n repeated n times. Another appeal to Theorem <28> gives

$$\mathbb{P}\{W_n \geq k\} = \mathbb{P}\{S_{n+2} \geq k + 1\} < \mathbb{P}\{W_{n+2} \geq k + 1\},$$

where $W_{n+2} \sim \text{Bin}(n + 2, \theta_{n+2})$ with $(n + 2)\theta_{n+2} = \mathbb{P}S_{n+2} = 1 + \mu$. Then define $S_{n+4} = 0 + W_{n+2} + 1$. And so on.

In this way we get a sequence of random variables $\{W_{n+2j} : j \in \mathbb{N}\}$ for which

$$W_{n+2j} \sim \text{Bin}(n + 2j, \theta_{n+2j}) \qquad \text{with } (n + 2j)\theta_{n+2j} = j + \mu \quad ,$$
$$\mathbb{P}\{W_{n+2j} \geq k + j\} < \mathbb{P}\{W_{n+2j+2} \geq k + j + 1\}.$$

Note that $\theta_{n+2j} = (\mu + j)/(n + 2j) \to 1/2$ as j increases. The variance of W_{n+2j} equals $\sigma^2_{n+2j} = (\mu+j)(1-\theta_{n+2j}) \to \infty$. By a central limit theorem for a sequence of Binomials (see Exercise 7.20 of Pollard, 2001, for example),

$$Z_j := \frac{W_{n+2j} - (\mu + j)}{\sigma_{n+2j}} \rightsquigarrow Z \sim N(0, 1) \qquad \text{as } j \to \infty.$$

It follows that

$$\mathbb{P}\{W_{n+2j} \geq k + j\} = \mathbb{P}\{Z_j \geq (k - \mu)/\sigma_{n+2j}\} \to \mathbb{P}\{Z \geq 0\} = 1/2.$$

The tail probability $\mathbb{P}\{S \geq k\}$ is bounded above by the limit of a strictly increasing sequence of numbers that converges to $1/2$, which implies the desired inequality $\mathbb{P}\{S \geq k\} < 1/2$.

An analogous argument with $n - S$, which is also Poisson-Binomial distributed, shows that

$$n - m_0(S) = m_1(n - S) \leq \lceil \mathbb{P}(n - S) \rceil = n - \lfloor \mu \rfloor,$$

which rearranges to $m_0(S_n) \geq \lfloor \mu \rfloor$. The first equality in the previous line comes from Problem [5] (v).

4.9 Problems

[1] Suppose $S \sim \text{PBin}(p_1, \dots, p_n)$ with $0 < p_j < 1$ for each j. Here is another proof that

$$(\mathbb{P}\{S = k\})^2 > \mathbb{P}\{S = k-1\}\mathbb{P}\{S = k+1\} \qquad \text{for } 0 < k < n,$$

as in Example <10>.

(i) Make an "independent copy" of the X_i's. That is, create $Y_i \sim \text{Ber}(p_i)$ such that the random variables $X_1, \dots, X_n, Y_1, \dots, Y_n$ are mutually independent. Define $T = Y_1 + \cdots + Y_n$. Show that the asserted inequality is equivalent, by independence, to $\mathbb{P}E_k > \mathbb{P}F_k$, where

$$E_k = \{S = k, T = k\} \quad \text{AND} \quad F_k = \{S = k+1, T = k-1\}.$$

(ii) Define $W_i = X_i + Y_i$. Explain why

$$\begin{aligned} &\mathbb{P}\{X_i = 1, Y_i = 1 \mid W_i = 2\} = 1 \\ &\mathbb{P}\{X_i = 0, Y_i = 0 \mid W_i = 0\} = 1 \\ &\mathbb{P}\{X_i = 1, Y_i = 0 \mid W_i = 1\} = 1/2 = \mathbb{P}\{X_i = 0, Y_i = 0 \mid W_i = 1\}. \end{aligned}$$

(iii) Define $B = \sum_{i=1}^n \{X_i = 1, Y_i = 0\}$ and $N_\alpha = \sum_{i \le n} \{W_i = \alpha\}$, for $\alpha = 0, 1, 2$. Show that

$$S = N_2 + B, \quad T = N_2 + N_1 - B, \quad \text{and } S + T = 2N_2 + N_1.$$

Show also that $\{S + T = 2k\} = \{2N_2 + N_1 = 2k\}$, which has nonzero probability only when the nonnegative integer N_1 is even and $2k - N_1 \ge 0$.

(iv) Define $A_\ell := \{N_1 = 2\ell, N_2 = k - \ell\}$ for $\ell = 0, 1, \dots, k$. Show that

$$\mathbb{P}(E_k \mid A_\ell) = \mathbb{P}\{B = \ell \mid A_\ell\} \quad \text{AND} \quad \mathbb{P}(F_k \mid A_\ell) = \mathbb{P}(B = \ell + 1 \mid A_\ell).$$

(v) Show that

$$2^{2\ell}\mathbb{P}(B = \ell \mid A_\ell) = \binom{2\ell}{\ell} > \binom{2\ell}{\ell+1} = 2^{2\ell}\mathbb{P}(B = \ell + 1 \mid A_\ell).$$

for $1 \le \ell \le k$. What happens if $\ell = 0$?

(vi) Average out to deduce that $\mathbb{P}E_k > \mathbb{P}F_k$.

[2] Suppose f is a convex function on $[0,1]$ and $\theta \in (0,1)$. Amongst probability measures P on $[0,1]$ with mean θ, show that Pf is maximized by the $\text{Ber}(\theta)$ distribution.

[3] (Hoeffding, 1956, comment at the foot of page 713) As in Section 4.7, think of S as the sum of coordinate maps on the set $\{0,1\}^n$ equipped with $\mathbb{P}_{\boldsymbol{p}}$, a product of $\text{Ber}(p_i)$ ditributions. For a given real valued function g on the set $\{0,1,\dots,n\}$ define a map $G : [0,1]^n \to \mathbb{R}$ by

$$G(\mathbf{p}) = \mathbb{P}_{\mathbf{p}} g(S) = \sum\nolimits_{k=0}^{n} g(k)\mathbb{P}_{\mathbf{p}}\{S=k\}.$$

(i) Show that $G(\mathbf{p}) = G(p_{\sigma(1)},\dots,p_{\sigma(n)})$ for each permutation σ of $[n]$ and that $p_j \mapsto G(\boldsymbol{p})$ is linear for each j (with the other coordinates held fixed). Hint: Look at equations <15> and <25>.

(ii) Conversely, suppose a map $H : [0,1]^n \to \mathbb{R}$ has the two properties (symmetry and linearity in each coordinate) identified in part (i). For $0 \le k \le n$ define $h(k) := H(\mathbf{1}_k, \mathbf{0}_{n-k})$, where the argument denotes a vector with 1 repeated k times followed by 0 repeated $n-k$ times. Show that $H(\mathbf{p}) = \mathbb{P}_{\mathbf{p}} h(S)$. Hint:

$$\begin{aligned}
H(\mathbf{p}) &= H(p_1 \times 1 + (1-p_1)\times 0, p_2 \dots, p_n) \\
&= p_1 H(1, p_2, \dots, p_n) + (1-p_1)H(0, p_2, \dots, p_n) \\
&= p_1\Big[p_2 H(1,1,p_3\dots,p_n) + (1-p_2)H(1,0,p_3\dots,p_n)\Big] \\
&\quad + (1-p_1)\Big[p_2 H(0,1,p_3\dots,p_n) + (1-p_2)H(0,0,p_3\dots,p_n)\Big],
\end{aligned}$$

and so on. Remember that the order of the arguments of H is irrelevant.

[4] (Chebyshev 1846; Hoeffding 1956, Corollary 2.1) Suppose $G(\mathbf{p}) = \mathbb{P}_{\mathbf{p}} g(S)$, as in the previous Problem. As a continuous function on the compact set $[0,1]^n$ it must achieve its maximum at some point $\boldsymbol{a}$, which is not unique. Amongst the maximizers choose a vector $\boldsymbol{a}$ of the form $(a_1,\dots,a_m,\mathbf{0}_{n_0},\mathbf{1}_{n_1})$ with $0 < a_1 \le \dots \le a_m < 1$ and $n_0 + n_1$ maximal.

(i) Define $b := (a_m + a_1)/2$ and $t_0 = (a_m - a_1)/2$ and $\epsilon := \min(b, 1-b)$. Show that $a_1 = b - t_0$ and $a_m = b + t_0$ and $0 \le t_0 < \epsilon$.

(ii) Define $T = S - X_1 - X_m$. Show that $G(\boldsymbol{a}) = a_1 a_m A + 2bB + C$ where

$$A = \mathbb{P}_{\boldsymbol{a}}\Big[g(2+T) - 2g(1+T) + g(T)\Big],$$
$$B = \mathbb{P}_{\boldsymbol{a}}\Big[g(1+T) - g(T)\Big], \quad \text{AND} \quad C = \mathbb{P}_{\boldsymbol{a}} g(T)$$

do not depend on a_1 or a_m.

(iii) Define $\boldsymbol{b} \in [0,1]^n$ by $b_1 = b - t$ and $b_m = b + t$ and $b_j = a_j$ otherwise, for a value of t with $|t| \le \epsilon$.

(iv) Show that $0 \le G(\boldsymbol{a}) - G(\boldsymbol{b}) = (t^2 - t_0^2)A$.

(v) Show that $A < 0$ by considering the choice $t = \epsilon$: if $A \ge 0$ then $G(\boldsymbol{b}) \ge G(\boldsymbol{a})$, which would mean that $\boldsymbol{b}$ is another maximizer with

$$\sum\nolimits_j (\{b_j = 0\} + \{b_j = 1\}) > n_0 + n_1.$$

(vi) Consider $t = 0$. Show that $G(\boldsymbol{b}) > G(\boldsymbol{a})$ if $t_0 \ne 0$.

(vii) Conclude that $a_1 = \cdots = a_m = b$. That is, $S - n_1$ has a $\text{B}\textsc{in}(m,b)$ distribution under $\mathbb{P}_{\boldsymbol{a}}$.

[5] Suppose Y is a real valued random variable. Define $F(y) = \mathbb{P}\{Y \le y\}$, the usual distribution function, and $G(y) = \mathbb{P}\{Y \ge y\}$.

(i) Show that F is a non-decreasing function that is continuous from the right. Show also that G is non-increasing and continuous from the left.

(ii) Define

$$\begin{aligned} \mathbb{U} = \mathbb{U}_Y &= \{y \in \mathbb{R} : F(y) \ge 1/2\} \\ \mathbb{L} = \mathbb{L}_Y &= \{y \in \mathbb{R} : G(y) \ge 1/2\} \end{aligned}$$

and define $m_0 = m_0(Y) = \inf \mathbb{U}_Y$ and $m_1 = m_1(Y) = \sup \mathbb{L}_Y$. Show that both m_i values are finite and $m_0 \le m_1$. (Hint: If $y > m_1$ then $\mathbb{P}\{Y \ge y\} < 1/2$, so that $\mathbb{P}\{Y < y\} > 1/2$.) Then use the results from part (i) to show that

$$\mathbb{L} = (-\infty, m_1] \quad \text{AND} \quad \mathbb{U} = [m_0, \infty).$$

(iii) Show that the set of all medians of Y is equal to $\mathbb{L} \cap \mathbb{U} = [m_0, m_1]$.

(iv) If $m_0 < y < m_1$ show that

$$\begin{aligned} 1/2 &\le F(m_0) \le \mathbb{P}\{Y \le y\} \\ 1/2 &\le G(m_1) \le \mathbb{P}\{Y > y\} \end{aligned}$$

Then deduce that $\mathbb{P}\{Y \le m_0\} = 1/2 = \mathbb{P}\{Y \ge m_1\}$ and, by subtraction, $\mathbb{P}\{m_0 < Y < m_1\} = 0$.

(v) Suppose c is a constant and $W = c - Y$. Show that $y \in \mathbb{L}_W$ iff $c - y \in \mathbb{U}_Y$. Deduce that $m_1(W) = c - m_0(Y)$. Argue similarly to show that $m_0(W) = c - m_1(Y)$.

(vi) Suppose Y takes values in a finite subset H of the real line. Show that $m_0(Y)$ and $m_1(Y)$ both belong to H. Hint: Draw pictures of F and G.

4.10 Notes

See Pitman (1997) for a well-written survey of various ways to characterize the Poisson-Binomial distribution, with probabilistic and combinatorial consequences.

The inequality <11>, in Section 4.3, was attributed to Isaac Newton by Hardy, Littlewood, and Pólya (1989, Sections 2.22, 4.3), who cited page 173 of Newton's *Arithmetica Universalis: sive de compositione et resolutione arithmetica liber [Opera, I.]*. Unfortunately I was unable to find the target (in Latin) of this citation, or an English translation thereof. At first I found it easier to follow the sketch proof by Lévy (1937, page 88). I was puzzled why HLP needed to replace the polynomial $p(x)$ by $y^n p(x/y)$ until I saw an argument (replacing $p(z)$ by $z^n p(1/z)$) by Vatutin and Mikhailov (1983), which revealed that the HLP trick was merely a device to change the order of the coefficients in the same way as operator τ in Lemma <8> does. I decided to include my more pedantic treatment of τ in that Lemma because I was worried about the role of roots of $p(z)$ that equal zero. This possibility caused me a lot of trouble. See my Remark in Example <9>.

Fortunately, my problems with Newton's inequality were solved when I looked at an unpublished paper by Stein (1990, page 5), which pointed me at Whiteside (2008, Volume 1, page 519), which contains a translation of some of Newton's work on roots of equations. Footnote (38) on page 523 describes Newton's method in modern algebraic notation. As explained in the General Introduction to Whiteside (2008, Volume 5), the *Arithmetica Universalis* evolved from a set of notes submitted to a Cambridge library by Newton, in fulfillment of an obligation attached to the chair he held. They were later published by another professor, much to Newton's displeasure. Interesting reading.

Section 4.6 is my reworking of material from a short (but most impressive) paper by Samuels (1965), with much help from Hoeffding (1956). As the reader might suspect, I had a lot of trouble with strict inequalities; counterexamples to my theorems kept popping up.

Samuels attributed the results that appear in my Section 4.7 to Hoeffding, although he also noted that similar results were obtained by Chebyshev (1846). The pairwise averaging method that I used in my Theorem <28> is due to Chebyshev. The method was later (apparently independently) reinvented by Hoeffding (1956). Chebyshev showed, for a fixed value of $\sum_{j=1}^{n} p_j$, that the tail probability $\mathbb{P}\{S \geq m\}$ for $S \sim \text{PBin}(p_1, \ldots, p_n)$ is maximized when there is a value θ such that every p_j belongs to $\{0, \theta, 1\}$. Hoeffding derived the same fact for a general $\mathbb{P}g(S)$. See Problems [3] and [4]. See

Seneta (1998, pages 206–207) and Maistrov (2014, pages 191,196) for very interesting discussions of Chebyshev (1846).

Section 4.8 is based on Jogdeo and Samuels (1968).

References

Chebyshev, P. L. (1846). Démonstration élémentaire d'une proposition générale de la théorie des probabilités. *Journal für die reine und angewandte Mathematik (Crelle's journal) 33*, 259–267. (The author's surname was spelled 'Tchebichef' in this paper.).

Hardy, G., J. Littlewood, and G. Pólya (1989). *Inequalities.* Cambridge University Press. (Reprint of the the 1952 second edition).

Hoeffding, W. (1956). On the distribution of the number of successes in independent trials. *The Annals of Mathematical Statistics 27*(3), 713–721.

Jogdeo, K. and S. M. Samuels (1968). Monotone convergence of binomial probabilities and a generalization of Ramanujan's theorem. *Annals of Mathematical Statistics 39*, 1191–1195.

Le Cam, L. (1983). A remark on empirical measures. In P. J. Bickel, K. Doksum, and J. L. Hodges (Eds.), *Festschrift for Erich Lehmann*, pp. 305–327. Belmont, California: Wadsworth.

Lévy, P. (1937). *Théorie de l'addition des variables aléatoires.* Paris: Gauthier-Villars. References from the 1954 second edition.

Maistrov, L. E. (2014). *Probability Theory: A Historical Sketch.* Academic Press. Translated from the 1967 Russian original by Samual Kotz.

Pitman, J. (1997). Probabilistic bounds on the coefficients of polynomials with only real zeros. *Journal of Combinatorial Theory, Series A 77*(2), 279–303.

Pollard, D. (2001). *A User's Guide to Measure Theoretic Probability.* Cambridge University Press.

Pólya, G. and G. Szegö (1976). *Problems and Theorems in Analysis II: Theory of Functions. Zeros. Polynomials. Determinants. Number Theory. Geometry.* Springer-Verlag.

Prasolov, V. V. (2004). *Polynomials*. Springer. (Translated from the 2001 Russian edition).

Samuels, S. M. (1965). On the number of successes in independent trials. *Annals of Mathematical Statistics 36*, 1272–1278.

Seneta, E. (1998). Early influences on probability and statistics in the Russian empire. *Archive for History of Exact Sciences 53*(3-4), 201–213.

Stein, C. (1990). Application of Newton's identities to a generalized birthday problem and to the Poisson binomial distribution. Technical Report 354, Stanford Department of Statistics. Available as report EFS NSF 354 at https://statistics.stanford.edu/resources/technical-reports.

Vatutin, V. and V. Mikhailov (1983). Limit theorems for the number of empty cells in an equiprobable scheme for group allocation of particles. *Theory of Probability and Its Applications 27*(4), 734–743.

Whiteside, D. T. (Ed.) (2008). *The mathematical papers of Isaac Newton*. Cambridge University Press. (A paperback reprint of the 8 volumes of an earlier (1967–?) hardcover printing. Described by CUP as "an annotated and critical edition, of all the known mathematical papers of Isaac Newton ... Translations of papers in Latin face the original text and notes are printed on the page-openings to which they refer, so far as possible. ").