Estimating uncertainties in histograms filled from Markov Chains

Christopher G. Lester

September 26, 2018

Abstract

'A rediscovery of the Central Limit Theorem for Markov Chains'

1 Setting the scene

We are working with a Markov chain on a basis of B-states that may be assumed to be bins of a histogram. Within such a framework, define

$$p_{sr}^{Nb}$$

to be the probability of scoring s in bin b with r being the last bin visited, given that there have been N draws from the Markov chain. I.e.

$$p_{sr}^{Nb} = p(s, r|N, b).$$

Define also

$$p_s^{Nb}$$

to be the probability of scoring s in bin b irresepctive of which bin was last visited, given again that there have been N draws from the Markov chain.

$$p_s^{Nb} = p(s|N, b).$$

Clearly the two are related:

$$p_s^{Nb} = \sum_{r=1}^{B} p_{sr}^{Nb} \tag{1}$$

if there are B bins. Since there is *some* score in bin b we can also see that it must be the case that

$$1 = \sum_{s=0}^{N} p_s^{Nb} = \sum_{s=0}^{N} \sum_{r=1}^{B} p_{sr}^{Nb}.$$

Furthermore, we use p_{ij} to indicate the Markov chain transition matrix element $p(i|j)$ which indicates the probability with which the chain will move to state i from position j. Note that the matrix P whose elements are $(P)_{ij} = p_{ij}$ is left-stochastic and so:

$$\sum_{i=1}^{B} p_{ij} = 1 \tag{2}$$

We take $\vec{\pi}$ to be the B-vector containing the limiting distribution of the Markov chain. As this is a vector of probabilities we take it to be normalised such that

$$\sum_{i=1}^{B} \pi_i = 1. \tag{3}$$

As it is the limiting distribution of the chain, $\vec{\pi}$ will satisfy

$$P\vec{\pi} = \vec{\pi}, \tag{4}$$

which is

$$\sum_{j=1}^{B} p_{ij}\pi_j = \pi_i$$

in component form. Defining U to be a $B \times B$ matrix composed entirely of ones, and $\vec{1}$ to be a column B-vector composed entirely of ones:

$$(U)_{ij} = 1 \qquad \text{and} \qquad (\vec{1})_i = 1$$

we would see from (3) that

$$U\vec{\pi} = \vec{1}. \tag{5}$$

Furthermore, (4) implies that

$$(\mathbb{1} - P)\,\vec{\pi} = 0 \tag{6}$$

which implies

$$(\mathbb{1} - P)\,\vec{\pi} + \vec{1} = \vec{1} \tag{7}$$

which by (5) implies

$$(\mathbb{1} - P)\,\vec{\pi} + U\vec{\pi} = \vec{1} \tag{8}$$

$$\implies \qquad (\mathbb{1} - P + U)\,\vec{\pi} = \vec{1} \tag{9}$$

and so

$$\vec{\pi} = (\mathbb{1} - P + U)^{-1} \cdot \vec{1} \tag{10}$$

and so

$$
\begin{aligned}
P^{\infty} &= \lim_{N\to\infty} P^N && (11)\\
&= \begin{pmatrix} \vec{\pi} & \vec{\pi} & \cdots & \vec{\pi} \end{pmatrix} && (12)\\
&= \vec{\pi} \cdot \vec{1}^T && (13)\\
&= (\mathbb{1} - P + U)^{-1} \cdot \vec{1} \cdot \vec{1}^T && (14)\\
&= (\mathbb{1} - P + U)^{-1} \cdot U. && (15)
\end{aligned}
$$

Note that above we have assumed that $\mathbb{1} - P + U$ is non singular, and therefore that its inverse exists. Allegedly Resnick (1992) 'Adventures in Stochastic Processes', Proposition 2.14.1, will prove that this is so in the cases which will matter to us. READ BEFORE INCLUDING

Note also that product in (15) does depend on the order of multiplication, since the other way round one has $U \cdot (\mathbb{1} - P + U)^{-1} = \frac{1}{B}U$ as demonstrated by the following argument:

$$
\begin{aligned}
& U = UP \qquad \text{(by the left-stochastic nature of } P \text{ seen in (2))} && (16)\\
\implies & 0 = U - UP && (17)\\
\implies & BU = U - UP + BU && (18)\\
\implies & BU = U - UP + U^2 && (19)\\
\implies & BU = U\,(\mathbb{1} - P + U) && (20)\\
\implies & BU\,(\mathbb{1} - P + U)^{-1} = U && (21)\\
\implies & U\,(\mathbb{1} - P + U)^{-1} = \frac{1}{B}U. && (22)
\end{aligned}
$$

1.1 Connection to an application: histogramming

The nature of the process by which the Markov chain generates scores may be described in the following recurrence relation:

$$p_{s+1,r}^{N+1,b} = \sum_{k=1}^{B} \left(\delta_{rb} p_{sk}^{Nb} + (1-\delta_{rb}) p_{s+1,k}^{Nb} \right) p_{rk} \tag{23}$$

which says, in words, that the probability for getting a particular score $s+1$ (at a particular time $N+1$, in a given bin b, when the last visited bin was r) can be broken down into a sum of the probabilites for reaching that state from the previous timestep. More specifically, it records that if the previous bin r was not the scoring bin b, the then the previous score was the same as the current score, while in other cases it must have been one lower. The boundary conditions for this recurence relation are:

$$p_{sr}^{0b} = \begin{cases} \pi_r & \text{if } s = 0, \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

$$= \delta_{s0}\pi_r \tag{25}$$

Note that the above boundary conditions state, in effect, that the starting point of the chain is 'typical'. One could choose alternative starting conditions (e.g. conditions that would fix the chain at a particular start point, rather than a typical one). This may be something interesting to pursue in the future.

It might also be nice to try to see if the above recurrence relation (23) is solvable for all indices N, b, s and r. I do not know whether it is solvable in a nice closed form way – I have not tried. However what I did next was convert (23) into a recurrence relation for expectations of s and s^2 as I will shortly describe. Note that I calculated the expectations for s and s^2 separately. I have wondered whether I could have done both at once using an appropriately chosen moment generating function. Trying to do so probably should be investigated.

2 Expectations

2.1 Proper expectations

Using $\langle s_b \rangle_N$ to denote the expected score s in a bin b at time N, we have from the very definition of an expectation that:

$$\langle s_b \rangle_N = \sum_{s=0}^{N} s \cdot p_s^{Nb} \tag{26}$$

and more generally, for $m \geq 1$, that

$$\langle s_b^m \rangle_N = \sum_{s=0}^{N} s^m \cdot p_s^{Nb}. \tag{27}$$

2.2 Pseudo-expectations

Later we will abuse the 'expectation' notation by making use of a quantity defined (again for $m \geq 1$) as follows:

$$\langle s_b^m \rangle_{Nr} = \sum_{s=0}^{N} s^m \cdot p_{sr}^{Nb} \tag{28}$$

which, on account of (1), satisfies:

$$\sum_{r=1}^{B} \langle s_b^m \rangle_{Nr} = \langle s_b^m \rangle_N . \tag{29}$$

3 A useful identity

For $m \geq 1$ we can put (28) together with the recurrence relation (23) as follows:

$$
\begin{aligned}
\langle s_b^m \rangle_{N+1,r} &= \sum_{s=0}^{N+1} s^m \cdot p_{sr}^{N+1,b} && (30)\\
&= \sum_{s=1}^{N+1} s^m \cdot p_{sr}^{N+1,b} \qquad \text{(since the first term is zero)} && (31)\\
&= \sum_{j=0}^{N} (j+1)^m \cdot p_{j+1,r}^{N+1,b} \qquad \text{(replacing } s \text{ with } j+1\text{)} && (32)\\
&= \sum_{s=0}^{N} (s+1)^m \cdot p_{s+1,r}^{N+1,b} \qquad \text{(replacing } j \text{ with } s\text{)} && (33)\\
&= \sum_{s=0}^{N} (s+1)^m \sum_{k=1}^{B} \left(\delta_{rb} p_{sk}^{Nb} + (1-\delta_{rb}) p_{s+1,k}^{Nb}\right) p_{rk} \qquad \text{(using recurrence (23))} && (34)\\
&= \delta_{rb} \sum_{s=0}^{N} (s+1)^m \sum_{k=1}^{B} p_{sk}^{Nb} p_{rk} + (1-\delta_{rb}) \sum_{s=0}^{N} (s+1)^m \sum_{k=1}^{B} p_{s+1,k}^{Nb} p_{rk} \qquad \text{(splitting)} && (35)\\
&= \delta_{rb} \sum_{s=0}^{N} (s+1)^m \sum_{k=1}^{B} p_{sk}^{Nb} p_{rk} + (1-\delta_{rb}) \sum_{s=0}^{N-1} (s+1)^m \sum_{k=1}^{B} p_{s+1,k}^{Nb} p_{rk} \qquad \text{(since } p_{N+1,b}^{Nb} = 0 \text{)} && (36)\\
&= \delta_{rb} \sum_{s=0}^{N} (s+1)^m \sum_{k=1}^{B} p_{sk}^{Nb} p_{rk} + (1-\delta_{rb}) \sum_{s=1}^{N} s^m \sum_{k=1}^{B} p_{sk}^{Nb} p_{rk} \qquad \text{(relabelling } s+1 \to s \text{ in second product)} && (37)\\
&= \delta_{rb} \sum_{s=0}^{N} (s+1)^m \sum_{k=1}^{B} p_{sk}^{Nb} p_{rk} + (1-\delta_{rb}) \sum_{s=0}^{N} s^m \sum_{k=1}^{B} p_{sk}^{Nb} p_{rk} \qquad \text{(since } s=0 \text{ contributes nothing)} && (38)\\
&= \delta_{rb} \sum_{s=0}^{N} ((s+1)^m - s^m) \sum_{k=1}^{B} p_{sk}^{Nb} p_{rk} + \sum_{s=1}^{N} s^m \sum_{k=0}^{B} p_{sk}^{Nb} p_{rk} \qquad \text{(collecting terms)} && (39)\\
&= \delta_{rb} \sum_{s=0}^{N} ((s+1)^m - s^m) \sum_{k=1}^{B} p_{sk}^{Nb} p_{rk} + \sum_{k=1}^{B} p_{rk} \sum_{s=0}^{N} s^m \cdot p_{sk}^{Nb} \qquad \text{(re-ordering)} && (40)\\
&= \delta_{rb} \sum_{s=0}^{N} ((s+1)^m - s^m) \sum_{k=1}^{B} p_{sk}^{Nb} p_{rk} + \sum_{k=1}^{B} p_{rk} \langle s_b^m \rangle_{Nk} \qquad \text{(by definition)} && (41)\\
&= \delta_{rb} \sum_{s=0}^{N} ((s+1)^m - s^m) \sum_{k=1}^{B} p_{sk}^{Nb} p_{bk} + \sum_{k=1}^{B} p_{rk} \langle s_b^m \rangle_{Nk} \qquad \text{(using the properties of } \delta_{rb}\text{).} && (42)
\end{aligned}
$$

The result just proved is very useful, so we re-state it in a single line:

$$\langle s_b^m \rangle_{N+1,r} = \delta_{rb} \sum_{s=0}^{N} ((s+1)^m - s^m) \sum_{k=1}^{B} p_{sk}^{Nb} p_{bk} + \sum_{k=1}^{B} p_{rk} \langle s_b^m \rangle_{Nk} \,. \qquad (43)$$

4 Results deriving from (43)

4.1 Moment expectation recurrence relations

By summing (43) over r we can generate a recurrence relation for the m-th moment of the score s in any bin b:

$$\langle s_b^m\rangle_{N+1} = \sum_{r=1}^{B} \langle s_b^m\rangle_{N+1,r} \qquad \text{(by the sum rule (29))} \tag{44}$$

$$= \sum_{r=1}^{B} \delta_{rb} \sum_{s=0}^{N}((s+1)^m - s^m) \sum_{k=1}^{B} p_{sk}^{Nb} p_{bk} + \sum_{r=1}^{B}\sum_{k=1}^{B} p_{rk} \langle s_b^m\rangle_{Nk} \qquad \text{(by (43))} \tag{45}$$

$$= \sum_{s=0}^{N}((s+1)^m - s^m) \sum_{k=1}^{B} p_{sk}^{Nb} p_{bk} + \sum_{k=1}^{B} \langle s_b^m\rangle_{Nk} \qquad \text{(summing over } r \text{ and using (2))} \tag{46}$$

$$= \sum_{s=0}^{N}((s+1)^m - s^m) \sum_{k=1}^{B} p_{sk}^{Nb} p_{bk} + \langle s_b^m\rangle_{N} \qquad \text{(by the sum rule (29)).} \tag{47}$$

Two special cases of (47) will be very useful to us. These are the $m = 1$ and $m = 2$ cases. We will consider $m = 1$ first, returning to the $m = 2$ case later in (70).

4.2 Results concerning the first moment

When $m = 1$ equation (47) reads:

$$\langle s_b\rangle_{N+1} - \langle s_b\rangle_{N} = \sum_{s=0}^{N}((s+1)^1 - s^1) \sum_{k=1}^{B} p_{sk}^{Nb} p_{bk} = \sum_{s=0}^{N}\sum_{k=1}^{B} p_{sk}^{Nb} p_{bk}. \tag{48}$$

We have a second means of calculating the LHS of (48) since we know that it is always the case that

$$\langle s_b\rangle_{N} = N\pi_b. \tag{49}$$

(Should probably have noted that earlier!) Equation (48) is thus telling us that

$$\sum_{s=0}^{N}\sum_{k=1}^{B} p_{sk}^{Nb} p_{bk} = \langle s_b\rangle_{N+1} - \langle s_b\rangle_{N} \tag{50}$$

$$= (N+1)\pi_b - N\pi_b \qquad \text{(by (49))} \tag{51}$$

$$= \pi_b \tag{52}$$

which we will make use of later.

4.3 Pseudo-expectation recurrence relations

By setting m to the value 1 in (43) we obtain an identity realating pseudo-expectations:

$$\langle s_b\rangle_{N+1,r} = \delta_{rb} \sum_{s=0}^{N}((s+1) - s) \sum_{k=1}^{B} p_{sk}^{Nb} p_{bk} + \sum_{k=1}^{B} p_{rk} \langle s_b\rangle_{Nk} \tag{53}$$

$$= \delta_{rb} \sum_{s=0}^{N}\sum_{k=1}^{B} p_{sk}^{Nb} p_{bk} + \sum_{k=1}^{B} p_{rk} \langle s_b\rangle_{Nk} \tag{54}$$

$$= \delta_{rb}\pi_b + \sum_{k=1}^{B} p_{rk} \langle s_b\rangle_{Nk} \qquad \text{(by (52))} \tag{55}$$

or equivalently

$$\langle s_b \rangle_{Nr} = \delta_{rb}\pi_b + \sum_{k=1}^{B} p_{rk} \langle s_b \rangle_{N-1,k} \tag{56}$$

which in vector form looks like:

$$\vec{s}^{N}_{(b)} = \vec{t}_{(b)} + P \cdot \vec{s}^{N-1}_{(b)} \tag{57}$$

if we define $\vec{s}^{N}_{(b)}$ to be the B-vector whose r-th component is $\langle s_b \rangle_{Nr}$ and the define $\vec{t}_{(b)}$ to be the B-vector whose r-th component is $\delta_{rb}\pi_b$:

$$\left(\vec{t}_{(b)}\right)_r = \delta_{rb}\pi_b. \tag{58}$$

Note that, for *any* $B \times B$ matrix A:

$$\begin{aligned}
\left(A \cdot \vec{t}_{(b)}\right)_b &= \sum_{r=1}^{B} (A)_{br} \left(\vec{t}_{(b)}\right)_r & (59)\\
&= \sum_{r=1}^{B} (A)_{br} \, \delta_{br}\pi_b \qquad \text{(by definion (58))} & (60)\\
&= (A)_{bb} \, \pi_b. & (61)
\end{aligned}$$

The advantage of the vectorial form of (57) is that it allows us to solve this recurrence relation in pesudo expectations. In particular, we can see from (57) that

$$\vec{s}^{1}_{(b)} = \vec{t}_{(b)} + P\vec{s}^{0}_{(b)} \tag{62}$$

and so

$$\vec{s}^{2}_{(b)} = \vec{t}_{(b)} + P\vec{t}_{(b)} + P^2\vec{s}^{0}_{(b)} \tag{63}$$

and in general

$$\vec{s}^{N}_{(b)} = \left(\mathbb{1} + P + P^2 + \cdots + P^{N-1}\right)\vec{t}_{(b)} + P^N\vec{s}^{0}_{(b)}. \tag{64}$$

To simplify the above we note that $\vec{s}^{0}_{(b)} = \vec{0}$ since

$$\begin{aligned}
\left(\vec{s}^{0}_{(b)}\right)_r &= \langle s_b \rangle_{0r} \qquad \text{(from the notation)} & (65)\\
&= \sum_{s=0}^{N} s \cdot p^{0b}_{sr} \qquad \text{(by definition (28))} & (66)\\
&= \sum_{s=0}^{N} s \cdot \delta_{s0} \cdot \pi_r \qquad \text{(from the bounary condition (25))} & (67)\\
&= 0. & (68)
\end{aligned}$$

Although we used the specific boundary condition (25) to show $\vec{s}^{0}_{(b)} = \vec{0}$, note that the only part of (25) that mattered was the δ_{s0} part. This is the part which enforces the notion that: at time $N = 0$ it is impossible to have a score that is greater than zero in any bin. We can safely say, therefore, that $\vec{s}^{0}_{(b)} = \vec{0}$ is zero for all possible legitimate boundary conditions – not only the specific set proposed in (25). In full generality, therefore, we can say that:

$$\vec{s}^{N}_{(b)} = \left(\mathbb{1} + P + P^2 + \cdots + P^{N-1}\right)\vec{t}_{(b)}. \tag{69}$$

4.4 Results concerning the second moment

The $m = 2$ case of (47) tells us

$$\left\langle s_b^2 \right\rangle_{N+1} - \left\langle s_b^2 \right\rangle_N = \sum_{s=0}^{N} ((s+1)^2 - s^2) \cdot \sum_{k=1}^{B} p_{sk}^{Nb} p_{bk} \tag{70}$$
$$= \sum_{s=0}^{N} (1+2s) \cdot \sum_{k=1}^{B} p_{sk}^{Nb} p_{bk} \tag{71}$$
$$= \sum_{s=0}^{N} \sum_{k=1}^{B} p_{sk}^{Nb} p_{bk} + \sum_{s=0}^{N} \sum_{k=1}^{B} 2s \cdot p_{sk}^{Nb} p_{bk} \tag{72}$$
$$= \sum_{s=0}^{N} \sum_{k=1}^{B} p_{sk}^{Nb} p_{bk} + \sum_{s=0}^{N} \sum_{k=1}^{B} 2s \cdot p_{sk}^{Nb} p_{bk} \tag{73}$$
$$= \pi_b + \sum_{s=0}^{N} \sum_{k=1}^{B} 2s \cdot p_{sk}^{Nb} p_{bk} \qquad \text{(using (52) as promised)} \tag{74}$$
$$= \pi_b + 2 \sum_{k=1}^{B} p_{bk} \sum_{s=0}^{N} s \cdot p_{sk}^{Nb} \qquad \text{(re-ordering)} \tag{75}$$
$$= \pi_b + 2 \sum_{k=1}^{B} p_{bk} \left\langle s_b \right\rangle_{Nk} \qquad \text{(by definition (28))} \tag{76}$$
$$= \pi_b + 2 \left(P \left(\mathbb{1} + P + P^2 + \cdots + P^{N-1} \right) \vec{t}_{(b)} \right)_b \qquad \text{(by (69))} \tag{77}$$
$$= \pi_b + 2 \left(\left(P + P^2 + P^3 + \cdots + P^N \right) \vec{t}_{(b)} \right)_b \tag{78}$$
$$= \pi_b + 2 \left(P + P^2 + P^3 + \cdots + P^N \right)_{bb} \pi_b \qquad \text{(by (61))} \tag{79}$$
$$= \pi_b + 2 \left(X_N \right)_{bb} \pi_b \tag{80}$$

if we define

$$X_N = P + P^2 + P^3 + \cdots + P^N. \tag{81}$$

Given that all boundary conditions imply that $\left\langle s_b^2 \right\rangle_0 = 0$ we therefore have

$$\left\langle s_b^2 \right\rangle_1 = 1\pi_b, \tag{82}$$
$$\left\langle s_b^2 \right\rangle_2 = 2\pi_b + 2(X_1)_{bb}\pi_b, \tag{83}$$
$$\left\langle s_b^2 \right\rangle_3 = 3\pi_b + 2(X_1 + X_2)_{bb}\pi_b, \tag{84}$$
$$\cdots \tag{85}$$
$$\left\langle s_b^2 \right\rangle_N = N\pi_b + 2(X_1 + X_2 + \cdots + X_{N-1})_{bb}\pi_b \tag{86}$$

which, using (81), gives

$$\left\langle s_b^2 \right\rangle_N = N\pi_b + 2 \left((N-1)P + (N-2)P^2 + \cdots + 2P^{N-2} + 1P^{N-1} \right)_{bb} \pi_b. \tag{87}$$

Having defined

$$\mathrm{Var}[s_b]_N = \left\langle s_b^2 \right\rangle_N - \left\langle s_b \right\rangle_N^2 \tag{88}$$

we can put (87) together with (49) to conclude that

$$\mathrm{Var}[s_b]_N = \left\langle s_b^2 \right\rangle_N - (N\pi_b)^2 \tag{89}$$
$$= N\pi_b + 2 \left((N-1)P + (N-2)P^2 + \cdots + 2P^{N-2} + 1P^{N-1} \right)_{bb} \pi_b - N^2 \pi_b^2. \tag{90}$$

5 Functions of P

We know that P has $\vec{\pi}$ as an eigenvector with unit eigenvalue. And we know that all other eigenvectors $\vec{v}_2, \ldots, \vec{v}_B$ have ei genvalues $\lambda_2, \ldots, \lambda_B$ each of whose magnitudes are less than one.

There is therefore a matrix

$$S = \begin{pmatrix} \vdots & \vdots & & \vdots \\ \vec{\pi} & \vec{v}_2 & \cdots & \vec{v}_B \\ \vdots & \vdots & & \vdots \end{pmatrix}$$

and a diagonal matrix

$$D = \mathrm{Diag}[1, \lambda_2, \cdots, \lambda_B]$$

which diagonalises P such that $P \cdot S = S \cdot D$. Assuming S is invertible CHECK we then can say that for any polynomial $f(x)$:

$$\begin{aligned} f(P) &= f(S \cdot D \cdot S^{-1}) && (91) \\ &= S \cdot f(D) \cdot S^{-1} && (92) \\ &= S \cdot \mathrm{Diag}[f(1), f(\lambda_2), \cdots, f(\lambda_B)] \cdot S^{-1} && (93) \\ &= S \cdot \mathrm{Diag}[f(1), 0, \cdots, 0] \cdot S^{-1} + S \cdot \mathrm{Diag}[0, f(\lambda_2), \cdots, f(\lambda_B)] \cdot S^{-1}. && (94) \end{aligned}$$

So far, so good. But to make further progress with (94) we need to work with a concrete $f(P)$. From (90) it is clear that to compute $\mathrm{Var}[s_b]_N$ we have interest in

$$f(P) = (N-1)P + (N-2)P^2 + \cdots + 2P^{N-2} + 1P^{N-1} \tag{95}$$

and thus in

$$f(x) = (N-1)x + (N-2)x^2 + \cdots + 2x^{N-2} + 1x^{N-1}.$$

If $x = 1$ it is readily apparent that

$$f(1) = \frac{1}{2}N(N-1). \tag{96}$$

For values of $x \neq 1$ we can instead say:

$$\begin{aligned} f(x) &= (N-1)x + (N-2)x^2 + \cdots + x^{N-1} && (97) \\ &= x + x^2 + \cdots + x^{N-2} + x^{N-1} + \\ &\quad x + x^2 + \cdots + x^{N-2} + \\ &\quad \cdots \\ &\quad x + x^2 + \\ &\quad x && (98) \\ &= \frac{x(1-x^{N-1}) + x(1-x^{N-2}) + \cdots + x(1-x^2) + x(1-x)}{1-x} && (99) \\ &= \frac{(N-1)x - (x^2 + x^3 + \cdots + x^N)}{1-x} && (100) \\ &= \frac{(N-1)x}{1-x} - \frac{x^2(1-x^{N-1})}{(1-x)^2} && (101) \\ &= \frac{(N-1)x}{1-x} - \frac{x^2 - x^{N+1}}{(1-x)^2} && (102) \\ &= N\frac{x}{1-x} - \frac{x}{1-x} - \frac{x^2 - x^{N+1}}{(1-x)^2} && (103) \\ &= N\frac{x}{1-x} - \frac{x - x^2}{(1-x)^2} - \frac{x^2 - x^{N+1}}{(1-x)^2} && (104) \\ &= N\frac{x}{1-x} - \frac{x - x^{N+1}}{(1-x)^2} && (105) \end{aligned}$$

which is well conditioned for all $N \geq 0$ even when $x = 0$.

With the above in mind, define

$$g(x) = \frac{1}{2}N(N-1)x, \tag{106}$$

and

$$h(x) = \begin{cases} \frac{Nx}{1-x} - \frac{x - x^{N+1}}{(1-x)^2} & \text{when } x \neq 1 \\ \text{undefined} & \text{otherwise.} \end{cases} \quad (107)$$

Note that on account of the trailing x in (106) which is not present in (96), the function $g(x)$ is not, in general, the same as $f(x)$, though they will co-incide when $x = 1$. Note also that $h(x)$ is not the same as $f(x)$, since $f(x)$ is well defined at $x = 1$ whereas $h(x)$ is not. Nonetheless, with those definitions: $g(0) = h(0) = 0$, $g(1) = f(1)$ and $f(x) = h(x)$ when $x \neq 1$, and so

$$f(P) = S \cdot \text{Diag}[f(1), 0, \cdots, 0] \cdot S^{-1} + S \cdot \text{Diag}[0, f(\lambda_2), \cdots, f(\lambda_B)] \cdot S^{-1} \quad \text{(from (94))} \quad (108)$$

$$= S \cdot \text{Diag}[g(1), g(0), \cdots, g(0)] \cdot S^{-1} + S \cdot \text{Diag}[h(0), h(\lambda_2), \cdots, h(\lambda_B)] \cdot S^{-1} \quad (109)$$

$$= S \cdot g(\text{Diag}[1, 0, \cdots, 0]) \cdot S^{-1} + S \cdot h(\text{Diag}[0, \lambda_2, \cdots, \lambda_B]) \cdot S^{-1} \quad \text{(CHECK OK NOW } h \text{ IS NOT POLYNO} \quad (110)$$

$$= g(S \cdot \text{Diag}[1, 0, \cdots, 0] \cdot S^{-1}) + h(S \cdot \text{Diag}[0, \lambda_2, \cdots, \lambda_B] \cdot S^{-1}) \quad \text{(CHECK OK NOW } h \text{ IS NOT POLYNO} \quad (111)$$

$$= g(G) + h(H) \quad (112)$$

provided that G and H are matrices which share the same eigenvalues $\{\vec{\pi}, \vec{v}_2, \cdots, \vec{v}_B\}$ as P, and so are simultaneously diagonalisable with P, but which have eigenvalues $\{1, 0, 0, \cdots, 0\}$ (for G) and $\{0, \lambda_2, \cdots, \lambda_B\}$ (for H).

It is trivial to show that the matrix P^n has the same eigenvectors as P but has eigenvalues $\{1, \lambda_2^n, \cdots, \lambda_B^n\}$. Since $\{\lambda_2, \cdots, \lambda_B\}$ all have modulus less than one, it is trivial therefore to identify

$$G = P^\infty.$$

We can also see that

$$H = P - P^\infty$$

since

$$H\vec{\pi} = (P - P^\infty)\vec{\pi} \quad (113)$$

$$= P\vec{\pi} - P^\infty\vec{\pi} \quad (114)$$

$$= \vec{\pi} - \vec{\pi} \quad (115)$$

$$= 0 \quad (116)$$

and

$$H\vec{v}_b = (P - P^\infty)\vec{v}_b \quad (117)$$

$$= P\vec{v}_b - P^\infty\vec{v}_b \quad (118)$$

$$= \lambda_b\vec{v}_b - 0\vec{v}_b \quad (119)$$

$$= \lambda_b\vec{v}_b. \quad (120)$$

Therefore, renaming H as Q, we conclude that

$$f(P) = g(P^\infty) + h(Q) \quad (121)$$

$$= \frac{1}{2}N(N-1)P^\infty + \frac{NQ}{\mathbb{1} - Q} - \frac{Q - Q^{N+1}}{(\mathbb{1} - Q)^2} \quad (122)$$

if

$$Q = P - P^\infty.$$

5.0.1 A cautionary aside

Note that

$$P = P^\infty + (P - P^\infty) = G + H$$

and

$$f(P) = g(G) + h(H)$$

and yet $f(x)$ need not be a linear function! The spliting of P into $G+H$, and the fact that $h(0)=0$, appear to be both 'fortunate' in some sense. It is not yet obvious to me why such a split need have been possible. Indeed, my method of finding a split seems very contrived, which suggests I am missing the bigger picture somwhere. We have certainly relied on some seemingly irrelevant but particular features of $f(x)$. For example: if $f(x)$ had been defined slightly differently as $\hat{f}(x)$ including a leading N:

$$\hat{f}(x) = N + (N-1)x + (N-2)x^2 + \cdots + x^{N-1}$$

then we would have concluded that $\hat{f}(1) = \frac{1}{2}N(N+1)$, and for values of $x \neq 1$:

$$\begin{aligned}
\hat{f}(x) &= N + (N-1)x + (N-2)x^2 + \cdots + x^{N-1} && (123)\\
&= 1 + x + x^2 + \cdots + x^{N-2} + x^{N-1} + \\
&\quad 1 + x + x^2 + \cdots + x^{N-2} + \\
&\quad \cdots \\
&\quad 1 + x + x^2 + \\
&\quad 1 + x + \\
&\quad 1 && (124)\\
&= \frac{(1-x^N) + (1-x^{N-1}) + \cdots + (1-x^2) + (1-x)}{1-x} && (125)\\
&= \frac{N - (x + x^2 + \cdots + x^N)}{1-x} && (126)\\
&= \frac{N}{1-x} - \frac{x(1-x^N)}{(1-x)^2}. && (127)
\end{aligned}$$

In such a scenario, although we might have attempted to define

$$\begin{aligned}
\hat{g}(x) &= \frac{1}{2}N(N+1)x \neq \hat{f}(x), \qquad \text{and} && (128)\\
\hat{h}(x) &= \hat{f}(x), && (129)
\end{aligned}$$

we would have found that the last of these definitions would have been incompatible with our last proof. The previous proof required $g(0) = h(0) = 0$, but above we have $\hat{h}(0) = N \neq 0$. This is a shame, as in many ways it would be a lot nicer to work with $\hat{f}(x)$ than with $f(x)$. There is probably a better way of approaching this problem that allows $\hat{f}(x)$ to be used.

6 Computing the variance

Putting (90) together with (95) and (122) yields the variance in b of our histogram after N draws:

$$\begin{aligned}
\mathrm{Var}[s_b]_N &= N\pi_b + 2\left(\frac{1}{2}N(N-1)P^\infty + \frac{NQ}{\mathbb{1}-Q} - \frac{Q-Q^{N+1}}{(\mathbb{1}-Q)^2}\right)_{bb}\pi_b - N^2\pi_b^2 && (130)\\
&= N\pi_b + N(N-1)\left(P^\infty\right)_{bb}\pi_b + 2\left(\frac{NQ}{\mathbb{1}-Q} - \frac{Q-Q^{N+1}}{(\mathbb{1}-Q)^2}\right)_{bb}\pi_b - N^2\pi_b^2 && (131)\\
&= N\pi_b + N(N-1)\pi_b^2 + 2\left(\frac{NQ}{\mathbb{1}-Q} - \frac{Q-Q^{N+1}}{(\mathbb{1}-Q)^2}\right)_{bb}\pi_b - N^2\pi_b^2 \qquad (\text{since } (P^\infty)_{bb} = \pi_b \text{ CITE}) && (132)\\
&= N\pi_b(1-\pi_b) + 2\pi_b\left(\frac{NQ}{\mathbb{1}-Q} - \frac{Q-Q^{N+1}}{(\mathbb{1}-Q)^2}\right)_{bb} \qquad (\text{after cancellation}). && (133)
\end{aligned}$$

7 Testing/Examples

7.1 Example 1

A 'sticky' Markov chain that flits between $B=2$ bins, resting in each with equal probability (on average) but which tends to 'stick' on the current bin for $O(1/\epsilon)$ events before moving off, might be expected to have an effective sample size per bin that is a factor of order $1/\epsilon$ smaller than would be expected for a chain that produces fully uncorrelated samples. We will see later that this is the case, though there are subtleties in how this statement must be made concrete.

A matrix P describing such a chain could be as follows:

$$P = \begin{pmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{pmatrix}.$$

It may be trivially checked that this P has invariant distribution $\vec{\pi} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$ and so satisfies

$$P^\infty = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \qquad \text{and} \qquad Q = P - P^\infty = \begin{pmatrix} \frac{1}{2}-\epsilon & \epsilon-\frac{1}{2} \\ \epsilon-\frac{1}{2} & \frac{1}{2}-\epsilon \end{pmatrix}.$$

A small amount of algebra then shows that the variance in either bin predicted by (133) is given by:

$$\mathrm{Var}[s_b|\text{example 1}]_N = \frac{N}{4}\left(\frac{1}{\epsilon}-1\right) + \frac{(1-2\epsilon)^N-1}{8\epsilon^2} - \frac{(1-2\epsilon)^N-1}{4\epsilon}. \tag{134}$$

For $\epsilon = \frac{1}{2}$ this chain becomes fully uncorrelated, at which point

$$\lim_{\epsilon\to\frac{1}{2}} \mathrm{Var}[s_b|\text{example 1}]_N = \frac{N}{4}$$

matching the standard $\mathrm{Var} = Npq$ formula for a bionomial distribution with $p = \frac{1}{2}$ as expected. What is more interesting, however, is the small-ϵ behaviour:

$$\begin{aligned} \mathrm{Var}[s_b|\text{example 1}]_N &= \frac{N}{4}\left(\frac{1}{\epsilon}-1\right) + \frac{-2N\epsilon + \frac{N(N-1)}{2}(-2\epsilon)^2 + O(\epsilon^3)}{8\epsilon^2} - \frac{-2N\epsilon + O(\epsilon^2)}{4\epsilon} & (135)\\ &= \frac{0}{\epsilon} + \left(-\frac{N}{4} + \frac{N(N-1)}{4} + \frac{N}{2}\right) + O(\epsilon) & (136)\\ &= \frac{N^2}{4} + O(\epsilon). & (137) \end{aligned}$$

Evidently the variance does not have the $O(N/\epsilon)$ behaviour we expected! Let us think more carefully about this. In the $\epsilon \ll 1$ case, our two options are that *either* all the samples land on state one, *or* all the samples land on state two. Both are equally likely. Therefore the mean score in any bin is $\langle s_b \rangle = N/2$ and the variance is therefore

$$\left\langle (s_b - \langle s_b \rangle)^2 \right\rangle = \frac{1}{2}(N - \frac{N}{2})^2 + \frac{1}{2}(0 - \frac{N}{2})^2 = \frac{N^2}{4}$$

which is the result we just found in (137). So the result was right, but our intuition needs qualifying more precisely. The reason is as follows. The variance cannot scale as N/ϵ in the small ϵ limit since this would lead to unboundedly large variances which are not possible when N is finite. The fininteness of N acts as a kind of 'regularizer' that prevents divergence in the variance. We must therefore be careful about the order in which we take our limits. If we take the $N \to \infty$ limit first, we can try to stay in the $\frac{1}{N} \ll \epsilon \ll \frac{1}{2}$ regime and find therein the originally expected behaviour. Let

$$\begin{aligned} \rho(\epsilon, N) &= \frac{\mathrm{Var}[s_b|\text{example 1 with } \epsilon = \epsilon]_N}{\mathrm{Var}[s_b|\text{example 1 with } \epsilon = \frac{1}{2}]_N} & (138)\\ &= \frac{\frac{N}{4}\left(\frac{1}{\epsilon}-1\right) + \frac{(1-2\epsilon)^N-1}{8\epsilon^2} - \frac{(1-2\epsilon)^N-1}{4\epsilon}}{\frac{N}{4}(2-1) + \frac{(1-1)^N-1}{2} - \frac{(1-1)^N-1}{2}} & (139)\\ &= \frac{\frac{N}{4}\left(\frac{1}{\epsilon}-1\right) + \frac{(1-2\epsilon)^N-1}{8\epsilon^2} - \frac{(1-2\epsilon)^N-1}{4\epsilon}}{\frac{N}{4}} & (140)\\ &= \left(\frac{1}{\epsilon} - 1 + \frac{(1-2\epsilon)^N-1}{2N\epsilon^2} - \frac{(1-2\epsilon)^N-1}{N\epsilon}\right) & (141)\\ &= \left(\frac{1}{\epsilon}-1\right) + \left(\frac{1}{2\epsilon^2} - \frac{1}{\epsilon}\right)\frac{(1-2\epsilon)^N-1}{N}. & (142) \end{aligned}$$

Then

$$\lim_{N\to\infty} \rho(\epsilon, N) = \left(\frac{1}{\epsilon} - 1\right) + \left(\frac{1}{2\epsilon^2} - \frac{1}{\epsilon}\right) \lim_{N\to\infty} \frac{(1-2\epsilon)^N - 1}{N} \tag{143}$$

$$= \frac{1}{\epsilon} - 1 \tag{144}$$

(since $0 \le \epsilon \le 1$ and so $0 \le 2\epsilon \le 2$ and so $-1 \le 2\epsilon - 1 \le 1$ and so $(1-2\epsilon)^2 \le 1$) in contrast to the

$$\lim_{\epsilon\to 0} \rho(\epsilon, N) = \frac{\frac{N^2}{4}}{\frac{N}{4}} = N \tag{145}$$

already found. Thus we see the sense in which the variance of a chain with

$$P_\epsilon = \begin{pmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{pmatrix}.$$

is bigger than the variance of the fully uncorrelated chain

$$P_{\frac{1}{2}} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

by a factor of $\frac{1}{\epsilon} - 1$ (previously guessed to be $O(1/\epsilon)$). This result requires the $N \to \infty$ limit to be taken before the $\epsilon \to 0$ limit, meaning that this result is only 'practically' valid under certain conditions, which might (perhaps) be something like $N \gg \frac{1}{\epsilon}$. Note also that, in words, the factor $\frac{1}{\epsilon} - 1$ carries the meanting of 'one less than the expected dwell time on each state' – something we might call the expected 'additional' dwell time.

7.2 Example 2

COPY IN THE EXAMPLE FROM THE PAPER HERE

7.3 Example with Weights

Suppose:

1. We create a set of bin counts s_b by drawing N samples from our general B-state Markov chain whose transition matrix is P.

2. That the Markov chain was constructed such that the 'desired' count in bin b is actually

   $$s'_b = w_b s_b$$

   rather than s_b. In other words, associated with each bin is a known weight correction factor w_b.

3. That each corrected bin count s'_b is intended to represent the count in a bin, of width δx_b, within a histogram.

4. That, if plotted as a histogram, the height y'_b within that bin would therfore be $y'_b = s'_b/\delta x_b$.

5. That normalising that histogram to unit area would requre generating a new set of heights

   $$y''_b = \frac{y'_b}{\sum_{i=1}^{B} y'_i \delta x_i}$$

   since then the resuling area A'' would be:

   $$A'' = \sum_{b=1}^{B} y''_b \delta x_b = \sum_{b=1}^{N} \frac{y'_b}{\sum_{i=1}^{B} y'_i \delta x_i} \delta x_b = 1.$$

6. Writing out MOO in terms of s', we see that

$$s''_b/\delta x_b = \frac{s'_b/\delta x_b}{\sum_{i=1}^{B}\left(s'_i/\delta x_i\right)\delta x_i} = \frac{s'_b/\delta x_b}{\sum_{i=1}^{B} s'_i}$$

and so

$$s''_b = \frac{s'_b}{\sum_{i=1}^{B} s'_i} = \frac{w_b s_b}{\sum_{i=1}^{B} w_i s_i} \tag{146}$$

and

$$\sum_{b=1}^{B} s''_b = \sum_{b=1}^{B} \frac{s'_b}{\sum_{i=1}^{B} s'_i} = 1. \tag{147}$$

We are therefore at liberty to ignore the widths δx_b and concentrate exclusively on using the weights so long as we renormalise such that $\sum_{b=1}^{B} s''_b = 1$. I don't really know why I'm saying this. This seems pretty obvious.

7. We are therefore primarily interested in the variance of s''_b since we can work out the variance in other things from it:

$$\mathrm{Var}[s''_b] = \left\langle (s''_b)^2 \right\rangle - \left(\langle s''_b \rangle\right)^2 \tag{148}$$

however neither of the quantities therein are easy to calculate. For example:

$$\langle s''_b \rangle = \left\langle \frac{s'_b}{\sum_{i=1}^{B} s'_i} \right\rangle \tag{149}$$

$$= \left\langle \frac{s_b w_b}{\sum_{i=1}^{B} s_i w_i} \right\rangle \tag{150}$$

which is not easy to simplify as the denominator inside the expectation is not constant and so cannot be taken outside the expectation. The denominator *would* be constant if the weights w_i all shared a common value w, because then $\sum_{i=1}^{B} s_i w_i = \sum_{i=1}^{B} s_i w = w \sum_{i=1}^{B} s_i = wN$. But alas the weights need not share a common value, and those that do are not particularly useful. We could Monte Carlo expectations of the above form relatively easily - and indeed have. But can we do anything else?

Recall that the weighted and normalised counts in each bin are defined to be:

$$s''_b = \frac{s_b w_b}{\sum_{i=1}^{B} s_i w_i}. \tag{151}$$

To empahsise that these weighted and re-normalised quantities are estimators for probabilities which sum to one (see (147)) we rename s''_b as p''_b:

$$p''_b = \frac{s_b w_b}{\sum_{i=1}^{B} s_i w_i} \tag{152}$$

admitting then the customary notation for complementary probabilities:

$$q''_b = 1 - p''_b = \sum_{i \neq B} p''_i. \tag{153}$$

With those definitions:

$$dp''_b = \sum_{k=1}^{B} \left(\frac{\delta_{bk} w_b}{\sum_{i=1}^{B} s_i w_i} - \frac{s_b w_b w_k}{\left(\sum_{i=1}^{B} s_i w_i\right)^2} \right) ds_k \tag{154}$$

and so

$$\frac{dp_b''}{p_b''} = \sum_{k=1}^{B} \left(\frac{\delta_{bk}}{s_b} - \frac{w_k}{\left(\sum_{i=1}^{B} s_i w_i\right)} \right) ds_k \tag{155}$$

$$= \sum_{k=1}^{B} \left(\delta_{bk} - \frac{s_k w_k}{\left(\sum_{i=1}^{B} s_i w_i\right)} \right) \frac{ds_k}{s_k} \qquad (\text{since } \frac{\delta_{bk}}{s_b} = \frac{\delta_{bk}}{s_k}) \tag{156}$$

$$= \sum_{k=1}^{B} \left(\delta_{bk} - p_k''\right) \frac{ds_k}{s_k}. \tag{157}$$

Defining the fractional differentials df_b and df_b'' by:

$$df_b = \frac{ds_b}{s_b} \qquad \text{and} \qquad df_b'' = \frac{dp_b''}{p_b''}$$

we therefore have:

$$df_b'' = \sum_{k=1}^{B} \left(\delta_{bk} - p_k''\right) df_k \tag{158}$$

which is the same as

$$df_1'' = +\underline{q_1''} df_1 - p_2'' df_2 - p_3'' df_3 - \ldots - p_B'' df_B \tag{159}$$
$$df_2'' = -p_1'' df_1 + \underline{q_2''} df_2 - p_3'' df_3 - \ldots - p_B'' df_B \tag{160}$$
$$df_3'' = -p_1'' df_1 - p_2'' df_2 + \underline{q_3''} df_3 - \ldots - p_B'' df_B \tag{161}$$
$$\vdots$$
$$df_B'' = -p_1'' df_1 - p_2'' df_2 - p_3'' df_3 - \ldots + \underline{q_B''} df_B. \tag{162}$$

7.4 Uncertainties in the linear regime

Prior to this point, statements have been rigorous, and no assumptions have been made about the size of any uncertainties. Hereafter, however, we make an assumption that need not always be valid, namely that all uncertainties are small enough that we may analyse them linearly, neglecting uncertainties on the coefficients of the differentials.

7.4.1 Correlations

Since $\sum_{b=1}^{B} s_b = N$, any up-fluctuations in some of the s_b will always be compensated for by down-fluctuations in some of the others: $\sum_{b=1}^{B} ds_b = 0$. For this reason, the $\delta f_b = \frac{\delta s_b}{s_b}$ cannot ever be completely independent. Additional correlations could arise from dependencies between adjancent samples generated by the Markov chain if $P^\infty \neq P$.

7.4.2 Assuming no correlations

If the fractional uncertainties of the raw counts in each bin were uncorrelated, we could say that:

$$\delta f_1'' = \underline{q_1''} \delta f_1 \oplus p_2'' \delta f_2 \oplus p_3'' \delta f_3 \oplus \ldots \oplus p_B'' \delta f_B \tag{163}$$
$$\delta f_2'' = p_1'' \delta f_1 \oplus \underline{q_2''} \delta f_2 \oplus p_3'' \delta f_3 \oplus \ldots \oplus p_B'' \delta f_B \tag{164}$$
$$\delta f_3'' = p_1'' \delta f_1 \oplus p_2'' \delta f_2 \oplus \underline{q_3''} \delta f_3 \oplus \ldots \oplus p_B'' \delta f_B \tag{165}$$
$$\vdots$$
$$\delta f_B'' = p_1'' \delta f_1 \oplus p_2'' \delta f_2 \oplus p_3'' \delta f_3 \oplus \ldots \oplus \underline{q_B''} \delta f_B. \tag{166}$$

in which we use $\oplus$ to mean summation in quadrature.

In the special case of an approximately uniform distribution ($p''_b = O(1/B)$) with approximately equal raw uncertainties in each bin ($\delta f_b = O(\delta f)$) then (163) would look like

$$\delta f''_b = \sqrt{\left[(1 - O(1/B))O(\delta f)\right]^2 + (B-1)\left[O(1/B)O(\delta f)\right]^2} \tag{167}$$

$$= O(\delta f) + O(1/B)O(\delta f) \tag{168}$$

7.4.3 Assuming worst possible correlations

If, instead, we assumed that the uncertainties in each of the raw counts were correlated so as to generate the largest possible fractional uncertainties in the f''_b quantities, we would have:

$$\delta f''_1 \le \underline{q''_1}|\delta f_1| + p''_2|\delta f_2| + p''_3|\delta f_3| + \ldots + p''_B|\delta f_B| \tag{169}$$

$$\delta f''_2 \le p''_1|\delta f_1| + \underline{q''_2}|\delta f_2| + p''_3|\delta f_3| + \ldots + p''_B|\delta f_B| \tag{170}$$

$$\delta f''_3 \le p''_1|\delta f_1| + p''_2|\delta f_2| + \underline{q''_3}|\delta f_3| + \ldots + p''_B|\delta f_B| \tag{171}$$

$$\vdots$$

$$\delta f''_B \le p''_1|\delta f_1| + p''_2|\delta f_2| + p''_3|\delta f_3| + \ldots + \underline{q''_B}|\delta f_B|. \tag{172}$$

In general these are over-estimates, since it is not possible for all δf_b to simultaneously fluctuate up (or down).

In the special case of an approximately uniform distribution ($p''_b = O(1/B)$) with approximately equal raw uncertainties in each bin ($\delta f_b = O(\delta f)$) then (169) would look like

$$\delta f''_b \le (1 - O(1/B))O(\delta f) + (B-1)O(1/B)O(\delta f) = O(\delta f) + O(1/B)O(\delta f) \tag{173}$$

7.4.4 Realistic correlations

What is really needed is a measure that is based on a realistic measure of potential correlations.

$$\mathrm{Var}[df''_b] = \mathrm{Var}\left[\sum_{k=1}^{B} \left(\delta_{bk} - p''_k\right) df_k\right] \tag{174}$$

$$= \sum_{i=1}^{B} \mathrm{Var}\left[\left(\delta_{bi} - p''_i\right) df_i\right] + \sum_{i \ne j} \mathrm{Cov}\left[\left(\delta_{bi} - p''_i\right) df_i, \left(\delta_{bj} - p''_j\right) df_j\right] \tag{175}$$

$$\approx \sum_{i=1}^{B} \left(\delta_{bi} - p''_i\right)^2 \mathrm{Var}\left[df_i\right] + \sum_{i \ne j} \left(\delta_{bi} - p''_i\right)\left(\delta_{bj} - p''_j\right) \mathrm{Cov}\left[df_i, df_j\right] \tag{176}$$

$$\approx \sum_{i=1}^{B} \frac{\left(\delta_{bi} - p''_i\right)^2}{s_i^2} \mathrm{Var}\left[ds_i\right] + \sum_{i \ne j} \frac{\left(\delta_{bi} - p''_i\right)}{s_i} \frac{\left(\delta_{bj} - p''_j\right)}{s_j} \mathrm{Cov}\left[ds_i, ds_j\right] \tag{177}$$

where the approximation in the last two lines come from our assumption (described earlier) that we are in an appropriate linear regime. This is usable if the covariances of the raw counts is a calculable quantity. Now,

$$\mathrm{Cov}(s_b, s_c) = \langle (s_b - \langle s_b \rangle)(s_c - \langle s_c \rangle) \rangle \tag{178}$$

$$= \langle s_b s_c \rangle - \langle s_b \rangle \langle s_c \rangle \tag{179}$$

$$= \langle s_b s_c \rangle - N^2 \pi_b \pi_c \tag{180}$$

so the only quantity needing to be calculated is $\langle s_b s_c \rangle$ which may be amenable to an approach that iterates on N.

Hypothesis

I hypothesise that the off-diagonal elements of the matrix buried in (133) are going to become relevant! (Update: they did!)

7.4.5 Can the existing arguments be generalised?

It does not matter whether or not they can, since the indicator function method used below surpasses them.

8 Using indicator functions

Let the random variable s_{bt} be 1 if the Markov chain sits in bin b at time t, and zero therwise. Therefore

$$s_b = \sum_{t=1}^{N} s_{bt}.$$

8.1 Variance

$$\mathrm{Var}[s_b] = \mathrm{Var}\left[\sum_{t=1}^{N} s_{bt}\right] \tag{181}$$

$$= \sum_{t=1}^{N}\sum_{u=1}^{N} \mathrm{Cov}\left[s_{bt}, s_{bu}\right] \tag{182}$$

$$= \sum_{t=1}^{N}\sum_{u=1}^{N} \left[\langle s_{bt}s_{bu}\rangle - \langle s_{bt}\rangle \langle s_{bu}\rangle\right] \tag{183}$$

$$= \sum_{t=1}^{N}\sum_{u=1}^{N} \left[\langle s_{bt}s_{bu}\rangle - \pi_b\pi_b\right] \tag{184}$$

$$= \sum_{t=1}^{N}\sum_{u=1}^{N} \langle s_{bt}s_{bu}\rangle - N^2\pi_b\pi_b \tag{185}$$

$$= \sum_{t=1}^{N}\sum_{u=1}^{N} p(b_t \cap b_u) - N^2\pi_b\pi_b \tag{186}$$

where $p(b_t \cap c_u)$ means the probability that the chain will be in bin b at time t and in bin c at time u. Thus

$$\mathrm{Var}[s_b] = \sum_{t=1}^{N} p(b_t \cap b_t) + 2 \sum_{t=1}^{N-1} \sum_{u=t+1}^{N} p(b_t \cap b_u) - N^2 \pi_b \pi_b \qquad (\text{since } p(b_u \cap b_t) = p(b_t \cap b_u)) \tag{187}$$

$$= \sum_{t=1}^{N} p(b_t) + 2 \sum_{t=1}^{N-1} \sum_{u=t+1}^{N} p(b_0 \cap b_{u-t}) - N^2 \pi_b \pi_b \tag{188}$$

$$= \sum_{t=1}^{N} \pi_b + 2 \sum_{t=1}^{N-1} \sum_{u=t+1}^{N} p(b_0 \cap b_{u-t}) - N^2 \pi_b \pi_b \tag{189}$$

$$= N\pi_b + 2 \sum_{t=1}^{N-1} \sum_{k=1}^{N-t} p(b_0 \cap b_k) - N^2 \pi_b \pi_b \qquad (\text{replacing } u \text{ with } k = u - t) \tag{190}$$

$$= N\pi_b + 2 \sum_{k=1}^{N-1} \sum_{t=1}^{N-k} p(b_0 \cap b_k) - N^2 \pi_b \pi_b \qquad (\text{reversing order of sum}) \tag{191}$$

$$= N\pi_b + 2 \sum_{k=1}^{N-1} (N-k) p(b_0 \cap b_k) - N^2 \pi_b \pi_b \tag{192}$$

$$= N\pi_b + 2 \sum_{k=1}^{N-1} (N-k) p(b_0) p(b_k | b_0) - N^2 \pi_b \pi_b \tag{193}$$

$$= N\pi_b + 2\pi_b \sum_{k=1}^{N-1} (N-k) p(b_k | b_0) - N^2 \pi_b \pi_b \tag{194}$$

$$= N\pi_b + 2\pi_b \sum_{k=1}^{N-1} (N-k) G_{bb}^{k} - N^2 \pi_b \pi_b. \tag{195}$$

in which we have defined G_{cb}^N to denote the probability of ending up at state c, after N moves, having started in state b:

$$G_{cb}^N = p(c_N | b_0).$$

Perhaps you might think of G as a bit like a Greens function. Clearly

$$G_{cb}^N = \sum_{i=1}^{B} p_{ci} G_{ib}^{N-1} \tag{196}$$

$$= \sum_{i,j}^{B} p_{ci} p_{ij} G_{jb}^{N-2} \tag{197}$$

$$= \sum_{i,j,\cdots,z}^{B} p_{ci} p_{ij} \cdots p_{yz} G_{zb}^{N-N} \tag{198}$$

$$= \sum_{i,j,\cdots,z}^{B} p_{ci} p_{ij} \cdots p_{yz} \delta_{zb} \tag{199}$$

$$= \sum_{i,j,\cdots,z}^{B} p_{ci} p_{ij} \cdots p_{yb} \tag{200}$$

$$= \left(P^N\right)_{cb}. \tag{201}$$

Hence

$$\mathrm{Var}[s_b] = N\pi_b + 2\pi_b \sum_{k=1}^{N-1} (N-k) (P^k)_{bb} - N^2 \pi_b \pi_b \tag{202}$$

$$= N\pi_b + 2\pi_b (f(P))_{bb} - N^2 \pi_b \pi_b \tag{203}$$

in terms of the $f(P)$ defined in (95).

8.2 Similarly, for the co-variance:

$$\begin{aligned}
\mathrm{Cov}\left[s_b, s_c\right] &= \langle (s_b - \langle s_b \rangle)(s_c - \langle s_c \rangle) \rangle && (204)\\
&= \langle s_b s_c \rangle - \langle s_b \rangle \langle s_c \rangle && (205)\\
&= \langle s_b s_c \rangle - N^2 \pi_b \pi_c && (206)\\
&= \left\langle \sum_{t=1}^{N} s_{bt} \sum_{u=1}^{N} s_{cu} \right\rangle - N^2 \pi_b \pi_c && (207)\\
&= \sum_{t=1}^{N} \sum_{u=1}^{N} \langle s_{bt} s_{cu} \rangle - N^2 \pi_b \pi_c && (208)\\
&= \sum_{t=1}^{N} \sum_{u=1}^{N} p(b_t \cap c_u) - N^2 \pi_b \pi_c && (209)
\end{aligned}$$

where $p(b_t \cap c_u)$ means the probability that the chain will be in bin b at time t and in bin c at time u. Thus

$$\begin{aligned}
\mathrm{Cov}[s_b, s_c] &= \sum_{t=1}^{N} p(b_t \cap c_t) + \sum_{t,u|t \neq u}^{N} p(b_t \cap c_u) - N^2 \pi_b \pi_c && (210)\\
&= \sum_{t=1}^{N} \delta_{bc} p(b_t) + \left(\sum_{t=1}^{N-1} \sum_{u=t+1}^{N} + \sum_{t=2}^{N} \sum_{u=1}^{t-1} \right) p(b_t \cap c_u) - N^2 \pi_b \pi_c && (211)\\
&= \sum_{t=1}^{N} \delta_{bc} \pi_b + \sum_{t=1}^{N-1} \sum_{u=t+1}^{N} p(b_0 \cap c_{u-t}) + \sum_{t=2}^{N} \sum_{u=1}^{t-1} p(c_0 \cap b_{t-u}) - N^2 \pi_b \pi_c && (212)\\
&= N \delta_{bc} \pi_b + \sum_{t=1}^{N-1} \sum_{u=t+1}^{N} p(b_0 \cap c_{u-t}) + \sum_{u=2}^{N} \sum_{t=1}^{u-1} p(c_0 \cap b_{u-t}) - N^2 \pi_b \pi_c \qquad \text{(re-labelling second double sum)} && (213)\\
&= N \delta_{bc} \pi_b + \sum_{t=1}^{N-1} \sum_{u=t+1}^{N} p(b_0 \cap c_{u-t}) + \sum_{t=1}^{N-1} \sum_{u=t+1}^{N} p(c_0 \cap b_{u-t}) - N^2 \pi_b \pi_c \qquad \text{(re-ordering second double sum)} && (214)\\
&= N \delta_{bc} \pi_b + \left(\sum_{t=1}^{N-1} \sum_{u=t+1}^{N} p(b_0 \cap c_{u-t}) + [b \leftrightarrow c] \right) - N^2 \pi_b \pi_c && (215)\\
&= N \delta_{bc} \pi_b + \left(\sum_{t=1}^{N-1} \sum_{k=1}^{N-t} p(b_0 \cap c_k) + [b \leftrightarrow c] \right) - N^2 \pi_b \pi_c \qquad \text{(replacing } u \text{ with } k = u - t\text{)} && (216)\\
&= N \delta_{bc} \pi_b + \left(\sum_{k=1}^{N-1} \sum_{t=1}^{N-k} p(b_0 \cap c_k) + [b \leftrightarrow c] \right) - N^2 \pi_b \pi_c \qquad \text{(reversing order of sum)} && (217)\\
&= N \delta_{bc} \pi_b + \left(\sum_{k=1}^{N-1} (N-k) p(b_0 \cap c_k) + [b \leftrightarrow c] \right) - N^2 \pi_b \pi_c && (218)\\
&= N \delta_{bc} \pi_b + \left(\sum_{k=1}^{N-1} (N-k) p(b_0) p(c_k | b_0) + [b \leftrightarrow c] \right) - N^2 \pi_b \pi_c && (219)\\
&= N \delta_{bc} \pi_b + \left(\pi_b \sum_{k=1}^{N-1} (N-k) p(c_k | b_0) + [b \leftrightarrow c] \right) - N^2 \pi_b \pi_c && (220)\\
&= N \delta_{bc} \pi_b + \left(\pi_b \sum_{k=1}^{N-1} (N-k) G_{cb}^{k} + [b \leftrightarrow c] \right) - N^2 \pi_b \pi_c. && (221)
\end{aligned}$$

in which we have defined G^N_{cb} to denote the probability of ending up at state c, after N moves, having started in state b:

$$G^N_{cb} = p(c_N|b_0).$$

Perhaps you might think of G as a bit like a Greens function. Clearly

$$G^N_{cb} = \sum_{i=1}^{B} p_{ci} G^{N-1}_{ib} \tag{222}$$

$$= \sum_{i,j}^{B} p_{ci} p_{ij} G^{N-2}_{jb} \tag{223}$$

$$= \sum_{i,j,\cdots,z}^{B} p_{ci} p_{ij} \cdots p_{yz} G^{N-N}_{zb} \tag{224}$$

$$= \sum_{i,j,\cdots,z}^{B} p_{ci} p_{ij} \cdots p_{yz} \delta_{zb} \tag{225}$$

$$= \sum_{i,j,\cdots,z}^{B} p_{ci} p_{ij} \cdots p_{yb} \tag{226}$$

$$= \left(P^N\right)_{cb}. \tag{227}$$

Hence

$$\mathrm{Cov}[s_b, s_c] = N\pi_b\delta_{bc} + \left(\pi_b \sum_{k=1}^{N-1} (N-k)(P^k)_{cb} + [b \leftrightarrow c]\right) - N^2\pi_b\pi_c \tag{228}$$

$$= N\pi_b\delta_{bc} + \Big(\pi_b (f(P))_{cb} + [b \leftrightarrow c]\Big) - N^2\pi_b\pi_c \tag{229}$$

in terms of the $f(P)$ defined in (95).

9 Computing the co-variance

Putting (229) together with (95) and (122) yields the variance in b of our histogram after N draws:

$$\mathrm{Cov}[s_b, s_c]_N = N\pi_b\delta_{bc} + \left\{ \left(\frac{1}{2} N(N-1) P^\infty + \frac{NQ}{\mathbb{1} - Q} - \frac{Q - Q^{N+1}}{(\mathbb{1} - Q)^2} \right)_{cb} \pi_b + [b \leftrightarrow c] \right\} - N^2\pi_b\pi_c \tag{230}$$

$$= N\pi_b\delta_{bc} + \left\{ \frac{1}{2} N(N-1) \left(P^\infty\right)_{cb} \pi_b + \left(\frac{NQ}{\mathbb{1} - Q} - \frac{Q - Q^{N+1}}{(\mathbb{1} - Q)^2} \right)_{cb} \pi_b + [b \leftrightarrow c] \right\} - N^2\pi_b\pi_c \tag{231}$$

$$= N\pi_b\delta_{bc} + \left\{ \frac{1}{2} N(N-1) \pi_c\pi_b + \left(\frac{NQ}{\mathbb{1} - Q} - \frac{Q - Q^{N+1}}{(\mathbb{1} - Q)^2} \right)_{cb} \pi_b + [b \leftrightarrow c] \right\} - N^2\pi_b\pi_c \qquad \text{(since } (P^\infty)_{cb} = \pi \tag{232}$$

$$= N\pi_b\delta_{bc} + N(N-1)\pi_b\pi_c + \left\{ \left(\frac{NQ}{\mathbb{1} - Q} - \frac{Q - Q^{N+1}}{(\mathbb{1} - Q)^2} \right)_{cb} \pi_b + [b \leftrightarrow c] \right\} - N^2\pi_b\pi_c \tag{233}$$

$$= N\pi_b(\delta_{bc} - \pi_c) + \left\{ \left(\frac{NQ}{\mathbb{1} - Q} - \frac{Q - Q^{N+1}}{(\mathbb{1} - Q)^2} \right)_{cb} \pi_b + [b \leftrightarrow c] \right\} \qquad \text{(after cancellation).} \tag{234}$$

9.1 Checks

9.1.1 Check number one

Try $P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$. This has $Q = 0$ and so has $\mathrm{Cov}[s_1, s_2]_N = N\frac{1}{2}\left(0 - \frac{1}{2}\right) = -\frac{N}{4}$. Compare this to a direct calculation of

$$\mathrm{Cov}[s_1, s_2] = \langle s_1 s_2 \rangle - \langle s_1 \rangle \langle s_2 \rangle \tag{235}$$
$$= \sum_{s=0}^{N} s(N-s)\frac{N!}{s!(N-s)!}(\frac{1}{2})^s(\frac{1}{2})^{N-s} - \frac{N}{2}\frac{N}{2} \tag{236}$$
$$= \sum_{s=0}^{N} s(N-s)\frac{N!}{s!(N-s)!}\left(\frac{1}{2}\right)^N - \frac{N^2}{4} \tag{237}$$
$$= \sum_{s=1}^{N-1} s(N-s)\frac{N!}{s!(N-s)!}\left(\frac{1}{2}\right)^N - \frac{N^2}{4} \tag{238}$$
$$= \left(\frac{1}{2}\right)^N N(N-1)\sum_{s=1}^{N-1}\frac{(N-2)!}{(s-1)!(N-s-1)!} - \frac{N^2}{4} \tag{239}$$
$$= \left(\frac{1}{2}\right)^N N(N-1)2^{N-2} - \frac{N^2}{4} \tag{240}$$
$$= \frac{1}{4}N(N-1) - \frac{N^2}{4} \tag{241}$$
$$= -\frac{N}{4} \tag{242}$$

as desired.

9.1.2 Check number two

Now try Try $P = \begin{pmatrix} 1-a & b \\ a & 1-b \end{pmatrix}$. This has $\vec{\pi} = (b, a)/(a+b)$ and so $P^\infty = \begin{pmatrix} b & b \\ a & a \end{pmatrix}/(a+b)$ and so

$$Q = P - P^\infty \tag{243}$$
$$= \begin{pmatrix} a - a^2 + b - ab & ab + b^2 \\ a^2 + ab & a - ab + b - b^2 \end{pmatrix}/(a+b) - \begin{pmatrix} b & b \\ a & a \end{pmatrix}/(a+b) \tag{244}$$
$$= \begin{pmatrix} a - a^2 - ab & ab + b^2 - b \\ a^2 + ab - a & -ab + b - b^2 \end{pmatrix}/(a+b) \tag{245}$$

...... I got bored and checked this one on mathematica, and it was fine for $N = 2$.

$$\tag{246}$$

10 Correlations between samples, introduced by the use of Markov chains, cannot be accounted for by a single correction that accounts for 'effective sample size'

It is easy to see that the correlations which a Markov chain introduces between its successive samples must tend to increase the variance in each bin of the histogram being estimated. For example, if $0 < \epsilon \ll 1$, then a Markov chain with left-stochastic matrix

$$\begin{pmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{pmatrix}$$

will visit two states equally frequently in the long run, much like fair coin tosses, but unlike fair coin tosses this Markov chain will tend to stay in any given state for $O(1/\epsilon)$ time-steps before moving on. Consequently, after N time-steps the Markov chain will have only made $O(\epsilon N)$ non-trival state transitions — much as $O(\epsilon N)$ independent coin-tosses will lead to $O(\epsilon N)$ transitions from head to tail (or vice versa). Accordingly, the N highly correlated samples of the original Markov chain are said to be about as useful as $O(\epsilon N)$ independent samples would have been. The history of this chain is said to have an 'effective sample size' of $O(\epsilon N)$, and the variance in the number of visits to a given state is $O(N/\epsilon)$ rather than $O(N)$ for independent coin tosses.

It is tempting to assume that for a more complicated Markov chain having more than two states, that the same will apply; namely that for any Markov history of length N there will exist a scale factor k, with $0 < k < 1$, such that the effective sample size of this chain is kN. If this were so then one might hope that the variance in the number of times state b is visited would be $O(n\pi_b(1-\pi_b)/k)$ instead of the $O(n\pi_b(1-\pi_b))$ variance that would pertain in the case that states had been filled independently. Alas, convenient though they would be, such hopes and desires are groundless. More precisely stated:

> When estimating uncertainties in a histogram generated by a Markov chain, one may not (i) first estimate per-bin uncertainties by assuming that the events in the Markov history were uncorrelated, and then (ii) inflate those uncertainties by a single global correction factor to account for sample-to-sample correlations.

The correctness of the above statement may be made manifest through the creation of a chain which visits each one of its states with equal probability, but which has one or more states whose frequency of visits has a variance which is different to that of one or more of the other states. It is not possible to demonstrate the above with a chain having fewer than three states. Let us consider, therefore, the matrix

$$P_\rho = \begin{pmatrix} 0 & 1-\rho & \rho \\ 1-\rho & 0 & \rho \\ \rho & \rho & 1-2\rho \end{pmatrix}$$

defined for some $0 < \rho < \frac{1}{2}$. It may easily be verified that the matrix is a left-stochastic matrix[1] representing a Markov chain M_ρ with a stationary distribution that is uniform on three discrete states.[2] If a history is generated from this Markov chain by starting at a randomly chosen state (each being equally likely), and then continuing until N locations have been visited, then since the stationary distribution is uniform the *expected* number of visits to each state b is simply $\langle n_b \rangle = N/3$ for all $b \in \{1,2,3\}$. Must the variance of these n_b values likewise be the same for all b? No. Concretely, for the case $\rho = 1/5$ we have

$$P_{1/5} = \begin{pmatrix} 0 & \frac{4}{5} & \frac{1}{5} \\ \frac{4}{5} & 0 & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{3}{5} \end{pmatrix}$$

and variances predicted by equation (**??**) are found to be:

$$\begin{pmatrix} \mathrm{Var}(n_1) \\ \mathrm{Var}(n_2) \\ \mathrm{Var}(n_3) \end{pmatrix} \equiv \begin{pmatrix} V_{1N}[M_{\frac{1}{5}}] \\ V_{2N}[M_{\frac{1}{5}}] \\ V_{3N}[M_{\frac{1}{5}}] \end{pmatrix} = \begin{pmatrix} \frac{100}{675}N - \frac{10}{243} + \epsilon_1(N) \\ \frac{100}{675}N - \frac{10}{243} + \epsilon_2(N) \\ \frac{350}{675}N - \frac{40}{81} + \epsilon_3(N) \end{pmatrix}.$$

For the particular value of $N = 675$, the specific values are

$$\begin{pmatrix} \mathrm{Var}(n_1, N=675) \\ \mathrm{Var}(n_2, N=675) \\ \mathrm{Var}(n_3, N=675) \end{pmatrix} = \begin{pmatrix} 100 - \frac{10}{243} + O(10^{-67}) \\ 100 - \frac{10}{243} + O(10^{-67}) \\ 350 - \frac{40}{81} + O(10^{-269}) \end{pmatrix} \approx \begin{pmatrix} 99.96 \\ 99.96 \\ 349.51 \end{pmatrix}.$$

The variance in the number of times the third state is visited is more than three times larger than the variance in the number of times the first or second states are visisted. The same result may be verified by computer simulation.

[1] It is a left-stochastic matrix if its columns sum to one and all elements are non-negative.

[2] The stationary distribution is uniform since $P_\rho \vec{\pi} = \vec{\pi}$ with $\vec{\pi} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)^T$.

10.1 A uniform symmetric check

Using $N = 675$ with the stochastic matrix

$$P_{1/5} = \begin{pmatrix} 0 & \frac{4}{5} & \frac{1}{5} \\ \frac{4}{5} & 0 & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{3}{5} \end{pmatrix} \qquad \text{which has} \qquad \vec{\pi} = \frac{1}{3}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

we find by simulation that

$$\begin{pmatrix} \mathrm{Cov}[n_1, n_1] & \mathrm{Cov}[n_2, n_1] & \mathrm{Cov}[n_3, n_1] \\ \mathrm{Cov}[n_1, n_2] & \mathrm{Cov}[n_2, n_2] & \mathrm{Cov}[n_3, n_2] \\ \mathrm{Cov}[n_1, n_3] & \mathrm{Cov}[n_2, n_3] & \mathrm{Cov}[n_3, n_3] \end{pmatrix} = \begin{pmatrix} 99.93 \pm 0.03 & 74.79 \pm 0.02 & -174.72 \pm 0.04 \\ 74.79 \pm 0.02 & 99.93 \pm 0.02 & -174.72 \pm 0.04 \\ -174.72 \pm 0.04 & -174.72 \pm 0.04 & 349.44 \pm 0.07 \end{pmatrix}$$

which is in good agreement with the theoretical predictions:

$$\begin{pmatrix} \mathrm{Cov}[n_1, n_1] & \mathrm{Cov}[n_2, n_1] & \mathrm{Cov}[n_3, n_1] \\ \mathrm{Cov}[n_1, n_2] & \mathrm{Cov}[n_2, n_2] & \mathrm{Cov}[n_3, n_2] \\ \mathrm{Cov}[n_1, n_3] & \mathrm{Cov}[n_2, n_3] & \mathrm{Cov}[n_3, n_3] \end{pmatrix} \approx \begin{pmatrix} 99.959 & 74.794 & -174.753 \\ 74.794 & 99.959 & -174.753 \\ -174.753 & -174.753 & 349.506 \end{pmatrix}.$$

10.2 A non-uniform, non-symmetric check

A non-symmetric, non-uniform test using $N = 10$ with the stochastic matrix

$$P = \begin{pmatrix} 0 & \frac{2}{5} & \frac{1}{5} \\ \frac{4}{5} & 0 & \frac{2}{5} \\ \frac{1}{5} & \frac{3}{5} & \frac{2}{5} \end{pmatrix}, \qquad \text{which has} \quad \vec{\pi} = \frac{1}{40}\begin{pmatrix} 9 \\ 14 \\ 17 \end{pmatrix}$$

was observed in simulation to have variances:

$$\begin{pmatrix} \mathrm{Cov}[n_1, n_1] & \mathrm{Cov}[n_2, n_1] & \mathrm{Cov}[n_3, n_1] \\ \mathrm{Cov}[n_1, n_2] & \mathrm{Cov}[n_2, n_2] & \mathrm{Cov}[n_3, n_2] \\ \mathrm{Cov}[n_1, n_3] & \mathrm{Cov}[n_2, n_3] & \mathrm{Cov}[n_3, n_3] \end{pmatrix} = \begin{pmatrix} 1.1502 \pm 0.0005 & 0.16478 \pm 0.0003 & -1.3149 \pm 0.0007 \\ 0.1648 \pm 0.0003 & 0.83845 \pm 0.0003 & -1.0032 \pm 0.0005 \\ -1.3149 \pm 0.0007 & -1.00322 \pm 0.0005 & 2.3182 \pm 0.0011 \end{pmatrix}$$

which is in agreement with the theoretical predictions:

$$\begin{pmatrix} \mathrm{Cov}[n_1, n_1] & \mathrm{Cov}[n_2, n_1] & \mathrm{Cov}[n_3, n_1] \\ \mathrm{Cov}[n_1, n_2] & \mathrm{Cov}[n_2, n_2] & \mathrm{Cov}[n_3, n_2] \\ \mathrm{Cov}[n_1, n_3] & \mathrm{Cov}[n_2, n_3] & \mathrm{Cov}[n_3, n_3] \end{pmatrix} \approx \begin{pmatrix} 1.15025 & 0.165032 & -1.31528 \\ 0.165032 & 0.838874 & -1.00391 \\ -1.31528 & -1.00391 & 2.31919 \end{pmatrix}.$$