

8. Pólya's Urn and the Beta-Bernoulli Process

In this section we will study two probability models that are interesting and important individually. The fact that there are deep connections between the two processes, of course, makes them even more important. Pólya's urn scheme is a dichotomous sampling model that generalizes the [hypergeometric model](#) (sampling without replacement) and the [Bernoulli model](#) (sampling with replacement). The beta-Bernoulli process is obtained by “randomizing” the success parameter p in the Bernoulli trials process with a [beta distribution](#). For certain values of the parameters, the two processes are equivalent, an interesting and surprising result..

Pólya's Urn Process

Suppose that we have an urn (what else!) that initially contains a red and b green balls, where a and b are positive integers. At each discrete time (trial), we select a ball from the urn and then return the ball to the urn along with c new balls of the same color. Ordinarily, the parameter c is a nonnegative integer. However, the model actually makes sense if c is a negative integer, if we interpret this to mean that we *remove* the balls rather than add them, and assuming that there are enough balls of the proper color in the urn. In any case, the random process is known as [Pólya's urn process](#), named for [George Pólya](#).

- ❖ 1. In terms of the colors of the selected balls, Pólya's urn scheme generalizes the standard models of sampling with and without replacement. Note that
- $c = 0$ corresponds to sampling with replacement.
 - $c = -1$ corresponds to sampling without replacement.

For the most part, we will assume that c is nonnegative so that the process can be continued indefinitely. Occasionally we consider the case $c = -1$ so that we can interpret the results in terms of sampling without replacement.

The Outcome Variables

Let X_i denote the color of the ball selected at time i , where 0 denotes green and 1 denotes red. Mathematically, our basic random process is the sequence of [indicator variables](#):

$$X = (X_1, X_2, X_3, \dots)$$

As with any random process, our first goal is to compute the [finite dimensional distributions](#) of X . That is, we want to compute the [joint distribution](#) of (X_1, X_2, \dots, X_n) for each $n \in \mathbb{N}_+$.

Some additional notation will really help. Recall the generalized permutation formula in our study of [combinatorial structures](#): for $r \in \mathbb{R}$, $s \in \mathbb{R}$, and $j \in \mathbb{N}$, we defined

$$r^{(s,j)} = r(r+s)(r+2s) \cdots (r+(j-1)s)$$

As usual, we adopt the convention that a product over an empty index set is 1. Hence $r^{(s,0)} = 1$ for every r and s .

2. Recall that

- $r^{(0,j)} = r^j$
- $r^{(-1,j)} = r^{(j)} = r(r-1) \cdots (r-j+1)$
- $r^{(1,j)} = r(r+1) \cdots (r+j-1)$
- $r^{(r,j)} = j! r^j$.
- $1^{(1,j)} = j!$.

The finite dimensional distributions are easy to compute using the multiplication rule of [conditional probability](#). If we know the contents of the urn at any given time, then the probability of an outcome at the next time is all but trivial.

3. Let $(x_1, x_2, \dots, x_n) \in \{0, 1\}^n$ and let $k = x_1 + x_2 + \cdots + x_n$. Show that

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \frac{a^{(c,k)} b^{(c,n-k)}}{(a+b)^{(c,n)}}$$

The joint probability in the previous exercise just depends on the number of red balls k . Thus, the joint distribution is invariant under a permutation of the coordinates, and hence \mathbf{X} is an [exchangeable sequence](#). Of course the joint distribution reduces to the formulas we have obtained earlier in the special cases of sampling with replacement ($c = 0$) or sampling without replacement ($c = -1$), although in the latter case we must have $n \leq a + b$.

4. Show that $\mathbb{P}(X_i = 1) = \mathbb{E}(X_i) = \frac{a}{a+b}$ for every i .

Thus \mathbf{X} is a sequence of identically distributed variables, quite surprising at first but of course inevitable for any exchangeable sequence. Compare the joint and marginal distributions. Note that \mathbf{X} is an [independent sequence](#) if and only if $c = 0$, when we have simple sampling with replacement. Pólya's urn is one of the most famous examples of a random process in which the outcome variables are exchangeable, but dependent (in general).

Next, let's compute the [covariance and correlation](#) of a pair of outcome variables.

5. Suppose that i and j are distinct indices. Show that

- $\mathbb{P}(X_i = 1, X_j = 1) = \mathbb{E}(X_i X_j) = \frac{a}{a+b} \frac{a+c}{a+b+c}$
- $\text{cov}(X_i, X_j) = \frac{a b c}{(a+b)^2 (a+b+c)}$

$$c. \text{ cor}(X_i, X_j) = \frac{c}{a+b+c}$$

Thus, the variables are positively correlated if $c > 0$, negatively correlated if $c < 0$, and uncorrelated (in fact, independent), if $c = 0$. These results certainly make sense when we recall the dynamics of Pólya's urn.

Pólya's urn is described by a sequence of indicator variables. We can study the same derived random processes that we studied with Bernoulli trials: the number of red balls in the first n trials, the trial number of the k^{th} red ball, and so forth.

The Number of Red Balls

The number of red balls selected in the first n trials is

$$Y_n = \sum_{i=1}^n X_i$$

6. Note that

- The number of green balls selected in the first n trials is $n - Y_n$.
- The number of red balls in the urn after the first n trials is $a + c Y_n$.
- The number of green balls in the urn after the first n trials is $b + c(n - Y_n)$.
- The number of balls in the urn after the first n trials is $a + b + c n$.

Of course, $\mathbf{Y} = (Y_0, Y_1, Y_2, \dots)$ is the **partial sum process** associated with \mathbf{X} . The basic analysis of \mathbf{Y} follows easily from our work with \mathbf{X} .

7. Show that

$$\mathbb{P}(Y_n = k) = \binom{n}{k} \frac{a^{(c,k)} b^{(c,n-k)}}{(a+b)^{(c,n)}}, \quad k \in \{0, 1, \dots, n\}$$

The distribution defined by this probability density function is known, appropriately enough, as the **Pólya distribution**. Of course, the distribution reduces to the **binomial distribution** in the case of sampling with replacement ($c = 0$) and to the **hypergeometric distribution** in the case of sampling without replacement ($c = -1$), although again in this case we need $n \leq a + b$. The case where all three parameters are equal is particularly interesting.

8. Suppose that $a = b = c$. Show that Y_n is uniformly distributed on $\{0, 1, \dots, n\}$.

In general, the Pólya family of distributions has a diverse collection of shapes.

9. Start the simulation of the **Pólya Urn Experiment**. Vary the parameters and note the shape of the probability density function. In particular, note when the function is skewed, when the function is symmetric, when the function is unimodal, when the function is monotone, and when the function is

U-shaped. For various values of the parameters, run the simulation 1000 times and note the apparent convergence of the empirical density function to the probability density function.

10. Solve the inequality $\mathbb{P}(Y_n = k) > \mathbb{P}(Y_n = k - 1)$ for k . In particular, show that

- The density function is unimodal if $a > b > c$ and $n > \frac{a-c}{b-c}$.
- The density function is unimodal if $b > a > c$ and $n > \frac{b-c}{a-c}$.
- The density function is U-shaped if $c > a > b$ and $n > \frac{c-b}{c-a}$.
- The density function is U-shaped if $c > b > a$ and $n > \frac{c-a}{c-b}$.
- The density function is increasing if $b < c < a$
- The density function is decreasing if $a < c < b$

Next, let's find the [mean](#) and [variance](#). As usual, our main tools are the facts that the expected value of a sum is the sum of the expected values and that the variance of a sum is the sum of all pairwise covariances. Curiously, the mean does not depend on the parameter c .

11. Show that

- $\mathbb{E}(Y_n) = n \frac{a}{a+b}$
- $\text{var}(Y_n) = n \frac{ab}{(a+b)^2} \left(1 + (n-1) \frac{c}{a+b+c} \right)$

12. Start the simulation of the [Pólya Urn Experiment](#). Vary the parameters and note the shape and location of the mean/standard deviation bar. For various values of the parameters, run the simulation 1000 times and note the apparent convergence of the empirical mean and standard deviation to their distributional counterparts.

13. Explicitly compute the probability density function, mean, and variance of Y_5 when $a = 6$, $b = 4$, and for the following values of $c \in \{-1, 0, 1, 2, 3, 10\}$. Sketch the graph of the density function in each case.

14. Fix a , b , and n , and let $c \rightarrow \infty$. Show that

- $\mathbb{P}(Y_n = 0) \rightarrow \frac{b}{a+b}$
- $\mathbb{P}(Y_n = n) \rightarrow \frac{a}{a+b}$
- $\mathbb{P}(Y_n = k) \rightarrow 0$ for $k \in \{1, 2, \dots, n-1\}$

Thus, the limiting distribution of Y_n is concentrated on 0 and n . The limiting probabilities are just the initial proportion of green and red balls, respectively. Interpret this result in terms of the dynamics of Pólya's urn scheme.

The Proportion of Red Balls

Suppose that c is nonnegative, so that the process continues indefinitely. The proportion of red balls selected in the first n trials is

$$M_n = \frac{Y_n}{n}$$

This is an interesting variable, since a little reflection suggests that it may have a [limit](#) as n increases. Indeed, if $c = 0$, then M_n is just the sample mean corresponding to n Bernoulli trials. Thus, by the [law of large numbers](#), M_n converges to the success parameter $\frac{a}{a+b}$ as $n \rightarrow \infty$ with probability 1.

On the other hand, the proportion of red balls in the urn after n trials is

$$Z_n = \frac{a + c Y_n}{a + b + c n}$$

When $c = 0$, of course, $Z_n = \frac{a}{a+b}$ so that Z_n and M_n have the same limiting behavior.

15. Suppose that $c > 0$. Show that M_n has a limit if and only if Z_n has a limit, and in this case, the limits are the same.

- If the limits are [with probability 1](#).
- If the limits are [in distribution](#).

16. Suppose that $a = b = c$. Show that the distribution of M_n converges to the uniform distribution on the interval $(0, 1)$ as $n \rightarrow \infty$.

More generally, it turns out that when $c > 0$, M_n and Z_n converge with probability 1 to a random variable U that has the [beta distribution](#) with left parameter $\frac{a}{c}$ and right parameter $\frac{b}{c}$. We need the theory of [martingales](#) to derive and understand this result.

The Trial Number of the k^{th} Red Ball

Suppose again that c is nonnegative, so that the process continues indefinitely. For $k \in \mathbb{N}_+$ let

$$V_k = \min \{n \in \mathbb{N}_+ : Y_n = k\}$$

17. The random processes V and Y are inverses of each other in a sense. Show that $V_k = n$ if and only if $Y_{n-1} = k - 1$ and $X_n = 1$, for $k \in \mathbb{N}_+$ and $n \in \mathbb{N}_+$.

18. Suppose that $n \in \mathbb{N}_+$ and $k \in \{0, 1, \dots, n - 1\}$. Show that

$$\mathbb{P}(X_n = 1 | Y_{n-1} = k) = \frac{a + k c}{a + b + (n - 1) c}$$

In particular, if $a = b = c = 1$ then

- $\mathbb{P}(X_n = 1 | Y_{n-1} = k) = \frac{k+1}{n+1}$
- $\mathbb{P}(X_n = 1 | Y_{n-1} = n-1) = \frac{n}{n+1}$

These last probabilities satisfy **Laplace's rule of succession**, another interesting connection. The rule is named for **Pierre Simon Laplace**, and is studied from a different point of view in the section on **Independence**.

▣ 19. Use [Exercises 7](#), [Exercise 17](#), [Exercise 18](#), and the multiplication rule for conditional probability to show that

$$\mathbb{P}(V_k = n) = \binom{n-1}{k-1} \frac{a^{(c,k)} b^{(c,n-k)}}{(a+b)^{(c,n)}}, \quad n \in \{k, k+1, k+2, \dots\}$$

Of course this probability density function reduces to the **negative binomial density function** with trial parameter k and success parameter $p = \frac{a}{a+b}$ when $c = 0$ (sampling with replacement).

▣ 20. Suppose that $a = b = c$. Show that

$$\mathbb{P}(V_k = n) = \frac{k}{n(n+1)}, \quad n \in \{k, k+1, k+2, \dots\}$$

▣ 21. Fix a, b , and k , and let $c \rightarrow \infty$. Show that

- $\mathbb{P}(V_k = k) \rightarrow \frac{a}{a+b}$
- $\mathbb{P}(V_k = n) \rightarrow 0$ for $n \in \{k+1, k+2, \dots\}$

Thus, the limiting distribution of V_k is concentrated on 0 and ∞ . The limiting probabilities at these two points are just the initial proportion of red and green balls, respectively. Interpret this result in terms of the dynamics of Pólya's urn scheme.

The Beta-Bernoulli Process

An interesting thing to do in almost any parametric probability model is to “randomize” one or more of the parameters. Done in a clever way, this often leads to interesting new models and unexpected connections between models. In this subsection we will randomize the success parameter in the Bernoulli trials model.

Suppose that W has the **beta distribution** on the interval $(0, 1)$ with left parameter $a > 0$ and right parameter $b > 0$. Thus, W has probability density function g given by

$$g(p) = \frac{1}{B(a, b)} p^{a-1} (1-p)^{b-1}, \quad 0 < p < 1$$

Next suppose that $X = (X_1, X_2, X_3, \dots)$ is a sequence of indicator random variables with the property that X is a conditionally independent sequence given $W = p$, with

$$\mathbb{P}(X_i = 1 | W = p) = p; \quad 0 < p < 1, i \in \mathbb{N}_+$$

In short, given $W = p$, X is a Bernoulli trials sequence with success parameter p . We will refer to X as the **beta-Bernoulli process** with parameters a and b .

For a statistical application, suppose that we have a Bernoulli trials process (coin tosses for example) with an unknown probability of success. We model the probability of success with the beta distribution; the parameters a and b are selected to incorporate our knowledge (if any) of this probability.

Distributions

What's our first step? Well, of course we need to compute the finite dimensional distributions of X .

22. Let $(x_1, x_2, \dots, x_n) \in \{0, 1\}^n$ and let $k = x_1 + x_2 + \dots + x_n$. Condition on W to show that

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \frac{B(a+k, b+(n-k))}{B(a, b)} = \frac{a^{(1,k)} b^{(1,n-k)}}{(a+b)^{(1,n)}}$$

Thus, if a and b are integers, then the beta-Bernoulli process X is equivalent to Pólya's urn process with parameters a, b , and $c = 1$, quite a beautiful result. In general, the processes are not equivalent. The beta-Bernoulli process is less restrictive in the sense that the parameters a and b need not be integers; it is more restrictive in the sense that c must be 1.

23. Verify that the basic mathematical results for the **Pólya process** also hold for the beta-Bernoulli process, except of course, that a and b can be any positive numbers (not just integers) and that c must be 1.

24. Use **Bayes' theorem** to show that the conditional distribution of W given $Y_n = k$ is beta with left parameter $a+k$ and right parameter $b+(n-k)$.

Thus, the left parameter increases by the number of successes while the right parameter increases by the number of failures. In the language of Bayesian statistics, the original distribution of W is the **prior distribution**, and the conditional distribution of W given $Y_n = k$ is the posterior distribution. The fact that the posterior distribution is beta whenever the prior distribution is beta means that the family of beta distributions is a **conjugate family**. These concepts are studied in more generality in the section on **Bayes Estimators** in the chapter on **Point Estimation**.

25. Run the simulation of the **beta coin experiment** for various values of the parameter. Note how the posterior density changes from the prior density, given the number of heads.