Stat 205B Lecture Notes #6

Lecturer: Yuval Peres, Scribe: Aaron Wagner

February 7, 2002

Disclaimer: These lecture notes have been only lightly proofread. Please inform peres@stat.berkeley.edu *of errors.*

In this lecture, we study exchangeable sequences of random variables. We will apply the results of this study to Pólya's urn, which was introduced in Lecture 2. Recall Pólya's urn begins with r red balls and g green balls at time zero. At each subsequent time step, we remove a ball from the urn, then replace it and add another ball of the same color.

Pólya's urn has the intriguing property that the probability of adding a particular finite sequence of ball colors does not depend on the order of the colors, only the number of reds and greens in the sequence. The following definition generalizes this notion to an arbitrary sequence of random variables.

Definition 1 *A sequence of random variables $\{X_n\}_{n\geq 1}$ (each taking values in an arbitrary space S) is* ***exchangeable*** *if for all $N \in \mathbb{N}$ and any permutation π of the integers $\{1, \ldots, N\}$, the distributions of $(X_1, \ldots, X_N)$ (in S^N) and $(X_{\pi(1)}, \ldots, X_{\pi(N)})$ are identical.*

Exercise 1 *Show that the sequence of ball colors picked from Pólya's urn is exchangeable. Here $S = \{R, G\}$. Durrett performs the calculation on page 241 of his text.*

A sequence of i.i.d. random variables is clearly exchangeable, and evidently exchangeable random variables are marginally identically distributed. However, the variables need not be independent. Suppose we have a bag containing two coins, one that lands heads with probability 1/2, and another that lands heads with probability 1/3. The coins are physically indistinguishable. We choose a coin uniformly at random, then flip it forever. The coin tosses are identically distributed but not independent: observing a head increases the posterior chance that we chose the fair coin. The coin flips are, however, exchangeable.

This fact holds more generally.

Proposition 1 *A mixture of exchangeable random variables is exchangeable. More precisely, if θ is a random variable in Θ with law ν, and $X_n = X_n^{(\theta)}$, where for each $\phi \in \Theta$, $X_n^{(\phi)}$ is exchangeable, then X_n is exchangeable.*

Proof.

$$
\begin{aligned}
P((X_1, \ldots, X_N) \in A) &= \int_\Theta P(X_1^{(\phi)}, \ldots, X_N^{(\phi)} \in A)\, d\nu(\phi) \\
&= \int_\Theta P(X_{\pi(1)}^{(\phi)}, \ldots, X_{\pi(N)}^{(\phi)} \in A)\, d\nu(\phi) \\
&= P(X_{\pi(1)}, \ldots, X_{\pi(N)} \in A). \qquad \square
\end{aligned}
$$

In words, the space of exchangeable sequences is convex. The next result shows that the "extreme points" of this convex set are contained in the space of i.i.d. sequences.

Theorem 1 (de Finetti) *(Finite-state variables) Let $\{X_n\}_{n=1}^{\infty}$ be an exchangeable sequence with values in a finite set S. Then $\{X_n\}$ is a mixture of i.i.d. sequences, i.e., there exist a space Θ_S, i.i.d. sequences $\{(X_n^{(\phi)})\}_{\phi\in\Theta_S}$, and a random variable θ taking values in Θ_S such that*

$$P((X_1,\dots,X_N)\in A)=\int_{\Theta_S} P(X_1^{(\phi)},\dots,X_N^{(\phi)}\in A)\,d\nu(\phi)$$

for all N, where ν is the law of θ.

Remarks. It will turn out that Θ_S is the space of probability measures on S. In particular, if $\#S=2$, then we can take $\Theta_S=[0,1]$. A remarkable corollary to de Finetti's Theorem is that Pólya's urn can be obtained as a mixture of independent coin flips.

Warning. Even though the notion of exchangeability is meaningful for finite sequences of random variables, de Finetti's Theorem requires an infinite sequence. Consider the pair of random variables (X_1,X_2) that equal $(0,1)$ with probability $1/2$ and $(1,0)$ with probability $1/2$. The pair (X_1,X_2) is exchangeable in that $(X_1,X_2)\stackrel{d}{=}(X_2,X_1)$, but it cannot be a mixture of i.i.d. sequences. To see this, suppose $(Y_1^{(p)},Y_2^{(p)})$ are i.i.d. on $\{0,1\}^2$ with $P(Y_1^{(p)}=1)=p$. Then

$$\begin{aligned}
P(Y_1^{(p)}=Y_2^{(p)})&=p^2+(1-p)^2\\
P(Y_1^{(p)}\neq Y_2^{(p)})&=2p(1-p).
\end{aligned}$$

Thus $P(Y_1^{(p)}=Y_2^{(p)})\geq P(Y_1^{(p)}\neq Y_2^{(p)})$. If (X_1,X_2) was a mixture of i.i.d. pairs then we would have $P(X_1=X_2)\geq P(X_1\neq X_2)$, since in that case,

$$\begin{aligned}
P(X_1=X_2)&=\int P(Y_1^{(p)}=Y_2^{(p)})\nu(dp)\\
&=\int p^2+(1-p)^2\nu(dp)\\
&\geq\int 2p(1-p)\nu(dp)\\
&=\int P(Y_1^{(p)}\neq Y_2^{(p)})\nu(dp)\\
&=P(X_1\neq X_2).
\end{aligned}$$

for some probability measure ν on $[0,1]$. But in the present case, $P(X_1\neq X_2)=1$.

We can use this calculation to construct more subtle counterexamples, such as the following. Suppose (X_1,X_2) take values $(0,1)$ and $(1,0)$ with probability

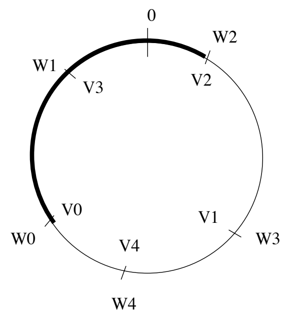

Figure 1: Placement of $V_0, \ldots, V_n$ around a circle.

1/3 each, and values $(0,0)$ and $(1,1)$ with probability 1/6 each. Then (X_1, X_2) are exchangeable but they cannot be a mixture of i.i.d. sequences.

De Finetti's Theorem is still useful in the context of finite sequences because it shows that if a finite sequence of exchangeable random variables can be extended to an infinite sequence of exchangeable random variables, then the infinite sequence, and hence also the finite sequence, is a mixture of i.i.d. sequences.

We return to Pólya's urn. Pólya himself showed that if $r = g = 1$, then the sequence of ball colors is a uniform mixture of i.i.d. binary sequences.

Theorem 2 (Pólya) *Let (X_n) be the sequence of ball colors chosen from Pólya's urn with $r = g = 1$. Let θ be uniformly distributed over $[0,1]$, and for each $\phi \in [0,1]$, let $\{X_n^{(\phi)}\}_{n=1}^{\infty}$ be an i.i.d. sequence of $\{R, G\}$-valued random variables that is independent of θ, with*

$$P(X_1^{(\phi)} = R) = \phi = 1 - P(X_1^{(\phi)} = G).$$

Then $\{X_n\}_{n=1}^{\infty} \stackrel{d}{=} \{X_n^{(\theta)}\}_{n=1}^{\infty}$.

Pólya's proof was entirely computational; now more transparent proofs are available. Here is one, which requires the following lemma.

Lemma 1 *If $V_1, \ldots, V_n$ are i.i.d. uniform on $[0,1]$ and $V^{(1)}, \ldots, V^{(n)}$ are their order statistics, then $V^{(1)} - 0, V^{(2)} - V^{(1)}, \ldots, 1 - V^{(n)}$ are identically distributed (in fact, exchangeable).*

Proof. One way to prove this is by manipulating densities. A more elegant approach is to add another independent uniform $[0,1]$ random variable, V_0, and to think of $V_0, \ldots, V_n$ as living on the circle with circumference 1, say with 0 at the top and increasing clockwise. Starting with V_0 and moving clockwise around the circle, label the first point W_0, the next W_1, etc., as shown in Figure 1. Write $V^{(0)} = 0$. Now $\{(V_i - V_0) \mod 1\}_{i=1}^n$ are i.i.d. uniform over $[0,1]$ and $(V_i - V_0) \mod 1$ is the length of the (clockwise) arc (V_0, V_i) (the clockwise arc (V_0, V_2) is shown in bold in Figure 1). The ith order statistic of $\{(V_i - V_0) \mod 1\}_{i=1}^n$ is the length of the arc (W_0, W_i). Then the distribution of $V^{(i)} - V^{(i-1)}$ is identical to the distribution of the length of the arc (W_{i-1}, W_i), for $i = 1, \ldots, n$, and $1 - V^{(n)}$ equals the length of the arc (W_n, V_0) in distribution. But these arc lengths are identically distributed since V_0 is indistinguishable from $V_1, \ldots, V_n$. □

After proving the theorem, we will provide an intuitive summary of the argument.

Proof of Theorem. Let $\{U_i\}_{i=1}^{\infty}$ be a sequence of i.i.d. random variables that are uniformly distributed over $[0,1]$. Let U be uniformly distributed over $[0,1]$ and independent of $\{U_i\}_{i=1}^{\infty}$. Let

$$Z_n = \begin{cases} G & \text{if } U_n \leq U \\ R & \text{if } U_n > U \end{cases} .$$

Evidently,

$$\{X_n^{(\theta)}\}_{n=1}^{\infty} \stackrel{d}{=} \{Z_n\}_{n=1}^{\infty}.$$

It suffices to show that $\{X_n\}_{n=1}^{N} \stackrel{d}{=} \{Z_n\}_{n=1}^{N}$ for all $N \in \mathbb{N}$. We show this by induction. It is true for $N = 1$ since X_1 and Z_1 are both uniformly distributed over $\{G, R\}$. Suppose it holds for N, and let $c = (c_1, \ldots, c_N)$ be a sequence in $\{G, R\}^N$, and let $\gamma(c) = \#\{i \in \{1, \ldots, N\} : c_i = G\}$. Write $Z = (Z_1, \ldots, Z_N)$ and $X = (X_1, \ldots, X_N)$. We must show that for all c,

$$P(Z_{N+1} = G, Z = c) = P(X_{N+1} = G, X = c)$$
$$P(Z_{N+1} = G | Z = c) = P(X_{N+1} = G | X = c)$$
$$= \frac{\gamma(c) + 1}{N + 2}$$

By Lemma 1, $\{Z_n\}_{n=1}^{\infty}$ is exchangeable. Thus,

$$\begin{aligned} P(Z_{N+1} = G | Z = c) &= \frac{P(Z_{N+1} = G, Z = c)}{P(Z = c)} \\ &= \frac{P(Z_{N+1} = G, \gamma(Z) = \gamma(c))/N!}{P(\gamma(Z) = \gamma(c))/N!} \\ &= P(Z_{N+1} = G | \gamma(Z) = \gamma(c)). \end{aligned}$$

Let $U^{(i)}$ be the ith order statistic of $\{U, U_1, \ldots, U_N\}$. Then

$$\begin{aligned} P(Z_{N+1} = G | \gamma(Z) = \gamma(c)) &= P(U_{N+1} \leq U | U = U^{(\gamma(c)+1)}) \\ &= P(U_{N+1} \leq U^{(\gamma(c)+1)} | U = U^{(\gamma(c)+1)}) \end{aligned}$$

It is simple to show, using symmetry arguments, that the events $U_{N+1} \leq U^{(\gamma(c)+1)}$ and $U = U^{(\gamma(c)+1)}$ are independent. This gives

$$P(Z_{N+1} = G | \gamma(Z) = \gamma(c)) = P(U_{N+1} \leq U^{(\gamma(c)+1)})$$

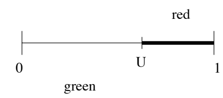

Figure 2: U partitions the unit interval into a green subinterval and a red one.

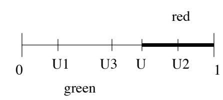

Figure 3: $U_1, U_2, U_3, \ldots$, further partition the $[0,1]$ interval.

Now U_{N+1} and $U^{(\gamma(c)+1)}$ are both continuous random variables, and independent, hence jointly continuous. Let $f(\cdot)$ be the density of $U^{(\gamma(c)+1)}$. Then

$$\begin{aligned} P(U_{N+1} \leq U^{(\gamma(c)+1)}) &= \int_0^1 \int_0^1 1(x \leq y) f(y)\ dx\ dy \\ &= \int_0^1 y f(y)\ dy = E[U^{(\gamma(c)+1)}]. \end{aligned}$$

The lemma shows that $E[U^{(i)}] - E[U^{(i-1)}] = 1/(N+2)$. Thus

$$\begin{aligned} U^{(\gamma(c)+1)} &= U^{(1)} - 0 + \sum_{i=2}^{\gamma(c)+1} U^{(i)} - U^{(i-1)} \\ E[U^{(\gamma(c)+1)}] &= \frac{\gamma(c)+1}{N+2}. \qquad \square \end{aligned}$$

Intuition. We can think of U as partitioning the unit interval into two subintervals, as shown in Figure 2. Label the interval to the left of U "green" and the one to the right "red." Now if U_i lands in the green (resp. red) interval then $Z_i = G$ (resp. $Z_i = R$). By sequentially labeling the U_i on the interval, we form a finer and finer partition of $[0,1]$, as shown in Figure 3. At each time step, the number of green (resp. red) intervals equals the number of green (resp. red) balls in the urn. The length of each of these subintervals corresponds to a difference between consecutive order statistics of $\{U, U_1, \ldots, U_N\}$, so we expect them all to have the same length by the lemma. Then the chance of U_{N+1} landing on the green interval is just the ratio of the number of green intervals to the total, i.e., the fraction of green balls in the urn.

Exercise 2 *Generalize Theorem 2 to Pólya's urn with an arbitrary (positive) initial number of red and green balls. Find the limiting distribution of* $G_k/(R_k + G_k)$.

Remark. In the $r = g = 1$ case, we see that the limiting distribution of $G_k/(R_k+G_k)$ is uniform over $[0,1]$. For general r and g, the limiting distribution will be an order statistic of several i.i.d. uniform $[0,1]$ random variables.