Chris Pyles
Statistics 140
September 4, 2018

Homework 2

1. (a) (ii). X and Y are equal in distribution because, while it is *not* the case that $\forall\omega\left(X(\omega)=Y(\omega)\right)$, the chance of each value of $X=x$ and $Y=y$ are the same (i.e. $\forall x,y\left(P(X=x)=P(Y=y)=\frac{1}{6}\right) \because$ the die is fair).

   (b) (iii). S and $2X$ are not equal in distribution because the possible values of $2X$ are only even, whereas S can have odd values (e.g. if $X=1$ and $Y=2$, then $S=3$). Because there are possible values of S that are not possible for $2X$, they cannot be equal in distribution (and $\therefore$ cannot be equal).

   (c) (i). There are three possible cases for $X=x$ and $Y=y$: 1) $x=y$, 2) $x<y$, and 3) $x>y$.

      (1) If $x=y$, then $S=x+y=2x=2y$. Also, $\max(x,y)=\min(x,y)=x=y$ and $\therefore$ $T=2x=2y$. This implies that $S=T$.

      (2) If $x<y$, then $S=x+y$ and $T=\max(x,y)+\min(x,y)=y+x$, which implies $S=T$ by the commutative property of addition.

      (3) If $x>y$, use similar reasoning as in Case 2 to show that $S=x+y=T \;\rightarrow S=T$.

2. (a) Any of the integers in the population (1, 2, ..., n) are possible values of M, because X and Y can be any integer in the population (even equal, because the population is sampled with replacement). There are n^2 possible combinations of X and Y, because each of the n possible choices of X has n possible choices of Y. Note that saying $M \leq k$ is the same as $(X \leq k) \wedge (Y \leq k)$ (since M is the maximum of both X and Y). This means that $P(M \leq k) = P(X \leq k, Y \leq k)$. If both X and Y must be $\leq k$, then there are k choices for X and Y, which means $P(M \leq k) = P(X \leq k, Y \leq k) = \frac{k^2}{n^2}$.

   (b) From part (a), we know that $P(M \leq k) = P(X \leq k, Y \leq k) = \frac{k^2}{n^2}$. In order to find $P(M = k)$, note that either X or Y must be k, and the other must be $\leq k$. There are three possible scenarios: $(X = k, Y \leq k)$, $(X \leq k, Y = k)$, and $(X = k, Y = k)$. Because $M = k$ is the union of the events $(X = k, Y \leq k)$ and $(X \leq k, Y = k)$, the intersection of which is $(X = k, Y = k)$, the inclusion-exclusion principle says:

   $P(M = k) = P(X = k, Y \leq k) + P(X \leq k, Y = k) - P(X = k, Y = k)$
   $= \frac{1}{n} \cdot \frac{k}{n} + \frac{k}{n} \cdot \frac{1}{n} - \frac{1}{n^2} = \frac{2k-1}{n^2}$

   (c) $\sum\limits_{i=1}^{n} \frac{2i-1}{n^2} = \frac{1}{n^2} \sum\limits_{i=1}^{n} 2i - 1 = \frac{n^2}{n^2} = 1$

3. (c) $N \stackrel{d}{=} \underline{7} - M$. If you look at `prob_table` from 3(a), it is symmetric about the diagonal (e.g. $P(M=3, N=6) = P(M=1, N=4)$). Note that these equal probabilities occur when $N = 7 - M$ and $M = 7 - N$. This is why N is equal in distribution to $7 - M$.

(e) The conditional distribution of M given $N = 5$ has only two nonzero components because if the minimum of the two rolls, N, is 5, then the maximum, M, can have only two possible values: 5 (if $D_1 = D_2 = 5$) or 6 (if $D_1 \neq D_2$). The reason why one of the conditional probabilities is twice as large as the other is because it is twice as likely that $D_1 \neq D_2$ than it is that $D_1 = D_2$, because there are 2 possible outcomes if $D_1 \neq D_2$ (($D_1 = 5, D_2 = 6$) and ($D_1 = 6, D_2 = 5$)), but only 1 if $D_1 = D_2 = 5$ (i.e. $P(D_1 = D_2) = \frac{1}{36}$ whereas $P(D_1 \neq D_2 \wedge N = 5) = \frac{1}{18}$).

4. Let S_k be the event that all faces appear in k rolls of the die and N_i be the even that face i does not appear in k rolls of the die. Because at each roll each face has a probability of appearing of $\frac{1}{6}$, the probability of any face i not appearing in k rolls is $P(N_i) = \left(\frac{5}{6}\right)^k$. The Bonferroni Method says to constrain the sums of the probabilities of the events that make up the complement of S_k, as Boole's Inequality shows that the probability of the union of these events is less than the sum of their individual probabilities. Note that the complement of S_k is the union of the N_i's, or $\overline{S_k} = N_1 \cup N_3 \cup N_3 \cup N_4 \cup N_5 \cup N_6$. This means that to make $P\left(S_k\right) \geq 0.9$, we need $0.10 = P\left(\bigcup\limits_{i=1}^{6} N_i\right) \leq \sum\limits_{i=1}^{6} P(N_i)$.

$$0.10 \leq \sum_{i=1}^{6} P(N_i) = 6 \cdot \left(\tfrac{5}{6}\right)^k$$
$$\log \tfrac{1}{60} = k \log \tfrac{5}{6}$$
$$k = \tfrac{\log 1 - \log 60}{\log 5 - \log 6} \approx 22.457$$

Therefore, in order to make sure that 99% of the time all six faces appear in k rolls, k needs to be at least 23.

5. (a) The complement of this event is the union of three events, that you don't get coupon i of the three coupons in b boxes; call these events N_i. Using the inclusion-exclusion principle, the probability of the union of these three events is:

$$P\left(\bigcup_{i=1}^{3} N_i\right) = P(N_1)+P(N_2)+P(N_3)-P(N_1\cap N_2)-P(N_2\cap N_3)-P(N_1\cap N_3) \\ +P(N_1\cap N_2\cap N_3)$$

The probability of the event that you get all 3 coupons in b boxes (here denoted event C_b) is the probability of its complement taken from 1, or $P(C_b) = 1 - P\left(\bigcup_{i=1}^{3} N_i\right)$. To calculate $P\left(\bigcup_{i=1}^{3} N_i\right)$, first note that $P(N_i) = \left(\frac{2}{3}\right)^b$, as each box contains 1 of 3 coupons, so the chance that you don't get a specific coupon is the $\frac{2}{3}$ chance that you don't get coupon i in a single box to the power of the number of boxes; in this case, b. Next, the probability of not getting coupon i and coupon j in a single box is $\frac{1}{3}$, since there is only 1 coupon of the 3 that can be in the box; this means that $P(N_i \cap N_j) = \left(\frac{1}{3}\right)^b$. Lastly, as every box must contain one of 3 coupons (i.e. there are no boxes without a coupon), the probability that b boxes contain no coupons, the event $N_1 \cap N_2 \cap N_3$, is 0. Using these values, it is possible to find the probability of the union of the N_i's,

$$P\left(\bigcup_{i=1}^{3} N_i\right) = 3\left(\tfrac{2}{3}\right)^b - 3\left(\tfrac{1}{3}\right)^b + 0$$

which means that the probability that you get all 3 coupons in b boxes is $P(C_b) = 1 - 3\left(\frac{2}{3}\right)^b + 3\left(\frac{1}{3}\right)^b$.

(b) Using the same logic as in part (a), finding the formula for the event D_b (that all 6 faces of a die turn up in b rolls) is the same as taking 1 from its complement, the union of the N_i's (the event defined the same as in question 4, but with b rolls instead of k). Using the principle of inclusion-exclusion,

$$P\left(\bigcup_{i=1}^{6} N_i\right) = \sum_{i=1}^{6} P(N_i) - \sum\sum_{1\le i<j\le 6} P(N_i\cap N_j) + \sum\sum\sum_{1\le i<j<k\le 6} P(N_i\cap N_j\cap N_k) \\ - \sum\sum\sum\sum_{1\le i<j<k<l\le 6} P(N_i\cap N_j\cap N_k\cap N_l) + \sum\sum\sum\sum\sum_{1\le i<j<k<l<m\le 6} P(N_i\cap N_j\cap N_k\cap N_l\cap N_m) \\ -P(N_1\cap N_2\cap N_3\cap N_4\cap N_5\cap N_6)$$

For each event, define f_k to be the number of faces that can show if there are k N_i's in the intersection (e.g. $f_4 = 6-4 = 2$ would be the value for the event $N_1\cap N_2\cap N_3\cap N_4$). The event of an intersection of k N_i's results in a probability of that intersection of events occurring of $\left(\frac{f_k}{6}\right)^b$ (e.g. $P\left(N_1 \cap N_2\right) = \left(\frac{4}{6}\right)^b$. Also, note that each term of the formula above as a summation that itself splits into $\binom{6}{k}$ terms; when putting the probabilities into the formula, these will be the coefficients. Using this formula and the other information, we can calculate the probability of the union of the N_i's:

$$\begin{aligned}
P\left(\bigcup_{i=1}^{6} N_i\right) &= \binom{6}{1}\left(\frac{f_1}{6}\right)^b - \binom{6}{2}\left(\frac{f_2}{6}\right)^b + \binom{6}{3}\left(\frac{f_3}{6}\right)^b - \binom{6}{4}\left(\frac{f_4}{6}\right)^b + \binom{6}{5}\left(\frac{f_5}{6}\right)^b - \binom{6}{6}\left(\frac{f_6}{6}\right)^b \\
&= 6\left(\tfrac{5}{6}\right)^b - 15\left(\tfrac{4}{6}\right)^b + 20\left(\tfrac{3}{6}\right)^b - 15\left(\tfrac{2}{6}\right)^b + 6\left(\tfrac{1}{6}\right)^b - 1\left(\tfrac{0}{6}\right)^b \\
&= 6\left(\tfrac{5}{6}\right)^b - 15\left(\tfrac{2}{3}\right)^b + 20\left(\tfrac{1}{2}\right)^b - 15\left(\tfrac{1}{3}\right)^b + 6\left(\tfrac{1}{6}\right)^b
\end{aligned}$$

This formula shows that

$$P(D_b) = 1 - 6\left(\tfrac{5}{6}\right)^b + 15\left(\tfrac{2}{3}\right)^b - 20\left(\tfrac{1}{2}\right)^b + 15\left(\tfrac{1}{3}\right)^b - 6\left(\tfrac{1}{6}\right)^b$$

(c) For a general n, we can use the same logic to find the formula. For this, continue to use the same definition of the event N_i, define k to be the number of events in an intersection, and define f_k to be the number of available options in an event with k N_i's (i.e. $f_k = n - k$). To find the probability of the union of the N_i's, utilize the same logic from part (b) to construct this formula:

$$P\left(\bigcup_{i=1}^{n} N_i\right) = \sum_{i=1}^{n} (-1)^{i+1} \binom{n}{i} \left(\frac{f_k}{n}\right)^b$$

This formula was constructed using the following facts: 1) the inclusion-exclusion principle follows an alternating series format, hence the $(-1)^{i+1}$, 2) the coefficient of the number of terms for a given summation of intersectional probabilities is the same as the number of ways to choose i events from a set of n, hence the $\binom{n}{i}$, and 3) every intersection of k N_i's has the same probability: the number of possible outcomes divided by the total number of outcomes to the power of b, hence the $\left(\frac{f_k}{n}\right)^b$. Finally, to construct the general formula for any n, we take the probability of the union and subtract it from 1:

$$P\left(\text{get all } n \text{ coupons from } b \text{ boxes}\right) = 1 - \sum_{i=1}^{n} (-1)^{i+1} \binom{n}{i} \left(\frac{f_k}{n}\right)^b$$

6. (a) Let A_k be the event that card k is an ace, and N_k the event that card k is not an ace. Using symmetry, the question of whether or not the 4th card is an ace given that the 7th is not is the same as the question of whether the 1st card is not an ace given that the 2nd card is. These probabilities, respectively, are $\frac{d-a}{d}$ and $\frac{a}{d-1}$. $\therefore$ the probability $P(A_4 \mid N_7)$ is $\frac{da-a^2}{d^2-d}$.

   (b) The question of whether or not there are k aces among the first n cards dealt is a question of counting good elements in a simple random sample. The formula for the probability of getting k of a aces in a sample of size n from a population of size d is $\frac{\binom{a}{k}\binom{d-a}{n-k}}{\binom{d}{n}}$.

   (c) The possible values of X are 1, 2, ..., $d-a+1$, because there are $d-a$ non-aces, so the latest possible time for the first ace to appear is at card $d-a+1$. The distribution of X is $P(X = x) = \frac{a\binom{d-a}{x-1}}{(d-x+1)\binom{d}{x-1}}$. This is because, for a specific value of X, there are $d-a$ non-aces, of which we need to choose $X-1$, out of $\binom{d}{X-1}$ possible choices; also, there are a aces when the first ace is dealt, but only $d-X+1$ cards left in the deck.

   (d) The probability $P(X \geq k)$ is $\sum_{i=k}^{d-a+1} \frac{a\binom{d-a}{i-1}}{(d-i+1)\binom{d}{i-1}}$. The reason for those limits is because they are the lowest and highest possible values of X given that $X \geq k$.

   (e) The event $\{X \geq k\}$ is equivalent to the event that the first $k-1$ cards are non-aces. The probability of this event is $\frac{\binom{d-a}{k-1}}{\binom{d}{k-1}}$.

   (f) $\sum_{i=k}^{d-a+1} \frac{a\binom{d-a}{i-1}}{(d-i+1)\binom{d}{i-1}} = \frac{\binom{d-a}{k-1}}{\binom{d}{k-1}}$