CS206 **Review Sheet 3** **October 24, 2018**

After our intense focus on counting, we continue with the study of some more of the basic notions from Probability (though counting will remain in our thoughts). An important concept is that of the

Random Variable: Consider a random experiment $\mathcal{E}$ that has probability space (S, P). A random variable X is a function from the sample space S to the reals. For example in the experiment of tossing a fair coin three times, let X be the profit if you receive a dollar for each Head and pay a dollar for each Tail.

$w \in S$	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
$X(w)$	3	1	1	-1	1	-1	-1	-3

The outcomes of $\mathcal{E}$ (*symbols*) are in the top row; the values of the random variable X (*real numbers*) are in the bottom. Define

$$\text{Range}(X) = \{\text{all possible values of } X\} = \{\text{distinct } X(w) : w \in S\};$$

notice that it is a *set*, not a multi-set. In the example above , $\text{Range}(X) = \{3, 1, -1, -3\} = \{a_1, \ldots, a_4\}$.

- Fact 1: $|\text{Range}(X)| \leq |S|$: the equality can occur only if X takes a distinct value for each $w \in S$; otherwise there are outcomes $w_1 \neq w_2$ in S for which $X(w_1) = X(w_2)$. In the example above $|S| = 8$, twice the size of $\text{Range}(X)$.

- Fact 2: Suppose $\text{Range}(X) = \{a_1, \ldots, a_k\}$. The events $A_i = \{w \in S : X(w) = a_i\}$, $i = 1, \ldots, k$, partition S. They form what is called *the partition induced by X*, and we write

  $$\mathcal{A}_X = \{A_1, \ldots, A_k.$$

  For each $a_i \in \text{Range}(X)$ we define

  $$f_X(a_i) = P(A_i) = P(\{w \in S : X(w) = a_i\}).$$

  Also, if $t \in R$ is NOT an element of $\text{Range}(X)$, we define $f_X(t) = 0$. The function f_X is called the frequency function of X (some books call it the probability mass function).

- Fact 3: Because $\mathcal{A}_X$ partitions S,

  $$\sum_{a_i \in \text{Range}(X)} f_X(a_i) = 1.$$

  Therefore f_X is a probability measure on $\text{Range}(X)$ and we now understand that *the random variable X has mapped the original probability space (S, P) into a new one $(\text{Range}(X), f_X)$*.

Independence: Random variables X and Z are independent iff their partitions $\mathcal{A}_X$ and $\mathcal{A}_Z$ are. For this we need $P(A \cap B) = P(A)P(B)$ for each $A \in \mathcal{A}_X$ and each $B \in \mathcal{A}_Z$ (so no hint about a value of X alters your assessment of probabilities for values of Z. Pairwise, k-wise, and mutual independence for a sequence $X_1, \dots, X_n$ of random variables is defined via the events in their partitions.

Independent Trials: We have an experiment $\mathcal{E}$ with sample space $T = \{t_1, \dots, t_k\}$ and we use probability P on T. $\mathcal{E}$ is performed n times in succession, each under identical conditions. We will refer to this sequence of repetitions of the experiment as a composite experiment $\mathcal{E}^{(n)} = \mathcal{E}_1, \dots, \mathcal{E}_n$, where $\mathcal{E}_i$ denotes the i^{th} performance. The composite sample space for the n repetitions

$$S^{(n)} = \{\underline{w} = (w_1, \dots, w_n) | w_i \in T \text{ is the outcome of } \mathcal{E}_i\} = \underbrace{T \times \cdots \times T}_{n \text{ times}}.$$

- Fact 1: $|S^{(n)}| = |T|^n = k^n$.

Product Probability: The points of $S^{(n)}$ can be assigned probabilities in infinitely many ways. We would like to do it in such a way that both

1. The original probability P is respected: For every trial $\mathcal{E}_i$ and every $t_j \in T$, the probability that t_j occurs on the i^{th} repetition should be $P(t_j)$; in other words, we want

$$\text{Prob}\{\underline{w} = (w_1, \dots, w_n) \in S^{(n)} : w_i = t_j\} = P(t_j), \tag{1}$$

   and

2. the repetitions are independent.

The product probability measure $P^{(n)}$ on $S^{(n)}$ is defined by

$$P^{(n)}(\underline{w}) = P(w_1)P(w_2)\dots P(w_n), \text{ for all } \underline{w} = (w_1, \dots, w_n) \in S^{(n)}. \tag{2}$$

It is not hard to show that

- Fact 2: $P^{(n)}$ really is a probability on $S^{(n)}$; i.e., $\sum P^{(n)}(\underline{w}) = 1$, the sum over all outcomes in $S^{(n)}$. Also
- Fact 3: Product probability respects P on T; i.e., (1) holds for Prob $= P^{(n)}$. Finally,
- Fact 4: $P^{(n)}$ "captures" the independence of the trials in a very strong way. If $A_1, \dots, A_n$ are events and A_i depends only on the outcome of the i^{th} repetition (i.e., $\underline{w} = (w_1, \dots, w_n) \in A_i$ depends on w_i and not the other coordinates), then the A_i are mutually independent. Also random variables $X_1, \dots, X_n$ are mutually independent as long as the value of $X_i(\underline{w})$ depends only on the w_i. In fact it can be shown that
- Fact 5: If a probability on $S^{(n)}$ satisfies (1) and if events $A_1, \dots, A_n$ are mutually independent as long as the occurrence of A_i depends only on the outcome of the i^{th} trial, then it is product probability $P^{(n)}$, defined in (2).

Bernoulli Trials: $\mathcal{E}$ has two outcomes, success (s) and failure (f). $\mathcal{P} = P(s) = 1 - P(f)$. $\mathcal{E}^{(n)}$ - the repetition of $\mathcal{E}$ n times - is called *n Bernoulli trials with success probability* $\mathcal{P}$. For each trial $i = 1, \ldots, n$ define

$$X_i = \begin{cases} 1 & \text{if trial } i \text{ is success} \\ 0 & \text{if trial } i \text{ is failure} \end{cases}$$

the indicator (of success) for the i^{th} trial, and

$$S_n = X_1 + \cdots + X_n$$

measures the number of successes that occur in the n trials.

- Fact 1: Range$(S_n) = \{0, 1, \ldots, n\}$. The equation

$$f_{S_n}(k) = \text{Prob}(S_n = k) = \binom{n}{k} \mathcal{P}^k (1 - \mathcal{P})^{n-k} \tag{3}$$

  defines the binomial frequency function, $k = 0, 1, \ldots, n$. [The product probability of a *particular* sequence of n trials that results in k successes is $\mathcal{P}^k(1-\mathcal{P})^{n-k}$. There are $\binom{n}{k}$ such sequences, one for each distinct way to choose which k trials result in success].

- Fact 2: It increases up to $n\mathcal{P}$ and decreases thereafter.

Infinite Sequences of Bernoulli Trials: Let $\mathcal{E}$ be a Bernoulli trial experiment with success probability $\mathcal{P}$.

1. ($k = 1$) Let $\mathcal{E}'$ be the experiment "repeat $\mathcal{E}$ until a succuss occurs". The composite sample space is $S' = \{s, fs, ffs, \ldots\}$. $|S'| = \infty$. The interesting random variable is W_1, the number of repetitions; i.e., *the waiting time for the first success.*

   - Fact 1: Range$(W_1) = \{1, 2, \ldots\}$. The equation

   $$f_{W_1}(n) = \text{Prob}(W_1 = n) = (1 - \mathcal{P})^{n-1} \mathcal{P} \tag{4}$$

     defines the geometric frequency function, $n = 1, 2, \ldots$.

2. (general case) Now let $\mathcal{E}'$ be the experiment "repeat $\mathcal{E}$ until the k^{th} succuss occurs", $k \geq 1$. The composite sample space is $S' = \{\underbrace{s \cdots s}_{k}, \underbrace{fs \cdots s}_{k+1}, \underbrace{sfs \cdots s}_{k+1}, \ldots, \underbrace{s \cdots sfs}_{k+1}, \ldots\}$. $|S'| = \infty$. The interesting random variable is W_k, the number of repetitions; i.e., *the waiting time for the k^{th} success.*

   - Fact 2: Range$(W_k) = \{k, k+1, \ldots\}$. The equation

   $$f_{W_k}(n) = \text{Prob}(W_k = n) = \binom{n-1}{k-1} (1 - \mathcal{P})^{n-k} \mathcal{P}^k \tag{5}$$

     defines the negative binomial frequency function, $n = k, k+1, \ldots$.

Expectation: The notion of random variable provides a basic and useful way to study aspects of a random experiment that are of particular interest. One of the most useful properties of a random variable is the *expectation.* Let X be a random variable defined on a probability space (S, P). Its **expectation** (expected value, mean) is defined by

$$E(X) = \sum_{w \in S} X(w)P(w). \tag{6}$$

This is a probability-weighted-average of values of X. Suppose $\text{Range}(X) = \{a_1, \ldots, a_k\}$, and that $\mathcal{A}_X = \{A_1, \ldots, A_k\}$ is the partition induced by X. Breaking the sum in (6) into sums over the events $A_i \in \mathcal{A}_X$, we get

- Fact 1:

$$\begin{aligned} E(X) &= \sum_{w \in A_1} X(w)P(w) + \cdots + \sum_{w \in A_k} X(w)P(w) \\ &= a_1 P(X = a_1) + \cdots + a_k P(X = a_k) \\ &= \sum_{a_i \in \text{Range}(X)} a_i f_X(a_i) \end{aligned}$$

- Fact 2: Expectation is linear; that is, $E(aX + bY + c) = aE(X) + bE(Y) + c$ for any random variables X and Y and reals a, b, c . It is interesting to take $a = b = 0$ and also to take $a = b = 1$, $c = 0$. This latter extends by induction to show

$$E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n). \tag{7}$$

- Fact 3 (mean of the geometric): Let W_1 be the wait for the first success in Bernoulli trials with success probability $\mathcal{P}$. Then $\boxed{E(W_1) = 1/\mathcal{P}}$. (You expect to wait twice as long for an event that is half as likely). This was proved by showing

$$\sum_{n=1}^{\infty} nP(W_1 = n) = \sum_{n=1}^{\infty} n\mathcal{P}(1 - \mathcal{P})^{n-1} = \frac{1}{\mathcal{P}}.$$

- Fact 4 (mean of the negative binomial): Let W_k be the wait for the k^{th} success in Bernoulli trials with success probability $\mathcal{P}$. Then $\boxed{E(W_k) = k/\mathcal{P}}$. This implies the indentity

$$\sum_{n=k}^{\infty} nP(W_k = n) = \mathcal{P}^k \sum_{n=k}^{\infty} n\binom{n-1}{k-1}(1 - \mathcal{P})^{n-k} = \frac{k}{\mathcal{P}}$$

  and is proved probabilistically by noting that $W_k = X_1 + \cdots + X_k$, where X_1 is the wait for the first success and X_{i+1} is the wait for the first success *after the i-th*; use (7) and note that each X_i is geometric.

- Fact 5 (mean of the binomial): Let S_n be the number of successes in n Bernoulli trials with success probability $\mathcal{P}$. Then $\boxed{E(S_n) = n\mathcal{P}}$. This implies the identity

$$\sum_{k=0}^{n} kP(S_n = k) = \sum_{k=0}^{n} k\binom{n}{k}\mathcal{P}^k(1 - \mathcal{P})^{n-k} = n\mathcal{P}$$

  and is proved using indicators: $S_n = X_1 + \cdots + X_n$ where X_i, the indicator (of success) for the i^{th} trial, has $E(X_i) = \mathcal{P}$. Now use (7).

Coupon Collecting: There are n coupon types, each type equally likely. You collect coupons (i.e., sample from the n coupons with replacement) until you have seen r of the types (so its likely you have sampled much more than r times). The expected wait needed to collect r different coupon types is

$$1 + \frac{n}{n-1} + \cdots + \frac{n}{n-r+1} = n\left[\frac{1}{n} + \cdots + \frac{1}{n-r+1}\right];$$

(after you have seen j types you are waiting for an event (a new type) which has probability $\mathcal{P} = (n-j)/n$). An interesting case is $r = n$. The expected wait to collect the whole set of n coupons is about $n \log n$ (natural log), using the fact that $1 + 1/2 + \cdots + 1/n \to \log_e n$.

The Method of Indicators: We already applied this method to establish the mean of the binomial random variable in Fact 5. Here is another example: Let $\mathcal{E}$ be the experiment of placing r balls randomly into n boxes, and let N count the number of empty boxes. The computation of $E(N)$ using

$$E(N) = \sum_{a_i \in \text{Range}(N)} a_i f_N(a_i)$$

could be difficult since you need to compute the entire frequency function f_N. However let X_i be the indicator of the event that box i is empty. $E(X_i) = 1 \cdot P(\text{box } i \text{ is empty}) = (1-1/n)^r$ (why?). Now notice that

$$N = X_1 + \cdots X_n$$

so $E(N) = E(X_1) + \cdots E(X_n)$ by linearity of expectation, and we easily obtain $E(N) = n(1-1/n)^r$.

For another example, in class we applied the same method to show that the expected number of people who get their own hats in the n hat experiment is 1 *independent of n (!!!)*: If X_i is the indicator of the event that person i gets her own hat (so $E(X_i) = 1/n$) and N is the total number who get their own hats, then $N = X_1 + \cdots + X_n$ and $E(N) = E(X_1) + \cdots + E(X_n) = 1$.