CIS6930/4930

Examples and Applications

Prof. Ye Xia

Trick of Indicator Random Variables

Suppose there are 25 different types of coupons. Suppose that each time one collects a coupon, it is equally likely to be any one of the 25 types. Compute the expected number of different types when 10 coupons are collected.

Let X denote the number of different types in the set of 10 coupons. For $i = 1, 2, \ldots, 25$, define

$$X_i = \begin{cases} 1, & \text{if type } i \text{ coupon shows up in the set of 10 coupons} \\ 0, & \text{otherwise.} \end{cases}$$

We have $X = \sum_{i=1}^{25} X_i$ and then

$$E[X] = \sum_{i=1}^{25} E[X_i] = \sum_{i=1}^{25} P(X_i = 1).$$

For coupon i, think about running IID Bernoulli trials 10 times. A success corresponds to coupon i being collected in a trial, which has the probability $1/25$.

$$\begin{aligned} P(X_i = 1) =& 1 - P(X_i = 0) \\ =& 1 - P(\text{type } i \text{ coupon is not in the set of 10 coupons}) \\ =& 1 - \left(\frac{24}{25}\right)^{10}. \end{aligned}$$

Therefore, $E[X] = 25\left(1 - \left(\frac{24}{25}\right)^{10}\right)$.

Another way is to think about the sample space Ω: Each element of the sample space is a set of 10 coupons collected. There are 25^{10} elements in Ω.

The probability assignment is uniform.

Fix a coupon type i. The number of elements in Ω that do not have a type-i coupon is equal to 24^{10}. Therefore,

$$P(X_i = 0) = \frac{24^{10}}{25^{10}}.$$

THE Coupon Collection Problem

Suppose there are n different types of coupons. Suppose that each time one collects a coupon, it is equally likely to be any one of the n types. What is the expected number of coupons that need to be collected before a complete set of n types is collected?

Let X be the number of coupons that need to be collected. We will try to break down X into a sum of random variables based on some significant events.

An interesting kind of events is when something new occurs: here, when a new type coupon is first collected.

Let X_i be the additional number of coupons needed to obtain a new type after $i-1$ types have been collected. Then, $X = \sum_{i=1}^{n} X_i$.

Each X_i is a geometric random variable with $p = \frac{n-(i-1)}{n}$. Why? After $i-1$ types have been collected, the collection of each additional coupon can be thought as a Bernoulli trial. A success means a new type is collected, which has the probability $\frac{n-(i-1)}{n}$. X_i is the number of trials till the first success.

We have $E[X_i] = \frac{1}{p} = \frac{n}{n-(i-1)}$. Therefore,

$$\begin{aligned} E[X] =& \sum_{i=1}^{n} \frac{n}{n-(i-1)} \\ =& n \sum_{i=1}^{n} \frac{1}{i}. \end{aligned}$$

For large n, $\sum_{i=1}^{n} \frac{1}{i} \approx \log n$; therefore, $E[X] \approx n \log n$.

Expectation of Nonnegative Random Variables

Suppose N a is nonnegative and integer-valued random variable. We will compute $E[N]$.

For $n = 1, 2, \ldots$, define indicator random variables I_n by

$$I_n = \begin{cases} 1, & \text{if } n \le N \\ 0, & \text{if } n > N. \end{cases}$$

In the old notation, each I_n is written as $I_{\{N \ge n\}}$. Things may get confusing here. The view of a random variable being a function is helpful. The definition of I_n is really, for each $\omega \in \Omega$,

$$I_n(\omega) = \begin{cases} 1, & \text{if } n \le N(\omega) \\ 0, & \text{if } n > N(\omega). \end{cases}$$

For each fixed ω, $N(\omega)$ is a fixed integer.

To continue, we have $\sum_{n=1}^{\infty} I_n = \sum_{n=1}^{N} I_n = N$.

$$
\begin{aligned}
E[N] =& E[\sum_{n=1}^{\infty} I_n] \\
=& \sum_{n=1}^{\infty} E[I_n] \\
=& \sum_{n=1}^{\infty} P(N \geq n).
\end{aligned}
$$

Unlike the case of a finite sum, the interchange of the expectation with an infinite sum always requires justification, e.g., by Fubini's Theorem.

If the summands are all nonnegative, the interchange is allowed; however, the result may be infinity.

Trapped Miner

A miner is trapped in a mine containing three doors. Door 1 leads to safety after 2 hours of traveling. Door 2 leads to where he started after 3 hours of traveling. Door 3 leads to where he started after 5 hours. Assume that the miner always chooses a door with equal probability (duh...). What is the expected length of time until he reaches safety?

Let X be the number hours until he reaches safety. We will compute $E[X]$ by conditioning on the door he chooses in the first time. Let Y represent the door he chooses.

$$\begin{aligned}E[X] =& E[X|Y = 1]P(Y = 1) + E[X|Y = 2]P(Y = 2)\\ &+ E[X|Y = 3]P(Y = 3)\\ =& \frac{1}{3}E[X|Y = 1] + \frac{1}{3}E[X|Y = 2] + \frac{1}{3}E[X|Y = 3].\end{aligned}$$

Now

$$\begin{aligned}E[X|Y = 1] =& 2\\ E[X|Y = 2] =& 3 + E[\tilde{X}|Y = 2]\\ E[X|Y = 3] =& 5 + E[\tilde{X}|Y = 3],\end{aligned}$$

where $\tilde{X}$ is the number of additional hours until reaching safety after returning where he started.

If return happens, the scenario has a **recurring** structure. After returning to where he started, the same process starts all over and it is independent of the past. Therefore, $\tilde{X}$ is independent of $Y = 2$ or $Y = 3$, and it has the same distribution as X. We have

$$E[\tilde{X}|Y = 2] = E[\tilde{X}] = E[X]$$

$$E[\tilde{X}|Y = 3] = E[\tilde{X}] = E[X].$$

Now,

$$E[X] = \frac{2}{3} + \frac{1}{3}(3 + E[X]) + \frac{1}{3}(5 + E[X]).$$

Solving for $E[X]$, we get $E[X] = 10$.

Sum of A Random Number of Random Variables

Suppose there is a random number N of accidents per year in a plant. Suppose the numbers of injured workers in each accident are independent random variables with a common mean μ. Suppose the number of accidents per year is independent of the number of injured workers in each accident. What is the expected number of injured workers per year?

Let X_i be the number of injured workers in accident i. The number of injured workers per year can be expressed as $\sum_{i=1}^{N} X_i$.

We will compute $E[\sum_{i=1}^{N} X_i]$ by first conditioning on N.

$$E[\sum_{i=1}^{N} X_i] = E[E[\sum_{i=1}^{N} X_i|N]].$$

$$
\begin{aligned}
E[\sum_{i=1}^{N} X_i | N = n] =& E[\sum_{i=1}^{n} X_i | N = n] \\
=& E[\sum_{i=1}^{n} X_i] \\
=& nE[X_i] = n\mu.
\end{aligned}
$$

Then,

$$
E[\sum_{i=1}^{N} X_i] = E[E[\sum_{i=1}^{N} X_i | N]] = E[N\mu] = \mu E[N].
$$

In the above, we assume the expectations involved all exist.

Gambler's Ruin

Consider a gambler who started with k dollars. At each play of the game, he wins or loses 1 dollar with probability $1/2$ each. His goal is to reach N dollars, where $N \geq k$. What is the probability that he goes broke before reaching the goal of N dollars?

Let p_k be the desired probability. We will condition on the outcome of the first play. If he wins the first game, he will face the same game setup but with an initial wealth of $k+1$. The probability of going broke before reaching N is p_{k+1}. Similarly, if he loses the first game, the probability of going broke before reaching N is p_{k-1}. Here, we assumed $0 < k < N$. We have

$$p_k = \frac{1}{2} p_{k+1} + \frac{1}{2} p_{k-1}.$$

This is a (second-order) linear difference equation, the discrete counterpart of second-order linear differential equations. There are systematic ways to solve p_k as a function of k. But, the problem here has special structure and it can be solved more simply.

When, $k = 0$ or N, we have the boundary conditions $p_0 = 1$ and $p_N = 0$.

The difference equation can be written as

$$p_{k+1} - p_k = p_k - p_{k-1}.$$

Let $b_k = p_k - p_{k-1}$, for $1 \le k \le N$. Then, we have $b_{k+1} = b_k$, and therefore, $b_k = b_1$ for $1 \le k \le N$. Then, we have

$$p_k = b_1 + p_{k-1} = 2b_1 + p_{k-2} = \cdots = kb_1 + p_0.$$

From $0 = p_N = Nb_1 + p_0 = Nb_1 + 1$, we get $b_1 = -1/N$. Therefore,

$$p_k = 1 - k/N, \;\; 0 \le k \le N.$$

As $N \to \infty$, $p_k \to 1$. When N is large, the gambler will very likely go broke.

This is a symmetric random walk with two absorbing barriers.

The Ballot Problem

In an election, candidate A received n votes and candidate B received m votes, where $n > m$. Suppose all sequences of votes are equally likely. Show that the probability that A is always ahead of B in the vote count is equal to $(n-m)/(n+m)$.

Let $p_{n,m}$ denote the desired probability. When using the conditioning trick on an event involving a sequence of actions, we often condition on either the first action or the last action. In this case, we will condition on which candidate received the last vote counted.

$$\begin{aligned} p_{n,m} =& P(A \text{ always head} \mid A \text{ receives the last vote}) \frac{n}{n+m} \\ &+ P(A \text{ always head} \mid B \text{ receives the last vote}) \frac{m}{n+m}. \end{aligned}$$

Argument for the probabilities $\frac{n}{n+m}$ and $\frac{m}{n+m}$: Let $Y_i = 1$ if vote i is for A, and $Y_i = 0$ otherwise. Then, $n = E[\sum_{i=1}^{m+n} Y_i] = \sum_{i=1}^{m+n} P(Y_i = 1)$. Since there is no real difference among the i's, $P(Y_i = 1)$ is the same for all i. Therefore,

$P(Y_i = 1) = \frac{n}{n+m}$ for every i. One can also argue this by counting the number of outcomes, i.e., different sequences of votes. Or, one can consider reverse sequence of voting one by one, starting from the last vote.

Given that A receives the last vote, the probability A is always ahead is the same as the event that A received $n-1$ votes, B received m votes and A is always ahead in the count of the $n-1+m$ votes. That probability is equal to $p_{n-1,m}$.

Given that B receives the last vote, the probability A is always ahead is the same as the event that A received n votes, B received $m-1$ votes and A is always ahead in the count of the $n+m-1$ votes. That probability is equal to $p_{n,m-1}$. Then, we have, for $m \geq 1$,

$$p_{n,m} = \frac{n}{n+m} p_{n-1,m} + \frac{m}{n+m} p_{n,m-1}. \tag{1}$$

We will show $p_{n,m} = (n-m)/(n+m)$ by induction on $n+m$.

For $m = 0$, the situation is trivial because $p_{n,0} = 1$ for any $n \geq 1$, and $\frac{n-m}{n+m} = 1$.

When $n+m = 1$, implying $n = 1$ and $m = 0$, it is clear that $p_{n,m} = 1 = (n-m)/(n+m)$.

Subsequently, we assume $m \geq 1$.

Now, suppose the result holds for $n + m = k$, where $k > 1$.

When $n + m = k + 1$, we have $n + m - 1 = k$, and hence, $p_{n-1,m} = \frac{n-1-m}{n-1+m}$ and $p_{n,m-1} = \frac{n-(m-1)}{n+m-1}$ by the induction hypothesis. Then, from (1), we get

$$
\begin{aligned}
p_{n,m} &= \frac{n}{n+m} \frac{n-1-m}{n-1+m} + \frac{m}{n+m} \frac{n-(m-1)}{n+m-1} \\
&= \frac{n^2 - n - m^2 + m}{(n+m)(n-1+m)} \\
&= \frac{(n+m-1)(n-m)}{(n+m)(n-1+m)} \\
&= \frac{n-m}{n+m}.
\end{aligned}
$$

Application of The Ballot Problem

The result of the ballot problem is quite useful.

Consider a sequence of coin tosses where the coin lands on heads with probability p in each toss. Let X be the first time the total number of heads is equal to the total number of tails. What is the probability mass function of X?

It is clear that we only need to consider even numbers for X. We will compute $P(X = 2n)$ for $n = 1, 2, \ldots$ We will condition on the number of heads in the first $2n$ tosses.

Let Y_{2n} be the number of heads in the first $2n$ tosses. Then,

$$P(X = 2n) = \sum_{k=1}^{2n} P(X = 2n | Y_{2n} = k) P(Y_{2n} = k).$$

For $k \neq n$, $P(X = 2n|Y_{2n} = k) = 0$. We have

$$\begin{aligned} P(X = 2n) =& P(X = 2n|Y_{2n} = n)P(Y_{2n} = n) \\ =& P(X = 2n|Y_{2n} = n)\binom{2n}{n}p^n(1-p)^n. \end{aligned}$$

The first equality above really says $P(X = 2n) = P(X = 2n, Y_{2n} = n)$. This must be so because the events $\{X = 2n\}$ and $\{X = 2n, Y_{2n} = n\}$ are the same. Each implies the other.

Important Side Note: In general, consider two events A and B. Suppose the occurrence of A implies that B must occur (we say A implies B, denoted by $A \implies B$). Then, as sets, $A \subseteq B$, and therefore, $P(A) \leq P(B)$.

This is so because when we say A occurs, we mean the experimental outcome ω could be any element of A, i.e., $\omega \in A$. When the two sets are such that $A \implies B$, any element in A must be also in B so that whichever $\omega \in A$ is the experimental outcome, it must be also in B to make B also occur; hence, $A \subseteq B$.

Many inequalities are derived this way. When it is difficult to compute $P(A)$, one often relaxes A to create another event B with $A \subseteq B$, with the requirements that (i) it is easier to compute $P(B)$ and (ii) $P(B)$ is close to $P(A)$.

To show two events A and B are the same (i.e., $A = B$), we usually show $A \implies B$ and $B \implies A$. In effect, we are showing $A \subseteq B$ and $B \subseteq A$.

To compute $P(X = 2n|Y_{2n} = n)$, we further condition on the outcome of the $2n$-th toss. If it is heads, to have $X = 2n$, tails must have been in the lead in all the first $2n - 1$ tosses, the probability of which is $p_{n,n-1}$ as in the ballot problem.

Similarly, if the $2n$-th toss lands on tails, then heads must have been in the lead in all the first $2n - 1$ tosses, the probability of which is also $p_{n,n-1}$.

Putting these two cases together (weighted average by some probabilities q and $1 - q$; in fact $q = 1/2$ but it doesn't matter), the result is

$P(X = 2n|Y_{2n} = n) = p_{n,n-1}$. Therefore,

$$P(X = 2n) = p_{n,n-1} \binom{2n}{n} p^n (1-p)^n$$
$$= \frac{1}{2n-1} \binom{2n}{n} p^n (1-p)^n.$$

Optimal Decision

Suppose that n prizes are sequentially presented. Each prize must be accepted or rejected upon its presentation and no more than one prize can be accepted. The ranking of the prizes seen are known. Suppose the sequence of presentation is a permutation drawn uniformly at random from all the $n!$ possibilities. The goal is to maximize the probability of getting the best prize among the following family of policies:

Let the first k prizes go by. Then, choose the first upcoming prize that is better than all the ones seen.

We will need to decide the optimal k, i.e., to choose the k that maximizes the desired probability. Let A_k be the event that we will get the best prize with a particular choice of k. We need to compute $P(A_k)$ by conditioning on the position of the best prize, which is denoted by X.

$$P(A_k) = \sum_{i=1}^{n} P(A_k|X = i)P(X = i) = \sum_{i=1}^{n} P(A_k|X = i)\frac{1}{n}.$$

Now, if $i \leq k$, we would have missed the best prize. Then, $P(A_k|X = i) = 0$ for $i \leq k$.

Consider each $i > k$. For A_k to happen, the best prize among the first $i - 1$ must be in the first k. Among the first $i - 1$ prizes, the best among them can happen anywhere with an equal probability. This is true whether we condition on $X = i$ or not. Therefore, $P(A_k|X = i) = k/(i-1)$. Then,

$$
\begin{aligned}
P(A_k) = &\sum_{i=k+1}^{n} \frac{k}{i-1}\frac{1}{n} \\
=&\frac{k}{n}\sum_{i=k}^{n-1}\frac{1}{i} \\
\approx&\frac{k}{n}\int_{k}^{n-1}\frac{1}{x}dx = \frac{k}{n}(\ln x)_k^{n-1} = \frac{k}{n}\ln\frac{n-1}{k} \\
\approx&\frac{k}{n}\ln\frac{n}{k}.
\end{aligned}
$$

Let $g(x) = x\ln\frac{1}{x} = -x\ln x$. Then, $g'(x) = -\ln x - 1$. $g'(x) = 0$ when $x = 1/e$.

$g(x)$ is maximized at $x = 1/e$

The maximum value of $P(A_k)$ is approximately $1/e$, which occurs when $k = n/e$.

Poisson Thinning

Each customer who enters a clothing store will purchase a suit with probability p. If the number of customers entering the store is a Poisson random variable with mean λ, what is the probability that the store will sell k suits?

Let X be the number of suits sold by the store. Let N be the number of customers entering the store. We are interested in $P(X = k)$. When conditioning on $N = n$, X is a binomial random variable with parameters n and p.

Clearly, when $n < k$, $P(X = k|N = n) = 0$. We only need to consider $n \geq k$.

$$
\begin{aligned}
P(X=k) &= \sum_{n=k}^{\infty} P(X=k|N=n)P(N=n) \\
&= \sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} e^{-\lambda} \frac{\lambda^n}{n!} \\
&= \sum_{n=k}^{\infty} \frac{n!}{k!(n-k)!} (\lambda p)^k (1-p)^{n-k} e^{-\lambda p} e^{-\lambda(1-p)} \frac{\lambda^{n-k}}{n!} \\
&= e^{-\lambda p} \frac{(\lambda p)^k}{k!} \sum_{n=k}^{\infty} e^{-\lambda(1-p)} \frac{(\lambda(1-p))^{n-k}}{(n-k)!} \\
&= e^{-\lambda p} \frac{(\lambda p)^k}{k!} \sum_{n=0}^{\infty} e^{-\lambda(1-p)} \frac{(\lambda(1-p))^{n}}{n!} \\
&= e^{-\lambda p} \frac{(\lambda p)^k}{k!}.
\end{aligned}
$$

In the last step, the summands are the probabilities of a Poisson random variable with mean $\lambda(1-p)$. Therefore, the sum is equal to 1.

The result is another Poisson distribution with mean λp.

Birthday Paradox and Poisson Approximation

Suppose there are $n = 23$ people in a room. What is the probability that some pair of them have the same birthday?

The probability can be expressed precisely for a general n with $2 \leq n \leq 365$. There is no need to consider $n > 365$ since the probability is equal to 1.

Let A_n be the event that no two people have the same birthday among n people. Then,

$$P(A_n) = \frac{364}{365}\frac{363}{365} \cdots \frac{365-(n-1)}{365}.$$

It can be computed that for $n = 23$, $P(A_{23}) = 0.492703$. Therefore, the probability that some pair of the people have the same birthday is equal to $1 - P(A_{23}) = 0.507297$.

The result may come as counter-intuitive, therefore, the term 'birthday paradox'.

One way to make the result less surprising is to consider the number of pairs of people who have a common birthday. Let the number be denoted by X.

The total number of people-pairs is $r(n) = \binom{n}{2}$ and let us index the pairs $1, \ldots, r(n)$.

Let X_i be the indicator random variable that is equal to 1 if and only if pair i have a common birthday, $i = 1, \ldots, r(n)$. Then, $X = \sum_{i=1}^{r(n)} X_i$ and $E[X] = \sum_{i=1}^{r(n)} P(X_i = 1)$.

The probability that a pair has a common birthday is equal to $p = 1/365$.

If we pretend the X_i's are IID, then X is a binomial random variable with parameters $r(n)$ and $p = 1/365$. (In fact, the X_i's are not independent of each other.)

$P(X > 0)$ is the probability that at least one pair have a common birthday.

For $n = 23$, $r(23) = 253$ and $E[X] = \frac{253}{365} \approx 0.69$. That is, the expected number of pairs with a common birthday is equal to 0.69.

We see that since the number of pairs is fairly large for the given p, it is not very surprising that there is good chance that two people have a common birthday.

Poisson Approximation: Recall that when $r(n)$ is large and p is small, the binomial random variable can be approximated by a Poisson random variable with mean $r(n)p$. Therefore,

$$P(X = k) \approx e^{-r(n)p} \frac{(r(n)p)^k}{k!}.$$

For $n = 23$, $P(X = 0) \approx e^{-253/365} = 0.499998$. Therefore, $P(X > 0) \approx 0.500002$.

List Access

- Consider a list with n items, $e_1, e_2, \ldots, e_n$. At each unit of time, a request is made for one of these items independent of the past, and the probability of requesting e_i is p_i. After an item e_i is requested, it will be moved to the front of the list.

  Problem: We are interested in the expected position of the element requested after this process has been in operation for a long time (i.e., in the steady state).

  (**Technical Note:** This is an aperiodic finite-state Markov chain, where each state is an ordering of the n items. There is a "steady state"; that is, the limit distribution exists, which is equal to the stationary distribution.)

- We will condition on which element is requested. We have

$$
\begin{aligned}
& E[\text{position of the element requested}] \\
&= \sum_{i=1}^{n} E[\text{position of the element requested} \mid e_i \text{ is requested}] p_i \\
&= \sum_{i=1}^{n} E[\text{position of } e_i \mid e_i \text{ is requested}] p_i \\
&= \sum_{i=1}^{n} E[\text{position of } e_i] p_i . \qquad (2)
\end{aligned}
$$

The removal of conditioning in the last step is because the position of e_i depends on the past requests whereas a new request for e_i is independent of the past.

The underlying probability corresponds to the steady-state (i.e., stationary) distribution of the Markov chain.

- To compute $E[\text{position of } e_i]$, we define indicator random variables I_j, where

$$I_j = \begin{cases} 1, & \text{if } e_j \text{ precedes } e_i \\ 0, & \text{otherwise} \end{cases}$$

- Then, position of $e_i = 1 + \sum_{j \neq i} I_j$, and

$$\begin{aligned} E[\text{position of } e_i] =& 1 + \sum_{j \neq i} E[I_j] \\ =& 1 + \sum_{j \neq i} P(e_j \text{ precedes } e_i) \end{aligned} \tag{3}$$

- Suppose the Markov chain has been in the steady-state for a long time. The Markov chain state is random with the steady-state distribution. Consider a particular realization of the state, which is an ordering of the n items.

  If you observe e_j precedes e_i, it must be that most recent request for either of them is for e_j. Conversely, if the most recent request for either of e_i or e_j is for e_j, then you will observe e_j precedes e_i.

- In short, e_j **precedes** e_i **if and only if the most recent request for either of**

them is for e_j. That is, the two events are the same.

- Since the sequence of requests forms an IID process, we don't have to look at the most recent request for either e_i or e_j. We only need to consider an arbitrary request that is for either e_i or e_j.
- Given that a request is for e_i or e_j, the probability it is for e_j is

$$P(\text{the request is for } e_j | \text{a request is for } e_i \text{ or } e_j) = \frac{p_j}{p_i + p_j}.$$

- Therefore, $P(e_j \text{ precedes } e_i) = \frac{p_j}{p_i+p_j}$.
- Then, by (2) and (3),

$$E[\text{position of the element requested}] = 1 + \sum_{i=1}^{n} p_i \sum_{j \neq i} \frac{p_j}{p_i + p_j}.$$

(**Technical Note:** A state of the Markov chain is a permutation of the n items. A transition occurs when a request is made for an item. Consider the steady state situation where the Markov chain makes transitions according to the stationary distribution. Split the discrete time into 'cycles' where a cycle starts when a request for either e_i or e_j is made, which causes either e_i or e_j to be moved the beginning of the list. The cycle lengths all have identical distribution; also they have the same distribution whether the cycle starts with a request for e_i or for e_j. The fractions of cycles starting with a request for e_i or with a request for e_j have a ratio of $p_i : p_j$. Over a long time, the amount of time that the Markov chain spends in the states with e_i preceding e_j and the amount of time it spends in the states with e_j preceding e_i have a ratio of $p_i : p_j$. Hence, for a state chosen from the stationary distribution, the probability that e_j precedes e_i is equal to $p_j/(p_i + p_j)$.)