1. Two-Population Sampling

We are conducting a public opinion poll to determine the fraction $p$ of people who will vote for Mr. Whatshisname as the next president. We ask $N_1$ college-educated and $N_2$ non-college-educated people, where $N_1$ and $N_2$ are positive integers. We assume that the votes in each of the two groups are i.i.d. Bernoulli($p_1$) and Bernoulli($p_2$), respectively in favor of Whatshisname. In the general population, the percentage of college-educated people is known to be $q$.

(a) What is a 95% confidence interval for $p$, using an upper bound for the variance?

(b) How do we choose $N_1$ and $N_2$ subject to $N_1 + N_2 = N$ to minimize the width of that interval? (You may ignore the constraint that $N_1$ and $N_2$ must be integers.)

Solution:

(a) If we let $\hat{p}_1$ and $\hat{p}_2$ denote the fraction of people who vote for Mr. Whatshisname in the two groups respectively, then an unbiased estimator for $p$ is

$$\hat{p} := q\hat{p}_1 + (1 - q)\hat{p}_2.$$

The variance of $\hat{p}$ is

$$\text{var } \hat{p} = \frac{q^2 p_1 (1 - p_1)}{N_1} + \frac{(1 - q)^2 p_2 (1 - p_2)}{N_2} \leq \frac{1}{4} \left( \frac{q^2}{N_1} + \frac{(1 - q)^2}{N_2} \right).$$

So, an approximate 95% confidence interval for $p$, using the CLT, is

$$\hat{p} \pm \sqrt{\frac{q^2}{N_1} + \frac{(1 - q)^2}{N_2}}.$$

(b) Minimizing the width of the interval is equivalent to minimizing the variance. We can explicitly enforce the constraint by writing $N_2 = N - N_1$, and then we have:

$$\frac{d}{dN_1} \left( \frac{q^2}{N_1} + \frac{(1 - q)^2}{N - N_1} \right) = -\frac{q^2}{N_1^2} + \frac{(1 - q)^2}{(N - N_1)^2}.$$

The second derivative is positive so the function is convex, and so the first-order condition tells us the minimizer. Setting the derivative to 0, we find that $q/N_1 = (1 - q)/(N - N_1)$. Therefore, the minimizer is $N_1 = qN$, $N_2 = (1 - q)N$.

2. Convergence of Exponentials

Let $X_1, X_2, \ldots$ be i.i.d. Exponential($\lambda$) random variables. Show that

$$\frac{X_n}{\ln n} \to 0 \quad \text{in probability as } n \to \infty.$$
Solution:

Fix \( \varepsilon > 0 \).

\[
\mathbb{P}\left( \frac{X_n}{\ln n} \geq \varepsilon \right) = \mathbb{P}(X_n \geq \varepsilon \ln n) = \exp(-\lambda \varepsilon \ln n) = n^{-\lambda \varepsilon} \rightarrow 0
\]
as \( n \rightarrow \infty \).

3. Twitch Plays Pokemon

After attending an EECS 126 lecture, you went back home and started playing Twitch Plays Pokemon. Suddenly, you realized that you may be able to analyze Twitch Plays Pokemon.

---

### Figure 1: Part (a)

(a) The player in the top left corner performs a random walk on the 8 checkered squares and the square containing the stairs. At every step the player is equally likely to move to any of the squares in the four cardinal directions (North, West, East, South) if there is a square in that direction. Find the expected number of moves until the player reaches the stairs in Figure 1.

---

### Figure 2: Part (b)

(b) The player randomly walks in the same way as in the previous part. Find the probability that the player reaches the stairs in the bottom right corner in Figure 2.

---

### Solution:

(a) Using symmetry, the 9 states can be grouped as follows.

```
    a  b  c
   b  d  e
   c  e  f
```

Now, observe that state \( d \) is equivalent to state \( c \).

```
    a  b  c
   b  c  e
   c  e  f
```
With the above states, one can write down the following first-step equations.

\[
\begin{align*}
T_a &= 1 + T_b \\
T_b &= 1 + \frac{1}{3} T_a + \frac{2}{3} T_c \\
T_c &= 1 + \frac{1}{2} T_b + \frac{1}{2} T_e \\
T_e &= 1 + \frac{2}{3} T_c + \frac{1}{3} T_f \\
T_f &= 0
\end{align*}
\]

Solving the above equations gives:

\[T_a = 18, T_b = 17, T_c = 15, T_e = 11\]

Thus, the player has to make 18 moves to go downstairs on average.

(b) Consider 9 initial states and corresponding probabilities of reaching the “good” stairs as follows. Using symmetry, one can obtain the following table.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1/2</td>
<td>1 - p</td>
</tr>
<tr>
<td>q</td>
<td>1/2</td>
<td>1 - q</td>
</tr>
<tr>
<td>0</td>
<td>1/2</td>
<td>1</td>
</tr>
</tbody>
</table>

With the above probabilities, one can write down the following first-step equations.

\[
\begin{align*}
p &= \frac{1}{2} q + \frac{1}{2} \cdot \frac{1}{2} \\
q &= \frac{1}{3} p + \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot 0
\end{align*}
\]

Solving the above equations gives:

\[p = 0.4, q = 0.3\]

Thus, we are going to reach the good stairs with probability 0.4.

4. Compression of a Markov Chain

Consider an irreducible Markov chain \((X_n)_{n \in \mathbb{N}}\) as shown below.

![Markov Chain Diagram]
Roughly how many bits are needed to represent \((X_0, X_1, \ldots, X_n)\)?

**Solution:**

Answer: \(nH(p)\) bits.

To keep track of the sequence, we need to keep track of the initial state \(X_0\), as well as \(Y_1, \ldots, Y_n\) where the \(Y_i\) represents the switches of the Markov chain (\(Y_i = 1\) if \(X_i\) and \(X_{i-1}\) are different, otherwise \(Y_i = 0\)). The \((Y_n)_{n \in \mathbb{N}}\) are i.i.d. Bernoulli\((p)\). You can take the sequence \((Y_n)_{n \in \mathbb{N}}\) and compress it down to \(nH(p)\) bits, where \(H(p) = -p \log_2 p - (1 - p) \log_2 (1 - p)\).

5. **Frogs**

Three frogs are playing near a pond. When they are in the sun they get too hot and jump in the lake at rate 1. When they are in the lake they get too cold and jump onto the land at rate 2. The rates here refer to the rate in exponential distribution. Let \(X_t\) be the number of frogs in the sun at time \(t \geq 0\).

(a) Find the stationary distribution for \((X_t)_{t \geq 0}\).

(b) Check the answer to (a) by noting that the three frogs are independent two-state Markov chains.

**Solution:**

(a) Let the states \(S = \{0, 1, 2, 3\}\) be the number of frogs in the sun. The Markov chain has \(\lambda_0 = 6, \lambda_1 = 4, \lambda_2 = 2, \mu_3 = 3, \mu_2 = 2, \) and \(\mu_1 = 1\). Here \(\lambda_i\) and \(\mu_i\) are the rates of jumping forward and backward from state \(i \in S\), respectively. Using detailed balance, we compute the stationary distribution to be 

\[
\pi = \frac{1}{27} \begin{bmatrix} 1 & 6 & 12 & 8 \end{bmatrix}.
\]

(b) The individual frogs follow independent Markov chains, each with stationary distribution

\[
\pi = \frac{1}{3} \begin{bmatrix} 2 & 1 \end{bmatrix}.
\]

The probability of being in state \(i \in S\) is therefore

\[
P(X_t = i) = \left(3 \choose i \right) \left(\frac{1}{3}\right)^{3-i} \left(\frac{2}{3}\right)^i, \quad i \in S.
\]

6. **Spatial Poisson Process**

A two-dimensional Poisson process of rate \(\lambda > 0\) is a process of randomly occurring special points in the plane such that (i) for any region of area \(A\) the number of special points in that region has a Poisson distribution with mean \(\lambda A\), and (ii) the number of special points in non-overlapping regions is independent. For such a process consider an arbitrary location in the plane and let \(X\) denote its distance from its nearest special point (where distance between two points \((x_1, y_1)\) and \((x_2, y_2)\) is defined as \(\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}\)). Show that:
(a) \( P(X > t) = \exp(-\lambda \pi t^2) \) for \( t > 0 \).

(b) \( \mathbb{E}[X] = \frac{1}{2\sqrt{\lambda}} \).

Solution:

(a) Given an arbitrary location, \( X > t \) if and only if there are no special points in the circle of radius \( t \) around the given point. The expected number in that circle is \( \lambda \pi t^2 \), and since the number in that circle is Poisson with expected value \( \lambda \pi t^2 \), the probability that number is 0 is \( \exp(-\lambda \pi t^2) \). (Recall that the Poisson distribution is parameterized by its mean.) Thus \( P(X > t) = \exp(-\lambda \pi t^2) \).

(b) Since \( X \geq 0 \), we have

\[
\mathbb{E}[X] = \int_0^\infty P(X > t) \, dt = \int_0^\infty e^{-\lambda \pi t^2} \, dt.
\]

We can look this up in a table of integrals, or recognize its resemblance to the Gaussian PDF. If we define \( \sigma^2 = 1/(2\pi\lambda) \), the above integral is

\[
\mathbb{E}[X] = \sigma \sqrt{2\pi} \int_0^\infty \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{t^2}{2\sigma^2}\right) \, dt = \frac{\sigma \sqrt{2\pi}}{2} = \frac{1}{2\sqrt{\lambda}}.
\]