Outline:
1. Geometric Distribution
2. Binomial Distribution
3. Birthday Problem
4. Administrivia

1 Geometric Distribution

Consider the probability distribution of

\[ n = 1, 2, 3, \ldots \]
\[ P(n) = p, p(1 - p), p(1 - p)^2, \ldots \]

In general, \( P(n) = p(1 - p)^{n-1} \) where \( n \geq 1 \).

Toss a coin of head probability \( p \). Then \( P(n) \) is the probability that the first head occurs in the \( n \)-th trial.

Sanity check: does the probability mass sum up to 1?

\[
\sum_{n=1}^{\infty} P(n) = p \cdot \sum_{n=1}^{\infty} (1 - p)^{n-1}
\]
\[
= p \cdot \frac{1}{1 - (1 - p)} \text{ by the summation of geometric series}
\]
\[
= 1
\]

What is the mean of the distribution? Recall the definition of “Expected Value” (or “Expectation”) in class. Let \( X \sim P(n) \),

\[
E[X] = \sum_{n=1}^{\infty} n \cdot \Pr[X = n]
\]
\[
= 1 \cdot p + 2 \cdot p(1 - p) + 3 \cdot p(1 - p)^2 + \ldots
\]
\[
\equiv I
\]

Each entry in the summation \( I \), i.e., \( np(1 - p)^{n-1} \), is a product of an arithmetic series \( (n) \) and a geometric series \( ((1 - p)^{n-1}) \). A trick for doing such summation is:

\[
I = 1p + 2p(1 - p) + 3p(1 - p)^2 + \ldots
\]
\[
(1 - p)I = 0 + 1p(1 - p) + 2p(1 - p)^2 + \ldots
\]
Subtract line 2 from line 1, it follows that \[1 - (1 - p)]I = p + p(1-p) + p(1-p)^2 + \ldots = p \frac{1}{1-(1-p)}\] (recall the summation of geometric series). Therefore, \(I = \frac{1}{p}\).

**Exercise:** Another way of summing \(I\), by way of calculus, is to express \(n(1 - p)^{n-1}\) as the derivative of \(-(1 - p)^n\). Then use the fact that differentiation and summation are inter-changeable. Verify that this method leads to the same answer.

**Exercise:** Verify that \(I\), as a summation of infinite series, converges. (Consult a calculus book, if necessary)

## 2 Binomial Distribution

Toss a coin (w/ head probability \(p\)) for \(n\) times. Let \(X\), a random variable (recall the definition?), be the number of heads that occurred. Then,

\[
\Pr[X = k] = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \ldots, n
\]

Why is it? Consider a sequence of \(n\) empty slots in which we can fill in an “H” or a “T”.

Given the number of “H”s being \(k\), there are \(\binom{n}{k}\) ways to fill in the sequence. Each way happens with probability \(p^k \cdot (1 - p)^{n-k}\) (head probability times tail probability).

Again, verify \(\sum_{k=0}^{n} \Pr[X = k] = 1\). Recall binomial coefficients.

Consider the mean of the distribution. Intuitively, what should it be?

By definition,

\[
EX = \sum_{k=0}^{n} k \Pr[X = k] = \sum_{k=0}^{n} k \binom{n}{k} p^k (1 - p)^{n-k}
\]

Aside: \(\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n \cdot (n-1)!}{(k-1)!(n-k)!} = \binom{n-1}{k-1}\)

\[
= \sum_{k=1}^{n} \binom{n-1}{k-1} (p^k (1 - p)^{n-k})
\]

\[
= np \sum_{k=1}^{n-1} \binom{n-1}{k-1} (1 - p)^{n-k}
\]

\[
= np \text{ by summation of binomial coefficients}
\]

Another way of computing the mean, by the simple but useful linearity of expectation: Write \(X = \sum_{i=1}^{n} X_i\), where \(X_i = 1\) if the \(i\)-th trial shows up head and \(X_i = 0\) otherwise. Then,

\[
EX = E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i] = np
\]

This method is simpler and more along the lines of the intuition.
3 Birthday Problem

Consider $Q(n, d = 365)$, the probability that no two people out of a group of $n$ have the same birthday.

The first person has a choice of all dates $\Pr = \frac{d}{d}$

The second person has a choice of all dates $\Pr = \frac{d - 1}{d}$

...The $n$-th person has a choice of all dates $\Pr = \frac{d - n + 1}{d}$

Therefore,

$$Q(n, d) = \frac{(d - 1)\ldots(d - n + 1)}{d^{n-1}} = \frac{d!}{d^{n-1}(d - n)!}$$

**Exercise:** Enumerate the total number of possible birthday combinations for a group of $n$ people, and the number of ways that their birthdays don’t collide. Verify that the division of the two numbers gives the same answer as above.

**Exercise:** Rephrase the birthday problem in terms of Balls and Bins. Relate this problem to problems that we covered in class.

Question: How many people does it take to have a 50%+ chance of identical birthdays? That is, we want the smallest $n$ such that $Q(n, d) < .5$. It turns out that $n = 23$. For a class room of size 50, the probability of identical birthdays is $1 - Q(n = 50, d = 365) \approx 0.9674$.

4 Administrivia

Send email to nhz@cs.berkeley.edu with

SUBJECT: cs174 signup
BODY: <full name>
   <email addr>
   <webpage, if any>

Separate the fields with new lines.