Lecture 23: some discrete distributions.

Everything that we've seen so far applies to distributions where the outcome space is not finite but countable, i.e. of the type \( \{1, 2, 3, \ldots\} \).

Example: consider repeated and independent trials of an experiment (or game) where, at each trial, we have probability of success = \( p \); also: \( q = 1 - p \).

(say we're rolling a die, and "success" is "rolling a 6", so \( p = \frac{1}{6} \)). Let:

\[ T = \# \text{ of trials needed for the 1st success} \]

\[
P(T = 1) = p \\
P(T = 2) = qp \\
P(T = 3) = q^2p \\
\vdots \\
P(T = k) = q^{k-1}p
\]

We have:

\[
P(\text{failure in first } k \text{ trials}) = q^k \\
P(\text{always failing}) = \lim_{k \to \infty} q^k = 0 \quad (\text{if } 1 < q < 1)
\]

Equivalently,

\[
P(\text{succeeding at some point}) = \sum_{k=1}^{\infty} P(T = k) \\
= \sum_{k=1}^{\infty} q^{k-1}p = p \sum_{j=0}^{\infty} q^j = p \frac{1}{1-q} = \frac{1}{p} = 1.
\]
Now let's compute:

\[ P(T \text{ is even}) = P(T = 2) + P(T = 4) + P(T = 6) + \ldots \]

\[ = \sum_{k=1}^{\infty} P(T = 2k) = \sum_{k=1}^{\infty} q^{2k-1} \frac{1}{p} = \sum_{k=1}^{\infty} q^{2k-2}q p \]

\[ = q p \sum_{k=1}^{\infty} (q^2)^{k-1} = q p \sum_{j=0}^{\infty} (q^2)^{j} = q p \frac{1}{1-q^2} = \]

\[ = q(1-q) \frac{1}{(1-q)(1+q)} = \frac{q}{1+q} = \frac{1-p}{2-p} \]

**Mean and variance of the geometric distribution.**

To compute \( E(T) \), in Lecture #18 we used a "derivative trick."

Now we use a different technique.

\[ E(T) = \sum_{k=1}^{\infty} k q^{k-1} \frac{1}{p} = \frac{1}{p} \sum_{k=1}^{\infty} k q^{k-1} \]

\[ \frac{d}{dq} [k q^{k-1}] = k q^{k-1} \]

"\( S_1 \)" let's compute it!

\( S_1 = 1 + 2q + 3q^2 + 4q^3 + \ldots \)

\[ q \quad S_1 = q + 2q^2 + 3q^3 + 4q^4 + \ldots \]

\( (1-q) \quad S_1 = 1 + q + q^2 + q^3 + \ldots = \sum_{k=0}^{\infty} q^k = \frac{1}{1-q} \]

\[ \Rightarrow S_1 = \frac{1}{(1-q)^2} \]

\[ \Rightarrow E(T) = \frac{1}{p(1-q)^2} = \frac{1}{p} \frac{1}{p^2} = \frac{1}{p} \]

**Interpretation:**

- average # of successes per trial = \( p \) (proportion of successes)
- average # of trials per success = \( \frac{1}{p} \)

If \( p \) small, \( E(T) \) is large!
Variance: \( \text{Var}(T) = E(T^2) - (E(T))^2 \), where \( E(T) = \frac{1}{p} \) and

\[
E(T^2) = \sum_{k=1}^{\infty} k^2 q^{k-1} p = p \sum_{k=1}^{\infty} k^2 q^{k-1}.
\]

Let's compute \( S_2 = \sum_{k=1}^{\infty} k^2 q^{k-1} \).

1st method: \( S_2 = 1 + 4q + 9q^2 + 16q^3 + \ldots + (k-1)^2 q^{k-2} + k^2 q^{k-1} + \ldots \)

\[
q S_2 = q + 4q^2 + 9q^3 + \ldots + (k-1)^2 q^{k-1} + \ldots
\]

\[
\Rightarrow (1-q) S_2 = 1 + 3q + 5q^2 + 7q^3 + \ldots + \left( \frac{k^2 - (k-1)^2}{q} \right) q^{k-1}
\]

So:

\[
(1-q) S_2 = \sum_{k=1}^{\infty} \left( 2k-1 \right) q^{k-1} = 2 \sum_{k=1}^{\infty} k q^{k-1} - \sum_{k=1}^{\infty} q^{k-1} = 2 S_1 - \frac{1}{1-q} = \frac{2}{(1-q)^2} - \frac{1}{1-q} = \frac{1+q}{(1-q)^2}.
\]

2nd method (by Jared - thanks!): by comparing

\[
q S_1 = \sum_{k=1}^{\infty} k q^{k-1} \quad \text{and} \quad S_2 = \sum_{k=1}^{\infty} k^2 q^{k-1},
\]

we note that: \( S_2 = \frac{d}{dq} \left( q S_1 \right) \), therefore:

\[
S_2 = \frac{d}{dq} \frac{q}{(1-q)^2} = \frac{(1-q)^2 + q 2(1-q)}{(1-q)^4} = \frac{1-q+2q}{(1-q)^3} = \frac{1+q}{(1-q)^3}.
\]

In conclusion: \( \text{Var}(T) = p S_2 - \left( \frac{1}{p} \right)^2 = p \frac{1+q}{p^3} - \frac{1}{p^2} = \frac{q}{p^2} \).

In summary: for a Geometric random variable with parameter \( p \),

\[
E(T) = \frac{1}{p}, \quad \text{Var}(T) = \frac{q}{p^2}, \quad \text{SD}(T) = \frac{\sqrt{q}}{p}
\]

Note: \( \text{Var}(T) = \frac{1-p}{p^2} \), so the variance (that measures how much the probability distribution is "spread out" around the mean) gets larger as \( p \) (prob. of success) gets smaller.
The Wait Until the $r^{th}$ Success.

So, we've seen that if

$T =$ # of trials until the 1st success

then the random variable $T$ has a geometric distribution:

$$P(T = k) = q^{k-1} p, \quad k = 1, 2, 3, \ldots$$

Fix a number $r \in \{1, 2, 3, \ldots \}$. Define:

$T_r =$ # of trials until the $r^{th}$ success.

(we have seen this in earlier homework). We have

$$P(T_r = k) = \begin{cases} \binom{k-1}{r-1} p^{r-1} q^{k-1-(r-1)} & \text{for } k = r, r+1, r+2, \ldots \\ \end{cases}$$

Problem: $E(T_r) =$ ? , $SD(T_r) =$ ?

Trick: $T_r = W_1 + W_2 + W_3 + \ldots + W_r$ (*)

where: $W_1 =$ waiting time for 1st success

$W_2 =$ " " for 2nd success after 1st success

$W_3 =$ " " " 3rd " " 2nd "
Now, the random variables $W_1, W_2, W_3, \ldots$ are:

- independent
- identically distributed: in fact they are all geometric
  \[ P(W_i = k) = q^{k-1} p^k, \]

So:
- \[ E(W_i) = \frac{1}{p}, \quad SD(W_i) = \sqrt{\frac{q}{p}}, \quad \text{for } i = 1, 2, \ldots, r \]

- By (*):
  \[ E(T_r) = E(W_1) + E(W_2) + \ldots + E(W_r) = \frac{r}{p}, \]
  \[ Var(T_r) = Var(W_1) + \ldots + Var(W_r) = r \frac{q}{p^2}, \]
  \[ SD(T_r) = \frac{\sqrt{rq}}{p}, \]

- By the central limit theorem, for $r$ large
  \[ T_r \sim \text{Gaussian } (\mu, \sigma^2), \quad \text{with } \mu = \frac{r}{p}, \quad \sigma^2 = \frac{rq}{p} \]