Chapter 4.11 Negative Binomial Distribution

Jim Albert and Monika Hu

Chapter 4 Discrete Distributions

A Baseball Story

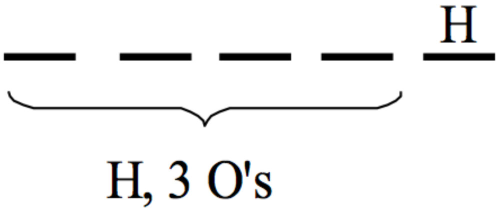
- In the 2004 baseball season, Ichiro Suzuki had the opportunity to break the season record for the most hits. Sisler's record was 257 hits and Suzuki had 255 hits before game on September 30. Was it likely that Suzuki would tie Sisler's record during this particular game?
- This can be considered a coin-tossing experiment. Assume the probability that he gets a hit on a single at-bat is $p = 0.372$.
- How many at-bats will it take him to get two hits?

Not a Binomial Experiment

- This is not a Binomial experiment since the number of trials is not fixed.
- Instead the number of successes (hits) is fixed in advance and the number of trials to achieve this is random.
- Consider $Y =$ number of at-bats to get two hits.
- One is interested in probabilities about the number of bats Y.

Probability Calculation 1/2

- It should be obvious that Y has be at least 2
- Consider the probability that $Y = 5$.
- Know the second hit must have occurred in the fifth trial (since Y=5). Also it is known that there must have been one hit and three outs in the first four trials – there are $\binom{4}{1}$ ways of arranging the H's and the O's in these trials.



Probability Calculation 1/2

- Also the probability of each possible outcome is $p^2(1-p)^3$, where p is the probability of a hit.
- So the probability that it takes 5 trials to observe 2 hits is

$$P(Y=5) = \binom{4}{1} p^2(1-p)^3.$$

- Since $p = 0.372$ in this case, we get

$$P(Y=5) = \binom{4}{1} 0.372^2(1-0.372)^3 = 0.1371.$$

Negative Binomial Experiment

- One has a sequence of independent trials where each trial can be a success (S) or a failure.
- The probability of a success on a single trial is p.
- The experiment is continued until one observes r successes, and $Y =$ number of trials one observes.

Probability that it takes y trials to observe r successes is

$$P(Y = y) = \binom{y-1}{r-1} p^r (1-p)^{y-r}, y = r, r+1, r+2, \ldots$$

Back to Baseball Example

- Consider our baseball example where $r = 2$ and $p = 0.372$. Table gives the probabilities for the number of at-bats $y = 2, 3, \ldots, 9$.

y	$P(Y = y)$
2	.1384
3	.1738
4	.1637
5	.1371
6	.1076
7	.0811
8	.0594
9	.0426

Comments

- Note that it is most likely that Suzuki will only need three at-bats to get his two additional hits
- But the probability of three at-bats is only 17%.
- Actually each of the values 2, 3, 4, 5, and 6 have probabilities exceeding 10%.
- The probability that Y is at most 9 is 0.904, so the probability that Y exceeds 9 is 1 - 0.904 = 0.096.

Mean of a Negative Binomial

- For a Negative Binomial experiment where Y is the number of trials needed to observe r successes, one can show that the mean value is

$$E(Y) = \frac{r}{p}.$$

- For the baseball example, $r = 2$ and $p = 0.372$, so the mean would be $E(Y) = 2/0.372 = 5.4$. So Suzuki would average over 5 at-bats to get 2 hits in many repetitions of this random experiment.

Negative Binomial Calculations on R

- The R function `dnbinom()` can be used to compute probabilities from Negative Binomial distributions.
- This function defines the random variable X to be the number of failures (instead of the total number of trials Y) until the r-th success.
- Have to translate question about number of trials Y to the number of failures X to use the R function.

Example

- Consider our baseball example where Y is the number of at-bats for Suzuki to get $r = 2$ hits where the probability of a hit on a single at-bat is $p = 0.372$.
- The probability $P(Y = 5)$ is the same as the probability $P(X = 3)$ where X is the number of failures until the 2nd success. One finds

```r
dnbinom(3, size = 2, prob = .372)
```

```
## [1] 0.137096
```

which is equivalent to the probability that $Y = 5$ computed earlier.