CHRISTOPHER PRENER, PH.D.
FALL, 2016

WEEK 06
LECTURE 08–EXTRA

QUANTITATIVE ANALYSIS

RANDOM VARIABLES

A NEGATIVE BINOMIAL DISTRIBUTION

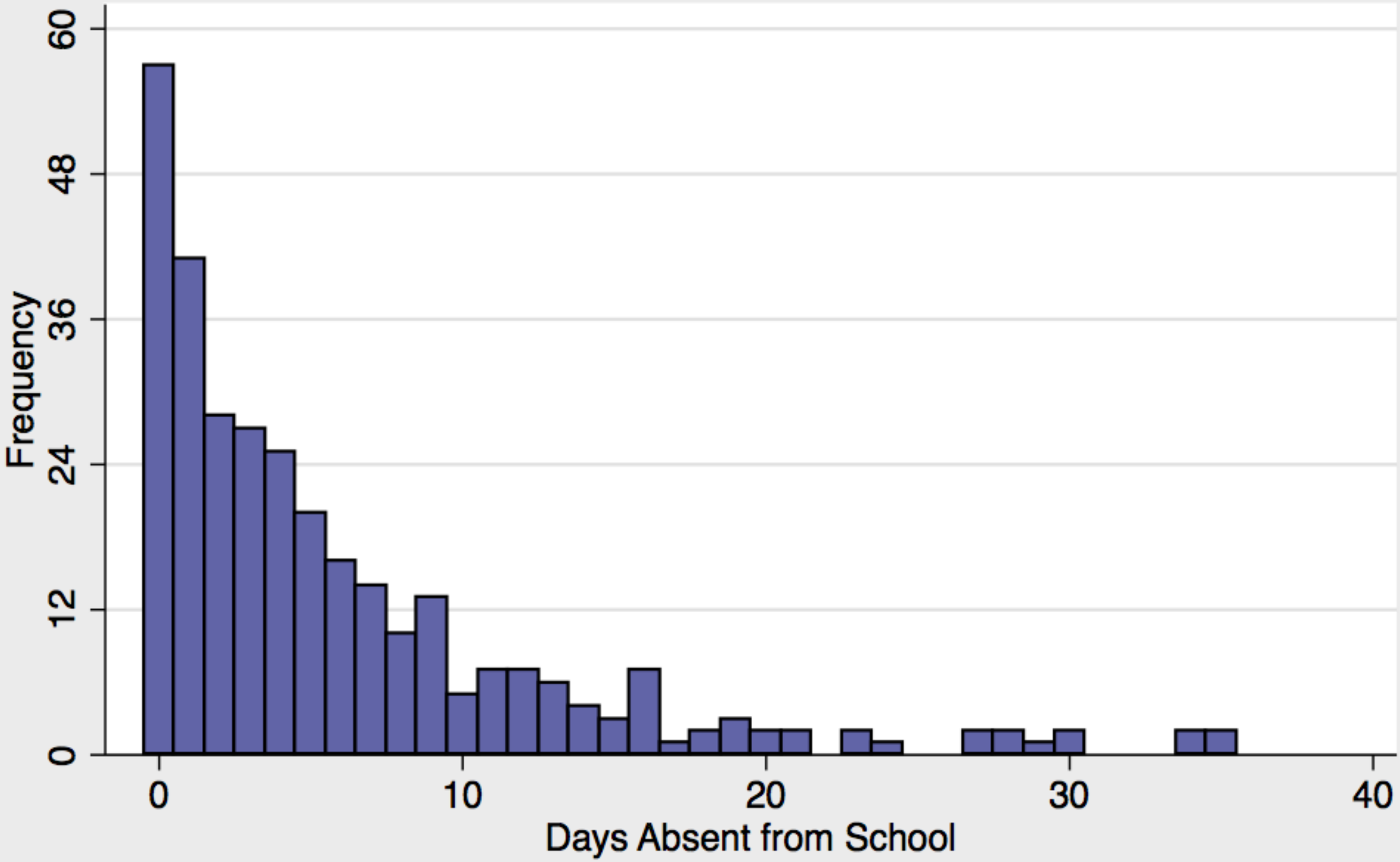
PROBLEM

- With the Poisson distribution, μ and σ^2 should be approximately equal.
- However, with many types of count data, σ^2 can be larger than μ. This is a condition known as *overdispersion*.

```
. use http://www.ats.ucla.edu/stat/stata/dae/nb_data, clear

. summarize daysabs

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
     daysabs |        314    5.955414    7.036958           0         35


. display 7.036958^2
49.518778
```

Example of Overdispersed Count Data

Days Absent from School



Data via UCLA ATS

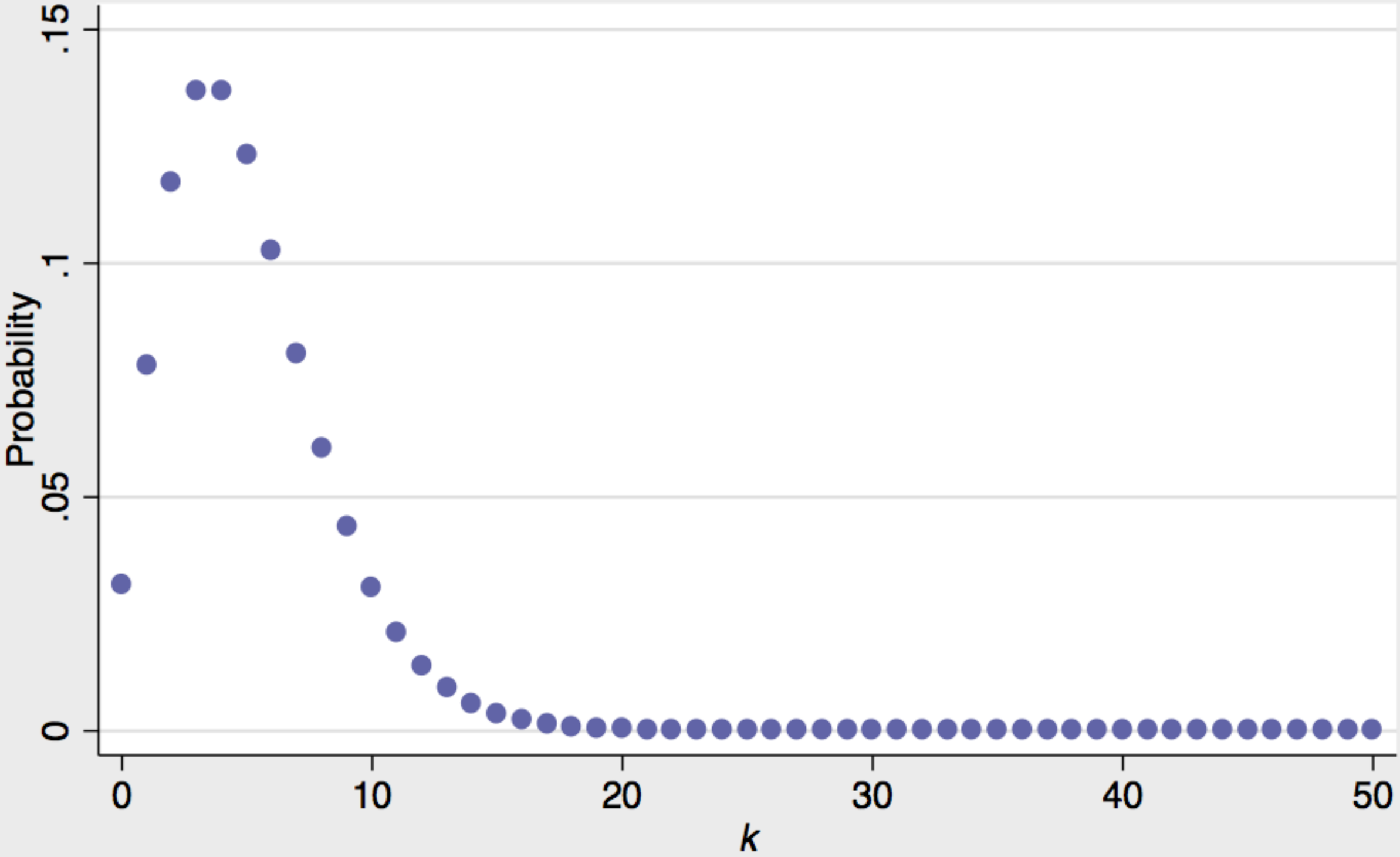
DEFINITION

- A sequence of independent trials with constant probability of success at each trial (p) where we are interested in the number of failures (k) required to produce a set number of successes (n).
- r is often used instead of n.
- Like the Poisson distribution, it is characterized by count data.
- Unlike the Poisson distribution, it is able to accommodate variances that differ from the mean.

$$P(X = x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r$$

Negative Binomial Distribution Probability Mass Function

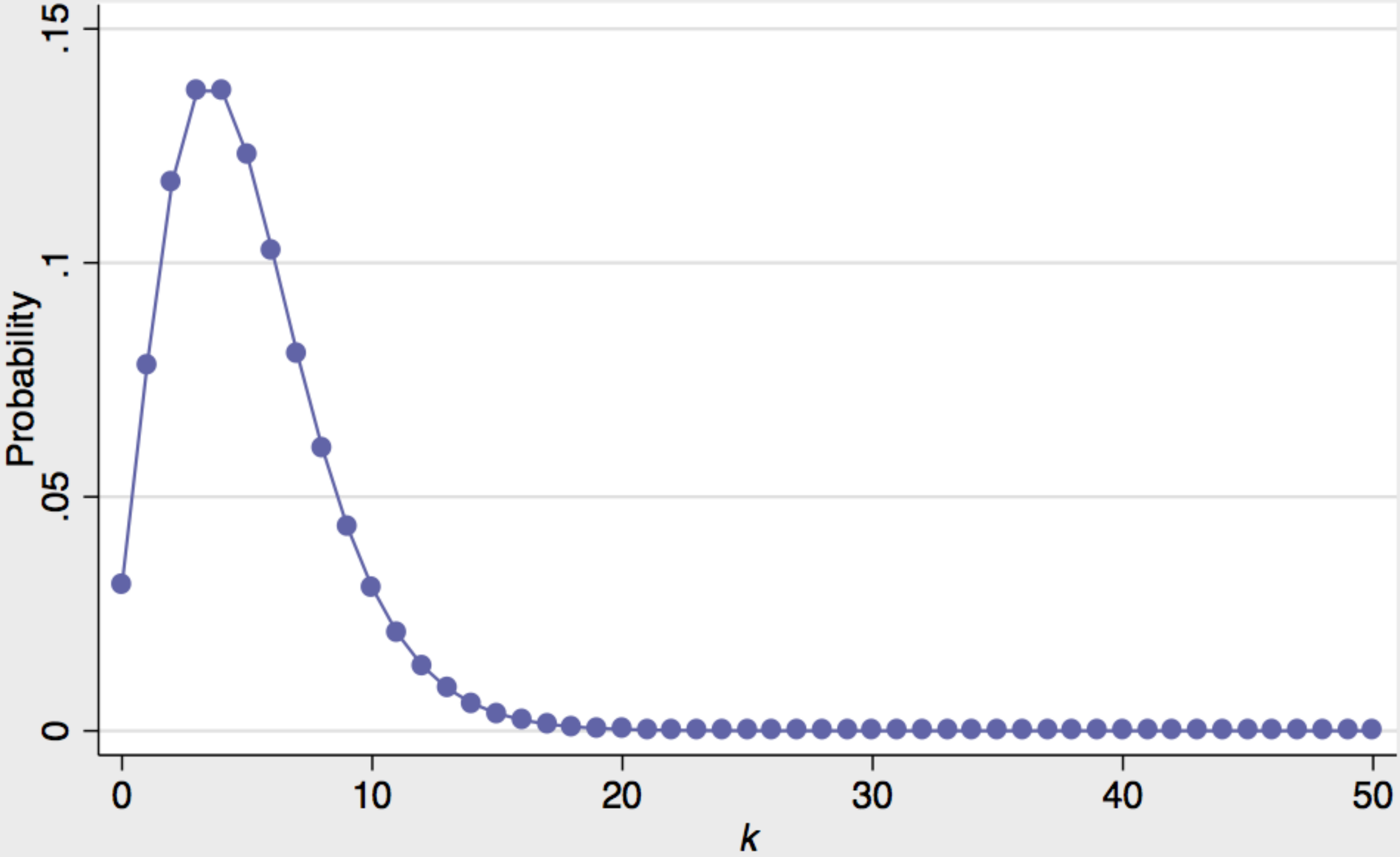
p=0.5 and n=5



Line between calculated probabilities included for visualization purposes only.

Negative Binomial Distribution Probability Mass Function

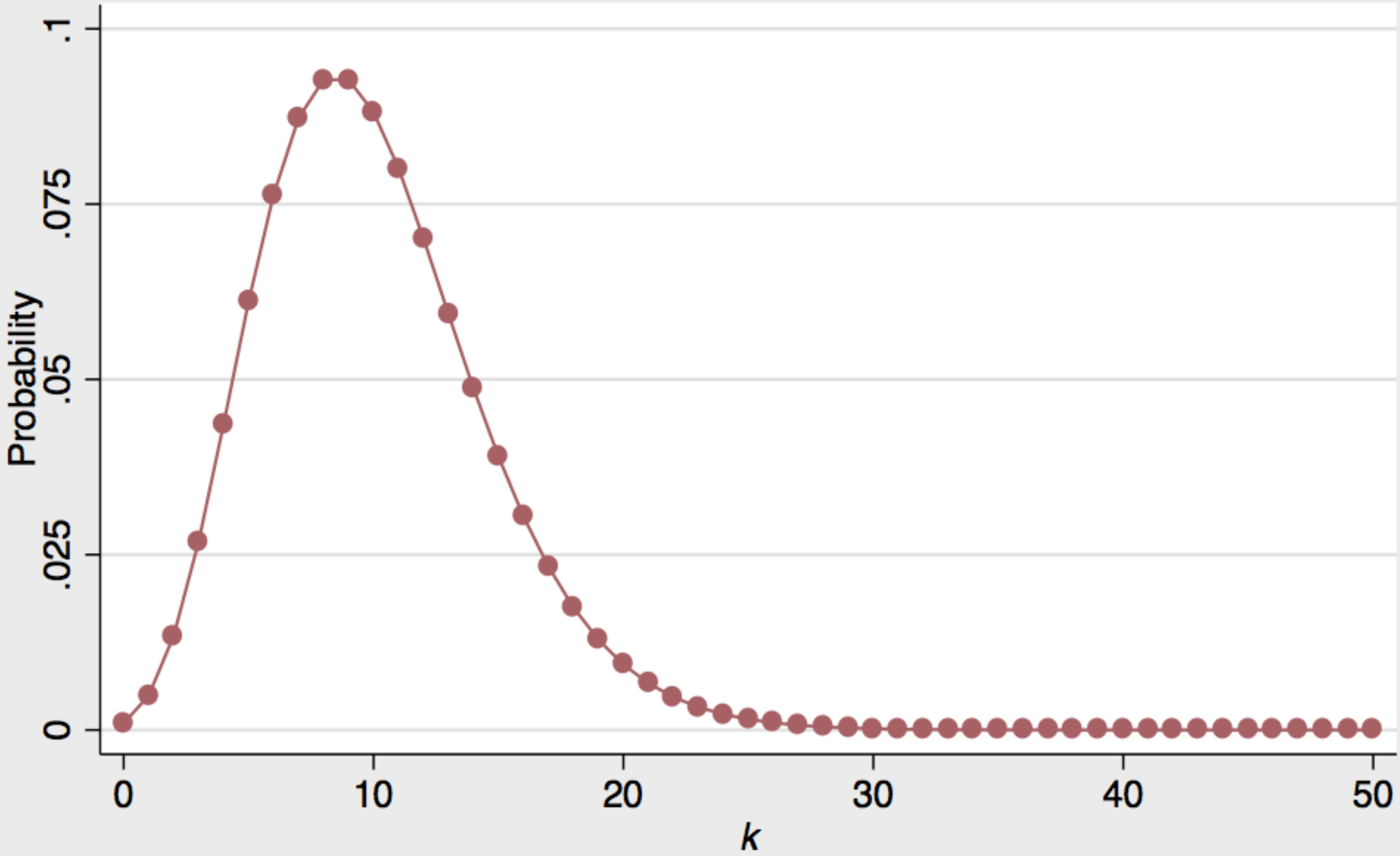
p=0.5 and n=5



Line between calculated probabilities included for visualization purposes only.

Negative Binomial Distribution Probability Mass Function

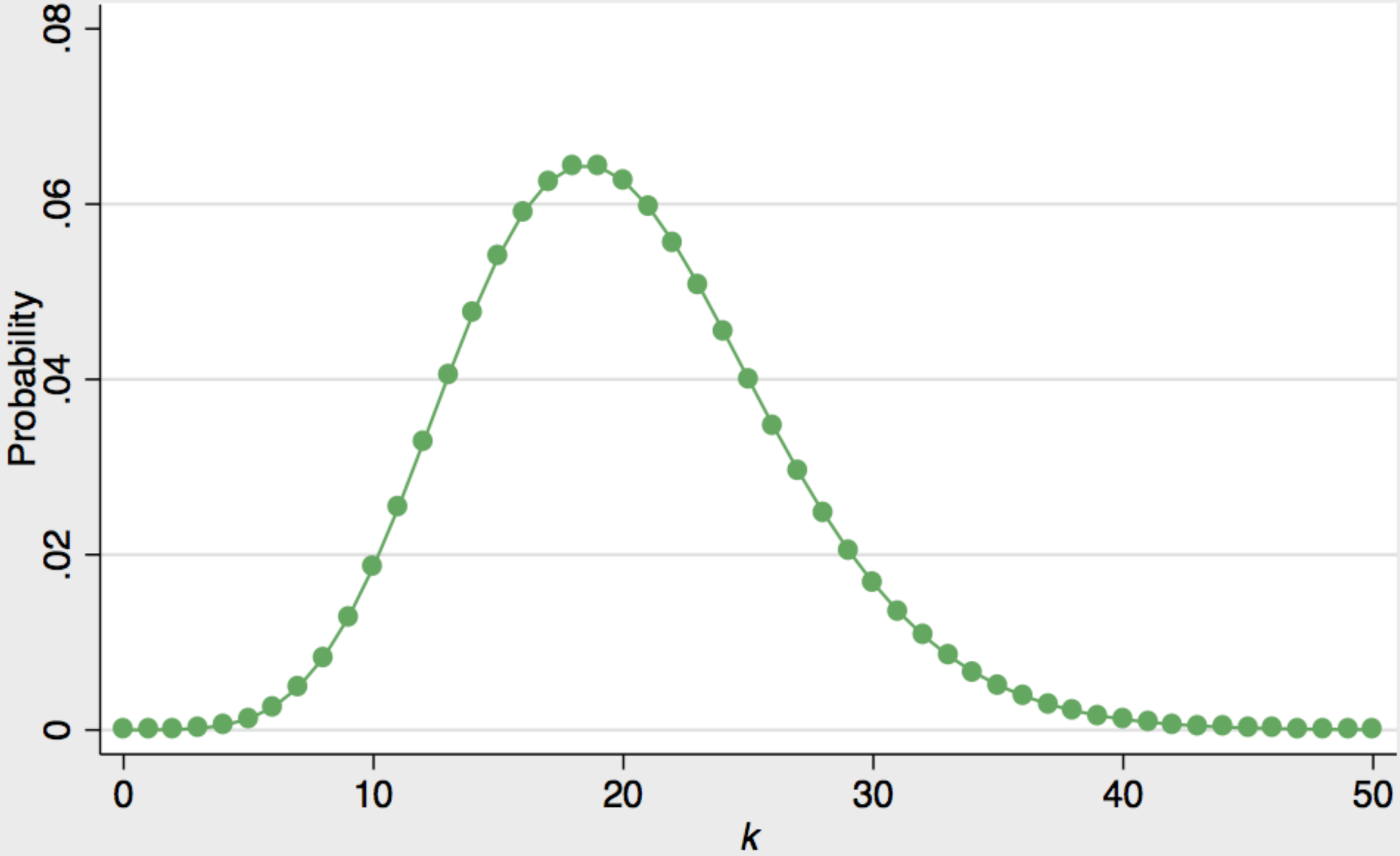
p=0.5 and n=10



Line between calculated probabilities included for visualization purposes only.

Negative Binomial Distribution Probability Mass Function

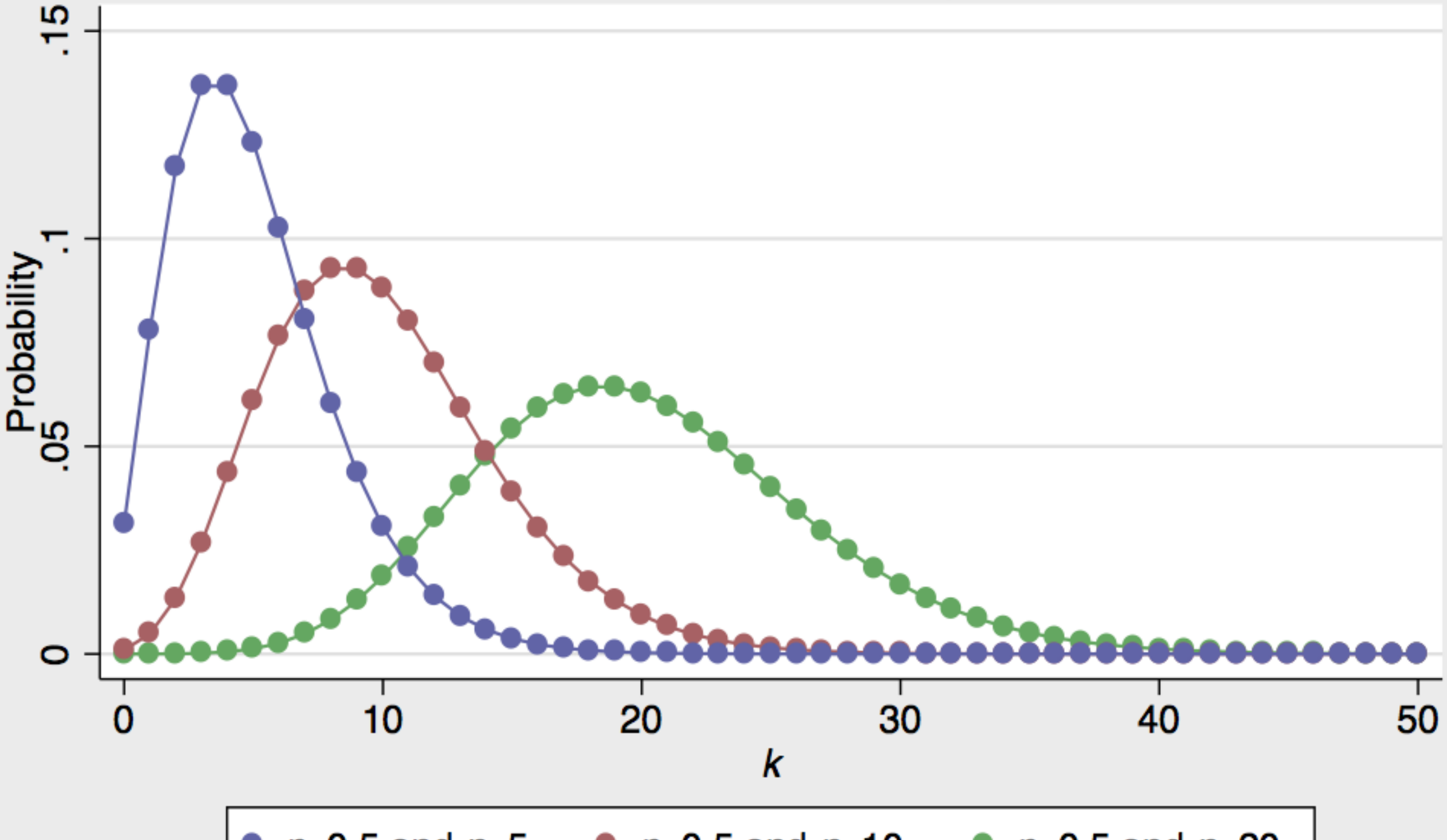
p=0.5 and n=20



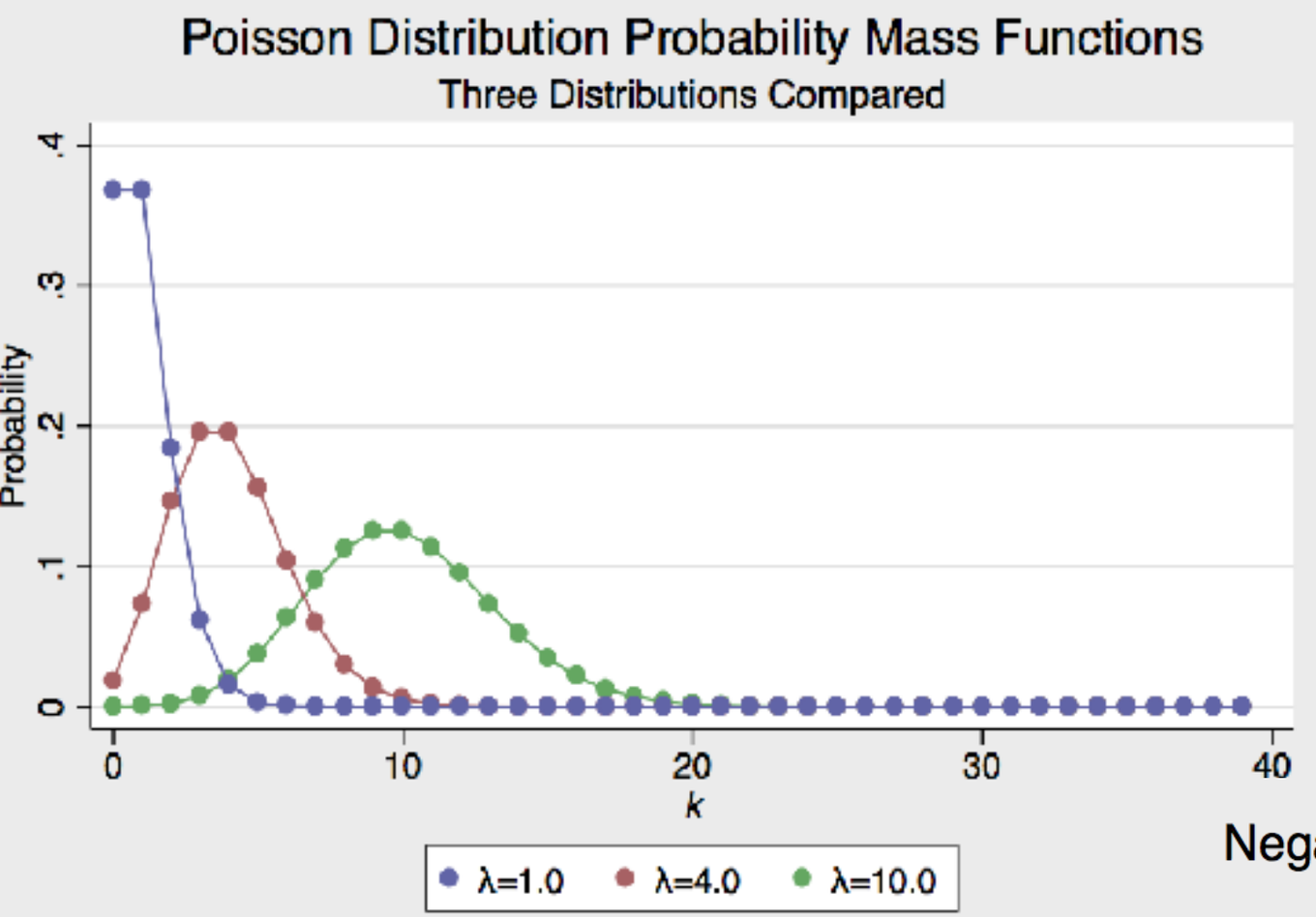
Line between calculated probabilities included for visualization purposes only.

Negative Binomial Distribution Probability Mass Functions

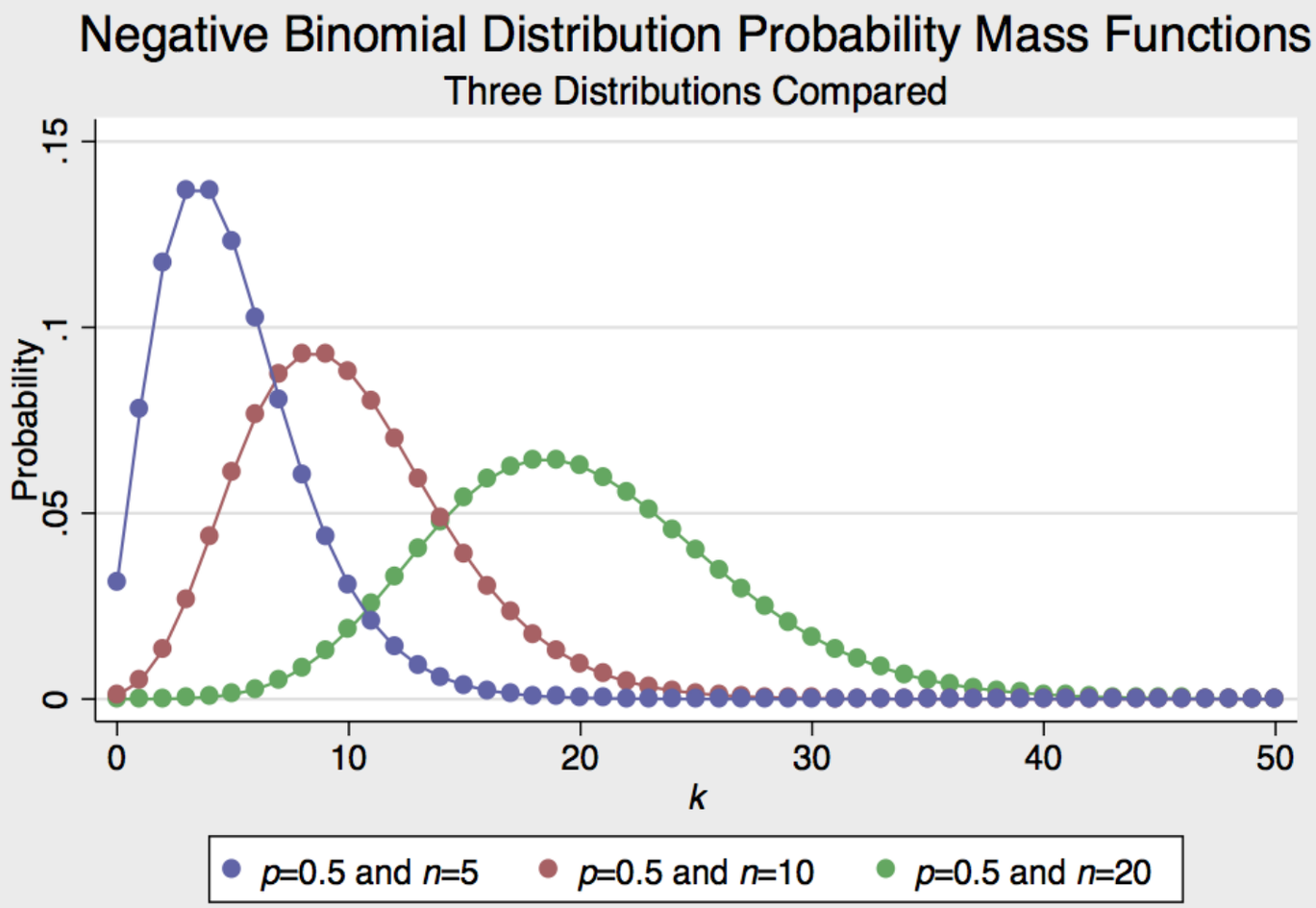
Three Distributions Compared



Line between calculated probabilities included for visualization purposes only.



Lines between calculated probabilities included for visualization purposes only.



Line between calculated probabilities included for visualization purposes only.

STATA FUNCTIONS

`nbinomialp(n,k,p)`	returns probability of observing ***k*** failures before the n^{th} success	P(x = k)
`nbinomial(n,k,p)`	returns probability of observing ***k* or fewer** failures before the n^{th} success	P(x ≤ k)
`nbinomialtail(n,k,p)`	returns probability of observing ***k* or more** failures before the n^{th} success	P(x ≥ k)

NEGATIVE BINOMIAL WORKFLOW

- A social service agency determines that 20% of clients approached on a given day will enroll in an intervention program. What is the probability that the first enrollment comes on the fifth client after four clients refuse (fail) to enroll?
- Is the negative binomial distribution appropriate?

NEGATIVE BINOMIAL WORKFLOW

- A social service agency determines that 20% of clients approached on a given day will enroll in an intervention program. What is the probability that the first enrollment comes on the fifth client after four clients refuse (fail) to enroll?
- Is the negative binomial distribution appropriate?

NEGATIVE BINOMIAL WORKFLOW

- A social service agency determines that 20% of clients approached on a given day will enroll in an intervention program. What is the probability that the first enrollment comes on the fifth client after four clients refuse (fail) to enroll?
- Is the negative binomial distribution appropriate?
- What is the appropriate Stata function?

NEGATIVE BINOMIAL WORKFLOW

- A social service agency determines that 20% of clients approached on a given day will enroll in an intervention program. What is the probability that the first enrollment comes on the fifth client after four clients refuse (fail) to enroll?
- Is the negative binomial distribution appropriate?
- What is the appropriate Stata function?

STATA FUNCTIONS

`nbinomialp(n,k,p)`	returns probability of observing ***k*** failures before the n^{th} success	P(x = k)
`nbinomial(n,k,p)`	returns probability of observing ***k* or fewer** failures before the n^{th} success	P(x ≤ k)
`nbinomialtail(n,k,p)`	returns probability of observing ***k* or more** failures before the n^{th} success	P(x ≥ k)

STATA FUNCTIONS

`nbinomialp(n,k,p)`	returns probability of observing ***k*** failures before the n^{th} success	P(x = k)
`nbinomial(n,k,p)`	returns probability of observing ***k* or fewer** failures before the n^{th} success	P(x ≤ k)
`nbinomialtail(n,k,p)`	returns probability of observing ***k* or more** failures before the n^{th} success	P(x ≥ k)

NEGATIVE BINOMIAL WORKFLOW

- A social service agency determines that 20% of clients approached on a given day will enroll in an intervention program. What is the probability that the first enrollment comes on the fifth client after four clients refuse (fail) to enroll?

- Is the negative binomial distribution appropriate?
- What is the appropriate Stata function?
- What is *n*? What is *k*? What is *p*?

NEGATIVE BINOMIAL WORKFLOW

- A social service agency determines that 20% of clients approached on a given day will enroll in an intervention program. What is the probability that the first enrollment comes on the fifth client after four clients refuse (fail) to enroll?

- Is the negative binomial distribution appropriate?
- What is the appropriate Stata function?
- What is *n*? What is *k*? What is *p*?

```
. display nbinomialp(1,4,.20)
.08192
```

DOCUMENT DETAILS

Document produced by Christopher Prener, Ph.D for the Saint Louis University course SOC 5050: QUANTITATIVE ANALYSIS - APPLIED INFERENTIAL STATISTICS. See the course wiki and the repository README.md file for additional details.