Choosing from the (a, b, 0) Class

15M

In this subsection, we will focus on determining the most appropriate $(a,\ b,\ 0)$ class distribution for a given set of data. There are two methods to help us determine the most appropriate distribution.

Method 1: Mean vs. Variance

This method compares the sample mean and unbiased sample variance to determine which distribution is the most suitable, as shown in the table below:

Distribution	Mean vs. Variance
Poisson	$\bar{x} = s^2$
Binomial	$\bar{x} > s^2$
Negative binomial	$\bar{x} < s^2$

For example, if the sample mean is less than the sample variance, a negative binomial distribution would likely fit the sample the best (among the three distributions above).

There is no need to memorize the table above, as the inequalities are based on the mean and variance formulas of each distribution, which are given in the exam table.

For **Poisson**,

$$\mathrm{E}[X] = \quad \lambda \quad = \mathrm{Var}[X]$$

For **binomial**,

$$\mathrm{E}[X] = mq \quad > \quad mq\,(1-q) = \mathrm{Var}[X]$$

For **negative binomial**,

$$\mathrm{E}[X] = r\beta \quad < \quad r\beta(1+\beta) = \mathrm{Var}[X]$$

Method 2: Slope of $\frac{k n_k}{n_{k-1}}$

This method uses the $(a,\ b,\ 0)$ class relationship:

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k}$$

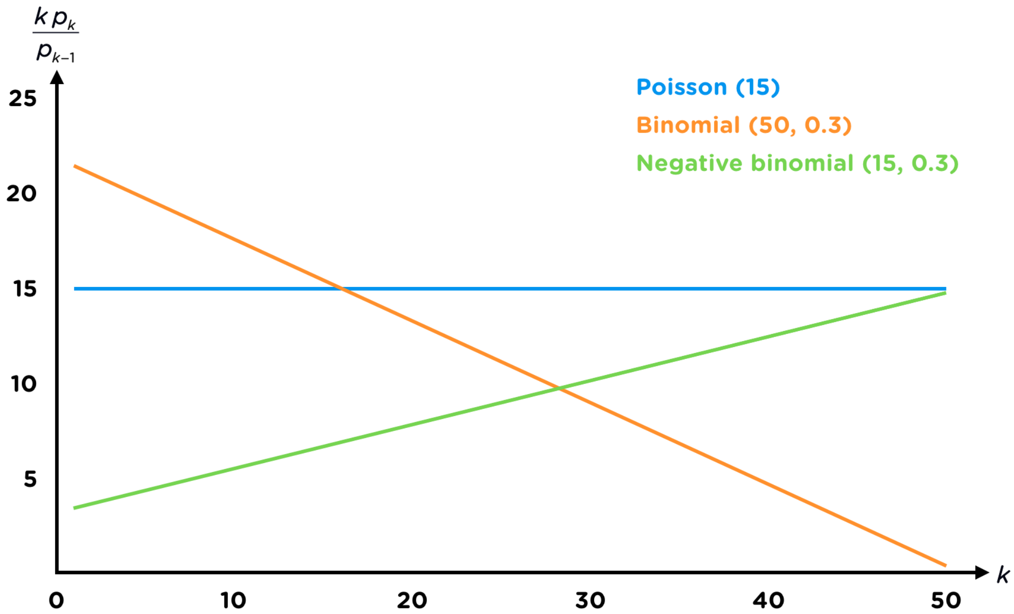
Multiply each side of the equation by k.

$$\frac{k\,p_k}{p_{k-1}} = a\,k + b$$

This gives us a linear equation for $\frac{kp_k}{p_{k-1}}$ in terms of k with a being the slope. Refer to the exam table and note the sign of a for each $(a,\ b,\ 0)$ class distribution.

Distribution	Slope, a
Poisson	0
Binomial	-
Negative binomial	+

The graph below plots $\frac{k\,p_k}{p_{k-1}}$ against k for each distribution.



Therefore, to determine the most appropriate distribution, calculate $\frac{k\,p_k}{p_{k-1}}$ for successive values of k and observe whether they are constant, decreasing, or increasing.

We may estimate p_k as

$$\hat{p}_k = \frac{n_k}{n}$$

where n_k is the number of observed k.

Then, we can multiply both the numerator and the denominator of $\frac{k\,\hat{p}_k}{\hat{p}_{k-1}}$ by n to get

$$\frac{n}{n} \cdot \frac{k\hat{p}_k}{\hat{p}_{k-1}} = \frac{k\,n_k}{n_{k-1}}$$

Therefore, observing the trend of $\frac{k\,p_k}{p_{k-1}}$ is similar to observing the trend of $\frac{k\,n_k}{n_{k-1}}$. This saves time because exam questions usually provide the data in counts rather than proportions.

Example S2.4.2.1

You are given the following auto accident data from 1,000 insurance policies:

Number of Accidents	Number of Policies
0	210
1	367
2	275
3	115
4	29
5	4
6+	0

Determine which of the $(a,\ b,\ 0)$ class distributions would be the most appropriate model for this data.

Solution

To determine the most appropriate distribution, compare the sample mean and variance.

$$\begin{aligned}\bar{x} &= \frac{210(0)+367(1)+275(2)+115(3)+29(4)+4(5)}{1{,}000} \\ &= 1.3980\end{aligned}$$

$$\begin{aligned}s^2 &= \frac{1{,}000}{999}\left[\frac{210(0^2)+367(1^2)+275(2^2)+115(3^2)+29(4^2)+4(\right. \\ &= 1.1127\end{aligned}$$

Since the mean is greater than the variance, we recommend the **binomial** distribution for this data.

Alternatively, we can calculate $\frac{kn_k}{n_{k-1}}$ for each k value.

$$\frac{1\cdot n_1}{n_0}=\frac{1\cdot 367}{210}=1.7476$$

$$\frac{2\cdot n_2}{n_1}=\frac{2\cdot 275}{367}=1.4986$$

$$\frac{3\cdot n_3}{n_2}=\frac{3\cdot 115}{275}=1.2545$$

$$\frac{4\cdot n_4}{n_3}=\frac{4\cdot 29}{115}=1.0087$$

$$\frac{5\cdot n_5}{n_4}=\frac{5\cdot 4}{29}=0.6897$$

$\frac{kn_k}{n_{k-1}}$ decreases as k increases. Therefore, we recommend the **binomial** distribution.

Coach's Remarks

The convention is to compare the mean to the unbiased sample variance, rather than the biased one. However, the difference between the two should be negligible, unless the sample size is extremely small. On the exam, small sample sizes for these problems should be unlikely.

Example S2.4.2.2

You are given the following accident data from 3,000 insurance policies:

Number of Accidents	Number of Policies
0	1,490
1	1,043
2	365
3	85
4	15
5	2
6+	0

Determine which of the $(a,\ b,\ 0)$ class distributions is the most appropriate model for this data.

Solution

To determine the most appropriate distribution, compare the sample mean and variance.

$$\begin{aligned}\bar{x} &= \frac{1{,}490(0)+1{,}043(1)+365(2)+85(3)+15(4)+2(5)}{3{,}000} \\ &= 0.6993\end{aligned}$$

$$\begin{aligned}s^2 &= \frac{3{,}000}{2{,}999}\left[\frac{1{,}490\left(0^2\right)+1{,}043\left(1^2\right)+365\left(2^2\right)+85\left(3^2\right)+15\left(4^2\right)+}{3{,}000}\right. \\ &= 0.6972\end{aligned}$$

Although $\bar{x} > s^2$, they are close enough that we can't confidently conclude that a binomial is the best fit. Let's try the second method and see if we reach the same conclusion.

Calculate $\frac{k n_k}{n_{k-1}}$ for each k value.

$$\frac{1 \cdot n_1}{n_0} = \frac{1 \cdot 1{,}043}{1{,}490} = 0.7000$$

$$\frac{2 \cdot n_2}{n_1} = \frac{2 \cdot 365}{1{,}043} = 0.6999$$

$$\frac{3 \cdot n_3}{n_2} = \frac{3 \cdot 85}{365} = 0.6986$$

$$\frac{4 \cdot n_4}{n_3} = \frac{4 \cdot 15}{85} = 0.7059$$

$$\frac{5 \cdot n_5}{n_4} = \frac{5 \cdot 2}{15} = 0.6667$$

$\frac{k n_k}{n_{k-1}}$ is not increasing or decreasing; it fluctuates around 0.7. Therefore, we recommend the **Poisson** distribution for this data.

Coach's Remarks

It is unreasonable, whether in practice or with made-up examples, to expect the data to yield a sample mean that is exactly equal to the sample variance, or to yield a perfectly constant $\frac{kn_k}{n_{k-1}}$. After all, values from a sample are not expected to perfectly match theoretical results.

Therefore, when using the method of comparing the sample mean and variance, we should be cautious not to make the decision too quickly when the values are unequal but close.

Some students may wonder how close is close enough. Our advice is to check the ratio of sample mean to sample variance, or vice versa. If the ratio is around 1.00, then it is close enough to conclude that Poisson is the best fit. Let's try this on the previous examples.

Example S2.4.2.1

$$\frac{\bar{x}}{s^2} = \frac{1.3980}{1.1127} = 1.26$$