

# **Lessons 013 - 015**

# **Named Discrete Distributions**

**Wednesday, October 11**

Whether a single coin shows up heads is a **Bernoulli random variable.**

The number of flips of a coin until a head is seen is a **Geometric** random variable.

If a coin is flipped a set number of times, the number of heads is a **Binomial random variable.**

# The Binomial Distribution

- If you have repeated, independent, Bernoulli trials, the total number of successes follows a **Binomial Distribution**.
- $X \sim \text{Bin}(n, p)$ , where  $n$  is the number of trials and  $p$  is the probability of success.
- For  $X$ , we have that  $E[X] = np$  and  $\text{var}(X) = np(1 - p)$ .
- The PMF is given by

$$p(x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad x = 0, 1, \dots, n$$

The number of flips of a coin until a set number of heads are seen is a **Negative binomial random variable.**

# The Negative Binomial Distribution

- If you have repeated, independent, Bernoulli trials, the number of failures until a set number of successes follows a **Negative Binomial Distribution**.
- $X \sim \text{NB}(r, p)$ , where  $r$  is the number of successes and  $p$  is the probability of success.

- For  $X$ , we have that  $E[X] = \frac{r(1-p)}{p}$  and  $\text{var}(X) = \frac{r(1-p)}{p^2}$ .

- The PMF is given by

$$p(x) = \binom{x+r-1}{r-1} p^r (1-p)^x \quad x = 0, 1, \dots$$

The number of hearts drawn in a set number of cards with replacement is a **Binomial random variable.**

The number of hearts drawn in a set number of cards without replacement is a **hypergeometric random variable.**



# The Hypergeometric Distribution

- If we sample from a finite population **without replacement**, the total number of successes follows a **Hypergeometric Distribution**.
- $X \sim \text{HyperGeo}(N, M, n)$ , where  $N$  is the population size,  $M$  is the number of successes, and  $n$  is the number of draws being made.

- For  $X$ , we have that  $E[X] = n \frac{M}{N}$  and  $\text{var}(X) = \frac{N-n}{N-1} \cdot n \cdot \frac{M}{N} \left(1 - \frac{M}{N}\right)$ .

- The PMF is given by

$$p(x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$$
$$x = \max\{0, n + M - N\}, \dots, \min\{n, M\}$$

# The Hypergeometric and Binomial Distributions

- A binomial distribution corresponds to sampling **with replacement**.

- We get  $E[X] = n \frac{M}{N}$  and  $\text{var}(X) = n \cdot \frac{N}{M} \left( 1 - \frac{N}{M} \right)$ .

- A hypergeometric distribution corresponds to sampling **without replacement**.

- We get  $E[X] = n \frac{M}{N}$  and  $\text{var}(X) = \frac{N - n}{N - 1} \cdot n \cdot \frac{M}{N} \left( 1 - \frac{N}{M} \right)$ .

- The term  $\frac{N - n}{N - 1}$  is known as the **finite population correction** factor.

- If  $N$  is large, this will be approximately 1 and the binomial can approximate the hypergeometric.

**What if we are not performing  
repeated trials?**

The number of events that occur over a fixed window of time is a **Poisson random variable.**

# The Poisson Distribution

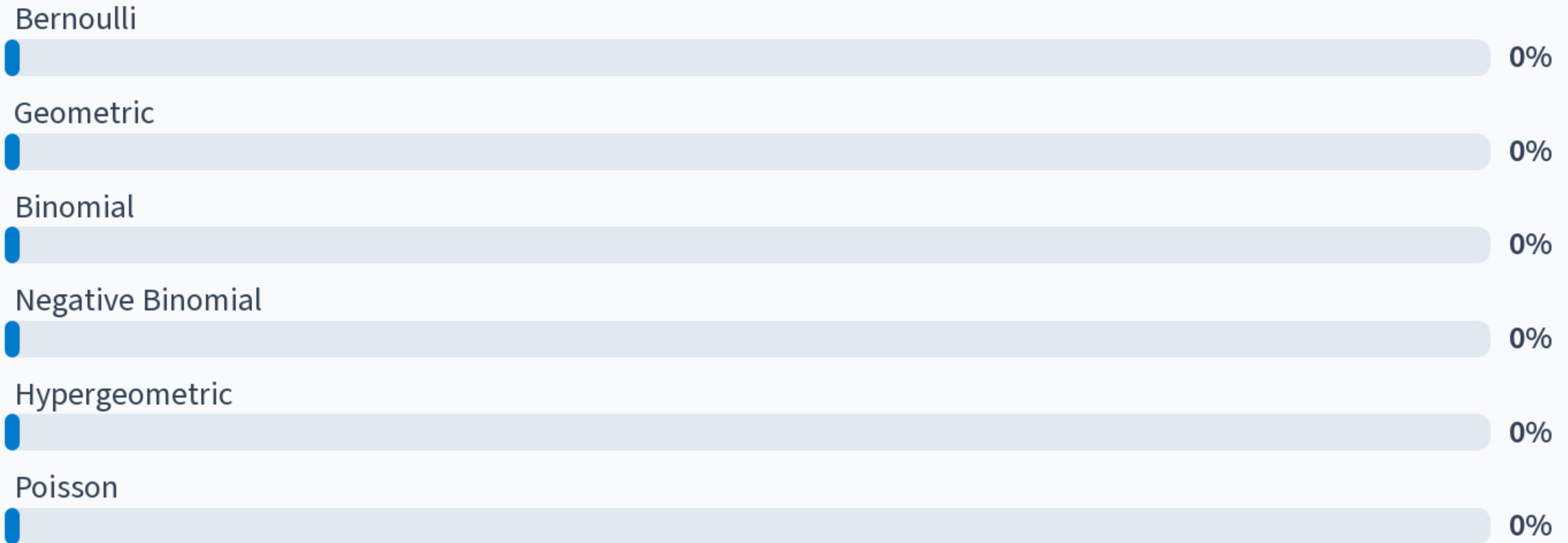
- If events occur at a constant rate over a period of time (or constant rate over region in space), the number of events that occur during a window of time follows a **Poisson random variable**.
- $X \sim \text{Poi}(\mu)$ , where  $\mu$  is the rate that events occur at.
- For  $X$ , we have that  $E[X] = \mu$  and  $\text{var}(X) = \mu$ .
- The PMF is given by

$$p(x) = \frac{e^{-\mu} \mu^x}{x!} \quad x = 0, 1, 2, \dots$$

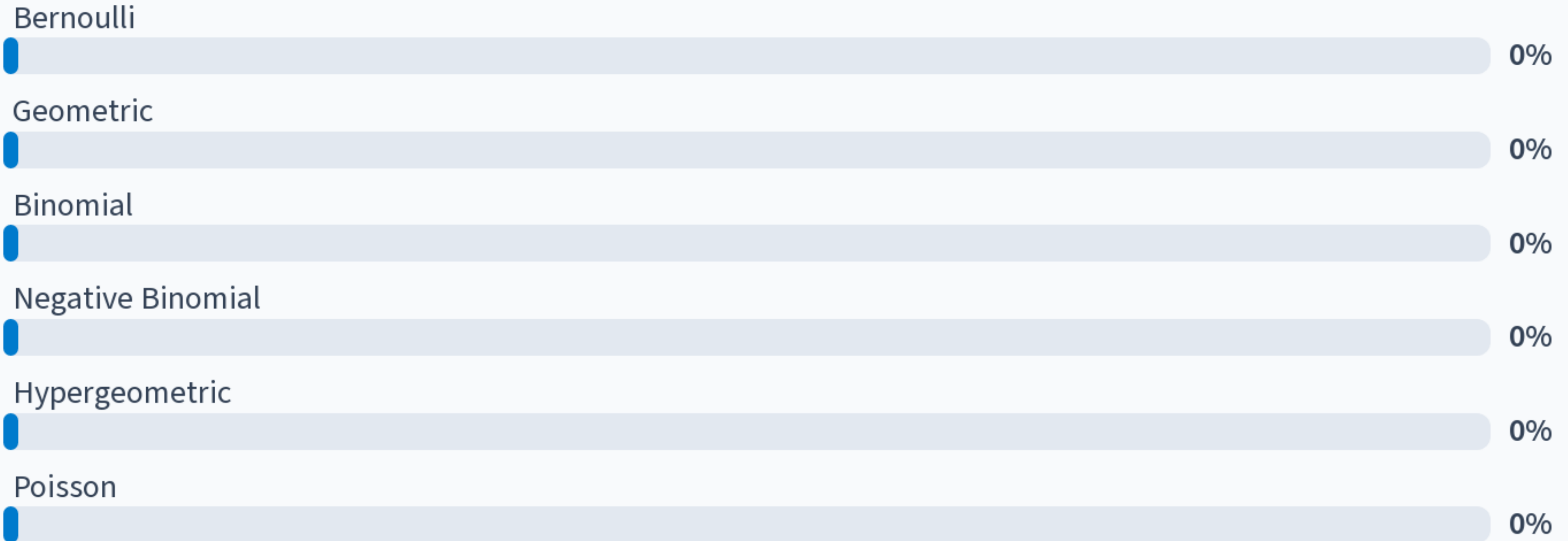
# The Poisson Process

- Suppose that events occur at a rate of  $\alpha$  per some unit of time.
  - The number of events that occur in  $t$  time is expected to be  $\alpha t$ .
- It is often reasonable to take  $X \sim \text{Poi}(\alpha t)$ .
- This is the **Poisson Process**.
- To use the Poisson Process we must assume:
  - Events occur at a constant rate;
  - Events are independent of each other;
  - No two events can occur simultaneously.

What is the most appropriate distribution for the following? In a manufacturing process, you're interested in the number of trials required to produce the first defective item.



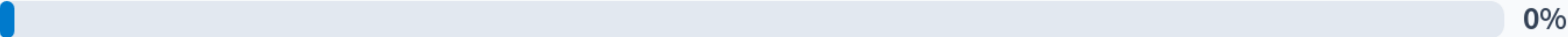
What is the most appropriate distribution for the following? A hospital wants to study the number of patients arriving at the emergency room in a fixed time interval, where the arrival rate is low and the events are rare.



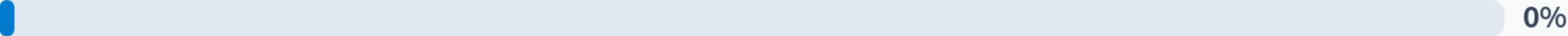


What is the most appropriate distribution for the following? A quality control team inspects a batch of items and classifies the full batch as either defective or non-defective.

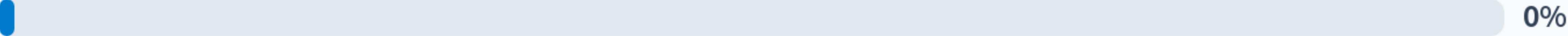
Bernoulli



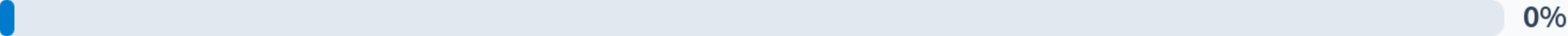
Geometric



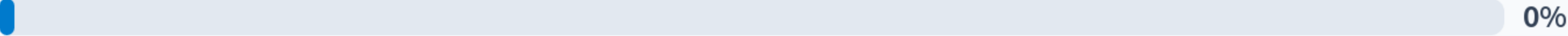
Binomial



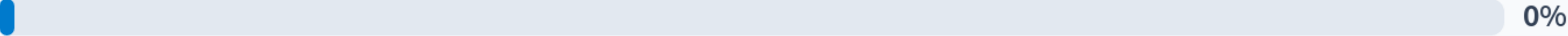
Negative Binomial



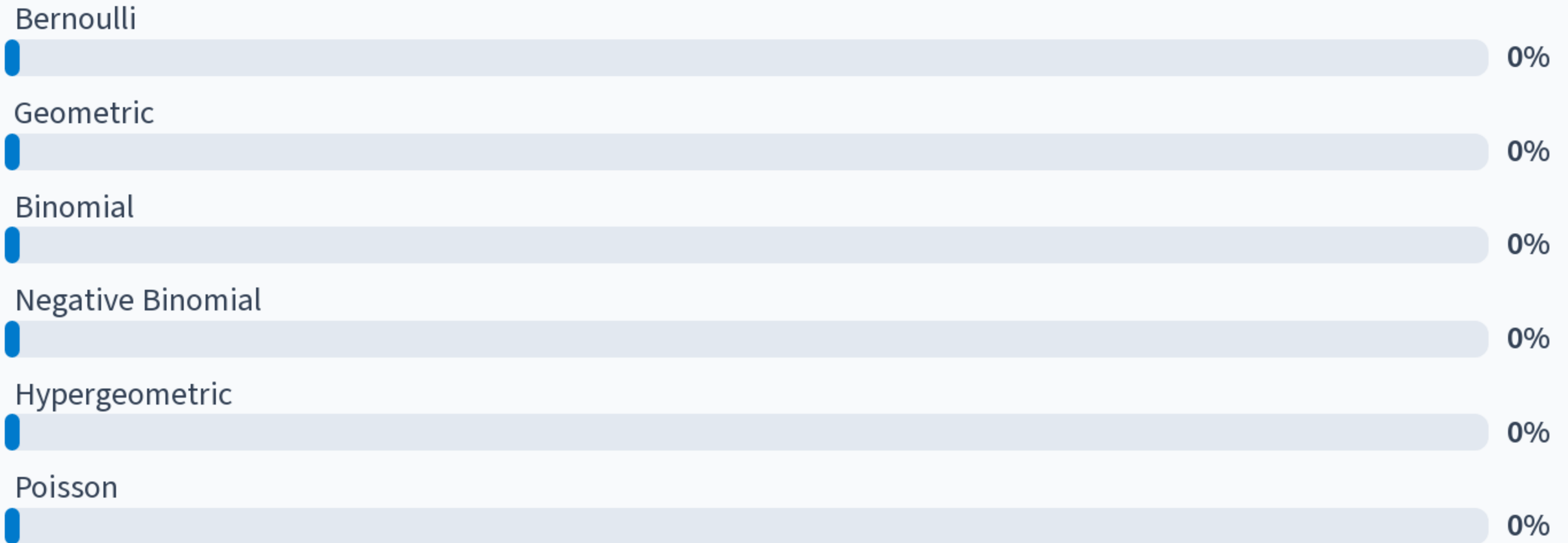
Hypergeometric



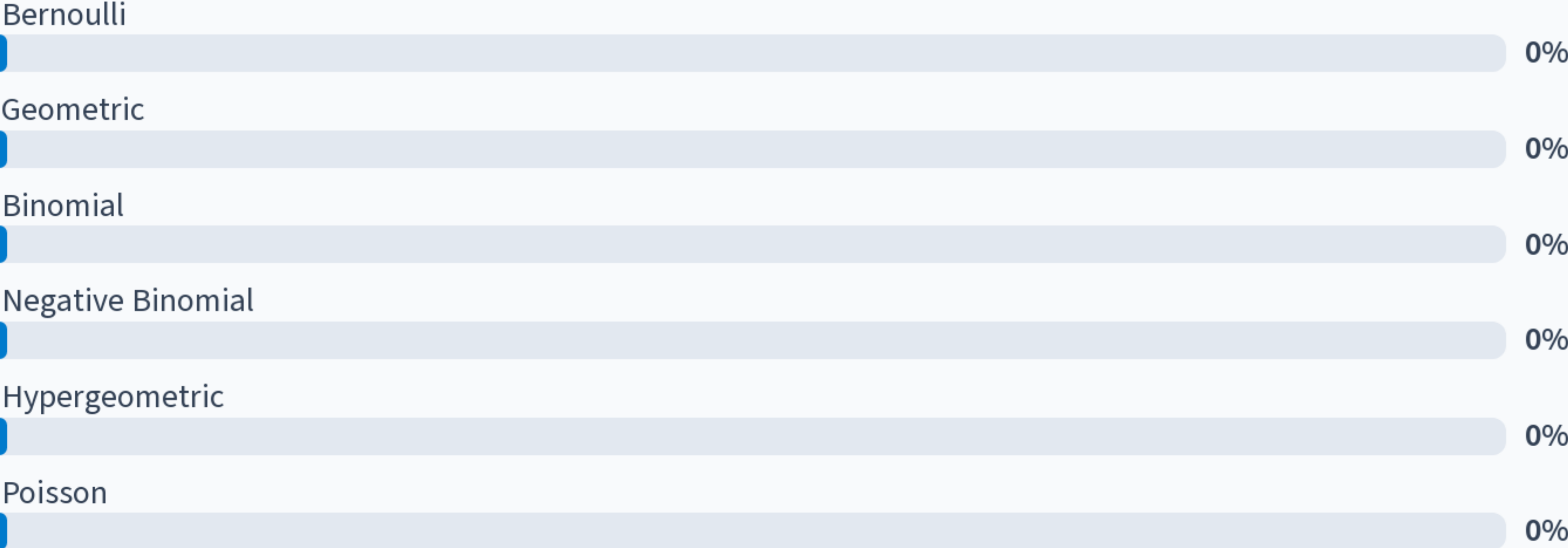
Poisson



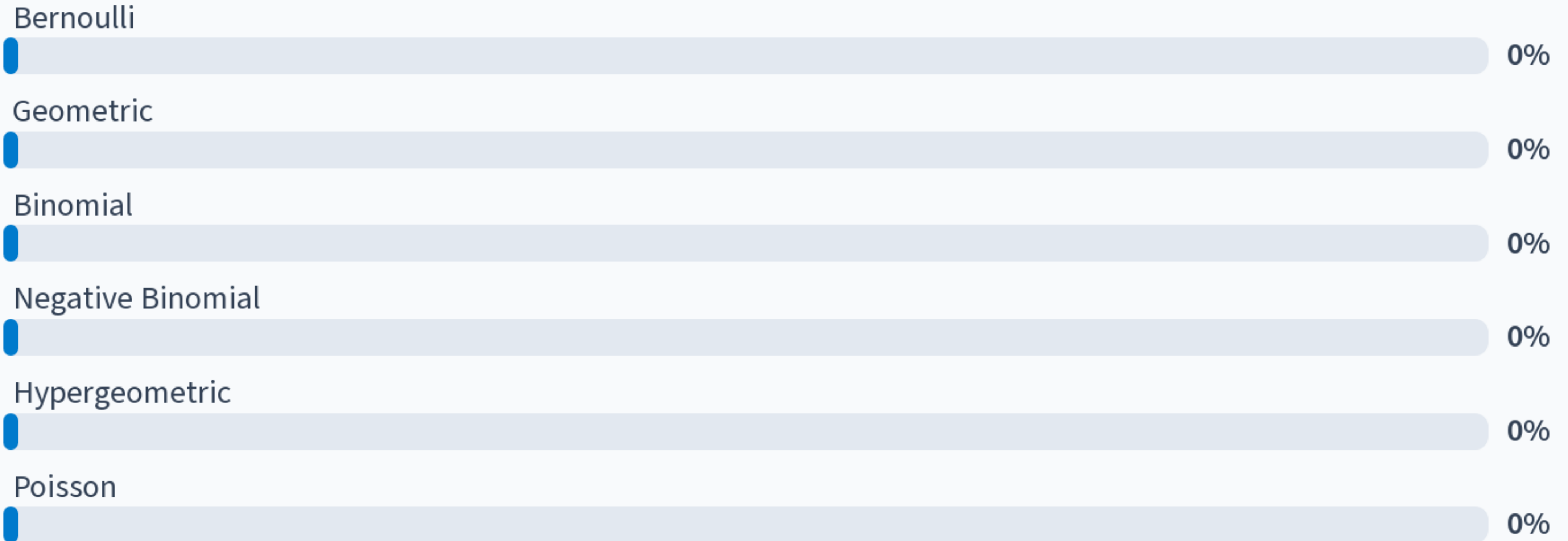
What is the most appropriate distribution for the following? In a survey, you want to determine the number of people who prefer online shopping over in-store shopping.



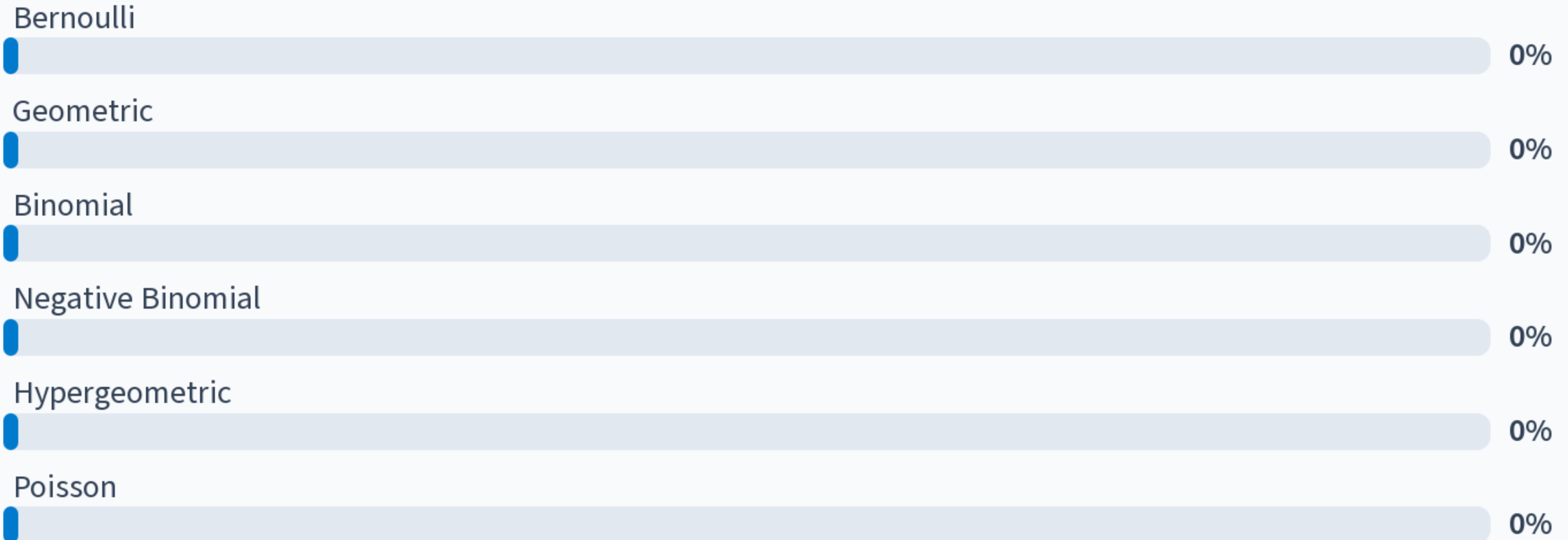
What is the most appropriate distribution for the following? An online streaming platform wants to model the number of times a specific video is watched in an hour, given a known average view rate.



What is the most appropriate distribution for the following? A baseball player wants is interested in the number of home runs they hit in a game, where the probability of hitting a home run is the same for each at-bat.



What is the most appropriate distribution for the following? A network remains functional as long as a certain number of backup routers are operational. A network engineer is interested in how many routers are needed to ensure network reliability.



What is the most appropriate distribution for the following? A lawyer is interested in the finalized jury composition for a trial. and is particularly concerned with the number of women. 13 jurors are selected from the 50 candidates.

