

# **Statistical and Mathematical Methods for Data Analysis**

**Dr. Syed Faisal Bukhari**

Associate Professor

Department of Data Science

Faculty of Computing and Information Technology

University of the Punjab

# Textbooks

- ❑ **Probability & Statistics for Engineers & Scientists**, Ninth Edition, Ronald E. Walpole, Raymond H. Myer
- ❑ **Elementary Statistics: Picturing the World**, 6<sup>th</sup> Edition, Ron Larson and Betsy Farber
- ❑ **Elementary Statistics**, 13<sup>th</sup> Edition, Mario F. Triola

# Reference books

- ❑ **Probability Demystified**, Allan G. Bluman
- ❑ **Schaum's Outline of Probability and Statistics**
- ❑ **MATLAB Primer**, Seventh Edition
- ❑ **MATLAB Demystified** by McMahan, David

# References

Readings for these lecture notes:

- ❑ Probability & Statistics for Engineers & Scientists, Ninth edition, Ronald E. Walpole, Raymond H. Myer

These notes contain material from the above book.

# Negative Binomial Distribution [1]

- Let us consider an experiment where the properties are the same as those listed for a binomial experiment, with the exception that the trials will be repeated **until a fixed number of successes occur**.
- Therefore, instead of finding the probability of  **$x$  successes** in  **$n$  trials**, where  $n$  is fixed, we are now interested in the probability that, the  **$k^{\text{th}}$  success** occurs on the  **$r^{\text{th}}$  trial**.
- Experiments of this kind are called **negative binomial experiments**.

# Negative Binomial Distribution [2]

- ❑ Consider the use of a drug that is known to be effective in **60%** of the cases where it is used. The drug will be considered a **success** if it is **effective** in bringing some degree of relief to the patient.
- ❑ We are interested in finding the **probability** that the **fifth patient** to experience relief is the **seventh patient** to receive the drug during a given week.

# Negative Binomial Distribution [3]

Designating a **success by  $S$**  and a **failure by  $F$** , a **possible order** of achieving the desired result is  **$SFSSSFS$** , which occurs with probability

$$p = 0.6 \text{ (S)}$$

$$q = 0.4 \text{ (F)}$$

$$(0.6)(0.4)(0.6)(0.6)(0.6)(0.4)(0.6) = (0.6)^5(0.4)^2$$

# Negative Binomial Distribution [4]

We could list all possible orders by rearranging the **F's** and **S's** except for the **last outcome**, which must be the **fifth success**. The total number of possible orders is equal to the number of partitions of the **first six trials** into two groups with **2 failures assigned** to the one group and **4 successes assigned** to the other group.

This can be done in  ${}_6C_4 = 15$  mutually exclusive ways.

Hence, if **X represents the outcome** on which the **fifth success** occurs, then

$$P(X = 7) = {}_6C_4 (0.6)^5(0.4)^2 = 0.1866.$$



# Negative Binomial Distribution [5]

If **repeated independent trials** can result in a success with probability **p** and **a failure** with probability **q = 1—p**, then the probability distribution of the random variable **x**, the number of the trial on which the **k<sup>th</sup>** success occurs, is

$$b^*(x; k, p) = {}_{x-1}C_{k-1} p^k q^{x-k}, x = k, k+1, k+2, \dots$$

# Negative Binomial Distribution [6]

**Example:** In an NBA (National Basketball Association) championship series, the team who wins **four games** out of **seven** will be the winner. Suppose that **team A** has probability **0.55** of winning over **team B** and both teams *A* and *B* face each other in the championship games.

**(a)** What is the probability that **team A** will win the series in **six games**?

**(b)** What is the probability that team **A will win the series**?

# Negative Binomial Distribution [7]

(c) If both teams face each other in a regional playoff series and the winner is decided by winning **three** out of **five games**, what is the probability that team A will win a playoff?

# Negative Binomial Distribution [8]

**Solution:**

$$b^*(x; k, p) = {}_{x-1}C_{k-1} p^k q^{x-k}, x = k, k+1, k+2, \dots$$

Let  $x$  denotes **number of matches** until the  **$k^{\text{th}}$  success** occur.

**Here  $x = 6$ ,  $k = 4$**

$$\begin{aligned} \text{(a) } b^*(6; 4, 0.55) &= {}_{6-1}C_{4-1} (0.55)^4 (0.45)^{6-4} \\ &= {}_5C_3 (0.55)^4 (0.45)^2 \\ &= 0.1853 \end{aligned}$$

# Negative Binomial Distribution [9]

$$b^*(x; k, p) = {}_{x-1}C_{k-1} p^k q^{x-k}, x = k, k+1, k+2, \dots$$

Let  $x$  denotes **number of matches** until the  **$k^{\text{th}}$  success** occur.

Here  $k = 4$

**(b)  $P(\text{team A wins the championship series}) = P(X \geq 4)$**

$$\begin{aligned} P(X \geq 4) &= b^*(4; 4, 0.55) + b^*(5; 4, 0.55) \\ &\quad + b^*(6; 4, 0.55) + b^*(7; 4, 0.55) \\ &= 0.0915 + 0.1647 + 0.1853 + 0.1668 \\ &= 0.6083. \end{aligned}$$

# Negative Binomial Distribution [10]

$$b^*(x; k, p) = {}_{x-1}C_{k-1} p^k q^{x-k}, x = k, k+1, k+2, \dots$$

Let  $x$  denotes **number of matches** until the  **$k^{\text{th}}$  success** occur.

Here  $k = 3$

**(c) P(team A wins the playoff) =**

$$\begin{aligned} P(X \geq 3) &= b^*(\mathbf{3}; \mathbf{3}, 0.55) + b^*(\mathbf{4}; \mathbf{3}, 0.55) + b^*(\mathbf{5}; \mathbf{3}, 0.55) \\ &= 0.1664 + 0.2246 + 0.2021 \\ &= 0.5931. \end{aligned}$$

# Negative Binomial Distribution [11]

□ If we consider the special case of the **negative binomial distribution where  $k = 1$** , we have a probability distribution for the number of trials required for a **single success**. An example would be the tossing of a **coin until a head occurs**. We might be interested in the probability that the first head occurs on the fourth toss. The **negative binomial distribution** reduces to the form

□ Here  $k = 1$

$$b^*(x; 1, p) = p q^{x-1}, x = 1, 2, 3, \dots$$

# Negative Binomial Distribution [12]

- Since the successive terms constitute a **geometric progression**, it is customary to refer to this special case as the **geometric distribution** and denote its values by  **$g(x; p)$** .



# Negative Binomial Distribution [13]

**Example** An oil drilling company ventures into various locations, and its success or failure is **independent** from one location to another. Suppose the probability of a success at any specific location is **0.25**.

**(a)** What is the probability that a driller drills **10 locations** and finds **1 success**?

**(b)** The driller feels that he will go **bankrupt** if he drills **10 times** before the **first success occurs**. What are the driller's prospects for bankruptcy?

**(c)** What is the probability that the driller gets the first success on the **11<sup>th</sup> trial**?

# Negative Binomial Distribution [14]

**Solution:** Let  $x$  denotes **number of successful drill**

$$(a) b(x; n, p) = {}_n C_x p^x q^{n-x}, x = 1, 1, 2, \dots, n$$

$$b(1; 10, 0.25) = {}_{10} C_1 0.25^1 0.75^{10-1} \\ = 0.1877$$

**(b) Since trials are independent**

**Probability of bankruptcy** =  $q \times q \times q \times q \times q \times q \times q \times q \times q \times q$

$$= q^{10}$$

$$= (0.75)^{10}$$

$$= 0.056$$

(c) Let x denotes **number of locations explore** until the **1<sup>th</sup> success**.

**Here x = 11**

$$g(x; p) = p q^{x-1}, x = 1, 2, 3, \dots$$

$$\begin{aligned} g(\mathbf{11}; 0.25) &= (0.25) (0.75)^{11-1} \\ &= (0.25) (0.75)^{10} \\ &= \mathbf{0.0141} \end{aligned}$$

# Negative Binomial Distribution [15]

**Example** Consider the information of Review **Exercise in previous question**. The driller feels that he will “**hit it big**” if the **second success** occurs on or before **the sixth attempt**. What is the probability that the driller will “**hit it big?**”

# Negative Binomial Distribution [16]

**Solution:**  $b^*(x; k, p) = {}_{x-1}C_{k-1} p^k q^{x-k}$ ,  $x = k, k+1, k+2, \dots$

Let  $x$  denotes **number of locations explore** to get  **$k^{\text{th}}$  successes**. Here  $x = 6$ ,  $k = 2$ ,  $p = 0.25$

$$\begin{aligned} b^*(6; 2, 0.25) &= {}_{6-1}C_{2-1} (0.25)^2 (0.75)^{6-2} \\ &= {}_5C_1 (0.25)^2 (0.75)^4 \\ &= 0.0989 \end{aligned}$$

# Negative Binomial Distribution [17]

**Example:** A couple decides they will continue to have children **until** they have **two males**. Assuming that  $P(\text{male}) = 0.5$ , what is the probability that their **second male** is their **fourth child**?

# Negative Binomial Distribution [18]

**Solution:**  $b^*(x; k, p) = {}_{x-1}C_{k-1} p^k q^{x-k}$ ,  $x = k, k+1, k+2, \dots$

Let  $x$  denotes **number of children** on which  **$k^{\text{th}}$  male** born.

Here  **$x = 4$** ,  $p = 0.5$ ,  **$k = 2$  (number of male)**

$$\begin{aligned} b^*(4; 2, 0.5) &= {}_{4-1}C_{2-1} (0.5)^2 (0.5)^{4-2} \\ &= {}_3C_1 (0.5)^2 (0.5)^2 \\ &= 0.1875 \end{aligned}$$

# Negative Binomial Distribution [19]

**Example** An oil company conducts a geological study that indicates that an exploratory oil well should have a 20% chance of striking oil. What is the probability that the **first strike** comes on the **third well drilled**?

To find the requested probability, we need to find  **$P(X = 3)$** . Note that  $X$  is technically a geometric random variable, since we are only looking for one success. Since a geometric random variable is just a special case of a negative binomial random variable, we'll try finding the probability using the negative binomial p.m.f.



# Negative Binomial Distribution [20]

Let  $x$  denotes **number of drills** on which  **$k^{\text{th}}$  strike** occur.

Here  $p = 0.20$ ,  $1-p = 0.80$

**$k = 1$** , and  **$x = 3$**

$$b^*(x; k, p) = {}_{x-1}C_{k-1} p^k q^{x-k}, \quad x = k, k+1, k+2, \dots$$

$$\begin{aligned} b^*(\mathbf{3}; \mathbf{1}, 0.2) &= {}_{3-1}C_{1-1} (0.2)^1 (0.8)^{3-1} \\ &= {}_2C_0 (0.2)^1 (0.8)^2 \\ &= 0.128 \end{aligned}$$

# Negative Binomial Distribution [21]

**Example:** What is the probability that the **third strike** comes on the **seventh well** drilled?

## Solution:

Let  $x$  denotes **number of drills** on which  **$k^{\text{th}}$  strike** occur.

$$p = 0.20, 1-p = 0.80, \mathbf{x} = \mathbf{7} \text{ and } \mathbf{k} = \mathbf{3}$$

$$b^*(\mathbf{x}; \mathbf{k}, p) = {}_{x-1}C_{k-1} p^k q^{x-k}, x = k, k+1, k+2, \dots$$

$$b^*(\mathbf{7}; \mathbf{3}, 0.2) = {}_{7-1}C_{3-1} (0.2)^3 (0.8)^{7-3}$$

$$= {}_6C_2 (0.2)^3 (0.8)^4$$

$$= 0.049$$