Lecture Notes EMSE 4765:
DATA ANALYSIS - Probability Review

Chapter 4: Discrete Random Variables

Version: 01/08/2021

Text Book: A Modern Introduction to Probability and Statistics, Understanding Why and How

By: F.M. Dekking. C. Kraaikamp, H.P.Lopuhaä and L.E. Meester

4 Discrete Random Variables

4.1 Random Variables...

Definition: Let Ω be a sample space. A discrete random variable is a function $X:\Omega \to \mathbb{R}$, that takes on a finite number of values $a_1, a_2, \ldots, a_n$ or an infinite number of values $a_1, a_2, \ldots$.

Die Example: $\Omega = \{1, 2, 3, 4, 5, 6\}$: Introducing **the random variable function** $X:\Omega \to \mathbb{R}$, one defines:

$$X(\{x\}) = x,\ x = 1, \ldots 6$$

The "function" argument part "$(\{x\})$" and **the wording "function"** in the above definition is **usually omitted** and one writes simply $\boldsymbol{X = x}$. However, that means **one ought to provide the definition in words** for the random variable X, i.e. in the die example:

$$X \equiv \text{"The outcome value when throwing a die once".}$$

4 Discrete Random Variables

4.2 The probability distribution function of a discrete random variable...

Coin Toss Example: $\Omega = \{\text{heads, tails}\}$: Introducing the random variable function $X : \Omega \to \mathbb{R}$, one defines: $X(\{\text{heads}\}) = 0$ and $X(\{\text{tails}\}) = 1$, or vice versa. When one omits the "$(\{\text{heads}\})$" part and the wording "function", one needs to define:

$$X = \begin{cases} 0 & \text{"heads" turns up on the single coin toss} \\ 1 & \text{"tails" turns up on the single coin toss} \end{cases}$$

Definition: The **probability mass function (pmf) $p(\cdot)$** of a discrete random variable $X : \Omega \to \mathbb{R}$ is the function $p : \mathbb{R} \to [0, 1]$, defined by

$$p(x) = Pr(X = x) \text{ for } -\infty < x < \infty$$

- If X is a discrete random variable that takes on values $x_1, x_2, \ldots$ then

$$p(x_i) > 0,\ p(x_1) + p(x_2) + \ldots = 1 \text{ and } p(x) = 0 \text{ for all other } x.$$

4 Discrete Random Variables

4.2 The probability distribution of a discrete random variable...

Die Example: $p(x) = Pr(X = x) = \frac{1}{6}$, $x = 1, 2, \ldots, 6$.

Coin Toss Example: $p(x) = Pr(X = x) = \frac{1}{2}$, $x = 0, 1$.

> **Definition:** The **cumulative distribution function (cdf)** $\boldsymbol{F(\,\cdot\,)}$ of a discrete random variable $X : \Omega \rightarrow \mathbb{R}$ is the function $F : \mathbb{R} \rightarrow [0, 1]$, defined by
>
> $$F(x) = Pr(X \leq x) \text{ for } -\infty < a < \infty$$

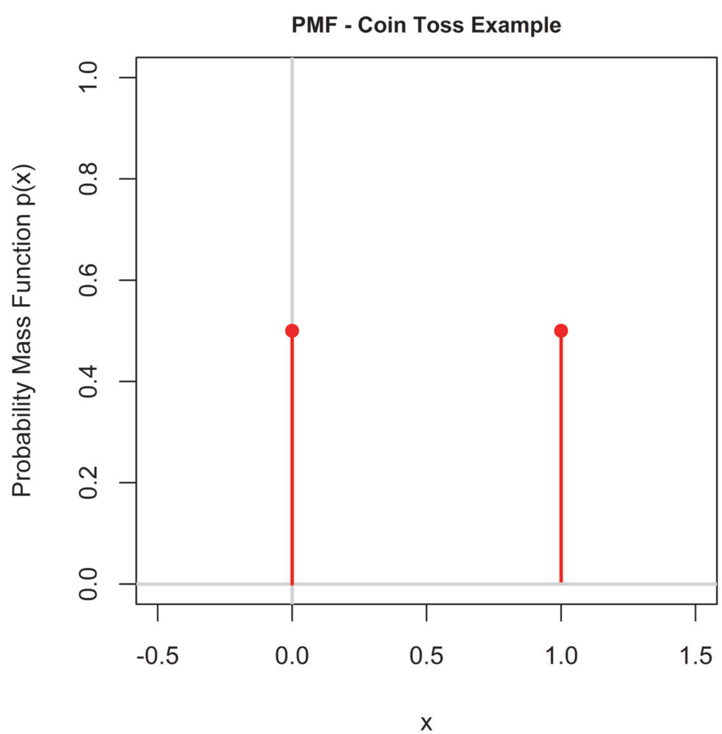
- If X is a discrete random variable that takes on values $a_1, a_2, \ldots$ then

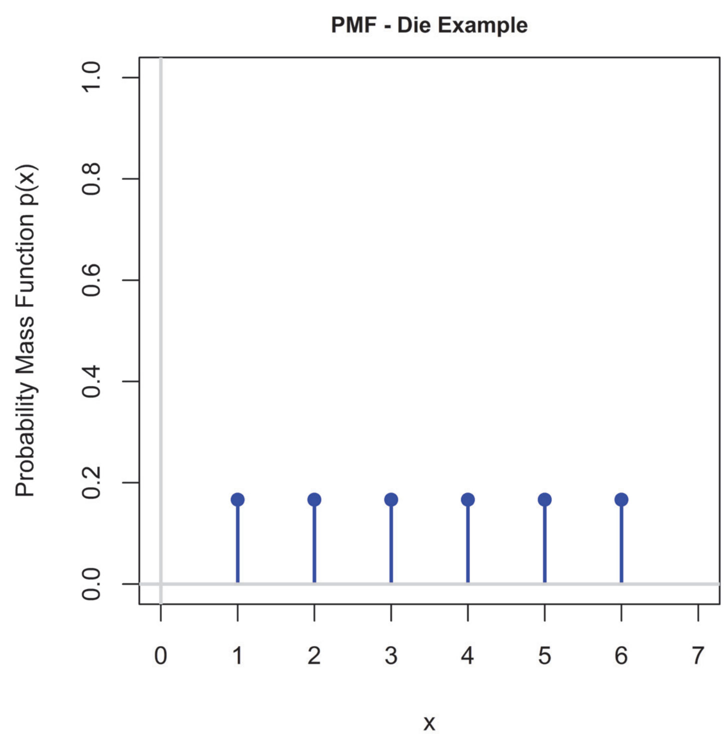
$$p(a_i) > 0,\ p(a_1) + p(a_2) + \ldots = 1 \text{ and } F(a) = \sum_{a_i \leq a} p(a_i).$$

4 Discrete Random Variables

4.2 The probability distribution of a discrete random variable. . .

- Both **the probability mass function** and **the cumulative distribution function** of a discrete random variable X contain **all the probabilistic information**; **The probability distribution of X** is determined by either.

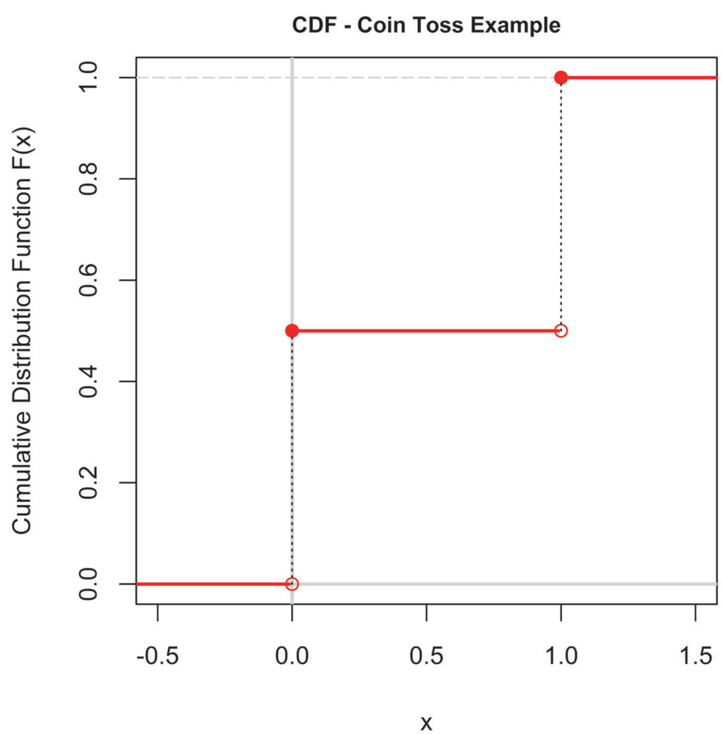


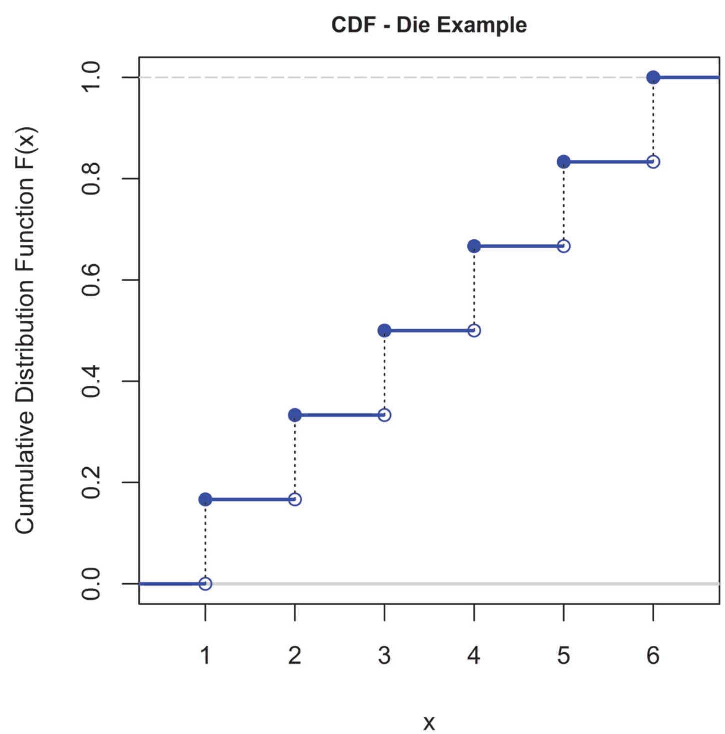


4 Discrete Random Variables

4.2 The probability distribution of a discrete random variable. . .

- Both **the probability mass function** and **the cumulative distribution function** of a discrete random variable X contain **all the probabilistic information**; **The probability distribution of X** is determined by either.





Three properties of distribution functions:

1. For $a \leq b$ we have $F(a) \leq F(b)$ since $\{X \leq a\} \subset \{X \leq b\}$

2. Since $F(a)$ is a probability, $F(a) \in [0, 1]$ and:

$$\lim_{a \to \infty} F(a) = \lim_{a \to \infty} Pr(X \leq a) = 1$$
$$\lim_{a \to -\infty} F(a) = \lim_{a \to -\infty} Pr(X \leq a) = 0$$

3. F is **right continuous**, i.e. $\lim_{\epsilon \downarrow 0} F(a + \epsilon) = F(a)$.

Think "coming from the right" we have continuity.

- **The Bernoulli distribution** models an experiment with only two possible outcomes "Success" $\equiv 1$ and "Failure" $\equiv 0$, respectively.

Definition: A discrete random variable X has a **Bernoulli distribution with parameter $\boldsymbol{p}$** $\in [0,1]$, if its probability mass function is given by

$$p_X(1) = Pr(X = 1) = p \text{ and } p_X(0) = Pr(X = 0) = 1 - p$$

We denote this distribution with $\boldsymbol{Ber(p)}$ and $\boldsymbol{X \sim Ber(p)}$.

- Assume that **per trial probability to win is p** and that **winning in a trial is independent from the previous trials.** Defining now the random variable

$$X \equiv \text{"\# of successes in } n \text{ trials"}$$

$$\binom{n}{k} = \textbf{the number of different ways one chooses } k \textbf{ out of a group of } n.$$

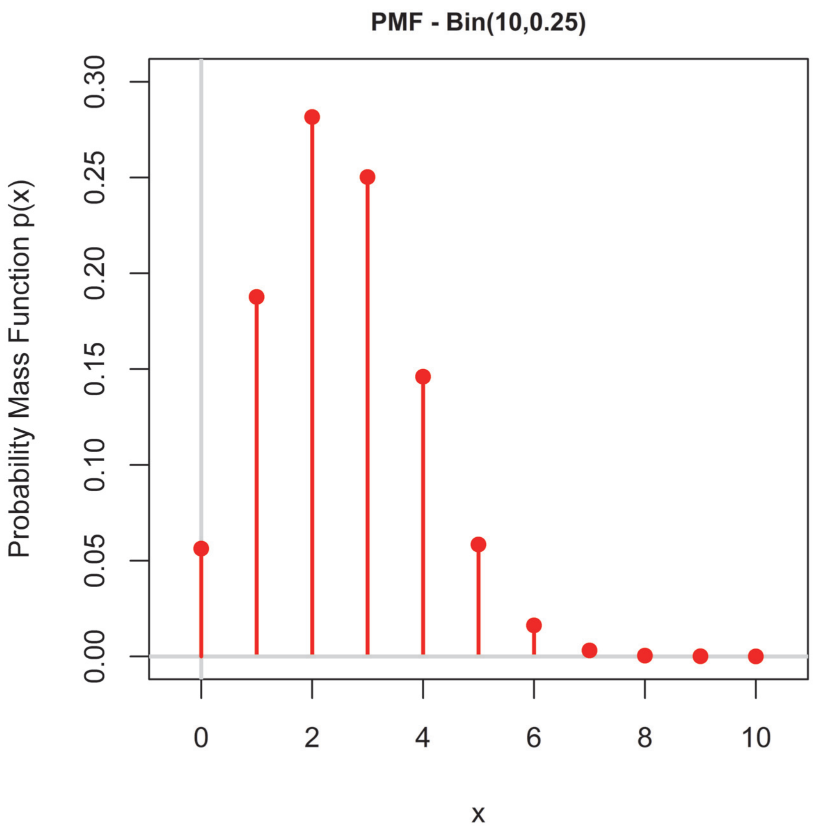
$$= \frac{n!}{k!(n-k)!} = \frac{n \cdot (n-1) \ldots (n-k+1)}{k!} \Rightarrow \binom{10}{2} = \frac{10 \cdot 9}{2} = 45$$

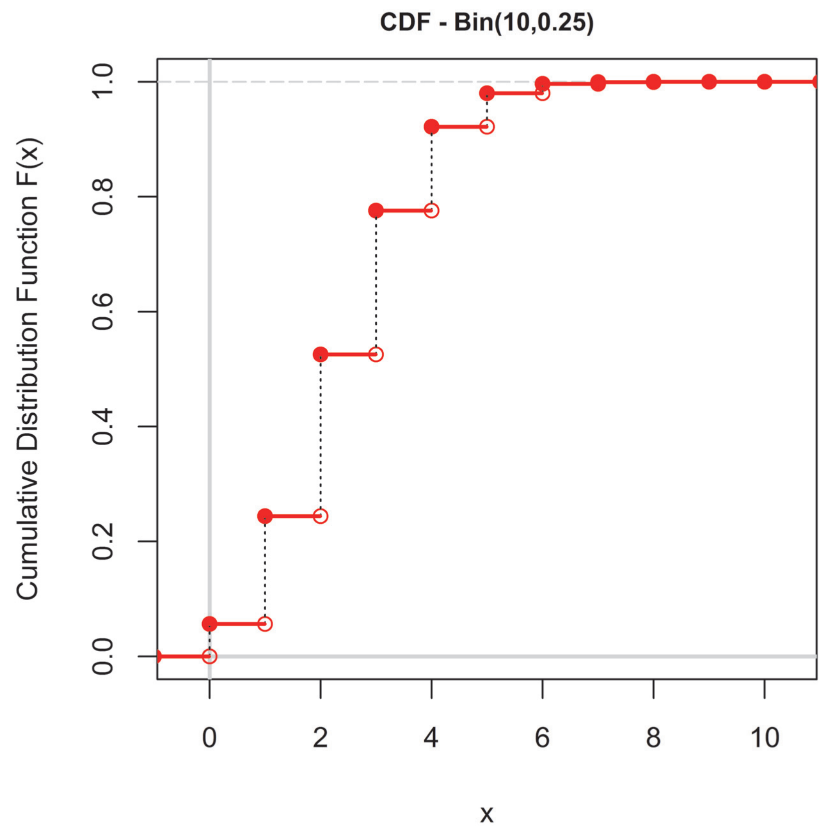
Definition: A discrete random variable X has a **Binomial distribution with parameters $n > 0$ and p** $\in [0, 1]$, if its probability mass function is given by

$$p_X(k) = Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \text{ for } k = 1, \ldots, n$$

We denote this distribution with $Bin(n, p)$ and $X \sim Bin(n, p)$

PMF $X \sim Bin(10, 0.25)$ CDF $X \sim Bin(10, 0.25)$





- Assume that **per trial probability to win is p** and that **winning in a trial is independent from the previous trials.** Define random variable $X \equiv$ "# of trials to win"

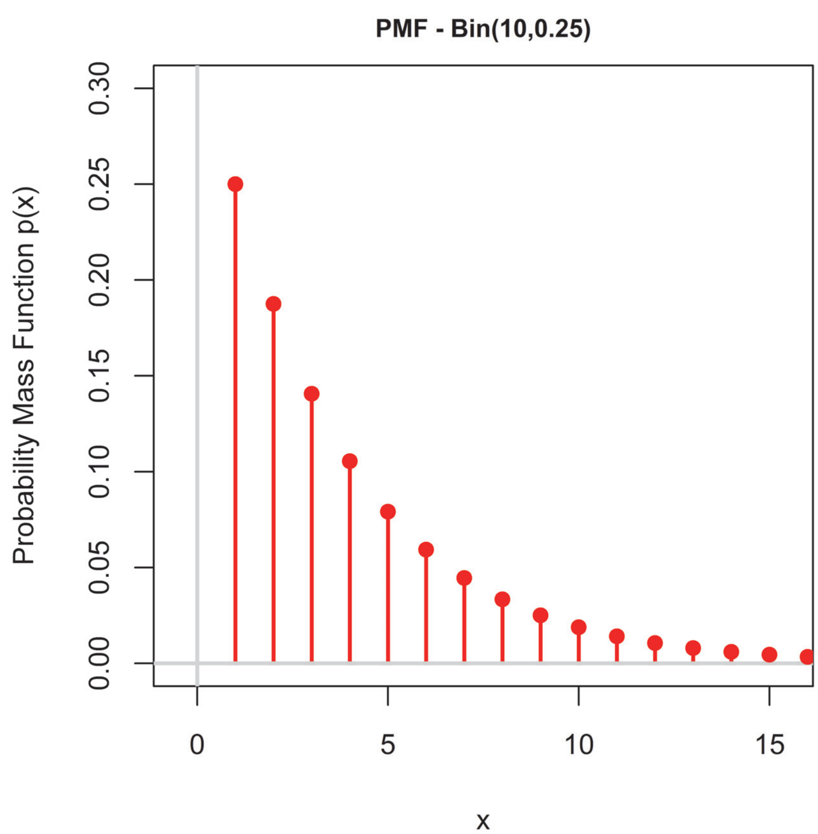
$$Pr(X = 1) = p,\ Pr(X = 2) = (1 - p)p,\ \ldots,\ Pr(X = k) = (1 - p)^{k-1}p$$

Definition: A discrete random variable X has a **Geometric distribution with parameter p** $\in [0, 1]$, if its probability mass function is given by

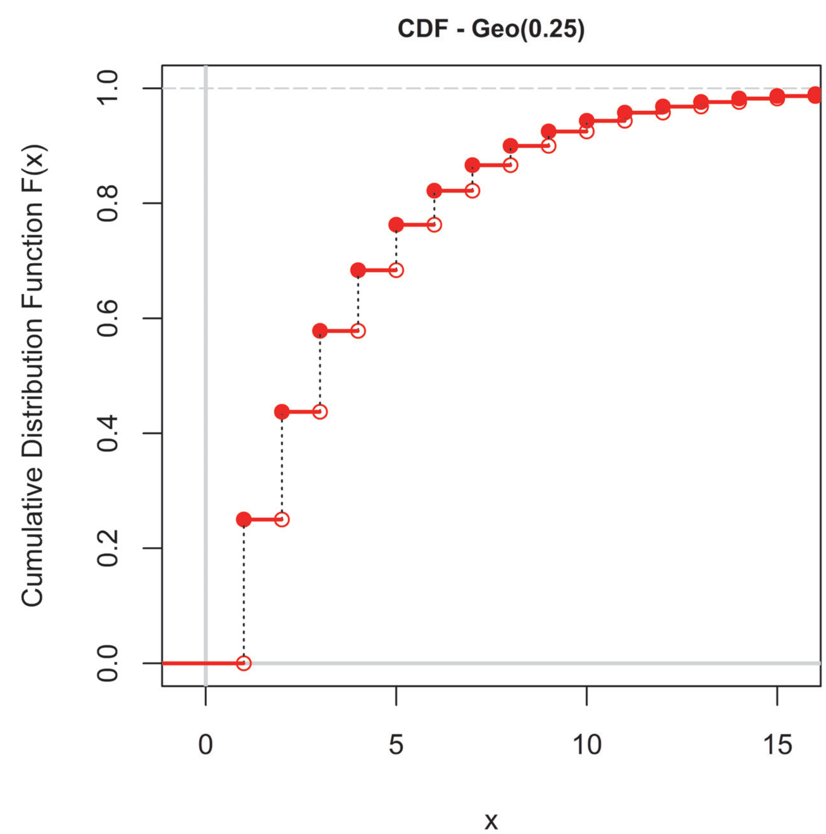
$$p_X(k) = Pr(X = k) = (1 - p)^{k-1}p \text{ for } k = 1, 2, 3$$

We denote this distribution with $Geo(p)$ and $X \sim Geo(p)$

PMF $X \sim Geo(0.25)$



CDF $X \sim Geo(0.25)$



Exercise: Let $X \sim Geo(p)$. Show that $Pr(X > n) = (1-p)^n$

Solution: $Pr(X > n) = Pr($"No success in n trials"$) = (1-p)^n$

- **The memoryless property of a $Geo(p)$ distribution:**

$$Pr(X > n + k | X > k) = Pr(X > n)$$

Proof:

$$Pr(X > n + k | X > k) = \frac{Pr(X > n + k \cap \; X > k)}{Pr(X > k)} = \frac{Pr(X > n + k)}{Pr(x > k)}$$
$$= \frac{(1-p)^{n+k}}{(1-p)^k} = (1-p)^n = Pr(X > n).$$

Conclusion: If you **did not win in the first k trials**, **the probability that you will not win in the next n trials** is the same as **the probability of not winning in the first n trials**.