

Probability and Discrete Probability Distributions

Daniel A. Menascé, Ph.D.
Department of Computer Science
George Mason University

1

© 2001 D. A. Menascé. All Rights Reserved.

Review of Probability Concepts

- Classical (theoretical) approach:

$$\frac{\text{No. Ways Event A Can Occur}}{\text{Total Number of Events}} \quad \textit{process has to be known!}$$

- Empirical approach (relative frequency):

$$\frac{\text{No. Times Result A Occurred in the Experiment}}{\text{Total Number of Observations}}$$

- The relative frequency converges to the probability for a large number of experiments.

2

© 2001 D. A. Menascé. All Rights Reserved.

Review of Probability Rules

1. A probability is a number between 0 and 1 assigned to an event that is the outcome of an experiment:

$$P[A] \in [0,1]$$

2. Complement of event A.

$$P[\bar{A}] = 1 - P[A]$$

3. If events A and B are mutually exclusive then

$$P[A \text{ or } B] = P[A] + P[B]$$

$$P[A \text{ and } B] = 0$$

3

© 2001 D. A. Menascé. All Rights Reserved.

Review of Probability Rules (cont'd)

4. If events A_1, \dots, A_N are mutually exclusive and collectively exhaustive then:

$$\sum_{i=1}^N P[A_i] = 1$$

5. If events A and B are not mutually exclusive then: $P[A \text{ or } B] = P[A] + P[B] - P[A \text{ and } B]$

6. Conditional Probability:

$$P[A | B] = \frac{P[A \text{ and } B]}{P[B]} = \frac{P[B | A]P[A]}{P[B]}$$

4

© 2001 D. A. Menascé. All Rights Reserved.

Review of Probability Rules (cont'd)

7. If events A and B are independent (i.e., $P[A] = P[A|B]$ and $P[B] = P[B|A]$) then:

$$P[A \text{ and } B] = P[A] \times P[B]$$

8. If events A and B are not independent then
- $$P[A \text{ and } B] = P[A | B]P[B] = P[B | A]P[A]$$

9. Theorem of Total Probability: if events A_1, \dots, A_N are mutually exclusive and collectively exhaustive then

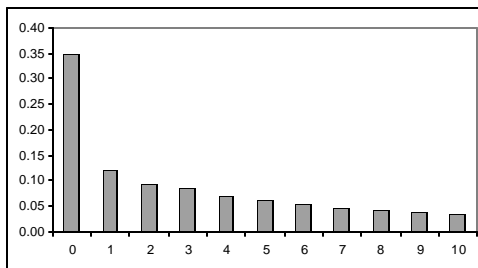
$$P[B] = \sum_{i=1}^N P[B | A_i]P[A_i]$$

5

© 2001 D. A. Menascé. All Rights Reserved.

Discrete Probability Distribution

- Distribution: set of all possible values and their probabilities.



Number of I/Os per Transaction	Probability
0	0.350
1	0.120
2	0.095
3	0.085
4	0.070
5	0.060
6	0.054
7	0.048
8	0.043
9	0.040
10	0.035
	1.000

6

© 2001 D. A. Menascé. All Rights Reserved.

Moments of a Discrete Random Variable

- Expected Value:

$$m = E[X] = \sum_{\forall i} X_i \times P[X_i]$$

- k-th moment:

$$m = E[X^k] = \sum_{\forall i} X_i^k \times P[X_i]$$

Number of I/Os per Transaction	Probability	For First Moment (average)	For Second Moment
0	0.350	0.000	0.000
1	0.120	0.120	0.120
2	0.095	0.190	0.380
3	0.085	0.255	0.765
4	0.070	0.280	1.120
5	0.060	0.300	1.500
6	0.054	0.324	1.944
7	0.048	0.336	2.352
8	0.043	0.344	2.752
9	0.040	0.360	3.240
10	0.035	0.350	3.500
	1.000	2.859	17.673

mean
second moment

7

© 2001 D. A. Menascé. All Rights Reserved.

Central Moments of a Discrete Random Variable

- k-th central moment:

$$E[(X - \bar{X})^k] = \sum_{\forall i} (X_i - \bar{X})^k \times P[X_i]$$

- The variance is the second central moment:

$$\begin{aligned} s^2 &= E[(X - \bar{X})^2] = E[X^2 + (\bar{X})^2 - 2X\bar{X}] \\ &= E[X^2] + (\bar{X})^2 - 2(\bar{X})^2 = \\ &= E[X^2] - (\bar{X})^2 \end{aligned}$$

8

© 2001 D. A. Menascé. All Rights Reserved.

Central Moments of a Discrete Random Variable

Number of I/Os per Transaction	Probability	For First Moment (average)	For Second Moment	For Second Central Moment
0	0.350	0.000	0.000	2.8609
1	0.120	0.120	0.120	0.4147
2	0.095	0.190	0.380	0.0701
3	0.085	0.255	0.765	0.0017
4	0.070	0.280	1.120	0.0911
5	0.060	0.300	1.500	0.2750
6	0.054	0.324	1.944	0.5328
7	0.048	0.336	2.352	0.8231
8	0.043	0.344	2.752	1.1365
9	0.040	0.360	3.240	1.5085
10	0.035	0.350	3.500	1.7848
	1.000	2.859	17.673	9.4991

average

variance

9

© 2001 D. A. Menascé. All Rights Reserved.

Properties of the Mean

- The mean of the sum is the sum of the means.

$$E[X + Y] = E[X] + E[Y]$$

- If X and Y are independent random variables, then the mean of the product is the product of the means.

$$E[XY] = E[X]E[Y]$$

10

© 2001 D. A. Menascé. All Rights Reserved.

Discrete Random Variables

- Binomial
- Hypergeometric
- Negative Binomial
- Geometric
- Poisson

11

© 2001 D. A. Menascé. All Rights Reserved.

The Binomial Distribution

- Distribution: based on carrying out independent experiments with two possible outcomes:
 - Success with probability p and
 - Failure with probability $(1-p)$.
- A binomial r.v. counts the number of successes in n trials.

$$P[X = k] = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

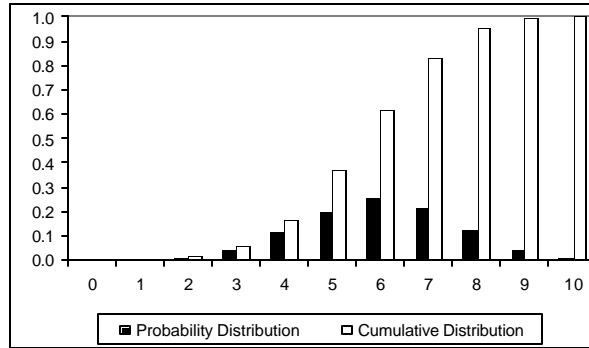
12

© 2001 D. A. Menascé. All Rights Reserved.

The Binomial Distribution

Success Probability 0.6 (p)
 Number of Attempts 10 (n)

Number of Attempts (k)	Probability k successful attempts in n	Cumulative
0	0.000105	0.000105
1	0.001573	0.001678
2	0.010617	0.012295
3	0.042467	0.054762
4	0.111477	0.166239
5	0.200658	0.366897
6	0.250823	0.617719
7	0.214991	0.832710
8	0.120932	0.953643
9	0.040311	0.993953
10	0.006047	1.000000



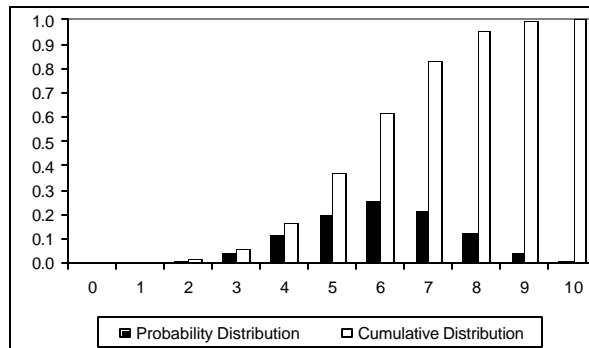
13

© 2001 D. A. Menascé. All Rights Reserved.

The Binomial Distribution

Success Probability 0.6 (p)
 Number of Attempts 10 (n)

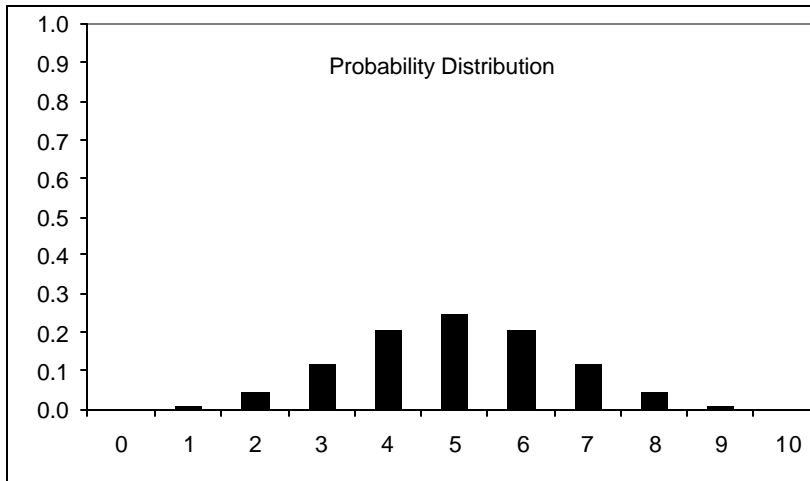
Number of Attempts (k)	Probability k successful attempts in n	Cumulative
0	0.000105	0.000105
1	0.001573	0.001678
2	0.010617	0.012295
3	0.042467	0.054762
4	0.111477	0.166239
5	0.200658	0.366897
6	0.250823	0.617719
7	0.214991	0.832710
8	0.120932	0.953643
9	0.040311	0.993953
10	0.006047	1.000000



14

© 2001 D. A. Menascé. All Rights Reserved.

Shape of the Binomial Distribution

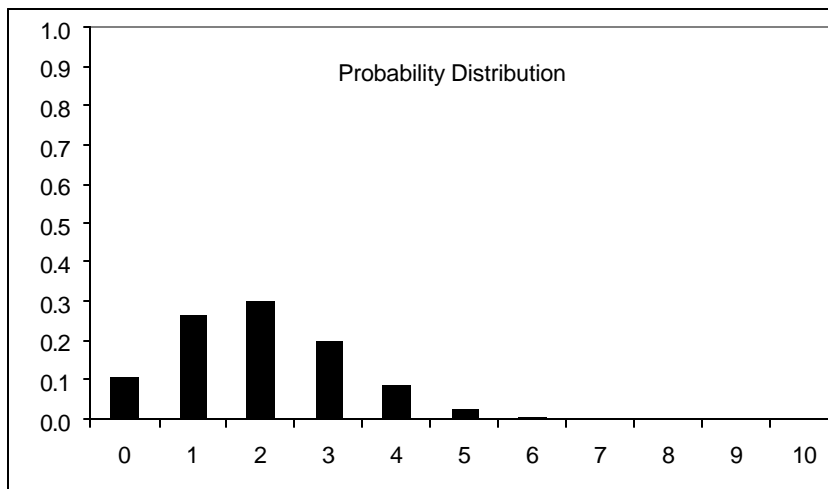


$p = 0.5$ symmetric for any n .

15

© 2001 D. A. Menascé. All Rights Reserved.

Shape of the Binomial Distribution

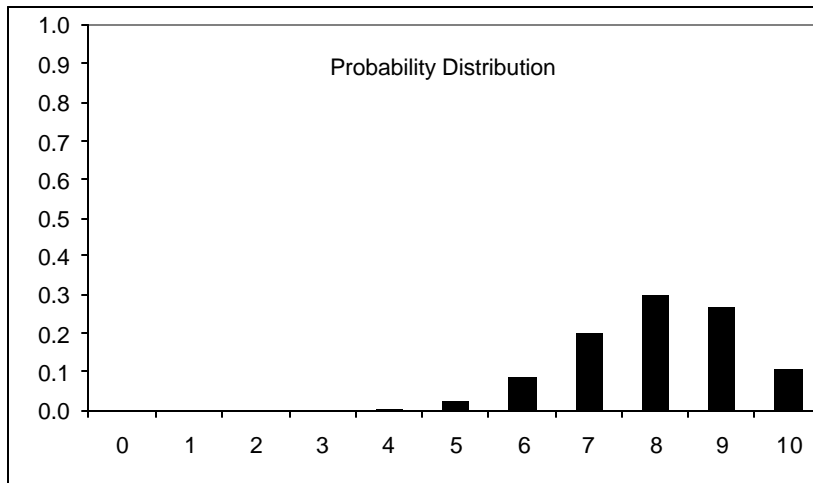


$p = 0.2$ right skewed

16

© 2001 D. A. Menascé. All Rights Reserved.

Shape of the Binomial Distribution



$p = 0.8$ left skewed

17

© 2001 D. A. Menascé. All Rights Reserved.

Moments of the Binomial Distribution

- Average: np
- Variance: $np(1-p)$
- Standard Deviation: $\sqrt{np(1-p)}$
- Coefficient of Variation:

$$\frac{\sqrt{np(1-p)}}{np} = \sqrt{\frac{1-p}{np}}$$

18

© 2001 D. A. Menascé. All Rights Reserved.

Hypergeometric Distribution

- Binomial was based on experiments with equal success probability.
- Hypergeometric: not all experiments have the same success probability.
- Given a sample size of n out of a population of size N with A known successes in the population, the probability of k successes is

choose k successes out of A successes in the population

choose $(n-k)$ failures from $N-A$ failures in the population

$$P[X = k] = \frac{\binom{A}{k} \binom{N-A}{n-k}}{\binom{N}{n}}$$

total # of possible samples

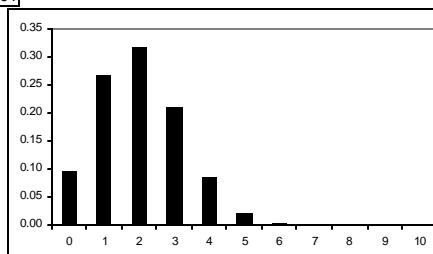
19

© 2001 D. A. Menascé. All Rights Reserved.

Hypergeometric Distribution

No. successes in sample k	sample size n	no. successes in population A	population size N	
0	20	10	100	0.09511627
1	20	10	100	0.26793316
2	20	10	100	0.31817063
3	20	10	100	0.20920809
4	20	10	100	0.08410730
5	20	10	100	0.02153147
6	20	10	100	0.00354136
7	20	10	100	0.00036793
8	20	10	100	0.00002300
9	20	10	100	0.00000078
10	20	10	100	0.00000001

In Excel:
 $\text{Pr}[X=k]=\text{HYPGEOMDIST}(k,n,A,N)$



20

© 2001 D. A. Menascé. All Rights Reserved.

Moments of the Hypergeometric

- Average: $\frac{nA}{N}$
- Standard Deviation: $\sqrt{\frac{nA(N-A)}{N^2}} \sqrt{\frac{N-n}{N-1}}$
- If the sample size is less than 5% of the population, the binomial is a good approximation for the hypergeometric.

21

© 2001 D. A. Menascé. All Rights Reserved.

Negative Binomial Distribution

- Probability of success is equal to p and is the same on all trials.
- Random variable X counts the number of trials until the k -th success is observed.

$$P[X = n] = \binom{n-1}{k-1} (1-p)^{n-k} p^k$$

$$\frac{S}{1} \quad \frac{F}{2} \quad \frac{F}{3} \quad \frac{S}{4} \quad \dots \quad \frac{F}{n-1} \quad \frac{S}{n}$$

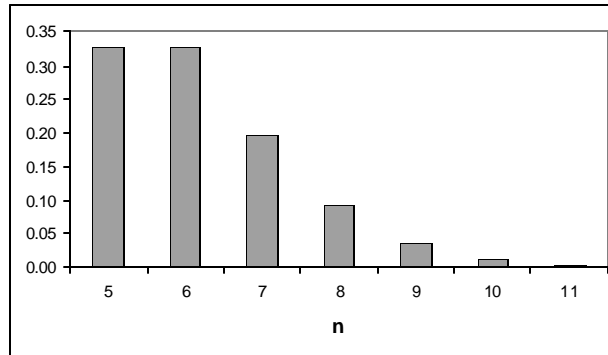
22

© 2001 D. A. Menascé. All Rights Reserved.

Negative Binomial Distribution

Success probability 0.8

k	n	Prob[X=n]
1	1	0.800000
1	2	0.160000
1	3	0.032000
1	4	0.006400
5	5	0.327680
5	6	0.327680
5	7	0.196608
5	8	0.091750
5	9	0.036700
5	10	0.013212
5	11	0.004404



In Excel:

Pr [X=n] = NEGBINOMDIST (n-k,k,p)

23

© 2001 D. A. Menascé. All Rights Reserved.

Moments of the Negative Binomial Distribution

- Average: $\frac{k}{p}$
- Standard Deviation: $\sqrt{\frac{k(1-p)}{p^2}}$
- Coefficient of Variation: $\sqrt{\frac{1-p}{k}}$

24

© 2001 D. A. Menascé. All Rights Reserved.

Geometric Distribution

- Special case of the negative binomial with $k=1$.
- Probability that the first success occurs after n trials is

$$p[X = n] = p(1 - p)^{n-1} \quad n = 1, 2, \dots$$

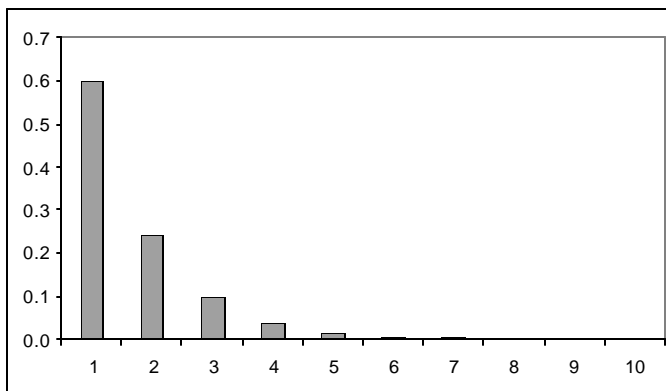
25

© 2001 D. A. Menascé. All Rights Reserved.

Geometric Distribution

Success probability

n	P[X=n]
1	0.6000
2	0.2400
3	0.0960
4	0.0384
5	0.0154
6	0.0061
7	0.0025
8	0.0010
9	0.0004
10	0.0002



26

© 2001 D. A. Menascé. All Rights Reserved.

Moments of the Geometric Distribution

- Average: $\frac{1}{p}$
- Standard Deviation: $\sqrt{\frac{1-p}{p^2}}$
- Coefficient of Variation: $\sqrt{1-p} \leq 1$

Poisson Distribution

- Used to model the number of arrivals over a given interval, e.g.,
 - Number of requests to a server
 - Number of failures of a component
 - Number of queries to the database.
- A Poisson distribution usually arises when arrivals come from a large number of independent sources.

Poisson Distribution

- Distribution: $P[X = k] = \frac{\lambda^k e^{-\lambda}}{k!} \quad k = 0, 1, \dots, \infty$

- Counting arrivals in an interval of duration t :

$$P[k \text{ arrivals in } [0, t)] = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \quad k = 0, 1, \dots, \infty$$

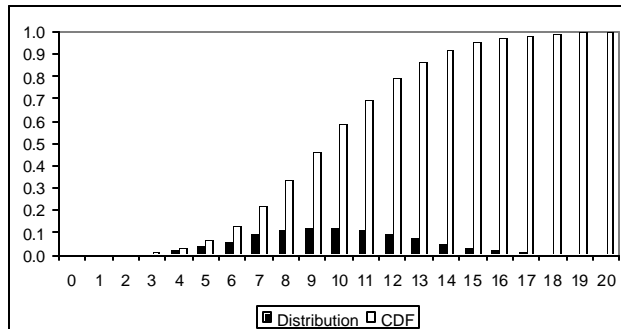
- Average=Variance= λ

29

© 2001 D. A. Menascé. All Rights Reserved.

Poisson Distribution

Lambda	10	
K	Poisson Distribution	CDF
0	0.00005	0.0000
1	0.00045	0.0005
2	0.00227	0.0028
3	0.00757	0.0103
4	0.01892	0.0293
5	0.03783	0.0671
6	0.06306	0.1301
7	0.09008	0.2202
8	0.11260	0.3328
9	0.12511	0.4579
10	0.12511	0.5830
11	0.11374	0.6968
12	0.09478	0.7916
13	0.07291	0.8645
14	0.05208	0.9165
15	0.03472	0.9513
16	0.02170	0.9730
17	0.01276	0.9857
18	0.00709	0.9928
19	0.00373	0.9965
20	0.00187	0.9984



In Excel:

$$P[X=k] = \text{POISSON}(k, \lambda, \text{FALSE})$$

$$P[X \leq k] = \text{POISSON}(k, \lambda, \text{TRUE})$$

30

© 2001 D. A. Menascé. All Rights Reserved.