University of Central Florida
School of Electrical Engineering and Computer Science
COT 4600 - Operating Systems.
Fall 2009 - dcm

Probability and Statistics Concepts

Random Variable: a rule that assigns a numerical value to each possible outcome of an experiment. All possible outcomes of the experiment constitute a sample space.

A random variable X on a sample space S is a function $X : S \mapsto \mathbb{R}$ which assigns a real number $X(s)$ to every sample point $s \in S$. This real number is called the probability of that outcome.

A discrete random variable maps events to values of a countable set e.g., the set of integers); each value in the range has a probability greater than or equal to zero.

Example 1. *the experiment is a coin toss; the outcome is either* 0 *(head) or* 1 *(tail).* If the coin is fair then $p_0 = p_1 = 0.5$; this means that in a large number of coin tosses we are likely to observe heads in about half of the cases and tails in the other half of the cases. Another example: when you throw throw a dice the outcome could be $1, 2, 3, 4, 5$, or 6; for a fair dice $p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = 1/6$.

A continuous random variable maps events to values of an uncountable set (e.g., the real numbers).

Example 2. *the experiment is to measure the speed of cars passing through an intersection:* the speed could be any value between 15 and 80 miles/hour. the probability of observing cars with a speed of 19.1 miles/hour could be zero but the probability of observing cars with a speed from 15 to 19.1 miles/hour could be $P_{19.1} = 0.3$ which means that 30% of the cars we observed have a speed in the range we considered.

A discrete random variable X has an associated probability density function, (also called probability mass function) $p_X(x)$ defined as:

$$p_X(x) = \text{Prob}(X = x)$$

and a probability distribution function also called cumulative distribution function, $P_X(x)$ defined as:

$$P_X(t) = \text{Prob}(X \leq t) = \sum_{x \leq t} p_X(x)$$

Example 3. *You have a binary random variable X (the outcome is either* 0 *or* 1*) and:*

$$p_0 = \text{Prob}(X = 0) = q \qquad \text{and} \qquad p_1 = \text{Prob}(X = 1) = p, \quad \text{with} \quad p + q = 1.$$

Bernouli trials: call the outcome of 1 a "success" and ask the question what is the probability Y_n that in n Bernoulli trials we have k successes:

$$p_k = \text{Prob}(Y_n = k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

The binomial cumulative distribution function is:

$$B(t:n,p) = \sum_{k=0}^{t} \binom{n}{k} p^k (1-p)^{n-k}$$

Example 4. *You have again Bernoulli trials and ask the question how many trails you need before the first "success".* If the first success occurs at the i-th trial then

$$p_Z(i) = q^{i-1}p$$

This is called a geometric distribution. It is easy to prove that:

$$\sum_{i=0}^{\infty} q^{i-1}p = \frac{p}{1-q} = 1.$$

A continuous random variable X has an associated probability density function, (also called probability mass function) $p_X(x)$ defined as:

$$f_X(x) = \text{Prob}(X = x)$$

and a probability distribution function also called cumulative distribution function, $F_X(x)$ defined as:

$$F_X(t) = \text{Prob}(X \leq t) = \int_{-\infty}^{t} f_X(x)dx$$

The expectation of random variable X: $E[X]$ is defined by

$$E[X] = \begin{cases} \sum_i x_i p_X(x_i) & \text{if X is discrete} \\ \\ \int_{-\infty}^{+\infty} x f(x)dx & \text{if X is continuous} \end{cases}$$

The variance Var[X] and standard deviation, σ of random variable X are defined by:

$$\text{Var}[X] = \sigma^2 = \begin{cases} \sum_i (x_i - E[X])^2 p_X(x_i) & \text{if X is discrete} \\ \\ \int_{-\infty}^{+\infty} (x - E[X])^2 f(x)dx & \text{if X is continuous} \end{cases}$$

The moment of order k of random variable X is defined as:

$$E[X^k] = \begin{cases} \sum_i x_i^k p_X(x_i) & \text{if X is discrete} \\ \\ \int_{-\infty}^{+\infty} x^k f(x)dx & \text{if X is continuous} \end{cases}$$

The centered moment of order k of random variable X is defined as the k-th moment of the random variable $x - E[X]$:

$$\mu_k = E\left[(X - E[X])^k\right] = \begin{cases} \sum_i (x_i - E[X])^k p_X(x_i) & \text{if X is discrete} \\ \\ \int_{-\infty}^{+\infty} (x - E[X])^k f(x)dx & \text{if X is continuous} \end{cases}$$

Examples of common distributions

1. Uniform distribution in the interval [a,b]: see Figure 1.

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if} \quad a \leq x \leq b \\ 0 & \text{if} \quad x < a \text{ or } x > b \end{cases}$$

$$F_X(x) = \begin{cases} 0 & \text{if} \quad x < a \\ \frac{x-a}{b-a} & \text{if} \quad a \leq x \leq b \\ 1 & \text{if} \quad x > b \end{cases}$$

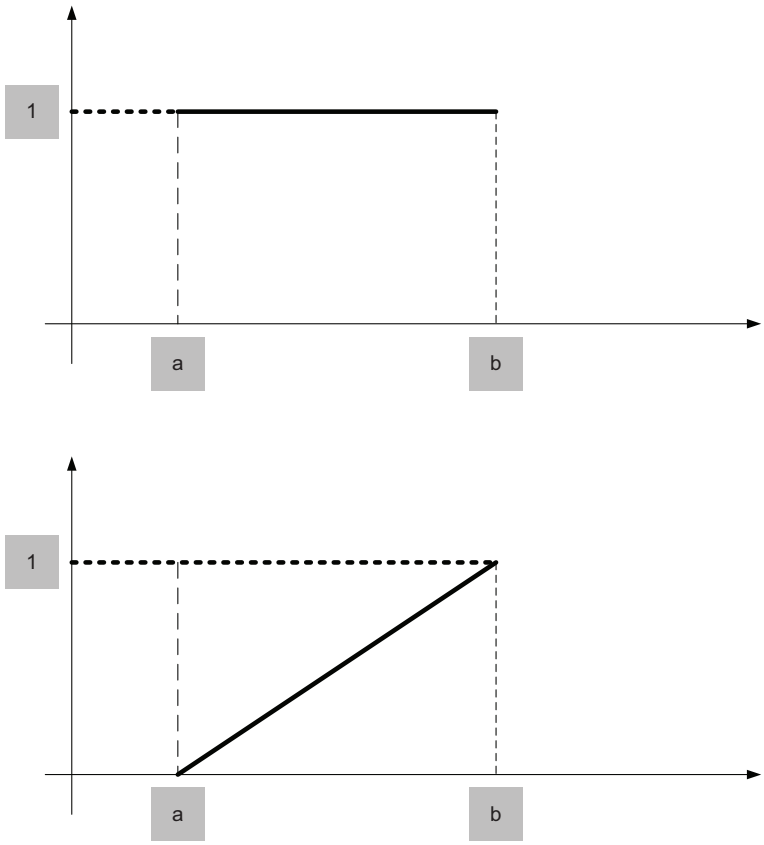


Figure 1: Probability density function (PDF) and cumulative distribution function of a uniform distribution

2. Standard normal distribution:

$$\phi(x) = \frac{1}{\sqrt{\pi}} e^{-x^2/2}$$

3. Normal distribution with mean μ and standard deviation σ:

$$f(x) = \frac{1}{\sigma 2\sqrt{\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

The probability density function (PDF) and the cumulative distribution function (CDF) of a normal distribution are displayed in Figures 2 and 3.

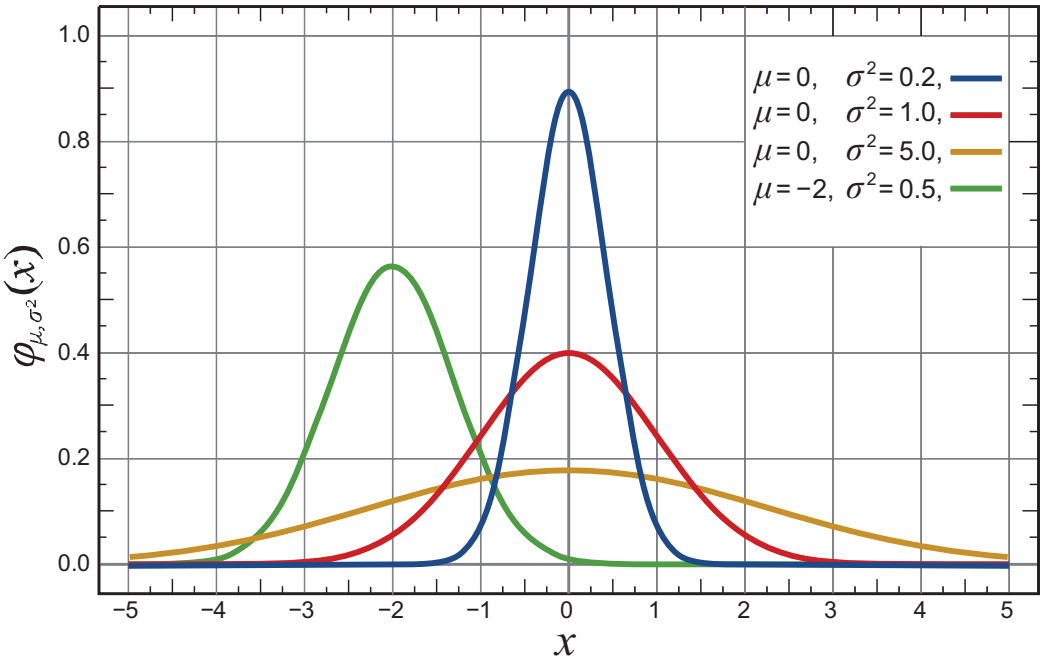


Figure 2: Probability density function (PDF) of a normal distribution

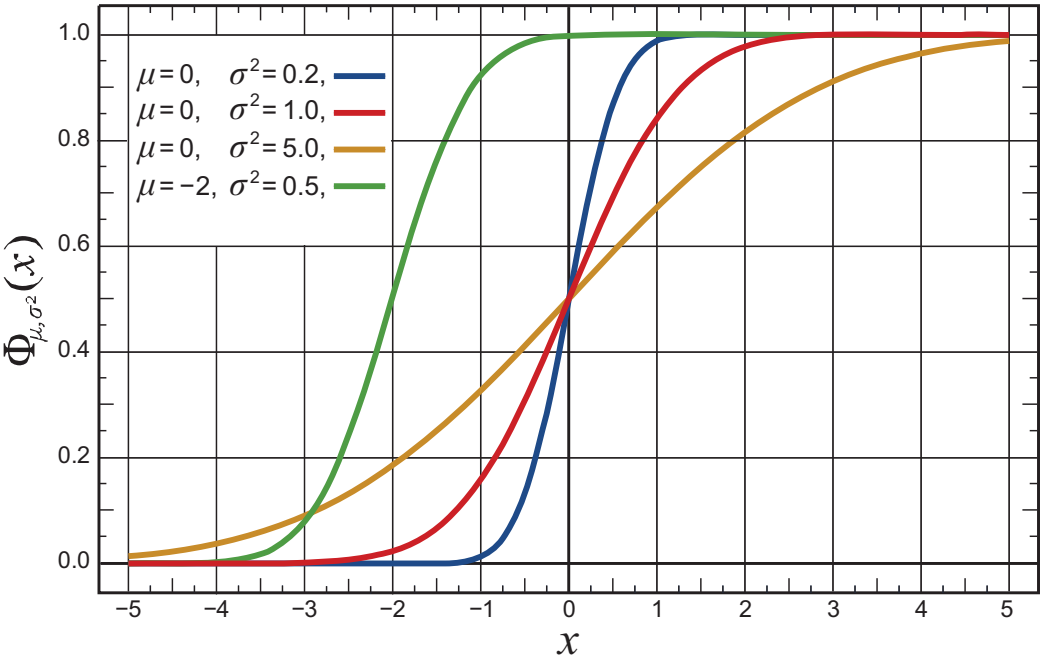


Figure 3: Cumulative distribution function (CDF) of a normal distribution

4. Exponential distribution with parameter λ. The probability density function (PDF) $f(x)$ and the cumulative distribution function (CDF), $F(x)$ of an exponential distribution are displayed in Figures 4 and 5.

$$f(x) = \begin{cases} \lambda x e^{-\lambda x} & \text{if} \quad x \geq 0 \\ 0 & \text{if} \quad X < 0 \end{cases}$$

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if} \quad x \geq 0 \\ 0 & \text{if} \quad X < 0 \end{cases}$$

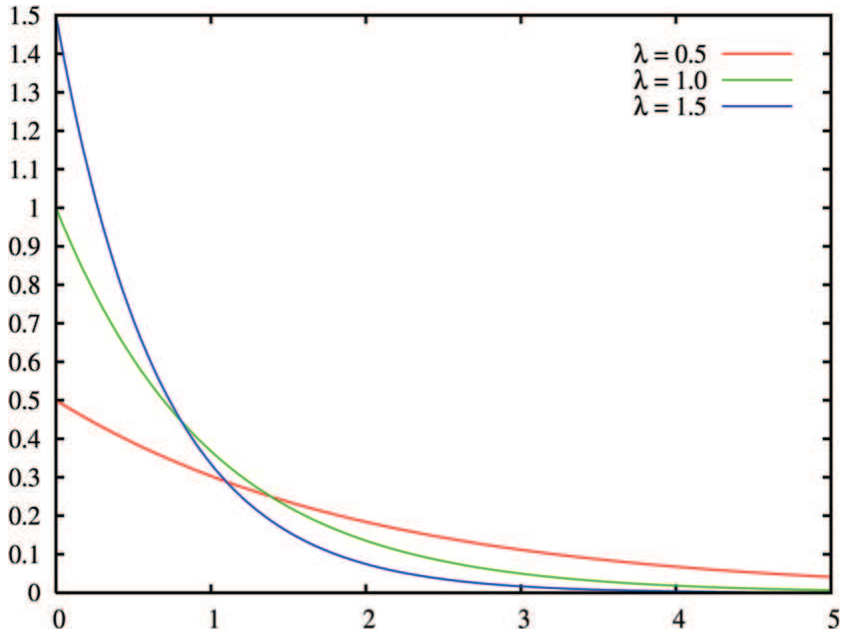


Figure 4: Probability density function (PDF) of an exponential distribution

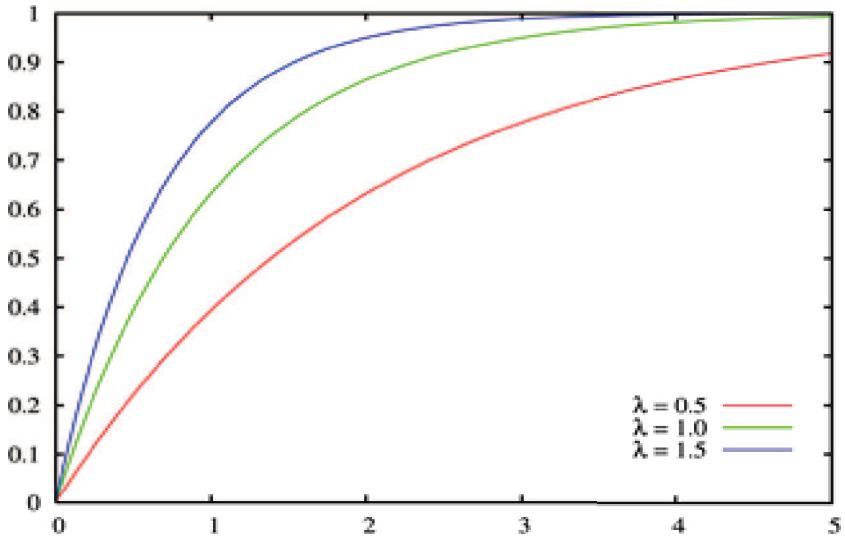


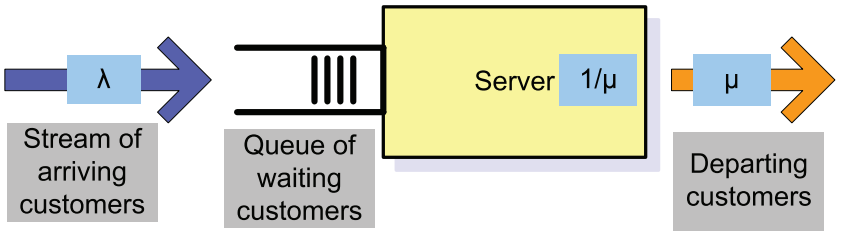
Figure 5: Cumulative distribution function (CDF) of an exponential distribution

Elements of Queuing Theory

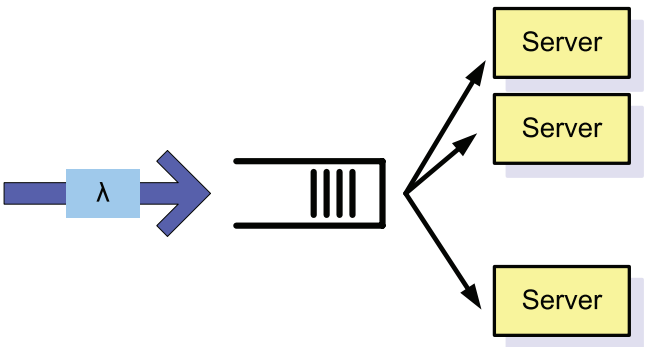
Queuing theory describes systems where you have a server and customers and both the arrival of the customers and the time it takes to service them are random variables with a certain distribution. There are two stochastic (aleatory) processes involved:

- The *arrival process* describes the pattern of the customers arrive. The arrival processes is characterized by its probability distribution function e.g., a uniform, exponential, hyper exponential. The distribution can described by one or more parameters such as the the average value of the random variables subject to that distribution, e.g., *the arrival rate* denoted as λ or the *the inter-arrival time* (the average time between two consecutive customer arrivals) $\frac{1}{\lambda}$. For example, if customers arrive in average at two minutes intervals the arrival rate is 1/2 customers/minute and the inter-arrival time is 2 minutes.

- The *service process* which describes how customers are served. The service processes is characterized by its probability distribution function e.g., a uniform, exponential, hyper exponential. The distribution can described by one or more parameters such as the the average value of the random variables subject to that distribution, e.g., *the service rate* denoted as μ or the *the service time* (the average time between two consecutive customer departures from the system) $\frac{1}{\mu}$. For example, if the service rate is $\mu = 10$ customers per hour, then the service time is $1/\mu = 60/10 = 5$ minutes.



(a) Single server system



(b) Multiple server system

Figure 6: The single- and multiple-server queuing systems

The pattern of customer behavior is illustrated in Figure 6:

1. The customer arrives in the system and joins a waiting queue. The time spent waiting is called *waiting time* and it is denoted as W.

2. When its turn arrives the customer enters service and the service time is $1/\mu$.

3. When the service is terminated the customer leaves the system. The time spent the customer is called *the time in system* $T = W + 1/\mu$.

   Other notations N is the total *number of customers in system*, one is in service and (N-1) are waiting in the queue.

A system is stable if the queue of waiting customers does not grow to infinity. Obviously, the service rate μ should be larger than the arrival rate λ, you should be able to process more customers per unit of time than they arrive, otherwise the queue of customers will grow in time. The number

$$\rho = \frac{\lambda}{\mu}$$

is called *server utilization* and $\rho \leq 1$ for the system to be stable (otherwise the queue of customers grows to infinity.

A queuing system is succinctly characterized as: $\mathcal{A}/\mathcal{S}/m$ with: $\mathcal{A}$ -the arrival process, $\mathcal{S}$ - the service process, and m the number of servers. The arrival and the service process are abbreviated as:

- M - exponential
- E - r-stage Erlangen
- D - deterministic
- G - general

Example 5. *M/M/*1*, M/M/m, M/G/1* are queueing systems when: (a) both the arrival and the service process are Markov (exponential) and we have only one server; (b) both the arrival and the service process are Markov (exponential) and we have m servers; (c) the arrival process is Markov (exponential) and the service process is general and we have only one server.

Little's Law. This law is general and realters the average number of customers in the system $\bar{N}$ with the average time spent by a customer in the system, $\bar{T}$, when the arrival rate is λ:

$$\bar{N} = \lambda \bar{T}$$

Rather than a rigorous proof, consider the following intuition. An "average customer" arrives at time t_a and leaves at time t_d. Being an "average customer" $t_d - t_a$ is equal to $\bar{T}$, the average time spent by a customer in the system; when our "average customer" leaves, he counts the number of customers in the system and finds it equal to the average, $\bar{N}$. But the customers arrive at a rate λ thus, during an interval of length $\bar{T}$ the total number is: $\bar{N} = \lambda \bar{T}$.

Poisson Process. The probability density function for a Poisson process is:

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

This equation gives the probability of seeing n arrivals in a period from 0 to t when λ is the arrival rate.

It follows that the probability of no arrivals (n=0) in a period from 0 to t is:

$$P_0(t) = e^{-\lambda t}$$

This equation shows that probability that no arrival takes place during an interval from 0 to t is negative exponentially related to the length of the interval.

Example 6. *Consider a highway with an average of 1 car arriving every 10 seconds $\lambda = 0.1$ cars/second arrival rate).*

The probability of not seeing a single car on the highway decreases dramatically with the observation period. If you observe the highway for a period of 1 second, there is 90% chance that no car will be seen during that period. If you monitor the highway for 20 seconds, there is only a 10% chance that you will not see a car on the highway. Put another way, there is only a 10% chance two cars arrive less than one second apart. There is a 90% chance that two cars arrive less than 20 seconds apart.

Birth and Death Process. We consider an $M/M/1$ system with arrival rate λ and service rate μ. We say that the system is in state k if there are k customers in the system. Such a system is modeled as a birth-death stochastic process with a potentially infinite number of states. The state transition diagram of a birth-death process is presented in Figure 7.

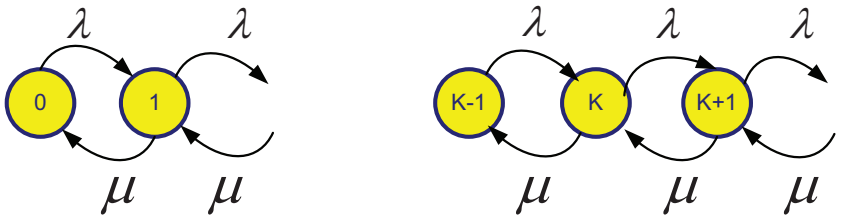


Figure 7: The states and transitions of a Birth-Death Stochastic Process. Note that state k can only be reached from state $k-1$ due to a birth or from state $k+1$ due to a death because of the property of the Poisson process which prohibits multiple birth or death occurring at the same time.

In steady-state the flow in and out of state k must be equal. The flow out of the state k corresponds to births and occurs at a rate λ thus, it is equal to λp_k; the flow into of the state k corresponds to deaths and occurs at a rate μ thus, it is equal to μp_{k+1}. We have the following relation between the probabilities p_k and p_{k+1} of states k and $k+1$, respectively:

$$\lambda p_k = \mu p_{k+1} \quad \Longrightarrow \quad p_{k+1} = \frac{\lambda}{\mu} p_k \quad \Longrightarrow \quad p_{k+1} = \rho p_k.$$

It follows that:

$$p_1 = \rho p_0, \qquad p_2 = \rho p_1 = \rho^2 p_0, \qquad \ldots, p_k = \rho p_{k-1} = \rho^k p_0, \ldots ...$$

But:

$$\sum_{i=0}^{\infty} p_i = 1 \quad \Longrightarrow \quad p_0 + \rho p_0 + \rho^2 p_0 + \ldots \rho^k p_0 + .. = 1$$

It follows that:

$$p_0 = \frac{1}{(1 + \rho + \rho^2 + \ldots \rho^k + \ldots)}$$

But $\rho \leq 1$ thus the sum of this geometric progression is equal to: $1/(1-\rho)$. We conclude that:

$$p_0 = (1-\rho), \qquad p_1 = (1-\rho)\rho, \qquad p_2 = (1-\rho)\rho^2 \qquad \ldots p_k = (1-\rho)\rho^k.$$

Knowing the probability of each state (recall that a state is characterized by the number of customers in the system) it is trivial to determine the average number in system:

$$\bar{N} = \sum_{k=0}^{\infty} k p_k = \sum_{k=0}^{\infty} k(1-\rho)\rho^k = (1-\rho)\sum_{k=0}^{\infty} k\rho^k$$

Recall again that $\rho \leq 1$; the sum is:

$$\sum_{k=0}^{\infty} k\rho^k = \frac{\rho}{1-\rho)^2}.$$

It follows that:

$$\bar{N} = \frac{\rho}{1-\rho)} = \frac{\lambda}{\mu - \lambda}.$$

If we apply Little's Law we find that:

$$\bar{T} = \frac{1}{\lambda}\bar{N} = \frac{1}{\mu - \lambda} = \frac{\frac{1}{\mu}}{1-\rho}$$

Figure 8 shows that $\bar{T} \to \infty$ when $\rho \to 1$.

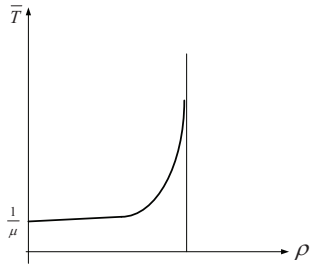


Figure 8: The average time in system $\bar{T}$ function of the utilization ρ.