Exponential Distributions and Expectations of Random Variables

ECE 313

Probability with Engineering Applications

Lecture 11

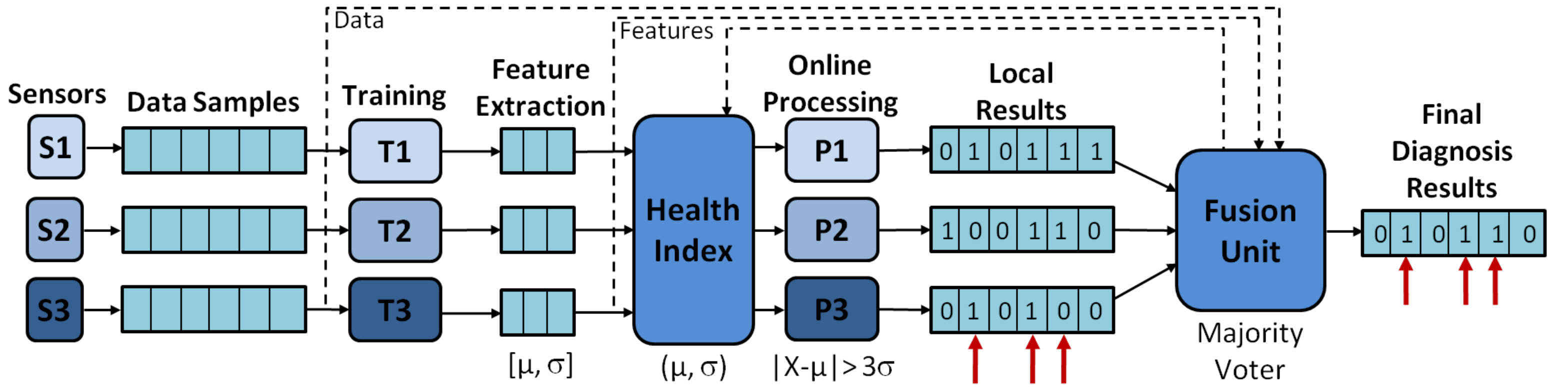
Professor Ravi K. Iyer

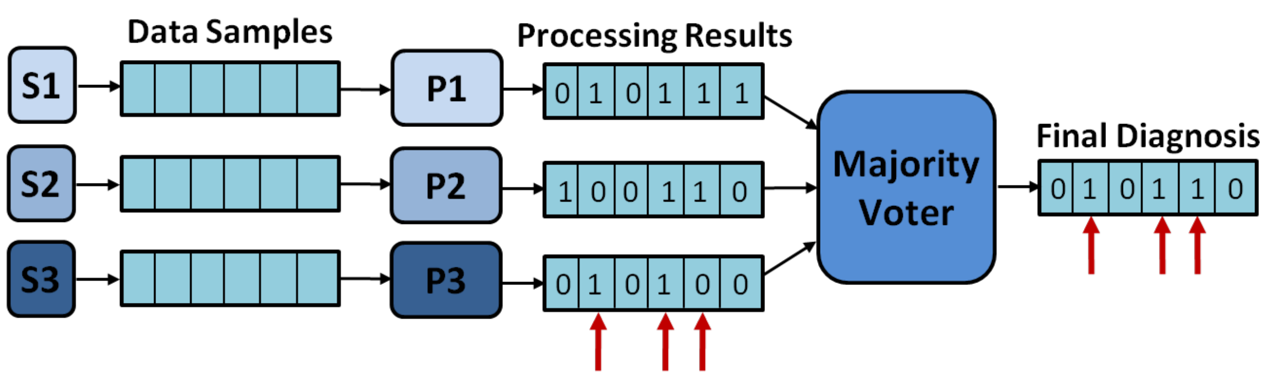
Dept. of Electrical and Computer Engineering

University of Illinois at Urbana Champaign

Mini Project 2: Monitoring Flow

- Multi-parameter Signal Analysis for Patient Monitoring



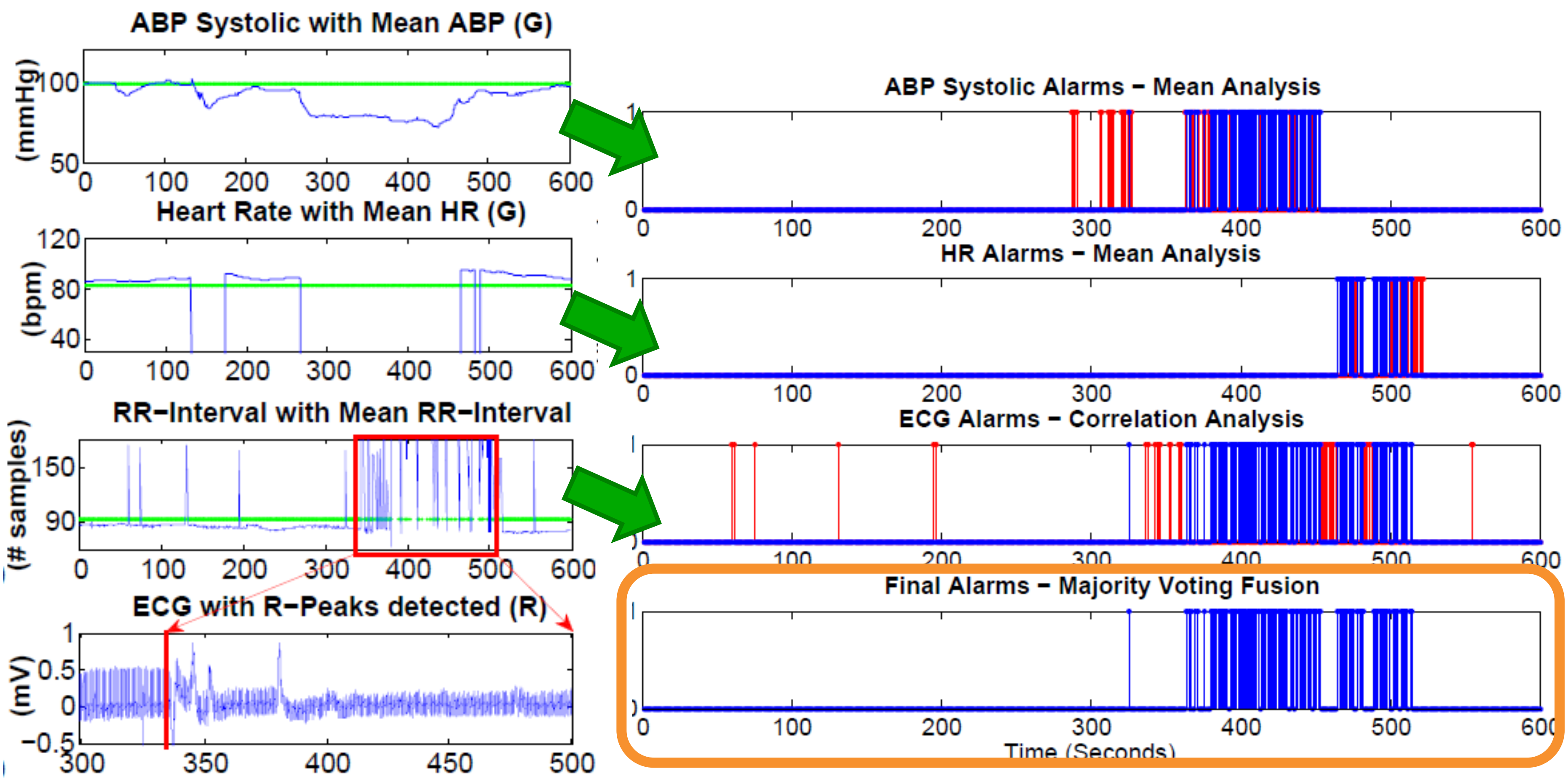


Mini Project 2: Summary

- The monitoring process starts with collecting multi-parameter inter-correlated physiological signals (such as Blood Pressure, Heart Rate, and Electrocardiogram) from biomedical sensors.
- Then it steps through an initial training phase, in which a physiological signature of the patient (Health Index), is compiled by aggregating (constructing a vector of) different statistical features (such as mean and standard deviation) from the input signals.
- During the monitoring phase, the obtained signature is used as a reference point (patient-specific threshold) for detecting abnormalities in each signal. At the end, a fusion technique (here a majority voter) is employed to reach at a final diagnostic decision.

H. Alemzadeh, C. D. Martino, Z. Jin, Z. Kalbarczyk, R. K. Iyer, "**Towards Resiliency in Embedded Medical Monitoring Devices,**" *Proceedings of DSN Workshop on Open Resilient Human-aware Cyber-physical Systems* ***(WORCS-2012)****, Boston, MA, July 2012.*

Mini Project 2: Majority Voter



Today's Topics

- Memoryless Property of Exponential Distribution
- Exponential Distribution Examples

- Uniform Distribution

- Expectation of Random Variables
  - Discrete case
  - Continuous case
  - Examples

Memoryless Property of the Exponential Distribution

- Assume that we know that X exceeds t (X > t). For example, X is the lifetime of a component.
- Assume we have observed the component to have been operating for ***t*** hours.
- We are then interested in the distribution of Y = X- t, which is the remaining (residual) lifetime of the component.
- The conditional probability of Y ≤ y can be denoted by $G_t(y)$.
- Thus for y ≥ 0:

$$G_t(y) = P(Y \le y \mid X > t)$$

$$= P(X - t \le y \mid X > t)$$

$$= P(X \le y + t \mid X > t)$$

$$= \frac{P(X \le y + t \text{ and } X > t)}{P(X > t)}$$

by the definition of conditional probability

$$= \frac{P(t < X \le y + t)}{P(X > t)}$$

Memoryless Property (cont.)

- Thus:

$$G_t(y) = \frac{\int_t^{y+t} f(x)dx}{\int_t^{\infty} f(x)dx}$$

$$= \frac{\int_t^{y+t} \lambda e^{-\lambda x} dx}{\int_t^{\infty} \lambda e^{-\lambda x} dx}$$

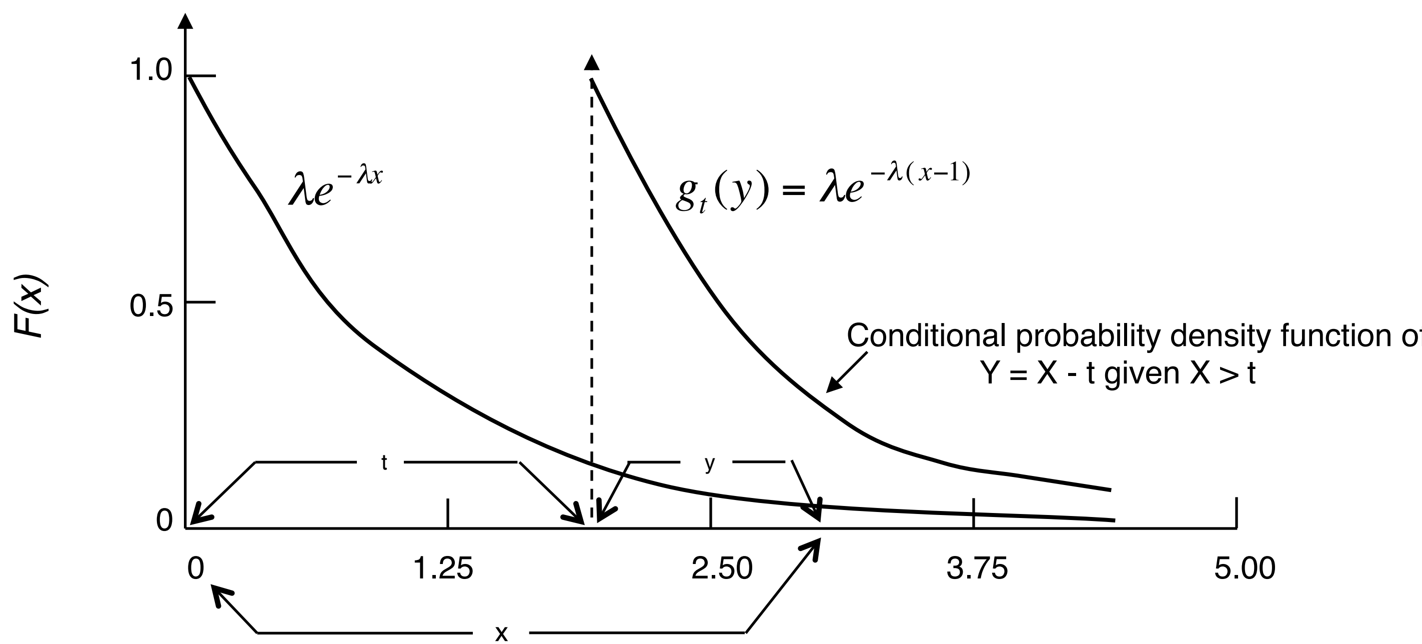
$$= \frac{e^{-\lambda t}(1 - e^{-\lambda y})}{e^{\lambda t}}$$

$$= 1 - e^{\lambda y}$$

Memoryless Property (cont.)

- Thus, $G_t(y)$ is independent of t and identical to the original exponential distribution of X.
- The distribution of the remaining life of the component does not depend on how long the component has been operating; the component does not "age." Its eventual breakdown results from a suddenly appearing failure, not from gradual deterioration.
- If inter-arrival times are exponentially distributed, the "memory-less" property (also known as the Markov property) says that "the wait time for a new arrival is statistically independent of how long we have already waited"!

Memoryless Property of the Exponential Distribution



Memoryless Property (cont.)

- If X is a nonnegative continuous random variable with the Markov property, then we can show that the distribution of X must be exponential:

$$\frac{P(t < X \leq y+t)}{P(X) > t} = P(X \leq y) = P(0 < X \leq y)$$

or

$$F_X(y+t) - F'_X(t) = [1 - F_X(t)][F_X(y) - F_X(0)]$$

- Since $F_x(0) = 0$, we rearrange the above equation to get:

$$\frac{F_X(y+t) - F_X(y)}{t} = \frac{F_X(t)[1 - F_X(y)]}{t}$$

Memoryless Property (cont.)

- Taking the limit as *t* approaches zero, we get:

$$F'_X(y) = F'_X(0)[1 - F_X(y)]$$

where F'_x denotes the derivative of F_x. Let $R_x(y) = 1 - F_x(y)$; then the above equation reduces to:

$$R'_X(y) = R'_X(0)R_X(y)$$

- The solution to this differential equation is given by:

$$R_X(y) = Ke^{R'_X(0)y}$$

where *K* is a constant of integration and $-R'_x(0) = F'_x(0) = f_x(0)$, the pdf evaluated at 0.

Memoryless Property (cont.)

- Noting that the $R_x(0) = 1$, and denoting $f_x(0)$ by the constant λ, we get:

$$R_x(y) = e^{-\lambda y}$$

  and hence

$$F_x(y) = 1 - e^{-\lambda y}, \qquad y > 0.$$

- Therefore X must have the exponential distribution.
- The exponential distribution can be obtained from the Poisson distribution by considering the interarrival times rather than the number of arrivals.

Uniform Distribution

The Uniform or Rectangular Distribution

- A continuous random variable *X* is said to have a uniform distribution over the interval (*a,b)* if its density is given by:

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b, \\ 0, & \text{otherwise}, \end{cases}$$

- And the distribution function is given by:

$$F(x) = \begin{cases} 0, & x < a, \\ \frac{x - a}{b - a}, & a \geq x < b, \\ 1, & x \geq b \end{cases}$$

Expectation of a Random Variable

Expectation of a Random Variable

- **The Discrete Case:** If *X* is a discrete random variable having a probability mass function *p(x)*, then the *expected value* of *X* is defined by

$$E[X] = \sum_{x:p(x)>0} xp(x)$$

The expected value of *X* is a weighted average of the possible values that *X* can take on, each value being weighted by the probability that *X* assumes that value. For example, if the probability mass function of *X* is given by

$$p(1) = \frac{1}{2} = p(2)$$

then

$$E[x] = 1\left(\frac{1}{2}\right) + 2\left(\frac{1}{2}\right) = \frac{3}{2}$$

is just an ordinary average of the two possible values 1 and 2 that *X* can assume.

Expectation of a Random Variable (Cont.)

Assume

$$p(1) = \frac{1}{3},\ p(2) = \frac{2}{3}$$

Then

$$E[X] = 1\left(\frac{1}{3}\right) + 2\left(\frac{2}{3}\right) = \frac{5}{3}$$

is a weighted average of the two possible values 1 and 2 where the value 2 is given twice as much weight as the value 1 since *p(*2*)* = 2*p(*1*)*.

- Find *E[X]* where *X* is the outcome when we roll a fair die.
- Solution: Since

$$p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}, we\ obtain$$

$$E[X] = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) = \frac{7}{2}$$

Expectation of a Random Variable (Cont.)

- **Expectation of a Bernoulli Random Variable:** Calculate $E[X]$ when X is a Bernoulli random variable with parameter p.

- Since: $p(0) = 1 - p, p(1) = p$
- We have: $E[X] = 0(1-p) + 1(p) = p$

Thus, the expected number of successes in a single trial is just the probability that the trial will be a success.

Expectation of a Random Variable (Cont.)

- **Expectation of a Binomial Random Variable:** Calculate $E[X]$ when X is a binomially distributed with parameters n and p.

$$E[X] = \sum_{i=0}^{n} ip(i)$$

$$= \sum_{i=0}^{n} i\binom{n}{i} p^i (1-p)^{n-i}$$

$$= \sum_{i=1}^{n} \frac{in!}{(n-i)!i!} p^i (1-p)^{n-i}$$

$$= \sum_{i=1}^{n} \frac{n!}{(n-i)!(i-1)!} p^i (1-p)^{n-i}$$

$$= np\sum_{i=1}^{n} \frac{(n-1)!}{(n-i)!(i-1)!} p^{i-1} (1-p)^{n-i}$$

$$= np\sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

$$= np[p + (1-p)]^{n-1}$$

$$= np$$

follows by letting $k = i - 1$. Thus, the expected number of successes in n independent trials is n multiplied by the probability that a trial results in a success.

Expectation of a Random Variable (Cont.)

- **Expectation of a Geometric Random Variable:** Calculate the expectation of a geometric random variable having parameter p.
- We have:

$$E[X] = \sum_{n=1}^{\infty} np(1-p)^{n-1}$$

$$= p\sum_{n=1}^{\infty} nq^{n-1}$$

$$where\, q = 1-p$$

$$E[X] = p\sum_{n=1}^{\infty} \frac{d}{dq}(q^n)$$

$$= p\frac{d}{dq}\left(\sum_{n=1}^{\infty} q^n\right)$$

$$= p\frac{d}{dq}\left(\frac{q}{1-q}\right)$$

$$= \frac{p}{(1-q)^2}$$

$$= \frac{1}{p}$$

The expected number of independent trials we need to perform until we get our first success equals the reciprocal of the probability that any one trial results in a success.

Expectation of a Random Variable (Cont.)

- **Expectation of a Poisson Random Variable**: Calculate $E[X]$ if X is a Poisson random variable with parameter λ.

$$E[X] = \sum_{i=0}^{\infty} \frac{ie^{-\lambda}\lambda^{i}}{i!}$$

$$= \sum_{i=1}^{\infty} \frac{e^{-\lambda}\lambda^{i}}{(i-1)!}$$

$$= \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!}$$

$$= \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k}}{k!}$$

$$= \lambda e^{-\lambda} e^{\lambda}$$

$$= \lambda$$

where we have used the identity:

$$\sum_{k=0}^{\infty} \lambda^{k} / k! = e^{\lambda}$$

The Continuous Case

- **The expected value of a continuous random variable:** If X is a continuous random variable having a density function $f(x)$, then the ***expected value of*** X is defined by:

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

- Example: **Expectation of a Uniform Random Variable,** Calculate the expectation of a random variable uniformly distributed over (α, β)

$$E[X] = \int_{\alpha}^{\beta} \frac{x}{\beta - \alpha} dx$$

$$= \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)}$$

$$= \frac{\beta + \alpha}{2}$$

The expected value of a random variable Uniformly distributed over the interval (α, β) is just the midpoint of the interval.

The Continuous Case (Cont.)

- ***Expectation of an Exponential Random Variable:*** Let X be exponentially distributed with parameter λ. Calculate $E[X]$.

$$E[X] = \int_0^\infty x\lambda e^{-\lambda x} dx$$

- Integrating by parts $(dv = \lambda e^{-\lambda x}, u = x)$ yields:

$$E[X] = -xe^{-\lambda x}\Big|_0^\infty + \int_0^\infty e^{-\lambda x} dx$$

$$= 0 - \frac{e^{-\lambda x}}{\lambda}\Big|_0^\infty$$

$$= \frac{1}{\lambda}$$

The Continuous Case Cont'd

- ***Expectation of a Normal Random Variable)***: X is normally distributed with parameters μ and σ^2:

$$E[X] = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} x e^{-(x-\mu)^2/2\sigma^2} dx$$

- Writing x as $(x-\mu) + \mu$ yields

$$E[X] = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} (x-\mu) e^{-(x-\mu)^2/2\sigma^2} dx + \mu \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2} dx$$

- Letting $y = x-\mu$ leads to

$$E[X] = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} y e^{-y^2/2\sigma^2} dy + \mu \int_{-\infty}^{\infty} f(x) dx$$

- Where $f(x)$ is the normal density. By symmetry, the first integral must be 0, and so

$$E[X] = \mu \int_{-\infty}^{\infty} f(x) dx = \mu$$

Example 1

- Consider the problem of searching for a specific name in a table of names. A simple method is to scan the table sequentially, starting from one end, until we either find the name or reach the other end, indicating that the required name is missing from the table. The following is a C program fragment for sequential search:

```
#define n 100

Example() {
 NAME Table[n+1];
 NAME myName;
 int I;

 \* myName has been initialized elsewhere *\
 Table[0] = myName; \\Table[0] is used as a sentinel or marker.
 I = n;

 while (myName != Table[I])
   I = I - 1;
 if (I > 0) {
   printf(``found!'');
   myName = Table[I];}
 else
   printf(``not found!'');
}
```

Example 1 Cont'd

- In order to analyze the time required for sequential search, let X be the discrete random variable denoting the number of comparisons *"myName≠Table[I]"* made. Clearly, the set of all possible values of X is $\{1,2, \ldots,n+1\}$, and $X=n+1$ for unsuccessful searches. Since the value of X is fixed for unsuccessful searches, it is more interesting to consider a random variable Y that denotes the number of comparisons for a successful search. The set of all possible values of Y is $\{1,2,\ldots,n\}$. To compute the average search time for a successful search, we must specify the pmf of Y. In the absence of any specific information, let us assume that Y is uniform over its range:

$$p_Y(i) = \frac{1}{n},\ 1 \le i \le n.$$

- Then $E[Y] = \sum_{i=1}^{n} i\, p_Y(i) = \frac{1}{n}\frac{n(n+1)}{2} = \frac{(n+1)}{2}.$
- Thus, on the average, approximately half the table needs to be searched

Example 2

- If α_i denotes the access probability for name *Table[i]*, then the average successful search time is *E[Y]* is minimized when names in the table are in the order of nonincreasing access probabilities; that is, $\alpha_1 \geq \alpha_2 \geq ... \geq \alpha_n$.

$$\alpha_1 = \frac{c}{i}, \; 1 \leq i \leq n,$$

- Where the constant c is determined from the normalization requirement $\sum_{i=1}^{n} \alpha_1 = 1$
- Thus, $c = \frac{1}{\sum_{i=1}^{n} \frac{1}{i}} = \frac{1}{H_n} \cong \frac{1}{\ln(n) + C},$
- Where H_n is the partial sum of a harmonic series; that is: $H_n = \sum_{i=1}^{n} (1/i)$ and $C(=0.577)$ is the Euler Constant.
- Now, if the names in the table are ordered as above, then the average search time is $E[Y] = \sum_{i=1}^{n} i\alpha_i = \frac{1}{H_n} \sum_{i=1}^{n} 1 = \frac{n}{H_n} \cong \frac{n}{\ln(n) + C}$
- Which is considerably less than the previous value $(n+1)/2$, for large n

Example 3

- Zipf's law has been used to model the distribution of Web page requests [BRES 1999]. It has been found that $p_Y(i)$ the probability of a request for the ith most popular page is inversely proportional to i [ALME1996, WILL 1996],

$$p_Y(i) = \frac{c}{i}, 1 \le i \le n,$$

- Where n is the total number of Web pages in the universe.

- We assume the Web page requests are independent and the cache can hold only m Web pages regardless of the size of each Web page. If we adopt a removal policy called "least frequently used", which always keeps the m most popular pages, then the hit ratio h(m)- the probability that a request can find its page in cache- is given by

$$h(m) = \sum_{i=1}^{m} p_Y(i) \cong cH_m = \frac{H_m}{H_n} \cong \frac{\ln(m) + C}{\ln(n) + C}$$

Example 4

- Let X be a continuous random variable with an exponential density given by

$$f(x) = \lambda e^{-\lambda x}, x > 0$$

- Then $E[X] = \int_{-\infty}^{\infty} xf(x)dx = \int_{0}^{\infty} \lambda x e^{-\lambda x} dx.$
- Let $u=\lambda x$, then $du= \lambda dx$, and $E[X] = \frac{1}{\lambda}\int_{0}^{\infty} ue^{-u}du = \frac{1}{\lambda}\Gamma 2 = \frac{1}{\lambda}$
- Thus, if a component obeys an exponential failure law with parameter λ (known as the failure rate), then its expected life, or its mean time to failure (MTTF), is 1/ λ. Finally, if the service time requirement of a job is an exponentially distributed random variable with parameter μ (known as the service rate), then the mean (average) service time is 1/μ.

Problem 1

- Using integration by parts, show (assuming that $E[X], \int_0^\infty [1 - F(x)]dx$ and $\int_{-0}^\infty F(x)dx$ are all finite) that for a continuous random variable X:

$$E[X] = \int_0^\infty [1 - F(x)]dx - \int_{-\infty}^0 F(x)dx.$$

- This result states that the expectation of a random variable X equals the difference of the areas of the right-hand and left-hand shaded regions in Figure 4.P.1.(This formula applies to the case of discrete and mixed random variables as well.)

Moments

- Let X be a random variable, and define another random variable Y as a function of X so that $Y = \phi(X)$. Suppose that we wish to compute $E[Y]$

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

(provided the sum or the integral on the right-hand side is absolutely convergent). A special case of interest is the power function $\phi(X) = X^k$ For k=1,2,3,…, $E[X^k]$ is known as the kth moment of the random variable X. Note that the first moment $E[X]$ is the ordinary expectation or the mean of X.

- We define the kth central moment, μ_k of the random variable X by

$$\mu_k = E[(X - E[X])^k]$$

- Known as the variance of X, Var[X], often denoted by σ^2
- Definition (Variance). The variance of a random variable X *is*

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

It is clear that Var[X] is always a nonnegative number.

Variance: 2nd Central Moment

- We define the kth central moment, μ_k of the random variable X by

$$\mu_k = E[(X - E[X])^k]$$

- μ^2 known as the variance of X, Var[X], often denoted by $\sigma^2 = E[(X - E[X])^2]$

- Definition (Variance). The variance of a random variable X *is*

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

- It is clear that Var[X] is always a nonnegative number.