

Continious Random Variables

Dr Ahmad Khonsari

ECE Dept.

The University of Tehran

Laplace transform of a sum

Let X and Y be independent random variables with L-transforms $F_X(s)$, $F_Y(s)$ (from now we use $f_X^*(s)$ and $f_Y^*(s)$).

$$f_{X+Y}^*(s) = E[e^{-s(X+Y)}]$$

$$= E[e^{-sX} e^{-sY}]$$

$$= E[e^{-sX}] E[e^{-sY}] \quad (\text{independence})$$

$$= f_X^*(s) f_Y^*(s)$$

thus

$$f_{X+Y}^*(s) = f_X^*(s) f_Y^*(s)$$

Perf Eval of Comp Systems

Calculating moments with the aid of Laplace transform

By derivation one sees

$$f^{*'}(s) = \frac{d}{ds} E[e^{-sX}] = E[-Xe^{-sX}]$$

Similarly, the n th derivative is

$$f^{*(n)}(s) = \frac{d^n}{ds^n} E[e^{-sX}] = E[(-X)^n e^{-sX}]$$

Evaluating these at $s = 0$ one gets

$$E[X] = -f^{*'}(0)$$

$$E[X^2] = +f^{*''}(0)$$

$$E[X^n] = (-1)^n f^{*(n)}(0)$$

Perf Eval of Comp Systems

Laplace transform of a random sum

Consider the random sum

$$Y = X_1 + \cdots + X_N$$

where the X_i are *i.i.d.* with the common L-transform $f_X^*(s)$ and $N \geq 0$ is a integer-valued r.v. with the generating function $\mathcal{G}_N(z)$.

$$\begin{aligned} f_Y^*(s) &= E[e^{-sY}] \\ &= E[E[e^{-sY} | N]] && \text{(outer expectation with respect to} \\ &\text{variations of } N) \\ &= E[E[e^{-s(X_1 + \cdots + X_N)} | N]] && \text{(in the inner expectation } N \text{ is fixed)} \\ &= E[E[e^{-s(X_1)}] \cdots E[e^{-s(X_N)}] | N] && \text{(independence)} \\ &= E[(f_X^*(s))^N] \\ &= \mathcal{G}_N(f_X^*(s)) && \text{(by the definition } E[z^N] = \mathcal{G}_N(z)) \end{aligned}$$

Perf Eval of Comp Systems

Uniform distribution $X \sim U(a, b)$ models the fact that any interval of length δ between a and b is equally likely.

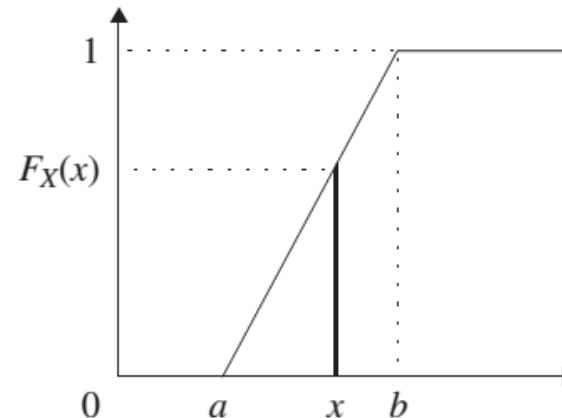
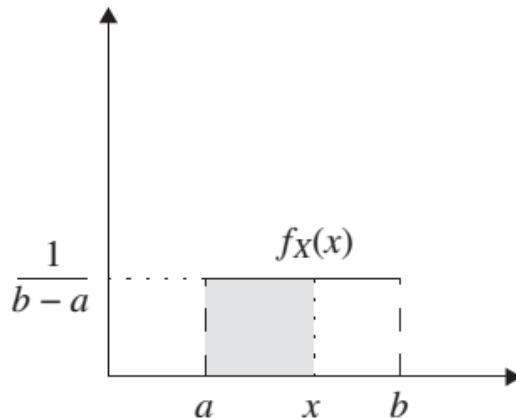
Definition: The pdf of X is constant in the interval (a, b) :

$$f_X(x) = \frac{1}{b-a} \quad a \leq x \leq b \quad F_X(x) = \frac{x-a}{b-a} \quad a \leq x < b, \quad F_X(x) = 1 \quad b \leq x$$

i.e. the value X is drawn randomly in the interval (a, b) (115)

Mean and variance are as follows:

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx = \frac{a+b}{2} \quad V[X] = \int_{-\infty}^{\infty} \left(x - \frac{a+b}{2}\right)^2 f(x) dx = \frac{1}{12} (b-a)^2 \quad (116)$$



Perf Eval of Comp Systems

Uniform distribution cntd.

Uniform distribution (continued)

Let U_1, \dots, U_n be independent uniformly distributed random variables, $U_i \sim U(0, 1)$.

- The number of variables which are $\leq x$ ($0 \leq x \leq 1$) is $\sim \text{Bin}(n, x)$
 - the event $\{U_i \leq x\}$ defines a Bernoulli trial where the probability of success is x

Uniform distribution (continued)

- Let $U_{(1)}, \dots, U_{(n)}$ be the ordered sequence of the values.

Define further $U_{(0)} = 0$ and $U_{(n+1)} = 1$.

It can be shown that all the intervals are identically distributed and

$$P\{U_{(i+1)} - U_{(i)} > x\} = (1 - x)^n \quad i = 1, \dots, n$$

– for the first interval $U_{(1)} - U_{(0)} = U_{(1)}$ the result is obvious because $U_{(1)} = \min(U_1, \dots, U_n)$

Perf Eval of Comp Systems

7.5. Exponential(λ) distribution

(Note that sometimes the shown parameter is $\frac{1}{\lambda}$, i.e. the mean of the distribution)

Definition: $X \sim \text{Exp}(\lambda)$, if:

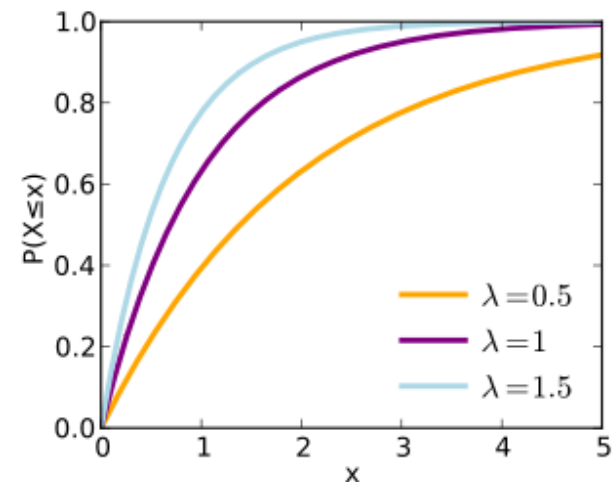
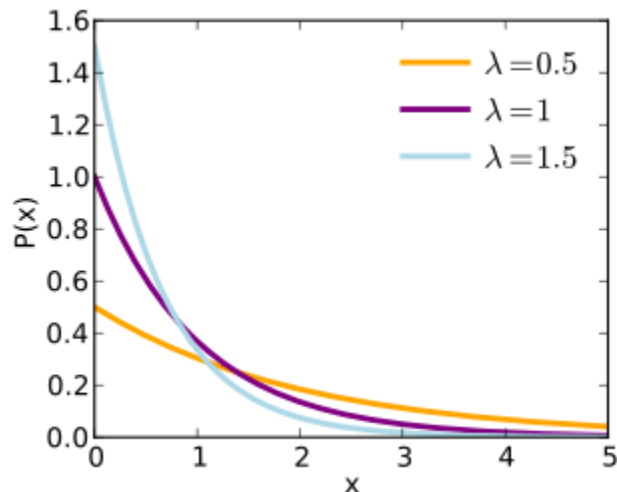
$$f_X(x) = \lambda e^{-\lambda x} \quad x \geq 0$$

$$F_X(x) = 1 - e^{-\lambda x}, \quad x \geq 0. \quad (111)$$

Mean and variance are as follows:

$$E[X] = \frac{1}{\lambda}$$

$$V[X] = \frac{1}{\lambda^2} \quad (112)$$



Perf Eval of Comp Systems

7.5. Exponential distribution cntd.

Laplace transform and moments of exponential distribution

The Laplace transform of a random variable with the distribution $\text{Exp}(\lambda)$ is

$$f^*(s) = \int_{-\infty}^{\infty} e^{-st} \lambda e^{-\lambda t} dt = \frac{\lambda}{\lambda + s}$$

With the aid of this one can calculate the moments:

$$E[X] = -f^{*'}(0) = \frac{\lambda}{(\lambda + s)^2} \Big|_{s=0} = \frac{1}{\lambda}$$

$$E[X^2] = +f^{*''}(0) = \frac{2\lambda}{(\lambda + s)^3} \Big|_{s=0} = \frac{2}{\lambda^2}$$

$$V[X] = E[X^2] - E[X]^2 = \frac{1}{\lambda^2}$$

$$E[X] = \frac{1}{\lambda} \qquad V[X] = \frac{1}{\lambda^2}$$

Perf Eval of Comp Systems

7.5. Exponential distribution cntd.

The memoryless property of exponential distribution

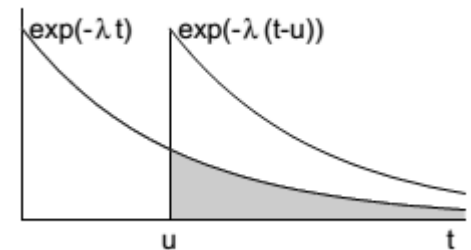
Assume that $X \sim \text{Exp}(\lambda)$ represents e.g. the duration of a call.

What is the probability that the call will last at least time x more given that it has already lasted the time t :

$$\begin{aligned} P\{X > t + x | X > t\} &= \frac{P\{X > t + x, X > t\}}{P\{X > t\}} \\ &= \frac{P\{X > t + x\}}{P\{X > t\}} \\ &= \frac{e^{-\lambda(t+x)}}{e^{-\lambda t}} = e^{-\lambda x} = P\{X > x\} \end{aligned}$$

$$P\{X > t + x, X > t\} = P\{X > x\}$$

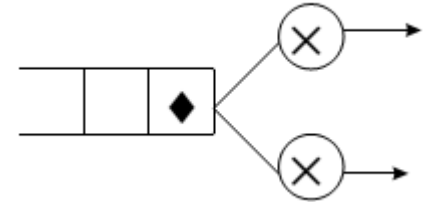
- The distribution of the remaining duration of the call does not at all depend on the time the call has already lasted has the same $\text{Exp}(\lambda)$ distribution as the total duration of the call.



Perf Eval of Comp Systems

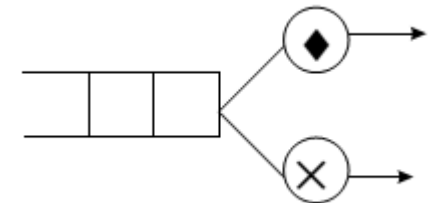
Example of the use of the memoryless property

A queueing system has two servers. The service times are assumed to be exponentially distributed (with the same parameter). Upon arrival of a customer (\diamond) both servers are occupied (\times) but there are no other waiting customers.



The question: what is the probability that the customer (\diamond) will be the last to depart from the system?

The next event in the system is that either of the customers (\times) being served departs and the customer enters (\diamond) the freed server.



By the memoryless property, from that point on the (remaining) service times of both customers (\diamond) and (\times) are identically (exponentially) distributed.

The situation is completely symmetric and consequently the probability that the customer (\diamond) is the last one to depart is $1/2$.

Perf Eval of Comp Systems

The ending probability of an exponentially distributed interval

Assume that a call with $\text{Exp}(\lambda)$ distributed duration has lasted the time t .

What is the probability that it will end in an infinitesimal interval of length h ?

$$P\{X \leq t + h \mid X > t\} = P\{X \leq h\} \quad (\text{memoryless})$$

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

$$= 1 - (1 - \lambda h + \frac{1}{2}(\lambda h)^2 - \dots)$$

$$= \lambda h + o(h)$$

The ending probability per time unit = λ (constant!)

Perf Eval of Comp Systems

The minimum and maximum of exponentially distributed random variables

Let $X_1 \sim \dots \sim X_n \sim \text{Exp}(\lambda)$ (*i.i.d.*)

The tail distribution of the minimum is

$$P\{\min(X_1, \dots, X_n) > x\} = P\{X_1 > x\} \cdot \dots \cdot P\{X_n > x\} \quad (\text{independence})$$

$$\left(e^{-\lambda x}\right)^n = e^{-n\lambda x}$$

The minimum obeys the distribution $\text{Exp}(n\lambda)$.

The cdf of the maximum is

$$P\{\max(X_1, \dots, X_n) \leq x\} = \left(1 - e^{-\lambda x}\right)^n$$

Perf Eval of Comp Systems

Calculating $E[\max(X_1, \dots, X_n)]$ Method 1:

$$E[N] = \sum_{k=0}^{\infty} P(N > k)$$

$$E[X] = \int_0^{\infty} (1 - F(x)) dx = \int_0^{\infty} R_X(t) dt$$

$$E[\max(X_1, \dots, X_n)] = \int_0^{\infty} R_X(t) dt = \int_0^{\infty} \left(1 - (1 - e^{-\lambda t})^n\right) dt =$$

$$\frac{1}{\lambda} \int_0^1 \frac{1 - u^n}{1 - u} du = \frac{1}{\lambda} \int_0^1 \left(\sum_{i=1}^n u^{i-1} \right) du =$$

$$\left(\frac{1}{\lambda} \sum_{i=1}^n \int_0^1 u^{i-1} du \right) = \frac{1}{\lambda} \sum_{i=1}^n \frac{u^i}{i} \Big|_0^1$$

$$= \frac{1}{\lambda} \sum_{i=1}^n \frac{1}{i} = \frac{1}{n\lambda} + \frac{1}{(n-1)\lambda} + \dots + \frac{1}{\lambda}$$

$$u = 1 - e^{-\lambda t} \quad \frac{(1-u)^n}{1-u} = \sum_{i=1}^n u^{i-1}$$

$$\frac{1 - e^{-\lambda t} \Big|_{t=0}^{\infty}}{1 - e^{-\lambda t} \Big|_{t=0}^{\infty}} \xrightarrow{1 - e^{-\lambda t} \Big|_{t=0}^{\infty} = 1} u \Big|_0^1$$

$$\frac{du = \lambda e^{-\lambda t} dt = \lambda(1-u) dt}{dt \Big|_0^{\infty}} \xrightarrow{du = \lambda(1-u) dt} \frac{du}{\lambda(1-u)} \Big|_0^1$$

Perf Eval of Comp Systems

Calculating $E[\max(X_1, \dots, X_n)]$ Method 2:

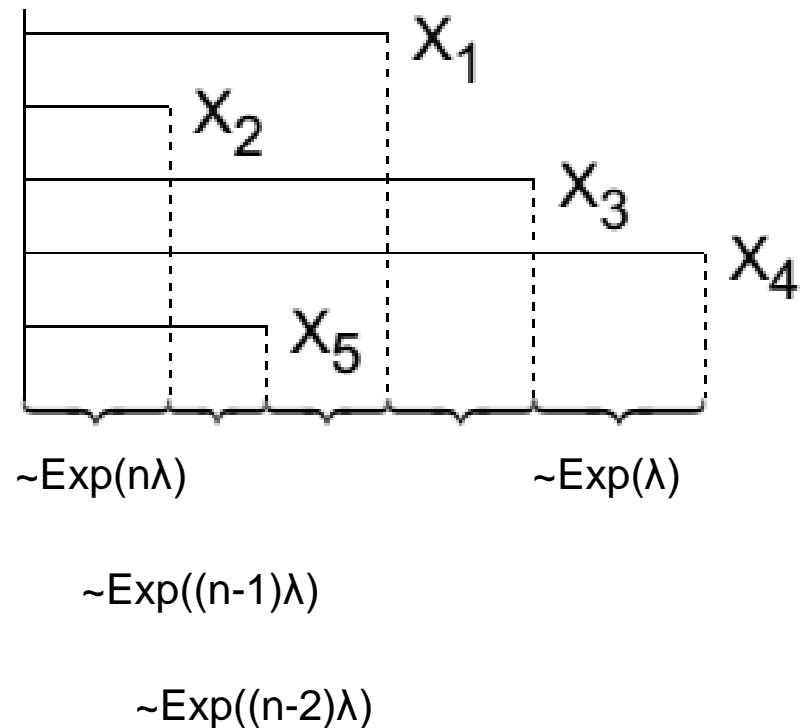
The ending intensity of the minimum = $n\lambda$

n parallel processes each of which ends with intensity λ independent of the others

The expectation can be deduced by inspecting the figure

$$E[\max(X_1, \dots, X_n)]$$

$$= \frac{1}{n\lambda} + \frac{1}{(n-1)\lambda} + \dots + \frac{1}{\lambda}$$



Performability : background

Q1: find distribution of $Y=aX+b$

$$A: f_Y(y) = \begin{cases} \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right) & \text{if } x \in Y=aX+b \\ 0 & \text{otherwise} \end{cases}$$

Q2: if $X \sim \text{Exp}(\lambda)$, what is distribution of $Y=rX$

$$A: f_X(x) = \lambda e^{-\lambda x}$$

Using Q1

$$f_Y(y) = \frac{1}{r} \lambda e^{-\lambda \frac{1}{r} y} \quad 06\text{-}Q2$$

Thus Y is an exponential with parameter λ/r

Performability : **Example**

Consider a system with n processors. In the beginning we let all n processors be active, performing different computations,

Thus the total computing capacity is n (where a unit of computing capacity corresponds to that of one active processor).

Performability : Example

Let X_1, X_2, \dots, X_n be the times to failure of the n processors.

After a period of time $Y_1 = \min \{X_1, X_2, \dots, X_n\}$ only $n - 1$ processors will be active and the computing capacity of the system will have dropped to $n - 1$.

The cumulative computing capacity that the system supplies until all processors have failed is then given by the random variable:

$$C_n = nY_1 + (n-1)(Y_2 - Y_1) + \dots + (n-j)(Y_{j+1} - Y_j) + \dots + (Y_n - Y_{n-1}).$$

Performability : Example

we note that C_n is the area under the curve.

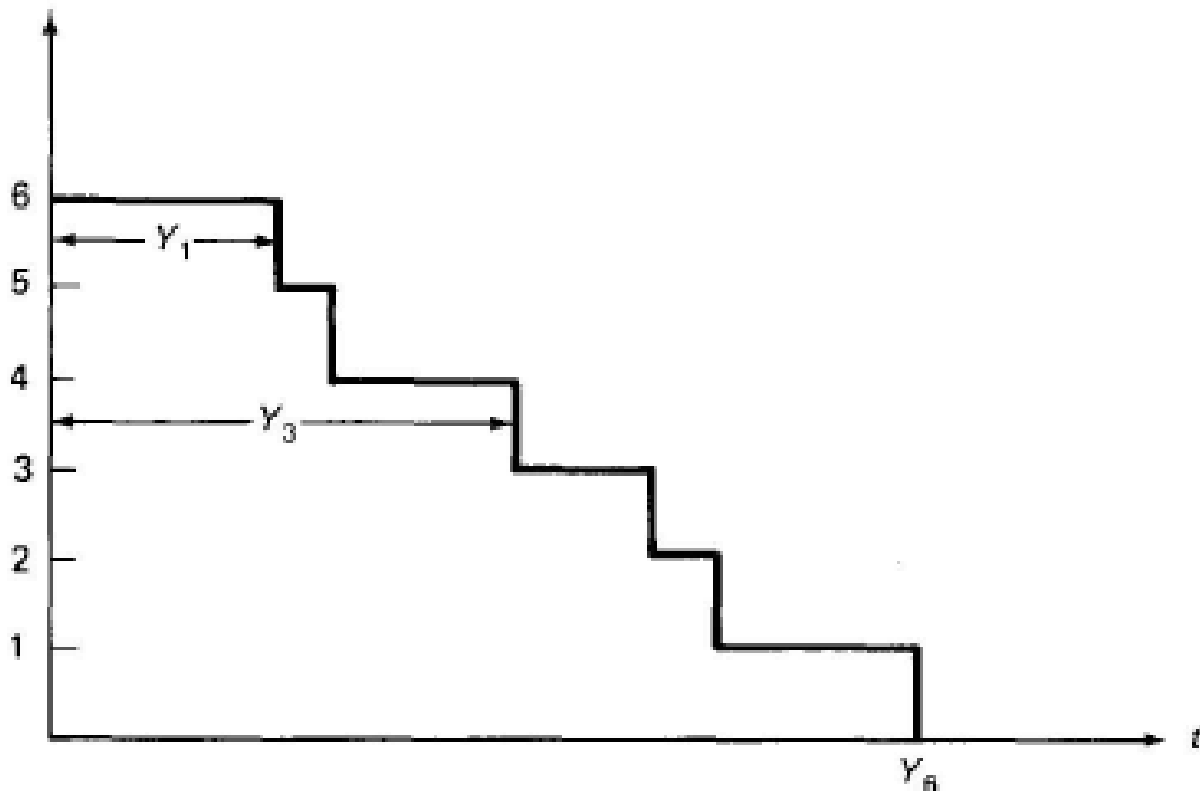


Fig: Computing capacity as a function of time

Performability : Example

C_n "computation before failure" Beaudry [BEAU 1978]

"performability" Meyer [MEYE 1980]

Distribution of C_n

first we obtain the distribution of $Y_{j+1} - Y_j$.
assume processor lifetimes are mutually
independent EXP (λ) RVs. we claim that the
distribution of

$Y_{j+1} - Y_j$ is $\text{Exp}((n - j)\lambda)$

Performability : Example

Define $Y_0 = 0$.

we know that $Y_1 = \min \{X_1, X_2, \dots, X_n\} \sim \text{Exp}(n\lambda)$,

Thus our claim holds for $j = 0$.

After j processors have failed, $(n-j)$ processors remain.

the residual lifetimes of the remaining $(n - j)$ processors, denoted by W_1, W_2, \dots, W_{n-j} are each exponentially distributed with parameter λ due to the memoryless property of the exponential distribution.

Performability : Example (cntd.)

Note that $Y_{j+1} - Y_j$ is simply the time between the $(j + 1)$ th and the j th failure; that is,

$$Y_{j+1} - Y_j = \min \{W_1, W_2, \dots, W_{n-j}\}$$

It follows that $Y_{j+1} - Y_j$ is $\sim \text{Exp}((n - j)\lambda)$.

Performability : Example (cntd.)

Hence, using 06-Q2 (i.e. if $X \sim \text{Exp}(\lambda)$, dist. of $Y=rX$ is

$$f_Y(y) = \frac{1}{r} \lambda e^{-\lambda \frac{1}{r}} \text{ we get:}$$

$X \sim \text{Exp}((n-j)\lambda)$, dist. of $Y=(n-j)X$ is ($r=n-j$)

$$f_Y(y) = \frac{1}{r} (n-j) \lambda e^{-(n-j) \lambda \frac{1}{r}} \Big|_{r=n-j}$$

Thus $(n-j) (Y_{j+1} - Y_j)$ is $\sim \text{Exp}(\lambda)$.

Therefore, C_n is the sum of n independent identically distributed exponential RVs,

That is C_n is **n -stage Erlang distributed with parameter λ** .

Perf Eval of Comp Systems

Erlang distribution $X \sim \text{Erlang}(n, \lambda)$ Also denoted Erlang- $n(\lambda)$.

X is the sum of n independent random variables with the distribution $\text{Exp}(\lambda)$

$$X = X_1 + \dots + X_n \quad X_i \sim \text{Exp}(\lambda) \text{ (i.i.d.)}$$

The Laplace transform is

$$f^*(s) = \left(\frac{\lambda}{\lambda + s} \right)^n$$

By inverse transform (or by recursively convoluting the density function) one obtains the pdf of the sum X

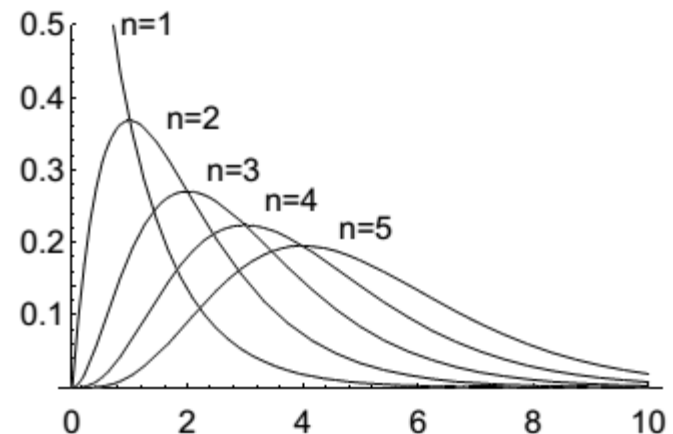
$$f(x) = \frac{(\lambda x)^{n-1}}{(n-1)!} \lambda e^{-\lambda x} \quad x \geq 0$$

The expectation and variance of Erlang are n times those of the $\text{Exp}(\lambda)$ distribution:

$$E[X] = \frac{n}{\lambda} \quad V[X] = \frac{n}{\lambda^2}$$

Note:

Erlang(1, λ) (Erlang-1(λ) = $E_{1, \lambda}$) is $\text{exp}(\lambda)$.



Erlang distribution (continued): gamma distribution

The formula for the pdf of the Erlang distribution can be generalized, from the integer parameter n , to arbitrary real numbers by replacing the factorial $(n - 1)!$ by the gamma function $\Gamma(n)$:

$$f(x) = \frac{(\lambda x)^{n-1}}{\Gamma(n)} \lambda e^{-\lambda x} \quad \text{Gamma}(n, \lambda) \text{ distribution}$$

Gamma function $\Gamma(p)$ is defined by

$$\Gamma(p) = \int_0^{\infty} e^{-u} u^{p-1} du$$

By partial integration it is easy to see that when p is an integer (i.e. $p=n$) then, indeed, $\Gamma(n) = (n - 1)!$

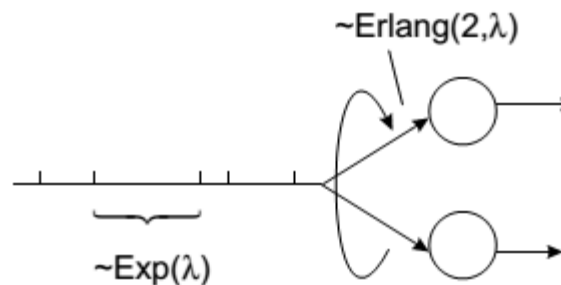
Perf Eval of Comp Systems

Erlang distribution (continued)

Example. The system consists of two servers. Customers arrive with $\text{Exp}(\lambda)$ distributed interarrival times.

Customers are alternately sent to servers 1 and 2.

The interarrival time distribution of customers arriving at a given server is $\text{Erlang}(2, \lambda)$.



Erlang distribution (continued)

Proposition. Let N_t , the number of events in an interval of length t , obey the Poisson distribution:

$$N_t \sim \text{Poisson}(\lambda t)$$

Then the time T_n from an arbitrary event to the n^{th} event thereafter obeys the distribution $\text{Erlang}(n, \lambda)$.

Proof.

$$F_{T_n}(t) = P\{T_n \leq t\} = P\{N_t \geq n\}$$

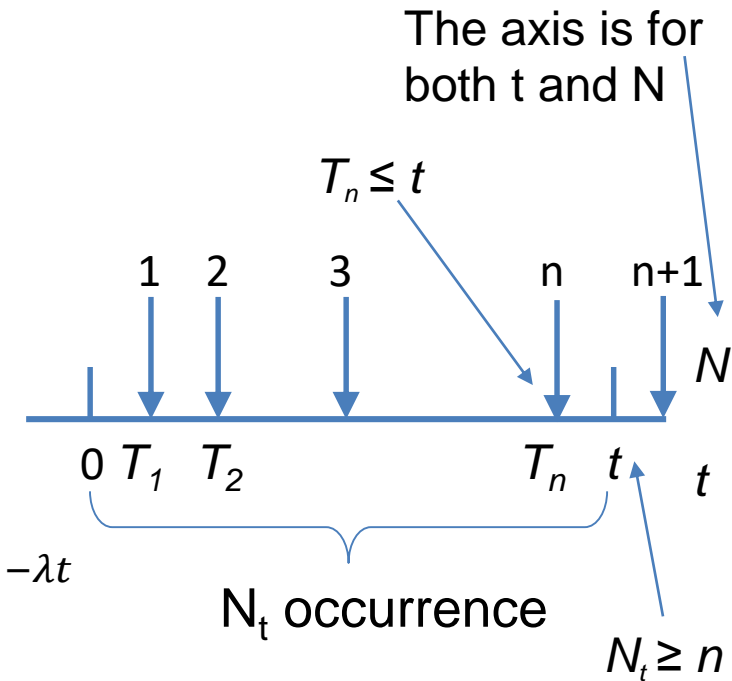
$$\sum_{i=n}^{\infty} P\{N_t = i\} = \sum_{i=n}^{\infty} \frac{(\lambda t)^i}{i!} e^{-\lambda t}$$

$$f_{T_n} = \frac{d}{dt} F_{T_n}(t) = \sum_{i=n}^{\infty} \frac{i \lambda (\lambda t)^{i-1}}{i!} e^{-\lambda t} - \sum_{i=n}^{\infty} \frac{(\lambda t)^i}{i!} \lambda e^{-\lambda t}$$

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

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

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



Perf Eval of Comp Systems

Normal distribution $X \sim N(\mu, \sigma^2)$

The pdf of a normally distributed random variable X with parameters μ and σ^2 is

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

Parameters μ and σ^2 are expectation and variance are of the distribution $E[X] = \mu$, $V[X] = \sigma^2$

Normal Standard $Z \sim N(0, 1)$

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

0 and 1 are expectation and variance are of the distribution $E[Z] = 0$, $V[Z] = 1$

Perf Eval of Comp Systems

Proposition: If $X \sim N(\mu, \sigma^2)$, then $Y = \alpha X + \beta \sim N(\alpha\mu + \beta, \alpha^2\sigma^2)$.

Proof:

$$F_Y(y) = P\{Y \leq y\} = P\left\{X \leq \frac{y-\beta}{\alpha}\right\} = F_X\left(\frac{y-\beta}{\alpha}\right)$$
$$= \int_{-\infty}^{\frac{y-\beta}{\alpha}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \quad z = \alpha x + \beta$$

Specifically if we set ($\alpha = 1/\sigma$, $\beta = -\mu/\sigma$)

Then we find $z = \frac{x-\mu}{\sigma} \sim N(0, 1)$

$$F_Y(y) = \int_{-\infty}^{\frac{y-\beta}{\alpha}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$
$$= \int_{-\infty}^y \frac{1}{\sqrt{2\pi}(\alpha\sigma)} e^{-\frac{1}{2}(z - (\alpha\mu + \beta))^2 / (\alpha\sigma)^2} dz$$

$$x = \frac{z - \beta}{\alpha} = \frac{z + \mu/\sigma}{1/\sigma} = \sigma z + \mu$$

$$x \Big|_{-\infty}^{\frac{y-\beta}{\alpha}} \rightarrow z \Big|_{-\infty}^{\alpha \frac{y-\beta}{\alpha} + \beta} \rightarrow z \Big|_{-\infty}^y$$

$$dx \rightarrow \frac{dz}{\alpha}$$

$$(x - \mu)^2 = \left(\frac{z - \beta}{\alpha} - \mu\right)^2 = \frac{(z - \beta - \alpha\mu)^2}{\alpha^2}$$

Perf Eval of Comp Systems

Proposition: If $X \sim N(\mu, \sigma^2)$, then $Y = \alpha X + \beta \sim N(\alpha\mu + \beta, \alpha^2\sigma^2)$.

Proof cntd.:

$$\frac{(z - (\alpha\mu + \beta))^2}{(\alpha\sigma)^2} = \frac{(z - ((1/\sigma)\mu + (-\mu/\sigma)))^2}{\left(\left(\frac{1}{\sigma}\right)\sigma\right)^2} = z^2$$

$$\begin{aligned} F_Y(y) &= \int_{-\infty}^y \frac{1}{\sqrt{2\pi}(\alpha\sigma)} e^{-\frac{1}{2}(z - (\alpha\mu + \beta))^2 / (\alpha\sigma)^2} dz \\ &= \int_{-\infty}^y \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \end{aligned}$$

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

Denote the PDF of a $N(0,1)$ random variable by $\Phi(x)$. Then

$$F_X(x) = P\{X \leq x\} = P\{\sigma Z + \mu \leq x\} = P\{Z \leq \frac{x - \mu}{\sigma}\} = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

Perf Eval of Comp Systems

Multivariate Gaussian (normal) distribution

Let X_1, \dots, X_n be a set of Gaussian (i.e. normally distributed) random variables with expectations μ_1, \dots, μ_n and covariance matrix

$$\mathbf{\Gamma} = \begin{bmatrix} \sigma_{11}^2 & \dots & \sigma_{1n}^2 \\ \dots & \dots & \dots \\ \sigma_{n1}^2 & \dots & \sigma_{nn}^2 \end{bmatrix} \quad \sigma_{ij}^2 = \text{Cov}[X_i, X_j] \quad \sigma_{ii}^2 = \text{V}[X_i]$$

Denote $\mathbf{X} = (X_1, \dots, X_n)^T$.

The probability density function of the random vector \mathbf{X} is

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{\Gamma}|}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{\Gamma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

where $|\mathbf{\Gamma}|$ is the determinant of the covariance matrix.

By a change of variables one sees easily that the pdf of the random vector $\mathbf{Z} = \mathbf{\Gamma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$ is $(2\pi)^{-n/2} \exp(-\frac{1}{2} \mathbf{z}^T \mathbf{z}) = \sqrt{2\pi} e^{-z_1^2/2} \dots \sqrt{2\pi} e^{-z_n^2/2}$

Thus the components of the vector \mathbf{Z} are independent $N(0,1)$ distributed random variables.

Conversely, $\mathbf{X} = \boldsymbol{\mu} + \mathbf{\Gamma}^{1/2} \mathbf{Z}$ by means of which one can generate values for \mathbf{X} in simulations.

Appendix

Power Laws

- Two quantities y and x are related by a power law if

$$y \approx x^{-\alpha}$$

- Compare it to $y \approx e^{-x}$ that decays fast.
- A (continuous) random variable X follows a **power-law** distribution if it has density function

$$f(x) = Cx^{-\alpha}$$

- Integrating above we find : Cumulative function

$$P[X \geq x] = \frac{C}{\alpha - 1} x^{-(\alpha - 1)}$$

Normalization constant

- Assuming a minimum value x_{\min}

$$C = (a - 1)x_{\min}^{a-1}$$

- The density function becomes

$$f(x) = \frac{(a-1)}{x_{\min}} \left(\frac{x}{x_{\min}} \right)^{-a}$$

Pareto distribution

- Pareto distribution is pretty much the same but we have

$$P[X \geq x] = C' x^{-\beta}$$

- and we usually we require

$$x \geq x_{\min}$$

Zipf's Law

- A random variable X follows Zipf's law if the r -th largest value x_r satisfies

$$x_r \approx r^{-\gamma}$$

- Same as requiring a Pareto distribution

$$P[X \geq x] \approx x^{-1/\gamma}$$

Zipf & Pareto:
what they have to do with power-laws

- Zipf
 - George Kingsley Zipf, a Harvard linguistics professor, sought to determine the 'size' of the 3rd or 8th or 100th most common word.
 - Size here denotes the frequency of use of the word in English text, and not the length of the word itself.
 - Zipf's law states that the size of the r 'th largest occurrence of the event is inversely proportional to its rank:

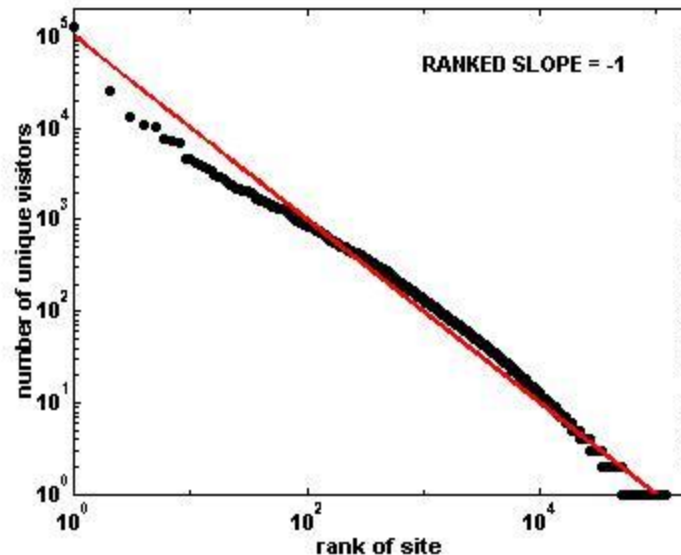
$$y \sim r^{-\beta}, \text{ with } \beta \text{ close to unity.}$$

So how do we go from Zipf to Pareto?

- The phrase "The r th largest city has n inhabitants" is equivalent to saying " r cities have n or more inhabitants".
- This is exactly the definition of the Pareto distribution, except the x and y axes are flipped. Whereas for Zipf, r is on the x-axis and n is on the y-axis, for Pareto, r is on the y-axis and n is on the x-axis.
- Simply inverting the axes, we get that if the rank exponent is β , i.e. $n \sim r^{-\beta}$ for Zipf, (n = income, r = rank of person with income n) then the Pareto exponent is $1/\beta$ so that $r \sim n^{-1/\beta}$ (n = income, r = number of people whose income is n or higher)

Zipf's law & AOL site visits

- Deviation from Zipf's law
 - slightly too few websites with large numbers of visitors:



Exponents and averages

- In general, power law distributions do not have an average value if $\alpha < 2$ (but the sample will!)
- This is because the average is given by (for integer values of k)

$$\sum_{k=k_{\min}}^{\infty} kp(k) = \sum_{k=k_{\min}}^{\infty} kk^{-\alpha} = \sum_{k=k_{\min}}^{\infty} \frac{1}{k^{\alpha-1}}$$

for a finite sample this
will only go to the largest
observed value

- n The harmonic series diverges...

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots$$

- n Same holds for continuous values of k

80/20 rule

- The fraction W of the wealth in the hands of the richest P of the the population is given by

$$W = p^{(\alpha-2)/(\alpha-1)}$$

- Example: US wealth: $\alpha = 2.1$
 - richest 20% of the population holds 86% of the wealth

Generative processes for power-laws

- Many different processes can lead to power laws
- There is no one unique mechanism that explains it all
- Next class: Yule's process and preferential attachment

What does it mean to be scale free?

- A power law looks the same no matter what scale we look at it on (2 to 50 or 200 to 5000)
- Only true of a power-law distribution!
- $p(bx) = g(b) p(x)$ – shape of the distribution is unchanged except for a multiplicative constant
- $p(bx) = (bx)^{-\alpha} = b^{-\alpha} x^{-\alpha}$

