Exponential random variables

- Exponential random variables arise in studies of waiting times, service times, etc.
- $X$ is called an exponential random variable with parameter $\lambda$ if its pdf is given by:
  \[ f(u) = \lambda \cdot \exp(-\lambda u) \quad \text{for} \quad u \geq 0 \]
  \[ f(u) = 0 \quad \text{for} \quad u < 0 \]
- Scale parameter $\lambda > 0$
- $E[X] = 1/\lambda$  
  $\text{var}(X) = 1/\lambda^2$

CDF of exponential RV

- $X$ is an exponential RV with parameter $\lambda$.
- $F(t) = P\{X \leq t\} = \text{area under pdf to left of } t = 1 - \exp(-\lambda t)$
- $P\{X > t\} = \exp(-\lambda t)$
- $E[X] = \int P\{X > t\} \, dt = 1/\lambda$
- $P\{X > t + \tau \mid X > t\} = P\{X > t + \tau\}/P\{X > t\} = \exp(-\lambda \tau)$
- Memoryless property of exponential RVs

What's this $\Gamma(t)$ stuff, anyway?

- The value of $\Gamma(t)$ is given by
  \[ \Gamma(t) = \int_0^\infty x^{t-1} \cdot \exp(-x) \, dx, \quad t > 0 \]
  \[ \Gamma(t) = (t-1) \cdot \Gamma(t-1) = (t-1) \cdot (t-2) \cdot \Gamma(t-2) = \ldots \]
- If $t$ is an integer, $\Gamma(t) = (t-1)!$
- If $t$ is not an integer, $\Gamma(t) = (t-1) \cdot (t-2) \cdot \ldots \cdot \Gamma(t-[t])$ where $t-[t]$ is the fractional part of $t$
- For $0 < t < 1$, numerical integration must be used to evaluate $\Gamma(t)$

Gamma random variables

- $X$ is called a gamma random variable with parameters $(t, \lambda)$ if its pdf is given by:
  \[ f(u) = \lambda \cdot \exp(-\lambda u) \cdot (\lambda u)^{t-1} / \Gamma(t) \quad \text{for} \quad u > 0 \]
  \[ f(u) = 0 \quad \text{for} \quad u \leq 0 \]
- $t > 0$ is the order parameter
- $\lambda > 0$ is the scale parameter
- $\Gamma(t)$ is a number whose value is the gamma function evaluated at $t$

Gamma pdfs: different values of $t$

- $\lambda = 1$
- $t = 1$
- $t = 2$
- $t = 3$
- $t = 4$
- $t = 5$

Gamma pdfs: different values of $\lambda$

- $t = 3$
- $\lambda = 1$
- $\lambda = 2$
- $\lambda = 3$
### Mean & variance of gamma RVs
- If $X$ is a gamma RV with parameters $(t, \lambda)$, then $\mathbb{E}[X] = t/\lambda$ and $\text{var}(X) = t/\lambda^2$.
- A gamma RV with order parameter $t = 1$ is an exponential RV with parameter $\lambda$.
- A gamma RV with order parameter $t = n$ is called an $n$-Erlang random variable.
- A gamma RV with $t = n/2, \lambda = 1/2$ is a chi-square RV with $n$ degrees of freedom.

### An analogy
- The exponential RV with parameter $\lambda$ is analogous to the geometric RV with parameter $p$ — both denote the waiting time till something occurs.
- $\mathbb{E}[X] = 1/\lambda \leftrightarrow \mathbb{E}[X] = 1/p$.
- The gamma RV with parameters $(n, \lambda)$ is analogous to the negative binomial RV with parameters $(n, p)$ — both denote the waiting time for the $n$-th occurrence.
- $\mathbb{E}[X] = n/\lambda \leftrightarrow \mathbb{E}[X] = n/p$.

### More on the analogy
- With geometric and negative binomial RVs, we are, in effect, measuring the waiting time till something occurs in discrete steps.
- With exponential and gamma RVs, we are modeling time as a continuous variable.
- With negative binomial and gamma RV, the waiting time for the $n$-th occurrence of something is the sum of $n$ (independent) waiting times for one thing to occur.

### What thing occurs?
- We are going to study occurrences of various interesting phenomena.
  - Telephone “off-hook” signals
  - Jobs arriving at a processor
  - Packets arriving at a router
  - $\alpha$-particle emissions
  - Gate failures in a TMR system
  - Cars passing a checkpoint.

### What’s common to all these?
- The occurrences of the phenomenon, e.g., a telephone going off-hook, a packet arriving at a router are not very frequent.
- Actually, this infrequency depends on the time scale being used.
- For a telephone, the chances that it will go off hook in the next millisecond are small.
- $P(\text{off-hook at some time during next day})$ is close to $1$.

### Occurrences at random
- The time interval between successive occurrences (called the inter-arrival time) varies at random.
- The inter-arrival time is modeled as a random variable.
- What is the probabilistic behavior of this random variable?
- Answer: geometric for discrete time, exponential for continuous time.
Independence of occurrences

- The observed value of a random variable depends on the outcome of an experiment.
- An experiment commences at time $t = 0$ and ends with the first arrival (i.e., occurrence) at some random time (say $t = X_1$) later.
- At this time, a new experiment is started, and it ends with the second arrival at some time $t = X_2$, and so on...
- These experiments are independent.

An illustrative diagram

- 1st experiment starts at $t = 0$.
- 2nd experiment starts at $t = X_1$.
- 3rd experiment starts at $t = X_2$.
- Inter-arrival time.

Long term average = expectation

- The inter-arrival time is a random variable.
- The successive inter-arrival times are the observed values of this random variable on independent trials.
- The long-term observed average $= (\text{total time till N-th arrival occurs})/N$ is roughly equal to the expected value of the inter-arrival time.

The arrival rate

- The inter-arrival time is a random variable with expected value (average) $= 1/\mu$.
- For large $N$, the total of $N$ inter-arrival times is roughly $N(1/\mu)$.
- The number of occurrences of the phenomenon of interest (arrivals) in a long time interval of duration $T$ is roughly $\mu T$.
- $\mu$ is called the arrival rate.
- On average, $\mu T$ arrivals in $T$ seconds.

Random arrivals in time

- We are considering some phenomenon that occurs at random times.
- Sometimes, these occurrences are called events.
- Here, event does not have its usual meaning in probability theory, viz. a subset of the sample space.
- To avoid confusion, we shall call these occurrences as arrivals or points.

Poisson process with arrival rate $\mu$

- The sequence of random arrivals or points (with independent inter-arrival times having average value $1/\mu$) is called a Poisson process with arrival rate $\mu$.
- Why Poisson?
- Let $N(t_1, t_2]$ denote the number of arrivals in the time interval $[t_1, t_2]$. Note endpoints.
- $N(t_1, t_2]$ is a Poisson random variable with parameter $\mu(t_2 - t_1) = \mu \cdot \text{duration of interval}$.
Basic assumptions

- At any time instant $t$, at most one arrival can occur.
- Assumption: For small values of $\Delta T$
  - $P(N(t, t+\Delta t) = 1) = \mu \cdot \Delta T$
  - $P(N(t, t+\Delta t) = 0) = 1 - \mu \cdot \Delta T$
- Approximation improves as $\Delta T \to 0$
- Approximation is nonsense if $\Delta T > 1/\mu$
- Small values of $\Delta T$ means $\Delta T \ll 1/\mu$

Is this consistent with Poissonity?

- We have not yet proved that $N(t_1, t_2]$ is a Poisson RV with parameter $\mu \cdot (t_2 - t_1)$
- But, if $N(t, t+\Delta t]$ were a Poisson random variable with parameter $\mu \cdot \Delta T << 1$, then
  - $P(N(t, t+\Delta t] = 0) = \exp(-\mu \cdot \Delta T) = 1 - \mu \cdot \Delta T + (\mu \cdot \Delta T)^2/2 - ...$
  - $P(N(t, t+\Delta t] = 1) = (\mu \cdot \Delta T) \approx \mu \cdot \Delta T$

The independence assumption

- Let $t_1 < t_2 < t_3 < t_4 < t_5 < t_6 ...$
- Independence assumption:
  - The random variables $N(t_1, t_2], N(t_3, t_4], N(t_5, t_6], ...$ are independent
  - The numbers of arrivals in disjoint intervals (i.e., non-overlapping intervals) of time are independent
  - Memoryless property of the process

Distribution of the first arrival time

- The first arrival occurs at random time $X_1$
- We set up and solve a differential equation for $P_0(t) = P(\text{no arrivals in } (0, t])$
  - $P_0(t) = P(\text{no arrivals in } (0, t]) = P(X_1 > t)$
  - $P_0(t+\Delta T) = P(\text{no arrivals in } (0, t+\Delta T])$
    - $P_0(0) = P(N(0, t+\Delta T) = 0)$
    - $P_0(0) = P(N(0, t) = 0) \cap P(N(t, t+\Delta T) = 0)$
    - $P_0(0) = P[N(0, t) = 0] \cdot P[N(t, t+\Delta T) = 0]$

The differential equation for $P_0(t)$

- $P_0(t + \Delta T) = P(\text{no arrivals in } (0, t+\Delta T])$
  - $P_0(t + \Delta T) = P(N(0, t+\Delta T) = 0)$
    - $P_0(t+\Delta T) = P[N(0, t+\Delta T) = 0]$ $\cap$ $P[N(t, t+\Delta T)]$
    - Independence of disjoint intervals
    - $P_0(t+\Delta T) = P_0(t) \cdot (1 - \mu \cdot \Delta T)$
  - $P_0(t+\Delta T) - P_0(t) = -\mu \cdot P_0(t)$

Solving the diff. eq. for $P_0(t)$

- $[P_0(t+\Delta T) - P_0(t)]/\Delta T = -\mu \cdot P_0(t)$
  - $dP_0(t)/dt = -\mu \cdot P_0(t)$
  - This is a first-order linear differential equation with constant coefficients
  - Initial condition:
    - $P_0(0) = P(\text{no arrivals in } (0,0])$
    - $P_0(0) = P(\text{no arrivals in } \emptyset) = 1$ (not zero!)
    - $P_0(t) = \exp(-\mu \cdot t)$ for $t \geq 0$
Thoughts about $P_0(t)$

- $P_0(t) = P\{\text{no arrivals in } (0, t]\}
  = P\{N(0, t] = 0\} = P\{X_1 > t\}
  = \exp(-\mu t)$
- $P_0(t)$ decays away exponentially as $t \to \infty$
- $P_0(\Delta t) = P\{\text{no arrivals in } (0, \Delta t]\} = \exp(-\mu \Delta t)$
  = $1 - \mu \Delta t + (\mu \Delta t)^2/2! - \ldots$
  $\approx 1 - \mu \Delta t$ for small values of $\Delta t$
- $P\{X_1 > t\} = \exp(-\mu t)$ for $t \geq 0$

Distribution of the first arrival time

- The first arrival occurs at random time $X_1$
- $P\{X_1 > t\} = \exp(-\mu t)$ for $t \geq 0$
- $P\{X_1 \leq t\} = F_{X_1}(t) = 1 - \exp(-\mu t)$ for $t \geq 0$
- $f_{X_1}(t) = \text{derivative of CDF } F_{X_1}(t)$
  = $\mu \exp(-\mu t)$ for $t \geq 0$
- The first arrival time $X_1$ is an exponential random variable with parameter $\mu$
- $E[X_1] = 1/\mu$

Distribution of the inter-arrival time

- The first arrival occurs at time $X_1 = \tau$
- At $t = \tau$, the experiment begins again
- $X_2$ is second arrival time
- $X_2 - X_1$ is the inter-arrival time
- Given $X_1 = \tau$, we set up a similar diff. eq. for $P\{\text{no arrivals in } (\tau, \tau + t]\}$
- The inter-arrival time $X_2 - X_1$ is an exponential RV with parameter $\mu$

Distribution of waiting times

- $X_1$ is an inter-arrival time if there was an arrival at $t = 0$
- Otherwise, $X_1$ is the waiting time for the next arrival after $t = 0$
- pdf of $X_1$ is same as pdf of inter-arrivals
- Generally, waiting time for the next arrival after $t = \tau$ also has the same pdf
- It is not necessary to have an arrival at $\tau$ for this result to hold

An illustrative diagram

- Experiment begins again after each arrival
- Successive experiments are repeated independent trials
- $X_k - X_{k-1}$ is k-th inter-arrival time
- Proceeding as before, all the inter-arrival times are independent exponential RVs with parameter $\mu$
- Is $X_1$ also an inter-arrival time?
  Yes, if there was an arrival at $t = 0$
An illustrative diagram

| X_1 | X_2 | X_3 |

1st experiment starts at t = 0

X_i is an inter-arrival time if there was an arrival at t = 0; a waiting time otherwise

Distribution of the k-th arrival time

1. The k-th arrival occurs at random time X_k
2. We set up and solve a differential equation for P_k(t) = P(\text{exactly } k \text{ arrivals in } (0, t])
   = P(N(0, t] = k)
   = P(N(0, t+\Delta T] = k)
   \begin{align*}
   &= P([N(0, t] = k) \cap \{N(t, t+\Delta T] = 0\}) \\
   &+ P([N(0, t] = k-1) \cap \{N(t, t+\Delta T] = 1\})
   \end{align*}

The differential equation for P_k(t)

\[
\frac{dP_k(t)}{dt} = -\mu P_k(t) + \mu P_{k-1}(t)
\]

Solving the diff. eq. for P_k(t)

- We can set k = 1 and solve for P_1(t) since we know that P_0(t) = \exp(-\mu t)
- Next, set k = 2 and solve for P_2(t) since we know P_1(t), and so on
- Alternatively, we can use LaPlace transforms

LaPlace transforms for P_k(t)

- Initial condition: P_k(0) = P(\text{exactly } k \text{ arrivals in } (0, 0]) = P([N(0, 0] = k) = 0 (not one!!)
- s\cdot\mathcal{L}[P_k(t)] = -\mu s\cdot\mathcal{L}[P_k(t)] + \mu s\cdot\mathcal{L}[P_{k-1}(t)]
- \mathcal{L}[P_k(t)] = \frac{\mu s^k}{(s+\mu)^k} \mathcal{L}[P_0(t)] = \frac{\mu s^k}{(s+\mu)^k} \left( \frac{1}{s+\mu} \right)^2
- \mathcal{L}[P_{k-2}(t)] = \frac{\mu s^{k-2}}{(s+\mu)^{k-2}} \mathcal{L}[P_0(t)] = \frac{\mu s^{k-2}}{(s+\mu)^{k-2}} \left( \frac{1}{s+\mu} \right)^2
- \mathcal{L}[P_{k-1}(t)] = \frac{\mu s^{k-1}}{(s+\mu)^{k-1}} \mathcal{L}[P_0(t)] = \frac{\mu s^{k-1}}{(s+\mu)^{k-1}} \left( \frac{1}{s+\mu} \right)^2
- \mathcal{L}[P_k(t)] = \frac{\mu s^k}{(s+\mu)^k} \mathcal{L}[P_0(t)] = \frac{\mu s^k}{(s+\mu)^k} \left( \frac{1}{s+\mu} \right)^2
- \mathcal{L}[P_0(t)] = \left( \frac{1}{s+\mu} \right)
- \mathcal{L}[P_k(t)] = \frac{(\mu t)^k}{k!} \exp(-\mu t) = P[N(0, t] = k)

Invert LaPlace transform to get P_k(t)

- For any fixed value of t, P[N(0, t] = k) = \frac{(\mu t)^k}{k!} \exp(-\mu t)
- N(0, t] is a Poisson RV with parameter \mu t

Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
It's a Poisson process!
- For any fixed value of $t$, $P(N(0, t] = k) = \frac{[\mu t]^k}{k!} e^{-\mu t}$
- $N(0, t]$ is a Poisson RV with parameter $\mu t$
- Start counting arrivals after time $t$
- Then $N(t_1, t_2]$ is a Poisson RV with parameter $\mu(t_2 - t_1) = \mu \cdot \text{duration of interval}$
- The RVs $N(t_1, t_2], N(t_3, t_4], N(t_5, t_6], \ldots$ are independent if the intervals $(t_1, t_2], (t_3, t_4], (t_5, t_6], \ldots$ don't overlap

Distribution of the $k$-th arrival time
- The $k$-th arrival occurs at random time $X_k$
- $P(X_k > t) = P(N(0, t] \leq k-1)$
  $= \exp(-\mu t) \cdot \left(1 + (\mu t) + (\mu t)^2/2! + \ldots + (\mu t)^{k-1}/(k-1)! \right)$
- $f_X(t)$ is the derivative of CDF $F_X(t)$
  $= -\text{derivative of } P(X_k > t)$
  $= \mu((\mu t)^{k-1}/(k-1)! \exp(-\mu t))$ for $t > 0$
  $= \mu\exp(-\mu t)(\mu t)^{k-1}/\Gamma(k)$
- The $k$-th arrival time $X_k$ is a gamma random variable with parameters $(k, \mu)$

Waiting time for next $k$ arrivals
- The $k$-th arrival time $X_k$ is a gamma random variable with parameters $(k, \mu)$
- Starting at any time $t = \tau$, the waiting time for the next $k$ arrivals has the same pdf
- It is not necessary to have an arrival at $\tau$ for this result to hold
- $E[X_k] = k/\mu \equiv kE[X_1]$
- Average waiting time for $k$ arrivals
  $= k \times \text{average waiting time for one arrival}$

A final observation...
- The $k$-th arrival time $X_k$ is a gamma random variable with parameters $(k, \mu)$
- But, $X_k = (X_k - X_{k-1}) + (X_{k-1} - X_{k-2}) + (X_{k-2} - X_{k-3}) + \ldots + (X_2 - X_1) + X_1$
- is the sum of $k$ inter-arrival times, i.e. $k$ independent $(1, \mu)$ gamma RV
- Special case of a general result: The sum of independent gamma RVs with same scale parameter is a gamma RV
- order = sum of the orders; scale is same

Summary
- We have discussed the Poisson process in some detail
- Basic assumptions:
  - Arrival rate is $\mu$
  - $P(\text{one arrival in } \Delta T \text{ interval}) = \mu \cdot \Delta T$
  - $P(\text{no arrival in } \Delta T \text{ interval}) = 1 - \mu \cdot \Delta T$
  - Arrivals in disjoint time intervals are independent (process is memoryless)

Consequences of assumptions
- Inter-arrival times are exponential random variables with parameter $\mu$
- Average inter-arrival time is $1/\mu$
- Time of $k$-th arrival/waiting time for next $k$ is a gamma RV with parameters $(k, \mu)$
- Average waiting time for k-th arrival is $k/\mu$
- Number of arrivals in a time interval of length $t$ is a Poisson random variable with parameter $\mu t$