Memoryless property:

- suppose time between session arrivals \( Z \) is exponentially distributed

\[
\Pr\{Z > y\} = \int_y^\infty be^{-bt} dt = e^{-by}
\]

- suppose a session has not arrived for \( y \) seconds

- what is the probability that a session will not arrive for another \( x \) seconds?

\[
\Pr\{Z > x + y \mid Z > y\}?
\]

By conditioning:

\[
\Pr\{Z > x + y \mid Z > y\} = \frac{\Pr\{Z > x + y\}}{\Pr\{Z > y\}}
\]

Hence:

\[
\Pr\{Z > x + y \mid Z > y\} = \frac{e^{-b(x+y)}}{e^{-by}} = e^{-bx}
\]

\[
\longrightarrow \text{ the past doesn’t impact the future!}
\]
Of course, not surprising since exponential distribution essentially comes from independent coin tossing

\[ \rightarrow \text{ independence over time is built in} \]

Another interpretation/application:

\[ \rightarrow \text{ view } Z \text{ as session lifetime} \]

\[ \ldots \text{ if a session has lasted for } y \text{ seconds, can we predict if it will last for another } x \text{ seconds?} \]

\[ \rightarrow \text{ no: it’s just } e^{-bx} \]

\[ \rightarrow \text{ knowing the past doesn’t help know the future} \]

\[ \rightarrow \text{ not good for gambling} \]
Important empirical fact: time between session arrivals has been observed to be approximately exponentially distributed

→ e.g., TCP sessions, Web (HTTP) requests

→ refinements: additional burstiness (why?)

However, session lifetimes are not exponentially distributed!

→ tend of have “heavier” tails

→ exponential distribution: “light” tail

→ where have we seen heavier tails?
Lastly, let’s count:

→ with exponential interarrivals, how many arrivals?

→ Poisson distribution

→ $x$ arrivals in unit time interval: $e^{-c}c^x / x!$

→ $c = 1/b$

→ tail: $\Pr\{X > x\} < e^{-c}(ec/x)^x$

→ mean of Poisson distribution: $c$ (so $x > c$)

→ very light

→ large deviations (“outliers”) are rare

→ reincarnation of what?
Session- and Packet-Level Resource Provisioning

Viewpoint: treating packets individually is ok but . . .

  \[\rightarrow\] more meaningful: groups of packets
  \[\rightarrow\] “packet train”
  \[\rightarrow\] e.g., TCP sends window full packets
  \[\rightarrow\] e.g., in multimedia frame is relevant unit

Thus:

  • packet train as “session” (micro-session)
    \[\rightarrow\] need to be careful about meaning of session
    \[\rightarrow\] session within session within session . . .

  • one user engages in multiple sessions over time
    \[\rightarrow\] e.g., HTTP client/server request (HTTP runs on top of TCP)
    \[\rightarrow\] persistent vs. non-persistent sessions: HTTP/1.1 vs. 1.0
    \[\rightarrow\] TCP connection set-up/tear-down overhead
Ex.: on/off model

$X(t)$

ON OFF ON OFF ON OFF

$X_2(t)$

$X_3(t)$

$X(t)$

$\rightarrow$ on-period: TCP file transfer

$\rightarrow$ on-period length: file transfer completion time

$\rightarrow$ ignore internal details within on-period: sawtooth

$\rightarrow$ on-period could be VoIP session: CBR

$\rightarrow$ not exactly: a user talks only $40\%$ of the time

$\rightarrow$ approximate view: ok by Amdahl’s law

$\rightarrow$ “don’t fret about small things”
We know session arrivals are (approximately) Poisson; what about session lifetimes?

Important fact: TCP session lifetimes are heavy-tailed

\[ \Pr\{ Z > x \} \approx x^{-\alpha} \]

\[ \text{as opposed to: } \Pr\{ Z > x \} \approx e^{-bx} \]

\[ \text{exponent: } 1 < \alpha < 2 \text{ (closer to 1)} \]

\[ \text{note: different from Internet connectivity power-law} \]

\[ \text{much more likely session will last a long time} \]

\[ \text{has finite mean but infinite variance} \]

\[ \text{cat has a very fat tail ("too fat to carry")} \]

Why would TCP session lifetimes be heavy-tailed?

\[ \text{TCP traffic makes up bulk of Internet traffic} \]

\[ \text{greater than 80\%} \]
Important fact: TCP session lifetimes are heavy-tailed because file sizes are heavy-tailed!

→ empirical fact from file server (incl. Web) studies
→ after all, TCP mostly transports files
→ write simple script to tabulate file sizes on arthur

Ex.: UNIX file system study (Gordon Irlam, 1993)

→ most are small, but a few are very large
How to check if files sizes are heavy-tailed?

Since $\Pr\{Z > x\} \approx x^{-\alpha}$, take logarithm on both sides:

$\rightarrow \log \Pr\{Z > x\} \approx -\alpha \log x$

$\rightarrow$ linear function with negative slope $-\alpha$

$\rightarrow$ holds true for large $x$

$\rightarrow$ what’s the slope $\alpha$?

$\rightarrow$ we don’t care about details of small sizes (why?)
More details: how “large” is large and “small” is small

range: from 0 bytes to \( \sim 2 \) GB

90\% of files are smaller than 10 KB

mean around \( \sim 2 \) kB

variation in small size range

in some file systems: bimodal

Disk space consumed:

10\% of files consume 90\% of space

same for bandwidth

“mice and elephants” metaphor
Consequences of heavy-tailed session lifetime:

→ for resource provisioning
→ as usual: good and bad

First, the good news!

→ what might be good?

Ex.: popular heavy-tailed distribution: Pareto

\[
\Pr\{Z > x\} = \left(\frac{k}{x}\right)^\alpha
\]

where

\(0 < \alpha < 2\): shape parameter;
\(k \leq x\): location parameter
mean: \(E[Z] = \alpha k/(\alpha - 1)\)
Good news: predictability

\[ \text{if session has lasted } y \text{ sec, will last for another } x \text{ sec} \]

Compute:

\[
\Pr\{Z > x + y \mid Z > y\} = \frac{\Pr\{Z > x + y\}}{\Pr\{Z > y\}} = \frac{(k/(x + y))^\alpha}{(k/y)^\alpha} = \left(\frac{y}{y + x}\right)^\alpha
\]

Knowing the past allows predicting the future:

\[
\Pr\{Z > x + y \mid Z > y\} \to 1 \text{ as } y \to \infty
\]

- by observing for longer period, can get more certainty
- average expected future duration:

\[
\to E\{Z \mid Z > y\} = y\alpha/(\alpha - 1)
\]

- find a casino with heavy-tailed roulette wheel

\[
\to \text{no money worries!}
\]