Outline

- $X \sim \text{exponential}(\lambda)$
  - [$P[X \leq rE[X]]$]
  - Memoryless property
- $X \sim \text{Pareto}(k, \alpha)$
- Mass-weighted distribution
- $X \sim \text{Weibull}(\lambda, \alpha)$
- $X \sim \text{geometric}(p)$
- $z$-transforms

Exponential CDF: memoryless property

$X \sim \text{exponential}(\lambda)$

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

$X \sim \text{exponential}(\lambda)$

$P[X \leq x + t | X > x] = \frac{F_X(x + t) - F_X(t)}{1 - F_X(t)}$

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

i.e., probability epoch will last at least $t$ more units of time is independent of how long it has been since the epoch began.

Exponential service times: $E[X]$ is the average remaining service time.
\( X \sim \text{Pareto}(k, \alpha) \)

\[
\begin{align*}
F_X(x) &= 1 - \left( \frac{k}{x} \right)^\alpha \\
\alpha &= \text{shape parameter} \\
k &= \text{min. value of } X
\end{align*}
\]

e.g. web file sizes

"power law distribution", heavy-tailed

\( \alpha = \frac{1}{\log x} \)

Mass-weighted Distribution Function

\[
\begin{align*}
f_{\text{mass-w}}(y) &= \frac{1}{E[X]} \int_0^\infty x f_x(x) \, dx \\
x = \text{exponential}: \quad f_{\text{mass-w}}(y) &= 1 - e^{-y(1+\lambda)} \\
x = \text{Pareto}: \quad f_{\text{mass-w}}(y) &= \frac{1}{1 + y^{1-\alpha}} \left( \frac{k}{y} \right)^{\alpha-1} \quad x \geq k, \quad \alpha > 1
\end{align*}
\]

Pareto has significantly more mass in the tail

Measuring a CDF

Select a period of stationary CDF

- approx. 100 requests/hour during peak hours (e.g., 7pm-1am)
Fitting a Measured CDF

Example CDF: time between arrivals to "BIBS" server (measured during period of "stable" number of requests/hr)

3/23, 6-11pm:

To fit exponential CDF: $\lambda = 1/E[X]$  
$1 - F(x) = e^{-\lambda x}$

To fit Pareto CDF: $E[X] = ku/(\alpha - 1)$ & $E[X^2] = ku/\alpha (\alpha - 2)$  
$1 - F(x) = \left(\frac{k}{x}\right)^{\alpha}$

Example Fitted CDFs

CDF of time between arrivals to "BIBS" media server (measured during period of "stable" number of requests/hr)

3/23, 6-11pm:

client interarrival times ~ exponential($\lambda$)

Pareto($k, \alpha$) vs Exponential($\lambda$)

Time between arrivals to eTeach media server

10/10, 9am-noon:

Which fit is best for modeling? 96% of client interarrival times ~ exponential($\lambda$)  
the body of the distribution is at least as important as the tail

Weibull($\lambda, \alpha$)

$F(x) = 1 - e^{-\lambda x}$  
$f(x) = \alpha \lambda \left(\frac{x}{\lambda}\right)^{\alpha - 1} e^{-\lambda x}$  
$x \geq 0; \ \alpha, \lambda > 0$

eTeach 10/10, 9am-noon

log 1-F(x)
**X ~ Zipf(\(\alpha\))**

\[ p_X(k) = \frac{1/k^{\alpha+1}}{\sum_{k=1}^{\infty} 1/k^{\alpha+1}}, \alpha > 0, \ k = 1,2,3,... \]

E.g., \(X ~ Zipf(0)\)

Example model: \(X = \) the video or file requested by a client

---

**X ~ geometric(\(p\))**

\[ p_X(k) = (1-p)^{k-1}p, \quad 0 < p \leq 1, \ k = 1,2,3,... \]

\[ \mathbb{E}(X) = \sum_{k=0}^{\infty} k(1-p)^{k-1}p = \frac{1/p}{1-(1-p)p} = \frac{1}{1-2p} \]

\(p_X(k) = (1-p) \cdot \mathbb{E}(X) = (1-p)k \cdot (1/p) = 1/p \)

Example models:
- \(X = \#\) times job visits the CPU
- \(X = \#\) processors requested by an O2K job

---

**z-transform: \(g_X(z)\)**

Let \(X\) be a non-negative integer-valued R.V.

\[ g_X(z) = \mathbb{E}[z^X] = \sum_{k=0}^{\infty} z^k p_X(k) \]

Example: \(X ~ geometric(p)\)

\[ g_X(z) = \sum_{k=0}^{\infty} z^k (1-p)^{k-1}p = pz \sum_{j=0}^{\infty} [(1-p)z]^j = \frac{pz}{1-(1-p)z} \quad |(1-p)z| < 1 \]

- Multiply each \(p_X(k)\) with a unique value, \(z^k\), and sum
- Compress infinite sequence to closed form expression

Be careful about allowed values of \(z\)
z-Transforms: $E[X]$

$g_X(z) = \sum_{k=0}^{\infty} z^k p_X(k)$

\[
\frac{d}{dz} g_X(z) \bigg|_{z=1} = \sum_{k=1}^{\infty} k z^{k-1} p_X(k) = \bar{X}
\]

Example: $X \sim \text{geometric}(p)$

$g_X(z) = \frac{pz}{1-(1-p)z}$, $|(1-p)z| < 1$

$\bar{X} = \frac{d}{dz} \frac{pz}{1-z-pz} \bigg|_{z=1} \frac{p(1-z+pz)-pq(p-1)}{(1-z+pz)^2} \bigg|_{z=1} \frac{p^2+pz-p^2}{p^2} \frac{1}{p}$

z-Transforms: $E[X^2], p_X(k)$

$g_X(z) = \sum_{k=0}^{\infty} z^k p_X(k)$

\[
\frac{d^2}{dz^2} g_X(z) \bigg|_{z=1} = \sum_{k=1}^{\infty} k(k-1) z^{k-2} p_X(k) = \bar{X^2} - \bar{X}
\]

\[
\frac{d^k}{dz^k} g_X(z) \bigg|_{z=0} = k! p_X(k) \quad \text{e.g.,} \quad \frac{d^2}{dz^2} g_X(z) \bigg|_{z=0} = -2p_X(2)
\]

(z-transform is also called probability generating function)