Conditional Expectation

\[ E[X|Y=y] = \sum_{i \in T} x_i p_{X|Y}(x_i|y) \text{ if } X \text{ is discrete} \]
\[ E[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) \, dx \text{ if } X \text{ is continuous} \]

example:
\[ Y = \text{job class} \in \{1, 2, \ldots, r\} \]
\[ X = \text{job service time}, \quad f_{X|Y}(x|i) = \lambda_i e^{-\lambda_i x} \]
\[ E[X | Y=i] = \frac{1}{\lambda_i} \]
also, \[ E[X^2 | Y=i] = \frac{2}{\lambda_i^2} \]
Conditional Expectation

\[ \bar{X}^{\text{Y}} = \sum E[X|Y = y] p_Y(y) \text{ if } Y \text{ is discrete} \]

\[ \bar{X}^{\text{Y}} = \int E[X|Y = y] f_Y(y)dy \text{ if } Y \text{ is continuous} \]

Example: G/G/1 queue

\[ N_a = \# \text{ customers in the queue at an arrival instant} \]

\[ E[R] = E[R|N_a = n] P[N_a = n] \]

\[ Z \sim \text{hyperexponential}(k,\{\alpha_i\},\{\lambda_i\}) \]

To compute \( E[Z] \), \( C_Z \):

\[ f_Z(z) = \sum_{i=1}^{k} \alpha_i e^{-\lambda_i x} \]

\[ F_Z(z) = \sum_{i=1}^{k} \alpha_i (1 - e^{-\lambda_i x}) \]

\( k \in \{1, 2, 3, \ldots\} \)

\( \alpha_i > 0 \), \( \lambda_i > 0 \), \( x > 0 \)

\( \alpha_1 + \alpha_2 + \cdots + \alpha_k = 1 \)

\( e.g., Z \sim \text{hyperexponential}(2, (0.95, 0.05), (1/10, 1/210)) \)

\( k=1: Z \sim \text{exponential}() \)

\( k>1: Z \) has a fatter tail than the exponential; (more variability)

\[ Z \sim \text{hyperexponential}(k,\{\alpha_i\},\{\lambda_i\}) \]

To compute \( E[Z] \), \( C_Z \):

\[ f_Z(z) = \sum_{i=1}^{k} \alpha_i e^{-\lambda_i x} \]

\[ F_Z(z) = \sum_{i=1}^{k} \alpha_i (1 - e^{-\lambda_i x}) \]

\( Y \in \{1, 2, \ldots, k\} \)

\( p_Y(i) = \alpha_i, \alpha_i > 0, \)

\[ Z = \alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_k X_k \]

\[ F_Z(z) = \sum_{i=1}^{k} p_Y(i) P[Z \leq z] = \sum_{i=1}^{k} \alpha_i (1 - e^{-\lambda_i z}) \]

\[ E[Z] = \sum_{i=1}^{k} p_Y(i) E[Z|Y=i] = \sum_{i=1}^{k} \frac{\alpha_i}{\lambda_i} \]

\( \text{"k-stage hyperexponential"} \)

\[ Z \sim \text{hyperexponential}(k,\{\alpha_i\},\{\lambda_i\}) \]

To compute \( E[Z], C_Z \):

\[ f_Z(z) = \sum_{i=1}^{k} \alpha_i e^{-\lambda_i x} \]

\[ F_Z(z) = \sum_{i=1}^{k} \alpha_i (1 - e^{-\lambda_i x}) \]

\( Y \in \{1, 2, \ldots, k\} \)

\( p_Y(i) = \alpha_i, \alpha_i > 0, \)

\[ Z = \alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_k X_k \]

\[ F_Z(z) = \sum_{i=1}^{k} p_Y(i) P[Z \leq z] = \sum_{i=1}^{k} \alpha_i (1 - e^{-\lambda_i z}) \]

\[ E[Z] = \sum_{i=1}^{k} p_Y(i) E[Z|Y=i] = \sum_{i=1}^{k} \frac{\alpha_i}{\lambda_i} \]

then

\[ E[Z] = \sum_{i=1}^{k} p_Y(i) E[Z|Y=i] = \sum_{i=1}^{k} \frac{\alpha_i}{\lambda_i} \]

\( C_Z \geq 1 \) (model: cpu service times)
Workload Characterization

Goal: complete characterization
all parameters needed for synthetic workload
e.g., O2K jobs for scheduling
  - job interarrival time
  - requested number of processors
  - requested memory
  - requested runtime
  - actual runtime
  - actual memory usage

Workload Characterization

Goal: avoid mixture distributions
measure distribution for population &
during period where the distribution is stationary
e.g.,
  - job or request interarrival times during period when the arrival rate is constant
  - file popularity during one day (not during one month)
  - peer-to-peer sessions in Europe, North America
  - session duration for passive peers, active peers
  - time until first request, between requests, after last request
Workload Characterization

- goal: use conditional distributions to capture correlations between parameters
  - e.g., size of memory request per O2K job

- fractional of jobs requested memory (MB)
  - P = 2
  - P = 3–5
  - P = 9–32
  - P = 33–64

- number of processors

Fraction of jobs with memory requests (GB)

80-percentile, average, 50-percentile, 20-percentile

Workload Characterization

- goal: complete characterization
  - all parameters needed for synthetic workload
  - e.g., peer-to-peer system workload

- fraction of sessions that are passive
- session duration for passive peers
- active sessions:
  - number of requests
  - time until first request
  - time between requests
  - time after last request
  - request popularity

Workload Characterization

- complete characterization
  - e.g., peer-to-peer system workload
  - session duration for passive peers

- Fraction of Sessions with Duration > x
  - Europe
  - North America
  - Asia

Workload Characterization

- complete characterization
  - e.g., peer-to-peer system workload
  - Number of queries per active session

- Fraction of Sessions with #Queries > x
  - Europe
  - North America
  - Asia
Workload Characterization

- complete characterization
  - e.g., peer-to-peer system workload
  - Distribution of number of queries per session, peers in North America

![Graph showing the distribution of number of queries per session.](image)

Questions?