For rest of course, assume M/G/1, where G has finite mean, variance. Assume $\rho < 1$, and a work-conserving scheduling policy.

Today: **Non-preemptive** scheduling policies.

1. Non-preemptive policies that don't make use of job size (Chpt 29)
2. Non-preemptive policies that make use of job size/class (Chpt 31)

1 Non-preemptive policies that don't use size

Example: **FCFS**, **LCFS**, **RANDOM**:

Theorem: All work-conserving, non-preemptive service orders that do not make use of job sizes have the same distribution on the number of jobs in the system.

PROOF:

Question: Why was it important that the scheduling policies not use job size?

2 Higher moments of response time

Question: What is $\mathbf{E}[T_Q]$ for all non-preemptive policies that don't use job size?

Question: How does $\widehat{N}(z)$ differ for FCFS, LCFS, and Random?

Question: Can we use DLL to convert $\widehat{N}(z)$ to $\widetilde{T}(s)$?

Question: Do FCFS, Random, and LCFS have the same $\mathbf{Var}(T)$?

GOAL: We will now derive $\widetilde{T_Q}^{LCFS}(s)$.

But first, it will help to review Busy Periods!

3 Reviewing Busy Periods

Given M/G/1 queue with arrival rate λ and job sizes S.

B_x = Busy period started by job of size x.

$B = B_S$ = Busy period started by arbitrary job

B_W = Busy period started by W work

Question: Derive $\widetilde{B_x}(s)$:

Question: Derive $\widetilde{B}(s)$:

Question: Derive $\widetilde{B_W}(s)$:

4 Derive $\widetilde{T_Q}^{LCFS}(s)$

After lots of differentiating we can show from the transforms for LCFS and FCFS that:

$$\mathbf{E}\left[T_Q^2\right]^{LCFS} = \frac{\lambda \mathbf{E}\left[S^3\right]}{3(1-\rho)^2} + \frac{\left(\lambda \mathbf{E}\left[S^2\right]\right)^2}{2(1-\rho)^3}$$

By contrast:

$$\mathbf{E}\left[T_Q^2\right]^{FCFS} = \frac{\lambda \mathbf{E}\left[S^3\right]}{3(1-\rho)} + \frac{\left(\lambda \mathbf{E}\left[S^2\right]\right)^2}{2(1-\rho)^2}$$

Question: How do $\mathbf{E}\left[T_Q^2\right]^{LCFS}$ and $\mathbf{E}\left[T_Q^2\right]^{FCFS}$ compare?

5 Review of tagged job analysis in M/G/1/FCFS

Question: Derive $\mathbf{E}[T_Q]$ for M/G/1/FCFS:

Vocabulary: T_Q : time in queue

N_Q : number in queue

N_Q^A : number in queue seen by arrival

S_i : service time of the i^{th} job in the queue

S : service time of a job, where $\mathbf{E}[S] = \frac{1}{\mu}$

S_e: excess of S. The remaining service time of the job in service, *given* there's a job in service.

We now turn to non-preemptive policies that make use of size.

6 The Non-Preemptive Priority Queue

DRAW PICTURE:

- $n =$
- $S_k =$
- $\lambda_k =$
- $\lambda =$
- $\rho_k =$

Question: What is $\mathbf{E}[S]$?

Question: What is $\sum_{k=1}^{n} \rho_k$?

Question: What is $\mathbf{E}\left[S^2\right]$?

Question: What is $\mathbf{E}[S_e]$?

GOAL: Derive $\mathbf{E}[T_Q(k)]$.

7 Deriving $\mathbf{E}[T_Q(1)]$

Question: Do type 1's care about other 1's arriving before them? _____

Question: Do type 1's care about other 1's that arrived after them? _____

Question: So can we say $\mathbf{E}[T_Q(1)]$ is the mean queueing time in an M/G/1 with arrival rate λ_1 and job size S_1? _____

<u>Tagged Job Argument</u>:

8 Deriving $\mathbf{E}[T_Q(2)]$

9 Deriving $\mathbf{E}[T_Q(k)]$

Question: What does type k arrival have to wait for?

1. ______________________________

2. ______________________________

3. ______________________________

10 Shortest-Job-First (SJF)

Definition: The **SJF** policy is a non-preemptive scheduling order. Whenever the server is free, it picks to run the job with the shortest original size.

GOAL: We want $\mathbf{E}[T_Q(x)]$, the mean delay (queueing time) for a job of size x.

Question: Can we think of SJF as something we just learned?

Question: What is $\rho(x)$, the load made up of jobs of size $< x$?

Question: What is $\mathbf{E}[T_Q(x)]^{SJF}$?

11 Comparison of SJF and FCFS

Question: How does $\mathbf{E}[T_Q(x)]^{SJF}$ compare with $\mathbf{E}[T_Q(x)]^{FCFS}$?

Question: What's the problem with all non-preemptive scheduling policies?