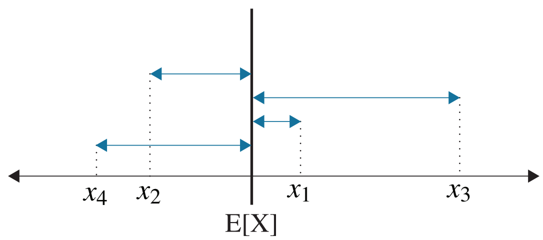
What is Variance?

<u>Defn</u>: The **variance** of a r.v. X, written as $\mathbf{Var}(\mathbf{X})$, is the expected squared difference of X from its mean (i.e., the square of how much we expect X to differ from its mean, $\mathbf{E}[X]$).

$$\mathbf{Var}(X) = \mathbf{E}\left[(X - \mathbf{E}[X])^2\right] .$$



Example: You collect all these samples of response time, T: 3, 4, 6, 2, 5

Question: Based on these, what is $\mathbf{E}[T]$?

Question: From the above definition, what is $\mathbf{Var}(T)$?

Why this definition?

Two equivalent definitions of variance

Defn 1:

$$\mathbf{Var}(X) = \mathbf{E}\left[(X - \mathbf{E}[X])^2\right] \ .$$

Defn 2:

$$\mathbf{Var}(X) = \mathbf{E}\left[X^2\right] - \mathbf{E}[X]^2 \ .$$

Question: Prove that these are the same.

Practice computing variance

From homework:

$$f(t) = \frac{5}{4}t^{-2} \qquad \text{where } 1 < t < 5$$

Practice computing variance

Kaiyang's distribution of memory allocations:

$$X = \begin{cases} 1 & \text{w.p. } \frac{9}{10} \\ 100 & \text{w.p. } \frac{1}{10} \end{cases}$$

Question: What is $\mathbf{Var}(X)$?

$$X = \begin{cases} Y & \text{w.p. } \frac{9}{10} \\ Z & \text{w.p. } \frac{1}{10} \end{cases}$$

Question: What is $\mathbf{Var}(X)$?

Practice computing variance

$X \sim \mathrm{Exp}(\mu)$

(Note you might want to use an integral calculator.)

Mutual Funds versus a Single Stock

Theorem: Let X and Y be random variables where $X \perp Y$. Then

$$\mathbf{Var}(X+Y) = \mathbf{Var}(X) + \mathbf{Var}(Y)\ .$$

If $X_1, X_2, \ldots, X_n$ are independent, then

$$\mathbf{Var}(X_1 + X_2 + \cdots + X_n) = \mathbf{Var}(X_1) + \mathbf{Var}(X_2) + \cdots + \mathbf{Var}(X_n)\ .$$

Question: Is it safer to split your money among 3 independent stocks, or buy 3 shares of the same stock?

Let X be the return of stock 1
Let Y be the return of stock 2
Let Z be the return of stock 3

Suppose that:
$X \perp Y \perp Z$ and $\mathbf{E}[X] = \mathbf{E}[Y] = \mathbf{E}[Z]$ and $\mathbf{Var}(X) = \mathbf{Var}(Y) = \mathbf{Var}(Z)$.

OPTION 1: $X + Y + Z$ vs. OPTION 2: $3X$

Squared Coefficient of Variation

Suppose that X and Y are measuring the same quantity, but X is measured in centimeters and Y is measured in millimeters. As a result we have that:

$$X = \begin{cases} 3 & \text{w.p. } \frac{1}{3} \\ 2 & \text{w.p. } \frac{1}{3} \\ 1 & \text{w.p. } \frac{1}{3} \end{cases} \qquad\qquad Y = \begin{cases} 30 & \text{w.p. } \frac{1}{3} \\ 20 & \text{w.p. } \frac{1}{3} \\ 10 & \text{w.p. } \frac{1}{3} \end{cases}.$$

Question: We want to say that X and Y have the same variance, but do they??

Question: Is there a normalized version of variance that is the same for X and Y?

Squared Coefficient of Variation

<u>Defn</u>: The squared coefficient of variation of X is denoted by C_X^2, where

$$C_X^2 = \underline{\hspace{5cm}} \, .$$

Question: Suppose $X = 5$. What is C_X^2?

Question: Suppose $X \sim \text{Exp}(\mu)$. What is C_X^2?

Question: What is Kaiyang's C_X^2?

************ BREAK ****************

The simplest queue

What is an M/G/1?

What is an M/M/1?

What is an M/D/1?

What is an D/D/1?

M/G/1 mean queueing time

$$\mathbf{E}[T_Q] = \frac{\rho}{1-\rho} \cdot \frac{\mathbf{E}[S^2]}{2\mathbf{E}[S]}$$

$$= \frac{\rho}{1-\rho} \cdot \frac{\mathbf{E}[S]}{2} \cdot \left(C_S^2 + 1\right)$$

Question: What does this mean for an M/D/1?

Question: What does this mean for an M/M/1?

Question: What changes in an M/G/1?

M/G/1 mean queueing time

$$\begin{aligned}\mathbf{E}\left[T_Q\right] &= \frac{\rho}{1-\rho} \cdot \frac{\mathbf{E}\left[S^2\right]}{2\mathbf{E}\left[S\right]} \\ &= \frac{\rho}{1-\rho} \cdot \frac{\mathbf{E}\left[S\right]}{2} \cdot \left(C_S^2 + 1\right)\end{aligned}$$

Question: What is the importance of load ρ?

Question: Why does C_S^2 come up?

Question: What is the meaning of the term that doesn't involve ρ?

Understanding the M/G/1

Question: Can you have a system with very low load but very high $\mathbf{E}[T_Q]$? (CombineNet story)

Question: What is $\mathbf{E}[T]$ for the M/G/1?

Important: We don't just know $\mathbf{E}[T]$. We can actually get all moments of response time. The formulas just get more complex.

For example: (see book)

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

Additionally, we can derive the tail of response time, $\mathbf{P}\{T > t\}$, for some particular distributions S.

*************** BREAK ***************

Application: Doubling the arrival rate & server speed?

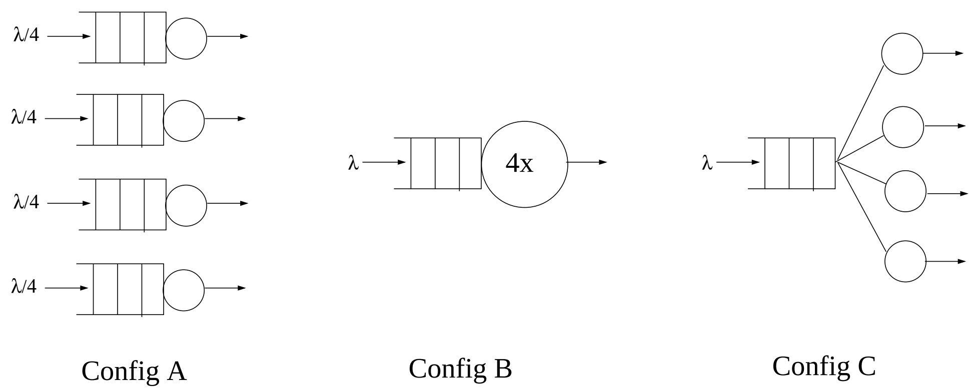
In homework, you looked at doubling the arrival rate and then doubling the server speed to compensate.

Question: Did that keep $\mathbf{E}[T]$ constant? Or did $\mathbf{E}[T]$ go up? Or down?

Let's try to understand this analytically:

Application: Comparison of three server organizations

Three server organizations. Outside arrivals occur according to a Poisson proces with rate λ. Job sizes denoted by r.v. S. When a job runs on a server of speed 4x, its service time is $S/4$.

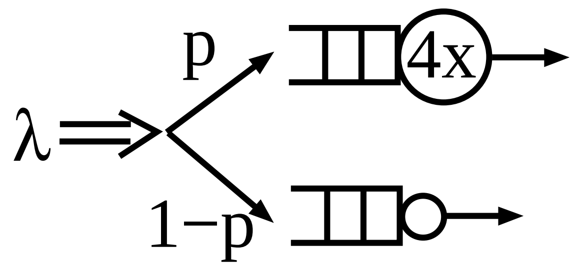


Question: Where does each configuration come up in practice?

Question: What is load ρ for each configuration?

Question: Which configuration is best for minimizing $\mathbf{E}[T]$?

Application: Load balancing



Question: What does "load balancing" mean?

Question: Should you balance load?