

Het $M/G/1$ model

In veel toepassingen is de aanname van exponentiële bedieningstijden niet realistisch (denk bijv. aan produktietijden). Daarom zullen we nu naar het model kijken met *willekeurig* verdeelde bedieningstijden.

Model:

- Aankomstproces is een Poisson proces met intensiteit λ .
- Bedieningstijden van klanten (Y_1, Y_2, \dots) zijn identiek verdeeld met een *willekeurige* verdelingsfunctie.

Gemiddelde bedieningstijd: $E(Y_1) = \tau$.

Variantie van de bedieningstijd: $E((Y_1 - E(Y_1))^2) = \sigma^2$.

Tweede moment van de bedieningstijd: $E(Y_1^2) = \sigma^2 + \tau^2 = s^2$.

- Er is één bediende en de capaciteit van de wachtrij is onbegrensd.

Helaas is voor dit model het proces $\{X(t) : t \geq 0\}$, het aantal klanten in het systeem geen CTMC meer. Een gedetailleerde analyse (de bepaling van de limietverdeling van het proces $\{X(t) : t \geq 0\}$) zullen we achterwege laten.

We zullen ons beperken tot zogenaamde *mean-value analyse* (bepaling van gemiddelde wachttijd, gemiddeld aantal klanten in het systeem,).

Stabiliteitsvoorwaarde:

Net als in het $M/M/1$ model geldt dat voorwaarde voor stabiliteit is dat de hoeveelheid werk die per tijdseenheid wordt aangeboden kleiner moet zijn dan de hoeveelheid werk die de bediende per tijdseenheid kan doen, i.e.,

$$\rho := \lambda\tau < 1.$$

Bezettingsgraad van de bediende:

Aangezien er gemiddeld $\rho < 1$ tijdseenheid werk per tijdseenheid aan de bediende wordt aangeboden zal gelden dat de fractie van de tijd dat de bediende bezig is (= de bezettingsgraad van de bediende) gelijk is aan ρ . De fractie van de tijd dat de bediende niet bezig is is dus gelijk aan $1 - \rho$.

Gemiddelde tijd W_q die klant in de wachtrij staat:

De tijd die een klant in de wachtrij staat bestaat uit 2 delen:

- de *resterende* bedieningstijd van de klant bij de bediende;
- de bedieningstijden van de klanten in de wachtrij.

Om W_q te kunnen berekenen zullen we eerst dus iets moeten zeggen over de verwachte resterende bedieningstijd van de klant bij de bediende.

Verwachte resterende bedieningstijd van klant bij bediende

Hieronder staat een plaatje van de resterende bedieningstijd van de klant bij de bediende als functie van de tijd.

Neem groot interval ter lengte T .

Verwachte aantal bediende klanten in $[0, T]$: λT .

Iedere klant zorgt voor verwachte oppervlakte: $E(Y_1^2/2) = s^2/2$.

=> Totale verwachte oppervlakte onder grafiek: $\lambda T \cdot s^2/2$.

=> Verwachte resterende bedieningstijd: $\lambda s^2/2$.

De gemiddelde tijd W_q dat een klant in de wachtrij staat kan nu bepaald worden met behulp van de volgende *mean-value relaties*:

$$W_q = \lambda s^2 / 2 + L_q \tau,$$

$$L_q = \lambda W_q.$$

Merk op dat bij de eerste relatie we de PASTA eigenschap gebruiken en dat de tweede relatie de formule van Little toegepast op de wachtrij is.

Er volgt dus

$$W_q = \frac{\lambda s^2}{2(1 - \lambda \tau)} = \frac{\lambda s^2}{2(1 - \rho)},$$

$$L_q = \lambda W_q = \frac{\lambda^2 s^2}{2(1 - \rho)}.$$

Als we W_q en L_q kennen, volgen W en L natuurlijk met behulp van

$$W = W_q + \tau \quad \text{en} \quad L = L_q + \rho.$$

Voorbeeld: $M/M/1$ model

In het geval dat de bedieningstijden exponentieel verdeeld zijn met parameter μ dan geldt:

$$\tau = \frac{1}{\mu}, \quad \sigma^2 = \frac{1}{\mu^2}, \quad s^2 = \frac{2}{\mu^2},$$

en dus is de verwachte resterende bedieningstijd gelijk aan

$$\frac{\lambda s^2}{2} = \frac{\lambda}{\mu^2} = \rho \cdot \frac{1}{\mu}.$$

Dit volgt ook uit geheugenloosheid van exponentiële verdeling (leg uit).

Voor de grootheden W_q en L_q vinden we dus (als voorheen)

$$W_q = \frac{1}{\mu} \frac{\rho}{1 - \rho}, \quad L_q = \frac{\rho^2}{1 - \rho}.$$

Voorbeeld: $M/D/1$ model

In het geval dat de bedieningstijden deterministisch zijn en gelijk aan τ dan geldt:

$$\sigma^2 = 0, \quad s^2 = \tau^2,$$

en dus is de verwachte resterende bedieningstijd gelijk aan

$$\frac{\lambda s^2}{2} = \frac{\lambda \tau^2}{2} = \rho \cdot \frac{\tau}{2}.$$

Voor de grootheden W_q en L_q vinden we dus

$$W_q = \frac{\tau}{2} \frac{\rho}{1 - \rho}, \quad L_q = \frac{\rho^2}{2(1 - \rho)}.$$

Merk op dat kleinere variantie van de bedieningstijden dus leidt tot kleinere wachtrijen en wachttijden.

Het $G/M/1$ model

We gaan nu kijken naar het model waarbij niet de bedieningstijden maar de tussenaankomsttijden willekeurig verdeeld zijn, het $G/M/1$ model.

Model:

- Het aankomstproces is een proces waarbij de tussenaankomsttijden (A_1, A_2, \dots) van klanten identiek verdeeld zijn met een willekeurige verdelingsfunctie $G(\cdot)$. De gemiddelde tijd tussen 2 aankomsten is $E(A_1) = 1/\lambda$. De functie $\tilde{G}(s)$ wordt gedefiniëerd als

$$\tilde{G}(s) = E(e^{-sA_1}).$$

- Bedieningstijden zijn exponentieel verdeeld met parameter μ .
- Er is één bediende en de capaciteit van de wachtrij is onbegrensd.

De functie $\tilde{G}(s) = E(e^{-sA_1})$ heet de *Laplace-Stieltjes getransformeerde* van de stochastische variabele A_1 en kan als volgt worden uitgerekend.

- Als A_1 een discrete stochastische variabele is met kansverdeling $p(x_i) = P(A_1 = x_i), i = 1, 2, \dots$, dan

$$\tilde{G}(s) = \sum_{i=1}^{\infty} e^{-sx_i} p(x_i).$$

- Als A_1 een continue stochastische variabele is met kansdichtheid $g(\cdot)$, dan

$$\tilde{G}(s) = \int_0^{\infty} e^{-sx} g(x) dx.$$

Voorbeelden:

- Als A_1 deterministisch is en gelijk aan $1/\lambda$, dan $\tilde{G}(s) = e^{-s/\lambda}$.
- Als A_1 exponentieel verdeeld met parameter λ , dan $\tilde{G}(s) = \lambda/(\lambda + s)$.

Helaas is ook voor dit model het proces $\{X(t) : t \geq 0\}$, het aantal klanten in het systeem geen CTMC meer. Ook de bij het $M/G/1$ model gepresenteerde mean-value analyse werkt niet meer, aangezien we de PASTA eigenschap hier niet meer mogen toepassen (het aankomstproces is immers geen Poisson proces meer).

We zullen ons beperken tot het noemen van resultaten voor de limietverdeling van het aantal klanten op aankomstmomenten $(\pi_j^*, j = 0, 1, 2, \dots)$ en op willekeurige momenten $(p_j, j = 0, 1, 2, \dots)$.

Stabiliteitsvoorwaarde:

Net als in het $M/G/1$ model geldt dat voorwaarde voor stabiliteit is dat de hoeveelheid werk die per tijdseenheid wordt aangeboden kleiner moet zijn dan de hoeveelheid werk die de bediende per tijdseenheid kan doen, i.e.,

$$\rho := \frac{\lambda}{\mu} < 1.$$

Limietverdeling van het aantal klanten op aankomstmomenten

De limietverdeling van het aantal klanten op aankomstmomenten wordt gegeven door

$$\pi_j^* = (1 - \alpha)\alpha^j, \quad j \geq 0,$$

waarbij α de unieke oplossing in het interval $(0,1)$ is van de vergelijking

$$u = \tilde{G}(\mu(1 - u)).$$

Voorbeeld:

Als de tussenaankomsttijden exponentieel verdeeld zijn met parameter λ , dan geldt $\alpha = \rho$ (ga na!) en dus

$$\pi_j^* = (1 - \rho)\rho^j, \quad j \geq 0,$$

Limietverdeling van het aantal klanten op willekeurige momenten

De limietverdeling van het aantal klanten op willekeurige momenten wordt gegeven door

$$p_0 = 1 - \rho, \quad p_j = \rho\pi_{j-1}^* = \rho(1 - \alpha)\alpha^{j-1}, \quad j \geq 1.$$

Idee bewijs:

Het aantal keer per tijdseenheid dat het aantal klanten in het systeem van $j - 1$ naar j springt is $\lambda\pi_{j-1}^*$.

Het aantal keer per tijdseenheid dat het aantal klanten in het systeem van j naar $j - 1$ springt is μp_j .

Aangezien deze twee intensiteiten gelijk moeten zijn geldt $p_j = \rho\pi_{j-1}^*$.

Gemiddeld aantal klanten in het systeem

Het gemiddeld aantal klanten in het systeem wordt gegeven door

$$L = \sum_{j=1}^{\infty} j p_j = \rho(1 - \alpha) \sum_{j=1}^{\infty} j \alpha^{j-1} = \frac{\rho}{1 - \alpha}.$$

Gemiddelde tijd dat klanten in het systeem

De gemiddelde tijd dat klanten in het systeem wordt gegeven door

$$W = \frac{L}{\lambda} = \frac{1}{\mu(1 - \alpha)}.$$

Alternatieve afleiding

$$W = \sum_{j=0}^{\infty} \pi_j^* \frac{j + 1}{\mu} = \frac{1 - \alpha}{\mu} \sum_{j=0}^{\infty} (j + 1) \alpha^j = \frac{1}{\mu(1 - \alpha)}.$$

Het $G/G/1$ model

Het laatste wachtrijmodel dat we bespreken is het $G/G/1$ model. In het model zijn zowel de tussenaankomsttijden als de bedieningstijden *willekeurig* verdeeld.

Voor dit model is een exacte analyse in het algemeen niet mogelijk. Daarom beperken we ons tot het geven van *benaderingen* voor prestatiematen als:

- de gemiddelde tijd W_q dat klanten in de wachtrij staan;
- de gemiddelde tijd W dat klanten in het systeem zijn;
- het gemiddeld aantal klanten L_q dat in de wachtrij staat;
- het gemiddeld aantal klanten L dat in het systeem is.

Model:

- Het aankomstproces is een proces waarbij de tussenaankomsttijden (A_1, A_2, \dots) van klanten identiek verdeeld zijn met een willekeurige verdelingsfunctie.

Gemiddelde tijd tussen 2 aankomsten: $E(A_1)$.

Variantie van de tijd tussen 2 aankomsten: $E((A_1 - E(A_1))^2) = \sigma_{A_1}^2$.

Variatiecoëfficiënt van de tijd tussen 2 aankomsten: $c_{A_1} = \frac{\sigma_{A_1}}{E(A_1)}$.

- Bedieningstijden van klanten (B_1, B_2, \dots) zijn identiek verdeeld met een *willekeurige* verdelingsfunctie.

Gemiddelde bedieningstijd: $E(B_1)$.

Variantie van de bedieningstijd: $E((B_1 - E(B_1))^2) = \sigma_{B_1}^2$.

Variatiecoëfficiënt van de bedieningstijd: $c_{B_1} = \frac{\sigma_{B_1}}{E(B_1)}$

- Er is één bediende en de capaciteit van de wachtrij is onbegrensd.

Stabiliteitsvoorwaarde:

Net als in het $M/M/1$, $M/G/1$ en $G/M/1$ model geldt dat voorwaarde voor stabiliteit is dat de hoeveelheid werk die per tijdseenheid wordt aangeboden kleiner moet zijn dan de hoeveelheid werk die de bediende per tijdseenheid kan doen, i.e.,

$$\rho := \frac{E(B_1)}{E(A_1)} < 1.$$

Benadering W_q :

Een veel gebruikte benadering voor de gemiddelde tijd dat klanten in de wachtrij staan is de volgende:

$$W_q \approx \frac{\rho}{1 - \rho} \cdot \frac{c_{A_1}^2 + c_{B_1}^2}{2} \cdot E(B_1)$$

Speciale gevallen:

Voor het $M/M/1$ model is de benadering exact:

$$W_q = \frac{\rho}{1 - \rho} \cdot \frac{1 + 1}{2} \cdot E(B_1) \quad (M/M/1)$$

Voor het $M/G/1$ model is de benadering exact:

$$W_q = \frac{\rho}{1 - \rho} \cdot \frac{1 + c_{B_1}^2}{2} \cdot E(B_1) \quad (M/G/1)$$

Voor het $G/M/1$ model is de benadering niet exact:

$$W_q \approx \frac{\rho}{1 - \rho} \cdot \frac{c_{A_1}^2 + 1}{2} \cdot E(B_1) \quad (G/M/1)$$

Benaderingen voor W , L_q en L :

Uit de benadering voor W_q

$$W_q \approx \frac{\rho}{1 - \rho} \cdot \frac{c_{A_1}^2 + c_{B_1}^2}{2} \cdot E(B_1)$$

volgen ook direct benaderingen voor W , L_q en L via de formules

$$W = W_q + E(B_1)$$

$$L_q = \frac{W_q}{E(A_1)} \quad (\text{Little})$$

$$L = \frac{W}{E(A_1)} \quad (\text{Little})$$

Voorbeeld:

- In een werkstation worden jobs aangeleverd met een intensiteit van 1 job per 8 uur.
- De spreiding van de tijd tussen twee opeenvolgende momenten dat een job wordt aangeleverd is 4 uur.
- De gemiddelde produktietijd van jobs op het werkstation is 6 uur en de spreiding van de produktietijd van jobs op het werkstation is 2 uur.

Vraag:

Hoe groot is de reductie van de tijd dat jobs bij het werkstation zijn als we in staat zijn de spreiding van de tijd tussen twee opeenvolgende momenten dat een job wordt aangeleverd te reduceren van 4 uur naar 1 uur?

Antwoord:

Deze tijd wordt gereduceerd van ongeveer 9.25 uur tot ongeveer 7 uur.