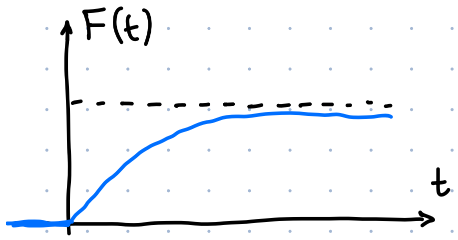
① Варка распределений

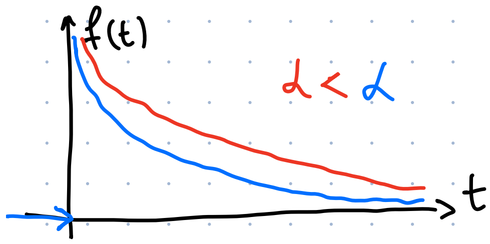
Упражнение

$X \sim Exp(\alpha)$

$$F_X(t) = \begin{cases} 1 - e^{-\alpha t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$



$$f_X(t) = \begin{cases} \alpha e^{-\alpha t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$



$$\mathbb{E}(X) = \frac{1}{\alpha}$$

$$Var(X) = \frac{1}{\alpha^2}$$

а) $Y = \sqrt{X}$

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(\sqrt{X} \leq y) = \mathbb{P}(X \leq y^2) = F_X(y^2) =$$

$$= \begin{cases} 1 - e^{-\alpha y^2}, & y \geq 0 \\ 0, & y < 0 \end{cases}$$

$F_X(x) \quad x \in [0; +\infty) \qquad Y = \sqrt{X}$

$F_Y(y) \quad y \in [\sqrt{0}; \sqrt{+\infty}) = [0; +\infty)$

б) $Y = X^2$

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y) = \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) =$$

$$= F_X(\sqrt{y}) - F_X(-\sqrt{y}) = 1 - e^{-\alpha\sqrt{y}} - 0, \quad y \geq 0$$

$F_X(x) \quad x \in [0; +\infty) \qquad Y = X^2$

$F_Y(y) \quad y \in [\,0^2; +\infty^2) = [0; +\infty)$

$y = 4$

$F_X(2) - \underbrace{F_X(-2)}_{0}$

в) $Y = 1 - e^{-\alpha X} = F_X(X)$

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}\left(1 - e^{-\alpha X} \leq y\right) = \mathbb{P}\left(X \leq -\frac{1}{\alpha}\ln(1-y)\right) =$$

$1 - y \leq e^{-\alpha X}$

$\ln(1-y) \leq -\alpha X$

$X \leq -\frac{1}{\alpha}\ln(1-y)$

X Y Z W - сл. вел.

x y z w - значения

$$= F_X\left(-\frac{1}{\alpha}\ln(1-y)\right) = 1 - e^{-\alpha \cdot \left[-\frac{1}{\alpha}\ln(1-y)\right]} =$$

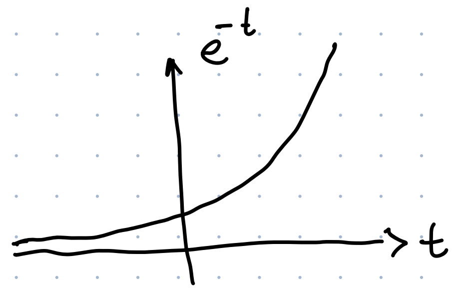
$$= 1 - (1-y) = y$$

$F_X(x) \quad x \in [0; +\infty) \qquad Y = 1 - e^{-\alpha X}$

$F_Y(y) \quad y \in [0; 1]$

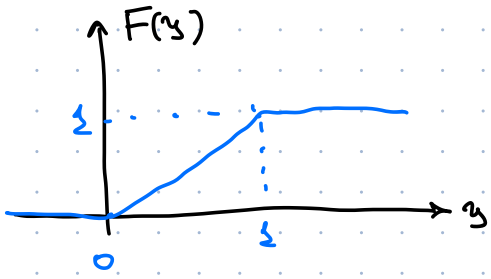
$\lim\limits_{x \to 0}(1 - e^{-\alpha x}) = 1 - e^0 = 0$

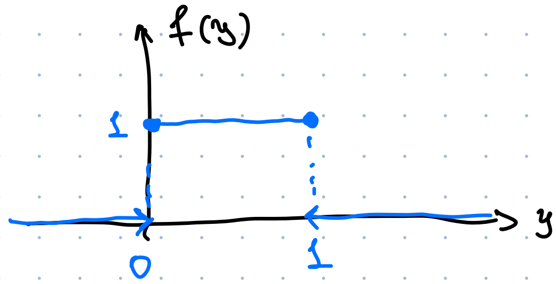
$\lim\limits_{x \to +\infty}\left(1 - e^{-\alpha x}\right) = 1 - e^{-\infty} = 1$



$$F_Y(y) = \begin{cases} y, & y \in [0; 1] \\ 0, & y < 0 \\ 1, & y \geq 1 \end{cases} \qquad f_Y(y) = \begin{cases} 1, & y \in [0; 1] \\ 0, & \text{иначе} \end{cases}$$

$P_Y(y)$





$X \sim Exp(\alpha)$

$Y = F_X(X) \sim U[0;1]$

Квантильное преобразование

Т. (Квантильное пр.-е.)

Пусть $F_X(x)$ непрерывна. Тогда сл. вел. $Y = F_X(X)$ имеет равномерное распределение

Док-во:

$$F_Y(y) = \mathbb{P}(F_X(X) \leq y) = \mathbb{P}(X \leq F_X^{-1}(y)) = F_X(F_X^{-1}(y)) = y$$

X	$-\infty$	$+\infty$
$F_X(X)$	0	1

ч.т.д. ■

Следствие:

$Y \sim U[0;1]$

хотим X с $F_X(x)$

$F^{-1}(y)$

Генерация сл. величин:

$y_1, \dots, y_n \sim iid\ U[0;1]$

$x_1, \dots, x_n \sim iid\ Exp(\alpha)$

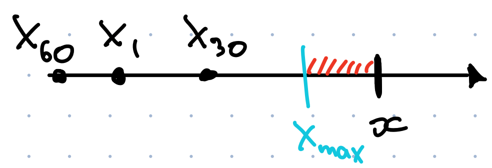
$$x_i = F_X^{-1}(y_i) = -\frac{1}{\alpha}\ln(1-y_i)$$

② Распределение максимума

$X_1, \dots, X_n \sim iid \; F_X(x)$

а) $X_{max} = \max(X_1, \dots, X_n) \sim ?$

$$F_{X_{max}}(x) = \mathbb{P}(X_{max} \leq x) = \mathbb{P}(X_1 \leq x \cap X_2 \leq x \cap \dots \cap X_n \leq x) =$$



$$= \underbrace{\mathbb{P}(X_1 \leq x)}_{F_X(x)} \cdot \underbrace{\mathbb{P}(X_2 \leq x)}_{F_X(x)} \cdot \ldots \cdot \underbrace{\mathbb{P}(X_n \leq x)}_{F_X(x)} = \left[F_X(x)\right]^n$$

$$f_{X_{max}}(x) = F'_{X_{max}}(x) = n \cdot \left[F_X(x)\right]^{n-1} \cdot f_X(x)$$

б) $X_{min} \sim ?$

$$F_{X_{min}}(x) = \mathbb{P}(X_{min} \leq x) = 1 - \mathbb{P}(X_{min} > x) =$$

$$= 1 - \mathbb{P}(X_1 > x \cap \dots \cap X_n > x) = 1 - \mathbb{P}(X_1 > x) \cdot \ldots \cdot \mathbb{P}(X_n > x) =$$

$$= 1 - (1 - \mathbb{P}(X_1 \leq x)) \cdot \ldots \cdot (1 - \mathbb{P}(X_n \leq x)) = 1 - (1 - F_X(x))^n$$

③ Метрика плохих показов

id	show
1	100k
2	3
....	....
42	15
....	....

p - доля плохих показов

$\mathbb{P}$(▶ - плохая | рандомный чел увидел)

$\underbrace{\boxed{1}\ \boxed{1} \dots \boxed{1}}_{100k}$

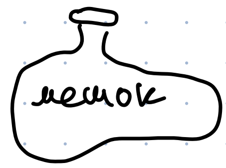
$\boxed{2}\ \boxed{2}\ \boxed{2}$

....

$\underbrace{\boxed{42}\ \boxed{42} \dots \boxed{42}}_{15}$

....

→



→ равномерно 100 карточек

↓

модераторы

$\boxed{i} \propto show_i$

$\frac{show_i}{\sum show_i}$

$\hat{p} = \frac{m}{100}$

np.random([1, 2, ..., 10^{100}], [$\frac{show_1}{\sum show_i}$, ...])

p_1, ...

Удачные совпадения:

→ всё влезло в память

→ все p_i не пробили float 64

Схема:

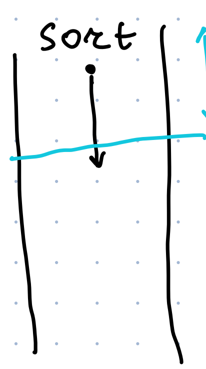
1 $X_1, \dots, X_{100k} \sim iid\ U[0;1]$

2 $X_1, X_2, X_3 \sim iid\ U[0;1]$

....

42 $X_1, X_2, \dots, X_{15} \sim iid\ U[0;1]$

...



разметка

X^{100k}_{max}

X^{3}_{max}

...

X^{15}_{max}

....

$F_{X_{max}}(x) = [F_X(x)]^n = x^n$

$Y \sim U[0;1]$

$X_{max} = \sqrt[n]{Y} \sim F_{max}(x)$

Новая проблема $\sqrt[n]{Y}$ плохая операция

$\sqrt[n]{Y} \to \boxed{\frac{1}{n} \ln \underbrace{Y}_{\in[0;1]}}$ ≤ 0

<u>Алгоритм сэмплирования:</u>

$n = show_i$

$-\frac{1}{n} \ln Y$

Идём по таблице с shows

1. Генерируем $Y \sim U[0;1]$
2. $\beta = \frac{1}{n} \ln Y$
3. Поддерживаем „список“ из 100 с наиб. β

„Куча“ - структура данных, которая хорошо
Heap поддерживает M min. элементов

можно взять интеграл и показать, что в такой процедуре каждый документ попадет в выборку с вер. $\frac{show_i}{\sum show_i}$

$\boxed{\text{Потом надо рассчитать } \hat{p} \text{ С этим будут проблемы}}$

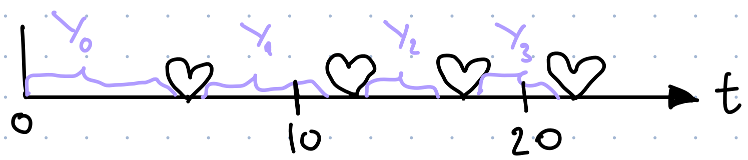
<u>To be continued</u>

④ Распределение Пуассона из воздуха

Даша

X_t – число ♡, которое Даша получила за время $[0; t]$

Y_n – время, которое прошло между n и n+1 лайком



$X_{10} = 1$

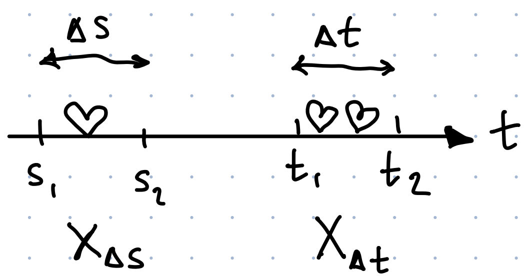
$X_{20} = 3$

$\boxed{\text{PA-0}}$ За нулевой промежуток времени ничего не происходит

$$\mathbb{P}(X_0 > 1) = 0$$

$$\mathbb{P}(X_0 = 0) = 1$$

$\boxed{\text{PA-1}}$ Отсутствие последействия



$X_{\Delta t}$ не зависит от $X_{\Delta s}$
если Δt и Δs
не пересекаются

$\boxed{\text{PA-2}}$ Стационарность

Появление k событий на промежутке Δt зависит только от длины этого промежутка и не зависит от точки отсчёта

$$\mathbb{P}(X_{\Delta t} = 1) = \lambda \cdot \Delta t$$

интенсивность потока событий

РА-3 Ординарность

Наступление более 1 соб. за ∞ малый промежуток времени – невозможно

$$o(\Delta) = -o(\Delta)$$

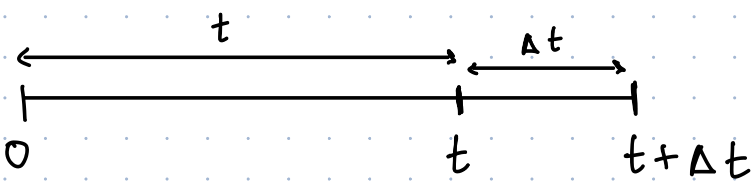
$$\mathbb{P}(X_{\Delta t} \geq 2) = o(\Delta t)$$

$$\mathbb{P}(X_{\Delta t} = 0) = 1 - [\lambda \Delta t + o(\Delta t)] = \underline{1 - \lambda \Delta t + o(\Delta t)}$$

Простейший поток событий

$$\mathbb{P}(X_t = k) - ?$$

$h_0(t) = \mathbb{P}(X_t = 0)$ $\quad h_1(t) = \mathbb{P}(X_t = 1)$ ......



$\Delta t \to 0$

$h_0(t)$ $\quad h_0(t+\Delta t)$ $\quad h_0(\Delta t)$ – <u>как связаны?</u>

РА-1 X_t $\quad X_{\Delta t}$ – нез.

$$h_0(t) = \mathbb{P}(X_t = 0)$$

$$h_0(\Delta t) = \mathbb{P}(X_{\Delta t} = 0)$$

$$\mathbb{P}(X_t = 0 \cap X_{\Delta t} = 0)$$

$$h_0(t+\Delta t) = \mathbb{P}(X_{t+\Delta t} = 0) = \mathbb{P}(X_t = 0) \cdot \mathbb{P}(X_{\Delta t} = 0)$$

$$h_0(t+\Delta t) = h_0(t) \cdot h_0(\Delta t) \quad \Big| \begin{array}{l} \div \Delta t \\ \ominus\, h_0(t) \end{array}$$

$$\frac{h_0(t+\Delta t) - h_0(t)}{\Delta t} = \frac{h_0(t)\cdot[h_0(\Delta t) - 1]}{\Delta t}$$

$\Delta t \to 0$ $\qquad h_0'(t) \qquad \dfrac{-\lambda\cdot\Delta t + o(\Delta t)}{\Delta t}$

$$h_0(\Delta t) = \mathbb{P}(X_{\Delta t} = 0) = 1 - \lambda\Delta t + o(\Delta t)$$

$$\lim_{\Delta t\to 0} h_0(t)\cdot\left(\frac{-\lambda\cancel{\Delta t}}{\cancel{\Delta t}} + \cancel{\frac{o(\Delta t)}{\Delta t}}^{\,0}\right) = -\lambda\cdot h_0(t)$$

$$h_0'(t) = -\lambda\cdot h_0(t) \qquad f'(t) = \cdot f(t) \qquad e^t$$

$$\frac{dh_0}{dt} = -\lambda h_0$$

$$\frac{dh_0}{h_0} = -\lambda\, dt$$

$\boxed{PA-0}$

$$\mathbb{P}(X_0 = 0) = h_0(0) = 1$$

$$\ln h_0 = -\lambda t + \ln(\text{const})$$

$$h_0(t) = \text{const}\cdot e^{-\lambda t}$$

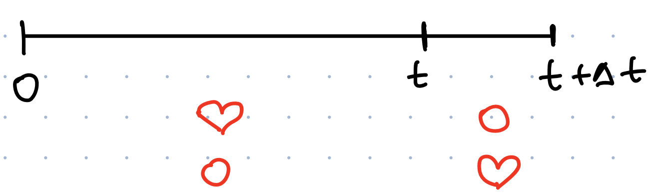
$$\text{const}\cdot e^{-\lambda\cdot 0} = 1$$

$$\text{const} = 1$$

$$\mathbb{P}(X_t = 0) = e^{-\lambda t}$$

$$\mathbb{P}(X_t = 1) = h_1(t)$$

$h_1(t) \qquad h_1(\Delta t) \qquad h_1(t+\Delta t)$



$\mathbb{P}(X_{t+\Delta t} = 1)$

$$h_1(t+\Delta t) = h_1(t)\cdot h_0(\Delta t) + h_0(t)\cdot h_1(\Delta t)$$

$$\frac{h_1(t+\Delta t) - h_1(t)}{\Delta t} = \frac{h_1(t)\left[h_0(\Delta t) - 1\right]}{\Delta t} + \frac{h_0(t)\cdot h_1(\Delta t)}{\Delta t}$$

$$\xrightarrow[\Delta t \to 0]{} h_1'(t) \qquad \frac{h_1(t)\left[-\lambda\Delta t + o(\Delta t)\right]}{\Delta t} \qquad \frac{h_0(t)\cdot \lambda\cdot\Delta t}{\Delta t}$$

$$-\lambda\cdot h_1(t) \qquad \lambda\cdot h_0(t)$$

$$h_1'(t) = -\lambda h_1(t) + \lambda \underbrace{h_0(t)}_{e^{-\lambda t}} \quad \Big| \otimes e^{\lambda t}$$

$$e^{\lambda t}\cdot h_1'(t) = -\lambda h_1(t)\, e^{\lambda t} + \lambda$$

$$e^{\lambda t} h_1'(t) + \lambda e^{\lambda t} h_1(t) = \lambda$$

$$\left[e^{\lambda t}\cdot h_1(t)\right]' = \lambda$$

$$e^{\lambda t}\cdot h_1(t) = \lambda t + \text{const}$$

$$h_1(t) = \lambda t\, e^{-\lambda t} + \text{const}\; e^{-\lambda t}$$

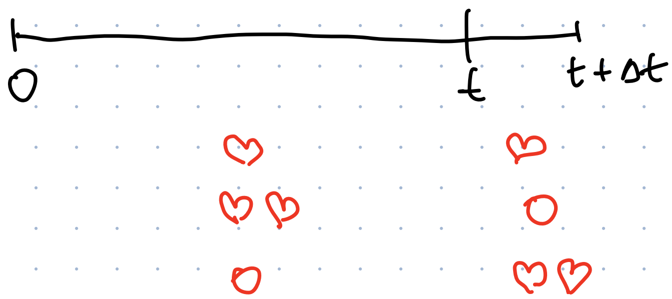
$$h_1(0) = \mathbb{P}(X_0 = 1) = 0$$

$$0 + \text{const}\cdot e^{-\lambda\cdot 0} = 0 \quad \Rightarrow \quad \text{const} = 0$$

$$h_1(t) = e^{-\lambda t} \cdot \lambda t$$

$$\mathbb{P}(X_t = 1) = e^{-\lambda t} \cdot \lambda t$$

$$\mathbb{P}(X_t = 2)$$



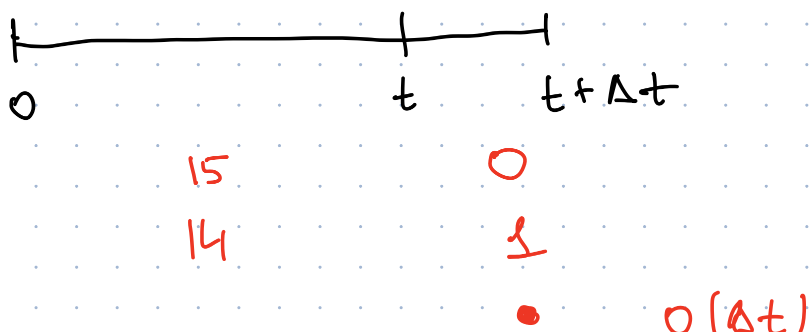
$h_1(t) \cdot h_1(\Delta t)$

$h_2(t) \cdot h_0(\Delta t)$

$h_0(t) \cdot \underbrace{h_2(\Delta t)}_{0}$ $\boxed{\text{PA-3}}$

$$h_2(t+\Delta t) = h_1(t)\, h_1(\Delta t) + h_2(t) \cdot h_0(\Delta t) + o(\Delta t)$$

$$\mathbb{P}(X_t = 15)$$



$$\mathbb{P}(X_t = k) = \frac{e^{-\lambda t} \cdot (\lambda t)^k}{k!}$$

$$X_t \sim \text{Poiss}(\lambda t)$$

$$\mathbb{E}(X_t) = \lambda t \qquad \text{Var}(X_t) = \lambda \cdot t$$

$\mathbb{E}(X_1) = \lambda$

Y_n – время между событиями

$X_{\Delta t}$ нез $X_{\Delta S}$ $\boxed{PA-2}$

$F_{Y_0}(t) = \mathbb{P}(Y_0 \leq t) = 1 - \mathbb{P}(Y_0 > t) = 1 - e^{-\lambda t}$

$\mathbb{P}(X_t = 0)$

$e^{-\lambda t}$

$Y_0 \sim Exp(\lambda)$

$\mathbb{E}(Y_0) = \frac{1}{\lambda}$

$Var(Y_0) = \frac{1}{\lambda^2}$

Упражнение (хозяйки кассы)

тётя Зина 5 кл./час

тётя Галя 7 кл./час

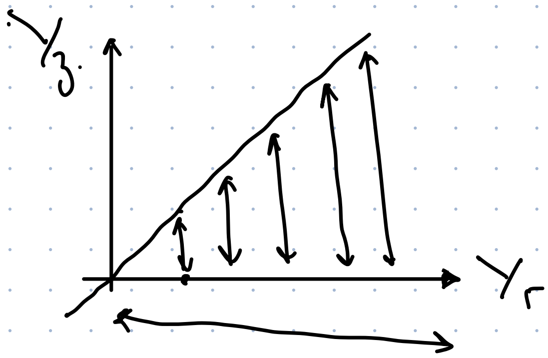
а) Вер. того, что Зина продаст быстрее Гали

$Y_{з.} \sim Exp(5)$ $Y_{г} \sim Exp(7)$

$\mathbb{P}(Y_{з.} < Y_{г}) - ?$

$f(t_{з.}, t_{г.}) = f(t_{з.}) \cdot f(t_{г.})$

Путь „простушки“

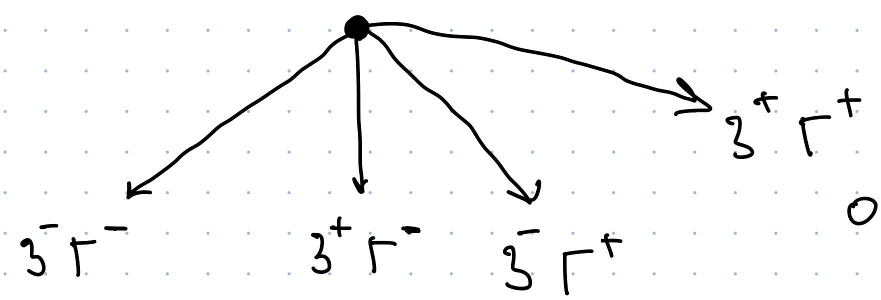


$$\int_0^{+\infty}\int_0^{t_г} 5e^{-5t_з}\, 7e^{-7t_г}\, dt_з\, dt_г$$

Путь „хитрюшки“ ~ метод первого шага

Δt – маленький

$X_{\Delta t}$ – число происшествий за Δt



$З^+Г^+$ $\quad (5\Delta + o(\Delta))(7\Delta + o(\Delta)) =$

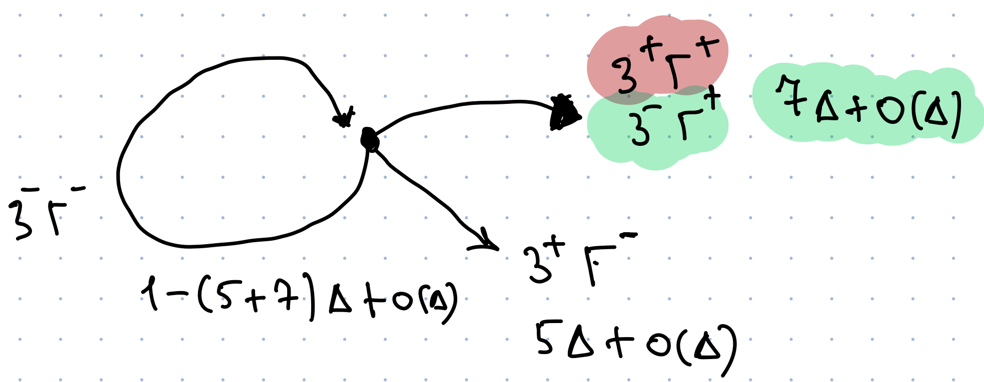
$$= \underbrace{35\Delta^2}_{o(\Delta)} + \underbrace{5\Delta\, o(\Delta)}_{o(\Delta)} + \underbrace{7\Delta\, o(\Delta)}_{o(\Delta)} + \underbrace{o(\Delta^2)}_{o(\Delta)} = o(\Delta)$$

$З^-Г^+$ $\quad (1 - 5\Delta + o(\Delta))(7\Delta + o(\Delta)) = 7\Delta$

$З^+Г^- = 5\Delta$ $\qquad$ $З^-Г^- = (1-5\Delta+o(\Delta))(1-7\Delta+o(\Delta)) =$

$= 1-5\Delta-7\Delta = 1-12\Delta$

$$\mathbb{P}(Y_з < Y_г) = 5\Delta + (1-(5+7)\Delta)\,\mathbb{P}(Y_з < Y_г) + o$$



$$p = 5\Delta + (1-(5+7)\Delta)\,p$$

$p = \dfrac{5}{5+7}$ $\qquad$ $\mathbb{P}(Y_з < Y_г) = \dfrac{\lambda_3}{\lambda_3+\lambda_2}$

8) Как распределено время обслуживания того клиента, который уйдет первым

$$F_{Z}(t) = \mathbb{P}(Z \le t) = \mathbb{P}(\min(Y_з, Y_2) \le t) =$$

$$= 1-(1-F_з(t))(1-F_2(t)) = 1 - e^{-(\lambda_з+\lambda_2)t}$$

$$Z_1 = \min(Y_z, Y_2) \sim \mathrm{Exp}(\lambda_z + \lambda_2)$$

$$\left.\begin{array}{l} X_t^A \sim \mathrm{Poiss}(\lambda_A) \\ X_t^B \sim \mathrm{Poiss}(\lambda_B) \end{array}\right| \text{kez.}$$

$$X_t = X_t^A + X_t^B \sim \mathrm{Poiss}(\lambda_A + \lambda_B)$$