Teoria probabilităților – refresher

$A, B, C \dots$ – evenimente

$P(A)$ – probabilitatea ca evenimentul A să se întâmple

$0 \leq P(A) \leq 1$

$P(\bar{A}) = 1 - P(A)$ – probabilitatea inversă

A și B → notăm $P(A \cdot B)$

$P(A \cdot B) = P(A|B) \cdot P(B)$

Dacă A și B nu sunt legate cauzal:

$P(A \cdot B) = P(A) \cdot P(B)$

A sau B → $P(A + B)$

$P(A+B) = P(A) + P(B) - P(A \cdot B)$

Dacă A și B se exclud mutual:

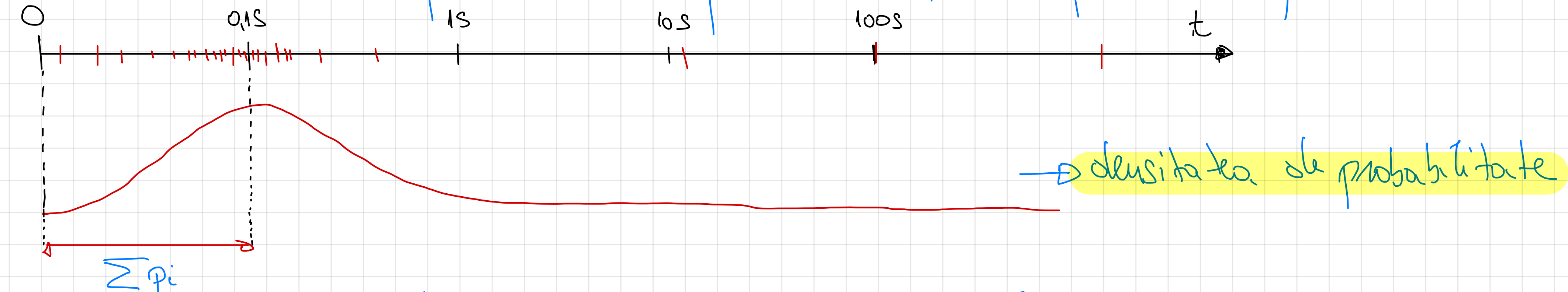
$P(A+B) = P(A) + P(B)$

Distribuții de probabilitate

Variabile aleatoare : X, Y, Z ... în spațiu de reprezentare < discret / continuu

Ex: timpul de răspuns al unui search engine → X

Dacă desenăm pe axa timpului toți timpii de răspuns :



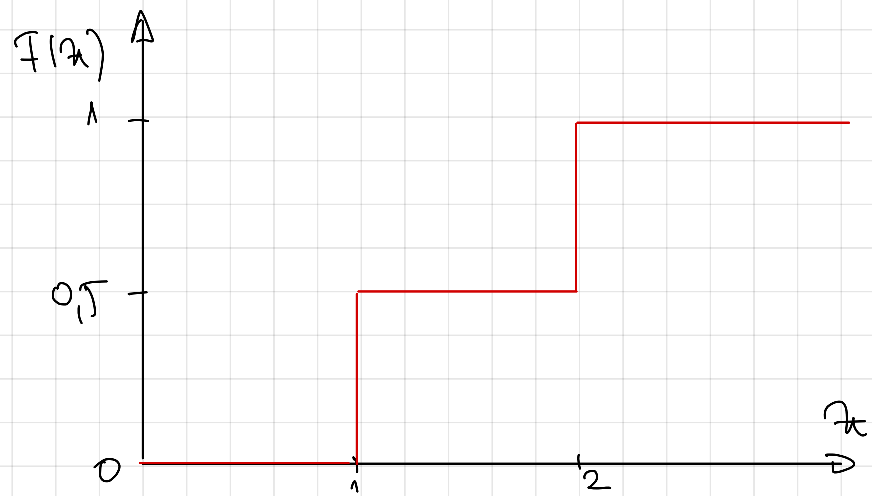
Care este probabilitatea ca timpul de răspuns să fie ≤ 0,1 s ?

Sumă de probabilități → funcție cumulativă de distribuție a prob. (CDF)

CDF : $F_X(x) = P(X \leq x)$

Ex: Aruncarea unei monede

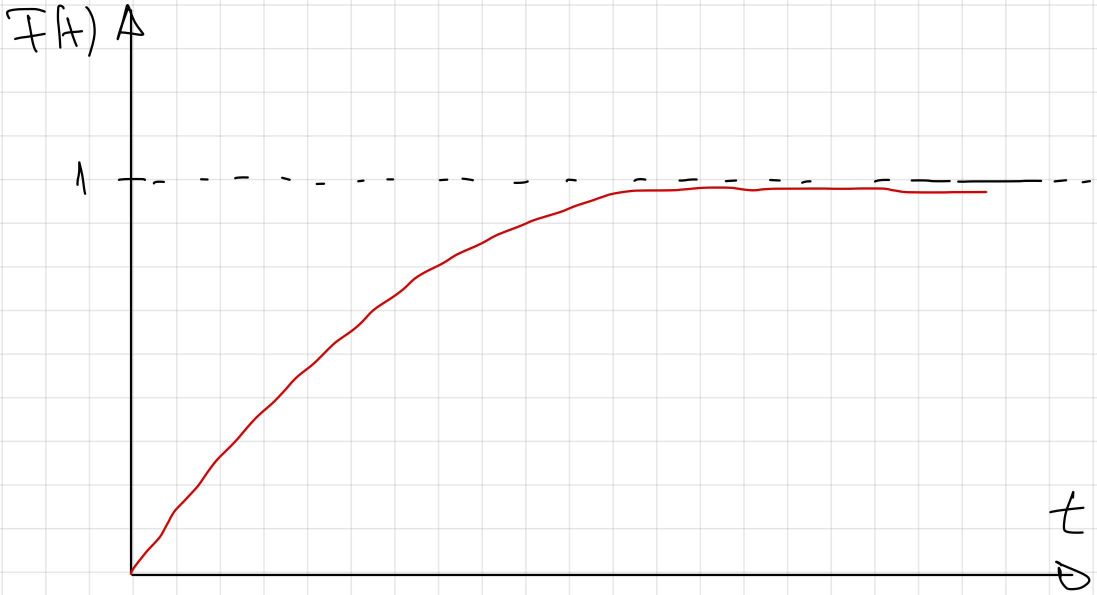
$$F(x) = \begin{cases} 0, & x < 1 \\ 0,5, & x \in [1, 2) \\ 1, & x > 2 \end{cases}$$



Este relevant să folosim timpul ca variabilă aleatoare

$F(x) = F(t)$, $t \in [0, \infty)$

Proprietăți CDF:



$0 \leq F(t) \leq 1$

$F(0) = 0$

$\lim_{t \to \infty} F(t) = 1$

$F(t)$ este monoton crescătoare

Ex: $F(t) = 1 + 2e^{-3t} - 3e^{-2t}$

Funcția de densitate de probabilitate PDF

Notată $f(x)$ – probabilitatea ca un eveniment să se petreacă în $[t, t+\Delta t]$

Proprietate evidentă: $\int_0^\infty f(x)\,dx = 1$ și $f(x) = \frac{dF(x)}{dx}$

$$F(t) = \int_0^t f(z)\,dz$$

$$P(X \geq t) = \int_t^\infty f(z)\,dz$$

$$P(a \leq X \leq b) = \int_a^b f(z)\,dz = F(b) - F(a)$$

Ex: $F(t) = 1 + 2e^{-3t} - 3e^{-2t} \Rightarrow$

$$f(t) = -6e^{-3t} + 6e^{-2t}$$

Valoarea așteptată

$$E[X] = p_1 x_1 + p_2 x_2 + \dots p_n x_n = \frac{p_1 x_1 + \dots p_n x_n}{1} = \frac{p_1 x_1 + p_2 x_2 + \dots p_n x_n}{p_1 + p_2 + \dots p_n}$$

Ex: Pentru un zar cu 6 fețe:

$$E[X] = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \dots + \frac{1}{6} \cdot 6 = 3,5$$

Dacă avem ca variabilă aleatoare timpul (var. continuă):

$$\boxed{E[t] = \int_0^\infty t \cdot f(t)\, dt, \ \forall\, t \geq 0}$$

Ex: $f(t) = -6e^{-3t} + 6e^{-2t} \Rightarrow E[t] = \int_0^\infty t \cdot \left(-6e^{-3t} + 6e^{-2t}\right) dt =$

$$= -6\int_0^\infty t e^{-3t}\, dt + 6\int_0^\infty t e^{-2t}\, dt = \dots = \frac{5}{6}$$

Distribuții de probabilitate

în funcție de tipul variabilei aleatoare — discrete / continue

Distribuția binomială

Funcția de probabilitate a distribuției: $p(x)$

$$p(x) = C_n^x \, p^x (1-p)^{n-x}$$

Exemplu: Rată de defect la becuri 5%. Dacă inspectezi 100 de becuri, care e prob. să găsești 2 becuri defecte?

$$P(x=2) = C_{100}^{2} \, 0{,}05^2 (1-0{,}05)^{98} \simeq 0{,}081 \quad (8{,}1\%)$$

Distribuția Poisson

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$, unde λ este rata de apariție a unui eveniment

$$\lambda = p \times n$$

Exemplu: pentru ex. cu becurile o să avem $\lambda = 100 \cdot 0{,}05 = 5$

$$P(x=2) = \frac{5^2 e^{-5}}{2!} = \ldots = 0{,}09 \ (9\%)$$

Distribuția normală (gaussiană)

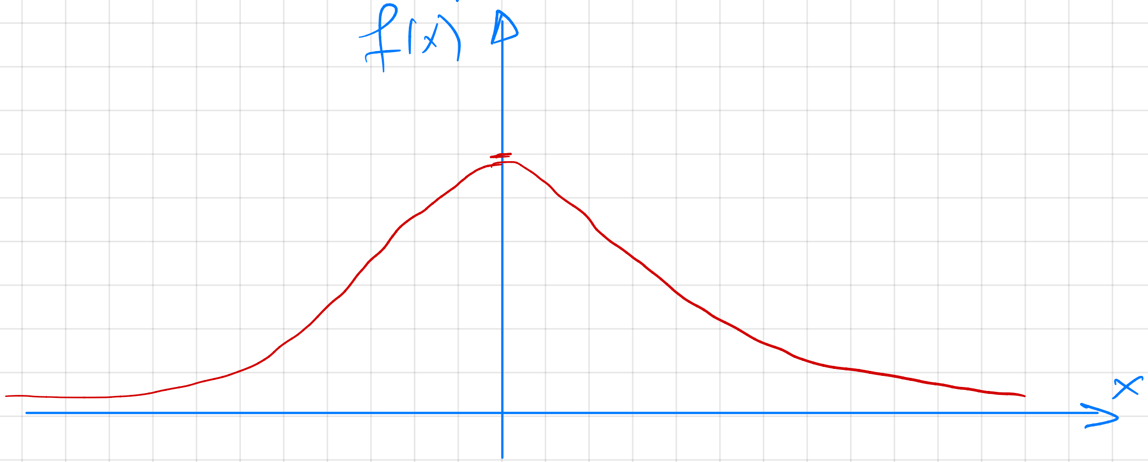
Distribuție continuă $\to$ funcție densitate probabilitate

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$, unde σ – deviația standard

μ – medie

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n-1}}$$

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n}$$

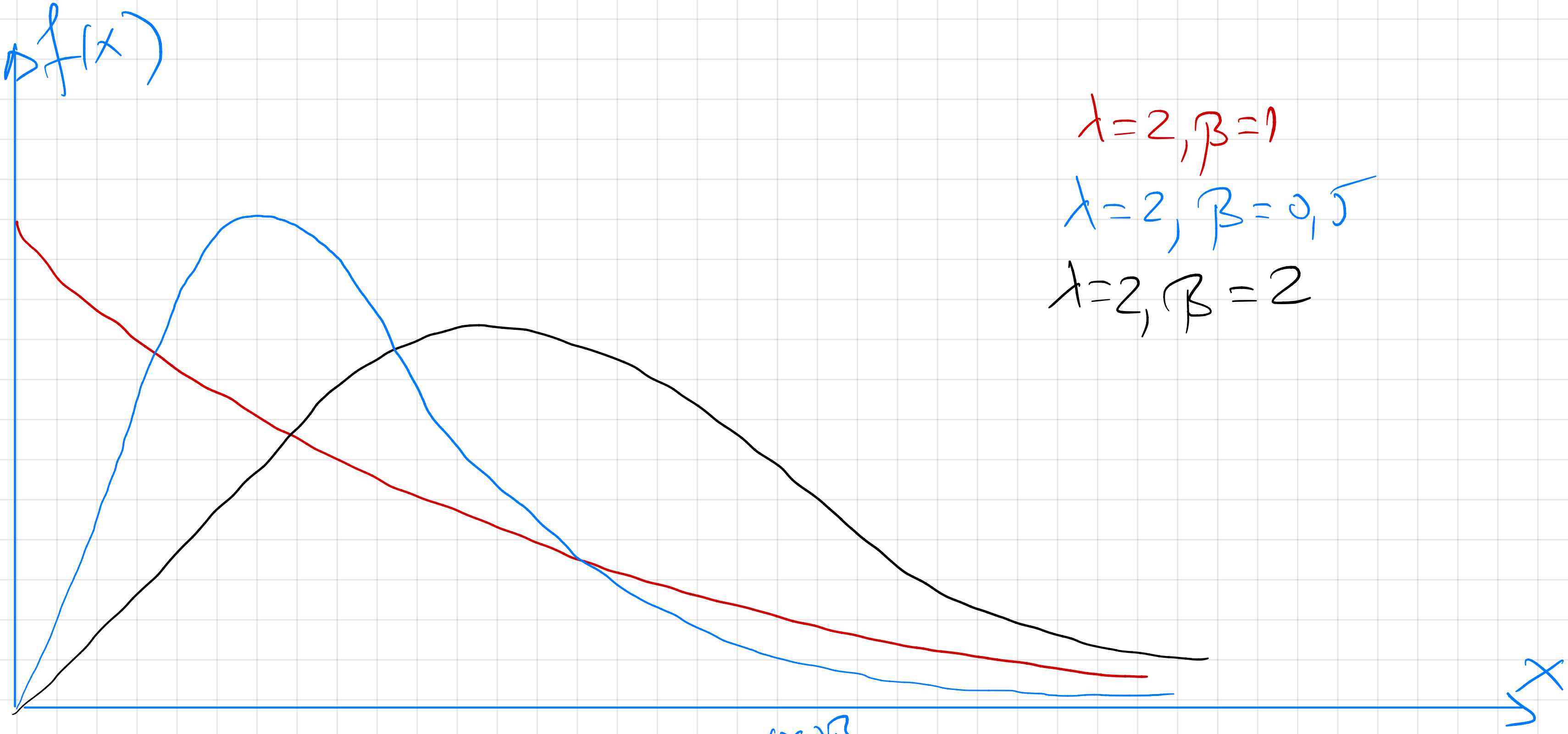


Distribuția Weibull

$$f(x) = \frac{\beta}{\lambda}\left(\frac{x}{\lambda}\right)^{\beta-1} e^{-\left(\frac{x}{\lambda}\right)^{\beta}},$$ unde

$\lambda \rightarrow$ parametrul de scală

$\beta \rightarrow$ parametrul de formă

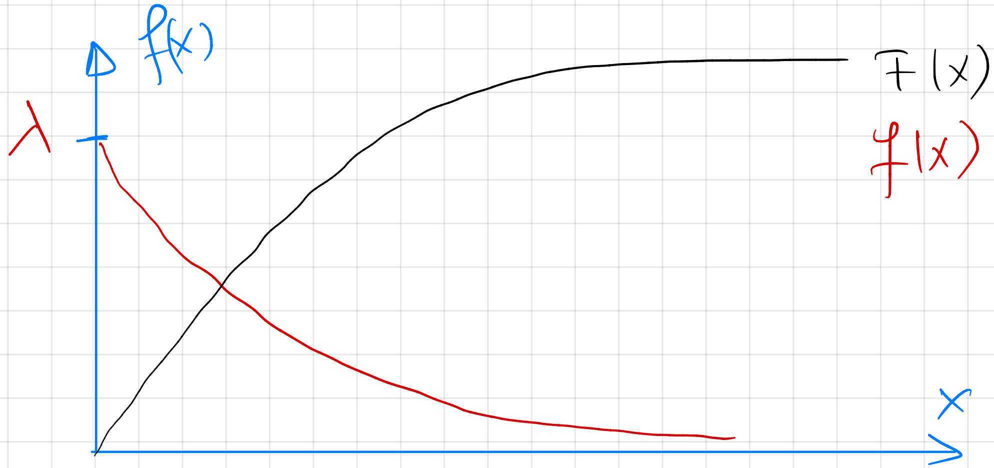


$$F(x) = \int f(x)\,dx = 1 - e^{-\left(\frac{x}{\lambda}\right)^{\beta}}$$

Distribuția exponențială

$$f(x) = \lambda e^{-\lambda x}$$ → Caz special al Weibull cu $\beta = 1$

$$F(x) = 1 - e^{-\lambda x}$$



Exemplu: server web, cererile vin aleatoriu. Avem o rată medie de 10 cereri pe minut ($\lambda = 10$).

Probabilitatea să primești o cerere în următorul minut:

$$F(1) = 1 - e^{-\lambda \cdot 1} = 1 - e^{-10} = 0{,}9995 \ (99{,}95\%)$$

$$f(1) = \lambda e^{-\lambda \cdot 1} = 10 e^{-10} \simeq 0{,}0005$$ → prob. cerere la minutul 1 (mică)

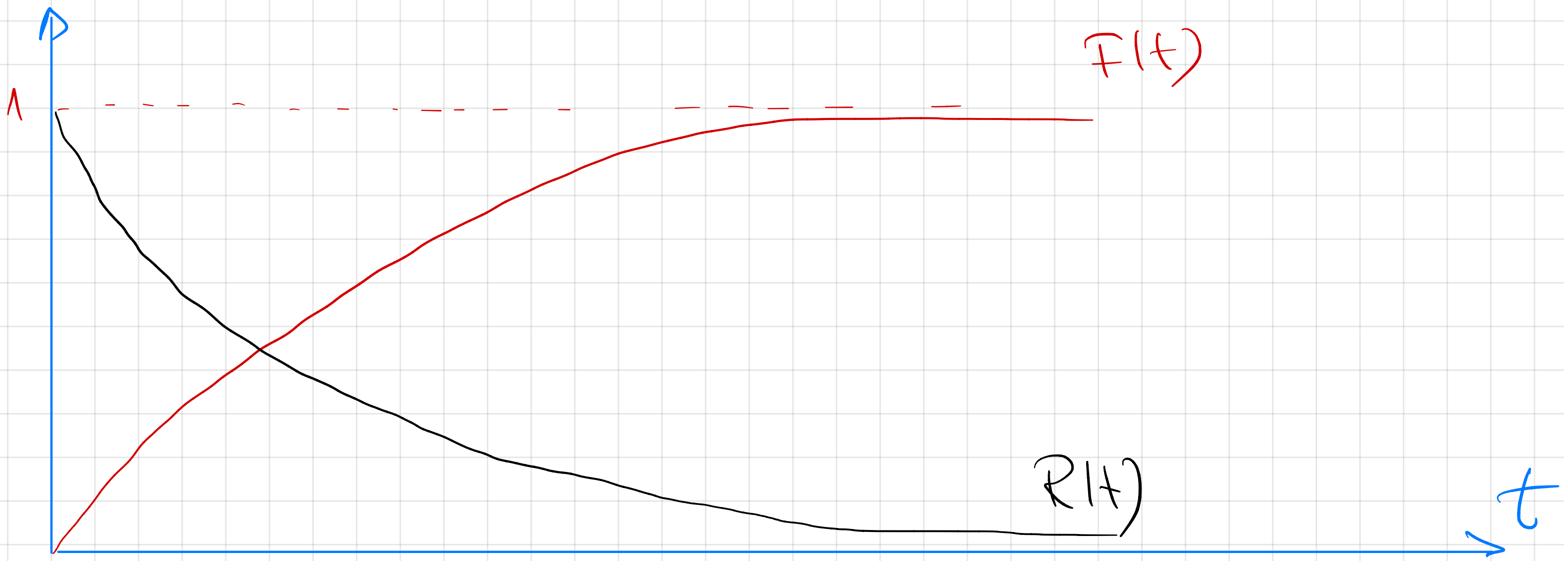
Modelarea fiabilității

Definim o funcție de **fiabilitate $R(t)$**

$R(t) = P($ sistemul este funcțional în $[0,t)$ $)$

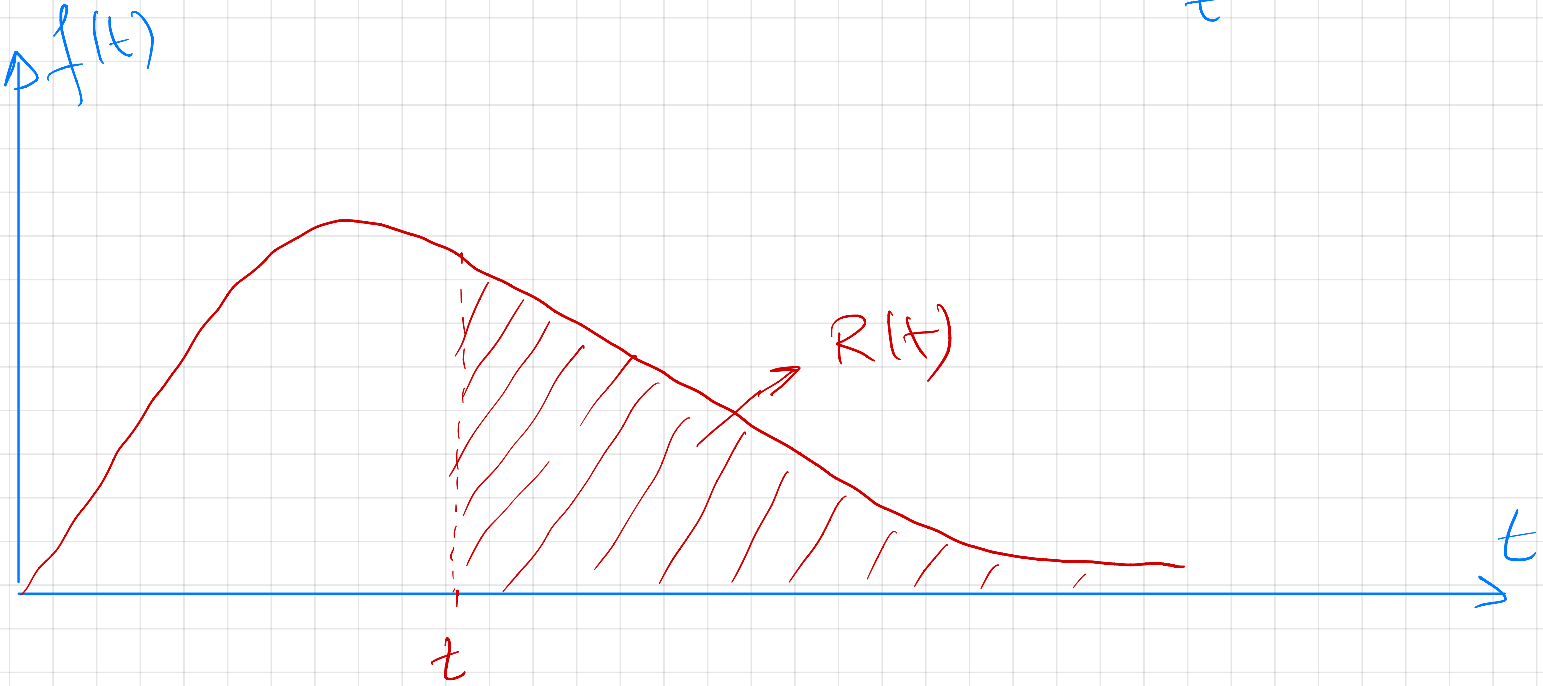
$F(t) \to$ CDF (prob. să apară un defect în $[0,t]$)

$$\left.\begin{array}{l} R(t) = P(X > t) \\ F(t) = P(X \le t) \end{array}\right| \Rightarrow R(t) = 1 - F(t)$$



Reprezentarea grafică a fiabilității

$$F(t) = \int_0^t f(\tau)\,d\tau \Rightarrow R(t) = \int_t^{\infty} f(\tau)\,d\tau$$



Funcția de fiabilitate poate fi interpretată ca aria de sub graficul funcției de densitate de probabilitate, de la un moment de timp la ∞.

Intensitatea defecțiunilor (failure rate)

Probabilitatea ca sistemul să se defecteze în $[t+\Delta t]$ cu condiția ca sistemul să funcționeze până la t.

$$P(t < X < t+\Delta t \mid X > t) = \frac{P(t<X<t+\Delta t)}{P(X>t)} = \frac{F(t+\Delta t)-F(t)}{1-F(t)}$$

Definim intensitatea defecțiunilor:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{F(t+\Delta t)-F(t)}{[1-F(t)]\Delta t} = \lim_{\Delta t \to 0} \frac{R(t)-R(t+\Delta t)}{R(t)\Delta t}$$

dar $R(t) = 1-F(t) \Rightarrow$

$$\lambda(t) = \frac{1}{R(t)} \lim_{\Delta t \to 0} \frac{R(t)-R(t+\Delta t)}{\Delta t} = \frac{1}{R(t)}\left[-\frac{dR(t)}{dt}\right] = \frac{f(t)}{R(t)}$$

Dar $f(t) = \dfrac{dF(t)}{dt} = \dfrac{d(1-R(t))}{dt} = -\dfrac{dR(t)}{dt}$

Deci $\boxed{\lambda(t) = \dfrac{f(t)}{R(t)}}$, sau $\boxed{\lambda(t) = -\dfrac{1}{R(t)} \dfrac{dR(t)}{dt}}$

Curba cadă de baie
(bath-tub curve)

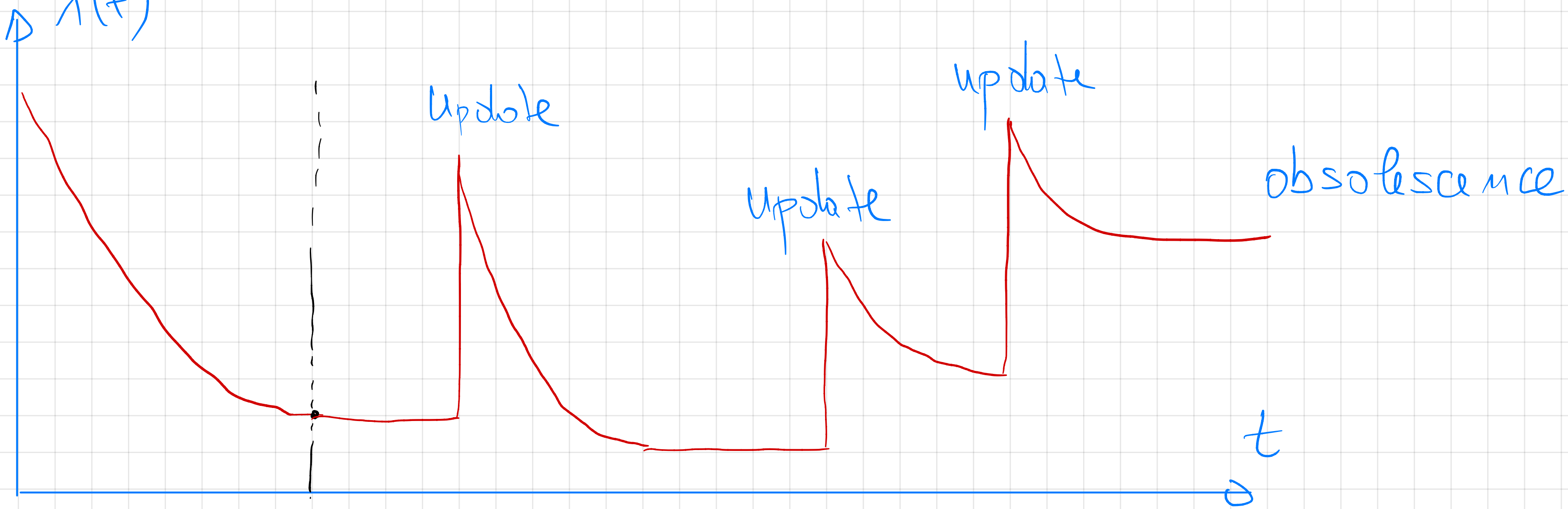
Zone distincte în profilul $\lambda(t)$:

I. Mortalitate infantilă

II. Zona de utilizare normală a produsului ($\lambda = ct$)

III. Îmbătrânire (wear-out)

Intensitatea defecțiunilor pentru software



Dacă $\lambda(t) = ct. = \lambda \Rightarrow \lambda(t) = \dfrac{f(t)}{R(t)} = -\dfrac{1}{R(t)} \dfrac{dR(t)}{dt} \Rightarrow$

$\Rightarrow \lambda = -\dfrac{1}{R(t)} \dfrac{dR(t)}{dt} \Rightarrow -\lambda\, dt = \dfrac{1}{R(t)} dR(t) \;\Big|\cdot\int \Rightarrow -\lambda \int dt = \int \dfrac{1}{R(t)} dR(t)$

$\Rightarrow -\lambda(t+c) = \ln(R(t)) \Rightarrow R(t) = e^{-\lambda(t+c)}$

Dar $R(0) = 1 \Rightarrow e^{-\lambda(0+c)} = 1 \Rightarrow e^{-\lambda \cdot c} = 1 \Big|_{\lambda \neq 0} \Rightarrow c = 0 \Rightarrow \boxed{R(t) = e^{-\lambda t}}$

Dacă $\lambda(t)$ nu este constant, modelăm folosind distribuția Weibull

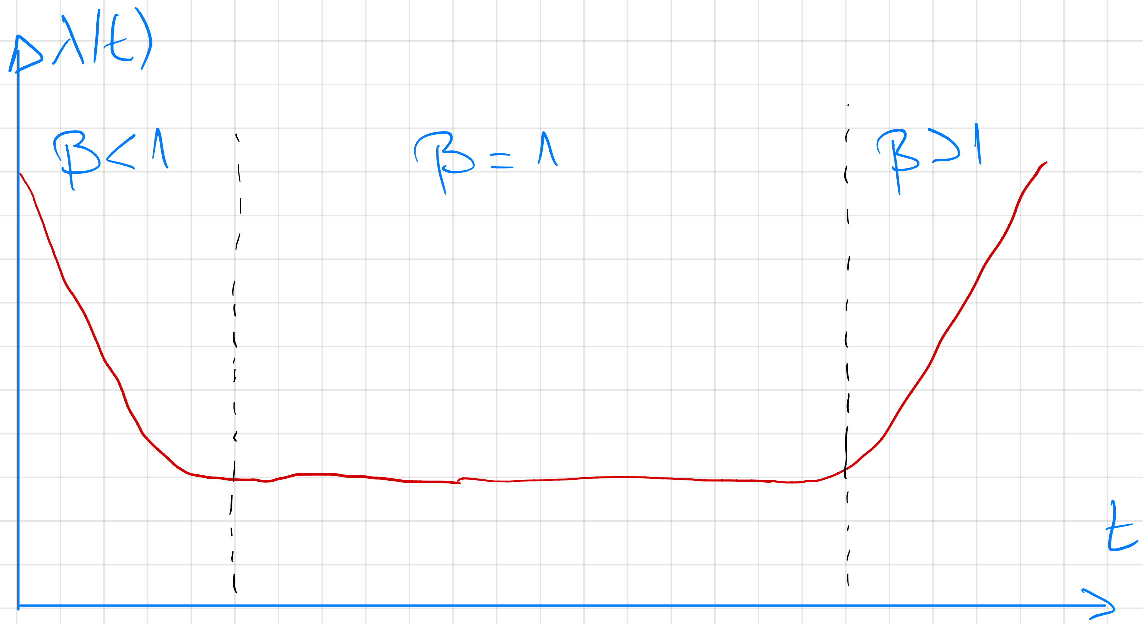
$$\lambda(t) = \lambda \beta t^{\beta-1}$$

$\beta > 1$: $\lambda(t) \nearrow$ — îmbătrânire

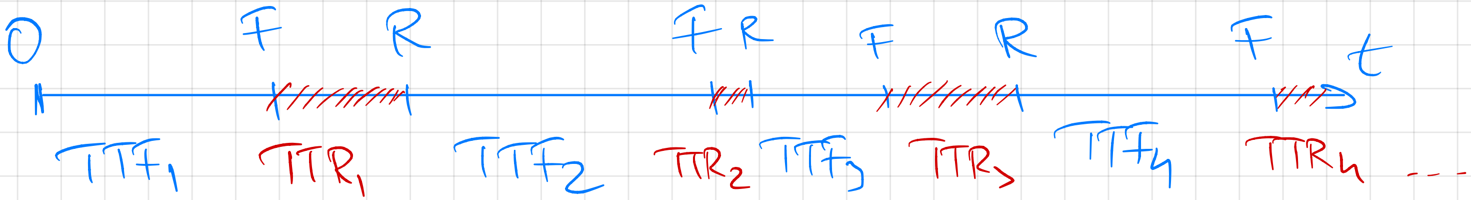
$\beta = 1$: $\lambda(t) = ct = \lambda$

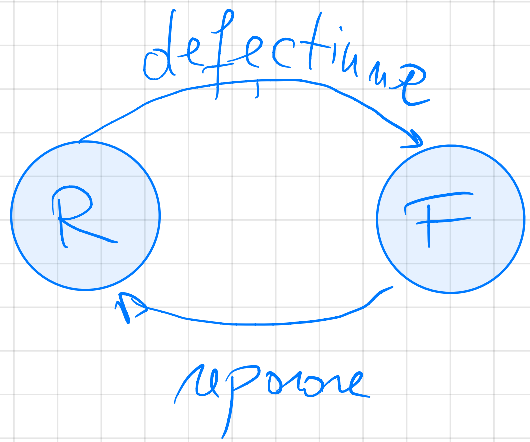
$\beta < 1$: $\lambda(t) \downarrow$ — mortalitate infantilă

$$R(t) = e^{-\lambda t^{\beta}}$$



Media timpului de bună funcționare





$$MTBF = \frac{\sum TTF_i}{n} \qquad MTTR = \frac{\sum TTR_i}{n}$$

A - disponibilitate

$$A = \frac{\sum TTF_i}{\sum TTF_i + \sum TTR_i} = \frac{MTBF}{MTBF + MTTR}$$

Exemplu:

MTBF = 2 ore (7200 s), MTTR = 3 s

$$A = \frac{7200}{7200+3} = 99{,}96\%$$

Exemplu:

Server email: 99,99% disponibilitate 4 de 9 → down 52,56 min/an

Sisteme critice: 99,9999%, 6 de 9, va fi indisponibil 31,5 s/an