Forgalmi modellezés, szimuláció

Mészáros András

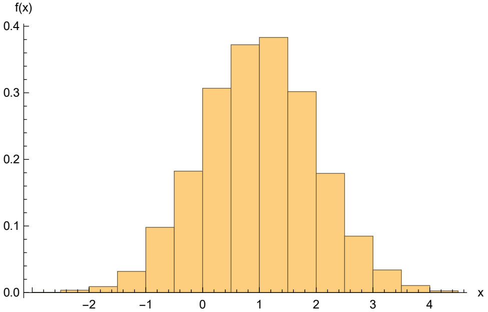
2020. október 19.

Áttekintés

Tartalom

Val. változó megadása

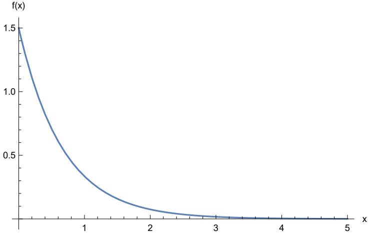
- Valószínűségi változó X: véletlentől függő érték (csak az eloszlása ismert)
- Megadása
  - eloszlásfüggvény: $F(x) = Pr(X < x)$
  - sűrűségfüggvény: $f(x) = \lim_{\Delta \to 0} \frac{Pr(x<X<x+\Delta)}{\Delta}$
    valóságban nem ismerjük→hisztogram mintákból



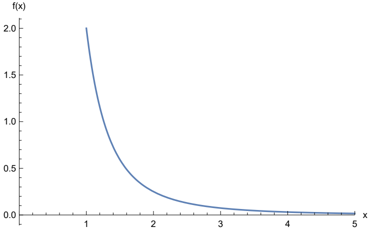
Nevezetes eloszlások

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

Pareto eloszlás - nehézfarkú
$F(x) = 1 - \left(\frac{c}{x}\right)^a$



(a) Exp. $\lambda = 1.5$



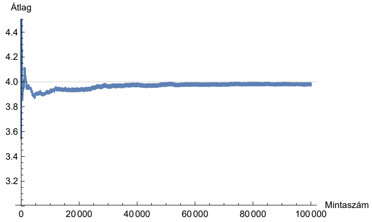
(b) Pareto $c = 1, a = 2$

Val. változók jellemzői I.

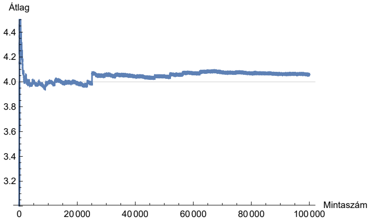
- várható érték: $\mu = E[X]$
- szórásnégyzet: $\sigma^2 = E[X^2] - E[X]^2$
- relatív szórás: $CV(X) = \frac{\sigma}{E[X]}$ - "átlagos relatív eltérés"
- farokeloszlás (nagy értékek valószínűségének lecsengése)
  könnyűfarkú - exponenciális
  nehézfarkú pl. $Pr(X > x) \sim x^{-2}$ - a nagy értékek ritkák, de...
  pl. flow* hossza hálózati kommunikációban

Val. változók jellemzői II.

- Nehézfarkú: átlag sokkal lassabban konvergál, CHT nem alkalmazható*



(a) Exp. átlag ($\mu = 4$)



(b) Pareto átlag ($\mu = 4$)

Véletlen pontfolyamatok

- $X_1, X_2, \ldots$ - pl. csomag érkezési időközök
- Általános megadás együttes sűrűségfüggvénnyel $Pr(X_1" = "x_1, X_2" = "x_2, \ldots)$
- Stat. jellemző: X_i eloszlása, autokorreláció, (együttes momentumok)
  - korreláció: $\rho(X, Y) = \frac{E[(X_k - E[X])(Y_k - E[Y])]}{\sigma_X \sigma_Y}$
  - autokorreláció: $\rho_i = \frac{E[(X_k - \mu)(X_{k+i} - \mu)]}{\sigma^2}$
  - $-1 \leq \rho_i \leq 1$
  - $\rho_i > 0$: "ha egy minta kicsi az utána i-vel jövő minta is kicsi"
- Feltételezzük, hogy a folyamat stacionárius = "nem változik" - reális?

(Hálózat) méretezés független igényekre I.

- Egy felhasználó, erőforrás (pl. sávszélesség) igénye X: mekkora kapacitás kell? pl. $Pr(X < K) > 0.99$
- Egyszerűsített méretezés: $E[X]\,(1 + \mathcal{R}(CV(X))$
- Mi történik sok felhasználónál? pl. egyetemi hálózat gerincvezetéke
- $Z = X + Y \Rightarrow E[Z]\,(1 + \mathcal{R}(CV(Z)))$

(Hálózat) méretezés független igényekre II.

- ha X és Y független(!!!):

$$\left.\begin{array}{r}E[X+Y]=E[X]+E[Y]\\ \sigma^2(X+Y)=\sigma^2(X)+\sigma^2(Y)\end{array}\right\}\Rightarrow CV(X+Y)=\frac{\sqrt{\sigma^2(X)+\sigma^2(Y)}}{E[X]+E[Y]}$$

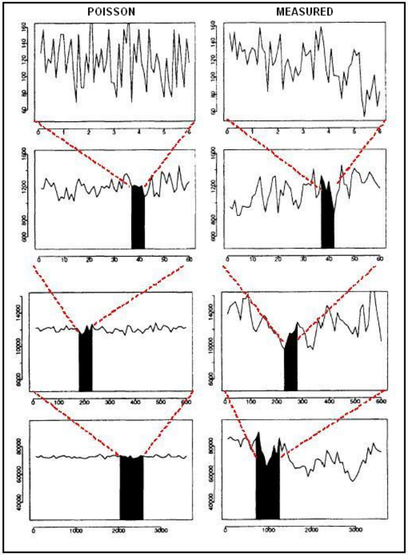
$CV(X+Y)\le \max(CV(X),CV(Y))$

pl. azonos eloszlásnál $CV(Z)=\sqrt{0.5}CV(X)=\sqrt{0.5}CV(Y)$

- Az aggregáció csökkenti a relatív szórást - kisebb túlméretezés

(Hálózat) méretezés összefüggő igényekre

- Klasszikus telefonhálózat - a hívások közti függetlenség jó feltételezés (hívás hossza exp. eloszlású)
- Internet - flow-k közti függetlenség nem teljesül - pozitív korreláció (flow hossza nehézfarkú)
  $Z = X + Y \Rightarrow E[Z]\,(1 + \mathcal{R}(CV(Z)))$
- Az aggregáció nem csökkenti annyira a relatív szórást



POISSON
MEASURED

Tartalom

Hálózati forgalom mérése

- Miért mérünk?
  - trendek megértése
  - bemenet szimulációnak
  - SLA* készítés
  - anomália észlelése (támadás, rendellenesség)
  - tervezési ökölszabályok
- Aktív mérés
  - throughput, késleltetés, jitter, csomagvesztés, csomagok átrendeződése
  - két pont között
  - hálózati teljesítmény - QoS, QoE
- Passzív mérés
  - forgalom mennyisége/összetétele és változása
  - adatkezelési vetület (pl. anonimizáció)

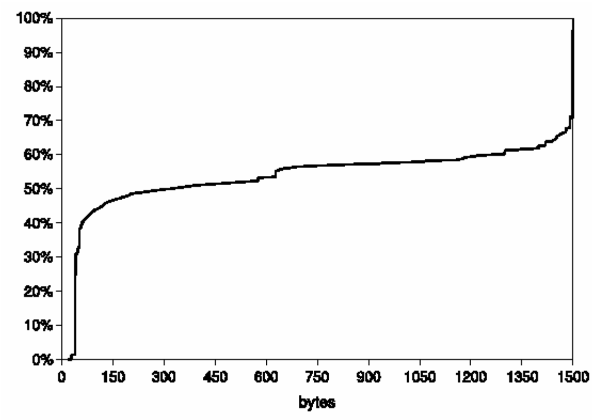
Forgalmi jellemzők

- Késleltetés és jitter
  - jelterjedés $\approx 2 * 10^5$km/s - 200ms (oda-vissza)
  - torlódás a csomópontban
  - feldolgozási idő a csomópontban
  - alkalmazás késleltetése
- Csomagvesztés
  - torlódás
  - zaj ($10^{-6} - 10^{-13}$)
  - meghibásodás
  - alkalmazási rétegbeli veszteség

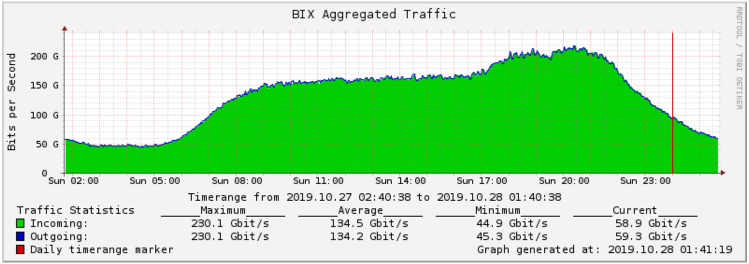
Internetes forgalom jellemzői

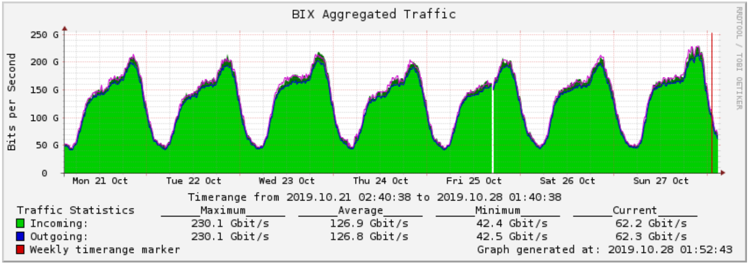
- Napi ciklus
- Heti ciklus (régen: hétvégén alacsonyabb)
- Óránként stacionárius
- Csomagok mérete: $\sim$ 40-50B, 1500B, kb 40-40%
- Sok rövid flow (50% rövidebb mint 2s)
- Kevés hosszú flow (1% hosszabb 15 percnél), de forgalom akár 50%-a
- A forgalom helytől is függ
- (Egyéb hálózatok forgalma eltérő lehet pl. szenzorhálózat)

Csomagméret eloszlása



Magyarországi forgalom





Erőforrás menedzsment

- Hogy biztosítjuk a szolgáltatás minőségét?
- Erőforrás megosztás - ütemezés
- Prioritásos kiszolgálás
  - nagy prioritás - kis késleltetés
  - kiéheztetés
- Súlyozott megosztás
  - nincs kiéheztetés
  - "prioritás" terhelés függő

Tartalom

Rendszervizsgálat

- Rendszer elemzése
  - Valós rendszer tesztelése - drága, élesben módosítani... (pl. QUIC)
  - Testbed - valós elemek (pl. egy routeren erőforrás megosztási algoritmus)
  - Szimuláció
  - Analitikus modell (pl. klasszikus telefonhálózat méretezése)
- Szimuláció
  - speciális vagy általános programnyelven írt program
  - modell ami a fő jellemzőket megragadja - mik azok?
  - sok paraméter - nehéz hangolni

Adatmodellek

- Adat generálására szolgáló modell (pl. forgalom modell)
- Konstruktív (white box) - rendszerből/viselkedésből modell
  - könnyen interpretálható
  - gyakran nem pontos
- Leíró (black box) - adatból modell
  - jól illeszkedik az adatra
  - rendszer változik $\Rightarrow$ újra adat kell
  - nincs interpretáció

Forgalom modell

Felépítése

- Session (pl. FTP szerverre bejelentkezés) - hálózattól független exponenciális időközű
- Flow* - csomagok sorozata (pl. FTP szerverről fájl letöltése) nehézfarkú hosszúságú, alkalmazásfüggő
- Csomag
  követési idő hálózat és protokollfüggő, mérete alkalmazásfüggő

A szimuláció tulajdonságai

Előnyök

- kevés megkötés (analitikus/testbed korlátok)
- nagy és összetett rendszer vizsgálható vele
- elrendezések sora vizsgálható megvalósítás előtt

Hátrányok

- pontos modell kialakítása nehéz és költséges lehet
- mindig csak közelítés marad
- ritka esemény nehezen szimulálható

Szimuláció lépései

- rendszer elemeinek sztochasztikus modellje - nehéz feladat
- realizációk generálása - mennyit? milyen hosszút?
- adatgyűjtés - mit mérjünk?
- adatok kiértékelése - mik a fontos jellemzők? tanulságok?

Szimulátorok fajtái

- Diszkrét idejű (DES - discrete event simulator)
  - egy lépés egy esemény
  - fix lépésközű ugrás
- Folytonos idejű - állapot folyamatosan változik (diffegyenlet)
- Trace driven - a bemenet egy konkrét adatsor (pl. csomagérkezések - prioritizálás hatása?)
- Monte Carlo szimuláció - ismételt véletlen minták
  (pl. mi a valószínűsége, hogy ad-hoc hálózatban két elem közt van kommunikációs út)

Véletlen számok generálása

- A legtöbb szimuláció tartalmaz véletlen bemenetet
- Valódi véletlen pl. termikus zaj
- Álvéletlen - generátor megadott seeddel - reprodukálható
- Legtöbb nyelv tartalmaz álvéletlen generátort - minőség számít
  pl. C srand() - Mersenne Twister

- Tetszőleges eloszlású véletlen szám generálása
  $y = F(x) \rightarrow x = F^{-1}(y)$, ahol y [0,1]-en egyenletes eloszlású
  pl. $y = 1 - e^{-\lambda x} \rightarrow x = \frac{ln(1-y)}{\lambda}$

Szimuláció eredményének megbízhatósága

- Több szimuláció eredményét átlagoljuk, $X - \mu_X, \sigma_X$
  $\bar{x}(n) = \frac{x_1 + \ldots + x_n}{n}$ - $E[\bar{x}] = \mu = E[X]$
  $\sigma^2(\bar{x}(n)) = \frac{\sigma_X^2}{n}$ - ha létezik!
- CHT: $n \to \infty$ esetén $\bar{x}(n)$ normális eloszlású μ várható értékkel és $\frac{\sigma}{\sqrt{n}}$ szórással $\frac{\bar{x}(n) - \mu}{\sigma(\bar{x}(n))} \sim N(0, 1)$
- $Pr(|\bar{x}(n) - \mu| < \Delta) = 1 - \alpha$ (pl. $\alpha = 0.05$)
  95% valószínűséggel a valódi várható értéktől való eltérés kisebb Δ-nál $\Delta = u_{\alpha/2}\sigma(\bar{x}(n))$

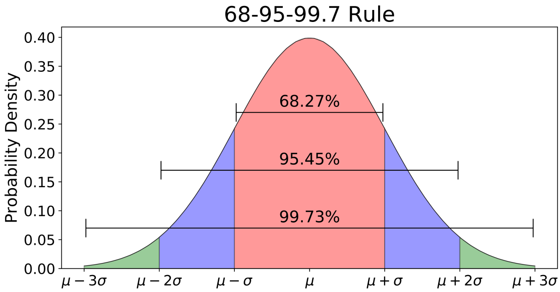
Szimuláció eredményének megbízhatósága - példa I.

Mi a valószínűsége, hogy ad-hoc hálózatban két elem közt van kommunikációs út?

$n = 100$ futtatásra $\bar{x}(100) = 0.63$ valószínűséggel van, az átlag tapasztalati szórása $\hat{\sigma}(\bar{x}(n)) = 0.06$

Szeretnénk 95%-os biztonsággal megmondani milyen tartományban lehet a valószínűség

Szimuláció eredményének megbízhatósága - példa II.



- $\Delta = u_{\alpha/2}\sigma(\bar{x}(n))$
- $u_{\alpha/2} \approx 2$, tehát $\Delta = 2 \times 0.06 = 0.12$
- A valószínűség 95% eséllyel a $[\bar{x}(100) - \Delta, \bar{x}(100) + \Delta]$=[0.51,0.75] tartományban van

Szimuláció hossza

- A stacionárius szakasz domináljon
- megfelelően hosszú (=?)
- inicializálás (pl. étterem foglaltsága)
- csonkolás pl. monoton szakasz levágása
- szimuláció elejének levágása - átlag nem változik (jelentősen)
- ablakozás - átlag az ablakon belül nem változik

Témalabor témák

- Prioritásos kiszolgálás hatása a szolgáltatás minőségére
  - Prioritásos kiszolgálás szimulációja
  - Mérés és programozás valódi eszközökön (router ütemezőjének programozása)
- Információ terjedésének szimulációja VANET (vehicular ad-hoc network) hálózatban
  - pl. baleset az úton - információ visszafelé küldése - milyen gyorsan terjed?