Departament d'Estadística i Investigació Operativa
UNIVERSITAT POLITÈCNICA DE CATALUNYA

Models de VA i simulació
Exercicis

Bloc B – Probabilitat i Estadística

2024

Índex

Exemple Binomial. Explicació funció de probabilitat

Suposem la VAD X: “número de uns en 4 tirades d’un dau” ~ B(n=4, p=1/6)

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k} \quad amb \quad k = 0, 1, \dots, n$$

resultat				X	probabilitat			
0	0	0	0	0	q	q	q	q
0	0	0	1	1	q	q	q	p
0	0	1	0	1	q	q	p	q
0	0	1	1	2	q	q	p	p
0	1	0	0	1	q	p	q	q
0	1	0	1	2	q	p	q	p
0	1	1	0	2	q	p	p	q
0	1	1	1	3	q	p	p	p
1	0	0	0	1	p	q	q	q
1	0	0	1	2	p	q	q	p
1	0	1	0	2	p	q	p	q
1	0	1	1	3	p	q	p	p
1	1	0	0	2	p	p	q	q
1	1	0	1	3	p	p	q	p
1	1	1	0	3	p	p	p	q
1	1	1	1	4	p	p	p	p

- Contarem 0 i 4, 1 vegada → $\binom{4}{0} = \binom{4}{4} = 1$
- Contarem 1 i 3 → 4 vegades → $\binom{4}{1} = \binom{4}{3} = 4$
- Contarem 2 → 6 vegades → $\binom{4}{2} = 6$
- Els resultats no són equiprobables excepte si p=0.5:
  - Obtenir 4 cares diferents de “1” té prob. de $(5/6)^4$=0.48
  - Obtenir 4 “1” té prob. de $(1/6)^4$=0.00077

Exercici Model Binomial

La cabina de discos d'un servidor conté 18 discs idèntics. Un disc de cada 5 necessita ser substituït al cap d'un any per ser reparat

VAD que compta substitucions:	
R: "número de discs que han de ser substituïts al cap de l'any"	
Model per a la VAD:	R ~ B(n=18, p=0.2)
Número esperat de substitucions anuals:	E(R)=3.6
Variància i desviació típica:	V(R)=2.88 → $\sigma_x = 1.697$
Prob. d'observar 4 casos:	P(R=4)=0.2153
Prob. sols una substitució:	P(R=1) =0.0811
Prob. de tenir-ne menys de 3:	P(R<3)=0.2713
Prob. que almenys 6 discs siguin substituïts:	P(R≥6)=0.1329

Exercici Model Geomètric

La probabilitat d’un disc dur defectuós d’una determinada marca és 0.05 Es durà a terme una inspecció sobre un lot de discs durs.

Quina distribució segueix la VAD X: “Nombre de discs durs a revisar fins a trobar el primer defectuós”?:

X~Geom(p=0.05)

Quina és la probabilitat d’haver-ne de revisar només 1? P(X=1) = 0.05

`dgeom(0,0.05)`

Quina és la probabilitat d’haver-ne de revisar 10? P(X=10) = 0.959 · 0.05 = 0.032

`dgeom(9,0.05)`

Quina és la probabilitat d’haver-ne de revisar més de 2?

P(X>2) = 1 – P(X=1) – P(X=2) = 1 – 0.05 – 0-95*0.05 = 0.9025 `1-pgeom(1,0.05)`

Quin és el nombre esperat de discos que s’hauran de revisar? E(X) = 1/0.05 = 20

Amb quin nombre de discos en el lot podrem estar segurs que trobarem algun defectuós amb una probabilitat almenys del 50% P(X<x) > 0.50 → x= 13

`qgeom(0.50,0.05)`

Exercici Model Binomial negatiu

Per aprovar un determinada assignatura, es requereix treure més d'un 5 en un problema d'estatus almenys 3 vegades. Un alumne concret té una probabilitat invariable de 0.6 de treure més d' un 5.

Quina distribució segueix la VAD X: "Nombre d'execucions per treure 3 notes superiors a 5"?:

X~BN(r = 3, p=0.6)

Quina és la probabilitat d'haver-ne de fer 1 execució?

P(X=1) = 0

Quina és la probabilitat d'haver-ne de fer 10 execucions?

$P(X = 10) = \binom{9}{2} \cdot 0.6^3 \cdot 0.4^7 = 0.013$ (dnbinom(7,3,0.6))

Quin és el nombre esperat de execucions que s'hauran de fer per aprovar?

E(X) = 3/0.6 = 5

Exercici Model Poisson

El centre de càlcul d'una empresa atén les incidències que sorgeixen als treballadors. S'ha observat que aquestes apareixen esporàdicament, encara que l'elevat número d'usuaris implica que el volum de problemes a tractar diàriament sigui considerable (s'ha suposat una mitjana de 2.35 incidències/dia).

Quina és la distribució de la v.a "número d'incidències diàries"?

X: "número d'incidències al dia" → $X \sim P(\lambda=2.35)$

Quina és la seva Esperança? I la seva desviació típus?

$E(X) = 2.35$ incidències i $\sigma_X=1.53$ incidències

Quina és la distribució de la v.a "número d'incidències en 5 dies"? I en 7?

X_5: "número d'incidències 5 dies" → $X_5 \sim P(\lambda=11.75)$; $X_7 \sim P(\lambda=16.45)$

Probabilitat que en un dia es produeixi 3 incidències

$P(X=3) = 0.2063$ `(dpois(3,2.35))`

Probabilitat d'observar menys de 3 incidències en un dia:

$P(X<3) = P(X\leq 2) = 0.5828$ `(ppois(2,2.35))`

Entre el dilluns i el dimarts s'han rebut sis incidències. Quina és la probabilitat que de dilluns a divendres es tractin no més de 15?

$P(X_3\leq 15-6) = P(X_3\leq 9) = 0.8254$ `(ppois(9,3*2.35))`

Quina és la probabilitat que cap dia de la setmana laboral (5 dies) presenti incidències?

$P(X_5=0) = P(X=0)^5 = 0.0000079$ `(dpois(0,5*2.35))` `((dpois(0,2.35))^5)`

Exemples Model Exponencial

Siguin les següents variables aleatòries:

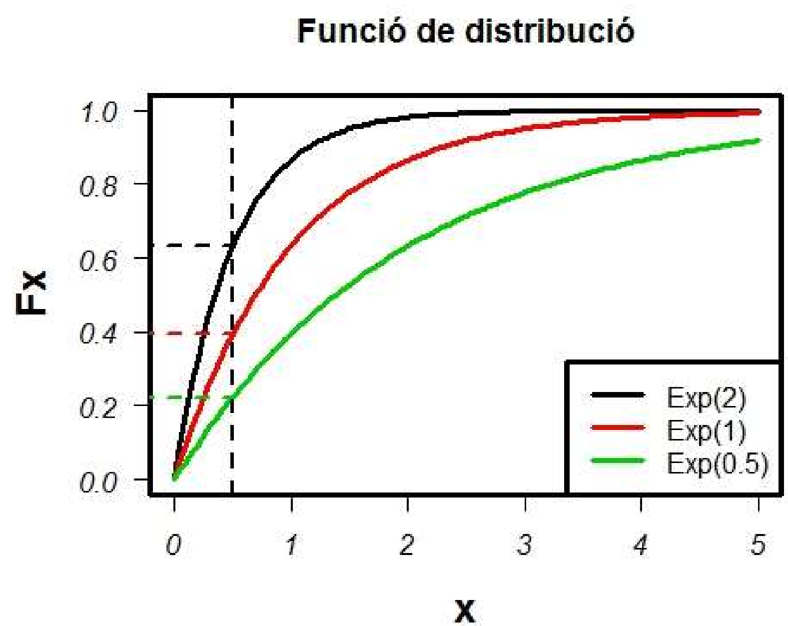
N: “número de peticions/seg a un servidor de BBDD” → N ~ P(λ)

T: temps (seg.) transcorregut entre dos peticions consecutives

Si λ = 2 arr./s. → P(T<0.5) = 0.632

Si λ = 1 arr./s. → P(T<0.5) = 0.393

Si λ = 0.5 arr./s. → P(T<0.5) = 0.221



Nota: Si les arribades són més freqüents (λ alt), el temps entre arribades són més curts. Per tant, és més probable trobar temps per sota de mig segon quan λ és major (*F_X creix més ràpid*).

Aplicacions del Model Exponencial

- **"Failure Rate and Reliability"** (Jane Horgan lecture 17)

  $X \sim Exp(\lambda)$ $F_X(x)=P(X \le x)=1 - e^{-\lambda \cdot x}$ $R_X(x)= P(X > x) = e^{-\lambda \cdot x}$

  – R és la funció de "reliability" o de fiabilitat (probabilitat de durar més de..)

  – λ és "failure rate" o taxa d'error

  – E(X) = 1/λ és "MTTF" o "Mean Time To Failure"

- **"Modelling Response Times: M/M/1"** (Jane Horgan lecture 17)

  – La teoria de cues permet estudiar sistemes en que interaccionen més d'una variable (per exemple temps de resposta en un sistema d'espera amb cues).

El model M/M/1 indica: 1 per contemplar una sola cua no finita, i dos M's pels dos temps d'arribada i servei exponencials (*Exp*(λ) i *Exp*(μ) o número d'arribades **P**(λ) i serveis **P**(μ)), que compleixen la propietat de Markov de no tenir memòria.

Amb l i m es defineix un indicador del sistema com factor de càrrega o "traffic intensity" (λ / μ):
- si λ /μ > 1 (λ > μ) → la taxa d'arribades és superior a la de sortides (sobreutilització del sistema)
- si λ /μ = 1 (λ = μ) →la taxa d'arribades s'iguala amb la taxa de sortides
- si λ /μ < 1 (λ < μ) → la taxa d'arribades és inferior a la de sortides (infrautilització del sistema)

(el model M/M/1 permet un tractament per teoria de cues o per simulació. Models més complexos poden permetre sols el tractament per simulació)

Exercici Model Exponencial i Poisson

El centre de càlcul d'una important empresa atén les incidències que els sorgeixen als treballadors. Se suposa una mitjana de 4 incidències/dia, i 8 hores laborables al dia.

Quina és la distribució de la v.a "número d'incidències diàries"?

X: "número d'incidències al dia" → $X \sim P(\lambda=4)$

Quina és la distribució de la v.a "número d'incidències per hora"?

Xh: "número d'incidències per hora" → $Xh \sim P(\lambda=4/8) = P(\lambda=1/2)$

Probabilitat de rebre 0 incidències en un dia:

$P(X = 0) = 0.0183$

Probabilitat de rebre 0 incidències en una hora:

$P(Xh = 0) = 0.607$

Quina és l'esperança de la variable temps (en hores) entre incidències?

T: "Temps en hores entre incidències" $\sim Exp(\lambda = 1/2)$ → $E(T) = 2$ hores

I la desviació tipus?

$V(T) = 1/\lambda^2$ → $V(T) = 4$ → $\sigma_T = 2$ hores

Probabilitat d'estar 8 o més hores sense rebre incidències:

$P(T > 8) = 0.0183$

Exercici Model Exponencial i Poisson. Fiabilitat

El centre de càlcul d'una important empresa garanteix treballar amb una mitjana de 2 hores entre incidències.

Quina és la distribució de la v.a "temps en hores entre incidències"?

Quina és la funció de distribució?

Quina és la funció de fiabilitat?

Quina és la taxa d'errors?

Quin és el MTTF?

Quin és el valor d'hores entre incidències que podem garantir que es superarà amb una fiabilitat del 78%?

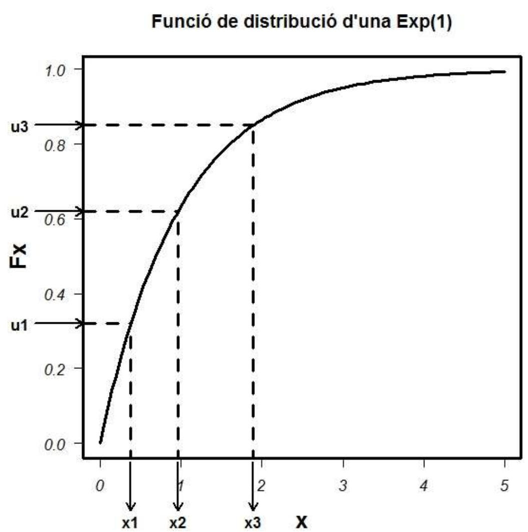
Exercici diversos models

Continuem analitzant el cas de l'aeroport, donant importància al procés d'arribades de passatgers als punts de facturació. Respon a les següents preguntes.

1. Un determinat punt de facturació es caracteritza perquè el número de viatgers que arriben per minut es distribueix segons una Poisson amb mitjana de 9.5. Calcula la probabilitat que aquest número sigui menor que 7 [0.165]
2. En aquest punt de facturació, quina és la probabilitat d'observar exactament 10 arribades en un minut? [0.123]
3. En el mateix punt de facturació, quina és l'esperança de la variable temps (*en segons*) entre dues arribades? [6.316 sg]
4. Al punt de facturació, quina és la probabilitat d'estar menys de 4 segons sense arribades? [0.469]
5. Considerant 18 punts de facturació caracteritzats per una probabilitat 0.8 d'observar exactament 0 arribades en un minut, quina és la probabilitat de tenir més de 14 punts amb 0 arribades? [0.501]

Aplicació model Uniforme i Exponencial

- Generar números pseudo-aleatoris amb distribució U[0,1] és relativament complicat
- No obstant, generar números aleatoris de qualsevol distribució, un cop tenim els nombres pseudo-aleatoris amb distribució U[0,1] és senzill:
  - Per generar $Y \sim U[a,b] \rightarrow Y = a + (b-a)\cdot u \sim U[a,b]$ on u són valors d'una U[0,1]
  - Per generar $T \sim \exp(\lambda) \rightarrow T = F^{-1}(u) = -\ln(1-u)/\lambda \sim \exp(\lambda)$ on u són valors d'una U[0,1]
- En general, el mètode de la transformació inversa (que empra F^{-1}) permet generar valors de qualsevol distribució (no cal que F tingui expressió analítica)



Generar valors d'una exponencial:
- Es parteix de $F_X(x) = 1 - e^{-\lambda\cdot x}$
- S'agafen valors $u_i \sim U[0, 1]$
- Es troben les x_i tals que $F_X(x_i) = u_i = 1 - e^{-\lambda\cdot x_i}$
- Els x_i segueixen una distribució Exponencial(λ)

R: Per generar valors psudoaleatoris tenim funcions per a cada distribució: rbinom, rpois, runif, rexp ....

Exemples model Normal

Exemple 1: Sigui Z ~ N(0, 1). Quin és el valor *z* que deixa una probabilitat per sota de 0.25?

$$P(Z < z) = 0.25 \quad qnorm(0.25) = -0.6745 \quad \dashrightarrow z = -0.6745$$

O per simetries:

$$P(Z < -z) = 0.75 \quad qnorm(0.75) = 0.6745 \dashrightarrow z = -0.6745$$

Exemple 2: Sigui X: "Increment diari espai disc" ~ N(10 MB, 3 MB) $\left(\mathrm{Z} = \frac{\mathrm{X}-10}{3} \sim \mathrm{N}(0,1)\right)$

1) Quant val P(*X*>15)?

$$P(X > 15) = 1 - P(X \leq 15) = 1 - F_X(15) = 1 - pnorm(15,10,3) = 1 - 0.9522 = 0.0478$$

$$P\left(Z > \frac{15-10}{3}\right) = P\left(Z > \frac{5}{3}\right) = 1 - P\left(Z \leq \frac{5}{3}\right) = 1 - F_Z\left(\frac{5}{3}\right) = 1 - pnorm\left(\frac{5}{3}\right) = 1 - 0.9522 = 0.0478$$

2) 1 de cada 10 dies, l'augment és inferior a quant? (Quant val *t* tal que P(*X* < *t*) = 0.1?)

$$\mathrm{P}(\mathrm{Z} < \mathrm{t}') = 0.1 \rightarrow \mathrm{t}' = \mathrm{F}_{\mathrm{Z}}^{-1}(0.1) = \mathrm{qnorm}(0.1) = -1.2816 \quad \rightarrow \quad \mathrm{t} = 10 + 3 \cdot \mathrm{t}' = 6.15\ \mathrm{MB}$$

(t'= $\frac{\mathrm{t}-10}{3}$ i per tant $t = 10 + 3 \cdot t'$)

Exemple Teorema Central del Límit (TCL)

El treball de CPU per fer un *backup* presenta unes característiques diàries de mitjana 30'/dia i una desviació de 15'/dia. Si volem calcular probabilitats sobre el consum de CPU total mensual (suposant independència entre els 30 dies del mes), haurem de plantejar-nos la variable: $S_n = X_1 + X_2 + \ldots + X_{30}$ (no ens donen la distribució de X_i, sols μ i σ)

1) Quina és la distribució de S_n?

$$S_n = \sum X_i \xrightarrow[n\ gran]{} N(n\mu, \sigma\sqrt{n}) = N(900\,, 82.15)\ min = N(15{,}1.37)\ hores$$

2) Quina és la probabilitat d'un consum total mensual de més de 18 hores de CPU?

$$P(S_n > 18) = P\left(\frac{S_n - 1}{1.37} > \frac{18-15}{1.37}\right) = P(Z > 2.19) = 1 - p(Z < 2.19) = 1 - 0.9857 = 0.0143$$

(`1-pnorm(18,15,1.37)` o bé `1-pnorm(2.19)`)

[1 / 0.0143 → 70 per tant un de cada 70 mesos]

3) I la probabilitat de, en un mes, un consum mitjà diari inferior a 36'?

$$\bar{X}_{30} \sim N\left(\mu, \frac{\sigma}{\sqrt{30}}\right) = N(30\,, 2.74)\ min \to P(\bar{X}_{30} < 36) = P\left(Z < \frac{36-30}{2.74}\right) = P(Z < 2.19) = 0.9857$$

(`pnorm(36,30,2.74)` o bé `pnorm(2.19)`)

Exercici Teorema Central del Límit (TCL)

Suposem que s'ha establert que el pes del equipatge d'un viatger segueix una distribució Normal amb mitjana 18.9 Kg. i desviació 4 Kg. Habitualment, si el equipatge d'un viatger sobrepassa els 20 Kg., llavors té un sobrepreu que depèn de l'excés de pes. Contesta les següents qüestions:

1. Troba la probabilitat que un viatger hagi de pagar sobrepreu per excedir el seu equipatge els 20 Kg. de pes. 0.392 [`1-pnorm(20,18.9,4)`]

2. Quin és el pes que podem assegurar , amb un 95% de probabilitat , que un equipatge no superarà? 25.479 [`qnorm(0.95,18.9,4)`]

3. Indica els paràmetres (μ , σ) de la variable pes <u>total</u> de 10 equipatges
   ($\mu = 189$, $\sigma = 4\sqrt{10} = 12.649$)

4. Indica els paràmetres (μ , σ) de la variable pes <u>mitjà</u> de 10 equipatges
   ($\mu = 18.9$, $\sigma = 4/\sqrt{10} = 1.265$)

5. Pel pes mitjà, indica l'interval de 90% de probabilitat més estret possible (el més estret és el centrat en μ) El 90% està entre 16.82 i 20.98

```
mu=18.9 sigma=1.265
qnorm(0.05,mu,sigma) → 16.82   qnorm(0.95,mu,sigma) → 20.98   (amplada: 4.16)
```

Si l'interval no és centrat en μ sinó desplaçat, resultarà més ample:

```
qnorm(0.025,mu,sigma) → 16.42   qnorm(0.925,mu,sigma) → 20.72   (amplada: 4.30)
```

Exercici Teorema Central del Límit (TCL)

L'error de mesura del temps d'un procediment és Normal amb σ = 1/4 seg. i mitjana 0. Es considera repetir les mesures de forma independent.

1) VAC d'una mesura de l'error: E_n: "error en la n-èssima mesura"
2) Model de la VAC: $E_n \sim N(\mu = 0, \sigma = 0.25)$
3) Prob. error menor de 0.1 s.:

$$P(|E_1| \leq 0.1) = P(-0.1 \leq E_1 \leq 0.1) = F_{E_1}(0.1) - F_{E_1}(-0.1) = 0.655 - 0.435 = 0.311$$

4) Error màxim (amb prob. 0.95) en una mesura:

$$P(-e \leq E_1 \leq e) = 0.95 \rightarrow P(E_1 \leq e) = 0.975 \rightarrow e = 0.49\ s.$$

5) Ídem per VAC de mitjana de 10 mesures:

$$X_{10} \sim N\left(\mu = 0, \sigma = \frac{0.25}{\sqrt{10}}\right) \rightarrow P(-f \leq X_{10} \leq f) = 0.95 \rightarrow f = 0.156\ \text{s.}$$

6) Número n mínim de mesures per tal que l'error màxim de la mitjana de les *n* mesures (amb prob. 0.95) sigui inferior a 0.1 s. És a dir n tal que

$$P(-0.1 \leq X_n \leq 0.1) \geq 0.95 \rightarrow P\left(\frac{-0.1-0}{\frac{0.25}{\sqrt{n}}} \leq Z \leq \frac{0.1-0}{\frac{0.25}{\sqrt{n}}}\right) \rightarrow \frac{0.1}{\frac{0.25}{\sqrt{n}}} = 1.956 \rightarrow n = 24.01 \rightarrow n = 25$$

Exemple aproximació entre distribucions

S’ha comprovat que la probabilitat que el temps de resposta d’una determinada pàgina web al llarg d’un dia sigui inadequat és d’un 5%. Per calcular probabilitats del número de dies en que el servei és inadequat durant una setmana (7 dies), un mes (30 dies) o cinc anys (1825 dies), necessitem les següents variables aleatòries per les quals podem identificar el model més adequat:

X_{set}=“nombre de dies en 1 setmana amb servei inadequat”

X_{set}~Bin(n=7,p=0.05)

X_{mes}=“nombre de dies en 1 mes amb servei inadequat“

X_{mes} ~ Bin(n=30,p=0.05)

X_{mes} ~ P(λ=1.5)

X_{5anys}=“nombre de dies en 5 anys amb servei inadequat”

X_{5anys} ~ Bin(n=1825, p=0.05)

X_{5anys} ~ P(λ=91.25)

X_{5anys} ~ N(μ=91.25, σ=9.55) [Encara que p és petita, n és molt gran]