7. Models de probabilitat de variables aleatòries

Part III: Probabilitats

Professor: Pablo Gregori
gregori@mat.uji.es
Dept. de Matemàtiques

Universitat Jaume I de Castelló

Estadística d'ETIG. Curs 2008-09

Guió

1. Experiments bàsics
   - Dos models d'experiment
   - Prova de Bernoulli de paràmetre p
2. Models de probabilitat per a vv.aa. discretes
   - Binomial de paràmetres n i p
   - Binomial negativa de paràmetres r i p
   - Hipergeomètrica de paràmetres N, K i n
   - Poisson de paràmetre λ
3. Models de probabilitat per a vv.aa. continues
   - Uniforme a l'interval (a, b)
   - Exponencial de paràmetre λ
   - Erlang de paràmetres λ i r
   - Normal o Gaussiana de paràmetres μ i σ

- Discret
  - Una urna conté boles idèntiques (excepte pel color).
  - Es trau 1 bola a l'atzar.
  - S'observa el color.
  - Si es vol crear un experiment amb probabilitat d'èxit $p = a/b$ ($a \leq b$, $a, b \in \mathbb{N}$), s'ha de compondre l'urna de a boles d'un color destacat, i la resta d'altre(s) color(s). S'dentificarà com a èxit que aparega el color destacat.
  - Es poden fer extraccions succesives sense o amb remplaçament.
- Continu
  - Una fletxa dins un disc pot rodar al voltant del seu centre.
  - Es fa rodar la fletxa a l'atzar.
  - S'observa el punt de la circunferència al qual apunta
  - Si es vol crear un experiment amb probabilitat d'èxit $p \in (0, 1)$, s'ha de pintar una porció p de la circunferència d'un color destacat. S'dentificarà com a èxit que la fletxa apunte sobre el color destacat.

Experiment Un esdeveniment **èxit** ocorre amb probabilitat p coneguda. La v.a. X pren el valor 1 si l'èxit té lloc i 0 en altre cas. **Aquest experiment s'anomena prova de Bernoulli de paràmetre** p.

Espai mostral $X \in \{0, 1\}$ Notació: $X \sim \mathrm{Be}(p)$.

Fn. prob. $$f(x) = \Pr(X = x) = \begin{cases} 1 - p & x = 0 \\ p & x = 1 \\ 0 & \text{altre cas} \end{cases}$$

Esper. i Var. $\mathbb{E}(X) = p \qquad \mathbb{V}(X) = p(1 - p)$

Exemple

Un estudi de transmissió informàtica en cert canal arriba a la conclusió de que un bit és transmés erròniament (és a dir, el seu valor canvia de l'emisor al receptor) en 1 de cada 20 enviaments, independentment del valor del bit emés i de la correcció en l'emissió del bit anterior. Es transmet un bit:

- *Si pensem que "èxit" és la transmissió correcta, quina és la probabilitat de que ho faça correctament?*
  $X_1 \sim \mathrm{Be}(p = \quad) \to \Pr(''\text{correcte}'') = \Pr(X_1 = \quad) =$
- *Si pensem que "èxit" és la transmissió incorrecta, quina és la probabilitat de que ho faça correctament?*
  $X_2 \sim \mathrm{Be}(p = \quad) \to \Pr(''\text{correcte}'') = \Pr(X_2 = \quad) =$

Experiment Es realitzen n proves de Bernouilli de paràmetre p independents i X = "**número d'èxits**".

Espai mostral $X \in \{0, 1, 2, \ldots, n\}$ Notació: $X \sim \mathrm{Bin}(n, p)$.

Fn. prob. $f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, \ldots, n \\ 0 & \text{altre cas} \end{cases}$

R $f(x) \equiv$ `dbinom(x=x, size=n, prob=p)`

Fn. acum. R $F(x) \equiv$ `pbinom(q=x, size=n, prob=p)`

Esper. i Var. $\mathbb{E}(X) = np$ $\mathbb{V}(X) = np(1-p)$

Propietat $X = \sum_{i=1}^{n} X_i$ on cada $X_i \sim Be(p)$ i són independents.

Exemple

Estem a l' estudi de transmissió informàtica d'abans (un bit és transmés erròniament en 1 de cada 20 enviaments). Ara l'experiment consisteix en transmetre una cadena de 8 bits i observar la qualitat de la transmissió pel número de bits correctes o incorrectes.

- *Defineix la v.a. X_1 ="......" de manera que s'adecue al perfil d'una v.a. amb distribució Binomial de paràmetres n=$\square$ i p=$\square$.*
- *Defineix la v.a. X_2 ="......" d'altra manera que s'adecue al perfil d'una v.a. amb distribució Binomial de paràmetres n=$\square$ i p=$\square$.*

Exemple (cont.)

Estem a l' estudi de transmissió informàtica d'abans (un bit és transmés erròniament en 1 de cada 20 enviaments). Ara l'experiment consisteix en transmetre una cadena de 8 bits i observar la qualitat de la transmissió pel número de bits correctes o incorrectes.

- *Número esperat de bits correctes?*
- *Prob. de que es transmeta la cadena perfectament?*
  $\Pr(X_1 \quad) =$ $\Pr(X_2 \quad) =$
- *Prob. de que es transmeta amb 2 ó menys errors?*
  $\Pr(X_1 \quad) =$ $\Pr(X_2 \quad) =$

Experiment Es fan proves de Bernoulli de paràmetre p independents fins que s'acumulen r èxits i X = "**número de proves fetes fins obtindre r èxits** (incloent-los)".

Espai mostral $X \in \{r, r+1, r+2, r+3, \ldots\}$.

Fn. prob. $f(x) = \begin{cases} \binom{x-1}{r-1} p^r (1-p)^{x-r} & x = r, r+1, \ldots \\ 0 & \text{altre cas} \end{cases}$

$f(x) \equiv$ `dnbinom(x=x-r, size=r, prob=p)`

Fn. acum. $F(x) \equiv$ `pnbinom(q=x-r, size=r, prob=p)`

Esper. i Var. $\mathbb{E}(X) = r/p \qquad \mathbb{V}(X) = r(1-p)/p^2$

Nota Quan $r = 1$ es coneix com **Geomètrica** i en eixe cas $F(x) = 1 - (1-p)^x$.

Exemple

Una (mala) impressora té una prob. d'espatllar-se de 0.05 a cada pàgina que imprimeix. S'estudia el núm. de pàgines impresses fins avariar-se (incloent la de l'avaria).

- *Fes un paralelisme entre l'experiment i l'extracció de boles d'una urna.*
- *Defineix la v.a. X =".....” de manera que s'adecue al perfil d'una v.a. amb distribució Binomial negativa de paràmetres $r = \square$ i $p = \square$.*

Exemple (cont.)

Una impressora té una prob. d'espatllar-se de 0.05 a cada pàgina que imprimeix. S'estudia el núm. de pàgines impresses fins avariar-se.

- *Expectativa del núm. de pàgines impresses fins la pròxima avaria.*
- *Prob. d'avariar-se a la 15ena pàgina.*
  $\Pr(X \qquad) =$
- *Prob. d'imprimir més de 15 pàg. abans de l'avaria.*
  $\Pr(X \qquad) =$
- *Prob. d'imprimir més de 50 pàg. abans de l'avaria.*
  $\Pr(X \qquad) =$

Exemple

En l'estudi de la transmissió de bits (5% erronis) es va rebent i comparant la cadena de bits rebuda amb l'original, analitzant la llargària de la cadena quan es conten 10 errors.

- *Defineix la v.a. X =".....” de manera que s'adecue al perfil d'una v.a. amb distribució Binomial negativa de paràmetres $r = \square$ i $p = \square$.*
- *Expectativa de la llargaria de la cadena.*
- *Prob. d'haver transmés més de 100 bits?*
  $\Pr(X \qquad) =$
- *Prob. d'haver transmés més de 1000 bits?*
  $\Pr(X \qquad) =$

Experiment Es trauen n boles (sense remplaçament!) d'una urna que conté N boles, de les quals només K són de tipus **èxit**. X = "**número d'èxits**"

Espai mostral $X \in \{\max(0, n-(N-K)), \ldots, \min(n, K)\}$.

Fn. prob. $f(x) = \begin{cases} \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} & x = \max(0, n-(N-K)), \ldots, \min(n, K) \\ 0 & \text{altre cas} \end{cases}$

$f(x) \equiv$ `dhyper(x=x, m=K, n=N-K, k=n)`

Fn. acum. $F(x) \equiv$ `phyper(q=x, m=K, n=N-K, k=n)`

Esper. i Var. $\mathbb{E}(X) = n\frac{K}{N}$ $\quad \mathbb{V}(X) = n\frac{K}{N}(1-\frac{K}{N})\frac{N-n}{N-1}$

Exemple

Un producte ve en lots de 500 unitats, de les quals 35 són defectuoses. El comprador pren 7 unitats per comprovar si compra el lot (el comprarà només si totes són correctes).

- *Defineix la v.a. X = "....." de manera que s'adecue al perfil d'una v.a. amb distribució Hipergeomètrica de paràmetres $N = \square$, $K = \square$ i $n = \square$.*
- *Expectativa de lots defectuosos trobats.*
- *Prob. de no triar cap producte defectuós (i que per tant el client compre el lot).*
  $\Pr(X \quad) =$
- *Idem, però si equivoquem la distribució de X per una binomial $Y \sim \text{Bin}(n = \quad, p = \quad)$.*
  $\Pr(Y \quad) =$

Experiment Un esdeveniment tipus **èxit** pot passar en cada punt d'un interval de temps o espai, amb independència, i amb una mitjana de λ aparicions a l'interval. X = "**número d'èxits observats en l'interval**"

Espai mostral $X \in \{0, 1, 2, 3, \ldots\}$ Notació: $X \sim \text{Po}(\lambda)$.

Fn. prob. $f(x) = \begin{cases} e^{-\lambda} \cdot \frac{\lambda^x}{x!} & x = 0, 1, 2, 3, \ldots \\ 0 & \text{altre cas} \end{cases}$

$f(x) \equiv$ `dpois(x=x, lambda=λ)`

Fn. acum. $F(x) \equiv$ `ppois(q=x, lambda=λ)`

Esper. i Var. $\mathbb{E}(X) = \lambda$ $\quad \mathbb{V}(X) = \lambda$

Exemple

Un servidor ftp rep un promedi de 3.5 peticions per minut. Calcula:

- *Expectativa de peticions per al pròxim minut.*
- *Prob. de no cap petició al següent minut?*
  $\Pr(X \quad) =$
- *Prob. de més de 5 peticions al següent minut?*
  $\Pr(X \quad) =$
- *Prob. d'entre 2 i 8 peticions al següent minut?*
  $\Pr(\quad X \quad) =$

Propietat (1,2: Sumes, 3:Aproximació de Binomial per Poisson)

1. *Si $X_1 \sim \mathrm{Bin}(n_1, p)$ i $X_2 \sim \mathrm{Bin}(n_2, p)$ són independents i $Y = X_1 + X_2$, aleshores:*

$$Y \sim \mathrm{Bin}(n_1 + n_2, p)$$

2. *Si $X_1 \sim \mathrm{Po}(\lambda_1)$ i $X_2 \sim \mathrm{Po}(\lambda_2)$ són independents, $\alpha > 0$ i $Y = X_1 + X_2$ i $Z = \alpha X_1$, aleshores:*

$$Y \sim \mathrm{Po}(\lambda_1 + \lambda_2) \qquad Z \sim \mathrm{Po}(\alpha\lambda_1)$$

3. *Si $X \sim \mathrm{Bin}(n, p)$, n és "gran" i p "menut", i $Y \sim \mathrm{Po}(\lambda = np)$, aleshores X i Y tenen distribucions molt semblants (raó per la qual es podria usar una per fer càlculs aproximats de l'altra).*

Exemple

Un servidor ftp rep un promedi de 3.5 peticions per minut. Calcula:

- *Def. i model de la v.a. per fer els següents apartats:*
  X = ".....", $X \sim$ [ ]
- *Expectativa de peticions per la pròxima hora.*
- *Prob. de no cap petició la següent hora?*
  $\Pr(X \quad) =$
- *Prob. de més de 600 peticions la següent hora?*
  $\Pr(X \quad) =$
- *Prob. d'entre 400 i 500 peticions la següent hora?*
  $\Pr(\quad X \quad) =$

Exemple

Aprox. 1 de cada 4500 discs durs eix defectuós. Tenim un lot de 10000 discs. Calcula:

- *X = "....."*, $X \sim$ [ ]
- *Expectativa del núm. de discs defectuosos.*
- *Prob. de no cap disc defectuosos?*
  $\Pr(X \quad) =$
- *Prob. d'un sol disc defectuosos?*
  $\Pr(X \quad) =$
- *Prob. de 2 ó més discs defectuosos?*
  $\Pr(X \quad) =$

Experiment Un número d'un interval (a, b) es triat completament a l'atzar i X = "**número triat**".

Espai mostral $X \in (a, b)$ Notació: $X \sim \mathrm{Unif}(a, b)$.

Fn. prob. $f(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b) \\ 0 & \text{altre cas} \end{cases}$

R $f(x) \equiv$ `dunif(x=x, min=a, max=b)`

Fn. acum. $F(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & x \in (a, b) \\ 1 & x \geq b \end{cases}$

R $F(x) \equiv$ `punif(q=x, min=a, max=b)`

Esper. i var. $\mathbb{E}(X) = \frac{a+b}{2}$ $\mathbb{V}(X) = \frac{(b-a)^2}{12}$

Exemple

Un virus informàtic tria completament a l'atzar una hora de cada dia (des de les 00:00:00 fins les 23:59:59.999...) i borra un arxiu en eixe moment.

- *Defineix la v.a. $X =$"....." de manera que s'adecue al perfil d'una v.a. amb distribució uniforme a l'interval (,).*
- *Calcula la probabilitat (o el % de voltes a llarg termini) de que l'arxiu triat siga eliminat en horari d'oficina (entre les 9h i les 17h).*
- *Si no s'elimina el virus en tot un any, quants arxius (més o menys) serien eliminats en la primera hora de cada dia?*

Experiment Un procés de Poisson de paràmetre λ ocorre en l'espai o el temps. Fixem un lloc en l'espai o un instant en el temps i X = "**Espai o temps transcorregut fins la nova aparició de l'esdeveniment o entre 2 aparicions succesives**".

Espai mostral $X \in (0, +\infty)$ Notació: $X \sim \mathrm{Exp}(\lambda)$.

Fn. prob. $f(x) = \begin{cases} 0 & x \leq 0 \\ \lambda e^{-\lambda x} & x > 0 \end{cases}$

R $f(x) \equiv$ `dexp(x=x, rate=λ)`

Fn. acum. $F(x) = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-\lambda x} & x > 0 \end{cases}$

R $F(x) \equiv$ `pexp(q=x, rate=λ)`

Esper. i var. $\mathbb{E}(X) = \frac{1}{\lambda}$ $\mathbb{V}(X) = \frac{1}{\lambda^2}$

Exemple

Un servidor rep una mitjana de 6.5 peticions per minut. Calcula:

- *La variable $X =$"Temps fins la pròxima petició (en minuts)" es distribueix com $X \sim \mathrm{Exp}(\lambda = \quad)$*
- *Temps mitjà que transcorre entre dues peticions consecutives?*
- *Prob. de que es reba una petició abans del 5 s. següents?*
- *Prob. de que passe 1 minut sense rebre cap nova petició?*

Exemple

Al analitzar la durabilitat d'un dispositiu, es fa un estudi on s'observa que el temps mitjà de duració dels aparells es de 1.35 anys.

- *La variable $X =$"durada del dispositiu (en anys)" pot considerar-se que es distribueix com $X \sim \mathrm{Exp}(\lambda = \quad)$.*
- *Prob. de que un dispositiu qualsevol dure més de 2 anys?*
- *Prob. de que un dispositiu qualsevol dure menys d' un mes?*
- *Si la garantia dels dispositius es de 3 anys, % d'aparells que es tornaran en periode de garantia?*

Experiment Un procés de Poisson de paràmetre λ ocorre en l'espai o el temps. Fixem un lloc en l'espai o un instant en el temps i X = "**Espai o temps transcorregut fins la nova aparició de r vegades l'esdeveniment o entre $r+1$ aparicions succesives**".

Espai mostral $X \in (0, +\infty)$ Notació: $X \sim \mathrm{Erl}(\lambda, r)$.

Fn. prob. $f(x) = \begin{cases} 0 & x \leq 0 \\ \frac{\lambda^r x^{r-1} e^{-\lambda x}}{(r-1)!} & x > 0 \end{cases}$

$f(x) \equiv$ `dgamma(x=x, shape=r, rate=λ)`

Fn. acum. (No hi ha expressió senzilla)

$F(x) \equiv$ `pgamma(q=x, shape=r, rate=λ)`

Esper. i var. $\mathbb{E}(X) = r/\lambda$ $\mathbb{V}(X) = r/\lambda^2$

Propietat $X = \sum_{i=1}^{n} X_i$ on cada $X_i \sim \mathrm{Exp}(\lambda)$ i són indep.

Exemple

Un node de la xarxa rep i emet paquets d'informació, de manera que els van rebent a raó de 100 paquets cada 2.76 segons, i espera a rebre 10 paquets abans de retransmetre'ls d'un sol cop. Calcula:

- *La v.a. X = "Temps des d'una emisió a altra" es distribueix com $X \sim \mathrm{Erl}(\lambda = \quad, r = \quad)$.*
- *Temps mitjà des d'una emisió a altra?*
- *Prob. de que el temps des d'una emisió a la següent siga de menys de 0.5 s.?*
- *Prob. de que passe més d'un segon sense enviar informació?*

Experiment En la natura o en processos industrials existeix una variabilitat natural i inevitable (o un error inherent en els instruments de mesura). Si ens fixem en una característica a mesurar i X = "**Valor observat o mesurat**".

Espai mostral $X \in (-\infty, +\infty)$ Notació: $X \sim \mathrm{N}(\mu, \sigma^2)$.

Fn. prob. $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

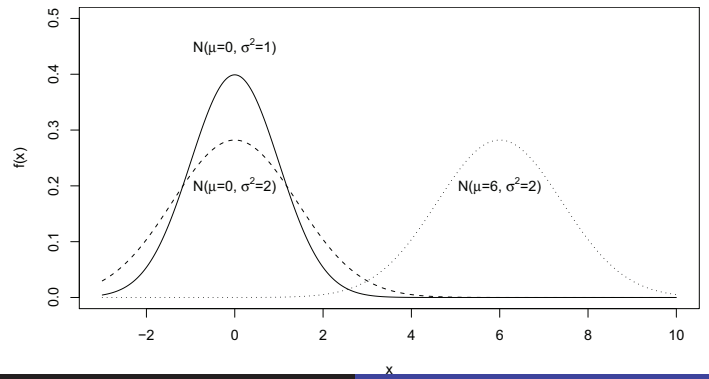
$f(x) \equiv$ `dnorm(x=x, mean=μ, sd=σ)`

Fn. acum. (No hi ha expressió senzilla)

$F(x) \equiv$ `pnorm(q=x, mean=μ, sd=σ)`

Esper. i var. $\mathbb{E}(X) = \mu$ $\mathbb{V}(X) = \sigma^2$

La funció de densitat de la normal té un aspecte que depén lògicament dels paràmetres μ i σ^2:
Atorga més versemblança als valors pròxims a la mitjana μ, i la versemblança va baixant (més o menys ràpid segons la dispersió donada per σ) a mesura que els valors s'allunyen de la mitjana.



Propietat (Aproximacions per la normal)

1. *Si* $X \sim \mathrm{N}(\mu_X, \sigma_X^2)$ *i definim una nova variable* $Y = a + bX$ *aleshores* $Y \sim \mathrm{N}(\mu = a + b\mu_X, \sigma^2 = b^2\sigma_X^2)$.
2. *Si* $X \sim \mathrm{N}(\mu_X, \sigma_X^2)$ *i* $Y \sim \mathrm{N}(\mu_Y, \sigma_Y^2)$ *i definim una nova variable* $W = X + Y$ *aleshores* $W \sim \mathrm{N}(\mu = \mu_X + \mu_Y, \sigma^2 = \sigma_X^2 + \sigma_Y^2)$.

Exercici

Els temps dels 4 *corredors de relleus* 4×100 *m. es distribueix normalment amb mitjanes* $11.3, 11.1, 10.6, 10.0$ *s i variàncies* $0.04, 0.02, 0.01, 0.01$ s^2*, respectivament. Quina és la distribució exacta de* X *= "temps total de la carrera"?*

Propietat (Tipificació)

Si $X \sim \mathrm{N}(\mu, \sigma^2)$ *i* $Z = \dfrac{X - \mu}{\sigma}$ *aleshores*

$$Z \sim \mathrm{N}(\mu = 0, \sigma^2 = 1)$$

i es diu NORMAL TIPIFICADA.

Exemple (Càlcul amb les taules)

Només hi ha taula de prob. per Z *normal tipificada. Per tant si, per exemple,* $X \sim \mathrm{N}(\mu = 10, \sigma^2 = 2)$ *i demanen* $\Pr(X < 12)$*:*

$$\Pr(X \leq 12) = \Pr\left(\frac{X - 10}{\sqrt{2}} \leq \frac{12 - 10}{\sqrt{2}}\right)$$
$$= \Pr(Z \leq 1.41) = (taules) = 0.9207$$

Propietat (Teorema del límit central)

Si $X_1, X_2, \ldots, X_n$ *són v.a. independents i de qualsevol distribució, aleshores, quan n és gran* $(n \to \infty)$

$$\sum_{i=1}^{n} X_i \sim_{aprox.} \mathrm{N}(\mu = \sum_{i=1}^{n} \mathbb{E}(X_i), \sigma^2 = \sum_{i=1}^{n} \mathbb{V}(X_i))$$

Exercici

Unes losetes de ceràmica es fabriquen d'una superficie mitjana de 3.15 cm^2 *i variància* 0.1 cm^4*. Si comprem una caixa de* 50000 *losetes:*

- *Si* X *= "Superfície total de les losetes",* $X \sim$ .....
- *Superficie total esperada?*
- *Prob. de que es puga cobrir un àrea de* 15.77 m^2*?*

Propietat (Relació Normal amb Binomial i Poisson)

- *Si* $X \sim \mathrm{Bin}(n, p)$ *i n es prou gran, aleshores* $X \sim_{aprox.} \mathrm{N}(\mu = np, \sigma^2 = np(1 - p))$.
- *Si* $X \sim \mathrm{Po}(\lambda)$*, i* λ *es prou gran, aleshores* $X \sim_{aprox.} \mathrm{N}(\mu = \lambda, \sigma^2 = \lambda)$.

Exercici

El 10% *de les reserves de bitllets d'avió són cancel·lades abans els* 5 *primers dies. Si hui han gestionat* 347 *reserves:*

- *Si* X *= "Núm. de cancel·lacions realitzades abans de* 5 *dies", aleshores* $X \sim$ .....
- *Núm. de cancel·lacions esperades durant els* 5 *dies?*
- *Prob. de que no es produeixca cap cancel·lació?*
- *Prob. de que es produeixen més de 5 cancel·lacions?*

Correcció per continuitat: Si X és una v.a. discreta (p.e. binomial o Poisson) i per qualsevol motiu es considera continua (p.e., si aproximem amb la normal), recorda que:

$$\Pr(X = x) = \begin{cases} f(x) & \text{si } X \text{ és v.a. discreta} \\ 0 & \text{si } X \text{ és v.a. continua} \end{cases}$$

Per tant, per calcular probabilitats cal identificar cada valor discret amb un interval (obert o tancat, dóna lo mateix). Per exemple:

V.a. discreta	V.a. continua
$\Pr(X = x)$	$\Pr(x - 0.5 < X < x + 0.5)$
$\Pr(X = 3)$	$\Pr(2.5 < X < 3.5)$
$\Pr(X = 0)$	$\Pr(-0.5 < X < 0.5)$
$\Pr(5 < X \leq 7)$	$\Pr(5.5 < X < 7.5)$