

## PRNG NORMALE E GAMMA

### Distribuzione normale

**RANNOR**(*seme*)

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-(x-\mu)^2 / 2\sigma^2\right]$$
$$\mu = 0, \sigma^2 = 1$$

Trasformazione Box-Muller da uniforme.

### Distribuzione gamma

**RANGAM**(*seme, a*)

$$f(x) = \left(\frac{x}{b}\right)^{a-1} \frac{\exp(-x/b)}{b\Gamma(a)} I_{(0,\infty)}(x)$$
$$b > 0, a > 0$$

- per  $b=1 \rightarrow$  Gamma standard
  - per  $2*a$  intero  $\rightarrow$  chi-quadrato con  $2*a$  g.l.
- Metodo di accettazione/rifiuto (Cheng, 1977; Fishman, 1978) (genera5.sas)

239

## PRNG BINOMIALE E ESPONENZIALE

### Distribuzione binomiale

**RANBIN**(*seme, n, p*)

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x} I_{\{0,1,\dots,n\}}(x)$$
$$0 \leq p \leq 1, n \in \mathcal{N}$$

- se  $n \leq 50$  o  $np \leq 5 \rightarrow$  trasformazione inversa da uniforme
- se  $n > 50$  e  $np > 5 \rightarrow$  approssimazione normale  $\rightarrow$  trasformazione Box-Muller da uniforme (genera4.sas)

### Distribuzione esponenziale

**RANEXP**(*seme*)

$$f(x) = e^{-x} I_{(0,\infty)}(x)$$

Metodo della trasformazione inversa da uniforme (genera3a.sas, genera3b.sas)

240

## CALCOLO DELLA FUNZIONE DI RIPARTIZIONE

POISSON	Poisson
PROBBETA	Beta
PROBBNML	Binomiale
PROBCHI	Chi-quadrato
PROBF	F di Fisher
PROBGAM	Gamma
PROBHYP	Ipergeometrica
PROBNEG	Binomiale negativa
PROBNORM	Normale standard (genera2b.sas, genera2d.sas)
PROBT	t di Student

## QUANTILI

BETAINV	Beta
CINV	Chi-quadrato
FINV	F di Fisher
GAMINV	Gamma
PROBIT	Normale standard (genera2e.sas)
TINV	t di Student

241

## CAMPIONAMENTO CASUALE

Un campione è probabilistico se valgono le seguenti condizioni:

- è possibile definire l'insieme  $S$  di tutti i campioni distinti estraibili dalla popolazione;
- a ogni campione  $s \in S$  è assegnabile a priori una probabilità di selezione  $p(s)$ ;
- a ciascuna unità della popolazione è associata una probabilità strettamente positiva di selezione;
- il campione è selezionato mediante un "meccanismo" casuale (schema di campionamento) tale che ogni campione  $s$  abbia una probabilità esattamente pari a  $p(s)$ ;
- stabilito un procedimento di stima  $c'$  è una corrispondenza univoca tra campione estratto e valore della stima che da esso si ricava.

242

## PIANO DI CAMPIONAMENTO

**Schema di campionamento:** stabilisce le regole attraverso le quali le singole unità possono entrare a far parte del campione

$$p(s) \geq 0 \text{ per ogni } s \in S, \sum p(s) = 1$$

**probabilità di inclusione:** probabilità che una generica unità della popolazione entri a far parte del campione

$$P(i \in s) = \pi_i = \sum_{s \ni i} p(s)$$

**base di campionamento** (*frame*): lista completa delle unità che compongono la popolazione obiettivo.

**lista** insieme ordinato di contrassegni (*labels*) delle unità della popolazione registrati su un supporto che ne consenta la consultazione. La lista identifica la popolazione di selezione.

243

## ESTRAZIONE E ANALISI DI CAMPIONI CASUALI IN SAS

SURVEYSELECT

selezione di campioni probabilistici da una popolazione data

SURVEYMEANS

SURVEYFREQ

SURVEYREG

SURVEYLOGISTIC

analisi di dati campionari incorporando il disegno di campionamento

244

## PROC SURVEYSELECT

Estrazione di campioni casuali.

*Selezione del campione:*

- input del SAS data set che contiene la *base di campionamento* o la lista delle unità da cui selezionare il campione.
- specificare il *metodo* di selezione
- Indicare la *dimensione* del campione o la frazione di campionamento

La procedura SURVEYSELECT

- Seleziona il campione
- Genera un SAS data set che contiene:
  - le unità selezionate
  - la loro probabilità di inclusione
  - i pesi campionari.

245

## PROC SURVEYSELECT

Campionamento con probabilità costanti:

- . c.c. semplice
- . c.c. con reintroduzione
- . c.c. sistematico
- . c.c. sequenziale

Campionamento con probabilità proporzionali (PPS):

- . PPS senza reintroduzione
- . PPS con reintroduzione
- . PPS sistematico
- . Algoritmo PPS per la selezione di 2 unità per strato
- . PPS sequenziale con reintroduzione

246

## SINTASSI

**PROC SURVEYSELECT** options ;  
**STRATA** variables ;  
**CONTROL** variables ;  
**SIZE** variable ;  
**ID** Variables ;

Opzioni di PROC SURVEYSELECT (selezione)

Task	Options
input data set	DATA=
output data sets	OUT=
	OUTSORT=
Metodo di campionamento	METHOD=
Dimensione campionria	SAMPSIZE=
	SELECTALL
Frazione di campionamento	SAMPRATE=
	NMIN=
	NMAX=
Seme numeri casuali	SEED=
contenuto OUT=	JTPROBS
	STATS

247

## ESEMPIO ESTRAZIONE CON SURVEYSELECT

L'archivio pop.sas7bdat contiene i dati relativi agli iscritti all'anagrafe di un certo comune.

Osservazioni 322697

Variabili 7

- Estrarre un c.c. semplice di dimensione 100 (sample1.sas)
- Estrarre un c.c. stratificato per quartiere e sesso con frazione di campionamento costante pari all'1% (sample2.sas)
- Estrarre un campione stratificato per quartiere, controllando per età e sesso dei rispondenti tramite estrazione sistematica all'interno di ciascuno strato, frazione di campionamento del 2% in ogni strato (sample3.sas).

Esempio di estrazione campione con probabilità variabili vedere sample4.sas

248

## C.C. SEMPLICE CON SURVEYSELECT

Ogni possibile campione ha la stessa probabilità di essere estratto

Il numero di possibili campioni di dimensione  $n$  selezionabili da una

popolazione di  $N$  unità è  $\binom{N}{n} = \frac{N!}{n!(N-n)!}$

⇓

- probabilità di ogni possibile campione

$$p(s) = \binom{N}{n}^{-1}$$

- probabilità di inclusione costanti  
 $\pi_i = n/N$
- pesi campionari  $w_i = 1/\pi_i = N/n$

249