Analisi statistica dell’output prodotto dalla simulazione

Teso di riferimento:

Steven Lavenberg "Computer Performance Modeling Handbook" Academic Press,

Random
Selection

|

Input 1 ------→ SIMULATORE ------→ output 1

Input 2 ------→ SIMULATORE ------→ output 2

.
.
.
.

Input M ------→ SIMULATORE ------→ output M

|

Inferenza
Statistica

Il simulatore realizza un adeguato modello del sistema da studiare;

Necessità di elaborare i dati prodotti dal simulatore

- quando i dati sono significativi?
-
- Quanti esperimenti (run) di simulazione sono necessari per ottenere stime significative?

- In che modo stimare gli indici di prestazione?

- Quali indici?

- Quale tipo di simulazione?

    - Terninating simulation

    - Steady State simulation

a) Terminating simulation: specifiche condizioni di inizio e termine della simulazione. La lunghezza del run è ben definita.

- Es: tempo medio di permanenza dei primi 100 clienti di un supermercato;

b) Steady-state simulation:

la lunghezza del run di simulazione è potenzialmente infinito (lung run, clock --→ ∝);

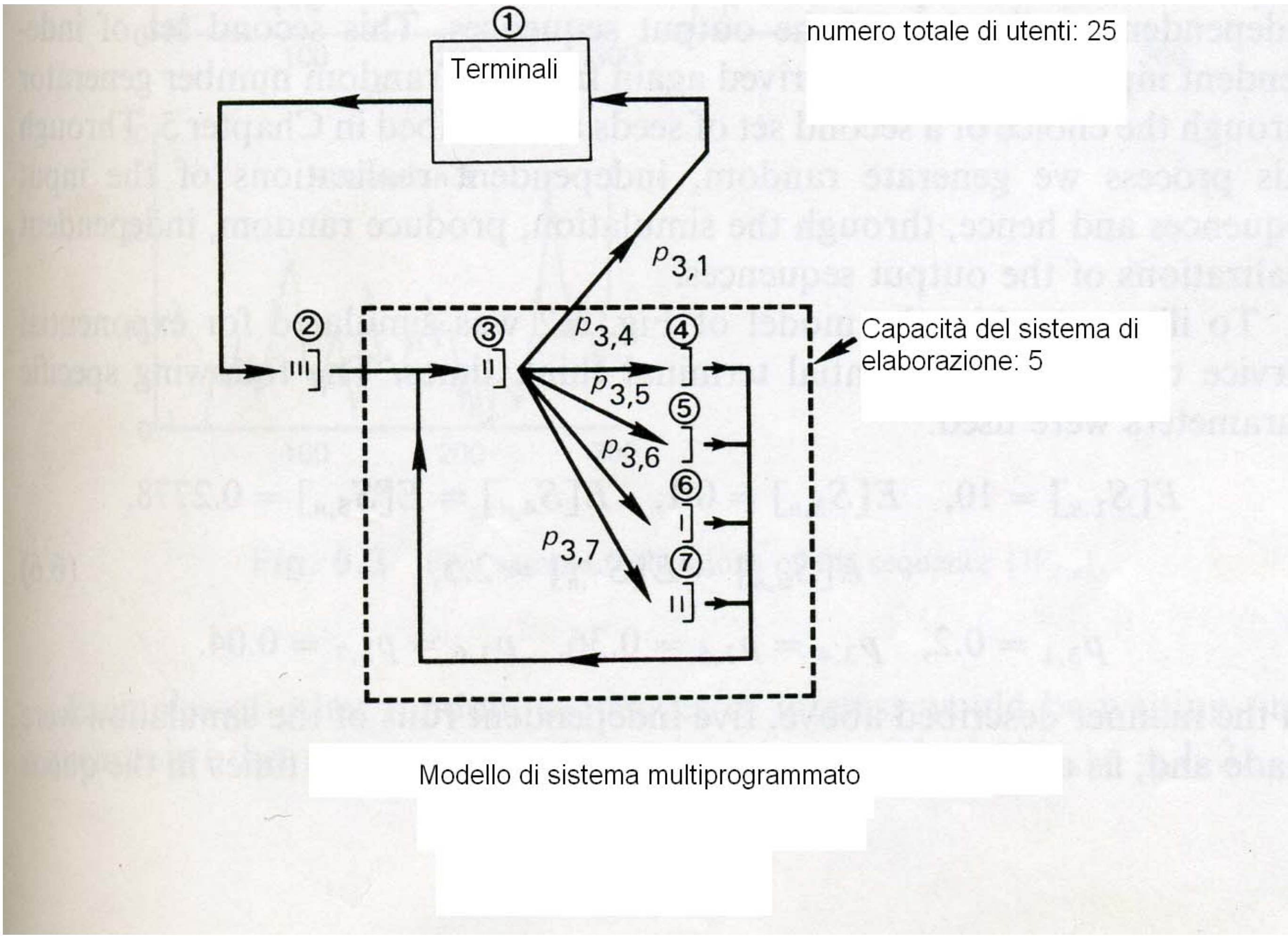
Le condizioni iniziali non sono determinati ai fini del risultato

Non esistono condizioni a priori per calcolare la fine del run di simulazione.

- Es:Tempo medio di permanenza nel sistema di un generico utente

Problemi: indipendenza dei dati e transiente iniziale dell'esperimento di simulazione (steady-state simulation)

ESEMPIO:



Modello di sistema multiprogrammato

Caratteristiche operative:

25 utenti nel sistema;

al sottosistema terminale non ci sono code: ogni utente ha a disposizione un terminale e quando lo utilizza sperimenta un tempo di soggiorno casuale;

il sottosistema di elaborazione può contenere al massimo 5 utenti;

Gli utenti non immediatamente ammessi al sottosistema di elaborazione sono posti nella CODA 2

La CODA 3 contiene gli utenti in attesa del processore;

CODE 4-7 sono riservate ai device di memoria secondaria.

L'INPUT al simulatore consiste di un insieme di 7 sequenze (indipendenti tra loro) di variabili casuali i.i.d:

$\{S_{1,n} : n = 1,2,\ldots.\}$ tempi di permanenza ai terminali;

$\{S_{3,n} : n = 1,2,\ldots.\}$ tempi di servizio al processore;

$\{S_{i,n} : n = 1,2,\ldots.\}$ $i = 4,\ldots 7$, Tempi di servizio ai device.

$\{\theta_n : n = 1,2,\ldots.\}$ indica il servente scelto dall'utente dopo aver lasciato il processore.

θ_n : variabile casuale discreta: $P\{\theta_n = k\} = p_{3,k}$ $k = 1,4,5,6,7$.

Sequenze diverse di input forniranno sequenze diverse di output.

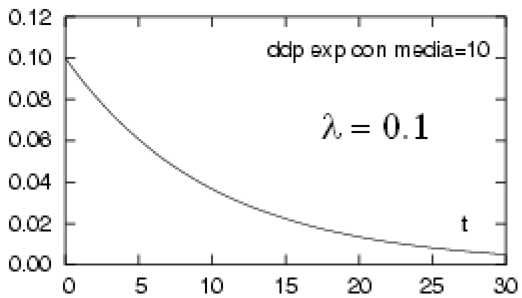
Assumiamo, per il modello in esame, le seguenti condizioni operative:

- $p_{3,1} = 0.2$; $p_{3,4} = p_{3,5} = 0.2$; $p_{3,6} = p_{3,7} = 0.04$.

- distribuzione dei tempi di servizio: **esponenziale negativa**;

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

$$f(t) = \lambda\, e^{-\lambda t}$$



- generatore di sequenze pseudocasuali:

$$U = 1- e^{-\lambda t}$$
$$e^{-\lambda t} = 1-U$$
$$t = \ln(1-U)/\lambda$$
$$t = \ln(U)/\lambda$$

- Parametri della distribuzione:

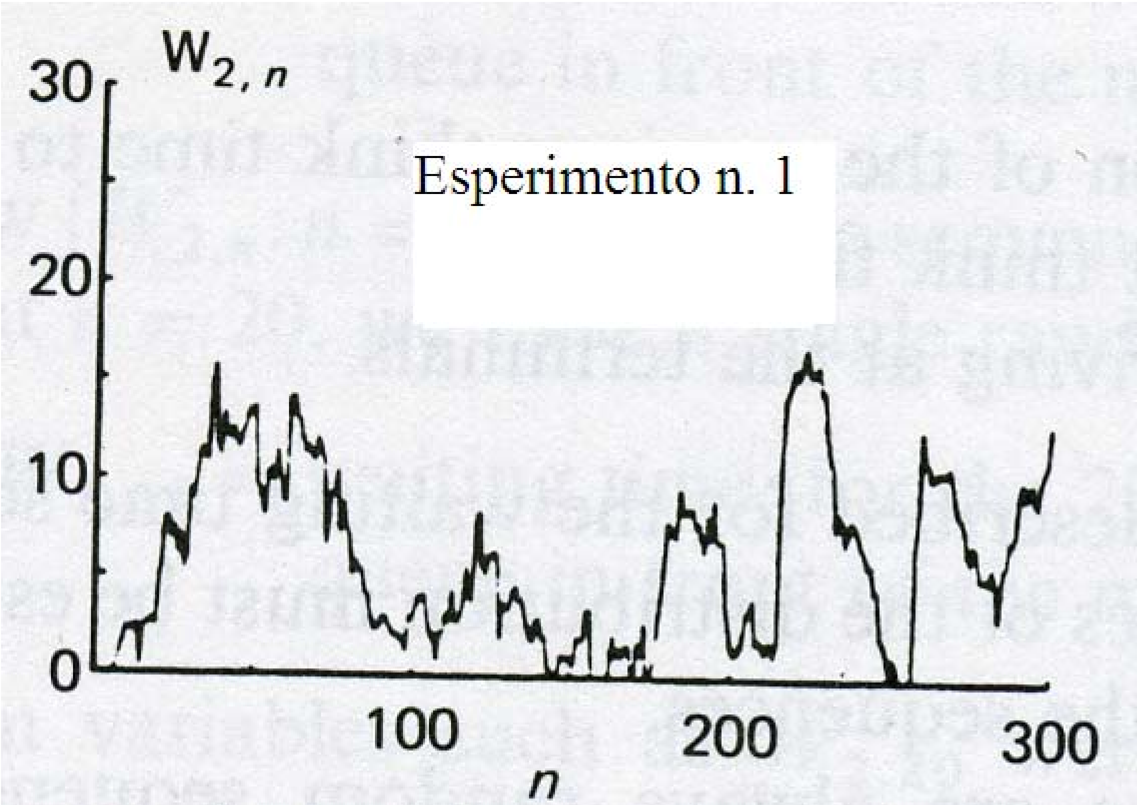
  - $E[S_{1,n}] = 10.0$; $E[S_{3,n}] = 0.1$; $E[S_{4,n}] = E[S_{5,n}] = 0.2778$;
  - $E[S_{6,n}] = E[S_{7,n}] = 2.5$;

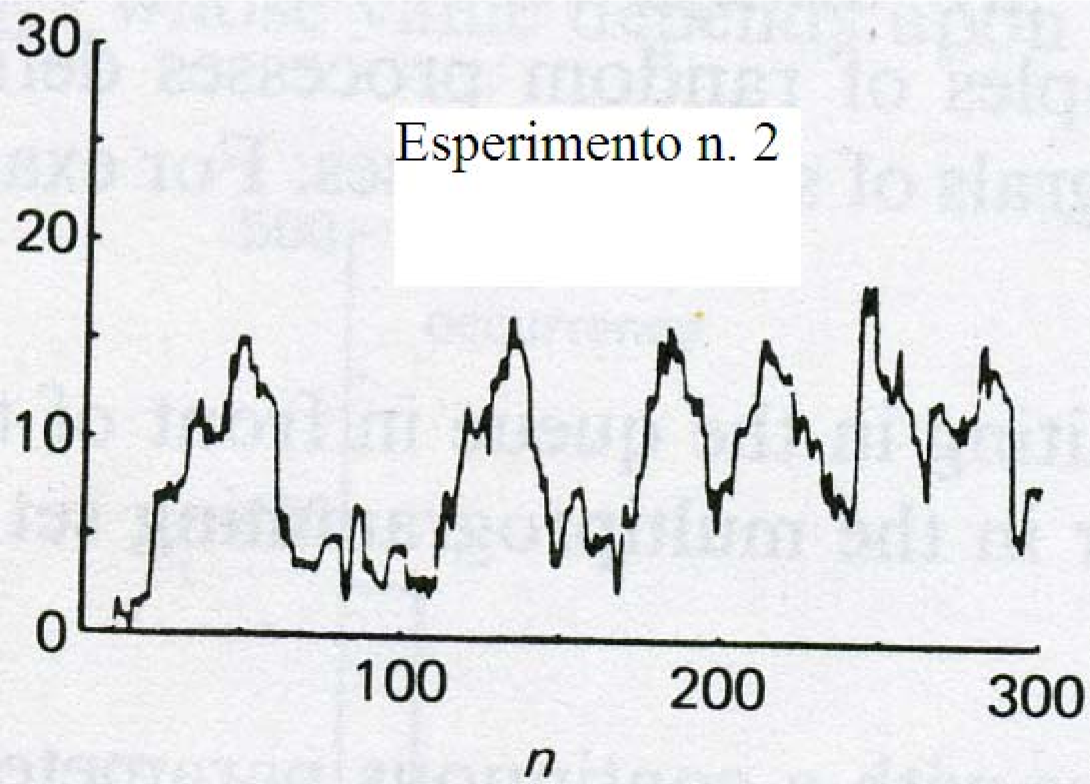
N.B il parametro della distribuzione esponenziale negativa è:

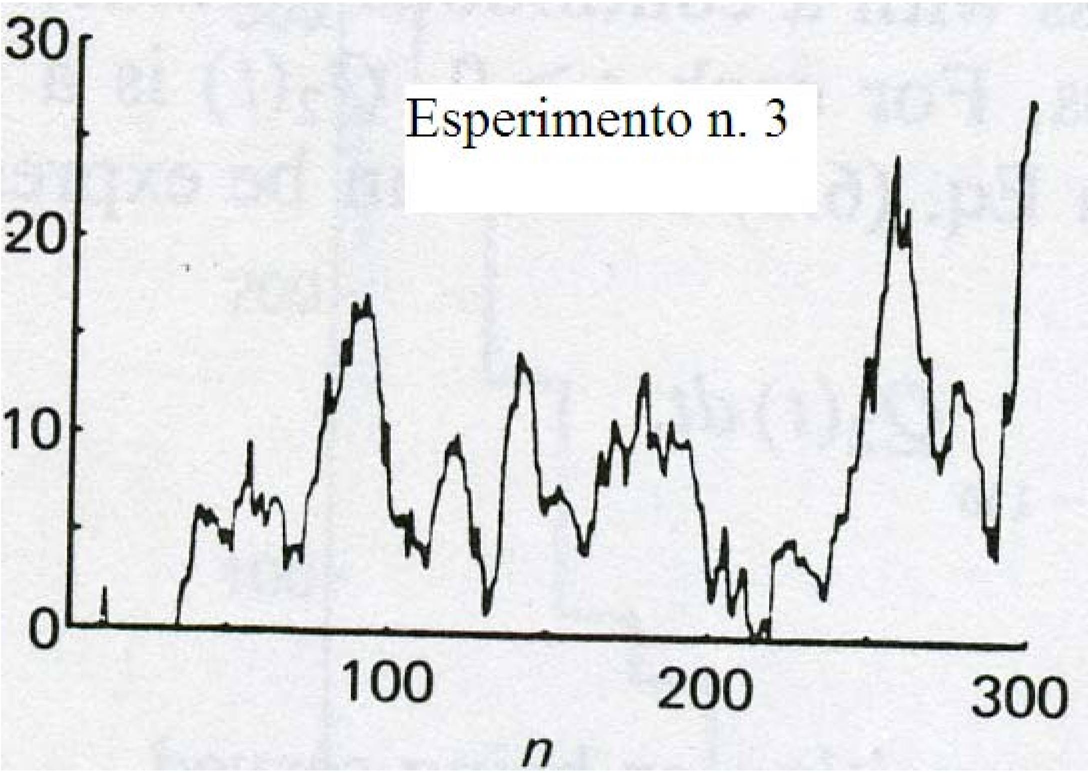
$$\lambda = 1/ E[S_{i,n}]$$

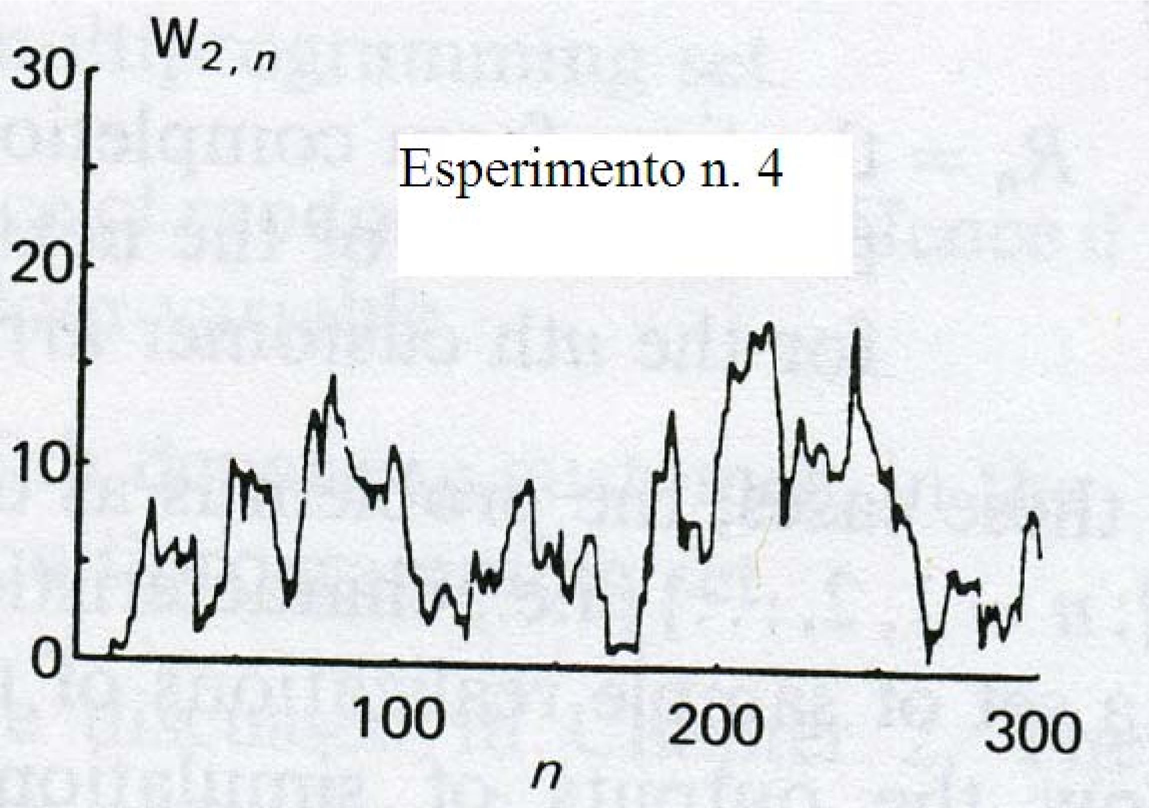
Eseguiamo 5 run diversi (con diverse sequenze di input) e campioniamo il tempo di permanenza degli utenti nella CODA 2:

$$\{W_{2,n} : n = 1,2,\ldots 300\}$$

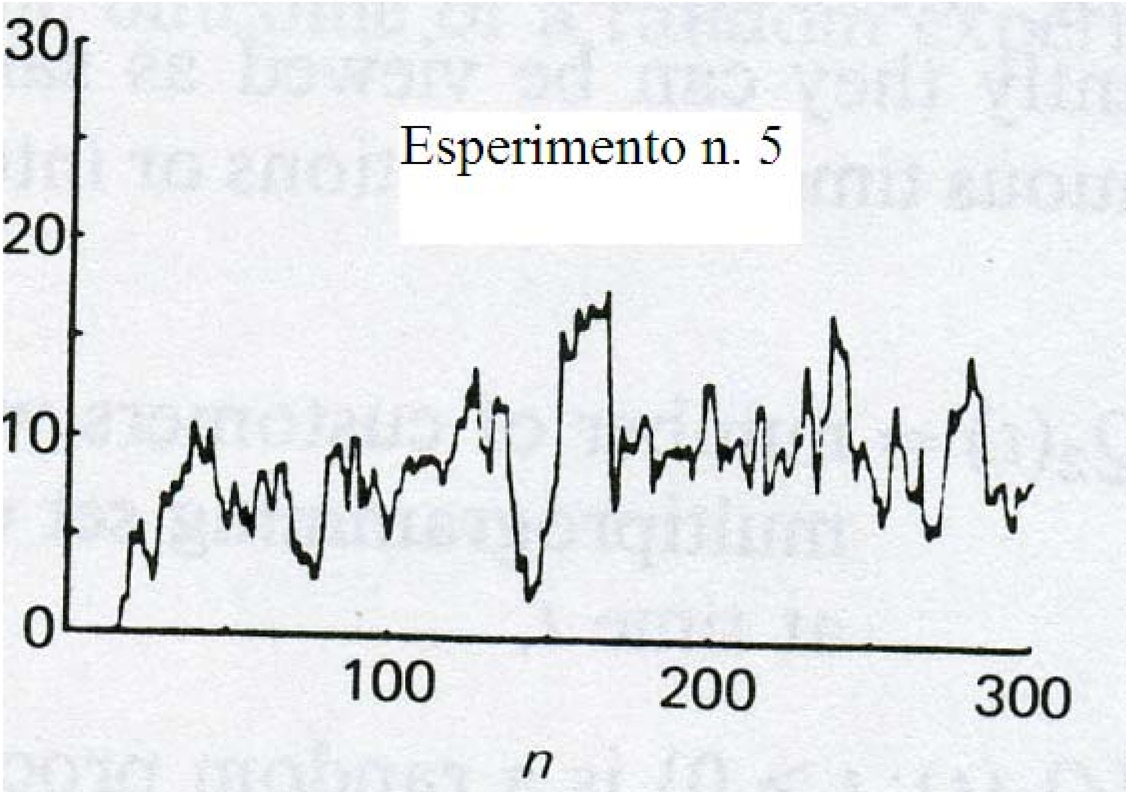



Esperimento n. 2
30
20
10
0
100
200
300
n


Esperimento n. 3
30
20
10
0
100
200
300
n



$W_{2,n}$
30
20
10
0
Esperimento n. 4
100
200
300
n



30
20
10
0
Esperimento n. 5
100
200
300
n

Altre sequenze di output interessanti:

- $\{R_n : n = 1,2,\ldots.\}$: tempo sperimentato dall' n-esimo utente che richiede servizio al terminale dopo aver completato un ciclo di elaborazione (terminale → terminale);

Fissato un intervallo di osservazione dove δ è un periodo di campionamento:

$$(n-1)\delta < t < n\delta$$

- $\{Q_{2,n} : n = 1,2,\ldots.\}$: lunghezza media della CODA 2 nei periodi di osservazione;

- $\{U_{3,n} : n = 1,2,\ldots.\}$: tempo di occupazione (fattore di utilizzo) del processore nei periodi di osservazione;

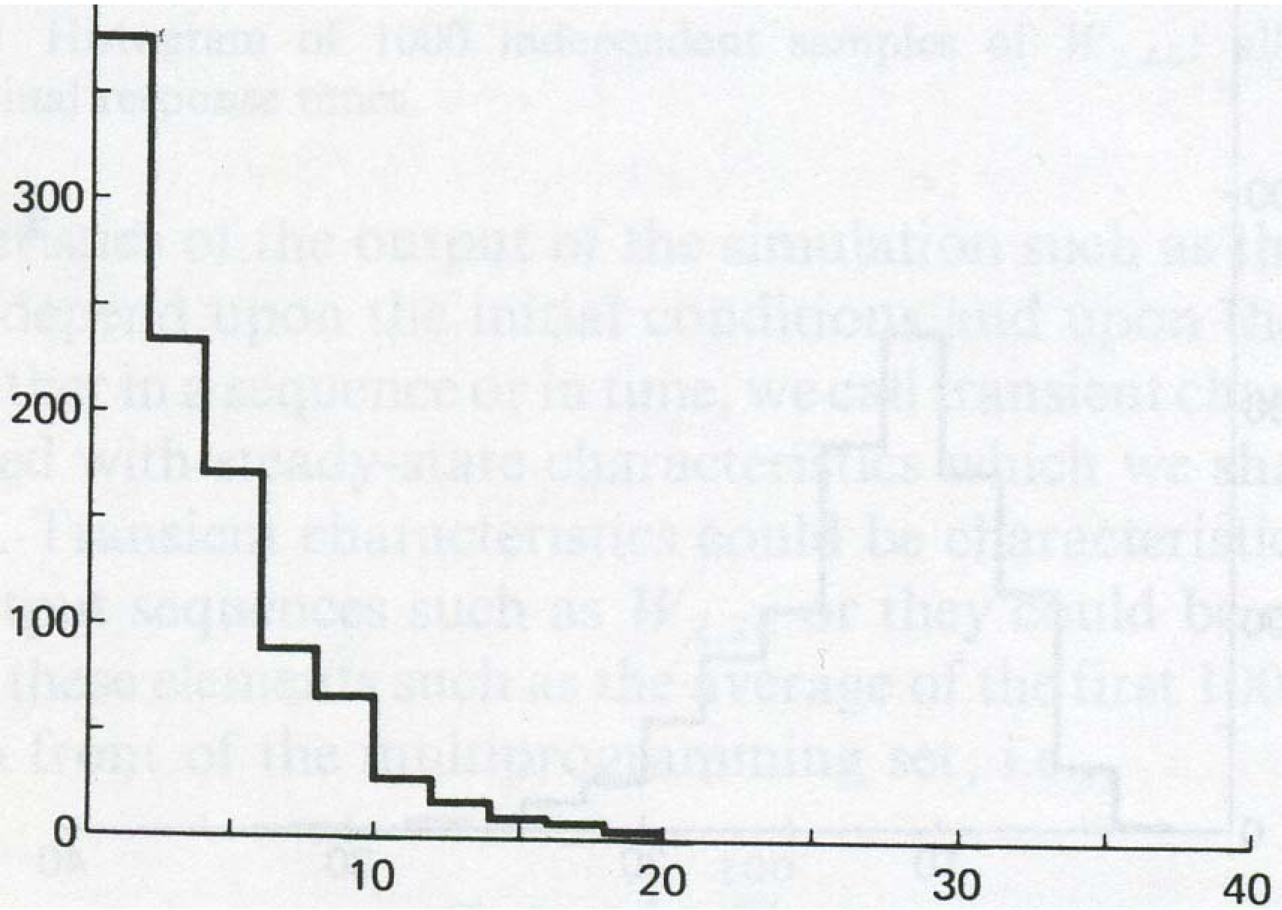
Come calcolare $Q_{2,n}$ e $U_{3,n}$?

FASE TRANSIENTE DEL SIMULATORE

Consideriamo $\{W_{2,n} : n = 1,2,\ldots.\}$ ed in particolare la v.c. $W_{2,20}$,

$W_{2,20}$: tempo di attesa in coda del 20-esimo utente che esce dalla CODA 2.

Eseguiamo 1000 distinti esperimenti assumendo come condizione iniziale la presenza di tutti gli utenti (25) nei terminali.



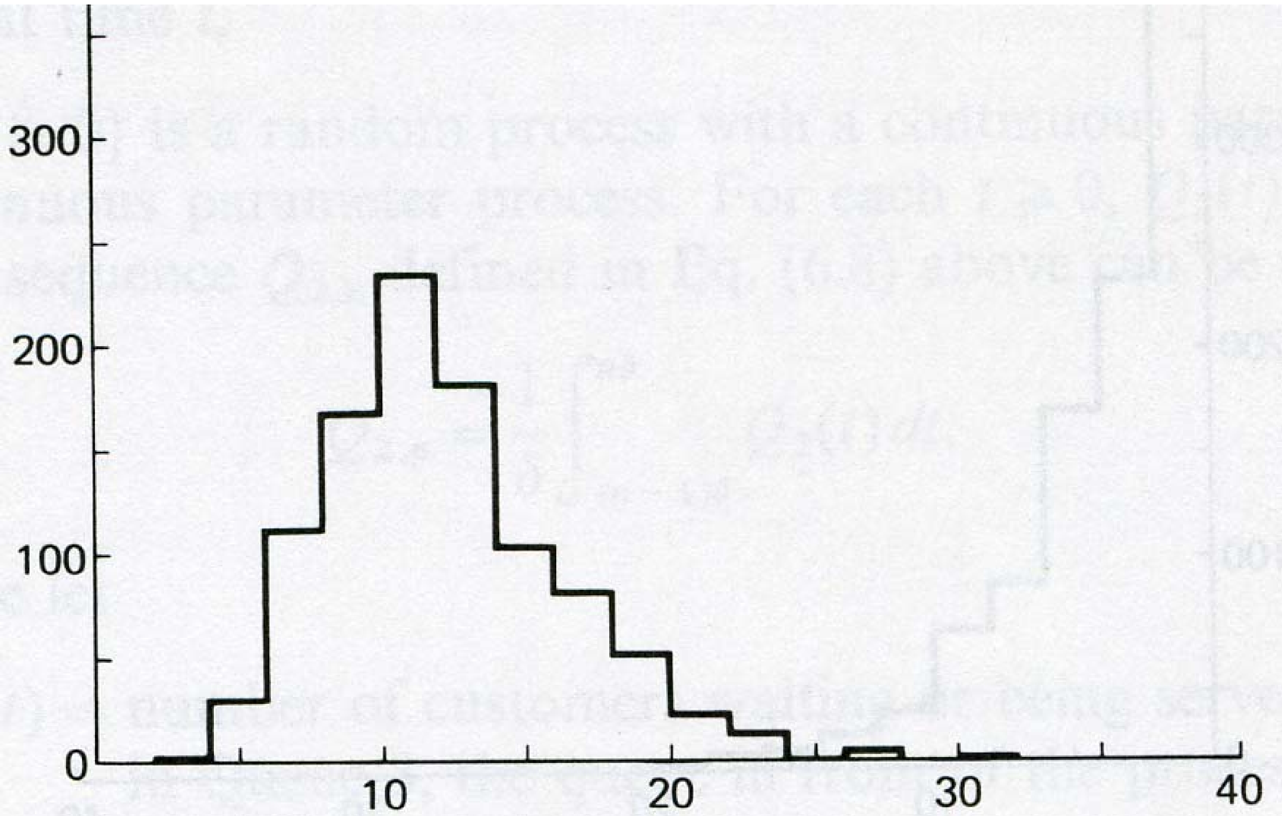
$W_{2,20}$

Istogramma della v.c $W_{2,20}$ con 1000 osservazioni e condizione iniziale con tutti gli utenti nei terminali.

Modifichiamo le condizioni iniziali:

- 20 utenti nei terminali;
- 5 utenti nella CODA 2.

Eseguiamo 1000 esperimenti:

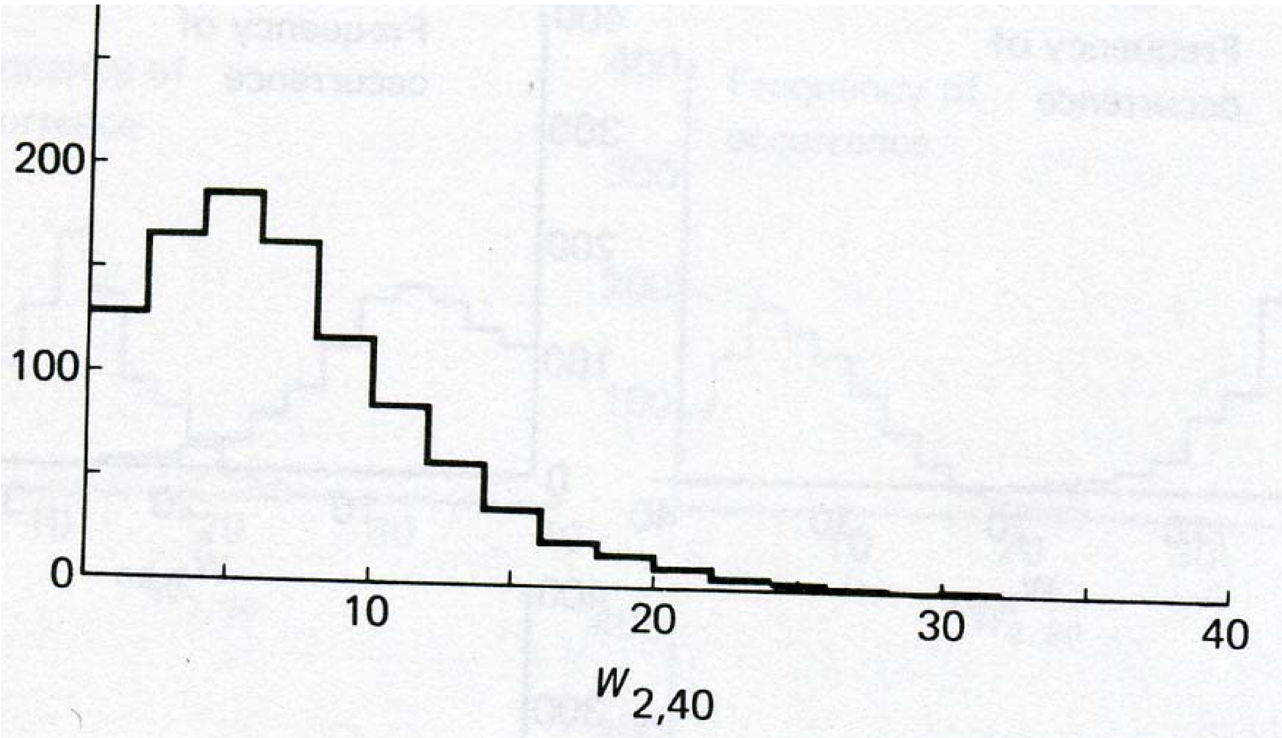


$W_{2,20}$

Istogramma della v.c $W_{2,20}$ con 1000 osservazioni e condizione iniziale con 5 utenti nella CODA 2 e i rimanti nei terminali

Consideriamo adesso la v.c $W_{2,40}$

- 1000 esperimenti;
- condizione iniziale con 25 utenti nei terminali.



Istogramma della v.c $W_{2,40}$ con 1000 osservazioni e condizione iniziale con tutti gli utenti nei terminali.

Fase transiente e terminating simulation fortemente condizionate dallo stato iniziale.

Stima di misure prestazionali nel transitorio:

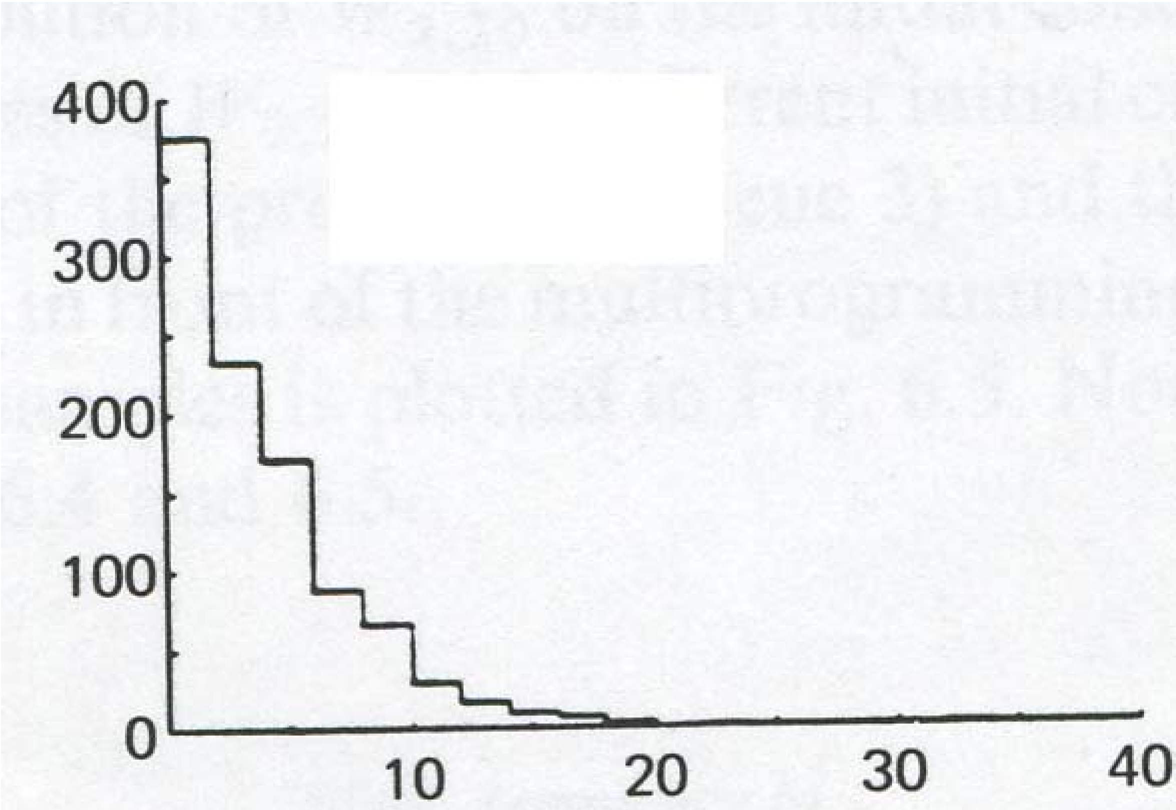
$$\bar{W}_2 = \frac{1}{100} \sum_{n=1}^{100} W_{2,n}.$$

Comportamento stazionario delle sequenze di output

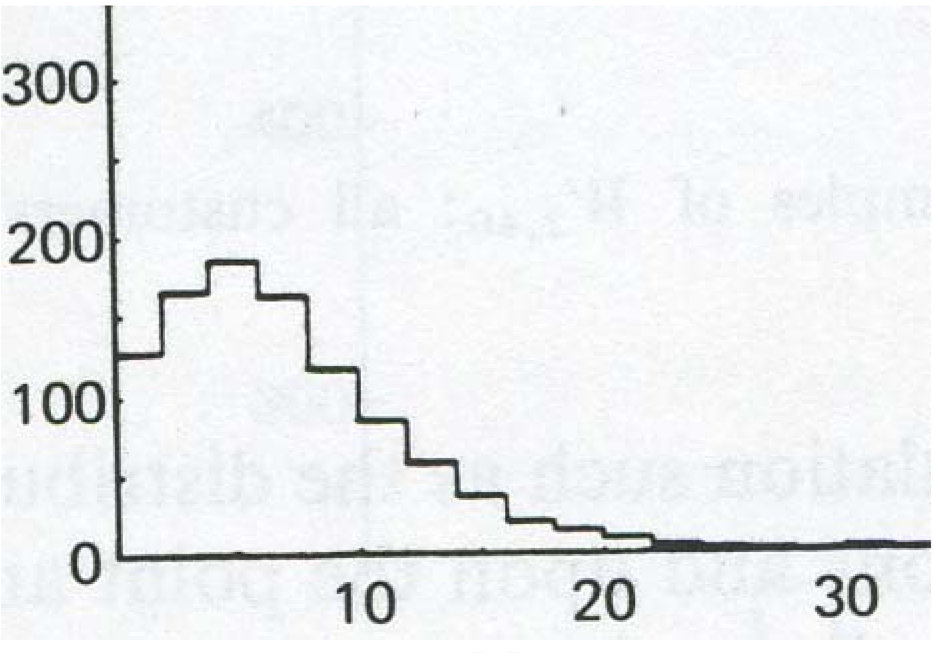
Consideriamo sempre la sequenza $\{W_{2,n} : n = 1,2,\ldots.\}$

Cosa succede per n→ oo?

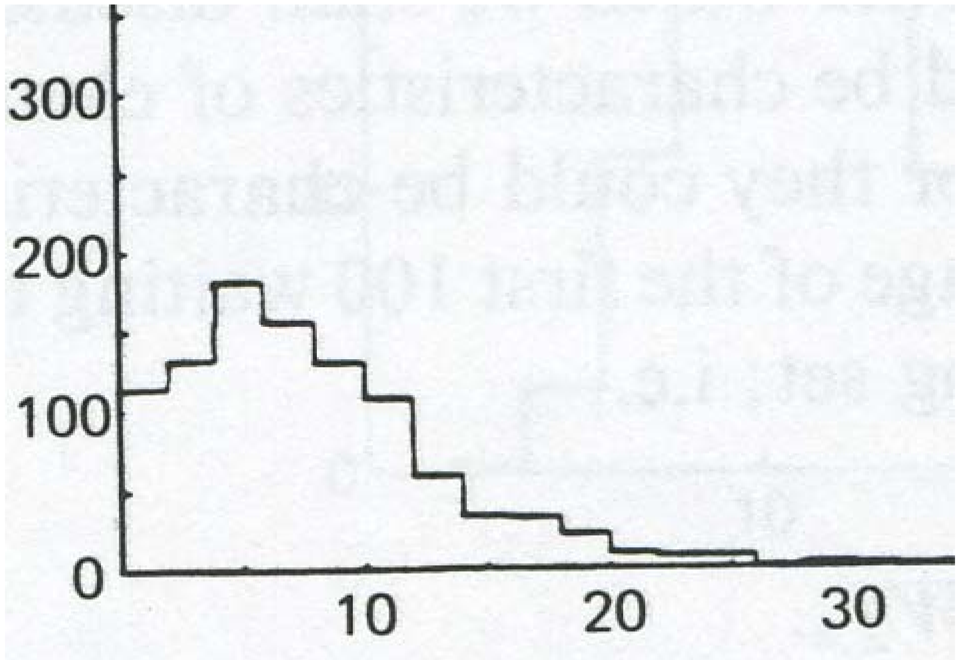
Eseguiamo 1000 esperimenti per le seguenti v.c. : $W_{2,20}$, $W_{2,40}$, $W_{2,60}$, $W_{2,80}$, $W_{2,100}$, $W_{2,120}$.



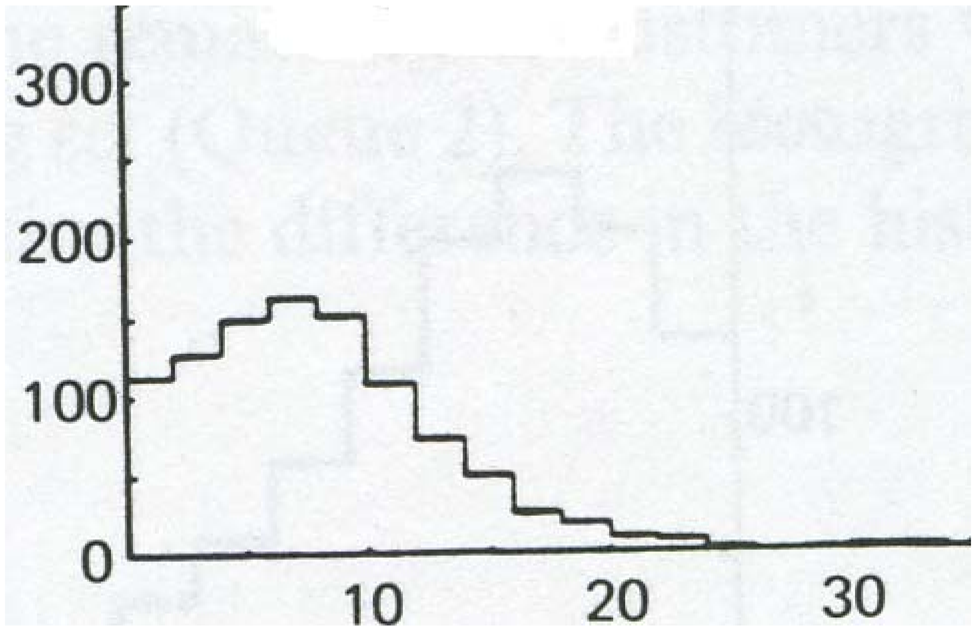
$W_{2,20}$



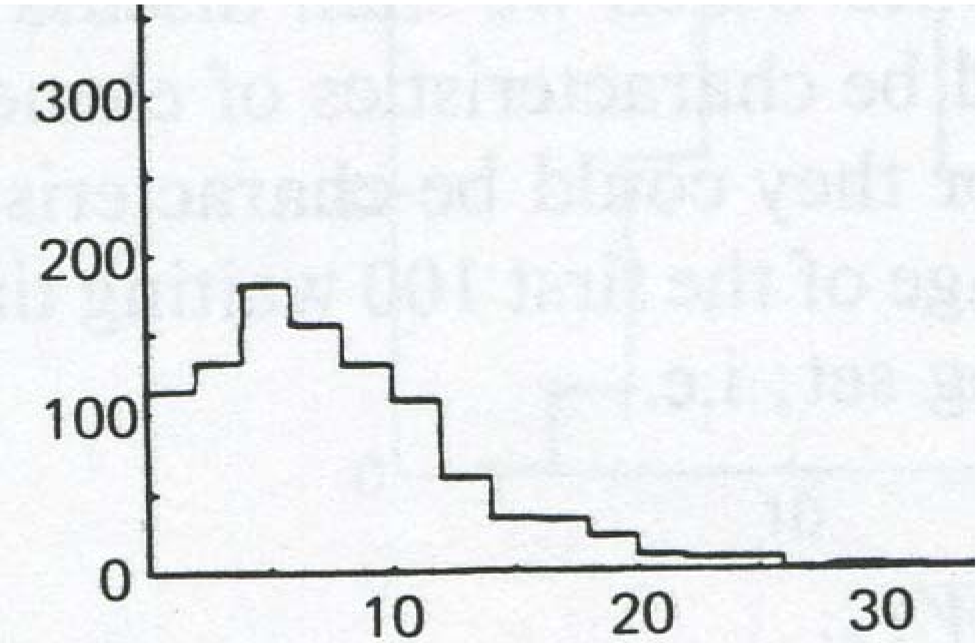
$W_{2,40}$



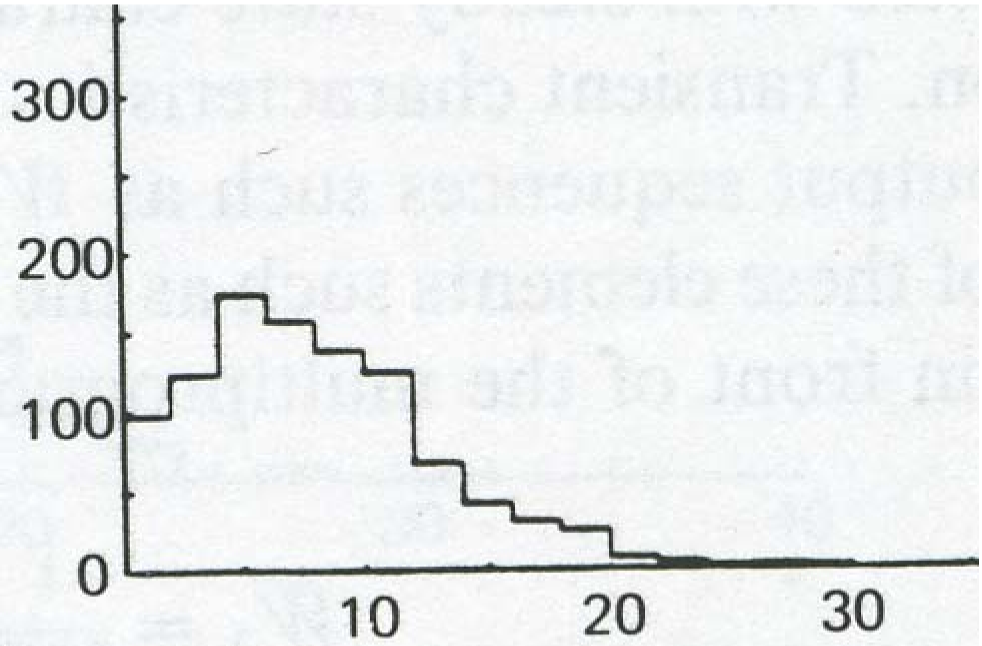
$W_{2,60}$



$W_{2,80}$



$W_{2,100}$

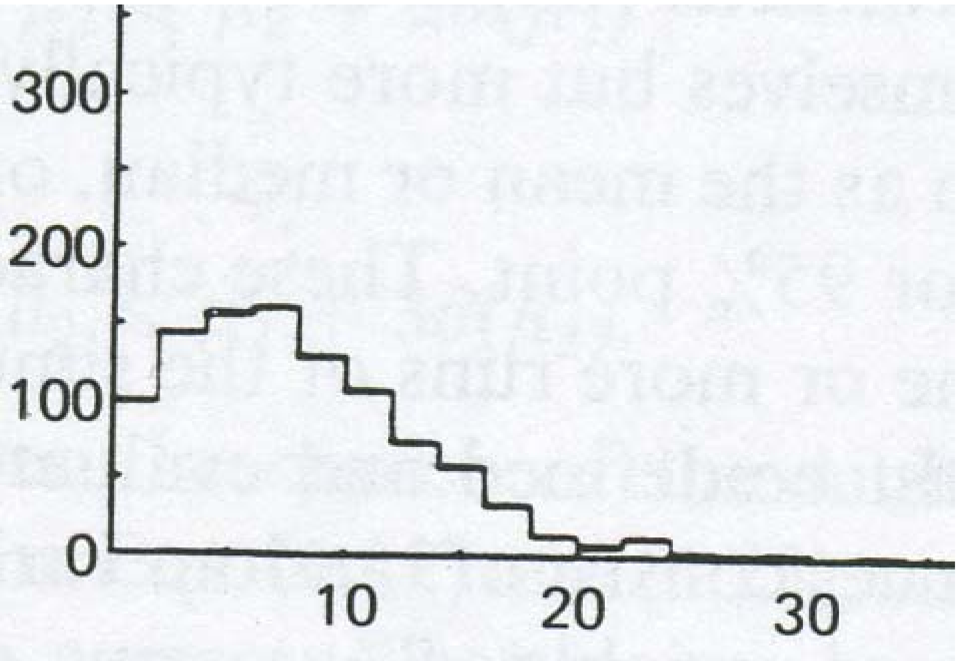


$W_{2,120}$

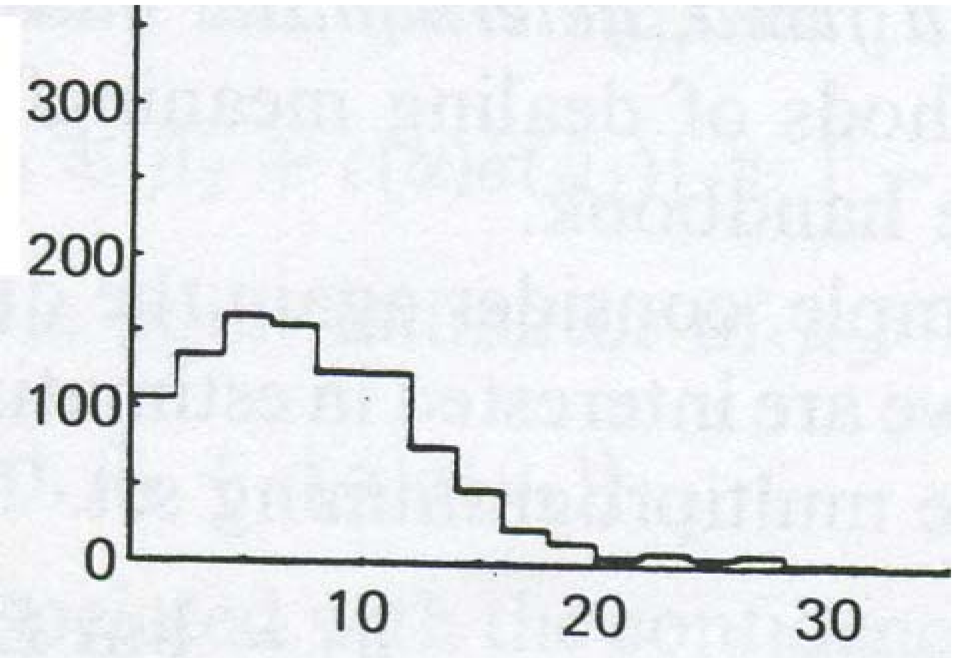
Notiamo la sostanziale convergenza degli istogrammi all'aumentare di n.

Quale influenza ha lo stato iniziale?

Eseguiamo 1000 esperimenti per le seguenti v.c. : $W_{2,100}$, $W_{2,120}$
Condizioni iniziali: 5 utenti in CODA 2 i rimanenti nei terminali.



$W_{2,100}$



$W_{2,120}$

Si osserva che lo stato iniziale non influenza le distribuzioni.

Stima puntuale degli indici di prestazione e intervalli di confidenza: