Sannolikhet och Statistik, F9

Rolf Larsson

Uppsala Universitet

23 september 2014

Idag

- 3.10.1 : Funktioner av en slumpvariabel
- 3.11 : Stora talens lag
- 3.12 : Centrala gränsvärdessatsen

Funktioner av en slumpvariabel

Antag att X är kontinuerligt likformigt fördelad på intervallet $(0, 1)$. Vilken är fördelningen för $Y = -\lambda \ln X$, där $\lambda > 0$?

Funktioner av en slumpvariabel

Theorem

Låt X vara en slumpvariabel med fördelningsfunktion $F_X(x)$. Antag att $g(x)$ är en strikt monoton kontinuerlig funktion. Bilda $V = g(X)$ med fördelningsfunktion $F_V(v)$.

1. *Om $g(x)$ är växande så är $F_V(v) = F_X\left\{g^{-1}(v)\right\}$.*
2. *Om $g(x)$ är avtagande och X är kontinuerlig så är $F_V(v) = 1 - F_X\left\{g^{-1}(v)\right\}$.*
3. *Om $g(x)$ är avtagande och X är diskret så är $F_V(v) = P\left\{X \geq g^{-1}(v)\right\}$.*

Stora talens lag

- Antag att $X_1, \ldots, X_n$ är okorrelerade med samma väntevärde μ och samma varians σ^2.
- Låt

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

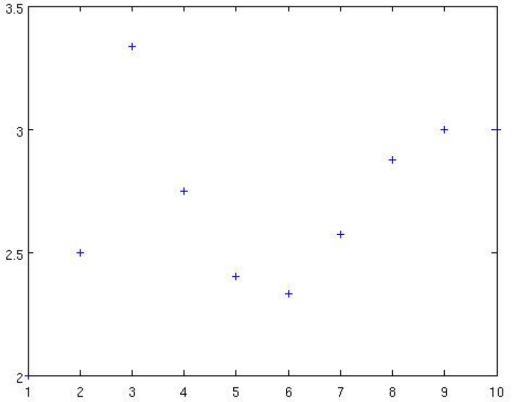
- Då gäller att

$$E\left(\overline{X}_n\right) = \mu$$

$$V\left(\overline{X}_n\right) = \frac{1}{n}\sigma^2$$

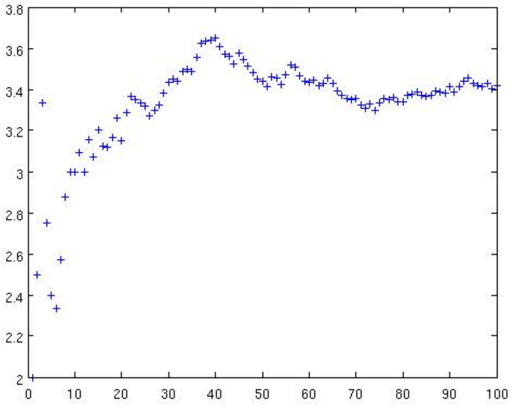
Stora talens lag

Kasta en välbalanserad tärning *n* gånger och beräkna det uppdaterade medelvärdet.
För n=10 :



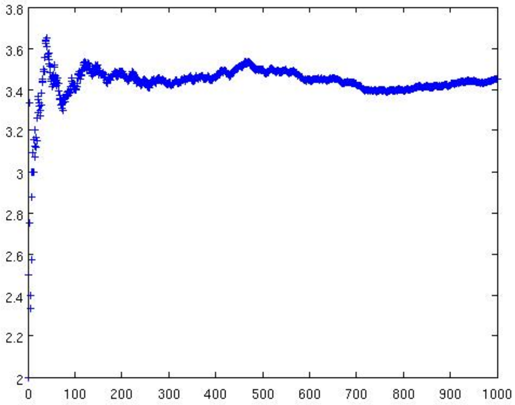
Stora talens lag

För n=100 :



Stora talens lag

För n=1000 :



Stora talens lag

Theorem (Stora talens lag)

Låt $X_1, X_2, \ldots$ *vara en följd av oberoende slumpvariabler, alla med samma väntevärde* μ *och standardavvikelse* σ.

Låt $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Då gäller för varje $\varepsilon > 0$ *att*

$$P\left(|\overline{X}_n - \mu| > \varepsilon\right) \to 0$$

då $n \to \infty$.

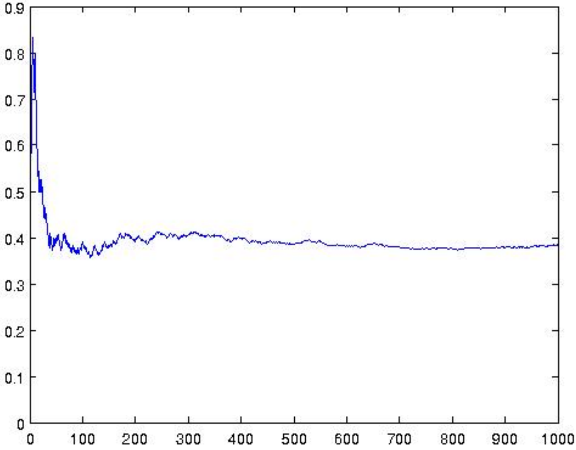
Stora talens lag

www.deskstore.com

Kasta ett häftstift upprepade gånger. Låt händelsen $A = \{$ hatten nedåt $\}$.

Stora talens lag

Plotta relativ frekvens av gångerna som A inträffar mot antalet kast :



Stora talens lag

Låt $X_i = 1$ om A är sann och 0 annars. Då är $\overline{X}_n$ den relativa frekvensen att A inträffar efter n försök och

$$\mu = E(X_i) = 1 * P(A) + 0 * P(A^c) = P(A).$$

Theorem (Bernoullis sats)

Betrakta en följd av oberoende upprepningar av ett försök där händelsen A inträffar med sannolikhet $P(A)$ i ett enskilt försök.
Då konvergerar den relativa frekvensen för A mot $P(A)$ då antalet försök går mot oändligheten.

Centrala gränsvärdessatsen

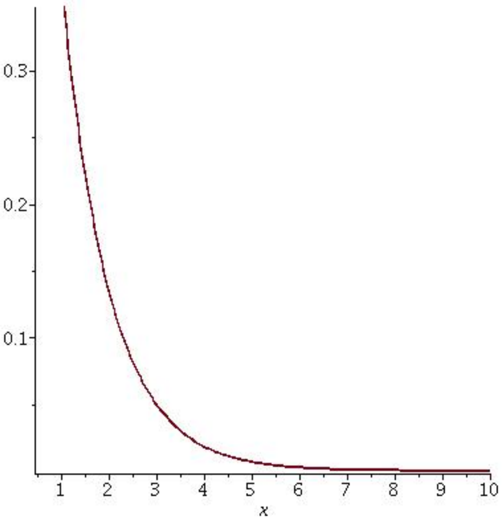
Låt $X_1, X_2, X_3, \ldots$ vara en följd av oberoende exponentialfördelade slumpvariabler med parameter 1. Bilda successiva medelvärden

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Plotta sannolikhetstätheten för $\overline{X}_n$ för olika n.

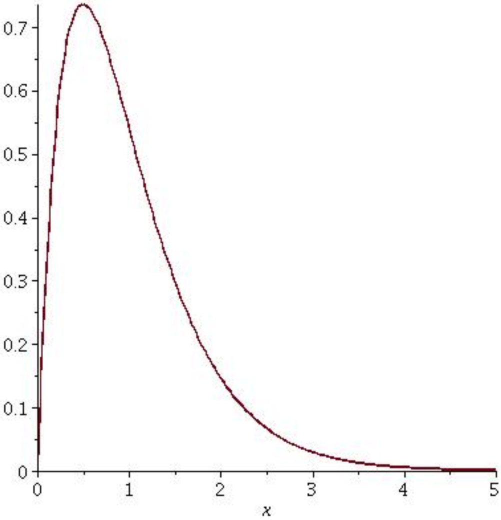
Centrala gränsvärdessatsen

$n = 1$:



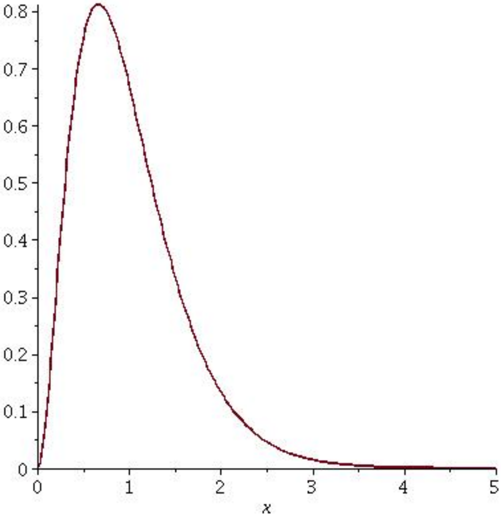
Centrala gränsvärdessatsen

$n = 2$:



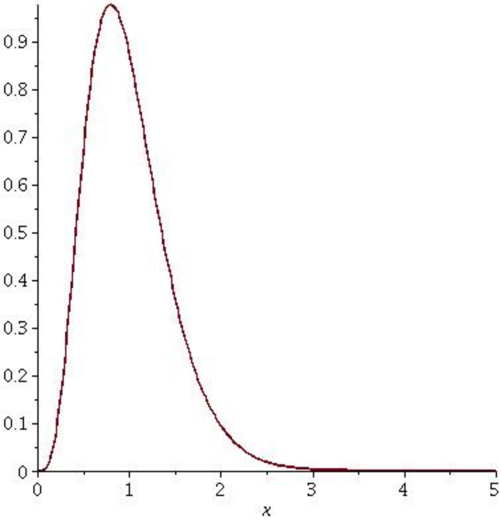
Centrala gränsvärdessatsen

$n = 3$:



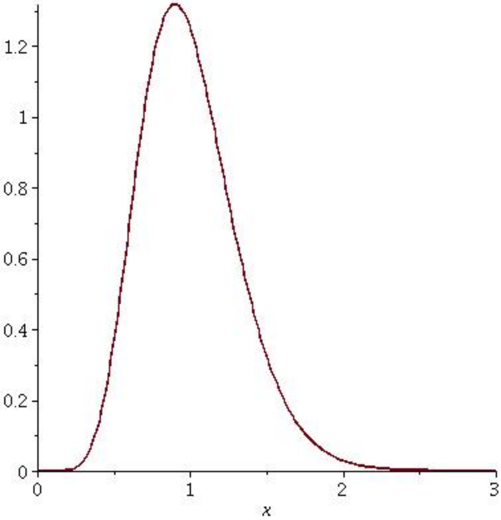
Centrala gränsvärdessatsen

$n = 5$:



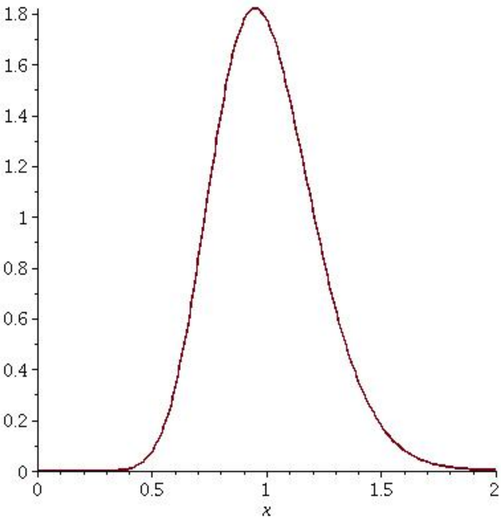
Centrala gränsvärdessatsen

$n = 10$:



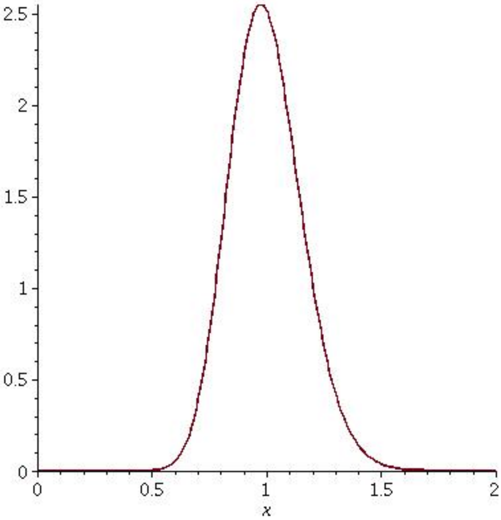
Centrala gränsvärdessatsen

$n = 20$:



Centrala gränsvärdessatsen

$n = 40$:



Centrala gränsvärdessatsen

- Låt $X_1, ..., X_n$ vara oberoende och likafördelade med väntevärde μ och varians σ^2.
- Låt $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.
- Då gäller att

$$E\left(\overline{X}_n\right) = \mu$$
$$V\left(\overline{X}_n\right) = \frac{1}{n}\sigma^2$$

- Det följer att

$$E\left\{\frac{\sqrt{n}}{\sigma}\left(\overline{X}_n - \mu\right)\right\} = 0$$
$$V\left\{\frac{\sqrt{n}}{\sigma}\left(\overline{X}_n - \mu\right)\right\} = 1$$

Centrala gränsvärdessatsen

Den standardiserade variabeln $\frac{\sqrt{n}}{\sigma}(\overline{X}_n - \mu)$ "konvergerar mot" $N(0,1)$ då $n \longrightarrow \infty$.

Theorem (Centrala gränsvärdessatsen)

Låt $X_1, X_2, \ldots$ vara oberoende och likafördelade slumpvariabler med väntevärde μ och standardavvikelse σ, där $0 < \sigma < \infty$, och låt $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.
För godtyckliga $a < b$ gäller då att, då $n \rightarrow \infty$,

$$P\left\{a < \frac{\sqrt{n}}{\sigma}(\overline{X}_n - \mu) < b\right\} \rightarrow \Phi(b) - \Phi(a)$$

där Φ är fördelningsfunktionen för $N(0,1)$.

Tolkning : För stora n är $\overline{X}_n$ approximativt $N\left(\mu, \frac{\sigma^2}{n}\right)$.

Centrala gränsvärdessatsen

Tio personers längder mäts och avrundas till hela centimeter. Sedan summeras de avrundade längderna. Hur stor är sannolikheten för att summan av de avrundade längderna avviker från summan av de verkliga längderna med mindre än två centimeter ?

1. Gissa först på ett av följande alternativ :
   0.20 0.57 0.97
2. Beräkna sannolikheten med hjälp av normalapproximation.

Centrala gränsvärdessatsen

Hur stort ska n vara för att $\overline{X}_n \approx N\left(\mu, \frac{\sigma^2}{n}\right)$ ska vara en bra approximation ?

Beräkning av $P(\overline{X}_n \leq a)$ där $\overline{X}_n$ är medelvärdet av $X_1, \ldots, X_n$ där alla $X_i \sim \mathrm{Re}(0, 1)$.

n	a	approximation	exakt
1	0.0252	0.050	0.025
2	0.1642	0.050	0.054
3	0.2258	0.050	0.052
5	0.2876	0.050	0.050
10	0.3498	0.050	0.050

Centrala gränsvärdessatsen

Beräkning av $P(\overline{X}_n \leq a)$ där $\overline{X}_n$ är medelvärdet av $X_1, \ldots, X_n$ där alla $X_i \sim \mathrm{Exp}(1)$.

n	a	approximation	exakt
3	0.0503	0.050	0.00005
5	0.2644	0.050	0.011
10	0.4798	0.050	0.025
30	0.6997	0.050	0.037
100	0.8355	0.050	0.044
300	0.9050	0.050	0.046
1000	0.9480	0.050	0.048

Centrala gränsvärdessatsen

Beräkning av $P(\overline{X}_n \leq a)$ där $\overline{X}_n$ är medelvärdet av $X_1, \ldots, X_n$ där alla $X_i \sim \mathrm{Exp}(1)$.

n	a	approximation	exakt
3	0.8538	0.400	0.472
5	0.8867	0.400	0.455
10	0.9199	0.400	0.439
30	0.9538	0.400	0.422
100	0.9747	0.400	0.412
300	0.9854	0.400	0.407
1000	0.9920	0.400	0.404