

SF1901: Demonstration VT 2017

Outline

- 1 Fördelnings- och täthetsfunktion
- 2 Multivariat normalfördelning
- 3 Simulering av slumpstal
- 4 Stora talens lag
- 5 Monte Carlo-simulering av väntevärden
- 6 Konfidensintervall

Mål med dagens föreläsning:

- Repetera några viktiga begrepp från kursen.
- Illustrera vissa koncept och resultat med hjälp av MATLAB.

Täthetsfunktionen f_X för en kontinuerlig stokastisk variabel X definieras av

$$P(X \in [a, b)) = \int_a^b f_X(x) dx,$$

och fördelningsfunktionen F_X ges av

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(y) dy.$$

Täthetsfunktionen f_X för en kontinuerlig stokastisk variabel X definieras av

$$P(X \in [a, b)) = \int_a^b f_X(x) dx,$$

och fördelningsfunktionen F_X ges av

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(y) dy.$$

MATLAB har kommandon för de vanligaste fördelningarna. Exempel (täthetsfunktioner):

- Normalfördelningen: `normpdf(x,mu, sigma)`.
- Exponentialfördelningen: `exppdf(x,mu)`; parametern är väntevärdet (jfr. Blom).
- Gammafördelningen: `gammampdf(x,a, b)`.
- ...

Exempel: Normalfördelningen

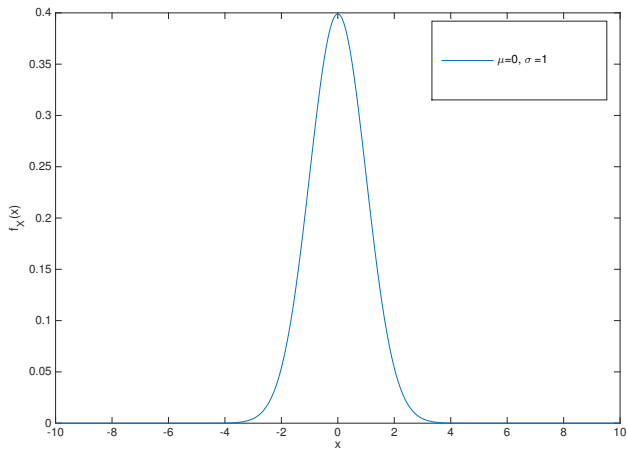
Plot av $N(\mu, \sigma)$ för olika μ, σ .

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

```
%% Tathetsfunktion for normalfordelning dx = 0.01;
x = -10:dx:10; % Skapar en vektor med dx som inkrement
y = normpdf(x, 0, 1);
plot(x,y), hold on
z = normpdf(x,-1,0.1);
plot(x,z,'r')
w = normpdf(x,2,2);
plot(x,w,'g')
```

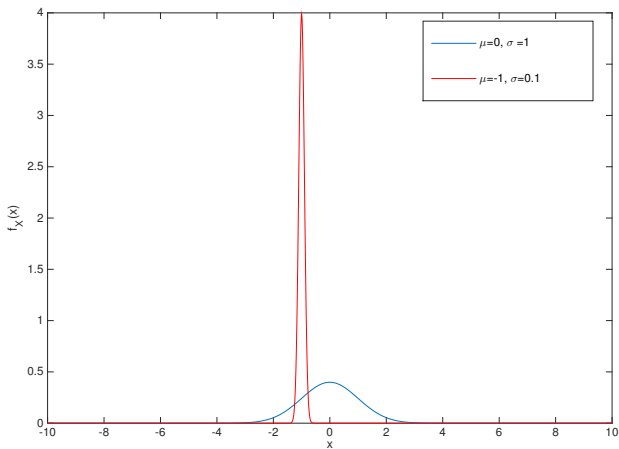
Exempel: Normalfördelningen

Plot av $N(\mu, \sigma)$ för olika μ, σ .



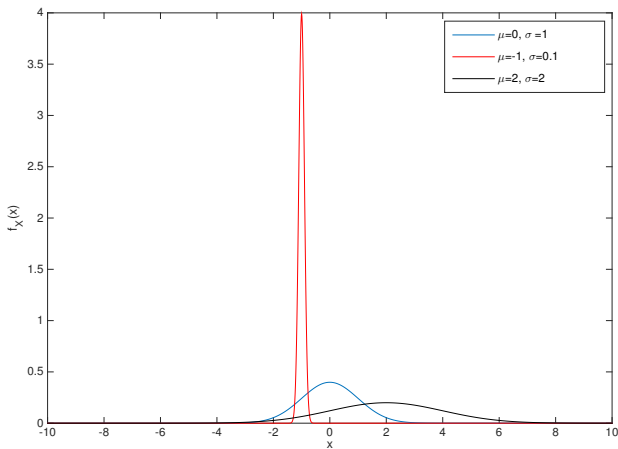
Exempel: Normalfördelningen

Plot av $N(\mu, \sigma)$ för olika μ, σ .



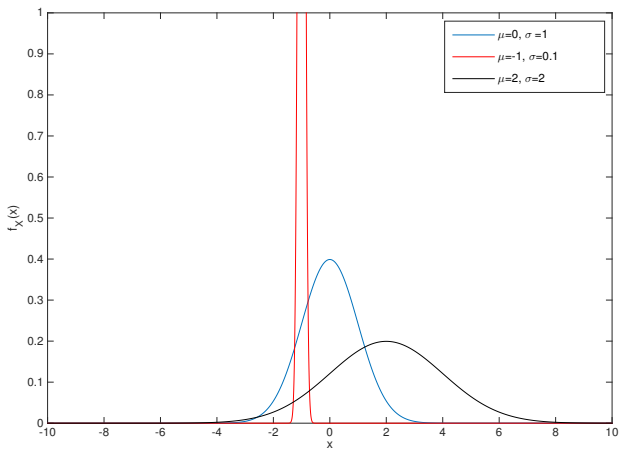
Exempel: Normalfördelningen

Plot av $N(\mu, \sigma)$ för olika μ, σ .



Exempel: Normalfördelningen

Plot av $N(\mu, \sigma)$ för olika μ, σ .



Exempel: Gammafördelningen

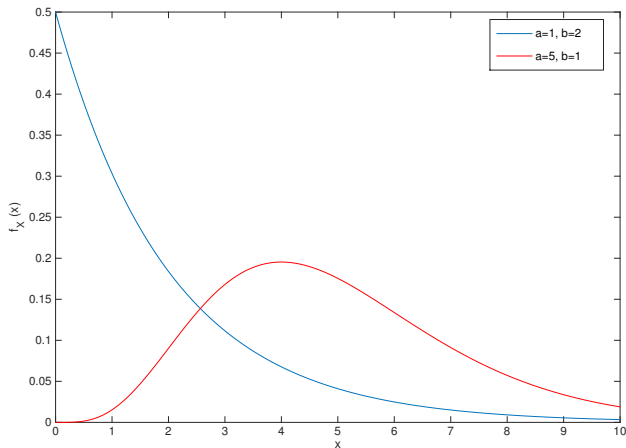
Liknande kod illustrerar Gammafördelningen för olika parametrar a, b :

$$f_X(x) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{-x/b}; \quad (\text{skillnad mot Blom})$$

```
%% Tathetsfunktion for gammafordelning dx = 0.01;
x = -0:dx:10; % Skapar en vektor med dx som inkrement
y = gampdf(x,1,2);
plot(x,y), hold on
z = gampdf(x,5,1);
plot(x,z,'r')
```

Exempel: Gammafördelningen

Liknande kod illustrerar Gammafördelningen för olika parametrar a, b :



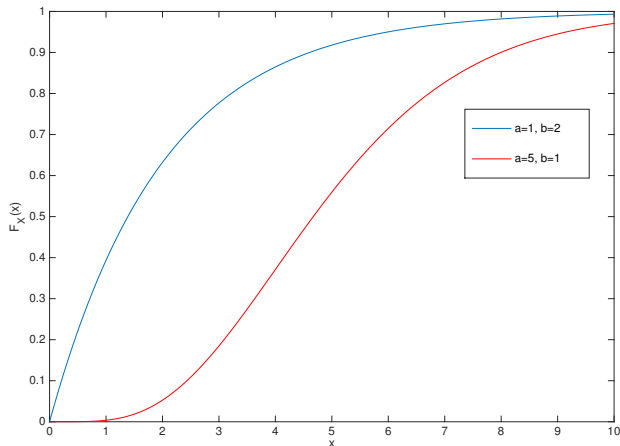
Exempel: Gammafördelningen

Finns även funktioner för de vanligaste fördelningsfunktionerna. För Gammafördelningen:

```
% Fördelningsfunktion för gammafördelning  
dx = 0.01;  
x = -0:dx:10; % Skapar en vektor med dx som inkrement  
y = gamcdf(x,1,2);  
plot(x,y), hold on  
z = gamcdf(x,5,1);  
plot(x,z,'r')
```

Exempel: Gammafördelningen

Finns även funktioner för de vanligaste fördelningsfunktionerna. För Gammafördelningen:



Betrakta nu en stokastisk variabel X med täthetsfunktion

$$f_X(x) = \lambda e^{-\frac{x}{\lambda}} + \frac{\lambda}{x}, \quad x \in [1, 10]$$

för ett specifikt λ . Möjliga värden på λ ?

Betrakta nu en stokastisk variabel X med täthetsfunktion

$$f_X(x) = \lambda e^{-\frac{x}{\lambda}} + \frac{\lambda}{x}, \quad x \in [1, 10]$$

för ett specifikt λ . Möjliga värden på λ ? Vill lösa ekvationen

$$\int_1^{10} f_X(x) dx = 1.$$

Betrakta nu en stokastisk variabel X med täthetsfunktion

$$f_X(x) = \lambda e^{-\frac{x}{\lambda}} + \frac{\lambda}{x}, \quad x \in [1, 10]$$

för ett specifikt λ . Möjliga värden på λ ? Vill lösa ekvationen

$$\int_1^{10} f_X(x) dx = 1.$$

Kan lösa den numeriskt, ger approximationen $\lambda = 0.4267$.

Betrakta nu en stokastisk variabel X med täthetsfunktion

$$f_X(x) = \lambda e^{-\frac{x}{\lambda}} + \frac{\lambda}{x}, \quad x \in [1, 10]$$

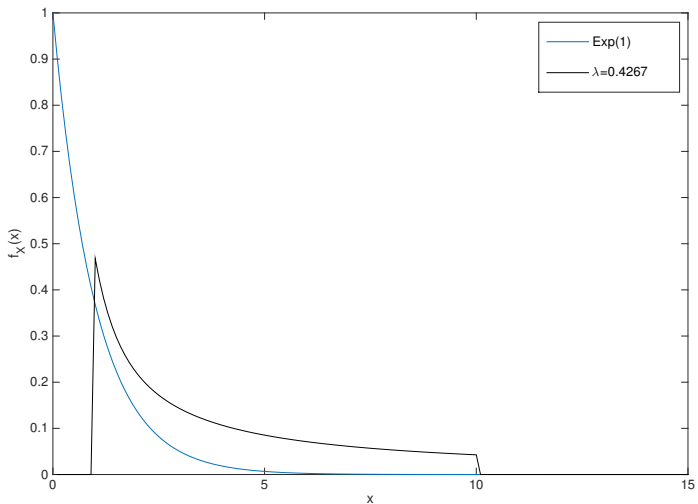
för ett specifikt λ . Möjliga värden på λ ? Vill lösa ekvationen

$$\int_1^{10} f_X(x) dx = 1.$$

Kan lösa den numeriskt, ger approximationen $\lambda = 0.4267$. Kan vidare jämföra täthetsfunktionen för X med t.ex. exponentialfördelningen.

```
%% Jämförelse av täthetsfunktioner dx = 0.1;
x = 0:dx:15; % Skapar en vektor med dx som inkrement
mu = 1;
y = exppdf(x, mu); % exponential-fordelningen
plot(x,y), hold on
lambda = 0.4267;
f=(lambda*exp(-x/lambda)+lambda./x).*(x >= 1 & x <= 10);
plot(x,f)
```

Jämförelse av täthetsfunktionen för $Exp(1)$ och den s.v. X ($\lambda = 0.4267$).
Skillnader?

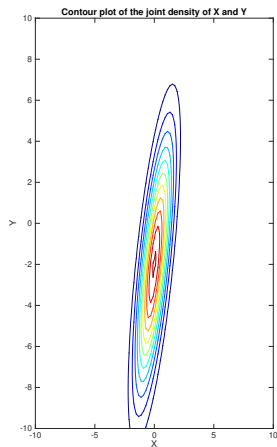
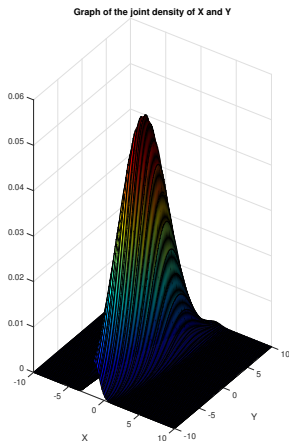


Multivariat normalfördelning (i tvådimensioner) bestäms av fem parametrar:

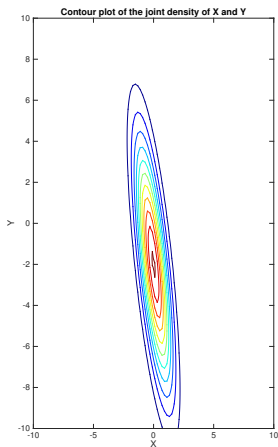
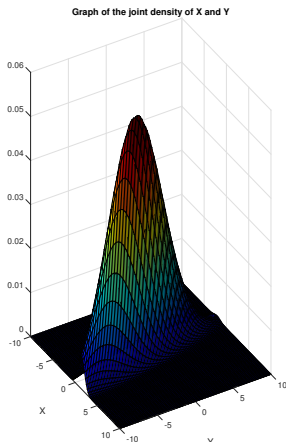
$$\mu_X, \sigma_X, \mu_Y, \sigma_Y, \rho,$$

där $\mu_X, \mu_Y \in \mathbb{R}$, $\sigma_X, \sigma_Y \in (0, \infty)$, $\rho \in [-1, 1]$. Kod för att rita upp fördelningen:

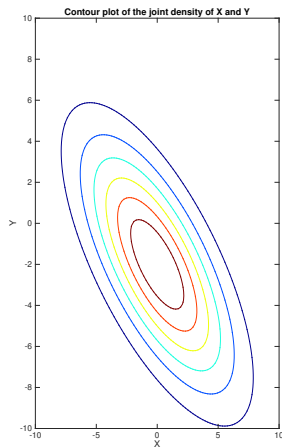
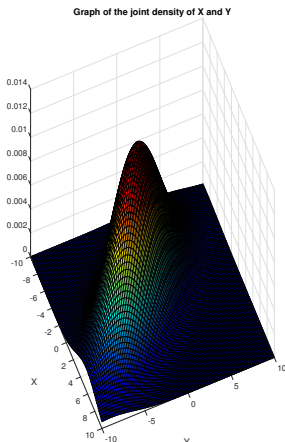
```
% Multivariat normal
mux = 0; muy = -2; sigmax = 1; sigmay = 4; rho = 0.7;
plot_mvnpdf(mux, muy, sigmax, sigmay, rho)
```



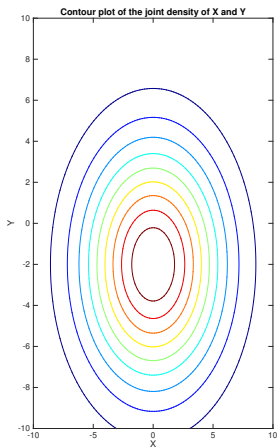
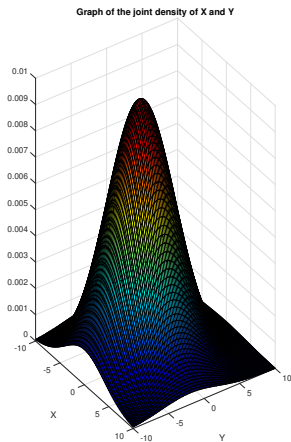
Vad svarar parametrarna mot? Hur påverkar olika värden plottens utseende?



Vad svarar parametrarna mot? Hur påverkar olika värden plottens utseende?



Vad svarar parametrarna mot? Hur påverkar olika värden plottens utseende?



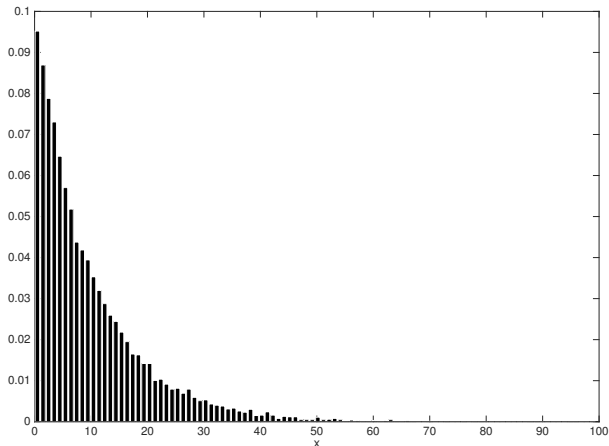
Vad svarar parametrarna mot? Hur påverkar olika värden plottens utseende?

Med hjälp av MATLAB (eller annan programvara) kan vi *simulera* slumptal.

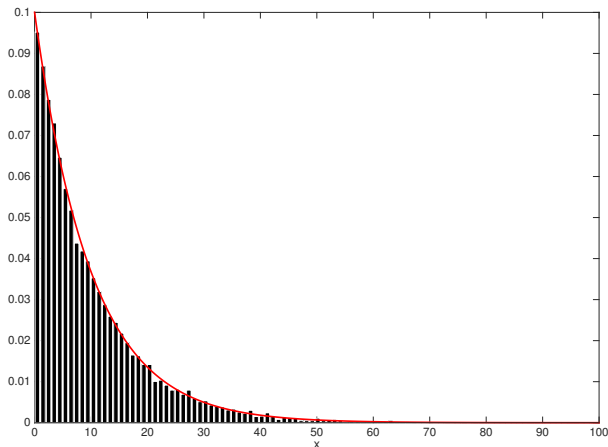
Med hjälp av MATLAB (eller annan programvara) kan vi *simulera* slumpstal. Följande experiment simulerar N slumpstal från $\text{Exp}(10)$, ritar det erhållna histogrammet samt den sanna täthetsfunktionen.

```
%% Simulering av slumpstal
mu = 10;
N = 1e4;
y = exprnd(mu, N, 1); % Genererar N exp-slumpstal
hist_density(y); % Skapar ett normaliserat histogram
t = linspace(0, 100, N/10); % Vektor med N/10 punkter
hold on
plot(t, exppdf(t, mu), 'r') % 'r' betyder rod linje
hold off
```

Med hjälp av MATLAB (eller annan programvara) kan vi *simulera* slumpstal. Följande experiment simulerar N slumpstal från $Exp(10)$, ritar det erhållna histogrammet samt den sanna täthetsfunktionen.

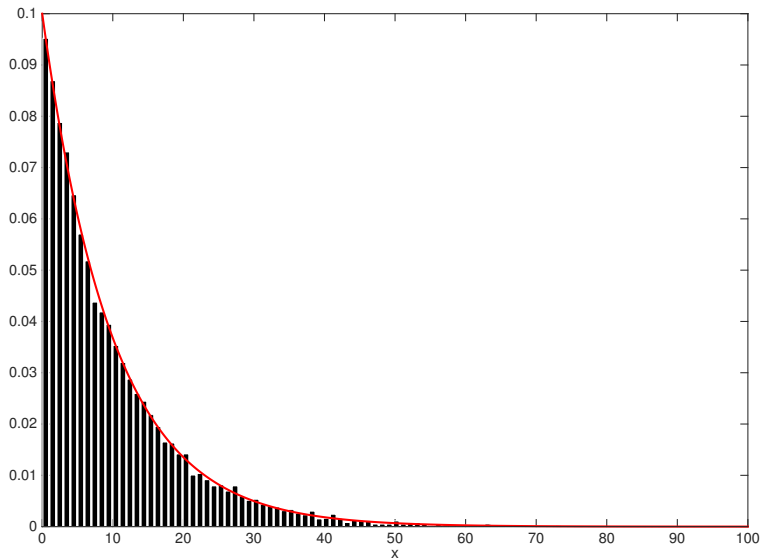


Med hjälp av MATLAB (eller annan programvara) kan vi *simulera* slumpstal. Följande experiment simulerar N slumpstal från $Exp(10)$, ritar det erhållna histogrammet samt den sanna täthetsfunktionen.



Upprepar experimentet.

- Hur förhåller sig histogrammet till den röda linjen (täthetsfunktionen)?
- Hur förklaras variationen kring linjen?



Theorem (Stora talens lag)

För oberoende, likafördelade s.v. $X_1, X_2, \dots$,

$$S_n := \frac{1}{n} \sum_{i=1}^n X_i \rightarrow E[X_1], \quad n \rightarrow \infty.$$

Theorem (Stora talens lag)

För oberoende, likafördelade s.v. $X_1, X_2, \dots$,

$$S_n := \frac{1}{n} \sum_{i=1}^n X_i \rightarrow E[X_1], \quad n \rightarrow \infty.$$

Hur ser konvergensen ut? Kan simulera de s.v. $X_1, X_2, \dots$ och studera S_n 's beteende.

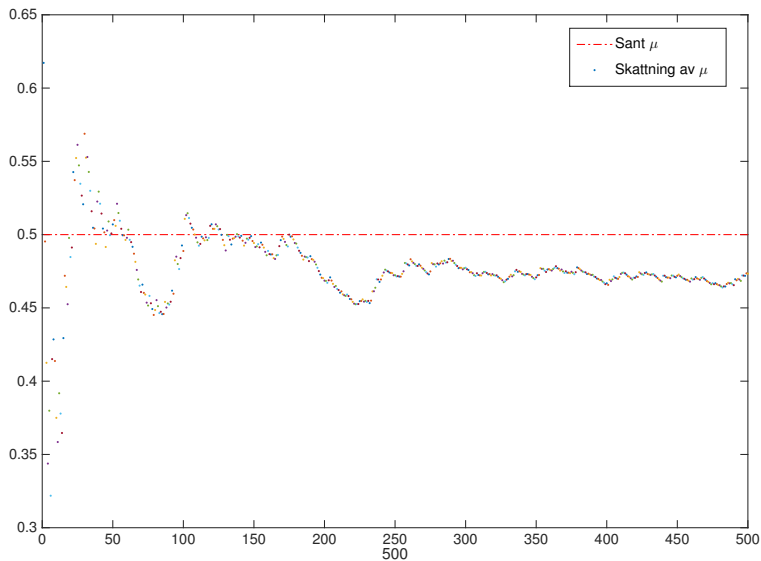
Theorem (Stora talens lag)

För oberoende, likafördelade s.v. $X_1, X_2, \dots$,

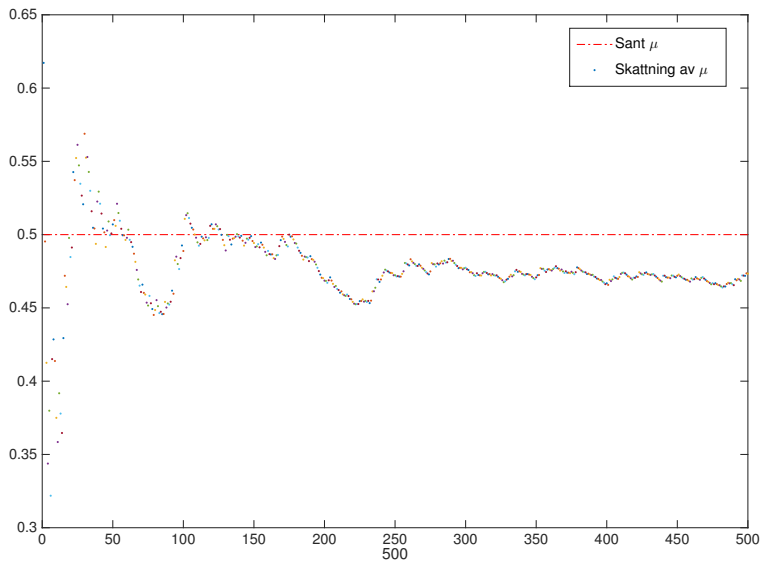
$$S_n := \frac{1}{n} \sum_{i=1}^n X_i \rightarrow E[X_1], \quad n \rightarrow \infty.$$

```
%% Stora talens lag
mu = 0.5;
M = 500;
X = exprnd(mu, M, 1);
plot(ones(M, 1)*mu, 'r-.')
hold on
for k = 1:M
    plot(k, mean(X(1:k)), '.')
    xlabel(num2str(k)), pause(0.001)
end
hold off
```

En illustration för $\text{Exp}(0.5)$, $n = 500$.



En illustration för $\text{Exp}(0.5)$, $n = 500$. Vad händer om vi ändrar n ?



Vi kan vara mer precisa gällande vad stora talens lag säger:

Theorem

För likafördelade $X_1, X_2, \dots$, med $E[X_1] = \mu$, gäller för varje $\epsilon > 0$

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \mu\right| < \epsilon\right) \rightarrow 1, \quad n \rightarrow \infty.$$

Vi kan använda detta för att med hjälp av en dator skatta väntevärden

¹http://homepages.rpi.edu/~angel/MULTISCALE/metropolis_u/lam1949.pdf

Vi kan vara mer precisa gällande vad stora talens lag säger:

Theorem

För likafördelade $X_1, X_2, \dots$, med $E[X_1] = \mu$, gäller för varje $\epsilon > 0$

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \mu\right| < \epsilon\right) \rightarrow 1, \quad n \rightarrow \infty.$$

Vi kan använda detta för att med hjälp av en dator skatta väntevärden:
Monte Carlo-metoden.

¹http://homepages.rpi.edu/~angel/MULTISCALE/metropolis_u/lam1949.pdf

Vi kan vara mer precisa gällande vad stora talens lag säger:

Theorem

För likafördelade $X_1, X_2, \dots$, med $E[X_1] = \mu$, gäller för varje $\epsilon > 0$

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \mu\right| < \epsilon\right) \rightarrow 1, \quad n \rightarrow \infty.$$

Vi kan använda detta för att med hjälp av en dator skatta väntevärden:

Monte Carlo-metoden.

- Den grundläggande idén har funnits inom matematiken sedan åtminstone 1700-talet.
- Stanislaw Ulam och John von Neumann utvecklade under 1940-talet metoder för att göra "tärningskast" på en dator.
- Första vetenskapliga artikeln av Ulam och Nicholas Metropolis ("The Monte Carlo Method" (1949)¹).
- Arbetet utfördes i Los Alamos i samband med Manhattanprojektet.

¹http://homepages.rpi.edu/~angel/MULTISCALE/metropolis_Ulam1949.pdf

Exempel: X, Y oberoende s.v. $X \sim \text{Exp}(4)$, $Y \sim N(0, 1)$. Vill beräkna $E[e^{X \cos(Y)}]$.

Exempel: X, Y oberoende s.v. $X \sim \text{Exp}(4)$, $Y \sim N(0, 1)$. Vill beräkna $E[e^{X \cos(Y)}]$.

Definitionen säger

$$E[e^{X \cos(Y)}] = \int_{x>0} \int_{y \in (-\infty, \infty)} e^{x \cos(y)} f(x, y) dx dy,$$

där $f = f_{X,Y}$ är tthetsfunktionen för den *gemensamma* fördelningen för paret (X, Y) . Oberoende $\rightarrow f_{X,Y}(x, y) = f_X(x)f_Y(y)$. Får därför

$$E[e^{X \cos(Y)}] = \int_0^\infty \int_{-\infty}^\infty e^{x \cos(y)} \frac{1}{4} e^{-\frac{x}{4}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy dx.$$

Exempel: X, Y oberoende s.v. $X \sim \text{Exp}(4)$, $Y \sim N(0, 1)$. Vill beräkna $E[e^{X \cos(Y)}]$.

Definitionen säger

$$E[e^{X \cos(Y)}] = \int_{x>0} \int_{y \in (-\infty, \infty)} e^{x \cos(y)} f(x, y) dx dy,$$

där $f = f_{X,Y}$ är tthetsfunktionen för den *gemensamma* fridelnigen för paret (X, Y) . Oberoende $\rightarrow f_{X,Y}(x, y) = f_X(x)f_Y(y)$. Får därför

$$E[e^{X \cos(Y)}] = \int_0^\infty \int_{-\infty}^\infty e^{x \cos(y)} \frac{1}{4} e^{-\frac{x}{4}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy dx.$$

Kan istället skatta det sökta väntevärdet med Monte Carlo

Tag N stort och simulera slumpvariables $X_1, \dots, X_N$ och $Y_1, \dots, Y_N$ från $\text{Exp}(4)$ resp. $N(0, 1)$.

För varje $i = 1, \dots, N$, betrakta Z_i :

$$Z_i = e^{X_i \cos(Y_i)},$$

och forma medelvärdet $\bar{z}$ från det erhållna stickprovet.

Resultat vid tre körningar ($N = 10^5$):

```
%% Monte Carlo
```

```
N = 1e5;
```

```
X = exprnd(1/4,N,1);
```

```
Y = randn(N,1);
```

```
mean(exp(X.*cos(Y)));
```

• 1.1974

Tag N stort och simulera slumpvariables $X_1, \dots, X_N$ och $Y_1, \dots, Y_N$ från $\text{Exp}(4)$ resp. $N(0, 1)$.

För varje $i = 1, \dots, N$, betrakta Z_i :

$$Z_i = e^{X_i \cos(Y_i)},$$

och forma medelvärdet $\bar{z}$ från det erhållna stickprovet.

Resultat vid tre körningar ($N = 10^5$):

```
%% Monte Carlo
N = 1e5;
X = exprnd(1/4,N,1);
Y = randn(N,1);
mean(exp(X.*cos(Y)));
```

- 1.1974
- 1.1950

Tag N stort och simulera slumpvariables $X_1, \dots, X_N$ och $Y_1, \dots, Y_N$ från $\text{Exp}(4)$ resp. $N(0, 1)$.

För varje $i = 1, \dots, N$, betrakta Z_i :

$$Z_i = e^{X_i \cos(Y_i)},$$

och forma medelvärdet $\bar{z}$ från det erhållna stickprovet.

Resultat vid tre körningar ($N = 10^5$):

```
%% Monte Carlo
N = 1e5;
X = exprnd(1/4,N,1);
Y = randn(N,1);
mean(exp(X.*cos(Y)));
```

- 1.1974
- 1.1950
- 1.1949

Exempel: Beräkning av π .

Exempel: Beräkning av π .

Tag U, V oberoende, likformigt fördelade på $[-1, 1]$. Paret (U, V) tar då värden i $[-1, 1] \times [-1, 1]$ och kan ses som koordinater i enhetskuben i två dimensioner. Sannolikheten att hamna i enhetscirkeln ges av

$$P(\sqrt{U^2 + V^2} \leq 1),$$

inses lätt vara

$$\frac{\text{area av enhetscirkeln}}{\text{area av enhetskuben}} = \frac{\pi}{4} \approx 0.785$$

Exempel: Beräkning av π .

Tag U, V oberoende, likformigt fördelade på $[-1, 1]$. Paret (U, V) tar då värden i $[-1, 1] \times [-1, 1]$ och kan ses som koordinater i enhetskuben i två dimensioner. Sannolikheten att hamna i enhetscirkeln ges av

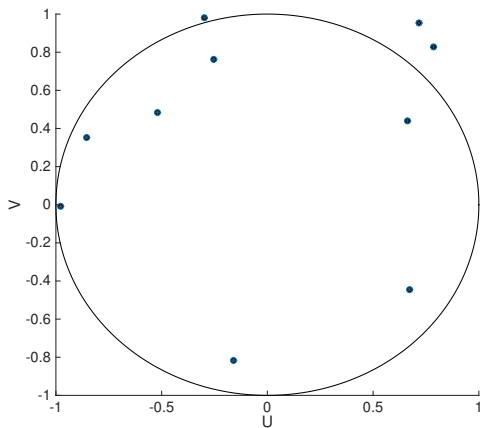
$$P(\sqrt{U^2 + V^2} \leq 1),$$

inses lätt vara

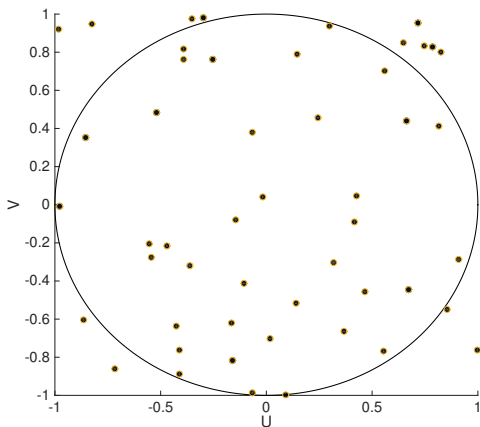
$$\frac{\text{area av enhetscirkeln}}{\text{area av enhetskuben}} = \frac{\pi}{4} \approx 0.785$$

Kan skatta π med Monte Carlo-metoden:

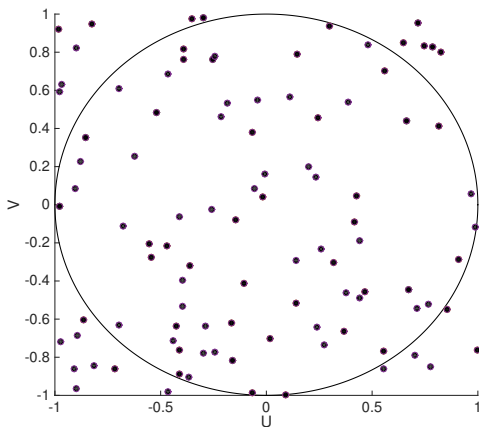
- Simulera $(U_1, V_1), \dots, (U_N, V_N)$ (N stort).
- För varje i kontrollera om $\sqrt{U_i^2 + V_i^2} \leq 1$ eller inte.
- Beräkna medelvärdet av motsvarande indikatorvariabler:
 $\frac{1}{N} \sum_{i=1}^N I\{\sqrt{U_i^2 + V_i^2} \leq 1\}.$



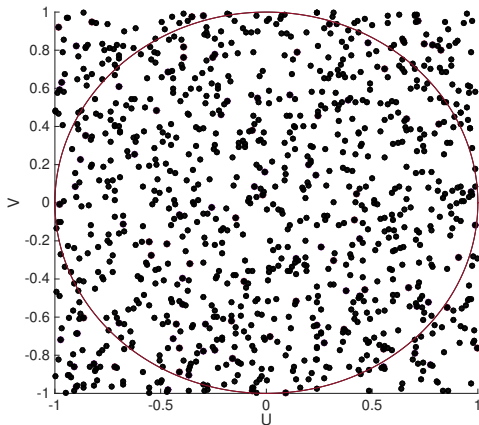
Estimate $\hat{\pi}$: 0.700 ($n = 10$),



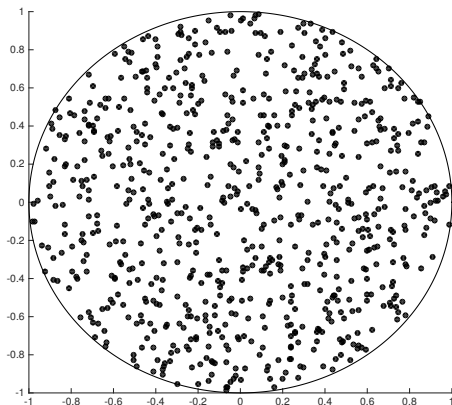
Estimate $\hat{\pi}$: 0.700 ($n = 10$), 0.720 ($n = 50$),



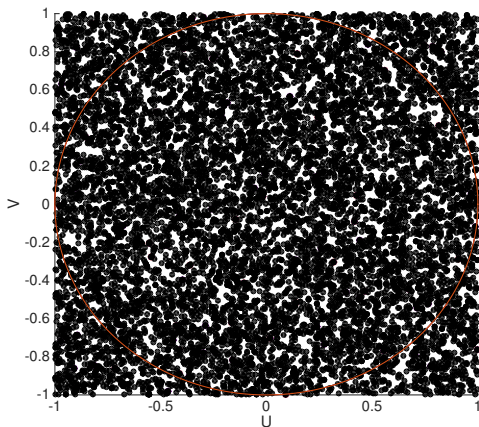
Estimate $\hat{\pi}$: 0.700 ($n = 10$), 0.720 ($n = 50$), 0.740 ($n = 100$),



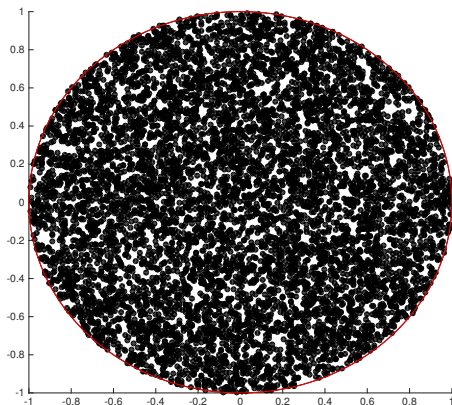
Estimate $\hat{\pi}$: 0.700 ($n = 10$), 0.720 ($n = 50$), 0.740 ($n = 100$),



Estimate $\hat{\pi}$: 0.700 ($n = 10$), 0.720 ($n = 50$), 0.740 ($n = 100$), 0.771 ($n = 10^3$),



Estimate $\hat{\pi}$: 0.700 ($n = 10$), 0.720 ($n = 50$), 0.740 ($n = 100$), 0.771 ($n = 10^3$),



Estimate $\hat{\pi}$: 0.700 ($n = 10$), 0.720 ($n = 50$), 0.740 ($n = 100$), 0.771 ($n = 10^3$), 0.783 ($n = 10^4$).

Ett konfidsensintervall med konfidsensgrad $1 - \alpha$ för en (okänd) parameter θ är ett intervall $I_\theta = (a(\mathbf{x}), b(\mathbf{x}))$ sådant att

$$P(a(\mathbf{X}) < \theta < b(\mathbf{X})) = 1 - \alpha,$$

där $\mathbf{x} = (x_1, \dots, x_n)$ är ett stickprov från $\mathbf{X} = (X_1, \dots, X_n)$, vars fördelning beror på den okända parametern.

Hur väl stämmer vår tolkning av konfidsensgraden i specifika exempel?

Ett konfidsensintervall med konfidsensgrad $1 - \alpha$ för en (okänd) parameter θ är ett intervall $I_\theta = (a(\mathbf{x}), b(\mathbf{x}))$ sådant att

$$P(a(\mathbf{X}) < \theta < b(\mathbf{X})) = 1 - \alpha,$$

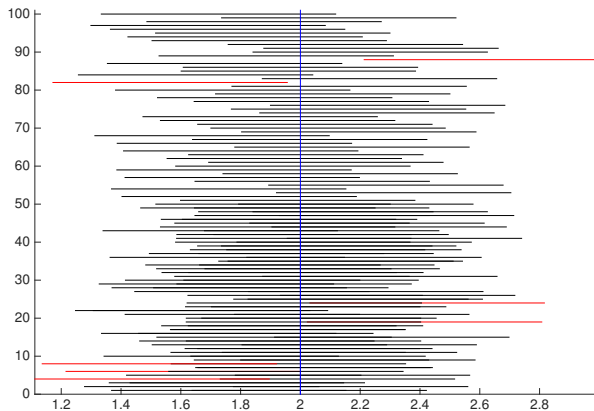
där $\mathbf{x} = (x_1, \dots, x_n)$ är ett stickprov från $\mathbf{X} = (X_1, \dots, X_n)$, vars fördelning beror på den okända parametern.

Hur väl stämmer vår tolkning av konfidsensgraden i specifika exempel?

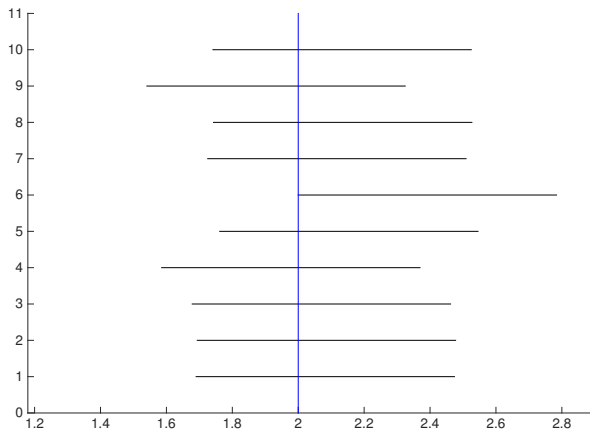
Experiment: Baserat på 100 stickprov av storlek $n = 25$, skatta 100 95%-iga konfidsensintervall för väntevärdet i en $N(\mu, \sigma)$ -fördelning (μ, σ givna).

- Hur många intervall innehåller det sanna värdet $\mu = 2$?
- Hur påverkar de olika parametrarna μ, σ, n och α resultatet?

Parametrar: $\mu = 2$, $\sigma = 1$, $n = 25$, $\alpha = 0.05$; $\sum_k I\{\mu \in I_\mu^k\} = 95$

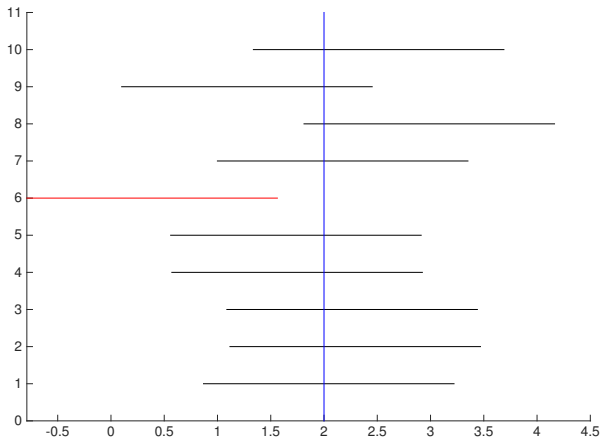


Parametrar: $\mu = 2$, $\sigma = 1$, $n = 25$, $\alpha = 0.05$; $\sum_k I\{\mu \in I_\mu^k\} = 95$



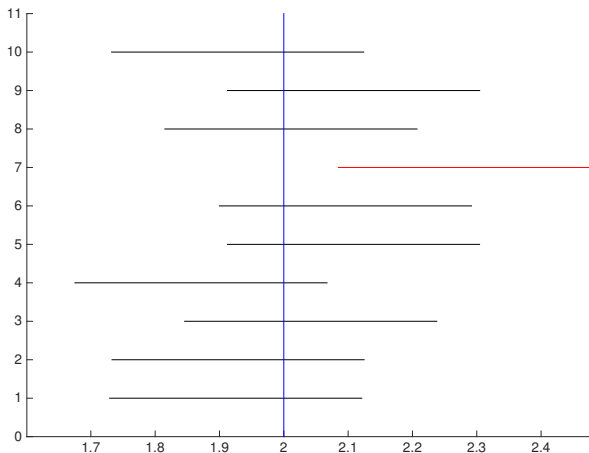
Parametrar: $\mu = 2$, $\sigma = 3$, $n = 25$, $\alpha = 0.05$;

Parametrar: $\mu = 2$, $\sigma = 3$, $n = 25$, $\alpha = 0.05$; $\sum_k I\{\mu \in I_\mu^k\} = 96$



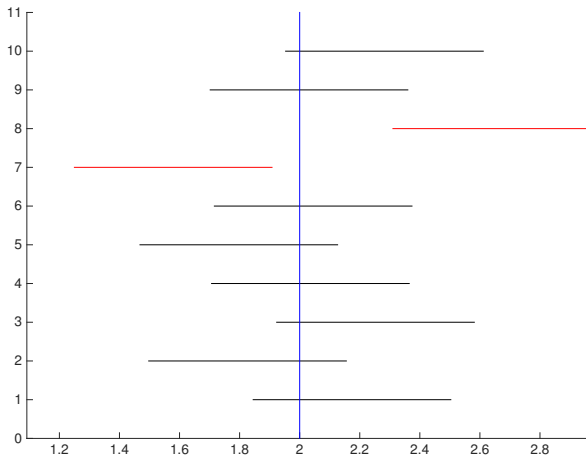
Parametrar: $\mu = 2$, $\sigma = 1$, $n = 100$, $\alpha = 0.05$;

Parametrar: $\mu = 2$, $\sigma = 1$, $n = 100$, $\alpha = 0.05$; $\sum_k I\{\mu \in I_\mu^k\} = 95$



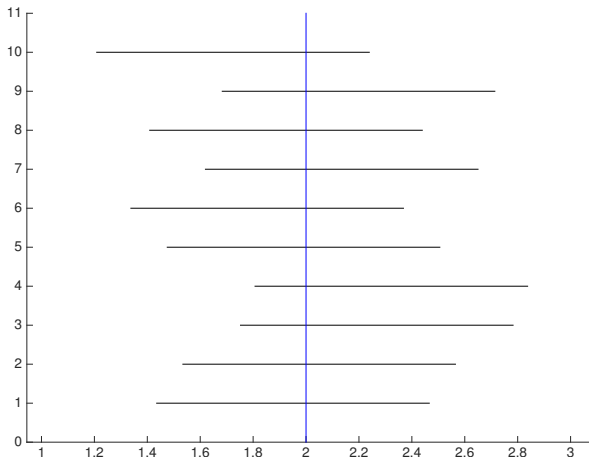
Parametrar: $\mu = 2$, $\sigma = 1$, $n = 25$, $\alpha = 0.10$;

Parametrar: $\mu = 2$, $\sigma = 1$, $n = 25$, $\alpha = 0.10$; $\sum_k I\{\mu \in I_\mu^k\} = 83$



Parametrar: $\mu = 2$, $\sigma = 1$, $n = 25$, $\alpha = 0.01$;

Parametrar: $\mu = 2$, $\sigma = 1$, $n = 25$, $\alpha = 0.01$; $\sum_k I\{\mu \in I_\mu^k\} = 98$



Intervallen beräknade under antagandet om en känd standardavvikelse.
Intervalllängd:

$$2\lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Intervallen beräknade under antagandet om en känd standardavvikelse.
Intervalllängd:

$$2\lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

- Större σ ger bredare intervall (med en faktor κ om $\tilde{\sigma} = \kappa\sigma$).

Intervallen beräknade under antagandet om en känd standardavvikelse.
Intervalllängd:

$$2\lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

- Större σ ger bredare intervall (med en faktor κ om $\tilde{\sigma} = \kappa\sigma$).
- Större n ger smalare intervall (med en faktor $1/\sqrt{\kappa}$, om $\tilde{n} = \kappa n$).

Intervallen beräknade under antagandet om en känd standardavvikelse.
Intervalllängd:

$$2\lambda_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

- Större σ ger bredare intervall (med en faktor κ om $\tilde{\sigma} = \kappa\sigma$).
- Större n ger smalare intervall (med en faktor $1/\sqrt{\kappa}$, om $\tilde{n} = \kappa n$).
- För $\tilde{\alpha} < \alpha$ gäller $\lambda_{\tilde{\alpha}/2} > \lambda_{\alpha/2}$. T.ex. $\lambda_{0.025} = 1.96$, $\lambda_{0.005} = 2.58$.
Minskat (ökat) α ger bredare (smalare) intervall.