

Stochastic variables  
(Stokastiska variabler)

Random variables  
slumpvariabler

Diskreta s.v.

Kontinuerliga s.v.

}

distribution



moment, väntevärde, varians,  
avvikelse

statistik vs sannolikhet

sannolikhet vs rimlighet  
(probability) (likelihood)

Exempel:

Låt  $X$  vara antalet brunnöga barn ett visst par föräldrar föder. Säg att de har två barn; då kan  $X$  anta värden  $0, 1, 2$ .

Vi skriver  $x=1$  om ett av barnen har bruna ögon.  
 $X$  är en diskret s.v.

---

Exempel:

Låt  $T$  vara klockslaget då strömanvändningen  
i stadsnötet när sin topp.

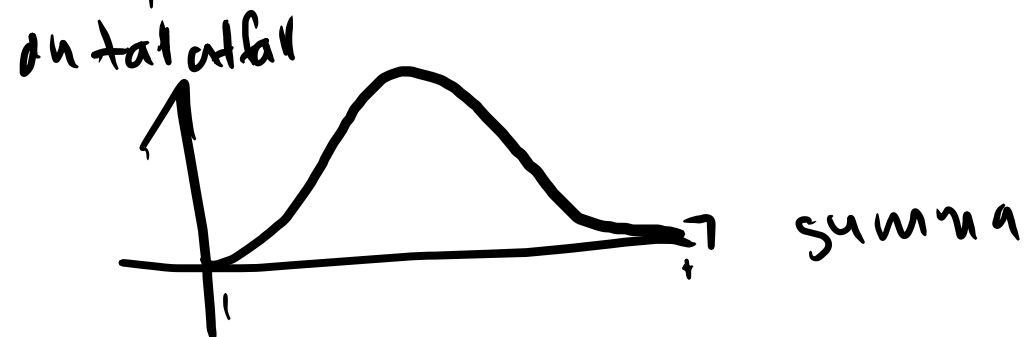
$T$  är en kontinuerlig s.v.

# Distributioner

$X$  är en s.v., Om vi upprepat drar värden från  $X$  och plottar i en graf:

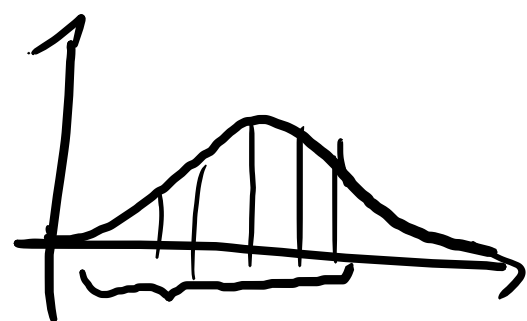


uniform distribution (tex en sex-sidig tärning)



normal fördelning (1e6 tärningar)

> 1e6 tärningar



area = sannolikhet

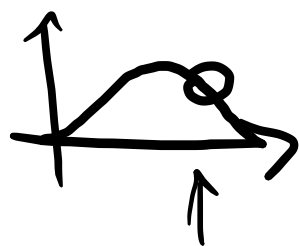
Låt  $X$  vara en diskret s.v.,

$$F(x) = P[X \leq x]$$

ä också  $f$  för kumulativ distributions-  
function (PDF)

Täthetsfunktion:

$$f(x) = P[X = x]$$



$$1. f(x) \geq 0 \quad \text{för alla } x$$

$$2. \sum_x f(x) = 1$$

$$F(x) = P[X \leq x]$$

$$F(x_0) = \sum_{x \leq x_0} f(x)$$

Kontinuerliga:

$$\int_x f(x) = 1$$

Väntevärde (Expectation)

"average", "mean", "expected"  
genomsnitt, medel, väntevärde  
 $= \mu$

$$E[H(x)] = \sum_x H(x) f(x)$$

$$E[X] = \sum_x x f(x)$$

$$E[c] = c$$

$$E[cx] = c E[X]$$

$$E[x+y] = E[x] + E[y]$$

Varians (Variance)

medelvärde  
↓

$$\begin{aligned}\text{Var } X &= \sigma^2 = E[(X - \mu)^2] \\ &= E[X^2] - (E[X])^2\end{aligned}$$

Avvikelse

$$\sigma = \sqrt{\text{Var } X} = \sqrt{\sigma^2}$$

1σ, 2σ, 3σ

68% - 95% - 99.7%

$$\text{Var } c = 0$$

$$\text{Var } cX = c^2 \text{Var } X$$

Om  $X, Y$  oberoende:

$$\text{Var } (X + Y) = \text{Var } X + \text{Var } Y$$

Gemensamma distributioner  
(joint distributions)

1.  $f_{XY}(X, Y) \geq 0$

2.  $\sum_x \sum_y f_{XY}(x, y) = 1$

Marginell täthet

↓

$$f_X(x) = \sum_y f_{XY}(x, y)$$

$$f_Y(y) = \sum_x f_{XY}(x, y)$$

$$f_{X|Y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

$$f_{Y|X}(y) = \frac{f_{XY}(x, y)}{f_X(x)}$$



$$f_{Y|X}(Y) = \frac{f_{XY}(X,Y)}{f_X(X)}$$

Givet  $X=x$  så räknar vi höj medlet av  $Y$  för det  $x$  et!

Om vi plottar varje  $x \in X$  mot  $Y$  får vi en regression av  $X$  på  $Y$ !

$$\mu_{Y|X} = \beta_0 + \beta_1 x \quad (kx + m)$$

Låt  $Y_i$  vara utfallet av en s.v.  $Y$ . Vi kallar den  
responsvariabel. Formad en tabell

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$E[\varepsilon_i] = 0$$

varje rad:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

← "rentlig", avvikelse från "sanna"  
linjen  
"fel"

efter uppskattning:

$$Y_i = b_0 + b_1 x_i + e_i$$

← känd, avvikelse från  
regressionslinjen!  
"residual"