



MVE051/MSG810 - Statistik och diskret matematik (LP4 2022-2023)

Chalmers och Göteborgs universitet

Föreläsning 3 – Kontinuerliga stokastiska variabler

March 28, 2023

Dagens föreläsning

- Kontinuerliga stokastiska variabler
 - Likformig fördelning
 - Normalfördelning
- Täthetsfunktion
- Väntevärde, varians och standardavvikelse
- Binomialapproximation
- Transformationer

Stokastiska variabler

Definition

En stokastisk variabel X (slumpvariabel, random variable) är en funktion som för varje utfall i ett slumpmässigt försök antar ett reellt tal, dvs.
 $X: \Omega \rightarrow \mathbb{R}$.

Diskret stokastisk variabel

En stokastisk variabel som antar ett ändligt eller uppreäkneligt (ex. $\mathbb{Z}$) antal värden kallas för en diskret stokastisk variabel.

Kontinuerlig stokastisk variabel

Ens stokastisk variabel som kan anta alla värden i ett intervall (i $\mathbb{R}$) kallas för en kontinuerlig stokastisk variabel.

Täthetsfunktion

Täthetsfunktion

Om $f: \mathbb{R} \rightarrow \mathbb{R}_+$ är en funktion sådan att

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

så kallas f för täthetsfunktionen (density function, probability density function, pdf) till (fördelningen av) X .

Konsekvenser

Om $a \in \mathbb{R}$ så är

$$P(X = a) = P(a \leq X \leq a) = \int_a^a f(x) dx = 0$$

och om $a \leq b$

$$P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b)$$

Täthetsfunktion och fördelningsfunktion

En funktion $f: \mathbb{R} \rightarrow \mathbb{R}$ är en täthetsfunktion om och endast om

1. $f(x) \geq 0$ för alla $x \in \mathbb{R}$, och
2. $\int_{\mathbb{R}} f(t) dt = 1$

Motsvarande fördelningsfunktion definieras som

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt.$$

Täthetsfunktion och fördelningsfunktion

Exempel

Låt X vara en kontinuerlig stokastisk variabel med täthetsfunktion

$$f(x) = \begin{cases} 3x^2 & \text{om } 0 < x < 1 \\ 0 & \text{annars.} \end{cases}$$

Bestäm $F(x)$ och beräkna $P(0.1 \leq X \leq 0.8)$.

Täthetsfunktion och fördelningsfunktion

Exempel

Visa att om $a < b$ så är funktionen

$$f(x) = \begin{cases} 1/(b-a) & \text{om } a < x < b \\ 0 & \text{annars} \end{cases}$$

en täthetsfunktion.

Likformig fördelning

Om en slumpvariabel X har täthetsfunktion

$$f(x) = \begin{cases} 1/(b-a) & \text{om } a < x < b \\ 0 & \text{annars} \end{cases}$$

så sägs X vara likformigt fördelad på intervallet (a, b) .

Vi skriver $X \sim \text{unif}(a, b)$.

Väntevärde, varians, standardavvikelse

Låt X vara en kontinuerlig stokastisk variabel ned täthetsfunktion f .

Väntevärde

Väntevärdet av X ges av

$$\mu = \mathbb{E}[X] = \int_{\mathbb{R}} x \cdot f(x) dx$$

Om $h: \mathbb{R} \rightarrow \mathbb{R}$ så definieras

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} h(x) \cdot f(x) dx$$

Varians och väntevärde

Variansen och väntevärdet definieras precis som för diskreta stokastiska variabler, dvs.

$$\text{Var}(X) = \mathbb{E}[(X - \mu)^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

och

$$\sigma = \sqrt{\text{Var}(X)}$$

Räknereglerna är samma som i det diskreta fallet.

Normalfördelning

Om $\mu \in \mathbb{R}$ och $\sigma > 0$ och en kontinuerlig stokastisk variabel X har täthetsfunktion

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

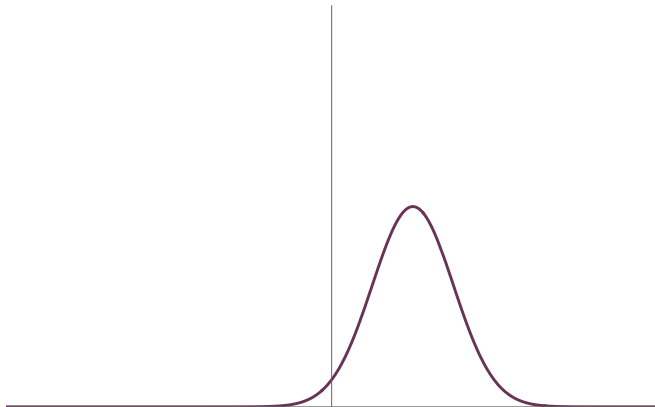
så är X normalfördelad med parametrar μ och σ . Vi skriver $X \sim N(\mu, \sigma^2)$.

Väntevärde och varians

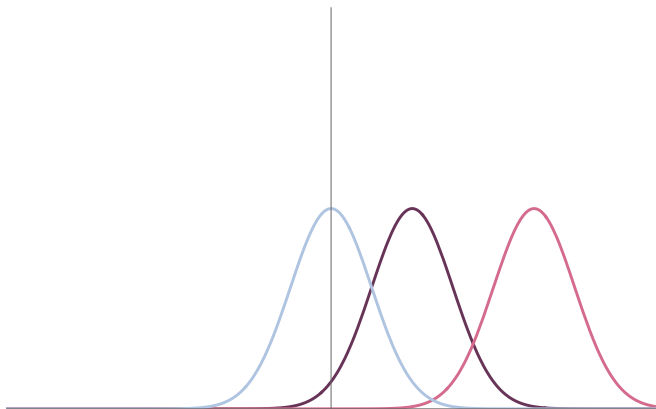
$$\mathbb{E}[X] = \mu$$

$$\text{Var}(X) = \sigma^2$$

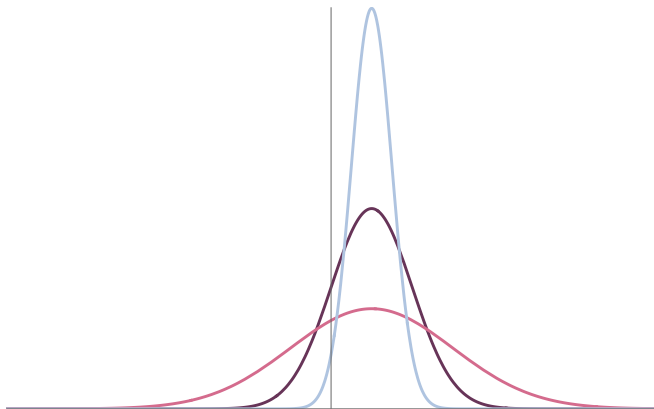
Normalfördelning



Normalfördelning



Normalfördelning



Normalfördelning

Standard normalfördelning

Om $Z \sim N(0, 1)$ så sägs Z ha en standard normalfördelning.

För en standard normalfördelning är

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

Man skriver ofta Z istället för X om $Z \sim N(0, 1)$ för att man ofta hanterar en annan stokastisk variabel $X \sim N(\mu, \sigma)$ samtidigt.

Normalfördelning

Sannolikhetsfunktion?

Det finns ingen sluten form för sannolikhetsfunktionen F till $Z \sim N(0, 1)$. Istället hittar man värden för $F(x)$ i en tabell (Tabell V, s. 697–698 i boken).

Ofta skriver man $\Phi(x)$ istället för $F(x)$ när $X \sim N(0, 1)$.

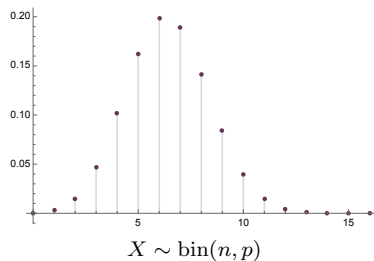
Några vanliga värden är

- $\Phi(0) = P(Z \leq 0) = 0.5$
- $\Phi(1) = P(Z \leq 1) = 0.841345$
- $\Phi(2) = P(Z \leq 2) = 0.97725$
- $\Phi(3) = P(Z \leq 3) = 0.99865$

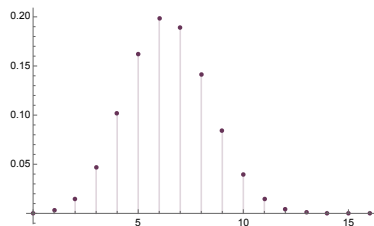
och

- $P(-1 \leq Z \leq 1) = 0.68269$
- $P(-2 \leq Z \leq 2) = 0.97725$
- $P(-3 \leq Z \leq 3) = 0.99865$

Normalfördelning som en approximation till binomialfördelning



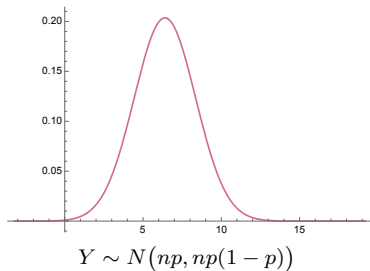
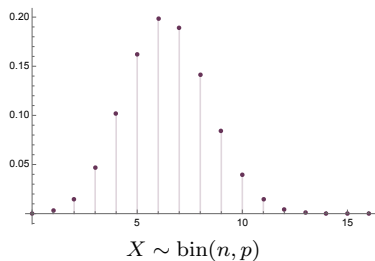
Normalfördelning som en approximation till binomialfördelning



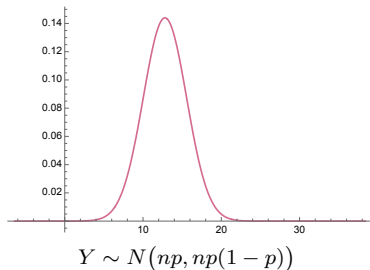
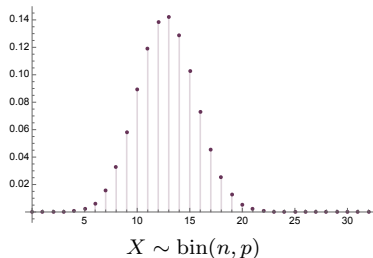
$$X \sim \text{bin}(n, p)$$

$$Y \sim N(\quad , \quad)$$

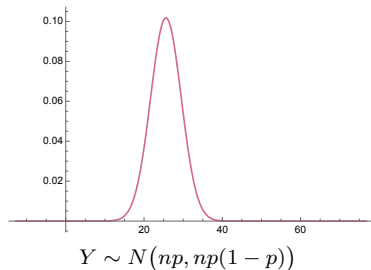
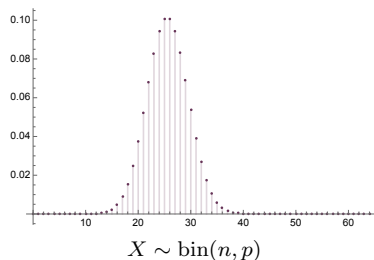
Normalfördelning som en approximation till binomialfördelning



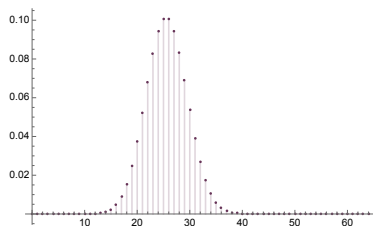
Normalfördelning som en approximation till binomialfördelning



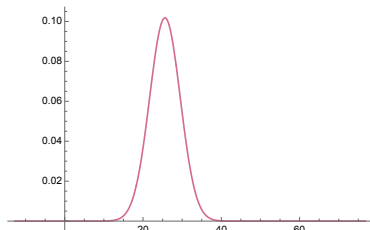
Normalfördelning som en approximation till binomialfördelning



Normalfördelning som en approximation till binomialfördelning



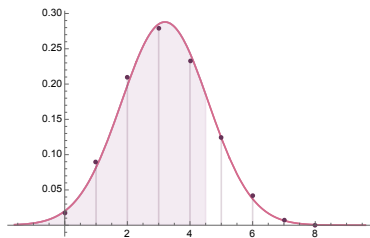
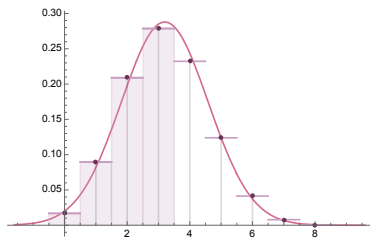
$$X \sim \text{bin}(n, p)$$



$$Y \sim N(np, np(1-p))$$

→ Om n är väldigt stort så kan X approximeras med $Y \sim N(np, np(1-p))$.

Normalfördelning som en approximation till binomialfördelning



Heltalskorrigering

Normalfördelnigen är kontinuerlig medan binomialfördelningen är diskret. Man approximerar därför ofta

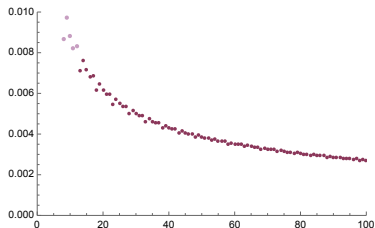
$$P(X \leq x) \approx P(Y < x + 1/2)$$

och

$$P(X < x) = P(X \leq x - 1) \approx P(Y < x - 1/2)$$

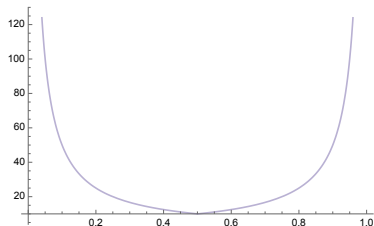
Normalfördelning som en approximation till binomialfördelning

Hur stort blir felet?



$$\max_{k \in \{0, \dots, n\}} |P(X \leq k) - P(Y \leq k)|$$

Hur stort stickprov behövs?



$$\max(5/p, 5/(1-p))$$

Tumregel

Approximationen kan användas om $np > 5$ och $n(1-p) > 5$.

Transformationer av kontinuerliga stokastiska variabler

Anta att X är en kontinuerlig stokastisk variabel med täthetsfunktion f_X .
Anta att $h: \mathbb{R} \rightarrow \mathbb{R}$ är strängt monotont och deriverbar, och låt
 $Y = h(X)$. Då är Y en kontinuerlig stokastisk variabel med
täthetsfunktion

$$f_Y(y) = f_X(h^{-1}(y)) \cdot (h^{-1})'(y).$$

Bevis:

Transformationer av kontinuerliga stokastiska variabler

Exempel

Låt X vara en kontinuerlig stokastisk variabel med täthetsfunktion $f_X(x)$, och låt $Y = aX + b$ för några $a \neq 0$ och $b \in \mathbb{R}$. Då är $h(x) = ax + b$ och $h^{-1}(y) = (y - b)/a$, så enligt satsen är

$$f_Y(y) = f_X(h^{-1}(y)) \cdot (h^{-1})'(y) = f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{a}.$$

Transformationer av kontinuerliga stokastiska variabler

Exempel

Om $X \sim N(\mu, \sigma)$ och $Y = \frac{X - \mu}{\sigma}$ så är $a = \frac{1}{\sigma}$ och $b = -\frac{\mu}{\sigma}$, så vi får

$$\begin{aligned} f_Y(y) &= f_X\left(\frac{y - (-\mu/\sigma)}{(1/\sigma)}\right) \cdot \frac{1}{(1/\sigma)} = f_X(\sigma y + \mu) \cdot \sigma \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{((\sigma y + \mu) - \mu)^2}{2\sigma^2}} \cdot \sigma = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}, \end{aligned}$$

dvs. $Y \sim N(0, 1)$.

Obs! Om $X \sim N(\mu, \sigma)$ så beror $P(X \leq \mu + k\sigma)$ alltså bara på k .

Normalfördelning

Exempel

Låt $X \sim N(12, 2^2)$ och $Z \sim N(0, 1)$. Då är

$$\begin{aligned} P(X \leq 13) &= P\left(\frac{X - 12}{2} \leq \frac{13 - 12}{2}\right) = P\left(Z \leq \frac{13 - 12}{2}\right) \\ &= P\left(Z \leq \frac{1}{2}\right) = \Phi(1/2) = 0.691462 \end{aligned}$$

enligt tabell.

$$P(X > 13) = 1 - P(X \leq 13) = 1 - 0.691462 = 0.308538$$

$$\begin{aligned} P(X < 11) &= P\left(Z \leq \frac{11 - 12}{2}\right) = P\left(Z \leq -\frac{1}{2}\right) = P\left(Z \geq \frac{1}{2}\right) \\ &= 1 - P\left(Z \leq \frac{1}{2}\right) = 1 - \Phi(1/2) = 1 - 0.691462 = 0.308538 \end{aligned}$$

och

$$\begin{aligned} P(11 < X < 13) &= P(X < 13) - P(X < 11) = 0.691462 - 0.308538 \\ &= 0.382925 \end{aligned}$$