

SF1920/SF1921 Sannolikhetsteori och statistik

6,0 hp

Föreläsning 7

Normalfördelningen

Linjärkombinationer av oberoende normalfördelade
stokastiska variabler

Centrala gränsvärdessatsen

Jörgen Säve-Söderbergh

Definition

Den stokastiska variabeln X är *normalfördelad* om dess täthetsfunktion ges som

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\}, \quad -\infty < x < \infty.$$

Symbolerna μ och σ är *parametrar* (fixerade reella tal) som uppfyller

$$-\infty < \mu < \infty \quad \text{och} \quad 0 < \sigma < \infty.$$

Vi använder beteckningen $X \in N(\mu, \sigma)$ för denna fördelning.

Definition

Fördelningsfunktionen för $X \sim N(\mu, \sigma)$ är

$$F_X(x) = P(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\} dt.$$

Standardiserad normalfördelning– $N(0, 1)$

Definition

Om vi låter $\mu = 0$ och $\sigma = 1$ erhåller vi den *standardiserade normalfördelningen*. Täthetsfunktion har då uttrycket

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{x^2}{2} \right\}, \quad -\infty < x < \infty,$$

och fördelningsfunktionen

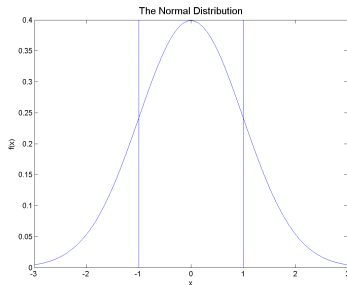
$$\Phi(x) = P(X \leq x) = \int_{-\infty}^x \varphi(t) dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp \left\{ -\frac{t^2}{2} \right\} dt.$$

$\varphi(x)$ är jämn, ty

$$\varphi(-x) = \varphi(x).$$

(Grafen till en jämn funktion är symmetrisk med avseende på y -axeln)

Standardiserad normalfördelning– $N(0, 1)$



- Maximum i $x = 0$, vilket är väntevärdet $\mu = 0$.
- I punkterna $x = -1$ och $x = 1$ har tätheten två inflexionspunkter. (De sitter i punkterna $\mu - \sigma$ och $\mu + \sigma$ för samtliga normalfördelningar, alltså en standardavvikelse från väntevärdet)

En *inflexionspunkt* är en punkt där kurvan byter från att vara konkav till konvex eller vice versa.

Fördelningsdenningsfunktionen för $N(0, 1)$ -fördelningen

$$\Phi(x) = P(X \leq x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{t^2}{2}\right\} dt.$$

Man kan bevisa att det inte går att hitta en primitiv funktion till $\exp\left\{-\frac{t^2}{2}\right\}$.

Genom numeriska metoder har tabeller skapats.

På sidan 397 finns en tabell för $x = 0, 0.01, 0.02, \dots, 4.0$

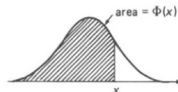
$$\Phi(0) = 0.5, \quad \Phi(1.64) = 0.9495 \quad \text{and} \quad \Phi(1.65) = 0.9505.$$

Tabell över den standardiserade normalfördelningens fördelningsfunktion

324

Tables

Table 1. Normal distribution. $\Phi(x) = P(X \leq x)$ where $X \sim N(0, 1)$. When x is negative, use the fact that $\Phi(-x) = 1 - \Phi(x)$.



x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9807	.9812	.9816
2.1	.9821	.9827	.9830	.9834	.9838	.9842	.9846	.9850	.9853	.9857
2.2	.9861	.9864	.9867	.9871	.9874	.9878	.9880	.9884	.9887	.9890
2.3	.9892	.9895	.9898	.9901	.9903	.9906	.9908	.9911	.9913	.9915
2.4	.9918	.9920	.9922	.9924	.9926	.9928	.9929	.9931	.9932	.9934

För $x < 0$ kan tabellen fortfarande användas tack vare sambandet

$$\Phi(-x) = 1 - \Phi(x) \quad \text{för alla reella } x.$$

Ett litet exempel

$$\Phi(-1) = 1 - \Phi(1) = 1 - 0.8413 = 0.1587$$

Den standardiserade normalfördelningen

Om X är en godtycklig stokastisk variabel med fördelningsfunktionen F , så gäller för a och b som uppfyller $a \leq b$

$$P(a < X \leq b) = F(b) - F(a). \quad (\text{Sats 3 i kapitel 3, sidan 67})$$

Alltså gäller för den standardiserade normalfördelningen

$$P(a < X \leq b) = \Phi(b) - \Phi(a).$$

Eftersom X är kontinuerlig så kan vi variera olikheterna i $P(a < X \leq b)$ som vi önskar – det spelar ingen roll.

α -kvantilen för en standardiserade normalfördelningen

... Om x bestäms så att arean under täthetsfunktionen till höger om x är lika med en given kvantitet α , $0 < \alpha < 1$, så definierar x den så kallade α -kvantilen för fördelningen. ...

α -kvantilen för den standardiserade normalfördelningen betecknas λ_α .

Arean till höger om λ_α är lika med α : $P(X > \lambda_\alpha) = \alpha$

α	0.0005	0.001	0.005	0.01	0.025	0.05	0.10
λ_α	3.29	3.09	2.58	2.33	1.96	1.64	1.28

Eftersom tätheten är symmetrisk kring 0 har vi

$$\lambda_{1-\alpha} = -\lambda_\alpha$$

Den standardiserade normalfördelningen – väntevärde och varians

Theorem

Om $X \sim N(0, 1)$, så är $E(X) = 0$ och $Var(X) = 1$.

Theorem

Om $X \in N(\mu, \sigma)$, så är den standardiserade variabeln

$$Y = \frac{X - \mu}{\sigma} \in N(0, 1).$$

$$Y = \frac{X - \mu}{\sigma} \Leftrightarrow X = \mu + \sigma Y$$

Theorem

Om $X \in N(\mu, \sigma)$, så är

$$E(X) = \mu, \quad \text{Var}(X) = \sigma^2, \quad D(X) = \sigma$$

Bevis.

$$\begin{aligned} E(X) &= E(\mu + \sigma Y) = \mu + \sigma E(Y) = \mu, \\ \text{Var}(X) &= \text{Var}(\mu + \sigma Y) = 0 + \sigma^2 \text{Var}(Y) = \sigma^2. \end{aligned}$$



Parametrarna i (den allmänna) normalfördelningen är

$$\mu = E(X) \qquad \sigma = D(X)$$

Normalfördelningen $N(\mu, \sigma)$ – Hur beräkna sannolikheten för olika händelser

Om $X \in N(\mu, \sigma)$, så

$$\begin{aligned}P(a \leq X \leq b) &= P\left(\frac{a - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right) \\&= P\left(\frac{a - \mu}{\sigma} \leq Y \leq \frac{b - \mu}{\sigma}\right) \\&= \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right).\end{aligned}$$

Vid en industri produceras järnbalkar som väger 2000 kg.

Vikten hos en producerad järnbalk kan beskrivas av en normalfördelning där $\mu = 2000$ och $\sigma^2 = 5.29$.

- A) Vissa balkar avviker med mer än 5 kg från den avsedda vikten. Dessa efterbehandlas för att få en vikt närmare 2000 kg. Bestäm vilken andel balkar som behöver efterbehandlas. (Svar: 0.03)
- B) Bestäm sannolikheten att en balk väger mindre än 1997 kg. (Svar: 0.0968)
- c) Bestäm sannolikheten att en balk väger mer än 2005 kg. (Svar: 0.015)

Normalfördelningen

Vi är ofta intresserade av att finna symmetriska intervall sådana att en stokastisk variabel $X \sim N(0, 1)$ befinner sig inom detta intervall med en föreskriven sannolikhet.

För vilket $x > 0$ har vi att $P(-x \leq X \leq x) = 0.95$?

Vi kan använda kvantiler för att hitta x .

Låt

$$x = \lambda_{0.025} = 1.96 \text{ och } -x = -\lambda_{0.025} = -1.96$$

vilket ger

$$P(-1.96 \leq X \leq 1.96) = 0.95.$$

En allmän formel för godtyckligt α : Om $X \in N(0, 1)$, så för godtyckligt val av α

$$P(-\lambda_{\alpha/2} \leq X \leq \lambda_{\alpha/2}) = 1 - \alpha.$$

För den (allmänna) normalfördelningen $X \in N(\mu, \sigma)$ kan vi uttrycka α -kvantilen x_α med hjälp av α -kvantilen för den standardiserade normalfördelningen λ_α .

$$P(X > x_\alpha) = P\left(\frac{X - \mu}{\sigma} > \frac{x_\alpha - \mu}{\sigma}\right) = P(Y > \lambda_\alpha) = \alpha,$$

d.v.s.

$$\lambda_\alpha = \frac{x_\alpha - \mu}{\sigma} \quad \Leftrightarrow \quad x_\alpha = \mu + \sigma \lambda_\alpha$$

Vidare

$$P\left(-\lambda_{\alpha/2} \leq \frac{X - \mu}{\sigma} \leq \lambda_{\alpha/2}\right) = 1 - \alpha$$

$$P(\mu - \lambda_{\alpha/2} \sigma \leq X \leq \mu + \lambda_{\alpha/2} \sigma) = 1 - \alpha$$

Normalfördelning – linjär transformation

Normalfördelningen har den (goda) egenskapen att efter en linjär tranformation är den nya variabeln fortfarande normalfördelad.

Theorem

Om $X \sim N(\mu, \sigma)$ och $Y = aX + b$, så är $Y \sim N(a\mu + b, |a|\sigma)$

Theorem

Om $X_1, X_2, \dots, X_n$ är oberoende och respektive normalfördelade $N(\mu_1, \sigma_1), N(\mu_2, \sigma_2), \dots, N(\mu_n, \sigma_n)$ och $a_1, a_2, \dots, a_n$ är givna konstanter, så är

$$a_1X_1 + a_2X_2 + \dots + a_nX_n + b \sim N\left(a_1\mu_1 + a_2\mu_2 + \dots + a_n\mu_n + b, \sqrt{a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \dots + a_n^2\sigma_n^2}\right)$$

Corollary

Ett specialfall; om $X_1, X_2, \dots, X_n$ är oberoende och likafördelade $N(\mu, \sigma)$, konstanterna är $a_i = 1/n$, $i = 1, 2, \dots, n$, samt $b = 0$, så är

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Ett annat specialfall är

Corollary

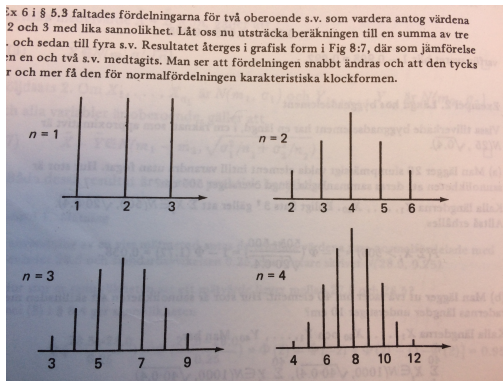
Om $X_1, X_2, \dots, X_{n_1}$ är $N(\mu_1, \sigma_1)$ och $Y_1, Y_2, \dots, Y_{n_2}$ är $N(\mu_2, \sigma_2)$ och samtliga variabler är oberoende, så är

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

Vi har just sett att summor av oberoende normalfördelade s.v. är normalfördelade.

Centrala gränsvärdessatsen uttalar att en summa av oberoende s.v. med en godtycklig fördelning är normalfördelad, om antalet termer i summan är tillräckligt många.

Ett försök att åskådliggöra den centrala gränsvärdessatsen



Fördelningen förändras snabbt och verkar anta formen av en normalfördelning.

I kursboken finns motsvarande illustration på sidan 159. De startar från en likformig fördelning över 1, 2, 3, 4, 5, 6.

Theorem

Om $X_1, X_2, \dots, X_n, \dots$ är en oändlig följd av oberoende likafördelade stokastiska variabler, var och en med väntevärdet μ och standardavvikelsen $\sigma > 0$.

och, låt

$$Y_n = X_1 + \dots + X_n$$

så gäller att

$$P\left(a < \frac{Y_n - n\mu}{\sigma\sqrt{n}} < b\right) \rightarrow \Phi(b) - \Phi(a), \text{ då } n \rightarrow \infty$$

$$E(Y_n) = E(X_1 + \dots + X_n) = n\mu$$

$$\text{Var}(Y_n) = \text{Var}(X_1 + \dots + X_n) = n\sigma^2$$

$\frac{Y_n - n\mu}{\sigma\sqrt{n}}$ standardiserad stokastisk variabel.

Centrala gränsvärdessatsen

Definition

Om $Z_n, n = 1, 2, \dots$, är en oändlig följd av s.v. och det är möjligt att hitta tal A_n och $B_n, n = 1, 2, \dots$, sådana att

$$P\left(a < \frac{Z_n - A_n}{B_n} < b\right) \rightarrow \Phi(b) - \Phi(a), \text{ när } n \rightarrow \infty$$

så säges Z_n vara *asymptotiskt normalfördelade* med parametrarna A_n och B_n . Betecknas med

$$Z_n \sim \text{AsN}(A_n, B_n)$$

Alltså kan centrala gränsvärdessatsen skrivas: $Y_n \sim \text{AsN}(n\mu, \sigma\sqrt{n})$

Corollary

Om $X_1, X_2, \dots, X_n, \dots$ är en oändlig följd av oberoende likafördelade stokastiska variabler, var och en med väntevärdet μ och standardavvikelsen $\sigma > 0$, så är $\bar{X} \sim \text{AsN}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

Illustration av centrala gränsvärdessaten

Antag att termerna är $X_i \sim \text{Exp}(\lambda)$, with $\lambda = 1$.

$$\mu = E(X) = \frac{1}{\lambda} = \frac{1}{1} = 1$$

$$\sigma = D(X) = \frac{1}{\lambda} = 1$$

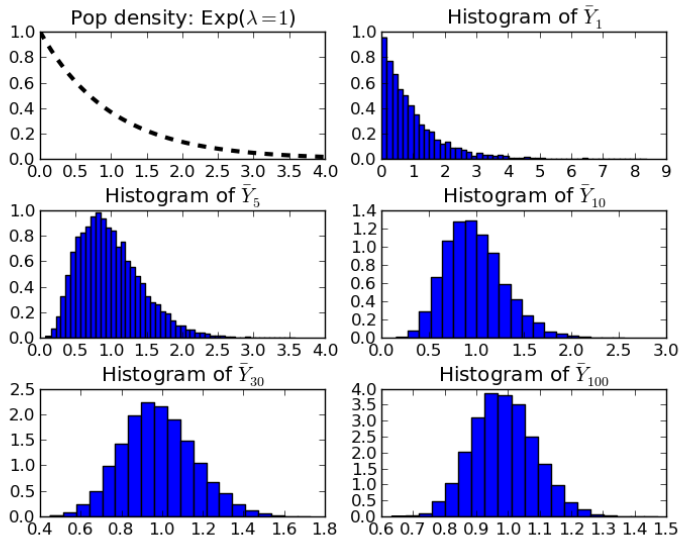
För dessa termer har vi ($Y_n = X_1 + \dots + X_n$)

$$\frac{Y_n - n\mu}{\sigma\sqrt{n}} = \frac{Y_n - n}{\sqrt{n}} \approx N(0, 1)$$

när n är stort. Vi skriver om så att det handlar om medelvärdet istället:

$$\frac{\frac{1}{n} \times \frac{Y_n - n\mu}{\sigma\sqrt{n}}}{\frac{1}{n}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X}_n - 1}{1/\sqrt{n}} = \sqrt{n}(\bar{X}_n - 1) \approx N(0, 1)$$

Fördelningen för medelvärdet $\bar{X}$ med datorsimulering



Fördelningen för medelvärdet $\sqrt{n}(\bar{X}_n - 1)$ med datorsimulering

