CHALMERS | GÖTEBORGS UNIVERSITET

Matematisk statistik
Tommy Norberg
26 augusti 2005

Formler och tabeller

till

Matematisk statistik

på universitet och tekniska högskolor

A Sannolikhetsteori

1 Likformig sannolikhetsfördelning på ett ändligt utfallsrum S

Om alla utfall är lika sannolika gäller för en händelse A att

$$P(A) = \frac{n(A)}{n(S)} = \frac{\text{antalet element i } A}{\text{antalet element i } S}$$

Detta är den klassiska sannolikhetsdefinitionen.

2 Kombinatorik

Antalet sätt att välja k objekt ur en mängd med n objekt är

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

om hänsyn ej tas till ordningen och

$$\binom{n}{k} k! = \frac{n!}{(n-k)!}$$

annars. Obs konventionen $0! = 1$. Notera generaliseringen

$$\binom{n}{k_1, \ldots, k_r} = \frac{n!}{k_1! \ldots k_r!}$$

för $k_i \geq 0$, $k_1 + \ldots + k_r = n$.

3 Bayes formel

Låt $A_1, \ldots, A_n$ vara en partition av utfallsrummet i händelser med strikt positiva sannolikheter och låt B vara en händelse, sådan att $P(B) > 0$. Då gäller att

$$P(A_j|B) = \frac{P(B|A_j)\,P(A_j)}{\sum_{i=1}^{n} P(B|A_i)\,P(A_i)}$$

Notera Lagen om total sannolikhet

$$\sum_{i=1}^{n} P(B|A_i)\,P(A_i) = P(B)$$

4 Positivt beroende

Händelserna A och B sägs vara positivt beroende då

$$P(A \cap B) \geq P(A)P(B)$$

B Stokastiska variabler

5 Momentgenererande funktion (mgf)

Momentgenererande funktion för den stokastiska variabel X existerar och är

$$m_X(t) = Ee^{tX}$$

om väntevärdet är ändligt för alla t i en öppen omgivning till 0. Under förutsättning att högerledet existerar, så gäller

$$E[X^k] = m_X^{(k)}(0)$$

6 Unikhet av mgf

Två stokastiska variabler X och Y har samma fördelning om, och endast om, $m_X(t) = m_Y(t)$ för alla t i en öppen omgivning till 0.

7 Bernoullifördelning

$X \sim \mathrm{Ber}(p)$ om X är diskret och tätheten är

$$f(k) = \begin{cases} p & \text{då } k = 1 \\ q & \text{då } k = 0 \end{cases}$$

där $q = 1 - p$. Parametern p är en sannolikhet $(0 < p < 1)$. Vi har

$$EX = p, \quad \mathrm{Var}[X] = pq \quad \text{och} \quad m_X(t) = q + pe^t$$

8 Geometrisk serie

För $r \neq 1$ och $n = 1, 2, \ldots$ gäller

$$\sum_{k=1}^{n} r^{k-1} = \frac{1 - r^n}{1 - r}$$

För $|r| < 1$ gäller

$$\sum_{k=1}^{\infty} r^{k-1} = \frac{1}{1 - r}$$

9 Geometrisk fördelning

$X \sim \mathrm{Geo}(p)$ om X är diskret och tätheten är

$$f(k) = p\,(1 - p)^{k-1}, \quad k = 1, 2, 3, \ldots$$

Parametern p är en sannolikhet $(0 < p < 1)$. Låt $q = 1 - p$. Vi har

$$EX = \frac{1}{p}, \quad \mathrm{Var}[X] = \frac{q}{p^2} \quad \text{och} \quad m_X(t) = \frac{pe^t}{1 - qe^t} \quad \text{för} \quad t < -\ln q$$

Låt $X_1, X_2, \ldots$ vara $\mathrm{Ber}(p)$ och oberoende. Då är $X = \min_n\{n : X_n = 1\} \sim \mathrm{Geo}(p)$.

10 Binomialsatsen

För reella a, b och positiva heltal n gäller

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k\, b^{n-k}$$

11 Binomialfördelningen

$X \sim \mathrm{Bin}(n, p)$ om X är diskret och tätheten är

$$f(k) = \binom{n}{k} p^k\,(1-p)^{n-k} \quad \text{för} \quad 0 \le k \le n$$

Parametern n är ett strikt positivt heltal och p är en sannolikhet $(0 < p < 1)$. Låt $q = 1-p$. Vi har

$$EX = np, \quad \mathrm{Var}[X] = npq \quad \text{och} \quad m_X(t) = (q + pe^t)^n$$

Låt $X_1, X_2, \ldots, X_n$ vara $\mathrm{Ber}(p)$ och oberoende. Då är $X = \sum_{k=1}^{n} X_k \sim \mathrm{Bin}(n, p)$.

12 Addition av oberoende Binomialvariabler

Låt $X \sim \mathrm{Bin}(n, p)$ och $Y \sim \mathrm{Bin}(m, p)$ vara oberoende. Då är $X + Y \sim \mathrm{Bin}(n + m, p)$.

13 Multinomialfördelningen

$X_1, \ldots, X_k$ är multinomialfördelade med parametrar n och $p_1, \ldots, p_k$, om tätheten är

$$f(x_1, \ldots, x_k) = \binom{n}{x_1, \ldots, x_k} p_1^{x_1} \cdot \ldots \cdot p_k^{x_k}$$

där $x_1, \ldots, x_k$ är icke-negativa heltal och $x_1 + \ldots + x_k = n$. Parametern n är ett strikt positivt heltal och $p_1, \ldots, p_k$ är sannolikheter, sådana att $p_1 + \ldots + p_k = 1$. Obs att $X_i \sim \mathrm{Bin}(n, p_i)$. Således är

$$E[X_i] = np_i \quad \text{och} \quad \mathrm{Var}[X_i] = np_i(1-p_i)$$

Dessutom gäller för $i \neq j$ att

$$\mathrm{Cov}[X_i, X_j] = -np_i p_j$$

Om $X \sim \mathrm{Bin}(n, p)$, så är $X, n - X$ multinomialfördelade med parametrar n och $p, 1-p$.

14 Hypergeometrisk fördelning

X är hypergeometriskt fördelad med parametrar N, n, r om X är diskret och tätheten är

$$f(k) = \frac{\binom{r}{k}\binom{N-r}{n-k}}{\binom{N}{n}} \quad \text{för} \quad \max[0, n-(N-r)] \le k \le \min(n, r)$$

Parametern N är ett strikt positivt heltal, och r, n är positiva heltal $\le N$. Vi har

$$EX = n\left(\frac{r}{N}\right) \quad \text{och} \quad \mathrm{Var}[X] = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$$

15 MacLaurin-utveckling av e^z

$$e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots \quad \text{för} \quad -\infty < z < \infty$$

16 Poissonfördelningen

$X \sim \mathrm{Poi}(\lambda)$ om X är diskret och tätheten är

$$f(k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad \text{för} \quad k = 0, 1, 2, \ldots$$

Parametern λ är ett strikt positivt reellt tal. Vi har

$$E[X] = \mathrm{Var}[X] = \lambda \quad \text{och} \quad m_X(t) = e^{\lambda(e^t-1)}$$

17 Poissonapproximation av Bin(n, p)

Låt $X \sim \mathrm{Bin}(n, p)$. Antag att $n \to \infty$ och att $p \to 0$ på ett sådant sätt att $E[X] = np \to \lambda$. Då

$$f_X(k) \to e^{-\lambda}\frac{\lambda^k}{k!}$$

M.a.o, $\mathrm{Bin}(n, p) \approx \mathrm{Poi}(np)$. Approximationen är hyfsad om $n \geq 20$ och $p \leq 0.5$. Den är klart tillfredsställande då $n \geq 100$ och $np \leq 10$. Om $p > 0.5$ approximera istället $Y = n - X \sim \mathrm{Bin}(n, q)$, där $q = 1 - p$.

18 Likformig fördelning på (a, b)

$X \sim \mathrm{U}(a, b)$ om X är kontinuerlig och tätheten är

$$f(x) = \frac{1}{b-a} \quad \text{för} \quad a < x < b$$

Parametrarna a, b är reella tal uppfyllande $a < b$. Vi har

$$EX = \frac{a+b}{2}, \quad \mathrm{Var}[X] = \frac{(b-a)^2}{12} \quad \text{och} \quad m_X(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

Obs $m_X(0) = 1 = \lim_{t\to 0} m_X(t)$. Obs även att $X \sim \mathrm{U}(a, b) \Rightarrow (X - a)/(b - a) \sim \mathrm{U}(0, 1)$.

19 Felbenägenhet

För en kontinuerlig icke-negativ stokastisk variabel X med täthet $f(x)$ och fördelningsfunktion $F(x) = 1 - R(x)$, är felbenägenheten (felintensiteten)

$$z(x) = \frac{f(x)}{R(x)} \quad \text{för} \quad x \geq 0$$

Obs att

$$R(x) = e^{-Z(x)} \quad \text{för} \quad x \geq 0$$

där $Z(x) = \int_0^x z(u)\, du$ är den under (tids-) intervallet $[0, x]$ ackumulerade risken.

20 Gammafunktionen

$$\Gamma(\alpha) = \int_0^\infty z^{\alpha-1}\, e^{-z}\, dz \quad \text{för} \quad \alpha > 0$$

Egenskaper:

1. $\Gamma(1) = 1$
2. $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$, då $\alpha > 1$
3. $\Gamma(1/2) = \sqrt{\pi}$

21 Gammafördelningen

$X \sim \text{Gamma}(\alpha, \beta)$ om X är kontinuerlig och tätheten är

$$f(x) = \frac{1}{\Gamma(\alpha)\,\beta^\alpha} x^{\alpha-1}\, e^{-x/\beta} \quad \text{för} \quad x > 0$$

Parametrarna α, β är strikt positiva reella tal. Vi har

$$EX = \alpha\beta, \quad \text{Var}[X] = \alpha\beta^2 \quad \text{och} \quad m_X(t) = (1-\beta t)^{-\alpha} \quad \text{för} \quad t < 1/\beta$$

Obs att $X \sim \text{Gamma}(\alpha, \beta) \Rightarrow X/\beta \sim \text{Gamma}(\alpha, 1)$.

22 Addition av oberoende Gammavariabler

Låt $X_1 \sim \text{Gamma}(\alpha_1, \beta)$ och $X_2 \sim \text{Gamma}(\alpha_2, \beta)$ vara oberoende. Då är $X_1 + X_2 \sim \text{Gamma}(\alpha_1 + \alpha_2, \beta)$.

23 χ^2-fördelningen

$X \sim \chi^2(\nu)$ om $X \sim \text{Gamma}(\nu/2, 2)$, d.v.s $\chi^2(\nu) = \text{Gamma}(\nu/2, 2)$. Parametern ν är ett positivt heltal. Låt $X_1, \ldots, X_\nu$ vara oberoende $\text{N}(0,1)$-variabler (se nedan). Då är

$$X = \sum_{i=1}^{\nu} X_i^2 \sim \chi^2(\nu)$$

Vi har

$$EX = \nu \quad \text{och} \quad \text{Var}[X] = 2\nu$$

24 Addition av oberoende χ^2-variabler

Låt $X \sim \chi^2(\nu)$ och $Y \sim \chi^2(\gamma)$ vara oberoende. Då är $X + Y \sim \chi^2(\nu + \gamma)$.

25 Exponentialfördelningen

$X \sim \text{Exp}(\lambda)$ om X är kontinuerlig och tätheten är

$$f(x) = \lambda e^{-\lambda x} \quad \text{för} \quad x > 0$$

Parametern λ är ett strikt positivt reellt tal. Vi har

$$EX = 1/\lambda, \quad \text{Var}[X] = 1/\lambda^2 \quad \text{och} \quad m_X(t) = \frac{1}{1 - t/\lambda} \quad \text{för} \quad t < \lambda$$

Obs att $\text{Exp}(\lambda) = \text{Gamma}(1, 1/\lambda)$, samt att $X \sim \text{Exp}(\lambda) \Rightarrow \lambda X \sim \text{Exp}(1)$. Notera också att Exponentialfördelningen ibland parametriseras av $\beta = 1/\lambda$ och då skriver man kanske $X \sim \text{Exp}(\beta)$, vilket isåfall betyder att tätheten är $f(x) = e^{-x/\beta}/\beta$, $x > 0$.

26 Addition av oberoende Exponentialvariabler

Låt $X_1, \ldots, X_n$ vara oberoende och $\text{Exp}(\lambda)$. Då är $\sum_{i=1}^n X_i \sim \text{Gamma}(n, 1/\lambda) \Rightarrow 2\lambda \sum_{i=1}^n X_i \sim \chi^2(2n)$.

27 Weibullfördelningen

$X \sim \text{Wei}(\alpha, \beta)$ om X är kontinuerlig och har tätheten

$$f(x) = \alpha\, \beta\, x^{\beta-1}\, e^{-\alpha x^\beta} \quad \text{för} \quad x > 0$$

Parametrarna α, β är strikt positiva reella tal. Vi har att

$$EX = \alpha^{-1/\beta}\, \Gamma(1 + 1/\beta) \quad \text{och} \quad \text{Var}[X] = \alpha^{-2/\beta}\, \Gamma(1 + 2/\beta) - (EX)^2$$

Obs att $\text{Exp}(\lambda) = \text{Wei}(\lambda, 1)$.

28 Paretofördelningen

$X \sim \text{Par}(\alpha, u)$ om X är kontinuerlig och

$$P(X > x) = \left(\frac{u}{x}\right)^\alpha \quad \text{för} \quad x \geq u$$

Dess täthet är

$$f(x) = \frac{\alpha}{u} \left(\frac{u}{x}\right)^{\alpha+1} \quad \text{för} \quad x \geq u$$

Parametrarna α, u är strikt positiva reella tal. Vi har för $\alpha > 1$, resp $\alpha > 2$, att

$$EX = \frac{\alpha}{\alpha - 1}\, u \quad \text{och} \quad EX^2 = \frac{\alpha}{\alpha - 2}\, u^2$$

och för $\alpha > 2$ att

$$\text{Var}[X] = \frac{\alpha}{(\alpha - 2)(\alpha - 1)^2}\, u^2$$

29 Generaliserad Paretofördelning

$X \sim \text{GP}(\xi, \sigma)$ om X är kontinuerlig och

$$P(X > x) = \left(1 - \frac{\xi x}{\sigma}\right)^{-1/\xi} \quad \text{för} \quad x \geq 0$$

30 Betafördelningen

$X \sim \text{Beta}(a, b)$ om X är kontinuerlig och har tätheten

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\, x^{a-1}\,(1-x)^{b-1} \quad \text{för} \quad 0 < x < 1$$

Parametrarna a, b är strikt positiva reella tal. Vi har att

$$EX = \frac{a}{a+b} \quad \text{och} \quad \text{Var}[X] = \frac{a\,b}{(a+b)^2\,(a+b+1)}$$

31 Normalfördelningen

$X \sim \text{N}(\mu, \sigma)$ om X är kontinuerlig och tätheten är

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-(x-\mu)^2/2\sigma^2} \quad \text{för} \quad -\infty < x < \infty$$

Obs att $\sigma > 0$. Vi har att

$$EX = \mu, \quad \text{Var}[X] = \sigma^2 \quad \text{och} \quad m_X(t) = e^{\mu t + \sigma^2 t^2/2}$$

Obs att $X \sim \text{N}(\mu, \sigma) \Rightarrow (X-\mu)/\sigma \sim \text{N}(0, 1)$. Dessutom gäller att linjärkombinationer av oberoende normalvariabler är normalfördelade, samt att

1. $P(-2\sigma < X - \mu < 2\sigma) = P(|X - \mu| < 2\sigma) \approx 0.95$
2. $P(-2.6\sigma < X - \mu < 2.6\sigma) = P(|X - \mu| < 2.6\sigma) \approx 0.99$

32 Standardiserad normalfördelning

$Z \sim \text{N}(0, 1)$ har tätheten

$$\varphi(z) = \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2} \quad \text{för} \quad -\infty < z < \infty$$

och fördelningsfunktionen

$$\Phi(z) = \int_{-\infty}^{z} \phi(u)\, du \quad \text{för} \quad -\infty < z < \infty$$

33 Normalapproximation av Bin(n, p)

Låt $X \sim \text{Bin}(n, p)$. Låt $q = 1 - p$. Då

$$P(l \leq X \leq k) = \sum_{j=l}^{k} \binom{n}{j} p^j\, q^{n-j}$$

$$\approx \Phi\left((k + \tfrac{1}{2} - np)/\sqrt{npq}\right) - \Phi\left((l - \tfrac{1}{2} - np)/\sqrt{npq}\right)$$

M.a.o, $\text{Bin}(n, p) \approx \text{N}(np, \sqrt{npq})$. Approximationen är tillfredsställande då $n \min(p, q) \geq 5$.

34 Normalapproximation av Poi(λ)

Låt $X \sim \text{Poi}(\lambda)$. Då

$$P(l \leq X \leq k) = e^{-\lambda} \sum_{j=l}^{k} \frac{\lambda^j}{j!}$$
$$\approx \Phi\left((k + \tfrac{1}{2} - \lambda)/\sqrt{\lambda}\right) - \Phi\left((l - \tfrac{1}{2} - \lambda)/\sqrt{\lambda}\right)$$

Detta följer från $\text{Poi}(\lambda) \approx \text{N}(\lambda, \sqrt{\lambda})$. Approximationen är hyfsad då $\lambda \geq 15$.

35 Lognormalfördelningen

X är lognormalfördelad med parametrar $\mu \in (-\infty, \infty)$ och $\sigma > 0$, om $\ln X \sim \text{N}(\mu, \sigma)$. Vi har att

$$EX = e^{\mu + \sigma^2/2} \quad \text{och} \quad \text{Var}[X] = e^{2\mu + \sigma^2}\left(e^{\sigma^2} - 1\right)$$

36 Bivariat Normalfördelning

$X, Y \sim \text{N}(\mu_x, \mu_y, \sigma_x \sigma_y, \rho)$ om tätheten är kontinuerlig och ges av

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}$$

Obs att $-1 < \rho < 1$ och $\rho = 0$ om och endast om X, Y är oberoende. Obs även att $X \sim \text{N}(\mu_x, \sigma_x)$, $Y \sim \text{N}(\mu_y, \sigma_y)$, $\text{Cov}[X, Y] = \rho\sigma_x\sigma_y$ och $Y|X = x \sim \text{N}\Big(\mu_y + \rho\sigma_y \frac{x-\mu_x}{\sigma_x}, \sigma_y\sqrt{1-\rho^2}\Big)$.

37 Chebyshevs olikhet

Om $EX = \mu$ och $\text{Var}[X] = \sigma^2$, så

$$P(|X - \mu| > k\sigma) \leq \frac{1}{k^2} \quad \text{för} \quad k > 0$$

38 Centrala gränsvärdessatsen (cgs)

Låt $X_1, \ldots, X_n$ vara oberoende stokastiska variabler, alla fördelade som X (alltså ett stickprov på X). Antag att $EX = \mu$ och $\text{Var}[X] = \sigma^2$. Då är variablernas medelvärde $\bar{X}$ approximativt normalfördelat med

$$E\bar{X} = \mu \quad \text{och} \quad \text{Var}[\bar{X}] = \frac{\sigma^2}{n}$$

Härur följer

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \overset{\text{ap}}{\sim} \text{N}(0, 1)$$

Approximationen blir bättre ju större n är och ju mer symmetrisk X:s fördelning är. Den är exakt då $X \sim \text{N}(\mu, \sigma)$.

39 t-fördelningen

$T \sim t(\nu)$ om

$$T = \frac{Z}{\sqrt{X/\nu}}$$

för oberoende $Z \sim \mathrm{N}(0,1)$ och $X \sim \chi^2(\nu)$. Obs att $t(\infty) = N(0,1)$, d.v.s $t(\nu) \to \mathrm{N}(0,1)$ då $\nu \to \infty$.

C Statistik

40 Ett stickprov

Låt $X_1, \ldots, X_n$ vara oberoende, alla med samma fördelning som X. Antag att $E[X] = \mu$ och $\mathrm{Var}[X] = \sigma^2$. Låt

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{och} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Då

$$E[\bar{X}] = \mu, \quad \mathrm{Var}[\bar{X}] = \frac{\sigma^2}{n} \quad \text{och} \quad E[S^2] = \sigma^2$$

Dessutom,

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \overset{\text{ap}}{\sim} t(n-1)$$

för n så stort att normalapproximationen av $\bar{X}$ är tillfredsställande.

41 Ett normalfördelat stickprov

Låt $X_1, \ldots, X_n$ vara oberoende $\mathrm{N}(\mu, \sigma)$-variabler. Då är medelvärdet $\bar{X}$ och stickprovsvariansen S^2 oberoende. Ur

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

följer därför

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

42 Ett exponentialfördelat stickprov

Låt $X_1, \ldots, X_n$ vara oberoende $\mathrm{Exp}(\lambda)$-variabler. Då är

$$2\lambda \sum_{i=1}^{n} X_i \sim \chi^2(2n)$$

43 F-fördelningen

$X \sim F(\nu_1, \nu_2)$ om

$$X = \frac{X^2_{\nu_1}/\nu_1}{X^2_{\nu_2}/\nu_2}$$

för oberoende $X^2_{\nu_i} \sim \chi^2(\nu_i)$, i = 1,2. Notera att $1/X \sim F(\nu_2, \nu_1)$.

44 Två oberoende normalfördelade stickprov. Allmänt

Låt $\bar{X}_1$, $\bar{X}_2$ och S_1^2, S_2^2 vara medelvärden resp stickprovsvarianser för två oberoende normalfördelade stickprov med resp väntevärden μ_1, μ_2 och standardavvikelser σ_1, σ_2. Vi har

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$$

45 Två oberoende normalfördelade stickprov. Varianserna lika

Då $\sigma_1 = \sigma_2 = \sigma$ gäller att

$$\frac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1)$$

För den sammanvägda stickprovsvariansen

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

gäller då att

$$\frac{(n_1 + n_2 - 2)S_p^2}{\sigma^2} \sim \chi^2(n_1 + n_2 - 2)$$

Ur oberoendet mellan $\bar{X}_1, \bar{X}_2$ och S_p^2 samt faktumet

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim \mathrm{N}(0, 1)$$

följer sedan

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

46 Två oberoende normalfördelade stickprov. Varianserna olika

Då $\sigma_1 \neq \sigma_2$ gäller

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim \mathrm{N}(0, 1)$$

samt

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \overset{\mathrm{ap}}{\sim} t(\gamma)$$

Antalet frihetsgrader γ fås medelst avrundning nedåt av Smith-Satterthwaites frihetsgradtal

$$\frac{\left[S_1^2/n_1 + S_2^2/n_2\right]^2}{\frac{\left(S_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(S_2^2/n_2\right)^2}{n_2 - 1}}$$

47 Skattning av proportioner

Antag att n oberoende försök görs, och låt f vara frekvensen för en händelse med sannolikheten p. Då är $f \sim \mathrm{Bin}(n, p)$. För stora n gäller därför

$$\frac{\widehat{p}-p}{\sqrt{\widehat{p}\,(1-\widehat{p})\,/n}} \approx \frac{\widehat{p}-p}{\sqrt{p(1-p)/n}} \overset{\text{ap}}{\sim} \mathrm{N}(0,1)$$

där $\widehat{p} = f/n$. Då $n < 100$, använd istället

$$\frac{\widehat{p}-p}{\sqrt{\widehat{p}\,(1-\widehat{p})\,/(n-1)}} \overset{\text{ap}}{\sim} t(n-1)$$

48 Jämförelse av proportioner

Låt $f_1 = n_1\widehat{p}_1 \sim \mathrm{Bin}(n_1, p_1)$ och $f_2 = n_2\widehat{p}_2 \sim \mathrm{Bin}(n_2, p_2)$ vara oberoende. För stora n_1 och n_2 (typiskt ≥ 100) gäller

$$\frac{(\widehat{p}_1-\widehat{p}_2)-(p_1-p_2)}{\sqrt{\widehat{p}_1\,(1-\widehat{p}_1)\,/n_1+\widehat{p}_2\,(1-\widehat{p}_2)\,/n_2}} \overset{\text{ap}}{\sim} \mathrm{N}(0,1)$$

Då $p_1 = p_2 = p$ är $f_1 \sim \mathrm{Bin}(n_1, p)$ och $f_2 \sim \mathrm{Bin}(n_2, p)$. Härur följer att $f = f_1 + f_2 \sim \mathrm{Bin}(n, p)$, där $n = n_1 + n_2$, och, då n_1 och n_2 är stora,

$$\frac{\widehat{p}_1-\widehat{p}_2}{\sqrt{\widehat{p}\,(1-\widehat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \approx \frac{\widehat{p}_1-\widehat{p}_2}{\sqrt{p(1-p)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \overset{\text{ap}}{\sim} \mathrm{N}(0,1)$$

där $\widehat{p} = f/n$.

49 Jämförelse mellan uppmätta och teoretiska frekvenser

Låt $f_1, \ldots, f_k$ vara uppmätta frekvenser för kategorierna $A_1, \ldots, A_k$ i totalt $n = \sum_i f_i$ oberoende försök. Låt vidare $p_i = P(A_i)$, $1 \leq i \leq k$. Då är $f_1, \ldots, f_k$ multinomialfördelade med parametrar n och $p_1, \ldots, p_k$. Vidare är χ^2-avståndet

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - np_i)^2}{np_i} \overset{\text{ap}}{\sim} \chi^2(n-1)$$

Om $\hat{p}_1, \ldots, \hat{p}_k$ är skattningar av $p_1, \ldots, p_k$ baserade på r parameterskattningar, så är

$$\hat{\chi}^2 = \sum_{i=1}^{k} \frac{(f_i - n\hat{p}_i)^2}{n\hat{p}_i} \overset{\text{ap}}{\sim} \chi^2(n-r-1)$$

50 Linjär regression

Modell: $Y_1, \ldots, Y_n$ är oberoende och sådana att $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, där ”felet” $\epsilon_i \sim \mathrm{N}(0, \sigma)$ och $x_i \in R$. Låt

$$S_{xx} = \sum_i (x_i - \bar{x})^2 = \sum_i x_i^2 - n\bar{x}^2$$

$$S_{xy} = \sum_i (x_i - \bar{x})(y_i - \bar{y}) = \sum_i x_i y_i - n\bar{x}\bar{y}$$

och

$$S_{yy} = \sum_i (y_i - \bar{y})^2 = \sum_i y_i^2 - n\bar{y}^2$$

Minsta kvadrat-skattningarna (och trolighetsskattningarna) av β_1 och β_0 är

$$B_1 = \widehat{\beta}_1 = \frac{S_{xY}}{S_{xx}} \quad \text{och} \quad B_0 = \widehat{\beta}_0 = \bar{Y} - B_1\bar{x}$$

En väntevärdesriktig estimator av σ^2 är

$$S^2 = \frac{1}{n-2}\sum_i (Y_i - B_0 - B_1 x)^2 = \frac{1}{n-2}\left(S_{YY} - \frac{S_{xY}^2}{S_{xx}}\right)$$

Notera att skattningarna B_0, B_1 och S^2 är oberoende, samt att

$$\frac{(n-2)S^2}{\sigma^2} \sim \chi^2(n-2)$$

och

$$B_0 \sim \mathrm{N}\left(\beta_0, \sigma\sqrt{\frac{\sum_{i=1}^n x_i^2}{nS_{xx}}}\right) \quad \Rightarrow \quad \frac{B_0 - \beta_0}{S\sqrt{\frac{\sum_{i=1}^n x_i^2}{nS_{xx}}}} \sim t(n-2)$$

$$B_1 \sim \mathrm{N}\left(\beta_1, \sigma/\sqrt{S_{xx}}\right) \quad \Rightarrow \quad \frac{B_1 - \beta_1}{S/\sqrt{S_{xx}}} \sim t(n-2)$$

Regressionslinjen $y|x = \beta_0 + \beta_1 x$ skattas av

$$\widehat{Y}|x = B_0 + B_1 x \sim \mathrm{N}\left(\beta_0 + \beta_1 x, \sigma\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}\right)$$

och vill man använda $\widehat{Y}|x = B_0 + B_1 x$ till att prediktera utfallet av en ny observation $Y|x$, så är prediktionsfelet

$$Y|x - (B_0 + B_1 x) \sim \mathrm{N}\left(0, \sigma\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}\right)$$

Motsvarande prediktionsintervall är

$$Y|x = b_0 + b_1 x \pm t_{\alpha/2,n-2}\, s\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}$$

51 Skattning av korrelationskoefficienten ρ

Modell: data $(x_1, y_1), \ldots, (x_n, y_n)$ är oberoende observationer ifrån $\mathrm{N}(\mu_x, \mu_y, \sigma_x\sigma_y, \rho)$. Pearsons korrelationskoefficient ρ skattas av

$$R = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}}$$

(j.f.r paragraf 50). Notera att

$$\frac{R\sqrt{n-2}}{\sqrt{1-R^2}} \sim t(n-2)$$

under $H_0 : \rho = 0$. Denna modell kan även tillämpas i linjär regression (paragraf 50). Då gäller att

$$\beta_0 = \mu_y - \rho\frac{\sigma_y}{\sigma_x}\,\mu_x \quad \text{och} \quad \beta_1 = \rho\frac{\sigma_y}{\sigma_x}$$

samt $\sigma = \sigma_y\sqrt{1-\rho^2}$ (se paragraf 36).

Normalfördelningen

Låt $Z \sim \mathrm{N}(0,1)$. Tabellen nedan ger $P(Z \leq z) = \Phi(z)$ med fyra decimalers noggranhet för $z \geq 0$. Då $z < 0$ använd att $\Phi(z) + \Phi(-z) = 1$.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

Normalkvantiler

I tabellen nedan finns lösningen z till ekvationen $\Phi(z) = p$, för diverse värden på $p \geq 0.9$. Obs att $\Phi(-z) = 1 - p$.

p	z	p	z	p	z
0.90	1.28155	0.95	1.64485	0.99	2.32635
0.91	1.34076	0.96	1.75069	0.995	2.57583
0.92	1.40507	0.97	1.88079	0.999	3.09023
0.93	1.47579	0.975	1.95996	0.9995	3.29053
0.94	1.55477	0.98	2.05375	0.9999	3.71902

t-kvantiler

I tabellen nedan finns lösningen t till ekvationen $P(T > t) = q$ för vissa värden på q och olika antal frihetsgrader ν (T betecknar en $t(\nu)$-fördelad stokastisk variabel). Obs att även $P(T < -t) = q$ samt att $t(\nu) \to \mathrm{N}(0, 1)$ då $\nu \to \infty$.

$\nu \diagdown q$	10%	5%	2.5%	1%	0.5%
1	3.07768	6.31375	12.7062	31.8205	63.6567
2	1.88562	2.91999	4.30265	6.96456	9.92484
3	1.63774	2.35336	3.18245	4.54070	5.84091
4	1.53321	2.13185	2.77645	3.74695	4.60409
5	1.47588	2.01505	2.57058	3.36493	4.03214
6	1.43976	1.94318	2.44691	3.14267	3.70743
7	1.41492	1.89458	2.36462	2.99795	3.49948
8	1.39682	1.85955	2.30600	2.89646	3.35539
9	1.38303	1.83311	2.26216	2.82144	3.24984
10	1.37218	1.81246	2.22814	2.76377	3.16927
11	1.36343	1.79588	2.20099	2.71808	3.10581
12	1.35622	1.78229	2.17881	2.68100	3.05454
13	1.35017	1.77093	2.16037	2.65031	3.01228
14	1.34503	1.76131	2.14479	2.62449	2.97684
15	1.34061	1.75305	2.13145	2.60248	2.94671
16	1.33676	1.74588	2.11991	2.58349	2.92078
17	1.33338	1.73961	2.10982	2.56693	2.89823
18	1.33039	1.73406	2.10092	2.55238	2.87844
19	1.32773	1.72913	2.09302	2.53948	2.86093
20	1.32534	1.72472	2.08596	2.52798	2.84534
21	1.32319	1.72074	2.07961	2.51765	2.83136
22	1.32124	1.71714	2.07387	2.50832	2.81876
23	1.31946	1.71387	2.06866	2.49987	2.80734
24	1.31784	1.71088	2.06390	2.49216	2.79694
25	1.31635	1.70814	2.05954	2.48511	2.78744
26	1.31497	1.70562	2.05553	2.47863	2.77871
27	1.31370	1.70329	2.05183	2.47266	2.77068
28	1.31253	1.70113	2.04841	2.46714	2.76326
29	1.31143	1.69913	2.04523	2.46202	2.75639
30	1.31042	1.69726	2.04227	2.45726	2.75000
40	1.30308	1.68385	2.02108	2.42326	2.70446
60	1.29582	1.67065	2.00030	2.39012	2.66028
120	1.28865	1.65765	1.97993	2.35782	2.61742
∞	1.28155	1.64485	1.95996	2.32635	2.57583

χ^2-kvantiler

I tabellen nedan finns lösningen x till ekvationen $P(X_\nu^2 < x) = q$, där $X_\nu^2 \sim \chi^2(\nu)$, för $q = 0.10, 0.05, 0.025, 0.01, 0.005$ och olika antal frihetsgrader ν.

$\nu \diagdown q$	10%	5%	2.5%	1%	0.5%
1	0.0157908	0.00393214	0.000982069	0.000157088	0.0000392704
2	0.210721	0.102587	0.0506356	0.0201007	0.0100251
3	0.584374	0.351846	0.215795	0.114832	0.0717218
4	1.06362	0.710723	0.484419	0.297109	0.206989
5	1.61031	1.14548	0.831212	0.554298	0.411742
6	2.20413	1.63538	1.23734	0.87209	0.675727
7	2.83311	2.16735	1.68987	1.23904	0.989256
8	3.48954	2.73264	2.17973	1.64650	1.34441
9	4.16816	3.32511	2.70039	2.08790	1.73493
10	4.86518	3.94030	3.24697	2.55821	2.15586
11	5.57778	4.57481	3.81575	3.05348	2.60322
12	6.30380	5.22603	4.40379	3.57057	3.07382
13	7.04150	5.89186	5.00875	4.10692	3.56503
14	7.78953	6.57063	5.62873	4.66043	4.07467
15	8.54676	7.26094	6.26214	5.22935	4.60092
16	9.31224	7.96165	6.90766	5.81221	5.14221
17	10.0852	8.67176	7.56419	6.40776	5.69722
18	10.8649	9.39046	8.23075	7.01491	6.26480
19	11.6509	10.1170	8.90652	7.63273	6.84397
20	12.4426	10.8508	9.59078	8.26040	7.43384
21	13.2396	11.5913	10.2829	8.89720	8.03365
22	14.0415	12.3380	10.9823	9.54249	8.64272
23	14.8480	13.0905	11.6886	10.1957	9.26042
24	15.6587	13.8484	12.4012	10.8564	9.88623
25	16.4734	14.6114	13.1197	11.5240	10.5197
26	17.2919	15.3792	13.8439	12.1981	11.1602
27	18.1139	16.1514	14.5734	12.8785	11.8076
28	18.9392	16.9279	15.3079	13.5647	12.4613
29	19.7677	17.7084	16.0471	14.2565	13.1211
30	20.5992	18.4927	16.7908	14.9535	13.7867
40	29.0505	26.5093	24.4330	22.1643	20.7065
50	37.6886	34.7643	32.3574	29.7067	27.9907
60	46.4589	43.1880	40.4817	37.4849	35.5345
70	55.3289	51.7393	48.7576	45.4417	43.2752
80	64.2778	60.3915	57.1532	53.5401	51.1719
90	73.2911	69.1260	65.6466	61.7541	59.1963
100	82.3581	77.9295	74.2219	70.0649	67.3276

χ^2-kvantiler (forts)

I tabellen nedan finns lösningen x till ekvationen $P(X_\nu^2 > x) = q$, där $X_\nu^2 \sim \chi^2(\nu)$, för $q = 0.10, 0.05, 0.025, 0.01, 0.005$ och olika antal frihetsgrader ν.

$\nu \diagdown q$	10%	5%	2.5%	1%	0.5%
1	2.70554	3.84146	5.02389	6.63490	7.87944
2	4.60517	5.99146	7.37776	9.21034	10.5966
3	6.25139	7.81473	9.34840	11.3449	12.8382
4	7.77944	9.48773	11.1433	13.2767	14.8603
5	9.23636	11.0705	12.8325	15.0863	16.7496
6	10.6446	12.5916	14.4494	16.8119	18.5476
7	12.0170	14.0671	16.0128	18.4753	20.2777
8	13.3616	15.5073	17.5345	20.0902	21.9550
9	14.6837	16.9190	19.0228	21.6660	23.5894
10	15.9872	18.3070	20.4832	23.2093	25.1882
11	17.2750	19.6751	21.9200	24.7250	26.7568
12	18.5493	21.0261	23.3367	26.2170	28.2995
13	19.8119	22.3620	24.7356	27.6882	29.8195
14	21.0641	23.6848	26.1189	29.1412	31.3193
15	22.3071	24.9958	27.4884	30.5779	32.8013
16	23.5418	26.2962	28.8454	31.9999	34.2672
17	24.7690	27.5871	30.1910	33.4087	35.7185
18	25.9894	28.8693	31.5264	34.8053	37.1565
19	27.2036	30.1435	32.8523	36.1909	38.5823
20	28.4120	31.4104	34.1696	37.5662	39.9968
21	29.6151	32.6706	35.4789	38.9322	41.4011
22	30.8133	33.9244	36.7807	40.2894	42.7957
23	32.0069	35.1725	38.0756	41.6384	44.1813
24	33.1962	36.4150	39.3641	42.9798	45.5585
25	34.3816	37.6525	40.6465	44.3141	46.9279
26	35.5632	38.8851	41.9232	45.6417	48.2899
27	36.7412	40.1133	43.1945	46.9629	49.6449
28	37.9159	41.3371	44.4608	48.2782	50.9934
29	39.0875	42.5570	45.7223	49.5879	52.3356
30	40.2560	43.7730	46.9792	50.8922	53.6720
40	51.8051	55.7585	59.3417	63.6907	66.7660
50	63.1671	67.5048	71.4202	76.1539	79.4900
60	74.3970	79.0819	83.2977	88.3794	91.9517
70	85.5270	90.5312	95.0232	100.425	104.215
80	96.5782	101.879	106.629	112.329	116.321
90	107.565	113.145	118.136	124.116	128.299
100	118.498	124.342	129.561	135.807	140.169