Matematisk statistik
KTH

Formel- och tabellsamling i Matematisk statistik, grundkurs

Vårterminen 2005

1. Kombinatorik

$\binom{n}{k} = \dfrac{n!}{k!(n-k)!}$. Tolkning: $\binom{n}{k}$ = antalet delmängder av storlek k ur en mängd med n element.

2. Stokastiska variabler

$V(X) = E(X^2) - (E(X))^2$

$C(X,Y) = E\big((X - E(X))(Y - E(Y))\big) = E(XY) - E(X)E(Y)$

$\rho(X,Y) = \dfrac{C(X,Y)}{D(X)D(Y)}$

3. Diskreta fördelningar

Binomialfördelningen

X är $\mathrm{Bin}(n,p)$ om $p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, 1, \ldots, n$, där $0 < p < 1$.

$E(X) = np$, $V(X) = np(1-p)$

”För-första-gången”-fördelningen

X är $\mathrm{ffg}(p)$ om $p_X(k) = p(1-p)^{k-1}$, $k = 1, 2, 3, \ldots$, där $0 < p < 1$.

$E(X) = \dfrac{1}{p}$, $V(X) = \dfrac{1-p}{p^2}$

Hypergeometriska fördelningen

X är $\mathrm{Hyp}(N,n,p)$ om $p_X(k) = \dfrac{\binom{Np}{k}\binom{N(1-p)}{n-k}}{\binom{N}{n}}$, $0 \le k \le Np$, $0 \le n-k \le N(1-p)$, där N, Np och n är positiva heltal samt $N \ge 2$, $n < N$, $0 < p < 1$. $E(X) = np$, $V(X) = \dfrac{N-n}{N-1} \cdot np(1-p)$

Poissonfördelningen

X är $\mathrm{Po}(\mu)$ där $\mu > 0$ om $p_X(k) = \dfrac{\mu^k}{k!} \cdot e^{-\mu}$, $k = 0, 1, 2, \ldots$

$E(X) = \mu$, $V(X) = \mu$

4. Kontinuerliga fördelningar

Likformig fördelning

X är $U(a,b)$ där $a < b$ om $f_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{för } a < x < b \\ 0 & \text{annars} \end{cases}$

$E(X) = \dfrac{a+b}{2}$, $V(X) = \dfrac{(b-a)^2}{12}$

Exponentialfördelningen

X är $\text{Exp}(\lambda)$ där $\lambda > 0$ om $f_X(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & \text{för } x > 0 \\ 0 & \text{annars} \end{cases}$

$E(X) = \dfrac{1}{\lambda}, \;\; V(X) = \dfrac{1}{\lambda^2}$

Normalfördelningen

X är $N(\mu, \sigma)$ om $f_X(x) = \dfrac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \;\; -\infty < x < \infty, \;\; \sigma > 0.$

$E(X) = \mu, \;\; V(X) = \sigma^2$

X är $N(\mu, \sigma)$ om och endast om $\dfrac{X - \mu}{\sigma}$ är $N(0,1)$.

Om Z är $N(0,1)$ så har Z fördelningsfunktionen $\Phi(x)$ enligt tabell 1 och täthetsfunktionen $\varphi(x) = \dfrac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2}, \;\; -\infty < x < \infty.$

En linjär sammansättning $\sum a_i X_i + b$ av oberoende, normalfördelade stokastiska variabler är normalfördelad.

5. Centrala gränsvärdessatsen

Om $X_1, X_2, \ldots, X_n$ är oberoende likafördelade stokastiska variabler med väntevärde μ och standardavvikelse $\sigma > 0$ så är $Y_n = X_1 + \cdots + X_n$ approximativt $N(n\mu, \sigma\sqrt{n})$ om n är stort.

6. Approximation

$\text{Hyp}(N, n, p) \sim \text{Bin}(n, p)$ om $\dfrac{n}{N} \le 0.1$

$\text{Bin}(n, p) \sim \text{Po}(np)$ om $p \le 0.1$

$\text{Bin}(n, p) \sim N\big(np, \sqrt{np(1-p)}\big)$ om $np(1-p) \ge 10$

$\text{Po}(\mu) \sim N(\mu, \sqrt{\mu})$ om $\mu \ge 15$

7. Tjebysjovs olikhet

Om $E(X) = \mu$ och $D(X) = \sigma > 0$ så gäller för varje $k > 0$ att

$P(|X - \mu| > k\sigma) \le \dfrac{1}{k^2}$

8. Statistiskt material

$$\overline{x} = \frac{1}{n} \sum_{j=1}^{n} x_j$$

$$s^2 = \frac{1}{n-1} \sum_{j=1}^{n} (x_j - \overline{x})^2 = \frac{1}{n-1} \left[\sum_{j=1}^{n} x_j^2 - \frac{1}{n} \Big(\sum_{j=1}^{n} x_j \Big)^2 \right]$$

9. Punktskattningar

9.1 Maximum-likelihood-metoden

Låt x_i vara en observation på X_i, $i = 1, 2, \ldots, n$, där fördelningen för X_i beror på en okänd parameter θ. Det värde θ^*_{obs} som maximerar L-funktionen

$$L(\theta) = \begin{cases} p_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\theta) = (\text{om oberoende}) = p_{X_1}(x_1;\theta)\cdots p_{X_n}(x_n;\theta) \\ f_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\theta) = (\text{om oberoende}) = f_{X_1}(x_1;\theta)\cdots f_{X_n}(x_n;\theta) \end{cases}$$

kallas *maximum-likelihood-skattningen (ML-skattningen)* av θ.

9.2 Minsta-kvadrat-metoden

Låt x_i vara en observation på X_i, $i = 1, 2, \ldots, n$, och antag att

$E(X_i) = \mu_i(\theta_1, \theta_2, \ldots, \theta_k)$ och $V(X_i) = \sigma^2$,

där $\theta_1, \theta_2, \ldots, \theta_k$ är okända parametrar och $X_1, X_2, \ldots, X_n$ är oberoende.

Minsta-kvadrat-skattningarna (MK-skattningarna) av $\theta_1, \theta_2, \ldots, \theta_k$

är de värden $(\theta_1)^*_{\text{obs}}, (\theta_2)^*_{\text{obs}}, \ldots, (\theta_k)^*_{\text{obs}}$ som minimerar kvadratsumman

$$Q = Q(\theta_1, \theta_2, \ldots, \theta_k) = \sum_{i=1}^{n} \big(x_i - \mu_i(\theta_1, \theta_2, \ldots, \theta_k)\big)^2.$$

9.3 Medelfel

En skattning av $D(\theta^*)$ kallas *medelfelet* för θ^* och betecknas $d(\theta^*)$.

9.4 Felfortplantning

Med beteckningar och förutsättningar enligt läroboken gäller

a) $E(g(\theta^*)) \approx g(\theta^*_{\text{obs}})$

$D(g(\theta^*)) \approx \left|g'(\theta^*_{\text{obs}})\right| \cdot D(\theta^*)$

b) $E(g(\theta^*_1, \ldots, \theta^*_n)) \approx g\big((\theta_1)^*_{\text{obs}}, \ldots, (\theta_n)^*_{\text{obs}}\big)$

$$V(g(\theta^*_1, \ldots, \theta^*_n)) \approx \sum_{i=1}^{n} \sum_{j=1}^{n} C(\theta^*_i, \theta^*_j) \cdot \left[\frac{\partial g}{\partial x_i} \cdot \frac{\partial g}{\partial x_j}\right]_{x_k = (\theta_k)^*_{\text{obs}}, k=1,\ldots,n}$$

10. Några vanliga fördelningar i statistiken

χ^2-fördelningen

Om $X_1, X_2, \ldots, X_f$ är oberoende och $N(0,1)$ så gäller att

$\sum_{k=1}^{f} X_k^2$ är $\chi^2(f)$-fördelad.

t-fördelningen

Om X är $N(0,1)$ och Y är $\chi^2(f)$ samt om X och Y är oberoende så gäller att $\dfrac{X}{\sqrt{Y/f}}$ är $t(f)$-fördelad.

11. Stickprovsvariablernas fördelningar vid normalfördelade stickprov

11.1 Låt $X_1, \ldots, X_n$ vara oberoende stokastiska variabler som alla är $N(\mu, \sigma)$.

Då gäller:

a) $\overline{X}$ är $N\Big(\mu, \dfrac{\sigma}{\sqrt{n}}\Big)$

b) $\dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2} = \dfrac{(n-1)S^2}{\sigma^2}$ är $\chi^2(n-1)$

c) $\overline{X}$ och S^2 är oberoende

d) $\dfrac{\overline{X} - \mu}{S/\sqrt{n}}$ är $t(n-1)$

11.2 Låt $X_1, \ldots, X_{n_1}$ vara $N(\mu_1, \sigma)$ och $Y_1, \ldots, Y_{n_2}$ vara $N(\mu_2, \sigma)$ och samtliga stokastiska variabler antas oberoende. Då gäller:

a) $\overline{X} - \overline{Y}$ är $N\Big(\mu_1 - \mu_2, \sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}\Big)$

b) $\dfrac{(n_1 + n_2 - 2)S^2}{\sigma^2}$ är $\chi^2(n_1 + n_2 - 2)$ där $S^2 = \dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$,

$$S_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_i - \overline{X})^2 \text{ och } S_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(Y_i - \overline{Y})^2$$

c) $\overline{X} - \overline{Y}$ och S^2 är oberoende

d) $\dfrac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ är $t(n_1 + n_2 - 2)$

11.3 Låt $X_1, \ldots, X_{n_1}$ vara $N(\mu_1, \sigma_1)$ och $Y_1, \ldots, Y_{n_2}$ vara $N(\mu_2, \sigma_2)$ och samtliga stokastiska variabler antas oberoende. Då gäller:

$\overline{X} - \overline{Y}$ är $N\Big(\mu_1 - \mu_2, \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}\Big)$

12. Konfidensintervall

12.1 λ-metoden

Låt θ^* vara $N(\theta, D)$ där D är känd och θ okänd. Då är

$\theta^*_{\text{obs}} \pm D \cdot \lambda_{\alpha/2}$

ett konfidensintervall för θ med konfidensgraden $1 - \alpha$.

12.2 t-metoden

Låt θ^* vara $N(\theta, D)$ där D och θ är okända och D inte beror på θ.

Låt D^*_{obs} vara en punktskattning av D sådan att $\dfrac{\theta^* - \theta}{D^*}$ är $t(f)$. Då är

$\theta^*_{\text{obs}} \pm D^*_{\text{obs}} \cdot t_{\alpha/2}(f)$

ett konfidensintervall för θ med konfidensgraden $1 - \alpha$.

12.3 Approximativa metoden

Låt θ^* vara approximativt $N(\theta, D)$.

Antag att D^*_{obs} är en lämplig punktskattning av D. Då är

$\theta^*_{\text{obs}} \pm D^*_{\text{obs}} \cdot \lambda_{\alpha/2}$

ett konfidensintervall för θ med den *approximativa* konfidensgraden $1 - \alpha$.

12.4 Metod baserad på χ^2-fördelning

Låt θ^*_{obs} vara en punktskattning av en parameter θ sådan att

$f \cdot \left(\dfrac{\theta^*}{\theta}\right)^2$ är $\chi^2(f)$. Då är

$$\left(\theta^*_{\text{obs}}\sqrt{\frac{f}{\chi^2_{\alpha/2}(f)}}\ ,\ \theta^*_{\text{obs}}\sqrt{\frac{f}{\chi^2_{1-\alpha/2}(f)}}\right)$$

ett konfidensintervall för θ med konfidensgraden $1 - \alpha$.

13. Linjär regression

13.1 Fördelningar

Låt Y_i vara $N(\alpha + \beta x_i, \sigma)$, $i = 1, 2, \ldots, n$, och oberoende. Då gäller:

a) $\beta^* = \dfrac{\sum_{i=1}^n (x_i - \overline{x})(Y_i - \overline{Y})}{\sum_{i=1}^n (x_i - \overline{x})^2}$ är $N\left(\beta, \dfrac{\sigma}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}}\right)$

b) $\alpha^* = \overline{Y} - \beta^* \overline{x}$ är $N\left(\alpha, \sigma\sqrt{\dfrac{1}{n} + \dfrac{(\overline{x})^2}{\sum_{i=1}^n (x_i - \overline{x})^2}}\right)$

c) $\alpha^* + \beta^* x_0$ är $N\left(\alpha + \beta x_0, \sigma\sqrt{\dfrac{1}{n} + \dfrac{(x_0 - \overline{x})^2}{\sum_{i=1}^n (x_i - \overline{x})^2}}\right)$

d) $\dfrac{(n-2)S^2}{\sigma^2}$ är $\chi^2(n-2)$ där $S^2 = \dfrac{1}{n-2}\displaystyle\sum_{i=1}^n (Y_i - \alpha^* - \beta^* x_i)^2$

e) S^2 är oberoende av α^* och β^*

13.2 Konfidensintervall

$$I_\alpha : \ \alpha^*_{\text{obs}} \pm t_{p/2}(n-2)\, s \sqrt{\frac{1}{n} + \frac{(\overline{x})^2}{\sum_{i=1}^n (x_i - \overline{x})^2}}$$

$$I_\beta : \ \beta^*_{\text{obs}} \pm t_{p/2}(n-2) \frac{s}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}}$$

$$I_{\alpha+\beta x_0} : \ \alpha^*_{\text{obs}} + \beta^*_{\text{obs}} x_0 \pm t_{p/2}(n-2)\, s \sqrt{\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{\sum_{i=1}^n (x_i - \overline{x})^2}}$$

13.3 Beräkningsaspekter

$$S_{xy} = \sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^n (x_i - \overline{x}) y_i = \sum_{i=1}^n x_i (y_i - \overline{y}) = \sum_{i=1}^n x_i y_i - n \overline{x}\,\overline{y}$$

$$S_{xx} = \sum_{i=1}^n (x_i - \overline{x})^2 = \sum_{i=1}^n x_i^2 - n \overline{x}^2$$

$$S_{yy} = \sum_{i=1}^n (y_i - \overline{y})^2$$

$$(n-2)s^2 = S_{yy} - (\beta^*_{\text{obs}})^2 S_{xx} = S_{yy} - \beta^*_{\text{obs}} \cdot S_{xy} = \min_{\alpha,\beta} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2$$

14. Hypotesprövning

14.1 Definitioner

Signifikansnivån (felrisken) α är (det maximala värdet av) P(förkasta H_0) då hypotesen H_0 är sann.

Styrkefunktionen $h(\theta) = P$(förkasta H_0) då θ är rätt parametervärde.

14.2 Konfidensmetoden

Förkasta $H_0 : \theta = \theta_0$ på nivån α om θ_0 ej faller inom ett lämpligt valt konfidensintervall med konfidensgraden $1 - \alpha$.

14.3 χ^2-test

Man gör n oberoende upprepningar av ett försök som ger något av resultaten $A_1, A_2, \ldots, A_r$ med respektive sannolikheter $P(A_1), P(A_2), \ldots, P(A_r)$. Låt för $j = 1, 2, \ldots, r$ den stokastiska variabeln X_j beteckna antalet försök som ger resultatet A_j.

Test av given fördelning: Vi vill testa $H_0 : P(A_1) = p_1, P(A_2) = p_2, \ldots,$ $P(A_r) = p_r$ för givna sannolikheter $p_1, p_2, \ldots, p_r$. Då blir

$Q = \sum_{j=1}^{r} \frac{(x_j - np_j)^2}{np_j}$ ett utfall av en approximativt $\chi^2(r-1)$-fördelad stokastisk variabel om H_0 är sann och $np_j \geq 5$, $j = 1, 2, \ldots, r$.

Om vi skattar k parametrar ur data, $\theta = (\theta_1, \ldots, \theta_k)$, för att skatta $p_1, p_2, \ldots, p_r$ med $p_1(\theta^*_{\text{obs}}), p_2(\theta^*_{\text{obs}}), \ldots, p_r(\theta^*_{\text{obs}})$ så är

$Q' = \sum_{j=1}^{r} \frac{\left(x_j - np_j(\theta^*_{\text{obs}})\right)^2}{np_j(\theta^*_{\text{obs}})}$ ett utfall av en approximativt

$\chi^2(r-k-1)$-fördelad stokastisk variabel.

Beräkningsaspekt: $Q = \sum_{j=1}^{r} \frac{x_j^2}{np_j} - n$, $Q' = \sum_{j=1}^{r} \frac{x_j^2}{np_j(\theta^*_{\text{obs}})} - n$

Homogenitetstest: Man vill testa om sannolikheterna för resultaten $A_1, A_2, \ldots, A_r$ är desamma i s försöksserier. Inför beteckningar enligt nedanstående tabell:

Serie	Antal observationer av					Antal försök
	A_1	A_2	A_3	$\ldots$	A_r	
1	x_{11}	x_{12}	x_{13}	$\ldots$	x_{1r}	n_1
2	x_{21}	x_{22}	x_{23}	$\ldots$	x_{2r}	n_2
$\vdots$	$\vdots$					$\vdots$
s	x_{s1}	x_{s2}	x_{s3}	$\ldots$	x_{sr}	n_s
Kolonnsumma	m_1	m_2	m_3	$\ldots$	m_r	N

Bilda $Q = \sum_{i=1}^{s} \sum_{j=1}^{r} \frac{\left(x_{ij} - \frac{n_i m_j}{N}\right)^2}{\frac{n_i m_j}{N}}$.

Q är ett utfall av en approximativt $\chi^2((r-1)(s-1))$-fördelad stokastisk variabel.

Kontingenstabell (test av oberoende mellan rader och kolonner):

Samma teststorhet och fördelning som ovan.

15. Markovprocesser

15.1 Markovkedjor i diskret tid

15.1.1 Grundläggande begrepp

Tillståndsrum: $\boldsymbol{E}$

Övergångsmatris: $\boldsymbol{P} = (p_{ij})_{i,j \in \boldsymbol{E}}$ där $p_{ij} = P(X_n = j \mid X_{n-1} = i)$

$\boldsymbol{p}^{(n)} = (p_1^{(n)}, p_2^{(n)}, \ldots)$ där $p_i^{(n)} = P(X_n = i)$.

Initialfördelning: $\boldsymbol{p}^{(0)}$

Stationär fördelning: $\boldsymbol{\pi}$

Gränsfördelning: $\boldsymbol{p}$

15.1.2 Chapman-Kolmogorovs ekvationer

a) $p_{ij}^{(m+n)} = \sum_{k \in \boldsymbol{E}} p_{ik}^{(m)} p_{kj}^{(n)}$

b) $\boldsymbol{P}^{(m+n)} = \boldsymbol{P}^{(m)} \boldsymbol{P}^{(n)}$

c) $\boldsymbol{P}^{(n)} = \boldsymbol{P}^n$

d) $\boldsymbol{p}^{(n)} = \boldsymbol{p}^{(0)} \boldsymbol{P}^{(n)} = \boldsymbol{p}^{(0)} \boldsymbol{P}^n$.

15.1.3 Stationära sannolikheter

Om $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_N)$ är en stationär sannolikhetsfördelning till en irreducibel markovkedja (ändlig eller oändlig), är $\pi_i = 1/E(T_i)$ där T_i är tiden mellan två besök i tillstånd i. π_j/π_i är förväntat antal besök i tillstånd j mellan två besök i tillstånd i.

15.1.4 Ergodicitet

1. En markovkedja kallas *ergodisk* om den har en asymptotisk fördelning som är oberoende av startfördelningen.
2. En ändlig markovkedja är ergodisk om och endast om dess tillståndsmängd $\boldsymbol{E}$ innehåller en enda sluten irreducibel deltillståndsmängd och denna är aperiodisk. Speciellt är en ändlig, irreducibel och aperiodisk markovkedja ergodisk.
3. En irreducibel, aperiodisk markovkedja (ändlig eller oändlig) är ergodisk om och endast om det existerar en stationär fördelning.

15.1.5 Absorption

En markovkedja kallas en A-kedja om $\boldsymbol{E} = \boldsymbol{A} \cup \boldsymbol{G}$, där alla tillstånd i $\boldsymbol{A}$ är absorberande och alla tillstånd i $\boldsymbol{G}$ leder till ett absorberande tillstånd i $\boldsymbol{A}$.

Sannolikheten för absorption i tillstånd $j \in \boldsymbol{A}$, då den startar i tillstånd i: a_{ij}.

Tiden till absorption, då den startar i tillstånd $i \in \boldsymbol{G}$: T_i, $t_i = E(T_i)$.

För en ändlig A-kedja gäller för alla $j \in \boldsymbol{A}$ att

$$a_{ij} = p_{ij} + \sum_{k \in \boldsymbol{G}} p_{ik} a_{kj}, \ i \in \boldsymbol{G}, \text{ och } t_i = 1 + \sum_{k \in \boldsymbol{G}} p_{ik} t_k, \ i \in \boldsymbol{G}.$$

15.2 Markovprocesser i kontinuerlig tid

15.2.1 Grundläggande begrepp

Övergångsmatris: $\boldsymbol{P}(t) = \big(p_{ij}(t)\big)_{i,j \in \boldsymbol{E}}$ där

$p_{ij}(t) = P(X(s+t) = j \mid X(s) = i)$

$\boldsymbol{p}(t) = (p_0(t), p_1(t), \ldots)$ där $p_i(t) = P(X(t) = i)$

Uthoppsintensitet från tillstånd i: q_i

Övergångsintensitet från tillstånd i till tillstånd j: q_{ij}

Stationär fördelning: $\boldsymbol{\pi}$

Gränsfördelning: $\boldsymbol{p}$

Intensitetsmatrisen (eller generatorn): $\boldsymbol{Q} = \lim\limits_{h \to 0+} \dfrac{\boldsymbol{P}(h) - \boldsymbol{I}}{h} = (q_{ij})_{i,j \in \boldsymbol{E}}$

där $q_{ii} = -q_i$.

15.2.2 Chapman-Kolmogorovs ekvationer

a) $p_{ij}(s+t) = \sum\limits_{k \in \boldsymbol{E}} p_{ik}(s) p_{kj}(t)$

b) $\boldsymbol{P}(s+t) = \boldsymbol{P}(s)\boldsymbol{P}(t)$

c) $\boldsymbol{p}(t) = \boldsymbol{p}(0)\boldsymbol{P}(t)$.

Kolmogorovs framåt- resp. bakåt-ekvation: $\boldsymbol{P}'(t) = \boldsymbol{P}(t)\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{P}(t)$

15.2.3 Stationära sannolikheter

1. $\boldsymbol{\pi}$ är en stationär fördelning till en markovprocess om och endast om $\boldsymbol{\pi}\boldsymbol{Q} = \boldsymbol{0}$.
2. Om $\boldsymbol{\pi} = (\pi_0, \pi_1, \ldots)$ är en stationär sannolikhetsfördelning till en ändlig (eller reguljär oändlig) irreducibel markovprocess, är $\pi_i = \dfrac{1/q_i}{E(T_i)}$ där T_i är tiden mellan två inträden i tillstånd i. $\dfrac{\pi_j}{q_i \pi_i}$ är förväntad tid i tillstånd j mellan två inträden i tillstånd i.

15.2.4 Ergodicitet

1. En ändlig och irreducibel markovprocess är ergodisk.
2. En reguljär, irreducibel markovprocess (ändlig eller oändlig) är ergodisk om och endast om det existerar en stationär fördelning.

15.2.5 Absorption

Definitioner se 15.1.5.

För en ändlig A-process gäller för alla $j \in \boldsymbol{A}$ att

$$a_{ij} = \frac{q_{ij}}{q_i} + \sum_{k \in \boldsymbol{G} \setminus \{i\}} \frac{q_{ik}}{q_i} a_{kj}\,,\; i \in \boldsymbol{G}, \text{ och } t_i = \frac{1}{q_i} + \sum_{k \in \boldsymbol{G} \setminus \{i\}} \frac{q_{ik}}{q_i} t_k\,,\; i \in \boldsymbol{G}.$$

15.2.6 Poissonprocesser

Om $\{N(t); t \geq 0\}$ är en poissonprocess med intensitet λ så är $N(t) - N(s) \in \mathrm{Po}(\lambda \cdot (t-s))$, för $t > s$, och processens ökningar i disjunkta tidsintervall är oberoende. $\{N(t); t \geq 0\}$ är en födelseprocess med alla $\lambda_i = \lambda$.

15.2.7 Födelseprocesser

En födelseprocess med födelseintensiteter $\lambda_0, \lambda_1, \ldots$ är en markovprocess med $X(0) = 0$ och med övergångsintensiteter

$$q_{ij} = \begin{cases} \lambda_i, & j = i+1,\, i = 0, 1, 2, \ldots \\ -\lambda_i, & j = i,\, i = 0, 1, 2, \ldots \\ 0, & \text{annars.} \end{cases}$$

En födelseprocess med födelseintensiteter $\lambda_0, \lambda_1, \ldots$ är reguljär, d.v.s. exploderar ej, om och endast om $\sum_{i=0}^{\infty} \frac{1}{\lambda_i} = \infty$.

15.2.8 Födelse-döds-processer

En födelse-döds-process med födelseintensiteter $\lambda_0, \lambda_1, \ldots$ och dödsintensiteter $\mu_1, \mu_2, \ldots$ är en markovprocess med övergångsintensiteter

$$q_{ij} = \begin{cases} \lambda_i, & j = i+1,\, i = 0, 1, 2, \ldots \\ \mu_i, & j = i-1,\, i = 1, 2, \ldots \\ -\lambda_i - \mu_i, & i = j,\, i = 1, 2, \ldots \\ -\lambda_0, & i = j = 0 \\ 0, & \text{annars.} \end{cases}$$

Låt $\rho_0 = 1$ och $\rho_i = \dfrac{\lambda_0 \cdot \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \cdot \mu_2 \cdots \mu_i}$.

En födelse-döds-process är reguljär om och endast om $\sum_{k=0}^{\infty} \frac{1}{\lambda_k \rho_k} \sum_{i=0}^{k} \rho_i = \infty$.

En födelse-döds-process har en stationär fördelning $\boldsymbol{\pi}$ om och endast om $\sum_{i=0}^{\infty} \rho_i < \infty$ och $\sum_{i=0}^{\infty} \frac{1}{\lambda_i \rho_i} = \infty$.

Den stationära fördelningen $\boldsymbol{\pi}$ är given av $\pi_j = \dfrac{\rho_j}{\sum_{i=0}^{\infty} \rho_i}$.

16. Köteori

16.1 Allmänt system

Vi förutsätter att inblandade gränsvärden existerar.

Beskrivning av systemet

$\lambda =$ ankomstintensitet.

$U =$ betjäningstiden för en godtycklig kund.

$B(t) = P(U \leq t) =$ betjäningstidens fördelningsfunktion.

$c =$ antalet betjäningsstationer.

Beteckningar

$\mu = 1/E(U) =$ betjäningsintensitet för en godtycklig kund.

$\rho = \dfrac{\lambda}{c\mu} =$ trafikintensiteten (betjäningsfaktorn). Vi förutsätter att $\rho < 1$.

$X(t) =$ antalet kunder i systemet vid tid t.

$\boldsymbol{p} = (p_0, p_1, \ldots) =$ gränsfördelningen för antalet kunder i systemet, d.v.s.
$p_i = \lim\limits_{t\to\infty} P(X(t) = i)$.

$\ell(t) = E(X(t)) =$ förväntat antal kunder i systemet vid tid t.

$\ell = \lim\limits_{t\to\infty} \ell(t) = \sum\limits_{i=0}^{\infty} ip_i =$ förväntat antal kunder i systemet (efter lång tid).

$X_q(t) =$ antalet kunder i kön vid tid t.

$\ell_q(t) = E(X_q(t)) =$ förväntat antal kunder i kön vid tid t.

$\ell_q = \lim\limits_{t\to\infty} \ell_q(t) =$ förväntat antal kunder i kön (efter lång tid).

$Q_n = n$:te kundens kötid.

$G_q(\tau) = \lim\limits_{n\to\infty} P(Q_n \leq \tau)$.

$U_n = n$:te kundens betjäningstid, $P(U_n \leq t) = B(t)$.

$S_n = Q_n + U_n = n$:te kundens tid i systemet.

$G(\tau) = \lim\limits_{n\to\infty} P(S_n \leq \tau)$.

Låt Q vara en s.v. med fördelningsfunktion $G_q(\tau)$.

$w_q = E(Q) =$ förväntad tid i kön.

$w = E(Q + U) =$ förväntad tid i systemet.

Under mycket allmänna villkor gäller att

$\ell = c\rho + \ell_q$, $w_q = \dfrac{\ell_q}{\lambda}$ och $w = \dfrac{\ell}{\lambda}$.

Då $c = 1$ gäller under mycket allmänna villkor att

sannolikheten att en kund måste stå i kö $= 1 - p_0 = \dfrac{\lambda}{\mu} = \rho$.

16.2 System med Poissonankomster och exponentialfördelade betjäningstider

System med $c \geq 1$

$$p_0 = \left(\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!} + \frac{(c\rho)^c}{c!(1-\rho)} \right)^{-1}.$$

$$p_n = \begin{cases} p_0 \cdot \dfrac{(c\rho)^n}{n!}, & n = 1, 2, \ldots, c, \\ p_c \cdot \rho^{n-c}, & n = c+1, c+2, \ldots \end{cases}$$

$$\ell_q = \frac{p_c \rho}{(1-\rho)^2}.$$

$$G_q(\tau) = 1 - \frac{p_c}{1-\rho} \cdot e^{-c\mu(1-\rho)\tau}, \quad \tau \geq 0.$$

Sannolikheten att en kund måste stå i kö $= \dfrac{p_c}{1-\rho}$.

System med $c = 1$

$$p_n = (1-\rho)\rho^n, \quad n = 0, 1, \ldots$$

$$\ell_q = \frac{\rho^2}{1-\rho}.$$

$$G_q(\tau) = 1 - \rho e^{-\mu(1-\rho)\tau}, \quad \tau \geq 0.$$

$$G(\tau) = 1 - e^{-\mu(1-\rho)\tau}, \quad \tau \geq 0.$$

System med $c = 2$

$$p_n = \begin{cases} \dfrac{1-\rho}{1+\rho}, & n = 0, \\ \dfrac{2(1-\rho)\rho^n}{1+\rho}, & n = 1, 2, \ldots \end{cases}$$

$$\ell_q = \frac{2\rho^3}{1-\rho^2}.$$

$$G_q(\tau) = 1 - \frac{2\rho^2}{1+\rho} \cdot e^{-2\mu(1-\rho)\tau}, \quad \tau \geq 0.$$

Förlustsystem med $c \geq 1$

Detta innebär att en kund, som inte får betjäning omedelbart, lämnar systemet.

$$p_n = \frac{\dfrac{(c\rho)^n}{n!}}{1 + \dfrac{c\rho}{1!} + \dfrac{(c\rho)^2}{2!} + \cdots + \dfrac{(c\rho)^c}{c!}}, \quad n = 0, 1, \ldots, c.$$

16.3 System med Poissonankomster och godtyckligt fördelade betjäningstider

System med $c = 1$

$$\ell_q = \frac{\rho^2}{2(1-\rho)} \cdot \left(1 + \frac{V(U)}{(E(U))^2}\right).$$

16.4 Jacksonnätverk

Ett könätverk som har m noder kallas ett *Jacksonnätverk* om följande villkor är uppfyllda:

1. Varje nod har identiska betjäningsstationer med exponentialfördelade betjäningstider. Nod i har c_i betjäningsstationer, vardera med betjäningsintensitet μ_i.
2. Kunder som kommer till nod i utifrån kommer enligt en Poissonprocess med intensitet λ_i. (Kunder kan även komma till nod i från andra noder i systemet.)
3. Så snart en kund är betjänad i nod i så går kunden till nod j med sannolikheten p_{ij} för $j = 1, \ldots, m$ eller lämnar systemet med sannolikheten $p_i = 1 - \sum_{j=1}^{m} p_{ij}$. Alla förflyttningar sker ögonblickligen.
4. Alla ankomstprocesser, betjäningstider och förflyttningar är oberoende av varandra och av systemet i övrigt.

Ankomstintensiteten till nod i ges av $\Lambda_i = \lambda_i + \sum_{j=1}^{m} p_{ji}\Lambda_j$ för $i = 1, \ldots, m$.

Låt $X_i(t)$ vara antalet kunder som befinner sig i nod i. För ett Jacksonnätverk med $\Lambda_i/(c_i\mu_i) < 1$ för $i = 1, \ldots, m$ gäller att

$$\lim_{t\to\infty} P(X_1(t) = n_1, X_2(t) = n_2, \ldots, X_m(t) = n_m) = p_{n_1}^{(1)} p_{n_2}^{(2)} \cdots p_{n_m}^{(m)},$$

där $(p_0^{(i)}, p_1^{(i)}, p_2^{(i)}, \ldots)$ är sannolikhetsfördelningen för ett $M/M/c_i$-system i stationärt tillstånd med ankomstintensitet Λ_i och förväntad betjäningstid $1/\mu_i$.