Sannolikhet och Statistik, F8

Rolf Larsson

Uppsala Universitet

22 september 2014

Idag

- 3.8.1 : Definition av flerdimensionella slumpvariabler
- 3.8.2 : Kovarians och korrelation
- 3.8.3 : Oberoende slumpvariabler

Flerdimensionella slumpvariabler

Fem kort dras slumpmässigt ur en kortlek.
Låt $X =$ antalet dragna hjärter
och $Y =$ antalet dragna ruter. Då är (X, Y) ett exempel på en tvådimensionell slumpvariabel.

Flerdimensionella slumpvariabler

Definition (Tvådimensionell slumpvariabel)

En *tvådimensionell slumpvariabel* $(X, Y) = \{X(u), Y(u)\}$ är en tvådimensionell funktion definierad på ett utfallsrum Ω, $(X, Y): \ \Omega \mapsto \mathcal{R} \times \mathcal{R}$.

Flerdimensionella slumpvariabler

Definition (Fördelningsfunktion)

Fördelningsfunktionen för en tvådimensionell slumpvariabel (X, Y) definieras som

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y).$$

Definition (Diskret slumpvariabel)

En tvådimensionell slumpvariabel (X, Y) sägs vara *diskret* om den endast kan anta ett ändligt eller uppräkneligt oändligt antal värden. *Sannolikhetsfunktionen* för en sådan slumpvariabel definieras av

$$p_{X,Y}(j, k) = P(X = j, Y = k).$$

Flerdimensionella slumpvariabler

Två kort dras slumpmässigt ur en kortlek.
Låt X = antalet dragna hjärter
och Y = antalet dragna ruter. Sannolikhetsfunktionen $p_{X,Y}(j,k)$ ges i tabell :

j \ k	0	1	2
0	25/102	26/102	6/102
1	26/102	13/102	0
2	6/102	0	0

1. Beräkna $F_{X,Y}(1,1)$
2. Beräkna $F_{X,Y}(2,1)$
3. Beräkna $F_Y(1)$
4. Beräkna $p_Y(1)$

Flerdimensionella slumpvariabler

Theorem

För en diskret tvådimensionell slumpvariabel (X, Y) *gäller att* den marginella fördelningsfunktionen *för* Y *ges av*

$$F_Y(y) = \lim_{x \to \infty} F_{X,Y}(x, y)$$

och den marginella sannolikhetsfunktionen *för* Y *ges av*

$$p_Y(k) = \sum_j p_{X,Y}(j, k).$$

Flerdimensionella slumpvariabler

Definition (Kontinuerlig slumpvariabel)

En tvådimensionell slumpvariabel (X, Y) sägs vara *kontinuerlig* om det finns en funktion $f_{X,Y}(x, y)$ så att för "alla" mängder A gäller

$$P\{(X, Y) \in A\} = \iint_A f_{X,Y}(x, y) dxdy.$$

Funktionen $f_{X,Y}(x, y)$ kallas för slumpvariabelns *täthetsfunktion*.

Flerdimensionella slumpvariabler

Exempel : (X, Y) har täthetsfunktion

$$f_{X,Y}(x,y) = \begin{cases} ye^{-xy}, & \text{om } x > 1, y > 0 \\ 0, & \text{annars} \end{cases}$$

Beräkna $F_{X,Y}(2,1)$

Flerdimensionella slumpvariabler

Theorem

För en kontinuerlig tvådimensionell slumpvariabel (X, Y) *gäller att* den marginella fördelningsfunktionen *för* Y *ges av*

$$F_Y(y) = \lim_{x \longrightarrow \infty} F_{X,Y}(x, y)$$

och den marginella täthetsfunktionen *för* Y *ges av*

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx.$$

Flerdimensionella slumpvariabler

Exempel : (X, Y) har täthetsfunktion

$$f_{X,Y}(x,y) = \begin{cases} ye^{-xy}, \text{ om } x > 1, y > 0 \\ 0, \text{ annars} \end{cases}$$

1. Beräkna $F_{X,Y}(x,y)$
2. Beräkna $F_Y(y)$
3. Beräkna $f_Y(y)$

Flerdimensionella slumpvariabler

Theorem

Låt (X, Y) vara en tvådimensionell slumpvariabel och $g(x, y)$ en reell funktion. Då gäller att

$$E\left\{g(X, Y)\right\} = \sum_j \sum_k g(j, k) p_{X,Y}(j, k)$$

om (X, Y) är diskret och

$$E\left\{g(X, Y)\right\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y) dxdy$$

om (X, Y) är kontinuerlig.

Flerdimensionella slumpvariabler

Theorem

För godtyckliga slumpvariabler X och Y gäller att

$$E(X+Y)=E(X)+E(Y).$$

Med induktion följer att, för godtyckliga slumpvariabler $X_1, X_2, ..., X_n$,

$$E(X_1+X_2+...+X_n)=E(X_1)+E(X_2)+...+E(X_n).$$

Kovarians och korrelation

Hur kan man mäta beroendet mellan två slumpvariabler X och Y?

Definition (kovarians, korrelationskoefficient)

Betrakta en tvådimensionell slumpvariabel (X, Y) med ändliga väntevärden μ_X och μ_Y och ändliga standardavvikelser σ_X och σ_Y. *Kovariansen* mellan X och Y definieras som

$$C(X, Y) = E\left\{(X - \mu_X)(Y - \mu_Y)\right\}.$$

Korrelationskoefficienten definieras som

$$\rho(X, Y) = \frac{C(X, Y)}{\sigma_X \sigma_Y}.$$

Man kan visa att $-1 \leq \rho \leq 1$.

Kovarians och korrelation

Theorem

För en tvådimensionell slumpvariabel (X, Y) *med ändliga väntevärden* μ_X *och* μ_Y *gäller att*

$$C(X, Y) = E(XY) - \mu_X \mu_Y .$$

Jämför :

$$V(X) = E\left\{(X - \mu_X)^2\right\} = C(X, X) = E(X^2) - \mu_X^2 .$$

Kovarians och korrelation

Två kort dras slumpmässigt ur en kortlek.
Låt X = antalet dragna hjärter
och Y = antalet dragna ruter.

Sannolikhetsfunktionen $p_{X,Y}(j,k)$ ges i tabell :

j╲ k	0	1	2
0	25/102	26/102	6/102
1	26/102	13/102	0
2	6/102	0	0

Beräkna

1. $C(X,Y)$
2. $\rho(X,Y)$

Kovarians och korrelation

Två kort dras slumpmässigt ur en kortlek.
Låt X = antalet dragna hjärter
och Y = antalet dragna äss.

Sannolikhetsfunktionen $p_{X,Y}(j,k)$ ges i tabell :

j \ k	0	1	2
0	210/442	36/442	1/442
1	144/442	24/442	1/442
2	22/442	4/442	0

Beräkna

1. $C(X, Y)$
2. $\rho(X, Y)$

Oberoende slumpvariabler

Definition (Oberoende och okorrelerade)

Två slumpvariabler X och Y sägs vara *oberoende* om deras simultana fördelning uppfyller

$p_{X,Y}(x,y) = p_X(x)p_Y(y)$ för alla (x,y) i det diskreta fallet och
$f_{X,Y}(x,y) = f_X(x)f_Y(y)$ för alla (x,y) i det kontinuerliga fallet.
Slumpvariablerna sägs vara *okorrelerade* om $C(X,Y) = 0$.

Observera : $C(X,Y) = 0$ är ekvivalent med $E(XY) = E(X)E(Y)$.

Theorem

Om X och Y är oberoende så är de också okorrelerade.

Oberoende slumpvariabler

Två kort dras slumpmässigt ur en kortlek.
Låt X = antalet dragna hjärter
och Y = antalet dragna ruter.

Sannolikhetsfunktionen $p_{X,Y}(j,k)$ ges i tabell :

j \ k	0	1	2
0	25/102	26/102	6/102
1	26/102	13/102	0
2	6/102	0	0

Är X och Y

1. oberoende ?
2. okorrelerade ?

Oberoende slumpvariabler

Två kort dras slumpmässigt ur en kortlek.
Låt X = antalet dragna hjärter
och Y = antalet dragna äss.

Sannolikhetsfunktionen $p_{X,Y}(j,k)$ ges i tabell :

j \ k	0	1	2
0	210/442	36/442	1/442
1	144/442	24/442	1/442
2	22/442	4/442	0

Är X och Y

1. oberoende ?
2. okorrelerade ?

Oberoende slumpvariabler

Theorem

För godtyckliga slumpvariabler X och Y gäller

$$V(X+Y) = V(X) + V(Y) + 2C(X,Y)$$
$$V(X-Y) = V(X) + V(Y) - 2C(X,Y)$$

Om X och Y är okorrelerade är alltså

$$V(X+Y) = V(X) + V(Y).$$

Med induktion följer för okorrelerade slumpvariabler $X_1, X_2, ..., X_n$ att

$$V(X_1 + X_2 + ... + X_n) = V(X_1) + V(X_2) + ... + V(X_n).$$

Oberoende slumpvariabler

- Antag att $X_1, X_2, \dots, X_n$ är okorrelerade med samma väntevärde μ och samma varians σ^2.
- Då gäller att

$$E(X_1 + X_2 + \dots + X_n) = n\mu$$
$$V(X_1 + X_2 + \dots + X_n) = n\sigma^2$$

- För $\overline{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$ gäller att

$$E(\overline{X}) = \mu$$
$$V(\overline{X}) = \frac{1}{n}\sigma^2$$

Oberoende slumpvariabler

- Låt $U_1, \dots, U_n$ vara oberoende $\mathrm{Be(p)}$, alltså
$$P(U_k = 1) = p$$
$$P(U_k = 0) = q = 1 - p$$
för $k = 1, \dots, n$.
- Då gäller att (varför ?) $X = U_1 + \dots + U_n$ är $\mathrm{Bin}(n, p)$.
- Vi har sett att $E(U_k) = p$, $V(U_k) = pq$ för $k = 1, \dots, n$.
- Det följer att (varför ?)
$$E(X) = np$$
$$V(X) = npq$$
- och att för $\overline{U} = X/n$
$$E(\overline{U}) = p$$
$$V(\overline{U}) = \frac{pq}{n}$$