6. Flerdimensionella stokastiska variabler

6.1 Simultana fördelningar

Den *simultana fördelningsfunktionen* av X och Y, vilka som helst två stokastiska variabler, definieras

$$F(a,b) = F_{X,Y}(a,b) = P(X \leq a, Y \leq b), \quad -\infty < a,b < \infty.$$

Fördelningen för X och Y kan härledas genom att använda den simultana fördelningen ovan:

$$F_X(a) = lim_{b\to\infty} F(a,b) = F(a,\infty)$$

och

$$F_Y(b) = lim_{a\to\infty} F(a,b) = F(\infty,b).$$

F_x och F_Y är de *marginala fördelningsfunktionerna* av X resp. Y.

Om X ovh Y är diskreta, är den simultana frekvensfunktionen av X och Y

$$p(x,y) = P(X = x, Y = y).$$

De *marginala frekvensfunktionerna* får man

$$p_X(x) = P(X = x) = \sum_{y:p(x,y)>0} p(x,y)$$

$$p_Y(y) = P(Y = y) = \sum_{x:p(x,y)>0} p(x,y)$$

Ex.1a: Det finns 3 röda, 4 vita och 5 blåa bollar i en urna. Tre bollar dras på måfå utan återläggning ur urnan. Låt

$$X = \text{antalet röda bollar dragna}$$

$$Y = \text{antalet vita bollar dragna}$$

Vad är $p(x,y)$?

De stokastiska variablerna X och Y är simultant kontinuerliga om det existerar en funktion $f(x,y)$, som är definierad för alla $x, y \in \mathbf{R}$ och som har egenskapen att för varje mängd C i $\mathbf{R}^2$

$$P((X,Y) \in C) = \int\int_{(x,y)\in C} f(x,y)\,dx\,dy.$$

Funktionen f kallas den *simultana täthetsfuktionen* av X och Y.

Om $C = \{(x, y) : x \in A, y \in B\}$, då är

$$P((X, Y) \in C) = P(X \in A, Y \in B) = \int_B \int_A f(x, y)\, dx\, dy.$$

Obs! $f(a, b) = \frac{\partial^2}{\partial a \partial b} F(a, b)$, när de partiella derivatorna är definierade.

Om X och Y är simultant kontinuerliga, då är de också individuelt kontinuerliga. Täthetsfunktionerna av X och Y kan härledas genom att använda den simultana täthetsfunktionen:

$$P(X \in A) = P(X \in A, Y \in (-\infty, \infty)) = \int_A \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = \int_A f_X(x)\, dx,$$

där $f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$ är täthetsfunktionen för X. På samma sätt får man att $f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$.

Ex.1c: Den simultana täthetsfunktionen av X och Y är

$$f(x, y) = \begin{cases} 2e^{-x}e^{-2y}, & 0 < x < \infty, 0 < y < \infty \\ 0 & \text{annars} \end{cases}.$$

Beräkna

a) $P(X > 1, Y < 1)$

b) $P(X < Y)$

c) $P(X < a)$

Ex.1e: Den simultana täthetsfunktionen av X och Y är

$$f(x, y) = \begin{cases} e^{-(x+y)}, & 0 < x < \infty, 0 < y < \infty \\ 0 & \text{annars} \end{cases}.$$

Vad är täthetsfunktionen av $\frac{X}{Y}$?

Den simultana täthetsfunktionen av n stokastiska variabler definieras på ett motsvarande sätt: Låt X_1, X_2, ..., X_n vara stokastiska variabler. Då är den simultana fördelningsfunktionen av X_1, X_2, ..., X_n

$$F(a_1, a_2, ..., a_n) = P(X_1 \leq a_1, X_2 \leq a_2, ..., X_n \leq a_n).$$

n stokastiska variabler är simultant kontinuerliga om det finns en funktion $f(x_1, x_2, ..., x_n)$, *simultan täthetsfunktion*, sådan att för vilken som helst C i $\mathbf{R}^n$

$$P((X_1, X_2, ..., X_n) \in C) = \int \int \cdots \int_{(x_1, x_2, ..., x_n) \in C} f(x_1, x_2, ..., x_n)\, dx_1 \cdots dx_n.$$

Dvs. att om man har n mängder i $\mathbf{R}$, A_1, A_2,..., A_n, då är

$$P(X_1 \in A_1, ..., X_n \in A_n) = \int_{A_n} \cdots \int_{A_1} f(x_1, x_2, ..., x_n)\, dx_1 \cdots dx_n.$$

Ex.1f: (Multinomialfördelning; en av de viktigaste simultana fördelningarna) Anta att man utför en följd av oberoende och identiska experiment och att vart och ett av dem har r möjliga utfall med resp. sannolikheter p_1, p_2,..., p_r, $\sum_{i=1}^{r} p_i = 1$. Låt X_i vara antalet experiment (bland de n) som resulterar utfallet i. Då är

$$P(X_1 = n_1, X_2 = n_2, ..., X_r = n_r) = \frac{n!}{n_1! n_2! \cdots n_r!} p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r},$$

där $\sum_{i=1}^{r} n_i = n$.

6.2 Oberoende stokastiska variabler

Två stokastiska variabler X och Y är *oberoende* om för vilka som helst två mängder A och B i $\mathbf{R}$

$$P(X \in A, Y \in B) = P(X \in A) P(Y \in B).$$

Dvs. att X och Y är oberoende om händelserna $\{X \in A\}$ och $\{Y \in B\}$ är oberoende.

Ovan är ekvivalent med

$$P(X \leq a, Y \leq b) = P(X \leq a) P(Y \leq b) \quad \text{för alla} \quad a, b,$$

och med

$$F(a, b) = F_X(a) F_Y(b) \quad \text{för alla} \quad a, b.$$

Om X och Y är diskreta stokastiska variabler, är oberoendet av dem ekvivalent med

$$p(x, y) = p_X(x) p_Y(y) \quad \text{för alla} \quad x, y.$$

I kontinuerliga fallet är oberoendet av X och Y ekvivalent med

$$f(x, y) = f_X(x) f_Y(y) \quad \text{för alla} \quad x, y.$$

Ex.2a: Anta att man utför $n + m$ oberoende försök, vart och ett av vilka har gemensam succésannolikhet p. Låt X vara antalet succéer bland de n första försöken och Y antalet succéer bland de m sista försöken. Är X och Y oberoende? Om Z är det totala antalet succéer bland de $n + m$ försöken, är X (eller Y) och Z oberoende?

Ex: Låt den simultana täthetsfunktionen av X och Y vara

$$f(x,y) = \begin{cases} 4x(1-y), & 0 < x < 1, 0 < y < 1 \\ 0 & \text{annars} \end{cases}.$$

Är X och Y oberoende?

Ex: Man tänder två nya glödlampor dess livslängder X och Y antas vara oberoende och $Exp(\frac{1}{10})$ resp. $Exp(\frac{1}{12})$ fördelade (enhet timmar). Hur stor är sannolikheten att både två lamporna slocknar innan de har används i 8 timmar?

Ex.2c: En man och en kvinna bestämmer sig att träffas på ett visst ställe. Var och en av dem kommer oberoende av varandra vid en tid, som är likformigt fördelad mellan 12 och 13. Vad är sannolikheten att den som kommer först måste vänta längre än 10 minuter?

Proposition 2.1: Kontinuerliga (diskreta) variabler X och Y är oberoende omm deras simultana täthetsfunktion (frekvensfunktion) kan skrivas

$$f_{X,Y}(x,y) = h(x)g(y), \quad -\infty < x < \infty, -\infty < y < \infty.$$

Ex.2f: Den simultana täthetsfunktionen av X och Y är

$$f(x,y) = \begin{cases} 24xy, & 0 < x < 1, 0 < y < 1, 0 < x+y < 1 \\ 0 & \text{annars} \end{cases}.$$

Är X och Y oberoende?

Oberoendet av fler än två stokastiska variabler: X_1, X_2,..., X_n är oberoende om för alla mängder A_1, A_2,..., A_n i $\mathbf{R}$

$$P(X_1 \in A_1, ..., X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i)$$

eller om

$$P(X_1 \leq a_1, ..., X_n \leq a_n) = \prod_{i=1}^{n} P(X_i \leq a_i) \quad \text{för alla } a_1, a_2, ..., a_n.$$

En oändlig samling av stokastiska variabler är oberoende om varje delsamling är oberoende.

Ex.2h: Låt X, Y och Z vara oberoende och $Lik(0,1)$-fördelade stokastiska variabler. Beräkna $P(X \geq YZ)$.

6.3 Summor av oberoende stokastiska variabler

Låt X och Y vara oberoende stokastiska variabler, som har täthetsfunktioner f_X resp. f_Y. Täthetsfunktionen av $X+Y$ är

$$f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a-y) f_Y(y)\, dy.$$

Ex.3a: Låt X och Y vara $Lik(0,1)$-fördelade. Vad är täthetsfunktionen av $X+Y$?

Proposition 3.1: Låt X och Y vara oberoende gammafördelade stokastiska variabler med resp. parametrar (s,λ) och (t,λ). Då är $X+Y$ gammafördelad med parametrar $(s+t,\lambda)$.

Låt X_1, X_2,..., X_n vara oberoende gammafördelade stokastiska variabler med resp. parametrar (t_1,λ),..., (t_n,λ). Då är $X_1 + X_2 + ... + X_n \sim Gamma(\sum_{i=1}^{n} t_i, \lambda)$.

Ex.3b: Låt X_1, X_2,..., X_n vara oberoende $Exp(\lambda)$-fördelade stokastiska variabler. Då är $X_1 + X_2 + ... + X_n \sim Gamma(n,\lambda)$.

Proposition 3.2: Låt X_1, X_2,..., X_n vara oberoende normalfördelade stokastiska variabler med resp. parametrar (μ_i, σ_i^2), $i = 1, ..., n$. Då är $X_1 + X_2 + ... + X_n \sim N(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2)$.

Ex.3c: Ett basketbollag spelar 44 matcher, 26 av vilka är mot A-lag och 18 mot B-lag. Laget vinner var och en av matcherna mot ett A-lag med sannolikhet 0.4 och mot ett B-lag med sannolikhet 0.7. Resultaten av de olika matcherna antas vara oberoende. Approximera sannolikheten att

a) laget vinner minst 25 av matcherna.

b) laget vinner fler matcher mot A-lagen än mot B-lagen.

Ex.3e: Låt X och Y vara oberoende Poissonfördelade stokastiska variabler med resp. parametrar λ_1 och λ_2. Hitta fördelningen för $X+Y$.

Ex.3f: Låt X och Y vara oberoende binomialfördelade stokastiska variabler med resp. parametrar (n,p) och (m,p). Hitta fördelningen för $X+Y$.

6.4 Betingade fördelningar: diskreta fallet

Låt X och Y vara diskreta stokastiska variabler. Då är den *betingade frekvensfunktionen* av X givet $Y=y$

$$p_{X|Y}(x|y) = P(X=x|Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)} = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

för y för vilka $p_Y(y) > 0$. På motsvarande sätt definierar man den *betingade fördelningsfunktionen* av X givet $Y = y$, dvs.

$$F_{X|Y}(x|y) = P(X \leq x|Y = y) = \sum_{a \leq x} p_{X|Y}(a|y).$$

Obs! Om X och Y är oberoende, då är

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} = \frac{p_X(x)p_Y(y)}{p_Y(y)} = p_X(x).$$

Ex.4b: Låt $X \sim Pois(\lambda_1)$ och $Y \sim Pois(\lambda_2)$ vara oberoende. Härleda den betingade fördelningen av X givet $X + Y = n$.

6.5 Betingade fördelningar: kontinuerliga fallet

Låt $f(x,y)$ vara den simultana täthetsfunktionen av kontinuerliga stokastiska variabler X och Y. Då är den betingade täthetsfunktionen av X givet $Y = y$

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

för y för vilka $f_Y(y) > 0$.

Om X och Y är simultant kontinuerliga, då är för vilken $A \in \mathbf{R}$ som helst

$$P(X \in A|Y = y) = \int_A f_{X|Y}(x|y)\, dx.$$

Om $A = (-\infty, a]$, kan man definiera den *betingade fördelningsfunktionen* av X givet $Y = y$

$$F_{X|Y}(a|y) = P(X \leq a|Y = y) = \int_{-\infty}^{a} f_{X|Y}(x|y)\, dx.$$

Ex.5b: Den simultana täthetsfunktionen av X och Y är

$$f(x,y) = \begin{cases} \frac{e^{-x/y}e^{-y}}{y}, & 0 < x < \infty, 0 < y < \infty \\ 0 & \text{annars} \end{cases}.$$

Hitta $P(X > 1|Y = y)$.

Obs! Om X och Y är oberoende, då är

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_X(x)f_Y(y)}{f_Y(y)} = f_X(x).$$

6.6 Ordningsstatistikor

Låt X_1, X_2,..., X_n vara oberoende likafördelade kontinuerliga stokastiska variabler med gemensam täthetsfunktion f och fördelningsfunktion F. Man definierar

$X_{(1)}$ = minst av X_1, X_2,..., X_n
$X_{(2)}$ = andra minst av X_1, X_2,..., X_n
....
$X_{(n)}$ = störst av X_1, X_2,..., X_n

Då är $X_{(1)} < X_{(2)} < ... < X_{(n)}$ och de kallas *ordningsstatistikor* som motsvarar de stokastiska variablerna X_1, X_2,..., X_n.

Täthetsfunktionen av $X_{(n)} = max\{X_1, ..., X_n\}$ är

$$f_{X_{(n)}}(x) = nf(x)(F(x))^{n-1}$$

och täthetsfunktionen av $X_{(1)} = min\{X_1, ..., X_n\}$ är

$$f_{X_{(1)}}(x) = nf(x)(1 - F(x))^{n-1}.$$

Ex: Låt X_1 och X_2 vara oberoende $Lik(0, 1)$-fördelade stokastiska variabler. Vad är väntevärdet av $max\{X_1, X_2\}$ och $min\{X_1, X_2\}$?