7 Egenskaper av väntevärdet

7.2 Väntevärdet av summor av stokastiska variabler

Proposition 2.1: Om stokastiska variabler X och Y har den simultana frekvensfunktionen $p(x, y)$, då är

$$\mathbf{E}[g(X,Y)] = \sum_y \sum_x g(x,y)p(x,y).$$

Om X och Y har den simultana täthetsfunktionen $f(x, y)$, då är

$$\mathbf{E}[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)f(x,y)\,dx\,dy.$$

Ex.2a: Det inträffar en olycka i en viss punkt X som är likformigt fördelad på en väg av längd L. När olyckan inträffar, finns det en ambulans på vägen i punkt Y, som är också likformigt fördelad på vägen. Anta att X och Y är oberoende. Vad är det förväntade avståndet mellan ambulansen och punkten, där olyckan inträffar?

Låt $g(X, Y) = X + Y$, $\mathbf{E}[X] < \infty$ och $\mathbf{E}[Y] < \infty$. Då är

$$\mathbf{E}[g(X,Y)] = \mathbf{E}[X+Y] = \mathbf{E}[X] + \mathbf{E}[Y].$$

Genom att använda induktion kan man visa att för X_1, X_2,..., X_n, $\mathbf{E}[X_i] < \infty$, $i = 1, ..., n$, är

$$\mathbf{E}[X_1 + ... + X_n] = \mathbf{E}[X_1] + ... + \mathbf{E}[X_n].$$

Ex.2c: LåtX_1, X_2,..., X_n vara oberoende och likafördelade stokastiska variabler som har fördelningsfunktionen F och väntevärdet μ. Man säger att en sådan följd av stokastiska variabler är ett *stickprov* (sample) från fördelningen F. Storheten

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

kallas ett *stickprovsmedelvärde*. Beräkna $\mathbf{E}[\bar{X}]$.

Ex.2e (väntevärdet av $Bin(n,p)$): Låt $X \sim Bin(n,p)$. Beräkna $\mathbf{E}[X]$.

Ex.2h: Anta att N personer kastar sina hattar på golvet. Man blandar hattarna och sedan tar var och en av personerna en av hattarna på måfå. Vad är väntevärdet för antalet personer som väljer sin egen hatt?

Ex.2k: 10 män jagar änder. När en flock av änder flyger över, skjuter jägarna samtidigt. Var och en av jägarna väljer var sin and på måfå och oberoende av varandra. Var och en av dem träffar sin and med sannolikhet p oberoende av de andra. Vad är det förväntade antalet änder som inte blir skjutna, när en flock av storlek 10 flyger över?

Ex.2j: Man samlar kuponger. Anta att det finns N olika kuponger och att varje gång man får en ny kupong är det lika sannolikt att den är vilken som helst av de N typerna.

a) Vad är det förväntade antalet kuponger som finns i en samling av n kuponger?

b) Vad är det förväntade antalet kuponger som man behöver ha tills man har en fullständig samling (minst en kupong av varje typ)?

7.3 Kovarians, varians av summor och korrelation

Proposition 3.1: Om X och Y är oberoende, då är för varje funktion h och g $(h(x), h(y) \in \mathbf{R})$

$$\mathbf{E}[g(X)h(Y)] = \mathbf{E}[g(X)]\mathbf{E}[h(Y)].$$

Definition: *Kovariansen* mellan X och Y, $Cov(X,Y)$, definieras

$$Cov(X,Y) = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y].$$

Obs! Om X och Y är oberoende är $Cov(X,Y) = 0$.

Ex: Låt X vara en stokastisk variabel som kan anta värdena 0, 1 och -1, var och ett med sannolikhet $\frac{1}{3}$ och låt

$$Y = \begin{cases} 0 & \text{om } X \neq 0 \\ 1 & \text{om } X = 0 \end{cases}.$$

Vad är kovariansen mellan X och Y? Är X och Y oberoende?

Proposition 3.2:

(i) $Cov(X,Y) = Cov(Y,X)$

(ii) $Cov(X,X) = Var(X)$

(iii) $Cov(aX,Y) = aCov(X,Y)$

(iv) $Cov(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j)$.

Varians av en summa av stokastiska variabler:

$$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Cov(X_i, X_j).$$

Ex.3a: Låt X_1, X_2,..., X_n vara oberoende och likafördelade stokastiska variabler med väntevärdet μ och variansen σ^2. Låt $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ vara stickprovsmedelvärdet. Den stokastiska variabeln

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

kallas *stickprovsvariansen*. Beräkna $Var(\bar{X})$ och $\mathbf{E}[S^2]$.

Ex.3c: Beräkna variansen av X, där X är antalet personer som väljer sin egen hatt i **Ex.2h**.

Definition: *Korrelationen* mellan två stokastiska variabler X och Y, $\rho(X, Y)$, definieras

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}},$$

om $Var(X) > 0$ och $Var(Y) > 0$.

Obs! $|\rho(X,Y)| \leq 1$ och $|\rho(X,Y)| = 1$ om och endast om $Y = a + bX$, där a och b är konstanter.

Ex.3g: Låt X_1, X_2,..., X_n vara oberoende och likafördelade stokastiska variabler med variansen σ^2. Visa att $Cov(X_i - \bar{X}, \bar{X}) = 0$.

Obs! $\bar{X}$ och $X_i - \bar{X}$ är inte nödvändigtvist oberoende.

7.4 Betingat väntevärde

7.4.1 Definitioner

Diskreta fallet: Låt X och Y vara diskreta stokastiska variabler med simultan frekvensfunktion $p(x, y)$. Då definieras den betingade frekvensfunktionen av X givet $Y = y$

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

för alla y sådana att $p_Y(y) > 0$. Det *betingade väntevärdet* av X givet $Y = y$ definieras

$$\mathbf{E}[X|Y = y] = \sum_x xP(X = x|Y = y) = \sum_x xp_{X|Y}(x|y), \quad p_Y(y) > 0.$$

Kontinuerliga fallet: Låt X och Y vara simultant kontinuerliga stokastiska variabler med den simultana täthetsfunktionen $f(x, y)$. Då definieras den betingade täthetsfunktionen av X givet $Y = y$

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

för y sådan att $f_Y(y) > 0$. Det *betingade väntevärdet* av X givet $Y = y$ definieras

$$\mathbf{E}[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, dx$$

för varje y för vilken $f_Y(y) > 0$.

Obs! Det betingade väntevärdet har alla egenskaper som det obetingade (vanliga) väntevärdet har.

Ex.4a: Om X och Y är oberoende $Bin(n, p)$-fördelade stokastiska variabler, vad är det betingade väntevärdet av X givet $X + Y = m$?

Ex.4b: Den simultana täthetsfunktionen av X och Y är

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

Vad är $\mathbf{E}[X|Y = y]$?

7.4.2 Beräkning av väntevärden genom betingning

Obs! $\mathbf{E}[X|Y]$ är en funktion av den stokastiska variabeln Y, som antar värdet $\mathbf{E}[X|Y = y]$ då $Y = y$.

Proposition 4.1:

$$\mathbf{E}[X] = \mathbf{E}[\mathbf{E}[X|Y]] = \begin{cases} \sum_y \mathbf{E}[X|Y = y] P(Y = y), & \text{om } Y \text{ diskret} \\ \int_{-\infty}^{\infty} \mathbf{E}[X|Y = y] f_Y(y)\, dy, & \text{om } Y \text{ kontinuerlig} \end{cases} .$$

Ex.4c: En gruvarbetare är i en gruva och vet inte hur han kan gå ut därifrån. Gruvan har tre dörrar: Om gruvarbetaren väljer

dörr 1, är han ute efter 3 timmar

dörr 2, är han tillbaka till gruvan efter 5 timmar

dörr 3, är han tillbaka till gruvan efter 7 timmar.

Varje gång väljer han en av dörrarna på måfå. Vad är då den förväntade tiden tills han är ute?

Ex.4f: Det finns a vita och b svarta bollar i en urna. Man drar en av bollarna åt gången tills man drar en vit boll för första gången. Vad är det förväntade antalet svarta bollar som dras?

7.4.3 Beräkning av sannolikheter genom betingning

Låt E vara en godtycklig händelse och låt

$$X = \begin{cases} 1, & \text{om } E \text{ inträffar} \\ 0 & \text{annars} \end{cases} .$$

Då är $\mathbf{E}[X] = P(X = 1) = P(E)$ och $\mathbf{E}[X|Y = y] = P(E|Y = y)$ för vilken som helst stokastisk variabel Y. Då är

$$P(E) = \begin{cases} \sum_y P(E|Y = y)P(Y = y), & \text{om } Y \text{ diskret} \\ \int_{-\infty}^{\infty} P(E|Y = y) f_Y(y)\, dy, & \text{om } Y \text{ kontinuerlig} \end{cases} .$$

Ex.4i (bästa priset -problemet/bästa sekreterare -problemet): Man presenterar n skilda priser i en följd. Efter att ett pris har presenterats, måste man antingen acceptera det eller överge det. Den enda informationen man får är rangordningen av det nya priset, som just presenterades, jämfört med de priserna man har redan sett. Om man överger ett pris, har man förlorat det. Man vill maximera sannolikheten att man får det bästa priset. Anta att var och en av de $n!$ ordningarna av priserna är lika sannolika. Hur väl kan man göra?

Ex.4l: Låt X och Y vara oberoende kontinuerliga stokastiska variabler. Vad är fördelningen för $X + Y$?

7.4.4 Betingad varians

Definition: Den *betingade variansen* av X givet $Y = y$ definieras

$$Var(X|Y) = \mathbf{E}[(X - \mathbf{E}[X|Y])^2|Y] = \mathbf{E}[X^2|Y] - (\mathbf{E}[X|Y])^2.$$

Proposition 4.2: $Var(X) = \mathbf{E}[Var(X|Y)] + Var(\mathbf{E}[X|Y])$.

Ex.4m: Antalet passagerare som kommer till järnvägstationen under en tidsperiod av längd t är $Poisson(\lambda t)$-fördelat. Tåget (som passagerarna vill åka med) ankommer till stationen vid en tidpunkt som är $Lik(0, T)$-fördelad (oberoende av när passagerarna kommer). Vad är väntevärdet och variansen av antalet passagerare som hinner med på tåget?

7.6 Momentgenererande funktion

Definition: Den *momentgenererande funktionen* $M(t)$ av en stokastisk variabel X definieras

$$M(t) = \mathbf{E}[e^{tX}] = \begin{cases} \sum\limits_x e^{tx} p(x) & \text{om } X \text{ diskret} \\ \int_{-\infty}^{\infty} e^{tx} f(x)\, dx & \text{om } X \text{ kontinuerlig} \end{cases}$$

för sådan $t \in \mathbf{R}$ att summan/integralen konvergerar.

Alla moment av X ($\mathbf{E}[X]$, $\mathbf{E}[X^2]$, $\mathbf{E}[X^3]$,...) kan beräknas genom att använda den momentgenerande funktionen $M(t)$.

Obs! Momentgenererande funktionen bestämmer fördelningen unikt.

Ex.6e: Momentgenererande funktionen av en stokastisk variabel X är $M(t) = exp(3(e^t - 1))$. Vad är $P(X = 0)$?

Låt X och Y vara oberoende stokastiska variabler med momentgenererande funktioner $M_X(t)$ och $M_Y(t)$. Då är

$$M_{X+Y}(t) = M_X(t)M_Y(t).$$

7.6.1 Den simultana momentgenererande funktionen

Låt X_1, X_2,..., X_n vara stokastiska variabler. Då är den *simultana momentgenererande funktionen* av X_1, X_2,..., X_n

$$M(t_1, t_2, ..., t_n) = \mathbf{E}[e^{t_1 X_1 + t_2 X_2 + ... + t_n X_n}]$$

för $(t_1, t_2, ..., t_n) \in \mathbf{R}^n$.

De individuella momentgenererande funktionerna av X_1, X_2,... och X_n är

$$M_{X_i}(t) = \mathbf{E}[e^{tX_i}] = M(0, ..., 0, t, 0, ..., 0).$$

Den simultana momentgenererande funktionen bestämmer den simultana fördelningen av X_1, X_2,..., X_n unikt.

De stokastiska variablerna X_1, X_2,... och X_n är oberoende om och endast om

$$M(t_1, t_2, ..., t_n) = M_{X_1}(t_1) M_{X_2}(t_2) \cdot ... \cdot M_{X_n}(t_n).$$

Ex.6m: Anta att antalet personer som kommer till ett postkontor är $Poisson(\lambda)$-fördelat och att en av personerna som kommer räknas med sannolikhet p oberoende av de andra. Visa att antalet personer som räknas och antalet personer som inte räknas är oberoende Poisson-fördelade variabler med resp. parametrar λp och $\lambda p(1 - p)$.