Kapitel 4

Stokastiska variabler

Ett utfall av ett slumpmässigt försök är ofta sådant som inte direkt kan mätas. T.ex. försöket “Kast med ett symmetriskt mynt” har utfallsrummet {krona, klave}. För att kvantitativt analysera försök av denna typ bör utfallsrummet avbildas t.ex. på reella axeln. Sådana avbildningar kallas stokastiska variabler.

Definition 4.1 En stokastisk variabel är en reellvärd funktion med ett utfallsrum som definitionsmängd.

Således är, trots benämningen, en stokastisk variabel i själva verket en funktion. Man brukar ofta beteckna dem med grekiska bokstäver, t.ex. ξ (ksi), η (eta), ζ (zeta), τ (tau),... Vi betecknar med Ω_ξ värdemängden för en stokastisk variabel ξ vars definitionsmängd är Ω.

Exempel 4.2

a) Låt Ω vara mängden av alla människor i världen. I detta utfallsrum kan vi t.ex. betrakta följande stokastiska variabler

$\omega \curvearrowright A(\omega) =$ människans ålder
$\omega \curvearrowright V(\omega) =$ människans vikt
$\omega \curvearrowright L(\omega) =$ människans längd

Om måttenheten för variabeln A är ett år är värdemängden av A en delmängd av de hela talen; A säges vara en heltalsvärd stokastisk variabel. Däremot är den lämpligaste värdemängden för V (och L) $\mathbb{R}^+ = \{x : x > 0\}$.

b) Låt Ω vara mänden av familjer med tre barn, samt ξ och η stokastiska variabler som anger antalet pojkar resp. flickor i familjen. Vi har då $\Omega_\xi = \Omega_\eta = \{0,\ 1,\ 2,\ 3\}$.

c) Betrakta försöket “6 kast med ett mynt” och låt ξ vara antalet krona. Då är $\Omega_\xi = \{0,\ 1,\ 2,\ 3,\ 4,\ 5,\ 6\}$.

I fortsättningen betraktas två fall. I det första antas att

Utfallsrummet är ändligt (eller högst numrerbart)

Låt $\Omega = \{\omega_1, ..., \omega_n\}$ och ξ vara en stokastisk variabel. Eftersom ξ är en funktion så är även Ω_ξ ändligt och $\mathrm{n}(\Omega_\xi) \leq \mathrm{n}(\Omega) = n$. Sätt $\Omega_\xi = \{x_1, ..., x_n\}$ $(k \leq n)$ och betrakta mängden $\mathrm{A}_i = \{\omega : \xi(\omega) = x_i\}$ (detta skrivs ofta kortare $\{\xi = x_i\}$). Enligt definitionen är A_i, $i = 1, ..., k$, en händelse och dess sannolikhet kan bestämmas:

$$\begin{aligned} \mathrm{P}(\mathrm{A}_i) = \mathrm{P}\big(\{\omega : \xi(\omega) = x_i\}\big) &= \sum_{\omega_u \in \mathrm{A}_i} \mathrm{P}\big(\{\omega_u\}\big) \\ &:= f(x_i) \end{aligned}$$

Definition 4.3 Talen $f(x_i), i = 1, ..., k$, bestämmer en funktion f som kallas **frekensfunktionen** för ξ.

Sats 4.4 En frekvensfunktion f uppfyller

(i) $f(x_i) \geq 0,\ i = 1, ...$

(ii) $\sum_{i=1}^{k} f(x_i) = 1$.

Bevis Eftersom $f(x_i) = \mathrm{P}(\mathrm{A}_i)$ och $\Omega = \bigcup_{i=1}^{k} \mathrm{A}_i,\ \mathrm{A}_i \cap \mathrm{A}_j = \emptyset,\ \ i \neq j$, följer påståendet ur Definition 3.3. □

Definition 4.5 Låt ξ vara en stokastisk variabel. Funktionen $F_\xi(x) = \mathrm{P}(\xi \leq x),\ x \in \mathbb{R}$, kallas **fördelningsfunktionen för** ξ.

Exempel 4.6 Låt ξ vara summan av poängtalen vid kast med två symmetriska tärningar. Ur figuren i Exempel 3.6 framgår att ξ har följande frekvensfunktion

Ω_ξ	2	3	4	5	6	7	8	9	10	11	12
$f_\xi(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

och fördelningsfunktionen

$$F_\xi(x) = \begin{array}{cc} 0 & x < 2 \\ \frac{1}{36} & 2 \le x < 3 \\ \frac{3}{36} & 3 \le x < 4 \\ \frac{6}{36} & 4 \le x < 5 \\ \frac{10}{36} & 5 \le x < 6 \\ \frac{15}{36} & 6 \le x < 7 \\ \frac{21}{36} & 7 \le x < 8 \\ \frac{26}{36} & 8 \le x < 9 \\ \frac{30}{36} & 9 \le x < 10 \\ \frac{33}{36} & 10 \le x < 11 \\ \frac{35}{36} & 11 \le x < 12 \\ 1 & 12 \le x \end{array}$$

Rita in dessa funktioner i ett koordinatsystem!

Låt φ vara en funktion från $\mathbb{R}$ till $\mathbb{R}$ och ξ en stokastisk variabel. Det är klart att även $\eta = \varphi(\xi)$ är en stokastisk variabel. Vi skall bestämma frekvensfunktionen för η. Låt $\Omega = \{\omega_1, ..., \omega_n\}$, $\Omega_\xi = \{x_1, ..., x_k\}$ och $\Omega_\eta = \{y_1, ..., y_m\}$. Då gäller att $\mathrm{n}(\Omega_\eta) \le \mathrm{n}(\Omega_\xi) \le \mathrm{n}(\Omega)$ $(m \le k \le n)$. Sätt $C_i = \{x : \varphi(x) = y_i\}$. Vi har

$$\{\omega_i : \eta(\omega_i) = y_j\} = \bigcup_{x_u \in C_j} = \{\omega_i : \xi(\omega_i) = x_u\}$$

och följaktligen,

$$\begin{aligned} \mathrm{P}(\eta = y_i) &= \mathrm{P}\Big(\bigcup_{x_u \in C_j} = \{\omega_i : \xi(\omega_i) = x_u\}\Big) \\ &= \sum_{x_u \in C_j} \mathrm{P}\{\omega_i : \xi(\omega_i) = x_u\} \\ &= \sum_{x_u \in C_j} f_\xi(x_u) := f_\eta(y_j). \end{aligned}$$

Funktioner som ofta förekommer i detta sammanhang är $|\xi|$, ξ^p (p är ett positivt heltal).

Betrakta en stokastisk variabel ξ och dess frekvensfunktion f. Frekvensfunktionen innehåller all information om ξ. Men för att bättre förstå beteenden av ξ karakteriserar man fördelningen av ξ genom att bestämma värden av vissa funktionaler (eller karakteristikor)(funktional = en funktion vars definitionsmängd och värdemängd är en funktionsklass (t.ex. frekvensfunktioner) resp. $\mathbb{R}$ (= de reella talen)). De viktigaste karakteristikorna är väntevärdet och variansen.

Vi skall först definiera väntevärdet för en stokastisk variabel ξ; detta betecknas med $\mathrm{E}(\xi)$.

Definition 4.7 Låt ξ vara en stokastisk variabel med frekvensfunktionen $f(x),\ x \in \Omega_\xi = \{x_1, ..., x_k\}$. Väntevärdet för ξ är då talet

$$\mathrm{E}(\xi) = \sum_{i=1}^{k} x_i\ f(x_i) = \sum_{x \in \Omega_\xi} x\ f(x).$$

Väntevärdet är m.a.o. ett vägt medeltal av de olika värdena på ξ.

Anmärkning 4.8 Då Ω_ξ är numrerbart, men inte ändligt, definieras väntevärdet som ovan, dvs $\mathrm{E}(\xi) = \sum_{x \in \Omega_\xi} x\ f(x)$. I detta fall säges väntevärdet existera om summan $\sum_{x \in \Omega_\xi} |x|\ f(x)$ är ändlig.

Exempel 4.9 Låt ξ vara likformigt fördelad över $\Omega_\xi = \{1, 2, ..., n\}$, dvs $\mathrm{P}(\xi = i) = \frac{1}{n}$. Då är

$$\mathrm{E}(\xi) = \sum_{i=1}^{n}\ i \cdot \frac{1}{n} = \frac{1}{n} \cdot n \cdot \frac{1+n}{2} = \frac{1+n}{2}.$$

I detta exempel sammanfaller således det aritmetiska medeltalet av elementen i Ω_ξ och väntevärdet av ξ.

Exempel 4.10 Låt ξ vara antalet erhållna krona vid tre kast med ett symmetrisk mynt. ξ har då följande frekvensfunktion $\big(\Omega_\xi = \{0, 1, 2, 3\}\big)$

x	0	1	2	3
$f(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Väntevärdet blir

$$\begin{aligned}\mathrm{E}(\xi) = \sum_{i=1}^{4} x_i\ f(x_i) &= 0 \cdot f(0) + 1 \cdot f(1) + 2\ f(2) + 3\ f(3) \\ &= 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8}\ \ =\ \ \frac{3}{2}.\end{aligned}$$

Sats 4.11 Väntevärdet för en sammansatt stokastisk variabel $\varphi(\xi)$ är

$$\mathrm{E}\big(\varphi(\xi)\big) = \sum_{i=1}^{k} \varphi(x_i)\ f(x_i) = \sum_{x \in \Omega_\xi} \varphi(x)\ f(x).$$

(Om Ω_ξ inte är oändligt bör summan vara absolutkonvergent, dvs $\sum_{x \in \Omega_\xi} |\varphi(x)|\ f(x) < \infty$. Se Anmärkning 4.8.)

Bevis Låt $\Omega = \{\omega_1, ..., \omega_n\}$, $\eta(\omega_i) = \varphi\big(\xi(\omega_i)\big)$, $\Omega_\xi = \{x_1, ..., x_k\}$ och $\Omega_\eta = \{y_1, ..., y_m\}$ $(m \leq k \leq n)$. Sätt $C_j = \{x_i : \varphi(x_i) = y_j\}$ Då gäller

$$\{\omega_i : \eta(\omega_i) = y_j\} = \bigcup_{x_u \in C_j} = \{\omega_i : \xi(\omega_i) = x_u\}.$$

Låt f_η och f_ξ vara frekvensfunktionerna för η resp. ξ. Vi har

$$\begin{aligned}
\mathrm{E}(\eta) = \mathrm{E}\big(\varphi(\xi)\big) &= \sum_{i=1}^{m} y_i f_\eta(y_i) \\
&= \sum_{i=1}^{m} y_i \mathrm{P}(\eta = y_i) \\
&= \sum_{i=1}^{m} y_i \sum_{x_u \in C_i} \mathrm{P}(\xi = x_u) \\
&= \sum_{i=1}^{m} \sum_{x_u \in C_i} y_i \mathrm{P}(\xi = x_u) \\
&= \sum_{i=1}^{m} \sum_{x_u \in C_i} \varphi(x_u) \mathrm{P}(\xi = x_u) \\
&= \sum_{u=1}^{k} \varphi(x_u) \mathrm{P}(\xi = x_u) \\
&= \sum_{u=1}^{k} \varphi(x_u) f_\xi(x_u).
\end{aligned}$$

$\square$

Exempel 4.12 Låt $\varphi(\xi) = a\xi + b$. Då fås

$$\begin{aligned}
\mathrm{E}\big(\varphi(\xi)\big) \;=\; \mathrm{E}(a\xi + b) \;&=\; \sum_{u=1}^{k} (a x_u + b) f_\xi(x_u) \\
&=\; a \sum_{u=1}^{k} x_u f_\xi(x_u) \;+\; b \sum_{u=1}^{k} f_\xi(x_u) \;=\; a\mathrm{E}(\xi) + b.
\end{aligned}$$

Definition 4.13 Låt ξ vara en stokastisk variabel med väntevärdet $\mathrm{E}(\xi)$. Variansen för ξ är talet

$$\mathrm{V}(\xi) = \mathrm{E}\Big(\big(\xi - \mathrm{E}(\xi)\big)^2\Big).$$

Den positiva kvadratroten av variansen kallas standardavvikelsen för ξ och betecknas $\mathrm{D}(\xi)$.

Variansen mäter variationen för ξ kring väntevärdet, dvs den är ett slags spridningsmått. Enligt Sats 4.11 gäller att

$$\mathrm{V}(\xi) = \sum_{i=1}^{k} (x_i - \mu)^2 f(x_i), \;\; \mathrm{E}(\xi) = \mu.$$

Sats 4.14

a) $\mathrm{V}(\xi) = \mathrm{E}(\xi^2) - \big(\mathrm{E}(\xi)\big)^2$.

b) $\mathrm{V}(c) = 0, \;\; c$ är en konstant.

c) $\mathrm{V}(a\xi + b) = a^2 \mathrm{V}(\xi), \;\; a, b$ konstanter.

Bevis

a) Sätt $\mathrm{E}(\xi) = \mu$; vi har

$$\begin{aligned} \mathrm{V}(\xi) &= \mathrm{E}\big((\xi - \mu)^2\big) = \sum_{i=1}^{k} (x_i - \mu)^2 f(x_i) \\ &= \sum_{i=1}^{k} (x_i^2 + \mu^2 - 2\mu x_i) f(x_i) \;=\; \sum_{i=1}^{k} x_i^2 f(x_i) + \mu^2 \sum_{i=1}^{k} f(x_i) - 2\mu \sum_{i=1}^{k} x_i f(x_i) \\ &= \sum_{i=1}^{k} x_i^2 f(x_i) + \mu^2 - 2\mu \cdot \mu \;=\; \mathrm{E}(\xi^2) - \big(\mathrm{E}(\xi)\big)^2. \end{aligned}$$

Fallen b) och c) lämnas till läsaren som övningsuppgifter.

$\square$

Exempel 4.15 Låt ξ vara **likformigt fördelad** över $\Omega = \{1, 2, ..., n\}$. Då är $\mathrm{E}(\xi) = \frac{1+n}{2}$ (enligt Exempel 4.9). Vi skall bestämma variansen för ξ:

$$\mathrm{E}(\xi^2) = \sum_{i=1}^{n} \; i^2 \frac{1}{n} \;=\; \frac{1}{n} \sum_{i=1}^{n} \; i^2.$$

För att beräkna summan $\sum_{i=1}^{n} \; i^2$ observera att $i^2 \equiv \frac{(i+1)^3 - i^3}{3} - i - \frac{1}{3}$ (verifiera detta!). Då fås (utför räkningen!)

$$\sum_{i=1}^{n} \; i^2 = \frac{n(n+1)(2n+1)}{6}.$$

Följaktligen

$$\begin{aligned} \mathrm{V}(\xi) = \mathrm{E}(\xi^2) - \big(\mathrm{E}(\xi)\big)^2 &= \frac{n(n+1)(2n+1)}{6} - \Big(\frac{n+1}{2}\Big)^2 \\ &= \frac{n^2-1}{12}. \end{aligned}$$

Exempel 4.16 (Binomialfördelning)

Låt n vara ett positivt heltal och $0 < p < 1$. Sätt

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k}, \;\; k \in \{0, 1, 2, ..., n\}.$$

Vi skall visa att f är en frekvensfunktion:

(i) Det är klart att $f(k) \geq 0, \;\; k = 0, 1, 2, ..., n.$

(ii) Vidare är

$$\begin{aligned} \sum_{k=0}^{n} f(k) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \big(p + (1-p)\big)^n = 1, \end{aligned}$$

enligt binomialteoremet.

Talen n och p kallas fördelningens **parametrar**. Att en stokastisk variabel är binomialfördelad med parametrarna n och p betecknas $\boldsymbol{\xi \sim} \mathbf{Bin}(\boldsymbol{n, p})$.

Vi bestämmer väntevärdet för ξ:

$$\begin{aligned} \mathrm{E}(\xi) = \sum_{k=0}^{n} k \; f(k) &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^{n} k \cdot \frac{n!}{k!(n-k)!} \; p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^{n} \frac{(n-1!)}{(k-1)!(n-k)!} \; p^{k-1} (1-p)^{n-k}. \end{aligned}$$

Sätt $m = k - 1$, då fås

$$\begin{aligned} \mathbf{E}(\boldsymbol{\xi}) = \; & np \sum_{m=0}^{n-1} \frac{n-1}{m!(n-1-m)!} \; p^m (1-p)^{n-1-m} \\ = \; & np \sum_{m=0}^{n-1} \binom{n-1}{m} \; p^m (1-p)^{n-1-m} \; = \; \boldsymbol{np}, \end{aligned}$$

enligt binomialteoremet.

För att bestämma variansen behövs $\mathrm{E}(\xi^2)$. Vi utgår från

$$\mathrm{E}\big(\xi(\xi-1)\big) \;=\; \sum_{k=0}^{n} k(k-1)\ f(k).$$

På samma sätt som ovan fås att $\mathrm{E}\big(\xi(\xi-1)\big) = n(n-1)p^2$. Således

$$\begin{aligned}\mathbf{V}(\boldsymbol{\xi}) &= \mathrm{E}(\xi^2) - \mathrm{E}(\xi)^2 = \mathrm{E}\big(\xi(\xi-1)\big) + \mathrm{E}(\xi) - \mathrm{E}(\xi)^2 \\ &= n(n-1)p^2 + np - n^2p^2 \;=\; \boldsymbol{np(1-p)}.\end{aligned}$$

Exempel 4.17 (Hypergeometrisk fördelning)

Låt N, N_1, N_2 och n vara positiva heltal sådana att $N = N_1 + N_2$ och $n \le \min(N_1, N_2)$. Sätt

$$f(k) = \frac{\binom{N_1}{k}\binom{N_2}{n-k}}{\binom{N}{n}},\ \ k \in \{0, 1, ..., n\}.$$

Vi skall visa att f är en frekvensfunktion:

(i) Det är klart att $f(k) \ge 0,\ \ k = 0, 1, ..., n$.

(ii) Vidare är $\sum_{k=0}^{n} f(k) = 1$, ty

$$\sum_{k=0}^{n} \binom{N_1}{k}\binom{N_2}{n-k} = \binom{N_1+N_2}{n} = \binom{N}{n},$$

(se övningsuppgift 9, kapitel 2).

Att en stokastisk variabel ξ är hypergeometriskt fördelad med parametrarna N, N_1, n betecknas $\boldsymbol{\xi \sim}$ **Hyp**$\boldsymbol{(N, N_1, n)}$. Vi bestämmer väntevärdet för ξ:

$$\begin{aligned}\mathrm{E}(\xi) \;&=\; \sum_{k=0}^{n} k\ f(k) \;=\; \sum_{k=0}^{n} k\frac{\binom{N_1}{k}\binom{N_2}{n-k}}{\binom{N}{n}} \\ &=\; \sum_{k=0}^{n} k \cdot \frac{\frac{N_1!}{k!(N_1-k)!}\ \frac{N_2!}{(n-k)!(N_2-n+k)!}}{\frac{N!}{n!(N-n)!}} \\ &=\; n \cdot \frac{N_1}{N} \cdot \sum_{k=1}^{n} \frac{\frac{(N_1-1)!}{(k-1)!(N_1-k)!}\ \frac{N_2!}{(n-k)!(N_2-n+k)!}}{\frac{(N-1)!}{(n-1)!(N-n)!}} \\ &=\; n \cdot \frac{N_1}{N} \cdot \sum_{k=1}^{n} \frac{\binom{N_1-1}{k-1}\binom{N_2}{n-k}}{\binom{N-1}{n-1}}.\end{aligned}$$

Sätt $m = k-1$; då fås

$$= \; n \cdot \frac{N_1}{N} \cdot \sum_{m=0}^{n-1} \frac{\binom{N_1-1}{m}\binom{N_2}{n-1-m}}{\binom{N-1}{n-1}} \; = \; n \, \frac{N_1}{N}.$$

Vid beräkningen av variansen utgår man från uttrycket

$$\mathrm{E}\big(\xi(\xi-1)\big) \; = \; \sum_{k=0}^{n} k(k-1) \, \frac{\binom{N_1}{k}\binom{N_2}{n-k}}{\binom{N}{n}} \; = ... =$$

$$= \; n(n-1) \, \frac{N_1(N_1-1)}{N(N-1)}.$$

Följaktligen

$$\begin{aligned} \mathrm{V}(\xi) \; &= \; \mathrm{E}(\xi^2) - \big(\mathrm{E}(\xi)\big)^2 \; = \; \mathrm{E}\big(\xi(\xi-1)\big) + \mathrm{E}(\xi) - \big(\mathrm{E}(\xi)\big)^2 \\ &= \; n \, \frac{N_1}{N} \, \frac{N-N_1}{N} \cdot \frac{N-n}{N-1}. \end{aligned}$$

Sätt $\frac{N_1}{N} = p, \; \frac{N-N_1}{N} = 1-p =: q$. Då fås $\boldsymbol{\mathrm{E}(\xi) = np}, \; \boldsymbol{\mathrm{V}(\xi) = npq \, \frac{N-n}{N-1}}$.

Exempel 4.18 (Poissonfördelning)

Låt $\lambda > 0$ och sätt

$$f(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \; k \in \{0, 1..., \}.$$

Vi visar att f är en frekvensfunktion:

(i) Det är klart att $f(k) \geq 0, \; k = 0, 1, ...$

(ii) Vidare är $\sum_{k=0}^{\infty} f(k) = 1$, ty

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \, e^{-\lambda} \; = \; e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \; = \; e^{-\lambda} \cdot e^{\lambda} = 1.$$

(Vi har använt MacLaurinutvecklingen för e^x; $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + ...$). Att en stokastisk variabel ξ är Poissonfördelad med parametern λ betecknas $\boldsymbol{\xi \sim \mathrm{Po}(\lambda)}$.

Vi bestämmer väntevärdet:

$$\begin{aligned} \boldsymbol{\mathrm{E}(\xi)} \; &= \; \sum_{k=0}^{\infty} k \; f(k) \; = \; \sum_{k=0}^{\infty} k \, \frac{\lambda^k}{k!} \, e^{-\lambda} \\ &= \; \sum_{k=0}^{\infty} \frac{\lambda^k}{(k-1)!} \, e^{-\lambda} \; = \; \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \; = \; \boldsymbol{\lambda}. \end{aligned}$$

Variansen är $\mathbf{V(\xi) = \lambda}$. Således sammanfaller väntevärdet och variansen för en Poissonfördelad stokastisk variabel.

Exempel 4.19 (Geometrisk fördelning)

Låt $0 < p < 1$ och sätt

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

Det gäller att

(i) $f(k) \geq 0,\ k = 1, 2, ...,$ och

(ii) $\sum_{k=1}^{\infty} f(k) = 1$, ty

$$\begin{aligned}\sum_{k=1}^{\infty} f(k) &= \sum_{k=1}^{\infty} (1-p)^{k-1}p = p\sum_{k=1}^{\infty} (1-p)^{k-1} \\ &= p \cdot \frac{1}{1-(1-p)} = 1.\end{aligned}$$

Vi beräknar väntevärdet.

$$\mathrm{E}(\xi) = \sum_{k=1}^{\infty} k\ f(k) = \sum_{k=1}^{\infty} k(1-p)^{k-1}p = p\sum_{k=1}^{\infty} k(1-p)^{k-1}.$$

För att beräkna summan $\sum_{k=1}^{\infty} k(1-p)^{k-1}$ sätt $h(x) = \sum_{k=1}^{\infty}(1-x)^k,\ 0 < x < 1.$

Vi kan derivera[1] $h(x)$ genom att derivera under summationstecknet; då fås

$$\frac{d}{dx}h(x) = -\sum_{k=1}^{\infty} k(1-x)^{k-1}. \tag{1}$$

Å andra sidan

$$\begin{aligned}h(x) &= \sum_{k=1}^{\infty} (1-x)^k = (1-x)\sum_{k=1}^{\infty} (1-x)^{k-1}. \\ &= (1-x)\ \frac{1}{1-(1-x)} = \frac{1-x}{x}\ ;\end{aligned}$$

följaktligen

$$\frac{d}{dx}h(x) = -\frac{1}{x^2}. \tag{2}$$

[1]Detta gäller inte i allmänhet; här gäller det eftersom h är en potensserie med konvergensradien 1 (Se t.ex. Sjöberg: Analytiska funktioner avsnitt 8.5)

Ur (1) och (2) följer

$$-\sum_{k=1}^{\infty} k(1-x)^{k-1} \;=\; -\frac{1}{x^2}.$$

Härav följer att

$$\mathbf{E}(\boldsymbol{\xi}) = p\sum_{k=1}^{\infty} k(1-p)^{k-1} \;=\; p\cdot\frac{1}{p^2} \;=\; \frac{\mathbf{1}}{\boldsymbol{p}}.$$

Variansen beräknas på ett analogt sätt:

$$\mathbf{V}(\boldsymbol{\xi}) = \frac{\mathbf{1}-\boldsymbol{p}}{\boldsymbol{p}^{\mathbf{2}}}\;.$$

Vi betraktar nu det andra fallet:

Stokastiska variabler som har en täthet

Definition 4.20 Låt ξ vara en stokastisk variabel vars värdemängd är ett intervall $\Omega_\xi = (a,b)$. Man säger att ξ har en **täthet $\boldsymbol{f}$** om

$$\mathrm{P}(\xi \in \mathrm{A}) = \int_{\mathrm{A}} f(x)dx,$$

där $\mathrm{A} \subset \Omega_\xi$ är en union av ett ändligt (eller högst numrerbart) antal öppna intervall, dvs $\mathrm{A} = \bigcup_{i=1}^{n} I_i,\ I_i \in \Omega_\xi = (a,b)$ (eller $\mathrm{A} = \bigcup_{i=1}^{\infty} I_i$).

Funktionen f kallas även frekvensfunktionen för ξ.

Ur Definition 4.19 följer att en täthet (frekvensfunktion) f uppfyller

(i) $f(x) \geq 0,\ x \in \Omega_\xi = (a,b)$

(ii) $\int_a^b f(x)dx = 1.$

Fördelningsfunktionen för ξ definieras på samma sätt som i det ändliga fallet, dvs

$$F(x) = \mathrm{P}(\xi \leq x),\ x \in \mathbb{R}.$$

Följaktligen gäller att

$$F(x) = \begin{cases} 0, & x \leq a, \\ \int_a^x f(x)dx, & a < x \leq b, \\ 1, & x > b. \end{cases}$$

Vidare är F kontinuerlig, icke-avtagande och $\frac{d}{dx}F(x) = f(x)$ i de punkter där F är deriverbar. Om $a = -\infty$ och $b = +\infty$ gäller det att $\lim_{x\to-\infty} F(x) = 0$ och $\lim_{x\to+\infty} F(x) = 1$. Ur definitionen följer också

$$\begin{aligned} &\mathrm{P}(x_1 \le \xi \le x_2) = \mathrm{P}(x_1 < \xi \le x_2) = \mathrm{P}(x_1 \le \xi < x_2) \\ &= \mathrm{P}(x_1 < \xi < x_2) = \int_{x_1}^{x_2} f(x)dx = F(x_2) - F(x_1). \end{aligned}$$

Exempel 4.21 Betrakta funktionen $f : x \curvearrowright cx,\ c > 0$. Eftersom $f \ge 0$ på $[0, \infty)$ kan vi bestämma konstanten c så att f blir en frekvensfunktion på $[0, 1]$.

Vi har

$$\int_0^1 f(x)dx = 1 \Longleftrightarrow \int_0^1 cx\ dx = 1$$

$$\Longleftrightarrow \left[\frac{c}{2}\ x^2\right]_0^1 = 1 \Longleftrightarrow \frac{c}{2} = 1 \Longleftrightarrow c = 2.$$

Fördelningsfunktionen är $F(x) = \int_0^x 2t\ dt = [t^2]_0^x = x^2,\ x \in [0, 1]$ och $F(x) = 0$ då $x < 0$ samt $F(x) = 1$ då $x > 1$.

Definition 4.22 Väntevärdet för ξ är

$$\mathrm{E}(\xi) = \int_a^b x f(x)dx$$

Då $a = -\infty$ och/eller $b = +\infty$ är väntevärdet en generaliserad integral och kan vara divergent. Om detta är fallet säges den stokastiska variabeln sakna väntevärde.

Utan bevis ger vi följande sats:

Sats 4.23 Låt φ vara en funktion från $\mathbb{R}$ till $\mathbb{R}$. Väntevärdet för den sammansatta variabeln $\varphi(\xi)$ är (förutsatt att det existerar)

$$\mathrm{E}\big(\varphi(\xi)\big) = \int_a^b \varphi(x)f(x)dx.$$

Definition 4.24 Låt $\mathrm{E}(\xi) = \mu$. Variansen för ξ är

$$\mathrm{V}(\xi) = \mathrm{E}\big((\xi-\mu)^2\big) = \int_a^b (x-\mu)^2 f(x)dx.$$

Sats 4.25 a) $\mathrm{V}(\xi) = \mathrm{E}(\xi^2) - \big(\mathrm{E}(\xi)\big)^2$

b) $\mathrm{V}(c) = 0,\ c$ är en konstant

c) $\mathrm{V}(a\xi + b) = a^2\mathrm{V}(\xi)$.

Bevis Jfr Sats 4.14. $\square$

Exempel 4.26 (Likformig fördelning)

En stokastisk variabel ξ är likformigt fördelad över intervallet (a, b) om den har tätheten

$$f(x) = \frac{1}{b-a}, \; x \in (a, b).$$

Funktionen f uppfyller

(i) $f \geq 0$ och

(ii) $\int_a^b f(x)dx = \int_a^b \frac{1}{b-a} \; dx = 1$.

$$\begin{aligned}
\mathrm{E}(\xi) &= \int_a^b x \; f(x)dx \;=\; \int_a^b x \; \frac{1}{b-a} \; dx \;=\; \frac{b+a}{2}. \\
\mathrm{E}(\xi^2) &= \int_a^b x^2 \; f(x)dx \;= ... =\; \frac{a^2+ab+b^2}{3}. \\
\mathrm{V}(\xi) &= \mathrm{E}(\xi^2) - \big(\mathrm{E}(\xi)\big)^2 \;= ... =\; \frac{(b-a)^2}{12}.
\end{aligned}$$

Exempel 4.27 (Exponentialfördelning)

En stokastisk variabel ξ är exponentialfördelad med parametern $\lambda > 0$ om den har tätheten

$$f(x) \;=\; \lambda e^{-\lambda x}, \; x \geq 0.$$

Då är

$$\begin{aligned}
\int_0^\infty f(x)dx &= \int_0^\infty \lambda e^{-\lambda x} dx \;= ... =\; 1, \\
\mathrm{E}(\xi) &= \int_0^\infty x \; f(x)dx \;= ... =\; \frac{1}{\lambda}, \\
\mathrm{E}(\xi^2) &= \int_0^\infty x^2 \; f(x)dx \;= ... =\; \frac{2}{\lambda^2}, \\
\mathrm{V}(\xi) &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} \;=\; \frac{1}{\lambda^2}.
\end{aligned}$$

Exempel 4.28 (Normalfördelning)

En stokastisk variabel ξ är normalfördelad med parametrarna μ och $\sigma > 0$ om den har tätheten (se fig. 4.1)

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\; e^{-\frac{(x-\mu)^2}{2\sigma^2}},\; x \in \mathbb{R}.$$

Att $\int_{-\infty}^{\infty} f(x)dx = 1$ kan bevisas genom att beräkna dubbelintegralen $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y)f(x)dydx$ (se t.ex. Björup & Edén: Analys i en och flera dimensioner s. 213 och övning 441).

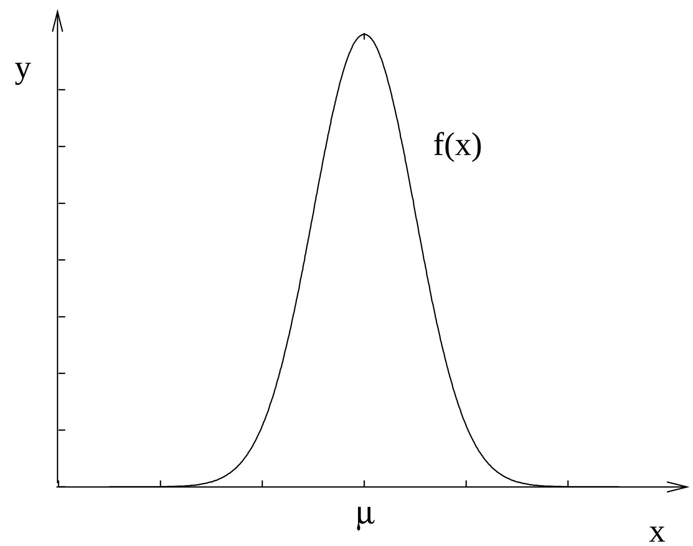

fig. 4.1

Observera att f är symmetrisk kring punkten μ. Det gäller

$$\mathrm{E}(\xi) = \int_{-\infty}^{\infty} x\; f(x)dx\; = ... = \;\mu$$
$$\mathrm{V}(\xi) = \int_{-\infty}^{\infty} x^2\; f(x)dx - \mu^2\; = ... = \;\sigma^2.$$

(Utför integrationerna!)

Att ξ är normalfördelad med parametrarna μ och σ betecknas $\boldsymbol{\xi \sim \mathrm{N}(\mu, \sigma)}$.

Betrakta variabeln $\eta = a\xi + b$, där a och b är konstanter. Då gäller $\mathrm{E}(\eta) = a\mathrm{E}(\xi) + b = a\mu + b$, $\mathrm{V}(\eta) = a^2\mathrm{V}(\xi) = a^2\sigma^2$ och vidare (utan bevis) $\boldsymbol{\eta = a\xi + b}$ **är normalfördelad med parametrarna** $\boldsymbol{a\mu + b,\; |a|\sigma}$ $\big(\boldsymbol{\eta \sim \mathrm{N}(a\mu + b, |a|\sigma)}\big)$.

Följaktligen gäller att **den standardiserade variabeln** $\boldsymbol{\zeta = \frac{\xi-\mu}{\sigma}}$ **är normalfördelad med parametrarna 0 och 1** $\big(\boldsymbol{\zeta \sim \mathrm{N}(0, 1)}\big)$.

Fördelningsfunktionen för den standardiserade normalfördelningen (dvs $\mu = 0,\ \sigma = 1$) betecknas ϕ:

$$\phi(x) = \mathrm{P}(\zeta \le x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{y^2}{2}}\, dy.$$

Låt $\xi \sim \mathrm{N}(100, 10)$. Vi bestämmer

$$\begin{aligned}
\mathrm{P}(\xi \le 125) &= \mathrm{P}\Big(\frac{\xi - 100}{10} \le \frac{125 - 100}{10}\Big) \\
&= \mathrm{P}(\zeta \le 2,5) \;=\; \phi(2,5) \;\simeq\; 0,9938, \\
\mathrm{P}(\xi \ge 92) &= \mathrm{P}\Big(\zeta \ge \frac{92 - 100}{10}\Big) \;=\; \mathrm{P}(\zeta \ge -0,8) \\
&= 1 - \phi(-0,8) \;=\; \phi(0,8) \;\simeq\; 0,7881, \\
\mathrm{P}(85 \le \xi \le 112) &= \mathrm{P}(\zeta \le 1,2) - \mathrm{P}(\zeta \le -1,5) \\
&= \phi(1,2) - \phi(-1,5) \\
&= \phi(1,2) + \phi(1,5) - 1 \;\simeq\; 0,8181.
\end{aligned}$$

Övningsuppgifter

1. En bilhandlare har femton bilar i ett lager. Av dessa är fem felfria och de övriga har mindre felaktigheter. Man väljer på måfå fyra bilar i lagret. Låt ξ vara antalet därvid erhållna felfria bilar. Bestäm frekvensfunktionen för ξ.

2. En urna innehåller fem kulor markerade 1, 3, 3, 4, 5. Man tar på måfå med återläggning två kulor ur urnan. Bestäm frekvensfunktionen för skillnaden mellan den största och minsta erhållna poängen.

3. Betrakta samma urna som i föregående övning. Man tar på måfå två kulor ur urnan utan återläggning. Bestäm frekvensfunktionen för den erhållna poängsumman.

4. Den stokastiska variabeln ξ, med $\Omega_\xi = \{1, 2, 3, 4\}$, har frekvensfunktionen $f_\xi(x) = ax$. Bestäm a samt frekvensfunktionen för variabeln $\eta = (\xi - 3)^2$.

5. En stokastisk variabel ξ har frekvensfunktionen

$$f(x) = 0,2, \; x \in \{-2, -1, 0, 1, 2\}.$$

   Bestäm utfallsrum och frekvensfunktion för a) $|\xi|$ b) ξ^2.

6. I ett lotteri finns 100 lotter, av vilka 50 inte ger någon vinst, 30 ger vinsten 2 mk, 10 vinsten 10 mk, 8 vinsten 20 mk och 2 vinsten 50 mk. Låt ξ vara vinsten om man köper en lott. Bestäm väntevärdet $\mathrm{E}(\xi)$. Vad är ett rimligt pris på lotterna?

7. Bestäm a) väntevärdet b) variansen för den stokastiska variabeln som infördes i övning 2.

8. Bestäm a) väntevärdet b) variansen för den stokastiska variabeln som infördes i övning 5 a) och b).

9. För en stokastisk variabel ξ gäller $\mathrm{E}(\xi) = 7$ och $\mathrm{D}(\xi) = 5$. Beräkna väntevärdet och standardavvikelsen för variabeln $\eta = 10 - 2\xi$.

10. Låt ξ vara en stokastisk variabel med väntevärdet $\mathrm{E}(\xi)$ och standardavvikelsen $\mathrm{D}(\xi)$. Man bildar variabeln $\eta = \frac{\xi - \mathrm{E}(\xi)}{\mathrm{D}(\xi)}$. Visa att $\mathrm{E}(\eta) = 0$ och $\mathrm{D}(\eta) = 1$.

11. Beräkna variansen för en stokastisk variabel med a) binomial b) hypergeometrisk c) Poisson d) geometrisk fördelning.

12. **(Pascalfördelning)** Låt n vara ett positivt heltal och $0 < p < 1$. Bevisa att

$$f(k) = \binom{k-1}{n-1} p^n (1-p)^{k-n}, \; k = \{n, n+1, n+2, ...\}$$

   är en frekvensfunktion samt bestäm väntevärdet och variansen.

13. Visa att följande funktioner är frekvensfunktioner:

    a) $f(x) = (n+1)x^n, \; x \in (0,1), \; n > 0$

    b) $f(x) = \frac{2}{\pi\sqrt{1-x^2}}, \; x \in (0,1)$

14. Bestäm konstanten k så att $f(x) = k\; e^{-|x-\theta|}, \; x \in \mathbb{R}$, blir en frekvensfunktion.

15. Visa att $F(x) = \frac{2}{\pi} \arcsin \sqrt{x}, \; x \in (0,1)$ är en fördelningsfunktion. Bestäm frekvensfunktionen samt beräkna följande sannolikheter

    $\mathrm{P}(0 < \xi < 0,5), \; \mathrm{P}(0,25 < \xi < 0,5)$ och $\mathrm{P}(|\xi - 0,5| > 0,25)$

16. Bestäm väntevärdet och variansen för en stokastisk variabel med den frekvensfunktion som infördes i övning 13 a) och b).

17. **(Cauchyfördelning)** Bstäm konstanten c så att $f(x) = \frac{c}{1+x^2}, \; x \in \mathbb{R}$, blir en frekvensfunktion. Visa att väntevärdet ej existerar (och således inte heller variansen).

18. En stokastisk variabel ξ är normalfördelad med parametrarna 50, 10 $\big(\xi \sim \mathrm{N}(50,10)\big)$. Bestäm följande sannolikheter

    a) $\mathrm{P}(\xi \leq 65)$ b) $\mathrm{P}(\xi \leq 25)$ c) $\mathrm{P}(\xi \geq 35)$ d) $\mathrm{P}(\xi \geq 70)$
    e) $\mathrm{P}(40 < \xi \leq 60)$ f) $\mathrm{P}(|\xi - 50| < 20)$ g) $\mathrm{P}(|\xi - 50| \geq 15)$.

19. Mellan vilka symmetriska värden på μ faller 95 %, 99 % respektive 99,9 % av den standardiserade normalfördelningen?

20. För en normalfördelad stokastisk variabel ξ gäller att $\mathrm{P}(-\infty < \xi < 40) = 0,3085$ och att $\mathrm{P}(40 < \xi < 60) = 0,3830$. Bestäm väntevärde och standardavvikelse för variabeln ξ.