3. Diskreta fördelningar

3.1 Stokastiska variabler

Några exempel av stokastiska variabler:

Ex: Man kastar två tärningar och är intresserad av två händelser:
$A =$ första tärningen ger en trea och
$B =$ summan av de två är 8.
Man kan i stället för händelserna A och B titta på stokastiska variabler:
$X =$ antalet ögon på den första tärningen och
$Y =$ summan av de två.

Ex. 3.1.1: Antag att X är en stokastisk variabel (s.v.) som ger antalet barn med bruna ögon av föräldrar som har en blå-gen och en brun-gen vardera. Totalt har de två barn.

Ex: Livslängden av en glödlampa,X, kan antas vara en stokastisk variabel.

Ex: Man kastar en tärning. Låt X vara antal kast tills man får en sexa för första gången.

3.2 Diskreta fördelningar

En stokastisk variabel är **diskret** om den kan anta ändligt många (eller uppräkneligt oändligt många) olika värden.

Låt X vara en diskret stokastisk variabel. Funktionen

$$f(x) = P(X = x), \quad x \in \mathbf{R},$$

kallas **frekvensfunktionen** (probability mass function; density function) för X.

f kan vara en frekvensfunktion om och endast om

1) $f(x) \geq 0$

2) $\sum_{\text{alla } x} f(x) = 1$

Fördelningsfunktionen (cumulative distribution function) för X definieras

$$F(x) = P(X \leq x) = \sum_{i \leq x} P(X = i) = \sum_{i \leq x} f(i)$$

för alla $x \in \mathbf{R}$.

3.3 Väntevärdet och fördelningsparametrar

Frekvensfunktionen (och fördelningsfunktionen) säger allt om fördelningen (hur värdena av en stokastisk variabel ser ut). För att snabbt kunna säga någonting om fördelningen, kan man titta på så-kallade fördelningsparametrar: väntevärdet μ, variansen σ^2 och standardavvikelsen σ, som kan räknas genom att använda frekvensfunktionen.

Väntevärdet av en diskret stokastisk variabel definieras

$$\mathbf{E}[X] = \mu = \sum_{\text{alla } x} x\, P(X = x).$$

Väntevärdet är tyngd- eller medelpunkten av fördelningen. Den kallas ofta en lokalisering- eller lägeparameter. Väntevärdet är det teoretiska medelvärdet.

Några regler angående väntevärdet:

1) $\mathbf{E}[c] = c$, om c är konstant

2) $\mathbf{E}[cX] = c\mathbf{E}[X]$, c konstant

3) $\mathbf{E}[X + Y] = \mathbf{E}[X] + \mathbf{E}[Y]$

Ex.3.3.3: Man vill jämföra ett nytt och ett gammalt läkemedel som används för att skaffa konstant hjärtslagshastighet efter en mild hjärtinfart. (Se tabell 3.6, s. 55 i boken.)

Variansen: Låt X vara en s.v. med väntevärde μ. **Variansen** av X definieras

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

Den kan räknas på följande sätt

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

Standardavvikelsen av X är

$$\sigma = \sqrt{Var(X)} = \sqrt{\sigma^2}$$

Några regler angående variansen:

1) $Var(c) = 0$, om c är konstant

2) $Var(cX) = c^2 Var(X)$, c konstant

3) Om X och Y är oberoende då är $Var(X+Y) = Var(X) + Var(Y)$

3.5 Binomialfördelning

När kan man använda binomialfördelning som en model?

1) Experimentet består av n $(n \geq 1)$ oberoende försök

2) Det finns två möjliga utfall, "success" och "failure"

3) Sannolikheten för succé är samma på varje försök. Om succésannolikheten är p, då är sannolikheten för "failure" $q = 1 - p$.

4) Låt X vara antalet succéer bland n försök. Då är X en stokastisk variabel som har binomialfördelning. Man säger att X är binomialfördelad med parametrar n (antalet oberoende försök) och p (succésannolikheten), $X \sim Bin(n, p)$.

Frekvensfunktionen av X är

$$f(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, ..., n$$

Ex: Låt X vara antalet barn med bruna ögon av föräldrar som har en blå-gen och en brun-gen vardera och som har två barn.

Väntevärdet och **variansen** av $X \sim Bin(n, p)$ är

1) $\mathbf{E}[X] = np$

2) $Var(X) = np(1-p)$

Ex: Antag att en maskin fungerar med sannolikhet 0.95 och att man har 10 sådana (oberoende) maskiner. Vad är sannolikheten att högst en av de inte fungerar?

Några andra diskreta fördelningar (ingår inte i tentan)

> **Geometrisk fördelning:** Man utför en följd av oberoende försök med succésannolikhet p. En stokastisk variabel X är antalet försök tills en succé inträffar för första gången
>
> **Poissonfördelning** kan användas för att approximera binomialfördelningen.