© 2005 Eric Järpe
Högskolan i Halmstad

Diskreta variabler

Ofta vill man beräkna sannolikheten för att något inträffar ett visst *antal* gånger. Och ofta återkommer samma resonemang men med nya siffror.

Antag t.ex. att vi har ett lotteri med 1000 lotter varav 100 är vinster. Då följer

- sannolikheten att få vinst på 1 lott *likformig fördelning*,
- sannolikheten att få 3 vinster bland 10 dragna lotter *binomialfördelning*,
- antalet lotter som måste tas innan sannolikheten för att någon av dem är vinst blir $\geq \frac{1}{2}$ *geometrisk fördelning.*

Därför vill man ha generella regler för hur olika situationer ska behandlas (dvs hur olika sannolikheter ska beräknas).

Definition 1 *Innan experimentet utförts kallas det hypotetiska antalet i fråga* **stokastisk variabel** *(betecknas vanligen med stora bokstäver X, Y, Z etc., i Vännman dock med grekiska bokstäver ξ, η etc.). Efter att experimentet utförts kallas det observerade antalet* **observation** *(betecknas med motsvarande liten bokstav x, y, z etc.).*

Definition 2 *En stokastisk variabel som bara kan anta ett uppräkneligt (numrerbart) antal utfall kallas* **diskret**.

Olika stokastiska variabler är olika på så sätt att sannolikheten för att variabeln antar olika värden *fördelar* sig olika. Namnet på den huvudregel för hur en stokastisk variabel uppträder kallas **fördelning**. För varje fördelning finns en sannolikhetsfunktion som matematiskt explicit uttrycker hur sannolikheter ska beräknas.

Läs Exempel 3.3 om binomialfördelning och Exempel 3.4 om Poissonfördelning.

Definition 3 *Låt X vara en diskret stokastisk variabel som antar med värdemängd $\{x_1, x_2, \ldots, x_k, \ldots\}$ där $x_1 < x_2 < \cdots < x_k < \cdots$* **Sannolikhetsfunktionen** *p för X är definierad $p(x_k) = P(X = x_k)$.* **Fördelningsfunktionen** *F för X är definierad $F(x_k) = P(X \leq x_k) = \sum_{i=1}^{k} p(x_i)$. Var och en av dessa två beskriver fullständigt fördelningen för X.*

Vanliga diskreta fördelningar

(Diskret) likformig fördelning

Betecknas $X \in Likf(n)$.
$p(x) = \begin{cases} \frac{1}{n} & x = 1, 2, \ldots, n \\ 0 & \text{annars} \end{cases}$
Ger lika stor sannolikhet åt alla utfall. Den som gäller om man ej har någon information. (Ex. kasta tärning.)

Binomialfördelning

Betecknas $X \in Bin(n, p)$.
$p(x) = \binom{n}{p} p^x (1-p)^{n-x},\ x = 0, 1, \ldots, n$
X är antalet lyckade delförsök i ett experiment bestående av n oberoende upprepningar av ett delförsök. Varje delförsök lyckas med sannolikhet p. (Ex. lotteri: dra x vinster då n lotter köps.)

Poissonfördelning

Betecknas $X \in Poi(\lambda)$.
$p(x) = e^{-\lambda} \frac{\lambda^x}{x!},\ x = 0, 1, 2, \ldots$
X är antalet gånger som en återkommande händelse inträffar under ett visst tidsintervall: i genomsnitt λ gånger. (Ex. Bilar som passerar på en väg, en partikel som sönderfaller.)

Geometrisk fördelning

Betecknas $X \in Geo(p)$.
$p(x) = (1-p)^{x-1} p,\ x = 1, 2, 3, \ldots$
X är antalet försök som behövs innan en händelse med sannolikhet p inträffar. (Ex. antal gånger man måste spela för att få kåk i given då man spelar poker, antal lotter man måste ta för att få en vinstlott.)

(**Obs** Behöver ej läsa om hypergeometrisk fördelning.)

Läs Sats 3A och Exempel 3.6.

Väntevärde och varians

Den teoretiska tyngdpunkten av en fördelning kallas *väntevärde*. Den teoretiska spridningsgraden av en fördelning anges av *variansen*.

Definition 4 *Låt X vara en diskret stokastisk variabel med sannolikhetsfunktion $p(x)$.* **Väntevärdet** *för X är $\mu = E(X) = \sum_x xp(x)$.* **Variansen** *för X är $\sigma^2 = V(X) = \sum_x (x-\mu)^2 p(x)$.*

Räkneregler

- $E(a + bX) = a + bE(X)$
- $V(a + bX) = b^2V(X)$
- $E(X + Y) = E(X) + E(Y)$
- $X \perp Y \;\Rightarrow\; V(X + Y) = V(X) + V(Y)$
- $V(X) = E(X^2) - E(X)^2$
- $E(g(x)) = \sum_x g(x)p(x)$

Dessa räkneregler kan som övning visas genom att använda defintionerna av väntevärde och varians. Gör det!

Lös uppgifterna i "Exempel på diskreta variabler" som kan laddas ned från kurshemsidan!

Exempel Antag att vi kastar häftstift och vinner 2 kr varje gång det landar med spik nedåt men förlorar 1 kr varje gång spiken hamnar uppåt. Häftstiftet hamnar med spiken nedåt i genomsnitt 37 gånger av 100. Vad är den förväntade vinsten inför *ett* ett kast?

Lösning: Den totala vinsten från 100 kast varav x kast hamnar nedåt är

$$2x + (-1)(100 - x) = 3x - 100$$

Den genomsnittliga vinsten per kast är

$$\frac{\text{Total vinst}}{\text{Totalt antal kast}} = \frac{3x - 100}{100} = 0.03x - 1$$

Eftersom $x = 37$ i detta exempel så är den förväntade vinsten på ett kast
$0.03 \cdot 37 - 1 = 0.11$ kr.

Vi kan också se det som $\frac{\text{Total vinst}}{\text{Totalt antal kast}} = \frac{2x+(-1)(100-x)}{100} = 2 \cdot \frac{x}{100} + (-1)(1 - \frac{x}{100})$. Eftersom $\frac{x}{100}$ är andelen kast av de 100 som hamnar med spiken nedåt så är $E(\frac{X}{100}) =$ $= P(\text{spiken nedåt på ett kast})$ varmed värdet av $P(\text{spiken nedåt på ett kast})$ är just $\frac{37}{100} = 0.37$. Låt nu Y vara vinsten vid ett kast med häftstiftet. Då har vi i allmänhet att

$$E(Y) = \sum_y yp(y) = 2P(\text{nedåt}) + (-1)P(\text{uppåt}) = 2 \cdot 0.37 - (1 - 0.37) = 0.11$$

Tolkningen av begreppet väntevärde är att det är "det värde man kan förvänta sig inför utförandet av experimentet". Denna formulering är dock i många fall inte helt rättvisande eftersom man kan förvänta sig att antingen förlora 1 kr eller vinna 2 kr men definitivt inte att vinna 0.11 kr. Dvs väntervärdet 0.11 återfinns inte i utfallsrummet för vinsten $\{-1, 2\}$. Inte desto mindre kan väntevärdet sägas vara "tyngdpunkten i fördelningen för Y". $\square$

Övning Visa att

- $X \in Likf(n) \;\Rightarrow\; E(X) = \frac{n+1}{2}$ och $V(X) = \frac{n^2-1}{12}$

- $X \in Bin(n,p) \;\Rightarrow\; E(X) = np$ och $V(X) = np(1-p)$
  (Tips: Använd $x\binom{n}{x} = n\binom{n-1}{x-1}$ för alla $n \geq x \geq 1$ och binomialsatsen.)

- $X \in Poi(\lambda) \;\Rightarrow\; E(X) = \lambda$ och $V(X) = \lambda$
  (Tips: $\sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{\lambda}$)

- $X \in Geo(p) \;\Rightarrow\; E(X) = \frac{1}{p}$ och $V(X) = \frac{1-p}{p^2}$