Ex. 13 I tärningsexemplet ovan har vi $E(X) = \sum_{x=-\infty}^{+\infty} x \cdot p_X(x)$ = (enl. ex. 11) =

$$= \sum_{x=1}^{6} x \cdot \frac{1}{6} = 3.5$$

Ex. 14 En roulettspelande person sätter 1 kr på rött. Bestäm väntevärdet av hans nettovinst.

Lösning: Låt X beteckna personens nettovinst. Om rött kommer upp (chansen för detta är $\frac{18}{37}$) får personen 2 kr tillbaka och hans nettovinst blir alltså 1 kr. Om rött inte kommer upp (chans: $\frac{19}{37}$) så förlorar personen sin satsade krona och hans netto-"vinst" blir -1 kr. Enl.(9) får vi $E(X) = 1 \cdot \frac{18}{37} + (-1) \cdot \frac{19}{37} = -\frac{1}{37} \approx -0.028$ kr. *I snitt* förlorar man alltså (i det långa loppet) 2.8 öre per spel.*

Om vi i det inledande tärningsexemplet hade intresserat oss för medeltalet av *kvadraterna* på antalet prickar hade vi på samma sätt letts till följande definition:

$$E(X^2) = \sum_{x=-\infty}^{+\infty} x^2 \cdot p_X(x)$$

Ännu mer allmänt gäller:

$$E(g(X)) = \sum_{x=-\infty}^{+\infty} g(x) \cdot p_X(x) \tag{10}$$

Ex. 15 (forts av tärningsexemplet) $E(X^2) = \sum_{x=-\infty}^{+\infty} x^2 \cdot p_X(x)$ = (enl. ex. 11) =

$$= \sum_{x=1}^{6} x^2 \cdot \frac{1}{6} = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + \ldots + 6^2 \cdot \frac{1}{6} = \frac{91}{6}$$

Varians och standardavvikelse.

Om vi i vårt tärningsexempel intresserar oss för *genomsnittliga avvikelsen* från väntevärdet 3.5, är det naturligt att studera $E(|X-3.5|)$ (absolutbelopp eftersom vi bara är intresserade av avvikelsens storlek och ej dess tecken). Men eftersom absolutbelopp är svårhanterligt ur matematisk synpunkt ser man istället på $E((X-3.5)^2)$. För att måttet på avvikelsen ska få samma dimension (dvs. samma sort) som X, drar man sedan (positiva) roten ur detta.

* Eftersom det vid roulettspel inte är medelvinsten utan *totalvinsten* som är intressant, betyder det låga värdet på väntevärdet dock *inte* att det är (relativt) riskfritt att spela roulett!

$\sqrt{E((X-3.5)^2)}$ kallas för *standardavvikelsen* för X och betecknas D(X). $E((X-3.5)^2)$ kallas för *variansen* för X (betecknas V(X)). Mera allmänt gäller följande:

(11) Def. $V(X) = E((X-m)^2)$ där $m=E(X)$

$D(X) = \sqrt{V(X)}$

Not. Observera att det här definierade sambandet mellan E, V och D gäller också i det kontinuerliga fallet (som snart ska behandlas).

Ex. 16 Vi ska senare visa (se nedan) att $V(X) = E(X^2) - m^2$.
För vårt tärningsexempel ger detta (enl. ex.13 och ex.15)
$V(X) = \frac{91}{6} - (3.5)^2 = \frac{35}{12}$ och $D(X) = \sqrt{\frac{35}{12}} \approx 1.71$. Så mycket ungefär avviker alltså *i snitt* de enskilda kastens resultat från 3.5.
(Ta en tärning och pröva; ju fler kast du gör desto bättre bör det stämma!)

Räkneregler för E och V.

1) $E(aX + bY) = a \cdot E(X) + b \cdot E(Y)$

2) $V(aX + bY) = a^2 \cdot V(X) + b^2 \cdot V(Y)$

3) $E(c) = c$, $V(c) = 0$

Här är a, b och c godtyckliga konstanter. I 2) men inte i 1) förutsätts att X och Y är ober. Formlerna gäller också i det kontinuerliga fallet.

Ex. $V(X) = E((X-m)^2) = E(X^2 + m^2 - 2mX) =$
$= E(X^2) + E(m^2) - 2m \cdot E(X) = E(X^2) + m^2$
$- 2m \cdot m = E(X^2) - m^2$. VSV

Den viktigaste av de diskreta fördelningarna är Binomialfördelningen. Den måste man lära sig känna igen. Två exempel:

1) X: antal sexor vid 5 kast med en symmetrisk tärning

X är Bin $(5, \frac{1}{6})$ d.v.s. (se FS)

$$P(X = x) = \binom{5}{x} \cdot \left(\frac{1}{6}\right)^{x} \cdot \left(1 - \frac{1}{6}\right)^{5-x} \text{ för } x = 0,1,2,3,4,5$$

2) 200 elever tentar ober. av varandra. De har alla 70% chans att klara tentan.

X: antal elever som klarar tentan

X är Bin $(\underbrace{200}_{n}, \underbrace{0.70}_{p})$

I båda dessa fall kan man se det som om ett antal försök upprepats ober. av varandra med samma sannolikhet att lyckas i varje försök och den stokastiska variabel man är intresserad av är totala antalet lyckade försök. Obs. att $E(X) = n \cdot p$ och att $V(X) = n \cdot p \cdot (1-p)$

Om $\bar{X}$ är Bin (n, p) så betecknar n antalet försök och p (P i F.S.!) sannolikheten att lyckas i ett enskilt försök. Om $p \to 0$ samtidigt som $n \to \infty$ (men np hela tiden hålls konstant $= \mu$) så går Bin-fördelningen över i den s.k. Poissonfördelning.

Denna har sannolikhetsfunktion:

$$P(\bar{X} = x) = \frac{\mu^x}{x!} e^{-\mu} \quad \text{för } x = 0, 1, 2, 3 \ldots$$

Här är $\mu = E(\bar{X}) = V(\bar{X})$

En viktig egenskap hos poisson-fördelningen är att summan av oberoende poissonfördelningar också är poissonfördelad.

Gå igenom ex. 4.10 på sid 168. (Satsen ovan används i 4.10.c)

Läs även stycket efter 4.10.c om hur poissonfördelningen kan användas.

Dragning med och utan återläggning.

Ex. Ett varuparti består av 100 batterier. 6 av dessa batterier är defekta.
X: antal defekta batterier i ett slumpmässigt urval av 5 batterier. Bestäm sannolikhetsfunktionen för X.

Det finns två svar på denna fråga beroende på hur urvalet görs. Gör man så att man drar ett batteri, undersöker om det är defekt eller ej och sedan lägger tillbaka det innan man slumpvis drar nästa batteri så gäller att X är Bin $(5, 0.06)$.
Gör man istället på det vanligare sättet d.v.s drar batterierna alla på en gång (eller en och en utan återläggning) så får man:

$$P(X = x) = \frac{g}{m} = \frac{\binom{6}{x} \cdot \binom{94}{5-x}}{\binom{100}{5}} \text{ för } x = 0, 1, 2, 3, 4, 5$$

(En s.k. Hypergeometrisk fördelning.)

Om antalet som man drar är högst 1/10 av partistorleken kan dock den hypergeometriska fördelningen approximeras med Bin-fördelningen.

Kontinuerliga fördelningar

För att kunna hantera dessa behövs den s.k. täthetsfunktionen. Denna är inte en sannolikhet utan en slags "hjälpfunktion" till sannolikhet. Egenskaper:

1) $f_X(x) \geq 0$ (≤ 1 ej nödvändigt!)

2) $P(a \leq X \leq b) = \int_a^b f_X(x)\,dx$

3) $1 = \int_{-\infty}^{+\infty} f_X(x)\,dx$

Tabeller över sannolikhetsfördelningar är vanligen uppställda i termer av den s.k. fördelningsfunktionen vilket gör att man (för det mesta) inte behöver integrera (eller summera i det diskreta fallet).

<u>Def.</u> Fördelningsfunktionen för X är $F_X(x) = P(X \le x)$.

Viktigt samband: $P(a < X \le b) = F_X(b) - F_X(a)$. (I det diskreta fallet är typen av olikhetstecken väsentlig, i det kontinuerliga spelar det ingen roll om det står $<$ el. $\le$)

Den viktigaste kontinuerliga fördelningen är den s.k. normalfördelningen. Om X är normalfördelad med $E(X) = \mu$ och $V(X) = \sigma^2$ skriver man detta som X är $N(\mu, \sigma)$ (Vi följer F.S., boken skriver X är $N(\mu, \sigma^2)$). Tabell 1 i boken är en tabell över fördelningsfunktionen för Z när Z är $N(0,1)$. Så för att klara andra parametrar än 0 och 1 så måste man kunna normera.

Ex. X är $N(5, 2)$. Då gäller (se F.S.) att $Z_1 = \frac{X-5}{2}$ är $N(0,1)$

Sökt: $P(X \leq 6) = P\left(\frac{X-5}{2} \leq \frac{6-5}{2}\right) =$
$= P(Z_1 \leq 0.5) = (\text{tabell } 1) = \underline{0.6915}$

En viktig sats är följande:

$X_1, X_2 \cdots X_n$ är ober och
X_i är $N(\mu_i, \sigma_i)$ $i = 1, 2, \cdots n$

$\Rightarrow \sum_{i=1}^{n} a_i X_i$ är $N\left(\sum_{i=1}^{n} a_i \mu_i, \sqrt{\sum_{i=1}^{n} a_i^2 \cdot \sigma_i^2}\right)$

(Här är a_i:na godtyckliga konstanter ej alla $= 0$ förstås)

Ex. Vikten (i kg) av barn är $N(20, 5)$
Vikten (i kg) av vuxna är $N(70, 10)$
Beräkna sannolikheten att 6 barn väger mer än 2 vuxna! (Samtliga personvikter antas vara ober.)

Med X_j: vikten i kg av barn nr j ($j = 1, 2 \cdots 6$) och Y_k = Vikten av vuxen nr k ($k = 1, 2$) blir det sökta:
$P(X_1 + \cdots + X_6 > Y_1 + Y_2) = (!) =$
$= P(\underbrace{X_1 + \cdots + X_6 - Y_1 - Y_2}_{Q} > 0) =$
satsen ovan $\Rightarrow$ Q är $N(-20, \sqrt{350})$

$$= P\left(\frac{Q-(-20)}{\sqrt{350}} > \frac{0-(-20)}{\sqrt{350}}\right) =$$

$$= P(Z_1 > 1.71) = 1 - P(Z_1 \le 1.71) =$$

$$= 1 - F_{Z_1}(1.71) = (\text{se tabell } 1) =$$

$$= 1 - 0.9564 \approx \underline{\underline{14\%}}$$

(((c.g.s)

Centrala gränsvärdessatsen

Antag att $X_1, X_2 \dots X_n$ är ober och att alla har precis samma fördelning. Då har de också samma väntevärde och standardavvikelse d.v.s.

$E(X_i) = \mu$, $D(X_i) = \sigma$ $i = 1, 2, \dots n$

Då gäller att $X_1 + \dots + X_n \approx N(n \cdot \mu, \sqrt{n} \cdot \sigma)$ om n är "stort".

Vad "stort" är beror på vad den gemensamma fördelningen på X_i är. Ju snedare fördelning ju större n krävs men $n \ge 30$ räcker nästan alltid!

Antag nu att $\underline{X}$ är $Bin(n, p)$

Om vi inför $\mathcal{U}_j = \begin{cases} 1 & \text{om } j\text{:te försöket lyckas} \\ 0 & \text{annars} \end{cases}$

så har vi $E(\mathcal{U}_j) = 1 \cdot p + 0 \cdot (1-p) = p$

och $E(\mathcal{U}_j^2) = 1^2 \cdot p + 0^2 \cdot (1-p) = p$. Detta ger $V(\mathcal{U}_j) = E(\mathcal{U}_j^2) - [E(\mathcal{U}_j)]^2 = p(1-p)$

Men $\underline{X} = \mathcal{U}_1 + \cdots + \mathcal{U}_n$ Således

$E(\underline{X}) = E(\mathcal{U}_1) + \cdots + E(\mathcal{U}_n) = n \cdot p$ och

$V(\underline{X}) = (\text{ober!}) = V(\mathcal{U}_1) + \cdots + V(\mathcal{U}_n) = n \cdot p(1-p)$

Man kan också tillämpa C.G.S. på $\underline{X} = \mathcal{U}_1 + \cdots + \mathcal{U}_n$ och får då $\underline{X} \approx N(np, \sqrt{np(1-p)})$ men här räcker det inte med n stort eftersom $\mathcal{U}_i$:nas fördelning blir väldigt sned om p (eller $1-p$) är mycket litet. Det villkor som också tar hänsyn till snedheten är $np(1-p) > 5$. Sammanfattning:

$\underline{X}$ är $Bin(n, p) \approx N(np, \sqrt{np(1-p)})$ om $np(1-p) > 5$.