Stokastiska variabler i två dimensioner

Definition

- En *stokastisk variabel*, (X,Y), är en reellvärd funktion definierad på ett utfallsrum (Ω). $(X,Y)\colon \Omega \to \mathcal{R}$

Definition

- Funktionen

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y)$$

kallas *fördelningsfunktionen* för (X,Y).

- Om (X,Y) båda antar ett uppräkneligt antal värden sägs (X,Y) vara *diskret*.

Definition

Med den *simultana sannolikhetsfunktionen* för en diskret tvådimensionell stokastisk variabel *(X,Y)* menas

$$p_{X,Y}(j,k) = P(X = j, Y = k) \quad , j \in \Omega_X , k \in \Omega_Y$$

Den *marginella sannolikhetsfuktionen* för *X* fås ur

$$p_X(j) = \sum_{k=0}^{\infty} p_{X,Y}(j,k)$$

Dessutom gäller

$$\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} p_{X,Y}(x,y) = 1$$

Definition

Om det finns en funktion, $f_{X,Y}(x,y)$,så att

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(s,t)dtds$$

gäller sägs *(X,Y)* vara kontinuerlig.

$f_{X,Y}(x,y)$ kallas *simultana täthetsfuktionen* för *(X,Y)*.
För *täthetsfunktionen* gäller

$$f_{X,Y}(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} \quad \text{och} \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1$$

Den *marginella sannolikhetsfuktionen* för *X* fås ur

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

Oberoende stokastiska variabler

<u>Definition</u>

X och Y sägs var oberoende stokastiska variabler om

$$p_{X,Y}(j,k) = p_X(j)p_Y(k) \quad diskreta$$
$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \quad kontinuerliga$$

gäller för alla j och k , x och y.

<u>Jmf definitionen av oberoende händelse</u>

$$P(A \cap B) = P(A)P(B)$$

Väntevärde, lägesmått (centralmått)

- Om X är en stokastisk variabel med utfallsrummet Ω definieras väntevärdet för *X*, E*[X]* ofta betecknat μ, av

$$E[X] = \sum_i x_i P(X = x_i), X\ diskret$$

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx, X\ kontinuerlig$$

- Väntevärdet anger tyngdpunkten i en fördelning
- Är ett teoretiskt medelvärde, ett medelvärde man kan förvänta sig om man gör upprepade försök

Sats 1 & 2

Om $Y = g(X)$ gäller

$$E(Y) = E(g(X)) = \sum_{x} g(x)p(x),\ X \text{diskret}$$

$$E(Y) = E(g(X)) = \int_{\Omega} g(x)f(x)dx,\ \text{om } X \text{kont.}$$

Om $Z = g(X,Y)$ gäller

$$E(Z) = E(g(X,Y)) = \sum_{x\in\Omega_X} \sum_{y\in\Omega_Y} g(x,y)p_{X,Y}(x,y) \ \text{om } X,Y \text{diskret}$$

$$E(Y) = E(g(X,Y)) = \int_{\Omega_X} \int_{\Omega_Y} g(x,y)f_{X,Y}(x,y)dydx \ \text{om } X,Y \text{kont.}$$

Sats 3

$$E(g(X)+ h(X)) = E(g(X)) + E(h(X))$$

dvs E är en linjär operator

Standardavvikelse och varians, spridningsmått

Definition

- Om X är stokastisk variabel med väntevärdet $E[X] = \mu$, definieras variansen för X, $V(X)$, ofta betecknad σ^2 , som

$$V[X] = E[(X-\mu)^2]$$

$$E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx, X \text{ kont.}$$

$$E[(X-\mu)^2] = \sum_x (x-\mu)^2 p(x), X \text{diskret}$$

- Standardavvikelsen för X, $\sigma = D(X) = \sqrt{V[X]}$

Andra läges och spridningsmått Median, kvartil och percentil

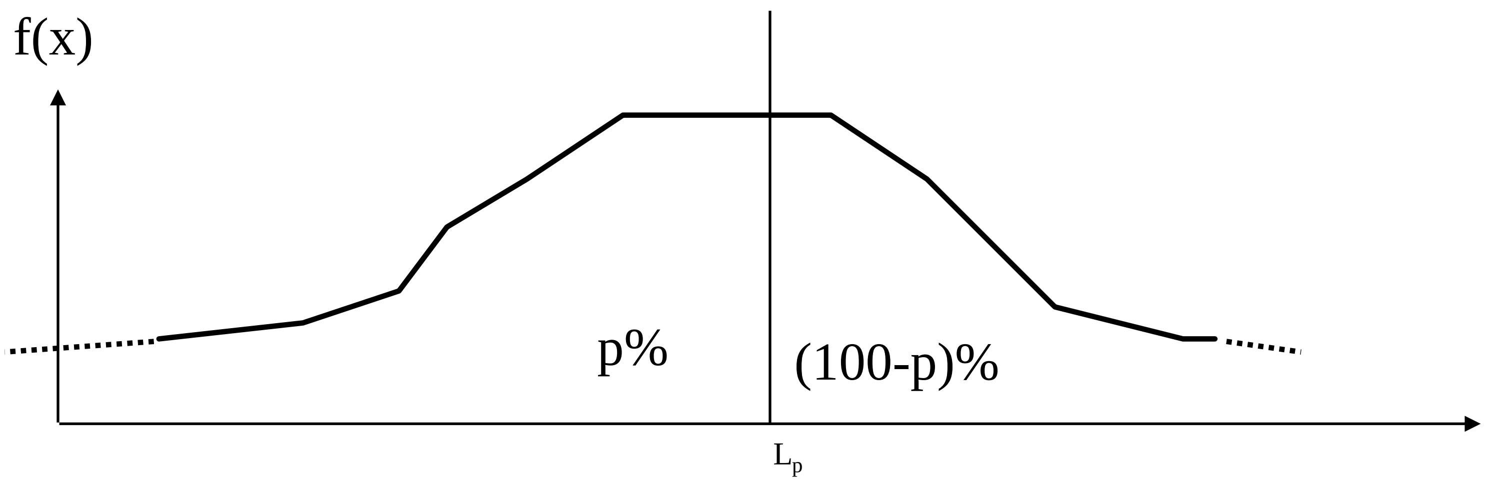
- Den stokastiska variabeln *X* har fördelningsfunktionen *F(x)*. Medianen definieras som det tal, *m*, som uppfyller

$$F(m) = 0{,}5$$

- Den stokastiska variabeln *X* har fördelningsfunktionen *F(x)*. Den p:te percentilen definieras som det tal L_p som uppfyller

$$F(L_p) = p\% = (p/100)$$

- Med kvartiler avses $Q_1 = L_{25}$, $Q_2 = L_{50}$ (medianen) och $Q_3 = L_{75}$.



Några räkneregler för väntevärde och varians

Sats 6

$$V(X) = E[(X-\mu)^2] = E(X^2) - \mu^2$$

Där $E(X^2) = \begin{cases} \displaystyle\sum_{k\in\Omega} x^2 p(x), X\ diskret \\ \displaystyle\int_{-\infty}^{\infty} x^2 f(x)dx, X\ kontinuerlig \end{cases}$

Sats 7

$E[aX+b] = aE[X]+b$ $\quad$ $V[aX+b] = a^2V[X]$

$a, b \in \mathbb{R}$ $\quad$ $D[aX+b) = |a|D[X]$

Beroendemått

Definition

Kovariansen för X och Y definieras av

- $C(X,Y) = E((X-\mu_1)(Y-\mu_2)) = E(XY)-E(X)E(Y)$

- Om $C(X,Y) = 0$ sägs X och Y okorrelerade
  dvs $E(XY) = E(X)E(Y)$

- *Korrelationskoefficienten* för X och Y ges av

$$\rho(X,Y) = \frac{C(X,Y)}{D(X)D(Y)}$$

det gäller att $\quad -1 \leq \rho \leq 1$

Fler räkneregler för väntevärden

Produkt (sats 4 & 9)

- Om X och Y är oberoende stokastiska variabler gäller

$$C(X,Y) = 0 \Leftrightarrow E(XY) - E(X)E(Y) = 0 \Leftrightarrow E(XY) = E(X)E(Y)$$

Summa och differens (sats 10)

- För alla stokastiska variabler X och Y gäller

$$E(X \pm Y) = E(X) \pm E(Y)$$
$$V(X \pm Y) = V(X) + V(Y) \pm C(X,Y)$$

- Om X och Y är oberoende stokastiska variabler gäller dessutom

$$V(X \pm Y) = V(X) + V(Y)$$

$$D(X \pm Y) = \sqrt{V(X) + V(Y)}$$

Fler räkneregler för väntevärden

Aritmetiskt medelvärde (sats 11.2 & 11.3)

Om $X_1, X_2, ..., X_n$ är oberoende och likafördelade stokastiska variabler dvs $E[X_i] = \mu$ *och* $V[X_i] = \sigma^2$, $i=1,..,n$

Låt $Y = \sum_{i=1}^{n} X_i$

Då gäller

$E[Y] = n\mu$ och

$V[Y] = n\sigma^2$

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

Då gäller

$E[\bar{X}] = \mu$ och $V[\bar{X}] = \sigma^2/n$

Sats 12 (Stora talens lag)

Låt $X_1, X_2, \ldots\ldots$ vara oberoende och likafördelade stokastiska variabler med $E(X_i) = \mu$. Sätt

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$ Då gäller, för alla $\varepsilon > 0$, att

$$P(\mu - \varepsilon < \bar{X} < \mu + \varepsilon) \to 1, n \to \infty$$

Sats 13 (Markovs olikhet)

För $a > 0$ och $Y \geq 0$ gäller

$$P(Y \geq a) \leq \frac{E(Y)}{a}$$

Tjebysjovs olikhet (sats 14)

För varje k > 0 gäller

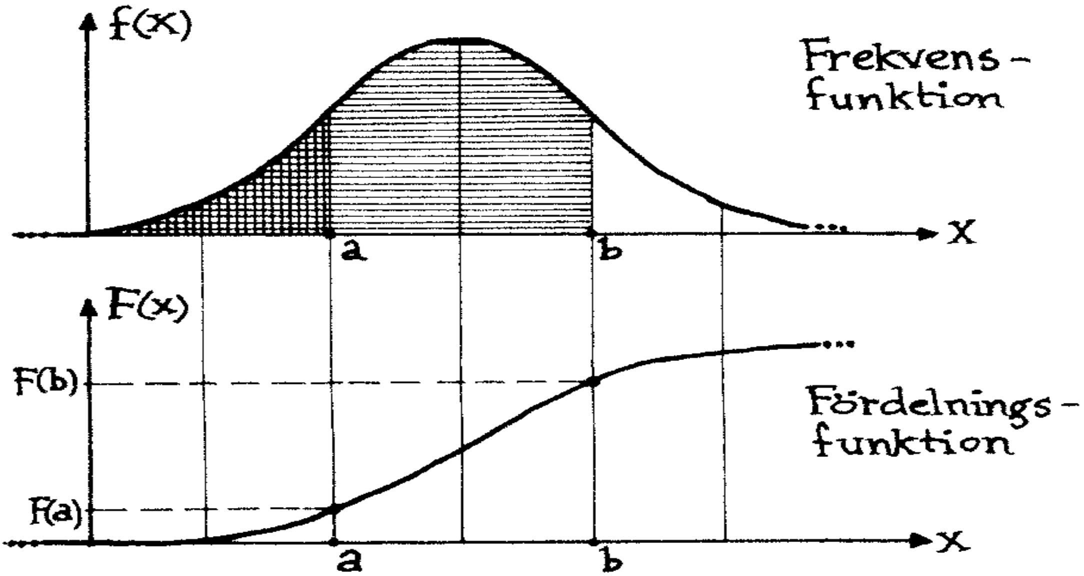
$$P(|X-\mu| \ \geq k\sigma) \leq 1/k^2$$

Normalfördelningen

Normalfördelningen används ofta till att beskriva variationen hos olika företeelser. Mycket viktig, stora delar av statistikteorin bygger på normalfördelningen.

– Normalfördelningen bestäms av två parametrar lägesparametern, μ , samt spridningsparametern, σ

$$\boxed{X \in N(\mu, \sigma)}$$



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

$$-\infty < x < \infty$$

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sigma \quad 2\pi} e^{-(t-\mu)^2/(2\sigma^2)} dt$$

$$-\infty < x < \infty$$

Standardnormalfördelning

Med μ = *0* och σ = *1* fås det viktiga specialfallet $\boxed{X \in N(0,1)}$

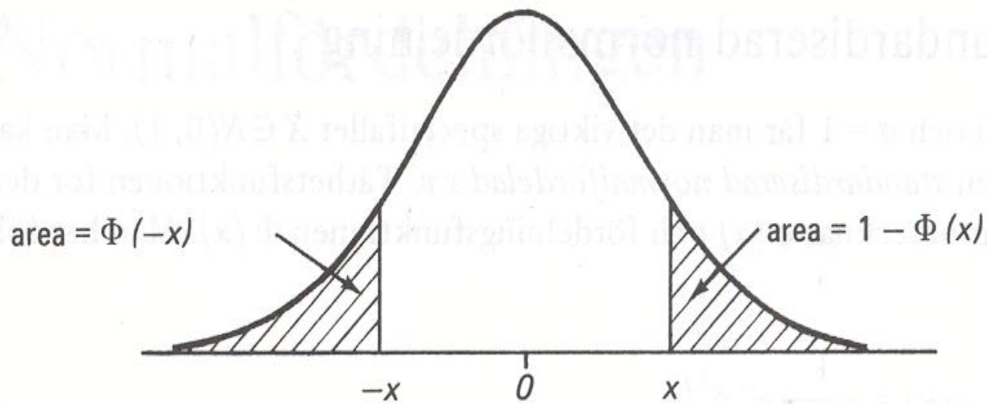
Man kallar *X* för en *standardnormalfördelad stokastisk variabel.*

Täthetsfunktionen betecknas med *φ(x)* och

fördelningsfunktionen med *Φ(x)*, vilken finns tabellerad i slutet av boken.

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \qquad -\infty < x < \infty$$

$$\Phi(x) = \int_{-\infty}^{x} \phi(x)dt = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/\sigma^2} dt$$



φ(x) är en jämn funktion och är symmetriskt kring 0.

Detta ger egenskapen: $\Phi(-x) = 1 - \Phi(x)$

Allmänna egenskaper

Sats 1

Om $X \in N(\mu,\sigma)$ och $Y = \dfrac{X-\mu}{\sigma}$

Då blir $Y \in N(0,1)$

Sats 2

Om $X \in N(\mu,\sigma)$ då är $E(X)=\mu$ och $D(X) = \sigma$.

Dessutom gäller

$Y = aX + b \in N(a\mu + b; |a|\sigma)$

Allmänna egenskaper forts.

För alla normalfördelningar gäller:

$$P(\mu-\sigma < X < \mu+\sigma) = 0.682$$

$$P(\mu-2\sigma < X < \mu+2\sigma) = 0.954$$

$$P(\mu-3\sigma < X < \mu+3\sigma) = 0.997$$

$$P(\mu-1.96\sigma < X < \mu+1.96\sigma) = 0.95$$

$$P(\mu-2.58\sigma < X < \mu+2.58\sigma) = 0.99$$

$$P(\mu-3.29\sigma < X < \mu+3.29\sigma) = 0.999$$

Fler egenskaper

Sats 4

Om $X_1 \in N(\mu_1; \sigma_1)$, Om $X_2 \in N(\mu_2; \sigma_2)$ och oberoende gäller

$$X_1 + X_2 \in N(\mu_1 + \mu_2; \sqrt{\sigma_1^2 + \sigma_2^2})$$

$$X_1 - X_2 \in N(\mu_1 - \mu_2; \sqrt{\sigma_1^2 + \sigma_2^2})$$

Sats 5 och 6

Om $X_i \in N(\mu_i; \sigma_\mathrm{i})$ och oberoende samt $c_i \in \mathbb{R}$ och är givna, $i = 1, \dots, n$ så gäller

$$\sum_{\mathrm{i}=1}^{\mathrm{n}} X_i \in N\left(\sum_{i=1}^{n} c_i\mu_i \,;\, \sqrt{\sum_{i=1}^{n} c_i^2\sigma_i^2}\right) \qquad \text{med} \quad \mu_i = \mu \text{ fås}$$

$$\sum_{\mathrm{i}=1}^{\mathrm{n}} X_i \in N(n\mu; \sqrt{n}\sigma) \quad \text{och} \quad \bar{X} \in N(\mu; \sigma/\sqrt{n})$$

Centrala gränsvärdessatsen

Låt $X_1, X_2, \ldots, X_n$ vara oberoende och likafördelade stokastiska variabler med väntevärdet μ och standardavvikelsen σ

$$P\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \leq x\right) \to \Phi(x), n \to \infty$$

Praktiskt: summan av antal slumpvariabler är approximativt normalfördelade om n är stort. (Tumregel n > 30)
Normalapproximationer är mycket användbara

Om n är stort så gäller $\bar{X}$ **är approximativt** $N(\mu, \sigma/\sqrt{n})$

$$\sum_{i=1}^{n} X_i$$ är approximativt normalfördelad $N(n\mu, \sigma\sqrt{n})$

Binomialfördelningen

- Ett försök består av n upprepningar av oberoende delförsök, (ex: <u>dragning med återlägg</u>). A är en speciell händelse som inträffar med sannolikheten p i varje delförsök. Slumpvariabeln X betecknar antalet gånger händelsen A inträffar i hela försöket. X blir då binomialfördelad

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \; x = 0, 1, 2, \ldots, n$$

$$\boxed{X \in Bin(n, p)}$$

”oberoende upprepningar med samma sannolikhet”

Binomialfördelningen

Egenskaper

Sats 2

Om $X \in Bin(n,p)$ *så är*

$E(X) = np \quad V(X) = np(1-p) \quad D(X) = \sqrt{np(1-p)}$

Sats 3

$Om\ X \in Bin(n_1; p)\ och\ Y \in Bin(n_2; p), X\ och\ Y\ oberoende,$
$\quad$ så $blir\ X + Y \in Bin(n_1 + n_2; p)$

Binomialfördelningen

Approximativa egenskaper.

Om $X \in Bin(n,p)$ *och* $V(X)>10$ *så är*

$X \approx N(np; \sqrt{np(1-p)}\,)$

Om $X \in Bin(n,p)$ *och* $p < 0.1$ *så är*

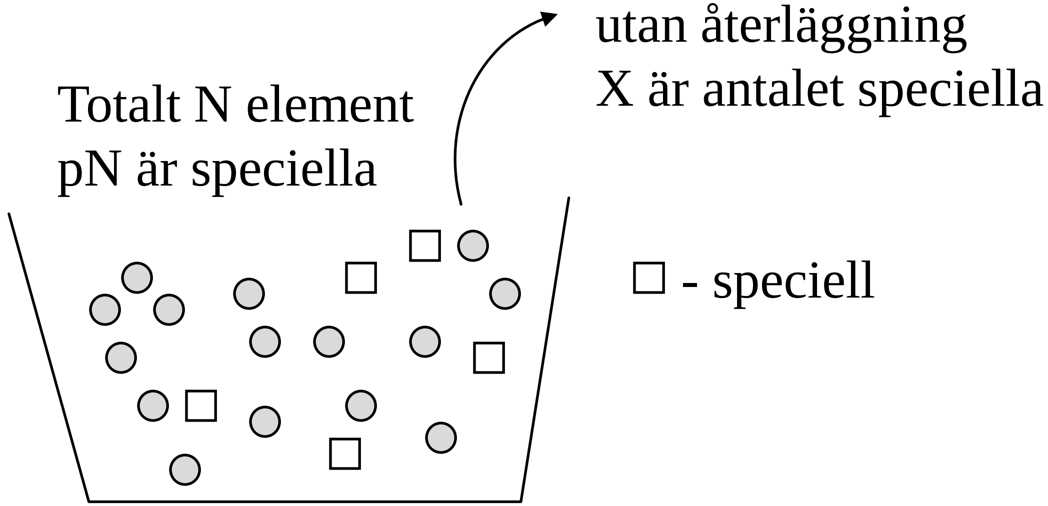
$X \approx Po(np)$

Hypergeometrisk fördelning

- En mängd innehåller totalt *N* element, av vilka *Np* är av speciellt slag (andelen speciella är *p*). Välj slumpmässigt, utan återläggning, ett urval av n element. Slumpvariabeln *X* betecknar antalet speciella element i urvalet. *X* blir då hypergeometriskt fördelad.

$$P(X = x) = \frac{\binom{Np}{x}\binom{N-Np}{n-x}}{\binom{N}{n}}$$

x är ett heltal sådant att $0 \leq x \leq Np$ *och* $0 \leq n-x \leq N-Np$



$$X \in Hyp(N, n, p)$$

Hypergeometrisk fördelning

Egenskaper Sats 6

Om $X \in Hyp(N,n,p)$ *så är*

$E(X) = np \quad V(X) = np(1-p)\frac{N-n}{N-1} \qquad D(X) = \sqrt{np(1-p)\frac{N-n}{N-1}}$

Approximativa egenskaper

Om $X \in Hyp(N,n,p)$

1. *och* $V(X) > 10$ *så är* $X \approx N(np; \sqrt{np(1-p)\frac{N-n}{N-1}})$
2. *och* $n/N < 0.1$ *så är* $X \approx Bin(n,p)$
3. *och* $p + n/N < 0.1$ *så är* $X \approx Po(np)$

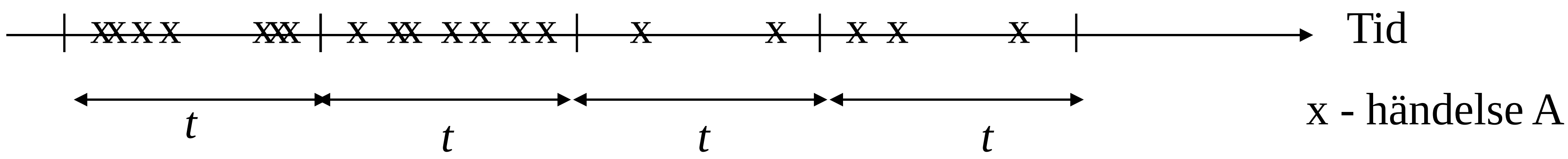
Poissonfördelning

- Betrakta händelser A som slumpmässig och oberoende av varandra inträffar i tiden. Slumpvariabeln X betecknar antalet händelser A som inträffar under ett tidsintervall av fix längd. X blir då poissonfördelad

$$P(X = x) = e^{-\mu}\frac{\mu^x}{x!},\ x = 0, 1, \ldots$$

μ – kallas intensitetsparameter

μ anger genomsnittligt antal händelser A under tiden t



$$\boxed{X \in Po(\mu)}$$

Poissonfördelning

Egenskaper

Sats 7 $Om\ X \in Po(\mu)\ s$å är
$E(X) = \mu, \quad V(X) = \mu$ och $D(X) = \sqrt{\mu}$

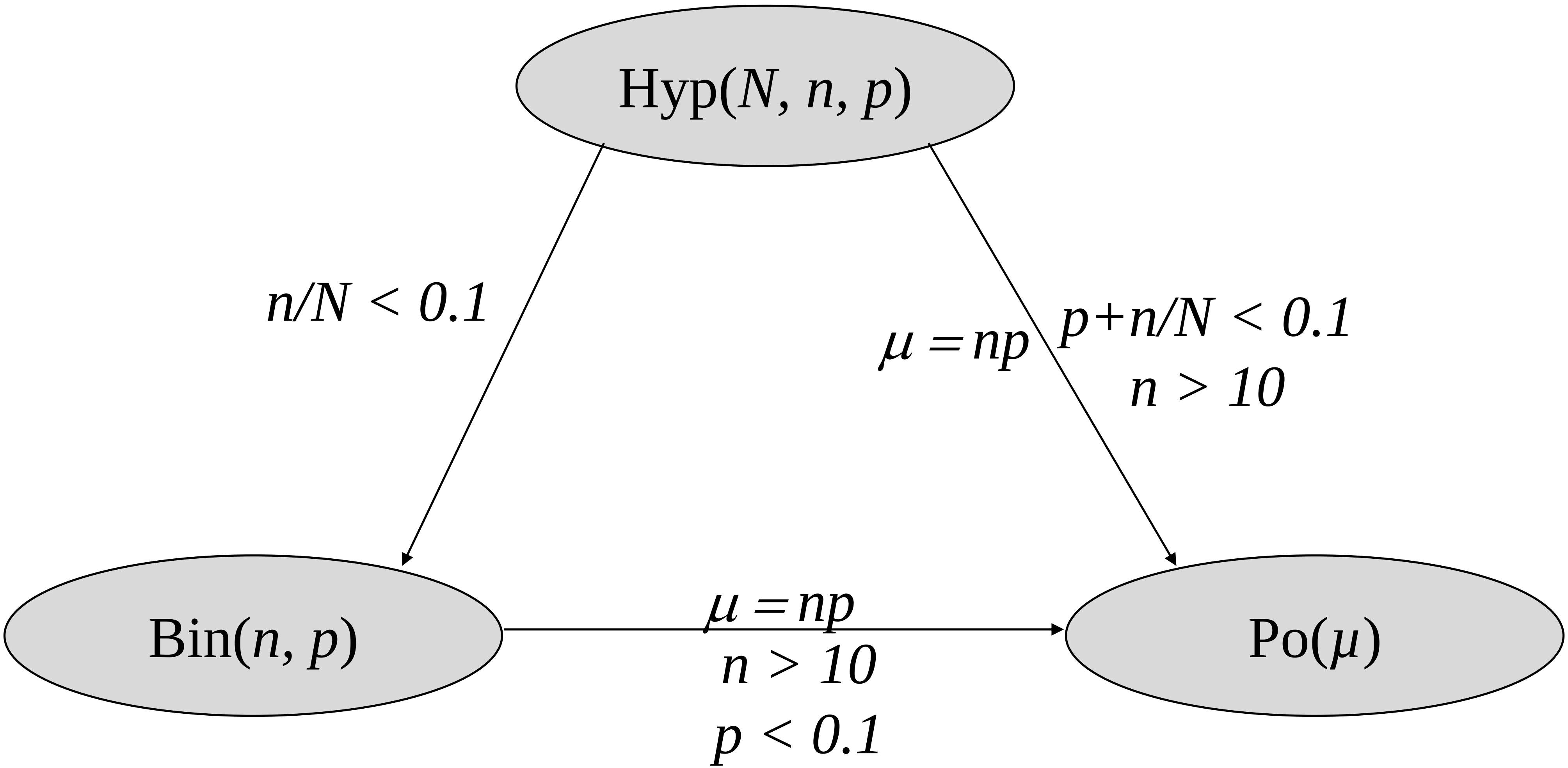
Sats 8 $Om\ X \in Po(\mu_1)\ och\ Y \in Po(\mu_2), X\ och\ Y\ oberoende,$
så är $X + Y \in Po(\mu_1 + \mu_2)$

Approximativa egenskaper

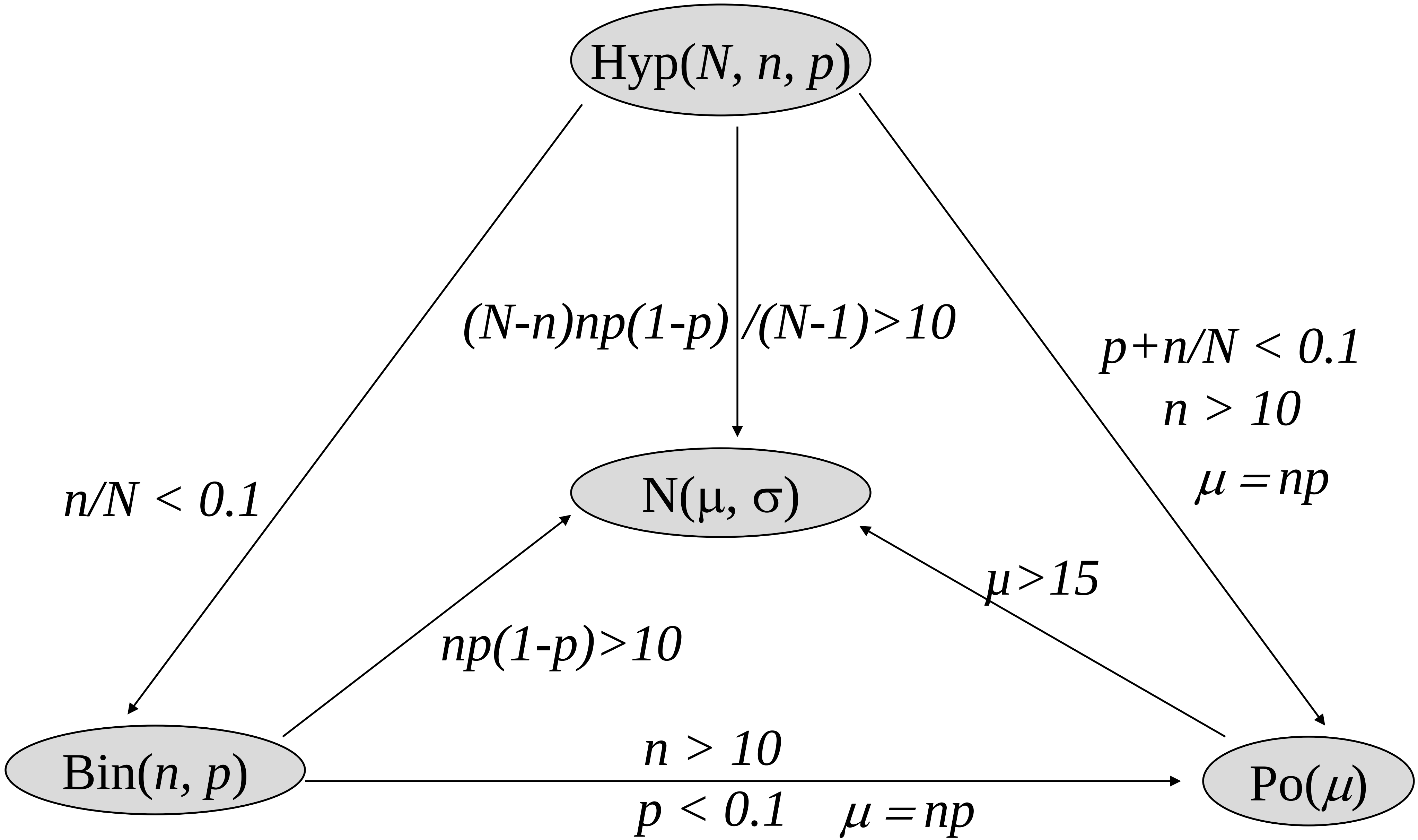
$Om\ X \in Po(\mu)\ och\ \mu > 15$
så är $X \approx N(\mu, \sqrt{\mu})$

Hypergeometrisk, binomial och poisson fördelning

Approximationsregler



Approximationsregler - centrala gränsvärdessatsen



Gauss approximationsformler

- X är en stokastisk variabel med vänte värde $E[X] = \mu$ och varians $V[X] = \sigma^2$ och g är en funktion med kontinuerlig derivata

Då gäller $E[g(X)] \approx g(\mu)$ och $V[g(X)] \approx (g'(\mu))^2 \sigma^2$

- X_1 och X_2 är oberoende stokastiska variabler med vänte värde m_1 och m_2 samt varianser σ_1^2 och σ_2^2
- g är en funktion av två variabler x_1 och x_2

Då gäller $E[g(X_1, X_2)] \approx g(\mu_1, \mu_2)$ och

$$V[g(X_1, X_2)] \approx \left(\frac{\partial g(\mu_1, \mu_2)}{\partial x_1}\right)^2 {\sigma_1}^2 + \left(\frac{\partial g(\mu_1, \mu_2)}{\partial x_2}\right)^2 {\sigma_2}^2$$