

SF1901: SANNOLIKHETSTEORI OCH  
STATISTIK  
FÖRELÄSNING 8  
CENTRALA GRÄNSVÄRDESSATSEN (CGS),  
NORMALFÖRDELNINGSAAPPROXIMATIONER.  
BINOMIAL- OCH POISSON FÖRDELNINGEN:  
APPROXIMATIVA EGENSKAPER.

Tatjana Pavlenko

21 september 2015



# PLAN FÖR DAGENS FÖRELÄSNING

- ▶ Egenskaper hos normalfördelningen (rep.)
- ▶ Linjärkombinationer av oberoende normalfördelade s.v. (rep.) (Kap. 6.5)
- ▶ Centrala gränsvärdessatsen och normalfördelningsapproximationer. (Kap. 6.7)
- ▶ Binomialfördelning och Poissionfördelningen. (rep.)
- ▶ Normalapproximation för  $Bin(n, p)$  och  $Po(\mu)$ . (Kap. 7.2, 7.4 )



# NORMALFÖRDELNING (REP.)

- ▶
- ▶ **Def:** En kontinuerlig s.v.  $X$  sägs vara *normalfördelad* med parametrar  $\mu$  och  $\sigma$ , ( $\sigma > 0$ ) om täthetsfunktionen ges av

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

- ▶ Beteckning:  $X \in N(\mu, \sigma)$ .
- ▶ Om  $\mu = 0$  och  $\sigma = 1$  sägs  $X$  vara standardiserad normalfördelad.



## EGENSKAPER HOS NORMALFÖRDELNING (REP.)

- ▶ **Sats:**  $X \in N(\mu, \sigma)$  om och endast om  $Z = \frac{X - \mu}{\sigma} \in N(0, 1)$ .
- ▶  $N(\mu, \sigma)$  är läge-skall familj (location-scale family)!
- ▶ Tolkning av  $\mu$  och  $\sigma$ : enligt sats får vi

$$E(X) = E(\sigma Z + \mu) = \sigma E(Z) + \mu = \mu,$$

$$V(X) = V(\sigma Z + \mu) = \sigma^2 E(Z) = \sigma^2,$$

dvs är parametrarna  $\mu$  och  $\sigma$  *väntevärde* respektive *standardavvikelse* för  $N(\mu, \sigma)$ -fördelad s.v.

- ▶ Vidare gäller att

$$f_X(x) = \frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right), \quad F_X(x) = \Phi\left(\frac{x - \mu}{\sigma}\right),$$

$$P(a < X \leq b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right).$$



# LINJÄR TRANSFORMATION AV NORMALFÖRDELNING.

- ▶ En viktig egenskap hos normalfördelningen är att den bevaras under linjära transformationer.

- ▶ **Sats:** Om  $X \in N(\mu, \sigma)$ , så gäller att

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

- ▶ **Sats:** Om  $X_1, X_2, \dots, X_n$  är oberoende och respektive  $N(\mu_1, \sigma_1), N(\mu_2, \sigma_2), \dots, N(\mu_n, \sigma_n)$  och konstanterna  $a_1, a_2, \dots, a_n, b$  är givna, så gäller att

$$\sum_{i=1}^n a_i X_i + b \in N\left(\sum_{i=1}^n a_i \mu_i + b, \sqrt{\sum_{i=1}^n a_i^2 \sigma_i^2}\right).$$

- ▶ Speciellt, om  $X_1, X_2, \dots, X_n$  är oberoende  $N(\mu, \sigma)$  och  $a_1 = a_2 = \dots = a_n = 1/n$  samt  $b = 0$ , så gäller att

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \in N(\mu, \sigma/\sqrt{n}).$$



# CENTRALA GRÄNSVÄRDESSATSEN (CGS).

- ▶ Satsen är den viktigaste resultaten inom sannolikhetsteorin:

*En summa av oberoende lika fördelade s.v. med godtycklig fördelning är ungefär normalfördelad, bara antalet komponenter i summan är tillräckligt stort.*

- ▶ **Sats (CGS):** Låt  $X_1, \dots, X_n, \dots$  vara en oändlig följd av oberoende, likafördelade s.v. med väntevärde  $\mu$  och standardavvikelse  $0 < \sigma < \infty$ . Sätt

$$Y_n = X_1 + \dots + X_n.$$

Då gäller för givna  $a < b$  att

$$P\left(a < \frac{Y_n - n\mu}{\sqrt{n}\sigma} \leq b\right) \rightarrow \Phi(b) - \Phi(a) \quad \text{då } n \rightarrow \infty.$$

- ▶ CGS uttalar sig alltså om *fördelningen av  $Y_n$  då antalet  $n$  växer mot oändligheten*:  $Y_n$  är ungefär  $N(n\mu, \sqrt{n}\sigma)$ -fördelad. Beteckning:

$$Y_n \in \text{AsN}(n\mu, \sqrt{n}\sigma).$$



# CENTRALA GRÄNSVÄTDESSATSEN (FORTS.)

- ▶ Observera att  $E(Y_n) = n\mu$  och  $D(Y_n) = \sqrt{n}\sigma$ . För varje givet  $n$  är

$$\frac{Y_n - n\mu}{\sqrt{n}\sigma}$$

en standardiserad s.v. Den har väntevärde lika med noll och standardavvikelse lika med 1 som en standardiserad normalfördelad s.v.

- ▶ Enligt CGS: när  $n$  går mot oändligheten kommer hela fördelningen för den angivna standardiserade s.v. att gå mot en *standardiserad normalfördelning*, dvs

$$\frac{Y_n - n\mu}{\sqrt{n}\sigma} \in AsN(0, 1).$$



- **Följdats:** För en oändlig följd av oberoende likafördelade s.v.  $X_1, \dots, X_n, \dots$  med  $E(X_i) = \mu$  och  $D(X_i) = \sigma$  ( $0 < \sigma < \infty$ ) gäller att

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \in AsN \left( \mu, \frac{\sigma}{\sqrt{n}} \right) \quad \text{då } n \rightarrow \infty,$$

dvs *aritmetisk medelvärde*  $\bar{X}_n$  är *approximativt normalfördelat* för *tillräckligt stort*  $n$ .

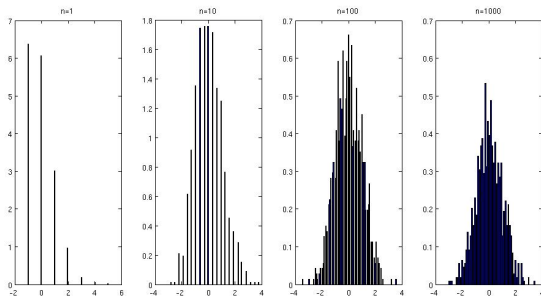
- **Normalfördelningsapproximation.** Enligt CGS:  $\sum_{i=1}^n X_i \in AsN(n\mu, \sqrt{n\sigma^2})$  och  $\bar{X}_n \in AsN \left( \mu, \sigma/\sqrt{n} \right)$ . Detta ger approximationerna

$$P \left( a < \sum_{i=1}^n X_i \leq b \right) \approx \Phi \left( \frac{b - n\mu}{\sqrt{n\sigma^2}} \right) - \Phi \left( \frac{a - n\mu}{\sqrt{n\sigma^2}} \right),$$

$$P(c < \bar{X}_n \leq d) \approx \Phi \left( \frac{d - \mu}{\sigma/\sqrt{n}} \right) - \Phi \left( \frac{c - \mu}{\sigma/\sqrt{n}} \right).$$



## CENTRALA GRÄNSVÄRDESSATSEN (FORTS.)



**FIGUR :** Fördelningen (sannolikhetsfunktion) för  $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$  för  $n = 1$ ,  $n = 10$ ,  $n = 100$  och  $n = 1000$ , där

$$\bar{X}_n = (X_1 + \dots + X_n)/n$$

och  $X_1, \dots, X_n, \dots$  är oberoende  $Po(1)$ -variabler (så att  $\mu = \sigma = 1$ ). Då  $n \rightarrow \infty$  liknar sannolikhetsfunktionen alltmer den standardiserade normalfördelningstäthet.

- ▶ Binomialfördelning omnämndes i Kap. 3. 4 (se föreläs. 3. Som tidigare nämnts, **Def:** om en s.v.  $X$  har sannolikhetsfunktionen

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n$$

där  $0 < p < 1$ , sägs  $X$  vara **binomialfördelad**. Beteckning:  $X \in \text{Bin}(n, p)$ .

- ▶ Förekomst:
  - ▶ Betrakta ett försök som utförs på *förhand bestämt antal gånger*  $n$ . Försöken antas vara oberoende, och varje försök kan lyckas (med slh  $p$ ) eller misslyckas. Låt s.v.  $X$  vara antalet lyckade försök av dessa  $n$ . Man är intresserad att finna  $P(X = k)$ , dvs sannolikheten för att antalet lyckade försök är  $k$ . Då är  $X \in \text{Bin}(n, p)$  och sannolikheterna  $P(X = k)$  ges av  $p_X(k)$ .

## BINOMIALFÖRDELNING (FORTS.)

- ▶ *Binomialfördelning* uppträder också som *fördelning för summan* av oberoende lika fördelade s.v. !
- ▶ Antag att för  $n$  st. oberoende försök är sannolikheten att lyckas i varje försök lika med  $p$ . Associera med vart och ett av de  $n$  försöken en s.v.  $I_i$ ,  $i = 1, \dots, n$  sådan att

$$I_i = \begin{cases} 1 & \text{om försök nr. } i \text{ lyckas} \\ 0 & \text{annars} \end{cases}$$

- ▶ Från tidigare vet vi att  $I_i$  är oberoende, Bernoulli-fördelade s.v. med samma  $p$ , dvs  $I_i \in Be(p)$  (se def. 3.3, s. 51).
- ▶ Låt  $X$  vara antalet lyckade försök bland de  $n$  utförda,

$$X = I_1 + I_2 + \dots + I_n.$$

Då har vi att  $X = \sum_{i=1}^n I_i \in Bin(n, p)$ . (Observera att  $I_i \in Bin(1, p)$ .)



# BINOMIALFÖRDELNINGENS EGENSKAPER (FORTS.)

- ▶ Vi använder summa framställningen ovan för att bestämma väntevärde och varians för  $X \in \text{Bin}(n, p)$ .

$$E(X) = E\left(\sum_{i=1}^n I_i\right) = \sum_{i=1}^n E(I_i) = np,$$

$$V(X) = V\left(\sum_{i=1}^n I_i\right) = |\text{ober. } I_i| = \sum_{i=1}^n V(I_i) = np(1-p),$$

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

- ▶ **Sats:** *Additionssatsen för binomialfördelningar.* Låt  $X_1 \in \text{Bin}(n_1, p)$  och  $X_2 \in \text{Bin}(n_2, p)$  vara oberoende s.v. (Obs! Samma  $p$  för  $X_1$  och  $X_2$ ). Då gäller att

$$Y = X_1 + X_2 \in \text{Bin}(n_1 + n_2, p).$$



## NORMALAPPROXIMATION FÖR $X \in \text{Bin}(n, p)$ .

- ▶ Av  $X$  representation som summa,  $X = \sum_{i=1}^n I_i$ , följer att centrala gränsvärdessatsen (CGS) kan tillämpas!
- ▶ För stora  $n$  är  $X$  ungefär normalfördelad med det väntevärde och varians som ges ovan, dvs,  $X \in \text{AsN}(np, \sqrt{np(1-p)})$ . Detta vidare ger

$$\begin{aligned} P(a < X \leq b) &= |\text{standardisera}| \\ &= P\left(\frac{a - np}{\sqrt{np(1-p)}} < \frac{X - np}{\sqrt{np(1-p)}} \leq \frac{b - np}{\sqrt{np(1-p)}}\right) \\ &\approx \Phi\left(\frac{b - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - np}{\sqrt{np(1-p)}}\right). \end{aligned}$$

- ▶ Hur stort  $n$ ? Approximation ger bra noggrannhet om  $np(1-p) > 10$ . Detta används som tumregel.
- ▶ Exempel på tavlan.



- ▶ **Def:** Om en s.v.  $X$  har sannolikhetsfunktionen

$$p_X(k) = e^{-\mu} \frac{\mu^k}{k!}, \quad k = 0, 1, 2, \dots$$

där  $\mu > 0$  sägs  $X$  vara **Poissonfördelad**. Beteckning:  $X \in Po(\mu)$ .

- ▶ **Förekomst:**
  - ▶ Används för att modellera *sällsynta händelser*. Antag att  $X \in Bin(n, p)$  där **ant. oberoende försök  $n$  är stort och sannolikheten  $p$  att lyckas i varje försök är liten**. Betrakta  $\mu = np$  som är "lagom". Då ges antalet lyckade försök approximativt av en s.v. som är Poissonfördelad med  $\mu = np$ . Detta kallas ibland *små talens lag*. Approximationen är rimlig om  $p < 0.1$  och  $n > 10$ .



# EGENSKAPER HOS $Po(\mu)$ .

- ▶ Exakta egenskaper hos  $Po(\mu)$ .
  - ▶ **Sats:** Om  $X \in Po(\mu)$  gäller att

$$E(X) = \mu, \quad V(X) = \mu, \quad D(X) = \sqrt{\mu}.$$

- ▶ **Sats:** *Additionssatsen för Poissonfördelningar.* Låt  $X_1 \in Po(\mu_1)$  och  $X_2 \in Po(\mu_2)$  vara oberoende s.v. Då gäller att

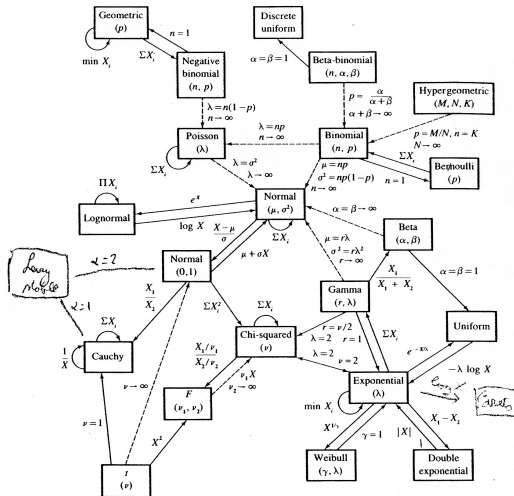
$$Y = X_1 + X_2 \in Po(\mu_1 + \mu_2).$$

- ▶ Approximativa egenskaper hos  $Po(\mu)$ .
  - ▶ Låt  $X \in Po(\mu)$  och låt  $\mu$  vara ett heltal. Då, enligt additionssats gäller att  $X = V_1 + \dots + V_\mu$  där  $V_i$  är oberoende  $Po(1)$ -fördelade s.v. Det följder nu av CGS att  $X \in AsN(\mu, \sqrt{\mu})$  då  $\mu$  är tillräckligt stort.
  - ▶ Approximation gäller även om  $\mu$  är inte heltal!
  - ▶ Tumregel: Approximation är ganska bra om  $\mu > 15$ .



# TABLE OF COMMON DISTRIBUTIONS

Table of Common Distributions



Relationships among common distributions. Solid lines represent transformations and special cases, dashed lines represent limits. Adapted from Leemis (1986).

