SF1914/SF1916: Sannolikhetsteori och statistik
Föreläsning 8
Centrala gränsvärdessatsen (CGS), normalfördelningsapproximationer. Binomial- och Poisson fördelningen: approximativa egenskaper.

Tatjana Pavlenko

18 september 2018

KTH Matematik

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 Possionfördelningen. (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 μ och σ, $(\sigma > 0)$ 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.

KTH Matematik

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 μ och σ: enligt sats får vi

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

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

  dvs är parametrarna μ och σ *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 $X_1, X_2, \ldots, X_n$ är oberoende och respektive $N(\mu_1, \sigma_1), N(\mu_2, \sigma_2), \ldots, N(\mu_n, \sigma_n)$ och konstanterna $a_1, a_2, \ldots, a_n, b$ 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 $X_1, X_2, \ldots, X_n$ är oberoende $N(\mu, \sigma)$ och $a_1 = a_2 = \cdots = a_n = 1/n$ samt $b = 0$, så gäller att

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

KTH Matematik

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 $X_1, \ldots, X_n, \ldots$ vara en oändlig följd av oberoende, likafördelade s.v. med väntevärde μ och och standardavvikelse $0 < \sigma < \infty$. Sätt

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

  Då gäller för givna $a < b$ att

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

- 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:

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

Centrala gränsvärdessatsen (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

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

KTH Matematik

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

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

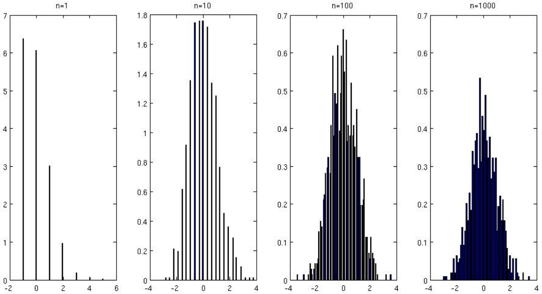
dvs *aritmetisk medelvärdet* $\bar{X}_n$ *är approximativt normalfördelat för tillräckligt stort* n.

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

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

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

CENTRALA GRÄNSVÄRDESSATSEN (FORTS.)



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

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

och $X_1, \ldots, X_n, \ldots$ är oberoende $Po(1)$-variabler (så att $\mu = \sigma = 1$). Då $n \to \infty$ liknar fördelningen alltmer den standardiserade normalfödelningstäthet.

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

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

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

- 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 $X \in Bin(n, p)$ och sannolikheterna $P(X = k)$ ges av $p_x(k)$.

- *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, $i = 1, \ldots, n$ sådan att

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

- 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,

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

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

Binomialfördelnings egenskaper (forts.)

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

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

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

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

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

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

NORMALAPPROXIMATION FÖR $X \in Bin(n, p)$.

- 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, $X \in AsN(np, \sqrt{np(1-p)})$. Detta vidare ger

$$P\left(a < X \leq b\right) = |\text{standardisera}|$$

$$= P\left(\frac{a - np}{\sqrt{np(1-p)}} < \frac{X - np}{\sqrt{np(1-p)}} \leq \frac{b - np}{\sqrt{np(1-p)}}\right)$$

$$\approx \Phi\left(\frac{b - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - np}{\sqrt{np(1-p)}}\right).$$

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

KTH Matematik

- Def: Om en s.v. X har sannolikhetsfunktionen

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

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 $n > 10$.

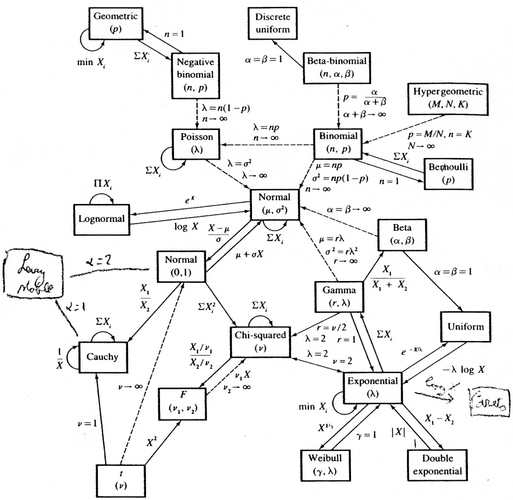
Egenskaper hos $Po(\mu)$.

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

KTH Matematik

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).