TDAB01 Sannolikhetslära och Statistik

Jose M. Peña
IDA, Linköpings Universitet

Föreläsning 3

Översikt

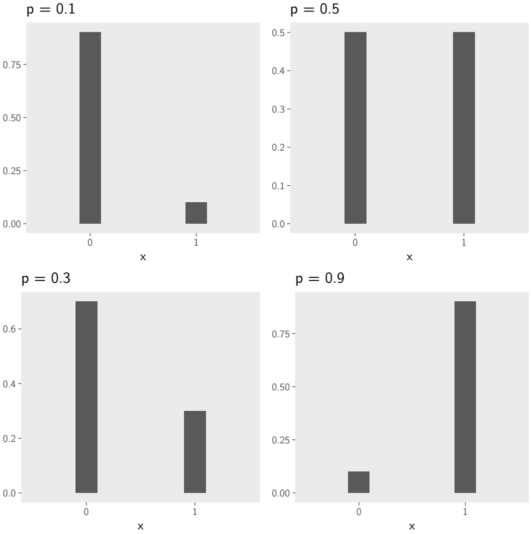
- **Fördelningsfamiljer för diskreta variabler**
- **Bernoulli, binomial, multinomial**
- **Geometrisk, negativ binomial**
- **Poisson**

Bernoullifördelningen

Definition. En **Bernoullivariabel** X kan anta två olika värden, 0 och 1.
Om X är **Bernoullifördelad**, dvs $X \sim \mathit{Bernoulli}(p)$, så gäller att
$\boldsymbol{P}(X = 1) = P(1) = p$ och $\boldsymbol{P}(X = 0) = P(0) = q = 1 - p$.

- Genom att ändra parametern p får vi en mängd olika sannolikhetsfördelningar på $\{0, 1\}$. En fördelningsfamilj är en mängd olika sannolikhetsfördelningar som indexeras med en eller flera parametrar. Några få fördelningsfamiljer räcker för att modellera en majoritet av experimenten.
- Se `ManipDistributions.R`.

Bernoullifördelningen



Bernoullifördelningen

- Pmf för $X \sim Bernoulli(p)$

$$P(x) = \begin{cases} q = 1 - p & \text{om x=0} \\ p & \text{om x=1} \end{cases}$$

- Om $X \sim Bernoulli(p)$

$$\mathbb{E}(X) = 0 \cdot q + 1 \cdot p = p$$
$$Var(X) = (0 - p)^2 \cdot q + (1 - p)^2 p = p - p^2 = p \cdot q$$

- En Benoullivariabel kallas också Bernoulliförsök.

Binomialfördelningen

Definition. Låt $X_1, X_2, \ldots, X_n$ vara en sekvens av n **oberoende** Bernoulliförsök med sannolikhet p. Låt X = **antalet lyckade försök i sekvensen**. Då är X **binomialfördelad** med parametrar n och p, dvs $X \sim Binomial(n,p)$ med pmf

$$P(x) = \binom{n}{x} p^x q^{n-x}$$

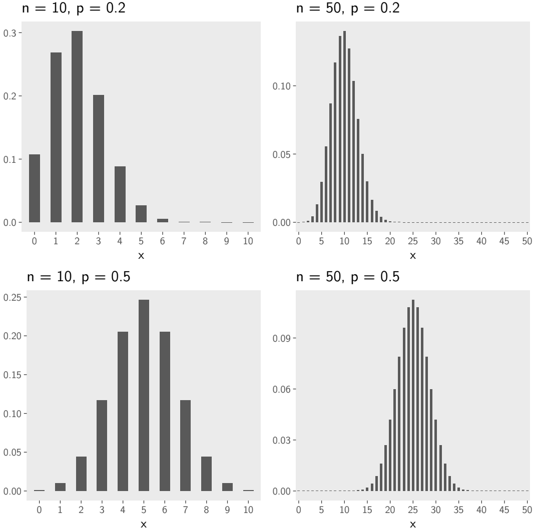
för $x = 0, 1, 2, \ldots, n$.

- $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ är antalet sekvenser av längd n med exakt x lyckade försök, så kallad **binomialkoefficienten**.
- Om t ex $n = 3$ och $x = 2$, så leder alla tre sekvenserna $(0,1,1)$, $(1,0,1)$ och $(1,1,0)$ till utfallet $x = 2$.
  - Sekvensen $(0,1,1)$ har sannolikheten $q \cdot p \cdot p = p^2 q$.
  - Sekvensen $(1,0,1)$ har sannolikheten $p \cdot q \cdot p = p^2 q$.
  - Sekvensen $(1,1,0)$ har sannolikheten $p \cdot p \cdot q = p^2 q$.
- Antalet misslyckade försök i sekvensen följer $Binomial(n,q)$.

Binomialfördelningen

- Binomialfördelningen passar data:
  - som är **diskreta icke-negativa heltal**.
  - som kan anta alla **heltal mellan** 0 **och** n.
- Passande: Hur många elever i klass 5A kan simma ?
- Inte passande: Hur många mål gör IFK Norrköping på lördag ? (pga ingen naturlig övre gräns) eller längdmätningar (kontinuerliga).
- Egenskaper för $X \sim Binomial(n, p)$
  - $\mathbb{E}(X) = n \cdot p$
  - $Var(X) = n \cdot p \cdot q$
- **Bevis**: $X \sim Binomial(n, p)$ innebär att $X = X_1 + X_2 + \cdots + X_n$, dvs X är en summa av n oberoende Bernoullivariabler med sannolikhet p. Dessutom, väntevärdet och variansen av en summa av oberoende variabler är summan av variablernas väntevärden och varianser.
- Se Example 3.17 i Baron.

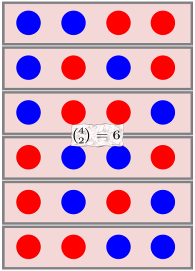
Binomialfördelningen



Multinomialfördelningen

- Bernoullidata: n personer utfrågas om vilket partiblock de föredrar (röd eller blå). n_1 personer svarar röd, n_2 personer svarar blå.
- Antal sätt vi kan få dessa data: $\binom{n}{n_1} = \frac{n!}{n_1!n_2!}$
- Sannolikheten för att få n_1 röda i n försök:

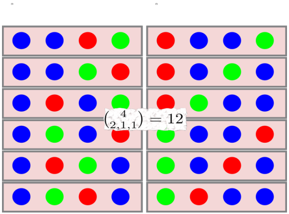
$$P(n_1) = \binom{n}{n_1} p^{n_1} q^{n_2}$$



Multinomialfördelningen

- Multinomiala data: n personer utfrågas om vilket partiblock de föredrar (röd, blå eller grön). n_1 personer svarar blå, n_2 personer svarar röd och n_3 personer svarar grön.
- Antal sätt vi kan få dessa data ges av **multinomialkoefficienten**: $\binom{n}{n_1\; n_2\; n_3} = \frac{n!}{n_1!n_2!n_3!}$ och

$$P(n_1, n_2, n_3) = \binom{n}{n_1\; n_2\; n_3} p_1^{n_1} p_2^{n_2} p_3^{n_3}$$



- Notera att multinomialfördelningen är en simultanfördelning för **tre** slumpvariabler: N_1, N_2 och N_3.

Geometriska fördelningen

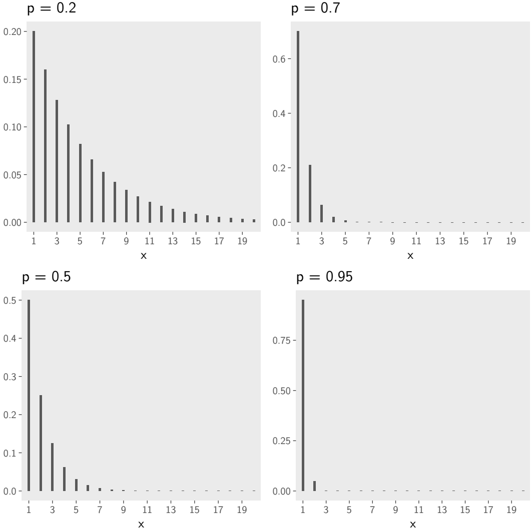
Definition. Låt $X_1, X_2, \ldots$ vara en sekvens av **oberoende** Bernoulliförsök med sannolikhet p. Låt X = **antalet Bernoulliförsök för att få ett lyckat försök**. Då är X **geometrisk fördelad**, dvs $X \sim Geo(p)$ med pmf

$$P(x) = (1-p)^{x-1}p$$

för $x = 1, 2, \ldots$

- Geometriska fördelningen passar data:
  - som antar **diskreta positiva heltal:** $1, 2, 3, \ldots$
  - som **inte har en övre gräns** (jfr binomial).
  - med **monotont avtagande pmf**.
- Egenskaper för $X \sim Geo(p)$
  - $\mathbb{E}(X) = 1/p$
  - $Var(X) = \frac{1-p}{p^2}$
  - Väntevärdet och variansen beräknas med hjälp av den geometriska serien.
- Slantsingling (lyckat=krona): $\mathbb{E}(X) = 2,\ Var(X) = 2$.
- Kasta tarning (lyckat=en prick): $\mathbb{E}(X) = 6,\ Var(X) = 30$.

Geometriska fördelningen



Negativa binomialfördelningen

Definition. Låt $X_1, X_2, \ldots$ vara en sekvens av **oberoende** Bernoulliförsök med sannolikhet p. Låt X = **antalet Bernoulliförsök för att få k lyckade försök**. Då är X **negativ binomialfördelad**, dvs $X \sim NegativBinomial(k, p)$ med pmf

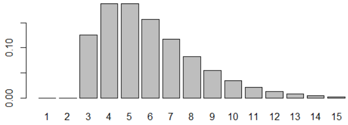
$$P(x) = \binom{x-1}{k-1}(1-p)^{x-k}p^k$$

för $x = 1, 2, \ldots$

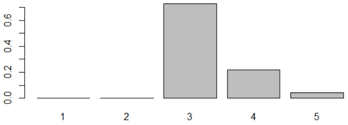
- $NegativBinomial(1, p) = Geo(p)$.
- Negativa binomialfördelningen är rak motsats till binomialfördelningen: Den sista modellerar hur många gånger man lyckas i en sekvens av n Bernoulliförsök, och den första modellerar antalet Benouilliförsök för att lyckas k gånger. Se Example 3.21 i Baron.
- Negativa binomialfördelningen passar data:
  - som antar **diskreta positiva heltal:** $1, 2, 3, \ldots$
  - som **inte har en övre gräns** (jfr binomial).
- Egenskaper för $X \sim NegativBinomial(k, p)$
  - $\mathbb{E}(X) = k/p$
  - $Var(X) = \frac{k(1-p)}{p^2}$
- **Bevis**: $X \sim NegativBinomial(k, p)$ innebär att X är en summa av k oberoende geometriskvariabler med sannolikhet p. Dessutom, väntevärdet och variansen av en summa av oberoende variabler är summan av variablernas väntevärden och varianser.

Negativa binomialfördelningen

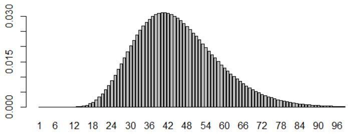
NegativBinomial(3,0.5)



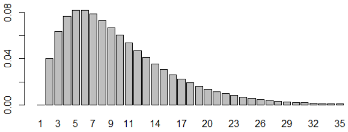
NegativBinomial(3,0.9)



NegativBinomial(9,0.2)



NegativBinomial(2,0.2)



Poissonfördelningen

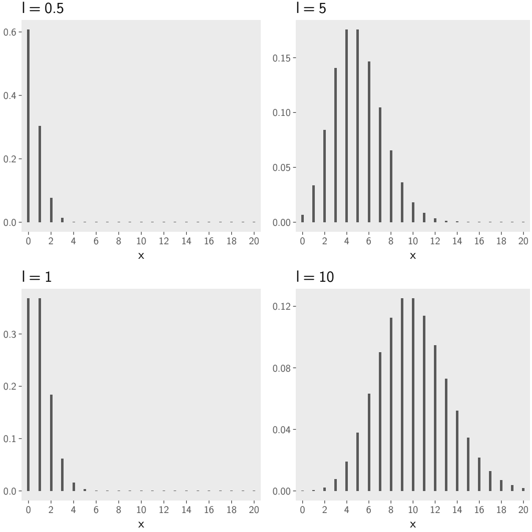
Definition. En **Poissonfördelad** slumpvariabel med frekvens λ, dvs $X \sim Po(\lambda)$, har pmf

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

för $x = 0, 1, 2, \ldots$

- Egenskaper för $X \sim Po(\lambda)$
  - $\mathbb{E}(X) = \lambda$
  - $Var(X) = \lambda$
  - Väntevärdet och variansen beräknas med Taylorutvecklingen.
- Poissonfördelningen passar data:
  - som antar **diskreta icke-negativa heltal:** $0, 1, 2, \ldots$
  - som **inte har en övre gräns** (jfr binomial).
  - vars väntevärde och varians är ungefär lika.
- Poissonfördelningen passar som modell av antalet ovanliga händelser i en tidsperiod, dvs osannolikt att flera händer samtidigt eller nära varandra i tiden. T ex
  - Antalet upptäckta buggar i en kod.
  - Antalet döda i trafiken under år 2014.
  - Se Example 3.22 i Baron.
- Poissonfördelningen med $\lambda = n \cdot p$ kan användas för att approximera binomialfördelningen när $n \geq 30$ and $p \leq 0.05$. Se `ManipDistributions.R`.

Poissonfördelningen



Poissonfördelningen

- Godtycklig modell att antal händelser i en tidsperiod ? Nej.
- Låt X_t vara händelser i tidsintervallet $[0, t]$. **Poisson postulat**:
  - $X_0 = 0$, dvs inga händelser i början.
  - $s < t \Rightarrow X_s$ och $X_t - X_s$ är oberoende, dvs oberoende antal för disjunkta tider.
  - X_s och $X_{t+s} - X_t$ har samma fördelning, dvs antal beror bara om längden av tidsintervallet.
  - $\lim_{t \to 0} \boldsymbol{P}(X_t = 1)/t = \lambda$, dvs sannolikhet proportionell till längden av tidsintervallet, för korta intervall.
  - $\lim_{t \to 0} \boldsymbol{P}(X_t > 1)/t = 0$, dvs ej samtidiga händelser.

Då $X_t \sim Po(\lambda t)$.

Översikt

- **Fördelningsfamiljer för diskreta variabler**
- **Bernoulli, binomial, multinomial**
- **Geometrisk, negativ binomial**
- **Poisson**