

Matematisk Statistik och Diskret Matematik, MVE051/MSG810, VT22

Timo Vilkas



Chalmers - Göteborgs Universitet

Föreläsning 4
31 mars 2022



Binomialfördelning

- n repetitioner av samma Bernoulli-experiment (n fast).
(utfallet av varje Bernoulli-experiment är antingen "succé" eller "misslyckande".)
- Repetitionerna är identiska och oberoende, dvs. sannolikheten av succé är alla gånger samma (man skriver p för denna)
- X räknar antalet lyckade försök bland alla n experiment. X antar då ett slumpvärde i mängden $\{0, \dots, n\}$.

Motsvarande sannolikhetsfunktion är

$$\mathbb{P}(X=k) =: f(k) = \binom{n}{k} (1-p)^{n-k} p^k$$

for $k = 0, \dots, n$.

Negativ Binomialfördelning

- Samma uppsättning med en följd av oberoende, identiska Bernoulli-experiment.
- Dock kör man nu inte ett förbestämt antal repetitioner, utan tills man har fått succé nummer r (där $r \in \mathbb{N}$ är givet).
- Slumpvariabeln X räknar nu **antalet försök** som behövs inklusive sista i vilket succé nummer r inträder.
- Observera att specialfallet $r = 1$ motsvarar den geometriska fördelningen $\text{Geom}(p)$.

- Sannolikhetsfunktionen till den negativa binomialfördelningen ges av

$$f(k) = \binom{k-1}{r-1} (1-p)^{k-r} p^r$$

för $r = 1, 2, 3, \dots$ samt $k = r, r+1, r+2, \dots$, och 0 annars.
 $p \in (0, 1)$ och $r \in \mathbb{N}$ är fördelningens parametrar.

- Observationen att X , som är negativ binomialfördelad med parametrar r och p , är en summa av r **oberoende** geometriska fördelade slumpvariabler ger oss direkt väntevärdet $\mathbb{E}[X] = \frac{r}{p}$ och variansen $\text{Var}[X] = \frac{rq}{p^2}$.

- Momentgenererande funktionen till X ges av

$$m_X(t) = \frac{(pe^t)^r}{(1 - qe^t)^r},$$

$X = X_1 + X_2 + \dots + X_r$
 $\sim \text{NB}(r, p)$ $\swarrow$ $\searrow$
 oberoende Geom(p)
 "man väntar r gånger på nästa succé"

där som vanligt $q = 1 - p$.

Navigation icons

Exempel

Vi rullar en symmetrisk tärning. Repetitionerna är identiska och oberoende.

- Vad är sannolikheten att vi får den första 6:an på tredje försök?
- Vad är sannolikheten att vi får den tredje 6:an på sjunde försök?
- Vad är väntevärdet och variansen av antalet försök som behövs för att ha fått en 6:a för tredje gången?

Navigation icons

Exempel (Lösning)

① X , antalet försök till första 6:a är $\text{Geom}(\frac{1}{6})$ -fördelad

$$\Rightarrow P(X=3) = (1-\frac{1}{6})^2 \cdot (\frac{1}{6}) = \frac{25}{216} \approx 0.11574$$

② Y , antalet försök till tredje 6:a är $\text{NB}(3, \frac{1}{6})$ -fördelad

(dvs. neg. binomial med parametrar $r=3, p=\frac{1}{6}$)

$$\Rightarrow P(Y=7) = \binom{6}{2} (1-\frac{1}{6})^4 \cdot (\frac{1}{6})^3 \approx 0.0349$$

③ $Y \sim \text{NB}(3, \frac{1}{6})$ ges $E(Y) = \frac{r}{p} = \frac{3}{\frac{1}{6}} = 18$ och

$$\text{var}(Y) = \frac{r}{p^2} = \frac{3 \cdot (\frac{5}{6})}{(\frac{1}{6})^2} = 90.$$

Hypergeometrisk fördelning

- Vi drar n objekt ur en mängd med $N \geq n$ objekt **utan återläggning**.
- Det finns en särskild egenskap som $r \leq N$ utav dessa objekt har (och resterande $N - r$ inte har).
- Låt slumpvariabeln X räkna antalet objekt (bland de n dragna) som har egenskapen.

OBS: Skillnaden mellan hypergeometrisk och binomialfördelning är att vi här inte lägger tillbaka det som drogs och på så sätt olika ronder varken är identiska eller oberoende.

T.ex. en urna med totalt N bollar, r röda och $N-r$ svarta.

Vi drar $n \leq N$ stycken och lägger dem bredvid urnan, inte tillbaka.

X räknar antalet röda bollar vi drog.

- Sannolikhetsfunktionen till den hypergeometrisk fördelningen ges av

$$f(k) = \frac{\binom{r}{k} \binom{N-r}{n-k}}{\binom{N}{n}}$$

för $\max[0, n - (N - r)] \leq k \leq \min(n, r)$, och 0 annars.
 N , r och n är fördelningens parametrar.

- En lite mödosam beräkning ger att X som är hypergeometrisk fördelad har väntevärde $\mathbb{E}[X] = \frac{nr}{N}$ och varians $\text{Var}[X] = n \left(\frac{r}{N} \right) \left(\frac{N-r}{N} \right) \left(\frac{N-n}{N-1} \right) = \frac{nr(N-r)(N-n)}{N^2(N-1)}$.

Exempel

Man har testat en grupp med 20 personer på SARS-CoV-2, och 5 utav dem hade ett positivt testresultat. Låt oss slumpmässigt välja ut 6 personer ur gruppen och X räkna antalet bland de utvalda som har viruset.

- 1 Bestäm sannolikhetsfunktionen till X .
- 2 Beräkna väntevärde $\mathbb{E}[X]$ och varians $\text{Var}[X]$.
- 3 Beräkna sannolikheten att maximalt en person av de utplockade har sjukdomen.

Exempel (Lösning)

- 1 X är hypergeometrisk fördelad med parametrar $N=20, r=5$ och $n=6$.

$$P(X=k) = \frac{\binom{5}{k} \binom{20-5}{6-k}}{\binom{20}{6}} \quad \text{för } 0 = \max\{0, 6-(20-5)\} \leq k \leq \min\{5, 6\} = 5,$$

0 annars.
- 2 $E(X) = \frac{n \cdot r}{N} = \frac{6 \cdot 5}{20} = \frac{3}{2}$ och $\text{var}(X) = \frac{n r (N-r) (N-1)}{(N-1) N^2}$

$$= \frac{6 \cdot 5 \cdot 14 \cdot 15}{19 \cdot 20^2} = \frac{63}{76} \approx 0.829$$
- 3 $P(X \leq 1) = P(X=0) + P(X=1)$

$$= \frac{\binom{5}{0} \binom{15}{6}}{\binom{20}{6}} + \frac{\binom{5}{1} \binom{15}{5}}{\binom{20}{6}} = \frac{1001}{1938} \approx 0.5165$$

Navigation icons

T. Vilkas

Matematisk statistik & diskret matematik, MVE051/MSG810

Poisson fördelning

- Poisson fördelningen är en diskret sannolikhetsfördelning som räknar antalet händelser som inträffar, oberoende av varandra, inom ett givet tidsintervall, när det genomsnittliga antalet per tidsperiod är känt.
- Sannolikhetsfunktionen till Poisson fördelningen är

$$f(k) = \frac{e^{-\lambda} \lambda^k}{k!} \quad \left[\begin{array}{l} \text{kom ihåg} \\ e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \end{array} \right]$$

för $k = 0, 1, 2, \dots$ (och fast $\lambda > 0$). λ är fördelningens parameter (som också kallas intensitet) och motsvarar det genomsnittliga antalet händelser per tidsenhet. Man skriver kort $X \sim \text{Poiss}(\lambda)$.

Navigation icons

T. Vilkas

Matematisk statistik & diskret matematik, MVE051/MSG810

- För $X \sim \text{Poiss}(\lambda)$ är $\mathbb{E}[X] = \text{Var}[X] = \lambda$.
- Momentgenererande funktionen till X är

$$m_X(t) = e^{\lambda(e^t - 1)}$$

- Om $X_1 \sim \text{Poiss}(\lambda_1)$ och $X_2 \sim \text{Poiss}(\lambda_2)$ är oberoende, så är $X_1 + X_2 \sim \text{Poiss}(\lambda_1 + \lambda_2)$.

Fördelingsfunktion: se tabell II (s. 692) i boken.

Exempel

Låt X vara antalet tryckfel på en boksida med i genomsnitt 3 fel per sida. Antag att tryckfel händer oberoende av varandra.

- 1 Beräkna sannolikheten att en slumpmässigt utvald sida innehåller minst ett tryckfel.
- 2 Vad är sannolikheten att tre slumpmässigt utvalda sidor innehåller fler än 8 tryckfel?

Låt X vara antalet tryckfel på en slumpm. utvald boksida, då gäller $X \sim \text{Poiss}(3)$, dvs. poissonfördelat med parameter $\lambda = 3$ och

$$\textcircled{1} \quad \mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - e^{-3} \cdot \frac{3^0}{0!} = 1 - e^{-3} \approx 0.95$$

$$\textcircled{2} \quad Y \text{ antalet tryckfel på 3 sidor} \rightarrow Y \sim \text{Poiss}(9)$$

$$\mathbb{P}(Y > 8) = 1 - \mathbb{P}(Y \leq 8) \approx 1 - 0.593 = 0.407$$

tabell II

Gamma fördelning

- Gamma funktion: $\Gamma(\alpha) = \int_0^\infty z^{\alpha-1} e^{-z} dz$.
- $\Gamma(1) = 1$, $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$. $\Rightarrow \Gamma(n) = (n-1) \cdot \Gamma(n-1) = \dots = (n-1)!$
- Slumpvariabeln X med täthetsfunktion för $n \in \mathbb{N}$.

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \quad x > 0,$$

där $\beta > 0$ och $\alpha > 0$, sägs vara gamma fördelad med parametrar α och β .

- Om X är gamma fördelad med parametrar α och β , så är $m_X(t) = (1 - \beta t)^{-\alpha}$ den momentgenererande funktionen och $\mathbb{E}[X] = \alpha\beta$ samt $\text{Var}[X] = \alpha\beta^2$.

Exponentialfördelning

- Specialfallet av en gamma fördelning med parameter $\alpha = 1$ kallas **exponentialfördelning**.
- Motsvarande täthetsfunktion ges då av

$$f(x) = \frac{1}{\beta} e^{-x/\beta} \quad x > 0,$$

eller ekvivalent (och mer standard) $f(x) = \lambda e^{-\lambda x}$, där $\lambda = \frac{1}{\beta} > 0$. Man skriver $X \sim \text{Exp}(\lambda)$.

- $\mathbb{E}[X] = \frac{1}{\lambda}$ och $\text{Var}[X] = \frac{1}{\lambda^2}$.
- Fördelningsfunktionen till exponentialfördelningen är

$$F(x) = 1 - e^{-\lambda x}. \quad \Rightarrow \quad \mathbb{P}(X \geq x) = 1 - F(x) = e^{-\lambda x}$$

Exp. fördelningen är minneslös: $\mathbb{P}(X \geq s+t \mid X \geq s) = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = \mathbb{P}(X \geq t)$

χ^2 -fördelning

- En slumpvariabel X kallas χ^2 -fördelad med γ frihetsgrader ($\gamma \in \mathbb{N}$) om den har en gamma fördelning med parametrar $\beta = 2$ och $\alpha = \frac{\gamma}{2}$.
- $\mathbb{E}[X] = \gamma$ och $\text{Var}[X] = 2\gamma$.
- Tabell IV på s.695-696 ger numeriska värden av fördelningsfunktionen till χ^2 -fördelningen med olika frihetsgrader. Observera dock att den är lite annorlunda då sannolikheter står i huvudraden (inte som i de andra tabeller i boken där de står i själva tabellen).
- Det som gör fördelningen ganska viktig i praktiken är följande: Om $X_1, \dots, X_n$ är oberoende standard normalfördelade slumpvariabler, så har $X_1^2 + \dots + X_n^2$ en χ^2 -fördelning med n frihetsgrader.

Det finns en intressant koppling mellan Poisson och exponentialfördelning: Om man har en såkallad Poisson process (som registrerar och räknar, t.ex. på tidslinjen, alla händelser som sker i ett Poisson experiment - och antalet händelser inom en given tidsenhet är då $\text{Poiss}(\lambda)$ fördelat), så är väntetiden mellan händelser exponentialfördelad med parameter λ :

Sats

Given en Poisson process med intensitet λ (alltså där antalet händelser per tidsenhet är $\text{Poiss}(\lambda)$ fördelat), låt W beteckna väntetiden tills första händelsen inträffar. Då är W exponentialfördelad med parameter λ .

← alltså $\text{Exp}(\lambda)$ i vår notation

Exempel

Låt oss anta att antalet studenter som besöker kurshemsidan kan beskrivas som en Poisson process med genomsnittligt 30 studenter per timme. Om man går in i hemsidans statistik nu, vad är sannolikheten att det dröjer minst 4 minuter tills nästa student ökar antalet besökare?

Låt X vara antalet besökare per minut, då gäller $X \sim \text{Poiss}(0.5)$, alltså Poisson fördelat med parameter $\lambda = \frac{30}{60} = 0.5$ (intensitet, per minut)

Räkneprocessen som räknar antalet besökare (på tidslinjen) är då en Poisson-process med intensitet 0.5.

Enligt satsen ovan är väntetiden W på första händelsen (eller nästa efter en fix tidpunkt: minnestöshet!) exponentialfördelat med parameter

$$\lambda = \frac{1}{2}. \Rightarrow P(W > 4) = e^{-\frac{1}{2} \cdot 4} = e^{-2} \approx 0.135.$$

Vi har lärt känna följande fördelningar:

Diskreta fördelningar:

- ① geometrisk, $\text{Geom}(p)$
- ② binomial, $\text{Bin}(n, p)$
- ③ negativ binomial, $\text{NB}(r, p)$
- ④ hypergeometrisk, med parametrar $(N, r; n)$
- ⑤ Poisson, $\text{Poiss}(\lambda)$

Kontinuerliga fördelningar

- ⑥ normal, $\mathcal{N}(\mu, \sigma^2)$
- ⑦ gamma, $\Gamma(\alpha, \beta)$
- ⑧ exponential, $\text{Exp}(\lambda)$
- ⑨ χ^2 , med γ frihetsgrader