

Vi vet att $Bin(n, p) \approx Po(np)$ om $p \leq 0.1$. Vi ska här visa följande feluppskattning för denna approximation

Om X är $Bin(n, p)$ och Y är $Po(np)$ så gäller att $|P(X \in A) - P(Y \in A)| \leq np^2$ där A är en godtycklig mängd av icke-negativa heltal.

(Specialfallet $A = \{k\}$ ger $|P(X = k) - P(Y = k)| \leq np^2$, $k = 0, 1, 2, \dots$).

Bevis: Man gör följande listiga konstruktion vars idé är att man konstruerar en tvådimensionell fördelning vars marginalfördelningar är $Bin(1, p)$ respektive $Po(p)$.

Låt därför (U, V) ha fördelningen

$$P(U = j; V = k) = \begin{cases} 1 - p & \text{om } j = 0 \text{ och } k = 0 \\ e^{-p} - 1 + p & \text{om } j = 1 \text{ och } k = 0 \\ \frac{p^k}{k!} e^{-p} & \text{om } j = 1 \text{ och } k = 1, 2, 3, \dots \\ 0 & \text{för övrigt} \end{cases}$$

Man kan notera att $P(U = 0) = P(U = 0; V = 0) = 1 - p$ och att

$$P(U = 1) = \sum_{k=0}^{\infty} P(U = 1; V = k) = e^{-p} - 1 + p + \sum_{k=1}^{\infty} \frac{p^k}{k!} e^{-p} = p,$$

(eller lättare genom att notera att $P(U = 1) = 1 - P(U = 0) = 1 - (1 - p)$) samt att $P(V = 0) = P(U = 0; V = 0) + P(U = 1; V = 0) = 1 - p + e^{-p} - 1 + p = e^{-p}$. Vidare är $P(V = k) = \frac{p^k}{k!} e^{-p}$, $k = 1, 2, 3, \dots$. Vi har alltså att U är $Bin(1, p)$ och V är $Po(p)$.

Notera att $P(U \neq V) = 1 - P(U = V) = 1 - P(U = 0; V = 0) - P(U = 1; V = 1) = 1 - (1 - p) - pe^{-p} = p(1 - e^{-p}) \leq p^2$ där den sista olikheten erhålls ur $1 - e^{-p} \leq p$ för $p > 0$.

Det fina med ovanstående konstruktion är alltså att U och V har rätta marginalfördelningar och att de nästan alltid är lika! Sannolikheten att de skiljer sig åt är ju högst p^2 .

Vi låter nu $(U_1, V_1), (U_2, V_2), \dots, (U_n, V_n)$, vara oberoende likafördelade två-dimensionella stokastiska variabler som alla har ovanstående fördelning. Eftersom summor av oberoende binomialfördelade (med samma p) är binomialfördelade och summor av oberoende Poissonfördelade är Poissonfördelade får vi att $X = \sum_{i=1}^n U_i$ är $Bin(n, p)$ och att $Y = \sum_{i=1}^n V_i$ är $Po(np)$.

Vidare är

$$\{X \neq Y\} \subseteq \bigcup_{i=1}^n \{U_i \neq V_i\}$$

(i ord : för att X och Y skall skilja sig åt måste U_i och V_i skilja sig för något i) och $P(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n P(A_i)$. Denna olikhet fås med upprepad användning av Boole's olikhet som innebär att $P(A \cup B) \leq P(A) + P(B)$ - rita figur! Detta ger

$$P(X \neq Y) \leq \sum_{i=1}^n P(U_i \neq V_i) \leq \sum_{i=1}^n p^2 = np^2$$

Vi har nu

$$\begin{aligned} & |P(X \in A) - P(Y \in A)| = \\ & |(P(X \in A; Y \in A) + P(X \in A; Y \notin A)) - (P(X \in A; Y \in A) + P(X \notin A; Y \in A))| = \\ & = |P(X \in A; Y \notin A) - P(X \notin A; Y \in A)| \end{aligned}$$

Båda dessa sannolikheter är $\leq P(X \neq Y)$ och alltså är $|\text{skillnaden}| \leq P(X \neq Y) \leq np^2$ vilket var vad vi skulle bevisa.

Om man granskar ovanstående bevis noterar man att man på precis samma sätt kan visa att om $P(U_i = 1) = p_i$, $P(U_i = 0) = 1 - p_i$ (- detta kallas för $Be(p_i)$ -fördelningen (Bernoulli-fördelningen)) så gäller att med $X = \sum_{i=1}^n U_i$ och Y $Po(\sum_{i=1}^n p_i)$ -fördelad så kan man visa att $|P(X \in A) - P(Y \in A)| \leq \sum_{i=1}^n p_i^2$. Detta kan ses som en förklaring till Poissonfördelningen dyker upp empiriskt i så många praktiska sammanhang. Så fort man har summor av 0-1-variabler som är oberoende och där de allra flesta är 0:or (dvs att p_i är litet) så blir summan approximativt Poissonfördelad. Man kan t o m visa att detta gäller då 0-1-variablerna är svagt beroende.

Den ovanstående bevistekniken är ett modernt påfund och knepet att skapa en lämplig tvådimensionell fördelning kallas koppling.