

På förra föreläsningen visade vi centrala gränsvärdessatsen som säger att en oändlig följd $X_1, X_2, \dots$ av oberoende, likafördelade s.v. med väntevärde μ och standardavvikelse σ uppfyller

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

vilket alltså gäller även om de s.v. X_i inte är normalfördelade. Vi kommer att undersöka ytterligare några approximationer idag, men börjar med att repetera binomialfördelningen. Betrakta ett försök där en händelse A inträffar med sannolikhet p . Om försöket upprepas n gånger och vi låter X beteckna antalet gånger som A inträffar, så är $X \in \text{Bin}(n, p)$. Vi kan också härleda binomialfördelningen på följande sätt. Till vart och ett av de n försöken kan vi koppla en Bernoullifördelad s.v. I_i som är ett om A inträffar och noll annars. Antalet gånger som A inträffar på n försök kan då skrivas

$$X = I_1 + \dots + I_n \in \text{Bin}(n, p).$$

Notera att de s.v. X_i är oberoende och likafördelade med väntevärde $E(I_i) = p \cdot 1 + (1-p) \cdot 0 = p$ och varians $V(I_i) = E(I_i^2) - (E(I_i))^2 = p \cdot 1^2 + (1-p) \cdot 0^2 - p^2 = p - p^2 = p(1-p)$. Väntevärdet och variansen av binomialfördelade s.v. blir därmed

$$E(X) = E(I_1 + \dots + I_n) = E(I_1) + \dots + E(I_n) = nE(I_1) = np$$

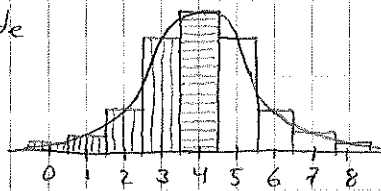
$$V(X) = V(I_1 + \dots + I_n) \stackrel{I_i \text{ oberoende}}{=} V(I_1) + \dots + V(I_n) = nV(I_1) = np(1-p)$$

Genom att betrakta binomialfördelade s.v. som summor av oberoende $\text{Ber}(p)$ -fördelade s.v., så är det uppenbart att om $X_1 \in \text{Bin}(n_1, p)$ och $X_2 \in \text{Bin}(n_2, p)$, så gäller $X_1 + X_2 \in \text{Bin}(n_1 + n_2, p)$. Med detta synsätt följer också att vi kan använda centrala gränsvärdessatsen på $X \in \text{Bin}(n, p)$.

Sats: Om $X \in \text{Bin}(n, p)$ och $np(1-p) \geq 10$, så gäller approximativt att $X \in \text{N}(np, \sqrt{np(1-p)})$, dvs

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

Vi approximerar här en diskret s.v. (för vilken $IP(X \leq k)$ är skilt från $IP(X < k)$ för alla $k = 0, 1, \dots, n$) med en kontinuerlig s.v. (för vilken $IP(X \leq k) = IP(X < k)$ för alla $k = 0, 1, \dots, n$). Approximationen förbättras om vi använder så kallad halukorrektion. Sannolikheten $IP(X \leq 4)$ motsvarar arean av de vertikalt streckade staplarna i figuren. Denna area är nästan lika stor som arean under normalfördelningens täthetsfunktion på intervallet $(-\infty, 3.5)$. På samma sätt är $IP(X \leq 4)$ lika med summan av areorna av de horisontellt och vertikalt streckade staplarna, dvs ungefär lika med arean under normalfördelningens täthetsfunktion på $(-\infty, 4.5)$. Med halukorrektion blir approximationen god för $np(1-p) \geq 3$.



Ex: En tärning kastas 60 gånger och X får beteckna antalet gånger som vi får en sexa. Vi vill bestämma $IP(X > 13)$. Eftersom $X \in \text{Bin}(60, 1/6)$, så är den sökta sannolikheten $\sum_{k=14}^{60} \binom{60}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{60-k} = 0,115$. Med normalapproximation får vi (notera att $np(1-p) = 8,33$ i detta exempel)

$$IP(X > 13) = 1 - IP(X \leq 13) \approx 1 - \Phi\left(\frac{13 - 60 \cdot \frac{1}{6}}{\sqrt{60 \cdot \frac{1}{6} \cdot \frac{5}{6}}}\right) = 0,149$$

$$\text{och med halukorrektion } IP(X > 13) = 1 - IP(X \leq 13) \approx 1 - \Phi\left(\frac{13 + \frac{1}{2} - 60 \cdot \frac{1}{6}}{\sqrt{60 \cdot \frac{1}{6} \cdot \frac{5}{6}}}\right) = 0,113.$$

Vi har tidigare sett att när n är stort och p litet, så är $X \in \text{Bin}(n, p)$ approximativt Poissonfördelat med parameter $\mu = np$. Approximationen är god för $p \leq 0,1$ (faktiskt oavsett n). Vi har tidigare visat att för $X \in \text{Po}(\mu)$, så gäller $E(X) = \mu$. För att bestämma $V(X)$, så beräknar vi först $E(X(X-1))$.

$$E(X(X-1)) = \sum_{k=0}^{\infty} k(k-1) \frac{\mu^k}{k!} e^{-\mu} = \mu^2 \sum_{k=2}^{\infty} \frac{\mu^{k-2}}{(k-2)!} e^{-\mu} = \mu^2 \sum_{l=0}^{\infty} \frac{\mu^l}{l!} e^{-\mu} = \mu^2$$

Termerna för $k=0$ och $k=1$ är noll. $\sum_{l=0}^{\infty} p_X(l) = 1$

och får sedan $V(X) = E(X^2) - (E(X))^2 = E(X(X-1)) + E(X) - (E(X))^2 = \mu^2 + \mu - \mu^2 = \mu$

Sats: Låt $X_1 \in \text{Po}(\mu_1)$ och $X_2 \in \text{Po}(\mu_2)$ vara oberoende. Då är $X_1 + X_2 \in \text{Po}(\mu_1 + \mu_2)$.

Bevis:
$$P(X_1 + X_2 = k) = \sum_{i=0}^k P(X_1 = i) P(X_2 = k-i) = \sum_{i=0}^k \frac{\mu_1^i}{i!} e^{-\mu_1} \frac{\mu_2^{k-i}}{(k-i)!} e^{-\mu_2}$$

$$= e^{-(\mu_1 + \mu_2)} \frac{1}{k!} \sum_{i=0}^k \frac{k!}{i!(k-i)!} \mu_1^i \mu_2^{k-i} = \frac{(\mu_1 + \mu_2)^k}{k!} e^{-(\mu_1 + \mu_2)}$$

$= (\mu_1 + \mu_2)^k$

Additionssatsen implicerar att om μ är ett heltal, så kan $X \in \text{Po}(\mu)$ skrivas som $X = Z_1 + Z_2 + \dots + Z_\mu$ där Z_i är oberoende $\text{Po}(1)$ -fördelade s.v. Eftersom $E(Z_i) = V(Z_i) = 1$, så följer det av centrala gränsvärdesatsen att X är approximativt $N(\mu, \sqrt{\mu})$ -fördelat för stora μ . Approximationen fungerar även när μ inte är ett heltal och är god för $\mu \geq 15$. Precis som vid normalapproximation av binomialfördelade s.v., så förbättras approximationen med halukorrektion.

Vi avslutar med att undersöka den hypergeometrisk fördelningen och återvänder därför till urnmodellerna som vi studerat tidigare. Antag att en urna innehåller s svarta och v vita kulor. Låt X beteckna antalet vita kulor som fås vid dragning av n kulor utan återläggning (Notera att med återläggning så gäller $X \in \text{Bin}(n, \frac{v}{s+v})$). Vi vet sedan tidigare att

$$P(X=k) = \frac{\binom{v}{k} \binom{s}{n-k}}{\binom{s+v}{n}}$$

Om vi nu betecknar totala antalet kulor $(s+v)$ med N och proportionen av vita kulor $(v/s+v)$ med p , så får vi, för $0 \leq k \leq Np$ och $0 \leq n-k \leq N(1-p)$, att

$$P(X=k) = \frac{\binom{Np}{k} \binom{N(1-p)}{n-k}}{\binom{N}{n}}$$

Vi säger då att X är en hypergeometrisk s.v. Detta betecknas $X \in \text{Hyp}(N, n, p)$. Precis som för binomialfördelade s.v., så kan vi skriva $X = I_1 + \dots + I_n$, där $I_i \in \text{Be}(p)$, eftersom vid var och en av de n dragningarna så är sannolikheten att få en vit kula p . I motsats till det binomialfördelade fallet, så är I_i na här beroende, eftersom dragningen sker utan återläggning (vad vi drar i omgång j påverkas av vad vi drog i omgång $1, 2, \dots, j-1$). Kovariansen är alltså nollskild och i boken visas att för $X \in \text{Hyp}(N, n, p)$, så gäller:

$$E(X) = np \quad \text{och} \quad V(X) = \frac{N-n}{N-1} np(1-p)$$

Korrektionsfaktor d_n^2 för ändliga populationer

Om $n \ll N$, så spelar inte återläggningen så stor roll och $d_n^2 \approx 1$. Mer specifikt gäller att om " $N \leq 0,1$ ", så är $X \in \text{Hyp}(N, n, p)$ approximativt $\text{Bin}(n, p)$ -fördelat. Om dessutom $np(1-p) \geq 10$, så är $X \in \text{Hyp}(N, n, p)$ approximativt $N(np, \sqrt{\frac{N-n}{N-1} np(1-p)})$ -fördelat. Notera att centrala gränsvärdesatsen här gäller trots att de s.v. I_i inte är oberoende.