Kapitel 7

Samplingfördelningar och Centrala gränsvärdessatsen

Statistikor och samplingfördelningar

I Kapitel 6 studerades metoder för att bestämma sannolikhetsfördelningen för funktioner av slumpvariabler.

Nu ska vi studera funktioner

$$U = h(X_1, X_2, \ldots, X_n)$$

där $X_1, X_2, \ldots, X_n$ utgör ett *slumpmässigt stickprov*, dvs de är *oberoende och likafördelade slumpvariabler (olfsv)*.

En sådan funktion av stickprovet kallas för en *statistika* (plural *statistikor*). En statistika är själv en slumpvariabel och dess sannolikhetsfördelning kallas för en *samplingfördelning*.

Statistikor och samplingfördelningar

Statistikor kan användas för att dra slutsatser (inferens) om populationens parametrar.

De två vanligaste statistikorna är

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \quad \text{och} \quad S^2 = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{n-1}$$

som ofta används vid inferens angående populations-medelvärdet μ och populationsvariansen σ^2.

Eftersom statistikor används vid inferens är det av stort intresse att finna deras samplingfördelningar.

UPPSALA UNIVERSITET

Samplingfördelningar vid normalfördelad population ($\overline{X}$)

Sats. Låt $X_1, X_2, \ldots, X_n$ vara oberoende och likafördelade $N(\mu,\sigma)$.

Då gäller att

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \quad \text{är} \quad N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

En direkt följd av detta är att

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

UPPSALA UNIVERSITET

Samplingfördelningar vid normalfördelad population (S^2)

Sats. Låt $X_1, X_2, \ldots, X_n$ vara oberoende och likafördelade $N(\mu,\sigma)$. Då gäller att

$$\mathbf{1}.\ \overline{X} \text{ och } S^2 \text{ är oberoende}$$

$$\mathbf{2}.\ U = \frac{(n-1)\,S^2}{\sigma^2} \quad \text{är} \quad \chi^2\,(n-1)$$

Samplingfördelningar vid normalfördelad population (S^2)

Bevis. Vi accepterar **1**. här. (För bevis se uppgift 13.94**.) För att bevisa **2**. behövs följande resultat från Kapitel 6 (se föreläsning Bild 9-11, 35).

$$Y = \sum_{i=1}^{n} Z_i^2$$

Det visar sig att Y uppvisar stora likheter med U.

$$U = \frac{(n-1)\,S^2}{\sigma^2}$$

För att finna sannolikhetsfördelningen för U kan vi alltså utgå från Y och sedan (på något sätt) ”byta ut” μ mot $\bar{X}$.

Samplingfördelningar vid normalfördelad population (S^2)

Bevis (forts). Det följer därmed att

$$Y = \sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2$$

Samplingfördelningar vid normalfördelad population (S^2)

<u>Bevis</u> (forts). Av Punkt **1**. följer att Z^2 och U är oberoende vilket innebär att vi med momentgenererande funktioner får att

$$m_Y(t) = m_U(t)\, m_{Z^2}(t)$$

Eftersom Y är $\chi^2(n)$ och Z^2 är $\chi^2(1)$ följer därmed att

$$(1-2t)^{-n/2} = m_U(t)\,(1-2t)^{-1/2}$$

dvs

$$m_U(t) = \frac{(1-2t)^{-n/2}}{(1-2t)^{-1/2}}$$

vilket är momentgenererande funktion för $\chi^2(n-1)$.

Samplingfördelningar vid normalfördelad population ($\overline{X}$)

Vi ska i Kapitel 8 se att eftersom

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \quad \text{är} \quad N(0,1)$$

är *Z* lämplig att använda vid inferens angående μ.

Problemet med *Z* är att populationsstandardavvikelsen σ ofta är okänd.

Vi söker därför en möjlighet att i detta uttryck ”byta ut” σ mot stickprovsstandardavvikelsen S. Frågan är hur detta påverkar samplingfördelningen?

t-fördelningen (definition)

Låt *Z* och *W* vara oberoende slumpvariabler där *Z* är *N(0,1)* och *W* är $\chi^2(\nu)$. Då är

$$T = \frac{Z}{\sqrt{W/\nu}}$$

t-fördelad med ν *frihetsgrader*. Vi skriver *t*(ν) eller t_ν.

t-fördelningen (användning)

Eftersom $\overline{X}$ och S^2 är oberoende samtidigt som att det vid normalfördelad population gäller

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \text{ är } N(0, 1)$$

får vi att

$$T = \frac{\frac{\overline{X} - \mu}{\sigma / \sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2} / (n-1)}}$$

Vi har lyckats ersätta σ med S och fått en statistika som följer en t-fördelning med n-1 frihetsgrader.

F-fördelningen (definition)

Låt W_1 och W_2 vara oberoende $\chi^2(\nu_1)$ respektive $\chi^2(\nu_2)$. Då är

$$F = \frac{W_1/\nu_1}{W_2/\nu_2}$$

F-fördelad med v_1 *frihetsgrader i täljaren* och v_2 *frihetsgrader i nämnaren*. Vi skriver $F(v_1,v_2)$ eller $F_{v1,v2}$.

F-fördelningen (användning)

Betrakta två oberoende stickprov om n_1 respektive n_2 observationer från två normalfördelade populationer.

$$W_1 = \frac{(n_1 - 1)\,S_1^2}{\sigma_1^2} \quad \text{och} \quad W_2 = \frac{(n_2 - 1)\,S_2^2}{\sigma_2^2}$$

W_1 och W_2 är då oberoende $\chi^2(n_1-1)$ respektive $\chi^2(n_2-2)$.

$$F = \frac{W_1/\nu_1}{W_2/\nu_2}$$

F följer en F-fördelning med n_1-1 frihetsgrader i täljaren och n_2-1 frihetsgrader i nämnaren.

UPPSALA UNIVERSITET

Centrala gränsvärdessatsen

Sats. Låt $X_1, X_2, \ldots, X_n$ vara oberoende och likafördelade slumpvariabler med väntevärde μ och standardavvikelse σ.

Om vi för fixt n låter

$$U_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

följer att

$$U_n \xrightarrow{d} Z \quad \text{där } Z \text{ är } N(0,1)$$

Vi säger att U_n *konvergerar i fördelning* mot Z med innebörden att fördelningsfunktionen för U_n konvergerar mot fördelningsfunktionen för N(0,1).
Alternativt sägs U_n vara *asymptotiskt normalfördelad*.

Uppgift 7.104

Låt Y_n vara $Bi(n,p_n)$ för $n=1,2,3,\ldots$, där $\lambda=np_n$ förutsätts vara en konstant för alla n. Då gäller att

$$M_{Y_n}(t)$$

som innebär att

$$\lim_{n\to\infty} M_{Y_n}(t) = \lim_{n\to\infty}\left(1+\frac{\lambda\left(e^t-1\right)}{n}\right)^n$$

vilket är momentgenererande funktion för $Po(\lambda)$.