February 1, 2018

1 Förel. VII

Punktskattningar av parametrar i fördelningar

1.1 Punktskattning

För att skatta väntevärdet för en fördelning är det lämpligt att använda Medelvärdet $\overline{\xi} = \frac{1}{n}\sum_{j=1}^{n} \xi_j$. Vi tar nu väntevärdet av denna stok. var. :

$$E(\overline{\xi}) = ... = \mu$$

1.1.1 V.v.r. och effektivitet

V.v.r. är en förkortning av **väntevärdesriktig**. Punktskattningen $\overline{\xi}$ är v.v.r. ty $E(\overline{\xi}) = \mu$.

Ex 1 Undersök om nedanstående stok. var. är v.v.r. för ξ_k, $k = 1, 2$ som är likafördelade med gemensamt v.v. μ.

(a) $0.5\xi_1 + 0.3\xi_2$

(b) $0.6\xi_1 + 0.4\xi_2$

Lösning:

(a) $E(0.5\xi_1 + 0.3\xi_2) = ... = 0.8\mu$. Alltså ej v.v.r.

(b) $E(0.6\xi_1 + 0.4\xi_2) = ... = \mu$. Alltså v.v.r.

■

Ex 2 För två oberoende stok. var. ξ_1 och ξ_2 med samma μ men med olika standardavvikelse 2.0 respektive 1.0. De observerade värdena är kända. Frågan är hur dessa skall användas för att få en så effektiv v.v.r. skattning som möjligt.

Lösning:

En v.v.r skattning är $t\xi_1 + (1-t)\xi_2$ med varians

$$V(t\xi_1 + (1-t)\xi_2) = t^2 \cdot 2.0^2 + (1-t)^2 \cdot 1.0^2 =: V(t).$$

Det gäller att finna det minsta värdet av $V(t)$.

$$V'(t) = 10t - 2 = 0 \Longleftrightarrow t = 0.2\,.$$

Detta motsvarar ett minimum.

$$V_{\min} = V(0.2) = 0.8 \text{ med motsvarande standardavvikelse } \sigma = 0.89\,.$$

■

1.2 Centrala gränsvärdessatsen

Vi har sett summan av två ober. stok. var. Som är diskreta.

Ex 3 I ett tidigare exempel har vi ξ som antal datorer som säljs under en vecka i affären Datornörd. Vi antar vidare att frekvensfunktionen är

x	0	1	2
$P(\xi = x)$	0.3	0.4	0.3

Vi låter vidare ξ_1 , ξ_2 och ξ_3 vara antal sålda datorer under vecka 1, 2 respektive 3. Vi antar att dessa är oberoende och har samma fördelning d.v.s. som i tabellen ovan.

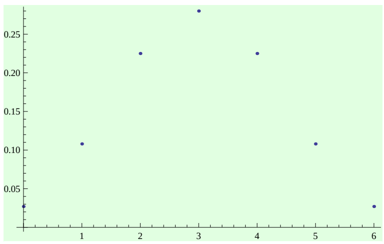
Lösning:

Det betyder att vi söker fördelningen hos summan $\eta = \xi_1 + \xi_2 + \xi_3$. Det är självklart att η kan anta värdena $0, 1, 2, 3, 4, 5, 6$. Hur fördelar sig sannolikheterna för η?

$$P(\eta = 0) = 0.3^3.$$

Fördelningen., d.v.s. frekvensfunktion (sannolikhetsfunktion) är från 0 till 6.

$$0.027, 0.108, 0.225, 0.28, 0.225, 0.108, 0.027$$



Genom att addera fler likafördelade stok. var., får man en "kurva", som allt mer liknar en normalfördelning.

∎

Sats 1 (Centrala gränsvärdessatsen)
Låt $\xi_1, \xi_2, ..., \xi_n$ vara oberoende likafördelade stokastiska variabler (d.v.s. med gemensamt väntevärde μ och varians σ^2). Sätt

$$\zeta = \sum_{k=1}^{n} \xi_k.$$

Då gäller för stora n att

$$\zeta \text{ approximativt har fördelningen } \mathrm{N}(n\mu, \sigma\sqrt{n}) \tag{1}$$

Kommentarer

- Det följer att $\bar{\xi}$ har approximativt $\mathrm{N}(\mu, \sigma/\sqrt{n})$.
- Tumregel för approximationen är $n \geq 30$.

Ex 4

(a) Vad är sannolikheten att man i datoraffären Datornörd säljer fler än 54 datorer under ett år?

(b) Vad är sannolikheten att det säljs i genomsnitt högst 0.9 datorer i veckan?

Förutsett att försäljningen vecka för vecka är ober.

Lösning:

(a) Väntevärdet $\mu = 1$ för försäljngen under en vecka. Motsvarande varians är

$$V = (0^2, 1^2, 2^2) \cdot (0.3, 0.4, 0.3) - 1^2 = 0.6,\ \sigma = \sqrt{0.6}.$$

Sannolikheten at man säljer fler än 54 datorer under ett år är

$$1 - \Phi\left(\frac{54 - 1 \cdot 52}{\sqrt{0.6} \cdot \sqrt{52}\cdot}\right) = 0.36$$

(b) Sannolikhetsfördelningen är $\mathrm{N}(1, \sqrt{0.6}/\sqrt{52}) \ni \zeta$ och vi söker sannolikheten

$$P := \Phi\left(\frac{0.9 - 1}{\sqrt{0.6}/\sqrt{52}}\right) = \Phi(-0.930949...) = 0.17594 \approx 0.18.$$

Svar (b) Sannolikheten att man säljer högst 0.9 datorer per vecka är 0.18.

∎

Ex 5 För en produkt har man fått följande värden:

47.72, 49.67, 46.7343, 49.15, 50.58, 49.29.

Det kan ses som en observerade mätvärden av en normalfördelad variabel $\xi \in \mathrm{N}(\mu, 1.1)$. Standardavvikelsen är alltså känd. Producenten påstår att medianvärdet är 50.00. Testa detta med ett symmetriskt konfidensintervall i (a) och (b) på signifikansnivå

(a) 95%

(b) 99%

1. Slutsatser?

Lösning:

(a) 95%: Vi får observerad punktskattning $\overline{\xi} = 48.8569$. Det symmetriska intervallet är

$$\left[\overline{x} - \frac{\lambda_{\alpha/2}\sigma}{\sqrt{6}}, \overline{x} + \frac{\lambda_{\alpha/2}\sigma}{\sqrt{6}}\right]$$

Med

$$\lambda_{\alpha/2} = \lambda_{0.025} = \{\text{tabell}\} = 1.96$$

blir konfidensintervallet

$$\overline{x} - \frac{1.96 \cdot 1.1}{\sqrt{6}}, \overline{x} + \frac{1.96 \cdot 1.1}{\sqrt{6}}] =$$

$$[48.8567 - \frac{1.96 \cdot 1.1}{\sqrt{6}}, 48.8567 + \frac{1.96 \cdot 1.1}{\sqrt{6}}] = [47.9765, 49.7368]$$

(b) 99%: Här är ”kvantilen”

$$\lambda_{\alpha/2} = \lambda_{0.005} = 2.57583.$$

Motsvarande intervall är

$$[47.6999,\ 50.0134]$$

(c) I (a) gäller att $50.00 \notin [47.9765, 49.7368]$ Det betyder att med 95% sannolikhet så är $\mu \neq 50.00$.
I (b) gäller att $50.00 \in [47.6999,\ 50.0134]$. Vi kan alltså med 99% **inte** påstå att $\mu \neq 50.00$.

Kommentarer

- Det som beskrivs i detta exempel är ett *hypotestest.* Att påstå att $\mu \neq 50.00$
- Producentens påstående att $\mu = 50.00$ är *nollhypotesen* H_0.
  Etthypotesen H_1 är $\mu \neq 50.00$.
- I exemplet *förkastas* H_0 på signifikansnivå 95% men förkastas inte på signifikansnivå 99%.
- *Innan* man ger konfidensintervallet har man en *intervallskattning*

$$\left[\overline{\xi} - \frac{\lambda_{\alpha/2} \cdot \sigma}{\sqrt{n}}, \overline{\xi} + \frac{\lambda_{\alpha/2} \cdot \sigma}{\sqrt{n}}\right] \tag{2}$$

- Innan nästa exempel tar vi upp *vilken fördelning* som en summa $\xi_1 + \xi_2$ kan ha.

1. Om ξ_k är binomialfördelade och oberoende samt $\xi_1 \in \mathrm{Bin}(n_1, p)$ och $\xi_2 \in \mathrm{Bin}(n_2, p)$, d.v.s. samma p, så är $\xi_1 + \xi_2 \in \mathrm{Bin}(n_1 + n_2, p)$ (Bevis m.h.a. indikatorvariabler).
2. Om $\xi_k \in \mathrm{N}(\mu_l, \sigma_k)$, så är $\xi_1 + \xi_2 \in \mathrm{N}(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$.

Ex 6 Ett byggprojekt av bostäder åt studenter betår av två moment, lägga grund och sätta upp väggar och tak. Dessa moment är $\mathrm{N}(2.0, 0.4)$ och $(N)(1.5, 0.3)$ i enheten 30 dagars perioder (d.v.s. 1.5 betyder $1.5 \cdot 30 = 45$ dagar etc). Var är sannolikheten att byggtiden totalt tar mer än 120 dagar? Antag att de två momentens tidsåtgång är oberoende.

Lösning:

Låt de två momentens tidsåtgång vara ξ resp. ζ och sätt $x = 4.0$ (eftersom $30 \cdot 4.0 = 120$).

Totala byggtiden är då

$$\xi + \zeta \in \mathrm{N}(2.0 + 1.5, \sqrt{0.4^2 + 0.3^2}).$$

Den sökta sannolikheten är

$$1 - \Phi\left(\frac{4.0 - 3.5}{0.5}\right) = 0.158655 \approx 16\%.$$

Sannolikheten att byggprojektet tar mer än 120 dagar är 16%.

■

1.3 Intervallskattning då σ okänd

Vi gör nu det i fallet få σ måste skattas, d.v.s. är okänd. Vi punktskattar då

$$\sigma^* = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(\xi_k - \bar{\xi})^2}$$

Vi antar att det finns ett c, sådant att vid en symmetrisk intervallskattning med konfidensgrad $1 - \alpha$, t

$$P(\bar{\xi} - k\frac{\sigma^*}{\sqrt{n}} < \mu) = 1 - \alpha/2.$$

Detta kan skrivas om till en jämn funktion

$$P\left(\frac{\bar{\xi} - \mu}{\sigma^*/\sqrt{n}} < k\right) = 1 - \alpha/2.$$

Den *stokastiska variabeln*, obs! Både säljare och nämnare är stok. var.

$$\frac{\bar{\xi} - \mu}{\sigma^*/n}$$

är, kan man visa, $t-$*fördelad med* $n-1$ *frihetsgrader*. Denna fördelningsfunktion är tabellerad.

Ex 7 Vid mätning av strålning från mobiltelefon har man följande stickprov (Enhet mr/h), som antas vara observerade värden från en normafördening. 0.50

mr/h är rekommenderat maxvärde. Ett mobiltelefonföretag påstår att medelstrålningen är 0.45 mr/h.

$$\{0.4, 0.48, 0.6, 0.15, 0.5, 0.8, 0.5, 0.36, 0.16, 0.89\}.$$

Ge ett nedåt begränsat 99% konfidensintervall för strålningen.

Lösning:

(Observerat) medelvärde och standardavvikelse är $\overline{x} = 0.484$ respektive $\sigma^*_{Obs} = s = 0.239824$. Intervallet är

$$\left(\overline{x} - t_{n-1,0.99} \cdot \frac{s}{\sqrt{n}}, +\infty\right].$$

$n = 10$, så att tabellvärdet $t_{9,0.99} = 2.9$ vilket ger intervallet $[0.280, \infty)$. Vi kan inte förkasta företagets hypotes att medelstrålningen är 0.45 mr/h, på 95% :s signifikansnivå. Om 0.45 hade varit t.v. om intervallets om intervallets nedre gräns hade vi förkastat företagets hypotes (nollhypotesen).

∎

Kommentarer

- Företagets hypotes, nollhypotesen H_0 att $\mu = 0.45$ ställs mot mothypotesen H_1, som här är att $\mu > 50.0$. H_0 kan inte förkastas på 95%.s signifikansnivå.

- $t-$fördelningenss frekvensfunktion beror på n och ser ut så här

  $$f_n(x) = c_n \left(1 + \frac{x^2}{n-1}\right)^{-n/2}.$$

  Med lite kunskaper om gränsvärde, kan man visa att

  $$f_n(x) \longrightarrow \frac{1}{\sqrt{2\pi}} e^{-x^2/2} = \phi(x) \text{ då } n \longrightarrow \infty$$

  Standardnormalfördelningens frekvensfunktion.

- Om n stort ($n \geq 30$) d.v.s. ett stickprov på en godtycklig fördelning blir det approximerande symmetriska konfidensintervallet

  $$\left[\overline{x} - \frac{\lambda_{\alpha/2} \cdot s}{\sqrt{n}}, \overline{x} + \frac{\lambda_{\alpha/2} \cdot s}{\sqrt{n}}\right].$$

  Man använder $\lambda_{\alpha/2}$ i ställer för $t_{n-1,\alpha/2}$ ty $f_n(x) \approx \phi(x)$ för $n \geq 30$.