1 Föreläsning V

1.1 Linjärkombination av stokastiska variabler

Givet två stokastiska variabler ξ_1 och ξ_2. Med en linjärkombination menas $\eta = a\,\xi_1 + b\,\xi_2$.

Ex 5.1 Givet ξ_1 och ξ_2, två tärningskast. Då är $\xi_1 + \xi_2 =: \eta$ tärningskastens sammanlagda poäng. $\eta = 2, , 3, \ldots, 11, 12$ och är inte likformig fördelning. Men rimligen borde $E(\eta) = 2 \cdot 3.5$.
Vidare är $E(2\xi_1) = 2 \cdot E(\xi_1) = 7.0$, alltså samma väntevärde. Vi beräkna v.v. och varians av $\xi_2 + \xi_2$ och av $2\xi_1$.
Väntevärde:

$$2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + \ldots + 12 \cdot \frac{1}{36} = \frac{1}{36}(2 + 6 + 12 + \ldots + 12 \cdot 1) = 7.0 = E(\xi_1) + E(\xi_2).$$

Lättare att visa är att

$$E(2\xi_1) = \frac{1}{6}(2 + 4 + 6 + 8 + 10 + 12) = 2E(\xi_1).$$

■

Ovanstående exempel illustrerar följande

$$E(\xi_1 + \xi_2) = E(\xi_1) + E(\xi_2), \quad E(a\,\xi) = a\,E(\xi). \tag{1}$$

Man kan också visa att, om de är oberoende, så är

$$\begin{aligned} V(\xi_1 + \xi_2) &= V(\xi_1) + V(\xi_2). \\ V(a\,\xi) &= a^2 V(\xi) \\ V(a\xi + b) &= a^2 V(\xi) \end{aligned} \tag{2}$$

där man i den sista likheten betraktar b som en konstant och har (således) variansen $= 0$.

Ex 5. 2 Beräkna variansen och standardavvikelsen för (a) summan av de två tärningskasten och för (b) dubblering av poäng på ett kast.

Lösning

(a)

$$V(\xi_1 + \xi_2) = V(\eta) = 2^2 \cdot \frac{1}{36} + 3^2 \cdot \frac{2}{36} + \ldots + 12^2 \cdot \frac{1}{36} - 7^2 = \ldots = \frac{35}{6} = 2 \cdot \frac{25}{12} = 2V(\xi_1) = V(\xi_1) + V(\xi_2).$$

(b)

$$V(2\xi_1) = \frac{1}{6}(2^2 + 4^2 + 6^2 + \ldots + 12^2) - 7^2 = \frac{35}{3} = 2^2 \cdot \frac{35}{12} = 2^2 \cdot V(\xi_1).$$

■

Vi behöver en linjärkombination till, av stokastiska variabler.
Anledningen till att man gör fler än en mätning (d.v.s. kontroll av mer än en produkt), är bl.a. att man då kan ta medelvärdet av dessa (innan mätning) och få ned variansens storlek. Ex.vis med n oberoende mätningar (se det som stok. var. innan vi får observerade värden) har medelvärdet som

$$\overline{\xi} = \frac{1}{n}(\xi_1+\xi_2+\ldots+\xi_n \text{ med v.v. } E\left(\frac{1}{n}(\xi_1+\xi_2+\ldots+\xi_n\right) = \frac{1}{n}(\mu+\mu+\ldots+\mu) = \mu.$$

Hur är det då med variansen och standardavvikelsen för medelvärdet $\overline{\xi}$? Vi utgår från att σ är dess gemensamma standardavvikelsen. Då är

$$\begin{aligned} V(\overline{\xi}) &= V\left(\frac{\xi_1+\xi_2+\ldots+\xi_n}{n}\right) = \frac{1}{n^2}(V(\xi_1)+V(\xi_2)+\ldots+V(\xi_n)) = \\ &= \frac{1}{n^2}(\underbrace{\sigma^2+\sigma^2+\ldots+\sigma^2}_{n \text{ termer}}) = \frac{1}{n^2}\cdot n\sigma^2. \end{aligned}$$

Alltså är

$$\sqrt{V(\overline{\xi})} = \frac{\sigma}{\sqrt{n}} =: \overline{\sigma}. \tag{3}$$

Kommentarer

- Var är likheten och skillnaden mellan $\xi_1+\xi_2$ och $2\xi_1$, där ju ändå ξ_1 och ξ_2 är likafördelade och oberoende? När det gäller väntevärde så är de lika. Men varians skiljer sig. Variansen är mindre för $\xi_1+\xi_2$ än för $2\xi_1$. Det förklaras med att genom i att i summan $\xi_1+\xi_2$ kan den ena ge ett för stort observerat värde medan den andra termen ger ett för litet observerat värde, så att värdena delvis tar ut varandra.

  $V(\xi_1)+V(\xi_2)$ och $V(2\xi_1)$

- Vid massproduktion blir det i bland fel på hela serier av ex.vis personbilar. Detta visar på $V(n\xi_1) = n^2V(\xi_1)$ snarare än $V(\sum_{k=1}^{n}\xi_k)$, d.v.s, är det fel på en komponenten i *en bil*, så är det fel på hela seriens komponenter.

- För vissa typer av summor av fördelningar är summans fördelning känd. När det gäller $\xi_1 \in \text{Bin}(n_1,p)$och $\xi_2 \in \text{Bin}(n_2,p)$, där ξ_1 och ξ_2 är oberoende, så är
$$\xi_1+\xi_2 \in \text{Bin}(n_1+n_2,p).$$
  Dessutom gäller att $\xi_1 \in \text{N}(\mu_1,\sigma_1)$och $\xi_2 \in \text{N}(\mu_2,\sigma_2)$, där ξ_1 och ξ_2 är oberoende, så är
$$\xi_1+\xi_2 \in \text{N}(\mu_1+\mu_2,\sigma).$$
  Att beräkna σ kräver att de två stokastisk variablerna är oberoende. Av ovan vet vi att då är
$$V(\xi_1+\xi_2) = V(\xi_1)+V(\xi_2) = \sigma_1^2+\sigma_2^2 = \sigma^2, \text{ d.v.s. } \sigma = \sqrt{\sigma_1^2+\sigma_2^2}.$$

- En orsak till att göra fler än en mätning/kontroll av en produkts egenskaper, är att ju fler mätningar (d.v.s. större n), erhålls en mindre standardavvikelse.

Ex 5.3 Lasse tapetserar ett rum på en tid $\xi \in \mathrm{N}(2.0, 0.40)$ och Börje lägger golvet på en tid $\zeta \in \mathrm{N}(2.5, 0.42)$, enhet timmar. Tiderna är oberoende.

(a) Beräkna sannolikheten att tiden för tapetsering och golvläggning tar totalt högst 5.0 h.

(b) Beräkna sannolikheten att golvläggningen går snabbare än tapetseringen.

Lösning

(a)

$$\xi + \zeta \in \mathrm{N}(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}) = \mathrm{N}(4.5, 0.58).$$

$$P(\xi + \zeta \leq 5.0) = \Phi\left(\frac{5.0 - 4.5}{0.58}\right) = \Phi(0.862) = 0.8051 = 0.8$$

Svar: Sannolikheten att de blir klara inom 5.0 h är 80%.

(b) Sannolikheten att tiden för golvläggningen tar längre tid än tapetseringen, kan tecknas enkelt som

$$P(\zeta < \xi) = P(\zeta - \xi < 0).$$

Nu gäller att $\zeta - \xi \in \mathrm{N}(\mu_2 - \mu_1, \sqrt{\sigma_1^2 + \sigma_2^2}) = \mathrm{N}(0.5, 0.58)$. Vi går över till standardnormalfördelningen och får

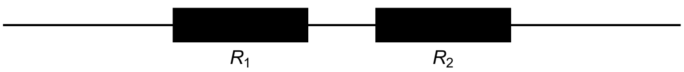
$$P(\zeta - \xi < 0) = \Phi\left(\frac{0 - 0.5}{0.58}\right) = \Phi(-0.862) = 1 - \Phi(0.862) = 0.1949.$$

∎

Andra möjligheter att kombinera två eller fler stokastiska variabler är att ta max eller min. I nästa exempel har vi en sådan variant.

Ex 5.4 Givet två seriekoppplade motstånd med exponentialfördelade livslängder ξ_1 och ξ_2. Förväntade livslängder är 200 respektive 250 h.

Systemet (alltså seriekopplingen) har sin livslängd given av att *båda* resistorerna fungerar. Antag att de två resistorernas livslängder är oberoende.



(a) Ge en sannolikhetsfördelning för det system som de utgör.

(b) Vad är sannolikheten att systemet har en livslängd på minst 150 h?

Lösning

(a) Vi har att systemets livslängd $= \eta$ är densamma som den minsta livslängden av de två resistorerna. D.v.s. $\eta = \min(\xi_1, \xi_2)$. Att $\eta > t$ (h) betyder alltså att *både* $\xi_1 > t$ och $\xi_2 > t$. Uttryckt med sannolikhet innebär det att

$$\begin{aligned} P(\eta > t) &= (P\xi_1 > t \text{ och } \xi_2 > t) = \{\text{ober.}\}) = P(\xi_1 > t) \cdot P(\xi_2 > t) = \\ &= (1 - F_1(t)) \cdot (1 - F_2(t)) = e^{-\lambda_1 t} \cdot e^{-\lambda_2 t} = e^{-(\lambda_1 + \lambda_2)t} \Longleftrightarrow \\ F(t) &= 1 - e^{-(\lambda_1 + \lambda_2)t} \rightarrow f(t) = \lambda e^{-\lambda t} \end{aligned}$$

där $F(t)$ och $f(t)$ är fördelningsfunktion respektive frekvensfunktion för $\min(\xi_1, \xi_2)$. Uppenbarligen är $\eta = \min(\xi_1, \xi_2) \in \mathrm{Exp}(\lambda_1 + \lambda_2)$. Vi skall till slut bara beräkna

$$\lambda_1 + \lambda_2 = \frac{1}{200} + \frac{1}{250} = 0.009.$$

(b) Sannolikheten att systemet har en livslängd på minst 150 h är

$$P(\eta > 150) = e^{-0.009 \cdot 150} = e^{-1.35} = 0.26 \text{ (svar)}$$

■

1.2 Punktskattning

Framgent använder vi våra kunskaper till att skatta parametrars värde för fördelningar, fr.a. μ och σ.

Punktskattning av μ

Ex 5.5 Ex Man vill punktskatta väntetvärdet och standardavvikelsen för en okänd fördelning och på hållfasthet (Maximal normalkraft i kN) för en typ aluminumrör och får 48.5, 50.5, 49.0, 54.5, 52.5. Man får, som medelvärdet av dessa, en *observerad punktskattning* 49.2.

■

- Vi tar om exemplet och
  1. tar ett stickprov av storlek $n = 5$ av en (okänd) fördelning och kallar motsvarande stokastisk variabel för ξ eller snarare $(\xi_1, \xi_2, \xi_3, \xi_4, \xi_5)$, där ξ och alla ξ_j, $j = 1, 2, 3, 4, 5$ är likafördelade och oberoende, med μ alltså med gemensamt väntevärde och σ^2 som gemensam varians *innan* vi får observerade värden.
  2. Observerade värden för ξ_1 till ξ_5 är
  $$(x_1, x_2, x_3, x_4, x_5) = (48.5, 50.5, 49.0, 54.5, 52.5)$$
  3. Dessutom inför vi en lämplig stickprovsvariabel, ex.vis medelvärdet
  $$\overline{\xi} =: \mu^* \text{ med } \textit{observerad punktskattning } \overline{x} =: \mu^*_{\text{obs}} = 51.0.$$
- $\overline{\xi} =: \mu^*$ kallas en *punktskattning* av μ.
- Vi antar att man har observerat värden $x_1, x_2, \ldots, x_n$ av dessa. Då kallas $\overline{x} =: \mu^*_{\text{obs}}$ en *observerad* punktskattning av μ.

  Kommentarer
- Man kan tänka sig andra typer av punktskattningar av μ, ex.vis genom att skriva dem i storleksordning (som stok. var.) och ta
  $$\mu^* = \frac{\xi_{\min} + \xi_{\max}}{2} \text{ med } \mu^*_{\text{obs}} = 51.5$$
  eller
  $$\mu^* = \text{det mittersta av } \xi_k \text{ med } \mu^*_{\text{obs}} = 50.5$$

- För två värden ξ_1 och ξ_2 har man följande stickprov

$$\mu_I^* = 0.2\xi_1 + 0.8\xi_2 \text{ respektive } \mu_{II}^* = 0.3\xi_1 + 0.9\xi_2.$$

  Lineariteten säger att μ_I^* men inte μ_{II}^* är *väntevärdesriktig*, d.v.s.

  $E(\mu_I^*) = 0.2 \cdot \mu + 0.8\,\mu = 1 \cdot \mu$ men $[E(\mu_{II}^*) = 0.3 \cdot \mu + 0.9\,\mu = 1.2 \cdot \mu \neq \mu.$

Definition 1 Givet en fördelning med parameter τ, som man vill få en observerad punktskattning av. En punktskattning τ^* av en parameter τ är väntevärdesriktig (v.v.r.), omm

$$E(\tau^*) = \tau. \tag{4}$$

Punktskattning av σ^2 och σ

- Med $\{\xi_1, \xi_2, \ldots, \xi_n\}$ som ovan använder man punktskattningen av σ^2, som ser ut så här

$$\sigma^{2*} = \frac{1}{n-1}\sum_{j=1}^{n}(\xi_j - \overline{\xi})^2 \equiv \frac{1}{n-1}\left[\sum_{j=1}^{n}\xi_j^2 - n\,(\overline{\xi})^2\right] \tag{5}$$

  Man kan visa att den är v.v.r. , d.v.s. $E(\sigma^{2^*}) = \sigma^2$.

- Med samma exempel som ovan är motsvarande observerade punktskattning av σ^2

$$s^2 := {\sigma^2}^*_{\text{obs}} = 6.25\,.$$

- Punktskattningen av σ får genom att dra roten ur (5) och p.s.s. med observerad punktskattning, som alltså är $\sqrt{{\sigma^2}^*_{\text{obs}}} = 2.5$.

Obs! σ^*_{obs} finns på miniräknaren $\boxed{\boxed{\sigma_{n-1}}}$.

Definition 2 Givet två punktskattningar τ_1^* och τ_2^* av samma parameter τ. τ_1^* är effektivare än τ_2^*, omm

$$V(\tau_1^*) \leq V(\tau_2^*).$$

Ex 5.6

En punktskattning av skjuvspänning för en kabel i en kabelbro görs med observerat värde 25.0 kN/m^2. Något decennium senare görs en ny mätning och man får 21.0 kN/m^2. Vid första mätningen använde man en metod med standardavvikelse $\sigma_1 = 2.0$ och vid den andra mätningen $\sigma_2 = 1.0$. Man funderar på att väga ihop de två mätresultaten på ett *effektivast* sätt. Innebär det att man inte skall ta hänsyn till det första mätresultatet (som ju har högre värde på σ)?

Lösning

Vi väljer en *effektivaste* metod av typ $\mu^* = t \cdot \xi_1 + (1-t)\xi_2$.Vi observerar först att den är v.v.r. Med effektivast menas att vi skall välja t, så att $V(\mu^*)$ minimeras. Variansen är

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

Minimum erhålls då derivatan $= 0$:

$$h'(t) = -2(1-t) + 8t = 0 \Longleftrightarrow t = 0.2.$$

Den effektivaste skattningen är $\mu^* = 0.2 \cdot \xi_1 + 0.8\xi_2$ och motsvarande observerade värde är

$$\mu^*_{\text{obs}} = 0.2 \cdot 25. + 0.8 \cdot 21.0 = 21.8.$$

■

- Varför göra flera mätningar $\xi_k, \quad k = 1, 2, 3, \ldots, n$ på samma typ av objekt (ex.vis längd, tyngd eller skjuvning)? Låt $\xi_1, \xi_2, \ldots, \xi_n$ vara ober. och likafördelade stok. variabler., med $\mu = E(\xi_k)$ och $V(\xi_k) = \sigma^2$. Vad vinner man på att ta medelvärdet av ett flertal mätningar?

  Vi betraktar medelvärde $\overline{\xi}$ och varians $V(\overline{\xi})$: och betecknar standardavvikelsen för medelvärdet $\overline{\xi}$ med $\overline{\sigma}$.

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

  och

  $$V(\overline{\xi}) = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n} \Longrightarrow \overline{\sigma} = \frac{\sigma}{\sqrt{n}}.$$

  Svaret är att man får en allt mindre standardavvikelse, ju större stickprovstorlek, (d.v.s. ju större n).

1.3 Intervallskattning

Vi gör bara intervallskattning på μ och σ för Normalfördelning. Vi erinrar oss om frekvens- och fördelningsfunktion för standardnormalfördelningen är

$$\begin{cases} \varphi(x) = & \dfrac{1}{\sqrt{2\pi}} e^{-x^2/2} \text{ och} \\ \\ \Phi(x) = & \displaystyle\int_{-\infty}^{x} \varphi(t)dt, \end{cases}$$

där $\Phi(x)$ finns tabellerad.

Ibland vill man skatta en parameter med ett intervall, såsom μ för en normalfördelning. Detta innebär att man med stor säkerhet, ex.vis 95%, vill påstå att $\mu \in [a, b]$, ett intervall som inte är alltför stort. Vi skall nedan intervallskatta μ och σ (bara) för en normalfördelning. För detta inför vi, för $\mathrm{N}(0, 1)$, *kvantiler*. Ex,vis är $x = \lambda_{0.05} = 1.6455$, det x, sådant att $0.95 = 95\%$ av arean under $\varphi(x)$ är t.v. om detta x.

1.3.1 Intervallskattning av μ i normalfördelning

Vi har sedan tidigare att för ξ_k, $k = 1, 2, \dots n$ likafördelade och oberoende och med v.v. μ och standardavvikelse σ, att

$$\begin{aligned} &\zeta := \textstyle\sum_{k=1}^n \xi_k \text{ har } E(\zeta) = n \cdot \mu \text{ och } V(\zeta) = n\sigma^2 \\ &\text{och} \\ &\eta := \overline{z} \text{ har } E(\eta) = n \cdot \mu \text{ och } V(\eta) = \frac{\sigma^2}{n}. \end{aligned} \tag{6}$$

μ då σ känd Givet ξ_k och η som i (6) där vi dessutom antar att $\xi_k \in \mathrm{N}(\mu, \sigma)$. Vi skall då utnyttja punktskattningen $\overline{z}$ för att göra en intervallskattning av μ. Vi vet att $\overline{z} \in \mathrm{N}(\mu, \sigma/\sqrt{n})$. Vi börjar med att bestämma a, så att

$$-a < \overline{\xi} - \mu < a$$

med, säg $1 - \alpha = 95\%$:s säkerhet, d.v.s. $\alpha = 5\% = 0.05$. Vi gör om mittenledet till en $\mathrm{N}(0, 1)$ stok. var. Vi skriver detta ekvivalent som

$$-\frac{a}{\sigma/\sqrt{n}} < \frac{\overline{\xi} - \mu}{\sigma/\sqrt{n}} < \frac{a}{\sigma/\sqrt{n}} \text{ och } \frac{\overline{\xi} - \mu}{\sigma/\sqrt{n}} \in \mathrm{N}(0, 1). \tag{7}$$

P.g.a. symmetri får vi vänster och höger gräns som $x = -\lambda_{\alpha/2}$ och $x = \lambda_{\alpha/2}$. Vi sätter vänsterled till $-\lambda_{\alpha/2}$ och HL till $\lambda_{\alpha/2}$ i (7) och får en tvåsidig *intervallskattning* av μ som

$$-\lambda_{\alpha/2} < \frac{\overline{\xi} - \mu}{\sigma/\sqrt{n}} < \lambda_{\alpha/2} \iff \overline{\xi} - \lambda_{\alpha/2}\sigma/\sqrt{n} < \mu < \overline{\xi} + \lambda_{\alpha/2}\sigma/\sqrt{n} \tag{8}$$

eller som intervall

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

Motsvarande *observerade intervallskattning* ges av

$$\overline{x} - \sigma/\sqrt{n} < \mu < \overline{x} + \sigma/\sqrt{n}. \tag{9}$$

eller som intervall

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

Ex 5.7 Ett *observerat* stickprov på μ är $\{48.5, 50.5, 49., 54.5, 52.5\}$ av storlek $n = 5$ av en normalförd. $\mathrm{N}(\mu, 0.5)$. Bestäm

(a) ett tvåsidigt 95% konfidensintervall för μ.

(b) ett ensidigt nedåt begränsat 95% konfidensintervall för μ.

Lösning

(a) ett tvåsidigt (symmetriskt) 95% konfidensintervall för μ: $\lambda_{\alpha/2} = \lambda_{0.025} = \{\text{tabell}\} = 1.96$. Nu är $\overline{x} = 51.0$. Konfidensintervallet är

$$[50.5617,\ 51.4383] = [50.5, 51.5].$$

(b) ett ensidigt nedåt begränsat 95% konfidensintervall för μ ser ut så här

$$[\overline{x} - \lambda_\alpha \cdot \frac{\sigma}{\sqrt{n}}, \infty) = [50.6, \infty).$$

■

μ då σ okänd I detta fall betraktar vi först variansen som en stokastisk variabel, en punktskattning:

$$\sigma^{2*} = \frac{1}{n-1}\sum_{k=1}^{n}(\xi_k - \overline{\xi})^2. \tag{10}$$

- Obs, man borde dividera med n och inte $n-1$ men divisionen med $n-1$ gör att σ^{2*} blir v.v.r., d.v.s.

  $$E(\sigma^{2*}) = \sigma^2.$$

  Den är dessutom den mest effektiva.

- Motsvarande observerade punktskattning är

  $$\sigma^{2*}{}_{\text{obs}} = s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \overline{x})^2. \tag{11}$$

  Vidare måste vi veta fördelningen för detta σ^{2*}.

  Det är en fördelning som liknar $\mathrm{N}(0,1)$ men med lite ”bredare” form.

Ex 5.8 Vi löser (a) och (b) i föregående exempel men med σ okänd. Med samma observerade stickprov som i föregående exempel men vi punktskattar σ genom att punktskatta σ^2 med (10) eller direkt med (11). Man får den v.v.r. observerade punktskattningen s^2 av σ^2, som man drar roten ur för att få s[1]. Beräkning ger observerad punktskattning s av σ som

$$s = 2.5$$

Med ett stickprov av storleken $n = 5$ talar man om antal ”frihetsgrader” $n-1 = 4$.

(a) Ett tvåsidigt 95% intervall. Frekvensfunktionen är $f_4(x) = 12\left(\frac{1}{x^2+4}\right)^{5/2}$ men vi inte behöver veta det eftersom vi ser i tabell. Vi ser i tabell för $t-$fördelningen att motsvarade kvantil (fraktil i boken) är $t_{\alpha/2}(4) = 2.776$. Vi har samma medelvärde 51.0. Det ger det tvåsidiga konfidensintervallet

$$[47.8963, 54.1037].$$

(b) Vi får kvantilen $t_\alpha = t_{0.05} = 2.132$ och motsvarande konfindensintervall

$$[48.6164, \infty) = [48.6, \infty)\,.$$

■

[1]Ej v.v.r. men konvergerar mot v.v.r. då $n \longrightarrow \infty$.

1.3.2 Konfidensintervall för σ^2 och σ

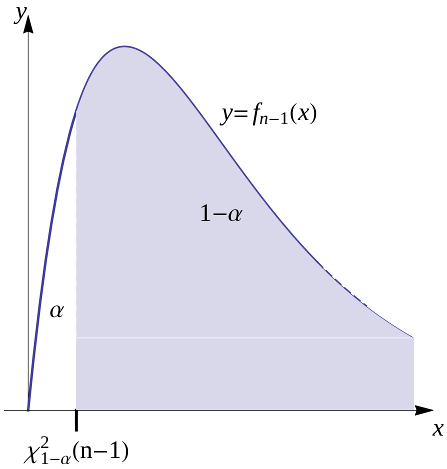
Vi börjar som vanligt med en intervallskattning, alltså vi börjar med skattning med stokastiska variabler. Här behandlas endast fallet μ okänd. Vi skall använda punktskattningen av σ^2, där $\xi_1, \xi_2, \ldots, \xi_n$ alla i $\mathrm{N}(\mu, \sigma)$.

Vi söker först ett intervall för σ^2 av typen $[0, k\,\sigma^{2^*}]$ av konfidensgrad $1-\alpha$, uttrycket med sannolikhet

$$P(\sigma^2 \le k \cdot \sigma^{2^*}) = 1 - \alpha$$

eller genom att gå över till komplementhändelsen

$$P(k \cdot \sigma^{2^*} < \sigma^2) = \alpha.$$



$$\begin{aligned}
P(k \cdot \sigma^{2^*} < \sigma^2) = \alpha \quad &= \quad P\left(\underbrace{\frac{(n-1)\sigma^{2^*}}{\sigma^2}}_{\in \chi^2_{n-1}} < \frac{n-1}{k}\right) \\
&\Longrightarrow \quad \frac{n-1}{k} = \chi^2_{1-\alpha}(n-1) \Longleftrightarrow \\
k \quad &= \quad \frac{n-1}{\chi^2_{1-\alpha}(n-1)}. \\
\sigma^{2^*} &\cdot \frac{n-1}{\chi^2_{1-\alpha}(n-1)}.
\end{aligned}$$

En upptå begränsad intervallskattning av σ^2 av konfidiensgrad $1-\alpha$ är

$$\left[0, \frac{(n-1)\sigma^{2^*}}{\chi^2_{1-\alpha}(n-1)}\right)$$

Eller skrivet som en olikhet

$$0 < \sigma^2 < \frac{(n-1)\sigma^{2^*}}{\chi^2_{1-\alpha}(n-1)}$$

med konfidengrad $1-\alpha$.

Ex 5.9 Samma som exempel 5.7, och 5.8 med σ okänt. Bestäm ett 99% uppåt begränsat konfidensintervall för σ^2 och σ.

Lösning

Sätt $\alpha = 0.01$. Här är

$$n-1 = 4,\ {\sigma^2}^* = 6.25\ldots \text{ och } \chi^2_{0.99}(4) = 0.30.$$

Intervallets övre gräns för σ^2 är

$$k \cdot {\sigma^2}^*_{\text{obs}} = \frac{n-1}{\chi^2_{1-\alpha}(n-1)} \cdot s^2 = 83.3$$

och för σ är övre gräns i konfidensintervallet är 9.1, d.v.s. konfidensintervallet för σ är $[0, 9.1]$ av konfidengrad 99%.

■

Ibland vill man ha ett nedåt begränsat konfidiensintervall för σ^2. Då utgår man från olikheten

$$k{\sigma^2}^* < \sigma^2$$

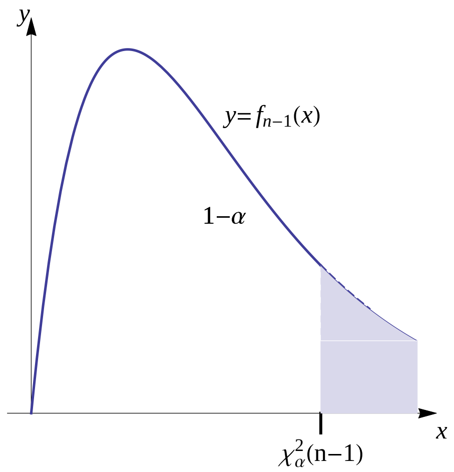
Ex 5.10 Ge ett nedåt begr. konfidensintervall för σ^2 med konf. grad 99% samma som i förågende exempel.

Lösning

Vi får konf. intervallet

$$\left[0, \frac{(n-1)s^2}{\chi^2_{0.99}(4)}\right) = \{\text{Tabell: } \chi^2_{0.99}(4) = 13.3\} = [1.883, \infty)$$

■



1.4 CGS

"CGS" står för Centrala gränsvärdessatsen och säger att summan av ett större antal likafördelade ober. stok. var., är approximativt normalfördelad. Som exempel kan vi ta summan av två, tre och fyra tärningskast poäng i föreläsning 1. Grafen av summan påminner om en N−fördelnings fefkvensfunktion. Låt $\eta_n = \xi_1 + \xi_2 + \ldots + \xi_n$, vara en sådan summa, där $E(\xi_k) = \mu$ samt $V(\xi_k) = \sigma^2$. Mer exakt säger den att

$$P\left(\frac{\eta_n - n\,\mu}{\sigma\sqrt{n}} \leq b\right) \approx \Phi(b)$$

där $n \geq 30$ är tumregeln.

Ex 5.11 En kontroll av utsläpp av kolmonoxid från ett visst bilmärke får inte överskrida 0.02%. Sannolikheten att en given bil har för hög halt (> 0.02%) beräknas till 0.05.
400 bilar kontrolleras slumpmässigt. Vad är sannolikheten att minst trettio av dessa bilar har ett för högt utsläpp? Antag att bilarnas utsläpp är likafördelade och ober.

Lösning

Låt $\eta = \sum_{k=1}^{400} \xi_k$ där ξ_k är utsläpp från bil nummer k. Mer exakt sätter vi

$$\xi_k = \begin{cases} 1, & \text{om } > 0.02\% \\ 0, & \text{om } \leq 0.02\% \end{cases}$$

Ett sådant ξ_k är en *indikatorvariabel* med parameter p.

Det följer, efter lite funderingar, att summan är

$$\boxed{\eta \in \mathrm{Bin}(400, 0.05),}$$

så vi kan beräkna sannolikheten utifrån det.

Men också approximera med normalfördelning, se bild i boken sidan 181. Den säger att normalfördelning fungerar, om $np(1-p) > 10$, i vårt fall $400 \cdot 0.05 \cdot 0.95 = 19 > 10$. Vi söker sannolikheten $P(\eta \geq 30)$. Approximationen säger att

$$\mathrm{Bin}(n, p) \approx \mathrm{N}(\mu, \sigma) \text{ med } \mu = np \text{ och } \sigma = \sqrt{np(1-p)}.$$

Vi har v.v. $\mu = n \cdot p = 20$ och varians $\sigma^2 = 19$.

$$P(\eta \geq 30) \approx 1 - F(30), \text{ där } F(30) = \Phi\left(\frac{30-20}{\sqrt{19}}\right) = \Phi(2.29) = 0.99.$$

Svar: Sannolikheten att det finns minst 30 bilar med för högt utsläpp av CO är 1%.

∎

Mer om indikatorvariabel

- En indikatorvariabel, med parameter p och $\mathcal{I} \in \mathrm{Bin}(1, p)$
- V.v. och varians är

  $E(\mathcal{I}) = 0{\cdot}(1-p)+1{\cdot}p = p$ respektive $V(\mathcal{I}) = 0^2(1-p)+1^2{\cdot}p-p^2 = p(1-p)$.

- Om $\mathcal{I}_k$ oberoende indikatorvariabler med samma p, så gäller att

$$\xi := \sum_{k=1}^{n} \mathcal{I}_k \in \mathrm{Bin}(n, p).$$

  Alltså kan vi få v.v. och varians för $\mathrm{Bin}(n, p)$ via räkneregler för summa:

$$E(\xi) = \sum_{k=1}^{n} p = n \cdot p \text{ och } V(\xi) = \sum_{k=1}^{n} p(1-p) = n\,p(1-p).$$

1.5 Fall då man inte har normalfördelning

För ξ_k ober. och likafördelade med $k = 1, 2, \ldots, n$ och $n \geq 30$ ser man

$$\zeta_n := \sum_{k=1}^{n} \xi_k$$

som normalfördelad:

$$\zeta_n \in \mathrm{N}(n\mu, \sqrt{n}\sigma) \text{ och } \overline{\xi} \in \mathrm{N}(\mu, \sigma/\sqrt{n}).$$

Ex 5.12 Vid vägning av enlitersförpackning av mjölk gjordes 50 mätningar (stickprovsstorlek 50) ξ_2, ξ_2, $\ldots, \xi_{50}$, enhet liter, och fick ett observerad summa $\zeta_{\text{obs}} = \sum x_k = 49.0$ (liter) och observerad standardavvikelse $s = \sqrt{{\sigma^2}^*_{\text{obs}}} = 0.1$. Gör en (a) punkt- och en (b) symmetrisk intervallskattning, (98%) av medelvärdet, d.v.s. bestäm ett dito konfidensintervall.

Lösning

(a)

$$\overline{x} = 0.98, \qquad \frac{s}{\sqrt{50}} = 0.0141421\ldots$$

(b) Man ser nu σ som känt och använder alltså inte $t-$fördelningens kvantiler. Tabell ger $\lambda_{0.01} = 2.32635$, så att konfidensintervallet är

$$[0.947, 1.013].$$

$\blacksquare$

Kommentarer

- Redan i exemplet med kolmonoxid från bilar, är det faktiskt CGS som används. Variabeln $\eta = \eta_n$ är en summa av ξ_k. Dessa är likafördelade indikatorvariabler: Vi tilldelar $\xi_k = 1$ om bil nummer k har för högt utsäpp och $\xi_k = 0$ om värdet är för lågt. Då gäller att

$$P(\xi_k = 1) = p = 0.05 \text{ och } P(\xi = 0) = 1 - p.$$

Då blir $\displaystyle\sum_{k=1}^{400} \xi_k =: \eta =\in \text{Bin}(400, 0.05)$ och denna summa är approximaivt $\in \text{N}(400p, 400p(1-p))$.

- Vi beräknar dens sökta sannolikheten för kolmonoxidutsläppet direkt med approximationen

$$\begin{aligned}
\sum_{k=1}^{400} \xi_k &\in \text{N}(20, 4.3589) \Longrightarrow \\
P(\eta > 30) &= 1 - \Phi\left(\frac{10}{4.3589}\right) 1 - \Phi(2.29) = \{\text{tabell}\} = 1 - 0.989 = 0.01.
\end{aligned}$$