Föreläsning 6: Statistisk inferens

Johan Thim (johan.thim@liu.se)

January 10, 2022

Vi är nu redo för att dyka ned i statistisk inferensteori! I sedvanlig ordning börjar vi med att definiera lite begrepp så vi är överens om vad det är vi diskuterar.

Stickprov

Definition. Låt de stokastiska variablerna X_1, X_2,..., X_n vara oberoende och ha samma fördelningsfunktion F. Följden $X_1, X_2, \ldots, X_n$ kallas ett **slumpmässigt stickprov** (av F). Ett **stickprov** x_1, x_2,..., x_n består av **observationer** av variablerna X_1, X_2,..., X_n. Samtliga möjliga observationer brukar kallas **populationen**. Vi säger att **stickprovsstorleken** är n.

Exempel

Antag att vi kastar en perfekt tärning 5 gånger och att dessa kast är oberoende. Före kasten representerar den stokastiska variabeln X_k resultatet vid kast k där alla X_k har samma fördelning; vi vet ännu inte vad resultatet blir, men känner sannolikhetsfördelningen. Följden X_k, $k = 1, 2, \ldots, 5$ är det *slumpmässiga stickprovet*.

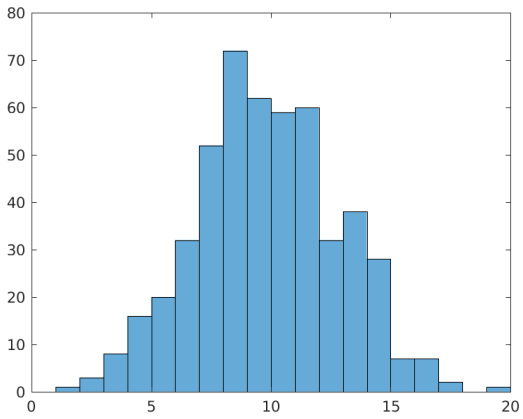
Efter kasten har vi erhållit observationer $x_1, x_2, \ldots, x_5$ av det slumpmässiga stickprovet. Detta är vårt *stickprov* och består alltså av utfallen vid kasten. Dessa observationer tillhör *populationen*. *Stickprovsstorleken* är 5.

Värt att notera är att språkbruket ibland är slarvigt där både stickprov och slumpmässigt stickprov används för att beskriva både följden av stokastiska variabler och följden av observationer (utfall). Det viktiga är att hålla koll på vad ni själva menar när ni genomför analyser.

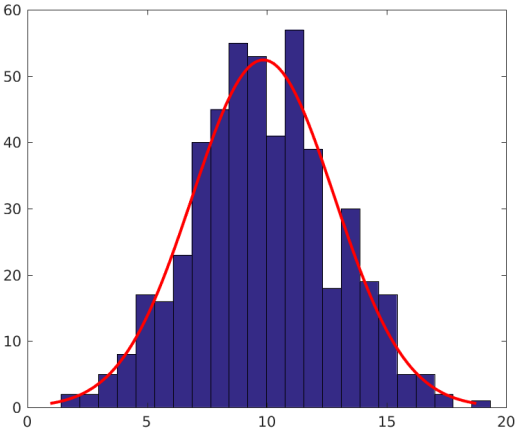
1 Representation av stickprov

Man kan representera statistiska data på en hel drös olika sätt med allt från tabeller till stolpdiagram till histogram till lådplottar. Läs avsnittet i boken om detta. Vi nöjer oss med att titta lite närmare på de verktyg vi kommer använda oss av i kursen. Ett mycket vanligt sätt att visualisera fördelningen för en mängd data är med hjälp av histogram. Vi genererar lite normalfördelad slumpdata i MATLAB och renderar ett histogram.

```
>> U = normrnd(10,3,500,1);
>> histogram(U);
```

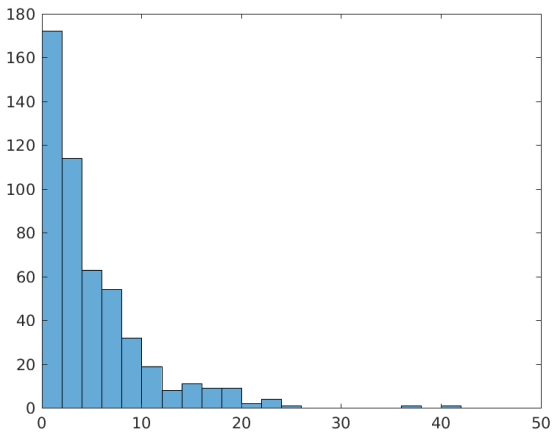



```
>> U = normrnd(10,3,500,1);
>> histfit(U)
```



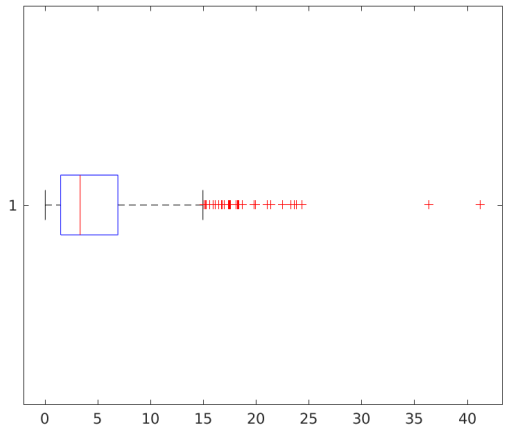
Om vi testar med exponentialfördelning istället blir resultatet enligt nedan.

```
>> U = exprnd(10,500,1);
> histogram(U);
```

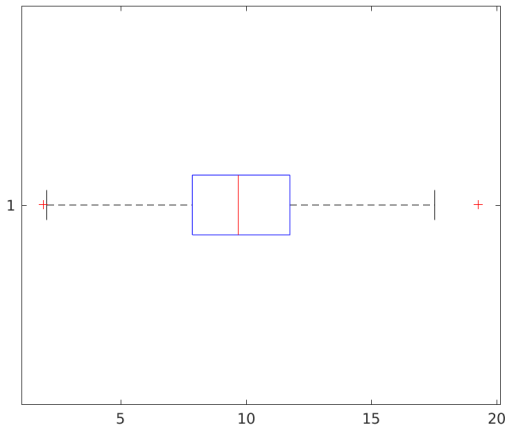


Ett lådagram (boxplot) representerar också materialet, kanske på ett enklare sätt för den oinsatte i sannolikhetsfördelningar. Lådan innehåller 50% av resultaten och den vänstra lådkanten är den undre kvartilen (25% till vänster om den) och den högra är den övre kvartilen (med 25% till höger om den). Medianen markeras med ett streck i lådan. Maximum och minimum markeras med små vertikala streck i slutet på en horisontell linje genom mitten på lådan. Värden som bedöms vara uteliggare markeras med kryss längs samma centrumlinje.

```
>> U = exprnd(5,500,1);
>> boxplot(U, 'orientation', 'horizontal');
```




```
>> U = normrnd(10,3,500,1);
>> boxplot(U, 'orientation', 'horizontal');
```

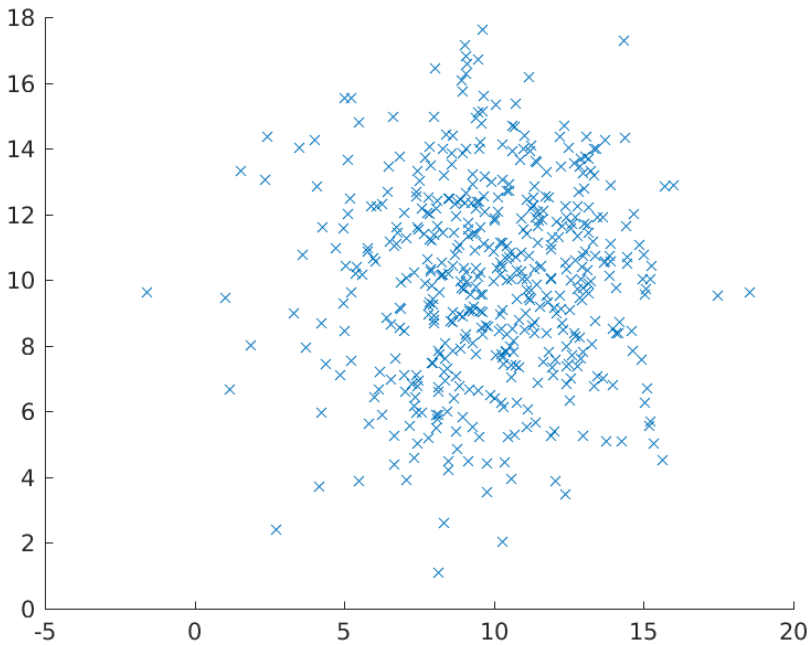


Vi kan tydligt se skillnad på hur mätvärden är spridda. Jämför även med motsvarande histogram ovan.

1.1 Två-dimensionell data

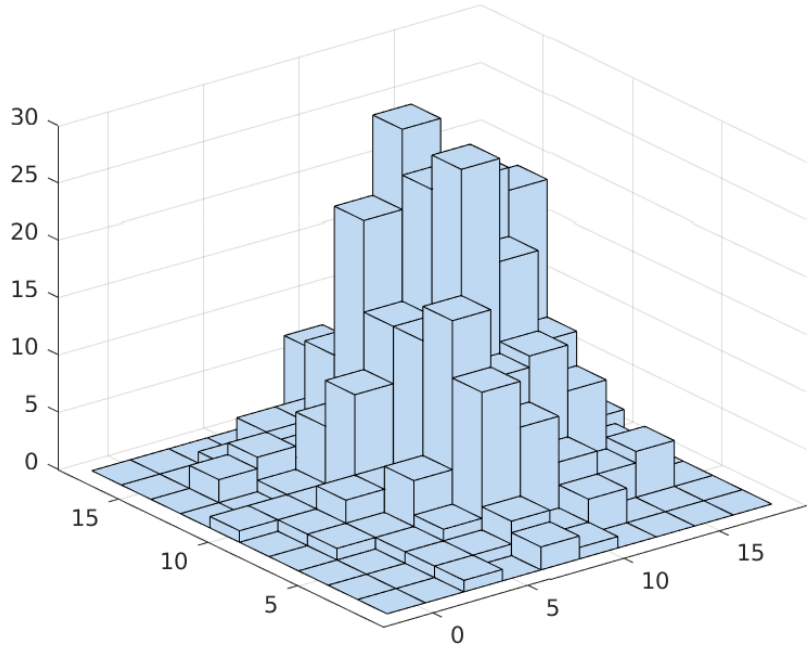
Vi kan även ha mätvärden i form av punkter (x, y) och den vanligaste figuren i dessa sammanhang är ett **spridningsdiagram** (scatter plot) där man helt enkelt plottar ut punkter vid varje koordinat (x_i, y_i).

```
>> U = normrnd(10,3,500,2);
>> scatter(U(:,1), U(:,2), 'x');
```



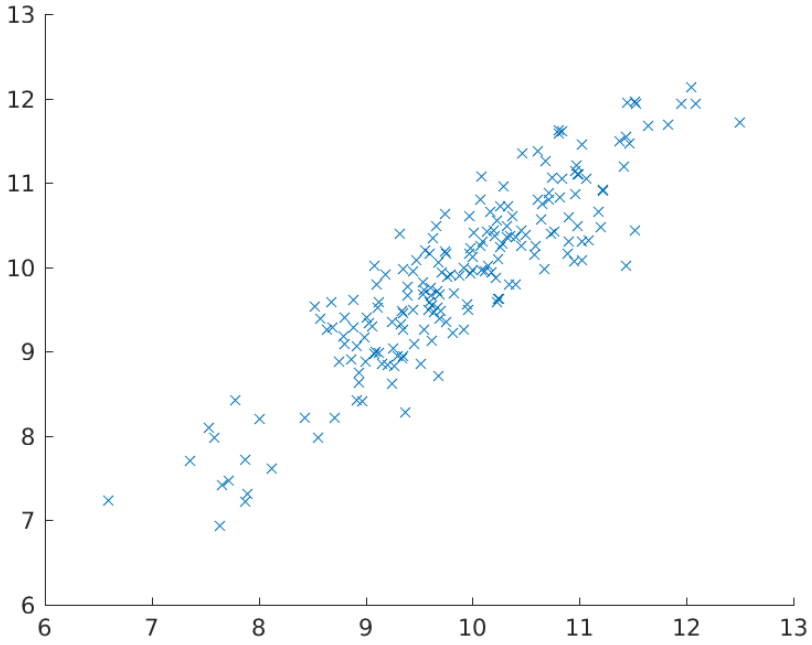
Vi ser från figuren att värdena verkar vara centrerade kring $(10, 10)$ och att det verkar föreligga någon form av cirkulär symmetri. Stämmer det för den bivariata normalfördelningen när komponenterna är oberoende? Vi kan även rendera ett två-dimensionellt histogram.

```
>> hist3(U);
```



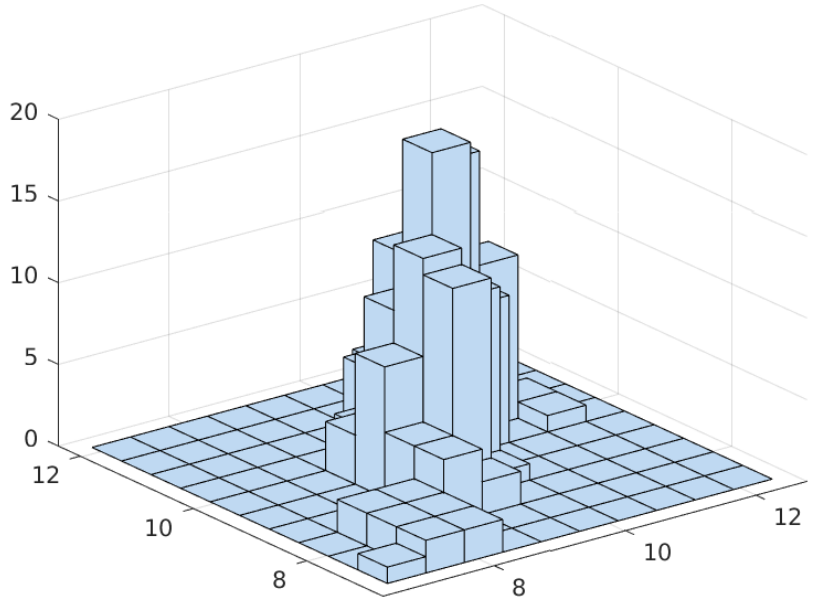
Vad gäller om variablerna i den bivariata normalfördelningen inte är oberoende?

```
>> mu = [10 10]; rho = 0.90; s1 = 1; s2 = 1;
>> Sigma = [s1*s1 s1*s2*rho; s1*s2*rho s2*s2]
>> R = chol(Sigma);
>> z = repmat(mu, 200, 1) + randn(200,2)*R;
>> scatter(z(:,1),z(:,2), 'x');
```



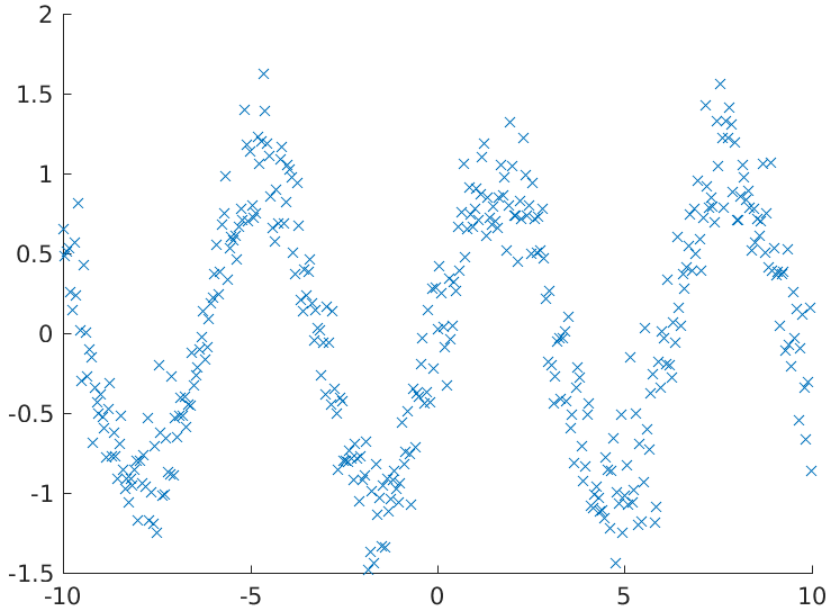
Värdena verkar fortfarande vara centrerade kring $(10, 10)$ (i någon mening) men symmetrin verkar nu utdragen diagonalt. Stämmer det för en bivariat normalfördelningen med korrelationen 0.90? Ett histogram kan genereras som ovan.

```
>> hist3(z);
```



Något konstigare? Visst.

```
>> x = (-10:0.05:10); y = sin(x) + normrnd(0,0.25,size(x));
>> scatter(x,y,'x');
```



Vi kan tydligt urskilja sinus-termen och något slags brus som gör att det inte blir en perfekt linje. Kan man få bort bruset?

2 Punktskattningar

Antag att en fördelning beror på en okänd parameter θ. Med detta menar vi att fördelningens täthetsfunktion (eller sannolikhetsfunktion) beror på ett okänt tal θ, och skriver $f(x\,;\,\theta)$ respektive $p(k\,;\,\theta)$ för att markera detta. Om vi har ett stickprov från en fördelning med en okänd parameter, kan vi *skatta* den okända parametern? Med andra ord, kan vi göra en ”gissning” på det verkliga värdet på parametern θ?

Punktskattning

Definition. En **punktskattning** $\widehat{\theta}$ av parametern θ är en funktion (ibland kallad **stickprovsfunktion**) av de observerade värdena $x_1, x_2, \ldots, x_n$:

$$\widehat{\theta} = g(x_1, x_2, \ldots, x_n).$$

Vi definierar motsvarande **stickprovsvariabel** $\widehat{\Theta}$ enligt

$$\widehat{\Theta} = g(X_1, X_2, \ldots, X_n).$$

Det är viktigt att tänka på att $\widehat{\theta}$ är en siffra, beräknad från de observerade värdena, medan $\widehat{\Theta}$ är en stokastisk variabel. Som vanligt använder vi stora bokstäver för att markera att vi syftar på en stokastisk variabel. Sambandet mellan $\widehat{\theta}$ och $\widehat{\Theta}$ är alltså att $\widehat{\theta}$ är en observation av den stokastiska variabeln $\widehat{\Theta}$. Förutom detta dras vi fortfarande med det okända talet θ, som inte är stokastiskt, utan endast en okänd konstant.

Exponentialfördelning

Betrakta en exponentialfördelning med okänt väntevärde. Formeln är välkänd: för alla $x \geq 0$ gäller att $f(x\,;\,\mu) = \mu^{-1}\exp(\mu^{-1}x)$. Parametern θ är alltså väntevärdet μ i detta fall. Ibland använder man exponentialfördelningen för att beskriva elektriska komponenters livslängd, och genom att betrakta ett stickprov kan man då uppskatta livslängden för en hel tillverkningsomgång.

Stokastiskt eller ej?

Var noggran med att tydligt visa och göra skillnad på vad som är stokastiskt eller inte i din redovisning! Vi har tre storheter:

(i) θ – verkligt värde. Okänt. Deterministiskt.

(ii) $\widehat{\theta}$ – skattat värde. Känt (beräknat från stickprovet). Deterministiskt.

(iii) $\widehat{\Theta}$ – stickprovsvariabeln. Denna är stokastisk!

Sannolikheter som beräknas bör använda sig av $\widehat{\Theta}$ då $\widehat{\Theta}$ beskriver variationen hos $\widehat{\theta}$ för olika stickprov. Om bara $\widehat{\theta}$ och θ ingår är sannolikheten alltid noll eller ett (varför?).

Så om vi har ett stickprov från en fördelning som beror på en okänd parameter, hur hittar vi skattningsfunktionen g? Fungerar vad som helst?

Exempel

Vid fyra dagar på en festival gjordes ljudnivåmätningar vid lunchtid. Följande mätdata erhölls: 107dB, 110dB, 117dB, 101dB. Vi antar att mätningarna är observationer av oberoende och likafördelade variabler med okänt väntevärde μ. Hur hittar vi en skattning $\widehat{\mu}$?

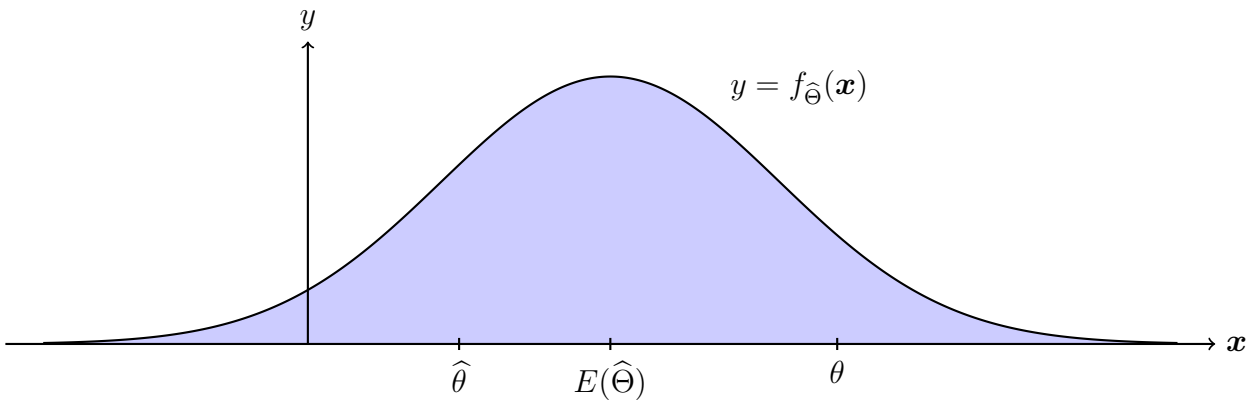
(i) $\widehat{\mu} = 100$dB är en skattning.

(ii) $\widehat{\mu} = 107$dB (den första dagen) är en skattning.

(iii) $\widehat{\mu} = (107 + 110 + 117 + 101)/4 = 108.75$dB (medelvärdet) är en skattning.

(iv) $\widehat{\mu} = \min\{107, 110, 117, 101\} = 101$dB är en skattning.

Så svaret är i princip "ja," alla värden $\widehat{\theta}$ som är tillåtna i modellen vi betraktar är punktskattningar. Hur väljer vi då den bästa, eller åtminstone en bra, punktskattning? Stickprovsvariabeln $\widehat{\Theta}$ är en stokastisk variabel, så normalt sett har den en täthetsfunktion (alternativt sannolikhetsfunktion). Vi skisserar en tänkbar täthetsfunktion för $\widehat{\Theta}$ (tänk dock på att $\widehat{\Theta}$ typisk är en flerdimensionell stokastisk variabel då $\widehat{\Theta} = g(X_1, X_2, \ldots, X_n)$).



Vi vet att $\widehat{\theta}$ beräknas från observerade siffror, så $\widehat{\theta}$ kan hamna lite var som helst. Dessutom vet vi inte om väntevärdet $E(\widehat{\Theta})$ sammanfaller med det okända värdet θ. Så hur kan vi då avgöra om en punktskattning är bra eller inte? Det finns två viktiga kriterier: *väntevärdesriktighet* och *konsistens*. Vi återkommer till dessa nästa föreläsning.

Väntevärdesriktig skattning

Definition. Stickprovsvariabeln $\widehat{\Theta}$ kallas **väntevärdesriktig** (vvr) om $E(\widehat{\Theta}) = \theta$.

Om en punktskattning inte är väntevärdesriktig pratar man ibland om ett systematiskt fel. Vi definierar detta som skillnaden $E(\widehat{\Theta}) - \theta$. En väntevärdesriktig skattning $\widehat{\Theta}$ har alltså inget systematiskt fel; i "medel" kommer den att hamna rätt (tänk på de stora talens lag).

Systematiskt fel; bias

Definition. Om $E(\widehat{\Theta}) - \theta \neq 0$ så säger vi att $\widehat{\Theta}$ har ett systematiskt fel (ett bias).

2.1 Vanliga punktskattningar

Vissa punktskattningar är så vanliga att de ha fått egna namn. Vi vill ofta skatta medelvärdet som positionsmått och stickprovsstandardavvikelsen är ett vanligt mått på spridningen.

Stickprovsmedelvärde

Definition. Stickprovsmedelvärdet $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ är en observation av $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ som skattar väntevärdet.

Stickprovsvarians och stickprovsstandardavvikelse

Definition. Stickprovsvariansen, definierad av $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$, är en observation av $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ som skattar variansen. Stickprovsstandardavvikelsen $s = \sqrt{s^2}$ är en skattning av σ.

Varför $n-1$? Vi återkommer till det nästa föreläsning.

3 Vilka skattningar är bra?

När vi har ett stickprov från en fördelning som beror på en okänd parameter så fungerar alltså i princip vad som helst som skattning på parametern. Funktionen g är således godtycklig. Vi vet att $\widehat{\theta}$ beräknas från observerade siffror, så $\widehat{\theta}$ kan hamna lite var som helst. Dessutom vet vi inte om väntevärdet $E(\widehat{\Theta})$ sammanfaller med det okända värdet θ. Så hur kan vi då avgöra om en punktskattning är bra eller inte? Det finns två viktiga kriterier: *väntevärdesriktighet* som vi såg ovan och *konsistens*.

Om en punktskattning inte är väntevärdesriktig pratar man ibland om ett systematiskt fel. Vi definierar detta som skillnaden $E(\widehat{\Theta}) - \theta$. En väntevärdesriktig skattning $\widehat{\Theta}$ har alltså inget systematiskt fel; i ”medel” kommer den att hamna rätt (tänk på de stora talens lag). Vi vill också gärna ha egenskapen att en punktskattning blir bättre ju större stickprov vi använder.

Konsistent skattning

Definition. Antag att vi har en punktskattning $\widehat{\Theta}_n$ för varje stickprovsstorlek n. Om det för varje $\epsilon > 0$ gäller att

$$\lim_{n\to\infty} P(|\widehat{\Theta}_n - \theta| > \epsilon) = 0,$$

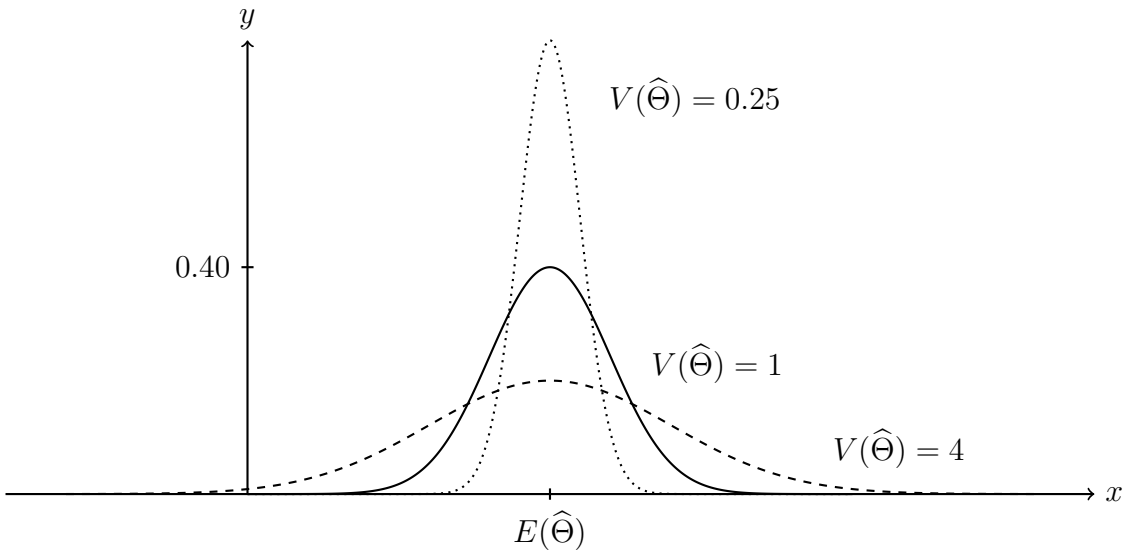
så kallar vi denna punktskattning för *konsistent*.

Teknisk definition, men innebörden bör vara klar. När stickprovsstorleken går mot oändligheten så är sannolikheten att skattningen befinner sig nära det okända värdet stor. Villkoret för konsistens kan vara lite jobbigt att arbeta med så följande sats är ofta användbar för att kontrollera konsistens.

Ett kriterium för konsistens

Om $E(\widehat{\Theta}_n) = \theta$ för alla n och $\lim_{n\to\infty} V(\widehat{\Theta}_n) = 0$ så är skattningen konsistent.

Bevisskiss: Här använder vi Tjebysjovs olikhet: om $a > 0$ och X är en stokastisk variabel så att $E(X) = \mu$ och $V(X) = \sigma^2 < \infty$, så gäller $P(|X - \mu| > a\sigma) \leq \dfrac{1}{a^2}$. Om vi låter $a = \epsilon/\sigma_n$ för fixt $\epsilon > 0$ så erhåller vi $P(|\widehat{\Theta} - \theta| > \epsilon) \leq \dfrac{\sigma_n^2}{\epsilon^2} \to 0$ då $n \to \infty$, eftersom $\sigma_n^2 = V(\widehat{\Theta}_n) \to 0$ då $n \to \infty$. □



Mindre varians för $\widehat{\Theta}$ medför att sannolikhetsmassan är mer centrerad kring väntevärdet (vid symmetrisk fördelning).

Exempel

Betrakta exemplet med ljudnivåerna igen, vi hade följande mätdata: 107dB, 110dB, 117dB, 101dB. Vi undersöker skattningarna lite närmare.

(i) $\widehat{\mu} = 100$dB är en fix siffra och kan varken vara väntevärdesriktig eller konsistent. Dålig skattning.

(ii) Den första siffran är en observation av den första variabeln X_1 i stickprovet. Alltså är $\widehat{M} = X_1$. Eftersom $E(\widehat{M}) = E(X_1) = \mu$ så är skattningen väntevärdesriktig. Med konstant varians oavsett stickprovsstorlek kan den dock inte vara konsistent.

(iii) Medelvärdet är både väntevärdesriktigt och konsistent; se nästa avsnitt!

När det gäller det sista förslaget så blir det lite klurigare när vi bildar minimum av observationerna. Om vi undersöker ett specialfall där variablerna är exponentialfördelade, säg till exempel att $X_i \sim \text{Exp}(\mu)$, så kan man faktiskt visa att (inte självklart)

$$\widehat{M} = \min\{X_1, X_2, X_3, X_4\} \sim \text{Exp}(\mu/4).$$

Således erhåller vi att $E(\widehat{M}) = \mu/4 \neq \mu$. Alltså är inte detta en väntevärdesriktig skattning av μ. Notera att vi här behövde använda fördelningen för variablerna X_i för att kunna svara på frågan dessutom. Går det att korrigera skattningen (så att den blir väntevärdesriktig) i detta fall?

3.1 Effektivitet – jämförelse mellan skattningar

Så om vi har två olika stickprovsvariabler $\widehat{\Theta}$ och Θ^*, hur avgör vi vilken som är ”bäst”? Om båda är väntevärdesriktiga och konsistenta, kan man säga att en är bättre?

Effektivitet

Definition. En skattning $\widehat{\Theta}$ kallas *effektivare* än en skattning Θ^* om $V(\widehat{\Theta}) \leq V(\Theta^*)$.

Den stickprovsvariabel med minst varians kallas alltså mer effektiv, och med mindre varians känns det rimligt att kalla den skattningen bättre (om den är någorlunda väntevärdesriktig).

4 Momentmetoden

Så kan man systematiskt finna lämpliga skattningar på något sätt om man känner till viss information om fördelningen? Svaret är ja, det finns många sådana metoder. Bland annat momentmetoden, MK-metoden (minsta kvadrat), och kanske den vanligaste, ML-skattningar (maximum likelihood). Vi börjar med att betrakta momentmetoden.

Momentmetoden (för en parameter)

Definition. Låt $E(X_i) = \mu(\theta)$ för alla i. Momentskattningen $\widehat{\theta}$ av θ fås genom att lösa ekvationen $\mu(\widehat{\theta}) = \overline{x}$.

Exempel

Låt $x_1, x_2, \ldots, x_n$ vara ett stickprov från en fördelning med täthetsfunktionen $f(x\,;\,\theta) = \theta e^{-\theta x}$ för $x \geq 0$. Använd momentmetoden för att punktskatta θ.

Lösning: Vi börjar med att beräkna väntevärdet, det vill säga funktionen $\mu(\theta)$. Alltså,

$$\mu(\theta) = \int_0^\infty x\theta e^{-\theta x}\,dx = \left[x\theta\frac{e^{-\theta x}}{-\theta}\right]_0^\infty + \int_0^\infty e^{-\theta x} = \theta^{-1}.$$

Vi löser nu ekvationen $\mu(\widehat{\theta}) = \overline{x}$, och erhåller då att

$$\widehat{\theta}^{-1} = \overline{x} \quad \Leftrightarrow \quad \widehat{\theta} = \frac{1}{\overline{x}},$$

så länge $\overline{x} \neq 0$. Momentskattningen av θ ges alltså av $\widehat{\theta} = (\overline{x})^{-1}$. Vad händer om $\overline{x} = 0$?

Om man har flera parametrar då? Här visar det sig varför metoden ovan kallas för just *momentmetoden*.

Moment

Definition. Låt X vara en stokastisk variabel X. För $k = 1, 2, \ldots$ definierar vi momenten m_k för X enligt $m_k = E(X^k)$.

Det första momentet m_1 är alltså inget annat än väntevärdet för X.

Momentskattning med flera parametrar

Definition. Låt $X \sim F(x\,;\theta_1, \theta_2, \ldots, \theta_j)$ bero på j okända parametrar $\theta_1, \theta_2, \ldots, \theta_j$ och definiera $m_i(\theta_1, \theta_2, \ldots, \theta_j) := E(X^i)$, $i = 1, 2, \ldots$ Momentskattningarna för θ_k, $k = 1, 2, \ldots, j$, ges av lösningen till ekvationssystemet

$$m_i(\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_j) = \frac{1}{n}\sum_{k=1}^{n} x_k^i, \quad i = 1, 2, \ldots, j.$$

Observera att det inte är säkert att en lösning finns eller att lösningen är entydig i de fall den existerar. Vidare kan det även inträffa att lösningen hamnar utanför det område som är tillåtet för parametern (i vilket fall vi givetvis inte kan använda den).

Exempel

Låt $X_k \sim N(\mu, \sigma^2)$, $k = 1, 2, \ldots, n$ vara ett stickprov. Hitta momentskattningarna för μ och σ^2.

Lösning. Vi vet att $E(X) = \mu$ och $E(X^2) = V(X) + E(X)^2 = \sigma^2 + \mu^2$, så

$$\begin{cases} \widehat{\mu} = \overline{x}, \\ \widehat{\sigma}^2 + \widehat{\mu}^2 = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} x_k^2. \end{cases}$$

Således erhåller vi direkt att $\widehat{\mu} = \overline{x}$. För $\widehat{\sigma}^2$ är

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n} x_k^2 - \overline{x}^2 = \cdots = \frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})^2.$$

Nästan stickprovsvariansen alltså.

Vektornotation för parametrar

Definition. När vi har en fördelning som beror på flera parametrar, säg $\theta_1, \theta_2, \ldots, \theta_j$, så skriver vi ibland $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_j) \in \mathbf{R}^j$ som en j-dimensionell vektor. Notationen blir då mer kompakt. Bokstäver typsatta i fet stil indikerar oftast en vektor i denna kurs.