

Sannolikhet och statistik – Läsanvisningar

för kursboken:

Sven Erick Alm & Tom Britton: *Stokastik — Sannolikhetsteori och statistikteori med tillämpningar*. Liber 2008.

Följande avsnitt ingår: 1, 2.1–2.5, 3.1–3.5, 3.6, 3.7.1–3.7.3, 3.8.1–3.8.3, 3.10.1, 3.10.3, 3.11–3.13, 6, 7.1, 7.2.1–7.2.3, 7.2.6, 7.3, 7.5.2, 7.6, 9.1, 9.2.1, 9.2.2.

Kapitel 2

Hela kapitlet utom Avsnitt 2.6 ingår. Mycket bör vara bekant sedan tidigare, men kanske inte i **Avsnitt 2.5**. Läs definitionerna av *betingad sannolikhet* och *oberoende* och studera *Lagen om total sannolikhet* som lämpligen används med hjälp av *träddiagram*!

Kapitel 3

Här ingår det mesta av avsnitten 3.1–3.5, 3.6, 3.7.1–3.7.3, 3.8.1–3.8.3, 3.10.1, 3.10.3, 3.11–3.13.

I de första avsnitten (**3.2–3.4**) definieras många viktiga begrepp: *diskret slumpvariabel*, *sannolikhetsfunktion*, *fördelningsfunktion*, *kontinuerlig slumpvariabel* och *täthetsfunktion*. Vidare ges egenskaper för dessa i satserna 3.1–3.3. Observera speciellt att för en godtycklig slumpvariabel X det gäller att

$$P(a < X \leq b) = F_X(b) - F_X(a).$$

Om Y är en kontinuerlig slumpvariabel så gäller det att

$$P(a < Y \leq b) = \int_a^b f_Y(t) dt,$$

(vilket också medför att $P(Y = y) = 0$ för alla y), medan om Y är diskret

$$P(a < Y \leq b) = \sum_{\{k: a < k \leq b\}} p_Y(k).$$

Allmänt gäller det att sannolikhetsfunktionen för en diskret slumpvariabel motsvaras av täthetsfunktionen för en kontinuerlig och att summation i det diskreta fallet svarar mot integration i det kontinuerliga.

I **Avsnitt 3.5** definieras ett antal *lägesmått* och *spridningsmått*: *väntevärde*, *kvantil* (viktigt i statistikdelen av kursen), *varians* och *standardavvikelse*.

Notera speciellt Sats 3.4 om $E(g(X))$, Sats 3.5 om beräkning av varians och Sats 3.6 med räkneregler för väntevärde och varians.

Avsnitt 3.5.3 Olikheter tas inte upp på föreläsningarna.

Avsnitt 3.6 ger exempel på diskreta fördelningar: $\text{Be}(p)$, $\text{Bin}(n, p)$, $\text{Hyp}(N, n, p)$, $\text{Po}(\lambda)$ och $\text{ffg}(p)$, som samtliga finns i formelsamlingen.

Notera speciellt Sats 3.12 om förekomst av binomialfördelningen!

Träna tabellanvändning: fördelningsfunktionerna för binomial- och Poisson-fördelningarna finns tabellerade (Tabell 2 och 3).

Negativ binomialfördelning ingår inte i kursen.

Avsnitt 3.7 ger exempel på kontinuerliga fördelningar: $\text{Re}(a, b)$, $\text{Exp}(\beta)$, $\text{N}(\mu, \sigma^2)$ samt speciellt den standardnormala fördelningen $\text{N}(0, 1)$, som samtliga finns i formelsamlingen.

Notera speciellt sambandet mellan allmän och standardiserad normalfördelning (Sats 3.24) och de speciella beteckningarna $\Phi(x)$ för fördelningsfunktionen och λ_α för kvantiler hos $\text{N}(0, 1)$.

Studera även egenskaper för normalfördelningen i Sats 3.25, Följdsats 3.2, Sats 3.47 och Följdsats 3.6 (de två senare i avsnitt 3.10).

Det är viktigt att kunna hantera normalfördelningstabellerna (Tabell 4 och 5).

Avsnitt 3.8 behandlar *flerdimensionella slumpvariabler* och behövs främst för att kunna definiera *oberoende slumpvariabler* vilket är väsentligt inte minst i statistikdelen av kursen. Vidare definieras *kovarians* och *korrelationskoefficient* för att mäta beroende mellan slumpvariabler.

Viktiga resultat ges i Följdsats 3.3 ($E(X + Y) = E(X) + E(Y)$), Sats 3.33 (varians för en summa) och Följdsats 3.4 (varians för en summa av oberoende slumpvariabler). Resultatet i Följdsats 3.5 används flitigt i statistikdelen.

I **Avsnitt 3.10** ingår definitionen av *moment* i Avsnitt 3.10.1 och i Avsnitt 3.10.3 endast de viktiga resultaten (om normalfördelningar) i Sats 3.47 och Följdsats 3.6.

Avsnitt 3.11 behandlar *Stora talens lag* (Sats 3.50) som man bör förstå innebörden av liksom *Bernoullis sats* (Följdsats 3.7) om relativa frekvensers stabilisering.

Avsnitt 3.12 behandlar den oerhört användbara *Centrala gränsvärdessatsen* (CGS, Sats 3.51) och dess viktiga användning som approximation (Metod 3.2).

I **Avsnitt 3.13** tillämpas normalapproximation med hjälp av CGS. Notera speciellt att man bör använda *halvkorrektur* då man approximerar en heltalsvärd fördelning med en kontinuerlig! Vidare ges en approximation av binomialfördelningen med Poisson.

Tumregler för när dessa approximationer kan användas finns i formelsamlingen.

Kapitel 6

I **Avsnitt 6.2** definieras *medelvärde*, $\bar{x}$, (*stickprovs*)*varians*, s^2 , och (*stickprovs*)*standardavvikelse*, s . Observera att $s^2 = S_{xx}/(n - 1)$, där $S_{xx} = \sum x_i^2 - (\sum x_i)^2/n$.

Skilj på varians för en slumpvariabel och varians i ett stickprov!

Vidare definieras *median* och *kvartiler* som behövs för att konstruera *lådagram* i **Avsnitt 6.3**, där även *stam-blad-diagram* beskrivs. Man bör även veta skillnaden mellan *stolpdia-gram* som används för att illustrera diskreta datamaterial och *histogram* som används för kontinuerliga material.

I **Avsnitt 6.4** bör man känna till *kovarians* och *korrelationskoefficient* för tvådimensionella datamaterial och även här se upp så att man inte förväxlar dessa med motsvarande storheter för slumpvariabler.

Kapitel 7

Här ingår 7.1, 7.2.1–7.2.3, 7.2.6, 7.3, 7.5.2 och 7.6. Tonvikten ligger på konfidensintervall, men man bör även känna till hypotesprövning (Avsnitt 7.6) och kunna utföra sådan med hjälp av konfidensmetoden.

I **Avsnitt 7.1** definieras *stickprov* och *slumpmässigt stickprov*. När det senare i texten talas om stickprov avses, om inte annat sägs, slumpmässigt stickprov. Skilj på observationerna i stickprovet, x , som är numeriska värden, och de slumpvariabler, X , som de är observationer av.

I **Avsnitt 7.2** definieras *skattning*, som är ett numeriskt värde, och *estimator*, som är motsvarande slumpvariabel. Om det inte finns risk för sammanblandning används skattning som benämning för båda.

Egenskaper för skattningar behandlas i **Avsnitt 7.2.1**. Man bör känna till vad som menas med att en skattning är *väntevärdesriktig* och även förstå att *medelfelet* är en skattning, dvs. en numerisk uppskattning, av standardavvikelsen för en estimator.

Begreppen i **Avsnitt 7.2.2** är av mer teoretiskt intresse. Man bör dock känna till begreppet *konsistens*.

I boken behandlas tre olika skattningsmetoder (i avsnitten 7.2.3–7.2.5). I kursen ingår **endast Momentmetoden** i **Avsnitt 7.2.3**. De övriga metoderna är Maximum-likelihoodmetoden (Avsnitt 7.2.4) och Minstakvadratmetoden (Avsnitt 7.2.5). Den senare används i Kapitel 9. **Avsnitt 7.2.3** ger också egenskaper för stickprovsmedelvärdet, $\bar{x}$, som skattning av väntevärdet, μ , och stickprovsvariansen, s^2 , som skattning av variansen, σ^2 .

I **Avsnitt 7.2.6** tillämpas detta på standardfördelningar: normalfördelning, binomialfördelning, hypergeometrisk fördelning och Poisson-fördelning.

Avsnitt 7.3 är viktigt. I **Avsnitt 7.3.1** definieras begreppet *konfidensintervall*. Observera att ett konfidensintervall är ett numeriskt intervall som är en observation av ett stokastiskt intervall. Egenskaperna hos konfidensintervallet, t.ex. *konfidensgraden*, bestäms av det stokastiska intervallet. Jämför med skillnaden mellan skattning och estimator.

I **Avsnitt 7.3.2** ges en allmän metod för att konstruera konfidensintervall baserade på en *referensvariabel*. Det är **viktigt** att förstå hur man använder en referensvariabel för att få fram konfidensintervallet eftersom formelsamlingen inte ger konfidensintervallen utan endast den lämpliga referensvariabeln och dess fördelning! Läs **Metod 7.1**!

Avsnitt 7.4 ingår bara orienterande och i **Avsnitt 7.5** ingår **endast** t -fördelningen i **Avsnitt 7.5.2**. Här räcker det att man känner till att fördelningen är symmetrisk och att man kan slå upp dess kvantiler i Tabell 6. Träna tabellanvändning!

I **Avsnitt 7.6** tillämpas inferensteorin på standardfördelningar. Läs i första hand det som handlar om konfidensintervall och bara orienterande om hypotesprövning och koncentrera dig på referensvariabler (och konfidensintervall). Följande normalfördelningssituationer (i avsnitten 7.6.1–7.6.3) bör man känna till (och behärska):

- Ett stickprov från normalfördelning, $N(\mu, \sigma^2)$,
 - konfidensintervall för μ då σ är känd,
 - konfidensintervall för μ då σ är okänd.

- Två stickprov från normalfördelningar, $N(\mu_1, \sigma_1^2)$ och $N(\mu_2, \sigma_2^2)$,
 - konfidensintervall för $\mu_1 - \mu_2$ då σ_1 och σ_2 är kända,
 - konfidensintervall för $\mu_1 - \mu_2$ då $\sigma_1 = \sigma_2 = \sigma$, där σ är okänd,
 - konfidensintervall för $\mu_1 - \mu_2$ då σ_1 och σ_2 är helt okända. I detta fall använder vi inte approximation med t -fördelning som i boken utan approximerar med $N(0, 1)$ som i formelsamlingen.
- Stickprov i par.

I **Avsnitt 7.6.4** behandlas *Normalapproximation*. Observera att oavsett om skattningens standardavvikelse är känd eller skattas med medelfelet approximeras den standardiserade skattningen med $N(0, 1)$ **utom** i det viktiga specialfallet att stickprovet kommer från en slumpvariabel X med $E(X) = \mu$ och $V(X) = \sigma^2$, där båda parametrarna är okända och σ^2 skattas med stickprovsvariansen s^2 . Då används att

$$R_\mu := \frac{\bar{X} - \mu}{s/\sqrt{n}} \approx t(n-1),$$

om stickprovet är av storlek n .

Vidare tillämpas normalapproximation på

- binomialfördelning, I_p och $I_{p_1-p_2}$,
- hypergeometrisk fördelning, I_p och $I_{p_1-p_2}$,
- Poisson-fördelning, I_μ ,
- Poisson-fördelning, $I_{\mu_1-\mu_2}$.

Kapitel 9

Här ingår 9.1, 9.2.1 och 9.2.2.

Avsnitt **Avsnitt 9.1** ger en introduktion via ett antal exempel medan **Avsnitt 9.2** ger en matematisk modell för *enkel linjär regression*. Observera att det **inte** är symmetri mellan y och x . x -värdena betraktas som fixa medan y -värdena är observationer av slumpvariabler, $y = \alpha + \beta \cdot x + \text{slumpf}$.

I avsnitt **Avsnitt 9.2.1** skattas parametrarna α och β och i **Avsnitt 9.2.2** ges statistiska egenskaper för skattningarna. Vidare införs begreppet *förklaringsgrad* som anger hur stor del av y -variationen som kan förklaras av x . Slutligen beskrivs hur man kan skatta variansen, σ^2 , för slumpfelen och konstruera konfidensintervall för α och β .