Sannolikhet och statistik sammanfattning

Johan Jonasson

Maj 2020

Innehållet i denna föreläsning är en något utökad version av de detaljerade lärandemålen på Canvas.

- Utfallsrum: mängden av allt som kan hända vid ett slumpförsök.
- Utfall: ett element i utfallsrummet.
- Händelse: delmängd av utfallsrummet.

Här ska man kunna förstå vad en händelse uttryckt informellt betyder sedd som delmängd till utfallsrummet.
Exempel. Singla slant tre gånger.
$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$. Om A är händelsen att man får exakt en krona (heads), gäller

$$A = \{HTT, THT, TTH\}.$$

□

En funktion $\mathbb{P} : \mathcal{P}(S) \to \mathbb{R}$ kallas för ett sannolikhetsmått om

- För alla A gäller $0 \leq \mathbb{P}(A) \leq 1$,
- $\mathbb{P}(S) = 1$,
- För alla $A_1, A_2, A_3, \ldots$ sådana att $A_i \cap A_j = \emptyset$ för alla $i \neq j$, gäller

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Några enkla följder av definitionen: För alla A och B gäller

- $\mathbb{P}(\emptyset) = 0$,
- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$,
- $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$,
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$,
- Om $A \subset B$: $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$,
- Om $A \subset B$: $\mathbb{P}(A) \leq \mathbb{P}(B)$.

Kombinatorik:

- Multiplikationsprincipen,
- Antalet sätt att välja k element ur en mängd med n element med återläggning och med hänsyn till ordning är n^k,
- utan återläggning med hänsyn till ordning $n(n-1)\cdots(n-k+1)$,
- utan återläggning utan hänsyn till ordning

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!}.$$

Det klassiska sannolikhetsmåttet: Om S är ändligt och alla utfall är lika sannolika gäller

$$\mathbb{P}(A) = \frac{|A|}{|S|}.$$

Två händelser A och B sägs vara oberoende om

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Mer generellt: Tre händelser A_1, A_2, A_3 sägs vara oberoende om $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$ för alla $i \neq j$ och om $\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)$.

Mest generellt: $A_1, A_2, \ldots$ sägs vara oberoende om det för alla n och alla $i_1, i_2, \ldots, i_n$ gäller att

$$\mathbb{P}\left(\bigcap_{k=1}^{n} A_{i_k}\right) = \prod_{k=1}^{n} \mathbb{P}(A_{i_k}).$$

Betingad sannolikhet: Om $\mathbb{P}(A) > 0$ definierar vi

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}.$$

Om $\mathbb{P}(A) > 0$ gäller att A och B är oberoende om och endast om $\mathbb{P}(B|A) = \mathbb{P}(B)$.

Totala sannolikhetslagen: Om $A_1, A_2, \ldots, A_n$ är en partion av S och B är en händelse gäller

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i).$$

Bayes formel:

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{j=1}^{n} \mathbb{P}(B|A_j)\mathbb{P}(A_j)}.$$

En stokastisk variabel är en funktion $X : S \to \mathbb{R}$.
Fördelningfunktionen för X är funktionen $F : \mathbb{R} \to [0, 1]$ given av

$$F(x) = \mathbb{P}(X \leq x).$$

X är diskret om den kan anta högst ett ändligt eller uppräkneligt antal olika värden: $V_X = \{x_1, x_2, \ldots\}$.
Om X är diskret, låt $p(x_k) = \mathbb{P}(X = x_k)$. Då kallas p för frekvensfunktionen till X. I så fall gäller för alla $B \subseteq V_X$ att $\mathbb{P}(X \in B) = \sum_{k:x_k \in B} p(x_k)$.

X kallas kontinuerlig om det finns en täthetsfunktion $f : \mathbb{R} \to [0, \infty)$ sådan att för alla $B \subseteq \mathbb{R}$

$$\mathbb{P}(X \in B) = \int_B f(x)dx.$$

I så fall gäller att $f(x) = F'(x)$ är en täthet.

Väntevärdet för en diskret sv X ges av

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} x_k \mathbb{P}(X = x_k).$$

Om X är kontinuerlig definerar vi

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x\, f(x) dx.$$

Väntevärdet är "medelvärdet i det långa loppet" eller tyngdpunkten hos p eller f.
Om X är positiv och kontinuerlig gäller att

$$\mathbb{E}[X] = \int_{0}^{\infty} \mathbb{P}(X > x) dx.$$

Motsvarande om X är diskret och $V_X \subseteq \{0, 1, 2, \ldots\}$ är

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} \mathbb{P}(X > k) = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k).$$

Variansen av en stokastisk variabel X med väntevärde μ ges av

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}[(X - \mu)^2].$$

Steiners formel:

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Standardavvikelsen är $Std(X) = \sqrt{\mathbb{V}\mathrm{ar}(X)}$.

Om X är en sv och $Y = g(X)$, bestäm fördelningen för Y genom att utnytta att

$$F_Y(y) = \mathbb{P}(g(X) \leq y)$$

och om möjligt lösa ut X i olikheten i sannolikhetsuttrycket.

OSL:

$$\mathbb{E}[g(X)] = \sum_{k=1}^{\infty} g(x_k)p(x_k)$$

eller

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx.$$

En kont sv är $likf(a, b)$-fördelad om

$$f(x) = \frac{1}{b-a},\ a < x < b.$$

$$\mathbb{E}[X] = \frac{a+b}{2}.$$

$$\mathbb{V}\mathrm{ar}(X) = \frac{(b-a)^2}{12}.$$

Diskreta fördelningar:
En sv som antar värdet 1 då en viss händelse A inträffar och 0 annars kallas för en indikatorvariabel för A. Den skrivs I_A. Det gäller att $p(1) = 1 - p(0) = \mathbb{P}(A) =: p$. Vi får $\mathbb{E}[I_A] = p$ och $Var(I_A) = p(1-p)$.

Upprepa försöket n ggr. Låt A_j vara händelsen att A inträffar vid försök nummer j. Låt

$$X = \sum_{j=1}^{n} I_{A_j}.$$

Då är $X \sim Bin(n, p)$ och

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k},\ 0 \leq k \leq n.$$

$$\mathbb{E}[X] = np,\ \mathbb{V}\mathrm{ar}(X) = np(1-p).$$

Om man upprepar försöket tills A inträffar för första gången och låter X vara antal försök som krävdes gäller att $X \sim Geo(p)$ och

$$\mathbb{P}(X = k) = p(1-p)^{k-1},\ k = 1, 2, \ldots.$$

Vi får

$$\mathbb{E}[X] = \frac{1}{p},\ \mathbb{V}\mathrm{ar}(X) = \frac{1-p}{p^2}.$$

Att säga att $X \sim Poi(\lambda)$ betyder att

$$\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!},\ k = 0, 1, 2, \ldots.$$

Uppstår som antalet incidenter av en typ som inträffar på ett sådant sätt att när en incident inträffar inte beror av när andra incidenter inträffar.

$$\mathbb{E}[X] = \lambda,\ \mathbb{V}\mathrm{ar}(X) = \lambda.$$

Kontinuerliga fördelningar:
$X \sim exp(\lambda)$ om $f(x) = \lambda e^{-\lambda x}$, $x > 0$. Detta är ekvivalent med $\mathbb{P}(X > x) = e^{-\lambda x}$, $x > 0$.

Exponentialfördelningen är den unika fördelningen som uppfyller glömskegeneskapen att $\mathbb{P}(X > x + t | X > x) = \mathbb{P}(X > t)$ för alla $x, t > 0$. Detta medför bl.a. att

$X_1 \sim exp(\lambda_1)$, $X_2 \sim exp(\lambda_2)$, X_1, X_2 oberoende
$\Rightarrow \min(X_1, X_2) \sim exp(\lambda_1 + \lambda_2)$.

$$\mathbb{E}[X] = \frac{1}{\lambda}, \mathbb{V}\mathrm{ar}(X) = \frac{1}{\lambda^2}.$$

Poissonprocessen:

Om $0 < \tau_1 < \tau_2 < \tau_3 < \ldots$ är slumpmässiga tidpunkter (då någon sorts impulser inträffar), sådana att tidsmellanrummen $\tau_1, \tau_2 - \tau_1, \tau_3 - \tau_2, \ldots$ är oberoende och $exp(\lambda)$-fördelade, kallas $\{\tau_1, \tau_2, \tau_3, \ldots\}$ för en Poissonprocess med intensitiet λ.

Skriv $X(s,t)$ för antalet impulser i (s,t). Det gäller att $X(s,t) \sim Poi(\lambda(t-s))$. Tack vare glömskeegenskapen är antalet impulser i disjunkta tidsintervall oberoende. En konsekvens av detta är att

$X_1 \sim Poi(\lambda_1)$, $X_2 \sim Poi(\lambda_2)$, X_1 och X_2 oberoende
$\Rightarrow X_1 + X_2 \sim Poi(\lambda_1 + \lambda_2)$.

En konsekvens av detta är att om man har två oberoende Poissonprocesser med int λ_1 och λ_2 gäller att processen man får genom att räkna impulserna från dem båda, får man en Poissonprocess med intensitet $\lambda_1 + \lambda_2$, den s.k. sammanvägda Poissonprocessen.

Om man i en Poissonprocess med int λ bara räknar impulserna med sannolikhet p, oberoende för olika impulser, får man en uttunnad Poissonprocess, vars intensitet är λp.

Om $T_1, T_2, \ldots$ är oberoende och $exp(\lambda)$-fördelade sägs

$$\sum_{k=1}^{n} T_k \sim \Gamma(n, \lambda).$$

Följd: $\tau_n \sim \Gamma(n, \lambda)$.

För $X \sim \Gamma(n, \lambda)$ gäller $\mathbb{E}[X] = n/\lambda$, $\mathbb{V}\mathrm{ar}(X) = n/\lambda^2$.

Gammafördelningen användes för att härleda resultat kring Poissonprocessen, men ingår i sig bara för TM.

Den allmänna gammfördelningen (bara TM): för $\alpha, \lambda > 0$ säger man att $X \sim \Gamma(\alpha, \lambda)$ om X har täthet

$$f(x) = Cx^{\alpha-1}e^{-\lambda x}.$$

Normalfördelning: $X \sim N(\mu, \sigma^2)$ om

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \; x \in \mathbb{R}.$$

$$\mathbb{E}[X] = \mu, \; \mathbb{V}\mathrm{ar}(X) = \sigma^2.$$

Om $Z \sim N(0,1)$ sägs Z vara standardnormalfördelad. Det gäller att $\sigma Z + \mu \sim N(\mu, \sigma^2)$. Omvänt: om $X \sim N(\mu, \sigma^2)$ gäller $\frac{X-\mu}{\sigma} \sim N(0,1)$.

Fördelningsfunktionen för Z betecknas Φ, dvs $\Phi(x) = \mathbb{P}(Z \leq x)$. Det följer att om $X \sim N(\mu, \sigma^2)$ gäller

$$\mathbb{P}(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Av symmetri: $\Phi(-x) = 1 - \Phi(x)$. Några percentiler $\Phi(1.96) \approx 0.975$, $\Phi(2.58) \approx 0.995$.

Summan av två normalfördelade sv X_1 och X_2 är normalfördelad. Om X_1 och X_2 också är oberoende är $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Endast TM: Om $Z_1, Z_2, \ldots$ är oberoende och standardnormalfördelade säger man att

$$\sum_k^n Z_k^2 \sim \chi_n^2.$$

$$X \sim \chi_n^2 \Rightarrow \mathbb{E}[X] = n,\ \mathbb{V}\mathrm{ar}(X) = 2n.$$

Vi visade att $Z_1^2 + Z_2^2 \sim exp(1/2)$. Det betyder att $\chi_n^2 = \Gamma(n/2, 1/2)$ då n är jämnt. Gäller även udda n.

Bivariata fördelningar: Fördelningsfunktionen för (X, Y)

$$F(x, y) = \mathbb{P}(X \leq x, Y \leq y).$$

(X, Y) är diskret om $V_X \times V_Y$ är ändlig eller uppräknelig.

$$p(x_j, y_k) = \mathbb{P}(X = x_j, Y = y_k).$$

Marginalfördelningar

$$p_X(x_j) = \sum_{k=1}^{\infty} p(x_j, y_k),\ p_Y(y_k) = \sum_{j=1}^{\infty} p(x_j, y_k).$$

(X, Y) är kontinuerlig om det finns en bivariat täthetsfunktion $f : \mathbb{R}^2 \to [0, \infty)$ sådan att

$$\mathbb{P}((X, Y) \in B) = \iint_B f(x, y) dx\, dy$$

för alla $B \subseteq \mathbb{R}^2$.

Om (X, Y) är kontinuerlig gäller att $f(x, y) = F''_{xy}(x, y)$ är en täthet.

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy,\ f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx.$$

Minns att X och Y är oberoende om $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$ för alla $A, B \subseteq \mathbb{R}$. Det gäller att

- (X, Y) diskret: X och Y är oberoende om och endast om $p(x_j, y_k) = p_X(x_j)p_Y(y_k)$ för alla j, k.
- (X, Y) kont: X och Y oberoende om och endast om $f(x, y) = f_X(x)f_Y(y)$ för alla x, y.

OSL i två dimensioner:

- (X, Y) diskret: $\mathbb{E}[g(X, Y)] = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} g(x_j, y_k)p(x_j, y_k)$.
- (X, Y) kont: $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f(x, y)dx\,dy$.

Exempel på viktiga tillämpningar:

- För alla sv X och Y gäller $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.
- Om X och Y är oberoende gäller $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Från den andra av dessa följer att om X och Y är oberoende är $\mathbb{V}\mathrm{ar}(X + Y) = \mathbb{V}\mathrm{ar}(X) + \mathbb{V}\mathrm{ar}(Y)$.

Ett specialfall av dessa räkneregler är att om $X_1, X_2, \ldots$ är oberoende och alla har vv μ och varians σ^2 gäller att

$$\mathbb{E}[\overline{X}_n] = \mu, \ \mathbb{V}\mathrm{ar}(\overline{X}_n) = \frac{\sigma^2}{n}.$$

Kovariansen av två sv:

$$\mathbb{C}\mathrm{ov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Kovariansen är bilinjär och en generalisering av varians eftersom $\mathbb{V}\mathrm{ar}(X) = \mathbb{C}\mathrm{ov}(X, X)$.

$$\mathbb{V}\mathrm{ar}(X + Y) = \mathbb{V}\mathrm{ar}(X) + \mathbb{V}\mathrm{ar}(Y) + 2\mathbb{C}\mathrm{ov}(X, Y).$$

Korrelationskoefficienten:

$$\rho_{XY} = \frac{\mathbb{C}\mathrm{ov}(X, Y)}{\sqrt{\mathbb{V}\mathrm{ar}(X)\mathbb{V}\mathrm{ar}(Y)}}.$$

Enligt Schwarz gäller att $-1 \leq \rho_{XY} \leq 1$.

Betingade fördelningar: För (X, Y) diskret, $p_{Y|X}(y_k|x_j) = \mathbb{P}(Y = y_k|X = x_j)$. För (X, Y) kont *definierar* man

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}.$$

Några följder är varianter av totala sannolikhetslagen:

- Diskret: $p_Y(y_k) = \sum_{j=1}^{\infty} p_{Y|X}(y_k|x_j)p_X(x_j)$.
- Kont: $f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)f_X(x)dx$.

och Bayes formel för tätheter:

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{\int_{-\infty}^{\infty} f_{Y|X}(y|t)f_X(t)dt}.$$

Betingade väntevärden:

- Diskret: $\mathbb{E}[Y|X = x_j] = \sum_{k=1}^{\infty} y_k \mathbb{P}(Y = y_k|X = x_j)$.
- Kont: $\mathbb{E}[Y|X = x] = \int_{-\infty}^{\infty} y\, f_{Y|X}(y|x)dy$.

Totala sannolikhetslagen för väntevärden följer

- Diskret: $\mathbb{E}[Y] = \sum_{j=1}^{\infty} \mathbb{E}[Y|X = x_j]\mathbb{P}(X = x_j)$.
- Kont: $\mathbb{E}[Y] = \int_{-\infty}^{\infty} \mathbb{E}[Y|X = x]f_X(x)dx$.

$\mathbb{E}[Y|X]$ är den sv som antar värdet $\mathbb{E}[Y|X = x]$ då $X = x$. Av OSL och TSL föjer direkt att

$$\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y].$$

Markovs olikhet: Om X är en ickenegativ sv gäller

$$\mathbb{P}(X > a) \leq \frac{\mathbb{E}[X]}{a}.$$

Chebyshevs olikhet: Om X har väntevärde μ och varians σ^2 gäller att

$$\mathbb{P}(|X - \mu| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2}.$$

Stora talens lag: Om $X_1, X_2, \ldots$ alla har samma väntevärde μ och samma varians $\sigma^2 < \infty$, gäller för alla $\epsilon > 0$ att

$$\lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mu| > \epsilon) = 0.$$

CGS: Låt $X_1, X_2, \ldots$ vara oberoende och likafördelade sv med vv μ och varians $\sigma^2 < \infty$. Skriv $S_n = \sum_{k=1}^{n} X_k$. Då gäller för alla $x \in \mathbb{R}$ att

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x).$$

Approximationen är oftast god för ickeextrema sannolikheter, men dålig i svansarna. Ju större n, desto mer extrema sannolikheter klarar man att approximera. Hur stort n behöver vara för god approximation beror också på hur fördelningen för X_k ser ut.

Punktskattningar: Om $\mathcal{X} = (X_1, \ldots, X_n)$ är ett stickprov på en sv vars fördelning beror av en okänd parameter θ (kan vara flerdimensionell) och $\hat{\theta}(\mathcal{X})$ är en funktion som används för att skatta θ, kallas $\hat{\theta}$ för en punktskattning av θ.

Om $\mathbb{E}[\hat{\theta}] = \theta$ oavsett vad det sanna värdet på θ är, kallas $\hat{\theta}$ för en väntevärdesriktig punktskattning.

Om $\mathbb{P}(|\hat{\theta} - \theta| > \epsilon) \to 0$ då $n \to \infty$ för alla $\epsilon > 0$, kallas $\hat{\theta}$ konsistent.

Proposition: Om $\hat{\theta}$ är vvr och $\mathbb{V}\mathrm{ar}(\hat{\theta}) \to 0$ då $n \to \infty$, är $\hat{\theta}$ konsistent.

Följd: Om $\mathcal{X}$ är ett stickprov på en sv X med vv μ och varians $\sigma^2 < \infty$ är $\overline{X}$ en vvr och konsistent skattning av μ.
Stickprovsvariansen

$$s^2 = \frac{1}{n-1} \sum_{k=1}^{n} (X_k - \overline{X})^2$$

är en vvr skattning av σ^2 och om $\mathbb{E}[X^4] < \infty$ är den också konsistent.

ML-principen: Punktskatta θ genom att välja det θ som maximerar sannolikheten/tätheten för att få de data vi faktiskt fick. M.a.o., maximera

$$L(\theta; \mathbf{x}) = \prod_{k=1}^{n} p_\theta(x_k)$$

sedd som funktion av θ. För kont förd $L(\theta; \mathbf{x}) = \prod_{k=1}^{n} f_\theta(x_k)$. Flera exempel. Ekvivalent att maximera $\ell(\theta; \mathbf{x}) = \ln L(\theta; \mathbf{x})$.

Konfidensintervall (endimensionella θ): Välj två funktioner $T_1(\mathcal{X})$ och $T_2(\mathcal{X})$ av data sådana att

$$\mathbb{P}(T_1(\mathcal{X}) \leq \theta \leq T_2(\mathcal{X})) = 1 - \alpha.$$

Då kallas $[T_1(\mathbf{x}), T_2(\mathbf{x})]$ för ett konfidensintervall av konfidensgrad $1 - \alpha$ (inte sannolikhet när data $\mathcal{X} = \mathbf{x}$ väl har observerats). Symmetriskt och uppåt/nedåt begränsat.

Vi gjorde konfidensintervall för θ i $likf(0,\theta)$, för μ i normalfördelning med känd eller okänd varians, för p i $Bin(n,p)$, och TM gjorde även för σ^2 i normalfördelning. Man behövde veta att $(\overline{X}-\mu)/(\sigma/\sqrt{n}) \sim N(0,1)$, alternativt att

$$\frac{\overline{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}.$$

För konfintervall för σ^2 behöver man veta att $(n-1)s^2/\sigma^2 \sim \chi^2_{n-1}$. För p behöver man veta att

$$\frac{\overline{X}-p}{\sqrt{\overline{X}(1-\overline{X})/n}} \approx N(0,1).$$

Exempelvis symmetriskt konfintervall för μ med okänd varians

$$\mu \in \overline{X} \pm F^{-1}_{t_{n-1}}(1-\alpha/2)\frac{s}{\sqrt{n}},\ (1-\alpha).$$

Vi gjorde också prediktionsintervall för en ny observation i normalfördelning.
Två stickprov: $X_1, \ldots, X_n \sim N(\mu_X, \sigma_X^2)$, $Y_1, \ldots, Y_m \sim N(\mu_Y, \sigma_Y^2)$.

Konfidensintervall för $\mu_X - \mu_Y$ med kända varianser och med okända varianser under antagandet att $\sigma_X = \sigma_Y$. I det senare fallet utnyttjas att

$$\frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{s_P \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{m+n-2}.$$

Den poolade stickprovsvariansen ges av

$$s_P^2 = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{m+n-2}.$$

Med kända varianser gäller

$$\frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim N(0,1).$$

Statistiska tester: Vi vill testa den neutrala hypotesen H_0 mot den alternativa hypotesen (som vi vill visa) H_A. Principen är att fixera en signifikansnivå α, finna lämplig mängd A sådan att $P_{H_0}(\mathcal{X} \in A) = 1 - \alpha$ och sådan att $\mathcal{X} \notin A$ talar för H_A. Om det sedan visar sig att $\mathbf{x} \notin A$ förkastar man H_0 till förmån för H_A på signnivå α, annars accepterar man H_0.

Tester av μ i n-förd och av p i binförd. TM kan även göra tester av σ^2 i n-förd. Konfidensintervall ger test: Förkasta $H_0 : \theta = \theta_0$ på signnivå α om $\theta_0 \notin [T_1, T_2] =$ konfintervall av grad $1 - \alpha$. Alltså, för endimensionella parametrar extremt likt konfintervall.

Du får aldrig använda testdata innan du bestämmer vad du ska göra för test. Det är ok att använda data för att bestämma vilka tester som ska göras, men i så fall måste ny data samlas in för att göra själva testet. Undvik att göra tester där du inte har goda skäl att tro på alternativhypotesen.

Linjär regression: Kovariater/ställvariabler/features $x_1, \ldots, x_n$, med respektive svarsvariabler $Y_1, \ldots, Y_n$. Modell

$$Y_k = a + bx_k + \epsilon_k$$

där a och b är de okända koeffiecienter vi vill skatta och $\epsilon_1, \ldots, \epsilon_n$ oberoende och $N(0, \sigma^2)$ med okänd varians σ^2.

ML-skattning sammanfaller med minsta-kvadrat-metoden och skattningarna blir

- $\hat{b} = \frac{S_{xy}}{S_{xx}}$,
- $\hat{a} = \overline{y} - \hat{b}\overline{x}$.

Konfintervall mest intressant för b och använder

$$\frac{\hat{b} - b}{s/\sqrt{S_{xx}}} \sim t_{n-2}.$$

Här är

$$s^2 = \frac{1}{n-2}\sum_{k=1}^{n}(y_k - \hat{a} - \hat{b}x_k)^2 = \frac{1}{n-2}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right).$$

Ett antal beräkningsformler som gör det lättare att räkna.

Prediktionsintervall för ny observation Y vid ställvariabel x baseras på

$$\frac{Y - \overline{Y}}{s\sqrt{1 + \frac{1}{n} + \frac{(x-\overline{x})^2}{S_{xx}}}} \sim t_{n-2}.$$

Resten endast för TM.

Tranformering mellan olika fördelningar: Om F_1 och F_2 är två (strikt växande) fördelningsfunktioner och $X \sim F_1$, gäller att $Y = F_2^{-1}(F_1(X)) \sim F_2$.

Fördelning av summor av oberoende sv:

- Diskret: $p_{X+Y}(z) = \sum_{x \in V_X} p_Y(z-x)p_X(x)$,
- Kont: $f_{X+Y}(z) = \int_{-\infty}^{\infty} f_Y(z-x)f_X(x)dx$.

Felintensitet för positiva sv:

$$r(t) = \frac{f(t)}{G(t)}.$$

Invertering

$$G(t) = e^{-\int_0^t r(s)ds}.$$

Momentgenererande funktion för X:

$$M_X(t) = \mathbb{E}[e^{tX}].$$

Om X och Y oberoende är

$$M_{X+Y}(t) = M_X(t)M_Y(t).$$

Mgf bestämmer fördelningen. Mer generellt om $M_{X_n}(t) \to M_X(t)$ för alla t gäller $F_{X_n}(x) \to F_X(x)$ för alla x.

Känn igen mgf för några fördelningar och bevisskissa CGS.

p-värde: p-värdet för ett test är den mista signifikansnivå som man hade kunnat förkasta nollhypotesen på (mer informativt än bara informationen acceptera/förkasta).

Styrka: Om man gör ett test $H_0 : \theta = \theta_0$ mot någon alternativhypotes, är styrkan

$$g(\theta_1) = \mathbb{P}_{\theta_1}(H_0 \text{ förkastas}).$$

Bayesiansk statistik: För en modell där data beror av en parameter θ (ofta högdimensionell), antar vi att θ har blivit slumpmässigt vald enligt en prior $f_\theta(t)$ och sedan är data vald enligt $f_{X|\theta}(x|t)$. Vi beräknar posterior

$$f_{\theta|X}(t|x) = Cf_{X|\theta}(x|t)f_\theta(t).$$

Betafördelning: $Y \sim \beta(b_1, b_2)$ om $f(y) = Cy^{b_1-1}(1-y)^{b_2-1}$, $0 < y < 1$.

Betafördelningen är konjugerande prior till binomialfördelningen: Om $\theta \sim beta(b_1, b_2)$ och sedan $X_1 \sim Bin(n, \theta)$ är posterior $\theta|X \sim \beta(b_1 + X_1, b_2 + X_2)$. (Här är $X_2 = n - X_1$.)

Mer generellt: $(Y_1, \ldots, Y_{K-1}) \sim Dir(b_1, \ldots, b_K)$ om

$$f(y_1, \ldots, y_{K-1}) = C \prod_{i=1}^{K} y_i^{b_i}$$

för positiva $y_1, \ldots, y_K$ som summerar till 1 och där $y_K = 1 - \sum_{i=1}^{K-1} y_i$.

Dirichlet är konjugerande till multinomial: Om $\theta \sim Dir(b_1, \ldots, b_K)$ och sedan $(X_1, \ldots, X_K) \sim Mult(\theta_1, \ldots, \theta_K)$ är $\theta|X \sim Dir(b_1 + X_1, \ldots, b_K + X_K)$. (Här är $X_K = 1 - \sum_{i=1}^{K-1} X_i$.)

Finns många andra konjugerande priors. Till exempel är normalfördelningen konjugerande prior som fördelning av väntevärdet i normalfördelningen med känd varians.