

Del 1 - Sannolikhetsteori

Sannolikhet och händelser

- Additionssatsen: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Betingad sannolikhet: $P(A | B) = \frac{P(A \cap B)}{P(B)}$
- A och B är oberoende $\iff P(A \cap B) = P(A)P(B)$.
- Bayes sats: $P(A | B) = \frac{P(B | A)P(A)}{P(B)}$
- Satsen om total sannolikhet: $P(A) = \sum_{i=1}^n P(A | H_i) P(H_i)$
om $H_i \cap H_j = \emptyset$ då $i \neq j$ och $\bigcup_{i=1}^n H_i = \Omega$.

Fördelningsfunktionen $F(x)$

- För en diskret stokastisk variabel ξ är
 $F(x_k) = P(\xi \leq x_k) = \sum_{i=1}^k P(\xi = x_i)$, då
 $x_1 < x_2 < \dots < x_k$, för sannolikhetsfunktion $P(\xi = x_i)$.
- För en kontinuerlig stokastisk variabel ξ är
 $F(x) = P(\xi \leq x) = \int_{-\infty}^x f(t)dt$, för frekvensfunktion
 $f(t) = F'(t)$.

Väntevärde, Varians och Standardavvikelse:

- Väntevärde: För en stokastisk variabel ξ och en funktion $g: \mathbb{R} \mapsto \mathbb{R}$ är:

$$E(g(\xi)) = \begin{cases} \sum_{i=1}^n g(x_i)P(\xi = x_i), & \xi \text{ diskret} \\ \int_{-\infty}^{\infty} g(t)f(t)dt, & \xi \text{ kontinuerlig} \end{cases}$$

- Varians: $V(\xi) = E((\xi - E(\xi))^2) = E(\xi^2) - E(\xi)^2$.
- Standardavvikelse: $D(\xi) = \sqrt{V(\xi)}$
- Linjärkombination av stokastiska variabler:

$$E\left(\sum_{i=1}^n a_i \xi_i + b\right) = \sum_{i=1}^n a_i E(\xi_i) + b$$

$$V\left(\sum_{i=1}^n a_i \xi_i + b\right) = \sum_{i=1}^n a_i^2 V(\xi_i), \quad \xi_1, \dots, \xi_n \text{ ober.}$$

- För medelvärdet $\bar{\xi}$, dvs. fallet $a_i = 1/n$, $i = 1, \dots, n$ och $b = 0$ från föreg. punkt, är

$$E(\bar{\xi}) = E(\xi), \quad V(\bar{\xi}) = \frac{V(\xi)}{n}, \quad \xi_1, \dots, \xi_n \text{ ober.}$$

Linjära funktioner av stokastiska variabler

- Om ξ_1 och ξ_2 är oberoende s.v. så gäller:

$$\begin{cases} \xi_1 \in \text{Bin}(n_1, p) \\ \xi_2 \in \text{Bin}(n_2, p) \end{cases} \Rightarrow \xi_1 + \xi_2 \in \text{Bin}(n_1 + n_2, p).$$

$$\begin{cases} \xi_1 \in \text{Po}(\mu_1) \\ \xi_2 \in \text{Po}(\mu_2) \end{cases} \Rightarrow \xi_1 + \xi_2 \in \text{Po}(\mu_1 + \mu_2).$$

- Om $\xi_1 \in N(\mu_1, \sigma_1)$ och $\xi_2 \in N(\mu_2, \sigma_2)$ är oberoende och a och b konstanter är:

$$a\xi_1 + b\xi_2 \in N\left(a\mu_1 + b\mu_2, \sqrt{a^2\sigma_1^2 + b^2\sigma_2^2}\right)$$

- För en normalfördelad s.v. $\xi \in N(\mu, \sigma)$ kan fördelningsfunktionen beräknas numeriskt med

$$F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right) = 1 - \Phi\left(\frac{-(x - \mu)}{\sigma}\right),$$

där Φ är fördelningsfunktionen för den standardiserade normalfördelningen, $\frac{\xi - \mu}{\sigma} \in N(0, 1)$, för vilken numeriska värden finns tabulerade.

Några icke-linjära funktioner av s.v.

Låt ξ_1, \dots, ξ_n vara oberoende och likafördelade s.v. med fördelningsfunktion $F(x)$. Då gäller för:

- Minsta värde, $\eta = \min(\xi_1, \dots, \xi_n)$, har fördelning
 $F_\eta(x) = 1 - (1 - F(x))^n$
- Största värde, $\omega = \max(\xi_1, \dots, \xi_n)$, har fördelning
 $F_\omega(x) = F(x)^n$

Centrala gränsvärdessatsen

Om $\xi_1, \xi_2, \dots, \xi_n$ är ett oberoende och likafördelat stickprov av ξ : $E(\xi) = \mu$, $V(\xi) = \sigma^2$ så är för stora n :

$$\eta = \sum_{i=1}^n \xi_i \in N(n\mu, \sqrt{n}\sigma) \Rightarrow \bar{\xi} = \frac{1}{n}\eta \in N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Tumregler för approximationer:

Binomial \rightarrow Normal om $np(1-p) \geq 10$.
Poisson \rightarrow Normal om $\mu \geq 15$.

Gauss approximationsformler:

För en funktion av stokastisk variabel, $g(\xi)$, där $\mu = E(\xi)$ och $\sigma^2 = V(\xi)$, kan funktionens väntevärde och varians approximeras med:

$$E(g(\xi)) \approx g(\mu), \quad V(g(\xi)) \approx [g'(\mu)]^2 \cdot \sigma^2.$$

Del 2 - Statistikteori

Punktskattningar

Ett stickprov

Låt x_1, \dots, x_n vara observationer av ett oberoende och likafördelat stickprov ξ_1, \dots, ξ_n från den s.v. ξ med okänt väntevärde μ och standardavvikelse σ . Väntevärdes- riktiga skattningar av μ och σ^2 är då

$$\mu_{obs}^* = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$
$$(\sigma^2)_{obs}^* = s^2 = \frac{Q}{n-1} = \frac{1}{n-1} \left(\sum_{i=1}^n (x_i^2) - n\bar{x}^2 \right)$$

Korrelationskoefficient

Låt x_1, \dots, x_n och y_1, \dots, y_n vara två uppsättningar observationer. Korrelationskoefficienten mellan dessa beräknas som:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Konfidensintervall under normalfördelning

Intervall för parameter θ

När parametern θ skattas med $\theta^* \in N(\theta, D(\theta^*))$ så ges ett konfidensintervall med konfidensgrad $1 - \alpha$ av:

$$I_\theta = \theta^* \pm \lambda_{\alpha/2} \cdot D(\theta^*), \quad \text{om } D(\theta^*) \text{ är känd}$$
$$I_\theta = \theta^* \pm \lambda_{\alpha/2} \cdot d(\theta^*), \quad \text{för skattad } d(\theta^*) \text{ då CGS använts.}$$
$$I_\theta = \theta^* \pm t_{\alpha/2}(f) \cdot d(\theta^*), \quad \text{för skattad } d(\theta^*) = c \cdot \sigma^*, \text{ där } \sigma_{obs}^* = s = \sqrt{Q/f}$$

Avrundning görs så att konfidensgraden är minst $1 - \alpha$.

Intervall för väntevärde

För väntevärdet μ , om det skattas med stickprovsmedel $\mu^* = \bar{x}$, blir det observerade konfidensintervallet:

$$(I_\mu)_{obs} = \bar{x} \pm \lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \quad \text{om } \sigma \text{ är känd}$$
$$(I_\mu)_{obs} = \bar{x} \pm t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}, \quad \text{då } \sigma \text{ skattas enl. ovan}$$

Jämförelse av väntevärde under normalfördelning

Stickprov i par

När observationer (x_i, y_i) kan grupperas parvis kan skillnaden i väntevärde studeras genom följande modell:

$$\left. \begin{array}{l} \xi_i \in N(\mu_i + \Delta, \sigma_i) \\ \eta_i \in N(\mu_i, \sigma_i) \end{array} \right\} \Rightarrow \zeta_i = \xi_i - \eta_i \in N(\Delta, \sigma),$$

med observationer $z_i = x_i - y_i$ för $i = 1, \dots, n$. Ett konfidensintervall för Δ kan konstrueras som ovan.

Två oberoende stickprov

För två oberoende stickprov; x_1, \dots, x_{n_1} från $\xi \in N(\mu_1, \sigma_1)$ och y_1, \dots, y_{n_2} från $\eta \in N(\mu_2, \sigma_2)$ blir förväntad skillnad:

$$\mu_1^* - \mu_2^* = \bar{x} - \bar{y} \in N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$
$$(I_{\mu_1 - \mu_2})_{obs} = \bar{x} - \bar{y} \pm \lambda_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \quad \sigma_1, \sigma_2 \text{ kända}$$
$$(I_{\mu_1 - \mu_2})_{obs} = \bar{x} - \bar{y} \pm t_{\alpha/2}(n_1 + n_2 - 2) \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},$$

om $\sigma_1 = \sigma_2 = \sigma$ är okänd och skattas med den poolade standardavvikelsen, $s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}}$, där s_1^2 och s_2^2 är punktskattningar för varians av respektive stickprov enligt ovan.

Hypotestest

Direktmetoden: Jämför p-värdet med signifikansnivån α :
 $p = P(\text{det vi fick eller längre från } H_0 \mid H_0 \text{ sann}).$

Konfidensintervall: Gör en lämplig intervallskattning och jmf. med θ_0 (värdet under H_0).

Teststorhet: Om skattningen θ^* är (approx.) normalfördelad, beräkna observationen av $T = \frac{\theta^* - \theta_0}{d(\theta^*)}$, jmf. med λ - eller $t(f)$ -kvantil.

Styrkefunktion: $h(x) = P(\text{Förkasta } H_0 \mid \theta = x \text{ är rätt värde})$

Signifikansnivå: $\alpha = P(\text{Förkasta } H_0 \mid H_0 \text{ är sann})$

Enkel linjär regression

• Modell: $\eta_i = \alpha + \beta x_i + \varepsilon_i$, $\varepsilon_i \in N(0, \sigma)$.

• Punktskattningar:

$$\alpha_{obs}^* = \bar{y} - \beta^* \bar{x}, \quad \beta_{obs}^* = \frac{S_{xy}}{S_{xx}}, \quad (R^2)_{obs}^* = \frac{S_{xy}^2}{S_{xx} S_{yy}}$$
$$(\sigma^2)_{obs}^* = s^2 = \frac{Q_0}{n-2}, \quad Q_0 = S_{yy} - \frac{S_{xy}^2}{S_{xx}}$$
$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2, \quad S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2,$$
$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}.$$
$$\alpha^* \in N\left(\alpha, \sigma \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}\right), \quad \beta^* \in N\left(\beta, \frac{\sigma}{\sqrt{S_{xx}}}\right).$$
$$\mu_0^* = \alpha^* + \beta^* x_0 \in N\left(\alpha + \beta x_0, \sigma \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}\right).$$
$$\eta_0^* = \alpha^* + \beta^* x_0 + \varepsilon_0 \in N\left(\alpha + \beta x_0, \sigma \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}\right).$$

Miniräknare

Funktioner för fördelningar finns med ändelserna **cdf/CD** *cumulative distribution function* (fördelningsfunktion, $F(x) = P(\xi \leq x)$), **pdf/PD** *probability density function* (sannolikhets-, $P(\xi = k) = P(\xi = k)$, eller täthetsfunktion, $f(x)$), eller **inv** *inverse cdf* (dvs $x = F^{-1}(p)$). Notera att kvantil och inverse funktionen skiljer sig ($\xi > x_\alpha$ eller $\xi \leq x_p$):

kvantil: Finn x_α så att $P(\xi > x_\alpha) = \alpha$.

inverse: $F^{-1}(p) = x_p$ eller finn x_p så att $P(\xi \leq x_p) = p$.

Fördelningsfunktionerna beräknar ibland sannolikheter i ett intervall $a < \xi \leq b$, använd $-1 \cdot 10^{99}$ eller $1 \cdot 10^{99}$ för värdena $\pm\infty$ (en mindre exponent kan eventuellt behövas men testa att den är "tillräckligt stor", dvs inte ändrar resultatet).

För att beräkna statistik och skattningar (μ_{obs}^* , s^2 , $\sum x_i^2$, etc) behöver värdena först sparas i listor.

	Texas instrument	Casio
$1 \cdot 10^x$	2nd → EE Ger $1Ex$ på räknaren.	SHIFT → log (10^x) Ger $1 * 10^x$ på räknaren.
$k!$ & $\binom{n}{k}$	MATH → PROB → ! MATH → PROB → nCr	OPTN → PROB → ! OPTN → PROB → nCr
Fördelningar	DISTR (2nd VARS) → välj lämplig fördelning normalpdf(x, μ, σ) eller normalcdf(a, b, μ, σ) binompdf(n, p, x) eller binomcdf(n, p, x) poissonpdf(λ, x) eller poissoncdf(λ, x) normalinv(1-α, μ, σ) för normal-kvantiler, λ_α tinvs(1-α, f) för t-kvantiler, $t_\alpha(f)$	OPTN → STAT → DISTR → välj lämplig fördelning Normal PD(x, σ, μ) eller Normal CD(a, b, σ, μ) Binomial PD(x, n, p) eller Binomial CD(x, n, p) Poisson PD(x, λ) eller Poisson CD(x, λ) Inverse Normal för normal-kvantiler, alternativet Tail anger vilket håll som inversen ska beräknas (left=inverse eller right=kvantil). Inverse Student-t för t-kvantiler.
Listor	STAT → EDIT För att lägga in värden listorna 2nd → 1 - 6 Hämta namn på listor för användning. Det går att räkna med och tilldela listor: Exempel: 2*L1-1 (STO►) L2	MENU → STAT(2) För att lägga in värden listorna SHIFT → 1(List) → 1 - 6 Ange namn på listor för användning. Det går att räkna med och tilldela listor: Exempel: 2*List 1-1 → List 2
Statistik	STAT → CALC → 1-Var Stats Beräkna sammanfattande statistik för angiven lista. Exempel: 1-Var Stats L1 STAT → CALC → 2-Var Stats Beräkna sammanfattande statistik för två angivna listor (beräknar t.ex. $\sum x_i y_i$ som behövs för regression och korrelation). Exempel: 2-Var Stats L1,L2 VARS → Statistics → Hämta variabler som innehåller statistik som beräknats ovan.	OPTN → CALC → 1VAR Beräkna sammanfattande statistik för angiven lista. OPTN → CALC → 2VAR Beräkna sammanfattande statistik för två angivna listor (beräknar t.ex. $\sum x_i y_i$ som behövs för regression och korrelation). VARS → STAT → Hämta variabler som innehåller statistik som beräknats ovan.

Användning av miniräknare på tentan: När miniräknaren används för beräkningar på tentan måste det tydligt framgå vad som beräknats (formler), vad resultatet är och vilka slutsatser som kan dras. Att bara repetera resultat från miniräknaren utan kommentarer utgör **INTE** väl motiverade lösningar.

Fördelningar

Fördelning			Väntevärde	Varians
Binomialfördelning, $Bin(n, p)$	$P(\xi = k) = \binom{n}{k} p^k (1-p)^{n-k}$	$k = 0, 1, \dots, n$	np	$np(1-p)$
Hypergeometrisk fördelning, $Hyp(N, n, p)$	$P(\xi = k) = \frac{\binom{Np}{k} \binom{N(1-p)}{n-k}}{\binom{N}{n}}$	$0 \leq k \leq Np,$ $0 \leq n-k \leq N(1-p)$	np	$\frac{N-n}{N-1} np(1-p)$
Poissonfördelning, $Po(\mu)$	$P(\xi = k) = \frac{\mu^k}{k!} e^{-\mu}$	$k = 0, 1, 2, \dots$	μ	μ
ffg-fördelning	$P(\xi = k) = p(1-p)^{k-1}$	$k = 1, 2, 3, \dots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Rektangelfördelning, $R(a, b), U(a, b)$	$f(x) = \frac{1}{b-a}$	$a \leq x \leq b$	$\frac{a+b}{2}$	$\frac{(a-b)^2}{12}$
Exponential- fördelning, $Exp(\lambda)$	$f(x) = \lambda e^{-\lambda x}$	$x \geq 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normalfördelning, $N(\mu, \sigma)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$x \in \mathbb{R}$	μ	σ^2
t -fördelning ¹ , $t(n)$	$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$	$x \in \mathbb{R}$	$0, n > 1$	$\frac{n}{n-2}, n > 2$
Lognormal- fördelning $\ln X \in N(m, \sigma)$	$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - m)^2}{2\sigma^2}}$	$x \geq 0$	$e^{m+\sigma^2/2}$	$e^{2m+\sigma^2} (e^{\sigma^2} - 1)$

1. Gamma-funktionen, $\Gamma(p)$, är en generaliserad faktulet:

$$\Gamma(p) = (p-1) \cdot \Gamma(p-1)$$

$$\Gamma(p) = (p-1)! \text{ om } p \text{ heltal}$$

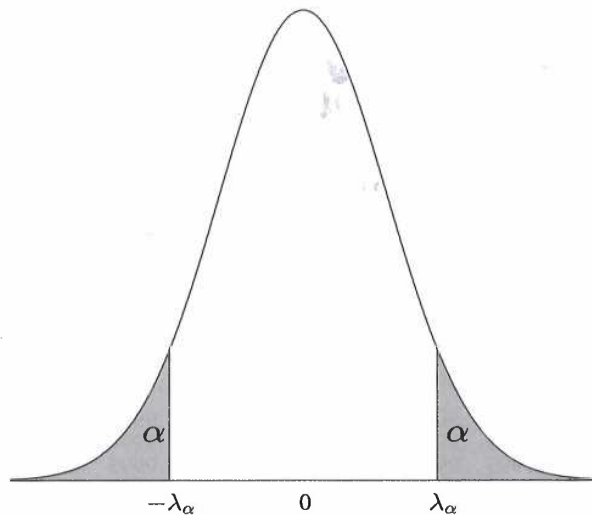
$$\Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x} dx, p > 0$$

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

A.5 Normalfördelningen (forts.)

Tabellen ger det λ_α -värde för vilket $P(\xi > \lambda_\alpha) = \alpha$, där $\xi \in N(0,1)$.

α	0.1	0.05	0.025	0.01	0.005	0.001
λ_α	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902
α	0.0005	0.0001	0.00005	0.00001	5.0×10^{-6}	1.0×10^{-6}
λ_α	3.2905	3.7190	3.8906	4.2649	4.4172	4.7534

A.6 t -fördelningen

Tabellen ger det x -värde för vilket $P(\xi > x) = \alpha$, där $\xi \in t(f)$.

f	α						
	.1	.05	.025	.01	.005	.001	.0005
1	3.078	6.314	12.706	31.821	63.657	318.309	636.619
2	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291