Formelsamling (*Statistik för Lärare*)

Diskreta fördelningar

Fördelning	Sannolikhetsfunktion	Väntevärde	Varians	Användning
Bernoulli, (p)	$p_X(k) = \begin{cases} p & , k=1, \\ 1-p & , k=0 \end{cases}$	p	$p(1-p)$	Ett försök som antigen lyckas eller misslyckas
Binomial, (n,p)	$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$, $0 \le k \le n$	np	$np(1-p)$	Antalet lyckade, då n stycken likadana (oberoende) försök utförs.
Hyp.geom, (N,n,p)	$p_X(k) = \frac{\binom{Np}{k}\binom{N(1-p)}{n-k}}{\binom{N}{n}}$, $0 \le k \le Np$, $0 \le n-k \le N(1-p)$	np	$np(1-p)\frac{N-n}{N-1}$	Vi drar n stycken utan återläggning från en ändlig population av storlek N, där Np stycken har en viss egenskap och räknar antalet med denna egenskap som finns bland de dragna.
Likformig, (N)	$p_X(k) = \frac{1}{N}$, $1 \le k \le N$	$\frac{N+1}{2}$	$\frac{(N+1)(N-1)}{12}$	
ffg, (p)	$p_X(k) = p(1-p)^{k-1}$, $k = 1, 2, 3, \ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	Antalet försök t.o.m. det första lyckade, alla försök är lika och har samma slh, p, att lyckas.
Poisson, (λ)	$p_X(k) = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}$, $k = 0, 1, 2, \ldots$	λ	λ	Används när något sker med en viss frekvens under ett givet intervall. T. ex. tryckfel per sida.

Kontinuerliga fördelningar

Fördelning	Täthetsfunktion	Fördelningsfunktion	Väntevärde	Varians	Användning
Likformig, (a,b)	$f_X(x) = \frac{1}{b-a}$, $a \le x \le b$	$\frac{x-a}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	
Exponential, (β)	$f_X(x) = \frac{1}{\beta}\mathrm{e}^{-\frac{x}{\beta}}$, $x \ge 0$	$1 - \mathrm{e}^{-\frac{x}{\beta}}$	β	β^2	Livslängder
Normal, (μ,σ)	$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $-\infty < x < \infty$		μ	σ^2	Läng,Vikt

För $X \sim \mathrm{N}(\mu, \sigma)$ gäller

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

där $\Phi(\cdot)$ är fördelningsfunktionen för standard normal (d.v.s. $\mu = 0$ och $\sigma = 1$).
$\Phi(-x) = 1 - \Phi(x)$.

Approximationer

$$\begin{array}{rcl} \text{Bin }(n,p) & \approx & \begin{cases} \mathrm{N}(np, \sqrt{np(1-p)}) & , \ np(1-p) > 10, \\ \mathrm{Poi}(np) & , \ p < 0.1, \ \ np(1-p) \approx np. \end{cases} \\ \text{Hyp }(N,n,p) & \approx & \left\{ \ \mathrm{N}(np, \sqrt{np(1-p)\frac{N-n}{N-1}}) \quad , \ np(1-p)\frac{N-n}{N-1} > 10. \right. \end{array}$$

Centrala Gränsvärdessatsen

Låt $X_1, X_2, X_3, \ldots$ vara en följd av oberoende stokastiska variabler med samma sannolikhetsfördelning s.a. $\mathbb{E}[X_i] = \mu$ och $\mathrm{V}(X) = \sigma^2$ för alla i. Då gäller

$$\mathbb{P}\left\{\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \leq x\right\} \to \Phi(x), \ \ \text{då } n \to \infty.$$

Punktskattning

$$\begin{aligned} \overline{X} &= \frac{1}{n}\sum_{i=1}^n X_i \\ S^2 &= \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2 = \frac{1}{n-1}\left(\sum_{i=1}^n X_i^2 - n\overline{X}^2\right) \\ R &= \frac{\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^n (X_i - \overline{X})^2 \sum_{i=1}^n (Y_i - \overline{Y})^2}} = \frac{\sum_{i=1}^n X_i Y_i - n\overline{X}\overline{Y}}{\sqrt{(\sum_{i=1}^n X_i^2 - n\overline{X})(\sum_{i=1}^n Y_i^2 - n\overline{Y})}} \end{aligned}$$

Konfidensintervall

- Andel i oändlig population: $I_p = \hat{p} \pm \lambda_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
- Andelsskillnad i oändlig population: $I_{p_2-p_1} = \hat{p}_2 - \hat{p}_1 \pm \lambda_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$
- Andel i ändlig population: $I_p = \hat{p} \pm \lambda_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}\frac{N-n}{N-1}}$
- Andelsskillnad i ändlig population: $I_{p_2-p_1} = \hat{p}_2 - \hat{p}_1 \pm \lambda_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}\frac{N_1-n_1}{N_1-1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}\frac{N_2-n_2}{N_2-1}}$
- Väntevärde (σ känd): $I_\mu = \hat{\mu} \pm \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
- Väntevärdesskillnad (samma varians): $I_{\mu_2-\mu_1} = \hat{\mu}_2 - \hat{\mu}_1 \pm \lambda_{\alpha/2}\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
- Väntevärde (σ okänd): $I_\mu = \hat{\mu} \pm t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}$
- Väntevärdesskillnad (samma varians): $I_{\mu_2-\mu_1} = \hat{\mu}_2 - \hat{\mu}_1 \pm t_{n_1+n_2-2,\alpha/2}s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
- Stickprov-i-par: $I_\Delta = \hat{\Delta} \pm t_{n-1,\alpha/2}\frac{s_z}{\sqrt{n}}$

Hypotesprövning

Under $H_0 : p = p_0$ (alternativt $\mu = \mu_0$, $\mu_{y0} = \mu_{x0}$, $\rho = 0$ etc.) gäller:

$$X \sim \mathrm{Bin}(n,p) \implies T_p = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim_{\text{appr.}} \mathrm{N}(0,1)$$

$$X \sim \mathrm{Hyp}(M,n,p) \implies T_p = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}\frac{N-n}{N-1}}} \sim_{\text{appr.}} \mathrm{N}(0,1)$$

$$X \sim \mathrm{N}(\mu,\sigma) \implies T_\mu = \frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}} \sim \mathrm{N}(0,1), \quad \sigma \text{ känd}$$

$$X \sim \mathrm{N}(\mu,\sigma) \implies T_\mu = \frac{\hat{\mu} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}, \quad \sigma \text{ okänd}$$

$$X \sim \mathrm{N}(\mu_x,\sigma),\ Y \sim \mathrm{N}(\mu_y,\sigma) \implies T_{\mu_y-\mu_x} = \frac{\hat{\mu}_y - \hat{\mu}_x - (\mu_{y0} - \mu_{x0})}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim \mathrm{N}(0,1)$$

$$\rho = 0 \implies T_R = \sqrt{n-2}\frac{R}{\sqrt{1-R^2}} \sim t_{n-2}$$

Homogenitetstest

r populationer, k klasser.

$$T = \sum_{i=1}^{r}\sum_{j=1}^{k} \frac{(X_{ij} - n_i p_j)^2}{n_i p_j} \sim_{\text{appr.}} \chi^2_{(r-1)(k-1)}$$

Oberoendetest

r variabler, k nivåer.

$$T = \sum_{i=1}^{r}\sum_{j=1}^{k} \frac{(X_{ij} - np_{\cdot i}p_{j\cdot})^2}{np_{\cdot i}p_{j\cdot}} \sim_{\text{appr.}} \chi^2_{(r-1)(k-1)}$$

Test av enkel hypotes

r klasser

$$T = \sum_{i=1}^{r} \frac{(X_i - np_i)^2}{np_i} \sim_{\text{appr.}} \chi^2_{r-1}$$