Formelsamling

L9MA30, LGMA30, matematisk statistik för lärare

Diskreta fördelningar

Fördelning	Sannolikhetsfunktion	μ	Varians	Användning
Bernoulli Ber(p)	$p_X(k) = \begin{cases} p & , k=1 \\ 1-p & , k=0 \end{cases}$	p	$p(1-p)$	Ett försök som antigen lyckas ($k=1$) eller misslyckas ($k=0$)
Binomial Bin(n,p)	$p_X(k) = \binom{n}{k} p^k(1-p)^{n-k}$, $k = 0, 1, \ldots, n$	np	$np(1-p)$	Antalet lyckade bland n stycken oberoende Bernoulli försök
Hyper-geometrisk Hyp(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	$np(1-p)\frac{N-n}{N-1}$	Dra n bollar utan återlägging från N bollar, där Np bollar är svarta, och sedan räkna antalet svarta bollar bland de dragna
Likformig U(N)	$p_X(k) = \frac{1}{N}$, $1 \le k \le N$	$\frac{N+1}{2}$	$\frac{(N+1)(N-1)}{12}$	Exempel med $N=6$: tärningsvärde
För-första-gången FFG(p)	$p_X(k) = p(1-p)^{k-1}$, $k = 1, 2, \ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	Antalet oberoende Bernoulli försök t.o.m. det första lyckade
Geometrisk Geom(p)	$p_X(k) = p(1-p)^k$, $k = 0, 1, 2, \ldots$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$	Antalet misslyckade försök innan det första lyckade
Poisson Po(λ)	$p_X(k) = \frac{\lambda^k}{k!}e^{-\lambda}$, $k = 0, 1, 2, \ldots$	λ	λ	Används när något sker med en viss frekvens under ett givet intervall. T.ex. tryckfel per sida

Kontinuerliga fördelningar

Fördelning	Täthetshetsfunktion	μ	Varians	Användning
Likformig U(a,b)	$f_X(x) = \frac{1}{b-a}$, $a \le x \le b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	Exempel med $a=0, b=1$: ett slumptal
Exponential Exp(β)	$f_X(x) = \frac{1}{\beta}e^{-\frac{x}{\beta}}$, $x \ge 0$	β	β^2	Livslängder
Normal N(μ,σ)	$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $-\infty < x < \infty$	μ	σ^2	Summan av många oberoende relativt små faktorer

Standardiserad slumpvariabel: om $Z = \frac{X-\mu}{\sigma}$, då $\mathrm{E}(Z) = 0$ och $\mathrm{Var}(Z) = 1$. För X med normalfördelning $\mathrm{N}(\mu, \sigma)$ får man $Z \sim \mathrm{N}(0,1)$ och

$$\mathrm{P}(X \le x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

där $\Phi(\cdot)$ är fördelningsfunktionen för standard normal fördelning $\mathrm{N}(0,1)$.

Varians, kovarians och korrelationskoefficient

$$\mathrm{Var}(X) = \mathrm{E}[(X-\mu)^2], \quad \mathrm{Cov}(X,Y) = \mathrm{E}[(X-\mu_X)(Y-\mu_Y)], \quad \rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y}$$

Additionsregel för varians

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$$

Centrala gränsvärdessatsen

Låt $X_1, X_2, X_3, \ldots$ vara en följd av oberoende slumpvariabler med samma sannolikhetsfördelning där väntevärdet är μ och standard avvikelse är σ. Då för varje x gäller

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

Approximationer

$$\mathrm{Bin}(n,p) \quad \approx \quad \begin{cases} \mathrm{N}(np, \sqrt{np(1-p)}) & , \text{ om } np(1-p) > 10 \\ \mathrm{Po}(\lambda) & , \text{ om } p < 0.1,\ \lambda = np \end{cases}$$

$$\mathrm{Hyp}(N,n,p) \quad \approx \quad \begin{cases} \mathrm{Bin}(n,p), & , \text{ om } \frac{n}{N} < 0.1 \\ \mathrm{N}(np, \sqrt{np(1-p)\frac{N-n}{N-1}}) & , \text{ om } np(1-p)\frac{N-n}{N-1} > 10 \\ \mathrm{Po}(\lambda) & , \text{ om } \frac{n}{N} + p < 0.1,\ \lambda = np \end{cases}$$

$$\mathrm{Po}(\lambda) \quad \approx \quad \mathrm{N}(\lambda, \sqrt{\lambda}), \text{ om } \lambda > 15.$$

Punktskattningar

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad \hat{p} = \bar{x} \quad \text{givet populationsfördelning är Ber}(p)$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}(x_i^2) - n\bar{x}^2\right)$$

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

$$r = \frac{s_{xy}^2}{s_x s_y}$$

Medelfel av en punktskattning

- medelfel av $\hat{p}$ i en oändlig respective ändlig population

$$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}} \text{ respective } s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}\frac{N-n}{N}}$$

- medelfel av $\bar{x}$

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Approximativa konfidensintervaller för andelar

- i en oändlig population med andel p

$$I_p = \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- i en ändlig population med andel p

$$I_p = \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}\frac{N-n}{N}}$$

Approximativa konfidensintervaller för andelsskillnader

- för på två oändliga populationer med andelar p_1 och p_2

$$I_{p_2-p_1} = \hat{p}_2 - \hat{p}_1 \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

- för på två ändliga populationer med andelar p_1 och p_2

$$I_{p_2-p_1} = \hat{p}_2 - \hat{p}_1 \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}\frac{N_1-n_1}{N_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}\frac{N_2-n_2}{N_2}}$$

Approximativt konfidensintervall för väntevärdet

$$I_\mu = \bar{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}$$

Exakta konfidensintervaller för väntevärde

Låt $(x_1, \ldots, x_n)$ vara ett stickprov från $\mathrm{N}(\mu, \sigma)$ fördelning

- känd σ

$$I_\mu = \bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

- okänd σ

$$I_\mu = \bar{x} \pm t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}$$

Exakta konfidensintervaller för väntevärdesskillnad

Låt $(x_1, \ldots, x_{n_1})$ och $(y_1, \ldots, y_{n_2})$ vara två stickprov från två populationer

- för två oberoende stickprov från $\mathrm{N}(\mu_1, \sigma)$ och $\mathrm{N}(\mu_2, \sigma)$ fördelningar med känd σ

$$I_{\mu_1-\mu_2} = \bar{x} - \bar{y} \pm z_{\alpha/2}\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

- för två oberoende stickprov från $\mathrm{N}(\mu_1, \sigma)$ och $\mathrm{N}(\mu_2, \sigma)$ fördelningar med okänd σ

$$I_{\mu_1-\mu_2} = \bar{x} - \bar{y} \pm t_{\alpha/2}(n_1+n_2-2)s_{\text{pool}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

med

$$s^2_{\text{pool}} = \frac{\sum_{i=1}^{n_1}(x_i-\bar{x})^2 + \sum_{i=1}^{n_2}(y_i-\bar{y})^2}{n_1+n_2-2}$$

- för stickprov-i-par då $n_1 = n_2 = n$ och skillnader $d_i = x_i - y_i$ har samma normal fördelning

$$I_{\mu_1-\mu_2} = \bar{d} \pm t_{\alpha/2}(n-1)\frac{s_d}{\sqrt{n}}$$

med stickprovsvarians för skillnader

$$s_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i-\bar{d})^2$$