Stats Methods formulae and distributions

Probability Calculus

Discrete

Addition rule: $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

Multiplication rule: $P(A \text{ and } B) = P(A|B)P(B)$

Law of total probability: $P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$

Bayes' theorem for single events: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

Bayes' theorem for multiple exhaustive mutually exclusive events:

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

Continuous

Bayes' theorem: $p(y|x) = \frac{p(x|y)p(y)}{p(x)} = \frac{p(x|y)p(y)}{\int_{-\infty}^{\infty} p(x|y)p(y)\mathrm{d}y}$

Random variates, expectation and variance

for variates X, Y, such that $Y = \sum_{i=1}^{n} a_i X_i$:

$E[Y] = \sum_{i=1}^{n} a_i \mu_i$

$V[Y] = \sum_{i=1}^{n} a_i^2 \sigma_i^2$

for variates X drawn from continuous distributions:

$E[X] = \mu = \int_{-\infty}^{+\infty} x p(x) \mathrm{d}x$

$V[X] = \sigma^2 = E[(X - \mu)^2)] = \int_{-\infty}^{+\infty} (x - \mu)^2 p(x) \mathrm{d}x$

Correlation coefficient:

$\rho(X, Y) = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$

Discrete probability distributions

Binomial

pmf: $p(x|n,\theta) = \frac{n!}{(n-x)!x!}\theta^x(1-\theta)^{n-x} \quad \text{for } x = 0, 1, 2, ..., n.$

variates: for $X \sim \text{Binom}(n,\theta)$, $E[X] = n\theta$ and $V[X] = n\theta(1-\theta)$

Poisson

pmf: $p(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$

variates: for $X \sim \text{Pois}(\lambda)$, $E[X] = \lambda$ and $V[X] = \lambda$

Continuous probability distributions

Uniform

pdf: $p(x|a,b) = 1/(b-a) \quad \text{for} \quad a \leq x \leq b$

variates: for $X \sim \text{U}(a,b)$, $E[X] = (b+a)/2$ and $V[X] = (b-a)^2/12$

Normal

pdf: $p(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/(2\sigma^2)}$

variates: for $X \sim \text{N}(\mu,\sigma)$, $E[X] = \mu$ and $V[X] = \sigma^2$

χ^2

pdf: $p(x|\nu) = \frac{(1/2)^{\nu/2}}{\Gamma(\nu/2)}x^{\frac{\nu}{2}-1}e^{-x/2}$

where for integers n, $\Gamma(n) = (n-1)!$ and $\Gamma(n+\frac{1}{2}) = \frac{(2n)!}{4^n n!}\sqrt{\pi}$.

variates: for $X \sim \chi^2_\nu$, $E[X] = \nu$, $V[X] = 2\nu$

Test statistics

Z-statistic: $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

t-statistic: $T = \frac{\bar{x}-\mu}{s_x/\sqrt{n}}$

χ^2-statistic: $\chi^2 = \sum_{i=1}^{n} \frac{(y_i-\mu_i)^2}{\sigma_i^2}$

Table of significance levels for different statistics

Distribution	1-σ	2-σ	3-σ	4-σ	5-σ
Standard normal p-values	0.3173	0.0455	0.0027	6.33×10^{-5}	5.73×10^{-7}
χ_1^2	1.0	4.0	9.0	16.0	25.0
χ_2^2	2.296	6.18	11.829	19.334	28.744
χ_3^2	3.527	8.025	14.156	22.061	31.812
χ_4^2	4.719	9.716	16.251	24.502	34.555
χ_5^2	5.888	11.314	18.205	26.766	37.095
t_1	1.837	13.968	235.801	10050	1110441
t_2	1.321	4.527	19.207	125.64	1320.7
t_3	1.197	3.307	9.219	32.616	156.68
t_4	1.142	2.869	6.62	17.448	56.848
t_5	1.111	2.649	5.507	12.281	31.847

Table 1: p-values and corresponding statistic values corresponding to standard normal significance levels. Subscripts give number of degrees of freedom. Note that for the one-sample t-test, d.o.f.=$n-1$ where n is the number of data points. For χ^2 statistics the significance levels correspond to those calculated for the survival function $(1-F(X))$ where $F(X)$ is the cdf value of the statistic value X. t is a 2-tailed distribution like the normal distribution, so statistics X correspond to p-values equal to $2 \times (1-F(X))$.