|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Prob | | Random experiment: can be repeated indep under identical conditions  Random vars: numerical outcomes of a random experiment  Prob: limiting relative frequency of an even occurrence as the experiment is repeated many times  Run experiment once to get x, a realisation of X | | | | | | r.v. X has a distribution:  1) Discrete: P(X=xi) = pi, i = 1,...,n, E(X) = ,  n can be ∞ yet E(X) be finite  2) Cts: density fn is f(x), x , E(X) =  More generally E{h(x)} = | | | |
| Expectation & variance | | | | For r.v. X, let = E(X).  E(X - ) = 0 E{X(X-)} = E(X2) - | | | | SD(X) = where var(X) = E{(X-)2} = E(X2) -  Var(X - c) = Var(X) | | | |
| Approx | | | = E(X), and = SD(X). Then "X is around , give or take or so"  For Z~N(0,1). E(Z2) = Var(Z) + [E(Z)]2 = 1 as n ∞  Law of Large Numbers: As n ∞, sample mean = E(X)  More generally: As n ∞, E{h(X)}  Monte Carlo approximation: E{h(X)} ≈ , i.e. use many simulations to find E{h(X)} | | | | | | | | Formulae is faster and easier to compute, no issue w random error unlike simulation methods such as Monte Carlo |
| Binomial RV | | | Let X ~ Binomial(100, 0.8) produces realisations x1,...,xn where n = 1000.  ≈ E(X) = np = 80. ≈ E(X2) = Var(X) + [E(X)]2 = 16 + 802 = 6416  ≈ E(√X). ≈ E{h(X)} = Pr(X < 80), where h(x) = | | | | | | | x <- rbinom(1000, 100, 0.8)  E(X) = mean(x), E(X2) = mean(x^2), E(√X) = mean(sqrt(x)), E{h(X)} = mean(x < 80) = pbinom(79, 100, 0.8) | |
| IID r.v | Independent and identically distributed. Independent: Pr(X1 = a, X2 = b) = Pr(X1=a)Pr(X2=b) for any a,b OR E(X3X4X5) = E(X3)E(X4)E(X5)  Identically distributed: For any a, Pr(Xi=a) is same across i  Let X1,...,Xn be IID r.v., where Xi X (equal in dist), we can view xi as a realisation of Xi, 1 ≤ i ≤ n  then x1 + x2 is around 2, give or take √2, x1+x2 is a realisation of X1+X2~(2, 2) since X1,X2 iid  ~(, /n), var() = var() = var()/n2 = []/n2 = /n, then is around , give or take /√n | | | | | | | | | | |
| Algebra of r.v. | | | Let X, Y be r.v. and a,b,c be constants. Z = aX + bY + c also an r.v.  If x and y are realisations of X and Y, then ax + by + c is a realisation of Z | | | | | | | E(Z) = aE(X) + bE(Y) + c  If a=b=0, z = c is also a r.v. | |
| 2 r.v. | | | X and Y have joint density f(x,y) = fX(x)fY(y|x)  fX(x) is the marginal density of X  fY(y|x) is the conditional density of Y given X = x  Generally, joint = marginal \* conditional  If X and Y are independent, f(x,y) = fX(x)fY(y)  E(X) = =  E{h(X,Y)} = | | | Let = E(X), = E(Y). X- aka deviation  cov(X,Y) = E{ (X-)(Y-) } = E(XY) - .  cov(X, X) = var(X). var(aX+bY+c) = a2var(X) + b2var(Y) + 2abcov(X,Y)  X and Y indep cov(X, Y) = 0. Converse may not be true  cov(X, Y) = 0 cor(X, Y) = 0 | | | | | |
| Multi-nomial dist | | Experiments has r events, E1, ..., Er w prob p1,..., pr. For 1 ≤ i ≤ r, let Xi be num of times Ei occurs in n indep trials. (X1,...,Xr) has the multinomial dist (generalization of binomial dist)  Pr(X1=x1, ..., Xr=xr) = , where x1,...xr are non-negative ints summing to n  = . X~multionomial dist(n, **p**) | | | | | E.g. let r = 3, X = (X1, X2, X3) has trinomial dist  X1 ~ Bin(n, p1), X1 + X2 ~ Bin(n, p1 + p2)  EX = , varX=  = n**p** =n{diag(**p**) - **pp**T} #for multionomial dist | | | | |
| Conditional Expectation & var | | | | | E[Y|x] = | | | | var[Y|x] = E[ (Y-E(Y|x))2 |x ] = E[Y2|x] - [E(Y|x)]2 | | |
| Extra | | Proven: = + n(z-)2  Let z = 0, then = + n()2 – (1)  Let **u** = (x1-, x2-,..., xn-) and **v** = (, ..., )  From (1), RHS = **u**T**u** + **v**T**v**, LHS = (**u**+**v**)T(**u**+**v**) = (**u**T+**v**T)(**u**+**v**) = **u**T**u** + **u**T**v** + **v**T**u** + **v**T**v** = **u**T**u** + **v**T**v** + 2**u**T**v**. Since LHS = RHS, **u**T**v** must be 0, i.e. **u** and **v** are orthogonal | | | | | | |  | | |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Mean sq error (MSE) | | MSE = E{(Y-c)2} = var(Y) + {EY-c}2 when using a constant c to predict Y.  c = EY minimises MSE, when MSE­min = var(Y) | | | | | If realisation x has been observed, MSE = E[(Y-c)2|x] = var[Y|x] + {E[Y|x]-c}2  In this case, c = E[Y|x] give MSEmin = var[Y|x] | | |
| In general var[Y] > var(Y|x). And MSEY - MSEY|x = (var[Y|x] + {E[Y|x]-EY}2) - (var[Y|x] + {E[Y|x]-E[Y|x]}2) = {E[Y|x]-EY}2 | | | | | | | |
| Random Conditional Expectations | | | | E(X) = E[E(X|Y)] | | | | var(X) = var(E[X|Y]) + E(var[X|Y]) | |
| Hence mean MSE = = E{var[Y|X}} ≤ var(Y) | | | | | |
| Standard Normal dist | | | Standard normal r.v. Z has density fn f(z) = , –∞< z < ∞  E(Z) = 0, var(Z) = 1. E(Z2) = 1. E(Z3) = 0. E(Z4) = 3  (x) = Pr(Z ≤ x) = | | | | | | Let Y = + Z ~ N(, )  Pr(Y ≤ y) = () (standardisation)  To get pdf, differentiate Pr(Y ≤ y) |
| dist | Let Z~N(0,1). X = Z2 has dist w degree of freedom (df) 1  Pr(X ≤ x) = Pr(Z2 ≤ x) = Pr( ≤ Z ≤ ) = 2Pr(Z ≤ ) = 2()  Density of X is , v > 0. E(X) = 1. Var(X) = 2. | | | | | Let V = has dist w df n, where Xi are IID dist w df 1  EV = nE(X) = n. Var(V) = 2n  V ~ Gamma(, ) | | | |
| Gamma dist | | | Let > 0 and > 0. Then Gamma(,) density is , t ≥ 0  = = | | | | | | (1/2) =  = , > 0 |
| t dist | | | Let Z ~ N(0,1) and V~ be indep, then tn = has t-dist w df n  t-dist is symmetric around 0. As n ∞, tn Z (since 1) | | | | | | |
| F dist | | | Let V ~ and W ~ be independent.  Fm,n = has F dist with (m,n) df | | | | | | If X~tn, X2 = tn2 = = ~ F1,n  As n ∞, Fm,n divided by m |
| IID r.v. | | | Let X1,...,Xn be IID r.v w mean . S2 = is known as the sample variance | | | | | | |
| and S2 normal | | | Let X1,...,Xn be IID N(, )   |  |  | | --- | --- | | A) and S2 are indep | B) ~ N(, ) | | C) = ~ | D) ~ tn-1 | | | For any i, and Xi – are uncorrelated, hence indep by normality  and (X1 – , ..., Xn – ) are indep, which implies A)  From tutorial 1, = + n  LHS ~ , rightmost term ~ , thus C) is likely (actual proof too diff) | | | | |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Population | | | Interested in variable v in a pop of N individuals, where individual i has fixed value vi  = and = . Both are called parameters, i.e. measured from all individuals in pop. | | | | |
| Sampling | Let X be the result of a random draw, i.e. every individual has equal prob of being chosen  E(X) = . var(X) = . Random sample is representative of the pop  Let M be the index of the chosen individual. Then X = vM. M ~ U(1,N). X = g(M), where g(m) = vm  "Proof": E(X) = = =  Pr(X = vm) = only if all the vm are unique | | | | | | |
| Simple random sampling (SRS) | | | | | SRS of size n: make n random draws w/o replacement. Let results be denoted Xi for i = 1,...,n. E(Xi) = . var(Xi) = . – (1)  For i ≠ j, cov(Xi, Xj) = (where N is pop size). X1 and X2 are negatively correlated – (2) | | |
| Justifica-tion of (1) and (2) | | Let M­1, ..., MN be successive random indices.  Theorem: (M1,...,MN) is uniformly distributed on all premutations of {1,...,N} i.e. order of ppl u pick from pop doesn't matter  Theorem implies Mi is uniformly distributed on {1,...,N}, hence (1) holds  Theorem implies (Mi, Mj) is uniformly distributed on {(a,b), 1 ≤ a ≤ N, 1 ≤ b ≤ N, a ≠ b}. (2) follows from calculating cov(X1, X2)  M1, ..., MN and X1,...,Xn are examples of exchangeable r.v. | | | | | |
| Exchangeable r.v | | | | R.v. Y1,...Yk are exchangeable if all reordered vectors have the same dist as (Y1,...,Yk), i.e. for any permutation p on {1,...,k}, (Yp(1),...,Yp(k)) (Y1,...,Yk)  IID r.v are exchangeable. Exchangeable rv might not be indep, but are identically distributed | | | |
| SRS mean | Let X1,...,Xn be SRS from a large pop of size N, w mean and var .  = . E = . Var() = (contains cov since not indep)  If n N, ≈ 1, so SRS is like draws w replacement | | | | | | Assume X1, ..., Xn are IID  To a high deg of accuracy, Var() = |
| Estimator | Given data x1,...,xn, estimate w the natural = .  is an estimate of . is a realisation of the estimator  has an error of – , which cannot be calculated nor estimated. | | | | | Since E() = , is an unbiased estimator  is an unbiased estimate | |
| Instead, we use the standard error to quantify error in : SE = (when n N)  SE of is defined as the SD of the estimator, i.e. SD() = | | | | | Note (n-1)s2 = n  More accurately, SD() = | |
| Intuitive estimate of : =  Sample variance: s2 = =  SE of is then estimated by = = | | | | | s2 better as E(S2) = (unbiased estimator)  But E() =  is estimated as , give or take (SE) | |
| Sampling Assumption | | | Estimation above works well for SRS, i.e. n N. Then X1,...,Xn are effectively IID r.v  If is relatively large, modify SE by correction factor | | | | |
| Estimating proportion (n N) | | Special case of SRS. In a 0-1 population, let p be the proportion of 1's  Population params: = p, = p(1-p)  Sample params: is the realised proportion of 1's in the sample which estimates p, SE is =  SE can be estimated by replacing p w its estimate =  Note is the estimator, which is the random proportion of 1's in the SRS.  E() = p, var() = , SE = SD() = (which is estimated by replacing p w the realisation of ) | | | | | |
| Hat notation | is an estimator of p, a r.v  Since is an estimator of , can write = . But no small letter to represent realisation of estimator  - hat symbol may denote realisation (try to avoid doing this) | | | | | | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Normal Approx | Let X1,...,Xn be IID r.v. w mean and var . (X can have any dist). As n ∞, dist of converges to N(0,1).  For sufficiently large n, ~ N(0,1) approximately (by CLT) | | | |
| Quantiles | Suppose X has a strictly increasing CDF F. For 0 < p < 1, the p-quantile of X is q = F-1(p) (i.e. Pr(X ≤ q) = p) | | Let Z~N(0,1). The p-quantile of Z is (p).  pnorm(): find p. qnorm(): find q | |
| Upper tail quantile | | For 0 < p < 0.5, let zp be s.t. Pr(Z > zp) = p  Then zp = (1-p) quantile of Z = (1 - p). | | zp = -z1-p  Pr(-zp ≤ Z ≤ zp) = 1-2p |
| Approx dist of | Let 0 < < 1. For large n, Pr ≈ 1 –  Consequently, Pr ≈ 1 – | | | Approximately, ~N(, ) |
| Random interval for | Let 0 < < 1. For large n, Pr ≈ 1 –  is a random interval. A realisation of gives the realised interval  If we generate many such interval, 100(1-)% of these intervals would contain | | | |
| CI for | Suppose is unknown but is known. An approximate (1-)-CI for is  If is also unknown. An approximate (1-)-CI for is OR  Note is a constant, not r.v. So writing Pr(c1 ≤ ≤ c2) ≈ ... is not correct.  For a real pop of size N, if n/N is is not small, use the adjusted SE (multiply by ) | | | |
| Normal data: exact CI for | | Let x1,...,xn be realisations from IID **N(, )** r.v. X1,...,Xn w mean and sample SD s  Let tn-1(/2) be s.t. Pr(tn-1 > tn-1(/2)) = /2  A (1-)-CI for is | | In practice, we don't know w certainty if pop is normal, so CI is also approx |
|  | CI does not take care of measurement bias.  MSE = SE2 + bias2 | | | As n ∞, MSE approaches bias2  If bias = 0, then MSE = SE2 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Poisson dist | Let > 0. The Poisson r.v. X has dist Pr(X = k) = , k = 0,1,2,...  = 1. E(X) = = = | | | E{X(X-1)} = .  So var(X) = E{X(X-1)} + EX – [EX]2 = | |
| Fitting Poisson dist to data | Assume counts x1,...,xn are realisations of IID Poisson() r.v. X1,...,Xn  is analogous to a param of a real pop, such as or  is used to estimate . SE = SD() = (since var() = )  CI for is () | | | Estimator of is =  SE = SD() = SD() | |
| Parameter space | Poisson: f(x|) = , x = 0,1,... and , the parameter space  Bernoulli: f(x|p) = px(1-p)1-x, x = 0,1 and p , the parameter space | | | | |
| Estimation problem | Assume data x1,...,xn are realisations of IID r.v. X1,...,Xn w density f(x|), estimate  Parameter lies in . is the parameter space.  From data, we estimate , calculate approximate SE (bootstrap), construct CI (if n large enough) | | | | Poisson: =, =  Bernoulli: = p, = (0,1)  Normal: = (, ), = |
| Moments | kth moment of an r.v. X is = E(Xk), k = 1,2,.... Var(X) =  Let X1, ..., Xn be IID w same dist as X. kth sample moment is =  is an estimator of , where E() = . i.e. is an estimate of | | | Poisson: = , = + , = .  Bernoulli: = p, =  Normal: = , = + ,  = = , = =  E() = , but E() = (tut3) | |
| General MOM estimator | Let X1,...,Xn be IID w density f(x|), where lies in , for some p ≥ 1  Suppose = g(,...,), the MOM estimator is = g(,...,) | | | Poisson: g(x) = x. = g()  Bernoulli: g(x) = x. p = g()  Normal: . (, ) = g(, ) | |
| Gamma dist | Let X1,...,Xn be IID w Gamma() dist. > 0 and > 0 are the shape and rate params  From tut2, . So , | | |  | |
| Example | Given = 0.224, = 0.1338, MOM estimates of , is 0.38, 1.67  SE of = SD() = SD, SE of = SD() = SD  But no exact formula for both SE, so cannot bootstrap. So approximate dist by an estimated dist. e.g. Gamma(0.38, 1.67) which are realisations of estimators | | Let be IID Gamma(0.38,1.67) r.v  Then = , and =  Then SE of 0.38: SD ≈ SD  SE of 1.67: SD ≈ SD | | |
| Monte Carlo | Aim: To get SD and SD  1. Generate realisations Gamma(0.38,1.67)  2. Calculate a realisation of and from 1.  3. Repeat 1 and 2 to get many realisations  4. Estimate SD and SD | Can also get bias for each parameter from Monte Carlo  Bias for estimated by: E  Bias for estimated by: E  1. Estimate and from realisation generated on the left  2. So is around (0.38 – bias for ) ± SD, around (1.67-bias) ± SD | | | |
| As n ∞ | By law of large numbers, . Hence  SE of estimates decrease by , while bias 0, i.e. consistent/asymptotically unbiased  Every MOM estimator is consistent: as n ∞, estimator parameter | | | | |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Poisson likelihood | | Let x1,...,xn be realisations of IID Poisson() r.v. X1, ..., Xn. f(x|) = , x = 0,1,2,...  Joint probability = P(X1 = x1, X2 = x2,..., Xn = xn) = f(x1|)\* f(x2|)\*...\* f(xn|) = = L() (likelihood). | | | | | | | | |
| Normal likelihood | | Let x1,...,xn be realisations of IID N(, ) r.v. X1, ..., Xn. f(x|,) = ,  Joint density = f(x1|,)\*...\* f(xn|,) = = L(,) (likelihood). | | | | | | | | |
| Likelihood fn | | Let x1,...,xn be realisations of IID r.v. X1, ..., Xn with density/mass fn f(x|),  Likelihood fn, L() = f(x1|)\*...\* f(xn|) = . (xi dont really matter here, only concerned w ) | | | | | | | | |
| Loglikelihood fn | | | | | = log L() = | | Poisson: =  Normal: , = | | | |
| MLE for Poisson | | =  = . Set = 0. = | | | | < 0 | | | | |
| MLE for Normal | | , =  = .  Set . . Set 0. | | | | | | .  (Hessian Matrix) | | |
| Estimator | | Poisson: ML Estimate = , Estimator:  Normal: ML Estimate = (,), Estimator: | | | | | | | | |
| Refined defn | | For Poisson: Instead of L() = . Use L() = (remove constant factors)  Use = (remove additive constants) | | | | | | | | Since only maximisers are of interest, and max values of fns seldom used. |
| MLE of Gamma | | Gamma Logarithm of density (, x ≥ 0):  loglikelihood: n  So ML estimators (, ) satisfy: and = 0, where | | | | | | | | . Set to 0 to get  Use numerical mtd to estimate |
| SE and bias | | Bootstrap approximation: SE of estimate = SD(estimator). bias of estimate = E(estimator) – parameter  E.g. ML estimate of and = 0.44,1.96. SE in 0.44 is SD() ≈ SD(), Bias in 0.44 is E() – ≈ E() –  SE in 1.96 is SD() ≈ SD(). Bias in 1.96 is E() – ≈ E() –  and are based on IID RV w Gamma(0.44,1.96)  Using Monte Carlo approximation (generate many and ): SD() ≈ 0.03. E() – ≈ 0.00. SD() ≈ 0.26. E() – ≈ 0.04  is around 0.44 ± 0.03. is around 1.96 ± 0.26. Less bbiased estimate is is around 1.92 ± 0.26  ML is better than MOM in terms of SE and bias | | | | | | | | |
| Multinomial | | | | Let (x1,...,xr) be a realisation of (X1,...,Xr) ~ Multinomial(n, (p1,...,pr))  L(p1,...,pr) = .  Since p1 + ... + pr = 1 (constraint imposed), so cannot just differentiate directly. Instead,  Define Lagrangian fn  , . Set . Then .  . (since = n). So pi = | | | | | ML estimate of pi is  ML estimator of pi is  SE of estimate is SD() =  Similar to estimating p from X~Bin(n,p) | |
| Genetics | | Chromosome comes in pairs, consisting of DNA bases: A,C,G,T. Locus is a subsequence of a chromosone. Alleles are diff versions of bases at a locus. An unordered pair of allels is a genotype  E.g. ABO locus has 3 alleles: a,b,o. Genotype are aa, ab, ao, bb, bo, oo  If k alleles: then num of possible genotypes are k + = | | | | | | | | |
| Hardy-Weinberg equilibrium | | | A population is in HWE at a locus if the genotype proportions are , where pi is the proportion of allele ai  ­E.g. locus in HWE has 2 alleles A and a, and the proportion of a is . The genotype proportions are AA: , Aa: 2(1– ), aa:  Num of a alleles in a random person ~ Binomial(2, ). Total num of a allels in an SES of size n: Binomial(2n, ) | | | | | | | |
| E.g. | Sample frequencies are AA: 342, Aa: 500, aa:187. Assume data come from SRS and HWE holds. Then (X1,X2,X3) ~ Multinomial(1029, (, 2(1-), ), where is population proportion of a  L() = .  ML estimate of is . Note (2x1 + x2) + (x2 + 2x3) = 2n. ML estimator is  X2 + 2X3 is num of a alleles: Binomial(2n, ). var() =  SE in 0.42, by bootstrap is | | | | | | | | | |
| Summary | | | |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | Dist | Method | Parameter | Estimator | SE | Monte Carlo | Bias | | simplest case |  | /p |  |  | No | No | | Gamma | MOM |  |  | ? | Yes | Yes | |  | ML |  | ? | ? | Yes | Yes | | | | | | | | |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Large-sample variance of ML estimator | | | | Let be the ML estimator of , based on IID r.v. X1,...,Xn w density f(x|)  As n ∞, var() ≈ , where the Fisher information I() is determined by f(x|)  For large n, Monte Carlo may not be required to estimate the SE of ML estimate. Bootstrap is still needed  MOM estimate is given by a formula more often than ML estimate, but opp is true for SE, for large n | | | | | | | |
| Fisher Info: Poisson | | f(x) = , x = 0,1,.... log f(X) = X log – – log X!  , < 0 | | | | | | | Fisher information is I() = ­–E | | |
| Fisher Info: Bernoulli | | f(x) = px(1-p)1-x, x=0,1. log random density: log f(X) = X log p + (1-X)log (1-p)  , < 0 | | | | | | | Fisher information is I() = ­–E | | |
| Fisher Info: Normal | | X~N(), . f(x) = , -∞ < x < ∞  log f(X) = | | | | | | | I() = –E() = | | |
| Fisher Info | | Let X have density f(x|), . The Fisher information is the p x p matrix I() = –E  I() is symmetric, w (i,j)-entry = –E which is – or – | | | | | | | | | |
| Interpre-tation | | I() measure the info about in one sample X~f(x|)  If X~Poisson(). I() = . The larger is, the less information is in X | | | | | | | | | |
| n IID samples | | IID X1,...,Xn w density f(x|) can be regarded as a sample from **X** = (X1,...,Xn) w density g(**X**|) = f(X1|)\*...\*f(Xn|).  The information in **X** = –E is nI(), where I() is the information in any one of the X's | | | | | | | | | |
| Binomial | | f(x) = px(1-p)n-x, x=0,1,...,n. log f(X) = log + x log p + (n-x) log (1-p)  , . I(p) = =  I(p) = information in Binomial(1,p) = Bernoulli(p) | | | | | | 1 binomial(n,p) sample has the same info about p as n IID Bernoulli(p) samples  Multinomial(n,(p1,...,pr)) has the same info about (p1,...,pr) as n IID Multinomial(1,(p1,...,pr)) samples | | | |
| HWE trinomial dist | | **X** = (X1, X2, X3) ~ Multionomial(n,**p**) where p1 = , p2 = 2(1– ), p3 = , 0 < < 1  P(X = x) =  log f(X) = (2X1 + X2)log (1-) + (X2 + 2X3)log (ignore constant coefficient)  log 'f(X) = , log ''f(X) =  X2 + 2X3 ~ Bin(2n,), 2X1 + X2 ~ Bin(2n,1-)  I() = | | | | | | | | | X1 = AA, X2 = Aa, X3 = aa  #density of trinomial  #X2 + 2X3 = num of a alleles in n samples  I() = , information in Trinomial(1,**p**()) |
| General Multinomial data | | | **X** ~ multionomial(n, (p1,...,pr)), = (p1,...,pr-1). pr = 1 - p1 - ... - pr-1  log f(**X**) =  , 1 ≤ i ≤ r-1 | | | | | , 1 ≤ i ≤ r-1, , 1 ≤ i ≠ j ≤ r-1  (i,j)-entry of I() = | | | |
| Multinomial Parameterisation | | | | | General trinomial dist has 2 params, p1,p2. HWE trinomial is defined by a single param  Any multionomial dist can be described as Multionomial(n,**p**()), **p**() = (p1(),..,pr()), , 1 ≤ k ≤ r-1 | | | | | | |
| Summary: Fisher info | | | | | | Let X have mass/density f(x|), . The fisher info at in X is the p x p matrix –E | | | | | |
| Variance of ML estimator | | Poisson: I() . variance of ML estimator : var() = =  Bernoulli: I(p) = , = = variance of ML estimator  Conjecture: var() =  Normal: I() ,  But variance of ML estimator = (, ): var() = var(, ) =  Revised conjecture: var() ≈ for large n | | | | | | | | = .  ~ . var() = 2(n-1)  var() = var = | |
| Gamma dist | X~Gamma(),  log f(X) =  = , | | | | | | , ,  I() = –E | | | | |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Main Result | | : ML estimator of , based on either  1. IID r.v. X1,...,Xn w density f(x|). I(): Fisher information in any Xi  2. (X1,...,Xr)~Multionomial(n,**p**())). I(): Fisher information in Multionomial(1,**p**())  As n ∞, the distribution of converges to N(**0**,**I**p)  For large n, approximately  ML estimators are asymptotically unbiased, and consistent: | | | | | | Note , meaning is standardized form | | | |
| Poisson and Bernoulli | | X­1­,..,X­n­ IID Poisson(). . I() = . For large n, approximately  X­1­,..,X­n­ IID Bernoulli(p). . I(p) = . For large n, approximately | | | | | | | | Note is discrete | |
| Normal | | X1,...,Xn IID N().  Variance of ML estimators: | | | | | | For large n, approximately  Note distribution of and independence are exact | | | |
| HWE trinomial | | **X** = (X1,X2,X3) ~ Trinomial(n, **p**), where p1 = , p2 = 2(1– ), p3 = .  Information in a Trinomial(1,**p**) distribution is I() = (same for Binomial(2n,)) | | | | | | | | | For large n, approximately, (same for Bin(2,)) |
| General Trinomial | **X** ~ Trinomial(n,(p1,p2,p3)). Let = (p1,p2). ML estimator:  Information in a Trinomial(1, (p1,p2,p3)) dist is I(p1,p2) = | | | | | For large n, approximately, , implying that is approximately normal  Note expectation and variance are exact  It is only approximate as using normal when is discrete | | | | | |
| Gamma | | X1,..,Xn IID Gamma(). ML estimators and cannot be expressed algebraically.  The Fisher information is I, where is the digamma function | | | | | | | For large n, approximately | | |
| Normal approx for ML estimator | | | | : ML estimator of (note p = 1 here). 0 < < 1  For large n, 1 – ≈ Pr | | | Hence 1 – ≈ Pr | | | | |
| CI | | For large n, the random interval  covers with probability of about 1 – | | | | From data collected, can get ML estimate of  SE is approximated by bootstrap: replace by its ML estimate in  Then (estimate - SE, estimate + SE) is an approximate (1 – ) CI for | | | | | |
| CI for Possion | | ML estimate of . I()-1 =.  Bootstrap approximation: SE = | | | | For large n, an approximate (1 – ) CI for is | | | | | |
| CI for Normal | | x1,...,xn realisations of IID N() r.v.s, n large. ML estimate of and are and  . SE of and estimated by bootstrap as and  Approximate (1 – ) CI: : . AND :  Note s is not used, but not much diff, since n is large | | | | | | | | | |
| Scope of asymptotic normality of ML estimators | | | | | Given IID normal r.vs, let be ML estimator of , so is ML estimator of  Both and are asymptotically normal (i.e. normal for large n),  More generally, let be ML estimator of . For any strictly increasing/decreasing h: , h() is the ML estimator of h(). For large n, h() is approximately normal | | | | | | |
| Multionomial Dist | | | **X** ~ Multinomial(n, (p1,...,pr)). ML estimator . = (p1,...,pr-1). I() is the information in a Multinomial(1,**p**) dist  For large n, approximately, , var = , with (i,j) entry:  Dist of implies also has an approximate normal dist, w expectation **p** and var() = (diag(**p**) – **pp**T) (var is r x r matrix) | | | | | | | | |
| Conclusion | | Both MOM and ML estimators are consistent: bias goes to 0 as n ∞  MOM uses only sample moments to estimate params. ML uses all info contained in density fn. Hence ML estimates tend to have smaller bias and SE  Asymptotic properties of ML estimators: For large n, SE can be estimated w/o Monte Carlo, and a good Ci for param is available  MOM estimators may not be asymptotically normal, so more diff to construct CI. | | | | | | | | | |
| SRS of size n from a large pop w mean and variance . . : ML estimator based on n IID r.v. or a multionomial r.v. w n trials   |  |  |  |  | | --- | --- | --- | --- | | Estimator | E | var | Distribution | |  |  | /n | ≈ normal | |  | ≈ | ≈ | ≈ normal |   Approx is better for larger n: | | | | | | | | | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Models as subsets | General die can be represented as = {(p1,...,p6): pi > 0, }  1. p1 = p2 = p3, p4 = p5 correspond to a subset of : {:}  2. p1 = p2 = p3, p4 = p5 = p6 correspond to a subset of : {:} | | | | |
| Multinomial Goodness of Fit | Let (X1,...,Xr)~Multionomial(n,**p**) w n,r fixed. Set of all distributions: = {(p1,...,pr): pi > 0, }  Consider a subset where **p** depends on , k < r-1:  Given realisations (x1,...,xr) to judge whether **p** | | | | |
| Genetic data | Assume (342,500,187) is a realisation of **X** ~ Trinomial(1029,**p**). = {(p1,...,p3): pi > 0, }  HWE says **p** is in , 2(1– ),  Goodness-of-fit test: 1. H0: **p** . 2. H1: **p** . 3. Calculate test statistic and p-value -> conclude  If H0 true, we expect X1 to ~Bin(n, p1). E(X) = np1  Calculate expected, then calculate (O-E)2/E for each Xi  The p-value is roughly Pr( ≥ ). If Pr > 0.05 -> do not reject H0 -> HWE seems to fit well | | | | |
| Pearson's X2 | Assumption: (X1,...,Xr)~Multionomial(n,**p**()), , k < r-1, i.e **p**  : ML estimator of . n**p**(): random expected counts. Chi-square statistic: X2 = . Note X2 is discrete  Thrm: As n ∞, the dist of X2 converges to  - k can be 0, in which case is a single point. Then expected counts are exact, not estimated (e.g. Tut 9 Q3, Pr( ≥ 14.2) ≈ 0.01 -> die is unfair) | | | | |
| X2 Goodness-of-fit test | (X1,...,Xr)~Multionomial(n,**p**), **p** w n large  .  1. H0: **p** . 2. H1: **p** . 3. Substitute each Xi w xi, and w ML estimate, we get a realisation x2 of X2  4. p-value: Pr(X2 ≥ x2) ≈ Pr( ≥ x2). 5. If p-value < 0.05, reject H0 | | | | |
| G statistic and likelihood ratio (LR) | G = . G is a likelihood ratio statistics (note natural log)  (X1,...,Xr)~Multinomial(n,**p**): (note pi = Xi/n)  Maximum of likelihood L(**p**) = over . L1 = L() =  Maximum of likelihood L() = over . L0 = L() =  L1/L0 ≥ 1. The larger the ratio, the more likely we reject H0  2log = G | | | | |
|  | For general trinomial dist, = {(p1,p2): pi > 0, p1 + p2 < 1}  For HWE model, = p1 = (1-t)2, p2 = 2t(1-t), 0 < t < 1  Let , and = (p1, p2). The Goodness of fit test of HWE model is H0: , H1: | | | | |
| LR statistic G | Assumption: n IID r.v. density defined by , with k1 indep parameters  L1: maximum likelihood value over . L0: maximum likelihood value over , with k0 < k1 indep parameters  Thrm: Suppose . As n ∞, the dist of G = 2log converges to  Letting l0 = log L0, l1 = log L1, G = 2(l1 - l0) | | | | |
| LR GOF test | 1. H0: . 2. H1: . 3. L0 and L1 are maximum likelihood values under and . g = 2log is a realisation of G  4. p-value: Pr(G ≥ g) ≈ Pr( ≥ g). 5. If p-value < 0.05, reject H0 | | | | |
| Multionomial GOF | | To test whether multinomial data (x1,...,xr) comes from a simpler model k < r-1 parameters, can use both G = or X2 = to test GOF. For large n = , under H0, both are approximately | | | |
| Poisson likelihood ratio | 1. : For i = 1,...,n, Xi ~ Poisson() are independent. log f(Xi) =  l(, ..., ) =  ML estimator of =  Maximum loglikelihood under : l1 = l(,...,) =  2. : Every  l() = . ML estimator of  l0 = l() =  G = 2(l1 - l0) = 2 = 2 ≈  Suppose every . For large n, G ~ approximately | | | | |
| Poisson dispersion test | | | 1. H0: rates are all equal. 2. H1: rates are not all equal. 3. Sample mean is 2.44 and sample var 4.59. g ≈ = ≈ 751  4. p-value is approx Pr( ≥ 751) ≈ 0. H0 is rejected | | |
| Normal dist LR | Based on N() realisations x1,...,xn. H0: = 0; H1: ≠ 0 | | | | |
| 1. If is known, under H0, ~ N(0,1)  P-value = Pr = Pr (for 2-tailed test)  Under H0, G = ~ exactly | | | | = , k1 = 1, , |
| = {0}, k0 = 0, (since = 0) |
| G = 2() = = ~ = |
| 2. If is unknown, under H0,  For large n, G ≈ ~ approximately | | | = {(): , > 0}, k1 = 2, , | |
| = {(0,): > 0}, k0 = 1,, | |
| G = 2() = = ~ approximately | |
| Conclusion | LR test applies in situations when testing GOF of model relative to a larger model. If n is large, P-value can be computed using a dist. Test assumes larger model is valid  P-value is not a prob that H0 is true. H0 is either true or false. P-value is computed assuming H0 is true | | | | |

|  |  |
| --- | --- |
| Sufficiency | **x** = (x1,...,xn): realisations of IID r.v. w density f(x|)  A function h(**x**) is a sufficient statistic for if the refined likelihood L() depends on **x** only through h(**x**) |
| Cramér-Rao inequality | Thrm: If is unbiased, then var() ≥ ,  is called the Cramér-Rao lower bound (CRLB) |
| Efficiency | Let be unbiased for . It is efficient if its variance is the CRLB: var() = |
| ML estimators | Since ML estimators are generally biased, efficiency is not so relevant  But they are asymptotically unbiased. And for large n, ML estimator has var() ≈  Hence is asymptotically efficient |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Distributions | X ~ Binomial(n, p). E(X) = np. var(X) = np(1-p) | Y ~ Normal(, ). E(X) = . Var(X) = | | | Z ~ Poisson(). E(Z) = . Var(Z) = |
|  | Law of Large Numbers: If x1,...,xn are realisations of X, then E(X), E(X2). Hence var(X)  Monte Carlo approximation: pick a large n, use to approximate E{h(X)}  SE of an estimate is the SD of the estimator. Bootstrap approximation  If X ~ N(, ), then ~ N(0,1) | | | | |
| Expectation & Variance | E(X) = , var(X) = E{(X-)2} = E(X2) -E(X)2, cov(X,Y) = E{ (X-)(Y-) } = E(XY) - = E(XY) - E(X)E(Y).  E(aX + bY + c) = aE(X) + bE(Y) + c. var(aX+bY+c) = a2var(X) + b2var(Y) + 2abcov(X,Y) | | | | |
| Let X1,...,Xn be IID N(, ) r.v. w mean . U = . V =  E(U) = = = n  ~ N(0,1). ~. var = 2. var = 2. var = 2.  Var(U) = (no covariance as IID) = 2n | | | S2 = . ~  ~ . E(V) = (n-1).  Var() = 2(n-1). var(V) = 2(n-1) | |
| Conditional | E[Y|X] is an r.v. w possible values/realisations E[Y|x]. var[Y|X] is an r.v. w possible values/realisations var[Y|x] | | | | |
| Let X~N(, ) and Y~N() be indep, and Z = X + Y. Z~N(, )  Given X = x, Z = x+Y ~ N(x+, )  E[Z|x] = x+. Var[Z|x] = . Hence, E[Z|X] = X+. Var[Z|X] =  E[Z|X] = X+ ~ N(, ). var{E[Z|X]} = | | So EZ = E{E[Z|X]} = E(X) + =  So var(Z) = var{E[Z|X]} + E{var[Z|X]} = | | |
| Multinomial | If (X1,...,Xr) are multinomial r.v., then the sum of any of k of them has a binomial dist, where k = 1,...,r-1. | | | | |
| (X1,...,X5) ~ Multinomial(n, (p1,...,p5)). ∑Xi = n. ∑pi = 1  var(∑Xi) = var(n) = 0. 0 = var(X1) + ... + var(X5) + 2cov(X1,X2) + ... + 2cov(X4,X5) = sum of all terms in covariance matrix  What is the Conditional dist of (X1,...,X4) given X5 = x5, where 0 ≤ x5 < n?  x1+x2+x3+x4 = n-x5  Pr(X1 = x1, ..., X4 = x4|X5 = x5) = = / = /  Given X5 = x5, (X1,...,X4)~Multinomial(n-x5, ) | | | | |