Content

$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx = E[(x-\mu)^2]$ variance

$\sigma_x^2 = E[x^2] - \mu_x^2$

$\cap$: And / intersect

$\cup$: union, or

$nP_r = \frac{n!}{(n-r)!}$ $nC_r = \frac{n!}{r!(n-r)!}$

$P(B'|A) = 1 - P(B|A)$

$P(A\cup B) = P(A) + P(B) - P(A\cap B)$

$P(A\cap B|C) = P(A|B\cap C)P(B|C)$

$P(A'\cap B') = 1 - P(A\cup B)$, $1 - P(A') = P(A)$

$P(C_n|A) = \frac{P(C_n)P(A|C_n)}{\sum^k P(C_i)P(A|C_i)}$

$P(B|A) = \frac{P(A\cap B)}{P(A)}$

Prob: $P(B|A)P(A) = P(B)P(A|B)$

indep: $P(A\cap B|C\ldots) = P(A)P(B)\ldots P(N) = P(A\cap B\cap\ldots)$

expectation + var

$E(x) = \sum x f(x) = \int x f(x)\,dx = \mu$

$E(g(x)) = \int g(x) f(x)$

$E(g(x,y)) = \iint g(x,y) f(x,y)\,dx\,dy$

$E[ax+b] = aE[x]+b$

$E(g(x,y)+h(x,y)) = E(g(x,y)) + E(h(x,y))$

marginal dist of x: $\int f(x,y)\,dy = g(x)$

covariance:

$\sigma_{xy} = \iint f(x,y)(x-\mu_x)(y-\mu_y)\,dx\,dy$

$\sigma_{XY} = E[XY] - E[X]E[Y] = \ldots - \mu_x\mu_y$

$\rho_{xy} = \sigma_{xy}/\sigma_x\sigma_y$, $\rho = 1$: correlated.

indep implies uncorrelated since $E[XY] = E[X]E[Y]$

$\sigma^2_{aX+bY+c} = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\,\sigma_{XY}$

Distributions

Binomial (constant w/ replacement)

$b(x;n,p) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x}$ ← pdf

n: # trials, p: prob success, x: # success

$\mu = E(x) = np$ $\sigma^2_X = np(1-p)$

Bernoulli process: repeated indep bool outcome

cdf: $B(x;n,p) = \sum_{x=0}^{x} b(x;n,p)$

$P(3\le x\le 6) = P(x\le 6) - P(x\le 3)$

Gamma

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

$\Gamma(\tfrac12) = \sqrt{\pi}$ $\Gamma(n) = (n-1)!$

$f(x;\alpha,\beta) = \begin{cases} \frac{1}{\beta^\alpha\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, & x>0 \\ 0, & ow \end{cases}$

$\mu = \alpha\beta$

$\sigma^2 = \alpha\beta^2$ → χ^2: $\alpha = v/2$, $\beta = 2$

Normal CDF: Φ

$n(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, std normal $\mu=0, \sigma=1$

$\mu = \mu$, $\sigma = \sigma$

Multinomial

$f(x_1\ldots x_n; p_1\ldots p_m, n) = \binom{n}{x_1\ldots x_n} p_1^{x_1} p_2^{x_2}\ldots p_m^{x_m}$

partition, prob of partition

$E[X_i] = np_i$

$\sigma^2 = npq$

Hypergeometric: w/o replacement, chance changes

$h(x;N,n,k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$ success, total, samples, total success

$\mu = nk/N$ $\sigma^2 = \frac{N-n}{N-1} n \frac{k}{N}(1-\frac{k}{N})$

Negative binom: chance of kth success on x trial

$b^*(x;k,p) = \binom{x-1}{k-1} p^k (1-p)^{x-k}$

Geometric: 1st success on x

$g(x;p) = p(1-p)^{x-1} = b^*(x;1,p)$ $\mu = 1/p$ $\sigma^2 = \frac{1-p}{p^2}$

Poisson dist: prob of x success in const. interval

$P(x;\lambda t) = \frac{e^{-\lambda}\lambda^x}{x!}$ $\mu = \lambda$ $\sigma^2 = \lambda$

$\lambda = rt$, t: length of interval, r: rate

thing to occur

$\lim_{n\to\infty} b(x;n,p) = p(x;\lambda)$, to mean/variance, Poisson.

Uniform

$f(x;A,B) = \begin{cases}\frac{1}{B-A} & A\le x\le B \\ 0 & ow\end{cases}$

$\mu = \frac{A+B}{2}$ $\sigma^2 = \frac{(B-A)^2}{12}$

Exponential = gamma, $\alpha = 1$, $\beta = \beta$

Chi^2 — gamma, $\alpha = v/2$, $\beta = 2$

MGF

rth moment: $\sum$ or $\int_{-\infty}^{\infty} x^r f(x)$ (discrete / cont.)

1st moment: mean, 2nd: $\mu^2 + \sigma^2$, recall: $\sigma^2 = E[x^2] - \mu^2$

MGF: $\sum$ or $\int_{-\infty}^{\infty} e^{tx} f(x) = M_x(t) = E[e^{tx}]$

when this is avail.

$\mu_r' = \frac{d^r M_x(t)}{dt^r}\Big|_{t=0}$ can deriv to find rth moment

X has $M_X(t)$, $Y = aX$ → $M_Y(t) = M_{aX}(t) = M_X(at)$

$Z = X+Y$, $M_Z(t) = M_X(t)M_Y(t)$

$M_{aX}(t) = M_X(at)$ → more on opp side

MATH

$\sum_{k=0}^{\infty} ar^k = \frac{a}{1-r}$

$\sum_{i=1}^{n} ar^{i-1} = \frac{a(1-r^n)}{1-r}$

$\sum_{k=1}^{\infty} \frac1k$ = divergent

$\frac{d}{dx}\ln(u) = \frac1u u'$

Lin comb

given $x, f(x)$, $Y = aX$, $h(y) = ?$

cont: $F(y) = P(Y\le y) = P(X\le Y/a) = \int_{-\infty}^{y/a} f(t)\,dt$ u sub, want μ

cont: $\frac{1}{|a|} f(Y/a)$, discrete $f(Y/a)$

CLT $t = \frac{\bar X - \mu}{\sigma/\sqrt n}$, from randomly sampled RV

$\mu = \mu$ of $\bar x$

$\sigma_{\bar x}^2 = \sigma^2/n$

$\bar X_1 + \bar X_2 \to \mu_1+\mu_2$, $\sigma^2_{\bar X_1+\bar X_2} = \ldots$ same h

$\bar X_1 - \bar X_2 \to \mu_1 - \mu_2$, $\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \sigma^2_{\bar x_1} + \sigma^2_{\bar x_2}$

Sample mean: $\mu_{\bar x} = \mu$

Sample variance: $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar X)^2$

given s^2

$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \sum_{i=1}^n \frac{(x_i-\bar x)^2}{\sigma^2}$, $v = n-1$

χ when μ, $v = n$; when $\bar x$, $v = n-1$

$Var(s^2) = var\left(\frac{\sigma^2\chi^2}{n-1}\right) = \frac{\sigma^4}{(n-1)^2}$ that $\chi^2 = \frac{2\sigma^4}{n-1}$

$E(s^2) = \sigma^2$

t-dist unknown variance ★ use table gives area to right

$T = \frac{\bar X - \mu}{S/\sqrt n}$, $v = n-1$ (dof), $\sigma^2 = v/(v-2)$, always >1, $\mu = 0$

good for $n\le 30$, 95% between $t_{0.025}$, $t_{0.975}$. Find for μ to val claim

$P(\bar X - t_{\alpha/2}\frac{s}{\sqrt n} < \mu < \bar X - t_{\alpha/2}\frac{s}{\sqrt n}) = 1-\alpha$

σ unknown.

$h(t) = \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{\pi v}}\left(1+\frac{t^2}{v}\right)^{-(v+1)/2}$

Normal Q-Q: y_i vs $q_{0,1}(f_i)$, $f_i = \frac{i-3/8}{n+1/4}$

Just quantile

% obs ≥ data value, empirical.

$f_i = \frac{i-3/8}{n+1/4}$

q = normal dist. $q_{\mu}(f) = 4.91[f^{0.14} - (1-f)^{0.14}]$

$q_{\mu,\sigma} = \mu + \sigma\{4.91[f^{0.14} - (1-f)^{0.14}]\}$

Confidence Intervals

$100(1-\alpha)$: CI. $1-\alpha$: conf. coef. $\alpha = 0.05 \Rightarrow$ 95% CI.

θ = true

$\hat\theta$ = observed value $P(\hat\Theta_L < \theta < \hat\Theta_U) = 1-\alpha$

$\hat\Theta$ = statistic ★ $z_{\alpha/2}$ is t w/ α → right (area), use table

for sample mean

$P(\bar x_L < \bar x < \bar x_U) =$

$P(-z_{\alpha/2} < z < z_\alpha) = P(-z_{\alpha/2} < \frac{\bar x - \mu}{\sigma/\sqrt n} < z_{\alpha/2}) \to (\bar x - z_{\alpha/2}\frac{\sigma}{\sqrt n} < \mu < \bar x + z_{\alpha/2}\frac{\sigma}{\sqrt n})$

gives area left, so $t_{\alpha=0.1}$ is reverse lookup for t w/ area $0.9 = 1-0.1$

i.e. for μ, $100(1-\alpha)$%

same error $< z_{\alpha/2}\sigma/\sqrt n$ ⇒ solve for n to get n for error.

one-sided $\bar x \pm z_\alpha \frac{\sigma}{\sqrt n}$ ★ standard err → stdev of estimate

or t_α, note it is α, not $\alpha/2$.

prediction interval (single value)

μ, σ^2 known

$\bar x - z_{\alpha/2}\sigma\sqrt{1+1/n} < x_0 < \bar x + z_{\alpha/2}\sigma\sqrt{1+1/n}$

σ^2 unknown

$\bar x - t_{\alpha/2} s\sqrt{1+1/n} < x_0 < \bar x + t_{\alpha/2} s\sqrt{1+1/n}$

tolerance interval $\mu \pm 1.96\sigma$

μ, σ known

otherwise $\bar x \pm ks$ (use table)

Assert w/ $100(1-\gamma)$ confidence that it contains $(1-\alpha)$ of elements.

F-dist $h(f) = \begin{cases} \frac{\Gamma[(v_1+v_2)/2](v_1/v_2)^{v_1/2}}{\Gamma(v_1/2)\Gamma(v_2/2)} \frac{f^{(v_1/2)-1}}{(1+v_1 f/v_2)^{(v_1+v_2)/2}}, & f>0, \\ 0, & ow \end{cases}$

given 2 dists, $n_1, n_2, \sigma_1^2, \sigma_2^2, S_1^2, S_2^2$

$F = (S_1^2/\sigma_1^2)/(S_2^2/\sigma_2^2)$, F has F-dist $v_1 = n_1 - 1$, $v_2 = n_2 - 1$

$\mu = \frac{v_2}{v_2 - 2}$, $v_2 > 2$

$\sigma^2 = (2v_2^2(v_2+v_1-2))/[v_1(v_2-2)^2(v_2-4)]$ $v_2 > 4$.

Unbiased estimator

if $E[\hat\Theta] = \theta$

most efficient → $\hat\Theta$ w/ smallest var

2 sample → difference

$\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ known: $(\bar x_1 - \bar x_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 \ldots +$

$\underbrace{\qquad}_{\text{same but } +}$

$Z = \frac{\bar X_1 - \bar X_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$

unknown variance: $T = \frac{(\bar x_1 - \bar x_2) - (\mu_1 - \mu_2)}{S_p\sqrt{1/n_1 + 1/n_2}}$, $S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$ ← note squared!

unknown but equal var $\sigma_1 = \sigma_2$

$\to (\bar x_1 - \bar x_2) - t_{\alpha/2} S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1 - \mu_2 <$ as before.

different var $\sigma_1 \ne \sigma_2$

$T' = \frac{\bar x_1 - \bar x_2 - (\mu_1 - \mu_2)}{\sqrt{(S_1^2/n_1) + (S_2^2/n_2)}}$ → $v = \frac{(S_1^2/n_1 + S_2^2/n_2)^2}{(S_1^2/n_1)^2/(n_1-1) + [(S_2^2/n_2)^2/(n_2-1)]}$ round down to integer.

then the expr is same as that for known σ_1, σ_2, but w/ $t_{\alpha/2}$, $v = v$, S_1 & S_2

Paired observations

ie before/after, a diet in an individual. not indep (the pairs)

$D_i = X_i - Y_i$, $Var(D_i) = Var(X_i - Y_i) = \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}$ expect >0 since correlated.

then apply CLT & typical t dist

$\mu_D = \mu_X - \mu_Y$, assume normal

ie $T = \frac{\bar D - \mu_D}{S_d/\sqrt n}$ ... same CI...

Proportions

$\hat p = X/n$ (estimate for p in bernoulli dist, success %), X: success, n: # trial

$\mu_{\hat p} = p$, $\sigma^2_{\hat p} = \frac{pq}{n}$

$P(-z_{\alpha/2} < z < z_{\alpha/2}) = 1-\alpha$, $z = \frac{\hat p - p}{\sqrt{pq/n}}$ → not good estimate when $p \approx 0$ or 1

$\hat p - z_{\alpha/2}\sqrt{\frac{\hat p\hat q}{n}} < p < \hat p + z_{\alpha/2}\sqrt{\frac{\hat p\hat q}{n}}$ ← n small: really long and complicated but more accurate.

can be CI% sure error $< z_{\alpha/2}\sqrt{\frac{\hat p\hat q}{n}}$, solve for n given CI

If $\hat p$ est. p, can be at least CI% confident that error will not exceed e if $n = \frac{z_{\alpha/2}^2}{4e^2}$

Examples unbiased

$E[S^2] = \sigma^2$ show it is an unbiased estimator.

$E[S^2] = \frac{1}{n-1}\left(-nE[\bar X^2] + \sum E[X_i^2]\right)$

(★ $S^2 = \frac{1}{n-1}\sum^n(\bar X - X_i)^2 = \sum \bar X^2 - 2\bar X X_i + X_i^2$ ...)

$= \frac{1}{n-1}\left(-n\left[E[\bar X]^2 + var\bar X\right] + \sum^n \sigma^2 + \mu^2\right)$

$= \frac{1}{n-1}\left(-n\left[\mu^2 + Var\left[\frac1n\sum^n X_i\right]\right] + n(\sigma^2 + \mu^2)\right)$

$\frac{1}{n^2}\sum^n Var[X_i]$

$\frac{1}{n^2}\sum\sigma^2$

another unbiased estimator

$\sum\frac{(x_i - \bar x)^2}{n}$, s^2 $\to \frac{n-1}{n}\left(\frac{1}{n-1}\sum\ldots\right)$

$\to \frac{n-1}{n}\sigma^2$ shrunk by $E[s^2]$ biased, underestimate

$\to \frac{1}{n-1}\left(-n\left(\mu^2 + \frac{1}{n^2}(n\sigma)\right) + n(\sigma^2 + \mu^2)\right)$

$= \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2$ ∴ it is an unbiased estimator

MSE

$MSE = E(\hat\Theta - \theta)^2$

$= E[\hat\Theta - E(\hat\Theta) + E(\hat\Theta) - \theta]^2$

$= E[\hat\Theta - E[\hat\Theta]]^2 + E[E\hat\Theta - \theta]^2 + \boxed{2E[\hat\Theta - E\hat\Theta]E[E\hat\Theta - \theta]}$ $= 0$

∴ $MSE = Var[\hat\Theta] + Bias[\hat\Theta]^2$

Max likelihood

$f(x|\mu) = \frac{e^{-\mu}\mu^x}{x!}$, poisson

$L(x_1\ldots|\mu) = \prod^n f(x_i|\mu) = \frac{e^{-n\mu}\mu^{\sum_i^n x_i}}{\prod_{i=1}^n x_i!}$

$\ln L = -n\mu + \sum x_i \ln\mu - \ln\prod x_i!$

$\frac{\delta}{\delta\mu}L = -n + \sum^n \frac{x_i}{\mu} = 0$ ← max, set deriv to 0, same for poisson, have it show

$\hat\mu = \sum_{i=1}^n \frac{x_i}{n} = \bar x$

max likelihood, bernoulli

p: $0 \to 1-p$, $1 \to p$, $\to p^{x_i} q^{1-x_i} \to p^{x_i}(1-p)^{1-x_i}$

∴ $L = \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i} = p^{nx_i}(1-p)^{n(1-x_i)}$.

$\ln L = n x_i \ln p + n(1-\bar x)\ln(1-p)$. $\frac{d}{\delta p}$ →

$\frac{d\ln L}{\delta p} = n\left(\frac{\bar x}{p} - \frac{1-\bar x}{1-p}\right) = 0$ → $\frac{\bar x}{p} - \frac{1-\bar x}{1-p} = 0$ ∴ $\hat p = \bar x$

★ s^2 vs $\hat\sigma_n^2$ ($\frac1n$ vs $\frac{1}{n-1}$) ... → $Var(s^2) = \frac{1}{(n-1)^2} Var\sum^n (x_i - \bar x)^2$

$= \frac{1}{(n-1)^2} Var\,\sigma^2\chi^2_{n-1}$

$\hat\sigma_n^2 \to Var(\hat\sigma^2) = \frac{2\sigma^4(n-1)}{n^2}$

$= \frac{1}{(n-1)^2}\sigma^4[2(n-1)]$

$= \frac{2\sigma^4}{n-1}$

Function of RV

$Y = u(X)$, $X = w(Y)$ → given $f(x)$ pdf, $w = u^{-1}$

$g(y) = f(w(y))$

if continuous, $g(y) = f(w(y))\left|\frac{dw}{dy}\right|$ ← abs

generally, $g(y_1, y_2) = f(w_1(y_1,y_2), w_2(y_1,y_2))|J|$

$J = \det\begin{bmatrix}\frac{\delta x_1}{\delta y_1} & \frac{\delta x_1}{\delta y_2} \\ \frac{\delta x_2}{\delta y_1} & \frac{\delta x_2}{\delta y_2}\end{bmatrix}$ → $ad - bc$

Max Likelihood

$L(x_1, \ldots x_n; \theta) = f(X,\theta) = f(x_1,\theta)f(x_2,\theta)\ldots f(x_n,\theta)$. joint dist

→ take ln → take deriv → solution!

MGF Part 2 $Z = X+Y$, $h(z)$

Discrete $h(z) = \sum_{w=-\infty}^{\infty} f(w)g(z-w)$ cont $h(z) = \int_{-\infty}^{\infty} f(w)g(z-w)\,dw$

MGF of $Z = X+Y$ is $M_X(t)M_Y(t)$

$aX+bY \to M_X(at)M_Y(bt)$

Est Variance

we have S^2 define $W^2 = \frac{(n-1)S^2}{\sigma^2}$, which has a χ^2 dist. $v = n-1$. PDF not symmetric

∴ $1-\alpha = P(\chi^2_{1-\alpha/2} < W^2 < \chi^2_{\alpha/2})$

$= P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}\right)$

→ area to the right.

ratio of 2 variances: use F-dist $F = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$, $\frac{S_1^2}{S_2^2}\frac{1}{f_{\alpha/2}(v_1,v_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} f_{\alpha/2}(v_2, v_1)$

2 samples

$(\hat p_1 - \hat p_2) - z_{\alpha/2}\sqrt{\frac{\hat p_1\hat q_1}{n_1} + \frac{\hat p_2\hat q_2}{n_2}} < p_1 - p_2 < \ldots + \ldots$

$z_{\alpha/2}$: area of $\alpha/2$ to right.

$Z = \frac{(\hat P_1 - \hat P_2) - (p_1 - p_2)}{\sqrt{p_1q_1/n_1 + p_2q_2/n_2}}$