MISC

Log

$\ln(mn) = \ln(m) + \ln(n)$

$\ln\left(\frac{m}{n}\right) = \ln(m) - \ln(n)$

$\ln(m^r) = r\ln(m)$

Exponent

$(ab)^x = a^x b^x$

$(a^x)^y = a^{xy}$

$a^x a^y = a^{x+y}$

Summation

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

Integrals

$\int \frac{1}{x}dx = \ln|x|$

$\int e^{ax}dx = \frac{1}{a}e^{ax}$

$\int a^x dx = \frac{a^x}{\ln(a)}$

$\int \ln(x)dx = x\ln(x) - x$

$\int \cos(x)dx = \sin(x)$

$\int \sin(x)dx = -\cos(x)$

Derivatives

$(e^x)' = e^x$

$(\ln(x))' = \frac{1}{x}$

$\sin(x) = \cos(x)$

$\cos(x) = -\sin(x)$

$(fg)' = fg' + f'g$

$\frac{1}{f} = -\frac{f'}{f^2}$

$(f(g(x)))' = f'(g(x))g'(x)$

$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$

Median

Middle number in sorted. If discrete distro, check up to where we have $p < 0.5$ and then $p > 0.5$, the number we have to add to cross threshold is median (see here)

Mode

Most likely value/appears most often.

COUNTING

n choose k

nb of combinations (any order)

$\binom{n}{k} = \frac{n!}{k!(n-k)!}$

permutations

nb of ways of ordering n elements (order matters)

$n!$

subsets of n elements

2^n

partitions

n objects into r groups

$\frac{n!}{n_1!n_2!\ldots n_r!}$

PROB BASICS

Properties

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

$P(A \cup B) \leq P(A) + P(B)$

Conditional

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

$P(A \cap B) = P(B)P(A \mid B)$

⚠ **Total Prob Theorem** ⚠

$P(B) = P(A_1 \cap B) + \ldots + P(A_n \cap B)$

$\quad = P(A_1)P(B \mid A_1) + \ldots$

Bayes

$P(A \mid B) = \frac{P(A)P(B \mid A)}{P(B)}$

Independence

$P(A \mid B) = P(A)$

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

CONT. DISTROS

$P(a \leq x \leq b) = \int_a^b f_X(x)dx$

Disjoint

$P(1 \leq x \leq 3 \text{ or } 4 \leq x \leq 5)$

$= P(1 \leq x \leq 3) + P(4 \leq x \leq 5)$

Properties

$f(x) \geq 0$ and $\int_{-\infty}^{\infty} f_X(x)dx = 1$

CDF Properties

· $\to_{x\to\infty} 1$ and $\to_{x\to-\infty} 0$

·increasing/monotonic

·right-continuous ___*---

Uniform CONT

$f_X(x) = \frac{1}{b-a}$ $\qquad \hat{b}^{\text{MLE}} = \max(X_i)$

$F_X(x) = \frac{x-a}{b-a}$

$E[X] = \frac{a+b}{2}$

$Var(X) = \frac{(b-a)^2}{12}$

Exponential

time to wait for something

$f_X(x) = \lambda e^{-\lambda x}$

$P(X \geq a) = \int_a^\infty \lambda e^{-\lambda x}dx = e^{-\lambda a}$

$F_X(x) = 1 - e^{-\lambda x}$

$E[X] = \frac{1}{\lambda}$ $\qquad I = \frac{1}{\lambda^2}$

$E\left[X^2\right] = \frac{2}{\lambda^2}$

$Var(X) = \frac{1}{\lambda^2}$

Beta

$f(x) = \frac{1}{k}x^{a-1}(1-x)^{b-1}\mathbf{1}\{x \in [0,1]\}$

$k = \int_0^1 t^{a-1}(1-t)^{b-1}dt$

$E[X] = \frac{a}{a+b}$

DISCRETE DISTROS

Bernouilli

$P(X=1) = p$ $\qquad I = \frac{1}{p(1-p)}$

$E[X] = p$ $\qquad \hat{p}^{\text{MLE}} = \overline{X_n}$

$Var(X) = p(1-p)$

Uniform DISCRETE

$p_X(x) = \frac{1}{b-a+1}$

$E[X] = \frac{a+b}{2}$

$Var(X) = \frac{(b-a+1)^2 - 1}{12}$

Binomial

k successes in n trials

$p_X(k) = \binom{n}{k}p^k(1-p)^{n-k}$

$E[X] = np$

$Var(X) = np(1-p)$

Geometric

number of trials until success

$p_X(k) = (1-p)^{k-1}p$

$E[X] = \frac{1}{p}$

$Var(X) = \frac{1-p}{p^2}$

Poisson

how many occurrences k in τ given rate λ

$P(k,\tau) = \frac{(\lambda\tau)^k e^{-\lambda\tau}}{k!}$ $\qquad I = \frac{1}{\lambda}$

$E[N_\tau] = \lambda\tau$

$Var(N_\tau) = \lambda\tau$

NORMALS

$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

$E[X] = \mu$ $\qquad I(\mu,\sigma^2) = \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 1/(2\sigma^4) \end{pmatrix}$

$Var(X) = \sigma^2$

Linear Functions

$Y = aX + b$ with $X \sim N\left(\mu,\sigma^2\right)$

$Y = N\left(a\mu + b, a^2\sigma^2\right)$ $\qquad \hat{\mu}^{\text{MLE}} = \overline{X_n}$

Indie Sum $\qquad \widehat{\sigma^2}^{\text{MLE}} = S_n$ (sample var)

$Z \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$

Tables

$\Phi(-2) = P(Y \leq -2) = 1 - P(Y \leq 2) = 1 - \Phi(2)$

Standardising

$X \sim N\left(\mu,\sigma^2\right)$ and $Y \sim N(0,1)$

$Y = \frac{X-\mu}{\sigma}$ $\quad X = \mu + \sigma Y$

Moments

1	μ	0
2	$\mu^2 + \sigma^2$	σ^2
3	$\mu^3 + 3\mu\sigma^2$	0
4	$\mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$	$3\sigma^4$
5	$\mu^5 + 10\mu^3\sigma^2 + 15\mu\sigma^4$	0
6	$\mu^6 + 15\mu^4\sigma^2 + 45\mu^2\sigma^4 + 15\sigma^6$	$15\sigma^6$
7	$\mu^7 + 21\mu^5\sigma^2 + 105\mu^3\sigma^4 + 105\mu\sigma^6$	0

CONDITIONAL VARS

same for PDF

$p_{X|Y}(x \mid y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$

Multiplication Rule

$p_{X,Y}(x,y) = p_Y(y)p_{X|Y}(x \mid y)$

$p_{X,Y,Z}(x,y,z) = p_X(x)p_{Y|X}(y \mid x)p_{Z|X,Y}(z \mid x,y)$

$p_{X,Y|Z}(x,y \mid z) = \frac{p_{X,Y,Z}(x,y,z)}{p_Z(z)}$

MULTIPLE VARS

$\sum_x \sum_y p_{X,Y}(x,y) = 1$

$P((X,Y) \in B) = \int\int_{(x,y)\in B} f_{X,Y}(x,y)dxdy$

Marginals / Total Probability

$p_X(x) = \sum_y p_{X,Y}(x,y) = \sum_y p_Y(y)p_{X|Y}(x \mid y)$

$f_X(x) = \int f_Y(y)f_{X|Y}(x \mid y)dy$

⚠ ranges: what values can Y take when X = x?

$\quad = \int f_{X,Y}(x,y)dx$

Expected Value Rule

$E[g(X,Y)] = \sum_x \sum_y g(x,y)p_{X,Y}(x,y)$

$E[g(X,Y)] = \int E[g(x,y) \mid Y = y]f_Y(y)dy$

$E[g(X,Y) \mid Y = y] = \int g(x,y)f_{X|Y}(x \mid y)dy$

CDF

$F_{X,Y}(x,y) = P(X \leq x, Y \leq y)$

$\quad = \int_{-\infty}^{y}\int_{-\infty}^{x} f_{X,Y}(s,t)dsdt$

VARIANCE

$Var(x) = E\Big[(x-\mu)^2\Big]$ and $\sigma = \sqrt{Var(X)}$

Properties

$Var(aX+b) = a^2Var(X)$

$Var(X) = E\Big[X^2\Big] - (E[X])^2$

Dependent Sum

$Var(X_1 + \ldots + X_n) = \sum_{i=1}^{n} Var(X_i) + \sum_{\{i,j\}:i\neq j} Cov(X_i,X_j)$

Independent Sum

$Var(X+Y) = Var(X-Y) = Var(X) + Var(Y)$

Law of Total Var

$Var(X) = E[Var(X \mid Y)] + Var(E[X \mid Y])$

Sample Variance

$S_n = \frac{1}{n}\sum^{n}\left(X_i - \overline{X_n}\right)^2$

$E[S_n] = \frac{n-1}{n}\sigma^2$

Unbiased Sample Variance

$\widetilde{S_n} = \frac{n}{n-1}S_n$

$E\left[\widetilde{S_n}\right] = \sigma^2$

STATISTICAL MODEL

$(E, (P_\theta)_{\theta\in\Theta})$

E: sample space ($X_1\ldots$)

P: family of prob measures on E

Θ: Param set

well specified if $\theta^\star \in \Theta$

⚠ sample space must not depend on parameter

⚠ sample space must be the support for the distribution. i.e. $\left([0,\infty), \left\{N\left(\mu,\sigma^2\right)\right\}\right)$ is not valid because the sample space for a N is all R

EXPECTATIONS

Expected Value

$E[g(x)] = \sum_x g(x)p_X(x)$

$E[g(x)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx$

Linearity of Expectations

$E[aX+b] = aE[X] + b$

$E[X+Y] = E[X] + E[Y]$

Total Expectation Th.

$E[X] = \sum_y p_Y(y)E[X \mid Y = y]$

$E[X] = \int_{-\infty}^{\infty} f_Y(y)E[X \mid Y = y]dy$

$E[X] = \sum_i P(A_i)E[X \mid A_i]$

Cond. Expectation

$E[g(x) \mid Y = y] = \sum_x g(x)p_{X|Y}(x \mid y)$

Iterated Expectation

$E[E[X \mid Y]] = E[X]$ ([ex.](ex.))

INDEPENDENCE

If Indie

$E[XY] = E[X]E[Y]$

$Var(X+Y) = Var(X) + Var(Y)$

$p_{X,Y|Z}(x,y \mid z) = p_{X|Z}(x \mid z)p_{Y|Z}(y \mid z)$

$f_{X,Y}(x,y) = f_X(x)f_Y(y)$

$f_{X|Y}(x \mid y) = f_X(x)$

$Cov(X,Y) = 0$

MIXED RV

$X = Y$ (discrete) w.p p

$\quad Z$ (continuous) w.p (1-p)

$F_X(x) = pF_Y(x) + (1-p)F_Z(x)$

$E[X] = pE[Y] + (1-p)E[Z]$

RANDOM NB OF RANDOM VARIABLES

N: nb of stores visited

X_i: money spent in store i

$Y = \sum^{N} X_i$

$E[Y] = E[N]E[X]$

$Var(Y) = E[N]var(X) + (E[X])^2var(N)$

DERIVED DISTROS

PMF function of discrete RV

$p_Y(y) = P(g(x) = y) = \sum_{x:g(x)=y} p_X(x)$

Linear Functions

$Y = aX + b$

$p_Y(y) = p_X\left(\frac{y-b}{a}\right)$

$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$

g is monotonic

$f_Y(y) = f_X(h(y))\left|\frac{dh}{fy}(y)\right|$

general case

1) find CDF: $F_Y(y) = P(g(x) \le y)$
2) derive CDF for PDF

CONVOLUTIONS

$Z = X + Y$

$p_Z(z) = \sum_x p_X(x) p_Y(z-x)$

$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx$

COVARIANCE

$Cov(X,Y) = E[(X - E[X])(Y - E[Y])]$

Direction

$Cov(X,Y) > 0$ same sign

If Indie

$Cov(X,Y) = 0$

⚠ inverse not usually true but true for Gaussians:

$Cov(X,Y) = 0$ → X,Y ~ N indie

Properties

$Cov(X,X) = Var(X)$

$Cov(X,Y) = E[XY] - E[X]E[Y]$

$Cov(aX+b, Y) = aCov(X,Y)$

$Cov(X, Y+Z) = Cov(X,Y) + Cov(Y,Z)$

CORRELATION COEF.

$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$

FRESH START/MEMORYLESSNESS

Exponential

$f_{X|X>t}(x \mid x > t) = f_X(x)$

Bernouilli/Poisson

$P(A \mid B) = P(A)$

i.e. prob of two arrivals (A) after 1 arrival (B) = prob of 2 arrivals (A)

MISC

$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \exp(\lambda)$

e limits

$\lim_{n\to\infty}\left(1 - \frac{t}{n}\right)^n = e^{-t}$

$\lim_{n\to\infty}\left(1 + \frac{t}{n}\right)^n = e^{t}$

BERNOUILLI PROCESS

requires indie, time homogen.

Properties

$S = X_1 + \dots + X_n$

$P(S = K) = \binom{n}{k} p^k (1-p)^{n-k}$

$E[S] = np$

$Var(S) = np(1-p)$

Time until 1st success

$T_1 = \min\{i : X_i = 1\}$

$P(T_1 = k) = (1-p)^{k-1} p$

Time of kth arrival

$p_{Y_k}(t) = \binom{t-1}{k-1} p^k (1-p)^{t-k}$

$E[Y_k] = \frac{k}{p}$

$Var(Y_k) = \frac{k(1-p)}{p^2}$

Merging

$Z_t = g(X_t, Y_t) \sim Ber(p + q - pq)$

⇒ prob either or both have arrival at time t

Splitting

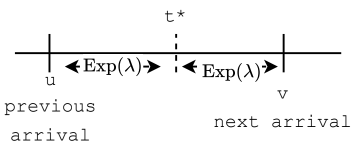
flip a coin with prob q

$A \sim Ber(qp)$

$B \sim Ber((1-q)p)$

⚠ these streams are not indie

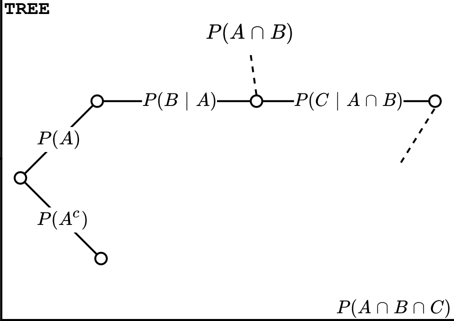
INTER-ARRIVAL TIMES / R.INCIDENCE



we arrive at t*
u,v are each $Exp(\lambda)$ away from t*

⇒ $E[V-U]$ is twice the expectation of $\mathrm{Exp}(\lambda)$

TREE



MIN/MAX

$P(\max > x) = 1 - P(\max < x) = 1 - [P(X_i < x)]^n$

$P(\min > x) = [P(X_i > x)]^n = [1 - P(X_i < x)]^n$

$P(\min < x) = 1 - P(\min > x)$

POISSON PROCESS

indie, time homogen. seq of exp

λ: arrival rate

$P(k,\tau) = \frac{(\lambda\tau)^k e^{-\lambda\tau}}{k!}$ $\qquad I = \frac{1}{\lambda}$

$E[N_\tau] = \lambda\tau$

$Var(N_\tau) = \lambda\tau$

$\lambda = \frac{E[N_\tau]}{\tau}$

Time of kth arrival / Erlang

$f_{Y_k} = \frac{\lambda^k y^{k-1} e^{-\lambda y}}{(k-1)!}$

$= Erlang(k)$

$= Erlang\left(\frac{k}{2}\right) + Erlang\left(\frac{k}{2}\right)$

Sum

⚠ must be indie

M: $Poisson(\mu)$ N: $Poisson(v)$

M+N: $Poisson(\mu + v)$

Merging

A: λ_A B: λ_B

$\lambda = \lambda_A + \lambda_B$

$P\left(k^{th}\text{arrival is A}\right) = \frac{\lambda_A}{\lambda_A + \lambda_B}$

$P(\text{k arrivals are A})$ is $\text{Binomial}\left(\frac{\lambda_A}{\lambda_A + \lambda_B}\right)$

Splitting

flip a coin with prob q

⚠ these streams <u>are</u> indie

A: $\lambda_A = \lambda q$

B: $\lambda_B = \lambda(1-q)$

Multiple Engine Example

3 engines with death rate λ_e
rate until 1st dies is $\lambda = 3\lambda_e$
then rate until 2nd dies $\lambda = 2\lambda_e$

Min

$P(\min\{X,Y,Z\} \ge t)$

$= P(X \ge t, Y \ge t, Z \ge t)$

$= e^{-3\lambda t}$

⇒ have 3 merged Poissons and want to know first arrival

⇒ $\min\{X,Y,Z\}$ is $\mathrm{Exp}(3\lambda)$

$E[\min\{X,Y,Z\}] = \frac{1}{3\lambda}$

Max

$P(\max(T_1,T_2,T_3) \le t)$

$= P(T_1 \le t) P(T_2 \le t) P(T_3 \le t)$

$= \left(1 - e^{-\lambda t}\right)^3$

then derive this to get PDF

LLN

req. iid and $E[|X_i|] < -\infty$

$\overline{X_n} = \frac{1}{n}\sum^n X_i = E[X]$

$E\left[\overline{X_n^2}\right] = Var\left(\overline{X_n}\right) + \left(E\left[\overline{X_n}\right]\right)^2$

⚠ because $\overline{X_n}$ is a RV like any other

CONT MAPPING TH.

$T_n \to T$ then $f(T_n) \to f(T)$

COVARIANCE MATRIX AND MV STUFF

$\Sigma = \begin{pmatrix} Cov(X,X) & Cov(X,Y) \\ Cov(Y,X) & Cov(Y,Y) \end{pmatrix}$

$= E\left[(X - E[X])(Y - E[Y])^T\right]$

$Var(\mathbf{X}) = Cov(\mathbf{X})$

$Cov(\mathbf{AX} + \mathbf{B}) = Cov(\mathbf{AX}) = \mathbf{A}Cov(\mathbf{X})\mathbf{A}^T = \mathbf{A}\Sigma\mathbf{A}^T$

Gaussian vector

defined by $\boldsymbol{\mu}$ and Σ, $x \in R^d$

$f_X(x) = \frac{1}{\sqrt{(2\pi)^d \det\Sigma}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)$

MV CLT

$X_i \sim R^d$ $E[\mathbf{X}_i] = \boldsymbol{\mu}$ $Cov(\mathbf{X}_i) = \Sigma$

MV Delta

$\sqrt{n}(g(T_n) - g(\theta)) \to N\left(0, \nabla g(\theta)^T \Sigma \nabla g(\theta)\right)$

CLT

req. iid, $E[X_i] < \infty$ and $Var(X_i) < \infty$

$\sqrt{n}\frac{\overline{X_n} - \mu}{\sigma} \to N(0,1)$

$\sqrt{n}\left(\overline{X_n} - \mu\right) \to N\left(0, \sigma^2\right)$

alt

$\frac{\left(\sum X_i\right) - n\mu}{\sqrt{n}\sigma} \to N(0,1)$

QUANTILES

$P(X \le q_\alpha) = 1 - \alpha$

$\alpha = .1 \to q_\alpha$ is 90th percentile

$P(|Z| > 1.96) = 0.05$

α	2.5%	5%	7.5%	10%
q_α	1.96	1.65	1.44	1.28

SLUTSKY TH.

$T_n \to T$ and $U_n \to u$

T is r.v. and u is real

$T_n + U_n \to T + u$

$T_n U_n \to Tu$

$\frac{T_n}{U_n} \to \frac{T}{u}$

LIKELIHOODS

Bernouilli $p^{\sum^n X_i}(1-p)^{n - \sum^n X_i}$

Poisson $\frac{\lambda^{\sum X_i}}{x_1! \dots x_n!} \exp(-n\lambda)$

Gaussian $\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2}\sum(x_i - \mu)^2\right)$

Exponential $\lambda^n \exp\left(-\lambda \sum X_i\right)$

Uniform $\frac{1}{b^n}\mathbf{1}\{\max X_i \le b\}$

⚠ a=0 here

IDENTIFIABILITY

θ identifiable iff mapping
$\theta \in \Theta \to P_\theta$ is injective
(injective: $\theta \ne \theta' \Rightarrow P_\theta \ne P_{\theta'}$)

ESTIMATORS

Asym. normal if

$\sqrt{n}\left(\widehat{\theta_n} - \theta\right) \to N\left(0, \sigma^2\right)$

Consistency

$\widehat{\theta_n} \to \theta$ as $n \to \infty$

Bias

$bias\left(\widehat{\theta_n}\right) = E\left[\widehat{\theta_n}\right] - \theta$

Quadratic Risk

$R\left(\widehat{\theta_n}\right) = E\left[\left|\widehat{\theta_n} - \theta\right|^2\right]$

Confidence Interval level $1 - \alpha$

conf.int. can't depend on unknown

$P\left(\overline{X_n} - q_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X_n} + q_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$

UNBIASED ESTIMATOR

we want $bias\left[\widehat{\theta_n}'\right] = 0$

find $\widehat{\theta_n}$ and use linear property of expectations to create a new estimator such that

$E\left[\widehat{\theta_n}'\right] = cE\left[\widehat{\theta_n}\right] = \theta$

1D DELTA METHOD

g: cont. differentiable

$\sqrt{n}(Z_n - \theta) \to N\left(0, \sigma^2\right)$

$\sqrt{n}(g(Z_n) - g(\theta)) \to N\left(0, (g'(\theta))^2 \cdot \sigma^2\right)$

P-VALUE

⚠ is a level α

what is the probability of observing a result more extreme than this one under H_0?

⚠ low p-value is bad → H_0 is unlikely

i.e. $P(|\widehat{a}| \ge \widehat{a}_{\text{obs}})$

KL DIVERGENCE

$KL(P_\theta, P_{\theta'}) = \sum_{x \in E} p_\theta(x) \log\left(\frac{p_\theta(x)}{p_{\theta'}(x)}\right)$

$KL(P_\theta, P_{\theta'}) = \int_E f_\theta(x) \log\left(\frac{f_\theta(x)}{f_{\theta'}(x)}\right) dx$

Properties

<u>not</u> symmetric
not negative definite
triangle ineq

TESTS

⚠ failing to reject H_0 does not mean accepting H_0

Errors

test / reality	H_0	H_1
H_0	✔	type 1 error (reject when shouldn't)
H_1	type 2 error (fail to reject when should)	✔

level α
max type 1 error rate
higher $\alpha \to$ more likely to reject H_0
power β
$\pi_\psi = \inf_{\theta\in\Theta_n}(1-\beta_\psi(\theta))$
example 2 sided
coin $H_0: p = \frac{1}{2}$ and $H_1: p \neq \frac{1}{2}$

$$\psi = 1\left\{\sqrt{n}\frac{\left|\overline{X_n}-\frac{1}{2}\right|}{\sqrt{\frac{1}{2}\left(1-\frac{1}{2}\right)}} > q_{\frac{\alpha}{2}}\right\}$$

stats diff between X and Y?
$\overline{X_n}\sim N(\mu_1,\sigma_1^2)$ and $\overline{Y_n}\sim N(\mu_2,\sigma_2^2)$
$H_0:\mu_1=\mu_2$ and $H_1:\mu_1\neq\mu_2$

$$\sqrt{n}\frac{\overline{X_n}-\overline{Y_n}}{\sqrt{\sigma_1^2\sigma_2^2}} \sim N(0,1)$$

single-sided
⚠ evaluate H_0 at boundary (see part c here)
$H_0 \mu \geq \sigma$ and $H_1 \mu < \sigma$
boundary is $\mu=\sigma$ for $g(\theta)$ or θ

TOTAL VARIATION DISTANCE

max dist between two distros
⚠ E is joint set of values of RVs

$$TV(P_\theta, P_{\theta'}) = \frac{1}{2}\sum_{x\in E}|p_\theta(x)-p_{\theta'}(x)|$$

$$TV(P_\theta, P_{\theta'}) = \frac{1}{2}\int_{-\infty}^{\infty}|f_\theta(x)-f_{\theta'}(x)|dx$$

Properties
symmetric: $TV(P_\theta,P_{\theta'}) = TV(P_{\theta'},P_\theta)$
positive: $0\leq TV\leq 1$
definite: if $TV(P_\theta,P_{\theta'})=0$ then $P_\theta = P_{\theta'}$
triangle ineq:
$TV(P_\theta,P_{\theta'}) \leq TV(P_\theta,P_{\theta''}) + TV(P_{\theta''},P_{\theta'})$
if disjoint: $TV=1$
if same: $TV=0$

MAXIMIZATION

global extremes on range
test critical points and end points
min/max
$h''(x)\leq 0 \to$ concave, maximum
$h''(x)<0 \to$ global max
$h''(x)\geq 0 \to$ convexe, minimum
MV min/max
$X^T Hh(\theta) X \leq 0$ concave, max
+1 top diag: convexe, minimum $\begin{pmatrix}+1 & ?\\ ? & ?\end{pmatrix}$

MLE

minimizes KL divergence
$\hat\theta_n^{MLE} = arg\max_{\theta\in\Theta}\log(L)$
⚠ MLE can be Biased
⚠ function must be cont. diff. to use derivative to find extremums. use a plot and think if not
Consistency and Asym. Norm.
if
- param is identifiable
- support of P_θ does not depend on θ
- $\theta^\star$ is not at boundary
- $I(\theta)$ is invertible
- more stuff

then
consistent: $\hat\theta_n^{MLE}\to\theta^\star$
A.normal: $\sqrt{n}\left(\hat\theta_n^{MLE}-\theta^\star\right)\to N\left(0, I(\theta^\star)^{-1}\right)$
Process to find extremum
- get l_n
- find crits with $l_n'(\theta)=0$
- check if crits are local min/max
- check values at endpoints

METHOD OF MOMENTS

$\widehat{m}_k = \overline{X_n^k} = \frac{1}{n}\sum X_i^k$
LLN $\widehat{m}_k \to m_k(\theta) = E_\theta[X_1^k]$
ASYM NORM $\sqrt{n}\left(\hat\theta-\theta\right)\to N(0,\Gamma(\theta))$

$$\Gamma(\theta) = \left[\frac{\delta M^{-1}}{\delta\theta}\right]^T\Sigma(\theta)\left[\frac{\delta M^{-1}}{\delta\theta}\right]$$

finding $\hat\theta$
write θ as function $E[X]$, $E\left[X^2\right]\dots$
then sub for $\overline{X_n}$, $\overline{X_n^2}$

M-ESTIMATION

Lecture 12, tab 2

FISHER INFORMATION

⚠ use ONE observation
not well defined if support depends on unknown (shifted exp)
⚠ $l''(\theta)$ must exist
$I(\theta) = Var(l'(\theta)) = -E[l''(\theta)]$
⚠ the E[] is of the observation X and not the unknown! $E[\theta X] = \theta E[X]$

χ^2 DISTRO

distro of sum of $Z_i \sim N(0,1)$
$E[V]=d$
$Var(V)=2d$

COCHRAN'S TH.

$\frac{nS_n}{\sigma^2}\sim\chi^2_{n-1}$ or $nS_n \sim \frac{\sigma^2}{n}\chi^2_{n-1}$

t DISTRO

for small nb of Gaussian samples w/
$Z\sim N(0,1)$ and $V\sim\chi^2_d$ and $\text{SampleVar}=\frac{V}{d}$
$\frac{Z}{\sqrt{\frac{V}{d}}}$ ⚠ Z and V must be indie

t TEST

- requires Gaussian samples
- is pivotal (q in tables)
- test is non-asymptotic

one sample two-sided
$H_0:\mu=\mu_0$ vs $H_1:\mu\neq\mu_0$

$$T_n = \frac{\sqrt{n}\overline{X_n}}{\sqrt{\tilde{S_n}}} = \frac{\sqrt{n}\frac{\overline{X_n}-\mu_0}{\sigma}}{\sqrt{\frac{\tilde{S_n}}{\sigma^2}}}\sim t_{n-1}$$

$$\psi_\alpha = \mathbf{1}\left\{|T_n| > q_{\frac{\alpha}{2}}\right\}$$

one sample one-sided
$H_0:\mu\leq\mu_0$ vs $H_1:\mu>\mu_0$

$$T_n = \frac{\sqrt{n}(\overline{X_n}-\mu_0)}{\sqrt{\tilde{S_n}}}\sim t_{n-1}$$

$\psi_\alpha = \mathbf{1}\{T_n > q_\alpha\}$
two sample
$\overline{X_n}\sim N\left(\Delta_d, \frac{\sigma_d^2}{n}\right)$ and $\overline{Y_m}\sim N\left(\Delta_c, \frac{\sigma_c^2}{m}\right)$

$$\frac{\overline{X_n}-\overline{Y_m}-(\Delta_d-\Delta_c)}{\sqrt{\frac{\hat\sigma_d^2}{n}+\frac{\hat\sigma_c^2}{m}}}\sim t_N$$

where N according to Welch-Satter:

$$N = \frac{\left(\frac{\hat\sigma_d^2}{n}+\frac{\hat\sigma_c^2}{m}\right)^2}{\frac{\hat\sigma_d^4}{n^2(n-1)}+\frac{\hat\sigma_c^4}{m^2(m-1)}} \geq \min(n,m)$$

WALD'S TEST

- test is asymptotic
- not invariant to change in rep of H_0
- only req est of unrestricted model, lower computation

⚠ MLE conditions must be satisfied
$H_0:\theta=\theta_0$ vs $H_1:\theta\neq\theta_0$ for $\theta\in\mathbb{R}^d$

$$n\left(\hat\theta^{\text{MLE}}-\theta_0\right)^T I\left(\hat\theta^{\text{MLE}}\right)\left(\hat\theta^{\text{MLE}}-\theta_0\right)\to\chi^2_d$$

equivalently

$$T_n = \left\|\sqrt{n}I(\theta_0)^{\frac{1}{2}}\left(\hat\theta^{\text{MLE}}-\theta_0\right)\right\|^2\to\chi^2_d$$

which gives test $\psi_\alpha = \mathbf{1}\{T_n > q_\alpha\}$
where q_α is the $(1-\alpha)$-quantile of χ^2_d

LIKELIHOOD RATIO TEST

- how diff is likelihood from null

X_i iid $\Theta\in\mathbb{R}^d$
$H_0:(\theta_{r+1},\dots,\theta_d) = \left(\theta_{r+1}^{(0)},\dots,\theta_d^{(0)}\right)$

$$T_n = 2\left(l_n\left(\hat\theta_n^{\text{MLE}}\right)-l_n\left(\theta_n^{(0)}\right)\right)$$

⚠ same n for both likelihoods
Wilk Th.
assuming H_0 is true and MLE conditions. is asymptotic
$T_n\to\chi^2_{d-r}$
$\psi_\alpha = \mathbf{1}\{T_n > q_\alpha\}$

CATEGORICAL LIKELIHOOD

i.e. are Zodiac signs uniformly distributed? $p_0 = \left(\frac{1}{12},\frac{1}{12}\dots\right)$
$L_n = p_1^{N}{}_{-}1\dots p_k^{N}{}_{-}1$
$N_j = \#\{X_i = a_j\}$
$\overset{\rightarrow\text{MLE}}{\hat p}: \hat p_j = \frac{N_j}{n}$ prob of obs. outcome j
$p_j = P(X=a_j) = \prod p_i^{\mathbf{1}(a_i=a_j)}$

χ^2 TEST

$H_0:\vec{p}=\vec{p}^0$ vs. $H1:\vec{p}\neq\vec{p}^0$

$$T_n = n\sum^k\left[\frac{\left(\hat p_j-p_j^0\right)^2}{p_j^0}\right]\to\chi^2_{k-1}$$

where k is nb of categories

χ^2 TEST FOR FAMILY OF DIST

$H_0: p\in\{\text{Bin}(k,\theta)\}_{\theta\in\Theta}$ vs $H_1: p\notin\{\text{Bin}(k,\theta)\}_{\theta\in\Theta}$

$$T_n = n\sum_{j=0}^k\frac{\left(\frac{N_j}{n}-f_{\hat\theta}(j)\right)^2}{f_{\hat\theta}(j)}\to\chi^2_{(k+1)-d-1}$$

⚠ $k-d-1$ if we start at j=1
$\Theta\in\mathbb{R}^d$
f_θ is PMF of $\text{Bin}(k,\theta)$
$\hat\theta$ is MLE here

EMPIRICAL CDF

$F_n(t) = \frac{1}{n}\sum\mathbf{1}\{X_i\leq t\}$
it is discontinuous
$\sqrt{n}(F_n(t)-F(t))\to N(0, F(t)(1-F(t)))$

DONSKER'S TH.

if F cont:
$\sqrt{n}\max_{t\in\mathbb{R}}|F_n(t)-F(t)|\to\max_{0\leq t\leq 1}|B(t)|$
where B is Brownian bridge

KS TEST (example)

X_i: real RV with unk CDF
$H_0:F=F^0$ vs $H1:F\neq F^0$
$\delta_\alpha^{\text{KS}} = \mathbf{1}\{T_n > q_\alpha\}$
$= \mathbf{1}\left\{\max_{t\in\mathbb{R}}\sqrt{n}|F_n(t)-F(t)| > q_\alpha\right\}$
p-value $P(Z > T_n \mid T_n)$
computation

$$\frac{T_n}{\sqrt{n}} = \max_{1\leq i\leq n}\left[\max\left(\left|F^0(X_i)-F_n(X_i)\right|, \left|F^0(X_i)-\frac{i}{n}\right|\right)\right]$$

$$= \max_{1\leq i\leq n}\left[\max\left(\left|F^0(X_i)-F_n(X_i)\right|, \left|F^0(X_i)-\frac{i-1}{n}\right|\right)\right]$$

CDF OF SAMPLE IS UNIFORM

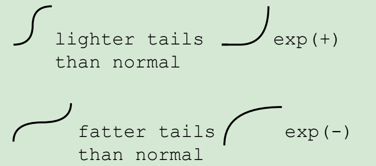
$Y = F_X(x)$
$F_Y\sim Un$ if $(0,1)$

KL TEST (example)

is my data Gaussian?
more likely to reject than KS test
$\max_{t\in\mathbb{R}}\left|F_n(t)-\Phi_{\hat\mu,\hat\sigma^2}(t)\right|$

QQ PLOT (example 1, 2)

$F_n^{-1}\left(\frac{i}{n}\right) = X_i$ (F_n is sample CDF)
points are $\left(F^{-1}\left(\frac{1}{n}\right), x_1\right), \left(F^{-1}\left(\frac{2}{n}\right), x_2\right)\dots$
to find inverse F^{-1}: "what input value gives output value t. we are looking for input value to F that gives $\frac{1}{n}$"



MARKOV INEQUALITY

$X\geq 0$ and $a>0$
$P(X\geq a)\leq\frac{E[X]}{a}$

CHEBYSHEV INEQUALITY (link)

probability of estimate of mean deviating from true mean by more than C
$P(|X-\mu|\geq c)\leq\frac{\sigma^2}{c^2}$

CONVERGENCE IN PROBABILITY

a seq. converges to a in probability if:
$\lim_{n\to\infty}P(|X_n-\mu|\geq\varepsilon) = 0$
another way to show convergence in prob is to determine expectation and variance. if $Var\to 0$ then convergence
properties
- if g is continuous then $g(X_n)\to g(a)$
- $X_n + Y_n\to a+b$
- but $E[X_n]$ doesn't need to converge to a

ESTIMATE BINOMIAL WITH NORMAL

PMF of # success in n trials w/p p approximates $N(np, np(1-p))$
with
$'P(X=19)' = P(18.5\leq X\leq 19.5)$

MOIVRE LAPLACE CORRECTION

when estimating an integer R.V. with the CLT, can do the "1/2 correction":
$P(S_n\leq 21)\to P(S_n\leq 21.5)$

is estimator consistent?
check lim as $n\to\infty$ against estimator
is estimator asym. normal?
start with CLT definition, then put in the estimator. also get aVar like this.
see examples.

WOLFRAM

Probability x>4.03, Chi Squared Distribution degrees of freedom 1
Chi Squared Distribution degrees of freedom 1
CDF[NormalDistribution[2, 1], 0.65] ⚠ CDF uses **STANDARD DEVIATION**

BAYESIAN STATS

$$\pi(\theta \mid X_1...X_n) = \frac{\pi(\theta)L_n(X_1...X_n \mid \theta)}{\int_\Theta \pi(\theta)L_n(X_1...X_n \mid \theta)} \propto \pi(\theta)L_n(X_1...X_n \mid \theta)$$

conjugate prior if post. distro. same as prior distro.

improper prior i.e. $\pi(\theta) = 1$, not a valid distro

Jeffrey's prior
non-informative prior, not always improper. reflects no prior belief, only stats model

$\pi_J(\theta) \propto \sqrt{\det I(\theta)}$

reparam. invariance
we have Jeff prior for θ, want $\eta = \Phi(\theta)$
· replace θ with $\Phi^{-1}(\eta)$
· multiply by $\frac{d\theta}{d\eta} = \frac{1}{\Phi'(\theta)}$

confidence region
$P(\theta \in \mathbb{R} \mid X_1...X_n) = 1 - \alpha$

BAYESIAN STATS - NORMALS

$f_X(x) = c\exp\Big(-\big(\alpha x^2 + \beta x + \gamma\big)\Big)$

$\mu = -\frac{\beta}{2\alpha}$ and $\sigma^2 = \frac{1}{2\alpha}$

the peak is min. of exponent:
· derive exponent and set to 0

$\widehat{\Theta}_{\text{MAP}} = \widehat{\Theta}_{\text{LMS}} = E[\Theta \mid X = x]$
(in general this is true if posterior is unimodal and symmetric)

MAP

$$\hat{\theta}^{\text{MAP}} = arg\max_\theta \pi(\theta \mid X_1...X_n) = arg\max_\theta L_n(X_1...X_n \mid \theta)\pi(\theta)$$

⚠ look at posterior PDF/PMF and ask "which <u>actual</u> possible values of θ make this result most likely, i.e. the mode

i.e. is $\left|\theta_1 - \hat{\theta}^{\text{Bayes}}\right| > \left|\theta_2 - \hat{\theta}^{\text{Bayes}}\right|$

⚠ if discrete, MAP is in set of possible values

BAYES ESTIMATOR

mean of posterior
also known as **LMS** "conditional expectation" $E[\Theta \mid X = x]$

⚠ MUST USE ACTUAL POSTERIOR, not the prop. one if we calculate it like below, else we may also use mean of the distribution if i.e. Beta without having to calculate denominator

$$\hat{\theta}^\pi = \int_\Theta \theta\pi(\theta \mid X_1...X_n)d\theta$$

$\text{aVar} = I^{-1}(\theta)$ of distro sampled

properties of LMS estimation error
let $\widehat{\Theta} = E[\Theta \mid X]$ and error $\widetilde{\Theta} = \widehat{\Theta} - \theta^\star$
· $E\big[\widetilde{\Theta} \mid X = x\big] = 0$
· $cov\big(\widetilde{\Theta}, \widehat{\Theta}\big) = 0$
· $Var(\Theta) = Var\big(\widehat{\Theta}\big) + Var\big(\widetilde{\Theta}\big)$

LLMS

unknown Θ, observation X
$\widehat{\Theta} = aX + b$
minimises $E\Big[(\Theta - aX - b)^2\Big]$

$a = \frac{Cov(\Theta, X)}{Var(X)}$
$b = E[\Theta] - aE[X]$

⚠ if all vars normals then LMS=LLMS