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

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

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

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

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

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

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

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

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

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

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

PROB BASICS

Properties

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

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

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

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

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

Bayes

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

Independence

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

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

DISCRETE DISTROS

Bernouilli

$P(X=1) = p$

$E[X] = p$

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

Uniform

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

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

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

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

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

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

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

CONT. DISTROS

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

Disjoint

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

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

Properties

$f(x) \ge 0$ 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

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

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

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

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

Exponential

time to wait for something

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

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

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

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

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

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

NORMALS

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

$E[X] = \mu$

$Var(X) = \sigma^2$

Linear Functions

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

$Y = N(a\mu + b, a^2\sigma^2)$

Indie Sum

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

Tables

$\Phi(-2) = P(Y \le -2)$

$\quad = 1 - P(Y \le 2) = 1 - \Phi(2)$

Standardising

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

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

$X = \mu + \sigma Y$

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

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

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

⚠ 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

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

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

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

$E[E[X \mid Y]] = E[X]$

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)

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

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

VARIANCE

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

Properties

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

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

Dependent Sum

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

Independent Sum

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

Law of Total Var

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

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: $F_Y(y) = P(g(x) \leq y)$

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)

BERNOUILLI PROCESS

requires indie, time homogen.

Properties

$S = X_1 + ... + X_n$

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

$E[S] = np$

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

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

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

⇒ prob either or both have arrival at time t

Splitting

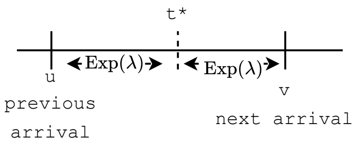
flip a coin with prob q

A~$Ber(qp)$

B~$Ber((1-q)p)$

⚠ these streams are not indie

INTER-ARRIVAL TIMES / R.INCIDENCE

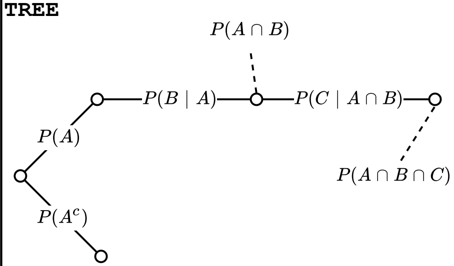


we arrive at t*

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

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

TREE



POISSON PROCESS

indie, time homogen. seq of exp

λ: arrival rate

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

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

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

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

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

$\lambda = \lambda_A + \lambda_B$

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

$P(\text{k arrivals are A})$ 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

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

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

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

⇒ 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

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

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

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

then derive this to get PDF

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

MV Delta

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

MISC

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

e limits

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

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

MIN/MAX

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

LLN

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

$$\overline{X_n}=\frac{1}{n}\sum^{n}X_i$$

CLT

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

$$\sqrt{n}\frac{\overline{X_n}-\mu}{\sigma}\to N(0,1)$$
$$\sqrt{n}(\overline{X_n}-\mu)\to N(0,\sigma^2)$$

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

$T_n+U_n\to T_u$

$T_nU_n\to T_u$

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

Continuous Mapping Th.

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

Statistical Model

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

E: sample space Θ: Param set

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

IDENTIFIABILITY

θ identifiable iff mapping $\theta\in\Theta\to P_\theta$ is injective

(injective: $\theta\neq\theta'\Rightarrow P_\theta\neq P_{\theta'}$)

ESTIMATORS

Asym. normal if

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

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 $E\left[\widehat{\theta_n}\right]=0$

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

$$E\left[\widehat{\theta_n}'\right]=\frac{1}{c}E\left[\widehat{\theta_n}\right]=0$$

1D DELTA METHOD

g: cont. differentiable

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

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. $P(|\hat{a}|\ge\hat{a}_{\text{obs}})$

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

not 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

$\pi_\psi=\inf_{\theta\in\Theta_1}(1-\beta_\psi(\theta))$

example 2 sided

coin $H_0:p=\frac{1}{2}$ and $H_1:p\neq\frac{1}{2}$

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

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 $H_1:\mu_1\neq\mu_2$

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

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: $0\le TV\le 1$

definite: if $TV(P_\theta,P_{\theta'})=0$ then $P_\theta=P_{\theta'}$

triangle ineq:

$TV(P_\theta,P_{\theta'})\le TV(P_\theta,P_{\theta''})+TV(P_{\theta''},P_{\theta'})$

if disjoint: $TV=1$

if same: $TV=0$

M-Estimation

Lecture 12, tab 2

LIKELIHOODS

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

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

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

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

Uniform $\dfrac{1}{b^n}\mathbb{1}\{\max X_i\le b\}$

MAXIMIZATION

global extremes on range

test critical points and end points

min/max

$h''(x)\le 0\to$ concave, maximum

$h''(x)<0\to$ global max

$h''(x)\ge 0\to$ convexe, minimum

MV min/max

$X^THh(\theta)X\le 0$ concave, max

+1 top diag: convexe, minimum

$$\begin{pmatrix}+1 & ?\\ ? & ?\end{pmatrix}$$

MLE

minimizes KL divergence

$$\hat{\theta}_n^{MLE}=arg\max_{\theta\in\Theta}\log(L)$$

⚠ function must be cont. diff. to use derivative to find extremums. use a plot and think if not

Gaussian check Wikipedia

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

$$\hat{\theta}_n^{MLE}\to\theta^\star$$
$$\sqrt{n}\left(\hat{\theta}_n^{MLE}-\theta^\star\right)\to N\left(0,I(\theta^\star)^{-1}\right)$$

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

Fisher Information

⚠ use ONE observation

not well defined if support depends on unknown (shifted exp)

⚠ $l''(\theta)$ must exist

$$I(\theta)=Var(l'(\theta))=-E[l''(\theta)]$$

Method of Moments

$$\widehat{m}_k=\overline{X_n^k}=\frac{1}{n}\sum X_i^k$$

LLN $\widehat{m}_k\to m_k(\theta)=E_\theta[X_1^k]$

Delta $\sqrt{n}\left(\hat{\theta}-\theta\right)\to N(0,\Gamma(\theta))$

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

finding $\hat{\theta}$

write θ as function $E[X]$, $E[X^2]$... then sub for $\overline{X_n}$, $\overline{X_n^2}$