SUMMARY TMA 4295

CHAPTER 2

Transformations

$$Y = g(X)$$

pmf: $f_Y(y) = P(Y=y) = \sum_{x \in g^{-1}(y)} P(X=x) = \sum_{x \in g^{-1}(y)} f_X(x) \qquad y \in \mathcal{Y}$

pdf. $f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| \qquad y \in \mathcal{Y}$

Moment generating functions

$M_X(t) = E[e^{tX}]$, $\qquad |t| < h$

$E[X^n] = M_X^{(n)}(0)$

$M_{aX+b}(t) = e^{bt} M_X(at)$

Th. $M_X(t) = M_Y(t) \Rightarrow F_X(x) = F_Y(x)$

$\lim_{n \to \infty} M_{X_n}(t) = M_X(t) \Rightarrow \lim_{n \to \infty} F_{X_n}(t) = F_X(t)$

$X \sim$ lognormal $\Rightarrow Y = \log X \sim N(\mu, \sigma^2)$

$X = e^Y$ and $E[X^n] = E[e^{nY}] = M_Y(n)$

2.4.1. Leibnitz Rule

$$\frac{d}{dt} \int_{a(t)}^{b(t)} f(x,t)\,dx = f(b(t),t)\frac{d}{dt}b(t) - f(a(t),t)\frac{d}{dt}a(t) + \int_{a(t)}^{b(t)} \frac{d}{dt} f(x,t)\,dx$$

Chapter 3. DISTRIBUTIONS

independent trials, Reg $A \mid A^c$, $P(A) = k$ } Bernoulli $B(1, p)$; Binomial $B(n, p)$; Geometric $NB(1, p)$; Negative Binomial $NB(k, p)$

Independence between disjoint intervals, Reg $A \mid A^c$, $P(A \text{ occurs once in } \Delta t) = \lambda \Delta t + o(\Delta t)$, $P(A \text{ occurs twice or more i } \Delta t) = o(\Delta t)$ } Poisson (λ), $\exp(\frac{1}{\lambda})$, $\Gamma(\alpha, \frac{1}{\lambda}) = \Gamma(\alpha, \beta)$

X gamma (α, β)

$P(X \leq x) = P(Y \geq \alpha)$, $Y \sim$ Poisson $\left(\lambda = \frac{x}{\beta}\right)$

$\Gamma(\alpha, \beta)$ } $\alpha = 1 \Rightarrow$ exponential (β); ~~[illegible]~~ $\beta = 2 \Rightarrow \chi^2(2\alpha)$

Beta distribution, $B(\alpha, \beta)$

$\alpha = \beta = 1 \Rightarrow$ uniform on $(0, 1)$

Exponential class

$$f(x \mid \underline{\theta}) = h(x)\, c(\underline{\theta})\, e^{\sum_i^k w_i(\theta) t_i(x)}$$

Includes { normal, gamma, beta; binomial, Poisson, negative binomial.

Chebychevs Inequality.

$$P(g(X) \geq r) \leq \frac{E[g(X)]}{r}.$$

In particular $g(x) = \left(\frac{X-\mu}{\sigma}\right)^2$

$$\Rightarrow \quad P\left(\left(\frac{X-\mu}{\sigma}\right)^2 \geq t^2\right) \leq \frac{1}{t^2}$$

$$P(X \geq a) \leq e^{-at} M_X(t)$$ if $M_X(t)$ exists. ~~over~~ $t > 0$.

Hölders Inequality

$$E|XY| \leq (E|X|^p)^{1/p} (E|Y|^q)^{1/q} \qquad \frac{1}{p} + \frac{1}{q} = 1$$

Cauchy - Schwartz

$$E|XY| \leq \left(E|X|^2\right)^{\frac{1}{2}} \left(E|Y|^2\right)^{\frac{1}{2}}$$

JENSENS Inequality

$g(x)$ convex

$$E[g(X)] \geq g(E[X])$$

Chapter 4

Bivariate Transformation

X, Y discrete. $U = g_1(X,Y)$, $V = g_2(X,Y)$

$$f_{U,V}(u,v) = \sum_{\{(x,y)\in A \,:\, g_1(x,y)=u,\ g_2(x,y)=v\}} f_{X,Y}(x,y).$$

X, Y continuous

$X = h_1(U,V)$, $Y = h_2(U,V)$

$$f_{U,V}(u,v) = f_{X,Y}(h_1(u,v), h_2(u,v))\,|J|$$

Hierachical Models, Mixture distribution.

$$\left.\begin{array}{l} X|Y \sim B(Y,p) \\ Y \sim \text{Poisson}(\lambda) \end{array}\right\} \quad X \sim \text{Poissom}(\lambda p)$$

Double expectation

$$E[X] = E[E(X|Y)]$$

Conditional variance

$$Var[X] = E[Var(X|Y)] + Var[E(X|Y)].$$

CONVERGENCE

Convergence in Probability

$$\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0 \qquad X_n \xrightarrow{P} X$$

$$X_n \xrightarrow{P} X \quad \xRightarrow{h \text{ continuous}} \quad h(X_n) \xrightarrow{P} h(X)$$

Convergence in Distribution

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \qquad X_n \xrightarrow{D} X$$

$$X_n \xrightarrow{P} X \implies X_n \xrightarrow{D} X$$

$$X_n \xrightarrow{D} \mu \implies X_n \xrightarrow{P} \mu$$

$$\left.\begin{array}{l} X_n \xrightarrow{D} X \\ Y_n \xrightarrow{P} a \end{array}\right\} \implies \left\{\begin{array}{l} X_n Y_n \xrightarrow{D} aX \\ X_n + Y_n \xrightarrow{D} X + a \end{array}\right.$$

$X_1, X_2, \ldots$ i.i.d. $E[X_i] = \mu$, $Var[X_i] = \sigma^2 < \infty$

WLN

$$\implies \bar{X}_n = \frac{\sum_i^n X_i}{n} \xrightarrow{P} \mu$$

$(M_{X_i}(t)$ exists for $|t| < h$, $Var[X_i] = \sigma^2 > 0)$ $(0 < Var(X_i) < \infty)$

$$\implies \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{D} Z \sim N(0,1) \qquad \text{CLT}$$

Delta Method

$$\sqrt{n}(Y_n - \theta) \xrightarrow{D} N(0, \sigma^2) \Rightarrow \sqrt{n}[g(Y_n) - g(\theta)] \xrightarrow{D} N(0, g'(\theta)^2 \sigma^2)$$

$$\sqrt{n}(Y_n - \theta) \xrightarrow{D} N(0, \sigma^2) \Longleftrightarrow \frac{Y_n - \theta}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{D} N(0, 1)$$

and $$Y_n - \theta = \underbrace{\frac{\sigma}{\sqrt{n}}}_{\to 0} \cdot \underbrace{\frac{\sqrt{n}}{\sigma}(Y_n - \theta)}_{\xrightarrow{D} N(0,1)} \xrightarrow{D} 0$$

$$\Rightarrow Y_n - \theta \xrightarrow{?} 0$$

Second order Delta Method

$$\sqrt{n}(Y_n - \theta) \xrightarrow{D} N(0, \sigma^2), \quad g'(\theta) = 0, \quad g''(\theta) \text{ exists}$$

$$\Rightarrow n[g(Y_n) - g(\theta)] \to \frac{\sigma^2}{2} \frac{g''(\theta)}{2} \chi^2(1)$$

Chapter 5 Random Sample

Random sample: $X_1,\ldots,X_n$ are iid.

Statistic: $T(X_1,\ldots,X_n)$

Some properties of Statistics

$X_1,\ldots,X_n$ are $N(\mu,\sigma^2)$

$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ and $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ are independent

$\bar{X}\sim N\left(\mu,\frac{\sigma^2}{n}\right)$, $\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$

T-statistic: $\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$, In general $T_p=\frac{N(0,1)}{\sqrt{\frac{\chi^2(p)}{p}}}$

$F_{p,q}$ statistic $=\frac{\frac{\chi^2(p)}{p}}{\frac{\chi^2(q)}{q}}$

$V\sim\chi^2(q)\Leftrightarrow V\sim\Gamma\left(\frac{q}{2},2\right)$

$$E\left(V^{-k}\right)=\frac{1}{\Gamma(q/2)2^{\frac{q}{2}}}\int_0^\infty v^{\frac{q}{2}-k-1}e^{-\frac{v}{2}}dv=\frac{\Gamma(\frac{q}{2}-k)}{\Gamma(\frac{q}{2})2^k}.$$

$$X_i\sim\chi^2(p_i)\Rightarrow\sum_{i=1}^{n}X_i\sim\chi^2\left(\sum_{i=1}^{n}p_i\right)$$

Chapter 6 Sufficiency and completeness

Sufficient statistics

A statistic $T(\boldsymbol{X})$ is a sufficient statistic for $\boldsymbol{\theta}$ if the conditional distribution of the sample $\boldsymbol{X}$ given the value of $T(\boldsymbol{X})$ does not depend on $\boldsymbol{\theta}$.

Theorem 6.2.2

If $p(\boldsymbol{x}|\boldsymbol{\theta})$ is the pdf/pmf of $\boldsymbol{X}$ and $q(t|\boldsymbol{\theta})$ is the pdf/pmf of $T(\boldsymbol{X})$, then $T(\boldsymbol{X})$ is a sufficient statistics for $\boldsymbol{\theta}$ if, for every $\boldsymbol{x}$ in the sample space the ratio $\dfrac{p(\boldsymbol{x}|\boldsymbol{\theta})}{q(T(\boldsymbol{x})|\boldsymbol{\theta})}$ is a constant as a function of $\boldsymbol{\theta}$.

Factorization Theorem

Let $f(\boldsymbol{x}|\theta)$ be the joint pdf/pmf for a sample $\boldsymbol{X}$. $T(\boldsymbol{X})$ is a sufficient statistics for θ if and only if for all $\boldsymbol{x}$ and all θ.

$$f(\boldsymbol{x}|\theta) = g(T(\boldsymbol{x}|\theta))h(\boldsymbol{x})$$

Minimal sufficient.

Definition 6.2.11. A sufficient statistics $T(\boldsymbol{X})$ is called a minimal sufficient statistics if for any other sufficient statistics $T'(\boldsymbol{X})$, $T(\boldsymbol{X})$ is a function of $T(\boldsymbol{X})$.

Theorem 6.2.3

Let $f(\boldsymbol{x}|\theta)$ be the joint pdf/pmf for a sample $\boldsymbol{X}$. Suppose there exists a $T(\boldsymbol{X})$ such that for every $\boldsymbol{x}$ and every $\boldsymbol{y}$, $f(\boldsymbol{x}|\theta)/f(\boldsymbol{y}|\theta)$ is a constant as a function of $\theta \Leftrightarrow T(\boldsymbol{x})=T(\boldsymbol{y})$. Then $T(\boldsymbol{X})$ is a minimal sufficient statistics for θ.

Chapter 7. Estimation

Maximum likelihood estimation

Likelihood:

$$L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta) \overset{iid}{=} \prod_{i=1}^{n} f(x_i|\theta)$$

$\hat{\theta}_e(\boldsymbol{x})$ maximizes $L(\theta|\boldsymbol{x})$

$\hat{\theta}(\boldsymbol{X})$ is the MLE

Candidates: For $\hat{\theta}$

$$\frac{\partial}{\partial\theta_i} L(\boldsymbol{\theta}) = 0 \Leftrightarrow \frac{\partial}{\partial\theta_i} \log L(\boldsymbol{\theta}) = 0$$

Invariance principle:

If $\hat{\theta}$ is the MLE of θ, $\tau(\hat{\theta})$ is the MLE of $\tau(\theta)$.

Bayes estimation:

Prior: $\pi(\theta)$ Posterior: $\pi(\theta|\boldsymbol{x})$

$$\pi(\theta|\boldsymbol{x}) = \frac{f(\boldsymbol{x},\theta)}{f(\boldsymbol{x})} = \frac{f(\boldsymbol{x}|\theta)\pi(\theta)}{\int f(\boldsymbol{x},\theta)\,d\theta}$$

$$\hat{\theta}_B = E(\theta|\boldsymbol{x})$$

Evaluation of Estimators

The mean square error

$$MSE = E\left[(W-\theta)^2\right] = Var[W] + \left(E[W]-\theta\right)^2$$

Score function

$$S(\boldsymbol{X}|\theta) = \frac{\partial}{\partial\theta}\log f(\boldsymbol{X}|\theta)$$

$$E\left[S(\boldsymbol{X}|\theta)\right] = 0$$

$$Var\left[S(\boldsymbol{X}|\theta)\right] = I_X(\theta) = -E\left[\frac{\partial}{\partial\theta}S(\boldsymbol{X}|\theta)\right] = -E\left[\frac{\partial^2}{\partial\theta^2}\log f(\boldsymbol{X}|\theta)\right]$$

Let $\tau(\theta) = E\left[W(\boldsymbol{X})\right]$

Cramer-Rao

$$Var\left[W(\boldsymbol{X})\right] \geq \frac{\left(\frac{\partial}{\partial\theta}\tau(\theta)\right)^2}{I_X(\theta)}$$

Cramer-Rao iid

$$Var\left[W(\boldsymbol{X})\right] \geq \frac{\left(\frac{\partial}{\partial\theta}\tau(\theta)\right)^2}{nI_X(\theta)}$$

Equality

If and only if $S(\boldsymbol{X}|\theta)=a(\theta)[W(\boldsymbol{X})-b(\theta)]$

The multiparameter case

$$Var[W(\boldsymbol{X})]\geq(\nabla\tau(\boldsymbol{\theta}))^{T}(I(\boldsymbol{\theta}))^{-1}(\nabla\tau(\boldsymbol{\theta})).$$

Completeness

Let $f(t|\theta)$ be a family of pdfs/pmfs for a statistic $T(\boldsymbol{X})$. The family is complete if

$E_{\theta}[g(T)]=0 \Rightarrow P_{\theta}(g(T)=0)=1$, for all θ.

Completeness and the exponential class

Let $X_1,\ldots,X_n$ be iid. from an exponential family i.e.

$$f(x|\boldsymbol{\theta})=h(x)c(\boldsymbol{\theta})e^{\sum_{i=1}^{k}w(\theta_i)t_i(x)}$$

Then $T(\boldsymbol{X})=\left(\sum_{i=1}^{n}t_1(X_i),\sum_{i=1}^{n}t_2(X_i),\ldots,\sum_{i=1}^{n}t_k(X_i),\right)$ is complete as long as the parameter space contains an open set in R^{k}.

Minimal sufficient if $w_i(\theta), i=1,2,\ldots n$ are not linearly dependent

Complete if no functional relationship exists between $w_i(\theta), i=1,2,\ldots n$

Then also the distribution of

$T(\boldsymbol{X})=\left(\sum_{i=1}^{n}t_1(X_i),\sum_{i=1}^{n}t_2(X_i),\ldots,\sum_{i=1}^{n}t_k(X_i),\right)$ is within the exponential family.

Sufficiency and Unbiasedness

W unbiased estimator of $\tau(\theta)$.

T a sufficient estimator, then $E[W|T] = \tau(\theta)$ and
$Var[W|T] \leq Var[W], \forall\theta$

T complete $\Rightarrow E[W|T]$ is the unique best unbiased estimator for $\tau(\theta)$

Chapter 8. Hypothesis Testing

Hypothesis testing.

$H_0 : \theta \in \Omega_0 \qquad H_1 : \theta \in \Omega_0^C$

LRT

$$\lambda(\boldsymbol{x}) = \frac{\sup\limits_{\Omega_0} L(\theta|\boldsymbol{x})}{\sup\limits_{\theta} L(\theta|\boldsymbol{x})} = \frac{\sup\limits_{\Omega_0} L(\theta|\boldsymbol{x})}{L(\hat{\theta}|\boldsymbol{x})} = \lambda^*(T(\boldsymbol{x}))$$

Reject if $\lambda(\boldsymbol{x}) \le c$.

Power function

$\beta(\theta) = P_\theta(X \in R)$

UMP

$\beta(\theta) \ge \beta'(\theta) \ \forall \theta \in \Omega^C$

Chaper 9. Interval estimation

Interval estimator $\left[L(\mathbf{X}),U(\mathbf{X})\right]$

Interval estimate $\left[L(\mathbf{x}),U(\mathbf{x})\right]$

Coverage probability: $P_\theta\left(\theta\in\left[L(\mathbf{X}),U(\mathbf{X})\right]\right)$

Methods of construction

Invertion of a test $H_0:\theta=\theta_0 \;\; H_1:\theta\neq\theta_0$

$$A(\theta_0)=\left\{\mathbf{x}:\mathbf{x}\in R^c\right\}$$

$$C(\mathbf{x})=\left\{\theta_0:\mathbf{x}\in A(\theta_0)\right\}$$

<u>Inverting LRT</u>

$$C(\mathbf{x})=\left\{\theta_0:\lambda(\mathbf{x})\geq k\right\}$$

<u>Pivotal Quantity</u>

The distribution of $Q(\mathbf{X},\theta)$ is independent of θ.

$$C(\mathbf{x})=\left\{\theta:\alpha_1\leq F_T(t|\theta)\leq 1-\alpha_2\right\}$$

<u>Credible sets</u>

$$P(\theta\in A|\mathbf{x})=\int_A \pi(\theta|\mathbf{x})d\theta$$

CHAPTER 10 ASYMPTOTICS

CONSISTENCY

$\hat{\theta}_n = \hat{\theta}_n(X_1, \ , X_n)$ $\quad\quad \hat{\theta}_n \xrightarrow{P} \theta$

EFFICIENCY

$\hat{\theta}$ unbiased, $Var(\hat{\theta})$ attains its lower bound

ASYMPTOTIC EFFICENCY

$W_n(X_1, \dots X_n)$ a.e. for $\tau(\theta)$ if.

$$\sqrt{n}\,|W_n - \tau(\theta)| \xrightarrow{D} N(0, v(\theta))$$

$$v(\theta) = \frac{|\tau'(\theta)|^2}{E\left(\left(\frac{d}{d\theta}\log f(X|\theta)\right)^2\right)} = \frac{|\tau'(\theta)|^2}{I(\theta)}$$ (one observation)

ASYMPTOTIC EFFICIENCY OF MLE

$\hat{\theta}$ MLE for θ, $\tau(\theta)$ ~~continuous~~ differentiable,

$$\sqrt{n}\,(\tau(\hat{\theta}) - \tau(\theta)) \xrightarrow{D} N(0, v(\theta))$$

$\tau(\hat{\theta})$ is consistent.

ASYMPTOTICS of LRT

$$-2\log\lambda(x) \xrightarrow{D} \chi^2(1)$$

EXAMS

SUFFICIENCY

EXPONENTIAL FAMILY, UMVUE

MLE, MSE, CRAMER-RAO

ASYMPTOTIC MLE. CONERGENCE

LRT, HYPOTHESIS, INTERVAL ESTIMATION

and, CONSTRUCTION, CONSISTENCY, BAYES

MOMENT GENERATING FUNCTIONS

Tuesday, December 5: 12-14
Wednesday, December 6: 10-12

~~Thursday, December 7: 12-14~~

Friday, December 8: 12-14