Chapter 1. Probability Theory

Sample space *S* - All possible outcomes of a particular experiment.

Event *A* – Subset of *S*

Probability – *P(A)*. $P(A): S \to \mathrm{R} \cap [0,1]$

σ - algebra (Definition 1.2.1)

A collection of subsets of *S, B,* that fulfills

1. $\phi \in B$
2. $A \in B \Rightarrow A^c \in B$
3. $A_1, A_2, \ldots \in B \Rightarrow \bigcup_{i=1}^{\infty} A_i \in B$

S finite or countable $\Rightarrow$ *B* is all subset of *S*

S not countable for instance. *S*=$(-\infty, \infty)$. B is all possible intervals of the type (a,b), (a, b], [a, b), [a,b]. (Borel *σ - algebraen*)

Probability function (Definition 1.2.4**)**

Given *S* and *B*, a probability function is a function that satisfies

1. $P(A) \geq 0 \ \forall \ A \in B$
2. $P(S) = 1$
3. $\left.\begin{matrix} A_1, A_2, \ldots \in B \\ A_i \cap A_j = \phi, \ i \neq j \end{matrix}\right\} \Rightarrow P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$

Calculus of probability

1. Addition rule (1.2.9)

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

2. Multiplication rule

$P(A \cap B) = P(A|B) \cdot P(B)$ (1.3.3)

3. The law of total probability (1.2.11)

$S = \bigcup_{i=1}^{\infty} C_i,\ C_i \cap C_j = \phi,\ \forall i \neq j$. Then

$$P(A) = \sum_{i=1}^{\infty} P(A \cap C_i) = \sum_{i=1}^{\infty} P(A|C_i) P(C_i).$$

4. Bayes rule (1.3.5)

$$P(C_i|A) = \frac{P(C_i \cap A)}{P(A)} = \frac{P(A|C_i)P(C_i)}{\sum_{j=1}^{\infty} P(A|C_j)P(C_j)}$$

Independence (1.3.12)

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

Random variables

X random variable. $X : S \to R$ (Definition 1.4.1)

Distribution function

$F_X(x) = P_X(X \le x), \; \forall \text{x}$ (Definition 1.5.1)

X is discrete if $F_X(x)$ is a step function

X is continuous if $F_X(x)$ is a continuous function

Probability mass function (X discrete**)**

$$f_X(x) = P_X(X = x) = P\left(\{s_j \in S : X\left(s_j\right) = x\}\right)$$

$$F_X(a) = \sum_{x \le a} P_X(X = x)$$

Support of X: All x for which $P_X(X = x) > 0$

Probability density function (X continuous**)**

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt, \; \forall \text{x}$$

$$f_X(x) = \frac{d}{dx} F_X(x)$$

Support of X: All x for which $f_X(x) > 0$

Identical distributed variables (Definition 1.5.8)

If $P(X \in A) = P(Y \in A) \forall A \in B$ then *X* and *Y* are identical distributed

Chapter 2. Transformations and Expectations

Distributions of Functions of a Random Variable (2.1)

X is defined on X og $Y = g(X)$ is defined on Υ.

$$P(Y \in A) = P(g(X) \in A) = P(\{x \in \mathrm{X} : g(x) \in A\}) = P\left(X \in g^{-1}(A)\right)$$

$$g^{-1}(A) = \{x \in \mathrm{X} : g(x) \in A\}$$

$$g^{-1}(y) = \{x \in \mathrm{X} : g(x) = y\}$$

X discrete

$$f_Y(y) = P(Y = y) = \sum_{x \in g^{-1}(y)} P(X = x), \text{ for } \mathrm{y} \in \Upsilon.$$

X continous

$$F_Y(y) = P(Y \le y) = P(g(X) \le y) = P(\{x \in \mathrm{X} : g(x) \le y) = \int_{\{x \in \mathrm{X}:\, g(x) \le y\}} f_X(x) dx$$

Monotone transformations (

g increasing if $u > v \Rightarrow g(u) > g(v)$

g decreasing if $u > v \Rightarrow g(u) < g(v)$

g increasing or decreasing $\Leftrightarrow$ g is monotone.

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} f_X\left(g^{-1}(y)\right)\left|\frac{d}{dy} g^{-1}(y)\right|, y \in \Upsilon \\ 0, \text{ elles} \end{cases}$$

Theorem 2.1.8

Let *X* have pdf $f_X(x)$, let $Y = g(X)$ and let χ be the sample space. Suppose there exist a partition, $A_0, A_1, \ldots, A_k$ of χ such that $P(X \in A_0) = 0$ and $f_X(x)$ is continuous on each A_i. Further suppose there exist functions $g_1(x), \ldots, g_k(x)$ defined on $A_1, \ldots, A_k$, repectively, satisfying:

i. $g(x) = g_i(x), \text{ for } x \in A_i$
ii. $g_i(x)$ is monotone on A_i
iii. The set $\Upsilon = \{y : y = g(x_i) \text{ for some } x \in A_i\}$ is the same for each $i = 1, 2, \ldots, k$,
iv. and $g_i^{-1}(y)$ has a continuous derivative on Υ, for each $i = 1, 2, \ldots, k$

Then $f_Y(y) = \begin{cases} \sum_{i=1}^{k} f_X\left(g_i^{-1}(y)\right)\left|\frac{d}{dy} g^{-1}(y)\right|, y \in \Upsilon \\ 0 \qquad \text{otherwise} \end{cases}$

Expected Value (2.2)

If $\begin{cases} \int_{-\infty}^{\infty} |x| f_X(x)\,dx < \infty \\ \sum_x |x| P(X = x) < \infty \end{cases}$ then $E[X] = \begin{cases} \int_{-\infty}^{\infty} x f_X(x)\,dx < \infty \\ \sum_x x P(X = x) < \infty \end{cases}$

Definition 2.2.1

$$E[g(X)] = \begin{cases} \int_{-\infty}^{\infty} g(x) f_X(x)\,dx \\ \sum_{x \in \mathrm{X}} g(x) P(X = x) \end{cases}$$

$$E\left[\sum_{i=1}^{n} g(X_i)\right] = \sum_{i=1}^{n} E[g(X_i)]$$

Momentgenerating function (2.3)

$$M_X(t) = E\left[e^{tX}\right] = \begin{cases} \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx, & \text{X continuous} \\ \sum_x e^{tx} P(X = x), & \text{X discrete} \end{cases}$$

$$M_X^n(t) = E\left[X^n e^{tX}\right]$$

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

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

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

$$X = e^Y \Rightarrow E\left[X^n\right] = M_Y(n)$$

Overview of some natural occurring distributions

Independent trials Register: $\mathrm{A}/\mathrm{A}^{\mathrm{c}}$ $P(A)=p$	Events in disjoint timeintervals are independent $P(\text{One event in }\Delta\text{t})=\lambda\Delta\text{t}+o(\Delta\text{t})$ $P(\text{More than one event in }\Delta\text{t})=o(\Delta\text{t})$
X=number of times A occurs in n trials $P(X=x)=\binom{n}{x}p^x(1-p)^{n-x},x=0,1,\ldots,n$	X=number of times A occur in [0,t] $P(X=x)=\frac{(\lambda t)^x e^{-\lambda t}}{x!},\ x=0,1,2,\ldots$
X=number of trials until A occurs for the first time $P(X=x)=(1-p)^{x-1}p,\ \mathrm{x}=1,2,\ldots$	X= time until A occurs for the first time $f_X(x)=\begin{cases}\lambda e^{-\lambda x}, \text{ x>0}\\ 0, \text{ otherwise}\end{cases}$
X=number of trials until A occurs for the r-th time $P(X=x)=\binom{x-1}{r-1}p^r(1-p)^{x-r},\ \mathrm{x}=r,r+1,\ldots$	X=time until A occurs the r-th time $f_X(x)=\begin{cases}\frac{\lambda^r}{\Gamma(r)}x^{r-1}e^{-\lambda x}, \text{ x>0}\\ 0, \text{ otherwise}\end{cases}$

Gamma distribution

$$f_X(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}},\ x > 0,\ \alpha>0,\ \beta>0.$$

$$X \sim \Gamma(\alpha,\beta) \Rightarrow Y = cX \sim \Gamma(\alpha, c\beta)$$

$$E\left[X^n\right] = \frac{\Gamma(\alpha+n)\beta^n}{\Gamma(\alpha)},\ n > \text{-}\alpha$$

$$\alpha = 1 \Rightarrow X \sim \exp\left(\frac{1}{\beta}\right)$$

$$\alpha = \frac{\nu}{2},\ \beta=2 \Rightarrow \mathrm{X} \sim \chi^2(\nu)$$

$$X_i \sim \Gamma(\alpha_i,\beta), i = 1,2,\ldots,n \Rightarrow \sum_{i=1}^{n} X_i \sim \Gamma\left(\sum_{i=1}^{n}\alpha_i, \beta\right)$$

Beta distribution

$$f\left(x \middle| \alpha, \beta\right) = \frac{\Gamma\left(\alpha+\beta\right)}{\Gamma\left(\alpha\right)\Gamma\left(\beta\right)} x^{\alpha-1}\left(1-x\right)^{\beta-1},\ 0<x<1,\ \alpha>0,\ \beta>0$$

$$E\left[X^n\right] = \frac{\Gamma\left(\alpha+n\right)\Gamma\left(\alpha+\beta\right)}{\Gamma\left(\alpha+\beta+n\right)\Gamma\left(\alpha\right)},\ n>\text{-}\alpha$$

Exponential Class of distributions

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

$$E\left(\sum_{i=1}^{k} \frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j} t_i(X)\right) = -\frac{\partial}{\partial \theta_j} \log c(\boldsymbol{\theta})$$

$$Var\left(\sum_{i=1}^{k} \frac{\partial w_i(\boldsymbol{\theta})}{\partial \theta_j} t_i(X)\right) = -\frac{\partial^2}{\partial \theta_j} \log c(\boldsymbol{\theta}) - E\left(\sum_{i=1}^{k} \frac{\partial^2 w_i(\boldsymbol{\theta})}{\partial^2 \theta_j} t_i(X)\right)$$

Location – Scale Families

$f(x)$ pdf. The family of pdfs: $\frac{1}{\sigma} f\left(\frac{x-u}{\sigma}\right),$

$\mu \in (-\infty, \infty),\ \sigma > 0$

The distribution of $Y = \mu + \sigma X$

Chebyshevs

$g(x) \geq 0,$ r>0

$$P(g(X) \geq r) \leq \frac{Eg(X)}{r}$$

Bivariate transformations

Monotone

$$\begin{aligned} U &= g_1(X,Y) \\ V &= g_2(X,Y) \end{aligned} \Rightarrow \begin{cases} X = h_1(U,V) \\ Y = h_2(U,V) \end{cases}$$

$$J = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix}$$

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

Hierarchical Models and Mixture Distributions

$$X \big| Y \sim B(Y, p)$$

$$Y \big| \Lambda \sim Po(\Lambda)$$

$$\Lambda \sim \exp(\beta)$$

$$E[X] = E\left[E[X|Y]\right]$$

$$Var[X] = E\left[Var[X|Y]\right] + Var\left[E[X|Y]\right]$$

Week 39

Hølders Inequality

$$\left|E[XY]\right| \le E|XY| \le \left(E|X|^p\right)^{\frac{1}{p}}\left(E|X|^q\right)^{\frac{1}{q}},\ \frac{1}{p}+\frac{1}{q}=1$$

Jensen's Inequality

$$E[g(X)] \ge g\left(E[X]\right),\ g(x)\,\text{convex}$$

Chapter 5 Random Sample

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

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

<u>Some properties of Statistics</u>

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

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

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

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

$$Var\left[T_p\right]=\frac{p}{p-2}$$

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

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

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

$$E[F] = \frac{q}{q-2}$$

$$Var[F] = \frac{2q^2(q+p-2)}{p(q-2)^2(q-4)}$$

Convergence concepts

<u>Convergence in probability:</u>

$$\{X_i\}_{i=1}^{\infty} \xrightarrow{P} X \text{ if } \forall \varepsilon > 0, \lim_{n\to\infty} \mathrm{P}\left(|X_n - X| \geq \varepsilon\right) = 0$$

Weak law of large numbers

$\{X_i\}_{i=1}^{\infty}\ iid,\ \mathrm{E}[X_i]=\mu$ and $\mathrm{Var}(X_i)=\sigma^2<\infty$. Then $\lim_{n\to\infty} P\left(\left|\bar{X}_n-\mu\right|<\varepsilon\right)=1$

$\{X_i\}_{i=1}^{\infty}\xrightarrow{P}X$ then $\{h(X_i)\}_{i=1}^{\infty}\xrightarrow{P}h(X)$ if h is continuous.

Convergence in distribution

$\{X_i\}_{i=1}^{\infty}\xrightarrow{D}X$ if $\lim_{n\to\infty} F_{X_n}(x)=F_X(x)$ at all x where $F_X(x)$ is continuous.

$\{X_i\}_{i=1}^{\infty}\xrightarrow{P}X \Rightarrow \{X_i\}_{i=1}^{\infty}\xrightarrow{D}X$

Central Limit Theorem

$\{X_i\}_{i=1}^{\infty}\ iid,\ \mathrm{E}[X_i]=\mu$ and $\mathrm{Var}(X_i)=\sigma^2<\infty$.

Define $X_n=\frac{1}{n}\sum_{i=1}^{n}X_i$. Then $\sqrt{n}\left(\frac{X_n-\mu}{\sigma}\right)\xrightarrow{D}X$ where $X\sim\mathrm{N}(0,1)$.

Slutsky's Theorem.

$X_n \xrightarrow{D} X,\ Y_n \xrightarrow{P} a,$ then

a) $X_n Y_n \xrightarrow{D} aX$

$b) X_n + Y_n \xrightarrow{D} X + a$

Delta method

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

$g'(\theta) = 0$

$$\sqrt{n}\left(Y_n - \theta\right) \xrightarrow{D} N\left(0, \sigma^2\right) \Rightarrow \sqrt{n}\left(g\left(Y_n\right) - g(\theta)\right) \xrightarrow{D} \frac{\sigma^2}{2}\left[g''(\theta)\right]\chi_1^2$$

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 $\boldsymbol{\theta}$.

A sufficient statistics for a parameter (-vector) $\boldsymbol{\theta}$ is a statistic that in a certain sense, captures all the information about $\boldsymbol{\theta}$ in the sample.

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 $\boldsymbol{\theta}$.

Theorem 6.2.6

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.

Definition 6.2.21

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

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

<u>Completeness and the exponential class</u>

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

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

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

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

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

<u>Invariance principle:</u>

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

Bayes estimation:

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

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

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

The mean square error

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

Score statistic

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

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

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

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

Cramer-Rao

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

Cramer-Rao iid

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

Equality

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

Cramer-Rao in the multiparameter case

$$\boldsymbol{\theta} = \left(\theta_1, \ldots \theta_k\right)^t$$

Define the Score function $\boldsymbol{S}\left(\boldsymbol{X}\,|\boldsymbol{\theta}\right) = \begin{bmatrix} \frac{\partial}{\partial\theta_1}\log f\left(\boldsymbol{x}|\boldsymbol{\theta}\right) \\ \vdots \\ \frac{\partial}{\partial\theta_k}\log f\left(\boldsymbol{x}|\boldsymbol{\theta}\right) \end{bmatrix} = \nabla \log f\left(\boldsymbol{x}|\boldsymbol{\theta}\right)$

Define the Fisher information $I\left(\boldsymbol{\theta}\right) = Cov\left[\boldsymbol{S}\left(\boldsymbol{X}\,|\boldsymbol{\theta}\right)\right]$

We have as in the univariate case that $E\left[\boldsymbol{S}\left(\boldsymbol{X}\,|\boldsymbol{\theta}\right)\right]=\boldsymbol{0}$ and
$I(\boldsymbol{\theta})=E\left[\boldsymbol{S}\left(\boldsymbol{X}\,|\boldsymbol{\theta}\right)\boldsymbol{S}\left(\boldsymbol{X}\,|\boldsymbol{\theta}\right)^{T}\right] = -E\left[H\left(\boldsymbol{X}\,|\boldsymbol{\theta}\right)\right]$ where
$h_{ij}=\dfrac{\partial}{\partial\theta_i}\dfrac{\partial}{\partial\theta_j}\log f\left(\boldsymbol{x}|\boldsymbol{\theta}\right)$.

If $\boldsymbol{W}\left(\boldsymbol{X}\right)$ is an unbiased estimator for $\boldsymbol{\theta}$. Then $\boldsymbol{I}\left(\boldsymbol{\theta}\right)^{-1}$ is taken as an approximation to $Cov\left[\boldsymbol{W}\left(\boldsymbol{X}\right)\right]$

Let $\tau=\tau(\boldsymbol{\theta})$ be univariate and let $\nabla\tau(\boldsymbol{\theta})=\begin{bmatrix}\dfrac{\partial}{\partial\theta_1}\tau(\boldsymbol{\theta})\\ \vdots \\ \dfrac{\partial}{\partial\theta_k}\tau(\boldsymbol{\theta})\end{bmatrix}$

Theorem. For an estimator $W\left(\boldsymbol{X}\right)$ with $E\left[W\left(\boldsymbol{X}\right)\right]=\tau$, we have under similar regularity conditions as in the univariate case that
$Var\left[W\left(\boldsymbol{X}\right)\right]\geq\left(\nabla\tau(\boldsymbol{\theta})\right)^{T}\left(I(\boldsymbol{\theta})\right)^{-1}\left(\nabla\tau(\boldsymbol{\theta})\right)$.

Sufficiency and Unbiasedness

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

T a sufficient statistic $E\left[W|T\right]=\tau(\theta)$ and $Var\left[W|T\right]\le Var\left[W\right], \ \forall\theta$

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

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

$\beta(\theta)\ge\beta'(\theta) \ \forall\theta\in\Omega_0^C$

Neyman-Pearson

$H_0 : \theta = \theta_0 \qquad H_1 : \theta = \theta_1$

UMP level α test.

$x \in R$ if $f\left(x \middle| \theta_1\right) > kf\left(x \middle| \theta_0\right)$

$x \in R^C$ if $f\left(x \middle| \theta_1\right) < kf\left(x \middle| \theta_0\right)$

for some $k \geq 0$ and $\alpha = P_{\theta_0}(X \in R)$

Interval Estimator

$\left[L\left(\boldsymbol{X}\right), U\left(\boldsymbol{X}\right)\right]$

Interval Estimate

$\left[L\left(\boldsymbol{x}\right), U\left(\boldsymbol{x}\right)\right]$

Coverage Probability

$P\left(\theta \in \left[L\left(\boldsymbol{X}\right), U\left(\boldsymbol{X}\right)\right]\right)$

Methods of Construction

Inverting a test $H_0\!:\theta=\theta_0 \quad H_1\!:\theta \neq \theta_0$

$$A(\theta_0) = \{\boldsymbol{x} : \boldsymbol{x} \in R^C\}$$

$$C(\boldsymbol{x}) = \{\theta_0 : \boldsymbol{x} \in A(\theta_0)\}$$

Inverting LRT

$$C(\boldsymbol{x}) = \{\theta_0 : \lambda(\boldsymbol{x}) \geq \mathrm{k}\}$$

Pivotal Quantity

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

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

Credible sets.

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

An example

$X_1, \ldots X_n$ iid Poisson$(\lambda) \Rightarrow Y = \sum_{i=1}^{n} X_i \sim$ Poisson$(n\lambda)$

$\pi(\lambda) = \text{gamma}(\alpha, \beta)$

$\pi\left(\lambda \middle| \sum_{i=1}^{n} x_i = y\right) = \text{gamma}\left(\alpha + y, \frac{\beta}{n\beta + 1}\right)$ which gives the $1-\alpha$ credibility interval

$$P\left(\frac{\beta}{2(n\beta+1)}\chi\left(2(y+\alpha)\right)_{1-\frac{\alpha}{2}} \leq \lambda \leq \frac{\beta}{2(n\beta+1)}\chi\left(2(y+\alpha)\right)_{\frac{\alpha}{2}}\right) = 1-\alpha$$

Which can be compared to the $1-\alpha$ confidence interval.

$$P\left(\frac{1}{2n}\chi(2y)_{1-\frac{\alpha}{2}} \leq \lambda \leq \frac{1}{2n}\chi\left(2(y+1)\right)_{\frac{\alpha}{2}}\right) = 1-\alpha$$

Asymptotics

Consistent estimator

$\hat{\theta}(X_1,\ldots X_n) \xrightarrow{P} \theta,\ \forall\theta.$

Efficient estimator

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

Asymptotic efficient estimator

$$\sqrt{n}\left(W_n - \tau(\theta)\right) \xrightarrow{D} N\left(0, v(\theta)\right),\ v(\theta) = \frac{\left|\tau'(\theta)\right|^2}{E\left[\left(\frac{d}{d\theta} log\ f(X|\theta)\right)^2\right]}$$

Asymptotic efficient and consistent MLE

$X_1,\ldots,X_n$ iid, $\hat{\theta}_n$ MLE $\Rightarrow \hat{\theta}_n \xrightarrow{P} \theta$ and

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

Asymptotics of LRT

$$-2 log\, \lambda(X_1,\ldots,X_n) \xrightarrow{D} \chi(1)$$