Kunnskap for ei betre verd

Institutt for matematiske fag

Eksamensoppgåve i **TMA4295 Statistisk inferens**

Fagleg kontakt under eksamen: Vaclav Slimacek

Tlf: 942 96 313

Eksamensdato: Tirsdag 2. desember 2014

Eksamenstid (frå–til): 09:00-–13:00

Hjelpemiddelkode/Tillatne hjelpemiddel: C:
Tabeller og Formler i Statistikk, Tapir
NTNU-godkjent kalkulator
Personleg, handskriven, gul hugselapp - A5-format

Annan informasjon:
Kopier av nokre viktige resultat frå læreboka er gitt som vedlegg

Målform/språk: nynorsk

Sidetal: 3

Sidetal vedlegg: 4

Kontrollert av:

Dato Sign

Oppgåve 1 *Terningane*

Det er mistanke om at terningane som ein brukar ved eit casino er manipulerte, slik at sjansane for å få 1-arar og 6-arar er endra, medan dei andre utfalla ikkje er påverka.

La p_i vere sannsynet for å få i "auge" med ein terning frå casinoet. Følgjande modell vil bli studert:

$$p_1 = \frac{1}{6} - \theta,\ p_2 = p_3 = p_4 = p_5 = \frac{1}{6},\ p_6 = \frac{1}{6} + \theta,$$

der θ er ein ukjend parameter med $|\theta| < 1/6$.

Ein gjer n kast med ein slik terning, der X_i av desse kasta ender med i auge $(i = 1, 2, \ldots, 6)$.

a) Kva for føresetnader må ein gjere for å sikre at den observerte vektoren $\boldsymbol{X} = (X_1, X_2, \ldots, X_6)$ er multinomisk fordelt? Anta i det følgjande at desse føresetnadene held.

Merk at resultat frå læreboka som gjeld uavhengige og identisk fordelte observasjonar også vil gjelde for multinomiske forsøk som i denne oppgåva.

Vis at den simultane sannsynsmassefunksjonen (*pmf*) for $\boldsymbol{X}$ kan verte skriven som ein eksponensiell familie på følgjande form:

$$f(\boldsymbol{x}|\theta) = h(\boldsymbol{x}) \exp\left\{x_1 \ln\left(\frac{1}{6} - \theta\right) + x_6 \ln\left(\frac{1}{6} + \theta\right)\right\} \quad \text{for } \boldsymbol{x} \in \boldsymbol{A}.$$

Spesifiser funksjonen $h(\boldsymbol{x})$ og mengda $\boldsymbol{A}$ av mulige verdiar for $\boldsymbol{X}$.

Gjer greie for kvifor den to-dimensjonale observatoren $T(\boldsymbol{X}) = (X_1, X_6)$ er suffisient for θ.

b) Vis at $T(\boldsymbol{X})$ er minimal-suffisient.

c) Vis at maximum likelihood estimatoren (*MLE*) for θ er gitt ved

$$\hat{\theta} = \frac{1}{6} \cdot \frac{x_6 - x_1}{x_6 + x_1} \quad \text{hvis } x_6 + x_1 > 0,$$

og at $\hat{\theta}$ er ubestemd dersom $x_6 + x_1 = 0$.

d) For å kunne påvise at terningen er manipulert, vil ein teste

$$H_0: \ \theta = 0 \text{ mot } H_1: \ \theta \neq 0.$$

Finn eit uttrykk for sannsynskvota (*likelihood ratio*) $\lambda(\boldsymbol{x})$ for dette problemet.

Kva blir konklusjonen på sannsynskvotetesten (*LRT*) dersom $n = 100$ og den observerte vektor er

$$(10, 14, 17, 21, 16, 22) \ \ ?$$

Du skal her bruke den asymptotiske fordelinga for sannsynskvota. Bruk signifikansnivå $\alpha = 0.05$.

e) Vis at Cramér-Raos nedre grense for variansen til forventingsrette estimatorar for θ er

$$\frac{1 - 36\theta^2}{12n}$$

Bruk denne til å setje opp den asymptotiske fordelinga for maximum likelihood estimatoren $\hat{\theta}$.

Finn eit tilnærma 95% konfidensintervall for θ, og berekn intervallet når observasjonane er som i det førre punktet.

Gjer også kort greie for, utan å gjere alle berekningane, korleis ein kan utleie ei tilnærma 95% konfidensmengde for θ ved å invertere ein sannsynskvotetest.

f) Vis at estimatoren

$$\tilde{\theta} = \frac{X_6 - X_1}{2n}$$

er forventingsrett for θ, og finn eit uttrykk for variansen.

Kvifor kan ikkje denne estimatoren bli forbetra ved hjelp av Rao-Blackwells teorem?

g) Estimatoren $\tilde{\theta}$ frå førre punkt er forventingsrett, og er dessutan ein funksjon av den suffisiente observator $T(\boldsymbol{X})$. Eit naturleg spørsmål er då om den er ein UMVU-estimator.

Kan du bruke Cramér-Raos nedre grense frå punkt e) til å avgjere dette? Grunngi svaret.

Dersom svaret er nei, er det naturleg å sjekke om Theorem 7.3.23 i vedlegget kan bli bruka til å avgjere om $\tilde{\theta}$ er UMVU. (Dette teoremet er essensielt det som i forelesningane er kalt *Lehmann-Scheffés teorem*).

Vis at $T(\boldsymbol{X})$ *ikkje* er komplett, ved å bruke definisjonen av komplettheit.

Kan du konkludere noko frå Theorem 7.3.23?

Oppgåve 2 *Tid til feil*

Ein maskin blir sett i gang ved begynninga av døger nr. 1 og blir observert til den feilar første gong. La $Y = y$ tyde at dette skjer i døger nr. y, $(y = 1, 2, \ldots)$.

a) Sett opp føresetnader for at Y er geometrisk fordelt med parameter p, $0 < p < 1$, dvs. at Y har pmf

$$f(y) = (1-p)^{y-1}p \quad \text{for } y = 1, 2, \ldots$$

I det følgjande skal du anta at desse føresetnadene er oppfylte.

Vis at den momentgenererande funksjon (*mgf*) for Y er gitt ved

$$M_Y(t) = \frac{pe^t}{1-(1-p)e^t} \quad \text{for } t < -\ln(1-p).$$

Vis korleis dette kan bli bruka til å vise at tida til feil, Y, har forventing $\mu = 1/p$ (døger).

For å kunne registrere feiltida meir nøyaktig deler ein døgeret i n delar (for heltall $n > 1$), slik at tida Y_n til feil nå blir målt med eining $1/n$ døger (t.d. i timar om $n = 24$).

Det er naturleg å anta at Y_n er geometrisk fordelt med parameter $p_n = p/n$, medan tida til feil, målt med eining *døger*, blir

$$X_n = \frac{1}{n}Y_n$$

b) Vis at X_n konvergerer i fordeling mot ein tilfeldig variabel X, dvs. $X_n \xrightarrow{d} X$. Kva for kjent fordeling har X? Kva blir $E(X)$?

(*Vink:* Berekn mgf for X_n og finn grensa når $n \to \infty$).

TMA 4295 Statistical Inference

IMF/IME/NTNU

Formulaes from Casella & Berger

Theorem 5.2.11 *Suppose $X_1, \dots, X_n$ is a random sample from a pdf or pmf $f(x|\theta)$, where*

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

is a member of an exponential family. Define statistics $T_1, \dots, T_k$ by

$$T_i(X_1, \dots, X_n) = \sum_{j=1}^{n} t_i(X_j), \quad i = 1, \dots, k.$$

If the set $\{(w_1(\theta), w_2(\theta), \dots, w_k(\theta)), \theta \in \Theta\}$ contains an open subset of $\Re^k$, then the distribution of $(T_1, \dots, T_k)$ is an exponential family of the form

$$f_T(u_1, \dots, u_k|\theta) = H(u_1, \dots, u_k)[c(\theta)]^n \exp\left(\sum_{i=1}^{k} w_i(\theta)u_i\right). \tag{5.2.6}$$

Definition 5.5.1 A sequence of random variables, $X_1, X_2, \dots$, *converges in probability* to a random variable X if, for every $\epsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0 \quad \text{or, equivalently,} \quad \lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1.$$

Definition 5.5.6 A sequence of random variables, $X_1, X_2, \dots$, *converges almost surely* to a random variable X if, for every $\epsilon > 0$,

$$P(\lim_{n\to\infty} |X_n - X| < \epsilon) = 1.$$

Theorem 5.5.9 (Strong Law of Large Numbers) *Let $X_1, X_2, \dots$ be iid random variables with $\mathrm{E}X_i = \mu$ and $\mathrm{Var}\, X_i = \sigma^2 < \infty$, and define $\bar{X}_n = (1/n)\sum_{i=1}^{n} X_i$. Then, for every $\epsilon > 0$,*

$$P(\lim_{n\to\infty} |\bar{X}_n - \mu| < \epsilon) = 1;$$

that is, $\bar{X}_n$ converges almost surely to μ.

Definition 5.5.10 A sequence of random variables, $X_1, X_2, \dots$, *converges in distribution* to a random variable X if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

at all points x where $F_X(x)$ is continuous.

Theorem 5.5.15 (Stronger form of the Central Limit Theorem) *Let $X_1, X_2, \dots$ be a sequence of iid random variables with $\mathrm{E}X_i = \mu$ and $0 < \mathrm{Var}\, X_i = \sigma^2 < \infty$. Define $\bar{X}_n = (1/n)\sum_{i=1}^{n} X_i$. Let $G_n(x)$ denote the cdf of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Then, for any x, $-\infty < x < \infty$,*

$$\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy;$$

that is, $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ has a limiting standard normal distribution.

Theorem 5.5.17 (Slutsky's Theorem) *If $X_n \to X$ in distribution and $Y_n \to a$, a constant, in probability, then*

a. *$Y_n X_n \to aX$ in distribution.*

b. *$X_n + Y_n \to X + a$ in distribution.*

Theorem 5.5.24 (Delta Method) *Let Y_n be a sequence of random variables that satisfies $\sqrt{n}(Y_n - \theta) \to \mathrm{n}(0, \sigma^2)$ in distribution. For a given function g and a specific value of θ, suppose that $g'(\theta)$ exists and is not 0. Then*

$$\sqrt{n}[g(Y_n) - g(\theta)] \to \mathrm{n}(0, \sigma^2[g'(\theta)]^2) \text{ in distribution.} \tag{5.5.10}$$

Definition 6.2.1 A statistic $T(\mathbf{X})$ is a *sufficient statistic for θ* if the conditional distribution of the sample $\mathbf{X}$ given the value of $T(\mathbf{X})$ does not depend on θ.

Theorem 6.2.2 *If $p(\mathbf{x}|\theta)$ is the joint pdf or pmf of $\mathbf{X}$ and $q(t|\theta)$ is the pdf or pmf of $T(\mathbf{X})$, then $T(\mathbf{X})$ is a sufficient statistic for θ if, for every $\mathbf{x}$ in the sample space, the ratio $p(\mathbf{x}|\theta)/q(T(\mathbf{x})|\theta)$ is constant as a function of θ.*

Theorem 6.2.6 (Factorization Theorem) *Let $f(\mathbf{x}|\theta)$ denote the joint pdf or pmf of a sample $\mathbf{X}$. A statistic $T(\mathbf{X})$ is a sufficient statistic for θ if and only if there exist functions $g(t|\theta)$ and $h(\mathbf{x})$ such that, for all sample points $\mathbf{x}$ and all parameter points θ,*

$$f(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x}). \tag{6.2.3}$$

Theorem 6.2.10 *Let $X_1, \dots, X_n$ be iid observations from a pdf or pmf $f(x|\boldsymbol{\theta})$ that belongs to an exponential family given by*

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

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_d)$, $d \leq k$. Then

$$T(\mathbf{X}) = \left(\sum_{j=1}^{n} t_1(X_j), \dots, \sum_{j=1}^{n} t_k(X_j)\right)$$

is a sufficient statistic for $\boldsymbol{\theta}$.

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

Theorem 6.2.13 *Let $f(\mathbf{x}|\theta)$ be the pmf or pdf of a sample $\mathbf{X}$. Suppose there exists a function $T(\mathbf{x})$ such that, for every two sample points $\mathbf{x}$ and $\mathbf{y}$, the ratio $f(\mathbf{x}|\theta)/f(\mathbf{y}|\theta)$ is constant as a function of θ if and only if $T(\mathbf{x}) = T(\mathbf{y})$. Then $T(\mathbf{X})$ is a minimal sufficient statistic for θ.*

Definition 6.2.21 Let $f(t|\theta)$ be a family of pdfs or pmfs for a statistic $T(\mathbf{X})$. The family of probability distributions is called *complete* if $E_\theta g(T) = 0$ for all θ implies $P_\theta(g(T) = 0) = 1$ for all θ. Equivalently, $T(\mathbf{X})$ is called a *complete statistic.*

Theorem 6.2.28 *If a minimal sufficient statistic exists, then any complete statistic is also a minimal sufficient statistic.*

Definition 7.2.4 For each sample point $\mathbf{x}$, let $\hat\theta(\mathbf{x})$ be a parameter value at which $L(\theta|\mathbf{x})$ attains its maximum as a function of θ, with $\mathbf{x}$ held fixed. A *maximum likelihood estimator* (MLE) of the parameter θ based on a sample $\mathbf{X}$ is $\hat\theta(\mathbf{X})$.

Theorem 7.2.10 (Invariance property of MLEs) *If $\hat\theta$ is the MLE of θ, then for any function $\tau(\theta)$, the MLE of $\tau(\theta)$ is $\tau(\hat\theta)$.*

Definition 7.3.7 An estimator W^* is a *best unbiased estimator* of $\tau(\theta)$ if it satisfies $E_\theta W^* = \tau(\theta)$ for all θ and, for any other estimator W with $E_\theta W = \tau(\theta)$, we have $\mathrm{Var}_\theta W^* \le \mathrm{Var}_\theta W$ for all θ. W^* is also called a *uniform minimum variance unbiased estimator* (UMVUE) of $\tau(\theta)$.

Theorem 7.3.9 (Cramér–Rao Inequality) *Let $X_1, \ldots, X_n$ be a sample with pdf $f(\mathbf{x}|\theta)$, and let $W(\mathbf{X}) = W(X_1, \ldots, X_n)$ be any estimator satisfying*

$$\frac{d}{d\theta} E_\theta W(\mathbf{X}) = \int_{\mathcal{X}} \frac{\partial}{\partial\theta}\left[W(\mathbf{x}) f(\mathbf{x}|\theta)\right] d\mathbf{x}$$

(7.3.4) *and*

$$\mathrm{Var}_\theta W(\mathbf{X}) < \infty.$$

Then

(7.3.5)
$$\mathrm{Var}_\theta\,(W(\mathbf{X})) \ge \frac{\left(\frac{d}{d\theta} E_\theta W(\mathbf{X})\right)^2}{E_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(\mathbf{X}|\theta)\right)^2\right)}.$$

Corollary 7.3.15 (Attainment) *Let $X_1, \ldots, X_n$ be iid $f(x|\theta)$, where $f(x|\theta)$ satisfies the conditions of the Cramér–Rao Theorem. Let $L(\theta|\mathbf{x}) = \prod_{i=1}^n f(x_i|\theta)$ denote the likelihood function. If $W(\mathbf{X}) = W(X_1, \ldots, X_n)$ is any unbiased estimator of $\tau(\theta)$, then $W(\mathbf{X})$ attains the Cramér–Rao Lower Bound if and only if*

(7.3.12)
$$a(\theta)[W(\mathbf{x}) - \tau(\theta)] = \frac{\partial}{\partial\theta}\log L(\theta|\mathbf{x})$$

for some function $a(\theta)$.

Theorem 7.3.17 (Rao–Blackwell) *Let W be any unbiased estimator of $\tau(\theta)$, and let T be a sufficient statistic for θ. Define $\phi(T) = E(W|T)$. Then $E_\theta\phi(T) = \tau(\theta)$ and $\mathrm{Var}_\theta\,\phi(T) \le \mathrm{Var}_\theta W$ for all θ; that is, $\phi(T)$ is a uniformly better unbiased estimator of $\tau(\theta)$.*

Theorem 7.3.23 *Let T be a complete sufficient statistic for a parameter θ, and let $\phi(T)$ be any estimator based only on T. Then $\phi(T)$ is the unique best unbiased estimator of its expected value.*

Definition 8.2.1 The *likelihood ratio test statistic* for testing $H_0: \theta \in \Theta_0$ versus $H_1: \theta \in \Theta_0^c$ is

$$\lambda(\mathbf{x}) = \frac{\sup_{\Theta_0} L(\theta|\mathbf{x})}{\sup_{\Theta} L(\theta|\mathbf{x})}.$$

A *likelihood ratio test* (LRT) is any test that has a rejection region of the form $\{\mathbf{x}: \lambda(\mathbf{x}) \le c\}$, where c is any number satisfying $0 \le c \le 1$.

Theorem 8.2.4 *If $T(\mathbf{X})$ is a sufficient statistic for θ and $\lambda^*(t)$ and $\lambda(\mathbf{x})$ are the LRT statistics based on T and $\mathbf{X}$, respectively, then $\lambda^*(T(\mathbf{x})) = \lambda(\mathbf{x})$ for every $\mathbf{x}$ in the sample space.*

Definition 8.3.1 The *power function* of a hypothesis test with rejection region R is the function of θ defined by $\beta(\theta) = P_\theta(\mathbf{X} \in R)$.

Definition 8.3.5 For $0 \le \alpha \le 1$, a test with power function $\beta(\theta)$ is a *size α test* if $\sup_{\theta\in\Theta_0} \beta(\theta) = \alpha$.

Definition 8.3.6 For $0 \le \alpha \le 1$, a test with power function $\beta(\theta)$ is a *level α test* if $\sup_{\theta\in\Theta_0} \beta(\theta) \le \alpha$.

Definition 8.3.9 A test with power function $\beta(\theta)$ is *unbiased* if $\beta(\theta') \ge \beta(\theta'')$ for every $\theta' \in \Theta_0^c$ and $\theta'' \in \Theta_0$.

Definition 8.3.11 Let $\mathcal{C}$ be a class of tests for testing $H_0: \theta \in \Theta_0$ versus $H_1: \theta \in \Theta_0^c$. A test in class $\mathcal{C}$, with power function $\beta(\theta)$, is a *uniformly most powerful* (UMP) *class $\mathcal{C}$ test* if $\beta(\theta) \ge \beta'(\theta)$ for every $\theta \in \Theta_0^c$ and every $\beta'(\theta)$ that is a power function of a test in class $\mathcal{C}$.

Definition 8.3.16 A family of pdfs or pmfs $\{g(t|\theta): \theta \in \Theta\}$ for a univariate random variable T with real-valued parameter θ has a *monotone likelihood ratio* (MLR) if, for every $\theta_2 > \theta_1$, $g(t|\theta_2)/g(t|\theta_1)$ is a monotone (nonincreasing or nondecreasing) function of t on $\{t: g(t|\theta_1) > 0 \text{ or } g(t|\theta_2) > 0\}$. Note that $c/0$ is defined as ∞ if $0 < c$.

Theorem 8.3.17 (Karlin–Rubin) *Consider testing $H_0: \theta \le \theta_0$ versus $H_1: \theta > \theta_0$. Suppose that T is a sufficient statistic for θ and the family of pdfs or pmfs $\{g(t|\theta): \theta \in \Theta\}$ of T has an MLR. Then for any t_0, the test that rejects H_0 if and only if $T > t_0$ is a UMP level α test, where $\alpha = P_{\theta_0}(T > t_0)$.*

Definition 8.3.26 A *p-value* $p(\mathbf{X})$ is a test statistic satisfying $0 \le p(\mathbf{x}) \le 1$ for every sample point $\mathbf{x}$. Small values of $p(\mathbf{X})$ give evidence that H_1 is true. A p-value is *valid* if, for every $\theta \in \Theta_0$ and every $0 \le \alpha \le 1$,

$$P_\theta(p(\mathbf{X}) \le \alpha) \le \alpha. \tag{8.3.8}$$

Theorem 8.3.27 *Let $W(\mathbf{X})$ be a test statistic such that large values of W give evidence that H_1 is true. For each sample point $\mathbf{x}$, define*

$$p(\mathbf{x}) = \sup_{\theta \in \Theta_0} P_\theta(W(\mathbf{X}) \ge W(\mathbf{x})). \tag{8.3.9}$$

Then, $p(\mathbf{X})$ is a valid p-value.

Definition 9.1.1 An *interval estimate* of a real-valued parameter θ is any pair of functions, $L(x_1, \ldots, x_n)$ and $U(x_1, \ldots, x_n)$, of a sample that satisfy $L(\mathbf{x}) \le U(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$. If $\mathbf{X} = \mathbf{x}$ is observed, the inference $L(\mathbf{x}) \le \theta \le U(\mathbf{x})$ is made. The random interval $[L(\mathbf{X}), U(\mathbf{X})]$ is called an *interval estimator*.

Definition 9.1.4 For an interval estimator $[L(\mathbf{X}), U(\mathbf{X})]$ of a parameter θ, the *coverage probability* of $[L(\mathbf{X}), U(\mathbf{X})]$ is the probability that the random interval $[L(\mathbf{X}), U(\mathbf{X})]$ covers the true parameter, θ. In symbols, it is denoted by either $P_\theta(\theta \in [L(\mathbf{X}), U(\mathbf{X})])$ or $P(\theta \in [L(\mathbf{X}), U(\mathbf{X})]|\theta)$.

Definition 9.1.5 For an interval estimator $[L(\mathbf{X}), U(\mathbf{X})]$ of a parameter θ, the *confidence coefficient* of $[L(\mathbf{X}), U(\mathbf{X})]$ is the infimum of the coverage probabilities, $\inf_\theta P_\theta(\theta \in [L(\mathbf{X}), U(\mathbf{X})])$.

Theorem 9.2.2 *For each $\theta_0 \in \Theta$, let $A(\theta_0)$ be the acceptance region of a level α test of $H_0: \theta = \theta_0$. For each $\mathbf{x} \in \mathcal{X}$, define a set $C(\mathbf{x})$ in the parameter space by*

$$C(\mathbf{x}) = \{\theta_0 : \mathbf{x} \in A(\theta_0)\}. \tag{9.2.1}$$

Then the random set $C(\mathbf{X})$ is a $1-\alpha$ confidence set. Conversely, let $C(\mathbf{X})$ be a $1-\alpha$ confidence set. For any $\theta_0 \in \Theta$, define

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

Then $A(\theta_0)$ is the acceptance region of a level α test of $H_0: \theta = \theta_0$.

Definition 9.2.6 A random variable $Q(\mathbf{X}, \theta) = Q(X_1, \ldots, X_n, \theta)$ is a *pivotal quantity* (or *pivot*) if the distribution of $Q(\mathbf{X}, \theta)$ is independent of all parameters. That is, if $\mathbf{X} \sim F(\mathbf{x}|\theta)$, then $Q(\mathbf{X}, \theta)$ has the same distribution for all values of θ.

Theorem 9.3.2 *Let $f(x)$ be a unimodal pdf. If the interval $[a, b]$ satisfies*

i. $\int_a^b f(x)\,dx = 1 - \alpha$,

ii. $f(a) = f(b) > 0$, *and*

iii. $a \le x^* \le b$, *where x^* is a mode of $f(x)$,*

then $[a, b]$ is the shortest among all intervals that satisfy (i).

Corollary 9.3.10 *If the posterior density $\pi(\theta|\mathbf{x})$ is unimodal, then for a given value of α, the shortest credible interval for θ is given by*

$$\{\theta : \pi(\theta|\mathbf{x}) \ge k\} \quad \text{where} \quad \int_{\{\theta:\pi(\theta|\mathbf{x})\ge k\}} \pi(\theta|\mathbf{x})\,d\theta = 1 - \alpha.$$

Definition 10.1.1 A sequence of estimators $W_n = W_n(X_1, \ldots, X_n)$ is a *consistent sequence of estimators* of the parameter θ if, for every $\epsilon > 0$ and every $\theta \in \Theta$,

$$\lim_{n\to\infty} P_\theta(|W_n - \theta| < \epsilon) = 1. \tag{10.1.1}$$

Theorem 10.1.3 *If W_n is a sequence of estimators of a parameter θ satisfying*

i. $\lim_{n\to\infty} \mathrm{Var}_\theta W_n = 0$,

ii. $\lim_{n\to\infty} \mathrm{Bias}_\theta W_n = 0$,

for every $\theta \in \Theta$, then W_n is a consistent sequence of estimators of θ.

Theorem 10.1.6 (Consistency of MLEs) *Let $X_1, X_2, \ldots,$ be iid $f(x|\theta)$, and let $L(\theta|\mathbf{x}) = \prod_{i=1}^n f(x_i|\theta)$ be the likelihood function. Let $\hat\theta$ denote the MLE of θ. Let $\tau(\theta)$ be a continuous function of θ. Under the regularity conditions in Miscellanea 10.6.2 on $f(x|\theta)$ and, hence, $L(\theta|\mathbf{x})$, for every $\epsilon > 0$ and every $\theta \in \Theta$,*

$$\lim_{n\to\infty} P_\theta(|\tau(\hat\theta) - \tau(\theta)| \ge \epsilon) = 0.$$

That is, $\tau(\hat\theta)$ is a consistent estimator of $\tau(\theta)$.

Definition 10.1.7 For an estimator T_n, if $\lim_{n\to\infty} k_n \mathrm{Var}\, T_n = \tau^2 < \infty$, where $\{k_n\}$ is a sequence of constants, then τ^2 is called the *limiting variance* or *limit of the variances*.

Definition 10.1.9 For an estimator T_n, suppose that $k_n(T_n - \tau(\theta)) \to \mathrm{n}(0, \sigma^2)$ in distribution. The parameter σ^2 is called the *asymptotic variance* or *variance of the limit distribution* of T_n.

Definition 10.1.11 A sequence of estimators W_n is *asymptotically efficient* for a parameter $\tau(\theta)$ if $\sqrt{n}[W_n - \tau(\theta)] \to \mathrm{n}[0, v(\theta)]$ in distribution and

$$v(\theta) = \frac{[\tau'(\theta)]^2}{\mathrm{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right)};$$

that is, the asymptotic variance of W_n achieves the Cramér–Rao Lower Bound.

Theorem 10.1.12 (Asymptotic efficiency of MLEs) *Let $X_1, X_2, \ldots,$ be iid $f(x|\theta)$, let $\hat\theta$ denote the MLE of θ, and let $\tau(\theta)$ be a continuous function of θ. Under the regularity conditions in Miscellanea 10.6.2 on $f(x|\theta)$ and, hence, $L(\theta|\mathbf{x})$,*

$$\sqrt{n}[\tau(\hat\theta) - \tau(\theta)] \to \mathrm{n}[0, v(\theta)],$$

where $v(\theta)$ is the Cramér–Rao Lower Bound. That is, $\tau(\hat\theta)$ is a consistent and asymptotically efficient estimator of $\tau(\theta)$.

Definition 10.1.16 If two estimators W_n and V_n satisfy

$$\sqrt{n}[W_n - \tau(\theta)] \to \mathrm{n}[0, \sigma_W^2]$$
$$\sqrt{n}[V_n - \tau(\theta)] \to \mathrm{n}[0, \sigma_V^2]$$

in distribution, the *asymptotic relative efficiency* (ARE) of V_n with respect to W_n is

$$\mathrm{ARE}(V_n, W_n) = \frac{\sigma_W^2}{\sigma_V^2}.$$

Theorem 10.3.1 (Asymptotic distribution of the LRT—simple H_0) *For testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, suppose $X_1, \ldots, X_n$ are iid $f(x|\theta)$, $\hat{\theta}$ is the MLE of θ, and $f(x|\theta)$ satisfies the regularity conditions in Miscellanea 10.6.2. Then under H_0, as $n \to \infty$,*

$$-2 \log \lambda(\mathbf{X}) \to \chi_1^2 \text{ in distribution,}$$

where χ_1^2 is a χ^2 random variable with 1 degree of freedom.

Theorem 10.3.3 *Let $X_1, \ldots, X_n$ be a random sample from a pdf or pmf $f(x|\theta)$. Under the regularity conditions in Miscellanea 10.6.2, if $\theta \in \Theta_0$, then the distribution of the statistic $-2 \log \lambda(\mathbf{X})$ converges to a chi squared distribution as the sample size $n \to \infty$. The degrees of freedom of the limiting distribution is the difference between the number of free parameters specified by $\theta \in \Theta_0$ and the number of free parameters specified by $\theta \in \Theta$.*