UofT-STA261-2015-Summer Note

hysw

Sunday 9$^{\text{th}}$ August, 2015
at 20:16

Contents

1 Basic Distributions

{ order statistics }

Distribution	repr	PDF,PMF	$E(X)$	$V(X)$	$M(t)$
Negative Binomial					
Hypergeometric					
Poisson					
Normal	$\mathcal{N}(\mu, \sigma^2)$	$\frac{\exp(-\frac{(x-\mu)^2}{2\sigma^2})}{\sigma\sqrt{2\pi}}$			
Gamma	Gamma(α, λ)	$\frac{\lambda^\alpha}{\Gamma\alpha} x^{\alpha-1} e^{-}\lambda x$			
Beta	Beta(α, β)	$\frac{x^{\alpha-1}(x-1)^{\beta-1}}{B(\alpha,\beta)}$			
Bernoulli		$[1-p, p]_k$	p	$p(1-p)$	
t					
Uniform	$U(a, b)$	$\frac{1}{b-a} I_{(a,b)}(x)$			
Exponential	Exp(λ)	$\lambda e^{-\lambda x}$			
Binomial	Binom(n, p)	$\binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$	
Geometric		$p(1-p)^{k-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	

Where the beta function is defined by $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

2 Definitions and Theorems

2.1 Random Variable

The **sample space** Ω is the set of all possible outcome. **Events** are the subset of Ω (denoted: $A, B, \ldots$). A **random variable** $X : \Omega \to E$ is a measurable function from the set of possible outcomes Ω to some set E(usually $E = \mathbb{R}$).

{ expected value }

{ variance }

{ standard deviation }

{ covariance }

{ correlation }

{ mgf - moment generating function }

{ the other basics }

Conditional probability:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

Two events A and B are independent if $P(A \cap B) = \Pr(A)\Pr(B)$.

Bayes' Rule

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}$$

{ PMF } { PDF } { CDF }

2.2 Convergence of random variables

Let $X_1, X_2, \ldots$ be a sequence of random variables. Let Y be another random variable.

The sequence $\{X_n\}$ **converges** to Y **in distribution** if

$$\forall x \in \mathbb{R} \left[\lim_{n\to\infty} \Pr(X_n \le x) = \Pr(Y \le x) \right]$$

Note that $\Pr(X_n \le x) = F_{X_n}(x)$ and $\Pr(Y \le x) = F_Y(x)$.

The sequence $\{X_n\}$ **converges** to Y **in probobility** if

$$\forall \varepsilon > 0 \left[\lim_{n\to\infty} \Pr(|X_n - Y| > \varepsilon) = 0 \right]$$

Suppose all random variable have a (finite) common mean.

The sequence $\{X_n\}$ **converges** to Y **almost surely** if

$$\Pr\left(\lim_{n\to\infty} X_n = Y \right) = 1$$

2.3 Law of large numbers (LLN)

Let $X_1, X_2, \ldots$ be a sequence of iid random variables with finite mean.

The **Weak law of large numbers (WLLN)** states that

$$\overline{X}_n \xrightarrow{p} \mu \text{ as } n \to \infty$$

The **Strong law of large numbers (SLLN)** states that

$$\overline{X}_n \xrightarrow{a.s.} \mu \text{ as } n \to \infty$$

2.4 Central limit theorem (CLT)

Let $X_1, X_2, \ldots$ be a sequence of iid random variables with finite mean and variance.

Let $Z_n = \frac{\overline{X} - \mu}{\sigma\sqrt{n}}$, then as $n \to \infty$, $Z_n \xrightarrow{d} Y$ where $Y \sim \mathcal{N}(0, 1)$

2.5 Statistic

A **statistic** $t(X)$ is a function of data X or sample.

Let $X_1, \ldots, X_n \sim F_\theta$ and let T be a function of $\mathbf{X} = \{X_1, \ldots, X_n\}$. We call $T = T(\mathbf{X})$ a **sufficient statistic** for θ if it is a statistic and the conditional probability $\Pr(\mathbf{X}|T)$ does not depend on θ.

An **ancillary statistic** is a pivot (which is also a statistic).

Note that a statistic is NOT the parameter.

2.6 Statistical model

A **statistical model** is a set of assertions that directly or indirectly specify the probobility distribution of observed data.

3 Statistical Inference

{ Bayesian or frequentist approach }

$\hat{\theta}$ is a random variable with probability distribution known as the **sampling distribution**. The standard deviation of the sampling distribution is known as the **standard error**.

3.1 Pivotal quantity

A **pivotal quantity** or **pivot** is a function of observations and unobservable parameters whose probability distribution does not depend on the unknown parameters.

3.2 Method of moments

The k^{th} moment is defined as

$$M^{(k)}(0) = \mu_k = \mathrm{E}(X^k)$$

if $X_1, X_2, \ldots, X_n$ are i.i.d. random variables from a distribution, the k^{th} **sample moment** is defined as

$$\hat{\mu}_k = \frac{1}{n}\sum_{i=0}^{n} X_i^k$$

$\hat{\mu}_k$ can be viewed as a **unbiased** estimate of μ_k.

{ application }

{ MME - Method of moment esitimator }

3.3 Maximum likelihood estimate

{ definition }

The log likelihood function is defined by $l(|\theta) = \ln L(|\theta)$.

3.4 Properties

An estimator $\hat{\theta}$ is **consistent** if $\hat{\theta}_n \xrightarrow{p} \theta$ as $n \to \infty$.

An estimator $\hat{\theta}$ is **umbiased** if $\mathrm{E}(\hat{\theta}) = \theta$.

The **mean square error**(MSE) of an estimator $\hat{\theta}$ is

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{E}[(\hat{\theta} - \theta)^2] = \mathrm{Var}(\hat{\theta}) + [\mathrm{B}(\hat{\theta})]^2$$

The **efficiency** of $\hat{\theta}$ relative to $\tilde{\theta}$ is the ration of their variances.

$$\mathrm{eff}(\hat{\theta}, \tilde{\theta}) = \frac{\mathrm{Var}(\tilde{\theta})}{\mathrm{Var}(\hat{\theta})}$$

{ Sufficiency }

{ MLE - Maximum likelihood esitimator }

{ Theorem 9.2 from p451 MSA }

{ p310 of PSTSCU }

3.5 Exponential family

{ Exponential family model and its properties }

A model in exponential family can be expressed as

$$f(x|\theta) = \exp(\pi(\theta)T(x) - g(\theta) + b(x))$$

where π is called **natural parameters**, and T is called **natural sufficient statistics**.

Examples of exponential family models: normal, bernoulli, χ^2, exponential, Poisson, gamma, binomial, negative binomial, beta, inverse normal, ...

Fact: $\hat{\theta}_{\mathrm{MLE}}$ is a function of sufficient statistic for exponential family.

Fact: If X is destributed according to a full rank exponential family, then $T(x)$ is minimal sufficient.

{ fact about T and complete }

Examples of 1-param exponential family models: bernoulli, binomial, possion, exponential.

For 1-param exponential family, method of moment estimator is equal to maximum likelihood estimator.

3.6 Score and Fisher information

The **score function** $S(\theta)$ is defined by:

$$S(\theta) = S(x|\theta) = \frac{\partial l(x|\theta)}{\partial \theta}$$

Fact: $\mathrm{E}[S(\theta)] = 0$

The **Fisher information** $I(\theta)$ is defined by:

$$I(\theta) = \mathrm{E}\left[-\frac{\partial^2 l(x|\theta)}{\partial^2 \theta}\right]$$

Fact: $\mathrm{Var}(S(\theta)) = I(\theta)$

Fact: $\mathrm{Var}(\sum_1^n S(\theta)) = nI(\theta)$

3.7 wtf

$\sum_{i=1}^n T(X_i)$ is sufficient statistic.

including normal, the binomial, the Poisson, and the gamma.

{ Score function and fisher information }

Under i.i.d. sampling from a model with Fisher information $I(\theta)$ the Fisher information for a sample of size n is given by $nI(\theta)$

{ Theorem 6.5.1 of PSTSCU }

The observed Fisher information is given by

$$\hat{I}(s|\theta) = \frac{\partial^2 l(s|\theta)}{\partial^2 \theta}\Big|_{\theta} = \hat{\theta}(s)$$

where $\hat{\theta}(s)$ is MLE.

{ sufficient at p460 }

{ factorization criterion at p461 }

{ Ex 9.66 }

{ 9.4, 9.5 and 9.7 }

{ Maximum Likelihood Estimator }

{ Rao-Blackwell Theorem }

{ Complete Statistic }

{ Lehmann-Scheffe Lemma }

4 Lecture Topics

4.1 Monday, June 29, 2015

- converges in distribution
- converges in probability
- converges almost surely
- law of large numbers (weak and strong)
- central limit theorem
- { normal distribution theorem }

4.2 Monday, July 6, 2015

- statistic
- { probobliity model and statistical model }
- { statistical inference }
- { estimator }
- { principal of minimizaing MSE }
- { Method of moment estimator }
- { parameter space }
- { biased estimator / unbiased estimator }
- { consistant estimator }
- { MVUE - Minimum-variance unbiased estimator }

4.2.1 Estimator

An *estimator* $\hat{\theta}$ is a rule, often expressed as a formula, that tells how to calculate the value of an estimate based on the measurements contained in a sample.

The *bias* of a point estimator $\hat{\theta}$ is given by $B(\hat{\theta}) = E(\hat{\theta}) - \theta$.

A estimator $\hat{\theta}$ is unbiased if $B(\hat{\theta}) = 0$.

A estimator $\hat{\theta}$ is consistant if $\hat{\theta} \xrightarrow{p} \theta$.

A square loose function is defined by $\mathcal{L}(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$

The risk function (expected lost) is defined by $R(\hat{\theta}, \theta) = E[(\hat{\theta} - \theta)^2] = \text{MSE}$

Principal of minimizing MSE $\hat{\theta}_1$ is prefered $\hat{\theta}_2$, if $\text{MSE}(\hat{\theta}_1) \leq \text{MSE}(\hat{\theta}_2)$. $\hat{\theta}$ is optimal if $\text{MSE}(\hat{\theta}) \leq \text{MSE}(\hat{\theta}')$ for all $\hat{\theta}'$. { check this }

$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + (B(\hat{\theta}))^2$

4.2.2 Statistical Inference

estimation($\hat{\theta}$, a estimator of θ)

hypothesis testing estimation($\hat{\theta}$)

estimator is a statistic $\hat{\theta}(X_1, \ldots, X_n)$ estimator $\hat{\theta}$ is a statistic of data sample $(X_1, \ldots, X_n)$

the set of all unbiased estimator D_0

Let $\hat{\theta}$ be a point estimator for a parameter θ. Then $\hat{\theta}$ is an *unbiased estimator* if $\text{E}(\hat{\theta}) = \theta$. If $E(\hat{\theta}) \neq \hat{\theta}$, $\hat{\theta}$ is said to be *biased*.

The *mean square error* of a point estimator $\hat{\theta}$ is

$$\text{MSE}(\hat{\theta}) = \text{E}[(\hat{\theta} - \theta)^2] = \text{Var}(\hat{\theta}) + [\text{B}(\hat{\theta})]^2$$

Define *standard error* $\sigma_{\hat{\theta}} = \sqrt{\sigma^2_{\hat{\theta}}}$.

The error of estimation ε is the distance between an estimator and its target parameter. That is, $\varepsilon = |\hat{\theta} - \theta|$.

Statistical Model and Inference Method of Moment Estimator

4.3 Wednesday, July 8, 2015

4.3.1 Confidence interval

An **interval estimator**(confidence intervals) is a rule specifying the method for using the sample measurements to calculate two numbers that form the endpoints of the interval.

Given $\Pr(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U) = 1 - \alpha$, the probability $(1 - \alpha)$ is the **confidence coefficient**.

Let X be a random sample from a distribution that depends on a parameter θ. Let $g(X, \theta)$ be a random variable whose distribution is the same for all θ. Then g is called a **pivotal quantity** (or simply a pivot).

8.5, 8.6 and 8.8

- { pivotal method at p407 }
- Large Sample CI concerning μ
- Likelihood Inference
- Likelihood Principle
- Suffcient Statistics
- factorization theorem

- minimal sufficient statistic

Suffcient Statistics is not unique

4.4 Monday, July 13, 2015

{ sufficient at p460 }

{ factorization criterion at p461 }

{ Ex 9.66 }

{ 9.4, 9.5 and 9.7 }

{ Maximum Likelihood Estimator }

{ Rao-Blackwell Theorem }

{ Complete Statistic }

{ Ancillary Statistic }

{ Lehmann-Scheffe Lemma }

4.5 Wednesday, July 15, 2015

{ Exponential family model and its properties }

{ Score function and fisher information }

$I(\theta) = \text{Var}(S(\theta))$

4.6 Wednesday, July 22, 2015

{ Observed fisher information }

{ Expected fisher information }

{ likelihood inverse }

4.6.1 Likelihood Asymptotic

Asymptotic properties of MLE

Let

$$S(\theta) = \sum_{i}^{n} S_i(\theta)$$

1. $\frac{1}{\sqrt{n}} S(\theta) \xrightarrow{d} N(0, I(\theta))$
2. $\sqrt{n}(\theta_{MLE} - \theta) \rightarrow N(0, I(\theta)^{-1})$
3. $\hat{\theta}_{MLE} \xrightarrow{p} \theta_{MLE}$

4.6.2 Delta method

If $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, [I(\theta)]^{-1})$ then g us a differentiable function and $g'(\theta) \neq 0$ then

$$\sqrt{n}(g(\hat{\theta}) - g(\theta)) \xrightarrow{d} \mathcal{N}(0, [g'(\theta)]^2 [I(\theta)]^{-1})$$

.

For $g'(\theta) = 0$ and $g''(\theta) \neq 0$,

$$n(g(\hat{\theta}) - g(\theta)) \sim \mathcal{X}_1^2 \frac{1}{2} g''(\theta) \frac{1}{I(\theta)}$$

{ proof } { Final Question }

4.6.3 Cramer-Rao lower bound

Suppose θ is unknown parameter which is to be estimated from $X = (X_1, \dots, X_n)$ distributed according to $f_e(X)$ then for any given unbiased estimator $\hat{\theta}$ $\text{Var}(\hat{\theta}) \geq [nI(\theta)]^{-1}$.

If $\hat{\theta}_{MLE}$ is unbiased and $\hat{\theta}_{MLE}$ achieves CR lower bound then $\hat{\theta}_{MLE}$ is MVUE.

{ retype } in exponential family $\hat{\theta}_{MLE}$ is a fn of complete mss then by lena-sheffe lemma $\hat{\theta}_{MLE}$ is MVUE

but not all MLEs are unbiased therefore MVUE does not always achieve Cramer-Rao lower bound.

Bernulli θ $\hat{\theta}_{MLE} = \bar{X}$ $\text{Var}(\bar{x}) = [nI(\theta)]^{-1}$

Bionomial (n, θ) $\hat{\theta}_{MLE} = \bar{X}$

Poisson $\hat{\theta}_{MLE} = \bar{X}$

$\exp(\beta)$ $\hat{\theta}_{MLE} = \bar{X}$

Normal - { check }

Uniform - { check non-exist }

Gamma - { check }

Beta - { check }

4.7 Monday, July 27, 2015

{ Intro to Hypothesis Testing }

Null hypothesis

Alternate hypothesis

Significance level

p-value

type I and type II error

One-side and two sided test

power of test

Neyman-Pearson Lemma

Hardy-Weinberg Equilibrium

Poisson Dispersion Test

4.8 Wednesday, July 29, 2015

{ Neyman-Pearson Lemma }

{ Likelihood Ratio Test }

4.9 Monday, August 5, 2015

{ P-value }

{ Relationship between Confidence Interval and Hypothesis Testing }

5 TODO

section 7.2, 7.3

Example 8.1

$\mathcal{X}^2$ distribution, t distribution, F distribution

L^2 convergence

MGF, PDF

All exercises in section 8.2, read 8.3 and 8.4. Ex 8.36, 8.37. 8.38. Read section 9.3. try exercise 9.15 –9.36 and exercise in section 9.6.

6 Distributions

Notation	Name	pdf	cdf
$\mathcal{N}(\mu, \sigma^2)$	normal	-	-
$B(n,p)$	binomial		
	T-distribution		
	$\mathcal{X}^2$ distribution		

7 Abbreviations

r.v. random variable

i.i.d. independent and identically distributed

SD standard deviation

CI Confidence Interval

- $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$
- $X_n \xrightarrow{d} Y$ converges in distribution
- $X_n \xrightarrow{p} Y$ converges in probobility
- $X_n \xrightarrow{a.s.} Y$ converges almost surely

Critical values

{ l for log-likelihood functinon }

8 Definitions

A **random variable** is a real-valued function for which the domain is a sample space.

A **statistic** is a function of the observable random variables in a sample and known constants.

Variance $\mathrm{Var}(X) = E\left[(X-\mu)^2\right]$

9 Basic Topics

10 Acknowledgement

Thanks to *Summer Bian* on Facebook for lecture notes.