Midterm Exam Review

Bruce M. Boghosian

Tufts UNIVERSITY School of Arts and Sciences

Department of Mathematics

Tufts University

Outline

Likelihood and maximum likelihood

- Definition of *likelihood* $L(\lambda) = \prod_{j=1}^{n} f_X(x_j; \lambda)$
- Maximum likelihood estimator $\hat{\lambda}(\vec{x})$ gives most likely value for parameter λ.
- Estimation of parameters if form of pdf is known a priori
- Can be used for discrete or continuous pdfs, discrete or continuous parameters
- The log likelihood $\log L(\lambda)$ is often useful
- Gives only a single result, no confidence interval

Method of moments

- *Method of moments* is another way to creating estimators
- Equate s theoretical moments to s sample moments

$$E(Y) = \int dy\ f_Y(y;\theta_1,\ldots,\theta_s)y = \frac{1}{n}\sum_{j=1}^{n} y_j$$

$$E(Y^2) = \int dy\ f_Y(y;\theta_1,\ldots,\theta_s)y^2 = \frac{1}{n}\sum_{j=1}^{n} y_j^2$$

$$\vdots$$

$$E(Y^s) = \int dy\ f_Y(y;\theta_1,\ldots,\theta_s)y^s = \frac{1}{n}\sum_{j=1}^{n} y_j^s$$

- Yields s simultaneous equations for the s parameters.
- May be different from MLE (e.g., the uniform distribution)
- May be used in combination with MLE.

Confidence intervals and interval estimation

- Moment-generating functions for proof of *CLT*
- Say $\overline{y} = \frac{1}{n}\sum_{j=1}^{n} y_j$ is normally distributed with known σ
- Form *standardized random variables* $z_j = \frac{y_j - \overline{y}}{\sigma/\sqrt{n}}$
- Standardized r.v.s distributed like standard normal $f_Z(z)$
- *Z tables* defined so $\int_{z_\alpha}^{\infty} dz\, f_Z(z) = \alpha$
- Confidence intervals can be symmetric or asymmetric
- Prob $\left(Y \in \left[\mu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \mu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]\right) = 1 - \alpha$
- *Margin of error*: Half maximum width of a (usually 95%) confidence interval
- How large does a trial have to be to achieve a certain confidence?

Properties of estimators: Unbiasedness

- If MLE and MM estimators different, which is “correct”?
- Estimators themselves are random variables.
- Estimators as functions of random variables have means and variances.
- For $f_X(x;\theta)$, an *unbiased* estimator has $E(\hat{\theta}(\vec{X})) = \theta$.
- If an estimator is biased, but the bias vanishes as $n \to \infty$, we say that it is *asymptotically unbiased*.
- Sometimes you can fix biased estimators by applying a correction for finite n.

Properties of estimators: Efficiency

Bruce M. Boghosian

- Cumulative distribution functions
- Order statistics for distribution of max and min
- *Efficiency* of estimators: Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators for parameter θ. If $\mathsf{Var}(\hat{\theta}_1) < \mathsf{Var}(\hat{\theta}_2)$, we say that $\hat{\theta}_1$ is *more efficient* than $\hat{\theta}_2$.
- *Relative efficiency* of estimators: The *relative efficiency* of $\hat{\theta}_1$ with respect to $\hat{\theta}_2$ is $\frac{\mathsf{Var}(\hat{\theta}_2)}{\mathsf{Var}(\hat{\theta}_1)}$.
- Cramér-Rao bound: Absolute efficiency for estimators

Properties of estimators: Cramér-Rao bound

- Proof of the Cramér-Rao bound
- *Cauchy-Schwarz inequality*
- *Pearson correlation coefficient*
- Two forms of Cramér-Rao bound
  - First-derivative form
  - Second-derivative form

Properties of estimators: Sufficiency and consistency

- Sufficiency and consistency
- Sufficiency defined by factorization theorem
- Later we learned a second factorization theorem
- Consistency of estmators
- Chebyshev's Theorem for establishing consistency

Bayesian estimation

- Bayes Theorem and examples

$$\boxed{P\left(A_j \mid B\right) = \frac{P\left(B \mid A_j\right) P(A_j)}{\sum_{k=1}^{n} P\left(B \mid A_k\right) P(A_k)}.}$$

- Updating priors to create new posterior distributions
- Bayesian search strategy
- Bayesian estimation

$$g_\Lambda(\lambda \mid W = w_s) = \frac{f_W(w_s \mid \lambda) f_\Lambda(\lambda)}{\int_{-\infty}^{+\infty} d\xi \; f_W(w_s \mid \xi) f_\Lambda(\xi)}$$

Hypothesis testing

- State in terms of $z := \frac{\overline{y}-\mu_0}{\sigma/\sqrt{n}}$
- Let $y_1, \ldots, y_n$ be a random sample from a normal distribution for which σ is known.
- To test $H_0 : \ \mu = \mu_0$ versus $H_1 : \ \mu > \mu_0$ at the α level of significance, reject H_0 if $z \geq z_\alpha$.
- To test $H_0 : \ \mu = \mu_0$ versus $H_1 : \ \mu < \mu_0$ at the α level of significance, reject H_0 if $z \leq -z_\alpha$.
- To test $H_0 : \ \mu = \mu_0$ versus $H_1 : \ \mu \neq \mu_0$ at the α level of significance, reject H_0 if either $z \leq -z_{\alpha/2}$ or $z \geq +z_{\alpha/2}$.

Testing binomial data

- Large-sample test if

$$0 < np_0 - 3\sqrt{np_0(1-p_0)} < np_0 + 3\sqrt{np_0(1-p_0)} < n$$

- Otherwise, small-sample test is necessary
- Type I versus Type II errors
- Power curves

Generalized likelihood ratio

- Sets of parameters ω and Ω
- *Generalized Likelihood Ratio (GLR)* is then defined to be

$$\lambda = \frac{\max_{\theta \in \omega} L(\theta)}{\max_{\theta \in \Omega} L(\theta)}$$

- Generalization to many parameters is straightforward.
- Hypothesis testing with the GLR – the GLRT
- Reject H_0 if λ is below a certain threshold.

χ^2 distribution

Bruce M. Boghosian

- Using the sample variance for estimation
- Reviewed gamma and beta functions
- Reviewed gamma and beta distributions
- Sums of gamma distributed r.v.s are gamma distributed
- Sums of squares of normally distributed r.v.s are χ^2 distributed
- Orthogonal matrices – $\overline{Y}$ and S_Y^2 are independent
- Showed that $\frac{(n-1)S^2}{\sigma^2}$ is chi square distributed

F and T distributions

- Finding pdf of quotient
- Quotient of two χ^2 r.v.s is F distributed.
- **Def.:** Suppose that U and V are independent chi squared r.v.s with n and m degrees of freedom, respectively. A random variable of the form $\frac{V/m}{U/n}$ is said to have an *F distribution with m and n degrees of freedom*.
- Student T distribution $T_n = \frac{Z}{\sqrt{U/n}}$
- Derived pdf of T_n – fat tails for small samples.
- Learned about T tables in appendices

Normally distributed data – μ and σ^2 both unknown

- Interval estimation of μ using Z ratio
- Interval estimation of μ using T ratio
- Hypothesis testing using Z ratio
- Hypothesis testing using T ratio: One-sample T test
- Let s^2 denote the sample variance from n observations drawn from a normal distribution with unknown mean μ and unknown variance σ^2. Let $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$.
  - To test $H_0 : \sigma^2 = \sigma_0^2$ versus $H_1 : \sigma^2 > \sigma_0^2$ at the α level of significance, reject H_0 if $\chi^2 \geq \chi^2_{1-\alpha,n-1}$.
  - To test $H_0 : \sigma^2 = \sigma_0^2$ versus $H_1 : \sigma^2 < \sigma_0^2$ at the α level of significance, reject H_0 if $\chi^2 \leq \chi^2_{\alpha,n-1}$.
  - To test $H_0 : \sigma^2 = \sigma_0^2$ versus $H_1 : \sigma^2 \neq \sigma_0^2$ at the α level of significance, reject H_0 if χ^2 is either (a) $\leq \chi^2_{\alpha/2,n-1}$ or (b) $\geq \chi^2_{1-\alpha/2,n-1}$.

Summary

We reviewed the following topics:

- Introduction to estimation
- Confidence intervals
- Properties of estimators
- Bayesian estimation
- Hypothesis testing
- GLR and GLRT
- Small-sample statistics