Parameter Estimation
Fitting Probability Distributions

MIT 18.443

Dr. Kempthorne

Spring 2015

Outline

1. Statistical Models
   - Definitions
   - General Examples
   - Classic One-Sample Distribution Models

Statistical Models: Definitions

Def: Statistical Model

- Random experiment with sample space Ω.
- Random vector $X = (X_1, X_2, \dots, X_n)$ defined on Ω.
  $\omega \in \Omega$: outcome of experiment
  $X(\omega)$: data observations
- Probability distribution of X
  $\mathcal{X}$: Sample Space $= \{\text{outcomes } x\}$
  $\mathcal{F}_X$: sigma-field of measurable events
  $P(\cdot)$ defined on $(\mathcal{X}, \mathcal{F}_X)$
- Statistical Model
  $\mathcal{P} = \{\text{family of distributions }\}$

Statistical Models: Definitions

Def: Parameters / Parametrization

- Parameter θ identifies/specifies distribution in $\mathcal{P}$.
- $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$
- $\Theta = \{\theta\}$, the Parameter Space

Outline

1. Statistical Models
   - Definitions
   - General Examples
   - Classic One-Sample Distribution Models

Statistical Models: General Examples

Example 1. One-Sample Model

- $X_1, X_2, \ldots, X_n$ i.i.d. with distribution function $F(\cdot)$.
  E.g., Sample n members of a large population at random and measure attribute X
  E.g., n independent measurements of a physical constant μ in a scientific experiment.
- Probability Model: $\mathcal{P} = \{$distribution functions $F(\cdot)\}$
- Measurement Error Model:
  $X_i = \mu + \epsilon_i,\ i = 1, 2, \ldots, n$
  μ is constant parameter (e.g., real-valued, positive)
  $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ i.i.d. with distribution function $G(\cdot)$
  (G does not depend on μ.)

Statistical Models: General Examples

Example 1. One-Sample Model (continued)

- Measurement Error Model:

  $X_i = \mu + \epsilon_i,\ i = 1, 2, \dots, n$

  μ is constant parameter (e.g., real-valued, positive)

  $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ i.i.d. with distribution function $G(\cdot)$

  (G does not depend on μ.)

$\Longrightarrow X_1, \dots, X_n$ i.i.d. with distribution function

$F(x) = G(x - \mu)$.

$\mathcal{P} = \{(\mu, G) : \mu \in R, G \in \mathcal{G}\}$

where $\mathcal{G}$ is ...

Example: One-Sample Model

Special Cases:

- Parametric Model: Gaussian measurement errors
  $\{\epsilon_j\}$ are i.i.d. $N(0, \sigma^2)$, with $\sigma^2 > 0$, unknown.
- Semi-Parametric Model: Symmetric measurement-error distributions with mean μ
  $\{\epsilon_j\}$ are i.i.d. with distribution function $G(\cdot)$, where $G \in \mathcal{G}$, the class of symmetric distributions with mean 0.
- Non-Parametric Model: $X_1, \ldots, X_n$ are i.i.d. with distribution function $G(\cdot)$ where
  $G \in \mathcal{G}$, the class of all distributions
  on the sample space $\mathcal{X}$ (with center μ)

Statistical Models: Examples

Example 2. Two-Sample Model

- $X_1, X_2, \ldots, X_n$ i.i.d. with distribution function $F(\cdot)$
- $Y_1, Y_2, \ldots, Y_m$ i.i.d. with distribution function $G(\cdot)$
  E.g., Sample n members of population A at random and m members of population B and measure some attribute of population members.
- Probability Model: $\mathcal{P} = \{(F, G),\ F \in \mathcal{F},\ \text{and}\ G \in \mathcal{G}\}$
  Specific cases relate $\mathcal{F}$ and $\mathcal{G}$
- Shift Model with parameter δ
  - $\{X_i\}$ i.i.d. $X \sim F(\cdot)$, response under Treatment A.
  - $\{Y_j\}$ i.i.d. $Y \sim G(\cdot)$, response under Treatment B.
  - $Y \doteq X + \delta$, i.e., $G(v) = F(v - \delta)$
  - δ is the difference in response with Treatment B instead of Treatment A.

Example 3. Regression Models

n cases $i = 1, 2, \ldots, n$

- 1 Response (dependent) variable
  $$y_i, \; i = 1, 2, \ldots, n$$
- p Explanatory (independent) variables
  $$\mathbf{x}_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,p})^T, \; i = 1, 2, \ldots, n$$

Goal of Regression Analysis:

- Extract/exploit relationship between y_i and $\mathbf{x}_i$.

Examples

- Prediction
- Causal Inference
- Approximation
- Functional Relationships

Example: Regression Models

General Linear Model: For each case i, the conditional distribution $[y_i \mid x_i]$ is given by

$$y_i = \hat{y}_i + \epsilon_i$$

where

- $\hat{y}_i = \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_{i,p} x_{i,p}$
- $\beta = (\beta_1, \beta_2, \ldots, \beta_p)^T$ are p regression parameters (constant over all cases)
- ϵ_i Residual (error) variable (varies over all cases)

Extensive breadth of possible models

- Polynomial approximation ($x_{i,j} = (x_i)^j$, explanatory variables are different powers of the same variable $x = x_i$)
- Fourier Series: ($x_{i,j} = sin(jx_i)$ or $cos(jx_i)$, explanatory variables are different sin/cos terms of a Fourier series expansion)
- Time series regressions: time indexed by i, and explanatory variables include lagged response values.

Note: *Linearity* of $\hat{y}_i$ (in regression parameters) maintained with non-linear x.

Outline

Classic One-Sample Distribution Models

Poisson Distribution Model

- Theoretical Properties
  - Data consists of counts of occurrences.
  - Events are independent.
  - Mean number of occurrences stays constant during data collection.
  - Occurrences can be over time or over space (distance/area/volume)
- Examples
  - Liability claims on a specific phramaaceutical marketed by a drug company
  - Telephone calls to a business service call center
  - Individuals diagnosed with a specific rare disease in a community
  - Hits to a website
  - Automobile accidents in a particular locale/intersection

Poisson Distribution Model

- $X_1, X_2, \ldots, X_n$ i.i.d. *Poisson*(λ) distribution

  X = number of occurrences ("successes")

  λ = mean number of successes

  $$f(x \mid \lambda) = P[X = x \mid \lambda] = \frac{e^{-\lambda}\lambda^x}{x!}, \; x = 0, 1, \ldots$$

- Expected Value and Standard Deviation of *Poisson*(λ) distribution:

  $$\begin{aligned} E[X] &= \lambda \\ StDev[X] &= \sqrt{\lambda} \end{aligned}$$

- Moment-Generating Function of *Poisson*(λ):

  $$\begin{aligned} M_X(t) &= E[e^{tX}] = \textstyle\sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} e^{tx} \\ &= e^{-\lambda} \textstyle\sum_{x=0}^{\infty} \frac{[\lambda e^t]^x}{x!} \\ &= e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)} \end{aligned}$$

  $$E[X^k] = \frac{d^k M_X(t)}{dx^k}\Big|_{t=0}, \; k = 0, 1, 2, \ldots$$

Berkson (1966) Data: National Bureau of Standards experiment measruing 10, 220 times between successive emissions of alpha particles from americium 241.

Observed Rate = 0.8392 emissions per second.

- Example 8.2: Counts of emissions in 1207 intervals each of length 10 seconds. Model as 1207 realizations of *Poisson* distribution with mean $\hat{\lambda} = 0.8392 \times 10 = 8.392$.
- Problem 8.10.1: Observed counts in 12, 169 intervals each of length 1 second.

n	Observed
0	5267
1	4436
2	1800
3	534
4	111
5+	21

Model as 12169 realizations of Poisson Distribution with $\hat{\lambda} = 0.8392 \times 1 = 0.8392$

Issues in Parameter Estimation

Statistical Modeling Issues

- Different experiments yield different parameter estimates $\hat{\lambda}$.
- A parameter estimate has a **sampling distribution**: the probability distribution of the estimate over independent, identical experiments.
- Better parameter estimates will have sampling distributions that are closer to the true parameter.
- Given a parameter estimate, how well does the distribution specified by the estimate fit the data?
  To evaluate "Goodness-of-Fit" compare *observed* data to *expected* data.

Classic Probability Models

Normal Distribution

- Two parameters:

  μ : mean
  σ^2 : variance

- Probability density function:

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}, -\infty < x < \infty.$$

- Moment-generating function:

$$M_X(t) = E[e^{tX}] = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

- Theoretical motivation
  - Central Limit Theorem
  - Sum of large number of independent random variables

Classic Probability Models

Gamma Distribution

- Two parameters:

  λ : rate

  α : shape

- Probability density function:

$$f(x \mid \alpha, \lambda) = \frac{1}{\Gamma(\alpha)} \lambda^{\alpha} x^{\alpha-1} e^{-\lambda x},\ 0 < x < \infty.$$

- Moment-generating function:

  $M_X(t) = (1 - \frac{t}{\lambda})^{-\alpha}$

- Theoretical motivation
  - Cumulative waiting time to α successive events, which are i.i.d. *Exponential*(λ).

LeCam and Neyman (1967) Rainfall Data

Objectives of Distribution Modeling

Basic Objectives

- Direct model of distribution based on scientific theory.
- Data summary/compression.
- Simulating stochastic variables for systems analysis.

Modeling Objectives

- Apply well-developed theory of parameter estimation.
- Use straight-forward methdologies for implementing parameter estimation in new problems.
- Understand and apply optimality principles in parameter estimation.

Important Methodologies

- Method-of-Moments
- Maximum Likelihood
- Bayesian Approach

18.443 Statistics for Applications
Spring 2015