Probability and Statistics

Part 2. More Probability, Statistics and their Application

Chang-han Rhee

Stanford University

Sep 20, 2011 / CME001

Outline

Statistics
- Estimation Concepts
- Estimation Strategies

More Probability
- Expectation and Conditional Expectation
- Interchange of Limit
- Transforms

Simulation
- Monte Carlo Method
- Rare Event Simulation

Further Reference
- Classes at Stanford
- Books

Outline

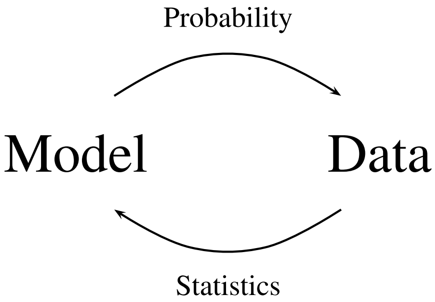
Statistics
- Estimation Concepts
- Estimation Strategies

More Probability
- Expectation and Conditional Expectation
- Interchange of Limit
- Transforms

Simulation
- Monte Carlo Method
- Rare Event Simulation

Further Reference
- Classes at Stanford
- Books

Probability and Statistics



Estimation

Making best guess of an unknown parameter out of sample data.

eg. Average height of west african giraffe

Estimator

An estimator (statistic) is a rule of estimation:

$$\hat{\theta}_n = g(X_1, \ldots, X_n)$$

Quality of an Estimator

- Bias

$$\mathbf{E}\hat{\theta} - \theta$$

- Variance

$$\text{var}(\hat{\theta})$$

- Mean Square Error (MSE)

$$\mathbf{E}[\hat{\theta} - \theta]^2 = (\text{bias})^2 + (\text{var})$$

Confidence Interval

Consider the sample mean estimator $\hat{\theta} = \frac{1}{n}S_n$. From the CLT,

$$\frac{S_n - n\mathbf{E}X_1}{\sqrt{n}} \stackrel{\mathcal{D}}{\to} \sigma N(0,1)$$

Rearranging terms, (note: this is not a rigorous argument)

$$\frac{1}{n}S_n \stackrel{\mathcal{D}}{\approx} \mathbf{E}X_1 + \frac{\sigma}{\sqrt{n}}N(0,1)$$

Outline

Maximum Likelihood Estimation

Finding most likely explanation.

$$\hat{\theta}_n = \arg\max_{\theta} f(x_1, x_2, \ldots, x_n|\theta) = f(x_1|\theta) \cdot f(x_2|\theta) \cdot f(x_n|\theta)$$

- Gold Standard: Gueranteed to be
- Often computationally challenging

Method of Moments

Matching the sample moment and the parametric moments.
If $\theta = (\theta_1, \ldots, \theta_k)$

$$\int x^j f_{\hat{\theta}_n}(x)dx = \frac{1}{n}\sum_{i=1}^{n} X_i^j \qquad \text{for } j = 1, \ldots, k$$

or

$$\sum x^j p_{\hat{\theta}_n}(x) = \frac{1}{n}\sum_{i=1}^{n} X_i^j \qquad \text{for } j = 1, \ldots, k$$

- Statistically, less efficient than MLE
- Often computationally efficient

Outline

Properties of Expectation

- Jensen's Inequality

$$g(\mathbf{E}X) \leq \mathbf{E}g(X) \qquad g(\cdot)\text{: convex}$$

- Markov's inequality

$$\mathbf{P}(|X| > x) \leq \frac{\mathbf{E}|X|}{x} \qquad x > 0$$

- Minkovsky's inequality

$$\left(\mathbf{E}|X+Y|^p\right)^{1/p} \leq \left(\mathbf{E}|X|^p\right)^{1/p} + \left(\mathbf{E}|Y|^p\right)^{1/p}$$

- Hölder's inequality

$$\mathbf{E}|XY| \leq \left(\mathbf{E}|X|^p\right)^{1/p}\left(\mathbf{E}|Y|^p\right)^{1/p} \qquad \text{for } 1/p + 1/q = 1$$

- If X and Y are independent,

$$\mathbf{E}g(X)h(Y) = \mathbf{E}g(X)\mathbf{E}h(Y)$$

Properties of Conditional Expectation

- Jensen's Inequality

$$g(\mathbf{E}[X|Y]) \leq \mathbf{E}[g(X)|Y] \qquad g(\cdot)\text{: convex}$$

- Markov's inequality

$$\mathbf{P}\big(|X| > x\big|Y\big) \leq \frac{\mathbf{E}\big[|X|\,\big|Y\big]}{x} \qquad x > 0$$

- Minkovsky's inequality

$$\big(\mathbf{E}\big[|X+Y|^p\big|Z\big]\big)^{1/p} \leq \big(\mathbf{E}\big[|X|^p\big|Z\big]\big)^{1/p} + \big(\mathbf{E}\big[|Y|^p\big|Z\big]\big)^{1/p}$$

- Hölder's inequality

$$\mathbf{E}\big[|XY|\big|Z\big] \leq \big(\mathbf{E}\big[|X|^p\big|Z\big]\big)^{1/p}\big(\mathbf{E}\big[|Y|^q\big|Z\big]\big)^{1/q} \qquad 1/p + 1/q = 1$$

Tower Property

Tower Property (Law of Iterated Expectation, Law of Total Expectation)

$$\mathbf{E}[X] = \mathbf{E}\big[\mathbf{E}[X|Y]\big]$$

i.e.,

$$\mathbf{E}[X] = \sum_{x \in S} \mathbf{E}[X|Y = y]\mathbf{P}(Y = y)$$

eg.

- $Y \sim \mathit{Unif}(0, 1)$ & $X \sim \mathit{Unif}(Y, 1)$. What is $\mathbf{E}X$?
- Mouse Escape

Bayes Rule

- The law of total probability:

$$\mathbf{P}(A) = \sum_i \mathbf{P}(A|B_i)\mathbf{P}(B_i)$$

- Bayes Rule

$$\mathbf{P}(A_i|B) = \frac{\mathbf{P}(B|A_i)\mathbf{P}(A_i)}{\sum_j \mathbf{P}(B|A_j)\mathbf{P}(A_j)}$$

where $A_1, A_2, \ldots, A_k$ is a disjoint partition of Ω.

More Properties of Conditional Expectation

$$
\begin{aligned}
\mathbf{E}(Xg(Y)|Y) &= g(Y)\mathbf{E}(X|Y) \\
\mathbf{E}(\mathbf{E}(X|Y,Z)|Y) &= \mathbf{E}(X|Y) \\
\mathbf{E}(X|Y) &= X, \qquad \text{if } X = g(Y) \text{ for some } g \\
\mathbf{E}(h(X,Y)|Y=y) &= \mathbf{E}h(X,y) \\
\mathbf{E}(X|Y) &= \mathbf{E}X, \qquad \text{if } X \text{ and } Y \text{ are independent}
\end{aligned}
$$

Outline

Monotone Convergence

Theorem (Monotone Convergence)

If $X_n \geq 0$ *and* $X_n \uparrow X_{n+1}$ *almost surely, then* $\mathbf{E}X_n \to \mathbf{E}X_\infty$.

Dominated Convergence

and bounded convergence as a corollary

Theorem (Dominated Convergence)

If $X_n \to X_\infty$ almost surely and $|X_n| \leq Y$ for all n and some Y such that $\mathbf{E}Y < \infty$, then $X_n \xrightarrow{L^1} X_\infty$.

Corollary (Bounded Convergence)

If $X_n \to X_\infty$ almost surely and $|X_n| \leq K$ for all n and some $K \in \mathbb{R}$, then $X_n \xrightarrow{L^1} X_\infty$.

and more

- Scheffe's Lemma
- Fatou's Lemma
- Uniform Integrability
- Fubini's Theorem

Outline

Moment Generating Function and Characteristic Function

Moment generating function and characteristic function chracterizes the distribution of the random variable.

- Moment Generating Function

$$M_X(\theta) = \mathbf{E}[\exp(\theta X)]$$

- Characteristic Function

$$\Phi_X(\theta) = \mathbf{E}[\exp(i\theta X)]$$

Outline

Monte Carlo Method

Computational algorithms that rely on repeated random sampling to compute their results.

Theoretical Bases

- Law of Large Numbers guarantees the convergence

$$\frac{1}{n}(\# X_i \in A) \to \mathbf{P}(X_1 \in A)$$

- Central Limit Theorem

$$\frac{1}{n}(\# X_i \in A) - \mathbf{P}(X_1 \in A) \stackrel{\mathcal{D}}{\approx} \frac{\sigma}{\sqrt{n}} N(0,1)$$

Outline

Challenges of Rare Event

Probability that a coin lands on its edge.

How many flips do we need to see at least one occurrence?

Importance Sampling (Change of Measure)

We can express the expectation of a random variable as an expectation of another random variable.

eg.
Two continuous random variable X and Y have density f_X and f_Y such that $f_Y(s) = 0$ implies $f_X(s) = 0$. Then,

$$\mathbf{E}g(X) = \int g(s) f_X(s) ds = \int g(s) \frac{f_X(s)}{f_Y(s)} f_X(s) ds = \mathbf{E}g(Y)L(Y)$$

where $L(X) = \frac{f_X(X)}{f_Y(X)}$.

$L(X)$ is called a likelihood ratio.

Outline

Probability

- Basic Probability
  STATS 116
- Stochastic Processes
  STATS 215, 217, 218, 219
- Theory of Probability
  STATS 310ABC

Statistics

- Intro to Statistics
  STATS 200
- Theory of Statistics
  STATS 300ABC

Application

- Applied Statistics
  STATS 191, 203, 208, 305, 315AB
- Stochastic Systems
  MS&E 121, 321
- Stochastic Control
  MS&E 322
- Stochastic Simulation
  MS&E 223, 323, STATS 362
- Little bit of Everything
  CME 308
- Econometrics, Finance, Bio and more
  http://explorecourses.stanford.edu

Outline

Books

- Sheldon Ross (2009). Introduction to Probability Models. Academic Press; 10th edition
- John A. Rice (2006). Mathematical Statistics and Data Analysis. Duxbury Press; 3rd edition
- Larry Wasserman (2004). All of Statistics : a concise course in statistical inference. Springer, New York