Math 408 - Mathematical Statistics

Lecture 8. Inequalities

February 4, 2013

Agenda

- Markov Inequality
- Chebyshev Inequality
- Hoeffding Inequality
- Cauchy-Schwarz Inequality
- Jensen Inequality
- Summary

Markov Inequality

Inequalities are useful for bounding quantities that might otherwise be hard to compute. They will be used in the large sample theory (next two Lectures) which is extremely important for statistical inference.

Markov Inequality

Let X be a non-negative random variable and suppose that $\mathbb{E}[X]$ exists. Then for any $a > 0$

$$\boxed{\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}}$$

Remark:

- This result says that the probability that X is much bigger than $\mathbb{E}[X]$ is small: Let

$$a = k\mathbb{E}[X]$$

Then

$$\mathbb{P}(X \geq k\mathbb{E}[X]) \leq \frac{1}{k}$$

Chebyshev Inequality

Chebyshev Inequality

Let X be a random variable with mean μ and variance σ^2. Then for any $a > 0$

$$\boxed{\mathbb{P}(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}}$$

Remarks:

- This result says that if σ^2 is small, then there is a high probability that X will not deviate much from μ.
- If $a = k\sigma$, then
$$\mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$
- If $Z = \frac{X-\mu}{\sigma}$, then
$$\mathbb{P}(|Z| \geq a) \leq \frac{1}{a^2}$$

Example

Suppose we test a prediction method on a set of n new test cases. Let

$$X_i = \begin{cases} 1, & \text{if the predictor is wrong;} \\ 0, & \text{if the predictor is right.} \end{cases}$$

Then

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

is the observed error rate. Let p be the true error rate. We hope that $\overline{X}_n \approx p$.

Question: Estimate the probability $\mathbb{P}(|\overline{X}_n - p| \geq \varepsilon)$

Answer:

$$\mathbb{P}(|\overline{X}_n - p| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2}$$

Hoeffding Inequality

Hoeffding inequality is similar in spirit to Chebyshev inequality but it is sharper. This is how it looks in a special case for Bernoulli random variables:

Hoeffding Inequality

Let $X_1, \ldots, X_n \sim \mathrm{Bernoulli}(p)$. *Then for any* $\varepsilon > 0$

$$\boxed{\mathbb{P}(|\overline{X}_n - p| \geq \varepsilon) \leq 2e^{-2n\varepsilon^2}}$$

Remark: Hoeffding inequality gives us a simple way to create a confidence interval for a binomial parameter p.

Definition

A $100(1-\alpha)\%$ confidence interval for a parameter p is an interval calculated from the sample $X_1, \ldots, X_n \sim \mathrm{Bernoulli}(p)$, which contains p with probability $1-\alpha$.

Example: Construct a $100(1-\alpha)\%$ confidence interval for p using Hoeffding inequality.

Answer: $\overline{X}_n \pm \sqrt{\frac{1}{2n} \ln\left(\frac{2}{\alpha}\right)}$

Cauchy-Schwarz and Jensen Inequalities

These are two inequalities on expected values that are often useful.

Cauchy-Schwarz Inequality

If X and Y have finite variances, then

$$\mathbb{E}[|XY|] \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}$$

Jensen Inequality

- *If g is convex $(x^2, e^x$, etc), then*

$$\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$$

- *If g is concave $(-x^2, \log x$, etc), then*

$$\mathbb{E}[g(X)] \leq g(\mathbb{E}[X])$$

Examples: $\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2$, $\mathbb{E}(1/X) \geq 1/\mathbb{E}[X]$, $\mathbb{E}[\log X] \leq \log \mathbb{E}[X]$.

Summary

- Markov inequality: If X is a non-negative random variable, then for any $a > 0$

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

- Chebyshev inequality: If X is a random variable with mean μ and variance σ^2, then for any $a > 0$

$$\mathbb{P}(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$$

- Hoeffding inequality: Let $X_1, \ldots, X_n \sim \mathrm{Bernoulli}(p)$, then for any $\varepsilon > 0$

$$\mathbb{P}(|\overline{X}_n - p| \geq \varepsilon) \leq 2e^{-2n\varepsilon^2}$$

- Cauchy-Schwarz inequality: If X and Y have finite variances, then

$$\mathbb{E}[|XY|] \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}$$

- Jensen Inequality:
  - If g is convex, then $\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$
  - If g is concave, then $\mathbb{E}[g(X)] \leq g(\mathbb{E}[X])$