EC381/MN308
Probability and Some Statistics

Yannis Paschalidis

yannisp@bu.edu, http://ionia.bu.edu/

Dept. of Manufacturing Engineering
Dept. of Electrical and Computer Engineering
Center for Information and Systems Engineering

Lecture 19 - Outline

1. **Inequalities (Markov, Chebyshev)**
2. **Weak Law of Large Numbers**
3. **Strong Law of Large Numbers**

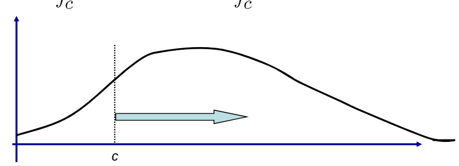
Estimating Probabilities

- A couple of inequalities allow us to estimate probabilities of events, based on statistics such as mean and variance
  - Markov Inequality
  - Chebyshev Inequality
- Used to derive some of the important limits we need later

Markov Inequality

- X is nonnegative RV ➔ $P(X \geq c) \leq \frac{E[X]}{c}, \quad c > 0$
- Proof

$$E[X] = \int_0^c x f_X(x)dx + \int_c^\infty x f_X(x)dx$$
$$\geq \int_c^\infty x f_X(x)dx \geq c\int_c^\infty f_X(x)dx = cP(X \geq c)$$



The Chebyshev Inequality

- Useful in estimating probabilities of events
- Chebyshev inequality: For any random variable X with mean μ and variance σ^2

$$P\{|X-\mu| \geq a\sigma\} \leq 1/a^2$$

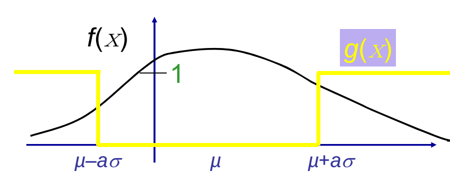
- The probability mass of X is concentrated around its mean
  - No more than $1/a^2$ is at a distance of $a\sigma$ (namely, a standard deviations) away from the mean!

Proof

- Let $g(x)$ be the indicator function of the values $|X - \mu| > a\sigma$:

$$g(x) = I(|x-\mu| \geq a\sigma) = \begin{cases} 0 & |x-\mu| \leq a\sigma \\ 1 & \text{otherwise} \end{cases}$$

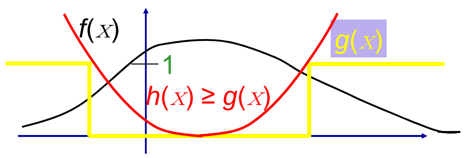
- Note: $P(|x-\mu| \geq a\sigma) = \int_{-\infty}^{\infty} g(x) f_X(x)dx$



Proof - 2

- Idea: Bound $g(x)$ by a quadratic centered at μ

$$g(x) \leq k(x-\mu)^2 = h(x)$$
$$g(\mu + a\sigma) = 1 \Rightarrow ka^2\sigma^2 = 1 \Rightarrow k = \frac{1}{a^2\sigma^2}$$



Proof - 3

- Since $h(x)$ is no less than $g(x)$,

$$\int_{-\infty}^{\infty} g(x) f_X(x)dx \leq \int_{-\infty}^{\infty} h(x) f_X(x)dx$$
$$\int_{-\infty}^{\infty} h(x) f_X(x)dx = \frac{1}{a^2\sigma^2}\int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)dx$$
$$= \frac{1}{a^2}$$
$$\Rightarrow \int_{-\infty}^{\infty} g(x) f_X(x)dx = P[|X-\mu| \geq a\sigma] \leq \frac{1}{a^2}$$

Example

- Quiz 7.2 variation: waiting time W is uniform in [0,10]. Estimate maximum probability that wait is at least 7.5 (we know the answer is 0.25!!)
  - Markov inequality:

$$P(W \geq 7.5) \leq \frac{E[W]}{7.5} = \frac{2}{3}$$

  - Chebyshev Inequality: mean is 5, variance is 100/12. 2.5 is .866 standard deviations from mean.

$$P\left\{|X-\mu| \geq a\sigma\right\} \leq 4/3$$

Dividing by 2 to get upper part yields 2/3 also!

Back to Average of n RVs - 5

- Var[Z_n] = Var[X_i]/n ➔

$$P[|Z_n - \mu| \geq a\frac{\sigma_X}{\sqrt{n}}] \leq \frac{1}{a^2}$$
$$\Rightarrow P[|Z_n - \mu| \geq a'] \leq \frac{\sigma_X^2}{na'^2} = \frac{Var(Z_n)}{a'^2}$$

  - Different statement of Chebyshev inequality

Chebyshev Inequality Provides Bound

- Compare to Gaussian $Q(\bullet)$ function
  - If Z_n were Gaussian, probability that Z_n is k standard deviations away from mean = 2 $Q(k)$
  - Chebyshev probability: ($1/k^2$)

Compare:	Chebyshev	Gaussian
k = 2	0.25	0.0455
k = 3	0.111	0.0027
k = 4	0.0625	0.0000633
k = 5	0.04	0.0000006

Bottom line: It is a bound for all types of RVs, but it is loose…

Weak Law of Large Numbers

- Numbers are not large…there are just a lot (large numbers) of them!
- Weak Law: If $X_1, \ldots, X_n$ are iid RVs with finite mean μ, then, for any $\epsilon > 0$

$$P(|\frac{1}{n}(X_1 + \ldots + X_n) - \mu| \geq \epsilon) \to 0 \text{ as } n \to \infty$$

  - Note: finite variance of X_i is not needed for the result, but proof is easier if we have finite variance (Chebyshev inequality!)

Simpler Version of WLLN

- Weak Law: If $X_1, \ldots, X_n$ are iid RVs with finite mean μ and finite variance σ_X^2 then, for any $\epsilon > 0$

$$\lim_{n\to\infty} P(|\frac{1}{n}(X_1 + \ldots + X_n) - \mu| < \epsilon) \to 1$$

Proof: Mean of $(X_1 + \ldots + X_n)/n = \mu$, Variance of $(X_1 + \ldots + X_n)/n$ is σ_X^2/n

Chebyshev: $$P(|\frac{1}{n}(X_1 + \ldots + X_n) - \mu| \geq \epsilon) \leq \frac{\sigma_X^2}{n\epsilon^2}$$

$$P(|\frac{1}{n}(X_1 + \ldots + X_n) - \mu| < \epsilon) \geq 1 - \frac{\sigma_X^2}{n\epsilon^2}$$

$$\to 1 \text{ as } n \to \infty$$

NOTE: True even if X_i are *uncorrelated* instead of independent!

Comments on WLLN

- Need existence of $E[X_i]$
- It is the key concept in statistics: observe an experiment many times, and one can estimate probabilities
- This is one concept of convergence of random variables
  - The sample mean Z_n "converges" to the true mean μ

$$\lim_{n\to\infty} P(|Z_n - \mu| < \epsilon) \to 1$$

  - There will be other concepts of "convergence" for RVs

Strong Law of Large Numbers

- If there is a weak version, there must be a stronger sibling…
- Strong Law of Large Numbers: Let $X_1, \ldots, X_n$ be iid random variables with finite mean μ. Then,

$$P(\lim_{n\to\infty} \frac{1}{n}(X_1 + \ldots + X_n) = \mu) = 1$$

- Note where the limit is located …

Interpreting the SLLN

- Let $Z_n = (X_1+X_2+\ldots+X_n)/n$ where the X_i are iid. RVs with finite mean μ
- For each fixed choice of ϵ,
  - P{event $|Z_n - \mu| < \epsilon$ occurs for infinitely many choices of n} = 1
  - P{event $|Z_n - \mu| > \epsilon$ occurs for finitely many choices of n}= 1
  - P{event $|Z_n - \mu| > \epsilon$ occurs for infinitely many choices of n}= 0

SLLN vs WLLN

SLLN	WLLN
$P(\lim_{n\to\infty} \frac{1}{n}(X_1 + \ldots + X_n) = \mu) = 1$	$\lim_{n\to\infty} P(\lvert\frac{1}{n}(X_1 + \ldots + X_n) - \mu\rvert < \epsilon) = 1$

- SLLN: for almost all outcomes, the sample mean approaches the true mean
- WLLN: The probability that the sample mean differs from the true mean by at least ϵ goes to zero

SLLN versus WLLN

- The Strong Law of Large Numbers says
  - P{event $|Z_n - \mu| > \varepsilon$ occurs for finitely many choices of n} = 1
  - P{event $|Z_n - \mu| > \varepsilon$ occurs for infinitely many choices of n} = 0
- The Weak Law does not require that the event $|Z_n - \mu| > \varepsilon$ occurs only for finitely many choices of n — the event might occur infinitely often: just so long as

$$P\{|Z_n - \mu| > \epsilon\} \to 0 \text{ as } n \to \infty$$

SLLN vs WLLN

- Subtle difference – SLLN says that, if you conduct the experiment and collect samples, the sample mean will converge to μ almost always
- WLLN does not say much about convergence of the sample mean for a specific sequence of samples
  - Refers primarily to the probability law, not to the specific samples collected for the experiment
- WLLN true if RVs are uncorrelated; SLLN requires independence

Application

- Probability estimation
  - X is a Bernoulli random variable, parameter p
  - Repeat experiment n times, observing X_i, $i = 1, \ldots, n$
  - Compute sample mean $Z_n = (X_1 + \ldots + X_n)/n$
  - SLLN: Z_n converges to p with probability 1

→ The observed frequency of an event converges to its true probability as the number of observations goes to ∞