INFO 6205 Program Structure and Algorithms

Nik Bear Brown

Chernoff bounds

Sampling Theorem

Linearity of Expectation

Union bound

Topics

- Chernoff bounds
- Sampling Theorem
- Linearity of Expectation
- Union bound

Basic Probability: Large Deviations

- Chebyshev inequality: For any random variable *X* with mean μ and standard deviation σ

$$\Pr[|X-\mu| \geq c] \leq \frac{\sigma^2}{c^2}$$

- Applies to any random variable
- Can be used to effectively bound large deviation for sum of pairwise independent random variables

Basic Probability

- Linearity of expectation: For any random variables X_1, X_2, …, X_n, we have
  - $E[\Sigma_i X_i] = \Sigma_i E[X_{i]}$
- Markov's inequality: For any random variable X
  - $\Pr[X \geq c] \leq E[X]/c$
- Union bound: For any sequence of events E_1, E_2, …, E_n, we have
  - $\Pr[U_i E_i] \leq \Sigma_i \Pr[E_i]$

Chernoff Bounds (above mean)

- Theorem. Suppose $X_1, \ldots, X_n$ are independent 0-1 random variables. Let $X = X_1 + \ldots + X_n$. Then for any $\mu \geq E[X]$ and for any $\delta > 0$, we have

sum of independent 0-1 random variables
is tightly centered on the mean

$$\Pr[X > (1+\delta)\mu] < \left[\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right]^{\mu}$$

- or

$$\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}$$

Chernoff Bounds (above mean)

- Pf. (cont)
  - Let $p_i = \Pr[X_i = 1]$. Then,

$$E[e^{t X_i}] \;=\; p_i e^t + (1 - p_i) e^0 \;=\; 1 + p_i (e^t - 1) \;\le\; e^{p_i (e^t - 1)}$$

for any $\alpha \ge 0$, $1+\alpha \le e^{\alpha}$

  - Combining everything:

$$\Pr[X > (1+\delta)\mu] \;\le\; e^{-t(1+\delta)\mu} \prod_i E[e^{t X_i}] \;\le\; e^{-t(1+\delta)\mu} \prod_i e^{p_i (e^t - 1)} \;\le\; e^{-t(1+\delta)\mu} \; e^{\mu (e^t - 1)}$$

previous slide

inequality above

$\Sigma_i \, p_i = E[X] \le \mu$

  - Finally, choose $t = \ln(1 + \delta)$. ▪

Chernoff Bounds (below mean)

- Theorem. Suppose $X_1, \ldots, X_n$ are independent 0-1 random variables. Let $X = X_1 + \ldots + X_n$. Then for any $\mu \leq E[X]$ and for any $0 < \delta < 1$, we have

$$\Pr[X < (1-\delta)\mu] \ < \ e^{-\delta^2 \mu / 2}$$

- or

$$\Pr[X < (1-\delta)\mu] \ \leq \ e^{-\delta^2 \mu / 2}$$

- or

$$\Pr[|X - \mu| \geq \delta\mu] \leq 2e^{-\mu\delta^2/3}$$

Chernoff Bound

- Let X_1, X_2, ..., X_n be n independent random variables in {0,1}
- For any nonnegative δ

$$\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}$$

- For any δ in [0,1]

$$\Pr[|X-\mu| \geq \delta\mu] \leq 2e^{-\mu\delta^2/3}$$

Coin Flips

$$n \geq \frac{1}{(p - \frac{1}{2})^2} \ln \frac{1}{\sqrt{\varepsilon}}.$$

Sampling Theorem

Sampling Theorem: Suppose we use independent, uniformly random samples to estimate p, the fraction of a population with some property. If the number of samples n we use satisfies

$$n \geq \frac{2+\epsilon}{\epsilon^2} \ln \frac{2}{\delta},$$

then we can assert that our estimate $\overline{X}$ satisfies

$$\overline{X} \in [p-\epsilon, p+\epsilon] \text{ with probability at least } 1-\delta.$$

Some comments:

- That range $[p-\epsilon, p+\epsilon]$ is sometimes called the *confidence interval.*
- Due to the slightly complicated statement of the bound, sometimes people will just write the slightly worse bounds

  $$n \geq \frac{3}{\epsilon^2} \ln \frac{2}{\delta},$$

  or even

  $$n \geq O\left(\frac{1}{\epsilon^2} \ln \frac{2}{\delta}\right).$$

- One beauty of the Sampling Theorem is that the number of samples n you need *does not depend on the size of the total population.* In other words, it doesn't matter how big the country is, the number of samples you need to get a certain accuracy and a certain confidence only depends on that accuracy and confidence.
- In the example we talked about earlier we were interested in accuracy $\epsilon = 2\%$ and confidence 95%, meaning $\delta = 1/20$. So the Sampling Theorem tells us we need at least

  $$n \geq \frac{2+.02}{(.02)^2} \ln \frac{2}{1/20} = 5050 \ln 40 \approx 18600.$$

  Not so bad: you only need to call 18600 or so folks! Er, well, actually, you need to get 18600 folks to respond. And you need to make sure that the events "person responds" and "person approves of the president" are independent. (Hmm... maybe being a pollster is not as easy as it sounds...)

Linearity of Expectation

$$E[X] = \sum_{j=0}^{\infty} j \Pr[X = j]$$

- Expectation. Given a discrete random variables X, its expectation E[X] is defined by:

$$E[X] \;=\; \sum_{j=0}^{\infty} j \cdot \Pr[X = j] \;=\; \sum_{j=0}^{\infty} j\,(1-p)^{j-1} p \;=\; \frac{p}{1-p} \sum_{j=0}^{\infty} j\,(1-p)^{j} \;=\; \frac{p}{1-p} \cdot \frac{1-p}{p^2} \;=\; \frac{1}{p}$$

j-1 tails   1 head

- Waiting for a first success. Coin is heads with probability p and tails with probability 1-p. How many independent flips X until first heads?

Expectation: Two Properties

$$E[X] = \sum_{j=0}^{\infty} j \cdot \Pr[X=j] = \sum_{j=0}^{1} j \cdot \Pr[X=j] = \Pr[X=1]$$

not necessarily independent

Linearity of expectation. Given two random variables X and Y defined over the same probability space, E[X + Y] = E[X] + E[Y].

- Decouples a complex calculation into simpler pieces.

Union bound

Union bound

The probability of a finite union of events is upper bounded by the sum of the probabilities of the individual events.

- Let A_{ki} denote that the observation vector $\mathbf{z}$ is closer to the symbol vector $\mathbf{s}_k$ than $\mathbf{s}_i$, when $\mathbf{s}_i$ is transmitted.
- $\Pr(A_{ki}) = P_2(\mathbf{s}_k, \mathbf{s}_i)$ depends only on $\mathbf{s}_i$ and $\mathbf{s}_k$
- Applying Union bounds yields

$$\boxed{P_e(m_i) \le \sum_{\substack{k=1 \\ k \ne i}}^{M} P_2(\mathbf{s}_k, \mathbf{s}_i)}$$

$$\boxed{P_E(M) \le \frac{1}{M} \sum_{i=1}^{M} \sum_{\substack{k=1 \\ k \ne i}}^{M} P_2(\mathbf{s}_k, \mathbf{s}_i)}$$