Lecture 2: Inequalities

Ziyu Shao

School of Information Science and Technology
ShanghaiTech University

March 9 & 11, 2020

Outline

Motivation

If you can not calculate a probability or expectation exactly, then you have three powerful strategies:

- Bounds (upper and lower bounds) on probability using inequalities.
- Approximations using limiting theorems
  - Poisson approximation: The Law of Small Numbers
  - Sample mean limit: The Law of Large Numbers
  - Normal approximation: The Central Limit Theorem
- Simulations using Monte Carlo

Outline

Cauchy-Schwarz Inequality

a, b.

$\cos\theta = \frac{a \cdot b}{\|a\| \cdot \|b\|}$.

$|\cos\theta| \le 1$

① $\forall t \in R$, $E[(Y-tX)^2] \ge 0$

$\Leftrightarrow E[(Y^2 - 2tXY + t^2X^2)] \ge 0$

$\Leftrightarrow f(t) = t^2 E[X^2] - 2t E(XY) + E(Y^2)$

$f(t) \ge 0$, $\forall t \in R$.

Theorem

For any r.v.s X and Y with finite variances,

$$|E(XY)| \le \sqrt{E(X^2)\,E(Y^2)}.$$

② $\Delta = (2E(XY))^2 - 4 \cdot E(X^2) \cdot E(Y^2) \le 0$

$\Rightarrow |E(XY)|^2 \le E(X^2) \cdot E(Y^2)$

$\Rightarrow |E(XY)| \le \sqrt{E(X^2) \cdot E(Y^2)}$

Revisit Correlation

By Cauchy-Schwarz inequality,

$$\left| E[(X-EX)(Y-EY)] \right| \le \sqrt{E\{(X-EX)^2\} \cdot E\{(Y-EY)^2\}}$$

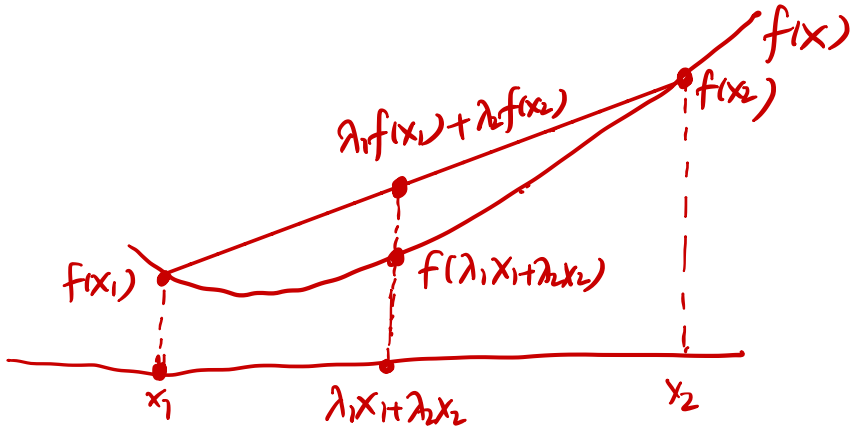
$$= \sqrt{Var(X) \cdot Var(Y)}$$

$$\Rightarrow \left| Corr(X,Y) \right| = \frac{\left| E[(X-EX)(Y-EY)] \right|}{\sqrt{Var(X) \cdot Var(Y)}} \le 1$$

Jensen's Inequality

If f is a convex function, $0 \leq \lambda_1, \lambda_2 \leq 1, \lambda_1 + \lambda_2 = 1$, then for any x_1, x_2,

$$f(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2).$$



Jensen's Inequality

$g''(x) < 0$ Concave.

Theorem

Let X be a random variable. If g is a convex function, then $E(g(X)) \geq g(E(X))$. If g is a concave function, then $E(g(X)) \leq g(E(X))$. In both cases, the only way that equality can hold is if there are constants a and b such that $g(X) = a + bX$ with probability 1.

Quick Examples

1^0. $g(x)=x^2$, $x\in\mathbb{R}$. Convex. $\mathbb{E}(X^2)\geq(\mathbb{E}X)^2$

2^0. $g(x)=\frac{1}{x}$, $x>0$. Convex $\mathbb{E}(\frac{1}{X})\geq\frac{1}{\mathbb{E}(X)}$

3^0. $g(x)=\log x$, $x>0$. Concave $\mathbb{E}(\log X)\leq\log(\mathbb{E}X)$

$g'(x)=\frac{1}{x}$

$g''(x)=-\frac{1}{x^2}<0$

Entropy

- Let X be a discrete r.v. whose distinct possible values are $a_1, a_2, \ldots, a_n$, with probabilities $p_1, p_2 \ldots, p_n$ respectively (so $p_1 + p_2 + \cdots + p_n = 1$).
- The *entropy* of X is defined as follows: $H(X) = \sum_{j=1}^{n} p_j \log_2(1/p_j)$.
- Using Jensen's inequality, show that the maximum possible entropy for X is when its distribution is uniform over $a_1, a_2, \ldots, a_n$, i.e., $p_j = 1/n$ for all j.
- This makes sense intuitively, since learning the value of X conveys the most information on average when X is equally likely to take any of its values, and the least possible information if X is a constant.

Proof

① Construct a random variable Y, s.t.

$$Y = \begin{cases} \frac{1}{p_1} & \text{w.p. } p_1 \\ \frac{1}{p_2} & \text{w.p. } p_2 \\ \vdots & \\ \frac{1}{p_n} & \text{w.p. } p_n \end{cases}$$

$$E(Y) = \frac{1}{p_1} \cdot p_1 + \frac{1}{p_2} \cdot p_2 + \cdots + \frac{1}{p_n} \cdot p_n = n$$

② $$H(X) = \sum_{j=1}^{n} p_j \log_2 \frac{1}{p_j} = E(\log_2 Y) \le \log_2 (E(Y)) = \log_2 n$$

equality holds $\Rightarrow p_1 = p_2 = \cdots = p_n = \frac{1}{n}$

$$H(\text{Unif}) = \log_2 n$$

Kullback-Leibler Divergence

Two perspectives

1°. information gain.

prior distribution r

posterior ..... p

Let $\mathbf{p} = (p_1, ..., p_n)$ and $\mathbf{r} = (r_1, ..., r_n)$ be two probability vectors (so each is nonnegative and sums to 1). Think of each as a possible PMF for a random variable whose support consists of n distinct values. The *Kullback-Leibler* divergence between $\mathbf{p}$ and $\mathbf{r}$ is defined as

$$D(\mathbf{p}, \mathbf{r}) = \sum_{j=1}^{n} p_j \log_2(1/r_j) - \sum_{j=1}^{n} p_j \log_2(1/p_j). = \sum_{j=1}^{n} p_j \log_2 \frac{p_j}{r_j}$$

Show that the Kullback-Leibler divergence is nonnegative.

2°. information loss.

desired distribution p.

approximation ..... r

Proof

① $D(P,r) = \sum_{j=1}^{n} p_j \log_2 \frac{p_j}{r_j} = -\sum_{j=1}^{n} p_j \log_2 \frac{r_j}{p_j}$.

② Construct a random variable Y, s.t.

$$P\left(Y = \frac{r_j}{p_j}\right) = p_j \quad , \quad j = 1, 2, \cdots, n$$

$$\Rightarrow \quad E(Y) = \sum_{j=1}^{n} p_j \cdot \frac{r_j}{p_j} = \sum_{j=1}^{n} r_j \overset{\text{valid pmf}}{=} 1$$

③ $D(P,r) = -E[\log_2 Y] \geq -\log_2[E(Y)] = -\log_2 1 = 0$

Norm Inequality

For a random variable X whose moment of order $r > 0$ is finite, we define the following norm

$$||X_|||_r = (\mathbb{E}(|X|^r))^{\frac{1}{r}}.$$

- **The Holder Inequality.** Let $\frac{1}{p} + \frac{1}{q} = 1$. If $\mathbb{E}(|X|^p), \mathbb{E}(|X|^q) < \infty$, then $|\mathbb{E}(XY)| \leq \mathbb{E}|XY| \leq ||X||_p \cdot ||X||_q$.
- **The Lyapunov Inequality.** For $0 < r \leq p$, $||X||_r \leq ||X||_p$.
- **The Minkowski Inequality**. Let $p \geq 1$, $\mathbb{E}(|X|^p), \mathbb{E}(|Y|^p) < \infty$, then $||X + Y||_p \leq ||X||_p + ||Y||_p$.

Markov's Inequality

Theorem

For any r.v. X and constant $a > 0$,

$$P(|X| \geq a) \leq \frac{E|X|}{a}.$$

Proof

(1) $Y = \frac{|X|}{a} \geq 0$

since $I(Y \geq 1) \leq Y$

	LHS		RHS	
$Y \geq 1$	1	$\leq$	Y	✓
$0 \leq Y < 1$	0	$\leq$	Y	✓

$\Rightarrow E[I(Y \geq 1)] \leq E(Y)$

$\Rightarrow P(Y \geq 1) \leq E(Y)$

$\Rightarrow P\left(\frac{|X|}{a} \geq 1\right) \leq E\left(\frac{|X|}{a}\right)$

$\Rightarrow P(|X| \geq a) \leq \frac{1}{a} E(|X|)$

Chebyshev's Inequality

$$P(|X-\mu| \geq a) = P(|X-\mu|^2 \geq a^2)$$

Markov's inequality

$$\leq \frac{1}{a^2} E(|X-\mu|^2) = \frac{1}{a^2}\sigma^2$$

Theorem

Let X have mean μ and variance σ^2. Then for any $a > 0$,

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

Proof

Chernoff's Inequality

$\forall t>0$, $P(X \geq a) = P(e^{tX} \geq e^{ta})$

$\leq \frac{E[e^{tX}]}{e^{ta}}$ → MGF of X.

Theorem

For any r.v. X and constants $a > 0$ and $t > 0$,

$$P(X \geq a) \leq \frac{E\left(e^{tX}\right)}{e^{ta}}.$$

Proof

Chernoff's Technique

1^0. $\forall t>0$, $P(X\geq a)\leq \frac{E[e^{tX}]}{e^{ta}}=f(t)$

$\Rightarrow P(X\geq a)\leq \inf_{t>0} f(t)$

Theorem

For any r.v. X and constants a,

$$P(X\geq a)\leq \inf_{t>0}\frac{E\left(e^{tX}\right)}{e^{ta}}$$

$$P(X\leq a)\leq \inf_{t<0}\frac{E\left(e^{tX}\right)}{e^{ta}}.$$

2^0. $\forall t<0$, $P(X\leq a)=P(e^{tX}\geq e^{ta})\leq \frac{E[e^{tX}]}{e^{ta}}=f(t)$

$\Rightarrow P(X\leq a)\leq \inf_{t<0} f(t)$

Proof

Example: Normal Distribution

① MGF of X is $M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$

$$P(X>a) \le \inf_{t>0} \frac{E[e^{tX}]}{e^{ta}} = \inf_{t>0} \frac{e^{\mu t+\frac{1}{2}\sigma^2 t^2}}{e^{ta}} = \inf_{t>0} e^{(\mu-a)t+\frac{1}{2}\sigma^2 t^2}$$

② $g(t) = \frac{1}{2}\sigma^2 t^2 + (\mu-a)t = \frac{1}{2}\sigma^2\left[t^2 + \frac{2(\mu-a)}{\sigma^2}t\right]$

Given $X \sim \mathcal{N}(\mu, \sigma^2)$, for arbitrary constant $a > \mu$, find the Chernoff bound on $P(X > a)$.

$$= \frac{1}{2}\sigma^2\left\{\left[t+\frac{\mu-a}{\sigma^2}\right]^2 - \frac{(\mu-a)^2}{\sigma^4}\right\}$$

$$\ge \frac{1}{2}\sigma^2 \cdot -\frac{(\mu-a)^2}{\sigma^4} = -\frac{(\mu-a)^2}{2\sigma^2} = g(t^*)$$

$$t^* = \frac{a-\mu}{\sigma^2} > 0,$$

③ $P(X>a) \le e^{g(t^*)} = e^{-\frac{(\mu-a)^2}{2\sigma^2}}$.

Solution

Example: Poisson Distribution

① $X \sim Pois(\lambda)$, $M_X(t) = e^{\lambda(e^t-1)}$

$$P(X>b) \le \inf_{t>0} \frac{E[e^{tX}]}{e^{tb}} = \inf_{t>0} e^{\lambda e^t - \lambda - bt}$$

② $g(t) = \lambda e^t - \lambda - bt$, $g'(t) = \lambda e^t - b$, $g''(t) = \lambda e^t > 0$

Given $X \sim Pois(\lambda)$, for arbitrary constant $b > 0$, find the Chernoff bound on $P(X > b)$.

$g'(t^*) = 0 \Rightarrow \lambda e^{t^*} = b \Rightarrow t^* = \ln\frac{b}{\lambda}$

$\Rightarrow g(t^*) = b - \lambda - b\ln\frac{b}{\lambda}$

③ $\Rightarrow P(X>b) \le e^{g(t^*)} = e^{b-\lambda-b\ln\frac{b}{\lambda}} = (\frac{\lambda}{b})^b \cdot e^{b-\lambda}$

Solution

Outline

X $E(X)$

$P(|X - E(X)| > \varepsilon)$ ↓

Hoeffding Lemma

$E(X)=0$, $a \le X \le b$

then if $a=b=0$, trivial. ($X=0$, LHS 1 = RHS 1)

thus we assume $a<0$, $b>0$, (in general)

Lemma

Let the random variable X satisfy $\mathbb{E}(X) = 0$ and $a \le X \le b$, where a and b are constants. Then for any $\lambda > 0$,

$$\mathbb{E}(e^{\lambda X}) \le e^{\frac{1}{8}\lambda^2(b-a)^2}.$$

Useful Analysis Tools

- Jensen's inequality: if f is convex, $0 \leq \lambda_1, \lambda_2 \leq 1, \lambda_1 + \lambda_2 = 1$, then for any x_1, x_2,

$$f(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2).$$

- Taylor's theorem or Taylor's expansion: If all the derivatives of a function $f(x)$ exist at point a, then for any positive integer k, there exist a real number θ between a and x such that

$$f(x) = f(a) + \cdots + \frac{f^{(k)}(a)}{k!}(x-a)^k + \frac{f^{(k+1)}(\theta)}{(k+1)!}(x-a)^{k+1}.$$

Proof

1°. $f(x)=e^{\lambda x}$, $x\in R$ is a convex function. $\Rightarrow \forall \alpha\in(0,1)$

$$f(\alpha\cdot a+(1-\alpha)\cdot b)\le \alpha f(a)+(1-\alpha)f(b)=\alpha\cdot e^{\lambda a}+(1-\alpha)e^{\lambda b}$$

For $x\in[a,b]$, let $\alpha=\frac{b-x}{b-a}\Rightarrow x=a\alpha+(1-\alpha)b$

$$\Rightarrow e^{\lambda x}\le \frac{b-x}{b-a}e^{\lambda a}+\frac{x-a}{b-a}e^{\lambda b}$$

$$\Rightarrow E[e^{\lambda x}]\le E[\frac{b-x}{b-a}]e^{\lambda a}+E[\frac{x-a}{b-a}]e^{\lambda b}$$

$$\overset{E(x)=0}{\Rightarrow} \frac{b}{b-a}e^{\lambda a}-\frac{a}{b-a}e^{\lambda b}=e^{\lambda a}[\frac{b}{b-a}-\frac{a}{b-a}e^{\lambda(b-a)}]$$

2°. Now let $\phi(t)=-\theta t+\ln(1-\theta+\theta e^t)$, for $\theta=\frac{-a}{b-a}>0$

$$\Rightarrow e^{\phi[\lambda(b-a)]}\overset{t=\lambda(b-a)}{=}e^{-\theta\lambda(b-a)}[1-\theta+\theta e^{\lambda(b-a)}]=e^{\lambda a}[1-\theta+\theta e^{\lambda(b-a)}]$$

$$[e^{\phi(t)}=e^{-\theta t}(1-\theta+\theta e^t)]=e^{\lambda a}[\frac{b}{b-a}-\frac{a}{b-a}e^{\lambda(b-a)}]$$

$$=\frac{b}{b-a}e^{\lambda a}-\frac{a}{b-a}e^{\lambda b}$$

3°. $E[e^{\lambda x}]\le e^{\phi[\lambda(b-a)]}$

Proof

4^0. We now focus on $\phi(t) = -\theta t + \ln(1-\theta+\theta e^t)$ $\quad \phi(0)=0$.

$$\phi'(t) = -\theta + \frac{\theta e^t}{1-\theta+\theta e^t}, \quad \phi'(0)=0$$

$$\phi''(t) = \frac{(1-\theta+\theta e^t)\cdot\theta e^t - \theta e^t(\theta e^t)}{(1-\theta+\theta e^t)^2}$$

$$= \frac{(1-\theta)\theta e^t}{(1-\theta+\theta e^t)^2}$$

$$\leq \frac{1}{4}.$$

[$m = 1-\theta > 0$, $n = \theta e^t > 0$, $\frac{m\cdot n}{(m+n)^2}$.]

Since $(m+n)^2 = (m-n)^2 + 4mn \geq 4mn$

$$\Rightarrow \frac{m\cdot n}{(m+n)^2} \leq \frac{1}{4}$$

Proof

5°. By Taylor's theorem, $\forall t > 0$, $\exists t' \in [0, t)$, s.t.

$$\begin{aligned}\phi(t) &= \phi(0) + \phi'(0) \cdot t + \phi''(t') \cdot \frac{1}{2} t^2 \\ &= 0 + 0 \cdot t + \frac{1}{2} t^2 \phi''(t') \\ &\le \frac{1}{8} t^2\end{aligned}$$

6°. Let $t = \lambda(b-a)$, $\phi[\lambda(b-a)] \le \frac{1}{8}\lambda^2(b-a)^2$

it follows that $E[e^{\lambda X}] \le e^{\phi[\lambda(b-a)]} \le e^{\frac{1}{8}\lambda^2(b-a)^2}$

Hoeffding Bound

Theorem

Let the random variables $X_1, X_2, \ldots, X_n$ be independent with $E(X_i) = \mu$, $a \leq X_i \leq b$ for each $i = 1, \ldots, n$, where a, b are constants. Then for any $\epsilon \geq 0$,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \epsilon\right) \leq 2e^{-\frac{2n\epsilon^2}{(b-a)^2}}.$$

Sample average.

exponentially decreasing.

Proof

① Let $Z_i = X_i - E(X_i) = X_i - \mu$, $E(Z_i) = 0$

$$Z = \frac{1}{n}\sum_{i=1}^n Z_i = \frac{1}{n}\sum_{i=1}^n X_i - \mu \qquad E(Z) = 0$$

For any $\lambda > 0$, $\varepsilon \geq 0$, we have

$$P(Z \geq \varepsilon) = P(e^{\lambda Z} \geq e^{\lambda\varepsilon}) \leq \frac{E[e^{\lambda Z}]}{e^{\lambda\varepsilon}} = \frac{E[e^{\lambda \cdot \frac{1}{n}\sum_{i=1}^n Z_i}]}{e^{\lambda\varepsilon}}$$

Z_i independent

$$= e^{-\lambda\varepsilon} \cdot \prod_{i=1}^n E[e^{\frac{1}{n}\cdot\lambda Z_i}] = e^{-\lambda\varepsilon} \cdot \prod_{i=1}^n E[e^{\lambda\cdot\frac{1}{n}Z_i}]$$

Since $E(\frac{1}{n}Z_i) = \frac{1}{n}E(Z_i) = 0$, $a \leq X_i \leq b \Rightarrow a-\mu \leq X_i - \mu = Z_i \leq b-\mu$.

$$\Rightarrow \frac{a-\mu}{n} \leq \frac{1}{n}Z_i = Z \leq \frac{b-\mu}{n}.$$

By Hoeffding Lemma, $\forall i = 1, 2, \ldots, n$,

$$E[e^{\lambda\cdot\frac{1}{n}Z_i}] \leq e^{\frac{1}{8}\lambda^2\cdot[\frac{b-\mu}{n} - \frac{a-\mu}{n}]^2} = e^{\frac{1}{2}\lambda^2\cdot\frac{(b-a)^2}{n^2}}$$

② $$P(Z \geq \varepsilon) \leq e^{-\lambda\varepsilon} \cdot \prod_{i=1}^n e^{\frac{1}{8}\lambda^2\cdot\frac{(b-a)^2}{n^2}}$$

$$= e^{-\lambda\varepsilon + \frac{1}{8}\lambda^2\cdot\frac{1}{n}(b-a)^2}$$

Proof

③ $\forall \lambda > 0$, $P(\bar{X} \geq \varepsilon) \leq e^{-\lambda\varepsilon + \frac{1}{8n}\lambda^2(b-a)^2}$

$\Rightarrow P(\bar{X} \geq \varepsilon) \leq \inf_{\lambda > 0} e^{-\lambda\varepsilon + \frac{1}{8n}\lambda^2(b-a)^2}$

Let $g(\lambda) = -\lambda\varepsilon + \frac{1}{8n}\lambda^2(b-a)^2$, $g'(\lambda) = -\varepsilon + \frac{\lambda}{4n}(b-a)^2$ [$g'(\lambda^*) = 0$

$g''(\lambda) = \frac{1}{4n}(b-a)^2 > 0$ $\Rightarrow \lambda^* = \frac{4n\varepsilon}{(b-a)^2}$]

$\Rightarrow g(\lambda^*) = -\lambda^*\varepsilon + \frac{1}{8n}\cdot(\lambda^*)^2(b-a)^2 = \frac{-2n\varepsilon^2}{(b-a)^2}$

④ thus $P(\bar{X} \geq \varepsilon) \leq e^{-\frac{2n\varepsilon^2}{(b-a)^2}}$

$\Rightarrow P\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu \geq \varepsilon\right) \leq e^{-\frac{2n\varepsilon^2}{(b-a)^2}}$

⑤ Applying the same argument. $P(\bar{X} \leq -\varepsilon) = P\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu \leq -\varepsilon\right)$

$\leq e^{-\frac{2n\varepsilon^2}{(b-a)^2}}$

Proof

(6) $P\left(\left|\frac{1}{n}\sum_{i=1}^{n}x_i-\mu\right|\geq\varepsilon\right)=P\left(\frac{1}{n}\sum_{i=1}^{n}x_i-\mu\geq\varepsilon\right)+P\left(\frac{1}{n}\sum_{i=1}^{n}x_i-\mu\leq-\varepsilon\right)$

$\leq 2e^{-\frac{2n\varepsilon^2}{(b-a)^2}}$

More General Hoeffding Bound

Theorem

Let the random variables $X_1, X_2, \ldots, X_n$ be independent, with $a_k \leq X_k \leq b_k$ for each k, where a_k, b_k are constants. Let $S_n = \sum_{k=1}^{n} X_k$ and let $\mu = \mathbb{E}(S_n)$. Then for any $t \geq 0$,

$$\mathbb{P}(|S_n - \mu| \geq t) \leq 2e^{-\frac{2t^2}{\sum_{k=1}^{n}(b_k - a_k)^2}}.$$

$P\left(\left|\frac{S_n}{n} - \frac{\mu}{n}\right| \geq \frac{t}{n}\right)$

Application: Parameter Estimation

Instead of predicting a single value for the parameter, we given an interval that is likely to contain the parameter:

Definition

A $1-\delta$ confidence interval for a parameter p is an interval $[\hat{p}-\epsilon, \hat{p}+\epsilon]$ such that

$$Pr\left(p \in [\hat{p}-\epsilon, \hat{p}+\epsilon]\right) \geq 1-\delta.$$

$\hat{p}$: estimation of p.

$\hat{p}-\epsilon \leq p \leq \hat{p}+\epsilon \Leftrightarrow -\epsilon \leq p-\hat{p} \leq \epsilon \Leftrightarrow |p-\hat{p}| \leq \epsilon$

$\delta = 0.05$ 95%.

$\Leftrightarrow Pr(|\hat{p}-p| \leq \epsilon) \geq 1-\delta$

$\Leftrightarrow Pr(|\hat{p}-p| > \epsilon) < \delta$

Application: Parameter Estimation

1^0. $E(\hat{p}) = \frac{1}{N}E(X_1+\cdots+X_N) = \frac{1}{N}\cdot Np = p$ Unbiased estimation

2^0. $0 \le X_i \le 1$, $a=0, b=1$, by Hoeffding Bound,

$$Pr(|\hat{p}-p| \ge \varepsilon) = Pr\left(\left|\frac{1}{N}\sum_{i=1}^{N} X_i - p\right| \ge \varepsilon\right) \le 2e^{-2N\varepsilon^2}$$

Tossing a coin with probability p landing heads and probability $1-p$ landing tails. p is unknown and we need to estimate its value from experiments results. We toss such coin N times, Let $X_i = 1$ if the ith result is head, otherwise 0. We estimate p by using $\hat{p} = \frac{X_1+\ldots+X_N}{N}$. Find the confidence interval for p.

3^0. Set $2e^{-2N\varepsilon^2} = \delta \Rightarrow \varepsilon = \sqrt{\frac{\ln(\frac{2}{\delta})}{2N}}$

4^0. $Pr(|\hat{p}-p| \ge \varepsilon) \le \delta \Rightarrow Pr(|\hat{p}-p| < \varepsilon) > 1-\delta$

$\Rightarrow Pr(p \in (\hat{p}-\varepsilon, \hat{p}+\varepsilon)) > 1-\delta$, $\hat{p} = \frac{1}{N}\sum_{i=1}^{N} X_i$, $\varepsilon =$

$(\hat{p}-\varepsilon, \hat{p}+\varepsilon)$: confidence interval

Solution

5°. $\forall \delta > 0$, w.p. $1-\delta$, $|\bar{p}-p| < \sqrt{\frac{\ln(\frac{2}{\delta})}{2N}}$

tradeoff among ε, δ, N.

$N\uparrow$, $\delta\uparrow$, $\varepsilon\downarrow$

Application: Monte Carlo Method for Estimation π

① $\frac{355}{113}$. (china). Buffon's needle.

② circle : $\{(x,y) : x^2+y^2 \le 1\}$

square : $[-1,1] \times [-1,1]$

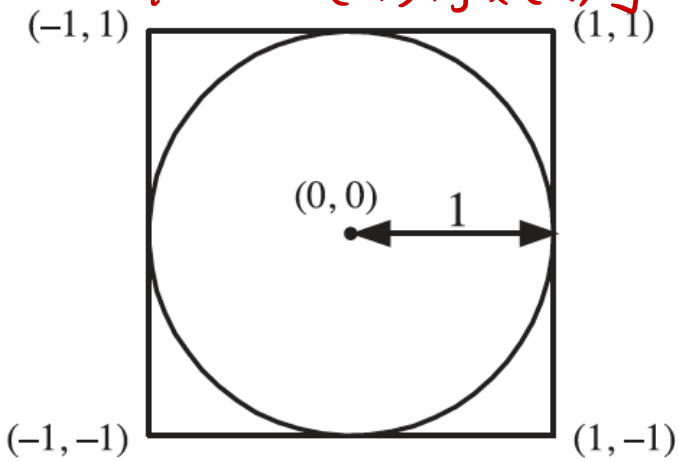


Figure 11.1: A point chosen uniformly at random in the square has probability $\pi/4$ of landing in the circle.

③ Pr(point landing in circle) = $\frac{\text{area(circle)}}{\text{area(square)}}$

$= \frac{\pi}{4}$.

Application: Monte Carlo Method for Estimation π

④ $\pi = 4\Pr(\text{point landing in the circle})$.

⑤. Z_i = indicator of the i^{th} point chosen uniformly at random in the square landing in the circle. , $Z_i = 1$ or 0

$P(Z_i = 1) = \frac{\pi}{4}$, Z_i . i.i.d. $E(Z_i) = \frac{\pi}{4} = \mu$.

$0 \le Z_i \le 1$.

Suppose we run the experiment m times,

Let $W = \frac{1}{m}\sum_{i=1}^{m} Z_i$, $E(W) = E(Z_i) = \frac{\pi}{4}$.

in fact, by SLLN. $W \longrightarrow E(Z_i) = \frac{\pi}{4}$. w.p.1.

$\Rightarrow 4W \longrightarrow \pi$. w.p.1.

$\hat{\pi} = 4W$ is a M.C. estimate of π.

Application: Monte Carlo Method for Estimation π

⑥. $P(|\hat{\pi}-\pi| \geq \varepsilon) = P(|W-\frac{\pi}{4}| \geq \frac{\varepsilon}{4})$

$= P(|\frac{1}{m}\sum_{i=1}^{m} Z_i - \mu| \geq \frac{\varepsilon}{4})$

$\leq$ Hoeffding Bound $2e^{-\frac{2m(\frac{\varepsilon}{4})^2}{(1-0)^2}} = 2e^{-\frac{1}{8}m\varepsilon^2}$

Let $\delta = 2e^{-\frac{1}{8}m\varepsilon^2} \Rightarrow \varepsilon = \sqrt{\frac{8\ln(\frac{2}{\delta})}{m}}$

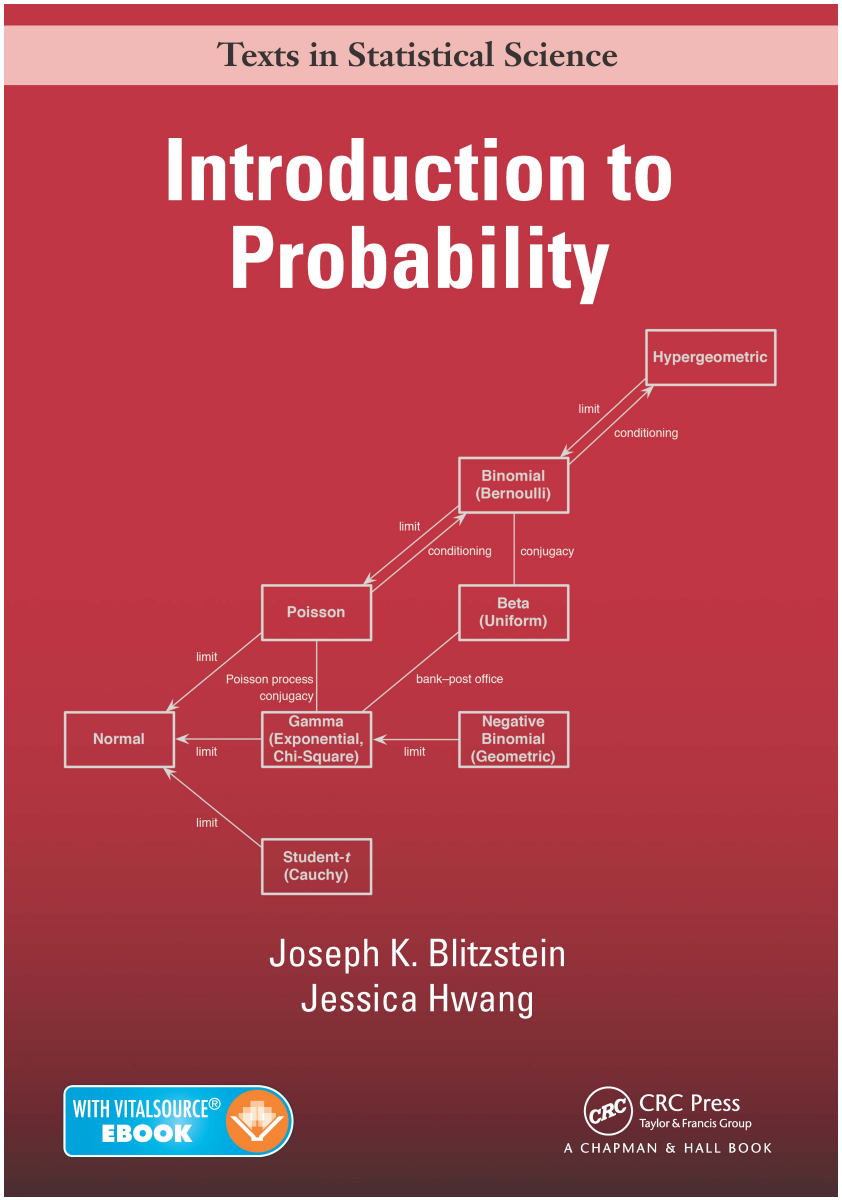
$\Rightarrow Pr(\pi \in (\hat{\pi} - \sqrt{\frac{8\ln(\frac{2}{\delta})}{m}}, \hat{\pi} + \sqrt{\frac{8\ln(\frac{2}{\delta})}{m}})) > 1-\delta$

Advanced Topics

- From independent case to dependent case
- Martingale inequalities
- Logarithmic Sobolev inequalities
- Transportation method

Outline

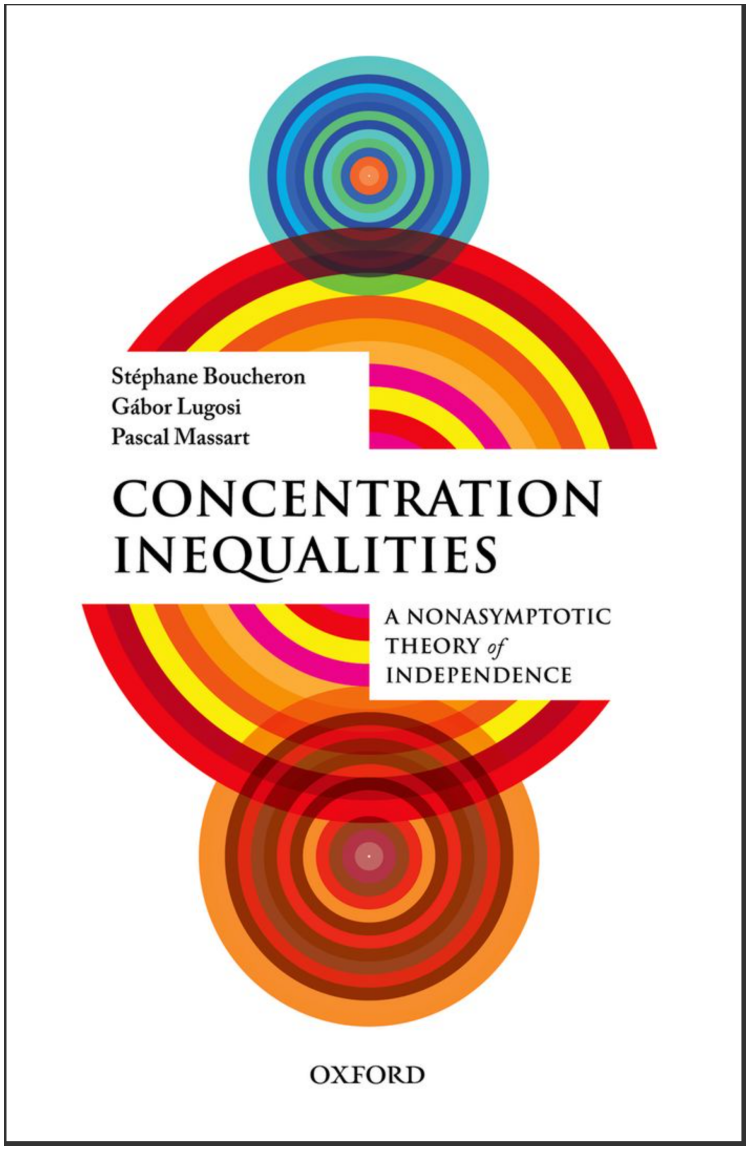
BH



Joseph K. Blitzstein & Jessica Hwang

- Introduction to Probability
- Chapman & Hall/CRC, 2014.
- Chapman & Hall/CRC, 2019.
- Chapter 10

BLM



Stephane Boucheron & Gabor Lugosi & Pascal Massart

- Concentration Inequalities
- Oxford, 2013