Basic Concentration Inequalities

Based on Section 2 of

Concentration Inequalities: A Nonasymptotic Theory of Independence

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart

Presented by Glenn Bevilacqua

Outline

- Markov's inequality
- Chebyshev's inequality
- Cramér-Chernoff method
- Heoffding's inequality
- Multiplicative Chernoff bound

Markov's Inequality

Theorem 1 (Markov's Inequality)

Given a non-negative random variable Y, for all $t \geqslant 0$,

$$\Pr[Y \geqslant t] \leqslant \frac{\mathbb{E}[Y]}{t}$$

Proof:

The following holds by construction.

$$t \cdot \mathbb{I}[Y \geqslant t] \leqslant Y \cdot \mathbb{I}[Y \geqslant t] \tag{1}$$

Apply expectation to both sides.

$$\mathbb{E}[t \cdot \mathbb{I}[Y \geqslant t]] \leqslant \mathbb{E}[Y \cdot \mathbb{I}[Y \geqslant t]] \tag{2}$$

$$\Pr[Y \geqslant t] \leqslant \frac{\mathbb{E}[Y \cdot \mathbb{I}[Y \geqslant t]]}{t} \leqslant \frac{\mathbb{E}[Y]}{t} \tag{3}$$

Final inequality follows from non-negativity of Y.

Markov's Inequality

Theorem 2 (Generalized Markov's Inequality)

Given a random variable $Y \in I$, where $I \subseteq \mathbb{R}$, if for all $t \in I$, $\varphi(t) \geqslant 0$ then,

$$\Pr[\varphi(Y) \geqslant \varphi(t)] \leqslant \frac{\mathbb{E}[\varphi(Y)]}{\varphi(t)}$$

Proof:

The following holds by construction.

$$\varphi(t) \cdot \mathbb{I}[\varphi(Y) \geqslant \varphi(t)] \leqslant \varphi(Y) \cdot \mathbb{I}[\varphi(Y) \geqslant \varphi(t)] \tag{5}$$

Apply expectation to both sides.

$$\varphi(t) \cdot \Pr[\varphi(Y) \geqslant \varphi(t)] \leqslant \mathbb{E}[\varphi(Y) \cdot \mathbb{I}[\varphi(Y) \geqslant \varphi(t)]] \tag{6}$$

By non-negativity of $\varphi(t)$.

$$\Pr[\varphi(Y) \geqslant \varphi(t)] \leqslant \frac{\mathbb{E}[\varphi(Y) \cdot \mathbb{I}[\varphi(Y) \geqslant \varphi(t)]]}{\varphi(t)} \leqslant \frac{\mathbb{E}[\varphi(Y)]}{\varphi(t)} \tag{7}$$

Chebyshev's Inequality

Theorem 3 (Chebyshev's Inequality)

Given a random variable Z, for all $t > 0$,

$$\Pr[|Z - \mathbb{E}[Z]| \geqslant t] \leqslant \frac{\mathrm{Var}(Z)}{t^2}$$

Proof:

The following holds by construction.

$$t^2 \cdot \mathbb{I}[(Z - \mathbb{E}[Z])^2 \geqslant t^2] \leqslant (Z - \mathbb{E}[Z])^2 \cdot \mathbb{I}[(Z - \mathbb{E}[Z])^2 \geqslant t^2] \tag{1}$$

Apply expectation to both sides.

$$t^2 \cdot \Pr[(Z - \mathbb{E}[Z])^2 \geqslant t^2] \leqslant \mathbb{E}[(Z - \mathbb{E}[Z])^2 \cdot \mathbb{I}[(Z - \mathbb{E}[Z])^2 \geqslant t^2]] \tag{2}$$

By non-negativity of $(Z - \mathbb{E}[Z])^2$.

$$\Pr[|Z - \mathbb{E}[Z]| \geqslant t] \leqslant \frac{\mathbb{E}[(Z - \mathbb{E}[Z])^2]}{t^2} = \frac{\mathrm{Var}(Z)}{t^2} \tag{3}$$

Chebyshev's Concentration Bound

Theorem 4 (Chebyshev's Concentration Bound)

Let $Z = X_1 + \ldots + X_n$ where the X_i's are independent then,

$$\Pr[|Z - \mathbb{E}[Z]| \geqslant t] \leqslant \frac{\mathsf{Var}(Z)}{t^2}$$

$$\Pr\left[\frac{1}{n}\left|\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i])\right| \geqslant t\right] \leqslant \frac{\sum_{i=1}^{n}\mathsf{Var}(X_i)}{n^2t^2}$$

The Cramér-Chernoff Method

Let Z, be a real-valued random variable. For $\lambda \geqslant 0$, using (Generalized) Markov's inequality with $\varphi(t) = \exp(\lambda t)$ gives,

$$\Pr[Z \geqslant t] = \Pr[\exp(\lambda Z) \geqslant \exp(\lambda t)] \leqslant \frac{\mathbb{E}[\exp(\lambda Z)]}{\exp(\lambda t)}$$

Since inequality holds for all $\lambda \geqslant 0$, choose λ to minimize the upper bound.

Let,

$$\psi_Z(\lambda) = \log \mathbb{E}[\exp(\lambda Z)] \qquad \text{(Moment generating function)}$$

$$\psi_Z^*(t) = \sup_{\lambda \in \mathbb{R}} (\lambda t - \psi_Z(\lambda)) \qquad \text{(Cramér transform of Z)}$$

then,

$$\Pr[Z \geqslant t] \leqslant \exp(-\psi_Z^*(t))$$

The Cramér-Chernoff Method (Bernoulli)

Let Y be a Bernoulli random variable with probability of success p. Denote $Z = Y - p$. Then,

$$\psi_Z(\lambda) = \log \mathbb{E}[\exp(\lambda Z)] = \log(pe^{\lambda} + 1 - p) - \lambda p$$
$$\psi_Z^*(t) = \sup_{\lambda \in \mathbb{R}} (\lambda t - \psi_Z(\lambda)) = \sup_{\lambda \in \mathbb{R}} (\lambda t - \log(pe^{\lambda} + 1 - p) - \lambda p)$$

Differentiate and solve for λ,

$$0 = (t - p - p\exp(\lambda))/(p\exp(\lambda) - p + 1)) \quad (1)$$
$$\lambda = \log \frac{(1-p)(p+t)}{p(1-p-t)} \quad (2)$$

$$\psi_Z^*(t) = (1-p-t)\log\frac{1-p-t}{1-p} + (p+t)\log\frac{p+t}{p}$$

for $t \in (0, 1-p)$.

Simplifying...

KL-Divergence: $D(x\|y) = (1-x)\log\frac{(1-x)}{(1-y)} + x\log\frac{x}{y}$

$$\psi_Z^*(t) = (1-p-t)\log\frac{1-p-t}{1-p} + (p+t)\log\frac{p+t}{p} \tag{1}$$

Let $a = t + p$

$$\psi_Z^*(a) = (1-a)\log\frac{1-a}{1-p} + a\log\frac{a}{p} \tag{2}$$

$$\psi_Z^*(a) = D(a\|p) \tag{3}$$

Recall... $Z = Y - p$ where Y is the Bernoulli random variable.

$$\Pr[Z \geqslant t] = \Pr[Y \geqslant p + t] \leqslant \exp(-\psi_Z^*(t+p)) = \exp(-D(p+t\|p))$$

Note: A upper bound follows for using $Z = p - Y$.

$$\Pr[Z \geqslant t] = \Pr[-Y \geqslant -p + t] = \Pr[Y \leqslant p - t] \leqslant \exp(-D(p-t\|p))$$

Sums of Independent random variables

Let $Z = X_1 + \ldots + X_n$ where the X_i's are independent and identically distributed (i.i.d.) then,

$$\psi_Z(\lambda) = \log \mathbb{E}[\exp(\lambda \sum_{i=1}^{n} X_i)] = \log \prod_{i=1}^{n} \mathbb{E}[\exp(\lambda X_i)] = n \cdot \psi_X(\lambda)$$

Consider a random variariable Y with binomial distirubtion with parameters n and p. Then,

$$\Pr[Y \geqslant t - np] \leqslant \exp(-n \cdot D(p + t/n \| p))$$

The Cramér-Chernoff Method (Normal random variables)

Let Z be a centered normal random variable with variance σ^2. Then,

$$\psi_Z(\lambda) = \log \mathbb{E}[\exp(\lambda Z)] = \frac{\lambda^2\sigma^2}{2}$$
$$\psi_Z^*(t) = \sup_{\lambda \in \mathbb{R}} (\lambda t - \psi_Z(\lambda)) = \lambda t - \frac{\lambda^2\sigma^2}{2}$$

Differentiate and solve for λ,

$$0 = t - \lambda\sigma^2 \tag{1}$$
$$\lambda = t/\sigma^2 \tag{2}$$

$$\psi_Z^*(t) = t^2/(2\sigma^2)$$
$$\Pr[Z \geqslant t] \leqslant \exp(-\psi_Z^*(t)) = \exp(-t^2/(2\sigma^2))$$

Hoeffding's Lemma (Bounded variables)

Lemma 1 (Hoeffding's Lemma)

Let Y be a random variable with $\mathbb{E}[Y] = 0$, taking values in a bounded interval $[a, b]$ then,

$$\psi_Y(\lambda) = \log \mathbb{E}[\exp(\lambda Y)] \leqslant \frac{1}{8}\lambda^2(b-a)^2$$

$$\psi_Z^*(t) = \sup_{\lambda \in \mathbb{R}} (\lambda t - \psi_Z(\lambda)) = \lambda t - \lambda^2(b-a)^2/8$$

Differentiate and solve for λ,

$$0 = t - \lambda(b-a)^2/4 \quad (1)$$

$$\lambda = 4t/(b-a)^2 \quad (2)$$

$$\psi_Z^*(t) = 2t^2/(b-a)^2$$

$$\Pr[Z \geqslant t] \leqslant \exp(-\psi_Z^*(t)) = \exp(-2t^2/(b-a)^2)$$

Hoeffding's Concentration Bound

Let $Z = X_1 + ... + X_n$, X_i's are independent, $X_i \in [a_i, b_i]$ and $\mathbb{E}[X_i] = 0$,

$$\psi_Z(\lambda) = \log \mathbb{E}[\exp(\lambda \sum_{i=1}^{n} X_i)] = \log \prod_{i=1}^{n} \mathbb{E}[\exp(\lambda X_i)] \leqslant \sum_{i=1}^{n} \frac{1}{8}\lambda^2 (b_i - a_i)^2$$

$$\psi_Z^*(t) = \sup_{\lambda \in \mathbb{R}} (\lambda t - \psi_Z(\lambda)) = \lambda t - \sum_{i=1}^{n} \frac{1}{8}\lambda^2 (b_i - a_i)^2$$

Differentiate and solve for λ,

$$0 = t - \textstyle\sum_{i=1}^{n} \lambda^2 (b_i - a_i)^2/4 \quad (1)$$

$$\lambda = 4t / \textstyle\sum_{i=1}^{n} (b_i - a_i)^2 \quad (2)$$

$$\psi_Z^*(t) = 2t^2 / \textstyle\sum_{i=1}^{n} (b_i - a_i)^2$$

$$\Pr[Z \geqslant t] \leqslant \exp(-\psi_Z^*(t)) = \exp(-2t^2 / \textstyle\sum_{i=1}^{n} (b_i - a_i)^2)$$

Multiplicative Chernoff Bound

Theorem 5 (Multiplicative Chernoff Upper Bound)

Let $X_1, ..., X_n$ be independent random Bernoulli random variables, where $p_i = \mathbb{E}[X_i]$. Then for $Z = X_1 + ... + X_n$, $\mu = \mathbb{E}[Z] = \sum_{i=1}^n p_i$ and any $\epsilon > 0$,

$$\Pr[Z \geqslant (1+\epsilon)\mu] \leqslant \exp\left(-\frac{\epsilon^2}{2(1+\epsilon/3)}\mu\right)$$

Proof:

$$\Pr[Z \geqslant (1+\epsilon)\mu] \leqslant \frac{\mathbb{E}[\exp(\lambda Z)]}{\exp(\lambda(1+\epsilon)\mu)}$$

$$\mathbb{E}[\exp(\lambda Z)] = \mathbb{E}[\exp(\lambda \sum_{i=1}^{n} X_i)] = \prod_{i=1}^{n} \mathbb{E}[\exp(\lambda X_i)]$$

Multiplicative Chernoff Bound

Proof Continued...

$$\begin{aligned} \mathbb{E}[\exp(\lambda X_i)] &\leqslant p_i \exp(\lambda) + 1 - p_i & (3) \\ &= 1 + p_i(\exp(t) - 1) & (4) \end{aligned}$$

Using $1 + x \leqslant \exp(x)$ with $x = p_i(\exp(\lambda) - 1)$

$$\leqslant \exp(p_i(\exp(t) - 1)) \quad (5)$$

$$\begin{aligned} \prod_{i=1}^{n} \mathbb{E}[\exp(\lambda X_i)] &\leqslant \prod_{i=1}^{n} \exp(p_i(\exp(\lambda) - 1)) & (6) \\ &= \exp(\sum_{i=1}^{n} p_i(\exp(\lambda) - 1)) = \exp(\mu(\exp(\lambda) - 1)) & (7) \end{aligned}$$

$$\Pr[Z \geqslant (1+\epsilon)\mu] \leqslant \frac{\exp(\mu(\exp(\lambda) - 1))]}{\exp(\lambda(1+\epsilon)\mu)}$$

Proof Continued...

Solving for λ that minimizes upper bound gives $\lambda = \log(1+\epsilon)$.

$$\Pr[Z \geqslant (1+\epsilon)\mu] \leqslant \frac{\exp(\mu\epsilon)]}{\exp(\log(1+\epsilon)(1+\epsilon)\mu)} \tag{8}$$

$$= \frac{\exp(\mu\epsilon)}{\exp(\log((1+\epsilon)^{(1+\epsilon)\mu}))} \tag{9}$$

$$= \frac{\exp(\mu\epsilon)}{(1+\epsilon)^{(1+\epsilon)\mu}} \tag{10}$$

$$= \left[\frac{\exp(\epsilon)}{(1+\epsilon)^{(1+\epsilon)}}\right]^{\mu} \tag{11}$$

Simplifying...

$$\Pr[Z \geqslant (1+\epsilon)\mu] \leqslant \left[\frac{\exp(\epsilon)}{(1+\epsilon)^{(1+\epsilon)}}\right]^{\mu} \tag{12}$$

$$= \exp\left(\mu \log\left[\frac{\exp(\epsilon)}{(1+\epsilon)^{(1+\epsilon)}}\right]\right) \tag{13}$$

$$= \exp\left(\mu(\epsilon - (1+\epsilon)\log(1+\epsilon))\right) \tag{14}$$

Using elementary inequality: $(1+u)\log(1+u) - u \geqslant \frac{u^2}{2(1+u/3)}$ for $u > 0$

$$\leqslant \exp\left(-\frac{\epsilon^2}{2(1+\epsilon/3)}\mu\right) \tag{15}$$

Multiplicative Chernoff Bound

Theorem 6 (Multiplicative Chernoff Lower Bound)

Let $X_1, ..., X_n$ be independent random Bernoulli random variables, where $p_i = \mathbb{E}[X_i]$. Then for $Z = X_1 + ... + X_n$, $\mu = \mathbb{E}[Z] = \sum_{i=1}^{n} p_i$ and any $\epsilon > 0$,

$$\Pr[Z \leqslant (1-\epsilon)\mu] \leqslant \exp\left(-\frac{\epsilon^2}{2}\mu\right)$$

Proof:

$$\Pr[Z \leqslant (1-\epsilon)\mu] = \Pr[-X \geqslant -(1-\epsilon)\mu] = \frac{\prod_{i=1}^{n} \mathbb{E}[\exp(-\lambda X_i)]}{\exp(-\lambda(1-\epsilon)\mu)}$$

(Skipping steps that are identical to upper bound)

$$\Pr[Z \geqslant (1-\epsilon)\mu] \leqslant \frac{\exp(\mu(\exp(-\lambda)-1))}{\exp(-\lambda(1-\epsilon)\mu)}$$

Multiplicative Chernoff Bound

Proof Continued...

Solving for λ that minimizes upper bound gives $\lambda = -\log(1-\epsilon)$.

$$\Pr[Z \geqslant (1-\epsilon)\mu] \leqslant \frac{\exp(-\mu\epsilon)}{\exp(\log(1-\epsilon)(1-\epsilon)\mu)} \tag{1}$$

$$= \frac{\exp(-\mu\epsilon)}{(1-\epsilon)^{(1-\epsilon)\mu}} \tag{2}$$

$$= \left[\frac{\exp(-\epsilon)}{(1-\epsilon)^{(1-\epsilon)}}\right]^{\mu} \tag{3}$$

$$\tag{4}$$

Simplifying...

$$\Pr[Z \geqslant (1-\epsilon)\mu] \leqslant \left[\frac{\exp(-\epsilon)}{(1-\epsilon)^{(1-\epsilon)}}\right]^{\mu} \tag{5}$$

$$= \exp\left(\mu \log\left[\frac{\exp(-\epsilon)}{(1-\epsilon)^{(1-\epsilon)}}\right]\right) \tag{6}$$

$$= \exp\left(\mu(-\epsilon - (1-\epsilon)\log(1-\epsilon))\right) \tag{7}$$

Using elementary inequality: $(1-u)\log(1-u) \geqslant -u + u^2/2$ for $u > 0$

$$\leqslant \exp\left(-\frac{\epsilon^2}{2}\mu\right) \tag{8}$$

(More) Questions?