EE5603:Concentration Inequalities

Sumohana Chennappayya and G V V Sharma[*]

1 MARKOV INEQUALITY

1.1 Let $X \geq 0$ be a positive random integer. Show that

$$E[X] = \sum_{m=0}^{\infty} \Pr(X \geqslant m) \tag{1.1}$$

Solution: By definition,

$$E[X] = \sum_{m=0}^{\infty} m \Pr(X = m) \tag{1.2}$$

$$= \Pr(X = 1) + 2\Pr(X = 2) + 3\Pr(X = 3)$$
$$+ \ldots \tag{1.3}$$

$$= \{\Pr(X = 1) + \Pr(X = 2) + \Pr(X = 3)$$
$$+ \ldots\} \tag{1.4}$$

$$+ \{\Pr(X = 2) + \Pr(X = 3) + \ldots\} \tag{1.5}$$

$$+ \{\Pr(X = 3) + \ldots\} + \ldots \tag{1.6}$$

$$= \Pr(X \geqslant 1) + 2\Pr(X \geqslant 2) + 3\Pr(X \geqslant 3)$$
$$+ \ldots \tag{1.7}$$

resulting in (1.2).

1.2 For a continuious r.v $X \geq 0$, show that

$$E[X] = \int_0^{\infty} \Pr(x \geqslant t)dt \tag{1.8}$$

1.3 For r.v $X \geq 0$ and $\varepsilon > 0$, show that

$$\Pr(X \geqslant \varepsilon) \leqslant \frac{E[X]}{\varepsilon} \tag{1.9}$$

Solution: $\because X \geqslant 0$,

$$E[X] = \int_0^{\infty} x p_X(x)\,dx \tag{1.10}$$

$$= \int_0^{\varepsilon} x p_X(x)\,dx + \int_{\varepsilon}^{\infty} x p_X(x)\,dx \tag{1.11}$$

$$\geqslant \int_{\varepsilon}^{\infty} x p_X(x)\,dx \tag{1.12}$$

which can be expressed as

$$E[X] \geqslant \int_{\varepsilon}^{\infty} \varepsilon p_X(x)\,dx \tag{1.13}$$

$$= \varepsilon \int_{\varepsilon}^{\infty} p_X(x)\,dx = \varepsilon \Pr(X \geqslant \varepsilon) \tag{1.14}$$

resulting in (1.9).

1.4 *Chernoff Bound :* For any r.v X with bounded variance, and for any $t > 0$, show using (1.9) that

$$\Pr(e^{tX} \geqslant e^{t\varepsilon}) \leqslant \frac{E\left(e^{tX}\right)}{e^{t\varepsilon}} \tag{1.15}$$

1.5 Show that

$$\Pr(X \geqslant \varepsilon) = \Pr(e^{tX} \geqslant e^{t\varepsilon}) \tag{1.16}$$

Solution: This is true for any monotonic function.

2 CHEBYSCHEV INEQUALITY

2.1 Let

$$Y = (X - E[X])^2 \tag{2.1}$$

and $\varepsilon > 0$. Show using (1.9) that

$$\Pr\left(Y \geqslant \varepsilon^2\right) \leqslant \frac{E(Y)}{\varepsilon^2} \tag{2.2}$$

2.2 Show that

$$\Pr\left(Y \geqslant \varepsilon^2\right) = \Pr\left(\sqrt{Y} \geqslant \varepsilon\right) + \Pr\left(\sqrt{Y} \leqslant -\varepsilon\right) \tag{2.3}$$

2.3 Show that

$$\Pr\left(\sqrt{Y} \leqslant -\varepsilon\right) = 0, \tag{2.4}$$

2.4 Show that

$$\Pr\left(\sqrt{Y} \geqslant \varepsilon\right) \leqslant \frac{E(Y)}{\varepsilon^2} \tag{2.5}$$

2.5 Show that

$$\Pr(|X - E[X]| \geqslant \varepsilon) \leqslant \frac{\text{Var}(X)}{\varepsilon^2} \tag{2.6}$$

3 LAW OF LARGE NUMBERS (LLN)

3.1 Let

$$S_n = \frac{1}{n} \sum_{i=1}^{n} X_i \qquad (3.1)$$

where X_i are i.i.d r.v. with mean μ and bounded variance σ^2. Show that

$$E(S_n) = \mu \qquad (3.2)$$

$$\text{Var}(S_n) = \frac{\sigma^2}{n} \qquad (3.3)$$

3.2 Using Chebyschev inequality in (2.6), show that

$$\Pr(|S_n - \mu| \geqslant \varepsilon) \leqslant \frac{\sigma^2}{n\varepsilon^2} \qquad (3.4)$$

3.3 Show that

$$\lim_{n \to \infty} \Pr(|S_n - \mu| \geqslant \varepsilon) = 0 \qquad (3.5)$$

4 HOEFFDING'S LEMMA

4.1 Consider a r.v. X such that $a < X < b$. If $\theta = \frac{a}{a-b}$ and $E[X] = 0$, show that

$$0 < \theta < 1 \qquad (4.1)$$

4.2 A convex function g is defined as

$$g(x) \leqslant f(x), \quad x \in (a, b) \qquad (4.2)$$

where f is the line joining the points $(a, g(a))$ and $(b, g(b))$. Show that

$$E[f(X)] \leq \frac{bg(a) - ag(b)}{b - a} \qquad (4.3)$$

4.3 Show that

$$g(x) = e^{sx}, s > 0 \qquad (4.4)$$

is convex.

4.4 Using (4.3) and (4.4), show that the moment generating function (MGF) of X

$$M_X(s) = E\left[e^{sX}\right] \leq e^{-u\theta}(1 - \theta + \theta e^u), \quad (4.5)$$

where $u = s(b - a)$.

4.5 Let

$$e^{-u\theta}(1 - \theta + \theta e^u) = e^{\psi(u)} \qquad (4.6)$$

Show that

$$\psi(0) = 0 \qquad (4.7)$$

$$\psi'(0) = 0 \qquad (4.8)$$

$$\psi''(v) = t(1 - t), \quad t = \frac{\theta e^v}{1 - \theta + \theta e^v}. \qquad (4.9)$$

4.6 Show that

$$\psi''(v) \leq \frac{1}{4} \qquad (4.10)$$

4.7 According to Taylor's theorem,

$$\psi(u) = \psi(0) + u\psi'(0) + \frac{u^2}{2}\psi''(v) \quad 0 < v < u. \qquad (4.11)$$

4.8 Using (4.7), (4.10) and (4.11), show that

$$\psi(u) \leq \frac{u^2}{8} \quad 0 < v < u. \qquad (4.12)$$

4.9 From (4.5), (4.6) and (4.12), show that

$$M_X(s) \leq e^{\frac{s^2(b-a)^2}{8}} \qquad (4.13)$$

5 HOEFFDING'S INEQUALITY

5.1 Let

$$S_n = \sum_{i=1}^{n} X_i, \quad X_i \in [a_i, b_i], E[X_i] = 0 \quad (5.1)$$

wherer X_i are independent and

$$Y = S_n - E[S_n]. \qquad (5.2)$$

5.2 Using (1.15) and (1.16), show that

$$\Pr(Y \geq t) \leq e^{-st} M_Y(s). \qquad (5.3)$$

5.3 Show that

$$M_Y(s) = \prod_{i=1}^{n} M_{X_i}(s) \qquad (5.4)$$

5.4 From (5.3), (5.5) and (4.13), show that

$$\Pr(Y \geq t) \leq \exp\left(-st + \frac{1}{8}s^2\|\mathbf{b} - \mathbf{a}\|^2\right) \quad (5.5)$$

5.5 Show that

$$\min_s\left(-st + \frac{1}{8}s^2\|\mathbf{b} - \mathbf{a}\|^2\right) = -\frac{2t^2}{\|\mathbf{b} - \mathbf{a}\|^2} \quad (5.6)$$

5.6 Show that

$$\Pr(S_n - E[S_n] \geq t) \leq \exp\left(-\frac{2t^2}{\|\mathbf{b} - \mathbf{a}\|^2}\right) \quad (5.7)$$

6 BENNET'S INEQUALITY

6.1 A real valued r.v X is said to be σ^2-sub Gaussian if there exists a σ such that

$$M_X(\lambda) < e^{\frac{\lambda^2 \sigma^2}{2}} \qquad (6.1)$$

6.2 Let

$$\phi(x) = e^x - x - 1. \qquad (6.2)$$

Show that $u^{-2}\phi(u)$ is non-decreasing.

6.3 Show that

$$(\lambda X_i)^{-2}\,\phi\,(\lambda X_i) \leqslant \lambda^{-2}\phi(\lambda), \quad \lambda > 0, X_i < 1. \qquad (6.3)$$

6.4 Show that

$$E[e^{\lambda X_i}] \leqslant 1 + \lambda E\,[X_i] + E\left[X_i^2\right]\phi(\lambda) \qquad (6.4)$$

6.5 Let

$$S = \sum_{1=1}^{n} X_i - E\,[X_i]. \qquad (6.5)$$

Show that

$$\log M_S(\lambda) \leqslant$$

$$\sum_{i=1}^{n} \log\left[1 + \lambda E\,(X_i) + E\left(X_i^2\right)\phi(\lambda)\right]$$

$$- \lambda \sum_{i=1}^{n} E\,(X_i) \qquad (6.6)$$

6.6 Given that log is convex, show that

$$\sum_{1=1}^{n} \log\left[1 + \lambda E\,(X_i) + E\left(X_i^2\right)\phi(\lambda)\right]$$

$$\leq n \log\left[\frac{1}{n}\sum_{1=1}^{n}\left(1 + \lambda E\,(X_i) + E\left(X_i^2\right)\phi(\lambda)\right)\right] \qquad (6.7)$$

6.7 Using the fact that $\log(1 + x) \leqslant x, x \geqslant 0$, show that

$$n \log\left[\frac{1}{n}\sum_{1=1}^{n}\left(1 + \lambda E\,(X_i) + E\left(X_i^2\right)\phi(\lambda)\right)\right]$$

$$- \lambda \sum_{i=1}^{n} E\,(X_i) \leq v\phi(\lambda) \qquad (6.8)$$

where

$$v = \sum_{1=1}^{n} E\left(X_i^2\right), \qquad (6.9)$$

$$\sum_{1=1}^{n} E\,(X_i) \geq 0, \qquad (6.10)$$

6.8 From (6.6)and (6.8), show that

$$\log M_S(\lambda) \leq \frac{v}{b^2}\phi(\lambda b), \qquad (6.11)$$

for $X_i \leq b$.

7 MC DIARMID'S INEQUALITY

7.1 Let

$$\mathbf{X}_i = \mathbf{X}_{1:i} = \begin{pmatrix} X_1 & X_2 & \dots & X_i \end{pmatrix}, \text{ (say)} \qquad (7.1)$$

$$\mathbf{X} \triangleq \mathbf{X}_n \qquad (7.2)$$

$$B_i \triangleq E\,[g\,(\mathbf{X})\,|\mathbf{X}_i] \qquad (7.3)$$

where X_i are all independent. Show that

$$B_n = g\,(\mathbf{X}) \qquad (7.4)$$

$$B_0 = E\,[g\,(\mathbf{X})] \qquad (7.5)$$

$$B_i = \sum_{\mathbf{a}_{i+1:n}} g\,(\mathbf{X}_i, \mathbf{a}_{i+1:n})\,\mathrm{Pr}\,(\mathbf{X}_{i+1:n} = \mathbf{a}_{i+1:n}) \qquad (7.6)$$

7.2 If

$$|g(x_1, x_2, \dots, x_n)$$
$$- g(x_1, x_2, \dots, x_{i-1}, \hat{x}_i, x_{i+1}, \dots, x_n)| \leq c_i$$
$$\text{for} \quad 1 \leq i \leq n, \qquad (7.7)$$

show that

$$|g\,(\mathbf{X}_i, \mathbf{a}_{i+1:n})$$
$$- \sum_{a_i'} g\,(\mathbf{X}_i, a_i', \mathbf{a}_{i+1:n})\,\mathrm{Pr}\,(X_i = a_i')|$$
$$\leq c_i \qquad (7.8)$$

Solution:

$$\because \sum_{a_i'} \mathrm{Pr}\,(X_i = a_i') = 1, \qquad (7.9)$$

(7.14) can be expressed as

$$\sum_{a_i'} |g\,(\mathbf{X}_i, \mathbf{a}_{i+1:n}) - g\,(\mathbf{X}_i, a_i', \mathbf{a}_{i+1:n})|$$

$$\times \mathrm{Pr}\,(X_i = a_i') \leq c_i \qquad (7.10)$$

7.3 If

$$V_i = B_i - B_{i-1}, \qquad (7.11)$$

show that

$$|V_i| \le c_i \qquad (7.12)$$

Solution:

$$\because V_i = B_i - B_{i-1}$$
$$= \sum_{\mathbf{a}_{i+1:n}} g\left(\mathbf{X}_i, \mathbf{a}_{i+1:n}\right) \Pr\left(\mathbf{X}_{i+1:n} = \mathbf{a}_{i+1:n}\right)$$
$$- \sum_{\mathbf{a}_{i:n}} g\left(\mathbf{X}_{i-1}, \mathbf{a}_{i:n}\right) \Pr\left(\mathbf{X}_{i:n} = \mathbf{a}_{i:n}\right) \quad (7.13)$$

we obtain

$$|V_i| \le \sum_{\mathbf{a}_{i+1:n}} |g\left(\mathbf{X}_i, \mathbf{a}_{i+1:n}\right)$$
$$- \sum_{a_i'} g\left(\mathbf{X}_i, a_i', \mathbf{a}_{i+1:n}\right) \Pr\left(X_i = a_i'\right)|$$
$$\Pr\left(\mathbf{X}_{i+1:n} = \mathbf{a}_{i+1:n}\right) \quad (7.14)$$

Using (7.8) in (7.14),

$$|V_i| \le c_i \sum_{\mathbf{a}_{i+1:n}} \Pr\left(\mathbf{X}_{i+1:n} = \mathbf{a}_{i+1:n}\right) = c_i \qquad (7.15)$$

7.4 Show that

$$E\left[V_i\right] = 0 \qquad (7.16)$$

7.5 Let

$$S_n = \sum_{i=1}^{n} V_i, \quad V_i \in [-c_i, c_i], E\left[V_i\right] = 0 \qquad (7.17)$$

Show that

$$S_n = B_n - B_0 = g\left(\mathbf{X}\right) - E\left[g\left(\mathbf{X}\right)\right] \qquad (7.18)$$

7.6 If

$$\mathbf{c}_i = \begin{pmatrix} c_1 & c_2 & \dots & c_n \end{pmatrix}, \qquad (7.19)$$

substituting $\mathbf{a} = -\mathbf{c}, \mathbf{b} = \mathbf{c}$ in (5.7), show that

$$\Pr\left(S_n - E\left[S_n\right] \ge t\right) \le \exp\left(-\frac{t^2}{2\|\mathbf{c}\|^2}\right) \qquad (7.20)$$

$$\Pr\left(S_n - E\left[S_n\right] \le -t\right) \le \exp\left(-\frac{t^2}{2\|\mathbf{c}\|^2}\right) \qquad (7.21)$$

- **Eform-stein inequality:** let

$$X_1, \dots, X_n \qquad (7.22)$$

be indepentent RVs. let $\delta : x^n \to |R$ be a square integrable function. let $z = f(x_1...x_n,)$

$$Var(z) \le \Sigma_{1=1}^{n} E(z - E^i(z))^2 \qquad (7.23)$$

$$E_i(z) = E(f(x^n)|x^1). \qquad (7.24)$$

To dos
- $\Delta_i = E_i(z) - E_{1=1}(z)$

  $\bullet Z - E[z] = \Sigma_{1=1}^{n} \Delta_i$

  $\bullet if E^i(z) = \int f(x_1...x_i - 1, x_{i+1}...x_n)$

  $dp(x_i) E_i[E^i(z)] = E_{i-1(z)}$

8 CONVERGENCE

8.1 Definitions

show that

9 REVIEW

10 EFROM-STEIN INQUALITY

let

$$X_1, \dots, X_n \qquad (10.1)$$

be independent RVs, let $f : x^n \to |R$ be a square integrable, $z = f(x_1....x_n)$.

$$Var(z) \le \Sigma_{1=1}^{n} E[(z - E^i(z))^2] dif \vartheta \qquad (10.2)$$

$\bullet E_i(z) = E[f(x_i...x_n)|x^i]; E_\circ = E \qquad (10.3)$

$\bullet E^i(z) = \int_{\lambda_i \in x} f(x_1...x_i, ...x_n) dp(xi) \qquad (10.4)$

$\bullet if \Delta_i = E_i(z) - E_{i-1}(z), \Sigma_{1=1}^{n} \Delta_i = z - E(z). \to 1 \qquad (10.5)$

$\bullet Var(z) = E[(z - E(z))^2](fromdefin) \qquad (10.6)$

$= E[(\Sigma_{1=1}^{n} \Delta_i)^2](from1) \qquad (10.7)$

$$= E[\Sigma_{1=1}^{n}\Delta_i^2 + 2\Sigma_{j>i}\Delta_i\Delta_j] \qquad (10.8)$$

$$E[\Sigma_{1=1}^{n}\Delta_i^2] + 2.\Sigma_{j>i}E[\Delta_j\Delta_i] \qquad (10.9)$$

Claim:if $j > i, E_i\Delta_i = 0$

$$E_i\Delta_j = E_i[E_j(z) - E_{j-1}(z)] \qquad (10.10)$$

$$= E[E_j(z) - E_{j-1}(z)|x^i] \qquad (10.11)$$

$$= E_i(z) - E_i(z) \leftarrow (since\, j > i) \qquad (10.12)$$

$$= 0 \qquad (10.13)$$

if$E_i\Delta_j = 0\, j > i$,then $2\Sigma_{j>i}E[\Delta_i\Delta_j] =? = 0$

$$\Rightarrow Var(z) = E(\Sigma_{1=1}^{n}\Delta_i^2) \rightarrow 2 \qquad (10.14)$$

• Recall. $E_i[E^i(z)] = E_{i-1}(z)$

$$E^i(z) = \int_{\lambda_i \epsilon x} f(x_1...x_{i-1}, xi, x_{i-1}...x_n)dp(xi) \quad (10.15)$$

$$E^i(z) = \int_{\lambda_{i+1}\epsilon x^{n-1}} f(x_1...x_{i-1}, x_{i+1}...x_n)d.p(x_{i+1}^n) \qquad (10.16)$$

$$\Rightarrow E_i[E^i(z)] = \int_{x_{i+1}\epsilon x^{n-i}} \int_{x_i \epsilon \chi} f(x_1...x_i, ...x_{i-1}...x_n).dp(x_i) \qquad (10.17)$$

$$= \int_{x_i \epsilon x^{n-(i-1)}} f(x_1...x_{i-1}, x_i, ...x_n).dp(x_i^n) \qquad (10.18)$$

$$= E_{i-1}(z) \rightarrow 3 \qquad (10.19)$$

$$\Delta_i = E_i(z) - E_{i-1}(z) \qquad (10.20)$$

$$\Delta_i = E_i[z - E^i(z)]. \qquad (10.21)$$

$$\Delta_i^2 = [Ei[z - E^i(z)]]^2. \qquad (10.22)$$

jensun's inequality if g is aconvex function and x is a random variable then

$$E[f(x)] \geqslant f(Ex) \qquad (10.23)$$

Apply this result to Δ_i

$$E[\Delta_i^2] \geqslant (E[\Delta_i])^2\, or \qquad (10.24)$$

$$(E[\Delta_i])^2 \leqslant E[\Delta_i^2] :\rightarrow 4 \qquad (10.25)$$

Recall Var $=E[\Sigma_{1=1}^{n}\Delta_i^2]$