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

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

Solution: By definition,

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

$$= \Pr\left(X = 1\right) + 2\Pr\left(X = 2\right) + 3\Pr\left(X = 3\right)$$

$$+ \dots \tag{1.3}$$

$$= \{\Pr\left(X = 1\right) + \Pr\left(X = 2\right) + \Pr\left(X = 3\right)$$

$$+ \dots\} \tag{1.4}$$

$$+ \{\Pr\left(X = 2\right) + \Pr\left(X = 3\right) + \dots\} \tag{1.5}$$

$$+ \{\Pr\left(X = 3\right) + \dots\} + \dots \tag{1.6}$$

$$= \Pr\left(X \geqslant 1\right) + 2\Pr\left(X \geqslant 2\right) + 3\Pr\left(X \geqslant 3\right)$$

$$+ \dots \tag{1.7}$$

resulting in (1.2).

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

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

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

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

Solution: $\because X \geqslant 0$,

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

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

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

which can be expressed as

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

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

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

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

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

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

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

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

2.5 Show that

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

*The author is with the Department of Electrical Engineering, IIT, Hyderabad 502285 India e-mail: gadepall@iith.ac.in.

3 Law of large numbers (LLN)

3.1 Let

$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{3.1}$$

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

$$E(S_n) = \mu \tag{3.2}$$

$$\mathrm{Var}(S_n) = \frac{\sigma^2}{n} \tag{3.3}$$

3.2 Using Chebyschev inequality in (2.6), show that

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

3.3 Show that

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

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

$$0 < \theta < 1 \tag{4.1}$$

4.2 A convex function g is defined as

$$g(x) \leqslant f(x), \quad x \in (a, b) \tag{4.2}$$

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

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

4.3 Show that

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

is convex.

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

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

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

4.5 Let

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

Show that

$$\psi(0) = 0 \tag{4.7}$$

$$\psi'(0) = 0 \tag{4.8}$$

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

4.6 Show that

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

4.7 According to Taylor's theorem,

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

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

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

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

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

5 Hoeffding's Inequality

5.1 Let

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

wherer X_i are independent and

$$Y = S_n - E[S_n]. \tag{5.2}$$

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

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

5.3 Show that

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

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

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

5.5 Show that

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

5.6 Show that

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

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

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

6.2 Let

$$\phi(x) = e^x - x - 1. \tag{6.2}$$

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

6.3 Show that

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

6.4 Show that

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

6.5 Let

$$S = \sum_{1=1}^{n} X_i - E\left[X_i\right]. \tag{6.5}$$

Show that

$$\log M_S(\lambda) \leqslant \sum_{i=1}^{n} \log\left[1 + \lambda E\left(X_i\right) + E\left(X_i^2\right)\phi(\lambda)\right] - \lambda\sum_{i=1}^{n} E\left(X_i\right) \tag{6.6}$$

6.6 Given that log is convex, show that

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

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

$$n\log\left[\frac{1}{n}\sum_{1=1}^{n}\left(1 + \lambda E\left(X_i\right) + E\left(X_i^2\right)\phi(\lambda)\right)\right] - \lambda\sum_{i=1}^{n} E\left(X_i\right) \leq v\phi(\lambda) \tag{6.8}$$

where

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

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

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

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

for $X_i \leq b$.

7 Mc Diarmid's inequality

7.1 Let

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

$$\mathbf{X} \triangleq \mathbf{X}_n \tag{7.2}$$

$$B_i \triangleq E\left[g\left(\mathbf{X}\right)|\mathbf{X}_i\right] \tag{7.3}$$

where X_i are all independent. Show that

$$B_n = g\left(\mathbf{X}\right) \tag{7.4}$$

$$B_0 = E\left[g\left(\mathbf{X}\right)\right] \tag{7.5}$$

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

7.2 If

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

show that

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

Solution:

$$\because \sum_{a_i'} \Pr\left(X_i = a_i'\right) = 1, \tag{7.9}$$

(7.14) can be expressed as

$$\sum_{a_i'} |g\left(\mathbf{X}_i, \mathbf{a}_{i+1:n}\right) - g\left(\mathbf{X}_i, a_i', \mathbf{a}_{i+1:n}\right)| \times \Pr\left(X_i = a_i'\right) \leq c_i \tag{7.10}$$

7.3 If

$$V_i = B_i - B_{i-1}, \tag{7.11}$$

show that

$$|V_i| \le c_i \tag{7.12}$$

Solution:

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

we obtain

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

Using (7.8) in (7.14),

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

7.4 Show that

$$E\left[V_i\right] = 0 \tag{7.16}$$

7.5 Let

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

Show that

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

7.6 If

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

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

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

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

8 Efrom-stein inquality

8.1 Let

$$\mathbf{X} = \begin{pmatrix} X_1 & X_2 & \dots & X_n \end{pmatrix} \tag{8.1}$$

be a random vector comprising of independent r.vs and f be square integrable.

$$\text{var}\left[f\left(\mathbf{X}\right)\right] \leqslant \frac{1}{2}\sum_{i=1}^{n} E\left[\left\{f\left(\mathbf{X}\right) - f\left(X_{1:i-1}, X_i', X_{i+1:n}\right)\right\}^2\right] \tag{8.2}$$

$$\bullet E_i(z) = E[f(x_i...x_n)|x^i]; E_\circ = E \tag{8.3}$$

$$\bullet E^i(z) = \int_{\lambda_i \epsilon x} f(x_1...x_i, ...x_n) dp(xi) \tag{8.4}$$

$$\bullet if \Delta_i = E_i(z) - E_{i-1}(z), \Sigma_{1=1}^n \Delta_i = z - E(z). \rightarrow 1 \tag{8.5}$$

$$\bullet Var(z) = E[(z - E(z))^2](fromdefin) \tag{8.6}$$

$$= E[(\Sigma_{1=1}^n \Delta_i)^2](from1) \tag{8.7}$$

$$= E[\Sigma_{1=1}^n \Delta_i^2 + 2\Sigma_{j>i}\Delta_i\Delta_j] \tag{8.8}$$

$$E[\Sigma_{1=1}^n \Delta_i^2] + 2.\Sigma_{j>i} E[\Delta_j \Delta_i] \tag{8.9}$$

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

$$E_i\Delta_j = E_i[E_j(z) - E_{j-1}(z)] \tag{8.10}$$

$$= E[E_j(z) - E_{j-1}(z)|x^i] \tag{8.11}$$

$$= E_i(z) - E_i(z) \leftarrow (since j > i) \tag{8.12}$$

$$= 0 \tag{8.13}$$

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

$$\Rightarrow Var(z) = E(\Sigma_{1=1}^n \Delta_i^2) \rightarrow 2 \tag{8.14}$$

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

$$E^i(z) = \int_{\lambda_i \epsilon x} f(x_1...x_{i-1}, xi, x_{i-1}...x_n) dp(xi) \tag{8.15}$$

$$E^i(z) = \int_{\lambda_{i+1} \epsilon \chi^{n-1}} f(x_1 ... x_{i-1}, x_{i+1} ... x_n) d.p(x_{i+1}^n) \tag{8.16}$$

$$\Rightarrow E_i[E^i(z)] = \int_{x_{i+1} \epsilon \chi^{n-i}} \int_{x_i \epsilon \chi} f(x_1 ... x_i, ... x_{i-1} ... x_n).dp(x_i) \tag{8.17}$$

$$= \int_{x_i \epsilon \chi^{n-(i-1)}} f(x_1 ... x_{i-1}, x_i, ... x_n).dp(x_i^n) \tag{8.18}$$

$$= E_{i-1}(z) \rightarrow 3 \tag{8.19}$$

$$\Delta_i = E_i(z) - E_{i-1}(z) \tag{8.20}$$

$$\Delta_i = E_i[z - E^i(z)]. \tag{8.21}$$

$$\Delta_i^2 = [Ei[z - E^i(z)]]^2. \tag{8.22}$$

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

$$E[f(x)] \geqslant f(Ex) \tag{8.23}$$

Apply this result to Δ_i

$$E[\Delta_i^2] \geqslant (E[\Delta_i])^2 or \tag{8.24}$$

$$(E[\Delta_i])^2 \leqslant E[\Delta_i^2] :\rightarrow 4 \tag{8.25}$$

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