

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

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

Solution: By definition,

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

$$= \Pr(X = 1) + 2 \Pr(X = 2) + 3 \Pr(X = 3) + \dots \quad (1.3)$$

$$= \{\Pr(X = 1) + \Pr(X = 2) + \Pr(X = 3) + \dots\} \quad (1.4)$$

$$+ \{\Pr(X = 2) + \Pr(X = 3) + \dots\} \quad (1.5)$$

$$+ \{\Pr(X = 3) + \dots\} + \dots \quad (1.6)$$

$$= \Pr(X \geq 1) + 2 \Pr(X \geq 2) + 3 \Pr(X \geq 3) + \dots \quad (1.7)$$

resulting in (1.2).

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

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

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

$$\Pr(X \geq \varepsilon) \leq \frac{E[X]}{\varepsilon} \quad (1.9)$$

Solution: $\because X \geq 0$,

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

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

$$\geq \int_{\varepsilon}^{\infty} x p_X(x) dx \quad (1.12)$$

which can be expressed as

$$E[X] \geq \int_{\varepsilon}^{\infty} \varepsilon p_X(x) dx \quad (1.13)$$

$$= \varepsilon \int_{\varepsilon}^{\infty} p_X(x) dx = \varepsilon \Pr(X \geq \varepsilon) \quad (1.14)$$

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

$$\Pr(e^{tX} \geq e^{t\varepsilon}) \leq \frac{E(e^{tX})}{e^{t\varepsilon}} \quad (1.15)$$

1.5 Show that

$$\Pr(X \geq \varepsilon) = \Pr(e^{tX} \geq e^{t\varepsilon}) \quad (1.16)$$

Solution: This is true for any monotonic function.

2 CHEBYSHEV INEQUALITY

2.1 Let

$$Y = (X - E[X])^2 \quad (2.1)$$

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

$$\Pr(Y \geq \varepsilon^2) \leq \frac{E(Y)}{\varepsilon^2} \quad (2.2)$$

2.2 Show that

$$\Pr(Y \geq \varepsilon^2) = \Pr(\sqrt{Y} \geq \varepsilon) + \Pr(\sqrt{Y} \leq -\varepsilon) \quad (2.3)$$

2.3 Show that

$$\Pr(\sqrt{Y} \leq -\varepsilon) = 0, \quad (2.4)$$

2.4 Show that

$$\Pr(\sqrt{Y} \geq \varepsilon) \leq \frac{E(Y)}{\varepsilon^2} \quad (2.5)$$

2.5 Show that

$$\Pr(|X - E[X]| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2} \quad (2.6)$$

*The author is with the Department of Electrical Engineering, IIT, Hyderabad 502285 India e-mail: gadepall@iith.ac.in. All material in the manuscript is released under GNU GPL. Free to use for all.

3 LAW OF LARGE NUMBERS (LLN)

3.1 Let

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

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

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

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

3.2 Using Chebyshev inequality in (2.6), show that

$$\Pr(|S_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \quad (3.4)$$

3.3 Show that

$$\lim_{n \rightarrow \infty} \Pr(|S_n - \mu| \geq \varepsilon) = 0 \quad (3.5)$$

4 Hoeffding's Lemma

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

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

4.2 A convex function g is defined as

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

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

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

4.3 Show that

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

is convex.

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

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

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

4.5 Let

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

Show that

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

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

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

4.6 Show that

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

4.7 According to Taylor's theorem,

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

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

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

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

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

5 Hoeffding's Inequality

5.1 Let

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

wherer X_i are independent and

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

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

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

5.3 Show that

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

5.4 From (5.3), (5.4) 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

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

6.2 Let

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

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

6.3 Show that

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

6.4 Show that

$$E[e^{\lambda X_i}] \leq 1 + \lambda E[X_i] + E[X_i^2] \phi(\lambda) \quad (6.4)$$

6.5 Let

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

Show that

$$\begin{aligned} \log M_S(\lambda) &\leq \\ &\sum_{i=1}^n \log \left[1 + \lambda E[X_i] + E[X_i^2] \phi(\lambda) \right] \\ &\quad - \lambda \sum_{i=1}^n E[X_i] \end{aligned} \quad (6.6)$$

6.6 Given that \log is convex, show that

$$\begin{aligned} &\sum_{i=1}^n \log \left[1 + \lambda E[X_i] + E[X_i^2] \phi(\lambda) \right] \\ &\leq n \log \left[\frac{1}{n} \sum_{i=1}^n \left(1 + \lambda E[X_i] + E[X_i^2] \phi(\lambda) \right) \right] \end{aligned} \quad (6.7)$$

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

$$\begin{aligned} &n \log \left[\frac{1}{n} \sum_{i=1}^n \left(1 + \lambda E[X_i] + E[X_i^2] \phi(\lambda) \right) \right] \\ &\quad - \lambda \sum_{i=1}^n E[X_i] \leq n \phi(\lambda) \end{aligned} \quad (6.8)$$

where

$$v = \sum_{i=1}^n E[X_i^2], \quad (6.9)$$

$$\sum_{i=1}^n E[X_i] \geq 0, \quad (6.10)$$

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

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

for $X_i \leq b$.

7 MC DIARMID'S INEQUALITY

7.1 Let

$$\mathbf{X}_i = \mathbf{X}_{1:i} = (X_1 \ X_2 \ \dots \ X_i), \text{ (say)} \quad (7.1)$$

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

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

where X_i are all independent. Show that

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

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

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

7.2 If

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

show that

$$\begin{aligned} &|g(\mathbf{X}_i, \mathbf{a}_{i+1:n}) \\ &\quad - \sum_{a'_i} g(\mathbf{X}_i, a'_i, \mathbf{a}_{i+1:n}) \Pr(X_i = a'_i)| \\ &\leq c_i \end{aligned} \quad (7.8)$$

Solution:

$$\because \sum_{a'_i} \Pr(X_i = a'_i) = 1, \quad (7.9)$$

(7.14) can be expressed as

$$\begin{aligned} &\sum_{a'_i} |g(\mathbf{X}_i, \mathbf{a}_{i+1:n}) - g(\mathbf{X}_i, a'_i, \mathbf{a}_{i+1:n})| \\ &\quad \times \Pr(X_i = a'_i) \leq c_i \end{aligned} \quad (7.10)$$

7.3 If

$$V_i = B_i - B_{i-1}, \quad (7.11) \quad 8.1 \text{ Let}$$

show that

$$|V_i| \leq c_i \quad (7.12)$$

Solution:

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

we obtain

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

Using (7.8) in (7.14),

$$|V_i| \leq c_i \sum_{\mathbf{a}_{i+1:n}} \Pr(\mathbf{X}_{i+1:n} = \mathbf{a}_{i+1:n}) = c_i \quad (7.15)$$

7.4 Show that

$$E[V_i] = 0 \quad (7.16)$$

7.5 Let

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

Show that

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

7.6 If

$$\mathbf{c}_i = (c_1 \ c_2 \ \dots \ c_n), \quad (7.19)$$

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

$$\Pr(S_n - E[S_n] \geq t) \leq \exp\left(-\frac{t^2}{2\|\mathbf{c}\|^2}\right) \quad (7.20)$$

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

8 EFROM-STIEN INEQUALITY

$$\mathbf{V} = \sum_{i=1}^n V_i = Z - E(Z) \quad (8.1)$$

Show that

$$\text{var}(Z) = \sum_{i=1}^n E[V_i^2] + 2 \sum_{i>j} E[V_i V_j] \quad (8.2)$$

8.2 If

$$\mathbf{X}_i = \mathbf{X}_{1:i} = (X_1 \ X_2 \ \dots \ X_i), \text{ (say)} \quad (8.3)$$

and

8.3 Let

$$\mathbf{X} = (X_1 \ X_2 \ \dots \ X_n) \quad (8.4)$$

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

$$\text{var}[f(\mathbf{X})] \leq \frac{1}{2} \sum_{i=1}^n E[\{f(\mathbf{X}) - f(X_{1:i-1}, X'_i, X_{i+1:n})\}^2] \quad (8.5)$$

$$\bullet E_i(z) = E[f(x_1 \dots x_n) | x^i]; E_o = E \quad (8.6)$$

$$\bullet E^i(z) = \int_{\lambda_i \in x} f(x_1 \dots x_i, \dots x_n) dp(x_i) \quad (8.7)$$

$$\bullet \text{if } V_i = E_i(z) - E_{i-1}(z), \sum_{i=1}^n V_i = z - E(z). \rightarrow 1 \quad (8.8)$$

$$\bullet \text{Var}(z) = E[(z - E(z))^2] (\text{from defin}) \quad (8.9)$$

$$= E[(\sum_{i=1}^n V_i)^2] (\text{from 1}) \quad (8.10)$$

$$= E[\sum_{i=1}^n V_i^2 + 2 \sum_{j>i} V_i V_j] \quad (8.11)$$

$$E[\sum_{i=1}^n V_i^2] + 2 \sum_{j>i} E[V_i V_j] \quad (8.12)$$

Claim: if $j > i, E_i V_j = 0$

$$E_i V_j = E_i[E_j(z) - E_{j-1}(z)] \quad (8.13)$$

$$= E[E_j(z) - E_{j-1}(z)|x^i] \quad (8.14)$$

$$= E_i(z) - E_i(z) \leftarrow (\text{since } j > i) \quad (8.15)$$

$$= 0 \quad (8.16)$$

if $E_i V_j = 0$ $j > i$, then $2 \sum_{j>i} E[V_i V_j] = ? = 0$

$$\Rightarrow \text{Var}(z) = E\left(\sum_{i=1}^n V_i^2\right) \rightarrow 2 \quad (8.17)$$

• Recall. $E_i[E^i(z)] = E_{i-1}(z)$

$$E^i(z) = \int_{\lambda_i \in x} f(x_1 \dots x_{i-1}, x_i, x_{i+1} \dots x_n) dp(x_i) \quad (8.18)$$

$$E^i(z) = \int_{\lambda_{i+1} \in x^{n-1}} f(x_1 \dots x_{i-1}, x_{i+1} \dots x_n) d.p(x_{i+1}^n) \quad (8.19)$$

$$\Rightarrow E_i[E^i(z)] = \int_{x_{i+1} \in x^{n-i}} \int_{x_i \in \mathcal{X}} f(x_1 \dots x_i, \dots x_{i-1} \dots x_n) . dp(x_i) \quad (8.20)$$

$$= \int_{x_i \in x^{n-(i-1)}} f(x_1 \dots x_{i-1}, x_i, \dots x_n) . dp(x_i^n) \quad (8.21)$$

$$= E_{i-1}(z) \rightarrow 3 \quad (8.22)$$

$$V_i = E_i(z) - E_{i-1}(z) \quad (8.23)$$

$$V_i = E_i[z - E^i(z)]. \quad (8.24)$$

$$V_i^2 = [E_i[z - E^i(z)]]^2. \quad (8.25)$$

jensen's inequality if g is a convex function and x is a random variable then

$$E[f(x)] \geq f(Ex) \quad (8.26)$$

Apply this result to V_i

$$E[V_i^2] \geq (E[V_i])^2 \text{ or } \quad (8.27)$$

$$(E[V_i])^2 \leq E[V_i^2] : \rightarrow 4 \quad (8.28)$$

Recall $\text{Var} = E[\sum_{i=1}^n V_i^2]$