Lecture 4 Moment Generating Function:

$$M_X(s) = \mathbb{E}\left[e^{sX}\right].$$

Note: We say MGF of X exists, if there exists a positive constant a s.t. $M_X(s)$ is finite for all $s \in [-a, a]$.

Property: If X and Y are independent.

$$M_{X+Y}(s) = M_X(s)\, M_Y(s)$$

$$= \mathbb{E}\left[e^{s(X+Y)}\right] = \mathbb{E}\left[e^{sX} e^{sY}\right] = \mathbb{E}\left[e^{sX}\right]\mathbb{E}\left[e^{sY}\right]$$

E.g. $X \sim N(\mu_x, \sigma_x^2)$, $Y \sim N(\mu_y, \sigma_y^2)$ independent

$$Z = aX + bY \Rightarrow Z \sim N(a\mu_x + b\mu_y, a^2\sigma_x^2 + b^2\sigma_y^2)$$

$$M_Z(s) = \mathbb{E}\left[e^{s(aX+bY)}\right]$$

$$= \mathbb{E}\left[e^{saX} e^{sbY}\right]$$

$$= \mathbb{E}\left[e^{saX}\right]\mathbb{E}\left[e^{sbY}\right] \quad \text{independence}$$

$$= M_X(sa)\, M_Y(sb)$$

Fact: If $X \sim N(\mu_x, \sigma_x^2)$ then

$$M_X(s) = e^{s\mu_x + s^2\sigma_x^2/2}$$

$$= e^{sa\mu_x + s^2a^2\sigma_x^2/2}\, e^{sb\mu_y + s^2b^2\sigma_y^2/2}$$

$$= e^{s(a\mu_x + b\mu_y) + s^2(a^2\sigma_x^2 + b^2\sigma_y^2)/2}$$

$$\Longrightarrow Z \sim N(a\mu_x + b\mu_y, a^2\sigma_x^2 + b^2\sigma_y^2)$$

Sub-Gaussian Random Variables

Def: A random variable X with $\mathbb{E}[X]=\mu$ is called sub-Gaussian with parameter σ^2 (variance proxy) if

$$\mathbb{E}\left[e^{s(X-\mu)}\right] \leq e^{s^2\sigma^2/2} \quad \forall s \in \mathbb{R}.$$

or equivalently $$\mathbb{E}\left[e^{sX}\right] \leq e^{s\mu + s^2\sigma^2/2} \quad \forall s \in \mathbb{R}$$

Example: $X \sim N(\mu, \sigma^2)$

Example 2: X is bounded, $a \leq X \leq b$, $\mathbb{E}[X]=\mu$

Hoeffding's Lemma: Let X be a r.v. s.t $a \leq X \leq b$ and $\mathbb{E}[X]=\mu$, then

$$\mathbb{E}\left[e^{s(X-\mu)}\right] \leq e^{s^2(b-a)^2/8}.$$

$\Longrightarrow$ X is sub-Gaussian with variance proxy $\frac{(b-a)^2}{4}$.

Property: X is sub-Gaussian with variance proxy σ^2

$$P(X-\mu \geq t) = P\left(e^{\lambda(X-\mu)} \geq e^{\lambda t}\right) \quad \text{for any } \lambda \geq 0$$

$$\leq \frac{\mathbb{E}\left[e^{\lambda(X-\mu)}\right]}{e^{\lambda t}} \quad \text{Markov's Inequality}$$

$$\leq e^{-\lambda t} e^{\lambda^2\sigma^2/2}.$$

$$= e^{\lambda^2\sigma^2/2 - \lambda t}. \quad \text{for all } \lambda \geq 0$$

$$2\lambda\frac{\sigma^2}{2} - t = 0$$

$$\lambda = t/\sigma^2.$$

minimizes the bound

$$= e^{t^2/2\sigma^2 - t^2/\sigma^2} = e^{-t^2/2\sigma^2}$$

Recall $X \sim \mathcal{N}(\mu, \sigma^2)$

$$P\left(\frac{X-\mu}{\sigma} \geq \frac{t}{\sigma}\right) = Q\left(\frac{t}{\sigma}\right)$$

$$\leq \frac{1}{\sqrt{2\pi}} e^{-t^2/2\sigma^2}.$$

The tail decay of a sub-Gaussian r.v is dominated (i.e. at least as fast as) the tail decay of a Gaussian r.v with the same variance.

Hoeffding's Inequality: Let $X_1, X_2, \ldots X_n$ be a sequence of iid rv's with $\mathbb{E}[X_i] = \mu$ and $a \leq X_i \leq b$. Then for any $t \geq 0$

$$P\left(|\bar{X}_n - \mu| \geq t\right) \leq 2e^{-2nt^2/(b-a)^2}$$

where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Proof: $$P\left(\bar{X}_n - \mu \geq t\right) \leq e^{-\lambda t}\Big(\underbrace{\mathbb{E}\left[e^{\frac{\lambda}{n}(X_i-\mu)}\right]}_{\leq e^{\left(\frac{\lambda}{n}\right)^2 (b-a)^2/8}}\Big)^n$$

if $X_1 \geq X_2 \geq 0$
$X_1^n \geq X_2^n \geq 0$.

by Hoeffding's Lemma

$$= e^{-\lambda t + \frac{\lambda^2}{n^2}(b-a)^2/8} \quad \forall \lambda \geq 0$$

choosing $\lambda = \frac{4nt}{(b-a)^2}$ $$= e^{-2nt^2/(b-a)^2}$$

$$P\left(\bar{X}_n - \mu \leq -t\right) = P\left(-\bar{X}_n - (-\mu) \geq t\right)$$

$$= P\Big(\frac{1}{n}\sum_{i=1}^{n}\underbrace{(-X_i)}_{\tilde{X}_i} - (-\mu) \geq t\Big) \leq e^{-\frac{2nt^2}{(b-a)^2}}$$

$\tilde{X}_i : \mathbb{E}[\tilde{X}_i] = -\mu, -b \leq \tilde{X}_i \leq -a$

Extensions:

- Sub-Gaussian random variables:

$$P\left(|\bar{X}_n - \mu| \geq t\right) \leq e^{-nt^2/2\sigma^2}.$$

- Non-identically distributed r.v's. Let $X_1, X_2, \dots X_n$ be independent r.v's s.t $a_i \leq X_i \leq b_i$.

Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \mathbb{E}[\bar{X}_n] = \mu.$

$$P\left(|\bar{X}_n - \mu| \geq t\right) \leq e^{-2n^2t^2 / \sum_{i=1}^{n}(b_i - a_i)^2}.$$

- Other functions:

<u>McDiarmid's Inequality:</u>

Let $X_1, X_2, \dots X_n$ be independent r.v's s.t $X_i \in \mathcal{X}_i \subseteq \mathbb{R}$ and let $f: \mathbb{R}^n \to \mathbb{R}$ be a function s.t

$$\max_{x_i, x_i' \in \mathcal{X}} \left| f(x_1, x_2, \dots x_{i-1}, x_i, x_{i+1}, \dots x_n) - f(x_1, x_2, \dots x_{i-1}, x_i', x_{i+1}, \dots x_n) \right| \leq c_i$$

for all $x_1 \in \mathcal{X}_1, \dots x_{i-1} \in \mathcal{X}_{i-1}, x_{i+1} \in \mathcal{X}_{i+1} \dots$

(Intuitively, f is not too sensitive to arbitrary changes in a single coordinate) Then

$$P\left(\left|f(X_1, X_2, \dots X_n) - \mathbb{E}[f(X_1, \dots X_n)]\right| \geq t\right) \leq e^{-2t^2 / \sum_{i=1}^{n} c_i^2}$$

Special case: $f(X_1, \dots X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $a_i \leq X_i \leq b_i$

$$c_i = \frac{b_i - a_i}{n} \Rightarrow \text{We recover Hoeffding's Inequality}$$