CAS CS 365
Lec 10

1. Moment Generating Function
    a. We define the MGF $\varphi_X(t)$ of a random variable X as the following function of a real variable t:

$$\boxed{\phi_X(t) = \mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[1 + tX + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \dots\right]}$$

    b. When X is Bernoulli (p), $\varphi_X(t)$ = E[e^tx] = (1-p)e^(t*0) + p*e^(t*1) = (1-p)+pe^t
    c. Moments derive when t > 0
        i. If X is discrete, then $$\phi_X(t) = \mathbb{E}\left[e^{tX}\right] = \sum_x e^{tx} p_X(x)$$
        ii. If X is continuous, then $$\phi_X(t) = \mathbb{E}\left[e^{tX}\right] = \int_{-\infty}^{+\infty} e^{tx} f_X(x) dx$$
    d. Why do we call it moment generating function?
        i. D
    e. Claim: $M_k(X) = \mathbb{E}\left[X^k\right] = \phi_X^{(k)}(0)$
    f. In other words, taking the k-th derivative, and setting t = 0 yields the k-th moment
    g. Proof sketch (induction)
        i. Base case:

$$\phi'(0) = \frac{d}{dt}\phi(t)\Big|_{t=0} = \mathbb{E}\left[Xe^{0X}\right] = \mathbb{E}[X]$$

        ii. IH: $\phi^{(k)}(t) = \mathbb{E}\left[X^k e^{tX}\right]$ so we obtain

$$\phi^{(k+1)}(t) = \frac{d}{dt}\phi^{(k)}(t) = \frac{d}{dt}\mathbb{E}\left[X^k e^{tX}\right] = \mathbb{E}\left[X^{k+1} e^{tX}\right]$$

        Substitute t = 0
2. Exercises on mgfs
    a. Let $\varphi_X(t)$ be the mgf of X. Let Y=aX+b. What is the mgf of Y?

$$\phi_Y(t) = \mathbb{E}\left[e^{t(aX+b)}\right] = e^{tb}\mathbb{E}\left[e^{(at)X}\right] = e^{tb}\phi_X(at)$$

b. Let X, Y be independent RVs with mgfs $\varphi_X(t)$, $\varphi_Y(t)$. What is the mgf $\varphi_{X+Y}(t)$ of the sum X+Y? What if we have the sum of n ind. variables?

$$\phi_{X+Y}(t) = \mathbb{E}\left[e^{t(X+Y)}\right] = \mathbb{E}\left[e^{tX}e^{tY}\right] = \mathbb{E}\left[e^{tX}\right]\mathbb{E}\left[e^{tY}\right] = \phi_X(t)\phi_Y(t)$$

c. What is the MGF of the normal N(0,1)?

$$\phi_{\mathcal{N}(0,1)}(X) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{tx}e^{-\frac{x^2}{2}}dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2-2tx)}dx = \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2-2tx+t^2)}dx = e^{-t^2/2}$$

PDF of e^(-0.5*(x-t)^2)) is sqrt(2π) since it is N ~ (t,1)

d. What is the mgf of X ~ Bernoulli (p)?

$$\boxed{\phi_X(t) = pe^t + (1-p)}$$

e. What is the mgf of X ~ Bin(n,p)?

∑ k = 0 to n → $_nC_k$ * p^k * (1-p)^(n-k) * e^(tk)

= ∑ k = 0 to n → $_nC_k$ * (p*e^t)^k * (1-p)^(n-k)

$$\boxed{\phi_X(t) = \prod_{i=1}^{n}\left(pe^t + (1-p)\right) = \left(1 + \left(e^t - 1\right)p\right)^n}$$

f. What is the mgf of X ~ Poisson(λ)?

$$\boxed{\phi_X(t) = \sum_{n=0}^{+\infty}\Pr(X=n)e^{tn} = \sum_{n=0}^{+\infty}\frac{e^{-\lambda}\lambda^n}{n!}e^{tn} = \sum_{n=0}^{+\infty}\frac{e^{-\lambda}\left(\lambda e^t\right)^n}{n!} = e^{\lambda(e^t-1)}}$$

3. General form of Chernoff bound

a. Let X be a random variable. Then for any $\epsilon \geq 0$, we have that

$$\boxed{\Pr(X \geq \epsilon) \leq e^{-\phi(\epsilon)}}$$ where

$$\boxed{\phi(\epsilon) = \max_{s>0}(s\epsilon - \log M_X(s)), \text{ where } M_X(s) = \mathbb{E}\left[e^{sX}\right]}$$

b. Proof (sketch)

i. We apply Markov's inequality as follows, and then we minimize LHS over all s:

$$\Pr(X \geq \epsilon) = \Pr\left(e^{sX} \geq e^{s\epsilon}\right) \leq \frac{M_X(s)}{e^{s\epsilon}} = e^{-s\epsilon + \log M_X(s)}$$

4. Chernoff bound

a. Let X_i = 1 with probability p_i, 0 with prob. 1-p_i

b. Define

$$X = \sum_{i=1}^{n} X_i, \text{ and } \mu = \mathbb{E}[X] = \sum_{i=1}^{n} p_i$$

c. Then, the following probability inequality holds

$$\boxed{\Pr(|X-\mu| \geq \delta\mu) \leq 2e^{-\frac{\delta^2}{3}\mu} \text{ for all } 0 < \delta < 1}$$

(much sharper bound by μ since it goes down exponentially) → when we know more information about the distribution unlike chebyshev (we do not specifically know the distribution and is more general)

5. Application:flipping a fair coin
    a. Suppose we flip a fair coin n times. Let X be the number of Heads
        i. Chebyshev's inequality:
        $$\Pr\left(\left|\frac{X}{n}-\frac{1}{2}\right| \geq \epsilon\right) = \Pr\left(\left|X-\frac{n}{2}\right| \geq \epsilon n\right) \leq \frac{\mathbb{V}\mathrm{ar}[X]}{n^2\epsilon^2} = \frac{1}{4n\epsilon^2}$$
        ii. Chernoff bound (for which ε is this valid? Always check the assumptions before applying a theorem)
        $$\Pr\left(\left|\frac{X}{n}-\frac{1}{2}\right| \geq \epsilon\right) = \Pr\left(\left|X-\frac{n}{2}\right| \geq \epsilon n\right) \leq 2e^{-(2\epsilon)^2\frac{n}{6}} = 2e^{-\frac{2n\epsilon^2}{3}}$$
6. Sampling theorem
    a. Chernoff bound comes in many variations
    b. A beautiful corollary of Chernoff bound is the following
        i. Sampling theorem: Given n independent 0-1 RVs Xi such that Pr(Xi=1) = p(i=1…n), where $n \geq \frac{3}{\epsilon^2}\ln\left(\frac{2}{\delta}\right)$ then the following holds:
        $$\Pr\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - p\right| \leq \epsilon\right) \geq 1-\delta$$
    c. Let's assume we want to be ε-accurate with a certain confidence 1-δ about the fraction of population that will vote for Biden:
    $$\boxed{n \geq \frac{3}{\epsilon^2}\ln\left(\frac{2}{\delta}\right)}$$
    The poll size does not depend on the size of the total population
7. The Thumbtack problem
    a. A billionaire from the suburbs of Boston asks you a question:
        i. He says: I have thumbtack, if I flip it, what's the probability it will fall with the nail up?
        ii. You say: Please flip it a few times
    b. Assumptions
        i. Pr(Heads) = p, Pr(Tails) = 1-p
        ii. Flips are iid
            1. Consider the sequence HTHHTHT → p^4(1-p)^3