DCP 1206: Probability
Lecture 20 — Moment Generating Functions and Concentration Inequalities

Ping-Chun Hsieh

November 27, 2019

How Many Weeks In a Semester?

獨 / 台灣一學期全球最長 台大、台師大、台科大擬縮短為15周

分享 分享 留言 列印 存新聞 A- A+

2019-11-25 09:36 聯合報 記者馮靖惠 / 台北即時報導 讚 2 萬 分享

國內法規定大學每學期長度是18周，校方普遍認為太僵化，無法彈性規畫課程。由台大、台師大、台科大組成的台大系統擬將一學期縮短為15周，以利與英、美、日等國際接軌，順利的話，最快109學年度上路。

台灣科技大學校長廖慶榮表示，台大系統的三校正規畫，一學期改為15周以及第三學期（暑期），主要是配合國際趨勢，讓學生可以出國當交換生，也方便國際生來台。

- In fact, we only have about 16 weeks this semester…
  - 9/13 (Friday): Mid-Autumn Festival
  - 12/4 (Wednesday): School Anniversary
  - 1/1 (Wednesday): New Year Holiday

3. Correlation Coefficient: $\rho(X_1, X_2) = \rho$

$$f_{X_1X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} \right)}{2(1-\rho^2)} \right]$$

$$\text{Cov}(X_1, X_2) = E[(X_1-\mu_1)(X_2-\mu_2)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x_1-\mu_1)(x_2-\mu_2) f_{X_1X_2}(x_1, x_2)\, dx_1 dx_2$$

- Question: Isn't this trivial by definition of bivariate normal?

Quick Review: Bivariate Normal

$$f_{X_1X_2} = \frac{1}{2\pi\sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$$

joint PDF (matrix form)

Marginal PDF

$X_1 \sim N(\mu_1, \sigma_1^2)$

$X_2 \sim N(\mu_2, \sigma_2^2)$

Conditional PDF

$$X_2 | X_1 \sim N\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(X_1-\mu_1),\ (1-\rho^2)\sigma_2^2\right)$$

X_1, X_2 bivariate normal

uncorrelated vs independence?

$\rho(X_1, X_2) = 0 \Rightarrow$ indep.

Construction from 2 independent standard normal r.v.s

<u>There is still one remaining question:</u>

Is it possible to construct
"<u>bivariate normal</u>" from "<u>normal</u>"?

Construction of Bivariate Normal R.V.

- Idea: Let Z, W be 2 independent standard normal r.v.s and define

$$
\begin{cases}
X_1 = \sigma_1 Z + \mu_1 \sim \mathcal{N}(\mu_1, \sigma_1^2) \\
X_2 = \sigma_2\left(\rho Z + \sqrt{1-\rho^2}\,W\right) + \mu_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)
\end{cases}
$$

$Var(X+Y) = Var(X) + Var(Y) + \cancel{2 \cdot Cov(X,Y)}$

- Result: X_1, X_2 are bivariate normal with joint PDF

$$
f_{X_1X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)}{2(1-\rho^2)} \right]
$$

Linear Transformation of 2 Random Variables

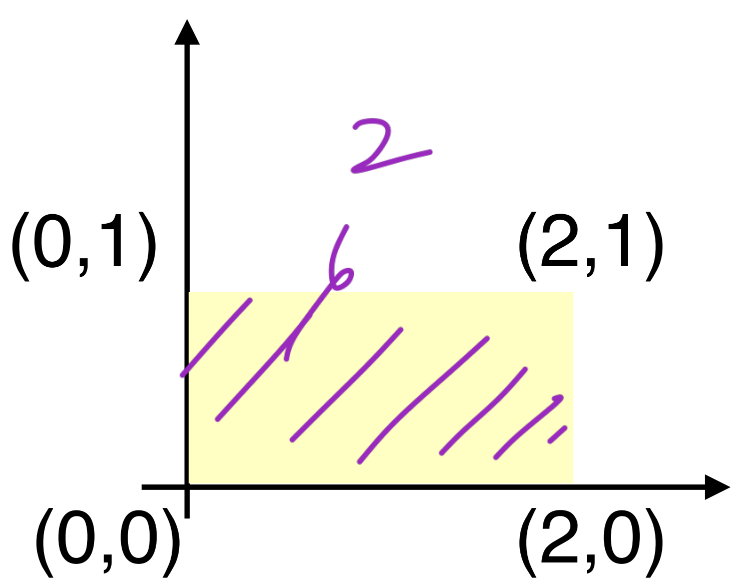
- **Theorem**: Let U_1, U_2, V_1, V_2 be random variables that satisfy $V_1 = aU_1 + bU_2$ and $V_2 = cU_1 + dU_2$. Define the matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then, we have

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}$$

$$f_{V_1V_2}(v_1, v_2) = \frac{1}{|\det(A)|} f_{U_1U_2}(A^{-1}[v_1, v_2]^T)$$

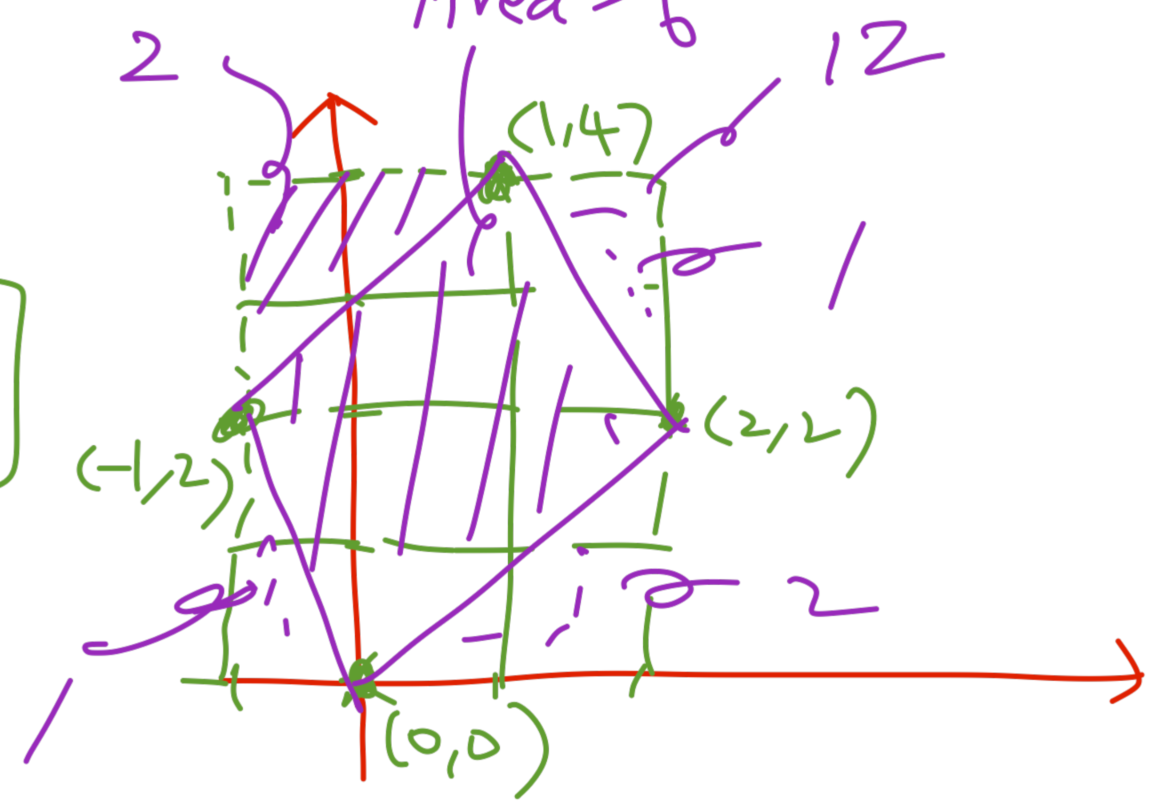
- **Intuition**:

det(A) = 3



$$\xrightarrow{A = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}} \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

Area = 6



Joint PDF of X_1 and X_2

- For simplicity, assume $\mu_1 = \mu_2 = 0$ (can be handled via translation)

$$\begin{cases} X_1 = \sigma_1 Z \\ X_2 = \sigma_2\left(\rho Z + \sqrt{1-\rho^2}W\right) \end{cases}$$

$$f_{X_1X_2}(x_1, x_2) = \frac{1}{|\det(A)|} f_{ZW}(A^{-1}[x_1, x_2]^T)$$

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \underbrace{\begin{bmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2 \cdot \sqrt{1-\rho^2} \end{bmatrix}}_{= A} \begin{bmatrix} Z \\ W \end{bmatrix}$$

$$\det(A) = \sigma_1 \sigma_2 \cdot \sqrt{1-\rho^2}$$

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} \sigma_2\sqrt{1-\rho^2} & 0 \\ -\rho\sigma_2 & \sigma_1 \end{bmatrix}$$

This Lecture

1. Sum of Independent Random Variables and Moment Generating Functions (MGF)

2. Concentration Inequalities

- Reading material: Chapter 8.4, 11.1-11.3

1. Sum of Independent Random Variables and Moment Generating Functions

$Z = X + Y$ and X, Y Independent: Discrete Case

- Question: X, Y are two independent discrete random variables.
  - Define $Z = X + Y$
  - What's the PMF of Z?

Convolution Theorem: Let X, Y be two independent discrete random variables with PMF $p_X(x)$ and $p_Y(y)$. Define $Z = X + Y$. Then, the PMF of Z is

$$p_Z(z) = P(Z = z) = \sum_x p_X(x) p_Y(z - x)$$

- Recall: $X \sim \text{Poisson}(\lambda_1, T)$ and $Y \sim \text{Poisson}(\lambda_2, T)$
  - What's the PMF of Z?

$Z = X + Y$ and X, Y Independent: Continuous Case

- Question: X, Y are two independent continuous random variables.
  - Define $Z = X + Y$
  - What's the PDF of Z?

Convolution Theorem: Let X, Y be two continuous independent random variables with PDF f_1 and f_2. Define $Z = X + Y$. Then, the PDF of Z is

convolution

$$f_Z(z) = \int_{-\infty}^{\infty} f_1(x) f_2(z - x) dx$$

Example: Sum of 2 Independent Normal R.V.s

- Example: $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$
  - X, Y are independent and define $Z = X + Y$
  - Question: What's the PDF of Z? Is Z also normal?

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) \cdot f_Y(z-x)\,dx = \int_{-\infty}^{+\infty} \left(\frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\right)\left(\frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{(z-x-\mu_2)^2}{2\sigma_2^2}}\right) dx$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sigma_1\sqrt{2\pi}} \frac{1}{\sigma_2\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{\sigma_2^2(x-\mu_1)^2 + \sigma_1^2(z-x-\mu_2)^2}{\sigma_1^2\sigma_2^2}\right]\right) dx$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sigma_1\sqrt{2\pi}} \cdot \frac{1}{\sigma_2\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{(\sigma_2^2+\sigma_1^2)x^2 + (-2\mu_1\sigma_2^2 - 2\sigma_1^2(z-\mu_2))x + (\sigma_2^2\mu_1^2 + \sigma_1^2(z-\mu_2)^2)}{\sigma_1^2\sigma_2^2}\right]\right) dx$$

$\ldots\ldots$

Any Issue With Convolution Theorem?

- Issue: Sometimes it is quite tedious to do convolution
- Question: Any other approach?

- Idea: Borrow ideas from signal processing — Laplace transform

$$x_{in}(\tau) \longrightarrow \boxed{\overset{h(\tau)}{\text{Channel}}} \longrightarrow x_{out}(\tau) \overset{\text{Laplace transform}}{\Longrightarrow}$$

$$\underline{X_{out}(t)} = \int_{-\infty}^{+\infty} x_{out}(\tau) e^{-t\tau} d\tau$$

$$\vdots$$

$$x_{out}(\tau) = \int_{-\infty}^{+\infty} x_{in}(s) \cdot h(\tau - s) ds$$

$$X_{out}(t) = X_{in}(t) \cdot H(t)$$

Moment Generating Function (Formally)

- **Moment Generating Function (MGF)**: For a random variable X, define

$$M_X(t) = E[e^{tX}], \; t \in \mathbb{R}$$

−0.0001, for any

If there exists $\delta > 0$ such that $M_X(t) < \infty$ for all $t \in (-\delta, \delta)$, then $M_X(t)$ is called the moment generating function of X

- Remark: If X is discrete with PMF $p_X(x)$, then

$$M_X(t) = \sum_{\text{all } x} \left(e^{tx} \right) \left(p_X(x) \right)$$

- Remark: If X is continuous with PDF $f_X(x)$, then

$$M_X(t) = \int_{-\infty}^{+\infty} \left(e^{tx} \right) \left(f_X(x) \right) dx$$

Example: MGF of Poisson Random Variables

$$e^z = \sum_{x=0}^{\infty} \frac{z^x}{x!}$$

- **Example**: Let $X \sim \text{Poisson}(\lambda, T)$
  - **Question**: What is the MGF of X, $M_X(t)$ = ?

$$M_X(t) = E[e^{tX}] = \sum_{x=0}^{\infty} e^{tx} \cdot p_X(x)$$

$$= \sum_{x=0}^{\infty} (e^t)^x \cdot \frac{e^{-\lambda T} (\lambda T)^x}{x!}$$

$$= e^{-\lambda T} \sum_{x=0}^{\infty} \frac{(\lambda T \cdot e^t)^x}{x!} = e^{\boxed{\lambda T \cdot e^t}}$$

$$= e^{\lambda T (e^t - 1)}, \quad t \in \mathbb{R}$$

Example: Find MGF of Normal Random Variables

- Example: Let $Z \sim \mathcal{N}(\mu, \sigma^2)$
  - Question: What is the MGF of Z, $M_Z(t) = ?$

$e^{t\mu + \frac{1}{2}\sigma^2 t^2}$

$$M_Z(t) = E\left[e^{tZ}\right] = \int_{-\infty}^{+\infty} e^{tz} \cdot f_Z(z)\, dz$$

$$= \int_{-\infty}^{+\infty} e^{tz} \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-\mu)^2}{2\sigma^2}} \right) dz$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{z^2 - 2(\mu + t\sigma^2) z + (\mu^2)}{2\sigma^2}} dz$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-(\mu+t\sigma^2))^2 + \mu^2 - (\mu+t\sigma^2)^2}{2\sigma^2}} dz$$

$$= \left(e^{-\frac{\mu^2 - (\mu+t\sigma^2)^2}{2\sigma^2}} \right) \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-(\mu+t\sigma^2))^2}{2\sigma^2}} dz$$

$$= e^{t\mu + \frac{1}{2}\sigma^2 t^2} \cdot 1$$

Nice Properties of MGF?

- **Let X_1, X_2 be two random variables**:

$e^{t\mu+\frac{1}{2}t^2\sigma^2}$

1. ______ Suppose $M_{X_1}(t) = M_{X_2}(t)$, for all $t \in \mathbb{R}$. Do X_1 and X_2 always have the same distribution (i.e. the same CDF)?

2. ______ Could we find moments $E[X_1^n]$ by using $M_{X_1}(t)$?

3. ______ Suppose X_1, X_2 are independent. Could we express $M_{X_1+X_2}(t)$ in terms of $M_{X_1}(t), M_{X_2}(t)$?

Nice Property (I): MGF Uniqueness Theorem

- **MGF Uniqueness Theorem**: Let X_1 and X_2 be two random variables with MGFs $M_{X_1}(t)$ and $M_{X_2}(t)$, respectively. If $M_{X_1}(t) = M_{X_2}(t)$ for all t in some interval $(-\alpha, \alpha)$, then X_1 and X_2 follow the same distribution, i.e.

$$P(X_1 \leq u) = P(X_2 \leq u), \text{ for all } u \in \mathbb{R}$$

- Remark: More details in the following reference
  - J. H. Curtiss, "A note on the theory of moment generating functions," 1942
  - https://projecteuclid.org/download/pdf_1/euclid.aoms/1177731541

Example: Find CDF from MGF

- Example: Suppose the MGF of a random variable X is

$$M_X(t) = \frac{1}{6}e^{-2t} + \frac{1}{3}e^{-t} + \frac{1}{4}e^{t} + \frac{1}{4}e^{2t}$$

  - Question: $P(|X| \leq 1) = ?$

$$M_X(t) = E[e^{tX}] = \sum p_X(x) e^{tx} \quad \text{or} \quad \int f_X(x) \cdot e^{tx} dx$$

$$p_X(x) = \begin{cases} \frac{1}{6}, & x = -2 \\ \frac{1}{3}, & x = -1 \\ \frac{1}{4}, & x = +1 \\ \frac{1}{4}, & x = +2 \end{cases}$$

(By Uniqueness Theorem).

MGF of Special Random Variables

Distribution	Moment-generating function $M_X(t)$
Degenerate δ_a	e^{ta}
Bernoulli $P(X=1)=p$	$1-p+pe^t$
Geometric $(1-p)^{k-1}\,p$	$\frac{pe^t}{1-(1-p)e^t}$ $\forall t < -\ln(1-p)$
Binomial $B(n,p)$	$(1-p+pe^t)^n$
Negative Binomial $NB(r,p)$	$\frac{(1-p)^r}{(1-pe^t)^r}$
Poisson $Pois(\lambda)$	$e^{\lambda(e^t-1)}$
Uniform (continuous) $U(a,b)$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$
Uniform (discrete) $DU(a,b)$	$\frac{e^{at}-e^{(b+1)t}}{(b-a+1)(1-e^t)}$
Laplace $L(\mu,b)$	$\frac{e^{t\mu}}{1-b^2t^2}$, $\lvert t\rvert < 1/b$
Normal $N(\mu,\sigma^2)$	$e^{t\mu+\frac{1}{2}\sigma^2t^2}$

$$e^{t\mu+\underline{\frac{1}{2}\sigma^2t^2}}$$

- **Example:** If $M_X(t) = \frac{1}{2} + \frac{1}{2}e^t$, then what kind of r.v. is X?

  $X \sim \text{Bernoulli}\left(\frac{1}{2}\right)$

- **Example:** If $M_Z(t) = e^{2t^2-t}$, then what kind of r.v. is Z?

  $Z \sim N(-1, 4)$

Nice Property (II): From Sum to Product

- **MGF and Sum of 2 Independent Random Variables**: Given 2 independent random variables X_1, X_2 with MGFs $M_{X_1}(t)$ and $M_{X_2}(t)$, the MGF of $X_1 + X_2$ is

$$M_{X_1+X_2}(t) = M_{X_1}(t) \cdot M_{X_2}(t)$$

cf: Laplace transform

$X_{out} = X_{in} \cdot H$

- Proof:

$$M_{X_1+X_2}(t) = E[e^{t(X_1+X_2)}]$$
$$= E[e^{tX_1} \cdot e^{tX_2}]$$

by independence

$$= E[e^{tX_1}] \cdot E[e^{tX_2}]$$
$$= M_{X_1}(t) \cdot M_{X_2}(t)$$

X, Y are indep.

$E[h(X) \cdot g(Y)]$

$= E[h(X)] \cdot E[g(Y)]$

Nice Property (II): From Sum to Product (Cont.)

- **MGF and Sum of n Independent Random Variables**: Given n independent random variables $X_1, X_2, \cdots, X_n$ with MGFs $M_{X_1}(t), \cdots, M_{X_n}(t)$, the MGF of $X_1 + X_2 + \cdots + X_n$ is

$$M_{X_1+X_2+\cdots+X_n}(t) = M_{X_1}(t)M_{X_2}(t)\cdots M_{X_n}(t)$$

$E[g_1(X_1) \cdot g_2(X_2) \cdots g_n(X_n)]$

$= E[g_1(X_1)] \cdot E[g_2(X_2)] \cdots E[g_n(X_n)]$

- Question: What property do we need to prove this?

Example: MGF of Sum of n Poisson R.V.s

- Example: $X_i \sim$ Poisson(λ_i, T), for $i = 1,2,\cdots,n$
  - $X_1, X_2, \cdots, X_n$ are assumed to be independent
  - Question: What is the MGF of $X_1 + X_2 + \cdots + X_n$?

$$M_{X_1}(t) = e^{\lambda_1 T(e^t - 1)}$$

$$M_{X_i}(t) = e^{\lambda_i T(e^t - 1)}$$

$$M_{X_1+X_2+\cdots+X_n}(t) = \left(e^{\lambda_1 t(e^t-1)}\right)\left(\cdots\right)\cdots\left(\quad\right)$$

$$= e^{(\lambda_1+\lambda_2+\cdots+\lambda_n)t\cdot(e^t-1)}.$$

Poisson$(\lambda_1+\lambda_2+\cdots+\lambda_n, T)$

(By Uniqueness Theorem)

Example: MGF of Sum of n Normal R.V.s

- Example: $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, for $i = 1,2,\cdots,n$
  - $X_1, X_2, \cdots, X_n$ are assumed to be independent
  - Question: What is the MGF of $X_1 + X_2 + \cdots + X_n$?

$$M_{X_1}(t) = e^{t\mu_1 + \frac{1}{2}\sigma_1^2 t^2}$$

$$M_{X_i}(t) = e^{t\mu_i + \frac{1}{2}\sigma_i^2 t^2}$$

$$M_{X_1+\cdots+X_n}(t) = e^{t(\mu_1+\mu_2+\cdots+\mu_n) + \frac{1}{2}t^2(\sigma_1^2+\cdots+\sigma_n^2)}$$

Nice Property (III): Why Is $M_X(t)$ Called the Moment Generating Function?

- Recall: What is the "n-th moment" of X?

- **Use MGF to Find Moments**: Let X be a random variable with MGF $M_X(t)$. Then, for every $n \in \mathbb{N}$, we have

$$E[X^n] = \frac{d^n}{dt^n} M_X(t)\big|_{t=0}$$

- Proof:

Example: Moments of Exp(λ)

- Example: Suppose $X \sim \text{Exp}(\lambda)$
  - What is the MGF of X?
  - Use MGF to verify that $E[X] = \dfrac{1}{\lambda}$ and $\text{Var}[X] = \dfrac{1}{\lambda^2}$?

2. Concentration Inequalities

Example: Tossing Moon Blocks

- 3 possible outcomes: Yes / No / Laughing
- $p = P(\text{outcome is "Yes"})$
- Each toss is <u>independent</u> from other tosses

- Question: Suppose p is unknown
  - How to learn p?
  - Could we learn anything useful after n experiments?

Concentration Inequalities

Markov's Inequality

- **Markov's Inequality**: Let X be a nonnegative random variable. Then, for any $t > 0$,

$$P(X \geq t) \leq \frac{E[X]}{t}$$

- Visualization:

Proof of Markov's Inequality

- **Markov's Inequality**: Let X be a <u>nonnegative</u> random variable. Then, for any $t > 0$,

$$P(X \geq t) \leq \frac{E[X]}{t}$$

- Proof:

Chebyshev's Inequality

- **Chebyshev's Inequality**: Let X be a random variable with mean μ and variance σ^2. Then, for any $t > 0$,

$$P(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}$$

- Proof:

Quick Review: Mean and Variance of Sum of Independent Random Variables

- Example: Each X_i has mean μ_i and variance σ_i^2
  - $X_1, X_2, \cdots, X_n$ are assumed to be independent
  - Question 1: $E[X_1 + X_2 + \cdots + X_n]$ =
  - Question 2: $E[\frac{1}{n}(X_1 + X_2 + \cdots + X_n)]$ =

Quick Review: Mean and Variance of Sum of Independent Random Variables (Cont.)

- Example: Each X_i has mean μ_i and variance σ_i^2
  - $X_1, X_2, \cdots, X_n$ are assumed to be independent
  - Question 3: Var$[X_1 + X_2 + \cdots + X_n]$ =
  - Question 4: Var$[\frac{1}{n}(X_1 + X_2 + \cdots + X_n)]$ =

Chebyshev's Inequality and Sample Mean

- Example: Tossing moon blocks

- Each toss X_i is Bernoulli with $P(\text{outcome is "Yes"}) = p$
- Each toss is <u>independent</u> from other tosses
- Question: Can we say anything about the sample mean of n tosses $\dfrac{1}{n}(X_1 + \cdots + X_n)$?

Chebyshev's Inequality and Sample Mean (Formally)

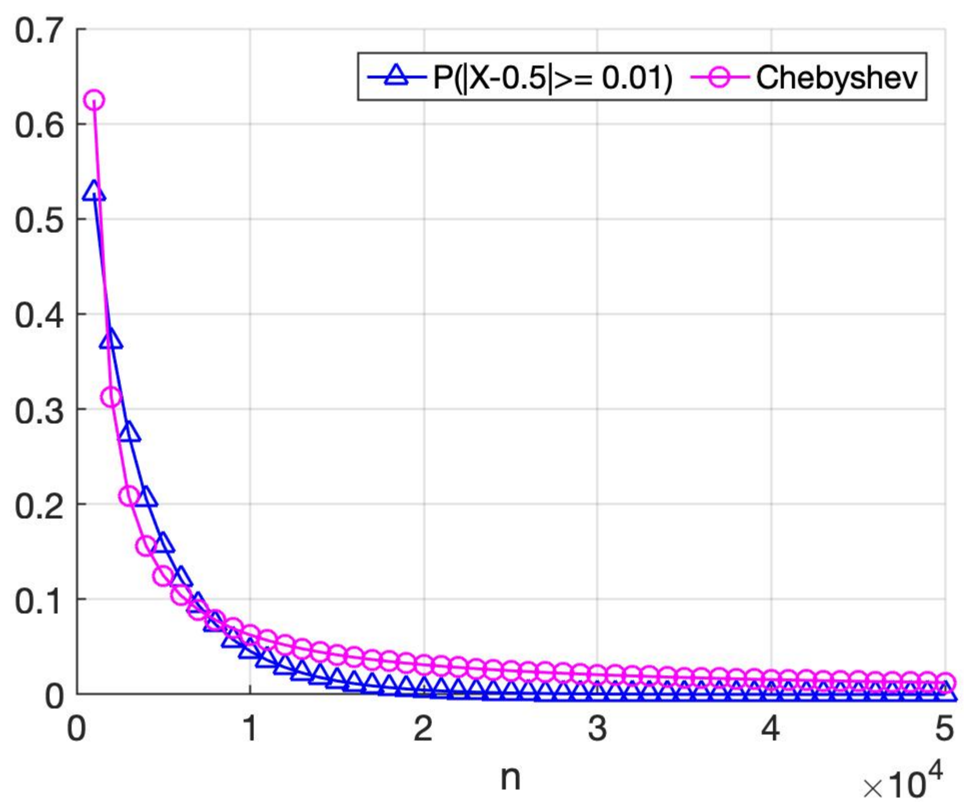
- **Chebyshev's and Sample Mean**: Let $X_1, \cdots, X_n$ be a sequence of independent and identically distributed (i.i.d.) random variables with mean μ and variance σ^2. Define $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$. Then, for any $\varepsilon > 0$, we have

$$P(|\bar{X} - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2 n}$$

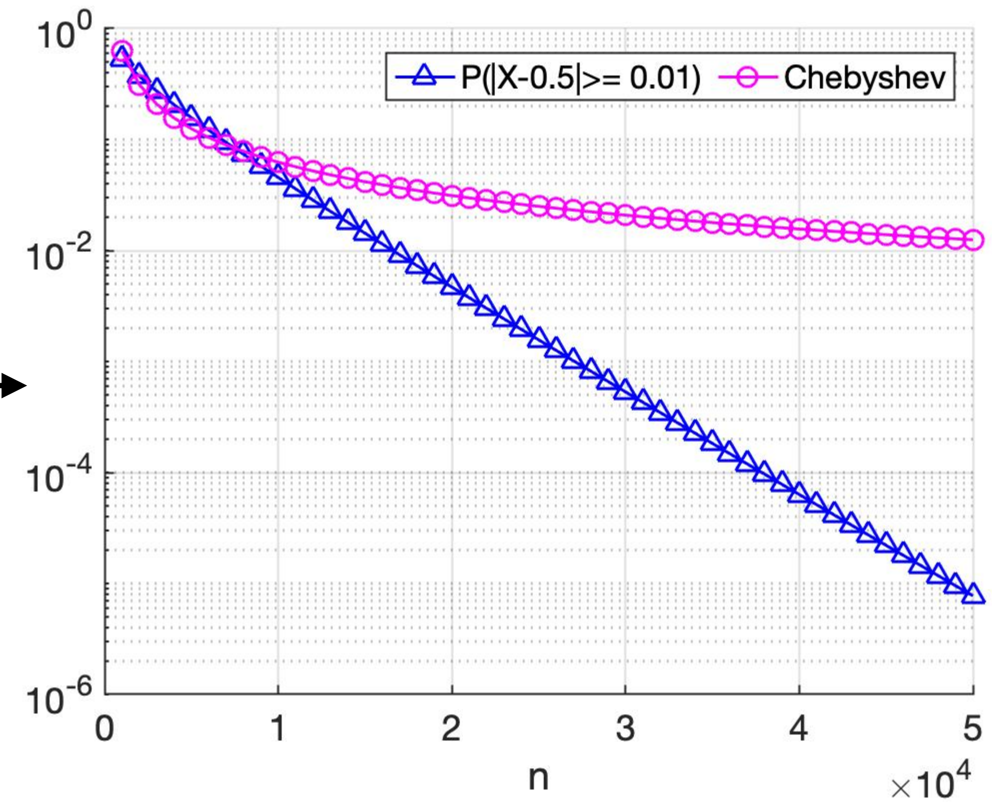
Any Issue With Chebyshev's Inequality?

- Example: $X_1, \cdots, X_n$ are i.i.d. Bernoulli with parameter 0.5
  - $E[X_i] =$ ________ and Var$[X_i] =$ _________
  - Chebyshev's: $P(|\bar{X} - \mu| \geq \varepsilon) \leq \dfrac{\sigma^2}{\varepsilon^2 n}$
  - Let's plot $P(|\bar{X} - \mu| \geq \varepsilon)$ for small ε

$\underline{\varepsilon = 0.01}$



log scale



Next Lecture

- Chernoff Bound
- Law of Large Numbers

1-Minute Summary

1. Sum of Independent Random Variables and Moment Generating Functions (MGF)

- Convolution theorem
- MGF = $E[e^{tX}]$
- MGF uniquely determines the CDF
- Find $E[X^n]$ using MGF

2. Concentration Inequalities

- Markov's and Chebyshev's Inequalities