Welcome to Stat 410!

- Personnel
  - Instructor: Liang, Feng
  - TA: Gan, Gary (Lingrui)
  - Instructors/TAs from two other sessions
- Websites: Piazza and Compass
- Homework
  - When, where and how to submit your homework
  - No late submissions will be accepted
  - Remember to list your section and instructor on the 1st page
  - Grading policy
- Three Exams, No Final.

Will keep you posted on

- where to pickup your graded hw;
- rules for reporting grading errors;
- rules for missing homework.

Communication

- For questions related to homework/lectures, please post your question on Piazza. By default, you are anonymous to your classmates, but not to the instructor.

- If you want to send email to my Illinois account, please

  1. Write from your Illinois email account (so I would know who you are)

  2. Start your subject line with the course number, e.g., "[stat410] HW1 missing" (since I'm teaching two courses this semester)

  3. Sign with your full name

  4. Don't send unexpected attachments.

General Expectations

- Review the notes/book
- Go through practice problems
- Finish homework independently

  You can discuss homework problems with other students but should write your answers independently using your own words.

- Feedback and questions
- How to do well: Exercise, Exercise, Exercise

Discrete Random Variables

- $X \sim$ Bern(p): $X = 0$ or 1

  $\mathbb{P}(X = 1) = p$, $\mathbb{P}(X = 0) = 1 - p$.

  Toss a fair coin, $X = 1$ if head; 0 if tail, then $X \sim$ Bern(0.5).

- $X \sim$ Bin(n, p): $X =$ number of 1's from n independent Bern(p) trials.

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, k = 0, 1, \ldots, n$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ is the binomial coefficient represents the number of ways to select (unordered) k objects out of n given objects.

Note

$$\binom{n}{k} = \binom{n}{n-k}, \quad \binom{n}{0} = \binom{n}{n} = 1, \quad \binom{n}{1} = \binom{n}{n-1} = n$$

Using Binomial Theorem:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{n-k} a^k b^{n-k},$$

check that $\sum_{k=0}^{n} \mathbb{P}(X = k) = (p + 1 - p)^n = 1$.

- $X \sim$ Geo(p): $X =$ number of 0's from independent Bern(p) trials before seeing the first 1.

$$\mathbb{P}(X = k) = (1-p)^k p, \quad k = 0, 1, \ldots$$

Sometimes Geo(p) is described as $\mathbb{P}(X = m) = (1-p)^{m-1}p$, $m = 1, 2, \ldots$, i.e., $X =$number of independent Bern(p) trials before seeing the first 1.

- $X \sim$ NB(r, p): $X =$ number of 0's from independent Bern(p) trials before seeing r 1's.

- $X \sim \text{Po}(\lambda)$: X can be viewed as the limit of Bern(n, p) when n is large and p is small and $np \approx \lambda$.

  For example, $X =$ number of Stat 410 students who have the same birthday as you.

$$
\begin{aligned}
& \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
= \quad & \frac{1}{k!} \frac{n!}{(n-k)!} \frac{p^k}{(1-p)^k} (1-p)^n \\
= \quad & \frac{1}{k!} [np] [(n-1)p] [(n-k+1)p] \frac{1}{(1-p)^k} (1-p)^n \\
\to \quad & \frac{1}{k!} \lambda^k e^{-\lambda}, \quad k = 0, 1, 2, \ldots
\end{aligned}
$$

- A jar contains 20 M&M candies and 7 red and 13 green.

- Sampling with replacement: randomly draw one M&M from the jar, record its color and put it back, and then repeat this 5 times. $X =$ number of red candies you've drew, which follows Bin$(5, \frac{7}{20})$.

- Sampling without replacement: randomly draw one M&M from the jar, record its color and repeat this 5 times, i.e., randomly draw 5 candies from the jar. $X =$ number of red candies you've drew, which follows a Hypergometric distribution $(20, 7, 5)$.

- Uniform distribution over a discrete set, e.g., tossing a fair die will result a uniform distribution over set $\{1, 2, \ldots, 6\}$.

Continuous Random Variables

- Uniform distribution, e.g., Unif$(0, 1)$.

- Exponential distribution

$$\text{pdf} \quad f(x) = \lambda e^{-\lambda x}, \quad x > 0,$$

  which is a special case of the Gamma distribution.

  Sometimes parameterized as $f(x) = \frac{1}{\beta} e^{-x/\beta}$.

- Normal distribution: THE most important distribution in Statistics.

  Student t dist, F dist, and Chi-squared dist are related to the Normal distribution.

- Beta distribution over $(0, 1)$. Unif$(0, 1)$ is a special case of Beta.

Random Variables

- How to describe a random variable? pmf/pdf or CDF.
- Properties of pmf/pdf $f(x)$:
  1. $f(x) \geq 0$;
  2. its integral/summation over the range is equal to 1.
- Properties of CDF $F(x) = \mathbb{P}(X \leq x)$ (Theorem 1.5.1 , p.35)
  1. F is nondecreasing, i.e., $F(a) \leq F(b)$ if $a < b$.
  2. $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$.
  3. F is right continuous, i.e., $\lim_{x_n \downarrow a} F(x_n) = F(a)$.
- Mean, median (or quantiles), and mode of a distribution

Expectations

$$\mathbb{E}(X) = \begin{cases} \sum_x x f(x) \\ \int x f(x) dx \end{cases}, \quad \mathbb{E}[g(X)] = \begin{cases} \sum_x g(x) f(x) \\ \int g(x) f(x) dx \end{cases},$$

provided that those integrals and summations exist (i.e., $\mathbb{E}|X|$ or $\mathbb{E}|g(X)|$ finite).

- $\mathbb{E}(a) = a$, where a is a constant.
- $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$ ($\mathbb{E}$ is a linear operator)
- k-th moment $= \mathbb{E}(X^k)$, $\mu = \mathbb{E}(X)$,

  $\text{Var}(X) = \sigma^2 = \mathbb{E}(X - \mu)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2$.
- What does a rv with zero variance look like?
- Usually $\mathbb{E}g(X) \neq g(\mathbb{E}X)$ when g is non-linear ($\mathbb{E}X^2 \neq \left[\mathbb{E}(X)\right]^2$)

Note and Tips

- Random variables are either discrete or continuous? NO
- In the discrete case, pmf $f(x) = \mathbb{P}(X = x)$, while in the continuous case, pdf $f(x)$ is NOT equal to $\mathbb{P}(X = x)$ that is equal to zero.
- When computing probabilities involving continuous rvs, "$\leq$" and "$<$" (or "$\geq$" and "$>$") are interchangeable, but that's not the case for discrete rvs.
- Range! When describing a rv, remember to specify its range; when computing expectations, keep track of the range.

Moment Generating Functions

$$M_X(t) = \mathbb{E}e^{tX} = \sum_x e^{tx} \cdot f(x) \quad \text{or} \quad \int e^{tx} f(x) dx,$$

provided that the expectation exists for t in a small neighborhood of 0, otherwise we say that the mgf does not exist.

We don't care about the mgf for any t, but t in an interval around zero, i.e., $-h < t < h$ for some $h > 0$.

The connection between moments and mgf:

$$\begin{aligned} M_X(t) &= \mathbb{E}e^{tX} = \mathbb{E}\Big[1 + \frac{tX}{1} + \frac{t^2X^2}{2!} + \frac{t^3X^3}{3!} + \cdots\Big] \\ &= 1 + t\frac{\mathbb{E}X}{1} + t^2\frac{\mathbb{E}X^2}{2!} + t^3\frac{\mathbb{E}X^3}{3!} + \cdots \end{aligned} \tag{1}$$

Then, we have

$$M_X^{(k)}(0) = \mathbb{E}(X^k), \quad k = 0, 1, ...$$

Another important use of mgf: it provides an alternative way to characterize a distribution.

If the mgf for X_1 and the one for X_2 exist, and $M_{X_1}(t) = M_{X_2}(t)$

$$\Longrightarrow X_1, X_2 \text{ follow the same distribution.}$$

That is, the mgf uniquely determines a distribution.

If we know all the moments of a r.v., then we know its distribution? Or in other words, do moments uniquely determine a distribution?

The answer is No. You might be tempted to say Yes. If the moments are not too large, the mgf defined via (1) exists, then moments determine a distribution.