Wednesday, November 29 **Fall 2023**

Section 10: Adam's & Eve's Laws, Inequalities, and Limit Theorems

Srihari Ganesh

Based on section note formatting template by Rachel Li and Ginnie Ma '23

Forms

- Attendance form: http://bit.ly/110attend
- Feedback form: https://bit.ly/SrihariFeedback

Attendance form

Section feedback form

1 Summary

1.1 Conditional Expectation, Adam's & Eve's Laws

Rehash:

Definition 1 (Conditional expectation given event)**.** The **conditional expectation** of a random variable X given an event A is

$$E(X|A) = \begin{cases} \sum_x P(X = x|A) & X \text{ discrete,} \\ \int_{-\infty}^{\infty} xf(x|A)dx & X \text{ continuous.} \end{cases}$$

Definition 2 (Law of total expectation)**.** Suppose $A_1, A_2, \ldots$ are events partitioning the sample space. Then the **law of total expectation (LOTE)** states that

$$E(X) = \sum_j E(X|A_j)P(A_j).$$

New content:

Definition 3 (Conditional expectation given random variable)**.** For random variables X, Y, $E(Y|X)$ is the **conditional expectation of Y given X**. If $E(Y|X = x) = g(x)$, then $E(Y|X) = g(X)$.

Definition 4 (Conditional variance given random variable)**. Conditional variance** can be defined through conditional expectation:

$$\text{Var}(Y|X) = E[(Y - E[Y|X])^2|X] = E[Y^2|X] - (E[Y|X])^2.$$

☣ **5.** Note that $E(Y|X)$ is a random variable, where X is the only source of randomness. $E(Y|X) = X^2$ is a valid conditional expectation, while $E(Y|X) = Y$ is not a valid conditional expectation. Similarly, $\text{Var}(Y|X)$ is a random variable that is a function of X alone.

Example 6. Let $X \sim \text{Pois}(\lambda)$ and $Y|X = x \sim \mathcal{N}(x^2 - 5, \sigma^2)$. Then since $E(Y|X = x) = x^2 - 5$, $E(Y|X) = X^2 - 5$, which is a random variable that is a function of X.

Example 7. Let $X \sim \text{Expo}(\lambda)$ and $Y|X = x \sim \mathcal{N}(\mu, X)$. Then since $\text{Var}(Y|X = x) = x$, $\text{Var}(Y|X) = X$.

Result 8 (Adam's Law)**.** For random variables X and Y, **Adam's law** (also known as law of total expectation, tower rule, law of iterated expectation, etc.) states that

$$E(Y) = E(E(Y|X)).$$

Result 9 (Eve's Law)**.** For ranodm variables X and Y, **Eve's law** (also known as EVVE, law of total variance, etc.) states that

$$\text{Var}(Y) = E(\text{Var}(Y|X)) + \text{Var}(E(Y|X))$$

1.2 Inequalities

For random variables X, Y,

Name	Formula	Conditions
Cauchy-Schwarz	$\lvert E(XY)\rvert \leq \sqrt{E(X^2)E(Y^2)}$	X, Y have finite variance
Jensen	$E(g(X)) \geq g(E(X))$	g convex
	$E(g(X)) \leq g(E(X))$	g concave
Markov	$P(\lvert X\rvert \geq c) \leq \frac{E\lvert X\rvert}{c}$	$c > 0$
Chebyshev	$P(\lvert X - E(X)\rvert \geq c) \leq \frac{\text{Var}(X)}{c^2}$	$c > 0$

Definition 10 (Convexity and concavity)**.** A function g (defined on an interval I) is **convex** by any of the following equivalent definitions:

- $g(px_1 + (1-p)x_2) \leq pg(x_1) + (1-p)g(x_2)$ for $x_1, x_2 \in I$ and $p \in (0, 1)$,
- Any line segment connecting two points on the graph of g is always *at or above* the graph of g,
- If the second derivative exists, $\frac{d^2 g(x)}{dx^2} \geq 0$.

A function g is **concave** if $-g$ is convex. Specifically,

- Any line segment connecting two points on the graph of g is always **at or below** the graph of g,
- If the second derivative exists, $\frac{d^2 g(x)}{dx^2} \leq 0$.

You may have previously seen "concave up" for convex and "concave down" for concave. Note that linear functions are both convex and concave because their second derivatives are 0 everywhere.

See A.2.4 in Blizstein & Hwang for details and Figure A.3 for a visual.

Remark 11 (On Jensen's inequality)**.** To remember the direction of Jensen's inequality, I use the fact that X^2 is a convex function of X. I know that $E(X^2) \geq (E(X))^2$ by the fact that variance is nonnegative, which means $E(g(X)) \geq g(E(X))$ for g convex.
Equality, $E(g(X)) = g(E(X))$, only holds if g is a linear function, like $g(X) = a + bX$. This is consistent with our previous results because linear functions are both convex and concave, and also follows from the linearity of expectation.

1.3 Limit theorems

Let $X_1, X_2, X_3, \ldots$ be i.i.d. with $E(X_1) = \mu$ and $\text{Var}(X_1) = \sigma^2$, where both mean and variance are finite. Notate $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$.

Remark 12. Note that $\bar{X}_n$ is a random variable with mean μ and variance $\frac{\sigma^2}{n}$.

Result 13 (Law of Large Numbers)**.** A **law of large numbers (LLN)** states that as n gets large, $\bar{X}_n$ converges to μ. Note that there are many LLNs (e.g., for some types of non-i.i.d. r.v.s), but we only discuss the following:

- **Strong LLN**: $P(\bar{X}_n \to \mu) = 1$. Since the random variables are defined on some sample space S, this is saying that $P(\{s \in S : X_n(s) \to \mu\}) = 1$. This is known as **almost sure** convergence.
- **Weak LLN**: For all $\epsilon > 0$, $P(|\bar{X}_n - \mu| > \epsilon) \to 0$ as n gets large. This is known as **convergence in probability**.

Result 14 (Central Limit Theorem)**.** A central limit theorem states that as n gets large,

$$\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \to \mathcal{N}(0,1) \text{ in distribution.}$$

As before, there are many variants of such results but we only discuss the i.i.d. case here. We can use this to get approximate distributions of $\bar{X}_n$ for large n:

$$\bar{X}_n \dot{\sim} \mathcal{N}(\mu, \frac{\sigma^2}{n}).$$

☣ **15.** Note that we cannot say $\bar{X}_n$ *converges* to $\mathcal{N}(\mu, \sigma^2/n)$ because that would leave an n in the limit (which is a problem because limits as n approaches as value should not involve n). We could, however, say that $\sqrt{n}(\bar{X}_n - \mu) \to \mathcal{N}(0, \sigma^2)$ in distribution.

2 Practice Problems

TBD