

## Section 10: Adam's &amp; 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

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

**Definition 2** (Law of total expectation). Suppose  $A_1, A_2, \dots$  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 random 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	$ E(XY)  \leq \sqrt{E(X^2)E(Y^2)}$	$X, Y$ have finite variance
Jensen	$E(g(X)) \geq g(E(X))$ $E(g(X)) \leq g(E(X))$	$g$ convex $g$ concave
Markov	$P( X  \geq c) \leq \frac{E X }{c}$	$c > 0$
Chebyshev	$P( X - E(X)  \geq c) \leq \frac{\text{Var}(X)}{c^2}$	$c > 0$

**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,  $\frac{d^2g(x)}{dx^2} \geq 0$ .

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,  $\frac{d^2g(x)}{dx^2} \leq 0$ .

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 Blitzstein & 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.

**Remark 12** (Identifying which inequality to use). The four inequalities in this section are essentially the only inequalities we assess (some specific cases like the boundedness of correlation and non-negativity of variance can come up as well). Here are some thoughts on my first-glance thoughts where I try to identify which of the four inequalities to use. Keep in mind that at the end of the day, there are few enough that you can try all of them out.

- When there is a single expectation on one side (e.g.,  $E(XYZ)$ ) and multiple expectations on the other side of a potential inequality (e.g.,  $E(X)E(Y)E(Z)$ ), I look to Cauchy-Schwarz
- If I see an expectation of fancy function on one side of an inequality (e.g.,  $E(\log Xe^X)$ ) and a fancy function of an expectation on the other side (e.g.,  $\log(E(X))e^{E(X)}$ ), I look to Jensen's inequality.
- If there is a probability on one side and an expectation/constant on the other side of an inequality, I look to Markov/Chebyshev

### 1.3 Limit theorems

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

**Remark 13.** Note that  $\bar{X}_n$  is a random variable with mean  $\mu$  and variance  $\frac{\sigma^2}{n}$ .

**Result 14** (Law of Large Numbers). A **law of large numbers (LLN)** states that as  $n$  gets large,  $\bar{X}_n$  converges to  $\mu$ . 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:**  $P(\bar{X}_n \rightarrow \mu) = 1$ . Since the random variables are defined on some sample space  $S$ , this is saying that  $P(\{s \in S : X_n(s) \rightarrow \mu\}) = 1$ . This is known as **almost sure** convergence, notated  $\bar{X}_n \xrightarrow{a.s.} \mu$ .
- **Weak LLN:** For all  $\epsilon > 0$ ,  $P(|\bar{X}_n - \mu| > \epsilon) \rightarrow 0$  as  $n$  gets large. This is known as **convergence in probability**, notated  $\bar{X}_n \xrightarrow{p} \mu$ .

**Result 15** (Central Limit Theorem). A central limit theorem states that as  $n$  gets large,

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

Convergence in distribution is notated as  $\xrightarrow{D}$ , so we could say  $\sqrt{n} \left( \frac{\bar{X}_n - \mu}{\sigma} \right) \xrightarrow{D} \mathcal{N}(0, 1)$ . 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  $n$ :

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

⚠ **16.** 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  $\sqrt{n}(\bar{X}_n - \mu) \rightarrow \mathcal{N}(0, \sigma^2)$  in distribution.

## 2 Practice Problems

1. (*inspired by HW 5.3*) [**Adam's & Eve's laws**] Suppose George has  $N \sim \text{Pois}(\lambda)$  children in his lifetime, and his  $i$ -th child has  $G_i \sim \text{Pois}(\lambda)$  children themselves, and  $G_1, \dots, G_N, N$  are all independent.

(a) Find the expected number of grandchildren that George has.

(b) Find the variance of the number of grandchildren George has.

2. **[Limit Theorems]** Let  $Y_n \sim \text{Bin}(n, p)$ .

- (a) Use the Central Limit Theorem to find the asymptotic distribution of  $\frac{1}{\sqrt{n}}Y_n - p\sqrt{n}$  as  $n \rightarrow \infty$ .

*Hint: recall that we can showed the central limit theorem for  $\sqrt{n}(\bar{X}_n - \mu)$  where  $\bar{X}_n$  is the average of i.i.d.  $X_1, \dots, X_n$  with  $\mu = E(X_1)$ . Apply the CLT to some expression involving  $Y_n$  using the story of the Binomial and rearrange it to get  $\frac{1}{\sqrt{n}}Y_n - p\sqrt{n}$ .*

- (b) Apply the Law of Large Numbers to find what  $Y_n/n$  converges to as  $n \rightarrow \infty$ .

3. (*William Chen and Sebastian Chiu 2013*) **[Inequalities]** Fill each inequality below with either  $=$ ,  $\leq$ ,  $\geq$ , or  $?$ . In all instances below, assume that  $X$  and  $Y$  are positive random variables, although not necessarily independent. Assume that the expected values exist.

(a)

$$E(X^4) \_ \sqrt{E(X^2)E(X^6)}$$

(b)

$$P(|X + Y| > 2) \_ \frac{1}{16}E((X + Y)^4)$$

(c)

$$\sqrt{E(X) + 50} \_ E(\sqrt{X + 50})$$

(d)

$$E(Y|10X) - E(Y|X)$$

(e)

$$E(\cos(X)) - \cos(E(X))$$