ECE 3100 - Functions, Formulas, and Definitions

Stephen Chin

Spring Semester 2019

Contents

1 Pre - Prelim 1

1.1 Lecture 1 - What is Probability?

Probability is a way of mathematically modelling situations involving uncertainty with the goal of making predications decisions and models. Probability can be understood in many ways, such as:

1. Frequency of Occurence: Or percentage of successes in a moderately large number of similar situations.

2. Subjective belief: Or ceratinty based on other understood facts about a claim.

For our Probability Models, we define the set of all outcomes to be Ω, better known as the **sample space** of an experiment. All subsets of Ω are called **events**. These are both sets and can be understood using default set notation.

1.2 Lecture 2 - Probability Law

Given Ω chosen, a **probability law** on Ω is a mapping $\mathbb{P}$ that assings a number for every event such that:

$$\boxed{\begin{array}{ll} \mathbb{P}(A) \geq 0 & \text{for every event A} \\ \mathbb{P}(\Omega) = 1 & \text{(normalization)} \end{array}} \tag{Kolmogorov's Axioms}$$

1.2.1 Additivity rules:

- If $A \cap B = \varnothing$, (A, B) events, then:

$$\boxed{\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)} \tag{1}$$

- If events $A_1, A_2, \ldots$ are all disjoint, then:

$$\boxed{\mathbb{P}(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)} \tag{2}$$

By these rules, we can surmise that $\boxed{\mathbb{P}(\varnothing) = 0}$.
For any events A, B:

$$\boxed{\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)} \tag{Event Union}$$

When we have a probability law on a finite Ω with all outcomes equally likely (i.e. $\mathbb{P}(\{s\}) = 1/size(\Omega)$), we call this probability law $\mathbb{P}$ a **(discrete) uniform probability law**.

1.3 Lecture 3 - Conditional Prob & Product Rule

1.3.1 Conditional Probability

Conditional Probability is defined $\mathbb{P}(A \mid B)$ = "Probability of A given B". It is understood as the likelyhood that event A occurs, given that B also occurs.

$$\boxed{\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}} \qquad \text{(Conditional Probability Def)}$$

If there is a finite number of different outcomes that are all equally likely, the conditional prbability can be written as follows:

$$\boxed{\mathbb{P}(A \mid B) = \frac{\text{number of elements of } A \cup B}{\text{number of elements of } B}} \qquad (3)$$

1.3.2 Product Rule

There are two main ways to write the product rule and they both have different setups.

- If we have events D_1 to D_n where $D_1 > D_2 > \cdots > D_n$ (D_1 largest, D_n smallest), then we can apply the first form of the product rule:

$$\boxed{\mathbb{P}(D_n) = \mathbb{P}(D_1)\mathbb{P}(D_2 \mid D_1)\mathbb{P}(D_3 \mid D_2)\ldots\mathbb{P}(D_n \mid D_{n-1})} \qquad \text{(Product Rule 1)}$$

- If we have events A_1 to A_n with non-empty intersection (i.e. $A_1 \cap A_2 \cap \cdots \cap A_n$), let $D_k = A_1 \cap A_2 \cap \cdots \cap A_k$, then $D_1 > D_2 > \cdots > D_n$. If we then write the product rule on the events D_n in terms of A_n, we get:

$$\boxed{\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_3 \mid A_1 \cap A_2)\ldots\mathbb{P}(A_n \mid A_1 \cap \ldots)} \qquad \text{(Product Rule 2)}$$

1.4 Lecture 5 - Bayes Law & Independence

1.4.1 Bayes Law

Bayes' Rule is defined by mixing the defintion of Condition Probability, and the Total Probability Theorem.

Given $\Omega, \mathbb{P}$, if $A_1, A_2, \ldots, A_n$ are events that partition Ω, and have nonzero $\mathbb{P}(A)$, then for any event B,

$$\boxed{\mathbb{P}(A_k \mid B) = \frac{\mathbb{P}(B \mid A_k)P(A_k)}{\mathbb{P}(B \mid A_1)\mathbb{P}(A_1) + \cdots + \mathbb{P}(B \mid A_n)}} \qquad \text{(Bayes' Law)}$$

1.4.2 Indpendence

Given $\Omega, \mathbb{P}$, any events A and B are **independent** when:

$$\boxed{\begin{array}{ll} \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) & \text{or} \\ \mathbb{P}(A \mid B) = \mathbb{P}(A), \mathbb{P}(B) > 0 & \end{array}} \qquad \text{(Independence Def)}$$

1.5 Lecture 6 - Conditional Dependence & Counting

1.5.1 Conditional Dependence

Given Ω and $\mathbb{P}$: say that events A and B are **conditionally independent** given C when:

$$\boxed{\begin{array}{ll}\mathbb{P}(A \cap B \mid C) = \mathbb{P}(A \mid C)\mathbb{P}(B \mid C) & \text{or} \\ \mathbb{P}(A \mid B \cap C) = \mathbb{P}(A \mid C), \mathbb{P}(B \cap C) > 0 & \end{array}} \tag{Conditional Independence Def}$$

1.5.2 Counting

Counting is the process of using the number of elements in the events to calculate probability. This technique mostly arises in situations where either:

- The sample space Ω has a finite number of equally likely outcomes. Then, for any event A,

$$\mathbb{P}(A) = \frac{\#\text{ of elements of } A}{\#\text{ of elements of } \Omega}$$

- An event A has a finite number of equally likely outcomes with probability p. Then for that event A:

$$\mathbb{P}(A) = p \cdot (\#\text{ of elements of } A)$$

1.6 Lecture 7 - Counting

Counting Principle: in a process with a sequence of stages $1, 2, \ldots, r$ with n_1 choices at stage 1 over to n_r at stage r; # of coutcomes is $n_1 n_2 \ldots n_k$.

Can be used to rederive (# subsets of Ω) = $2^{\#(elem)}$.

1.6.1 k-permutations of n objects

We are given n distinct objects and a number $k \leq n$, and we want to find out the number of ways we could take k distinct objects from the group of n objects and arrange them in a sequence. By using the Counting Principle, we can find that the **number of k-permutations** of this set is:

$$\boxed{\begin{aligned} n(n-1)\ldots(n-k+1) &= \frac{n(n-1)\ldots(n-k+1)(n-k)\ldots 2 \cdot 1}{(n-k)\ldots 2 \cdot 1} \\ &= \frac{n!}{(n-k)!} \end{aligned}} \tag{K-permutations}$$

Special Case: If $k = n$, then the number of k-permutations of n objects is simply $n!$.

1.6.2 k-combinations of n objects

For finding the number of k-combinations, we can look back to our k-permutations and reason about them. Say we have the same setup as before but we are not arranging the items in a sequence. For each combination, we have $k!$ "duplicate" permutations. Thus, we can look at the number permutations and reason that the number of k-combinations should be that over $k!$, making the **number of k-combinations** of this set is:

$$\boxed{\frac{n!}{k!(n-k)!} = \binom{n}{k}} \tag{K-combinations}$$

1.7 Lecture 8 - Discrete Random Variables

1.7.1 Random Variables

Given Ω and $\mathbb{P}$, a **discrete random variable (r.v.)** is a real valued function with domain Ω that takes on only finite or countably infinite number of different values (i.e. $X : \Omega \to \mathbb{R}$).

1.7.2 Probability Mass Functions

Given $\Omega, \mathbb{P}$, associated with any discrete rv $X : \Omega \to \mathbb{R}$ is X's **probability mass function (pmf)** - notation p_X.

$$\boxed{\forall x \text{ of } X, p_X(x) = \mathbb{P}(A_X) \text{ where } A_X = \{s \in \Omega : X(s) = x\}} \tag{pmf Def}$$

Things to Note:

- $\mathbb{P}(A_X)$ can also be written as $\mathbb{P}(\{X = x\})$ or $\mathbb{P}(X = x)$.
- $p_X(x) \geq 0$ for all possible values of X.
- A pmf is essentially a probability law on the different values in the codomain of a random variable, so the same laws that apply to probability laws apply to pmfs:

$$\begin{aligned} p_X(x) &\geq 0 \quad \text{for every } x \in X \\ \sum_{x \in X} p_X(x) &= 1 \quad \text{(normalization)} \end{aligned}$$

- If V is any finite or countably inf. set of possible values of X, then if we set $B = \{$the event "$X \in V$"$\}$, (i.e. $B = \{s \in \Omega : X(s) \in V\}$), then $\mathbb{P}(B) = \sum_{x \in V} p_X(x)$.

Note: for a given pmf, there are multiple Ω's, $\mathbb{P}$'s, X's that lead to that PMF.

1.7.3 Common PMFs

- **Discrete uniform pmf of interval** $a \leq k \leq b$, $a, b \in \mathbb{N}$:

$$p_X(k) = \begin{cases} \frac{1}{b-a+1} & \text{when } a \leq k \leq b \\ 0 & \text{all over } k \end{cases}$$

- Let $p \in [0,1]$; the **Bernoulli p pmf** be defined by:

$$p_X(k) = \begin{cases} p & \text{when } k = 1 \\ 1-p & \text{when } k = 0 \end{cases}$$

- Given positive integer n, some $p \in [0,1]$, the **Binomial(n,p) pmf** is defined as:

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad 0 \leq k \leq n$$

  This pmf tends to show up in situations involving sequences of independent trials, such as coin flips. Useful if you are trying to find the probability of k heads in n coin flips.

- Given $p \in (0,1)$ the **geometric pmf** defined by:

$$p_X(k) = p(1-p)^{k-1} \quad \text{for all } 1 \leq k \leq \infty \text{ positive integers}$$

  This pmf tends to show up in situations such as $\mathbb{P}$(it takes k flips to flip a heads).

- Poisson(X):

$$p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad 0 \leq k \leq \infty (k \in \mathbb{N})$$

1.8 Lecture 9 - Expectation, Variance

1.8.1 Function of a Random Variable

Given a random variable X and any function $g : \mathbb{R} \to \mathbb{R}$, can define another r.v. $Y = g(x)$:

$$\forall s \in \Omega, Y(s) = g(X(s))$$

The function of a discrete r.v. is another discrete r.v.. Generally, it is non-trivial to get the pmf of $Y = g(X)$, but it is sometimes easy. (See examples)

1.8.2 Expected Value

Given a discrete r.v. X with $p_X(x)$ pmf, we define the **expected value (expectation)**:

$$\boxed{\mathbb{E}(X) = \sum_{x \in X} x p_X(x)} \qquad \text{(Expected Value Definition)}$$

Given $X, Y = g(X)$, what is $\mathbb{E}(Y)$? One way is to figure out $p_Y(g)$ for all possible values of $y \in Y$ and then find it through:

$$\mathbb{E}(Y) = \sum_{y \in Y} y p_Y(y)$$

and get P_y through p_x, though that is generally a non-trivial solution. However, another possible solution is to use the **Expected Value Rule**.

1.8.3 Expected Value Rule

Given X, p_X, and $Y = g(X)$,

$$\boxed{\mathbb{E}(Y) = \sum_{x \in X} g(X) p_X(x)} \tag{Expected Value Rule}$$

Special Case: $Y = \alpha X + \beta$

$$\begin{aligned}
\mathbb{E}(Y) &= \sum_{x \in X} g(x) p_X(x) \\
&= \sum_{x \in X} (\alpha x + \beta) p_X(x) \\
&= \alpha \sum_{x \in X} x p_X(x) + \beta \sum_{x \in X} p_X(x) \\
&= \alpha \mathbb{E}(X) + \beta
\end{aligned}$$

1.8.4 Variance

Given an rv X with pmf p_X, we define **Variance** to be:

$$\boxed{Var(X) = \mathbb{E}((X - \mathbb{E}(X))^2)} \tag{Variance Def}$$

Off of this definition, we also define standard deviation to be $\sigma_X = \sqrt{Var(X)}$.

1.9 Lecture 10 - Expected Value and Variance Examples

- X is **Bernoulli(p)**:

$$\mathbb{E}(X) = p; Var(X) = p(1-p)$$

- X is **discrete uniform on** $a \leq k \leq b$:

$$\mathbb{E}(X) = \frac{b+a}{2}$$

- X is **Poisson(**λ**)**:

$$\mathbb{E}(X) = \lambda; Var(X) = \lambda$$

- X is **Geometric(p)**:

$$\mathbb{E}(X) = \frac{1}{p}$$

---

2 Post Prelim 1 - Pre Prelim 2

2.1 Lecture 11 - Multiple Discrete RVs, Joint pmf's, Conditionals

2.1.1 Joint pmf's

Given $\Omega, \mathbb{P}$, and two discrete rv's X, Y defined in Ω, define the **joint pmf** of $X\&Y$ as:

$$\boxed{\forall x \in X, y \in Y, \quad p_{X,Y}(x,y) = \mathbb{P}(\{X = x\} \cap \{Y = y\}) = \mathbb{P}(A_x \cap B_y)} \tag{Joint pmf}$$

For any set V of possible value pairs x, y, we have that

$$\sum_{x,y \in V} p_{X,Y}(x,y) = \mathbb{P}(\text{event that}(X,Y) \in V)$$

From the joint pmf $p_{X,Y}(x,y)$, we can derive the **marginal pmfs** $p_X(x)$ and $p_Y(y)$ from the joint pmf as:

$$\boxed{\begin{aligned} \forall x, \quad p_X(x) &= \sum_y p_{X,Y}(x,y) \\ \forall y, \quad p_Y(y) &= \sum_x p_{X,Y}(x,y) \end{aligned}} \tag{Marginal Pmfs}$$

Given X, Y with joint pmf $p_{X,Y}(x,y)$ and some real valued function $Z = g(X,Y)$, we have:

$$\boxed{\mathbb{E}(Z) = \sum_{x \in X} \sum_{y \in Y} g(x,y) p_{X,Y}(x,y)} \tag{Joint Expected Value Rule}$$

Special g choice: $g(X,Y) = \alpha X + \beta Y = \gamma$

$$\mathbb{E}(z) = \sum_x \sum_y (\alpha x + \beta y + \gamma) p_{X,Y}(x,y) = \cdots = \alpha \mathbb{E}(X) + \beta \mathbb{E}(Y) + \gamma$$

Can be generalized to: $\mathbb{E}(\alpha_1 z_1 + \alpha_2 z_2 + \cdots + \alpha_n z_n + \beta) = \alpha_1 \mathbb{E}(z_1) + \alpha_2 \mathbb{E}(z_2) + \cdots + \alpha_n \mathbb{E}(z_n) + \beta$

2.1.2 Conditional pmf

Given $\Omega, \mathbb{P}$, a discrete rv X defined on Ω, an event $A \subset \Omega, \mathbb{P}(A) > 0$, and a possible value $x \in X$, the **conditional pmf** of X given A is defined as:

$$\boxed{p_{X|A}(x) = \frac{\mathbb{P}(\{X = x\} \cap A)}{\mathbb{P}(A)} = \text{"}\mathbb{P}(B \mid A) \text{ where } B \text{ is the event } \{X = x\}\text{"}} \tag{Conditional pmf on event}$$

Observe that for any A with $\mathbb{P}(A) > 0, p_{X|A}(x)$ as x ranges over X's values defines a pmf: i.e. $p_{X|A}(x) \leq 0, \ \forall x$ and $\sum_{x \in X} p_{X|A}(x) = 1$.

2.1.3 RVs Conditional on RVs

Given X, Y defined on some $\Omega, \mathbb{P}$, conditional pmf of X given Y is defined on $\forall x$ & $\forall y$, with $\mathbb{P}(\{Y = y\}) = p_Y(y) > 0$ as:

$$\boxed{p_{X|Y}(x \mid y) = \frac{\mathbb{P}(\{X = x\} \cap \{Y = y\})}{\mathbb{P}(\{Y = y\})} = \frac{p_{X,Y}(x,y)}{p_Y(y)}} \tag{Conditional pmf on rv}$$

Observe that for any fixed y with $p_Y(y) > 0, p_{X|Y}(x \mid y)$ as x ranges over X values defines a pmf: i.e. $p_{X|Y} \geq 0$ and $\sum_{x \in X} p_{X|Y}(x \mid y) = 1$.

2.2 Lecture 12 - Conditional Probability for RV's, Conditional Expectation

2.2.1 Conditional Probability

Given events $A_1, A_2, \ldots, A_n$ that partition Ω and $\mathbb{P}(A_k) > 0,\ 0 \leq k \leq n$, then for any discrete rv X on Ω,

$$\boxed{p_X(x) = \sum_{k=1}^{n} p_{X|A_k} \mathbb{P}(A_k)} \qquad \text{(Conditional Total Probability)}$$

There are also ways of expressing the joint pmfs in terms of the marginals and vice versa:

$$\boxed{\begin{aligned} &p_{X,Y}(x,y) = p_Y(y)p_{X|Y}(x \mid y) \quad \forall x \in X \qquad \text{or,} \\ &p_{X,Y}(x,y) = p_X(x)p_{Y|X}(y \mid x) \quad \forall y \in Y \end{aligned}} \qquad \text{(Product Rule of Sorts)}$$

$$\boxed{\begin{aligned} &p_X(x) = \sum_{y \in Y} p_Y(y)p_{X|Y}(x \mid y) \quad \forall x \in X \qquad \text{or,} \\ &p_Y(y) = \sum_{x \in X} p_X(x)p_{Y|X}(y \mid x) \quad \forall y \in Y \end{aligned}} \qquad \text{(Total-Prob Rule of Sorts)}$$

These also generalize to > 2 rvs:

$$p_{X|Y,Z}(x \mid y,z) = \frac{\mathbb{P}(\{X = x\} \cap \{Y = y\} \cap \{Z = z\})}{\mathbb{P}(\{Y = y\} \cap \{Z = z\})} = \frac{p_{X,Y,Z}(x,y,z)}{p_{Y,Z}(y,z)}$$

$$p_{X,Y,Z}(x,y,z) = p_Z(z)p_{Y|Z}(y \mid z)p_{X|Y,Z}(x \mid y,z)$$

2.2.2 Conditional Expectation

Given $\Omega, \mathbb{P}$, a discrete rv X, and an event A, we define the **Conditional Expectation** of X given event A as:

$$\boxed{\mathbb{E}(X \mid A) = \sum_{x \in X} x p_{X|A}(x)} \qquad \text{(Conditional Expectation)}$$

We know that given events $A_1, \ldots, A_n$ that partition Ω, $p_X(x) = \sum_{k=1}^{n} p_{X|A_k}(x \mid A_k)\mathbb{P}(A_k)$. From this we can derive:

$$\mathbb{E}(X) = \sum_{x \in X} x p_X(x) = \sum_x \sum_k x p_{X|A_k}(x)\mathbb{P}(A_k) = \sum_k (\sum_x x p_{x|A_k}(x))\mathbb{P}(A_k)$$

Given $\Omega, \mathbb{P}$, rvs X, Y, event $A = \{Y = y\}$:

$$\mathbb{E}(X \mid A) = \sum_{x \in X} x p_{X|A}(x) = \sum_{x \in X} x p_{X|A}(x \mid y)$$

Since all events $\{Y = y\}$ partition Ω, we get:

$$\mathbb{E}(X) = \sum_{y \in Y} \mathbb{E}(X \mid Y = y)\mathbb{P}(\{Y = y\})$$

2.3 Lecture 13 & 14 - Indpendence of RVs

Given $\Omega, \mathbb{P}$, a discrete rv X defined of Ω, and an event A, say X is **independent (event)** of A when every event $\{X = x\}$ is independent of A (event-wise), i.e:

$$\boxed{\mathbb{P}(\{X = x\} \cap A) = p_X(x)\mathbb{P}(A)} \tag{RV Event Independence}$$

Note that when $\mathbb{P}(A) > 0$, it is the same as stating $p_{X|A}(x) = p_X(x),\ \forall x$.
Say two rvs X and Y are **independent (rvs)** when:

- X is indpendent of every event $\{Y = y\}$.
- Y is independent of every event $\{X = x\}$.
- $p_{X,Y}(x, y) = p_X(x)p_Y(y) \quad \forall x \in X, y \in Y$.

This extends to multiple rvs $X_1 \dots X_n$. These rvs are independent when:

$$\underbrace{p_{X_1,\dots,X_n}(x_1, \dots, x_n)}_{\text{joint}} = \underbrace{p_{X_1}(x_1)\dots p_{X_n}(X_n)}_{\text{product of marginals}} \quad \forall x_1, \dots, x_n$$

Important facts about rv independence:

- If X, Y are indpendent, then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$
- If X, Y are indpendent, then $Var(X + Y) = Var(X) + Var(Y)$.

2.4 Lecture 15 - "Randomness" of RVs

2.4.1 Binary Entropy

We can quantify the "randomness" of an rv through the use of binary entropy. Suppose an rv X has N possible values, and that $p_X(x_k) = p_k, 1 \le k \le N$, ($p_K = 0$ is allowed). We define the **binary entropy** of discrete rvs as:

$$\boxed{H(X) = \sum_{k=1}^{N} p_k log_2(p_k)} \tag{Binary Entroy of RVs}$$

$H(X)$ then becomes a good measure of "how random X is".

<u>Important Facts About Entropy:</u>

- If $p_k = 1$ for some k and 0 for all other k, X is "least random".
- If $p_k = \frac{1}{N} \forall k, X$ is "maximally random".
- $0 \le H(X) \le log_2 N$.
- $H(X) = 0 \Leftrightarrow p_k = 1$ for some k.

2.4.2 Source Coding Theorem

Given X with pmf $p_X(x_k) = p_k, 1 \leq k \leq N$, any guaranteed successful Y/N 20 questions scheme for determining value of X has a mean number of questions $\mathbb{E}(L)$ such that:

$$\boxed{\mathbb{E}(L) \leq H(X)} \tag{Source Coding Theorm}$$

We compute $\mathbb{E}(L)$ as follows: let $l_k = \#(\text{questions you need to ask when } X = x_k)$, then:

$$\mathbb{E}(L) = \sum_{k=1}^{N} p_k l_k$$

2.5 Lecture 16 - Continuous Random Variables

2.5.1 Continous Random Variables

Given $\Omega, \mathbb{P}, X : \Omega \to \mathbb{R}$ is a **continous rv** when there's a "reasonable" function $f_X(x)$ such that for every $V \subset \mathbb{R}$, we have:

$$\boxed{\mathbb{P}(\{X \in V\}) = \int_V f_X(x)dx} \tag{Continuous rv}$$

That function $f_X(x)$ is called the **probability density function (pdf)** of X. It can be interpreted as the "probability 'mass' per unit 'length"' of an rv.

Special Case of V: $V = [a, b]$ or $[a, b), (a, b], (a, b)$ we have:

$$\mathbb{P}(\{X \in V\}) = \int_a^b f_X(x)dx$$

Some properties of $f_X(x)$:

- $f_X(x) \geq 0 \; \forall x$ (need to ensure $\mathbb{P}(\{X \in V\}) \geq 0$ for all $V \subset R$)
- $\lim\limits_{R\to\infty} \int\limits_{-R}^{+R} f_X(x)dx = 1 \to \int\limits_{-\infty}^{\infty} f_X(x)dx = \mathbb{P}(\{X \in (-\infty, \infty)\}) = 1$
- Given $x \in R$, $f_X(x)$ is NOT $\mathbb{P}$(some event) — in particular, $f_X(x) \neq \mathbb{P}(\{X = x\})$.
- Turns out $\mathbb{P}(\{X = x\}) = 0 \quad \forall x \in R$ when X is a continuous random variable.
- Since $f_X(x)$ isn't $\mathbb{P}$(some event), need not have $f_X(x) \leq 1$! In fact, $f_X(x)$ can take on arbitrarily large values!

2.5.2 Expected Value

The **expected value** of a continous rv X with pdf $f_X(x)$ is defined as (Note: expected value not always defined, integral might fail to exist):

$$\boxed{\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f_X(x)dx} \tag{Expected Value: Continous rv}$$

Given continous rv X with pdf $f_X(x)$ and $Y = g(X)$, we define the **Expected Value Rule** as:

$$\boxed{\mathbb{E}[Y] = \int_{-\infty}^{+\infty} g(X) f_X(x)dx} \tag{Expected Value Rule}$$

Special Case: $g(X) = \alpha X + \beta$

$$\mathbb{E}[g(X)] = \alpha\mathbb{E}[X] + \beta$$

2.5.3 Variance

Given continous rv X w/ defined expected value $\mathbb{E}[X]$, we define the **Variance** as:

$$\boxed{Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]} \qquad \text{(Variance: Continous rv)}$$

By expected value rule, we also have:

$$Var(X) = \int_{-\infty}^{\infty} (X - \mathbb{E})^2 f_X(x)dx$$

Also, as before:

$$Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

2.5.4 Interpretations

$\mathbb{E}(X)$ can interpreted as the center of the "probability mass" defined by $f_X(x)$.
$Var(X)$ can be interpreted as the spread of the "proability mass" about its center.

2.5.5 Common Continuous RVs

Some common continuous rvs are:

- **X uniform on [a,b]:** Given $a, b \in \mathbb{R};\ a < b$; let the pdf of the "uniform on [a,b]" rv be defined as:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{when } x \in [a, b] \\ 0 & \text{else} \end{cases}$$

  For the uniform on [a,b] rv, we have: $\mathbb{E}[X] = \frac{b+a}{2}$ and $Var(X) = \frac{(b-a}{2}12$.

- **X exponential(λ):** Given $\lambda \in \mathbb{R}$; let the pdf of the "exponential(λ)" rv be defined as:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{when } x \geq 0 \quad \forall \lambda > 0 \\ 0 & \text{when } x < 0 \end{cases}$$

  For the exponential(λ) rv, we have: $\mathbb{E}[X] = \frac{1}{\lambda}$ and $Var(X) = \frac{1}{\lambda^2}$.

- **X piecewise uniform:** TODO

2.6 Lecture 17 - Cumulative Distribution Function

2.6.1 Cumulative Distribution Function

For any rv X (discrete of continous), the **cumulative distribution function (cdf)** if defined as:

$$\boxed{F_X(x) = \mathbb{P}(\{X \leq x\}) \quad \forall x \in \mathbb{R}} \qquad \text{(Cdf Definition: Continuous)}$$

If X is a continuous rv w/ pdf $f_X(x)$, then since $\mathbb{P}(\{X \leq x\}) = \int_{-\infty}^{x} f_X(t)dt$, we have:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt \quad \text{and} \quad f_X(x) = \frac{d}{dx}F_X(x)$$

Discrete Version: If X is a discrete rv with pmf $p_X(x)$, we have:

$$\boxed{F_X(x) = \sum_{\{x_k | x_k \leq x\}} p_X(x_k)} \qquad \text{(Cdf Definition: Discrete)}$$

This formula can also be inverted to get $p_X(x)$ in terms of $F_X(x)$:

$$p_X(x_k) = F_X(x_k) - F_X(x_{k-1})$$

where x_{k-1} is the "next largest value" of X below x_k.

2.6.2 General Properties of cdfs

1. $\lim_{x\to-\infty} F_X(x) = 0$ and $\lim_{x\to\infty} F_X(x) = 1$.
2. When X is a continuous rv, $F_X(x)$ is continuous in x and differentiable "almost everywhere" (corners in $F_X(x)$ correspond to jump in $f_X(x)$)
3. X is a discrete rv iff $F_X(x)$ is peice wise constant.
4. $F_X(x)$ is monotomically increasing in x, i.e.

$$x_1 \leq x_2 \quad \Rightarrow \quad F_X(x_1) \leq F_X(x_2)$$

Cdfs are alse useful for getting the pdf of X by first computing $F_X(x)$, then taking d/dx.

2.6.3 Gaussian rv

Another inportant continuous rv is the **Gaussian rv**. The pdf of the Gaussian rv is:

$$f_X(x) = \frac{1}{\sqrt{2x\sigma^2}} \exp\bigg(- \frac{(x-M)^2\Big)}{2\sigma^2}$$

With this pdf, we can see that $\mathbb{E}(X) = M$ and $Var(X) = \mathbb{E}((X-M)^2) = \sigma^2$. The cdf of a Gaussian rv is:

$$F_X(x) = \frac{1}{\sqrt{2x\sigma^2}} \int_{-\infty}^{x} \exp - \frac{(t-m)^2}{2\sigma^2} dt$$

The **standard normal** pdf is a specific type of Gaussian pdf:

$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \therefore M = 0, \ \sigma = 1$$

The cdf of the standard normal pdf is:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

The standard normal pdf can be created from a Gaussian rv with mean M and var σ^2 as $Y = \frac{X-M}{\sigma}$ (Example):

$$\mathbb{P}(\{X > 17\}) = \mathbb{P}(\{\sigma + Y + M > 17\}) = \mathbb{P}\Big(Y > \frac{17-M}{\sigma}\Big)$$

Gaussian rvs are important because the sum of many different independent rvs that have the same pdf "converges" to a Gaussian rv.

2.7 Lecture 18 - Multiple Continous Random Variables

2.7.1 Joint Continuous Random Variables

Say X, Y rvs defined in some Ω, $\mathbb{P}$ are **jointly continuous** with **joint pdf** $f_{X,Y}(x,y)$ when:

$$\boxed{\mathbb{P}(\{(X,Y) \in V\}) = \iint_{-V} f_{X,Y}(x,y)dxdy \quad \forall V \subset \mathbb{R}^2} \tag{Jointly Continuous Definition}$$

Special Case of V: $V : [a_1, b_1] \times [a_2, b_2]$

$$\mathbb{P}(\{X, Y) \in V\}) = \int_{a_1}^{b_2} dx \int_{a_2}^{b_2} dy(f_{X,Y}(x, Y))$$

From the joint pdf, we can also derive the marginals:

$$\boxed{\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy \\ f_Y(y) &= \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx \end{aligned}} \tag{Joint pdf Marginals}$$

Other Properties:

- $\int\limits_{-\infty}^{+\infty} dx \int\limits_{-\infty}^{+\infty} dy(f_{X,Y}(x,y)) = 1$
- Joint CDF: $F_{X,Y}(x,y) = \mathbb{P}(\{X = x\} \cap \{Y = y\}) = \int\limits_{-\infty}^{x} ds \int\limits_{-\infty}^{y} dy(f_{X,Y}(s,t))$
- $f_{X,Y}(x,) = \frac{\delta}{\delta x}\frac{\delta}{\delta y}F_{X,Y}(x,y)$
- Generalization to > 2 rv pretty "straightforward"

For discrete rvs, joint determines marginals, but not vice-versa.

2.7.2 Conditional for Continuous rvs: Events

Given a continuous rv X on Ω, $\mathbb{P}$, and some event $A \subset \Omega$, the conditional pdf of X given A "defined" as follows:

$$\boxed{\mathbb{P}(\{X \in V\} \mid A) = \int_V f_{X|A}(x)dx \quad \forall V \subset \mathbb{R}} \tag{Conditional pdf: Event}$$

In general, no decent formula for $f_{X|A}(x)$ in terms of $f_X(x)$. One way to compute it is to take the conditional cdf of X given A ($F_{X|A} = \mathbb{P}(\{X \leq x\} \mid A)$) and take d/dx to find $f_X(x)$. However, if A is an event of the form $\{X \in W\}$ and $\mathbb{P}(A) > 0$, we have:

$$f_{X|A}(x) = \begin{cases} \frac{f_X(x)}{\mathbb{P}(\{X \in W\})} & \text{when } X \in W \\ 0 & \text{otherwise} \end{cases}$$

We derive this by defining the indication function of W as:

$$\chi_W(x) = \begin{cases} 1 & \text{when } x \in W \\ 0 & \text{when } x \notin W \end{cases}$$

And using conditional functions for continuous rvs:

$$
\begin{aligned}
\mathbb{P}(\{X \in W\} \mid A) &= \frac{\mathbb{P}(\{X \in (V \cap W)\})}{\mathbb{P}(\{x \in W\})} \\
&= \frac{\int_{V \cap W} f_X(x)dx}{\mathbb{P}(\{X \in W\})} \\
&= \int_V \left(\frac{f_X(x)\chi_W(x)}{\mathbb{P}(\{X \in W\})}\right)dx \quad = \begin{cases} \frac{f_X(x)}{\mathbb{P}(\{X \in W\})} & \text{when } X \in W \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

2.8 Lecture 19 - Total Probability Theorem

2.8.1 Total Probability Theorem

In the context of $F_{X|A}$: If X is a continuous rv and $A_1, \dots, A_n$ are events of positive probability that partition Ω, then:

$$\boxed{f_X(x) = \sum_{k=1}^{n} f_{X|A_k}(x)\mathbb{P}(A_k)} \tag{Total Probability Theorem}$$

To see this: go via cdfs:

$$F_{X|A_k} = \frac{\mathbb{P}(\{X \le x\} \cap A_k)}{\mathbb{P}(A_k)} \frac{d}{dx} F_{X|A_k} = f_{X|A_k}(x)$$

$$F_X(x) = \mathbb{P}(\{X \le x\}) = \sum_{k=1}^{n} f_{X|A_k}\mathbb{P}(A_k) \to \sum_{k=1}^{n} f_{X|A_k}\mathbb{P}(A_k) = f_X(x)$$

2.8.2 Conditional for Continuous rvs: Other rvs

Given two continuous rvs X and Y, the **conditional pdf of X given $Y = y$** is defined as:

$$\boxed{f_{X|Y}(x, y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}} \tag{Conditional pdf: Other rvs}$$

Note: that for fixed y, this as a function of x is a legit pdf;

$$\int_{-\infty}^{\infty} f_{X|Y}(x \mid y)dx = \int_{-\infty}^{\infty} \frac{f_{X,Y}(x, y)}{f_Y(y)}dy = \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} f_{X,Y}(x, y)dx = \frac{f_Y(y)}{f_Y(y)} = 1$$

2.9 Lecture 20 - Conditional Expectance, Independence, Continuous Bayes' Rule

2.9.1 Conditional Expected Value

We define the **Conditional Expected Value** for an rv to be:

$$\boxed{\begin{aligned} &\text{Given } X, A, \mathbb{E}(X \mid A) = \int_{-\infty}^{\infty} x f_{X|A}(x)dx \\ &\text{Given } X, Y, \mathbb{E}(X \mid Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x \mid y)dx \quad \forall y \end{aligned}} \tag{}$$

From this, we can see that the **Expected Value Rule** holds as usual:

$$\boxed{\begin{aligned} \mathbb{E}(g(X) \mid A) &= \int_{-\infty}^{\infty} g(X) f_{X|A}(x) dx \\ \mathbb{E}(g(X) \mid Y = y) &= \int_{-\infty}^{\infty} g(X) f_{X|Y}(x \mid y) dx \quad \forall y \end{aligned}} \tag{Conditional Expected Value Rule}$$

We can also see these "total expectation theorems":

$$\mathbb{E}(X) = \sum_{k=1}^{n} \mathbb{E}(X \mid A_k)\mathbb{P}(A_k) = \int_{-\infty}^{\infty} \mathbb{E}(X \mid Y = y) f_Y(y) dy$$

2.9.2 Indpendence

For any pair X, Y, both continuous rvs with densities $f_X(x)$, $f_Y(y)$, $f_{X,Y}(x, y)$, X and Y are independent iff:

$$\boxed{f_{X,Y}(x, y) = f_X(x) f_Y(y) \quad \forall x, y} \tag{Independence: Continuous}$$

Note: When X and Y are independent, we have:

- $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$
- $\mathbb{E}(g(X)h(Y)) = \mathbb{E}(g(X))\mathbb{E}(h(Y))$
- $Var(X + Y) = Var(X) + Var(Y)$

2.9.3 Continuous Bayes' Rule

The **Continuous Bayes' Rule** becomes important when we have a good understanding of $f_X(x)$, a good model for $f_{Y|X}(y \mid x)$, and we want to find $f_{X|Y}(x \mid y)$.

$$\boxed{f_{X|Y}(x \mid y) = \frac{f_{Y|X}(y \mid x) f_X(x)}{\int_{-\infty}^{\infty} f_{Y|X}(y \mid x) f_X(x) dx}} \tag{Continuous Bayes' Rule}$$

2.10 Lecture 21 - Derived Distribution

The basis of Derived Distributions stems from having a continuous rv X with a pdf $f_X(x)$ and a function $Y = g(X)$. If we want to find $f_Y(y)$, in many cases it is easier to find $F_Y(y)$ and find $f_Y(y)$ by $f_Y(y) = \frac{dF_Y(y)}{dy}$.

2.11 Lecture 22 - Examples of Derived Distribution

---

3 Post Prelim 2

3.1 Lecture 23 - Covariance, Conditional Expectation Revisited

3.1.1 Covariance

3.1.2 Conditional Expectation Revisited

3.2 Lecture 24 - Conditional Expectation + Covariance

3.2.1 Law of Iterated Expectations

3.2.2 Conditional Variance

3.3 Lecture 25 - Law of Total Variance, Moment Generating Functions

3.3.1 Law of Total Variance

3.3.2 Moment Generating Functions

3.4 Lecture 26 - MGF Examples, Limit Theorems

3.4.1 Limit Theorems

3.5 Lecture 27 - More Limit Theorems (Central Limit Theorem)

3.5.1 Chebychev's Inequality

3.5.2 Markov Inequality

3.5.3 Central Limit Theorem

3.6 Lecture 28 - Caution with CLT, SLLN

3.6.1
