ECE180J Advanced Renewable Energy Sources

Lecture 9: Probability Theory Revisited

Dr. Yu Zhang

ECE Department
University of California, Santa Cruz

- H. Stark and J. Woods, "Probability, Random Processes, and Estimation Theory for Engineers".
- A. Papoulis, "Probability, Random Variables, and Stochastic Processes".

Probability Space

Probability space is a three-tuple $(\Omega, \mathcal{F}, P)$:

- **Sample space** Ω: the set of all outcomes of a random experiment.
  - Ω may be finite ($\{H, T\}$), countably infinite ($\Omega = \{1,2,3,...\}$), or uncountably infinite ([0,1]).

- **Event space** $\mathcal{F}$: a set whose elements $A \in \mathcal{F}$ are subsets of Ω.

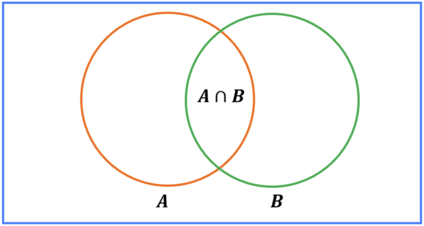
- **Probability function** P: satisfies three axioms:
  - $P(A) \geq 0$, $\forall A \in \mathcal{F}$.
  - $P(\Omega) = 1$.
  - If $A_1, A_2, ...$ are mutually exclusive events ($A_i \cap A_j = \emptyset$, for $i \neq j$), then $P(\bigcup_i A_i) = \sum_i P(A_i)$.

Properties of Probability

Consider events $A, B, C \subseteq \Omega$. Let $P(AB)$ denote $P(A \cap B)$.

- $A \subseteq B \Rightarrow P(A) \leq P(B)$.
- $P(AB) \leq \min\{P(A), P(B)\}$.
- $P(A \cup B) = P(A) + P(B) - P(AB)$.
- $P(\bar{A}) = 1 - P(A)$: $\bar{A} = \Omega \backslash A$.
- Complement rule (De Morgan's Law): $\overline{AB} = \bar{A} \cup \bar{B}$; $\overline{A \cup B} = \bar{A}\bar{B}$
- $P(AB) = P(A|B)P(B) = P(B|A)P(A)$.
- **Conditional probability**: $P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$ (Bayes' theorem).
- If A and B are independent, then $P(AB) = P(A)P(B)$.
- Distributive law: $(A \cup B) \cap C = (AC) \cup (BC)$.

Venn Diagram



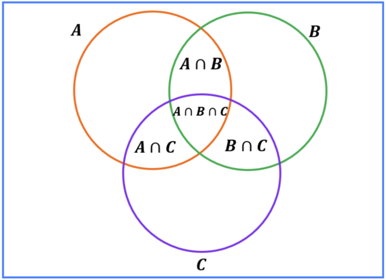


Figure: A Venn diagram is a diagram that shows the relationship between and among a finite collection of sets.

Conditional vis-à-vis Unconditional Probability

Q: $P(A|B) \lesseqqgtr P(A)$?

A: It can be any case in general.

Take a standard deck of cards (52 cards without Jokers). Remove all black queens and kings. When picking a card, define 3 events:

- A: The card picked is a face card.
- B_1: The card picked is a heart.
- B_2: The card picked is a spade.

Question: Find the values of $P(A), P(A|B_1), P(A|B_2)$.

- A: 48 cards left with 8 face cards: $P(A) = \frac{1}{6}$.
- B_1: 13 cards of hearts left with 3 face cards: $P(A|B_1) = \frac{3}{13}$.
- B_2: 11 card of spades left with only 1 face card (the Jack of Spades): $P(A|B_2) = \frac{1}{11}$.

We have $P(A|B_2) < P(A) < P(A|B_1)$.

Random Variables

Definition (Random Variable (RV))

Let Ω be the sample space of an experiment, and $\mathbb{R}$ denote the set of real numbers. Then, a *random variable* $X : \Omega \mapsto \mathbb{R}$ associated with this experiment is a function that assigns each outcome in Ω to a real number. The range of X is denoted as $\mathrm{val}(X)$.

Example. Flip a coin 5 times. Let X denote the random variable for the number of times the coin came up heads. Then $X(\omega_0) = 3$, for the outcome $\omega_0 = \{\mathrm{HHTHT}\}$.

There are two different types of random variables that are often studied: *discrete* and *continuous*.

Random Variables (Cont'd)

If X is a discrete random variable, we use the notation

$$\Pr(X = k) \coloneqq \Pr(\{\omega : X(\omega) = k\})$$

for the probability of the event $X = k$.

If X is a continuous random variable, we use the notation

$$\Pr(a \leq X \leq b) \coloneqq \Pr(\{\omega : a \leq X(\omega) \leq b\})$$

for the probability of the event $a \leq X \leq b$.

Cumulative Distribution Function (CDF)

Definition

Let X be a random variable associated with an experiment. Then,

$$\boxed{F_X(x) \coloneqq \Pr(X \leq x)}$$

is a *cumulative distribution function*.

Properties of CDFs:

(a) $0 \leq F_X(x) \leq 1$.

(b) $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to +\infty} F_X(x) = 1$.

(c) F_X is nondecreasing, namely if $x \leq y$ then $F_X(x) \leq F_X(y)$.

(d) F_X is right-continuous, i.e. $\lim_{x \to a^+} F_X(x) = F_X(a)$.

CDF for Discrete and Continous RVs

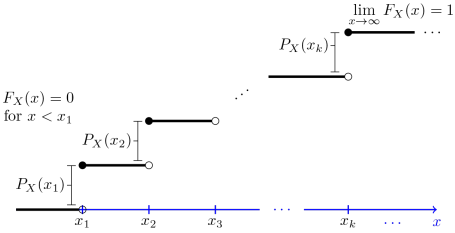


Figure: CDF of a discrete random variable.

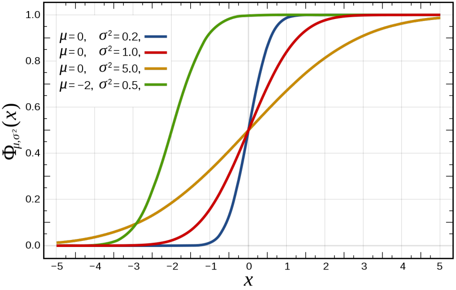


Figure: CDF of a continuous random variable.

PMF and PDF

For discrete random variables, we have the *probability mass function* (PMF):

$$p_X(x) \coloneqq \Pr(X = x),$$

where $\sum_{x \in \mathrm{val}(X)} p_X(x) = 1.$

For continuous random variables, we instead consider *probability density function* (PDF),

$$f_X(x) \coloneqq \frac{dF_X(x)}{dx} = F'_X(x),$$

provided that F_X is differentiable at x.[1] Notice that $f_X(x)$ and $\Pr(X = x)$ are two different concepts, which can be related by

$$\Pr(x \le X \le x + \Delta x) \approx f_X(x)\Delta x$$

and

$$\Pr(X \in A) = \int_{x \in A} f_X(x) dx,$$

where $A \subseteq \mathrm{val}(X)$.

[1] Note that F_X may not be everywhere differentiable even for continuous random variable X.

Common Distributions

Name of the probability distribution	Probability distribution function	Mean	Variance
Binomial distribution	$\Pr(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$
Geometric distribution	$\Pr(X=k) = (1-p)^{k-1} p$	$\frac{1}{p}$	$\frac{(1-p)}{p^2}$
Normal distribution	$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
Uniform distribution (continuous)	$f(x \mid a, b) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b, \\ 0 & \text{for } x < a \text{ or } x > b \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential distribution	$f(x \mid \lambda) = \lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

Figure: PDF or PMF of commonly used random variables (Wiki).

Gaussian (Normal) Distribution

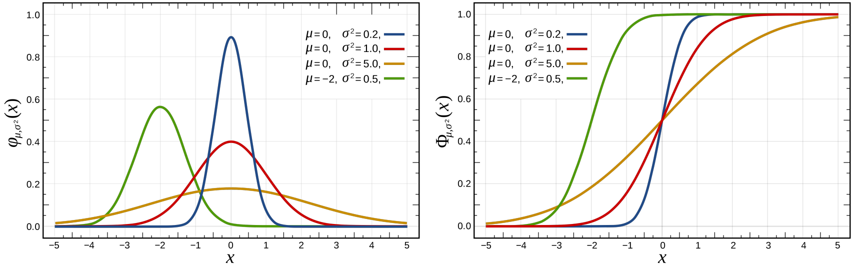


Figure: PDF and CDF of Gaussian distribution.

PDF of Gaussian distribution:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

CDF of standard Gaussian distribution ($\mu = 0$ and $\sigma = 1$):

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

Expectation

Definition (Expectation)

Let $g : \mathbb{R} \to \mathbb{R}$ be a function and X be a random variable, discrete or continuous, then the *expectation* of the random variable $g(X)$ is

$$\mathrm{E}[g(X)] := \sum_{x \in \mathrm{val}(X)} g(x) p_X(x) \quad \text{or} \quad \mathrm{E}[g(X)] := \int_{x \in \mathrm{val}(X)} g(x) f_X(x) dx,$$

respectively.

Linearity of the expectation operator:

(a) $\mathrm{E}[ag(X) + bh(X)] = a\mathrm{E}[g(X)] + b\mathrm{E}[h(X)]$ for any constants a, b, and arbitrary functions $g(\cdot), h(\cdot)$.

(b) $X \perp\!\!\!\perp Y \Rightarrow \mathrm{E}[XY] = \mathrm{E}[X]\mathrm{E}[Y]$.

The *indicator function*: $\mathbf{1}_A := \begin{cases} 1, & \text{if } A \text{ is true} \\ 0, & \text{otherwise.} \end{cases} \Rightarrow$

$\mathrm{E}[\mathbf{1}_A] = \Pr(A) \times 1 + \Pr(\bar{A}) \times 0 = \Pr(A)$, $F_X(x) = \mathrm{E}[\mathbf{1}_{\{X \leq x\}}]$.

Variance

Given a distribution of a random variable, its variance to measure how concentrated that distribution is around the expectation (mean). Formally, we have

Definition (Variance)

$\mathrm{Var}[X] := \mathrm{E}[(X - \mathrm{E}[X])^2] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$, where $\mathrm{E}[X^2]$ is the second moment of X.

The following can be derived immediately from the definition:

(a) $\mathrm{Var}[cX] = c^2\mathrm{Var}[X]$.

(b) $\mathrm{Var}[c] = 0$ for any constant c.

(c) $\mathrm{Var}[aX \pm bY] = a^2\mathrm{Var}[X] + b^2\mathrm{Var}[Y] \pm 2ab \times \mathrm{Cov}[X, Y]$.

Examples

Let $X \sim \exp(\lambda)$ whose density function is $f_X(x) = \lambda e^{-\lambda x}$. Find $\mathrm{E}[X]$ and $\mathrm{Var}[X]$.

Solution: From the definition of expectation and integration by parts, we have

$$
\begin{aligned}
E(X) &= \int_0^\infty x f_X(x)\,dx \\
&= \lambda \int_0^\infty x e^{-\lambda x}\,dx \\
&= -x e^{-\lambda x}\Big|_0^\infty + \int_0^\infty e^{-\lambda x}\,dx \\
&= 0 + \frac{e^{-\lambda x}}{-\lambda}\Big|_0^\infty = \frac{1}{\lambda}\,.
\end{aligned}
$$

$$
\begin{aligned}
V(X) &= \int_0^\infty x^2 f_X(x)\,dx - \frac{1}{\lambda^2} \\
&= \lambda \int_0^\infty x^2 e^{-\lambda x}\,dx - \frac{1}{\lambda^2} \\
&= -x^2 e^{-\lambda x}\Big|_0^\infty + 2\int_0^\infty x e^{-\lambda x}\,dx - \frac{1}{\lambda^2} \\
&= -x^2 e^{-\lambda x}\Big|_0^\infty - \frac{2x e^{-\lambda x}}{\lambda}\Big|_0^\infty - \frac{2}{\lambda^2} e^{-\lambda x}\Big|_0^\infty - \frac{1}{\lambda^2} = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}\,.
\end{aligned}
$$

Multivariate Random Variables

Definition (Joint CDF)

Let X, Y be two random variables associated with an experiment. Then the *joint cumulative distribution function* of X and Y is

$$F_{XY} := \Pr(X \leq x, Y \leq y)$$

and the *marginal cumulative distribution function* of X is

$$F_X(x) := \lim_{y \to +\infty} \Pr(X \leq x, Y \leq y).$$

Similarly, the *joint* PMF $p_{XY}(x, y) := \Pr(X = x, Y = y)$
the *marginal* PMF of X, $p_X(x) := \sum_{y \in \mathrm{val}(Y)} \Pr(X = x, Y = y)$.
The *joint* PDF of continuous X and Y is $f_{XY}(x, y) := \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y}$
the *marginal* PDF of X: $f_X(x) := \int_{y \in \mathrm{val}(Y)} f_{XY}(x, y) dy$.

Conditional Distribution

The above summation or integral operation is called *marginalization*. Note that

$$\iint\limits_A f_{XY}(x, y)dxdy = \Pr\Big((x, y) \in A\Big).$$

Definition (Conditional Distribution for Discrete RVs)

Let X and Y be discrete random variables. The *conditional distribution* of Y given $X = x$ is

$$\boxed{p_{Y|X}(y \mid x) := \frac{p_{XY}(x, y)}{p_X(x)}},$$

provided that $p_X(x) \neq 0$.

We say that X and Y are *independent* if $p_{Y|X}(y \mid x) = p_Y(y)$.

Note that $X \perp\!\!\!\perp Y \Leftrightarrow p_{XY}(x, y) = p_X(x)p_Y(y)$.

Q: If $X \perp\!\!\!\perp Y$. For arbitrary functions $g(\cdot)$ and $h(\cdot)$, are $g(X) \perp\!\!\!\perp h(Y)$?

A: Yes.

Covariance & Correlation

Definition (Covariance and Correlation)

Let X and Y be two random variables. Their *covariance* is defined as

$$\mathrm{Cov}[X, Y] := \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y].$$

Their *correlation* is defined as

$$\mathrm{Corr}[X, Y] := \frac{\mathrm{Cov}[X, Y]}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$

1. $\mathrm{Corr}[X, Y] \in [-1, 1]$ is a measure of linear association between X and Y.
2. X and Y are uncorrelated if $\mathrm{Corr}[X, Y] = 0$.
3. $\mathrm{Corr}[X, Y] = \pm 1 \Leftrightarrow Y = aX + b$ for some constants a and b.
4. If $X \perp\!\!\!\perp Y \implies$ they are uncorrelated.
5. X and Y can be uncorrelated yet dependent due to a nonlinear relationship.

Example: $X \sim N(0, 1), Y = X^2 \implies \mathrm{Corr}[X, Y] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y] = \mathrm{E}[X^3] = 0$ (all odd-order moments of X are equal to zero). Hence, X and Y are uncorrelated. But they're clearly dependent since Y is a function of X.

Example 1 of Covariance

Let X and Y be discrete random variables, with joint probability function $p_{X,Y}$ given by

$$p_{X,Y}(x, y) = \begin{cases} 1/2 & x = 3, y = 4 \\ 1/3 & x = 3, y = 6 \\ 1/6 & x = 5, y = 6 \\ 0 & \text{otherwise.} \end{cases}$$

Then $E(X) = (3)(1/2) + (3)(1/3) + (5)(1/6) = 10/3$, and $E(Y) = (4)(1/2) + (6)(1/3) + (6)(1/6) = 5$. Hence,

$$\begin{aligned} \operatorname{Cov}(X, Y) &= E((X - 10/3)(Y - 5)) \\ &= (3 - 10/3)\,(4 - 5)/2 + (3 - 10/3)\,(6 - 5)/3 + (5 - 10/3)\,(6 - 5)/6 \\ &= 1/3. \blacksquare \end{aligned}$$

Example 2 of Covariance

Let X be any random variable with $\mathrm{Var}(X) > 0$. Let $Y = 3X$, and let $Z = -4X$. Then $\mu_Y = 3\mu_X$ and $\mu_Z = -4\mu_X$. Hence,

$$\begin{aligned} \mathrm{Cov}(X, Y) &= E((X - \mu_X)(Y - \mu_Y)) = E((X - \mu_X)(3X - 3\mu_X)) \\ &= 3\,E((X - \mu_X)^2) = 3\,\mathrm{Var}(X), \end{aligned}$$

while

$$\begin{aligned} \mathrm{Cov}(X, Z) &= E((X - \mu_X)(Z - \mu_Z)) = E((X - \mu_X)((-4)X - (-4)\mu_X)) \\ &= (-4)E((X - \mu_X)^2) = -4\,\mathrm{Var}(X). \end{aligned}$$

Note in particular that $\mathrm{Cov}(X, Y) > 0$, while $\mathrm{Cov}(X, Z) < 0$. Intuitively, this says that Y increases when X increases, whereas Z decreases when X increases. ∎

PDF of Sum of Two Random Variables

For independent random variables X and Y, the distribution f_Z of $Z = X + Y$ equals the **convolution** of f_X and f_Y:

$$f_Z(z) = \int_{-\infty}^{\infty} f_Y(z - x) f_X(x) \, dx$$

LLN and CLT

Theorem (Law of Large Numbers (LLN))

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed (i.i.d.) random variables so that $\mathrm{E}[X_1] = \mathrm{E}[X_2] = \cdots = \mathrm{E}[X_n] < \infty$. Let

$$\bar{X}_n := \frac{X_1 + X_2 + \cdots + X_n}{n}$$

denote the sample mean of those n random variables. Then $\bar{X}_n \to \mathrm{E}[X_1]$ as $n \to \infty$ almost surely (a.s., strong law) and in probability (i.p., weak law).

Theorem (Central Limit Theorem (CLT))

Let $X_1, X_, \ldots, X_n$ be a sequence of i.i.d. random variables, and assume that $\mathrm{E}[X_i] = \mu$ and $\mathrm{Var}[X_i] = \sigma^2$, for all i. Let $S_n := X_1 + X_2 + \cdots + X_n$. Then, $\mathrm{E}[S_n] = n\mu$, $\mathrm{Var}[S_n] = n\sigma^2$ and we have the standardization of S_n,

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\text{i.p.}} N(0,1) \text{ as } n \to \infty$$

where $N(0,1)$ denotes the standard normal random variable.

Thank You!

Email: `<zhangy@ucsc.edu>`

Homepage: `https://people.ucsc.edu/~yzhan419/`