Probability Theory Review

Axiom (Basic Counting Principle)

Suppose we perform two experiments, one after the other. If the first experiment has m_1 outcomes and the second experiment has m_2 outcomes, then there are $m_1 m_2$ outcomes for the two-step experiment.

Definition (Permutation)

An ordering of n distinct objects $x_1, x_2, ..., x_n$ is called a permutation of $x_1, x_2, ..., x_n$.

Definition (Factorial)

We define $0! = 1$

n factorial $= n! = n(n-1)...(2)(1)$ for $n = 1,2,3,...$

Proposition

There are $n!$ permutations of n distinct objects

Proposition

Suppose that a sequence S of n items has n_1 identical items of type 1, n_2 identical items of type 2, …, and n_k identical items of type k. Then, the number of orderings of S is $\dfrac{n!}{n_1!n_2!...n_k!}$.

Definition (k-combination)

Let $x_1, x_2, ..., x_n$ be n distinct elements. A k-combination of $x_1, x_2, ..., x_n$ is a k-element subset of the set $\{x_1, x_2, ..., x_n\}$.

Further, there are $\binom{n}{k} = \dfrac{n!}{(n-k)!k!}$ k-combinations of n distinct elements.

Definition (Binomial Coefficient)

For $k \le n$, we define the binomial coefficient as follows:

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

For $k > n$, we set $\binom{n}{k} = 0$

Identities

(a) $\quad \binom{n}{k} = \binom{n}{n-k}$

(b) $\quad \binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$

Theorem (The Binomial Theorem)

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}, \quad n = 1,2,3,\ldots$$

Theorem (Trinomial Theorem)

$$(x+y+z)^n = \sum_{\substack{i,j,k \geq 0 \\ i+j+k=n}} \frac{n!}{i!\,j!\,k!} x^i y^j z^k$$

Definition (Sample Space)

The <u>sample space</u> is the set of all possible outcomes. Denoted S.

Definition (Event)

An <u>event</u> is a subset of the sample space.

Definition (Probability Function)

A <u>probability function</u> is a function $P : \{events\} \to [0,1]$ with the following properties:

(a) For all events $E \subset S$, $0 \leq P(E) \leq 1$

(b) $P(S) = 1$

(c) For all mutually exclusive $E_1, E_2, \ldots \subset S$ $(i.e.\ E_i \cap E_j = \emptyset \ \ \forall i \neq j)$, we have $P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$ (additivity)

Proposition

Let S be a finite sample space, and suppose that all outcomes are equally likely. Then, for all $E \subset S$, $P(E) = \frac{|E|}{|S|}$.

Proposition

If $E \subset F \subset S$, $P(E) \leq P(F)$

Proposition

For all events $E \subset S$, we have that $P(E^C) = 1 - P(E)$ where $E^C = S \setminus E$ is the complement of E.

Proposition

For all events $E, F \subset S$ $P(E \cup F) = P(E) + P(F) - P(EF)$

Proposition (General Inclusion-Exclusion Formula)

For any events $E_1, E_2, ..., E_n \subset S$,

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i) - \sum_{i \le i_1 < i_2 \le n} P(E_{i_1} \cap E_{i_2}) + \sum_{i \le i_1 < i_2 < i_3 \le n} P(E_{i_1} \cap E_{i_2} \cap E_{i_3})$$

$$- + ... + (-1)^{n+1} P\left(\bigcap_{i=1}^{n} E_i\right)$$

Definition (Conditional Probability)

If $P(B) > 0$, then we define $P(A | B)$, the <u>conditional probability of A given B</u> by the following:

$$P(A | B) = \frac{P(AB)}{P(B)}$$

Proposition (Multiplication Rule)

Let $E_1, E_2, ..., E_n \subset S$ be events such that $P\left(\bigcap_{i=1}^{n-1} E_i\right) > 0$. Then,

$$P\left(\bigcap_{i=1}^{n} E_i\right) = P(E_1)P(E_2 | E_1)P(E_3 | E_1E_2)...P(E_n | E_1E_2...E_{n-1})$$

Proposition

Let $A, B \subset S$ be events such that $0 < P(B) < 1$. Then,

$$P(A) = P(A | B)P(B) + P(A | B^c)P(B^c)$$

Proposition (Generalization of Previous)

Let $A, B_1, B_2, ..., B_n \subset S$ be events such that

(a) $0 < P(B) < 1$

(b) $B_i \cap B_j = \emptyset$ for all $i \neq j$

(c) $\bigcup_{i=1}^{n} B_i = S$.

Then, $P(A) = P(A | B_1)P(B_1) + P(A | B_2)P(B_2) + ... + P(A | B_n)P(B_n)$

$$= \sum_{i=1}^{n} P(A | B_i)P(B_i)$$

Definition (Partition)

Let $B_1, B_2, ..., B_n \subset S$ be events such that

(a) $0 < P(B) < 1$

(b) $B_i \bigcap B_j = \emptyset$ for all $i \neq j$

(c) $\bigcup_{i=1}^{n} B_i = S$.

$\{B_1, ..., B_n\}$ is called a partition of S.

Proposition (Bayes' Formula)

Let $P(A) > 0$ and $B_1, ..., B_n$ be a partition of S with $0 < P(B_i) < 1$ for all i. Then,

$$P(B_j \mid A) = \frac{P(A \mid B_j)P(B_j)}{P(A \mid B_1)P(B_1) + ... + P(A \mid B_n)P(B_n)}.$$

Definition (Independent)

Events $A, B \subset S$ are called independent if $P(AB) = P(A)P(B)$.

Proposition

Let $A, B \subset S$ be events such that $P(B) > 0$. Then, A and B are independent if and only if $P(A \mid B) = P(A)$.

Proposition

If A and B are independent events, then A and B^c are independent.

Definition (Pairwise Independent)

Events $A_1, A_2, ..., A_n$ are said to be pairwise independent if $P(A_i A_j)$ $= P(A_i)P(A_j)$ for all $i \neq j$.

Definition (Independent)

Events $A_1, A_2, ..., A_n$ are said to be independent if $P(A_{i_1} A_{i_2} ... A_{i_k})$ $= P(A_{i_1})P(A_{i_2})...P(A_{i_k})$ for all $1 \leq i_1 < i_2 < ... < i_k \leq n$, $2 \leq k \leq n$.

Definition (Discrete Random Variable)

A discrete random variable is a function $X : S \to \Delta$ where Δ is a countable set.

Definition (Probability Mass Function)

The probability mass function of X is defined as follows: $p : \Delta \to [0,1]$.

$p(x) = P(\{s \in S \mid X(x) = x\})$ for $x \in \Delta$. We abbreviate $p(x) = P(X = x)$.

Definition (Cumulative Distribution Function)

The cumulative distribution function of a real-valued discrete random variable X is the function $F:\mathbf{R}\to[0,1]$ defined by $F(x)=P(X\le x)=P(\{s\in S \mid X(s)\le x\})$

Fact

Let X have distribution F. Then

(a) $P(X>x)=1-P(X\le x)=1-F(x)$ for all $x\in\mathbf{R}$.

(b) $P(x<X\le y)=F(y)-F(x)$ for all $y>x$

Definition (Expected Value)

The expectation (expected value) of a random variable X is defined by

$E[X]=\sum_{x:p(x)>0} xp(x)$

Proposition

Let X be a discrete random variable with probability mass function p.

For any function $g:\mathbf{R}\to\mathbf{R}$, $E[g(X)]=\sum_{x\in D(p)} g(x)p(x)$

This is true whenever the sum makes sense.

Definition (Variance)

If X is a random variable with a finite mean μ, we define the variance of X by:

$Var(X)=E[(X-\mu)^2]$.

Proposition

$Var(X)=E[X^2]-\left(E[X]\right)^2$

Proposition

For all $a,b\in\mathbf{R}$, $\mathrm{var}(aX+b)=a^2\,\mathrm{var}(X)$

Definition (Standard Deviation)

The standard deviation of a random variable X is defined by $SD(X)=\sqrt{Var(X)}$

Definition (Bernoulli Random Variable)

A random variable is called a Bernoulli random variable if its probability mass function is given by: $p(0)=1-p$ and $p(1)=p$ for some $0\le p\le 1$.

Definition (Binomial Random Variable)

A random variable with the following probability mass function is called a Binomial random variable with parameters n and p:

$$p(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \text{ for all } k = 0,1,\ldots,n.$$

Proposition

If $X \sim Binomial(n, p)$, then $E(X) = np$ and $Var(X) = np(1-p)$

Definition (Poisson Random Variable)

A Poisson random variable approximates a $Binomial(n, p)$ random variable with n very large and mean $\lambda = np$, $\lambda > 0$

Proposition

Let $X_n \sim Binomial(n, \lambda/n)$. Then, for all $k \geq 0$, $\lim_{n\to\infty} P(X_n = k) = e^{-\lambda}\frac{\lambda^k}{k!}$

Fact

$p(k) = e^{-\lambda}\frac{\lambda^k}{k!}$ for $k = 0,1,2,\ldots$ defines a probability mass function.

Definition (Poisson(λ)-distributed)

Let $\lambda > 0$. A random variable with the probability mass function given by $p(k) = e^{-\lambda}\frac{\lambda^k}{k!}$ is called $Poisson(\lambda)$-distributed.

Proposition

If $X \sim Poisson(\lambda)$, then $E[X] = \text{var}(X) = \lambda$.

Fact

$p(k) = (1-p)^{k-1} p$, $k \geq 1$ defines a probability mass function

Definition (Geometric(p)-distributed)

Let $p \in (0,1)$. A random variable with the probability mass function $p(k) = (1-p)^k p$ is called $geometric(p)$-distributed with parameter p.

Proposition

If $X \sim geometric(p)$, then $E[X] = \frac{1}{p}$ and $\text{var}(X) = \frac{1-p}{p^2}$.

Definition (Continuous Random Variable)

We say $X : S \to \mathbf{R}$ is a continuous random variable if there exists a function $f : \mathbf{R} \to [0,\infty)$ with the following properties:

(a) $f(x) \geq 0$ for all $x \in \mathbf{R}$.

(b) $\int_{-\infty}^{\infty} f(x)dx = 1$

(c) $P(X \in A) = \int_{x \in A} f(x)dx$ for all intervals $A \subset \mathbf{R}$

Definition (Density Function)

The function f defined above is called the <u>density function</u> of X.

Fact

Let X be a continuous random variable. Then, $P(X = a) = 0$ for all $a \in \mathbf{R}$.

Definition (Uniform(α, β)-Distributed)

A random variable is <u>uniformly distributed on the interval</u> (α, β) if its density function is given by $f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & x \in (\alpha, \beta) \\ 0 & otherwise \end{cases}$

Proposition

Let X be a continuous random variable. Its cumulative distribution function F has the following properties:

(a) F is increasing

(b) $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$

(c) $F'(x) = f(x)$ where f is the density of X.

Definition (Expectation)

We defined the <u>expectation</u> of a continuous random variable X with density f by:

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx \text{ whenever this integral makes sense.}$$

Proposition

Let X be a continuous random variable with density function f_X, and let $g : \mathbf{R} \to \mathbf{R}$. Then $E[g(X)] = \int_{-\infty}^{\infty} g(X) f_X(x)dx$ whenever the integral makes sense.

Lemma

For a non-negative random variable Y, $E[Y] = \int_0^{\infty} P(Y > y)dy$.

Corollary

For all $a, b \in \mathbf{R}$, $E[aX + b] = b + aE[X]$

Definition (Exponential (λ)-Distributed)

Let $\lambda > 0$. A random variable is said to be an exponential(λ) distribution if it has the following density function:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & else \end{cases}$$

Proposition

Let $X \sim exponential(\lambda)$. Then, $E[X] = \frac{1}{\lambda}$, $Var(X) = \frac{1}{\lambda^2}$.

Proposition (Memoryless Property)

Let X be exponential (λ)-distributed. Then, for all $s, t > 0$,

$$P(X > s + t \mid X > t) = P(X > s)$$

Theorem

Let X be a continuous random variable with density f_X. Let g be a continuous, strictly monotone function. Then the random variable $Y = g(X)$ has the following density:

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \frac{d}{dy}\left[g^{-1}(y)\right] \right| & y = g(x) \text{ for some } x \\ 0 & else \end{cases}$$

Definition

We say that X is <u>normally distributed</u> with parameters μ and σ^2 if the density is given by the following:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \; x \in \mathbf{R}.$$

Fact

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \; x \in \mathbf{R} \text{ is a probability density}$$

Theorem (de Moivre and Laplace)

Let $X_1, X_2, \ldots$ be independent random variables with $P(X_i = 1) = p$ and $P(X_i = 0) = 1 - p$. Let $S_n = X_1 + \ldots + X_n$. Then,

$$\lim_{n\to\infty} P\left(a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b \right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

Proposition

Let $X \sim N(\mu,\sigma^2)$. Then, $E[X] = \mu$, $\mathrm{var}(X) = \sigma^2$.

Definition (Moment-Generating Function)

Let X be a random variable. The <u>moment-generating function</u> of X is defined by

$$M(t) = E[e^{tX}] \text{ for } t \in \mathbf{R}.$$

Theorem

Suppose $M(t)$ is finite on some open interval containing the origin. Then,

(a) $E[X] = M'(0)$

(b) For all $k \ge 1$, $E[X^k] = M^{(k)}(0)$

Definition (Joint Cumulative Distribution Function)

The <u>joint-cumulative distribution function</u> of random variables X and Y is defined by $F : \mathbf{R} \times \mathbf{R} \to [0,1]$:

$$F(a,b) = P(X \le a, Y \le b)$$

Definition (Joint Probability Mass Function)

If X and Y are discrete random variables, then we define their <u>joint probability mass function</u> by:

$$p(x,y) = P(X = x, Y = y)$$

Definition (Marginal Distribution)

Given $p(x,y)$, the <u>marginal distributions</u> of X and Y are given by:

$$p_X(x) = \sum_{y\in\Delta} p(x,y)$$

$$p_Y(y) = \sum_{x\in\Delta} p(x,y)$$

Definition (Jointly Continuous)

We say that X and Y are <u>jointly continuous</u> if there is a function:

$$f : \mathbf{R} \times \mathbf{R} \to [0,\infty) \text{ such that } P(X \in A, Y \in B) = \int_A \int_B f(x,y)\,dy\,dx$$

for all intervals $A, B \subset \mathbf{R}$.

Definition (Joint Probability Density Function)

The function f defined above is called the joint probability density function of X and Y.

Fact

If X and Y are jointly continuous, then both X and Y are continuous random variables.

Definition (Independent)

Two random variables X and Y are called independent if for all intervals $A, B \subset \mathbf{R}$,

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

Proposition

(a) Discrete random variables X and Y are independent if and only if $p(x, y) = p_X(x)p_Y(y)$ for all x, y.

(b) Random variables X and Y with a joint density function $f(x, y)$ are independent if and only if $f(x, y) = f_X(x)f_Y(y)$ for all x, y.

Proposition

Let X and Y be two independent, continuous random variables with densities f_X, f_Y. Then, $X + Y$ is a continuous random variable with density function:

$$f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a - y)f_Y(y)dy$$

Proposition

Let X_1 and X_2 be normally distributed with parameters (μ_1, σ_1^2) and (μ_2, σ_2^2) respectively. If X and Y are independent, then $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Proposition

If X and Y are independent and Poisson-distributed with parameters λ and μ respectively, then $X + Y \sim Poisson(\lambda + \mu)$.