

Random Experiment

A procedure with several possible outcomes.

Sample Space

The set of all possible outcomes of a random experiment.

Sample Point

An individual possible outcome from a random experiment.

Event

A subset of the sample space.

Probability (Kolmogorov Definition)

Let S be a sample space. A probability \mathbb{P} is a function that assigns to each event E a positive number with the following properties:

Axioms of Probability

1. $0 \leq \mathbb{P}(E) \leq 1$
2. $\mathbb{P}(S) = 1$
3. If E_1, E_2, E_3, \dots are events with $E_i \cap E_j = \emptyset$ for all $i \neq j$, then $\mathbb{P}(E_1 \cup E_2 \cup \dots) = \mathbb{P}(E_1) + \mathbb{P}(E_2) + \dots$

Probability Space

A pair (S, \mathbb{P}) , where S is a sample space and \mathbb{P} is a probability.

Complement

Let E be an event. Then the complement of E , denoted \overline{E} or E^c , is the set which contains all elements in S which are not in E .

$$\mathbb{P}(\overline{E}) = 1 - \mathbb{P}(E).$$

Discrete Uniform Distribution

Suppose $|S| = n$, so $S = \{x_1, x_2, \dots, x_n\}$. The discrete uniform distribution is the probability \mathbb{P} which assigns to each simple event the number $\frac{1}{n}$.

$$\text{Thus, } \mathbb{P}(E) = \frac{|E|}{|S|}.$$

Permutation

Assume you have n items. The ways these items can be arranged in order are called permutations.

There are $\frac{n!}{(n-k)!}$ permutations of n unique items.

Choose

Without caring about order, the number of ways to choose k items from a group of n items is $\binom{n}{k}$.

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

Conditional Probability

Let (S, \mathbb{P}) be a probability space, and A, B be two events with $\mathbb{P}(B) > 0$.

The conditional probability of A given B , denoted $\mathbb{P}(A|B)$, is defined as $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.

Independence

A and B are independent, denoted $A \perp\!\!\!\perp B$ if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
(Also true for more than 2 events)

Random Variable

A random variable is a function which assigns a real number to each sample point:

$$X : S \rightarrow \mathbb{R}$$

$$(X(\text{sample point}) = \#)$$

$$\text{Notation: } \mathbb{P}(X = k) = P_X(k).$$

Probability Mass Function (Distribution Function)

$P_X(k) : \mathbb{R} \rightarrow [0, 1]$ is called the probability mass function.

Note that $0 \leq P_X(k) \leq 1$.

$$\sum_k P_X(k) = 1.$$

(Assuming that X takes a countable number of possible values.)

When X has a countable number of possible values, it is called a discrete random variable.

Binomial Random Variable

Let $X = \#$ of successes in n independent trials of an experiment with two possible outcomes. Let p denote the probability of success in each independent trial. (often $q = 1 - p$ denotes the probability of failure) Then X is a binomial random variable if $P_X(k) = \mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$ for all integers $0 \leq k \leq n$.

Example: Let X be a binomial random variable with parameters n and p . (Denoted $X \sim \text{Binomial}(n, p)$, where \sim means “is distributed as”)

The PMF of X is then $P_X(k) = \binom{n}{k} p^k (1 - p)^{n-k}$. ($0 \leq n \leq k$) This is called the binomial distribution.

Geometric Random Variable

Let $X = \#$ of independent trials of an experiment with two possible outcomes up to and including the first success. Then, X is a geometric random variable with parameter p , denoted $X \sim \text{Geometric}(p)$.

$$P_X(k) = \mathbb{P}(X = k) = (1 - p)^{k-1} p.$$

Bayes' Theorem

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

Expected Value

Let $X : S \rightarrow I = \{x_1, x_2, \dots\} \subseteq \mathbb{R}$ be a discrete random variable. The expected value of X , denoted $\mathbb{E}[X] = \mathbb{E}X$, is defined as

$$\mathbb{E}X = \sum_{x_k \in I} \mathbb{P}(X = x_k) \cdot x_k.$$

If Y is a binomial random variable with parameters n, p , then $\mathbb{E}[Y] = np$.

If Z is a geometric random variable with parameter p , then $\mathbb{E}[Z] = \frac{1}{p}$.

Linearity of Expectation

Let X_1, X_2, \dots, X_n be random variables and $\alpha_1, \alpha_2, \dots, \alpha_n \in \mathbb{R}$.

Then define $X = \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n = \sum_{k=1}^n \alpha_k X_k$.

Then, $\mathbb{E}X = \mathbb{E}[\alpha_1 X_1 + \dots + \alpha_n X_n] = \alpha_1 \mathbb{E}X_1 + \dots + \alpha_n \mathbb{E}X_n$.

A Helpful Trick

When you're trying to do a complicated expected value, try breaking up the complicated variable into a sum of many simple variables. (Hat check problem, sock problem)

Variance

Let $X : S \rightarrow \mathbb{R}$ be a discrete random variable. The variance of X , denoted $\mathbb{V}[X]$, is a number describing the spread of possible values:

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

The standard deviation is given by

$$\sqrt{\mathbb{V}[X]}.$$

This will sometimes be denoted σ_X .

To compute variance directly:

$$\mathbb{V}[X] = \sum_k \mathbb{P}(X = k) \cdot (k - \mathbb{E}[X])^2.$$

But the actual way you will calculate variance is this:

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

Variance of Geometric Random Variable

Let Y be a geometric random variable with parameter p . Then,

$$\mathbb{V}[Y] = \frac{1-p}{p^2}.$$

Variance of Binomial Random Variable

Let X be a binomial random variable with parameters n, p . Then,

$$\mathbb{V}[X] = np(1-p).$$

Independence of Random Variables

Let $X, Y : S \rightarrow \mathbb{R}$ be discrete random variables. We say X and Y are independent if $\mathbb{P}(X = a \text{ and } Y = b) = \mathbb{P}(X = a)\mathbb{P}(Y = b)$ for all $a, b \in \mathbb{R}$.

Breaking up sums of independent random variables in variances

If X and Y are independent, then $\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y]$.

Breaking up products of independent random variables in expected values

If X and Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Scaling and Shift

Let $a \in \mathbb{R}$ and X be a discrete random variable. Then,

$$\begin{aligned}\mathbb{V}[aX] &= a^2\mathbb{V}[X] \\ \mathbb{V}[a + X] &= \mathbb{V}[X]\end{aligned}$$

Markov's Inequality

Let X be a nonnegative random variable. Then, $\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}X}{a}$ for all $a > 0$.

Chebyshev's Inequality

Let X be a random variable. Then, $\mathbb{P}(|X - \mathbb{E}X| \geq a) \leq \frac{\mathbb{V}X}{a^2}$ for all $a > 0$.

Weak Law of Large Numbers

Let X_1, X_2, \dots be independent, identically-distributed random variables. Then, for all $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - \mathbb{E}[X_n]\right| \geq \epsilon\right) = 0.$$

In other words, the average of the first n trials of an experiment approach the expected value of the n^{th} trial in the long run.

Poisson Random Variable

Let X be the number of events which occur in a fixed time interval $(0, T)$.

If X satisfies:

1. The number of events in two disjoint subintervals are independent.
2. The number of successes per unit of time, usually referred to as the rate λ , is independent of time.
3. In an infinitesimal interval, $[t, t + \delta t)$, there is at most one event.

then X is a Poisson distributed random variable with parameters λ, T .

$$\mathbb{P}(X = k) = \frac{e^{-\lambda T} (\lambda T)^k}{k!} \text{ for } k \in \mathbb{N}_0.$$

$$\mathbb{E}X = \lambda T.$$

$$\mathbb{V}X = \lambda T.$$

Sometimes, since λ and T are always a product, this is shortened to having only one parameter, λ , which is actually λT .

Hypergeometric Random Variable

Suppose you have N items of which K are special. From N total items, you pick n uniformly at random. Let X be the number of special items picked. Then, X is said to be a hypergeometric random variable with parameters N, K, n .

$$\mathbb{P}(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}} \text{ for } \max\{0, n + K - N\} \leq k \leq \min\{K, n\}$$

Fact about continuous random variables

If $X : S \rightarrow \mathbb{R}$ is continuous, then $\mathbb{P}(X = x) = 0$ for all $x \in \mathbb{R}$.

Cumulative Distribution Function

The cumulative distribution function of a random variable X is the function $F(x) = \mathbb{P}(X \leq x)$.

Every CDF is non-decreasing.

The CDF of a discrete random variable is a step function.

$$\mathbb{P}(a \leq X \leq b) = F(b) - F(a) \text{ for any } a < b.$$

$$\lim_{x \rightarrow \infty} F_X(x) = 1, \text{ and } \lim_{x \rightarrow -\infty} F_X(x) = 0.$$

Continuous Random Variable

A random variable is continuous if its CDF is continuous.

Probability Density Function (PDF)

The probability density function of a continuous random variable X with CDF $F_X(x)$ is the function

$$f(x) = \frac{d}{dx} F_X(x)$$

for all x such that $F_X(x)$ is differentiable.

$$f_X(x) \geq 0$$

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

$$\int_a^b f_X(x) dx = F_X(b) - F_X(a) = \mathbb{P}(a < X \leq b)$$

Expected value and variance of a continuous random variable

Let X be a continuous random variable with PDF f_X . The expected value of X is given by

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

The variance of X is given by the same formula as before. All previously-stated properties of expectation and variance still hold.

CDF of Geometric Random Variables

$$1 - (1 - p)^{\lfloor x \rfloor}$$

Uniform Distribution

A random variable X is uniformly distributed on (Θ_1, Θ_2) , denoted

$$X \sim \text{Uniform}(\Theta_1, \Theta_2),$$

if density is given by

$$f_X(x) = \begin{cases} \frac{1}{\Theta_2 - \Theta_1} & \Theta_1 < x < \Theta_2, \\ 0 & \text{else.} \end{cases}$$

Then,

$$F_X(x) = \begin{cases} 0 & x < \Theta_1, \\ \frac{x - \Theta_1}{\Theta_2 - \Theta_1} & \Theta_1 \leq x \leq \Theta_2, \\ 1 & x > \Theta_2. \end{cases}$$

If (c, d) is a subinterval of (Θ_1, Θ_2) , then $\mathbb{P}(c < X < d) = \int_c^d \frac{1}{\Theta_2 - \Theta_1} dx = \frac{d - c}{\Theta_2 - \Theta_1}$. This is also obvious.

$$\mathbb{E}X = \frac{\Theta_1 + \Theta_2}{2}.$$

$$\mathbb{V}X = \frac{1}{12}(\Theta_2 - \Theta_1)^2.$$

Normal (Gaussian) Distribution

We say X is a Gaussian (normal) random variable if its density (PDF) is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

We say $X \sim \text{Normal}(\mu, \sigma^2)$.

$$\mathbb{E}X = \mu.$$

$$\mathbb{V}X = \sigma^2.$$

If $X \sim \text{Normal}(\mu_X, \sigma_X^2)$ and $Y \sim \text{Normal}(\mu_Y, \sigma_Y^2)$ are independent, then $X+Y \sim \text{Normal}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

$$aX + b \sim \text{Normal}(a\mu_X + b, a^2\sigma_X^2).$$

If you combine many sources of randomness, you get a normal distribution (Central Limit Theorem)

Normal distribution maximizes entropy (???)

If you don't know anything about a random variable, Normal is a decent guess.

~ 68% of probability is within one standard deviation.

~ 95% for two standard deviations.

~ 99.7% for three.

Standard Normal Distribution

If X is a Gaussian random variable, then

$$Z = \frac{X - \mu_X}{\sigma_X} \sim \text{Normal}(\mu = 0, \sigma^2 = 1)$$

is a standard normal random variable. To compute probabilities for X , convert to probabilities involving Z and use precomputed values for standard normal.

Exponential Random Variable

The random variable X is exponentially distributed with parameter $\lambda > 0$, if its density is

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0, \\ 0 & \text{else.} \end{cases}$$

Then, its CDF is

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

Sometimes parameterized using $\beta = \frac{1}{\lambda}$.

X is the waiting time until first event when events occur at rate λ per unit time.

This is kind of like the continuous version of the geometric distribution.

$$\mathbb{E}X = \frac{1}{\lambda}$$

$$\mathbb{V}X = \left(\frac{1}{\lambda}\right)^2$$

The exponential distribution is the only continuous distribution with memorylessness:

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

In other words, if no event occurs by time t , the probability of the event occurring s time units in the future is the same as the probability of it occurring after waiting s from the outset.

Gamma Random Variable

X is said to be Gamma distributed with parameters $r \in \mathbb{N}$ and $\lambda > 0$ if its density is

$$f_X(x) = \begin{cases} \frac{\lambda^r}{(r-1)!} x^{r-1} e^{-\lambda x} & x > 0, \\ 0 & \text{else.} \end{cases}$$

X is the waiting time until r events have occurred when events occur at a constant rate λ .

$$\mathbb{V}X = \frac{r}{\lambda^2}$$

Joint Distribution for Discrete Random Variables

Let $X : S \rightarrow \{x_1, x_2, x_3, \dots\}$ and $Y : S \rightarrow \{y_1, y_2, y_3, \dots\}$ be discrete random variables on the same sample space S . The joint distribution of the random vector (X, Y) is given by

$$\mathbb{P}(X = x_k, Y = y_l),$$

sometimes denoted $P_{X,Y}(x_k, y_l)$.

CDF of a Random Vector

Let $X, Y : S \rightarrow \mathbb{R}$ be random variables. The CDF of the random vector (X, Y) is given by

$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y).$$

Joint Distribution for Continuous Random Variables

Let $X, Y : S \rightarrow \mathbb{R}$ be continuous random variables. The joint distribution or joint density of (X, Y) is the function

$$f_{X,Y}(x, y) = \frac{\delta^2}{\delta x \delta y} F_{X,Y}(x, y).$$

$$f_{X,Y}(x, y) \geq 0 \forall x, y \in \mathbb{R}$$

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

Marginal Distribution of Discrete a Random Variable

Let X, Y be discrete random variables with joint distribution $P_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y)$. The marginal distribution of Y , denoted $P_Y(y)$, is defined by

$$P_X(x) = \sum_y P_{X,Y}(x, y).$$

Conditional Distribution of Discrete a Random Variable

The conditional distribution of X given $Y = y$ is

$$P_{X|Y}(x) = \frac{P_{X,Y}(x, y)}{P_Y(y)}.$$

Independence for Joint Distributions of Discrete Random Variables

X and Y are independent if and only if $P_{X,Y}(x, y) = P_X(x)P_Y(y)$ for all x, y .

Marginal Distribution of a Continuous Random Variable

Let X, Y be continuous random variables with density $f_{X,Y}(x, y)$. The marginal distribution of X is

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy,$$

and the marginal distribution of Y is

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx.$$

Conditional Distribution of Continuous Variables

The conditional distribution of X given $Y = y$ is

$$P_{X|Y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

Independence for Joint Distributions of Continuous Random Variables

X and Y are independent if and only if $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ for all $x, y \in \mathbb{R}$.

Expected Value of a Function of Discrete Random Variables

Let (X, Y) be a discrete random vector and $P_{X,Y}(x, y)$ their joint distribution. Let

$$g(x, y) \rightarrow \mathbb{R}$$

be a function. Then,

$$\mathbb{E}[g(X, Y)] = \sum_x \sum_y P_{X,Y}(x, y)g(x, y).$$

Expected Value of a Function of Continuous Random Variables

