Prob-Stats Workshop

by

Prepleaf

https://www.prepleaf.com

Section 1

Random Experiment

Outcomes, Events, Sample Space

Probability

Random Experiment

A **random experiment** is a process characterized by the following **properties**:

(i) It can be repeated arbitrarily often under same conditions,

(ii) The outcome of each experiment depends on chance and hence cannot be predicted with certainty.

Example: Rolling a dice, tossing a coin etc. etc.

Outcomes, Events and Sample Space

- Outcome: An **outcome** is the result of a single trial of an experiment.
- Sample Space (Ω): The universal set of all possible outcomes.
- Event ($\mathcal{E}$): An **event** is one or more outcomes of an experiment.
- We are interested in $\mathcal{E} \subseteq \Omega$, so we perform the experiment N times and see how many times $\mathcal{R}$ actually Occurred and call is relative frequency.

$$\# \mathcal{E} = f/N$$

$$Probability(\mathcal{E}) = \lim_{n \to \infty} f/N$$

Probability

Probability is the measure of the likelihood that an event will occur.

Axioms of Probability:

1: For any event A, P(A)≥0.

2: Probability of the sample space SS is P(S)=1.

3: If A1,A2,A3,⋯A1,A2,A3,⋯ are disjoint events, then

P(A1 ∪ A2 ∪ A3 …) = P(A1) + P(A2) + P(A3) + ……

Puzzle 1

Three ants are sitting at the three corners of an equilateral triangle. Each ant starts randomly picks a direction and starts to move along the edge of the triangle. What is the probability that none of the ants collide? Assume that they move at same rate.

Puzzle: 2

You have a biased coin. You need to create a random experiment such that the experiment behaves like a random coin toss experiment, which means, there will be 2 outcomes and there will be equal probability for occurrence of both.

Section 2

Random Variable

Probability Mass Function

Probability Density Function

Cumulative Density Function

Random Variable

A Random Variable X is a function which maps the Sample Space to Real Line.

$$X : \Omega \ \rightarrow \mathcal{R}$$

Typically, random variables are denoted using using upper case letters X(w) or more simply X.

Discrete Random Variable

If the random variable, $X(w)$ takes only a finite or countable infinite values.

Example: Coin Toss

- $X(\text{head}) = 0$
- $X(\text{tails}) = 1$

Continuous Random Variable

If X(w) takes uncountable infinite number of possible values, then it’s called Continuous Random Variable.

Example,

Height of Students in a class,

Time required to reach the office

Probability Mass Function (PMF)

It is a function that gives the probability that a discrete random variable X exactly equals to value x.

$$P\ (X = x)$$

In our example before, if we assume equally likely probabilities, then

$P(X = 0) = 0.4$ implies that probability of heads is 0.4

$P(X = 1) = 0.6$ implies 0.2 probability of tails is 0.6

Probability Density Function (PDF)

- It is a function analogous to PMF but it is defined for a continuous random variable. The functional value at any given point can be interpreted as providing a **RELATIVE LIKELIHOOD** that the value of the random variable would equal that sample.

Question: What is the probability that $X = 0, X$ follows a standard normal distribution.

Some Standard Distributions

Distribution	PMF/PDF and Support	Expected Value	Variance	MGF
Bernoulli Bern(p)	$P(X=1)=p$ $P(X=0)=q=1-p$	p	pq	$q+pe^t$
Binomial Bin(n,p)	$P(X=k)=\binom{n}{k}p^k q^{n-k}$ $k \in \{0,1,2,\dots n\}$	np	npq	$(q+pe^t)^n$
Geometric Geom(p)	$P(X=k)=q^k p$ $k \in \{0,1,2,\dots\}$	q/p	q/p^2	$\frac{p}{1-qe^t}$, $qe^t<1$
Negative Binomial NBin(r,p)	$P(X=n)=\binom{r+n-1}{r-1}p^r q^n$ $n \in \{0,1,2,\dots\}$	rq/p	rq/p^2	$(\frac{p}{1-qe^t})^r$, $qe^t<1$
Hypergeometric HGeom(w,b,n)	$P(X=k)=\binom{w}{k}\binom{b}{n-k}/\binom{w+b}{n}$ $k \in \{0,1,2,\dots,n\}$	$\mu=\frac{nw}{b+w}$	$\left(\frac{w+b-n}{w+b-1}\right)n\frac{\mu}{n}(1-\frac{\mu}{n})$	messy
Poisson Pois(λ)	$P(X=k)=\frac{e^{-\lambda}\lambda^k}{k!}$ $k \in \{0,1,2,\dots\}$	λ	λ	$e^{\lambda(e^t-1)}$

Uniform $\mathrm{Unif}(a,b)$	$f(x) = \frac{1}{b-a}$ $x \in (a,b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$
Normal $\mathcal{N}(\mu,\sigma^2)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$ $x \in (-\infty,\infty)$	μ	σ^2	$e^{t\mu + \frac{\sigma^2 t^2}{2}}$
Exponential $\mathrm{Expo}(\lambda)$	$f(x) = \lambda e^{-\lambda x}$ $x \in (0,\infty)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda-t}, t < \lambda$
Gamma $\mathrm{Gamma}(a,\lambda)$	$f(x) = \frac{1}{\Gamma(a)}(\lambda x)^a e^{-\lambda x}\frac{1}{x}$ $x \in (0,\infty)$	$\frac{a}{\lambda}$	$\frac{a}{\lambda^2}$	$\left(\frac{\lambda}{\lambda-t}\right)^a, t < \lambda$
Beta $\mathrm{Beta}(a,b)$	$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$ $x \in (0,1)$	$\mu = \frac{a}{a+b}$	$\frac{\mu(1-\mu)}{(a+b+1)}$	messy
Log-Normal $\mathcal{LN}(\mu,\sigma^2)$	$\frac{1}{x\sigma\sqrt{2\pi}} e^{-(\log x-\mu)^2/(2\sigma^2)}$ $x \in (0,\infty)$	$\theta = e^{\mu+\sigma^2/2}$	$\theta^2(e^{\sigma^2}-1)$	doesn't exist
Chi-Square χ^2_n	$\frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}$ $x \in (0,\infty)$	n	$2n$	$(1-2t)^{-n/2}, t < 1/2$
Student-t t_n	$\frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)}(1+x^2/n)^{-(n+1)/2}$ $x \in (-\infty,\infty)$	0 if $n > 1$	$\frac{n}{n-2}$ if $n > 2$	doesn't exist

Cumulative Density Function (CDF)

It is a function which maps the value of a Random Variable to [0,1], i.e. it specifies a probability measure for that value. It is defined as

$$F_X(x) = P(X<=x).$$

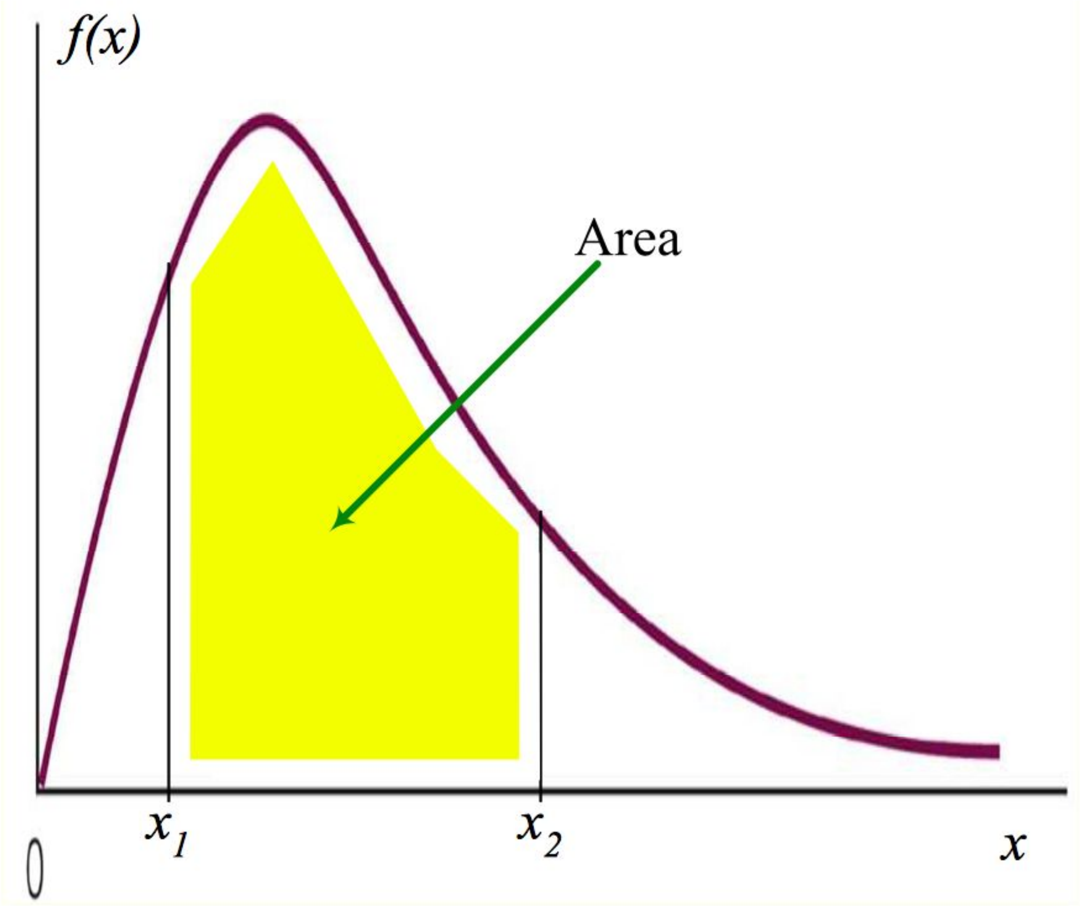
For discrete random variable, it is defined as

$$F(x) = \mathrm{P}(X \leq x) = \sum_{x_i \leq x} \mathrm{P}(X = x_i) = \sum_{x_i \leq x} p(x_i).$$

For continuous random variable, it is defined as

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt.$$

Interpreting CDF



Properties of CDF

Properties:

- $0 <= F_X(x) <= 1$
- $\lim_{x \to -\inf} F_X(x) = 0$
- $\lim_{x \to \inf} F_X(x) = 1$
- $x <= y$ =) $F_X(x) <= F_X(y)$.

Puzzle 3

A rod of length L is cut into three halves. What is the probability that the three cut pieces of rod will form a triangle?

Hint

- Lets assume the lengths of three halves are $a, b\ and\ L - a - b$
- What is the sample space?
- What are the condition under which they form triangle?

Section 3

Conditional Probability

Bayes Theorem

Independence of Events

Conditional Probability

- Conditional probability formalizes the notion of "having information" about the outcome of a probabilistic experiment.

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Bayes Theorem

Bayes' theorem describes the probability of an event, based on prior knowledge of conditions that might be related to the event. Bayes Rule comes from the law of total probability.

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)} \cdot$$

$$P(B) = \sum_j P(B \mid A_j) P(A_j),$$

$$\Rightarrow P(A_i \mid B) = \frac{P(B \mid A_i)\, P(A_i)}{\sum_j P(B \mid A_j)\, P(A_j)} \cdot$$

Continued

For proposition A and evidence B,

- P (A), the prior, is the initial degree of belief in A.
- P (A | B) is the "posterior," is the degree of belief having accounted for B.
- The quotient P(B |A)/P(B) represents the support B provides for A.

Independence of Events

Two events X and Y are called independent if $P(X|Y) = P(X)$.

Another way to write the above condition is $P(XY) = P(X)P(Y)$.

In simple language, knowledge of the event Y doesn't help us in giving more information about event X.

Section 4

Expectation

Variance

Covariance

Correlation

Expectation

The expected value of a random variable, intuitively, is the long-run average value of repetitions of the experiment it represents.

For discrete random variables, it is defined as

$$\mathrm{E}[X] = \sum_{i=1}^{\infty} x_i \, p_i$$

For continuous random variables, it is defined as

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x) \, dx.$$

Expectation (More Precise Definition)

If X *is discrete, then the expectation of* $g(X)$ *is defined as, then*

$$E[g(X)] = \sum_{x \in \mathcal{X}} g(x) f(x),$$

where f *is the probability mass function of* X *and* $\mathcal{X}$ *is the support of* X.

If X *is continuous, then the expectation of* $g(X)$ *is defined as,*

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx,$$

where f *is the probability density function of* X.

Linearity of Expectation

- This property will be exploited the most while doing probability puzzles.

$$\mathrm{E}[X+Y]=\mathrm{E}[X]+\mathrm{E}[Y],$$

$$\mathrm{E}[aX]=a\,\mathrm{E}[X],$$

where X and Y are (arbitrary) random variables, and a is a scalar.

Variance

Variance is the expectation of the squared deviation of a random variable from its mean. Informally, it measures how far a set of (random) numbers are spread out from their average value.

Formally, it is defined as follows:

$$\begin{aligned}
\operatorname{Var}(X) &= \operatorname{E}\left[(X - \operatorname{E}[X])^2\right] \\
&= \operatorname{E}\left[X^2 - 2X\operatorname{E}[X] + \operatorname{E}[X]^2\right] \\
&= \operatorname{E}\left[X^2\right] - 2\operatorname{E}[X]\operatorname{E}[X] + \operatorname{E}[X]^2 \\
&= \operatorname{E}\left[X^2\right] - \operatorname{E}[X]^2
\end{aligned}$$

Standard Deviation of a random variable is the square root of Variance.

Covariance

Covariance is a measure of the joint variability of two random variables. The sign of the covariance therefore shows the tendency in the linear relationship between the variables. Formally, it is defined as,

$$
\begin{aligned}
\operatorname{cov}(X,Y) &= \mathrm{E}[(X - \mathrm{E}[X])\,(Y - \mathrm{E}[Y])] \\
&= \mathrm{E}[XY - X\,\mathrm{E}[Y] - \mathrm{E}[X]Y + \mathrm{E}[X]\,\mathrm{E}[Y]] \\
&= \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] - \mathrm{E}[X]\,\mathrm{E}[Y] + \mathrm{E}[X]\,\mathrm{E}[Y] \\
&= \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y].
\end{aligned}
$$

Covariance (… continued)

If the greater values of one variable mainly correspond with the greater values of the other variable, and the same holds for the lesser values, i.e., the variables tend to show similar behavior, the covariance is positive, and vice-versa.

Note: The magnitude of the covariance is not easy to interpret because it is not normalized. So, it is not used generally for interpretation.

Correlation

Cor(X,Y) is a measure of the linear correlation between two variables X and Y. It has a value between +1 and −1, where 1 is total positive linear correlation, 0 is no linear correlation, and −1 is total negative linear correlation. This is a normalized version of covariance.

Correlation is represented using rho, it is calculated as the ratio of covariance to the standard deviation of X and Y.

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y}$$

Property of Correlation

- If two variables X and Y are independent, then they have zero correlation but this DOESN'T hold other way round.
- It might happen that two variables have 0 correlation but they are still not independent.

Section 5

Chebyshev’s Inequality

Weak Law of Large numbers

Central Limit Theorem

Chebyshev's Inequality

- Let X be a random variable certain distribution with finite variance σ^2, then

$$P(|X - \mu| \geq k\sigma) \leq 1/k^2$$

Or, alternatively

$$P(|X - \mu| \geq \epsilon) \leq \sigma^2/\epsilon^2$$

Weak Law of Large numbers

- Let $\{X_n\}$ be a sequence of iid random variables with mean μ and variance $\sigma^2 < \infty$. And let, $\bar{X}_n = \frac{1}{n}\sum_i X_i$ Then,

$$P[|\bar{X}_n - \mu| \geq \epsilon] \leq \frac{\sigma^2}{n\epsilon^2}$$

We also say that $\bar{X}_n \longrightarrow \mu$ in probability

Central Limit Theorem

- Let $X_1, X_2, \cdots, X_n$ is a random sample from a distribution (**any**) that has mean μ and variance $\sigma^2 > 0$.
- Further let, $Y_n = \frac{\sqrt{n}\,(\bar{X}_n - \mu)}{\sigma}$
- Y_n converges in distribution to a standard normal distribution.

$$Y_n \rightarrow N(0,1)$$

Section 6

Puzzles!

Puzzle 4

There is a family in India. They try to conceive until they receive a Girl. What is the expected number of children will the family have?

Solution

- Assume P(boy) = p and P(Girl) = q

- Realize that family will keep having boy until they conceive a Girl

- Expectation $\sum n * p(b_1 b_2 b_{n-1} g_n)$

- Solve the AGG series to get the expectation

Puzzle 5

There is a square table of side 5 inches. There is a coin whose diameter is half inch. The coin is tossed and it is known that it will land on table. What is the probability that it will land completely on table.

Puzzle 5: Hint

- Think about the sample space; it is given that coin lands on the table
- Find the conditions under which it completely lands on the table
- Get the probability

Puzzle 6

Suppose there are 3 curtains, behind one door is a grand prize and the other two are worthless prizes. A contestant is asked to chose one door and the Monte Hall opens one of other two doors which contains worthless prize. Hall provides the contestant to switch the curtain after he has revealed one worthless prize. Should he switch or not?

Possibilities

Door 1	Door 2	Door 3	Switch/Not Switch
Goat	Goat	Car	Switch
Goat	Car	Goat	Switch
Car	Goat	Goat	Do not Switch

Puzzle 7

A coin is tossed till the time there are two consecutive heads. What is the expected number of tosses.

Puzzle 8

An urn contains n balls numbered $1,2,3,\ldots,n$. We remove k balls at random and add up their numbers. Find the expected value of this final number.

Puzzle 8: Solution

- Lets say you pick balls numbered $X_1, X_2, \ \dots, X_k$
- What is the expectation of X_i
- Calculate the Expectation of the $S = \ \sum X_i$
- Hint: Use Linearity of the Expectations

Puzzle 9

We throw m balls randomly, uniformly and independently into n bins. What is the expected number of empty bins.

Solution: Puzzle 9

- Calculate the probability of a particular ball not falling into particular bin

- From that, calculate probability that a particular bin is empty

- The expectation becomes trivial since bins are independent

Puzzle 10

You have 50 red marbles, 50 blue marbles and 2 jars. One of the jars is chosen at random and then one marble will be chosen from that jar at random. How would you maximize the chance of drawing a red marble? What is the probability of doing so? All 100 marbles should be placed in the jars.

Puzzle 11(Try this at home)

Player M has $1, and Player N has $2. Each play gives one of the players $1 from the other. Player M is enough better than enough Player N that he wins 2/3 of the plays. They play until one is bankrupt. What is the chance that player M wins.

Answers to Puzzles

- Puzzle 1: P(Not Colliding) =1/4
- Puzzle 2: Toss the biased coin twice. Define HT= New Head, TH= New Tails, Discard HH or TT and toss again
- Puzzle 3: P(Triangle) = ¼
- Puzzle 4: E(#Children)=1/q
- Puzzle 5: P(Falling inside the table) = (a-d)^2/a^2
- Puzzle 6: Always Switch as P(win|switch)=2/3
- Puzzle 7: E(# tosses until 2 Heads) =6

Answers to puzzles

- Puzzle 8: E(Sum of K balls) = k(n+1)/2
- Puzzle 9: P = $\mathrm{n} * \left(1 - \frac{1}{n}\right)^{m}$
- Puzzle 10: Put only 1 red marble in Jar A and rest all 49 red and 50 marbles in Jar B
- Puzzle 11: P(A winning)= 4/7

Resources

- Introduction to Mathematical Statistics- Hogg, Craig and Mckean
- Fifty challenging problems in probability by Frederick Mosteller
- https://static1.squarespace.com/static/54bf3241e4b0f0d81bf7ff36/t/55e9494fe4b011aed10e48e5/1441352015658/probability_cheatsheet.pdf
- http://www.madandmoonly.com/doctormatt/mathematics/dice1.pdf

Thank You!