Probability Study Guide

May 8, 2020

1 Combinatorics

If r experiments are to be performed such that each has n outcomes, there is a total of $n_1, n_2, ..., n_r$ possible outcomes of the r experiments. We use the factorial to find how many different ordered arrangements are possible. When order does not matter, we use the combination formula:

$$C(n,k) = \frac{n!}{(n-r)!r!}$$

2 Axioms

- The sample space is the set of all possible outcomes
- An event is a subset of the sample space:

$$P(E) = \frac{|E|}{|S|}$$

- $E \cup F$ consists of all outcomes of E and F
- $E \cap F = EF$ consists of all outcomes that are in both E and F
- if $EF = \emptyset$, E and F do not have any like outcomes and are mutually exclusive
- E^c is the complement of E that contains everything in sample space not in E

- DeMorgan's Laws $(\cup E_i)^c = \cap E_i^c$
- $0 \leq P(E) \leq 1$
- $P(S) = 1$
- $P(\cap E_i) = \sum P(E_i)$

3 Propositions

- $P(E^c) = 1 - P(E)$
- if $E \subset F$, then $P(E) \leq P(F)$
- $P(E \cup F) = P(E) + P(F) - P(EF)$
- $P(\bigcup^n E_i) = \Sigma P(E_i) - \Sigma P(E_{i_1} E_{i_2}) + ... + (-1)^{n+1} P(E_1 E_2 ... E_n)$

4 Bayes' Theorem

Bayes' theorem describes the probability of an event based on prior knowledge of conditions related to the event.

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

where $P(A)$ is the prior initial belief in event A, $P(A|B)$ is the posterior belief accounting for B, and $\frac{P(B|A)}{P(B)}$ is the support that B provides for A. The $P(B)$ can be expanded to

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

5 Discrete Random Variables

A random variable X is a function from a sample space to the real number. With discrete RV's, X take s in countably infinite set of numbers. The probability mass function gives the probability a discrete RV is exactly equal to some value, represented $P(X = i)$. The cumulative distribution function,$F(x) = P(X \leq x)$, is the probability a random variable is random variable is equal to or less than some value.

5.1 Expected Value

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

$$E(x) = \sum_{i=1}^{n} x_i p_i$$

$$E(g(x)) = \sum g(x_i)p(x_i)$$

$$E(ax + b) = aE(x) + b$$

5.2 Variance

Variance measures how far a set of random numbers are spread out from their mean.

$$var(x) = \sum_{i}^{2} = E(x^2) - [E(x)]^2$$

$$var(x + b) = v(x)$$

$$var(ax) = a^2 var(x)$$

5.3 Moment Generating Function

The expected values $E(X), E(X^2), ..., E(X^n)$ are called moments. The moment generating function is defined as:

$$M(t) = E(e^{tx}) = \sum e^{tx} f(x)$$

$$m^{'}(0) = E(X)$$

$$m^{''}(0) = E(X^2)$$

If $M_1(t) = M_2(t)$ then random variables $X_1 = X_2$

5.4 Discrete Distributions

5.4.1 Binomial (Bernoulli) Random Variable

The binomial distribution expresses the number of successes in a sequence of n independent binary experiments.

$$x_i = \begin{cases} 1 success \\ 0 failure \end{cases}$$

$$P(X_i = 1) = p, P(X_i = 0) = 1 - p$$

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$E(x) = np, var(x) = npq, \text{ where } q = (1-p)$$

5.4.2 Poisson Distribution

The Poisson distribution expresses the probability of a given number of events occurring in a fixed interval of time if the events occur with a known average rate, λ, and independently.

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

$$E(x) = var(x) = \lambda$$

5.4.3 Geometric Distribution

The geometric distribution expresses the probability of n Bernoulli trials yields the first success.

$$P(X = n) = (1-p)^{n-1} p$$

$$E(x) = \frac{1}{p}$$

$$var(x) = \frac{1-p}{p^2}$$

The cumulative distribution function express the probability of at least k trials before first success.

$$P(X \geq k) = (1-p)^{k-1}$$

5.4.4 Negative Binomial Distribution

The negative binomial distribution expresses the probability it takes k trials until rth success.

$$
\begin{aligned}
P(X = k) &= \binom{k-1}{r-1} p^r (1-p)^{k-r} \\
E(x) &= \frac{r}{p} \\
var(x) &= \frac{r(1-p)}{p^2}
\end{aligned}
$$

5.4.5 Hypergeometric Distribution

The hypergeometric distribution expresses the probability of k successes in n draws, without replacement, from a population of size N and k successes.

$$
\begin{aligned}
P(X = k) &= \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \\
E(X) &= \frac{nm}{N} \\
var(X) &= npq\binom{N-m}{N-1}
\end{aligned}
$$

6 Continuous Random Variable

X is a continuous random variable if there exists a function $f(x)$ such that $f(x)$ is the non-negative domain of $x \in (-\infty, \infty)$ such that $P(X \in B) =$

$\int_B f(x)dx$.

$$P(-\infty \le x \le \infty) = 1$$
$$P(X = a) = \int_a^a f(x)dx = 0$$
$$P(X < a) = \int_{-\infty}^a f(x)dx = F(a)$$
$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$
$$var(X) = \int_{-\infty}^{\infty} x^2 f(x)dx - \left(\int_{-\infty}^{\infty} xf(x)dx\right)^2$$
$$M(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f(x)dx$$

6.1 Continuous Distribution

6.1.1 Uniform Distribution

The uniform distribution is uniformly distributed on interval $[a, b]$ if it assumes all values on $[a, b]$. It is represented by the probability density function:

$$f(x) = \begin{cases} \frac{1}{b-a}, a \le x \le b \\ 0, \text{otherwise} \end{cases}$$

and the cumulative distribution function:

$$F(x) = P(X \le x) = \begin{cases} 0, x \le a \\ \frac{x-a}{b-a}, a \le x \le b \\ 1, otherwise \end{cases}$$

6.1.2 Exponential Distribution

The exponential distribution describes the time between events in the Poisson process, a process in which events occur continuously and independently at a constant average rate.

$$\begin{aligned}
f(x) &= \begin{cases} \lambda e^{-\lambda x}, x \geq 0, \lambda > 0 \\ 0, x < 0 \end{cases} \\
F(a) &= P(X \leq a) = 1 - e^{-\lambda a} \\
E(X) &= \frac{1}{\lambda} \\
var(X) &= \frac{1}{\lambda^2}
\end{aligned}$$

6.1.3 Normal Random Distribution

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

- if $\sigma = 1, \mu = 0$ then $f(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$ and z is the standard normal random variable
- $Y = aX + b$ then Y is a normal random variable with parameters $a\mu + b, a^2\sigma^2$
- $m(t) = e^{\frac{t^2}{2}}$ is the moment generating function for the standard normal RV
- the CDF for Z is $\Phi(Z) = P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x^2}{2}} dx$
- $P(a \leq x \leq b) = \Phi(b^*) - \Phi(a^*)$

7 Limit Theorems

- The law of large numbers says that the average of a sequence of random variables converges to the expected average
- The central limit theorem says the the sum of a large number of independent and identically distributed random variables has a distribution $\approx$ normal
- Markov's inequality: if X is a random variable such that $R_x \geq 0$ then for any $a > 0, P(X \geq a) \leq \frac{E(x)}{a}$

- Chebyshev's inequality: if X is a random variable with finite mean and variance then for any $k > 0, P(|x - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$

8 Joint Random Variables