DRAFT (SEPTEMBER 21)

Formulas for use on STATS 2141 tests and exams

Chapter 2

$$P_r^n = n(n-1)(n-2)...(n-r+1) = \frac{n!}{(n-r)!}$$

$$\frac{n!}{n_1!n_2! \ \dots \ n_r!}$$

Combinations

The number of combinations, subsets of r elements that can be selected from a set of n elements, is denoted as $\binom{n}{r}$ or C_r^n and

$$C_r^n = \binom{n}{r} = \frac{n!}{r!(n-r)!} \quad (2.4)$$

Multiplication Rule

$$P(A \cap B) = P(B \mid A)P(A) = P(A \mid B)P(B) \quad (2.10)$$

Total Probability Rule (Multiple Events)

Assume $E_1, E_2, \dots, E_k$ are k mutually exclusive and exhaustive sets. Then

$$\begin{aligned} P(B) &= P(B \cap E_1) + P(B \cap E_2) + \cdots + P(B \cap E_k) \\ &= P(B \mid E_1)P(E_1) + P(B \mid E_2)P(E_2) + \cdots + P(B \mid E_k)P(E_k) \end{aligned} \quad (2.12)$$

Bayes' Theorem

If $E_1, E_2, \dots, E_k$ are k mutually exclusive and exhaustive events and B is any event,

$$P(E_1 \mid B) = \frac{P(B \mid E_1)P(E_1)}{P(B \mid E_1)P(E_1) + P(B \mid E_2)P(E_2) + \cdots + P(B \mid E_k)P(E_k)} \quad (2.16)$$

for $P(B) > 0$

Chapter 3

$$\mu = E(X) = \sum_x x f(x)$$

$$\sigma^2 = V(X) = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x) = \sum_x x^2 f(x) - \mu^2$$

Discrete Uniform Distribution

A random variable X has a **discrete uniform distribution** if each of the n values in its range, $x_1, x_2, \ldots, x_n$, has equal probability. Then

$$f(x_i) = \frac{1}{n} \tag{3.5}$$

Mean and Variance

Suppose that X is a discrete uniform random variable on the consecutive integers a, $a+1$, $a+2, \ldots, b$, for $a \le b$. The mean of X is

$$\mu = E(X) = \frac{b+a}{2}$$

The variance of X is

$$\sigma^2 = \frac{(b-a+1)^2 - 1}{12} \tag{3.6}$$

Geometric Distribution

In a series of Bernoulli trials (independent trials with constant probability p of a success), the random variable X that equals the number of trials until the first success is a **geometric random variable** with parameter $0 < p < 1$ and

$$f(x) = (1-p)^{x-1} p \qquad x = 1, 2, \ldots \tag{3.9}$$

Mean and Variance

If X is a geometric random variable with parameter p,

$$\mu = E(X) = 1/p \qquad \text{and} \qquad \sigma^2 = V(X) = (1-p)/p^2 \tag{3.10}$$

Negative Binomial Distribution

In a series of Bernoulli trials (independent trials with constant probability p of a success), the random variable X that equals the number of trials until r successes occur is a **negative binomial random variable** with parameters $0 < p < 1$ and $r = 1, 2, 3, \ldots$, and

$$f(x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r \qquad x = r, r+1, r+2, \ldots \tag{3.11}$$

Mean and Variance

If X is a negative binomial random variable with parameters p and r,

$$\mu = E(X) = r/p \quad \text{and} \quad \sigma^2 = V(X) = r(1-p)/p^2 \tag{3.12}$$

Hypergeometric Distribution

A set of N objects contains

K objects classified as successes

$N - K$ objects classified as failures

A sample of size n objects is selected randomly (without replacement) from the N objects where $K \leq N$ and $n \leq N$.

The random variable X that equals the number of successes in the sample is a **hypergeometric random variable** and

$$f(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} \qquad x = \max\{0, n+K-N\} \text{ to } \min\{K, n\} \tag{3.13}$$

Mean and Variance

If X is a hypergeometric random variable with parameters N, K, and n, then

$$\mu = E(X) = np \qquad \text{and} \qquad \sigma^2 = V(X) = np(1-p)\left(\frac{N-n}{N-1}\right) \tag{3.14}$$

where $p = K/N$.

Poisson Distribution

The random variable X that equals the number of events in a Poisson process is a **Poisson random variable** with parameter $0 < \lambda$, and

$$f(x) = \frac{e^{-\lambda T}(\lambda T)^x}{x!} \quad x = 0, 1, 2, \ldots$$

Mean and Variance

If X is a Poisson random variable over an interval of length T with parameter λ, then

$$\mu = E(X) = \lambda T \quad \text{and} \quad \sigma^2 = V(X) = \lambda T \tag{3.16}$$

Chapter 4

Probability Density Function

For a continuous random variable X, a **probability density function** is a function such that

(1) $f(x) \geq 0$

(2) $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1$

(3) $\displaystyle P(a \leq X \leq b) = \int_a^b f(x)\,dx = \text{area under } f(x) \text{ from } a \text{ to } b \text{ for any } a \text{ and } b$ (4.1)

Cumulative Distribution Function

The **cumulative distribution function** of a continuous random variable X is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\,du \tag{4.3}$$

for $-\infty < x < \infty$.

Probability Density Function from the Cumulative Distribution Function

Given $F(x)$,

$$f(x) = \frac{dF(x)}{dx}$$

as long as the derivative exists.

Mean and Variance

Suppose that X is a continuous random variable with probability density function $f(x)$. The **mean** or **expected value** of X, denoted as μ or $E(X)$, is

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx \tag{4.4}$$

The **variance** of X, denoted as $V(X)$ or σ^2, is

$$\sigma^2 = V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx = \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2$$

The **standard deviation** of X is $\sigma = \sqrt{\sigma^2}$.

Continuous Uniform Distribution

A continuous random variable X with probability density function

$$f(x) = 1/(b-a), \qquad a \le x \le b \tag{4.6}$$

is a **continuous uniform random variable**.

Mean and Variance

If X is a continuous uniform random variable over $a \le x \le b$,

$$\mu = E(X) = \frac{a+b}{2} \quad \text{and} \quad \sigma^2 = V(X) = \frac{(b-a)^2}{12} \tag{4.7}$$

Normal Distribution

A random variable X with probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty \tag{4.8}$$

is a **normal random variable** with parameters μ where $-\infty < \mu < \infty$ and $\sigma > 0$. Also,

$$E(X) = \mu \quad \text{and} \quad V(X) = \sigma^2 \tag{4.9}$$

and the notation $N(\mu, \sigma^2)$ is used to denote the distribution.

Normal Approximation to the Binomial Distribution

If X is a binomial random variable with parameters n and p,

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \tag{4.12}$$

is approximately a standard normal random variable. To approximate a binomial probability with a normal distribution, a **continuity correction** is applied as follows:

$$P(X \le x) = P(X \le x + 0.5) \approx P\left(Z \le \frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right)$$

and

$$P(x \le X) = P(x - 0.5 \le X) \approx P\left(\frac{x - 0.5 - np}{\sqrt{np(1-p)}} \le Z\right)$$

The approximation is good for $np > 5$ and $n(1-p) > 5$.

Normal Approximation to the Poisson Distribution

If X is a Poisson random variable with $E(X) = \lambda$ and $V(X) = \lambda$,

$$Z = \frac{X - \lambda}{\sqrt{\lambda}} \tag{4.13}$$

is approximately a standard normal random variable. The same continuity correction used for the binomial distribution can also be applied. The approximation is good for

$$\lambda > 5$$

Exponential Distribution

The random variable X that equals the distance between successive events from a Poisson process with mean number of events $\lambda > 0$ per unit interval is an **exponential random variable** with parameter λ. The probability density function of X is

$$f(x) = \lambda e^{-\lambda x} \quad \text{for} \quad 0 \le x < \infty \tag{4.14}$$

Mean and Variance

If the random variable X has an exponential distribution with parameter λ,

$$\mu = E(X) = \frac{1}{\lambda} \quad \text{and} \quad \sigma^2 = V(X) = \frac{1}{\lambda^2} \tag{4.15}$$