Formula for AS 388: SOA Exam P preparation
Chapter 0 ACTEX manual Broverman (2010) Review of Algebra and Calculus

SET THEORY: *Set Relationships:* 1) $A \cup B = B \cup A$; $A \cap B = B \cap A$; $A \cup A = A$; $A \cap A = A$;
2) $A \cup \varnothing = A$; $A \cap \varnothing = \varnothing$; $A - \varnothing = A$;
3) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
4) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
5) If $A \subset B$, then $A \cup B = B$ and $A \cap B = A$
6) For any sets A and B, $A \cap B \subset A \subset A \cup B$ and $A \cap B \subset B \subset A \cup B$
7) $(A \cup B)' = A' \cap B'$ and $(A \cap B)' = A' \cup B'$
8) For any set A, $\varnothing \subset A$ (the empty set is a subset of any other set A)

PROPERTIES OF FUNCTIONS: Exponential and logarithmic functions:
Some important properties of exponential and logarithmic functions are:

$b^0 = 1$	$\log_b(1) = 0$
domain $(f) = \mathbb{R} =$ range (f^{-1})	range $(f) = (0, +\infty) =$ domain (f^{-1})
$b^{\log_b(y)} = y$ for $y > 0$	$\log_b(b^x) = x$ for all x
$b^x = e^{x \cdot \ln\ b}$	$\log_b(y) = \dfrac{\ln y}{\ln b}$
$(b^x)^y = b^{xy}$	$\log_b(y^k) = k \cdot \log_b(y)$
$b^x b^y = b^{x+y}$	$\log_b(xy) = \log_b(x) + \log_b(y)$
$b^x / b^y = b^{x-y}$	$\log_b(x/y) = \log_b(x) - \log_b(y)$

Inverse functions, $f^{-1}()$: $f^{-1}()$ exists.only if $f()$ is a one-to-one function. $y = f(x)$. $x = f^{-1}(y)$.

Quadratic functions and equations: $p(x) = ax^2 + bx + c$.
Roots of $p(x) = ax^2 + bx + c = 0$ are $r_1 = \dfrac{-b - \sqrt{b^2 - 4ac}}{2a}$ and $r_2 = \dfrac{-b + \sqrt{b^2 - 4ac}}{2a}$.
Cases: (1) If $b^2 - 4ac = 0$, then $r_1 = r_2$. (2) If $b^2 - 4ac > 0$, then $r_1 \neq r_2$ and $r_1, r_2 \in \mathbb{R}$.
(3) If $b^2 - 4ac < 0$, then $r_1 \neq r_2$ and $r_1, r_2 \notin \mathbb{R}$.

DIFFERENTATION: Some important differentiation rules.

Rules of differentiation	$f(x)$	$f\prime(x) = \dfrac{df(x)}{dx}$
	c (a constant)	0
Power rule	$cx^n \quad (n \in \mathbb{R})$	cnx^{n-1}
	$g(x) + h(x)$	$g'(x) + h'(x)$
Product rule	$g(x) \cdot h(x)$	$g'(x) \cdot h(x) + g(x) \cdot h'(x)$
	$u(x)v(x)w(x)$	$u'vw + uv'w + uvw'$
Quotient rule	$\dfrac{g(x)}{h(x)}$	$\dfrac{h(x)g'(x) - g(x)h'(x)}{[h(x)]^2}$
Chain rule	$g(h(x))$	$g'(h(x)) \cdot h'(x)$
	$e^{g(x)}$	$g'(x) \cdot e^{g(x)}$
	$\ln(g(x))$	$\dfrac{g'(x)}{g(x)}$
	$a^x \quad (a > 0)$	$a^x \ln a$
	e^x	e^x
	$\ln x$	$\dfrac{1}{x}$
	$\log_b(x)$	$\dfrac{1}{x \ln b}$
	$\sin x$	$\cos x$
	$\cos x$	$-\sin x$

L'Hospital's rule for calculating limits: when $\lim_{x \to c} \dfrac{f(x)}{g(x)}$ is indeterminate

1. **IF** $= \begin{cases} \text{(i)} & \lim_{x\to c} f(x) = \lim_{x\to c} g(x) = 0 \textbf{ and} \\ \text{(ii)} & f'(c) \text{ exists, } \textbf{and} \\ \text{(iii)} & g'(c) \text{ exists, and is} \neq 0 \end{cases}$ **THEN** $\lim_{x\to c} \frac{f(x)}{g(x)} = \frac{f'(c)}{g'(c)}$

2. **IF** $= \begin{cases} \text{(i)} & \lim_{x\to c} f(x) = \lim_{x\to c} g(x) = 0 \textbf{ and} \\ \text{(ii)} & f \text{ and } g \text{ are differentiable near } c, \textbf{ and} \\ \text{(iii)} & \lim_{x\to c} \frac{f'(x)}{g'(x)} \text{ exists,} \end{cases}$ **THEN** $\lim_{x\to c} \frac{f(x)}{g(x)} = \frac{f'(x)}{g'(x)}$

INTEGRATION: Antiderivatives of some frequently used functions:

$f(x)$	$\int f(x)dx$ (antiderivatives)
$g(x)+h(x)$	$\int g(x)dx + \int h(x)dx + c$
$x^n \quad (n \neq -1)$	$\frac{x^{n+1}}{n+1} + c$
$\frac{1}{x}$	$\ln x + c$
e^x	$e^x + c$
$a^x \quad (a > 0)$	$\frac{a^x}{\ln a} + c$
xe^{ax}	$\frac{xe^{ax}}{a} - \frac{e^{ax}}{a^2} + c$
$\sin x$	$-\cos x + c$
$\cos x$	$\sin x + c$

The method of Substitution: *Target:* Rewrite the integral in a *standard form* which the antiderivative is well known.
In general, to find $\int f(x)\ dx$, (1) we may make the substitution $u = g(x)$ for an "appropriate" function $g(x)$. (2) Then we define the "differential" du to be $du = g'(x)dx$, and (3) we try to rewrite $\int f(x)\ dx$ as an integral with respect to the variable u.

Integration by parts:
$\int f(x)\cdot g'(x)\ dx = f(x)\cdot g(x) - \int f'(x)\cdot g(x)\ dx$

$\int xe^{bx}\ dx = \frac{bxe^{bx} - e^{bx}}{b^2} = -\frac{xe^{bx}}{b} - \frac{e^{bx}}{b^2}.$

NOTE: for integer $n \geq 0$ and $b > 0$ $\quad \int_0^\infty x^n e^{-bx}\ dx = \frac{n!}{b^{n+1}}.$

Geometric progression: $a, ar, ar^2, ar^3, \ldots$

$$a + ar + ar^2 + ar^3 + \ldots + ar^{n-1} = a(1 + r + r^2 + r^3 + \ldots + r^{n-1}) = a\frac{r^n - 1}{r - 1} = a\frac{1 - r^n}{1 - r}.$$

if $|r| < 1$, then $a + ar + ar^2 + ar^3 + \ldots = \frac{a}{1-r}.$

Arithmetic progression: $a, a+d, a+2d, a+3d, \ldots$

$$a + (a+d) + (a+2d) + (a+3d) + \ldots + (a+(n-1)d) = na + d\cdot\frac{n(n-1)}{2}.$$

Special case: $1 + 2 + 3 + 4 + \ldots + n = \frac{n(n+1)}{2}.$

Chapter 10 Broverman (2010) Risk Management Concepts

Pure premium=expected claim=$E[X]$.

Unitized risk (Coefficient of variation): $CV = \frac{\sqrt{Var(X)}}{E[X]} = \frac{\sigma}{\mu}.$

Ordinary Deductible Insurance with deductible amount d: $Y = \begin{cases} 0 & \text{if } X \leq d \\ X - d & \text{if } X > d \end{cases} = Max(X-d, 0) = (X-d)_+$

For $X > 0$ only, $E[(X-d)_+] = \int_d^\infty (x-d)f_X(x)dx = \int_d^\infty (1 - F_X(x))dx.$

Franchise deductible insurance: $Y = \begin{cases} 0 & \text{if } X \leq d \\ X & \text{if } X > d \end{cases} = Max(X, 0) = (X)_+$

Disappearing deductible insurance: $Y = \begin{cases} 0 & \text{if } X \le d \\ d'\dfrac{X-d}{d'-d} & d' < X \le d \\ X & \text{if } X > d \end{cases}$

Insurance with Policy limit u: $Y = \begin{cases} X & \text{if } X \le u \\ u & \text{if } X > u \end{cases} = Min(X,u)$

For $X > 0$ only, $E[Min(X,u)] = \int_0^u x f_X(x)dx + u \cdot [1 - F_X(u)] = \int_0^u (1 - F_X(x))dx$.

Deductible Insurance with policy limit: $Y = \begin{cases} 0 & \text{if } X \le d \\ X - d & d < X \le u \\ u - d & \text{if } X > u \end{cases}$

For $X > 0$ only, $E[(X-d)_+] = \int_d^\infty (x-d) f_X(x)dx = \int_d^\infty (1 - F_X(x))$

Proportional Insurance: $Y = \alpha X$. $E[\alpha X] = \alpha E[X]$

Models.for random loss X:

1) $X = \begin{cases} 0 & \text{with probability } 1-q \\ k & \text{with probability } q. \end{cases}$ where k=constant.

2) $X = \begin{cases} 0 & \text{with probability } 1-q \\ B & \text{with probability } q. \end{cases}$ and B a conditional distribution of X given loss occurred.

if $E[B]$ and $Var(B)$ given, $E[B^2] = Var(B) + E[B]^2$ is needed for $Var(X)$

$E[X] = E[E[B|I(x)] = (1-q)E[B|I(x)=0] + qE[B|I(x)=1] = qE[B]$.

$E[X^2] = E[E[B^2|I(x)] = (1-q)E[B^2|I(x)=0] + qE[B^2|I(x)=1] = qE[B^2]$.

$Var[X] = qE[B^2] - (qE[B])^2 = qVar(B) + q(1-q)(E[B])^2$.

Individual Risk Model of aggregate claims in insurance portfolio:

$S = \sum_{i=1}^{n} X_i$. $E[S] = \sum_{i=1}^{n} E[X_i]$ and $Var[S] = \sum_{i=1}^{n} Var[X_i]$.

Normal *approximation* to aggregate claims:

$P[S \le k] = 0.95 \to P\left(\dfrac{S - E[S]}{\sqrt{Var[S]}} \le \dfrac{k - E[S]}{\sqrt{Var[S]}}\right) = 0.95 \to k = E[S] + 1.645\sqrt{Var[S]}$.

Formula for STAT 301: Exam preparation
Chapter 1 Ross(2010) Combinatorial Analysis

1.2 The basic principle of counting

Let m = any one of possible outcomes of experiment 1. And let n be any one of possible outcomes of experiment 2. Then total possible outcomes of the two experiments together is mn.

The generalized basic principle of counting

Let n_1 = possible outcomes of experiment 1, n_2 = possible outcomes of experiment 2, n_3 = possible outcomes of the third experiment 3, and so on so forth.then there is a total of $n_1 \cdot n_2 .... n_r$ possible outcomes of r experiments.

1.3 Permutations (*keywords:* arrangements, order)

all distinct objects: $n(n-1)(n-2).....3 \cdot 2 \cdot 1 = n!$

n_i like objects, $i = 1, 2, \ldots, r$ (See Example 3d): $\dfrac{n!}{n_1! n_2! ..... n_r!}$.

1.4 Combinations (*keywords:* select, choose)**:** for $r \le n$, $\binom{n}{r} = \frac{n!}{(n-r)!r!}$

meaning: $\binom{n}{r}$ represents the number of possible combinations of n objects taken r at a time

The binomial theorem: $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$

1.5 Multinomial Coefficients: If $n_1 + n_2 + ... + n_r = n$ then $\dbinom{n}{n_1, n_2, ...n_r} = \dfrac{n!}{n_1! n_2! ..... n_r!}$.

meaning: $\dbinom{n}{n_1, n_2, ...n_r}$ represents the number of possible divisions of n distinct objects into r distinct groups of respective sizes $n_1, n_2, ...n_r$.

The multinomial theorem: $(x_1 + x_2 + ... + x_r)^n = \sum_{(n_1, ..., n_r): n_1 + ... + n_r = n} \dbinom{n}{n_1, n_2, ...n_r} x_1^{n_1} x_2^{n_2} ..... x_r^{n_r}$ that is, the sum is over all nonnegative integer-valued vectors $(n_1, n_2, ...n_r)$ such that $n_1 + n_2 + ... + n_r = n$

1.6 On the distribution of balls in urns

Proposition 6.1: There are $\binom{n-1}{r-1}$ distinct positive integer-valued vectors $(x_1, x_2, ..., x_r)$ satisfying

$$x_1 + x_2 + ... + x_r = n \qquad x_i > 0, i = 1, ...r$$

Proposition 6.2: There are $\binom{n+r-1}{r-1}$ distinct nonnegative integer-valued vectors $(x_1, x_2, ..., x_r)$ satisfying $x_1 + x_2 + ... + x_r = n$.

There are $n! = n(n-1)...3 \cdot 2 \cdot 1$ possible linear orderings of n items. $0! = 1$.

Let $\binom{n}{i} = \dfrac{n!}{(n-i)!i!}$ $\qquad (x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}$

For nonnegative integers $(n_1, ..., n_r)$ summing to n, $\binom{n}{n_1, n_2, ...n_r} = \dfrac{n!}{n_1!n_2!.....n_r!}$ is the number of ways of dividing up n items into r distinct nonoverlapping subgroups of sizes $n_1, n_2, ...n_r$

——

Chapter 2 Ross (2010) Axioms of Probability

Axioms of Probability: ***Axiom 1**. $0 \le P(E) \le 1$ ***Axiom 2.** $P(S) = 1$

***Axiom 3**. For any sequence of mutually exclusive events $E_1, E_2, ...$(i.e., events for which $E_1E_2 = \emptyset$ when $i \ne j$),

$$P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$$

We refer to *P(E)* as the probabilty of the event E

2.4 *Some Simple Propositions*

Proposition 4.1. $P(E^c) = 1 - P(E)$

Proposition 4.2. If $E \subset F$, then $P(E) \le P(F)$

Proposition 4.3. $P(E \cup F) = P(E) + P(F) - P(EF)$

Proposition 4.4.

$$P(E_1 \cup E_2 \cup .... \cup E_n) = \sum_{i=1}^{n} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1}E_{i_2}) + ... + (-1)^{r+1} \sum_{i_1 < i_2 < ... < i_r} P(E_{i_1}E_{i_2}....E_{i_r}) + .. + (-1)^{n+1} P(E_1E_2...E_n).$$

The summation $\sum_{i_1 < i_2 < ... < i_r} P(E_{i_1}E_{i_2}....E_{i_r})$ is taken over all of the $\binom{n}{r}$ possible subsets of size r of the set {1, 2 ....., n }

2.6 *Probability as a continuous set function*

Proposition 6.1. If $\{E_n, n \ge 1\}$ is an increasing (or decreasing) sequence of events, then $\lim_{n\to\infty} P(E_n) = P(\lim_{n\to\infty} E_n)$.

——

Chapter 3 Conditional Probability and Independence

Conditional Probabilities: If $P(F) > 0$, then $P(E|F) = \dfrac{P(EF)}{P(F)}$

The multiplication rule: $P(E_1E_2E_3....E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2).....P(E_n|E_1......E_{n-1})$

The odds ratio of an event A is defined by $\dfrac{P(A)}{P(A^c)} = \dfrac{P(A)}{1-P(A)}$

Proposition 3.1 (*Bayes' Formula*): $P(F_j|E) = \dfrac{P(EF_j)}{P(E)} = \dfrac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)}$

Independent Event

Two events E and F are said to be *independent* if $P(EF) = P(E)P(F)$.

Two events E and F that are *not independent* are said to be *dependent*.

Proposition 4.1. If E and F are independent, then so are E and F^c.

Independent Event (3 events): The three events *E, F* and G are said to be *independent* if

$$\begin{aligned} P(EFG) &= P(E)P(F)P(G) \\ P(EF) &= P(E)P(F) \\ P(EG) &= P(E)P(G) \\ P(FG) &= P(F)P(G) \end{aligned}$$

Proposition 5.1. $P(.|F)$ *is a Probability*

(a) $0 \le P(E|F) \le 1$ (b) $P(S|F) = 1$

(c) If $E_i, i = 1, 2, \ldots$ are mutually exclusive events, then $P(\bigcup_1^{\infty} E_i|F) = \sum_1^{\infty} P(E_i|F)$

——————————————————————————————————————

Chapter 4 Random Variables (rv)

4.5 Expectation of a function of a random variable

Proposition 5.1. If X is a discrete rv that takes on one of the values x_i, $i \ge 1$ with respective probabilities $p(x_i)$, then for any real-valued function g, $E[g(X)] = \sum_i g(x_i)p(x_i)$

Corollary 5.1. If a and b are constants, then $E[aX + b] = aE[X] + b$

4.6 Variance: If X is a rv with mean μ , then variance of X is defined by $Var(X) = E[(X - \mu)^2]$.

4.7 The Bernoulli and Binomial Random Variables

If X is a *bernoulli* rvs with parameters p, then $E[X] = p$ $Var(X) = p(1 - p)$.

If X is a *binomial* rvs with parameters n and p, then $E[X] = np$ $Var(X) = np(1 - p)$.

Proposition 7.1. If X is a *binomial* rv with parameters (n, p), where $0 < p < 1$, then as k goes from 0 to n, $P\{X = k\}$ first increases monotonically and then decreases monotonically, reaching its largest value when integer$(k) \le (n + 1)p$.

4.8 The Poisson Random Variable

If X is a *Poisson* rv, then $E[X] = Var(X) =$its parameter λ.

——————————————————————————————————————

Chapter 5 Continuous Random Variables

A rv X is called *continuous* if there is a nonnegative function f, called the *probability density function* of X, such that for any set B, $P\{X \in B\} = \int_B f(x)dx$.

If X is continuous, then its distribution function $F(x)$ will be differentiable and $\frac{d}{dx}F(x) = f(x)$.

The expected value of a continuous rv X is defined by $E[X] = \int_{-\infty}^{\infty} xf(x)dx$.

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

The variance of X is defined by $Var(X) = E[(X - E[X])^2]$.

A random variable X is said to be *uniform* over the interval $(a,\ b)$ if its probability density function is given by

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

Its expected value and variance are $E[X] = \dfrac{a + b}{2}$ $Var(X) = \dfrac{(b - a)^2}{12}$.

A random variable X is said to be *normal* with parameters μ and σ^2 if its probability density function is given by $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$ $-\infty < x < \infty$.

Its expected value and variance are $\mu = E[X]$ $\sigma^2 = Var(X)$.

If X is normal with mean μ and variance σ^2, then $Z = \dfrac{X - \mu}{\sigma}$ is normal with mean 0 and variance 1. Z is said to be a *standard* normal rv.

The probability distribution function of a binomial random variable with parameters n and p can, when n is large, be approximated by that of a normal random variable having mean np and variance $np(1 - p)$.

A rv whose probability density function is of the form $f(x) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & \text{otherwise} \end{cases}$ is said to be an *exponential* random variable with parameter λ.

Its expected value and variance are $E[X] = \dfrac{1}{\lambda}$ $Var(X) = \dfrac{1}{\lambda^2}$.

Exponential rvs possessed a key property, *memoryless property,* in the sense that for positive s and t, $P\{X > s + t|X > t\} = P\{X > s\}$.

If X represents the life of an item, then the **memoryless property** states that for any t, the remaining life of a t-year-old item has the same probability distribution as the life of a new item. Thus the age of an item do not need to be remembered to know its distribution of remaining life.

Let X be a nonnegative continuous rv with distribution function $F(x)$ and density function $f(x)$. The *hazard rate* (or *failure rate*) function is $\lambda(t) = \dfrac{f(t)}{1 - F(t)}$ $t \ge 0$.

If we interpret X as being the life of an item, then for small values of dt, $\lambda(t)dt$ is approximately the probability that a t unit old item will fail within an additional time dt.

If F is the exponential distribution with parameter λ then $\lambda(t) = \lambda \qquad t \geq 0$.
In addition, the exponential is the unique distribution having a constant failure rate.

A rv is said to have a *gamma* distribution with parameters α and λ if its probability density function is equal to $f(x) = \dfrac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} \qquad x \geq 0$.

The quantity $\Gamma(\alpha)$ is called gamma function and is defined by $\Gamma(\alpha) = \int_0^\infty e^{-x}x^{\alpha-1}dx$.

The expected value and variance of a gamma random variable are $\quad E[X] = \dfrac{t}{\lambda} \qquad Var(X) = \dfrac{t}{\lambda^2}$.

A rv is said to have a *beta* distribution with parameters (a, b) if its probability df is equal to $f(x) = \dfrac{1}{B(a,b)}x^{a-1}(1-x)^{b-1} \qquad 0 \leq x \leq 1$.

The constant $B(a, b)$ is given by $B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx$.

The mean and variance of such a rv are $\quad E[X] = \dfrac{a}{a+b} \qquad Var(X) = \dfrac{ab}{(a+b)^2(a+b+1)}$.

−−

Chapter 6 Jointly Distributed Random Variables

The *joint cummulative probability distribution function* of the pair of rv X and Y is defined by

$$f(x,y) = P\{X \leq x, Y \leq y\} \qquad\qquad -\infty < x, y < \infty.$$

All probabilities regarding the pair can be obtained from F. To obtain the individual probability distribution functions of X and Y, use $F_X(x) = \lim_{y\to\infty} F(x,y) \qquad F_Y(y) = \lim_{x\to\infty} F(x,y)$.

If X and Y are both *discrete* random variables, then their *joint probability mass function* is defined by

$$p(i,j) = P\{X = i, Y = j\}$$

The individual mass functions are $P\{X = i\} = \sum_j p(i,j) \qquad\qquad P\{Y = j\} = \sum_i p(i,j)$.

The rvs X and Y are said to be *jointly continuous* if there is a function $f(x, y)$, called the *joint probability density function*, such that for any two-dimensional set C, $P\{(X,Y) \in C\} = \int\int_C f(x,y)\ dx\ dy$

It follows that $P\{x < X < x + dx, y < Y < y + dy\} \approx f(x,y)\ dx\ dy$.
If X and Y are jointly continuous, then they are individually *continuous* with density functions

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy \qquad\qquad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

The rvs X and Y are *independent* if for all sets A and B, $P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\}$.
If the joint distribution function (or the joint probability mass function in the discrete case, or the joint density function in the continuous case) factors into a part depending only on x and a part depending onlt on y, then X and Y are **independent.**
In general the rvs $X_1, ..., X_n$ are independent if for all sets of real numbers $A_1, ..., A_n$
$P\{X_1 \in A_1, ...X_n \in A_n\} = P\{X_1 \in A_1\} \cdot P\{X_2 \in A_2\}....P\{X_n \in A_n\}$.

Sums of independent random variables

If X and Y are *independent continuous* rvs, then the distribution function of *their sum* can be obtained from the convolution identity $F_{X+Y}(a) = \int_{-\infty}^{\infty} F_X(a-y)\ f_Y(y)\ dy$.

If $X_i, i = 1, ...n$, are *independent normal* rvs with respective parameters μ_i and $\sigma_i^2, i = 1, ...n$, then $\sum_{i=1}^n X_i$ is *normal* with parameters $\sum_{i=1}^n \mu_i$ and $\sum_{i=1}^n \sigma_i^2$.

If $X_i, i = 1, ...n$, are *independent Poisson* rvs with respective parameters $\lambda_i, i = 1, ...n$, then $\sum_{i=1}^n X_i$ is *Poisson* with parameter $\sum_{i=1}^n \lambda_i$.

If X and Y are *discrete* rv, then the *conditional probability mass function* of X given that $Y = y$ is defined by

$P\{X = x|Y = y\} = \dfrac{p(x,y)}{p_Y(y)}$ where p is their joint probability mass function.
Also, if X and Y are jointly *continuous* with joint density function f, then the *conditional probability density function* of X given that $Y = y$ is given by $f_{X|Y}(x|y) = \dfrac{f(x,y)}{f_Y(y)}$.
The ordered values $X_{(1)} \le X_{(2)} \le ...... \le X_{(n)}$ of a set of *independent and identically distributed* (iid) random variable are called the *order statistics* of that set.

If the rvs are *continuous* with density function f, then the joint density function of the *order statistics* is $f(x_1, ..., x_n) = n!f(x_1)....f(x_n) \qquad x_{(1)} \le x_{(2)} \le ...... \le x_{(n)}$.
The rvs $X_1, ..., X_n$ are *exchangeable* if the joint distribution of $X_{i_1}, ..., X_{i_n}$ is the same for every permutation $i_1, ...i_n$ of $1, ..., n$.

Chapter 7 Properties of Expectation

If X and Y have a joint probability *mass* function $p(x,y)$, then $E[g(X,Y)] = \sum_y \sum_x g(x,y)p(x,y)$
whereas if they have a joint *density* function $f(x,y)$, then $E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)p(x,y)dx\ dy$
$E[X+Y] = E[X] + E[Y]$
Generalization: $E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$
The *covariance* between rvs X and Y is given by

$$Cov(X,Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$

$Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j)$.
When $n = m$, and $Y_i = X_i$, $i = 1, ..., n$, the $Var\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} Var(X_i) + 2\sum\sum_{i<j} Cov(X_i, X_j)$.
The correlation between X and Y, denoted by $\rho(X,Y)$, is defined by $\rho(X,Y) = \dfrac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$.
If X and Y are jointly *discrete* rvs, then the conditional expected value of X given that $Y = y$ is defined by $E[X|Y = y] = \sum_x xP(X = x|Y = y)$
If they are jointly *continuous* rvs, then $E[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)$
where $f_{X|Y}(x|y) = \dfrac{f(x,y)}{f_Y(y)}$ is the conditional probability density of X given that $Y = y$.

Conditional expectations, which are similar to ordinary expectations except that all probabilities are now computed conditional on the event that $Y = y$, satisfy all the properties of ordinary expectations.

Let $E[X|Y]$ denote that function of Y whose value at $Y = y$ is $E[X|Y = y]$. A very useful identity is that $E[X] = E[E[X|Y]]$
In the case of *discrete* rvs, this reduces to the identity $E[X] = \sum_y E[X|Y = y]P\{Y = y\}$
and, in the *continuous* case, to $E[X] = \int_{-\infty}^{\infty} E[X|Y = y]\ f_Y(y)dy$
The preceding equations can be often applied to obtain $E[X]$ by first "conditioning" on the value of some other random variable Y. In addition, since for any event A, $P(A) = E[I_A]$, where I_A is 1 if A occurs and 0 otherwise, we can also use them to compute probabilities.

The conditional variance of X is given that $Y = y$ is defined by $Var(X|Y = y) = E[(X - E|Y = y])^2|Y = y$.
Let $Var(X|Y)$ be that function of Y whose value at $Y = y$ is $Var(X|Y = y)$. The following is known as the *conditional variance formula*: $Var(X) = E[Var(X|Y)] + Var(E[X|Y])$.

Suppose that the rv X is to be observed and, based on its value, one must then predict the value of the random variable Y. In such a situation, it turns out that, among all predictors, $E[Y|X]$ has the *smallest expectation* of the square of the difference between it and Y.

The *moment generating function* (mgf)of the rv X is defined by $M(t) = E[e^{tX}]$.
The moments of X can be ontained by successively differentiating $M(t)$ and then evaluating the resulting quantity at $t = 0$. Specifically, we have that $E[X^n] = \dfrac{d^n}{dt^n} M(t)|_{t=0} \qquad n = 1, 2....$

Two useful results concerning mgf

1) mgf uniquely determines the function of the random variable, and

2) X_i are iid and $S = \sum_{i=1}^{n} X_i$. The mgf of the sum of independent rvs is equal to the product of their mgfs. $E[e^{tS}] = E[e^{tX_1}] \cdot E[e^{tX_2}] \cdot E[e^{tX_3}] \cdots E[e^{tX_n}] = \prod_{i=1}^{n} E[e^{tX_i}]$.

These results lead to simple proofs that the sum of independent normal (Poisson) [gamma] rvs remains a normal (Poisson) [gamma] rv.

If $X_1, ..., X_m$ are all linear combinations of a finite set of independent standard normal rvs, the they are said to have a *multivariate normal distribution*. Their joint distribution is specified by the values of $E[X_i]$, $Cov(X_i, X_j), i, j = 1, ..., m$.

If $X_1, ..., X_n$ are iid normal rvs, then their *sample mean* $\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$ and their *sample variance* $S^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1}$ are independent.

The sample mean $\bar{X}$ is a normal rv with mean μ and variance σ^2/n; The rv $(n-1)S^2/\sigma^2$ is a chi-squared rv with $n-1$ degrees of freedom.

Discrete Dsn	pmf $p(x)$	mgf $M(t)$	Mean $E[X]$	Variance $Var(X)$
Binomial(n, p) $0 \le p \le 1$	$\binom{n}{x} p^x (1-p)^{n-x}$ $x = 0, 1, \ldots, n$	$(pe^t + 1 - p)^n$	np	$np(1-p)$
Poisson(λ) $\lambda > 0$	$e^{-\lambda}\lambda^x/x!$ $x = 0, 1, \ldots$	$exp[\lambda(e^{-t} - 1)]$	λ	λ
Geometric(p)	$p^x(1-p)^{x-1}$ $x = 0, 1, \ldots$	$\frac{pe^t}{1-(1-p)e^t}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative Binomial(r, p)	$\binom{n-1}{r-1} p^r (1-p)^{n-r}$ $n = r, r+1, \ldots$	$\left[\frac{pe^t}{1-(1-p)e^t}\right]^r$	$\frac{r}{p}$	$r\frac{1-p}{p^2}$

Continuous Dsn	pdf $f(x)$	mgf $M(t)$	Mean $E[X]$	$Var(X)$
Uniform(a, b)	$f(x) = \begin{cases} (b-a)^{-1} & a < x < b \\ 0 & \text{otherwise} \end{cases}$	$\frac{e^{tb} - e^{ta}}{t(b-a)}$	$\frac{b+a}{2}$	$\frac{(b-a)^2}{12}$
Exponential(λ) $\lambda > 0$	$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & x < 0 \end{cases}$	$\frac{\lambda}{\lambda - t}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma(α, λ)	$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & x \ge 0 \\ 0 & x < 0 \end{cases}$	$\left(\frac{\lambda}{\lambda - t}\right)^\alpha$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
Normal(μ, σ^2) $-\infty < x < \infty$	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$	$exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$	μ	σ^2

−−

Chapter 8 Limit Theorems

Two useful probability bounds are provided by the *Markov* and *Chebyshev* inequalities.

The *Markov* inequality: for $X > 0$ and $a > 0$, $P\{X \ge a\} \le \frac{E[X]}{a}$

The *Chebyshev* inequality, (special case of Markov inequality): If X has mean μ and variance σ^2, then for every positive k, $P\{|X - \mu| \ge k\sigma\} \le \frac{1}{k^2}$.

The two most important theoretical results in probability are the *central limit theorem* and the s*trong law of large numbers*. Both are concerned with a sequence of iid rvs.

The *central limit theorem* says that if the rvs have a finite mean μ and a finite variance σ^2, then the distribution of the sum of the first n of then is, for large n, *approximately* a normal rv with mean $n\mu$ and variance $n\sigma^2$. That is, if X_i, $i = 1$ is the sequence, the the *central limit theorem* states that for every real number a,
$\lim_{n\to\infty} P\left\{\frac{X_1 + ... + X_n - n\mu}{\sigma\sqrt{n}} \le a\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2} dx$.
The s*trong law of large numbers* requires only that the rvs in the sequence have a finite mean μ. It states that with probability 1, the average of the first n of them will converge to μ as n goes to infinity.

The implies that if A is any specified event of an experiment for which independent replications are performed, then the limiting proportion of experiments whose outcomes are in A will, with probability 1, equal $P(A)$. Therefore, if we accept the interpretation that "with probability 1" means "with certainty", we obtain the theoretical justification for the long run relative frequency interpretation of probabilities.

____