Actuarial Science Exam 1/P

Ville A. Satopää

December 5, 2009

Contents

1 Review of Algebra and Calculus

1. The complement of the set B: This can be denoted B', $\bar{B}$, or $\sim B$.

2. For any two sets we have: $A = (A \cap B) \cup (A \cap B')$

3. The inverse of a function exists only if the function is one-to-one.

4. The roots of a quadratic equation $ax^2 + bx + c$ can be deduced from

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

   Recall that the equation has distinct roots if $b^2 - 4ac > 0$, distinct complex roots if $b^2 - 4ac < 0$, or equal real roots if $b^2 - 4ac = 0$.

5. Exponential functions are of the form $f(x) = b^x$, where $b > 0$, $b \neq 1$. The inverse of this function is denoted $\log_b(y)$. Recall that

$$\begin{aligned} b^{\log_b(y)} &= y \text{ for } y > 0 \\ b^x &= e^{x\log(b)} \end{aligned}$$

6. A function f is continuous at the point $x = c$ if $\lim_{s \to c} f(x) = f(c)$.

7. The algebraic definition of $f'(x_0)$ is

$$\frac{d^1 f}{dx^1}\Big|_{x=x_0} = f^{(1)}(x_0) = f'(x_0) \quad = \quad \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

8. Differentation: Product and quotient rule

$$\begin{aligned} g(x) \times h(x) &\rightarrow g'(x) \times h(x) + g(x) \times h'(x) \\ \frac{g(x)}{h(x)} &\rightarrow \frac{h(x)g'(x) - g(x)h'(x)}{[h(x)]^2} \end{aligned}$$

   and some other rules

$$\begin{aligned} a^x &\rightarrow a^x \log(a) \\ \log_b(x) &\rightarrow \frac{1}{x\log(b)} \\ \sin(x) &\rightarrow \cos(x) \\ \cos(x) &\rightarrow -\sin(x) \end{aligned}$$

9. L'Hopital's rule: if $\lim_{x \to c} f(x) = \lim_{x \to c} f(x) = 0$ or $\pm\infty$ and $\lim_{x \to c} f'(x)/g'(x)$ exists, then

$$\lim_{x \to c} \frac{f(x)}{g(x)} = \lim_{x \to c} \frac{f'(x)}{g'(x)}$$

10. Integration: Some rules

$$\begin{aligned} \frac{1}{x} &\rightarrow \log(x) + c \\ a^x &\rightarrow \frac{a^x}{\log(a)} + c \\ xe^{ax} &\rightarrow \frac{xe^{ax}}{a} - \frac{e^{ax}}{a^2} + c \end{aligned}$$

11. Recall that

$$\int_{-\infty}^{\infty} f(x)dx = \lim_{a\to\infty}\int_{-a}^{a} f(x)dx$$

This can be useful if the integral is not defined at some point a, or if f is discontinuous at $x = a$; then we can use

$$\int_{a}^{b} f(x)dx = \lim_{c\to a}\int_{c}^{b} f(x)dx$$

Similarly, if $f(x)$ has discontinuity at the point $x = c$ in the interior of $[a, b]$, then

$$\int_{a}^{b} f(x)dx = \int_{a}^{c} f(x)dx + \int_{c}^{b} f(x)dx$$

Let's do one example to clarify this a little bit:

$$\int_{0}^{1} x^{-1/2}dx = \lim_{c\to 0}\int_{c}^{1} x^{-1/2}dx = \lim_{c\to 0}\left[(1/2)x^{1/2}\Big|_{c}^{1}\right] = \lim_{c\to 0}[2 - 2\sqrt{c}] = 2$$

More examples on page. 18

12. Some other useful integration rules are:

    (i) for integer $n \geq 0$ and real number $c > 0$, we have $\int_0^\infty x^n e^{-cx}dx = \frac{n!}{c^{n+1}}$

13. Geometric progression: The sum of the first n terms is

$$a + ar + ar^2 + ... + ar^{n-1} = a[1 + r + r^2 + ... + r^{n-1}] = a \times \frac{r^n - 1}{r - 1} = a \times \frac{1 - r^n}{1 - r}$$

and

$$\sum_{k=0}^{\infty} ar^k = \frac{a}{1 - r}$$

14. Arithmetic progression: The sum of the first n terms is

$$a + (a + d) + (a + 2d) + ... + (a + nd) = na + d \times \frac{n(n - 1)}{2}$$

2 Basic Probability Concepts

1. Outcomes are exhaustive if they combine to be the entire probability space, or equivalently, if at least on of the outcomes must occur whenever the experiment is performed. In other words, if $A_1 \cup A_2 \cup ... \cup A_n = \Omega$, then $A_1, A_2, ..., A_n$ are referred to as exhaustive events.

2. Example 1-1 on page 37 summarizes many key definitions very well.

3. Some useful operations on events:

    (i) Let $A, B_1, B_2, ..., B_n$ be any events. Then

$$A \cap (B_1 \cup B_2 \cup ... \cup B_n) = (A \cap B_1) \cup (A \cap B_2) \cup ... \cup (A \cap B_n)$$

and

$$A \cup (B_1 \cap B_2 \cap ... \cap B_n) = (A \cup B_1) \cap (A \cup B_2) \cap ... \cap (A \cup B_n)$$

(ii) If $B_1 \cup B_2 = \Omega$, then $A = (A \cap B_1) \cup (A \cap B_2)$. This of course applies to a partition of any size. To put in this in the context of probabilities we have $P(A) = P(A \cap B_1) + P(A \cap B_2)$

3. An event A consists of a subset of sample points in the probability space. In the case of a discrete probability space, the probability of A is $P(A) = \sum_{a_1 \in A} P(a_i)$, the sum of $P(a_1)$ over all sample points in event A.

4. An important inequality:

$$P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$$

Notice that the equality holds only if they are mutually exclusive. Any overlap reduces the total probability.

3 Conditional Probability and Independence

1. Recall the multiplication rule: $P(B \cap A) = P(B|A)P(A)$

2. When we condition on event A, we are assuming that event A has occurred so that A becomes the new probability space, and all conditional events must take place within event A. Dividing by $P(A)$ scales all probabilities so that A is the entire probability space, and $P(A|A) = 1$.

3. A useful fact: $P(B) = P(B|A)P(A) + P(B|A')P(A')$

4. Bayes' rule:

   (i) The basic form is $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$. This can be expanded even further as follows.

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

   (ii) The extended form. If $A_1, A_2, ..., A_n$ form a partition of the entire probability space Ω, then

$$P(A_i|B) = \frac{P(B \cap A_i)}{P(B)} = \frac{P(B \cap A_i)}{\sum_{i=1}^{n} P(B \cap A_i)} = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)} \text{ for each } i = 1, 2, ..., n$$

5. If events $A_1, A_2, ..., A_n$ satisfy the relationship

$$P(A_1 \cap A_2 \cap ... \cap A_n) = \prod_{i=1}^{n} P(A_i)$$

then the events are said to be mutually exclusive.

6. Some useful facts:

   (i) If $P(A_1 \cap A_2 \cap ... \cap A_{n-1}) > 0$, then

$$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1) \times P(A_2|A_1) \times P(A_3|A_1 \cap A_2) \times ... \times P(A_n|A_1 \cap A_2 \cap ... \cap A_{n-1})$$

   (ii) $P(A'|B) = 1 - P(A|B)$

   (iii) $P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C)$

4 Combinatorial Principles, Permutations and Combinations

1. We say that we are choosing an ordered subset of size k without replacement from a collection of n objects if after the first object is chosen, the next object is chosen from the remaining $n-1$, the next after that from the remaining $n-2$, etc. The number of ways of doing this is $\frac{n!}{(n-k)!}$ and is denoted ${}_nP_k$

2. Given n objects, of which n_1 are of Type 1, n_2 are of Type 2, ..., and n_t are of Type t, and $n = n_1+n_2+...+n_t$, the number of ways of ordering all n objects is $\frac{n!}{n_1!n_2!...n_t!}$

3. Remember that $\binom{n}{k} = \binom{n}{n-k}$

4. Given n objects, of which n_1 are of Type 1, n_2 are of Type 2, ..., and n_t are of Type t, and $n = n_1+n_2+...+n_t$, the number of ways choosing a subset of size $k \leq n$ with k_1 objects of Type 1, k_2 objects of Type 2, ..., and k_t objects of Type t, where $k = k_1 + k_2 + ... + k_t$ is $\binom{n_1}{k_1} \times \binom{n_2}{k_2} \times ... \times \binom{n_t}{k_t}$

5. Recall the binomial theorem:

$$(1+t)^N = \sum_{k=0}^{\infty} \binom{N}{k} t^k$$

6. Multinomial Theorem: In the power series expansion of $(t_1 + t_2 + ... + t_s)^N$ where N is a positive integer, the coefficient of $t_1^{k_1} \times t_2^{k_2} \times ... \times t_s^{k_s}$ (where $k_1 + k_2 + ...k_s = N$) is $\binom{N}{k_1 k_2 ... k_s} = \frac{N!}{k_1!k_2!...k_s!}$. For example, in the expansion of $(1+x+y)^4$, the coefficient of xy^2 is the coefficient of $1^1x^2y^2$, which is $\binom{4}{1\ 1\ 2} = \frac{4!}{1!1!2!} = 12$.

5 Random Variables and Probability Distributions

1. The formal definition of a random variable is that it is a function on a probability space Ω. This function assigns a real number $X(s)$ to each sample point $s \in \Omega$.

2. The probability function of a discrete random variable can be described in a probability plot or in a histogram.

3. Note that for a continuous random variables X, the following are all equal: $P(A < X < b)$, $P(A < X \leq b)$, $P(A \leq X < b)$, and $P(A \leq X \leq b)$.

4. Mixed distribution: A random variable that has some points with non-zero probability mass, and with a continuous pdf on one or more intervals is said to have a mixed distribution. The probability space is a combination of a set of discrete points of probability for the discrete part of the random variable long with one ore more intervals of density for the continuous part. The sum of the probabilities at the discrete points of probability plus the integral of the density function on the continuous region for X must be 1. For example, suppose that X has a probability of 0.5 at $X = 0$, and X is a continuous random variable on the interval $(0,1)$ with density function $f(x) = x$ for $0 < x < 1$, and X has no density or probability elsewhere. It can be checked that this satisfies the conditions of a random variable.

5. Survival function is the complement of the distribution function, $S(x) = 1 - F(x) = P(X > x)$.

6. Recall that for a continuous distribution $F(x)$ is continuous, differentiable, non-decreasing function such that

$$\frac{d}{dx}F(x) = F'(x) = -S'(x) = f(x)$$

7. For a continuous random variable, the hazard rate or failure rate is

$$h(x) = \frac{f(x)}{1-F(x)} = -\frac{d}{dx}\log[1-F(x)]$$

8. Conditional distribution of X given even A: Suppose that $f_X(x)$ is the density function or probability function of X, and suppose that A is an event. The conditional pdf of "X given A" is

$$f_{X|A}(x|A) = \frac{f(x)}{P(A)}$$

if x is an outcome in event A; otherwise $f_{X|A}(x|A) = 0$. For example, suppose that $f_X(x) = 2x$ for $0 < x < 1$, and suppose that A is the event that $X \leq 1/2$. Then $P(A) = P(X \leq 1/2) = 1/4$, and for $0 < x \leq 1/2$, $f_{X|A}(x|X \leq 1/2) = \frac{2x}{1/4} = 8x$ and for $x > 1/2$, $f_{X|A}(x|X \leq 1/2) = 0$.

6 Expectation and Other Distribution Parameters

1. The general increasing geometric series relation $1 + 2r + 3r^2 + ... = \frac{1}{(1-r)^2}$. This can be obtained by differentiating both sides of the equation $1 + r + r^2 + ... = \frac{1}{1-r}$
2. The n-th moment of X is $E[X^n]$. The n-th centered moment of X is $E[(X - \mu)^n]$
3. If $k \geq 0$ is an integer and $a > 0$, then by repeated applications of integration by parts, we have

$$\int_0^\infty t^k e^{-at} dt = \frac{k!}{a^{k+1}}$$

4. Recall that $Var[X] = E[(X - \mu_X)^2] = E[X^2] - (E[X])^2$
5. The standard deviation of the random variable X is the square root of the variance, and is denoted $\sigma_X = \sqrt{Var[X]}$. In addition, the coefficient of variation of X is $\frac{\sigma_X}{\mu_x}$.
6. The moment generating function of X is denoted $M_X(t)$, and it is defined to be $M_X(t) = E[e^{tX}]$. Moment generating functions have some important properties.
   (i) $M_X(0) = 1$.
   (ii) The moments of X can be found from successive derivatives of $M_X(t)$. In other words, $M_X'(t) = E[X]$, $M_X''(t) = E[X^2]$, etc. In addition, $\frac{d^2}{dt^2} \log[M_X(t)]\big|_{t=0} = Var[X]$
   (iii) The moment generating function of X might not exist for all real numbers, but usually exists on some interval of real numbers.
7. If the mean of random variable X is μ and the variance is σ^2, then the skewness is defined to be $E[(X-\mu)^3]/\sigma^3$. If skewness is positive, the distribution is said to be skewed to the right, and if skewness is negative it is skewed to the left.
8. If X is a random variables defined on the interval $[a, \infty)$, then $E[X] = a + \int_a^\infty [1 - F(x)]dx$, and if X is defined on the interval $[a, b]$, where $b < \infty$, then $E[X] = a + \int_a^b [1 - F(x)]dx$. This result is valid for any distribution.
9. Jensen's inequality: If h is a function and X is a random variable such that $\frac{d^2}{dx^2} h(x) = h''(x) \geq 0$ at all points x with non-zero density, then $E[h(X)] \geq h(E[X])$, and if $h''(x) > 0$, then $E[h(X)] > h(E[X])$. The inequality reverses if $h''(x) \leq 0$. For example, if $h(x) = x^2$, then $h''(x) = 2 \geq 0$ for any x, so that $E[h(X)] \geq h(E[X])$.
10. Chebyshev's inequality: If X is a random variable with mean μ_X and standard deviation of σ_X, then for any real number $r > 0$, $P[|X = \mu_x| > r\sigma_X] \leq \frac{1}{r^2}$.
11. The Taylor series expansion of $M_X(t)$ expanded about the point $t = 0$ is

$$M_X(t) = \sum_{k=0}^\infty \frac{t^k}{k!} E[X^k] = 1 + \frac{t}{1!}E[X] + \frac{t^2}{2!}E[X^2] + \frac{t^2}{3!}E[X^3] + ...$$

12. If X_1 and X_2 are random variables, and $M_{X_1}(t) = M_{X_2}(t)$ for all values of t in an interval containing $t = 0$, then X_1 and X_2 have identical probability distributions.

13. The distribution of the random variable X is said to be symmetric about the point c if $f(c+t) = f(c-t)$ for any value t. It then follows that the expected value of X and the median of X is c. Also, for a symmetric distribution, any odd-order central moments about the mean are 0, this means that $E[(X-\mu)^k] = 0$ if $k \geq 1$ is an odd integer.

7 Frequently Used Discrete Distributions

Uniform distribution on N points, $p(x) = \frac{1}{N}$

1. $E[X] = \frac{N+1}{2}$
2. $Var[X] = \frac{N^2-1}{12}$
3. $M_X(t) = \sum_{j=1}^{N} e^{jt}\frac{1}{N} = \frac{e^t(e^{Nt}-1)}{N(e^t-1)}$

Binomial distribution with parameters n and p, $p(x) = \binom{n}{x}p^x(1-p)^{n-x}$

1. $E[X] = np$
2. $Var[X] = np(1-p)$
3. $M_X(t) = (1-p+pe^t)^n$

Poisson distribution with parameter $\lambda > 0$, $p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$

1. $E[X] = Var[X] = \lambda$
2. $M_X(t) = e^{\lambda(e^t-1)}$
3. For the Poisson distribution with mean λ, we have the following relationship between successive probabilities: $P(X = n+1) = P(X = n) \times \frac{\lambda}{n+1}$

Geometric distribution with parameter p, $f(x) = (1-p)^x p$

1. $E[X] = \frac{1-p}{p}$
2. $Var[X] = \frac{1-p}{p^2}$
3. $M_X(t) = \frac{p}{1-(1-p)^{e^t}}$
4. The geometric distribution is called memoryless, i.e. $P(X = n+k|X \geq n) = P(X = k)$

Negative binomial distribution with parameters r and p ($r > 0$ and $0 < p \leq 1$), $p(x) = \binom{r+x-1}{x}p^r(1-p)^x$.

1. If r is an integer, then the negative binomial can be interpreted as follows. Suppose that an experiment ends in either failure or success, and the probability of success for a particular trial of the experiment is p. Suppose further that the experiment is performed repeatedly (independent trials) until the r-th success occurs. If X is the number of failures until the r-th success occurs, then X has a negative binomial distribution with parameters r and p.
2. $E[X] = \frac{r(1-p)}{p}$
3. $Var[X] = \frac{r(1-p)}{p^2}$
4. $M_X(t) = \left[\frac{p}{1-(1-p)e^t}\right]^r$
5. Notice that the geometric distribution is a special cage of the negative binomial with $r = 1$.

Hypergeometric distribution with integer parameters M, K, and n ($M > 0$, $0 \leq K \leq M$, and $1 \leq n \leq M$), $p(x) = \frac{\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}}$

1. In a group of M objects, suppose that K are of Type 1 and $M - K$ are of Type 2. If a subset of n objects is randomly chosen without replacement from the group of M objects, let X denote the number that are of Type 1 in the subset of size n. X is said to have a hypergeometric distribution.
2. $E[X] = \frac{nK}{M}$
3. $Var[X] = \frac{nK(M-K)(M-n)}{M^2(M-1)}$

Multinomial distribution with parameters $n, p_1, p_2, ..., p_k$ (where n is a positive integer and $0 \leq p_1 \leq 1$ for all $i = 1, 2, ..., k$ and $p_1 + p_2 + ... + p_k = 1$.)

1. $P(X_1 = x_1, X_1 = x_1, ..., X_k = x_k) = \frac{n!}{x_1!x_2!...x_k!} p_1^{x_1} p_2^{x_2} ... p_l^{x_k}$
2. For each i, X_i is a random variable with a mean and variance similar to the binomial mean and variance: $E[X_i] = np_i$ and $Var[X_i] = np_i(1 - p_i)$
3. $Cov[X_i, X_j] = -np_i p_j$

8 Frequently Used Continuous Distributions

Uniform distribution on the interval (a, b), $f(x) = \frac{1}{b-a}$

1. The mean and the variance are $E[X] = \frac{a+b}{2}$ and $Var[X] = \frac{(b-a)^2}{12}$
2. $M_X(t) = \frac{e^{bt} - e^{at}}{t(b-a)}$
3. The n-th moment of X is $E[X^n] = \frac{b^{n+1} - a^{n+1}}{(n+1)(b-a)}$
4. The median is $\frac{a+b}{2}$

The normal distribution

1. The standard normal: $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ for which the moment generating function is $M_X t = e^{t^2/2}$
2. The more general form: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$ for which the moment generating function is $M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$
3. Notice also that for a normal distribution mean = median = mode = μ.
4. Given ad random variable (not normal) X with mean μ and σ^2, probabilities related to the distribution of X are sometimes approximated by assuming the distribution of X is approximately $N(\mu, \sigma^2)$.
   (i) If n and m are integers, the probability $P(n \leq X \leq m)$ can be approximated by using a normal random variable Y withe same mean and variance as X, and then finding the probability $P(n - 1/2 \leq Y \leq m + 1/2)$. For example, to approximate $P(X = 3)$ where X is a binomial distribution with $n = 10$ and $p = 0.5$, we could use $Y \sim N(5, 2.5)$ and calculate $P(2.5 \leq Y \leq 3.5)$.
5. If X_1 and X_2 are independent normal random variables with mean μ_1 and μ_2, and variances σ_1^2 and σ_2^2, then $W = X_1 + X_2$ is also a normal random variables, and has mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$.

Exponential distribution with mean $\frac{1}{\lambda} > 0$, $f(x) = \lambda e^{-\lambda x}$

1. The mean and variance are $E[X] = 1/\lambda$ and $Var[X] = 1/\lambda^2$
2. $M_X(t) = \frac{\lambda}{\lambda - t}$ for $t < \lambda$.
3. The k-th moment is $E[X^k] = \int_o^\infty x^k \lambda e^{-\lambda x} dx = \frac{k!}{\lambda^k}$
4. Lack of memory property: $P(X > x + y | X > x) = P(X > y)$

5. Suppose that independent random variables $X_1, X_2, ..., X_n$ each have exponential distributions with means $\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, ..., \frac{1}{\lambda_n}$. Let $X = min\{X_1, X_2, ..., X_n\}$. Then X has an exponential distribution with mean $\frac{1}{\lambda_1+\lambda_2+...+\lambda_n}$

Gamma distribution with parameter $\alpha > 0$ and $\beta > 0$, $f(x) = \frac{1}{\Gamma(\alpha)}\beta^\alpha x^{\alpha-1}e^{-\beta x}$

1. $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}e^{-y}$. If n is a positive integer it can be shown that $\Gamma(n) = (n-1)!$.
2. The mean, variance, and moment generating function of X are $E[X] = \alpha/\beta$, $Var[X] = \alpha/\beta^2$, and $M_X(t) = \left(\frac{\beta}{\beta-t}\right)^\alpha$ for $t < \beta$.

Pareto distribution with parameters $\alpha, \beta > 0$, $f(x) = \frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}}$

1. The mean and variance are $E[X] = \frac{\theta}{\alpha-1}$ and $Var[X] = \frac{\alpha\theta^2}{(\alpha-2)(\alpha-1)^2}$
2. Another version of the Pareto distribution is the single parameter Pareto, which still has α and θ, but has a pdf $f(x) = \frac{\alpha\theta^\alpha}{x^{\alpha+1}}$. The the mean and variance become $E[X] = \frac{\alpha\theta}{\alpha-1}$ and $Var[X] = \frac{\alpha\theta^2}{(\alpha-2)(\alpha-1)^2}$. Overall the one parameter Pareto pdf is the same as the two parameter Pareto shifted to the right θ units.

Beta distribution with parameters $\alpha > 0$ and $\beta > 0$, $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$ for $0 < x < 1$

1. The mean and variance are $E[X] = \frac{\alpha}{\alpha+\beta}$ and $Var[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
2. Note that the uniform distribution on the interval $(0, 1)$ is a special case of the beta distribution with $\alpha = \beta = 1$.

Lognormal distribution with parameters μ and $\sigma^2 > 0$, $f(x) = \frac{1}{x\sigma\sqrt{2\pi}}e^{-(\log(x)-\mu)^2/2\sigma^2}$

1. If $W \sim N(\mu, \sigma^2)$, then $X = e^W$ is said to have a lognormal distribution with parameters μ and σ^2. This is the same as saying that the natural log of X has a normal distribution $N(\mu, \sigma^2)$.
2. The mean and variance are $E[X] = e^{\mu+\frac{1}{2}\sigma^2}$ and $Var[X] = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2}$
3. Notice that μ and σ^2 are the mean and variance of the underlying random variable W.

Weibull distribution with parameters $\theta > 0$ and $\tau > 0$, $f(x) = \frac{\tau(x/\theta)^\tau e^{-(x/\theta)^\tau}}{x}$

1. The mean and variance involve the gamma function and therefore are not listed here.
2. Note that the exponential with mean θ is a special case of the Weibull distribution with $\tau = 1$.

Chi-square with k degrees of freedom, $f(x) = \frac{1}{\Gamma(k/2)}(1/2)^{k/2}x^{(k-2)/2}e^{-x/2}$

1. The mean, variance, and moment generating function are $E[X] = k$, $Var[X] = 2k$, and $M_X(t) = \left(\frac{1}{1-2t}\right)^{k/2}$ for $t < 1/2$.
2. If X has a Chi-square distribution with 1 degree of freedom then $P(X < a) = 2\Phi(\sqrt{a}) - 1$. where Φ is the cdf of the standard normal. A Chi-square distribution with 2 degrees of freedom is the same as an exponential distribution with mean 2. If $Z_1, Z_2, ..., Z_m$ are independent standard normal random variables, and if $X = Z_1^2 + Z_2^2 + ... + Z_m^2$, then X has a Chi-square distribution with m degrees of freedom.

9 Joint, Marginal, and Conditional Distributions

1. Cumulative distribution function of a joint distribution: If X and Y have a joint distribution, then the cumulative distribution function is $F(x,y) = P((X \leq x) \cap (Y \leq y))$.
   (i) In the continuous case we have $F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s,t)dtds$
   (ii) In the discrete case we have $F(x,y) = \sum_{s=-\infty}^{x} \sum_{t=-\infty}^{y} f(s,t)$
2. Let X and Y be jointly distributed random variables, then the expected value of $h(X,Y)$ is defined to be
   (i) In the continuous case we have $E[h(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(X,Y)f(x,y)dxdy$
   (ii) In the discrete case we have $E[h(X,Y)] = \sum_x \sum_y h(X,Y)f(x,y)$
3. If X and Y have a joint distribution with joint density or probability function $f(x,y)$, then the marginal distribution of X has a probability function or density function denoted $f_X(x)$, which is equal to $f_X(x) = \sum_y f(x,y)$ in the discrete case, and is equal to $f_X(x) = \int_{-\infty} \infty f(x,y)dy$ in the continuous case. Remember to be careful with the limits or integration/summation.
4. If the cumulative distribution function of the joint distribution of X and Y is $F(x,y)$, then the cdf for the marginal distributions of X and Y are

$$F_X(x) = \lim_{y\to\infty} F(x,y)$$

$$F_Y(y) = \lim_{x\to\infty} F(x,y)$$

5. Random variables X and Y with density functions $f_X(x)$ and $f_Y(y)$ are said to be independent, if the probability spare is rectangular ($a \leq x \leq b$, $c \leq y \leq d$, where the end points can be infinite) and if the joint density function is of the form $f(x,y) = f_X(x)f_Y(y)$. This definition is equivalent in terms of the joint cumulative distribution function.
6. Many properties of the conditional probability apply to probability density functions. For instance, recall that $f_{Y|X}(y|X=x) = \frac{f(x,y)}{f_X(x)}$. Similarly, we can find the conditional mean of$Y$ given $X = x$, which is $E[Y|X=x] = \int y \times f_{Y|X}(y|X=x)dy$. Another useful fact is $f(x,y) = f_{Y|X}(y|X=x) \times f_X(x)$
7. Covariance and correlation are defines as follows:

$$\begin{aligned} Cov[X,Y] &= E[(X - E[X])(Y - E[Y])] \\ &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

$$corr[X,Y] = \rho(X,Y) = \frac{Cov[X,Y]}{\sigma_X \sigma_Y}$$

7. Moment generating function of a joint distribution: Given jointly distributed random variables X and Y, the moment generating function of the joint distribution is $M_{X,Y}(t_1,t_2) = E[e^{t_1X+t_2Y}]$. This definition can be extended to the joint distribution of any number of random variables. It can be shown that $E[X^nY^m]$ is equal to the multiple partial derivative evaluated at 0, i.e.

$$E[X^nY^m] = \frac{\partial^{n+m}}{\partial^n t_1 \partial^m t_2} M_{X,Y}(t_1,t_2)\bigg|_{t_1=t_2=0}$$

8. Suppose that X and Y are normal random variables with means and variances $E[X] = \mu_X$, $Var[X] = \sigma_X^2$, $E[Y] = \mu_Y$, $Var[Y] = \sigma_Y^2$, and with correlation coefficient ρ_{XY}. X and Y are said to have a bivariate normal distribution. The conditional mean and variance of Y given $X = x$ are

$$\begin{aligned} E[Y|X = x] &= \mu_Y + \rho_{XY} \times \frac{\sigma_Y}{\sigma_X} \times (x - \mu_x) \\ &= \mu_Y + \times \frac{Cov[X, Y]}{\sigma_X^2} \times (x - \mu_x) \\ Var[Y|X = x] &= \sigma_Y^2 \times (1 - \rho_{XY}^2) \end{aligned}$$

The pdf of the bivariate normal distribution is

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \times \exp\left[-\frac{1}{2(1-\rho^2)} \times [(\frac{x-\mu_X}{\sigma_X})^2 + (\frac{x-\mu_Y}{\sigma_Y})^2 - 2\rho(\frac{x-\mu_X}{\sigma_X})(\frac{x-\mu_Y}{\sigma_Y})] \right]$$

9. Some additional useful facts relating to this section:

   (i) If X and Y are independent, then for any functions g and h, $E[g(X) \times h(Y)] = E[g(X)] \times E[h(Y)]$, and in particular, $E[XY] = E[X]E[Y]$

   (ii) For constants a, b, c, d, e, f and random variables X, Y, Z, and W, we have

   $$Cov[aX + bY + c, dZ + eW + f] = adCov[X, Z] + aeCov[X, W] + bdCov[Y, Z] + beCov[Y, W]$$

   (ii) If X and Y have a joint distribution which is uniform on the two dimensional region R, then the pdf of the joint distribution is $\frac{1}{\text{Area of R}}$ inside the region R. The probability of any event A is the proportion $\frac{\text{Area of A}}{\text{Area of R}}$. Also the conditional distribution of Y given $X = x$ has a uniform distribution on the line segment defined by the intersection of the region R and the line $X = x$. The marginal distribution of Y might or might not be uniform.

   (iii) $P((x_1 < X < x_2) \cap (y_1 < Y < y_2)) = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1)$

   (iv) $P((X \le x) \cup (Y \le y)) = F_X(x) + F_Y(y) - F(x, y) \le 1$. A nice form of the inclusion-exclusion theorem.

   (v) $M_{X,Y}(t_1, 0) = E[e^{t_1X}] = M_X(t_1)$ and $M_{X,Y}(0, t_2) = E[e^{t_2Y}] = M_Y(t_2)$

   (vi) Recall that

   $$\left.\frac{\partial}{\partial t_1} M_{X,Y}(t_1, t_2)\right|_{t_1=t_2=0} = E[X]$$
   $$\left.\frac{\partial}{\partial t_2} M_{X,Y}(t_1, t_2)\right|_{t_1=t_2=0} = E[Y]$$

   (vii) If $M(t_1, t_2) = M(t_1, 0)M(0, t_2)$ for t_1 and t_2 in the region about $(0, 0)$, then X and Y are independent.

   (viii) if $Y = aX + b$, then $M_Y(t) = e^{bt}M_X(at)$.

10 Transformations of Random Variables

1. Suppose that X is a continuous random variable with $f_X(x)$ and cdf $F_X(x)$, and suppose that $u(x)$ is a one-to-one function. As a one-to-one function, u has an inverse function, so that $u^{-1}(u(x)) = x$. The random variables $Y = u(X)$ is referred to as a transformation of X. The pdf of Y can be found in one of the two ways:

   (i) $f_Y(y) = f_X(u^{-1}(y)) \times |\frac{d}{dy}u^{-1}(y)|$

(ii) If u is a strictly increasing function, then

$$F_Y(y) = P(Y \leq y) = P(u(X) \leq y) = P(X \leq u^{-1}(y)) = F_X(u^{-1}(y))$$

and $f_Y(y) = F'_Y(y)$.

2. Suppose that X is a discrete random variables with probability functions $f_X(x)$. If $u(x)$ is a function of x, and Y is random variables defined by the equation $Y = u(X)$, then Y is a discrete random variable with probability functions $g(y) = \sum_{y=u(x)} f(x)$. In other words, given a value of y, find all values of x for which $y = u(x)$, and then $g(y)$ is the sum of those $f(x_i)$ probabilities.

3. Suppose that the random variables X and Y are jointly distributed with joint density function $f(x, y)$. Suppose also that u and v are functions of the variables x and y. Then $U = u(X, Y)$ and $V = v(X, Y)$ are also random variables with a joint distribution. We wish to find the joint density function of U And V, say $g(u, v)$. In order to do this, we must be able to find inverse functions $h(u, v)$ and $k(u, v)$ such that $x = h(u(x, y), v(x, y))$ and $y = k(u(x, y), v(x, y))$. The joint density of U and V is then

$$g(u, v) = f(h(u, v), k(u, v)) \times \left| \frac{\partial h}{\partial u} \times \frac{\partial k}{\partial v} - \frac{\partial h}{\partial v} \times \frac{\partial k}{\partial u} \right|$$

4. If $X_1, X_2, ..., X_n$ are random variables, and the random variable Y is defined to be $Y = \sum_{i=1}^{n} X_i$, then

$$\begin{aligned} E[Y] &= \sum_{i=1}^{n} E[X_i] \\ Var[Y] &= \sum_{i=1}^{n} Var[X_i] + 2\sum_{i=1}^{n} \sum_{j=i+1}^{n} Cov[X_i, X_j] \end{aligned}$$

In addition, if $X_1, X_2, ..., X_n$ are mutually independent random variables, then

$$\begin{aligned} Var[Y] &= \sum_{i=1}^{n} Var[X_i] \\ M_Y(t) &= \prod_{i=1}^{m} M_{X_i}(t) \end{aligned}$$

5. Suppose that $X_1, X_2, ..., X_n$ are independent random variables and $Y = \sum_{i=1}^{n} X_i$, then

Distribution of X_i	Distribution of Y
Bernoulli $B(1, p)$	Binomial $B(k, p)$
Binomial $B(n_i, p)$	Binomial $B(\sum n_i, p)$
Poisson λ_i	Poisson $\sum \lambda_i$
Geometric p	Negative binomial k, p
Negative Binomial r_i, p	Negative binomial $\sum r_i, p$
Normal $N(\mu_i, \sigma_i^2)$	Normal $N(\sum \mu_i, \sum \sigma_i^2)$
Exponential with mean μ	Gamma with $\alpha = k$ and $\beta = 1/\mu$
Gamma with α_i, β	Gamma with $\sum \alpha_i, \beta$
Chi-square with k_i df	Chi-square with $\sum k_i$ df

6. Suppose that X_1 and X_2 are independent random variables. We define two new random variables related to X_1 and X_2: $U = max\{X_1, X_2\}$ and $V = min\{X_1, X_2\}$. We wish to find the distributions of U and V. Suppose that we know that the distribution functions of X_1 and X_2 are $F_{X_1}(x) = P(X_1 \leq x)$ and $F_{X_2}(x) = P(X_2 \leq x)$. We can formulate the distribution functions of U and V in terms of F_{X_1} and F_{X_2} as follows.

$F_U(u) = P(U \le u) = P(max\{X_1, X_2\} \le u) = P((X_1 \le u) \cap (X_2 \le u)$. Since X_1 and X_2 are independent, we have that $P((X_1 \le u) \cap (X_2 \le u) = P((X_1 \le u)) \times P((X_2 \le u) = F_{X_1}(u) \times F_{X_2}(u)$. Thus the distribution function of U is $F_U(u) = F_{X_1}(u) \times F_{X_2}(u)$.

$F_V(v) = P(V \le v) = 1 - P(V > v) = 1 - P(min\{X_1, X_2\} > v) = 1 - P((X_1 > v) \cap (X_2 > v)$. Since X_1 and X_2 are independent, we have that $P((X_1 > v) \cap (X_2 > v) = P((X_1 > v)) \times P((X_2 > v) = (1 - F_{X_1}(v)) \times (1 - F_{X_2}(v))$. Thus the distribution function of V is $F_V(v) = 1 - (1 - F_{X_1}(v)) \times (1 - F_{X_2}(v))$

6. Let $X_1, X_2, ..., X_n$ are independent random variables , and Y_i's be the same collection of numbers as X's, but they have been put increasing order. The density function of Y_k can be described in terms of $f(x)$ and $F(x)$. For each $k = 1, 2, ..., n$ the pdf of Y_k is

$$g_k(t) = \frac{n!}{(k-1)!(n-k)!}[F(t)]^{k-1}[1 - F(t)]^{n-k} f(t)$$

   In addition, the joint density of $Y_1, Y_2, ..., Y_n$ is $g(y_1, y_2, ..., y_n) = n! f(y_1) f(y_2) ... f(y_n)$.

7. Suppose X_1 and X_2 are random variables with density functions $f_1(x)$ and $f_2(x)$, and suppose a is a number with $0 < a < 1$. We define a new random variables X by defining a new density functions $f(x) = a \times f_1(x) + (1 - a) \times f_2(x)$. This newly defined density function will satisfy the requirements for being a properly defined density function. Furthermore, all moments, probabilities and the moment generating function of the newly defined random variables are of the form

$$\begin{aligned} E[X] &= aE[X_1] + (1-a)E[X_2] \\ E[X^2] &= aE[X_1^2] + (1-a)E[X_2^2] \\ F_X(x) &= aF_1(x) + (1-a)F_2(x) \\ M_X(t) &= aM_{X_1}(t) + (1-a)M_{X_2}(t) \end{aligned}$$

   Notice that this relationship does not apply to variance. If we wanted to calculate variance, we would have to use the formula $Var[X] = E[X^2] - E[X]^2$.

11 Risk Management Concepts

1. When someone is subject to the risk of incurring a financial loss, the loss is generally modeled using a random variables or some combination of random variables. Once the random variable X representing the loss has been determined, the expected value of the loss, $E[X]$, is referred to as the pure premium for the policy. $E[X]$ is also the expected claim on the insurer. For a random variable X a measure of the risk is $\sigma^2 = Var[X]$. The unitized risk or coefficient of variation for the random variable X is defined to be $\frac{\sqrt{Var[X]}}{E[X]} = \frac{\sigma}{\mu}$

2. There are several different ways to model loss:

   Case 1: The complete description of X is given. In this case, if X is continuous, the density function $f(x)$ or distribution function $F(x)$ is given. In X is discrete, the probability function is given. One typical example of the discrete case is a loss random variable such that

$$\begin{aligned} P(X = K) &= q \\ P(X = 0) &= 1 - q \end{aligned}$$

   This could, for instance, arise in a one-year term life insurance in which the death benefit is K.

   Case 2: The probability of a of a non-negative loss is given, and the conditional distribution of B of loss amount given that loss has occurred is given. The probability of no loss occurring is $1 - q$, and the loss amount X is 0 if no loss occurs; thus, $P(X=0)=1-q$. If a loss does occur, the loss amount is the random variable B, so that $X=B$. The random variable B is the loss amount given that a loss has occurred, so that B is really the conditional distribution of the loss amount X given that a loss occurs. The random variable V might be described in detail, or only the mean and variance of B might be known. Note that if $E[B]$ and $Var[B]$ are given, then $E[B^2]=Var[B]+(E[B])^2$.

3. The individual risk model assumes that the portfolio consists of a specific number, say n, of insurance policies, with the claim for one period on policy i being the random variable X_i. X_i would be modeled in one of the was described above for an individual policy loss random variable. Unless mentioned otherwise, it is assumed that the X_i's are mutually independent random variables. Then the aggregate claim is the random variable

$$\begin{aligned} S &= \sum_{i=1}^{n} X_i \text{ with} \\ & E[S]=\sum_{i=1}^{n} E[X_i] \text{ and } Var[S]=\sum_{i=1}^{n} Var[X_i] \end{aligned}$$

An interesting fact is to notice that if $E[X_i]=\mu$ and $Var[X_i]=\sigma^2$ for each $i=1,2,...,n$, then the coefficient of variation of the aggregate claim distribution S is $\frac{\sqrt{Var[X]}}{E[S]}=\frac{\sqrt{nVar[X]}}{nE[X]}=\frac{\sigma}{\mu\sqrt{n}}$ which goes to 0 as $n\to\infty$.

4. Many insurance policies do not cover the full amount of the loss that occurs, but only provide partial coverage. There are a few standard types of partial insurance coverage that can be applied to a basic ground up loss (full loss) random variables X:

(i) Deductible insurance: A deductible insurance specifies a deductible amount, say d. If a loss of amount X occurs, the insurer pays nothing if the loss is less than d, and pays the policyholder the amount of the loss in excess of d if the loss is greater than d. The amount paid by the insurer can be described as

$$\begin{aligned} Y &= 0 \text{ if } X\le q \\ Y &= X-d \text{ if } X>d \end{aligned}$$

This is also denoted $(X-d)_+$. The expected payment made by the insurer when a loss occurs would be $\int_d^\infty (x-d)f_X(x)dx$ in the continuous case.

(ii) Policy limit: A policy limit of amount u indicates that the insurer will pay a maximum amount of u when a loss occurs. Thus the amount paid by the insurer is

$$\begin{aligned} Y &= X \text{ if } X\le u \\ Y &= u \text{ if } X>u \end{aligned}$$

The expected payment made by the insurer per loss would be $\int_0^u xf_X(x)dx+u[1-F_X(u)]$ in the continuous case.

(iii) Proportional insurance: Proportional insurance specifies a fraction $0<\alpha<1$, and if a loss of amount X occurs, the insurer pays the policyholder αX, the specified fraction of the full loss.