Exam P Review Sheet

$\log_b(b^x) = x$ $\quad$ $\log_b(y^k) = k\log_b(y)$ $\quad$ $\log_b(y) = \frac{\ln(y)}{\ln(b)}$

$\log_b(yz) = \log_b(y) + \log_b(z)$ $\quad$ $\log_b(y/z) = \log_b(y) - \log_b(z)$

$\ln(e^x) = x$ $\quad$ $e^{\ln(y)} = y$ for $y > 0$.

$\frac{d}{dx}a^x = a^x \ln(a)$ $\quad$ $\int a^x dx = \frac{a^x}{\ln(a)}$ for $a > 0$.

$\int_0^\infty x^n e^{-cx} dx = \frac{n!}{c^{n+1}}$. $\quad$ In particular, for $a > 0$, $\int_0^\infty e^{-at} dt = \frac{1}{a}$

$$a + ar + ar^2 + ar^3 + \cdots + ar^n = a\frac{1 - r^{n+1}}{1 - r} = a\frac{r^{n+1} - 1}{r - 1}$$

if $|r| < 1$ then $a + ar + ar^2 + ar^3 + \cdots = \sum_{i=0}^\infty ar^i = \dfrac{a}{1 - r}$

Note: (derivative of above) $a + 2ar + 3ar^2 + \cdots = \dfrac{a}{(1 - r)^2}$.

If $A \subseteq B$, then $P(A) \leq P(B)$.

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

$P(A) = P(A \cap B) + P(A \cap B')$

If $P(A) > 0$, then $P(B|A) = \dfrac{P(A \cap B)}{P(A)}$, and $P(A \cap B) = P(B|A)P(A)$.

$P(B) = P(B|A)P(A) + P(B|A')P(A')$ (Law of Total Probability)

$P(A|B) = \dfrac{P(B|A)P(A)}{P(B)}$ (Bayes)

$P(A_j|B) = \dfrac{P(A_j \cap B)}{P(B)} = \dfrac{P(B|A_j)P(A_j)}{\sum_{i=1}^n P(B|A_i)P(A_i)}$ (Note that the A_i's form a partition)

A, B are **independent** iff $P(A \cap B) = P(A)P(B)$ iff $P(A|B) = P(A)$ iff $P(B|A) = P(B)$.

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

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

If A, B are independent, then $(A', B), (A, B')$, and (A', B') are also independent (each pair).

Given n distinct objects, the number of different ways in which the objects may be ordered (or permuted) is $n!$.

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 $n - 1$, then next after that from the remaining $n - 2$ and so on. The number of ways of doing this is $\frac{n!}{(n-k)!}$, and is denoted ${}_nP_k$ or $P_{n,k}$ or $P(n, k)$.

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 $n = n_1 + n_2 + \cdots + n_t$, the number of ways of ordering all n objects (where objects of the same type are indistinguishable) is
$\dfrac{n!}{n_1!n_2!\cdots n_t!}$ which is sometimes denoted as $\dbinom{n}{n_1 n_2 \cdots n_t}$.

Given n distinct objects, the **number of ways of choosing a subset of size $k \le n$ without replacement and without regard to the order in which the objects are chosen** is $\frac{n!}{k!(n-k)!}$, which is usually denoted $\binom{n}{k}$. Remember, $\binom{n}{k} = \binom{n}{n-k}$.

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 $n = n_1 + n_2 + \cdots + n_t$, the number of ways of choosing a subset of size $k \le n$ (without replacement) with k_1 objects of Type 1, k_2 objects of Type 2, and so on, where $k = k_1 + k_2 + \cdots + k_t$ is $\binom{n_1}{k_1}\binom{n_2}{k_2}\cdots\binom{n_t}{k_t}$

Binomial Theorem: In the power series expansion of $(1+t)^N$, the coefficient of t^k is $\binom{N}{k}$, so that $(1+t)^N = \sum_{k=0}^{\infty} \binom{N}{k} t^k$.

Multinomial Theorem: In the power series expansion of $(t_1 + t_2 + \cdots + t_s)^N$ where N is a positive integer, the coefficient of $t_1^{k_1} t_2^{k_2} \cdots t_s^{k_s}$ (where $k_1 + k_2 + \cdots + k_s = N$) is $\binom{N}{k_1 k_2 \cdots k_s} = \frac{N!}{k_1! k_2! \cdots k_s!}$.

Discrete Distributions:
The **probability function (pf)** of a discrete random variable is usually denoted $p(x)$, $f(x)$, $f_X(x)$, or p_x, and is equal to $P(X = x)$, the probability that the value x occurs. The probability function must satisfy:
(i) $0 \le p(x) \le 1$ for all x, and (ii) $\sum_x p(x) = 1$.

Given a set A of real numbers (possible outcomes of X), the probability that X is one of the values in A is $P(X \in A) = \sum_{x \in A} p(x) = P(A)$.

Continuous Distributions:
A continuous random variable X usually has a **probability density function (pdf)** usually denoted $f(x)$ or $f_X(x)$, which is a continuous function except possibly at a finite number of points. Probabilities related to X are found by integrating the density function over an interval. The probability that X is in the interval (a, b) is $P(X \in (a, b)) = P(a < X < b)$, which is defined to be $\int_a^b f(x)dx$.
Note that for a continuous random variable X, $P(a < X < b) = P(a \le X < b) = P(a < X \le b) = P(a \le X \le b)$. Recall that if we have a mixed distribution that this is not always the case.

The pdf $f(x)$ must satisfy:
(i) $f(x) \ge 0$ for all x, and (ii) $\int_{-\infty}^{\infty} f(x)dx = 1$.

Mixed Distributions:
A random variable that has some points with non-zero probability mass, and with a continuous pdf on one ore 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 along with one or 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. If a pdf looks to have discontinuities then we likely have a mixed distribution.

Cumulative distribution function (and survival function): Given a random variable X, the **cumulative distribution function** of X (also called the distribution function, or cdf) is $F(x) = P(X \le x)$ (also denoted $F_X(x)$). $F(x)$ is the cumulative probability to the left of (and including) the point x. The survival function is the complement of the distribution function, $S(x) = 1 - F(x) = P(X > x)$.

For a discrete random variable with probability function $p(x), F(x) = \sum_{w \le x} p(w)$, and in this case $F(x)$ is a step function (it has a jump at each point with non-zero probability, while remaining constant until the next jump).

If X has a continuous distribution with density function $f(x)$, then $F(x) = \int_{-\infty}^{x} f(t)dt$ and $F(x)$ is a continuous, differentiable, non-decreasing function such that $\frac{d}{dx}F(x) = F'(x) = -S'(x) = f(x)$. If X has a mixed distribution, then $F(x)$ is continuous except at the points of non-zero probability mass, where $F(x)$ will have a jump.

For any cdf $P(a < X < b) = F(b) - F(a), \quad \lim_{x \to \infty} F(x) = 1, \quad \lim_{x \to -\infty} F(x) = 0.$

Some results and formulas from this section:

(i) For a continuous random variable, the hazard rate or failure rate is
$h(x) = \dfrac{f(x)}{1 - F(x)} = -\dfrac{d}{dx} \ln[1 - F(x)]$
(ii) If X, Y are independent, then $P[(a < X \le b) \cap (c < Y \le d)] = P(a < X \le b)P(c < Y \le d)$. In general, what we mean by saying that X and Y are independent is that if A is any event involving only X (such as $a < X \le b$), and B is any event involving only Y, then A and B are independent events.

(iii) **Conditional distribution of X given event 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 or pf of X given A is
$f_{X|A}(x|A) = \dfrac{f(x)}{P(A)}$ if x is an outcome in A, and 0 otherwise.

Expected value of a random variable: For a random variable X, the expected value of X (also called the **mean of** X or the **expectation of** X) is denoted $E[X]$, or μ_x or μ. The mean is interpreted as the average of the random outcomes.

For a discrete random variable, the expected value of X is $\sum xp(x) = x_1 p(x_1) + x_2 p(x_2) + \cdots$ where the sum is taken over all points x at which X has non-zero probability.

For a continuous random variable, the expected value of X is $\int_{-\infty}^{\infty} xf(x)dx$. Note that although the integral is written with limits involving infinity, we actually integrate over the interval(s) of non-zero density for X.

Expectation of $h(x)$: If h is a function, then $E[h(X)]$ is equal to $\sum_x h(x)p(x)$ in the discrete case, and is equal to $\int_{-\infty}^{\infty} h(x)f(x)dx$ in the continuous case.

Moments of a random variable: If $n \ge 1$ is an integer, the $n-$th moment of X is $E[X^n]$. If the mean of X is μ, then the $n-$th central moment of X (about the mean μ) is $E[(X - \mu)^n]$.

Symmetric distribution: If X is a continuous random variable with pdf $f(x)$, and if c is a point for which $f(c + t) = f(c - t)$ for all $t > 0$, then X is said to have a symmetric distribution about the point $x = c$. For such a distribution, the mean will be the point of symmetry, $E[X] = c$.

Variance of X: The variance of X is denoted $Var[X], V[X], \sigma_x^2$, or σ^2. It is defined to be equal to $Var[X] = E[(X-\mu)^2] = E[X^2] - (E[X])^2$.
The **standard deviation of** X is the square root of the variance of X, and is denoted $\sigma_x = \sqrt{Var[X]}$.
The **coefficient of variation** of X is $\dfrac{\sigma_x}{\mu_x}$.

Moment generating function (mgf) or a random variable X: The moment generating function of X (mgf) is defined to be $M_x(t) = E[e^{tX}]$. The mfg is also denoted as $M_X(t), m_x(t), M(t)$, or $m(t)$.

If X is a discrete random variable then $M_x(t) = \sum e^{tx} p(x)$.

If X is a continuous random variable then $M_x(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$.

Some important properties that moment generating functions satisfy are

(i) It is always true that $M_X(0) = 1$

(ii) The moments of X can be found from the successive derivatives of $M_x(t)$.

$M_x'(0) = E[X],\ M_X''(0) = E[X^2],\ M_X^{(n)}(0) = E[X^n]$ and $\dfrac{d^2}{dt^2} \ln[M_x(t)]\Big|_{t=0} = Var[X]$.

(iii) The moment generating function of X might not exist for all real numbers, but usually exists on some interval of real numbers.

Percentiles of a distribution: If $0 < p < 1$, then the $100p-$th percentile of the distribution of X is the number c_p which satisfies both of the following inequalities:
$P(X \le c_p) \ge p$ and $P(X \ge c_p) \ge 1 - p$.

For a continuous random variable, it is sufficient to find the c_p for which $P(X \le c_p) = p$.

If $p = .5$ the $50-$th percentile of a distribution is referred to as the median of the distribution, it is the point M for which $P(X \le M) = .5$.

The mode of a distribution: The mode is any point m at which the probability or density function $f(x)$ is maximized.

The skewness of a distribution: 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 the skewness is positive, the distribution is said to be skewed to the right, and if the skewness is negative, then the distribution is said to be skewed to the left.

Some formulas and results:
(i) Note that the mean of a random variable X might not exist.

(ii) For any constants a_1, a_2 and b and functions h_1 and h_2,

$E[a_1 h_1(X) + a_2 h_2(X) + b] = a_1 E[h_1(X)] + a_2 E[h_2(X)] + b$. As a special case, $E[aX+b] = aE[X]+b$.

(iii) If X is a random variable defined on the interval $[a, \infty)(f(x) = 0$ for $x < a)$, 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 relationship is valid for any random variable, discrete, continuous or with a mixed distribution. As a special case, if X is a non-negative random variable (defined on $[0, \infty)$, or $(0, \infty)$), then $E[X] = \int_0^{\infty} [1 - F(x)] dx$.

(iv) **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 or probability for X, then $E[h(X)] \geq h(E[X])$, and if $h'' > 0$, then $E[h(X)] > h(E[X])$.

(v) If a and b are constants, then $Var[aX + b] = a^2 Var[X]$.

(vi) **Chebyshev's inequality:** If X is a random variable with mean μ_x and standard deviation σ_x, then for any real number $r > 0, P[|X - \mu_x| > r\sigma_x] \leq \frac{1}{r^2}$.

(vii) Suppose that for the random variable X, the moment generating function $M_x(t)$ exists in an interval containing the point $t = 0$. Then,

$\frac{d^n}{dt^n} M_X(t)\Big|_{t=0} = M_X^{(n)}(0) = E[X^n]$, the $n-$th moment of X, and

$\frac{d}{dt} \ln[M_X(t)]\Big|_{t=0} = \frac{M'_X(0)}{M_X(0)} = E[X]$, and $\frac{d}{dt^2} \ln[M_X(t)]\Big|_{t=0} = Var[X]$.

The Taylor series expansion of $M_X(t)$ expanded about the point $t = 0$ is

$M_X(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} E[X^k] = 1 + tE[X] + \frac{t^2}{2} E[X^2] + \frac{t^3}{6} E[X^3] + \cdots$.

Therefore, if we are given a moment generating function and we are able to formulate the Taylor series expansion about the point $t = 0$, we can identify the successive moments of X.

If X has a discrete distribution with probability space $\{x_1, x_2, \ldots\}$ and probability function $P(X = x_k) = p_k$, then the moment generating function is $M_X(t) = e^{tx_1} p_1 + e^{tx_2} p_2 + e^{tx_3} p_3 + \cdots$.
Conversely, if we are given a moment generating function in this form (a sum of exponential factors), then we can identify the points of probability and their probabilities.

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.

(xi) **A mixture of distributions:** Given any finite collection of random variables, $X_1, X_2, \ldots, X_k$ with density or probability functions, say $f_1(x), f_2(x), \ldots, f_k(x)$, where k is a non-negative integer, and given a set of weights, $\alpha_1, \alpha_2, \ldots, \alpha_k$, where $0 \leq \alpha_i \leq 1$ for each i and $\sum_{i=1}^{k} \alpha_i = 1$, it is possible to construct a new density function: $f(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x) + \cdots + \alpha_k f_k(x)$ which is a weighted average of the original density functions. It then follows that the resulting distribution X, whose density/probability function is f, has moments and mgf which are weighted averages of the original distribution moments and mgf:

$E[X^n] = \alpha_1 E[X_1^n] + \alpha_2 E[X_2^n] + \cdots + \alpha_k E[X_k^n]$ and

$M_X(t) = \alpha_1 M_{X_1}(t) + \alpha_2 M_{X_2}(t) + \cdots + \alpha_k M_{X_k}(t)$.

Normal approximation for binomial: If X has a binomial (n, p) distribution, then we can use the approximation Y with a normal $N(np, np(1-p))$ distribution to approximate X in the following way:

With integer correction: $P(n \leq X \leq m) = P(n - 1/2 \leq Y \leq m + 1/2)$ and then convert to a standard normal.

Some properties of the exponential distribution:
Lack of memory property: For $x, y > 0$, $P[X > x + y | X > x] = P[X > y]$

Link between the exponential distribution and Poisson distribution: Suppose that X has an exponential distribution with mean $\frac{1}{\lambda}$ and we regard X as the time between successive occurrences of some type of event, where time is measured in some appropriate units. Now, we imagine that we choose some starting time (say $t = 0$), and from now we start recording times between successive events. Let N represent the number of events that have occurred when one unit of time has elapsed. Then N will be a random variable related to the times of the occurring events. It can be shown that the distribution of N is Poisson with parameter λ.

The minimum of a collection of independent exponential random variables: Suppose that independent random variables $Y_1, Y_2, \ldots, Y_n$ each have exponential distributions with means $\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \ldots, \frac{1}{\lambda_n}$, respectively. Let $Y = \min\{Y_1, Y_2, \ldots, Y_n\}$. Then Y has an exponential distribution with mean $\dfrac{1}{\lambda_1 + \lambda_2 + \cdots + \lambda_n}$.

Joint Distribution of random variables X and Y:
If X and Y are discrete random variables, then $f(x, y) = P[(X = x) \cap (Y = y)]$ is the joint probability function, and it must satisfy:
(i) $0 \leq f(x, y) \leq 1$ and (ii) $\sum_x \sum_y f(x, y) = 1$.
If X and Y are continuous random variables, then $f(x, y)$ must satisfy:
(i) $f(x, y) \geq 0$ and (ii) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dy dx = 1$.

Cumulative distribution function of a joint distribution: If random variables X and Y have a joint distribution, then the cumulative distribution function is $F(x, y) = P[(X \leq x) \cap (Y \leq y)]$.

In the discrete case: $F(x, y) = \sum_{s=-\infty}^{x} \sum_{t=-\infty}^{y} f(s, t)$.

In the continuous case: $F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s, t) dt ds$, and $\dfrac{\partial^2}{\partial x \partial y} F(x, y) = f(x, y)$.

Expectation of a function of jointly distributed random variables: If $h(x, y)$ is a function of two variables, and X and Y are jointly distributed random variables, then the expected value of $h(X, Y)$ is defined to be

$E[h(X, Y)] = \sum_x \sum_y h(x, y) f(x, y)$ in the discrete case, and

$E[h(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) f(x, y) dy dx$ in the continuous case.

Marginal distribution of X found from a joint distribution of X and Y:
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.

Note $F_X(x) = \lim_{y \to \infty} F(x, y)$ and similarly $F_Y(y) = \lim_{x \to \infty} F(x, y)$.

Independence of random variables X and Y: Random variables X and Y with density function $f_X(x)$ and $f_Y(y)$ are said to be independent (or stochastically independent) if the probability space is rectangular ($a \le x \le b, c \le y \le d$, where the endpoints can be infinite) and if the joint density function is of the form $f(x,y) = f_X(x)f_Y(y)$. Independence of X and Y is also equivalent to the factorization of the cumulative distribution function $F(x,y) = F_X(x)F_Y(y)$ for all (x,y).

Conditional distribution of Y given $X = x$: The way in which a conditional distribution is defined follows the basic definition of conditional probability, $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$. In fact, given a discrete joint distribution, this is exactly how a conditional distribution is defined. Also,
$E[X|Y=y] = \sum_x x f_{X|Y}(x|Y=y)$ and
$E[X^2|Y=y] = \sum_x x^2 f_{X|Y}(x|Y=y)$. Then the conditional variance would be,
$Var[X|Y=y] = E[X^2|Y=y] - (E[X|Y=y])^2$.

The expression for conditional probability used in the discrete case is $f_{X|Y}(x|Y=y) = \dfrac{f(x,y)}{f_Y(y)}$. This can be also applied to find a conditional distribution of Y given $X = x$, so that we define $f_{Y|X}(y|X=x) = \dfrac{f(x,y)}{f_X(x)}$.

We also apply this same algebraic form to define the conditional density in the continuous case, with $f(x,y)$ being the joint density and $f_X(x)$ being the marginal density. In the continuous case, the conditional mean of Y given $X = x$ would be

$E[Y|X=x] = \int y f_{Y|X}(y|X=x)dy$, where the integral is taken over the appropriate interval for the conditional distribution of Y given $X = x$.

If X and Y are independent random variables, then $f_{Y|X}(y|X=x) = f_Y(y)$ and similarly $f_{X|Y}(x|Y=y) = f_X(x)$, which indicates that the density of Y does not depend on X and vice-versa.

Note that if the marginal density of X, $f_X(x)$, is known, and the conditional density of Y given $X = x$, $f_{Y|X}(y|X=x)$, is also known, then the joint density of X and Y can be formulated as $f(x,y) = f_{Y|X}(y|X=x)f_X(x)$.

Covariance between random variables X and Y: If random variables X and Y are jointly distributed with joint density/probability function $f(x,y)$, the covariance between X and Y is

$Cov[X,Y] = E[XY] - E[X]E[Y]$.

Note that now we have the following application:

$Var[aX + bY + c] = a^2Var[X] + b^2Var[Y] + 2abCov[X,Y]$.

Coefficient of correlation between random variables X and Y: The coefficient of correlation between random variables X and Y is

$\rho(X,Y) = \rho_{X,Y} = \dfrac{Cov[X,Y]}{\sigma_x \sigma_y}$, where σ_x, σ_y are the standard deviations of X and Y, respectively.

Note that $-1 \le \rho_{X,Y} \le 1$ always!

Moment generating functions 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. And,

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

Some results and formulas:

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

(ii) The density/probability function of jointly distributed variables X and Y can be written in the form $f(x,y) = f_{Y|X}(y|X=x)f_X(x) = f_{X|Y}(x|Y=y)f_Y(y)$.

(iii) $Cov[X,Y] = E[XY] - \mu_x\mu_y = E[XY] - E[X]E[Y]$. $Cov[X,Y] = Cov[Y,X]$. If X, Y are independent then $Cov[X,Y] = 0$.

For constants a, b, c, d, e, f and random variables X, Y, Z, W,

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

(iv) If X and Y are independent then, $Var[X+Y] = Var[X] + Var[Y]$. In general (regardless of independence), $Var[aX + bY + c] = a^2Var[X] + b^2Var[Y] + 2abCov[X,Y]$.

(v) If X and Y have a joint distribution which is uniform (constant density) on the two dimensional region R, then the pdf of the joint distribution is $\frac{1}{\text{Area of } R}$ inside the region R (and 0 outside R). The probability of any event A (where $A \subseteq R$) 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(s) defined by the intersection of the region R with the line $X = x$. The marginal distribution of Y might or might not be uniform.

(vi) $E[h_1(X,Y) + h_2(X,Y)] = E[h_1(X,Y)] + E[h_2(X,Y)]$, and in particular, $E[X+Y] = E[X] + E[Y]$ and $E[\sum X_i] = \sum E[X_i]$.

(vii) $\lim_{x\to-\infty} F(x,y) = \lim_{y\to-\infty} F(x,y) = 0$.

(viii) $P[(x_1 < X \le x_2) \cap (y_1 < Y \le y_2)] = F(x_2,y_2) - F(x_2,y_1) - F(x_1,y_2) + F(x_1,y_1)$.

(ix) $P[(X \le x) \cup (Y \le y)] = F_X(x) + F_Y(y) - F(x,y)$

(x) For any jointly distributed random variables X and Y, $-1 \le \rho_{XY} \le 1$.

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

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

In general, $E[X^rY^s] = \frac{\partial^{r+s}}{\partial^r t_1 \partial^s t_2} M_{X,Y}(t_1,t_2)\Big|_{t_1=t_2=0}$.

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

(xiv) If $Y = aX + b$ then $M_Y(t) = E[e^{tY}] = E[e^{atX+bt}] = e^{bt}M_X(at)$.

(xv) If X and Y are jointly distributed, then $E[E[X|Y]] = E[X]$ and $Var[X] = E[Var[X|Y]] + Var[E[X|Y]]$.

Distribution of a transformation of a continuous random variable X: Suppose that X is a continuous random variable with pdf $f_X(x)$ and cdf $F_X(x)$ and suppose that $u(x)$ is a one-to-one function (usually u is either strictly increasing or strictly decreasing). As a one-to-one function, u has an inverse function v, so that $v(u(x)) = x$. The random variable $Y = u(X)$ is referred to as a **transformation of** X. The pdf of Y can be found in one of two ways:

(i) $f_Y(y) = f_X(v(y))|v'(y)|$

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

$F_Y(y) = P[Y \le y] = P[u(X) \le y] = P[X \le v(y)] = F_X(v(y))$, and then $f_Y(y) = F'_Y(y)$.

Distribution of a transformation of a discrete random variable X: Suppose that X is a discrete random variable with probability function $f(x)$. If $u(x)$ is a function of x, and Y is a random variable defined by the equation $Y = u(X)$, then Y is a discrete random variable with probability function $g(y) = \sum_{y=u(x)} f(x)$. Given a value of y, find all values of x for which $y = u(x)$ (say $u(x_1) = u(x_2) = \cdots = u(x_t) = y$), and then $g(y)$ is the sum of those $f(x_i)$ probabilities.

If X and Y are independent random variables, and u and v are functions, then the random variables $u(X)$ and $v(Y)$ are independent.

Transformation of jointly distributed random variables X and Y: 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 $g(u, v)$. This is a two-variable version of the transformation procedure outlined above. In the one variable case we required that the transformation had an inverse. In the two variable case 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 $g(u, v) = f(h(u, v), k(u, v))\left|\frac{\partial h}{\partial u}\frac{\partial k}{\partial v} - \frac{\partial h}{\partial v}\frac{\partial k}{\partial u}\right|$.

This procedure sometimes arises in the context of being given a joint distribution between X and Y, and being asked to find the pdf of some function $U = u(X, Y)$. In this case, we try to find a second function $v(X, Y)$ that will simplify the process of finding the joint distribution of U and V. Then, after we have found the joint distribution of U and V, we can find the marginal distribution of U.

The distribution of a sum of random variables:

(i) If X_1 and X_2 are random variables, and $Y = X_1 + X_2$, then $E[Y] = E[X_1] + E[X_2]$ and $Var[Y] = Var[X_1] + Var[X_2] + 2Cov[X_1, X_2]$.

(ii) If X_1 and X_2 are discrete non-negative integer valued random variables with joint probability function $f(x_1, x_2)$, then for any integer $k \ge 0$,

$P[X_1 + X_2 = k] = \sum_{x_1=0}^{k} f(x_1, k - x_1)$ (this considers all combinations of X_1 and X_2 whose sum is k).

If X_1 and X_2 are independent with probability functions $f_1(x_1), f_2(x_2)$ respectively, then,

$P[X_1 + X_2 = k] = \sum_{x_1=0}^{k} f_1(x_1)f_2(k - x_1)$ (this is the **convolution method** of finding the distribution of the sum of independent discrete random variables).

If X_1 and X_2 are continuous random variables with joint probability function $f(x_1, x_2)$, then the density function of $Y = X_1 + X_2$ is $f_Y(y) = \int_{-\infty}^{\infty} f(x_1, y - x_1)dx_1$

If X_1 and X_2 are independent continuous random variables with density functions $f_1(x_1)$ and $f_2(x_2)$, then the density function of $Y = X_1 + X_2$ is $f_Y(y) = \int_{-\infty}^{\infty} f_1(x_1) f_2(y - x_1) dx_1$ (this is the continuous version of the convolution method).

(iv) If $X_1, X_2, \ldots, X_n$ are random variables, and the random variable Y is defined to be $Y = \sum_{i=1}^{n} X_i$, then $E[Y] = \sum_{i=1}^{n} E[X_i]$ and $Var[Y] = \sum_{i=1}^{n} Var[X_i] + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} Cov[X_i, X_j]$.

If $X_1, X_2, \ldots, X_n$ are mutually independent random variables, then $Var[Y] = \sum_{i=1}^{n} Var[X_i]$ and $M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t) = M_{X_1}(t) M_{X_2}(t) \cdots M_{X_n}(t)$.

(v) If $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ are random variables and $a_1, a_2, \ldots, a_n, b, c_1, c_2, \ldots, c_m$, and d are constants, then

$$Cov[\sum_{i=1}^{n} a_i X_i + b, \sum_{j=1}^{m} c_j Y_j + d] = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i c_j Cov[X_i, Y_j]$$

(vi) **The Central Limit Theorem:** Suppose that X is a random variable with mean μ and standard deviation σ and suppose that $X_1, X_2, \ldots, X_n$ are n independent random variables with the same distribution as X. Let $Y_n = X_1 + X_2 + \cdots + X_n$. Then $E[Y_n] = n\mu$ and $Var[Y_n] = n\sigma^2$, and as n increases, the distribution of Y_n approaches a normal distribution $N(n\mu, n\sigma^2)$.

(vii) **Sums of certain distributions:** Suppose that $X_1, X_2, \ldots, X_k$ are independent random variables and $Y = \sum_{i=1}^{k} X_i$

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 $N(\mu_i, \sigma_i^2)$	$N(\sum \mu_i, \sum \sigma_i^2)$
exponential with mean μ	gamma with $\alpha = k$, $\beta = 1/\mu$
gamma with α_i, β	gamma with $\sum \alpha_i, \beta$
Chi-square with k_i df	Chi-square with $\sum k_i$ df

The distribution of the maximum or minimum of a collection of independent random variables: 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 $F_1(x) = P(X_1 \leq x)$ and $F_2(x) = P(X_2 \leq x)$, respectively. We can formulate the distribution functions of U and V in terms of F_1 and F_2 as follows

$$F_U(u) = F_1(u)F_2(u) \quad \text{and} \quad F_V(v) = 1 - [1 - F_1(v)][1 - F_2(v)]$$

This can be generalized to n independent random variables $(X_1, \ldots, X_n)$. Here we would have

$$F_U(u) = F_1(u)F_2(u)\cdots F_n(u) \quad \text{and} \quad F_V(v) = 1 - [1 - F_1(v)][1 - F_2(v)]\cdots[1 - F_n(v)].$$

Order statistics: For a random variable X, a **random sample of size** n is a collection of n independent X_i's all having the same distribution as X. Now let us define new random variables $Y_1, Y_2, \ldots, Y_n$ where $Y_1 = \min\{X_1, X_2, \ldots, X_n\}$, Y_2 is the second smallest element in the random sample, and so on, until $Y_n = \max\{X_1, X_2, \ldots, X_n\}$.
The probability density function (pdf) of Y_k can be described as follows:

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

The joint density of $Y_1, \ldots, Y_n$ is $g(y_1, \ldots, y_n) = n!f(y_1)\cdots f(y_n)$.

Mixtures of Distributions: Suppose that X_1 and X_2 are random variables with density (or probability) functions $f_1(x)$ and $f_2(x)$, and suppose a is a number with $0 < a < 1$. We define a new random variable X by defining a new density function $f(x) = af_1(x) + (1 - a)f_2(x)$. This newly defined density function will satisfy the requirements for being a properly defined density function. Furthermore, all moments, probabilities and moment generating function of the newly defined random variable are of the form:

$E[X] = aE[X_1] + (1 - a)E[X_2]$, $E[X^2] = aE[X_1^2] + (1 - a)E[X_2^2]$,

$F_X(x) = P(X \le x) = aP(X_1 \le x) + (1 - a)P(X_2 \le x) = aF_1(x) + (1 - a)F_2(x)$,

$M_X(t) = aM_{X_1}(t) + (1 - a)M_{X_2}(t)$, $Var[X] = E[X^2] - (E[X])^2$.

Loss Distributions and Insurance
Let X be a loss random variable. Then $E[X]$ is referred to as the **pure premium** or **expected claim**.

The **unitized risk** or **coefficient of variation** for the random variable X is defined to be
$\frac{\sqrt{Var[X]}}{E[X]} = \frac{\sigma}{\mu}$.

Models for a loss random variable X:
Case 1: The complete distribution of X is given. $X = K$ with probability q, $X = 0$ with probability $1 - q$.

Case 2: The probability q of a non-negative loss is given, and the **conditional distribution** B **of loss amount given that a 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 $X = B$.

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

Keep in mind that B is the loss amount **given that a loss has occurred**, whereas X is the overall loss amount.

Modeling the aggregate claims in a portfolio of insurance policies: 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 ways 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

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

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

The normal approximation to aggregate claims: For an aggregate claims distribution S, if the mean and variance of S are known $(E[S], Var[S])$, it is possible to approximate probabilities for S by using the normal approximation. The 95-th percentile of aggregate claims is the number Q for which $P(S \le Q) = .95$. If S is assumed to have a distribution which is approximately normal, then by standardizing S we have

$P[S \le Q] = P\left[\frac{S - E[S]}{\sqrt{Var[S]}} \le \frac{Q - E[S]}{\sqrt{Var[S]}}\right] = .95$, so that $\frac{Q - E[S]}{\sqrt{Var[S]}}$ is equal to the 95-th percentile of the standard normal distribution (which is 1.645), so that Q can be found; $Q = E[S] + 1.645\sqrt{Var[S]}$. If the insurer collects total premium of amount Q, then (assuming that it is reasonable to use the approximation) there is a 95% chance (approx.) that aggregate claims will be less than the premium collected, and there is a 5% chance that aggregate claims will exceed the premium. Since S is a sum of many independent individual policy loss random variables, the Central Limit Theorem suggests that the normal approximation is not unreasonable.

Partial Insurance Coverage

(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 $Y = X - d$ if $X > d$ and 0 if $X \le d$, so $Y = \max\{X - d, 0\}$. This is also denoted $(X - d)_+$. The expected payment made by the insurer when a loss occurs would be $\int_d^\infty (x - d) f(x) dx$ in the continuous case (this is also called the **expected cost per loss**). The above integral is also equal to $\int_d^\infty [1 - F_X(x)] dx$. This type of policy is also referred to as an ordinary deductible insurance.

(ii) **Policy Limit:** A **policy limit of amount** u indicates that the insurer will pay a maximum amount of u when a loss occurs. Therefore, the amount paid by the insurer is X if $X \le u$ and u if $X > u$. The expected payment made by the insurer per loss would be $\int_0^u f_X(x) dx + u[1 - F_X(u)]$ in the continuous case. This is also equal to $\int_0^u [1 - F_X(x)] dx$.

Note: if you were to create a RV $X = Y_1 + Y_2$ where Y_1 has deductible c, and Y_2 has policy limit c, then the sum would just be the total loss amount X.