SECTION 9 - FUNCTIONS AND TRANSFORMATIONS OF RANDOM VARIABLES

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, such as $u(x) = e^x$, $\sqrt{x}$ or $ln\, x$, or u is strictly decreasing, such as $u(x) = e^{-x}$ or $\frac{1}{x}$). 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 (they are actually equivalent):

(i) $\boldsymbol{f_Y(y) = f_X(v(y)) \cdot |v'(y)|}$, or

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

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.

Example 9-1:

The random variable X has an exponential distribution with a mean of 1. The random variable Y is defined to be $Y = 2\,ln\,X$. Find $f_Y(y)$, the pdf of Y.

Solution:

$F_Y(y) = P[Y \le y] = P[2\,ln\,X \le y] = P[X \le e^{y/2}]$.

We can now use the cdf of X, $F_X(t) = 1 - e^{-t}$, so that
$F_Y(y) = P[X \le e^{y/2}] = F_X(e^{y/2}) = 1 - e^{-e^{y/2}}$.

Then $f_Y(y) = F'_Y(y) = \frac{d}{dy}\,(1 - e^{-e^{y/2}}) = \frac{1}{2}e^{y/2} \times e^{-e^{y/2}}$.

Alternatively, $Y = 2\,ln\,X$. We see that $y = 2\,ln\,x$ is a strictly increasing function of x with inverse function $x = v(y) = e^{y/2}$ and $X = e^{Y/2}$. It follows that
$f_Y(y) = f_X(v(y)) \times |v'(y)| = f_X(e^{y/2}) \times \left|\frac{d}{dy}\,e^{y/2}\right| = e^{-e^{y/2}} \times \frac{1}{2}e^{y/2}$. □

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 on the previous page. In the one variable case we required that the transformation had an inverse. There is a similar requirement 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

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

The factor $|\frac{\partial h}{\partial u} \times \frac{\partial k}{\partial v} - \frac{\partial h}{\partial v} \times \frac{\partial k}{\partial u}|$ is referred to as the Jacobian of the transformation.

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.

Example 9-2:

Suppose that X and Y are independent exponential random variables, each with mean 1. Suppose that $U = \frac{Y}{X}$ and $V = X$. Find the joint distribution of U and V and the marginal distribution of U.

Solution:

$U = u(X,Y) = \frac{Y}{X}$ and $V = v(X,Y) = X$, so that $u(x,y) = \frac{y}{x}$ and $v(x,y) = x$. We can invert these transformations in the following way. $x = v = h(u,v)$, and $y = \frac{y}{x} \cdot x = u \cdot v = k(u,v)$. Since X and Y are independent, the joint density of X and Y is $f(x,y) = f_X(x) \times f_Y(y) = e^{-x} \times e^{-y} = e^{-(x+y)}$. According to the two-variable transformation method outlined above, the joint density of U and V is

$$\begin{aligned} g(u,v) &= f(h(u,v), k(u,v)) \times |\frac{\partial h}{\partial u} \times \frac{\partial k}{\partial v} - \frac{\partial h}{\partial v} \times \frac{\partial k}{\partial u}| = f(v, uv) \times |0 \times u - 1 \times v| \\ &= e^{-(v+uv)} \times v = ve^{-v(u+1)} \text{ (since } \frac{\partial h}{\partial u} = 0 \text{, } \frac{\partial k}{\partial v} = u \text{, } \frac{\partial h}{\partial v} = 1 \text{ and } \frac{\partial k}{\partial u} = v). \end{aligned}$$

We also note that X and Y are defined on the region $x > 0$ and $y > 0$, so U and V are defined on the region $u > 0$ and $v > 0$. The marginal density of U is

$$g_U(u) = \int_0^\infty g(u,v)\,dv = \int_0^\infty ve^{-v(u+1)}\,dv = -\frac{ve^{-v(u+1)}}{(u+1)} - \frac{e^{-v(u+1)}}{(u+1)^2}\Big|_{v=0}^{v=\infty} = \frac{1}{(u+1)^2}$$

We used the integration by parts rule $\int ve^{-av}\,dv = -\frac{ve^{-av}}{a} - \frac{e^{-av}}{a^2}$, with $a = u+1$ and the fact that $\lim_{v\to\infty} \frac{ve^{-v(u+1)}}{(u+1)} = 0$. □

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
$\boldsymbol{E[Y] = E[X_1] + E[X_2]}$ and $\boldsymbol{Var[Y] = Var[X_1] + Var[X_2] + 2Cov[X_1, X_2]}$

(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 an integer $k \geq 0$
$P[X_1 + X_2 = k] = \sum_{x_1=0}^{k} f(x_1, k - x_1)$ (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 $f_1(x_1)$ and $f_2(x_2)$, respectively, then
$P[X_1 + X_2 = k] = \sum_{x_1=0}^{k} f_1(x_1)\text{G}f_2(k - x_1)$ (this is the **convolution method** of finding the distribution of the sum of independent discrete random variables).

(iii) If X_1 and X_2 are continuous random variables with joint density 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 $f_Y(y) = \int_{-\infty}^{\infty} f_1(x_1) \times f_2(y - x_1)\, dx_1$.
If $X_1 \geq 0$ and $X_2 \geq 0$, then $f_Y(y) = \int_0^y f(x_1, y - x_1)\, dx_1$.
This is the continuous version of the convolution method.

(iv) 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 $E[Y] = \sum_{i=1}^{n} E[X_i]$ 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 $\boldsymbol{X_1, X_2, ..., X_n}$ are mutually independent random variables, then

$$\boldsymbol{Var[Y] = \sum_{i=1}^{n} Var[X_i]} \text{ and } \boldsymbol{M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t) = M_{X_1}(t) \times M_{X_2}(t) \times \cdots \times M_{X_n}(t)}$$

and $\boldsymbol{P_Y(t) = \prod_{i=1}^{n} P_{X_i}(t) = P_{X_1}(t) \times P_{X_2}(t) \times \cdots \times P_{X_n}(t)}$

(v) If $X_1, X_2, ..., X_n$ and $Y_1, Y_2, ..., Y_m$ are random variables and $a_1, a_2, ..., a_n$, b, $c_1, c_2, ..., c_m$ 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 $X_1, X_2, ..., X_n$ 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 $N(n\mu, n\sigma^2)$. This is a justification for using the normal distribution as an approximation to the distribution of a sum of random variables. **When an exam question asks for a probability involving a sum of a large number of independent random variables, it is usually asking for the normal approximation to be applied. As mentioned earlier in Section 7 of these notes, when**

applying the normal approximation to an integer random variable, we may be asked to use the integer correction.

(vii) Sums of certain distributions: Suppose that $X_1, X_2, ..., X_k$ 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 $B(\Sigma n_i, p)$
Poisson λ_i	Poisson $\Sigma\lambda_i$
geometric p	negative binomial k, p
negative binomial r_i , p	negative binomial Σr_i , p
normal $N(\mu_i, \sigma_i^2)$	$N(\Sigma\mu_i, \Sigma\sigma_i^2)$
exponential with mean μ	gamma with $\alpha = k$, $\beta = 1/\mu$
gamma with α_i , β	gamma with $\Sigma\alpha_i$, β
Chi-square with k_i df	Chi-square with Σk_i df

Example 9-3:

Suppose that X and Y are independent discrete integer-valued random variables with X uniformly distributed on the integers 1 to 5, and Y having the following probability function: $f_Y(0) = 0.3$, $f_Y(1) = 0.5$, $f_Y(3) = 0.2$. Let $Z = X + Y$. Find $P[Z = 5]$.

Solution:

Using the fact that $f_X(x) = 0.2$ for $x = 1, 2, 3, 4, 5$, and using the convolution method for independent discrete random variables, we have

$$f_Z(5) = \sum_{i=1}^{5} f_X(i) \times f_Y(5-i) \ = 0.2 \times 0 + 0.2 \times 0.2 + 0.2 \times 0 + 0.2 \times 0.5 + 0.2 \times 0.3 = 0.20$$

□

Example 9-4:

X_1 and X_2 are independent exponential random variables each with a mean of 1. Find $P[X_1 + X_2 < 1]$.

Solution:

Using the convolution method, the density function of $Y = X_1 + X_2$ is

$$f_Y(y) = \int_0^y f_{X_1}(t) \times f_{X_2}(y-t)\, dt = \int_0^y e^{-t} \times e^{-(y-t)} dt = ye^{-y} \text{ , so that}$$
$$P[X_1 + X_2 < 1] = P[Y < 1] = \int_0^1 ye^{-y} dy = [-ye^{-y} - e^{-y}]\Big|_{y=0}^{y=1} = 1 - 2e^{-1}$$

(the last integral required integration by parts). □

For trial use only. Content may not represent the most current version

Example 9-5:

Given n independent random variables $X_1, X_2, ..., X_n$ each having the same variance of σ^2, and defining
$U = 2X_1 + X_2 + \cdots + X_{n-1}$ and
$V = X_2 + X_3 + \cdots + 2X_n$, find the coefficient of correlation between U and V.

Solution:

$\rho_{UV} = \frac{Cov[U,V]}{\sigma_U \sigma_V}$; $\sigma_U^2 = (4+1+1+\cdots+1)\sigma^2 = (n+2)\sigma^2 = \sigma_V^2$.

Since the X's are independent, if $i \neq j$ then $Cov[X_i, X_j] = 0$. Then, noting that
$Cov[W,W] = Var[W]$, we have

$Cov[U,V] = Cov[2X_1, X_2] + Cov[2X_1, X_3] + \cdots + Cov[X_{n-1}, 2X_n]$
$= Var[X_2] + Var[X_3] + \cdots + Var[X_{n-1}] = (n-2)\sigma^2$.

Then, $\rho_{UV} = \frac{(n-2)\sigma^2}{(n+2)\sigma^2} = \frac{n-2}{n+2}$. □

Example 9-6:

Independent random variables X, Y and Z are identically distributed. Let $W = X + Y$. The moment generating function of W is $M_W(t) = (.7 + .3e^t)^6$.

Find the moment generating function of $V = X + Y + Z$.

Solution:

For independent random variables, $M_{X+Y}(t) = M_X(t) \times M_Y(t) = (0.7 + 0.3e^t)^6$.

Since X and Y have identical distributions, they have the same moment generating function.

Thus, $M_X(t) = (0.7 + 0.3e^t)^3$, and then $M_V(t) = M_X(t) \times M_Y(t) \times M_Z(t) = (0.7 + 0.3e^t)^9$.

Alternatively, note that the moment generating function of the binomial $B(n,p)$ is
$(1 - p + pe^t)^n$. Thus, $X + Y$ has a $B(6, 0.3)$ distribution, and each of X, Y and Z has
a $B(3, 0.3)$ distribution, so that the sum of these independent binomial distributions is
$B(9, 0.3)$, with mgf $(0.7 + 0.3e^t)^9$. □

Example 9-7:

The birth weight of males is normally distributed with mean 6 pounds, 10 ounces, standard deviation 1 pound. For females, the mean weight is 7 pounds, 2 ounces with standard deviation 12 ounces. Given two independent male/female births, find the probability that the baby boy outweighs the baby girl.

Solution:

Let random variables X and Y denote the boy's weight and girl's weight, respectively. Then, $W = X - Y$ has a normal distribution with mean $6\frac{10}{16} - 7\frac{2}{16} = -\frac{1}{2}$ lb. and variance $\sigma_X^2 + \sigma_Y^2 = 1 + \frac{9}{16} = \frac{25}{16}$.

Then, $P[X > Y] = P[X - Y > 0] = P\left[\frac{W-(-\frac{1}{2})}{\sqrt{25/16}} > \frac{-(-\frac{1}{2})}{\sqrt{25/16}}\right] = P[Z > 0.4]$,

where Z has standard normal distribution (W was standardized). Referring to the standard normal table, this probability is 0.34. □

Example 9-8:

If the number of typographical errors per page typed by a certain typist follows a Poisson distribution with a mean of λ, find the probability that the total number of errors in 10 randomly selected pages is 10.

Solution:

The 10 randomly selected pages have independent distributions of errors per page. The sum of m independent Poisson random variables with parameters $\lambda_1, \lambda_2 ...., \lambda_m$ has a Poisson distribution with parameter $\Sigma\lambda_i$. Thus, the total number of errors in the 10 randomly selected pages has a Poisson distribution with parameter 10λ. The probability of 10 errors in the 10 pages is $\frac{e^{-10\lambda}\times(10\lambda)^{10}}{10!}$. □

Example 9-9:

Smith estimates his chance of winning a particular hand of blackjack at a casino is 0.45, his probability of losing is 0.5, and his probability of breaking even on a hand is 0.05. He is playing at a \$10 table, which means that on each play, he either wins \$10, loses \$10 or breaks even, with the stated probabilities. Smith plays 100 times. What is the approximate probability that he has won money on the 100 plays of the game in total? Solve with and without an integer correction.

Solution:

Suppose that X is the gain on a particular play of the game. Then

$E[X] = 10 \times 0.45 + (-10) \times 0.5 = -0.5$ is his expected gain on each play, and

$E[X^2] = 100 \times 0.45 + 100 \times 0.5 = 95 \Rightarrow Var[X] = E[X^2] - (E[X])^2 = 94.75.$

$W = \sum_{i=1}^{100} X_i \rightarrow E[W] = -50, Var[W] = 9475.$

The use of "approximate" in the context of the sum of a large number of independent random variables (the X's) indicates that we are to apply the normal approximation to find the probability.

$P[W > 0] = P\left[\frac{W-E[W]}{\sqrt{Var[W]}} > \frac{0-E[W]}{\sqrt{Var[W]}}\right]$. We assume that W has an approximate normal distribution. Then

$P[W > 0] = P\left[\frac{W-E[W]}{\sqrt{Var[W]}} > \frac{0-E[W]}{\sqrt{Var[W]}}\right] = P\left[Z > \frac{50}{\sqrt{9475}}\right]$

$= P[Z > 0.51] = 1 - P[Z \leq 0.51] = 0.305$. This is the solution without an integer correction.

Since the amount won (or lost) must be a multiple of 10, the integer correction form for $P[W > 0]$ is $P[W > 5]$. This is because each possible multiple of 10 is replaced by an interval of length 10 centered at that point. So, for example, we could find the approximate probability $P[W = 20]$ as $P[15 < W \leq 25]$ and apply the normal approximation to that interval. In the case being considered,

$P[W > 5] = P\left[Z > \frac{55}{\sqrt{9475}}\right] = P[Z > 0.565] = 0.286.$ □

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 $F_1(x) = P[X_1 \le x]$ and $F_2(x) = P[X_2 \le x]$, respectively.

We can formulate the distribution functions of U and V in terms of F_1 and F_2 as follows.

$F_U(u) = P[U \le u] = P[max\{X_1\ ,\ X_2\} \le u] = P[(X_1 \le u) \cap (X_2 \le u)]$

(if the larger of X_1 and X_2 is $\le u$, then so is the smaller one, so both are $\le u$).

Since X_1 and X_2 are independent, we have

$P[(X_1 \le u) \cap (X_2 \le u)] = P[X_1 \le u] \times P[X_2 \le u] = F_1(u) \times F_2(u)$.

Therefore, the distribution function of U is $F_U(u) = F_1(u) \times F_2(u)$.

$F_V(v) = P[V \le v] = 1 - P[V > v] = 1 - P[min\{X_1\ ,\ X_2\} > v] = 1 - P[(X_1 > v) \cap (X_2 > v)]$

(if the smaller of X_1 and X_2 is $> v$, then so is the larger one, so both are $> v$).

Since X_1 and X_2 are independent, we have

$P[(X_1 > v) \cap (X_2 > v)] = P[X_1 > v] \times P[X_2 > v] = [1 - F_1(v)] \times [1 - F_2(v)]$.

Therefore, the distribution function of V is $F_V(v) = 1 - [1 - F_1(v)] \times [1 - F_2(v)]$.

Example 9-10:

A homeowner is accepting sealed bids from two prospective buyers on their offering price to purchase his home. The homeowner assumes that the two bidders will formulate their bids independently of one another. The homeowner assumes a probability distribution for the bid that will be offered by each of the two bidders. For one of the bidders, the homeowner assumes that the bid will be uniformly distributed between 100,000 and 120,000. For the other bidder, the homeowner assumes that the bid will be uniformly distributed between 90,000 and 140,000. Find the probability that the larger of the two bids is over 110,000.

Solution:

Let us denote the two bids as X_1 and X_2, so that X_1 has a uniform distribution on the interval $(100,000\ ,$ $120,000)$, and X_2 has a uniform distribution on the interval
$(90,000\ ,\ 140,000)$. The distribution function of X_1 and X_2 are

$F_1(x) = \frac{x-100,000}{20,000}$ for $100,000 < x < 120,000$

$F_2(x) = \frac{x-90,000}{50,000}$ for $90,000 < x < 140,000$.

The larger of the two bids is $U = max\{X_1\ ,\ X_2\}$. Then

$P[U > 110,000] = 1 - P[U \le 110,000] = 1 - P[(X_1 \le 110,000) \cap (X_2 \le 110,000)]$

$= 1 - F_1(110,000) \times F_2(110,000) = 1 - \frac{110,000-100,000}{20,000} \times \frac{110,000-90,000}{50,000}$

$= 1 - \frac{1}{2} \times \frac{2}{5} = \frac{4}{5}$ □

It is possible to extend the case of the max or min of two random variables to the max or min of any collection of independent random variables. For instance, if $X_1X_2, ..., X_n$ are independent random variables with cdf's $F_1(x)$, $F_2(x)$, $\ldots$, $F_n(x)$, and if $U = max\{X_1X_2, ..., X_n\}$,
then the cdf of U is

$$\begin{aligned} F_U(u) &= P[U \leq u] = P[max\{X_1X_2, ..., X_n\} \leq u] \\ &= P[(X_1 \leq u) \cap (X_2 \leq u) \cap \cdots \cap (X_n \leq u)] = F_1(u) \cdot F_2(u) \cdots F_n(u) \ . \end{aligned}$$

If $V = min\{X_1X_2, ..., X_n\}$, then the cdf of V is

$$\begin{aligned} F_V(v) &= P[V \leq v] = 1 - P[V > v] = 1 - P[min\{X_1X_2, ..., X_n\} > v] \\ &= 1 - P[(X_1 > v) \cap (X_2 > v) \cap \cdots \cap (X_n > v)] \\ &= 1 - [1 - F_1(v)] \cdot [1 - F_2(v)] \cdots [1 - F_n(v)] \ . \end{aligned}$$

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. For instance, if X is the outcome that results from tossing a fair die, and the die is tossed independently 10 times, then the outcomes $X_1, X_2, ..., X_{10}$ form a random sample of size 10. We can think of the X_i's as 10 separate independent random variables (when we actually toss the die, we will have 10 numerical outcomes, but in advance of tossing the die we can still think of the outcomes as random variables). When we toss the die 10 times, we will get values between 1 and 6, and they will occur in a random order. For instance, the outcomes might be 5 , 2 , 4 , 4 , 1 , 5 , 2 , 6 , 3 , 1.

Suppose in advance of actually tossing the die, we decide that we will summarize the 10 outcomes by placing them in increasing order. So the first actual outcome X_1 might not be the smallest numerical outcome, etc. We define 10 new variables Y_1 , Y_2 , $\ldots$, Y_{10}, so that the Y's are the same collection of numbers as the X's, but they have been put in increasing order.
Y_1 is the smallest of the X's, Y_2 is the next smallest, . . . , Y_{10} is the largest.
In general Y_k is the k-th from the smallest of the X_i's.

We can imagine that we will do this even before the die is actually tossed, so that we can think of the Y's as random variables as well. In fact, Y_1 is just the minimum of the X's, $Y_1 = min\{X_1, X_2, ..., X_{10}\}$, and Y_{10} is the maximum of the X's (and we also have the Y's that are in between).

We saw in the previous example and comments how to find the distribution of the max and the min of a collection of independent random variables, and that would apply to Y_1 and Y_{10}. The Y_i's that we get in this procedure are called the order statistics of the random sample of X's.

In Example 9-10 we had X_1 and X_2 with different distributions. We are assuming now that although the X_i's are independent, they all have the same distribution (such as the outcome of tossing a die), say with density function $f(x)$ and distribution function $F(x)$.

We now wish to describe the distribution of each of the order statistics $Y_1, Y_2, ..., Y_n$. The density function of Y_k can be described in terms $f(x)$ and $F(x)$, the density function and distribution function of X. For each $k = 1, 2, ..., n$ the pdf of Y_k is

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

We will not give the general derivation of this density, but the derivation of the density $g_1(t)$ for Y_1 is not difficult to find. If we consider Y_1(the "first order" statistic of the sample of $X's$), its pdf according to the expression above with $k = 1$ is $g_1(t) = n[1-F(t)]^{n-1} \times f(t)$. We saw on the previous page that the cdf of the minimum of $X_1 X_2, ..., X_n$ (it was called V) was

$F_V(v) = 1 - [1 - F_1(v)] \times [1 - F_2(v)] \times \cdots \times [1 - F_n(v)]$.

Since V is the first order statistic, $Y_1 = V$ and then

$F_{Y_1}(t) = 1 - [1 - F_1(t)] \times [1 - F_2(t)] \times \cdots \times [1 - F_n(t)] = 1 - [1 - F(t)]^n$

(since each F_i is the cdf of X). Then the pdf of Y_1 is

$g_1(t) = \frac{d}{dt} F_{Y_1}(t) = \frac{d}{dt}(1 - [1 - F(t)]^n) = n[1 - F(t)]^{n-1} \times f(t)$.

Since Y_k is one of the X's, it takes on the same possible values as X, so the probability space for each Y is the same as the probability space for X.

The largest order statistic Y_n is the same as the random variable $U = max\{X_1 X_2, ..., X_n\}$ described on the previous page. The cdf of Y_n is $F_{Y_n}(t) = [F(t)]^n$, and the pdf is

$$f_{Y_n}(t) = \frac{d}{dt} F_{Y_n}(t) = \frac{d}{dt}[F(t)]^n = n\,[F(t)]^{n-1} f(t)$$

which can be found from the general form of the pdf of Y_k noted above.

For the other order statistics, Y_2 , Y_3 , $\ldots Y_{n-1}$, the cdf's tend to be more complicated (but we do have the pdf of $g_k(t)$ of Y_k for $k = 1, 2, ..., n$ described above). It is possible to formulate the joint distribution of the order statistics Y_1 , Y_2 , $\ldots Y_n$. The joint density of $Y_1, Y_2, ..., Y_n$ is

$$g(y_1, y_2, ..., y_n) = n! \times f(y_1) \times f(y_2) \times \cdots \times f(y_n)$$

Example 9-11:

An airport shuttle service driver is waiting for three passengers to arrive. The passengers will be arriving on three separate flights. The shuttle driver assumes that the times until arrival of the three flights are independent of one another, but each time until arrival has an exponential distribution (as measured from now) with a mean of 1 hour. Find the expected time until the 2nd arriving flight.

Solution:

We let $X_1,\ X_2$ and X_3 be the three arrival times. The time until the 2nd arriving flight is Y_2, the second order statistics of the 3 X's. We wish to find the expected value of Y_2. The pdf of each X is $f(t) = e^{-t}$, and the cdf of each X is $F(t) = 1 - e^{-t}$. The pdf of Y_2 can be found from the general form described earlier: $n = 3$, $k = 2$

$$\begin{aligned} g_2(t) &= \tfrac{3!}{(2-1)!(3-2)!}\,[F(t)]^{2-1} \times [1 - F(t)]^{3-2} \cdot \times (t) \\ &= 6(1 - e^{-t}) \times e^{-t} \times e^{-t} = 6(e^{-2t} - e^{-3t})\ ,\ t > 0 \end{aligned}$$

The expected value of Y_2 is

$$\begin{aligned} E[Y_2] = &\int_0^\infty t\, g_2(t)\, dt = \int_0^\infty t \times 6(e^{-2t} - e^{-3t})\, dt = 6[\int_0^\infty te^{-2t}\, dt - \int_0^\infty te^{-3t}\, dt] \\ &= 6\left[-\tfrac{te^{-2t}}{2} - \tfrac{e^{-2t}}{4}\Big|_{t=0}^{t=\infty} - \left(-\tfrac{te^{-3t}}{3} - \tfrac{e^{-3t}}{9}\Big|_{t=0}^{t=\infty} \right) \right] = 6(\tfrac{1}{4} - \tfrac{1}{9}) = \tfrac{5}{6} \end{aligned}$$

The reader might recall that near the end of Section 7 of this study guide there was a summary of some properties of the exponential distribution. In particular, it was pointed out that the minimum of a collection of independent exponential random variables is also exponential. In this example, the order statistic is Y_1, the minimum of three independent exponential random variables, each with a mean of 1. According to the comments in Section 7 (and also, using the methods of order statistics developed in this section) the distribution of Y_1 will be exponential with a mean of $\frac{1}{3}$. We can expect the first flight arrival to occur in 20 minutes.

Also, recall the exponential integration formula, for an integer $k \geq 0$, $\int_0^\infty t^k \cdot e^{-ct}\, dt = \frac{k!}{c^{k+1}}$.
This can be used to calculate the integrals above. □

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 $f(x) = a \times f_1(x) + (1-a) \times f_2(x)$. 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 variable are of the following "weighted-average" form:

$E[X] = aE[X_1] + (1-a)E[X_2]\ ,\ E[X^2] = aE[X_1^2] + (1-a)E[X_2^2]$
$F_X(x) = P[X \leq x] = aP[X_1 \leq x] + (1-a)P[X_2 \leq x] = aF_1(x) + (1-a)F_2(x)$
$M_X(t) = aM_{X_1}(t) + (1-a)M_{X_2}(t)$

The random variable X is called a mixture of X_1 and X_2, and a and $1-a$ are referred to as mixing weights. As mentioned in Section 5, this notion of mixture can be extended to a mixture of any number of random variables. One place where this "weighted average" relationship **does not work** is in the formulation of the variance of X.
WE DO NOT USE $Var[X] = aVar[X_1] + (1-a)Var[X_2]$ **, it is incorrect**.

We must use the earlier relationship above to get the second and first moments of X, and then $\boldsymbol{Var[X] = E[X^2] - (E[X])^2}$, and we would find $E[X]$ and $E[X^2]$ using the weighted-average approach described above.

Another point to note is the following. It appears that the mixture random variable X is equal to $aX_1 + (1-a)X_2$. **This is incorrect.** X is **not** a sum of random variables. X is totally defined by the definition of the pdf $f(x) = a \times f_1(x) + (1-a) \times f_2(x)$.

A special case of a mixture occurs when X_1 is the constant 0. This situation can be described in the following way. Suppose there is probability a that a loss does not occur, and probability $1-a$ that a loss does occur, and if the loss does occur, the loss amount is a random variable X_2.

The overall loss amount is $X = \begin{cases} 0 & \text{if loss does not occur, prob. } a \\ X_2 & \text{if loss does occur, prob. } 1-a \end{cases}$.

This is a mixture of the constant "random" variable X_1 which is always 0, and the loss random variable X_2, with mixing weight a applied to 0 and mixing weight $1-a$ applied to X_2.
Then the expected value of X will be $a(0) + (1-a)E[X_2] = (1-a)E[X_2]$
and the second moment of X will be $a(0^2) + (1-a)E[X_2^2] = (1-a)E[X_2^2]$.

Example 9-12:

Suppose there are two urns containing balls. Urn I contains 5 red and 5 blue balls and Urn II contains 8 red and 2 blue balls. A die is tossed, and if the number turning up is even then a ball is picked from Urn I, and if the number turning up is odd then a ball is picked from Urn II. X is the number of red balls chosen (0 or 1). We can formulate the distribution of X as a mixture of X_1 and X_2, where random variable X_1 is the number of red balls chosen from Urn I and X_2 is the number of red balls chosen from Urn II. Since each urn is equally likely to be chosen, the mixing weights are $a = 0.5$, $1-a = 0.5$. Then

$P[X=1] = aP[X_1=1] + (1-a)P[X_2=1] = 0.5 \times 0.5 + 0.5 \times 0.8 = 0.65$, and
$P[X=0] = aP[X_1=0] + (1-a)P[X_2=0] = .5 \times 0.5 + .5 \times 0.2) = 0.35$ □

Example 9-13:

An insurer has three risk classifications for policies: low, medium and high. 25% of the company's policies are low risk, 70% are medium risk and 5% are high risk. An individual policy loss is exponentially distributed with the following mean: low risk has mean 1, medium risk has mean 2 and high risk has mean 5. A policy is chosen from the insurer's portfolio of policies, but the risk class is not known. Find the expected loss that will be experienced by the policy, and find the probability that the policy will experience a loss of at least 1.

Solution:

We define three loss random variables. X_1 (low risk) has an exponential distribution with a mean of 1, X_2 (medium risk) has an exponential distribution with a mean of 2, and X_3 (high risk) has an exponential distribution with a mean of 5.

Since there is a 25% chance that the chosen policy is low risk, and a 70% chance that it is medium risk and a 5% chance that it is high risk, the distribution of the loss from the chosen policy is a mixture of X_1, X_2 and X_3 , with mixing weights of .25 applied to X_1, .70 applied to X_2 and .05 applied to X_3 . The pdf of X is
$f(x) = .25f_1(x) + .70f_2(x) + .05f_3(x) = .25 \times e^{-x} + .70 \times \frac{1}{2}e^{-x/2} + .05 \times \frac{1}{5}e^{-x/5}$.

The expected value of X is
$E[X] = 0.25 \times E[X_1] + 0.70 \times E[X_2] + 0.05 \times E[X_3] = 0.25 \times 1 + 0.70 \times 2 + 0.05 \times 5 = 1.90.$

The cdf of X is

$$\begin{aligned} F_X(x) &= 0.25F_1(x) + 0.70F_2(x) + 0.05F_3(x) \\ &= 0.25 \times (1 - e^{-x}) + 0.70 \times (1 - e^{-x/2}) + 0.05 \times (1 - e^{-x/5}), \text{ so} \end{aligned}$$

$$\begin{aligned} P[X > 1] &= 1 - F_X(1) = 1 - [0.25 \times (1 - e^{-1}) + 0.70 \times (1 - e^{-1/2}) + 0.05 \times (1 - e^{-1/5})] \\ &= 1 - 0.44 = 0.56 \end{aligned}$$

□

PROBLEM SET 9

Functions and Transformations of Random Variables

1. (SOA) The profit for a new product is given by $Z = 3X - Y - 5$. X and Y are independent random variables with $Var(X) = 1$ and $Var(Y) = 2$. What is the variance of Z?
A) 1 B) 5 C) 7 D) 11 E) 16

2. Let X_1, X_2, X_3 be independent discrete random variables, with the probability function $P[X_i = k] = \binom{n_i}{k} p^k (1-p)^{n_i - k}$ for $k = 0, 1, ..., n_i$ for $i = 1, 2, 3$ and $0 < p < 1$.
Determine the probability function of $S = X_1 + X_2 + X_3$, $P[S = s]$.
A) $\binom{n_1+n_2+n_3}{s} p^s (1-p)^{n_1+n_2+n_3-s}$ B) $\sum_{i=1}^{3} \frac{n_i}{n_1+n_2+n_3} \binom{n_i}{s} p^s (1-p)^{n_i-s}$
C) $\prod_{i=1}^{3} \binom{n_i}{s} p^s (1-p)^{n_i-s}$ D) $\sum_{i=1}^{3} \binom{n_i}{s} p^s (1-p)^{n_i-s}$
E) $\binom{n_1 n_2 n_3}{s} p^s (1-p)^{n_1 n_2 n_3 - s}$

3. (SOA) The time, T, that a manufacturing system is out of operation has cumulative distribution function

$$F(t) = \begin{cases} 1 - \left(\frac{2}{t}\right)^2 & \text{for } t > 2 \\ 0 & \text{otherwise.} \end{cases}$$

The resulting cost to the company is $Y = T^2$.
Determine the density function of Y, for $y > 4$.
A) $\frac{4}{y^2}$ B) $\frac{8}{y^{3/2}}$ C) $\frac{8}{y^3}$ D) $\frac{16}{y}$ E) $\frac{1024}{y^5}$

4. Let X and Y be two independent random variables with moment generating functions
$M_X(t) = e^{t^2+2t}$, $M_Y(t) = e^{3t^2+t}$
Determine the moment generating function of $X + 2Y$.
A) $e^{t^2+2t} + 2e^{3t^2+t}$ B) $e^{t^2+2t} + e^{12t^2+2t}$ C) e^{7t^2+4t} D) $2e^{4t^2+3t}$ E) e^{13t^2+4t}

5. Let X_1 and X_2 be random variables with joint moment generating function
$M(t_1, t_2) = .3 + .1e^{t_1} + .2e^{t_2} + .4e^{t_1+t_2}$. What is $E[2X_1 - X_2]$?
A) $-.1$ B) $.4$ C) $.8$ D) $.2e + .4e^2$ E) $.3 + .1e^{3t_1} + .2e^{-t_2} + .4e^{3t_1-t_2}$

6. (SOA) An investment account earns an annual rate R that follows a uniform distribution on the interval $(0.04\,,\,0.08)$. The value of a 10,000 initial investment in this account after one year is given by $V = 10,000e^R$. Determine the cumulative distribution function, $F(v)$, of V for values of v that satisfy $0 < F(v) < 1$.

A) $\frac{10{,}000e^{v/10{,}000} - 10{,}408}{425}$ B) $25e^{v/10{,}000} - 0.04$

C) $\frac{v - 10{,}408}{10{,}833 - 10{,}408}$ D) $\frac{25}{v}$ E) $25\left[ln(\frac{v}{10{,}000}) - 0.04\right]$

7. Let X and Y be discrete random variables with joint probability function $f(x, y)$ given by the following table:

		$x=0$	$x=1$	$x=2$
y	0	0	0.40	0.20
	1	0.20	0.20	0

What is the variance of $Y - X$?
A) 0.16 B) 0.64 C) 1.04 D) 1.25 E) 1.4

8. Let X_1 and X_2 be two independent observations from a normal distribution with mean and variance 1. If $E[c|X_1 - X_2|] = 1$, then $c =$

A) $\sqrt{\pi}$ B) $\frac{1}{\sqrt{\pi}}$ C) $\frac{\sqrt{2\pi}}{4}$ D) $\frac{2}{\sqrt{\pi}}$ E) $\frac{\sqrt{\pi}}{2}$

9. Let X, Y and Z be independent Poisson Random variables with $E[X] = 3$, $E[Y] = 1$, and $E[Z] = 4$. What is $P[X + Y + Z \leq 1]$?

A) $12e^{-12}$ B) $9e^{-8}$ C) $\frac{13}{12}e^{-1/12}$ D) $9e^{-1/8}$ E) $\frac{9}{8}e^{-1/8}$

10. (SOA) The monthly profit of Company I can be modeled by a continuous random variable with density function f . Company II has a monthly profit that is twice that of Company I.
Determine the probability density function of the monthly profit of Company II.

A) $\frac{1}{2}f(\frac{x}{2})$ B) $f(\frac{x}{2})$ C) $2f(\frac{x}{2})$ D) $2f(x)$ E) $2f(2x)$

11. (SOA) An actuary models the lifetime of a device using the random variable $Y = 10X^{0.8}$ where X is an exponential random variable with mean 1 year. Determine the probability density function $f(y)$, for $y > 0$, of the random variable Y.

A) $10y^{0.8}e^{-y^{-0.2}}$ B) $8y^{-02\cdot}e^{-10y^{0.8}}$

C) $8y^{-0.2}e^{-(0.1y)^{1.25}}$ D) $(0.1y)^{1.25}e^{-0.125(0.1y)^{0.25}}$

E) $0.125(0.1y)^{0.25}e^{-(0.1y)^{1.25}}$

For trial use only.
Content may not
represent the most
current version

12. Let X, Y and Z have means 1, 2 and 3, respectively, and variances 4, 5 and 9, respectively. The covariance of X and Y is 2, the covariance of X and Z is 3, and the covariance of Y and Z is 1. What are the mean and variance, respectively, of the random variable $3X + 2Y - Z$?

    A) 4 and 31 B) 4 and 65 C) 4 and 67 D) 14 and 13 E) 14 and 65

13. (SOA) A device containing two key components fails when, and only when, both components fail. The lifetimes, T_1 and T_2, of these components are independent with common density function $f(t) = e^{-t}$, $t > 0$. The cost, X, of operating the device until failure is $2T_1 + T_2$. Which of the following is the density function of X for $x > 0$?

    A) $e^{-x/2} - e^{-x}$ B) $2\left(e^{-x/2} - e^{-x}\right)$ C) $\dfrac{x^2 e^{-x}}{2}$ D) $\dfrac{e^{-x/2}}{2}$ E) $\dfrac{e^{-x/3}}{3}$

14. (SOA) A company has two electric generators. The time until failure for each generator follows an exponential distribution with mean 10. The company will begin using the second generator immediately after the first one fails. What is the variance of the total time that the generators produce electricity?

    A) 10 B) 20 C) 50 D) 100 E) 200

15. (SOA) A company offers earthquake insurance. Annual premiums are modeled by an exponential random variable with mean 2. Annual claims are modeled by an exponential random variable with a mean of 1. Premiums and claims are independent. Let X denote the ratio of claims to premiums. What is the density function of X?

    A) $\dfrac{1}{2x+1}$ B) $\dfrac{2}{(2x+1)^2}$ C) e^{-x} D) $2e^{-2x}$ E) xe^{-x}

16. (SOA) Let T denote the time in minutes for a customer service representative to respond to 10 telephone inquiries. T is uniformly distributed on the interval with endpoints 8 minutes and 12 minutes. Let R denote the average rate, in customers per minute, at which the representative responds to inquiries. Which of the following is the density function of the random variable R on the interval $\frac{10}{12} \le R \le \frac{10}{8}$?

    A) $\frac{12}{5}$ B) $3 - \frac{5}{2r}$ C) $3r - \frac{5\,ln(r)}{2}$ D) $\frac{10}{r^2}$ E) $\frac{5}{2r^2}$

17. (SOA) A charity receives 2025 contributions. Contributions are assumed to be independent and identically distributed with mean 3125 and standard deviation 250. Calculate the approximate 90th percentile for the distribution of the total contributions received.

    A) 6,328,000 B) 6,338,000 C) 6,343,000 D) 6,784,000 E) 6,977,000

18. (SOA) An insurance company issues 1250 vision care insurance policies. The number of claims filed by a policyholder under a vision care insurance policy during one year is a Poisson random variable with mean 2. Assume the numbers of claims filed by distinct policyholders are independent of one another. What is the approximate probability that there is a total of between 2450 and 2600 claims during a one-year period?
A) 0.68 B) 0.82 C) 0.87 D) 0.95 E) 1.00

19. The number of claims received each day by a claims center has a Poisson distribution. On Mondays, the center expects to receive 2 claims but on other days of the week, the claims center expects to receive 1 claim per day. The numbers of claims received on separate days are mutually independent of one another. Find the probability that the claims center receives at least 3 claims in a 5 day week (Monday to Friday).
A) 0.90 B) 0.92 C) 0.94 D) 0.96 E) 0.98

20. In analyzing the risk of a catastrophic event, an insurer uses the exponential distribution with mean α as the distribution of the time until the event occurs. The insurer has n independent catastrophe policies of this type. Find the expected time until the insurer will have the first catastrophe claim.
A) $n\alpha$ B) α/n C) α^n D) $\alpha^{1/n}$ E) n/α

21. (SOA) In an analysis of healthcare data, ages have been rounded to the nearest multiple of 5 years. The difference between the true age and the rounded age is assumed to be uniformly distributed on the interval from -2.5 years to 2.5 years. The healthcare data are based on a random sample of 48 people. What is the approximate probability that the mean of the rounded ages is within 0.25 years of the mean of the true ages?
A) 0.14 B) 0.38 C) 0.57 D) 0.77 E) 0.88

22. (SOA) A city has just added 100 new female recruits to its police force. The city will provide a pension to each new hire who remains with the force until retirement. In addition, if the new hire is married at the time of her retirement, a second pension will be provided for her husband. A consulting actuary makes the following assumptions:
(i) Each new recruit has a 0.4 probability of remaining with the police force until retirement.
(ii) Given that a new recruit reaches retirement with the police force, the probability that she is not married at the time of retirement is 0.25.
(iii) The number of pensions that the city will provide on behalf of each new hire is independent of the number of pensions it will provide on behalf of any other new hire.
Determine the approximate probability that the city will provide at most 90 pensions to the 100 new hires and their husbands using the integer correction.
A) 0.60 B) 0.67 C) 0.75 D) 0.93 E) 0.99

23. An insurer has a portfolio of 1000 independent one-year insurance policies. For any particular policy there is a probability of .01 of a loss occurring within the year. For any particular policy, if a loss occurs, the expected loss is \$2000 with a standard deviation of \$1000. Find the standard deviation of the insurer's aggregate payout for the year (nearest 1000).

A) 6000 B) 7000 C) 8000 D) 9000 E) 10,000

24. (SOA) Claims filed under auto insurance policies follow a normal distribution with mean 19,400 and standard deviation 5,000. What is the probability that the average of 25 randomly selected claims exceeds 20,000?

A) 0.01 B) 0.15 C) 0.27 D) 0.33 E) 0.45

25. (SOA) You are given the following information about N, the annual number of claims for a randomly selected insured:

$$P(N=0) = \frac{1}{2}$$
$$P(N=1) = \frac{1}{3}$$
$$P(N>1) = \frac{1}{6}$$

Let S denote the total annual claim amount for an insured. When $N=1, S$ is exponentially distributed with mean 5. When $N>1, S$ is exponentially distributed with mean 8. Determine $P(4<S<8)$.

A) 0.04 B) 0.08 C) 0.12 D) 0.24 E) 0.25

26. (SOA) A company manufactures a brand of light bulb with a lifetime in months that is normally distributed with mean 3 and variance 1. A consumer buys a number of these bulbs with the intention of replacing them successively as they burn out. The light bulbs have independent lifetimes. What is the smallest number of bulbs to be purchased so that the succession of light bulbs, produces light for at least 40 months with probability at least 0.9772?

A) 14 B) 16 C) 20 D) 40 E) 55

27. A financial analyst tracking the price of a particular stock uses the uniform distribution between 1 and 2 as the model for the distribution of the stock price P one year from now. A second analyst analyzing the same stock price uses the uniform distribution on the interval from 10 to 100 as the model for the distribution of 10^Q one year from now (Q is the stock price one year from now). Find $m_P - m_Q$, the difference in the median stock price one year from now as estimated by the first and second analyst.

A) 0.24 B) 0.12 C) 0 D) -0.12 E) -0.24

28. An actuary is reviewing a study she performed on the size of claims made ten years ago under homeowners insurance policies. In her study, she concluded that the size of claims followed an exponential distribution and that the probability that a claim would be less than \$1,000 was 0.250. The actuary feels that the conclusions she reached in her study are still valid today with one exception: every claim made today would be twice the size of a similar claim made ten years ago as a result of inflation. Calculate the probability that the size of a claim made today is less than \$1,000.

A) 0.063 B) 0.125 C) 0.134 D) 0.163 E) 0.250

29. An automobile insurance company divides its policyholders into two groups: good drivers and bad drivers. For the good drivers, the amount of an average claim is 1400, with a variance of 40,000. For the bad drivers, the amount of an average claim is 2000, with a variance of 250,000. Sixty percent of the policyholders are classified as good drivers.

Calculate the variance of the amount of a claim for a policyholder.

A) 124,000 B) 145,000 C) 166,000 D) 210,400 E) 235,000

30. An insurance company designates 10% of its customers as high risk and 90% as low risk. The number of claims made by a customer in a calendar year is Poisson distributed with mean θ and is independent of the number of claims made by that customer in the previous calendar year. For high risk customers $\theta = 0.6$, while for low risk customers $\theta = 0.1$. Calculate the probability that a customer of unknown risk profile who made exactly one claim in 1997 will make exactly one claim in 1998.

A) 0.08 B) 0.12 C) 0.16 D) 0.20 E) 0.24

31. (SOA) Let X and Y be the number of hours that a randomly selected person watches movies and sporting events, respectively, during a three-month period. The following information is known about X and Y:

$$\begin{aligned} E(X) &= 50 \\ E(Y) &= 20 \\ Var(X) &= 50 \\ Var(Y) &= 30 \\ Cov(X,Y) &= 10 \end{aligned}$$

One hundred people are randomly selected and observed for these three months. Let T be the total number of hours that these one hundred people watch movies or sporting events during this three-month period. Approximate the value of $P(T < 7100)$.

A) 0.62 B) 0.84 C) 0.87 D) 0.92 E) 0.97

32. For a certain type of insurance policy, the actual loss amount has an exponential distribution with a mean of λ. An insurer will pay 75% of the loss that occurs. Find the moment generating function for the random variable representing the amount paid by the insurer.

A) $\frac{.75}{.75-s\lambda}$ B) $\frac{1}{.75-.75s\lambda}$ C) $\frac{.75}{1-.75s\lambda}$ D) $\frac{1}{1-.75s\lambda}$ E) $\frac{1}{1-1s\lambda}$

33. (SOA) The total claim amount for a health insurance policy follows a distribution with density function

$$f(x) = \frac{1}{1000}e^{-x/1000}\ ,\ x > 0$$

The premium for the policy is set at 100 over the expected total claim amount. If 100 policies are sold, what is the approximate probability that the insurance company will have claims exceeding the premiums collected?

A) 0.001 B) 0.159 C) 0.333 D) 0.407 E) 0.460

34. A company finds that the time it takes to process a randomly selected insurance claim has a uniform distribution on the interval from 1 to 2 hours. A claims adjuster has developed a new method for processing claims such that if the claim processing time under the current method is t hours, then the claim processing time under his new method is $\ln t$ hours. Find the density function $f(t)$ for the claim processing time under the new method.

A) $\ln t$ B) $t \ln t$ C) t D) te^t E) e^t

35. X and Y are random variables with correlation coefficient .75, and with $E[X] = Var[X] = 1$, and $E[Y] = Var[Y] = 2$. Find $Var[X + 2Y]$.

A) 9 B) $9 + \sqrt{2}$ C) $9 + 2\sqrt{2}$ D) $9 + 3\sqrt{2}$ E) $9 + 4\sqrt{2}$

36. (SOA) Claim amounts for wind damage to insured homes are independent random variables with common density function

$$f(x) = \begin{cases} \frac{3}{x^4} & \text{for } x > 1 \\ 0 & \text{otherwise} \end{cases}$$

where x is the amount of a claim in thousands. Suppose 3 such claims will be made. What is the expected value of the largest of the three claims?

A) 2025 B) 2700 C) 3232 D) 3375 E) 4500

37. (SOA) A company agrees to accept the highest of four sealed bids on a property. The four bids are regarded as four independent random variables with common cumulative distribution function

$$F(x) = \frac{1}{2}(1+\sin \pi x) \qquad \text{for} \qquad \frac{3}{2} \le x \le \frac{5}{2}$$

Which of the following represents the expected value of the accepted bid?

A) $\pi \int_{3/2}^{5/2} x \cos \pi x \, dx$ B) $\frac{1}{16}\int_{3/2}^{5/2} (1+\sin \pi x)^4 \, dx$ C) $\frac{1}{16}\int_{3/2}^{5/2} x(1+\sin \pi x)^4 \, dx$

D) $\frac{1}{4}\pi \int_{3/2}^{5/2} \cos \pi x(1+\sin \pi x)^3 \, dx$ E) $\frac{1}{4}\pi \int_{3/2}^{5/2} x \cos \pi x(1+\sin \pi x)^3 \, dx$

38. (SOA) A device runs until either of two components fails, at which point the device stops running. The joint density function of the lifetimes of the two components, both measured in hours, is
$f(x,y) = \frac{x+y}{27}$ for $0 < x < 3$ and $0 < y < 3$.
Calculate the probability that the device fails during its first hour of operation.
A) 0.04 B) 0.41 C) 0.44 D) 0.59 E) 0.96

39. $X_1, ..., X_n$ is a random sample from a uniform distribution on the interval $(0,2)$. Let $Y_1, \ldots, Y_n$ be the order statistics of that sample $X_1, ..., X_n$. What is $P\left[Y_1 < \frac{1}{2} < Y_n\right]$?
A) $\frac{3^n - 1}{4^n}$ B) $\frac{3^n + 1}{4^n}$ C) $\frac{4^n - 3^n - 1}{4^n}$ D) $\frac{4^n - 3^n + 1}{4^n}$ E) $\frac{4^n + 3^n + 1}{4^n}$

40. The random variables $X_1, X_2, X_3, X_4,$ and X_5 are independent and identically distributed. The random variable $Y = X_1 + X_2 + X_3 + X_4 + X_5$ has moment generating function $M_Y(t) = e^{15e^t - 15}$. Find the variance of X_1.
A) $\sqrt{3}$ B) 3 C) $\sqrt{15}$ D) 15 E) 225

41. (SOA) X and Y are independent random variables with common moment generating function $M(t) = e^{t^2/2}$. Let $W = X + Y$ and $Z = Y - X$.
Determine the joint moment generating function $M(t_1, t_2)$ of W and Z.
A) $e^{2t_1^2 + 2t_2^2}$ B) $e^{(t_1 - t_2)^2}$ C) $e^{(t_1 + t_2)^2}$ D) $e^{2t_1 t_2}$ E) $e^{t_1^2 + t_2^2}$