PROBABILITY NOTES - PR4

JOINT, MARGINAL, AND CONDITIONAL DISTRIBUTIONS

Joint distribution of random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$: A joint distribution of two random variables has a probability function or probability density function $f(x, y)$ that is a function of two variables (sometimes denoted $f_{X,Y}(x, y)$).
If X and Y are discrete random variables, then $f(x, y)$ 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) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dy\, dx = 1$.
It is possible to have a joint distribution in which one variable is discrete and one is continuous, or either has a mixed distribution. The joint distribution of two random variables can be extended to a joint distribution of any number of random variables.
If A is a subset of two-dimensional space, then $P[(X, Y) \in A]$ is the double summation (discrete case) or double integral (continuous case) of $f(x, y)$ over the region A.

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

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 $\boldsymbol{h(X, Y)}$** is defined to be
$E[h(X, Y)] = \sum_x \sum_y h(x, y) \cdot f(x, y)$ in the discrete case, and
$E[h(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) \cdot f(x, y)\, dy\, dx$ 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. The density function for the marginal distribution of Y is found in a similar way ; $f_Y(y)$ is equal to either $f_Y(y) = \sum_x f(x,y)$ or $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$.
If the cumulative distribution function of the joint distribution of X and Y is $F(x,y)$, then $F_X(x) = \lim_{y\to\infty} F(x,y)$ and $F_Y(y) = \lim_{x\to\infty} F(x,y)$.
This can be extended to define the marginal distribution of any one (or subcollection) variable in a multivariate distribution.

Independence of random variables X and Y: Random variables X and Y with cumulative distribution functions $F_X(x)$ and $F_Y(y)$ are said to be independent (or stochastically independent) if the cumulative distribution function of the joint distribution $F(x,y)$ can be factored in the form $F(x,y) = F_X(x) \cdot F_Y(y)$ for all (x,y). This
definition can be extended to a multivariate distribution of more than 2 variables.
If X and Y are independent, then $f(x,y) = f_X(x) \cdot f_Y(y)$, (but the reverse implication is not always true, i.e. if the joint distribution probability or density function can be factored in the form $f(x,y) = f_X(x) \cdot f_Y(y)$ then X and Y are usually, but not always, independent).

Conditional distribution of Y given $X = x$: Suppose that the random variables X and Y have joint density/probability function $f(x,y)$, and the density/probability function of the marginal distribution of X is $f_X(x)$. The density/probability function of the conditional distribution of Y given $X = x$ is $f_{Y|X}(y|X=x) = \frac{f(x,y)}{f_X(x)}$, if $f_X(x) \neq 0$.
The conditional expectation of Y given $X = x$ is $E[Y|X=x] = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y|X=x)\,dy$ in the continuous case, and $E[Y|X=x] = \sum_x y \cdot f_{Y|X}(y|X=x)$ in the discrete case.
The conditional density/probability is also written as $f_{Y|X}(y|x)$, or $f(y|x)$.
If X and Y are independent random variables, then $f_{Y|X}(y|X=x) = f_Y(y)$ and $f_{X|Y}(x|Y=y) = f_X(x)$.

Covariance between random variables X and Y: If random variables X and Y are jointly distributed with joint density/probability function $f(x,y)$, then the covariance between X and Y is $Cov[X,Y] = E[(X - E[X])(Y - E[Y])] = E[(X - \mu_X)(Y - \mu_Y)]$.
Note that $Cov[X,X] = Var[X]$.

Coefficient of correlation between random variables X and Y:

The coefficient of correlation between random variables X and Y is
$\rho(X,Y) = \rho_{X,Y} = \frac{Cov[X,Y]}{\sigma_X \sigma_Y}$, where σ_X and σ_Y are the standard deviations of X and Y respectively.

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.

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

Suppose that an experiment has k possible outcomes, with probabilities $p_1,\ p_2,\ldots,p_k$ respectively. If the experiment is performed n successive times (independently), let X_i denote the number of experiments that resulted in outcome i, so that
$X_1 + X_2 + \cdots + X_k = n$. The multivariate probability function is
$f(x_1,x_2,\ldots,x_k) = \frac{n!}{x_1!\cdot x_2!\cdots x_k!} \cdot p_1^{x_1} \cdot p_2^{x_2} \cdots p_k^{x_k}$.
$E[X_i] = np_i$, $Var[X_i] = np_i(1-p_i)$, $Cov[X_iX_j] = -np_ip_j$.
For example, the toss of a fair die results in one of $k=6$ outcomes, with probabilities $p_i = \frac{1}{6}$ for $i = 1,2,3,4,5,6$. If the die is tossed n times, then with X_i = # of tosses resulting in face "i" turning up, the multivariate distribution of $X_1, X_2, \ldots, X_6$ is a multinomial distribution.

Some results and formulas related to this section are:

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

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

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

(iv) $P[(X \le x) \cap (Y \le y)] = F_X(x) + F_Y(y) - F(x,y) \le 1$

(v) If X and Y are independent, then for any functions g and h,
$E[g(X) \cdot h(Y)] = E[g(X)] \cdot E[h(Y)]$, and in particular, $E[X \cdot Y] = E[X] \cdot E[Y]$.

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

(vii) $Cov[X, Y] = E[X \cdot Y] - \mu_X \cdot \mu_Y = E[XY] - E[X] \cdot E[Y]$.
$Cov[X, Y] = Cov[Y, X]$. If X and Y are independent, then $E[X \cdot Y] = E[X] \cdot E[Y]$ and $Cov[X, Y] = 0$. For constants a, b, c, d, e, f and random variables X, Y, Z and W, $Cov[aX + bY + c\ ,\ dZ + eW + f]$

$$= adCov[X, Z] + aeCov[X, W] + bdCov[Y, Z] + beCov[Y, W]$$

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

(ix) $Var[X + Y] = E[(X + Y)^2] - (E[X + Y])^2$

$$= E[X^2 + 2XY + Y^2] - (E[X] + E[Y])^2$$
$$= E[X^2] + E[2XY] + E[Y^2] - (E[X])^2 - 2E[X]E[Y] - (E[Y])^2$$
$$= Var[X] + Var[Y] + 2 \cdot Cov[X, Y]$$

If X and Y are independent, then $Var[X + Y] = Var[X] + Var[Y]$.
For any X, Y, $Var[aX + bY + c] = a^2 Var[X] + b^2 Var[Y] + 2ab \cdot Cov[X, Y]$

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

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

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

(xiii) If $Y = aX + b$ then $M_Y(t) = e^{bt} M_X(at)$.

(xiv) If X and Y are jointly distributed, then for any y, $E[X|Y = y]$ depends on y, say $E[X|Y = y] = h(y)$. It can then be shown that $E[h(Y)] = E[X]$; this is more usually written in the form $E[\ E[X|Y]\ ] = E[X]$.
It can also be shown that $Var[X] = E[\ Var[X|Y]\ ] + Var[\ E[X|Y]\ ]$.

(xv) A random variable X can be defined as a combination of two (or more) random variables X_1 and X_2, defined in terms of whether or not a particular event A occurs.

$X = \begin{cases} X_1 & \text{if event } A \text{ occurs (probability } p) \\ X_2 & \text{if event } A \text{ does not occur (probability } 1-p) \end{cases}$

Then, Y can be the indicator random variable $I_A = \begin{cases} 1 & \text{if } A \text{ occurs (prob. } p) \\ 0 & \text{if } A \text{ doesn't occur (prob. } 1-p) \end{cases}$

Probabilities and expectations involving X can be found by "conditioning" over Y:

$P[X \le c] = P[X \le c | A \text{ occurs}] \cdot P[A \text{ occurs}] + P[X \le c | A' \text{ occurs}] \cdot P[A' \text{ occurs}]$

$= P[X_1 \le c] \cdot p + P[X_2 \le c] \cdot (1-p)$,

$E[X^k] = E[X_1^k] \cdot p + E[X_2^k] \cdot (1-p)$, $M_X(t) = M_{X_1}(t) \cdot p + M_{X_2}(t) \cdot (1-p)$

This is really an illustration of a mixture of the distributions of X_1 and X_2, with $\alpha_1 = p$ and $\alpha_2 = 1-p$.

As an example, 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 2 balls are picked from Urn I, and if the number turning up is odd then 3 balls are picked from Urn II. X is the number of red balls chosen. Event A would be $A =$ die toss is even. Random variable X_1 would be the number of red balls chosen from Urn I and X_2 would be the number of red balls chosen from Urn II, and since each urn is equally likely to be chosen, $\alpha_1 = \alpha_2 = \frac{1}{2}$.

(xvi) If X and Y have a joint distribution which is uniform on the two dimensional region R (usually R will be a triangle, rectangle or circle in the (x, y) plane), then the conditional distribution of Y given $X = x$ has a uniform distribution on the line segment (or segments) defined by the intersection of the region R with the line $X = x$. The marginal distribution of Y might or might not be uniform.

Example 116: If $f(x, y) = K(x^2 + y^2)$ is the density function for the joint distribution of the continuous random variables X and Y defined over the unit square bounded by the points $(0,0)$, $(1,0)$, $(1,1)$ and $(0,1)$, find K.

Solution: The (double) integral of the density function over the region of density must be 1, so that $1 = \int_0^1 \int_0^1 K(x^2 + y^2)\, dy\, dx = K \cdot \frac{2}{3} \rightarrow K = \frac{3}{2}$. □

Example 117: The cumulative distribution function for the joint distribution of the continuous random variables X and Y is $F(x, y) = (.2)(3x^3y + 2x^2y^2)$, for $0 \le x \le 1$ and $0 \le y \le 1$. Find $f(\frac{1}{2}, \frac{1}{2})$.

Solution: $f(x, y) = \frac{\partial^2}{\partial x\, \partial y} F(x, y) = (.2)(9x^2 + 8xy) \rightarrow f(\frac{1}{2}, \frac{1}{2}) = \frac{17}{20}$. □

Example 118: X and Y are discrete random variables which are jointly distributed with the following probability function $f(x,y)$:

		X		
		-1	0	1
	1	$\frac{1}{18}$	$\frac{1}{9}$	$\frac{1}{6}$
Y	0	$\frac{1}{9}$	0	$\frac{1}{6}$
	-1	$\frac{1}{6}$	$\frac{1}{9}$	$\frac{1}{9}$

Find $E[X \cdot Y]$.

Solution: $E[XY] = \sum_x \sum_y xy \cdot f(x,y) = (-1)(1)(\frac{1}{18}) + (-1)(0)(\frac{1}{9}) + (-1)(-1)(\frac{1}{6})$

$+ (0)(1)(\frac{1}{9}) + (0)(0)(0) + (0)(-1)(\frac{1}{9})$

$+ (1)(1)(\frac{1}{6}) + (1)(0)(\frac{1}{6}) + (1)(-1)(\frac{1}{9}) \ = \frac{1}{6}$. $\square$

Example 119: Continuous random variables X and Y have a joint distribution with density function $f(x,y) = \frac{3(2-2x-y)}{2}$ in the region bounded by $y = 0$, $x = 0$ and $y = 2 - 2x$. Find the density function for the marginal distribution of X for $0 < x < 1$.

Solution: The region of joint density is illustrated in the graph at the right. Note that X must be in the interval $(0,1)$ and Y must be in the interval $(0,2)$. Since $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$, we note that given a value of x in $(0,1)$, the possible values of y (with non-zero density for $f(x,y)$) must satisfy $0 < y < 2 - 2x$, so that

$f_X(x) = \int_0^{2-2x} f(x,y)\,dy$

$= \int_0^{2-2x} \frac{3(2-2x-y)}{2}\,dy = 3(1-x)^2$. $\square$

Example 120: Suppose that X and Y are independent continuous random variables with the following density functions - $f_X(x) = 1$ for $0 < x < 1$ and $f_Y(y) = 2y$ for $0 < y < 1$. Find $P[Y < X]$.

Solution: Since X and Y are independent, the density function of the joint distribution of X and Y is $f(x,y) = f_X(x) \cdot f_Y(y) = 2y$, and is defined on the unit square. The graph at the right illustrates the region for the probability in question. $P[Y < X] = \int_0^1 \int_0^x 2y\,dy\,dx = \frac{1}{3}$. $\square$

Example 121: Continuous random variables X and Y have a joint distribution with density function $f(x,y) = x^2 + \frac{xy}{3}$ for $0 < x < 1$ and $0 < y < 2$.
Find $P[X > \frac{1}{2}|Y > \frac{1}{2}]$.

Solution: $P[X > \frac{1}{2}|Y > \frac{1}{2}] = \frac{P[(X>\frac{1}{2})\cap(Y>\frac{1}{2})]}{P[Y>\frac{1}{2}]}$.

$P[(X > \frac{1}{2}) \cap (Y > \frac{1}{2})] = \int_{1/2}^{1}\int_{1/2}^{2}[x^2 + \frac{xy}{3}]\,dy\,dx = \frac{43}{64}$.

$P[Y > \frac{1}{2}] = \int_{1/2}^{2} f_Y(y)\,dy = \int_{1/2}^{2}[\int_{0}^{1} f(x,y)\,dx]\,dy = \int_{1/2}^{2}\int_{0}^{1}[x^2 + \frac{xy}{3}]\,dx\,dy = \frac{13}{16} \rightarrow$

$P[X > \frac{1}{2}|Y > \frac{1}{2}] = \frac{43/64}{13/16} = \frac{43}{52}$. $\square$

Example 122: Continuous random variables X and Y have a joint distribution with density function $f(x,y) = \frac{\pi}{2}\big(sin\,\frac{\pi}{2}y\big)e^{-x}$ for $0 < x < \infty$ and $0 < y < 1$.
Find $P[X > 1|Y = \frac{1}{2}]$.

Solution: $P[X > 1|Y = \frac{1}{2}] = \frac{P[(X>1)\cap(Y=\frac{1}{2})]}{f_Y(\frac{1}{2})}$.

$P[(X > 1\big) \cap (Y = \frac{1}{2})] = \int_{1}^{\infty} f(x,\frac{1}{2})\,dx = \int_{1}^{\infty}\frac{\pi}{2}\big(sin\,\frac{\pi}{4}\big)e^{-x}dx = \frac{\pi\sqrt{2}}{4}\cdot e^{-1}$.

$f_Y(\frac{1}{2}) = \int_{0}^{\infty} f(x,\frac{1}{2})\,dx = \int_{0}^{\infty}\frac{\pi}{2}\big(sin\,\frac{\pi}{4}\big)e^{-x}dx = \frac{\pi\sqrt{2}}{4}$

$\rightarrow$ $P[X > 1|Y = \frac{1}{2}] = e^{-1}$. $\square$

Example 123: X is a continuous random variable with density function $f_X(x) = x + \frac{1}{2}$ for $0 < x < 1$. X is also jointly distributed with the continuous random variable Y, and the conditional density function of Y given $X = x$ is
$f_{Y|X}(y|X = x) = \frac{x+y}{x+\frac{1}{2}}$ for $0 < x < 1$ and $0 < y < 1$. Find $f_Y(y)$ for $0 < y < 1$.

Solution: $f(x,y) = f(y|x)\cdot f_X(x) = \frac{x+y}{x+\frac{1}{2}}\cdot(x + \frac{1}{2}) = x + y$.

Then, $f_Y(y) = \int_{0}^{1} f(x,y)\,dx = y + \frac{1}{2}$. $\square$

Example 124: Find $Cov[X,Y]$ for the jointly distributed discrete random variables in Example 118 above.

Solution: $Cov[X,Y] = E[XY] - E[X]\cdot E[Y]$. In Example 118 it was found that $E[XY] = \frac{1}{6}$. The marginal probability function for X is $P[X = 1] = \frac{1}{6} + \frac{1}{6} + \frac{1}{9} = \frac{4}{9}$, $P[X = 0] = \frac{2}{9}$ and $P[X = -1] = \frac{1}{3}$, and the mean of X is
$E[X] = (1)(\frac{4}{9}) + (0)(\frac{2}{9}) + (-1)(\frac{1}{3}) = \frac{1}{9}$.
In a similar way, the probability function of Y is found to be $P[Y = 1] = \frac{1}{3}$, $P[Y = 0] = \frac{5}{18}$, and $P[Y = -1] = \frac{7}{18}$, with a mean of $E[Y] = -\frac{1}{18}$.
Then, $Cov[X,Y] = \frac{1}{6} - (\frac{1}{9})(-\frac{1}{18}) = \frac{14}{81}$. $\square$

Example 125: The coefficient of correlation between random variables X and Y is $\frac{1}{3}$, and $\sigma_X^2 = a$, $\sigma_Y^2 = 4a$. The random variable Z is defined to be $Z = 3X - 4Y$, and it is found that $\sigma_Z^2 = 114$. Find a.

Solution: $\sigma_Z^2 = Var[Z] = 9Var[X] + 16Var[Y] - 2 \cdot (3)(4)\, Cov[X,Y]$.

Since $Cov[X,Y] = \rho[X,Y] \cdot \sigma_X \cdot \sigma_Y = \frac{1}{3} \cdot \sqrt{a} \cdot \sqrt{4a} = \frac{2a}{3}$, it follows that

$114 = \sigma_Z^2 = 9a + 16(4a) - 24(\frac{2a}{3}) = 57a \rightarrow a = 2$. $\square$

Example 126: Suppose that X and Y are random variables whose joint distribution has moment generating function $M(t_1, t_2) = \left(\frac{1}{4}e^{t_1} + \frac{3}{8}e^{t_2} + \frac{3}{8}\right)^{10}$, for all real t_1 and t_2.

Find the covariance between X and Y.

Solution: $Cov[X,Y] = E[XY] - E[X] \cdot E[Y]$.

$$E[XY] = \frac{\partial^2}{\partial t_1\, \partial t_2} M_{X,Y}(t_1,t_2)\Big|_{t_1=t_2=0}$$

$$= (10)(9)\left(\frac{1}{4}e^{t_1} + \frac{3}{8}e^{t_2} + \frac{3}{8}\right)^8 \left(\frac{1}{4}e^{t_1}\right)\left(\frac{3}{8}e^{t_2}\right)\Big|_{t_1=t_2=0} = \frac{135}{16} ,$$

$$E[X] = \frac{\partial}{\partial t_1} M_{X,Y}(t_1,t_2)\Big|_{t_1=t_2=0} = (10)\left(\frac{1}{4}e^{t_1} + \frac{3}{8}e^{t_2} + \frac{3}{8}\right)^9 \left(\frac{1}{4}e^{t_1}\right)\Big|_{t_1=t_2=0} = \frac{5}{2} ,$$

$$E[Y] = \frac{\partial}{\partial t_2} M_{X,Y}(t_1,t_2)\Big|_{t_1=t_2=0} = (10)\left(\frac{1}{4}e^{t_1} + \frac{3}{8}e^{t_2} + \frac{3}{8}\right)^9 \left(\frac{3}{8}e^{t_2}\right)\Big|_{t_1=t_2=0} = \frac{15}{4} ,$$

$$\rightarrow\ Cov[X,Y] = \frac{135}{16} - (\frac{5}{2})(\frac{15}{4}) = -\frac{15}{16} . \square$$

Example 127: Suppose that X has a continuous distribution with p.d.f. $f_X(x) = 2x$ on the interval $(0,1)$, and $f_X(x) = 0$ elsewhere. Suppose that Y is a continuous random variable such that the conditional distribution of Y given $X = x$ is uniform on the interval $(0,x)$. Find the mean and variance of Y.

Solution: This problem can be approached in two ways.

(i) The first approach is to determine the unconditional (marginal) distribution of Y. We are given $f_X(x) = 2x$ for $0 < x < 1$, and $f_{Y|X}(y|X = x) = \frac{1}{x}$ for $0 < y < x$.

Then, $f(x,y) = f(y|x) \cdot f_X(x) = \frac{1}{x} \cdot 2x = 2$ for $0 < x < 1$ and $0 < y < x$.

The unconditional (marginal) distribution of Y has p.d.f.

$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\, dx = \int_y^1 2\, dx = 2(1-y)$ for $0 < y < 1$ (and $f_Y(y)$ is 0 elsewhere). Then $E[Y] = \int_0^1 y \cdot 2(1-y)\, dy = \frac{1}{3}$, $E[Y^2] = \int_0^1 y^2 \cdot 2(1-y)\, dy = \frac{1}{6}$, and $Var[Y] = E[Y^2] - (E[Y])^2 = \frac{1}{6} - (\frac{1}{3})^2 = \frac{1}{18}$.

(ii) The second approach is to use the relationships $E[Y] = E[\,E[Y|X]\,]$ and $Var[Y] = E[\,Var[Y|X]\,] + Var[\,E[Y|X]\,]$.

From the conditional density $f(y|X=x) = \frac{1}{x}$ for $0 < y < x$, we have $E[Y|X=x] = \int_0^x y \cdot \frac{1}{x}\,dy = \frac{x}{2}$, so that $E[Y|X] = \frac{X}{2}$, and, since $f_X(x) = 2x$, $E[\,E[Y|X]\,] = E[\frac{X}{2}] = \int_0^1 \frac{x}{2} \cdot 2x dx = \frac{1}{3} = E[Y]$.

In a similar way, $Var[Y|X=x] = E[Y^2|X=x] - (E[Y|X=x])^2$, where $E[Y^2|X=x] = \int_0^x y^2 \cdot \frac{1}{x}\,dy = \frac{x^2}{3}$, so that $E[Y^2|X] = \frac{X^2}{3}$, and since $E[Y|X] = \frac{X}{2}$, we have $Var[Y|X] = \frac{X^2}{3} - (\frac{X}{2})^2 = \frac{X^2}{12}$.

Then $E[\,Var[Y|X]\,] = E[\frac{X^2}{12}] = \int_0^1 \frac{x^2}{12} \cdot 2x\,dx = \frac{1}{24}$, and

$Var[E[Y|X]] = Var[\frac{X}{2}] = \frac{1}{4}Var[X] = \frac{1}{4} \cdot [E[X^2] - (E[X])^2] = \frac{1}{4} \cdot [\frac{1}{2} - (\frac{2}{3})^2] = \frac{1}{72}$

so that $E[\,Var[Y|X]\,] + Var[\,E[Y|X]\,] = \frac{1}{24} + \frac{1}{72} = \frac{1}{18} = Var[Y]$. $\square$

FUNCTIONS AND TRANSFORMATIONS OF RANDOM VARIABLES

Distribution of a function of a continuous random variable $\boldsymbol{X}$: Suppose that X is a continuous random variable with p.d.f. $f_X(x)$ and c.d.f. $F_X(x)$, and suppose that $u(x)$ is a one-to-one function (usually u is either strictly increasing, such as $u(x) = x^3,\ e^x,\ \sqrt{x}$ or $ln\,x$, or u is strictly decreasing, such as $u(x) = e^{-x}$). As a one-to-one function, u has an inverse function v, so that $v(u(x)) = x$. Then the random variable $Y = u(X)$ (Y is referred to as a **transformation of $\boldsymbol{X}$**) has p.d.f. $f_Y(y)$ found as follows: $f_Y(y) = f_X(v(y)) \cdot |v'(y)|$. If u is a strictly increasing function, then
$F_Y(y) = P[Y \leq y] = P[u(X) \leq x] = P[X \leq v(y)] = F_X(v(y))$.

Distribution of a function of a discrete random variable $\boldsymbol{X}$: 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.

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 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 $P[X_1 + X_2 = k] = \sum_{x_1=0}^{k} f_1(x_1) \cdot f_2(k - x_1)$ (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
$f_Y(y) = \int_{-\infty}^{\infty} f_1(x_1) \cdot f_2(y - x_1)\, dx_1$

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

(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.

(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 $B(\sum n_i, p)$
geometric p	negative binomial $k,\ p$
negative binomial n_i, p	negative binomial $\sum n_i, p$
Poisson λ_i	Poisson $\sum \lambda_i$
$N(\mu_i, \sigma_i^2)$	$N(\sum \mu_i, \sum \sigma_i^2)$

Example 128: 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 p.d.f. of Y.
Solution: $F_Y(y) = P[Y \le y] = P[2\,ln\,X \le y] = P[X \le e^{y/2}] = 1 - e^{-e^{y/2}}$
$\rightarrow f_Y(y) = F'_Y(y) = \frac{d}{dy}\,(1 - e^{-e^{y/2}}) = \frac{1}{2}e^{y/2} \cdot e^{-e^{y/2}}$.
Alternatively, since $Y = 2\,ln\,X$ ($y = u(x) = 2\,ln\,x$, and ln is a strictly increasing function with inverse $x = v(y) = e^{y/2}$), and $X = e^{Y/2}$, it follows that
$f_Y(y) = f_X(e^{y/2}) \cdot \left|\frac{d}{dy}\,e^{y/2}\right| = \frac{1}{2}e^{y/2} \cdot e^{-e^{y/2}}$. □

Example 129: 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 - $f_Y(0) = .3$, $f_Y(1) = .5$, $f_Y(3) = .2$. Let $Z = X + Y$.
Find $P[Z = 5]$.
Solution: Using the fact that $f_X(x) = .2$ for $x = 1, 2, 3, 4, 5$, and the convolution method for independent discrete random variables, we have

$$f_Z(5) = \sum_{i=0}^{5} f_X(i) \cdot f_Y(5 - i)$$
$$= (0)(0) + (.2)(0) + (.2)(.2) + (.2)(0) + (.2)(.5) + (.2)(.2) = .20\,.$$ □

Example 130: 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
$f_Y(y) = \int_0^y f_{X_1}(t) \cdot f_{X_2}(y - t)\,dt = \int_0^y e^{-t} \cdot e^{-(y-t)}dt = ye^{-y}$, 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). □

Example 131: 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 $i \ne j$ 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 132: 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,
$M_{X+Y}(t) = M_X(t) \cdot M_Y(t) = (.7 + .3e^t)^6$. Since X and Y have identical distributions, they have the same moment generating function. Thus,
$M_X(t) = (.7 + .3e^t)^3$, and then $M_V(t) = M_X(t) \cdot M_Y(t) \cdot M_Z(t) = (.7 + .3e^t)^9$.
Alternatively, note that the moment generating function of the binomial $B(n, p)$ is $(1 - p + pe^t)^n$. Thus, $X + Y$ has a $B(6, .3)$ distribution, and each of X, Y and Z has a $B(3, .3)$ distribution, so that the sum of these independent binomial distributions is $B(9, .3)$, with m.g.f. $(.7 + .3e^t)^9$. $\square$

Example 133: 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, $P[X > Y] = P[X - Y > 0] = P\left[\frac{W-(-\frac{1}{2})}{\sqrt{25/16}} > \frac{-(-\frac{1}{2})}{\sqrt{25/16}}\right] = P[Z > .4]$,
where Z has standard normal distribution (W was standardized). Referring to the standard normal table, this probability is $.34$. $\square$

Example 134: If the number of typographical errors per page type 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 $\sum \lambda_i$. 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 $\frac{e^{-10\lambda}(10\lambda)^{10}}{10!}$. $\square$